https://doi.org/10.46389/rjd-2022-0905
www.rjdnmd.org

Rom J Diabetes Nutr Metab Dis
2022; volume 29, issue 3, pages 293-305

Original Article

Forecasting of COVID-19 confirmed cases in different countries with ARIMA models

Tania Dehesh [1], Heydar Ali Mardani-Fard [2], Paria Dehesh [3] *

[1] Department of Biostatistics and Epidemiology, Kerman University of Medical Sciences, Kerman, Iran
[2] Department of Mathematics and Statistics, Yasuj University, Yasuj, Iran
[3] Department of Epidemiology, School of Public Health, Iran University of Medical Sciences, Tehran, Iran

* Correspondence to: Paria Dehesh, Department of Epidemiology, School of Public Health, Iran University of Medical Sciences, Hemat Highway next to Milad Tower, Tehran, Iran. Phone: +98-21-31325069; Fax: +98-34-31325070; E-mail: Paria_dehesh@yahoo.com

Received: 15 December 2021 / Accepted: 11 March 2022

Abstract

The epidemic of a new coronavirus disease (COVID-19) has emerged as a global threat. Many countries and their health care systems were caught off guard. This study aims to predict the prevalence of COVID-19 in the most infected countries in the World Health Organization (WHO) regions in order to have better preparedness in health systems. The Auto-Regressive Integrated Moving Average (ARIMA) model was used to predict the pattern of confirmed cases based on epidemiological data from Johns Hopkins from February 25 to July 19, 2020. Mean incremental and logarithmic transfers were carried out to stabilize the series. Based on the ACF (AutoCorrelation Function) and PACF (Partial AutoCorrelation Function) charts, the first parameters of the model have been identified. The best model was chosen based on the likelihood ratio test and the least performance criteria value among all ARIMA models. Stata software version 12 was used. A number of ARIMA models have been formulated with various parameters. ARIMA (6,2,1) for South Africa, ARIMA (6,2,2) for U.S.A, ARIMA (2,1,1) for Iran, ARIMA (2,1,1) for Russia, ARIMA (5,2,2) for India, and ARIMA (3,1,2) for Australia were chosen based on the likelihood ratio tests and the values of the lower performance criteria. This research demonstrates that ARIMA models are sufficiently effective in predicting the prevalence of COVID-19 in the future. Predicting trends in COVID-19 prevalence in these countries can convince other countries to use this model in their future studies. The analysis results can help governments and health systems understand the patterns of this pandemic and plan for future waves of patients.

Keywords: COVID-19, forecast, predict, ARIMA, time series.

Introduction

A new virus belonging to the family of coronaviruses passed from animals to humans was identified in Wuhan, China, in December 2019. The virus can cause serious illness and death [1]. It has since been identified as a zoonotic coronavirus, similar to the Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) and the Middle East Respiratory Syndrome Coronavirus (MERS-CoV), and referred to as 2019-nCoV [2]. A total of 4515 cases, including 106 deaths, were confirmed on 27 January 2020 [3]. In the initial research, a number of cases visited a local seafood market in Wuhan, indicating that a common zoonotic exposure could cause this new disease [4]. The extent of the prevalence of this disease is unclear since the prevalence of this disease is currently very dynamic [1]. The capacity to monitor epidemiologically and detect suspected cases varies from country to country [5]. Several cases of COVID-19 infection have also been reported outside China, other Asian nations, the United States, Italy, Australia and Iran. In this situation, when the illness has no specific treatment, preventing and preparing for the disease in the health services is very important. Modeling and predicting future daily case counts can help the treatment system provide services to new patients. The statistical

prediction models could be helpful in forecasting and controlling this global epidemic threat. The Automatic-Regressive Integrated Moving Average (ARIMA) model has been used in the health domain with accurate predictions because of its simple explanation and rapid estimation in the correlated dataset [6].

Since the beginning of the COVID19 pandemic, different prediction models have been used for confirmed cases and deaths in China. For example, Li *et al*. developed a function to predict the pandemic trend with data-driven analysis in China [7]. Roosa *et al*. forecasted a short-term used number of confirmed cases with validated phenomenological models in Hubei, China [8]. The temporal dynamics of the COVID-19 pandemic in mainland China, Italy, and France were analyzed [9]. A standard SIR and SEIR framework were used in another study to model COVID-19 in Wuhan Province, China [10]. The Adaptive NeuroFuzzy Inference System (ANFIS) was used to estimate the number of confirmed COVID-19 cases in China by applying an Enhanced Flower Pollination Algorithm [11]. Information Based Algorithm of the Patient was applied to estimate the death rate of COVID-19 using publicly available data [12].

In summary, there are many studies in the literature to predict the spread of COVID-19 in China. However, this epidemic is growing rapidly throughout the world. It is necessary to look at this epidemic globally and simultaneously predict cases in all the countries involved with Covid-19. The global geographic regions in this study are according to six World Health Organization (WHO) regions. The countries with the highest cumulative confirmed cases were chosen in each region during the study period. The daily confirmed cases of COVID-2019 from February 25, 2020 to July 19, 2020 were collected from Johns Hopkins University's official website to build these models. This study aims to find the best predicting model by applying different ARIMA models for the most infected countries during the study period in six WHO regions (South Africa, U.S.A, Iran, Russia, India and Australia) and also to estimate the prevalence of COVID-19. These predictive models can help patients plan for improved preparation of treatment personnel in these countries in the near future.

Material and methods

Data source

The prevalence data of COVID-19 was taken from the Johns Hopkins epidemiological data website (https://covid19.who.int/data), and MS Excel was used to build a time-series database. To create a stable and effective ARIMA model, at least 30 observations are required [13]. Therefore, a time series containing at least 146 data from 25 February to 19 July was used in this study to predict COVID-19 prevalence in the most infected countries in WHO regions.

ARIMA models

A time series is a set of data points ordered in time [14]. Box and Jenkins introduced the ARIMA model for the first time in the 1970s [13]. ARIMA model is generally explained with a three parameter-argument, ARIMA (p, d, q), where p is the order of autoregression, d is the degree of difference, and q is the order of moving average [15].

