Available online at www.alphanumericjournal.com

alphanumeric journal

The Journal of Operations Research, Statistics, Econometrics and Management Information Systems

Volume 11, Issue 1, 2023

Received: December 02, 2022
Accepted: July 12, 2023
Published Online: July 18, 2023

AJ ID: 2023.11.01.STAT.01
DOI: 10.17093/alphanumeric.1213585
Research Article

Time Series Forecasting of the Covid-19 Pandemic: A Critical Assessment in Retrospect

Murat Güngör, Ph.D. *

Assist. Prof., Faculty of Engineering and Natural Sciences, İstanbul Medeniyet University, İstanbul, Turkiye, murat.gungor@medeniyet.edu.tr

* İstanbul Medeniyet Üniversitesi, Kuzey Yerleşkesi H Blok; Ünalan Mah. Ünalan Sk. D100 Karayolu Yanyol 34700 Üsküdar / İstanbul,Türkiye

ABSTRACT

The COVID-19 pandemic is perceived by many to have run its course, and forecasting its progress is no longer a topic of much interest to policymakers and researchers as it once was. Nevertheless, in order to take lessons from this extraordinary two and a half years, it still makes sense to have a critical look at the vast body of literature formed thereon, and perform comprehensive analyses in retrospect. The present study is directed towards that goal. It is distinguished from others by encompassing all of the following features simultaneously: (i) time series of 10 of the most affected countries are considered; (ii) forecasting for two types of periods, namely days and weeks, are analyzed; (iii) a wide range of exponential smoothing, autoregressive integrated moving average, and neural network autoregression models are compared by means of automatic selection procedures; (iv) basic methods for benchmarking purposes as well as mathematical transformations for data adjustment are taken into account; and (v) several test and training data sizes are examined. Our experiments show that the performance of common time series forecasting methods is highly sensitive to parameter selection, bound to deteriorate dramatically as the forecasting horizon extends, and sometimes fails to be better than that of even the simplest alternatives. We contend that the reliableness of time series forecasting of COVID-19, even for a few weeks ahead, is open to debate. Policymakers must exercise extreme caution before they make their decisions utilizing a time series forecast of such pandemics.

Keywords: Time Series Forecasting, Coronavirus, Exponential Smoothing, Autoregressive Integrated Moving Average, Neural Network Autoregression

.

1. Introduction

The COVID-19 pandemic is a global outbreak of coronavirus, an infectious disease caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus. It has been affecting millions of lives worldwide since the World Health Organization (WHO) characterized the outbreak as a pandemic on March 11, 2020. Regarding daily new cases, peak figures were observed at the end of January 2022. Subsequently, many countries began scrapping restrictions, thanks to high vaccination rates and low case incidence (Anadolu Agency, 2022). In September 2022, the United Nations agency reported that weekly deaths from COVID-19 were the lowest since March 2020, and Tedros Adhanom Ghebreyesus, head of the WHO, told at a virtual press conference that the end of the pandemic is "in sight"(World Health Organization, 2022). Indeed, the outbreak is perceived by many to have run its course, and forecasting its progress is no longer a topic of much interest to policymakers and researchers as it once was. Nevertheless, in order to take lessons from this extraordinary two and a half years, it still makes sense to have a critical look at the vast body of literature formed thereon, and perform comprehensive analyses in retrospect. The present study is directed towards that goal.

Two of the most common model types to forecast the course of epidemics are compartmental epidemiological models and time series models. The former are based on modeling the actual infection process. The simplest such models classify living individuals in the population as susceptible, infectious, or recovered; hence they are called SIR models. They describe how people move between groups using differential equations. As an example in the COVID-19 context, see Aslan et al. (2022). Time series models, on the other hand, work by taking a series of historical observations and extrapolating patterns into the future. Exponential smoothing (ETS), autoregressive integrated moving average (ARIMA), and neural network autoregression (NNAR) are widely used time series models.

The literature on time series forecasting of COVID-19 is vast. Authors generally compare a few methods on one or more countries' statistics, and announce the best method with respect to some measure of point forecast accuracy. On an early stage of the pandemic, Ceylan (2020) develops several ARIMA models with different parameters to predict the epidemiological trend of COVID-19 prevalence of Italy, Spain, and France, using the data from February 21, 2020 to April 15, 2020. Kırbaş et al. (2020) model confirmed COVID-19 cases of eight European countries with ARIMA, nonlinear NNAR, and long short-term memory (LSTM) networks approaches. Six model performance metrics were used to select the most accurate model. LSTM was found the most favorable. Eroğlu (2020) tests NNAR and LSTM to forecast COVID-19 cases in Turkey. Zeroual et al. (2020) present a comparative study of five deep learning methods to forecast the number of new and recovered cases, based on six countries' data. Nikolopoulos et al. (2021), using data from the USA, India, UK, Germany, and Singapore up to mid-April 2020, provide predictive analytics tools for forecasting and planning during a pandemic. Ahmad et al. (2021) aim to provide guidance by forecasting the cumulative COVID-19 cases up to four weeks ahead for 187 countries, using four data-driven methodologies: ARIMA, ETS, and random walk forecasts with and without drift. ArunKumar et al. (2021) use ARIMA models to generate a 60-day forecast of cumulative COVID-19 cases for 16 countries. Initial combinations of model parameters are selected using a specific algorithm, followed by an optimization of model parameters. For each country, they consider the date on which the first case was reported as the starting day of the time series. The data collected spans a duration of seven months, from January 22, 2020 to August 3, 2020. The scripts are written in Python. Ballı (2021) analyzed COVID-19 data consisting of 35 weeks up until September 18, 2020 for USA, Germany and the global. Linear regression, multi-layer perceptron, random forest and support vector machines methods were used. The latter achieved the best performance. He estimates the global pandemic to peak at the end of January 2021. Devaraj et al. (2021) model the prediction of cumulative confirmed and recovered global cases and deaths with ARIMA, LSTM, stacked LSTM and Prophet approaches. Guleryuz (2021) aims to develop a forecasting model, considering the data for Turkey. ARIMA, Brown's ETS model and recurrent neural networks are employed. ARIMA possesses the lowest Akaike's Information Criterion (AIC) values for a number of statistics. ARIMA and ETS analyses are performed using Statgraphics Centurion 18 software, and recurrent neural networks are implemented using Python. Talkhi et al. (2021) apply nine models including NNAR, ARIMA, Holt-Winters', and Prophet on a six-month data of Iran until August 2020. Toğa et al. (2021) inspect the COVID-19 prevalence of Turkey. The number of infected cases, deaths, and recovered cases are predicted with ARIMA and NNAR. They use one year's data until March, 2021. Rahimi et al. (2021) present a review and brief analysis of the most important machine learning forecasting models against COVID-19.

