

Letter To Editor

Asian Pacific Journal of Tropical Medicine

journal homepage: www.apjtm.org



doi: 10.4103/1995–7645.291407

Impact Factor: 1.94

ARIMA models forecasting the SARS–COV–2 in the Islamic Republic of Iran

Nayereh Esmaeilzadeh¹, Mohammadtaghi Shakeri^{2✉}, Mostafa Esmaeilzadeh³, Vahid Rahmani⁴¹Department of Epidemiology, School of Health, Mashhad University of Medical Sciences, Mashhad, Iran²Department of Biostatistics, School of Health, Mashhad University of Medical Sciences, Mashhad, Iran³Department of Mechanical Engineering, Mashhad Branch, Azad University, Mashhad, Iran⁴Zoonoses Research Center, Jahrom University of Medical Sciences, Jahrom, Iran

Currently, the COVID-19 epidemic has spread to more than 210 countries, with 3 272 202 confirmed cases and 230 104 deaths globally as of 3rd May 2020. Iran is found as the hotspot region of COVID-19 in the Eastern Mediterranean with more than 93 thousands confirmed cases and 5 957 deaths until 30th April[1]. Suppression strategies especially case isolation, elective home quarantine, and other mitigation approaches such as the closure of educational centers and transmission control by the lockdown of social activities are applied to reduce the basic reproduction number to less than 1[2]. The strategies have achieved varying degrees of success in different countries[3]. The autoregressive integrated moving average (ARIMA) models were developed to determine the temporal patterns and short-term prediction[4].

This approach is useful for forecasting and evaluation of confronting measures and a number of studies have confirmed it[5–7]. The results of this study can help to make an informative decision by the government and set proper policy to adopt interventions for this infectious disease.

The daily new laboratory-confirmed, recovered and death due to COVID-19 cases between 20th February 2020 and 30th April 2020 extracted from World Health Organization website[1].

Firstly, we developed the ARIMA model for each series. This model includes single regression, multiple regression, and the moving averages. It can remove the confounding effect of time. Therefore, the time series model ARIMA (p, d, q) consists of several components. The order of p, d, and q is explained as the autoregressive part of the model, the integrated part of the model, and the moving average parameter[8].

This linear combination is formulated as:

$$y_t = \delta + \lambda_1 y_{t-1} + \lambda_2 y_{t-2} + \dots + \lambda_p y_{t-p} + \gamma_1 \varepsilon_{t-1} + \gamma_2 \varepsilon_{t-2} + \dots + \gamma_q \varepsilon_{t-q} + e_t$$

$$e_t \sim \text{i.i.d.}(0, \sigma^2) \quad (1)$$

Where y is a dependent variable (daily cases of COVID19), λ_p is an autoregressive operator coefficient γ_q is the moving average operator coefficient, y_{t-p} is the value of the cases of COVID-19 in an earlier time, ε_{t-q} is the value of the cases of COVID-19 deviation in q time, and e_t is a random error term with the white-noise distribution.

The assumption of this model is based on the stationary data, so we performed the Bartlett and the unit-root tests for determination stationery for variance and mean value of data and then transformed them if needed. For estimation of the number of autoregressive and moving average parameters, we used autocorrelation functions and partial autocorrelation functions correlograms, after which possible models were identified[9].

In the next step, we evaluated the goodness-of-fit of the end model through checking white noise residuals with Ljung-Box (Q) test and the best- model which was fitted to the data was selected based on least value of the Akaike Information Criterion (AIC)[5].

Then, the best ARIMA models were applied to the prediction of the events of COVID-19, and the forecasting precision was estimated by the root-mean-square error (RMSE). This is computed using the following formula:

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^N (Y_t - \hat{Y}_t)^2}{N}} \quad (2)$$

✉ To whom correspondence may be addressed. E-mail: shakerimt@mums.ac.ir

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-Non Commercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: reprints@medknow.com

©2020 Asian Pacific Journal of Tropical Medicine Produced by Wolters Kluwer-Medknow. All rights reserved.

How to cite this article: Esmaeilzadeh N, Shakeri M, Esmaeilzadeh M, Rahmani V. ARIMA models for forecasting the SARS-COV-2 in the Islamic Republic of Iran. Asian Pac J Trop Med 2020; 13(11): 521-524.