The ARIMA model can also be expressed with other summary forms such as ARMA model, AR model, I model or MA model. In AR (p) model, the current value of the time series y_t is linearly related to its p previous values v_{t-1}, y_{t-2},..., y_{t-p} and the current residuals ε_t. In MA (q) model, the current value of the time series y_t is linearly related to its current and q previous residual series ε_{t-1}, ε_{t-2},..., ε_{t-q}. The statistical form of AR (p) and MA (q) models are defined as follows:

$$y_t=\varphi_1 y_{t-1}+ \varphi_2 y_{t-2}+\cdots+ \varphi_p y_{t-p}+ \varepsilon_t$$

$$y_t=\theta_1 \varepsilon_{t-1}- \theta_2 \varepsilon_{t-2}-\cdots- \theta_q \varepsilon_{t-q}+ \varepsilon_t$$

Where ϕ and θ are the autoregressive and moving average parameters, respectively. y_t is the response value at time t (daily number of confirmed cases) and ε_t is the random error at time t. The random errors are assumed to be independently and identically distributed with a mean of zero and a constant variance of σ^2. If AR and MA models become composed, then we will have ARMA (p, q) model. In ARMA (p, q), the current response of the time series is related linearly to its previous values as well as the current and previous residual series. The statistical ARMA (p, q) model can be presented as follows:

$$y_t=\alpha + \varphi_1 y_{t-1}+ \varphi_2 y_{t-2}+\cdots+ \varphi_p y_{t-p}+ \varepsilon_t- \theta_1 \varepsilon_{t-1}- \theta_2 \varepsilon_{t-2}-\cdots- \theta_q \varepsilon_{t-q}$$

Where α is a constant, and ε_{t-1} is the value of the previous random error. The differentiation could make it station if the time series was not stationary in the mean. These models are called ARIMA models. In ARIMA (p, d, q), d refers to the degree of differentiation [16]. Four steps must be completed during the development of an ARIMA model, including fixed time series (average and

https://doi.org/10.46389/rjd-2022-0905

variance stationary), model identification, parameter estimation and diagnostic verification [17].

Assessment and identification

Before analyzing, the time series must become a station in mean and variance. An average and variance stationary time series means that the mean and variance of the series are constant over time. The Dickey-Fuller test [17] to recognize the mean stationary values and the Box-Cox test were used to determine if the time series are stationary in variance. Log transformation and differences are remedial approaches to stabilize the time series for variance and mean, respectively [18].

Seasonal differences were used to stabilize the series from the seasonality trend. After reaching a stationary series, the orders of autoregressive terms (AR) and moving mean (MA) must be identified using the autocorrelation function (ACF) and the partial autocorrelation function (PACF).

Model parameter estimation

The model parameters were estimated with the maximum likelihood approach. As mentioned above, many ARIMA models were examined and the likelihood ratio test was used to compare different ARIMA models. This test provides a comparison of nested ARIMA models. The nested model means the full model has only one parameter more than the reduced model. Besides the likelihood ratio test, the lowest Bayesian information criterion (BIC) and Akaic information criterion (AIC) were used to select the best model from all significant ARIMA models. The BIC and AIC are expressed as follows [19]:

$$BIC = n.\ ln(^{RSS}/_{n}) + k.\ ln\ (n)$$

$$AIC = 2k - 2ln\ (\acute{L})$$

Where n is the number of observations, k is the number of parameters in the model, RSS is the residual sum of the square, and $\acute{L}$ is the maximum likelihood value.

Diagnostic checking

The usual procedure to diagnose the goodness of fit in a model is to compare actual values with the predicted values. In this study, three performance criteria, namely Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) were used to check the predictive accuracy of chosen ARIMA models. The mathematical formulas of these criteria are expressed in Eqs [5–7].

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} e_t^2}$$

$$MAE = \frac{1}{2}\sum_{t=1}^{n} |e_t|$$

$$MAPE = \frac{100\%}{n}\sum_{k=1}^{n} \left|\frac{e_t}{y_t}\right|$$

Where y_t is the observed response at time t, e_t is the residual at time t and n is the number of observations. Lower RMSE, MAE, and MAPE values show a better prediction model.

The adequacy of the prediction model was checked using the residuals ACF and PACF and the Ljung Box statistics (Q*). These statistics were introduced for adequacy checking of ARIMA models by Ljung and Box in 1978 [6]. The Q* statistics are obtained as follows:

$$Q^* = n(n+2)\sum_{j=1}^{p} \frac{r_j^2}{n-j}$$

Where r_j is the residual autocorrelation at lag j, n is the number of residuals, and P is the number of time lags in the test. The p-value associated with the Q* statistic should be bigger than the specified α (p>α) in order to have an adequate model.

The methodology of the current study was based on a previous study as a reference [20]. Excel 2016 was used to build the daily database of Covid-19 in the world, and STATA version 12 software was adopted to develop the ARIMA model. The statistical significance level was set at 0.05

Ethics

Since no primary data collection was undertaken, no patient or public was involved; no formal ethical assessment or informed consent was required. All data were collected from the official website, and all data were fully anonymized.

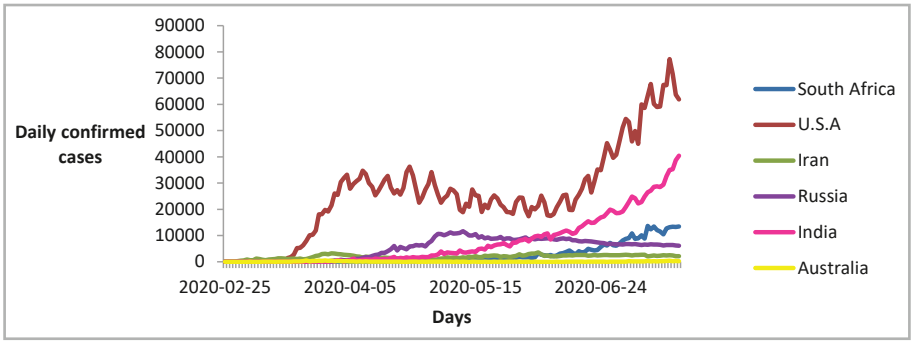


Figure 1: The number daily confirmed COVID-19 cases in different country.

Results

As shown in Figure 1, the COVID-19 outbreak in South Africa and India started later than in other countries. Descriptive statistics of confirmed cases in countries during the study period are shown in Table 1.