More recently, Daniyal et al. (2022) compare different time series models with NNAR, using COVID-19 data in Pakistan spanning two years. Based on the mean absolute scaled error and root mean squared error, NNAR outperforms ARIMA and other competing models. Atchade and Sokadjo (2022) test automatic ETS and ARIMA models for a 10-month time series ending by November 2020. The variable of interest is the cumulative number of global laboratory-confirmed infections. They use the R software with the forecast library. Their results are in favor of ETS. Fernandes et al. (2022) use stacked LSTM to forecast the growth of the number of contaminations and deaths in one of Brazil's states. Mohanraj et al. (2022) propose a novel model mixing ETS and LSTM networks. Petropoulos et al. (2022), accepting the limitations of forecasting to predict the long-term trajectory of an outbreak, propose a statistical, time series approach to modeling and predicting the short-term behavior of COVID-19. They narrow their focus on nonseasonal ETS models with multiplicative trends, aiming to capture the continuation of the two variables they predict, namely global confirmed cases and deaths, as well as their uncertainty. They present the timeline of producing and evaluating 10-day-ahead forecasts over a period of four months. They conclude that more computationally intensive and data-hungry models do not necessarily perform better. Tan et al. (2022) develop seasonal ARIMA models to generate a 28-day forecast of COVID-19 cases during the third wave in Malaysia, using a data spanning 20 months until September 2021. Wang et al. (2022) propose seasonal and nonseasonal ARIMA and Prophet models to predict daily new cases and cumulative confirmed cases in the USA, Brazil and India over the next 30 days, based on a 17-month dataset ending by November 2021. Coroneo et al. (2022) test the predictive accuracy of forecasts of the number of COVID-19 fatalities produced by several forecasting teams and collected by the United States Centers for Disease Control and Prevention for the epidemic in the United States. They conclude that collecting a wide range of forecasts and combining them in an ensemble forecast may be a superior approach for health authorities, rather than relying on a small number of forecasts.

There are also many studies apart from mainstream. Castillo and Melin (2020) propose a hybrid approach combining the fractal dimension and fuzzy logic for enabling an efficient and accurate forecasting of COVID-19 time series. Forecasting windows of 10 and 30 days ahead were used to test the proposed approach. Niazkar et al. (2020) apply three explicit mathematical prediction models to the outbreaks in Iran and Turkey. Samanta et al. (2022) point out that most traditional time series models are parametric in nature and use the predicted values to generate forecasts for future time steps. This leads to error accumulation in each step of the forecasting horizon, resulting in increasingly poorer forecasts in the long term. They propose a nonparametric method which uses statistical representations such as trend, linearity, entropy etc. to cluster series from a predefined repository, and the series from same cluster are tagged as similar series. The method is validated empirically with a rich set of experiments involving COVID-19 data. Abbasimehr et al. (2022) use time series augmentation techniques to create new time series that take into account the characteristics of the original series, which are then utilized to generate enough samples to fit deep learning models properly. The proposed method improves the performance of LSTM and convolutional neural networks. The primary aim of Doornik et al. (2022) is to provide short-term forecasts that might aid policy makers, where these forecasts could serve as a useful guide to what might happen in the week ahead. The forecasts are based on extracting trends from windows of data using machine learning and then computing the forecasts by applying some constraints to the flexible extracted trend. They are complementary to the forecasts obtained from epidemiological models. Drews et al. (2022) use comparative and retrospective analyses to illuminate the aggregated effect of systematic biases on ensemble-based model forecasts. They compare the actual progression of active infections across 10 of the most affected countries in the world until late November 2020 with re-forecasts produced by a compartmental model and Holt-Winters' model. They specifically examine the sensitivity of the model parameters, estimated systematically from different subsets of the data and thereby different time windows, to illustrate the associated implications for short- to medium-term forecasting. Their findings portray considerable variations in forecasting skill in between the 10 countries and demonstrate that individual model predictions are highly sensitive to parameter assumptions. Markeviciute et al. (2021) propose an attention-based method combining machine learning techniques and statistical methods, and evaluate its effectiveness on Lithuanian data in comparison with the classical nonseasonal ARIMA model for short-term forecasts. The idea is to utilize the data of other countries with a longer history of the disease to forecast trends in Lithuania. Zhang and Yi (2022) establish analytical results for quantifying the biases of the parameter estimation in autoregressive models if the measurement error effects are neglected.

Adopting a totally different point of view, Luo (2021) criticizes mechanical, accuracy-oriented models to forecast COVID-19 infection cases and deaths. Forecasting future developments in the pandemic is fundamentally challenged by the intrinsic uncertainty rooted in many unknowns, not just about the biologically novel and evolving virus itself but also about the intertwined human, social, and political factors, which keep the future of the pandemic open-ended. These unknowns make the time series forecasting misleading. For instance, Chimmula and Zhang (2020), based on a deep learning approach, predict in the spring of 2020 that the outbreak will end at the beginning of the same summer, an unfortunate prophecy in retrospect. Therefore, in order to address this ``wicked problem'' (Rittel and Webber, 1973), Luo advocates a heuristic and exploratory approach that synthesizes prediction and monitoring, to make government policies, organization planning, and individual mentality future-informed despite the extreme uncertainty.

The present study is distinguished from other similar studies in the literature by encompassing all of the following features simultaneously:

- time series of 10 of the most affected countries are considered
- forecasting for two types of periods, namely days and weeks, are analyzed
- a wide range of ETS, ARIMA, and NNAR models are compared by means of automatic selection procedures
- basic methods for benchmarking purposes as well as mathematical transformations for data adjustment are taken into account
- several test and training data sizes are examined.