Article history: Received 3 May 2020
Accepted 15 June 2020

Revision 9 June 2020
Available online 14 August 2020

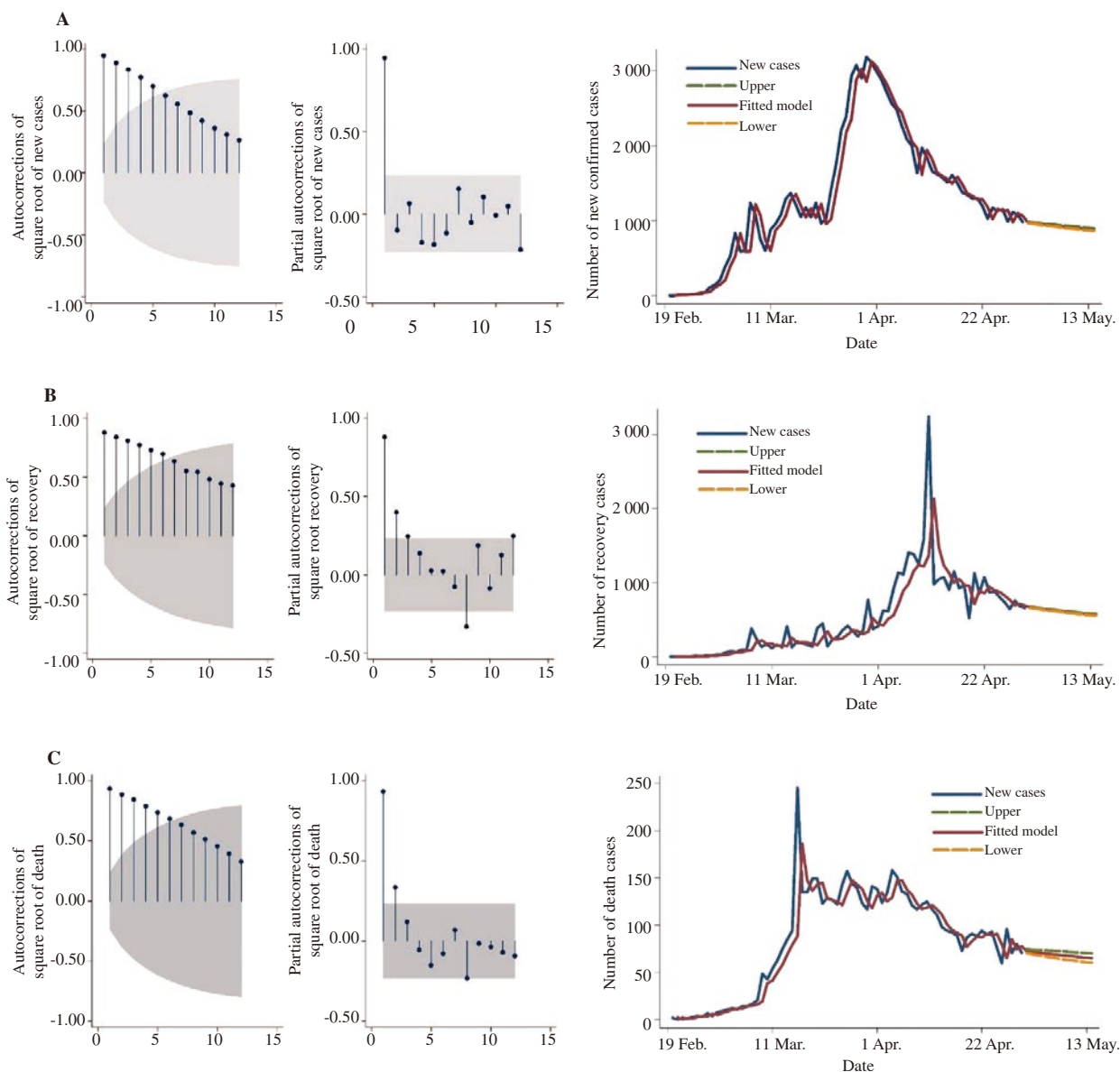


Figure 1. Autocorrelation and partial autocorrelation functions plots and daily observed numbers of series of COVID-19, fitted values (20th February to 30th April) and 1-step ahead predicted values (14 days ahead).

Where Y_t of events is the observed number, Υt is the forecast values at time t , and N is the number of events[10]. The statistical significance level was set at 0.05. Stata (ver.14) was used as the software for the statistical analysis.

The trend of the actual and predicted number of cases for each series of COVID-19 including new cases, recovery, and death cases for 71 days from 20th February 2020 to 30th April 2020 is presented in Figure 1. Also, these graphs indicate the forecast numbers for the 14 days ahead as shown in see Table 1.

After the stationary tests, the square root transformation was used for the new cases, death cases, and recovery cases, and no one needed a regular difference. All statistical methods were performed on the transformed data. Autocorrelation functions and partial autocorrelation functions plots were drawn for each series of

COVID-19 cases. In these charts, the grey zone displays the 95% confidence interval and the lines that are continuously out of range is considered as significantly different (Figure 1).

The potential ARIMA models for the new cases of COVID-19 cases were ARIMA (1, 0, 0) and ARIMA (1, 0, 1), and for the recovery cases were ARIMA (1, 0, 1) and ARIMA (2, 0, 1). Finally, ARIMA (1, 0, 1) and ARIMA (1, 0, 0) were recruited for the death cases.

The goodness-of-fit of the models was evaluated by using the Ljung-Box (Q) test and AIC. The ARIMA (1, 0, 0), ARIMA (1, 0, 1), and ARIMA (1, 0, 1) were selected for determining the new confirmed cases, the death cases, and the recovery cases as the best ARIMA models, respectively.