The data includes four iterative steps to adapt the ARIMA models to the time series: model assessment and identification, parameter estimation, diagnostic verification and prediction. The first step in model assessment is to control whether the mean and variance are constant over time (stationary). The Dickey-Fuller and Box-Cox tests were conducted for mean and variance stationary checking. First, Box-Cox test was performed for all confirmed case series in different countries. The appropriate transmission such as logarithm or inverse was used if they did not show stationary in variance.

After data transmission, the Dickey-Fuller test was done on transmitted data. If the P-value of the Dickey-Fuller test is bigger than 0.05, the series is non-station in mean. Then, the difference was taken, and the Dickey-Fuller test was done on the first difference data. The results of the Dickey-Fuller test on the original series and after difference are shown in Table 2. As observed, Iran, Russia and Australia became mean stations after the first difference. Other countries needed a second difference.

Table 1: Descriptive statistics on the number of confirmed cases of COVID-19 in different countries.

Country	Number of days	Mean	Std. Dev.	Min	Max
South Africa	146	2495.40	3835.90	0	13674
USA	146	25844.14	17485.30	0	77255
Iran	146	1874.84	763.82	34	3574
Russian	146	5276.09	3824.32	0	11656
India	146	7658.93	9660.99	0	40425
Australia	146	82.562	115.51	0	497

Table 2: The Dickey-Fuller Test of the Number of confirmed cases of COVID-19.

Country	Original series		First Difference		Second difference	
	Z-Statistic	P-value	Z-Statistic	P-value	Z-Statistic	P-value
South Africa	0.867	0.9926	0.754	0.8452	-17.572	0.000
USA	-0.744	0.8351	-0.857	0.9936	-11.836	0.000
Iran	-2.626	0.0877	-12.979	0.000	-	-
Russia	-1.367	0.5979	-14.012	0.000	-	-
India	5.760	0.999	3.875	0.0957	-9.998	0.000
Australia	-2.299	0.0749	-22.974	0.000	-	-

https://doi.org/10.46389/rjd-2022-0905

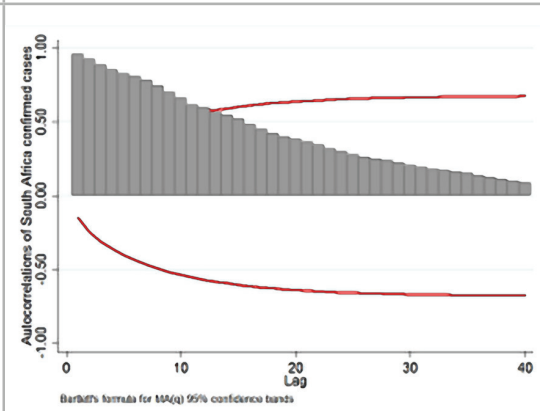
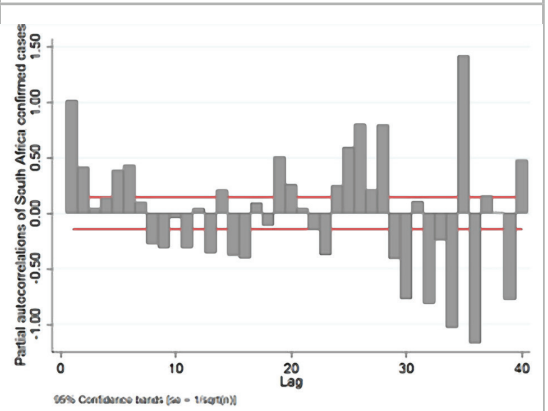
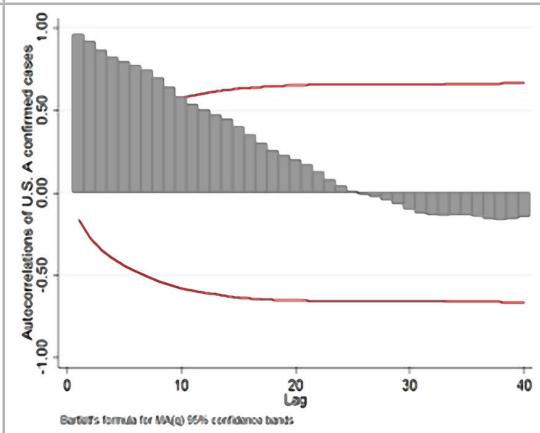
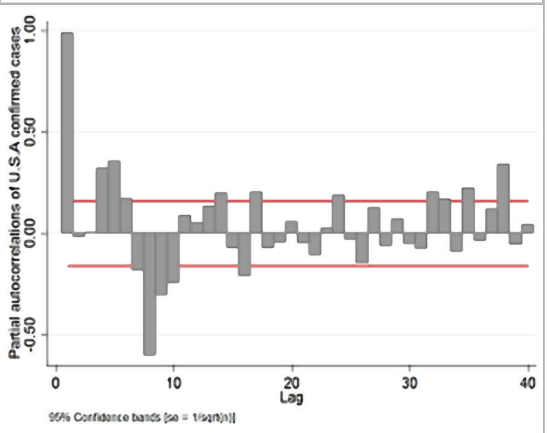
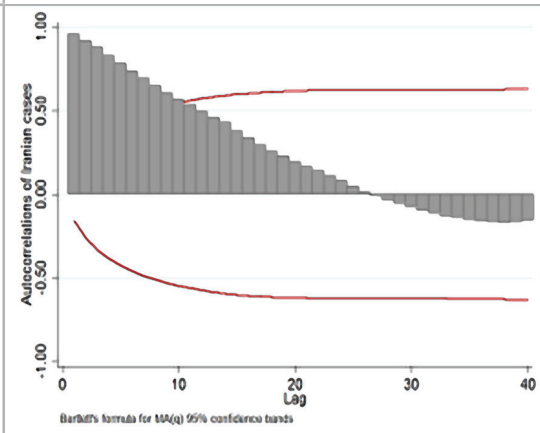
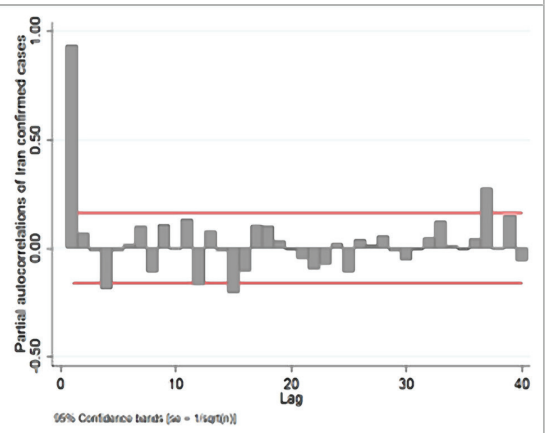
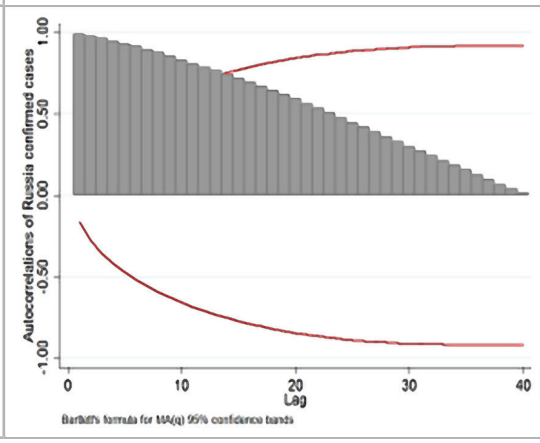
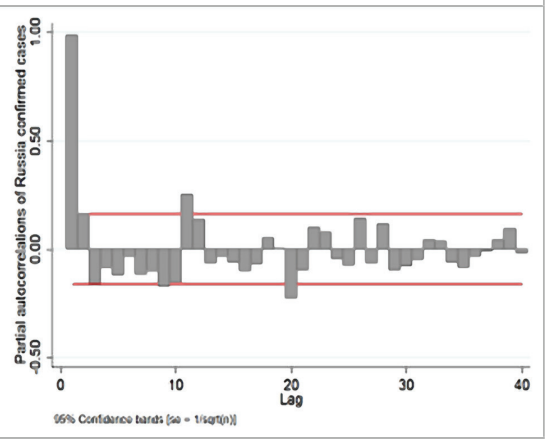
Country	Original series	
	ACF	PACF
South Africa		
U.S.A.		
Iran		
Russian		