Our main goal is to draw some general conclusions about time series forecasting of COVID-19, if possible, rather than to identify some so-called best method.

Outline of the paper is as follows: in the next section, we present the details of our experimental setting. Specifically, we cite our data source and discuss its preprocessing in §2.1, list all parameters and their sets of values in §2.2, give a concise information about the forecasting methods to be compared in §2.3, and mention some implementation issues in §2.4. Then we tabulate and interpret the results in §3. Finally, in §4, we summarize our main conclusions.

2. Materials and methods

2.1. Data

Our study is essentially based on the complete Our World in Data (OWID) dataset, including all historical data on the pandemic up to August 30, 2022 (Our World in Data, 2022). We focus on the time series of new cases in 10 of the most affected countries in the world, namely United States, India, Brazil, Germany, United Kingdom, South Korea, Italy, Russia, Japan, and Turkey. Actually, we consider (sub)series starting on the first Monday in which the number of daily new cases exceeds 1000 and ending on a Sunday such that the missing or misleading information in between is nonexistent or minimal. Indeed, possibly due to several reasons, many days' new cases data in the OWID dataset simply do not exist or are entered wrongly as zero. To fill in these gaps in order to obtain the longest meaningful series for each country, we tried to benefit from other sources whenever available. The French data was formidably dirty, so we chose to investigate Turkey instead of France; otherwise, our list would be made up of the 10 countries most affected by the pandemic with respect to the cumulative number of cases. We note that Turkey is the 11th country in that sense as of August 30, 2022.

We shall use the data for an analysis of daily forecasting as well as weekly forecasting, so each series starts on a Monday and ends on a Sunday, enabling a smooth weekly aggregation. The preceding paragraph clarifies why the actual dates and lengths differ for each country. The starting date, the ending date, the length in days, and the length in weeks of the time series to be investigated is given in Table 1. We adopt precisely the OWID dataset with the following exceptions: There were two missing values for United Kingdom, namely of April 9, 2021 and May 18, 2021, which we filled in according to the "cases by specimen date" information in the website Coronavirus in the UK. For Turkey, we corrected three values, of December 10, 2020 and May 20-21, 2022, according to the website of TurkishMinistry of Health. Finally, for India we revised two successive values, of January 8-9, 2021, as 18434 and 18433 from 0 and 36867.

Country	Starting date	Ending date	Length in days	Length in weeks
United States	March 16, 2020	August 28, 2022	896	128
India	April 6, 2020	November 28, 2021	602	86
Brazil	June 22, 2020	September 12, 2021	448	64
Germany	March 16, 2020	March 27, 2022	742	106
United Kingdom	March 23, 2020	January 30, 2022	679	97
South Korea	December 28, 2020	April 24, 2022	483	69
Italy	June 22, 2020	August 28, 2022	798	114
Russia	April 13, 2020	August 28, 2022	868	124
Japan	November 23, 2020	August 28, 2022	644	92
Turkiye	November 30, 2020	May 29, 2022	546	78

Table 1. Starting date, ending date, length in days, and length in weeks of countries' time series to be investigated.

We were more specific on the choice of dates defining the Turkey series. First of all, we discarded the data prior to November 25, 2020 right away because Turkish authorities previously announced only symptomatic coronavirus cases (Reuters, 2020). Indeed, OWID dataset for this country exhibits an abrupt increase on November 26, 2020. Second, Turkey announces weekly statistics rather than daily statistics since June 2022, so the latest possible ending point of the series is May 31, 2022. This explains the prevalence of zeroes in the OWID dataset for the last few months, as a consequence of which we would be cutting off the original series anyway.

Difference in the dates corresponding to various series enables us to experiment with forecasting at different phases of the pandemic. As our chief aim is to make some general inferences about time series forecasting of COVID-19, this is an advantage rather than a disadvantage.

2.2. Parameters

We consider two versions of each of the 10 time series given in Table 1: the original series to analyze daily forecasting, and their aggregated weekly versions to analyze weekly forecasting. In other words, time periods are either days or weeks. We partition each series into two sets, namely training and test data, to evaluate point forecast accuracy. For each period type, we examine three distinct training and test data sizes: for daily data, we analyze forecasting one day, five days, and 10 days ahead; for weekly data, we analyze forecasting one week, two weeks, and four weeks ahead. For daily analysis, we use a training data of 40 days, 180 days, and of largest possible size depending on the particular series in question; for weekly analysis, we use a training data of 16 weeks, 52 weeks, and similarly of largest possible size. We simply call these training data as small, medium, and large, respectively.

We compare exponential smoothing, autoregressive integrated moving average (ARIMA), and neural network autoregression together with the simpler methods of naive, seasonal naive, drift, and linear regression for benchmarking. Also, we apply square root and logarithmic transformations for all methods except for naive and seasonal naive. For weekly forecasting, seasonal naive method is not taken into account since one-year's lag is esteemed too large to be of any practical use for COVID-19. Table 2 summarizes possible parameter values in our experimental setting.

Parameter	**Values**
Countries	United States, India, Brazil, Germany, United Kingdom, South Korea, Italy, Russia, Japan, Turkey
Period types	day, week
Test data sizes for daily forecasting	1, 5, 10
Test data sizes for weekly forecasting	1, 2, 4
Training data sizes for daily forecasting	40, 180, largest possible size
Training data sizes for weekly forecasting	16, 52, largest possible size
Methods for daily forecasting	naive, seasonal naive, drift, linear regression, exponential smoothing, ARIMA, neural network autoregression, and these methods with a square root and logarithmic transformation (except for naive and seasonal naive)
Methods for weekly forecasting	the same methods as in daily forecasting except seasonal naive

Table 2. Parameter values in the experimental setting.

2.3. Methods

By exponential smoothing in Table 2, we mean the model selected by the ETS()function in the R package fable (O'Hara-Wild et al., 2022). Candidate models can be represented by an ordered triple, where the coordinates stand for error, trend, and seasonal components. Errors are either additive or multiplicative; trend either nonexistent, additive, or damped additive; and seasonality either nonexistent, additive, or multiplicative. The 18 possible combinations therefrom include simple exponential smoothing, Holt's linear method, additive and multiplicative Holt-Winters' methods, and Holt-Winters' damped method among others. ETS()selects the model with the minimum AIC corrected for small sample bias. For details, we refer the reader to Hyndman and Athanasopoulos (2021) and Hyndman et al. (2008).

By ARIMA in Table 2, we mean the model chosen by the ARIMA() function in the R package fable. Candidate models are possibly seasonal, so a total of six parameters—order of the autoregressive part, degree of the first differencing involved, order of the moving average part, and the seasonal counterparts of these three—determine them. ARIMA()uses a variation of the Hyndman-Khandakar algorithm (Hyndman and Khandakar, 2008), which performs a stepwise search to traverse the model space. Again, the criterion is corrected AIC. For more information, we refer the reader to

Hyndman and Athanasopoulos (2021)and Box et al. (2015). We note that ARIMA()reverts to a nonseasonal model unless there are at least two full seasons of data. More precisely, for daily and weekly series, the training data size must be greater than or equal to 15 and 105 periods, respectively, for seasonal components to exist.

By neural network autoregression in Table 2, we mean the network returned by the NNETAR() function in the R package fable. This is necessarily a feed-forward neural network with a single hidden layer, defined by three parameters: the numbers of lagged inputs, lagged seasonal inputs, and nodes in the hidden layer. For nonseasonal time series, the number of lagged inputs is the optimal number of lags for a linear autoregressive model according to AIC. For seasonal time series, there is one lagged seasonal input by default, and the autoregressive model is fitted to the seasonally adjusted data. For details, the reader is referred to Hyndman and Athanasopoulos (2021). We note that NNETAR()regards daily and weekly series as seasonal as long as the training data size is greater than or equal to 9 and 54 periods, respectively.

For the simpler methods of naive, seasonal naive, drift, and linear regression, and for more information on the use of mathematical transformations in forecasting context, see Hyndman and Athanasopoulos (2021).

2.4. Implementation

For each of the 10 time series given in Table 1, we computed for all the parameter combinations in Table 2 the mean absolute percentage error (MAPE). We chose the criterion MAPE instead of root mean squared error to be able to combine statistics coming from multiple series and to facilitate the interpretation of results. Experiments are done with the programming language R on a computer with Intel(R) Core(TM) i5-2450M CPU (2.50GHz) processor and 4 GB RAM, running a 64-bit Windows 7 operating system.

We used the functions ETS(), ARIMA(), and NNETAR()in the R package fable with default settings. ETS()function takes two to three seconds on a typical combination, ARIMA()one to two seconds, NNETAR()some two minutes. All other models take negligible time. Thus, for each country and for each specific combination of training and test data sizes, calculation of MAPE for all models take about seven minutes, including the additional mathematical transformations carried out. Consequently, the experiment takes approximately two hours for each country, and 20 hours in total for all 10 countries.

Interestingly, the default ARIMA()function failed to find a model twice: for daily forecasting in India with medium training data and a test data of five and 10 days. In these two cases, we used the ARIMA model that is returned by the same function for the corresponding large training data, which is ARIMA(1,0,0)(0,1,1)7.

3. Results and discussion

Results of the computational study are summarized in Tables 3-9. Tables 3 and 4 show countries' smallest daily and weekly mean absolute percentage error (MAPE) for each combination of test and training data sizes. Tables 5 and 6 show forecasting methods' daily and weekly MAPE averaged over all countries for each combination of test and training data sizes. Table 7 is derived from Tables 3 and 4, showing the average of small-med(ium)-large columns for each period type and test data size. Tables 8 and 9 are derived from Tables 5 and 6,respectively, showing the average of small-med(ium)-large columns for daily and weekly forecasting.

	Smallest daily MAPE (%)								
test data size	1			5			10		
training data size	small	medium	large	small	medium	large	small	medium	large
United States	9.93	10.86	13.45	18.34	21.22	18.25	53.54	48.79	42.52
India	1.03	1.37	2.15	4.83	4.87	5.92	6.38	5.81	5.52
Brazil	6.94	5.96	13.19	22.60	29.85	28.89	27.91	34.99	41.33
Germany	12.53	9.51	6.69	7.40	7.40	7.18	9.58	11.30	9.76
United Kingdom	0.80	9.03	0.98	9.84	7.69	8.05	7.52	7.52	7.52
South Korea	0.71	1.14	0.60	9.75	9.07	6.53	7.79	8.46	13.44
Italy	1.62	0.68	0.92	8.28	18.75	18.75	19.07	10.19	10.36
Russia	0.23	0.02	0.58	2.44	1.40	1.09	3.26	4.38	3.00
Japan	2.68	2.26	0.52	15.25	15.25	10.43	10.84	13.26	13.38
Turkiye	0.75	3.87	2.03	7.04	6.60	6.09	10.54	11.12	11.97
Average	3.72	4.47	4.11	10.58	12.21	11.12	15.64	15.58	15.88

Table 3. Countries' smallest daily mean absolute percentage error (MAPE) for each combination of test and training data sizes (med: medium).

	Smallest weekly MAPE (%)								
test data size	1			2			4		
training data size	small	medium	large	small	medium	large	small	medium	large
United States	0.13	2.31	1.33	17.32	13.49	21.08	15.80	12.10	9.33
India	2.21	3.54	5.95	0.13	11.74	10.75	5.86	4.35	10.58
Brazil	15.61	36.10	35.16	12.38	32.79	33.45	11.28	31.89	32.18
Germany	0.75	0.61	0.55	3.78	3.51	2.88	6.82	5.14	5.05
United Kingdom	0.35	0.46	1.40	8.25	4.49	2.84	24.01	21.01	19.80
South Korea	0.91	3.60	0.53	5.12	4.01	1.96	44.16	7.50	41.67
Italy	0.16	1.95	0.80	2.23	1.70	5.29	1.47	11.27	8.29
Russia	4.46	0.12	0.25	2.05	2.29	3.56	9.43	4.72	9.31
Japan	7.24	4.89	5.44	6.26	6.26	6.26	5.51	5.51	5.51
Turkiye	1.85	3.42	0.75	6.86	11.85	7.93	12.61	27.67	20.80
Average	3.37	5.70	5.22	6.44	9.21	9.60	13.70	13.12	16.25

Table 4. Countries' smallest weekly mean absolute percentage error (MAPE) for each combination of test and training data sizes (med: medium).