Figure 2: Estimated autocorrelations and partial autocorrelations for orginal series and after differences for different country.

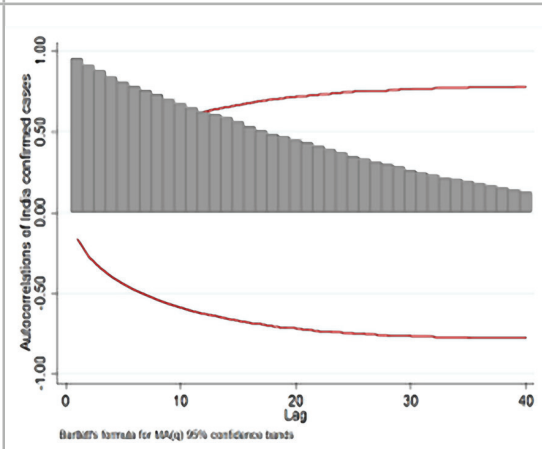
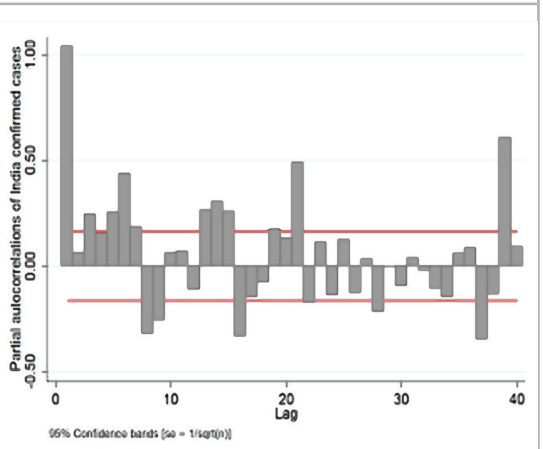
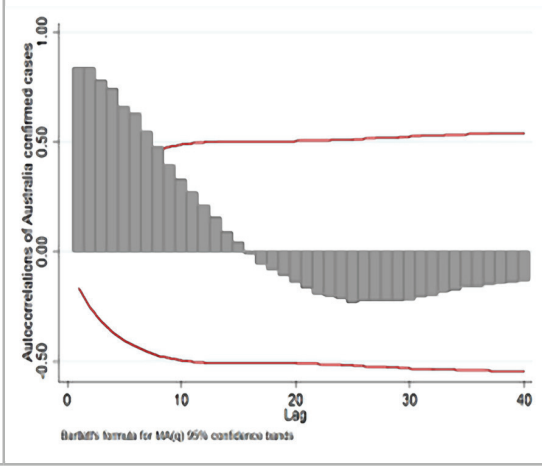
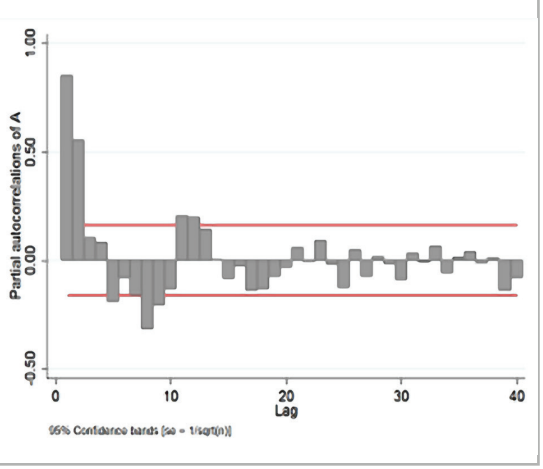
Country	Original series	
	ACF	PACF
India		
Australia		

Figure 2: Continued.

Figure 2 shows the ACF and PACF plots for model identification. These plots demonstrate that seasonality did not affect confirmed cases of COVID-19.

Straight vertical lines on the graph are 95 percent confidence interval bounds. The significant Bars in ACF and PACF plots that extend beyond the lines determine the order of q and p in the ARIMA model. In order to find the best final ARIMA model, different models with different P and q parameters were also created. Likelihood ratio tests were used for comparing nested ARIMA models. Table 3 shows the result of different ARIMA models for each country. Besides the likelihood ratio test, the model with minimum MAPE, MAEP, RMSE, AIC, BIC, and higher log-likelihood was selected as the best model.