	Average daily MAPE (%)								
test data size	1			5			10		
training data size	small	medium	large	small	medium	large	small	medium	large
Naive	45.3	45.3	45.3	71.0	71.0	71.0	104.0	104.0	104.0
Seasonal naive	22.8	22.8	22.8	23.2	23.2	23.2	41.3	41.3	41.3
Drift	36.2	43.6	44.8	59.2	72.7	70.1	112.6	108.7	102.9
Drift w sqrt	77.9	72.0	60.1	122.6	111.6	93.0	217.7	172.6	141.2
Drift w log	65.2	62.9	55.1	245.9	201.4	120.3	718.5	417.1	225.5
LR	149.7	381.9	805.5	66.7	222.0	543.7	102.6	273.0	555.8
LR w sqrt	161.4	418.2	730.0	64.7	226.1	492.9	92.9	249.6	509.0
LR w log	170.8	481.4	709.2	65.7	230.9	483.1	98.2	245.2	511.9
ETS	19.4	25.9	18.3	25.8	23.8	17.3	26.0	27.8	25.0
ETS w sqrt	11.4	14.5	17.7	21.2	18.7	19.3	32.5	22.9	24.6
ETS w log	14.6	16.7	17.8	21.8	18.9	20.9	36.1	25.8	27.2
ARIMA	45.7	32.9	27.7	30.3	35.0	26.9	34.6	38.3	28.9
ARIMA w sqrt	12.8	11.4	16.1	27.7	22.2	23.6	33.7	21.3	25.9
ARIMA w log	17.7	11.8	13.2	21.3	27.4	20.9	35.6	23.1	22.5
NNAR	54.4	33.1	50.5	26.6	42.2	45.0	45.1	63.7	66.5
NNAR w sqrt	40.0	33.4	32.8	19.3	31.6	32.1	41.8	47.8	38.3
NNAR w log	28.2	39.7	52.6	20.5	39.8	41.9	44.4	49.4	39.1

Table 5. Forecasting methods' daily mean absolute percentage error (MAPE) averaged over all countries for each combination of test and training data sizes (med: medium, LR: linear regression, ETS: exponential smoothing, ARIMA: autoregressive integrated moving average, NNAR: neural network autoregression, w: with, sqrt: square root transformation, log: logarithmic transformation).

	Average weekly MAPE (%)								
test data size	1			2			4		
training data size	small	medium	large	small	medium	large	small	medium	large
Naive	18.5	18.5	18.5	31.3	31.3	31.3	55.4	55.4	55.4
Seasonal naive	-	-	-	-	-	-	-	-	-
Drift	76.2	17.0	18.6	94.1	28.8	30.8	135.0	52.9	54.3
Drift w sqrt	22.7	23.2	22.3	32.1	39.7	39.0	65.8	69.9	69.1
Drift w log	24.2	23.6	23.2	42.8	40.1	40.5	108.0	82.6	82.1
LR	333.7	468.5	465.5	334.4	465.6	449.6	255.5	474.6	443.9
LR w sqrt	87.2	444.6	429.2	86.5	449.3	421.3	80.3	471.1	426.1
LR w log	198.9	422.6	433.8	198.2	444.4	443.2	173.8	499.3	475.1
ETS	17.4	15.3	11.2	29.7	27.1	17.7	70.0	49.6	59.1
ETS w sqrt	14.2	14.7	15.1	22.6	22.3	22.5	72.3	51.3	49.0
ETS w log	16.4	16.1	17.3	23.5	26.3	26.3	108.2	74.4	76.9
ARIMA	9.6	25.3	36.7	26.5	72.7	68.2	86.4	117.7	110.1
ARIMA w sqrt	11.5	17.0	17.5	25.1	35.4	47.5	54.3	70.3	76.9
ARIMA w log	13.8	15.8	11.5	19.5	31.4	33.5	50.0	90.0	81.8
NNAR	37.4	24.3	42.1	36.6	61.1	166.9	53.8	144.1	363.5
NNAR w sqrt	17.9	18.2	22.0	32.2	45.6	115.7	53.0	120.1	410.8
NNAR w log	15.7	29.0	15.1	32.3	53.2	36.5	47.9	83.0	79.6

Table 6. Forecasting methods' weekly mean absolute percentage error (MAPE) averaged over all countries for each combination of test and training data sizes (med: medium, LR: linear regression, ETS: exponential smoothing, ARIMA: autoregressive integrated moving average, NNAR: neural network autoregression, w: with, sqrt: square root transformation, log: logarithmic transformation).

	Average of the smallest MAPE (%)					
period type	day			week		
test data size	1	5	10	1	2	4
United States	11.42	19.27	48.28	1.26	17.29	12.41
India	1.52	5.21	5.90	3.90	7.54	6.93
Brazil	8.70	27.11	34.74	28.96	26.21	25.12
Germany	9.58	7.33	10.22	0.64	3.39	5.67
United Kingdom	3.60	8.53	7.52	0.74	5.20	21.60
South Korea	0.82	8.45	9.90	1.68	3.70	31.11
Italy	1.07	15.26	13.21	0.97	3.07	7.01
Russia	0.28	1.65	3.55	1.61	2.63	7.82
Japan	1.82	13.64	12.50	5.85	6.26	5.51
Turkiye	2.22	6.58	11.21	2.01	8.88	20.36
Average	4.10	11.30	15.70	4.76	8.42	14.35

Table 7. Countries' smallest mean absolute percentage error (MAPE) averaged over all training data sizes for each combination of period type and test data size.