Accordingly, the ARIMA (6,2,1), ARIMA (6,2,2), ARIMA (2,1,1), ARIMA (2,1,1), ARIMA (5,2,2), and ARIMA (3,1,2) models were chosen as the best models for South Africa, U.S.A, Iran, Russia, India and Australia (Table 3).

Table 4 shows the coefficients of chosen models for each country. Similar to releases, all the coefficients in each model were meaningful.

The overall adequacy of the models checked using Ljung-Box (Q*) statistic, (last column in Table 3) confirmed that the models were adequate and good fited for the confirmed cases of COVID-19 data in different countries during the study period. Moreover, the p-values computed for each country were greater than the alpha value (α=0.05). The plots of residuals ACF are shown in the second column of Figure 3. As observed, the residuals are not significant at any lag. This means that serial correlation was not significant between the error terms and confirms the adequacy of the models. The predicted 14 days (from 20 July to 2 August) of confirmed cases with a 95% confidence interval are presented in Table 5.

Figure 3 shows the forecast plots of ARIMA models. The closest of predicted plots with actual confirmed data could be observed in these plots. This shows

https://doi.org/10.46389/rjd-2022-0905

Table 3: Comparison of tested ARIMA models in different countries.

Country	ARIMA	Model Fit Statistics						Ljung-Box Q*	
		Log-likelihood	AIC	BIC	RMSE	MAE	MAEP	Q-Statistics	P-value
South-Africa	(6,2,2)	-1096.261	2212.522	2242.22	475.90	254.94	38.30	25.91	0.96
	(5,2,1)	-1102.607	2221.214	2244.973	498.26	273.66	43.35	35.82	0.66
	(6,2,0)	-1110.842	2237.685	2261.443	529.69	291.14	37.67	47.28	0.20
	(6,2,1)	**-1096.376**	**2210.752**	**2237.480**	**476.36**	**255.14**	**38.48**	**25.55**	**0.96**
U.S.A	(7,2,1)	-1330.548	2681.096	2710.794	2424.76	1780.30	18.60	31.83	0.82
	(6,2,1)	-1330.937	2679.875	2706.603	2417.57	1765.50	17.76	32.05	0.81
	(5,2,1)	-1345.959	2707.918	2731.677	2706.23	2009.20	17.23	70.55	0.00
	(6,2,2)	**-1329.069**	**2678.137**	**2707.836**	**2389.9**	**1730.90**	**16.74**	**28.25**	**0.92**
Iran	**(2,1,1)**	**-988.6762**	**1987.352**	**2002.236**	**220.46**	**169.80**	**10.97**	**37.17**	**0.59**
	(2,1,2)	-989.093	1988.186	2003.070	221.14	169.00	10.95	38.648	0.53
	(3,1,2)	-987.2049	1988.410	2009.247	218.24	166.96	10.96	32.523	0.79
	(2,1,3)	-987.1974	1988.395	2009.232	218.19	164.27	10.84	37.638	0.58
Russia	(1,1,1)	-1084.851	2177.702	2189.609	428.02	267.63	28.21	54.51	0.06
	(2,1,1)	**-1081.548**	**2173.097**	**2187.981**	**418.11**	**254.79**	**15.74**	**52.52**	**0.08**
	(3,1,1)	-1081.543	2175.087	2192.947	418.10	254.67	15.74	52.49	0.08
	(2,1,2)	-1081.540	2175.080	2192.940	418.09	254.64	15.70	52.45	0.08
India	(7,2,1)	-1131.507	2283.014	2312.712	611.55	430.13	257.32	58.92	0.02
	(7,2,0)	-1131.530	2281.060	2307.788	611.65	430.35	257.56	58.72	0.02
	(6,2,2)	-1131.058	2282.117	2311.815	608.30	420.92	242.34	53.63	0.07
	(5,2,2)	**-1131.153**	**2280.306**	**2307.034**	**608.79**	**422.02**	**236.61**	**53.09**	**0.08**
Australia	(1,1,1)	-777.8179	1563.636	1575.543	51.44	26.69	30.39	27.11	0.94
	(2,1,2)	-766.5273	1545.055	1562.915	47.55	26.60	30.92	24.03	0.98
	(3,1,2)	**-763.2184**	**1538.437**	**1556.297**	**46.54**	**24.89**	**28.99**	**13.69**	**0.99**
	(3,1,3)	-761.9119	1537.824	1558.661	46.096	25.23	30.05	13.98	0.99

Note: The bold model is the best-selected model. The selected model has the lowest performance criteria (BIC, AIC, RMSE, MAE and MAPE). RMSE – Root Mean Square Error; MAE – Mean Absolute Error; MAPE – Mean Absolute Percentage Error; BIC – Bayesian Information Criterion; AIC – Akaike's Information Criterion.

Table 4: Parameters of best ARIMA models of different countries.

Country	Best model	parameters	Coefficient	Standard error	Z-Statistic	P-value
South Africa	ARIMA (6,2,1)	AR (1)	-0.9640	0.0742	-12.99	0.0000
		AR (2)	-0.6558	0.0943	-6.95	0.0000
		AR (3)	-0.6221	0.0618	-10.07	0.0000
		AR (4)	-0.83750	0.0736	-11.38	0.0000
		AR (5)	-0.8322	0.0934	-8.91	0.0000
		AR (6)	-0.3816	0.0747	-5.11	0.0000
		MA (1)	-0.7022	0.0760	-9.24	0.0000
USA	ARIMA (6,2,2)	AR (1)	-0.5104	0.1452	-3.51	0.000
		AR (2)	-0.6059	0.1019	-5.94	0.0000
		AR (3)	-0.7008	0.0805	-8.70	0.0000
		AR (4)	-0.7099	0.1007	-7.05	0.0000
		AR (5)	-0.6530	0.1030	-6.34	0.0000
		AR (6)	-0.3705	0.1188	-3.12	0.0002
		MA (1)	-0.9036	0.1492	-6.06	0.0000
		MA (2)	0.4288	0.1237	3.47	0.0001
Iran	ARIMA (2,1,1)	AR (1)	-0.9872	0.0993	-9.94	0.0000
		AR (2)	-0.1542	0.0841	-1.83	0.051
		MA (1)	0.9333	0.0747	12.50	0.0000
Russia	ARIMA (2,1,1)	AR (1)	0.6170	0.1228	5.03	0.0000
		AR (2)	0.2729	0.0533	5.12	0.0000
		MA (1)	-0.8006	0.1092	-7.34	0.0000
India	ARIMA (5,2,2)	AR (1)	0.1499	0.1026	1.46	0.017
		AR (2)	-0.5424	0.0747	-7.26	0.0000
		AR (3)	-0.3147	0.0947	-3.32	0.0000
		AR (4)	-0.3705	0.0830	-4.46	0.0000
		AR (5)	-0.4152	0.1171	-3.54	0.0000
		MA (1)	-1.332	0.0747	-17.83	0.0000
		MA (2)	0.7897	0.0625	12.63	0.0000
Australia	ARIMA (3,1,2)	AR (1)	0.7593	0.0576	13.16	0.0000
		AR (2)	0.02100	0.1410	0.15	0.042
		AR (3)	-0.2155	0.0942	-2.29	0.022
		MA (1)	-1.6522	0.0571	-28.96	0.0000
		MA (2)	0.99999	0.0553	18.10	0.0000

https://doi.org/10.46389/rjd-2022-0905

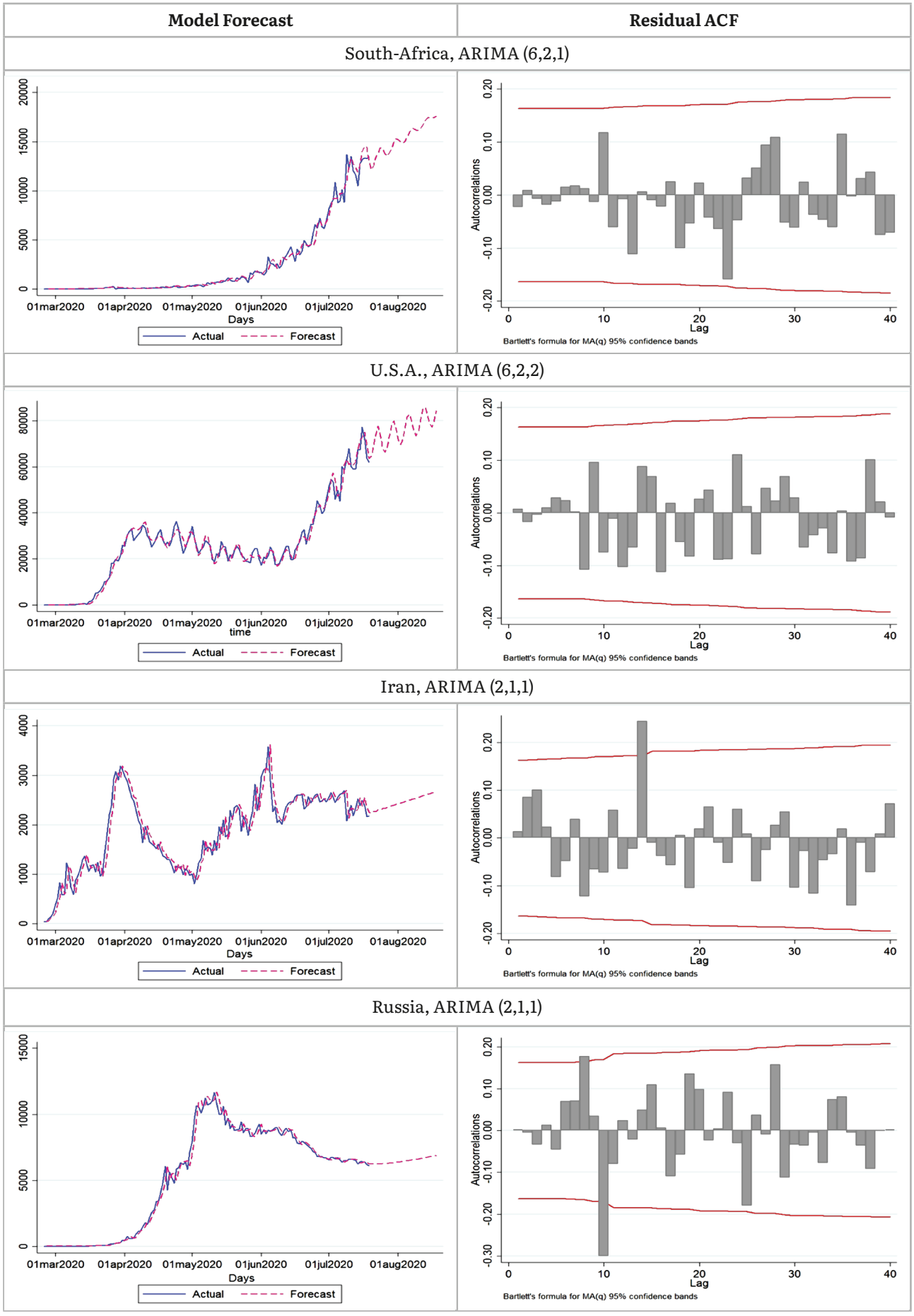


Figure 3: Time-series forecast plots and autocorrelation plots of residual.