	Average daily MAPE (%)			
test data size	1	5	10	Average
ETS w sqrt	14.5	19.7	26.7	20.3
ARIMA w log	14.2	23.2	27.1	21.5
ARIMA w sqrt	13.4	24.5	27.0	21.6
ETS w log	16.4	20.5	29.7	22.2
ETS	21.2	22.3	26.3	23.3
Seasonal naive	22.8	23.2	41.3	29.1
ARIMA	35.4	30.7	33.9	33.4
NNAR w sqrt	35.4	27.7	42.6	35.2
NNAR w log	40.1	34.1	44.3	39.5
NNAR	46.0	37.9	58.4	47.5
Drift	41.5	67.4	108.1	72.3
Naive	45.3	71.0	104.0	73.4
Drift w sqrt	70.0	109.1	177.2	118.7
Drift w log	61.1	189.2	453.7	234.7
LR w sqrt	436.5	261.2	283.8	327.2
LR w log	453.8	259.9	285.1	332.9
LR	445.7	277.5	310.4	344.5

Table 8. Forecasting methods' daily mean absolute percentage error (MAPE) for each test data size averaged over all countries and training data sizes (LR: linear regression, ETS: exponential smoothing, ARIMA: autoregressive integrated moving average, NNAR: neural network autoregression, w: with, sqrt: square root transformation, log: logarithmic transformation).

COVID-19 data exhibits various patterns throughout its progress, so it might be tempting to truncate it in the hope of being able to produce better forecasts. In terms of smallest MAPE, Tables 3 and 4 show that for one-step forecasts, truncation is beneficial on average; however, for multi-step forecasts, it does not seem to make a significant difference. Indeed, when training data size is small instead of medium, average MAPE falls from %4.47 to %3.72 in daily one-step forecasting, and it falls from %5.70 to %3.37 in weekly one-step forecasting. For five-day and two-week forecasts, average MAPE is again lower for small training data, but the percent difference is less. For 10-day and four-week forecasts, on the other hand, training data size appears to have no effect. Moreover, this general conclusion regarding averages does not hold true for many specific series. For example, while forecasting two weeks ahead, medium training data performs notably better than small training data for United Kingdom and United States.

	Average weekly MAPE (%)			
test data size	1	2	4	Average
ETS w sqrt	14.7	22.5	57.5	31.6
ETS	14.6	24.8	59.5	33.0
Naive	18.5	31.3	55.4	35.1
ARIMA w log	13.7	28.1	73.9	38.6
ARIMA w sqrt	15.3	36.0	67.1	39.5
Drift w sqrt	22.7	37.0	68.3	42.6
ETS w log	16.6	25.4	86.5	42.8
NNAR w log	19.9	40.6	70.1	43.6
Drift w log	23.7	41.1	90.9	51.9
Drift	37.2	51.2	80.7	56.4
ARIMA	23.8	55.8	104.7	61.5
NNAR w sqrt	19.4	64.5	194.7	92.8
NNAR	34.6	88.2	187.1	103.3
LR w sqrt	320.3	319.0	325.9	321.7
LR w log	351.8	361.9	382.7	365.5
LR	422.6	416.5	391.3	410.1

Table 9. Forecasting methods' weekly mean absolute percentage error (MAPE) for each test data size averaged over all countries and training data sizes (LR: linear regression, ETS: exponential smoothing, ARIMA: autoregressive integrated moving average, NNAR: neural network autoregression, w: with, sqrt: square root transformation, log: logarithmic transformation).

Tables 5 and 6 show that, with respect to different forecasting methods, truncation does not seem to yield consistent results. For example, average MAPE for ARIMA decreases for one-step daily forecasting as training data gets larger, being %45.7, %32.9, and %27.7 in order, but it increases for one-step weekly forecasting, being %9.6, %25.3, and %36.7 in order. Exponential smoothing displays rises as well as falls. Neural network autoregression performs best with medium training data for one-step daily and weekly forecasting; however, for multi-step forecasting, it performs best with small training data, quite contrary to intuition.

Table 7 shows countries' smallest MAPE averaged over all training data sizes for each combination of period type and test data size. Overall, average MAPE for five-day forecasting almost triples that of one-day. For 10-day forecasting, it is even larger, being %15.70. Similarly, average MAPE for two-week forecasting nearly doubles that of one-week. For four-week forecasting, it is %14.35, about three times as large as the corresponding one-step accuracy. Average MAPE for five-day forecasting is larger than that of one-day for all countries with the exception of Germany. For two-week forecasting it is larger than that of one-week in all cases but Brazil. In view of Table 7, we can say that the methods under investigation produce definitely poorer forecasts as the horizon extends.

Table 8 shows the methods in ascending order with respect to MAPE for daily forecasting for each test data size averaged over all countries and training data sizes. Altogether, exponential smoothing with a square root transformation is the best with

MAPE %20.3. ARIMA models with logarithmic and square root transformations follow with %21.5 and %21.6. In general, exponential smoothing and ARIMA together with transformations outperform other models. We note that the simple benchmarking method seasonal naive has an average MAPE of %29.1, less than that of ARIMA without a transformation as well as all three neural network autoregression models.

Table 9 is the weekly counterpart of Table 8. Collectively, exponential smoothing with a square root transformation is once again at the top of the list with MAPE %31.6. Exponential smoothing without a transformation comes next in sequence. Naive, arguably the simplest of all forecasting methods, shows the third best performance in our weekly setting, leaving all three ARIMA and neural network autoregression models behind. Furthermore, it singles out as the best method for four-week forecasting. This outcome alone may be enough to question the validity of medium- and long-term time series forecasting of COVID-19-related data.

4. Conclusion

Forecasting the progress of COVID-19 is challenged by many unknowns. The factors that contribute to it are not very well-understood. Future is usually not similar to the past. This makes it hard for time series models to produce good forecasts, as they work by taking a series of historical observations and extrapolating patterns into the future. The multiparameter empirical study carried out in the present paper supports this viewpoint. Our experiments show that the performance of common time series forecasting methods is highly sensitive to parameter selection, bound to deteriorate dramatically as the forecasting horizon extends, and sometimes fails to be better than that of even the simplest alternatives. Indeed, in forecasting four weeks ahead, the naive method outperformed all others including exponential smoothing, autoregressive integrated moving average, and neural network autoregression with respect to overall mean absolute percentage error.

To sum up, reliableness of time series forecasting of COVID-19, even for a few weeks ahead, is open to debate. Policymakers must exercise extreme caution before they make their decisions utilizing a time series forecast of such pandemics.

Statements and Declarations

Competing interests

The author has no relevant financial or nonfinancial interests to disclose.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability

The findings of this paper is based on the publicly available dataset of Our World in Data (2022) with the few exceptions listed in §2.1.

Code availability

The R code used for the computational study is available from the author upon request.

References

Abbasimehr, H., Paki, R., & Bahrini, A. (2022). A novel approach based on combining deep learning models with statistical methods for COVID-19 time series forecasting. Neural Computing and Applications, 34, 3135–3149. doi:10.1007/s00521-021-06548-9

Ahmad, G., Ahmed, F., Rizwan, M. S., Muhammad, J., Fatima, S. H., Ikram, A., & Zeeb, H. (2021). Evaluating data-driven methods for short-term forecasts of cumulative SARS-CoV2 cases. PLoS ONE, 16. doi:10.1371/journal.pone.0252147

Anadolu Agency. (2022). Many countries scrapping COVID-19 restrictions, thanks to high vaccination rates, low case incidence. Many countries scrapping COVID-19 restrictions, thanks to high vaccination rates, low case incidence. https://www.aa.com.tr/en/latest-on-coronavirus-outbreak/many-countries-scrapping-covid-19-restrictions-thanks-to-high-vaccination-rates-low-case-incidence/2500190 adresinden alındı

ArunKumar, K. E., Kalaga, D. V., Sai Kumar, C. M., Chilkoor, G., Kawaji, M., & Brenza, T. M. (2021). Forecasting the dynamics of cumulative COVID-19 cases (confirmed, recovered and deaths) for top-16 countries using statistical machine learning models: Auto-Regressive Integrated Moving Average (ARIMA) and Seasonal Auto-Regressive Integrated Moving Averag. Applied Soft Computing, 103. doi:10.1016/j.asoc.2021.107161

Aslan, I. H., Demir, M., Wise, M. M., & Lenhart, S. (2022). Modeling COVID-19: Forecasting and analyzing the dynamics of the outbreaks in Hubei and Turkey. Mathematical Methods in the Applied Sciences, 45, 6481–6494. doi:10.1002/mma.8181

Atchade, M. N., & Sokadjo, Y. M. (2022). Overview and cross-validation of COVID-19 forecasting univariate models. Alexandria Engineering Journal, 61, 3021–3036. doi:10.1016/j.aej.2021.08.028

Ballı, S. (2021). Data analysis of Covid-19 pandemic and short-term cumulative case forecasting using machine learning time series methods. Chaos, Solitons and Fractals, 142. doi:10.1016/j.chaos.2020.110512

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time Series Analysis: Forecasting and Control.

Castillo, O., & Melin, P. (2020). Forecasting of COVID-19 time series for countries in the world based on a hybrid approach combining the fractal dimension and fuzzy logic. Chaos, Solitons and Fractals, 140. doi:10.1016/j.chaos.2020.110242

Ceylan, Z. (2020). Estimation of COVID-19 prevalence in Italy, Spain, and France. Science of the Total Environment, 729. doi:10.1016/j.scitotenv.2020.138817

Chimmula, V. K., & Zhang, L. (2020). Time series forecasting of COVID-19 transmission in Canada using LSTM networks. Chaos, Solitons and Fractals, 135. doi:10.1016/j.chaos.2020.109864

Coronavirus in the UK. (2022). Cases in United Kingdom. Cases in United Kingdom. https://coronavirus.data.gov.uk/details/cases?areaType=overview&areaName=United Kingdom adresinden alındı

Coroneo, L., Iacone, F., Paccagnini, A., & Santos Monteiro, P. (2022). Testing the predictive accuracy of COVID-19 forecasts. International Journal of Forecasting. doi:10.1016/j.ijforecast.2022.01.005

Daniyal, M., Tawiah, K., Muhammadullah, S., & Opoku-Ameyaw, K. (2022). Comparison of Conventional Modeling Techniques with the Neural Network Autoregressive Model (NNAR): Application to COVID-19 Data. Journal of Healthcare Engineering. doi:10.1155/2022/4802743

Devaraj, J., Madurai Elavarasan, R., Pugazhendhi, R., Shafiullah, G. M., Ganesan, S., Jeysree, A. K., . . . Hossain, E. (2021). Forecasting of COVID-19 cases using deep learning models: Is it reliable and practically significant? Results in Physics, 21. doi:10.1016/j.rinp.2021.103817

Doornik, J. A., Castle, J. L., & Hendry, D. F. (2022). Short-term forecasting of the coronavirus pandemic. International Journal of Forecasting, 38, 453–466. doi:10.1016/j.ijforecast.2020.09.003

Drews, M., Kumar, P., Singh, R. K., De La Sen, M., Singh, S. S., Pandey, A. K., . . . Srivastava, P. K. (2022). Model-based ensembles: Lessons learned from retrospective analysis of COVID-19 infection forecasts across 10 countries. Science of the Total Environment, 806. doi:10.1016/j.scitotenv.2021.150639

Eroğlu, Y. (2020). Forecasting Models for Covid-19 Cases of Turkey Using Artificial Neural Networks and Deep Learning. Journal of Industrial Engineering, 31, 354–372.

Fernandes, F., Stefenon, S. F., Seman, L. O., Nied, A., Ferreira, F. C., Subtil, M. C., . . . Leithardt, V. R. (2022). Long short-term memory stacking model to predict the number of cases and deaths caused by COVID-19. Journal of Intelligent and Fuzzy Systems, 42, 6221–6234. doi:10.3233/JIFS-212788

Guleryuz, D. (2021). Forecasting outbreak of COVID-19 in Turkey; Comparison of Box–Jenkins, Brown's exponential smoothing and long short-term memory models. Process Safety and Environmental Protection, 149, 927–935. doi:10.1016/j.psep.2021.03.032

Hewamalage, H., Bergmeir, C., & Bandara, K. (2021). Recurrent Neural Networks for Time Series Forecasting: Current status and future directions. International Journal of Forecasting, 37, 388–427. doi:10.1016/j.ijforecast.2020.06.008

Hyndman, R. J. (2021). Package `fpp3'. Package `fpp3'. https://cran.r-project.org/web/packages/fpp3/index.html adresinden alındı

Hyndman, R. J., & Athanasopoulos, G. (2021). Forecasting: Principles and Practice.