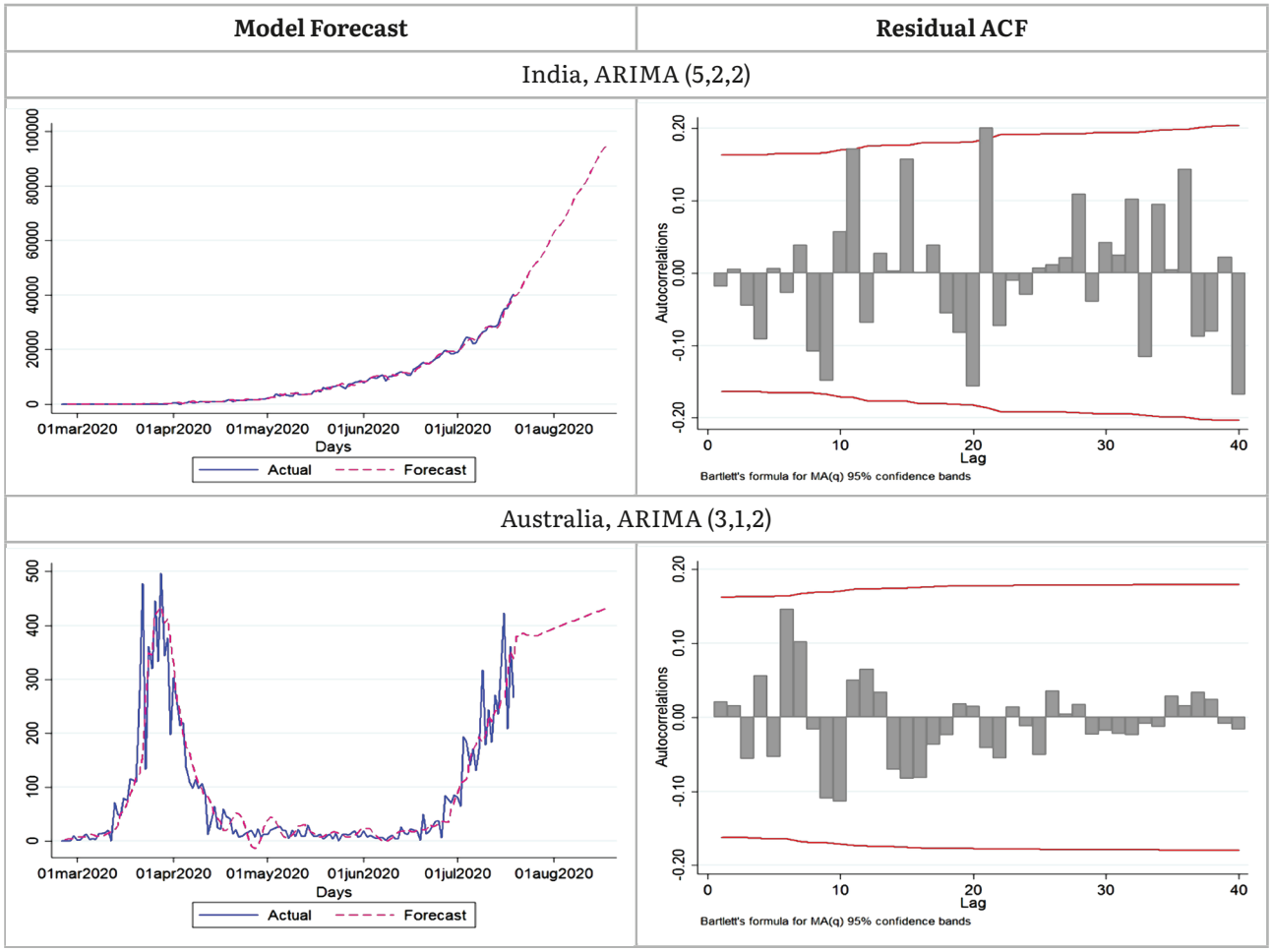


Figure 3: Continued.

the precision of models in forecasting. Figure 3 also shows that all countries will have an increasing trend in the future after this study.

Discussion

Prevention and control of the outbreak in huge epidemics need effective strategies. Finding simple procedures for estimating the prevalence trend is necessary to prepare medical equipment, finance, and prepare for unexpected situations. Thus, creating an effective forecasting model is important to help healthcare systems and governments decide on suitable strategies before being surprised. Up to now, any complex models have been used for predicting epidemics. A famous procedure used for predicting auto-correlated data is time series analysis. This method is an applicable instrument in medicine, especially in forecasting the prevalence trend of various diseases.

ARIMA model is one of the most popular methods in time series due to its simplicity and acceptable forecasting performance [12]. In the current study, the prevalence of the COVID-19 pandemic among high-prevalent countries in WHO's six regions was foreseen with ARIMA model. To the best of our knowledge, this study is the first to implement ARIMA models to simultaneously predict the prevalence of COVID-19 in the most prevalent countries in WHO regions. The number of confirmed cases in all six countries in WHO regions is still increasing. While the world has spent six months with COVID-19 and the healthcare systems of many countries have become tired, now there is great concern that the healthcare system capacity of these countries can respond to the referral wave of COVID-19-infected patients in the future.

Despite the latest outbreak compared to other countries in South Africa, the confirmed cases had increasing trends from the beginning of June within a short time. The new cases seem to have an increasing trend in this country, with 1500 new cases daily. The U.S.A had an increasing trend with fluctuation. The number of infected patients had a decreasing trend up to mid-June, but after that, it became increasing.

https://doi.org/10.46389/rjd-2022-0905

Table 5: Prediction of confirmed cases of COVID-19 for the next four days according to best ARIMA models of different countries with a 95% confidence interval.

Country	Forecast	20/ 07/20	21/ 07/20	22/ 07/20	23/ 07/20	24/ 07/20	25/ 07/20	26/ 07/20	27/ 07/20	28/ 07/20	29/ 07/20	30/ 07/20	31/ 07/20	01/ 08/20	02/ 08/20
	Forecast	13130	12562	8340	13877	13376	14273	12772	12547	13827	7345	12933	12312	12212	11957
South Africa	Lower	11189	10735	7254	11800	12299	12182	11670	11438	11714	7015	11805	11179	10074	10818
	upper	14070	14389	10430	15953	14453	16363	14874	13656	15940	7595	14061	16445	15351	15096
	Forecast	64388	69243	64082	67532	70954	70455	66193	67914	72897	55010	60858	67787	68006	69677
U.S.A	Lower	59672	61074	55519	65950	68372	69871	57542	59248	64077	53189	59035	64961	63163	60806
	upper	69105	77412	82645	86115	72537	75040	74843	76579	81717	61832	69681	71612	70848	71548
	Forecast	2231	2274	2367	2299	2542	2326	2332	2353	2363	2581	2793	2510	2523	2439
Iran	Lower	2031	2174	2167	2099	2322	2226	2132	2286	2268	2381	2393	2410	2423	2242
	upper	2665	2708	2704	2738	2641	2767	2774	2796	2806	2825	2836	2853	2866	2882
	Forecast	6153	6111	6065	6020	5874	5928	5882	5835	5787	5539	5691	5642	5593	5544
Russian	Lower	5317	4832	4759	4712	4666	4620	4573	4526	4479	4431	4383	5334	4285	4235
	upper	6988	7390	7371	7328	6283	7237	6190	7143	7096	5948	6999	5951	6902	6852
	Forecast	39974	41520	44180	46148	48324	49450	49737	51686	54297	49419	53728	56086	60115	55 835
India	Lower	38772	39660	42318	44242	46419	48533	48780	50729	52314	44434	50740	54098	59124	52582
	upper	41175	43381	46042	48053	50230	52368	52694	54643	56280	50404	59716	63073	65105	60572
	Forecast	380	380	386	483	382	280	285	382	385	388	390	493	595	398
Australia	Lower	293	291	288	276	257	243	123	209	199	193	188	285	382	180
	upper	467	468	484	589	506	517	338	556	570	582	593	601	609	615