Hyndman, R. J., & Khandakar, Y. (2008). Automatic Time Series Forecasting: The forecast Package for R. Journal of Statistical Software, 27, 22. doi:10.18637/jss.v027.i03

Hyndman, R. J., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). Forecasting with exponential smoothing: The state space approach.

Kırbaş, İ., Sözen, A., Tuncer, A. D., & Kazancıoğlu, F. Ş. (2020). Comparative analysis and forecasting of COVID-19 cases in various European countries with ARIMA, NARNN and LSTM approaches. Chaos, Solitons and Fractals, 138. doi:10.1016/j.chaos.2020.110015

Luo, J. (2021). Forecasting COVID-19 pandemic: Unknown unknowns and predictive monitoring. Technological Forecasting & Social Change, 166. doi:10.1016/j.techfore.2021.120602

Markeviciute, J., Bernataviciene, J., Levuliene, R., Medvedev, V., Treigys, P., & Venskus, J. (2021). Attention-based and time series models for short-term forecasting of COVID-19 spread. Computers, Materials and Continua, 70, 695–714. doi:10.32604/cmc.2022.018735

Mohanraj, G., Mohanraj, V., Marimuthu, M., Sathiyamoorthi, V., Luhach, A. K., & Kumar, S. (2022). Epidemic Prediction using Machine Learning and Deep Learning Models on COVID-19 Data. Journal of Experimental and Theoretical Artificial Intelligence. doi:10.1080/0952813X.2022.2058618

Niazkar, M., Türkkan, G. E., Niazkar, H. R., & Türkkan, Y. A. (2020). Assessment of three mathematical prediction models for forecasting the covid-19 outbreak in Iran and Turkey. Computational and Mathematical Methods in Medicine. doi:10.1155/2020/7056285

Nikolopoulos, K., Punia, S., Schäfers, A., Tsinopoulos, C., & Vasilakis, C. (2021). Forecasting and planning during a pandemic: COVID-19 growth rates, supply chain disruptions, and governmental decisions. European Journal of Operational Research, 290, 99–115. doi:10.1016/j.ejor.2020.08.001

O'Hara-Wild, M., Hyndman, R. J., & Wang, E. (2022). Package `fable'. Package `fable'. https://cran.r-project.org/web/packages/fable/index.html adresinden alındı

Our World in Data. (2022). Data on COVID-19. Data on COVID-19. https://github.com/owid/covid-19-data/tree/master/public/data adresinden alındı

Petropoulos, F., Makridakis, S., & Stylianou, N. (2022). COVID-19: Forecasting confirmed cases and deaths with a simple time series model. International Journal of Forecasting, 38, 439–452. doi:10.1016/j.ijforecast.2020.11.010

Rahimi, I., Chen, F., & Gandomi, A. H. (2021). A review on COVID-19 forecasting models. Neural Computing and Applications, 8. doi:10.1007/s00521-020-05626-8

Reuters. (2020). Turkey announces asymptomatic coronavirus case numbers for first time since July. Turkey announces asymptomatic coronavirus case numbers for first time since July. https://www.reuters.com/article/us-health-coronavirus-turkey-cases-idUSKBN2852W3 adresinden alındı

Rittel, H. W., & Webber, M. M. (1973). Dilemmas in a General Theory of Planning. Policy Sciences, 4, 155–169. doi:10.1007/BF01405730

Samanta, S., Prakash, P. K., & Chilukuri, S. (2022). MLTF: Model less time-series forecasting. Information Sciences, 593, 364–384. doi:10.1016/j.ins.2022.02.007

Talkhi, N., Akhavan Fatemi, N., Ataei, Z., & Jabbari Nooghabi, M. (2021). Modeling and forecasting number of confirmed and death caused COVID-19 in IRAN: A comparison of time series forecasting methods. Biomedical Signal Processing and Control, 66. doi:10.1016/j.bspc.2021.102494

Tan, C. V., Singh, S., Lai, C. H., Zamri, A. S., Dass, S. C., Aris, T. B., . . . Gill, B. S. (2022). Forecasting COVID-19 Case Trends Using SARIMA Models during the Third Wave of COVID-19 in Malaysia. International Journal of Environmental Research and Public Health, 19, 1–12. doi:10.3390/ijerph19031504

Toğa, G., Atalay, B., & Toksari, M. D. (2021). COVID-19 prevalence forecasting using Autoregressive Integrated Moving Average (ARIMA) and Artificial Neural Networks (ANN): Case of Turkey. Journal of Infection and Public Health, 14, 811–816. doi:10.1016/j.jiph.2021.04.015

Turkish Ministry of Health. (2022). Genel Koronavirüs Tablosu (in Turkish). Genel Koronavirüs Tablosu (in Turkish). https://covid19.saglik.gov.tr/TR-66935/genel-koronavirus-tablosu.html adresinden alındı

Wang, Y., Yan, Z., Wang, D., Yang, M., Li, Z., Gong, X., . . . Wang, Y. (2022). Prediction and analysis of COVID-19 daily new cases and cumulative cases: times series forecasting and machine learning models. BMC Infectious Diseases, 22, 1–12. doi:10.1186/s12879-022-07472-6

World Health Organization. (2022). The COVID-19 pandemic is nowhere near over. The COVID-19 pandemic is nowhere near over. https://unric.org/en/the-covid-19-pandemic-is-nowhere-near-over-who adresinden alındı

World Health Organization. (2022). The end of the COVID-19 pandemic is in sight. The end of the COVID-19 pandemic is in sight. https://news.un.org/en/story/2022/09/1126621 adresinden alındı

Zeroual, A., Harrou, F., Dairi, A., & Sun, Y. (2020). Deep learning methods for forecasting COVID-19 time-series data: A comparative study. Chaos, Solitons and Fractals, 140. doi:10.1016/j.chaos.2020.110121

Zhang, Q., & Yi, G. Y. (2022). Sensitivity analysis of error-contaminated time series data under autoregressive models with the application of COVID-19 data. Journal of Applied Statistics. doi:10.1080/02664763.2022.2034760.