The government should explore the change that happened during that time. They can bring everything to that time to better control the situation. However, the number of total confirmed cases in Iran is still increasing. The trend of confirmed cases in Iran decreased from the 1st of April until mid-May and started increasing till Jun 1st. After June 1st, the trend increased, but with the slightest slope. The May median was a critical time for Iran because then the growing trend continued. Confirmed daily cases in Russia had a declining trend compared to mid-May but increased with a gentle slope. Indian daily confirmed cases had a similar trend to those of South Africa. Confirmed cases trend became increasing with a severe slope from mid-point of June. Australia also showed an upward trend with a rapid trend starting at the beginning of June. It seems that all countries except Russia started the first or second increasing trends from mid-June.

These increasing trends may be due to governments' plans to return to normal life gradually from that time or may be due to warmer weather approaching summertime. However, there is no downward trend in new confirmed cases in all of these countries. It appears that people are tired of observing health protocols and more days are required to reach the plateau. As a result, if some limitations do not return by governments, the number of daily cases will be expected to increase.

Conclusions

Forecasting the disease's prevalence is important to have more ready healthcare services and better allocate medical resources. A time series model is an important statistical procedure in predicting disease. In the current study, ARIMA time series models were applied to the prevalence of COVID-19 in six countries most affected by COVID-19 in WHO regions: South Africa, the U.S.A, Iran, Russia, India, and Australia. The study results can help governments and healthcare services plan and manage medical equipment effectively in these countries over the next few days. These models could have a real-time update to be useful for more days in the future.

However, there are some limitations to this study. First, the data of this study came from a government report. Some countries may not find all infected individuals. Some factors may influence the diagnosis of COVID-19, such as the lack of diagnostic kits. Therefore the daily confirmed cases may account for smaller than actual confirmed cases. Second, the only number of confirmed cases with time was considered, and the influence of other possible factors such as medical conditions and environment were ignored.

Conflict of interest

The authors declare no conflict of interest.

References

1. Paules CI, Marston HD, Fauci AS. Coronavirus infections–more than just the common cold. JAMA. 2020;323(8):707-8.
2. Liu Y, Gayle AA, Wilder-Smith A, Rocklöv J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. Journal of Travel Medicine. 2020.
3. Jung S-m, Akhmetzhanov AR, Hayashi K, Linton NM, Yang Y, Yuan B, *et al*. Real-Time Estimation of the Risk of Death from Novel Coronavirus (COVID-19) Infection: Inference Using Exported Cases. Journal of Clinical Medicine. 2020;9(2):523.
4. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, *et al*. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The Lancet. 2020;395(10223):497-506.
5. Niehus R, De Salazar PM, Taylor A, Lipsitch M. Quantifying bias of COVID-19 prevalence and severity estimates in Wuhan, China that depend on reported cases in international travelers. medRxiv. 2020.
6. Cao L-t, Liu H-h, Li J, Yin X-d, Duan Y, Wang J. Relationship of meteorological factors and human brucellosis in Hebei province, China. Science of The Total Environment. 2020;703:135491.
7. Li Q, Feng W, Quan Y-H. Trend and forecasting of the COVID-19 outbreak in China. Journal of Infection. 2020;80(4):469-96.
8. Roosa K, Lee Y, Luo R, Kirpich A, Rothenberg R, Hyman J, *et al*. Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020. Infectious Disease Modelling. 2020;5:256-63.
9. Fanelli D, Piazza F. Analysis and forecast of COVID-19 spreading in China, Italy and France. Chaos, Solitons & Fractals. 2020;134:109761.
10. Roda WC, Varughese MB, Han D, Li MY. Why is it difficult to accurately predict the COVID-19 epidemic. Infectious Disease Modelling. 2020.
11. Al-Qaness MA, Ewees AA, Fan H, Abd El Aziz M. Optimization method for forecasting confirmed cases of COVID-19 in China. Journal of Clinical Medicine. 2020;9(3):674.
12. Wang Y, Xu C, Wang Z, Zhang S, Zhu Y, Yuan J. Time series modeling of pertussis incidence in China from 2004 to 2018 with a novel wavelet based SARIMA-NAR hybrid model. PloS one. 2018;13(12):e0208404.
13. Box GE, Jenkins GM, Reinsel GC, Ljung GM. Time series analysis: forecasting and control: John Wiley & Sons; 2015.
14. Fanoodi B, Malmir B, Jahantigh FF. Reducing demand uncertainty in the platelet supply chain through artificial neural networks and ARIMA models. Computers in biology and medicine. 2019;113:103415.

 https://doi.org/10.46389/rjd-2022-0905

15. Li X, Zhang C, Zhang B, Liu K. A comparative time series analysis and modeling of aerosols in the contiguous United States and China. Science of The Total Environment. 2019;690:799-811.
16. He Z, Tao H. Epidemiology and ARIMA model of positive-rate of influenza viruses among children in Wuhan, China: A nine-year retrospective study. International Journal of Infectious Diseases. 2018;74:61-70.
17. Cao S, Wang F, Tam W, Tse LA, Kim JH, Liu J, *et al*. A hybrid seasonal prediction model for tuberculosis incidence in China. BMC medical informatics and decision making. 2013;13(1):56.
18. Cheung Y-W, Lai KS. Lag order and critical values of the augmented Dickey–Fuller test. Journal of Business & Economic Statistics. 1995;13(3):277-80.
19. Priestley MB. Spectral analysis and time series: probability and mathematical statistics1981.
20. Wang Y-w, Shen Z-z, Jiang Y. Comparison of ARIMA and GM (1, 1) models for prediction of hepatitis B in China. PloS one. 2018;13(9).