Table 1. The forecast values (95% CI) according to fitted models of COVID-19 for the period from 1st May to 14th May 2020.

Events	Date	Forecast values	Lower values	Upper values
New confirmed cases	1st May	974.83	969.20	980.46
	2nd May	966.81	958.90	974.71
	3rd May	958.93	949.32	968.55
	4th May	951.20	940.18	962.22
	5th May	943.61	931.37	955.85
	6th May	936.15	922.84	949.47
	7th May	928.83	914.55	943.11
	8th May	921.64	906.47	936.80
	9th May	914.57	898.60	930.55
	10th May	907.64	890.91	924.36
	11th May	900.82	883.40	918.25
	12th May	894.13	876.05	912.21
	13th May	887.56	868.86	906.25
	14th May	881.10	861.83	900.37
Recovery cases	1st May	1 321.58	1 308.54	1 334.62
	2nd May	1 303.54	1 289.07	1 318.01
	3rd May	1 285.95	1 270.21	1 301.6
	4th May	1 268.77	1 251.91	1 285.63
	5th May	1 252.0	1 234.13	1 269.89
	6th May	1 235.66	1 216.84	1 254.47
	7th May	1219.69	1 200.02	1 239.36
	8th May	1204.11	1 183.64	1 224.58
	9th May	1188.90	1 167.69	1 210.10
	10th May	1174.06	1 152.16	1 195.95
	11th May	1159.56	1 137.02	1 182.10
	12th May	1145.41	1 122.26	1 168.55
	13th May	1131.59	1 107.87	1 155.30
	14th May	1118.09	1 093.83	1 142.34
Death cases	1st May	72.34	70.32	74.36
	2nd May	71.72	69.27	74.17
	3rd May	71.12	68.32	73.92
	4th May	70.53	67.42	73.63
	5th May	69.94	66.57	73.32
	6th May	69.37	65.76	72.99
	7th May	68.82	64.98	72.65
	8th May	68.27	64.23	72.31
	9th May	67.73	63.50	71.96
	10th May	67.20	62.80	71.60
	11th May	66.68	62.12	71.25
	12th May	66.18	61.46	70.89
	13th May	65.68	60.82	70.54
	14th May	65.19	60.20	70.18

Table 2. Characteristics of the best ARIMA fitted models series of COVID-19 from 20th February to 30th April.

Events	The best model	Ljung-Box test ^a		AIC [#]	Training RMSE [†]
		P-value	Statistics		
New confirmed cases	ARIMA (1, 0, 0)	0.92	22.08	360.85	2.86
Recovery cases	ARIMA (1, 0, 1)	0.69	28.34	480.87	6.63
Death cases	ARIMA (1, 0, 1)	0.81	25.62	216.78	1.16

^aThe portmanteau test for residuals white noise. [#]Akaike Information Criterion. [†]The root means square error.

The results of the goodness-of-fit of the models are presented in Table 2. Note that this is only for the best-fitted models. For the residuals of the selected models, it is shown that the data were completely modelled.

We found models based on the best models that fit for each series of COVID-19 between 20th February 2020 and 30th April 2020 and then forecasted them for 14 days ahead (Figure 1A and Table 1). Next, we compared the actual data of COVID-19 events with the predicted cases. The predicted models are approximately in line with the real death and new confirmed cases, but the recovery cases are less precise than others as shown in Figure 1. The formula of models is as follows:

The equation of daily new laboratory-confirmed cases is:

$$\text{SQRT New cases}_t = 22.06 + 0.98 \text{ SQRT New cases}_{t-1} \quad (3)$$

Eq. (3) indicates that an increment in the square number of new cases at this time leads to increase of 98% in square number for new cases, one day ahead ($P < 0.001$), and the Wald test is significant. After an exponential increase in the middle of the epidemic period, the situation is converted as shown in Figure 1A. In a short time, we have predicted a declining trend in the occurrence of new cases.

The equation of the recovery cases is:

$$\text{SQRT Recovery cases}_t = 22.20 + 0.98 \text{ SQRT Recovery cases}_{t-1} - 0.50 \epsilon_{t-1} \quad (4)$$

Eq. (4) shows that the rising square number of recovery cases at this time results in a significant increase of 98% in the square number of recovery cases one day ahead, and a negative correlation with the deviation in one time ago ($P < 0.001$). The Wald test is significant. This character is shown in Figure 1B, where we expect to see a somewhat decreased trend over time.

The equation of death cases is as follows:

$$\text{SQRT Death cases}_t = 6.12 + 0.98 \text{ SQRT Death case}_{t-1} - 0.30 \epsilon_{t-1} \quad (5)$$

Similarly, Eq. (5) shows that the increasing square number of death cases at this time leads to an increase in the square number of death cases one day ahead, and a negative correlation with the deviation in one time ago ($P < 0.001$). The Wald test is significant. Figure 1C shows that after an exponential increase in the early stage of the epidemic period, the situation converted. In a short time, we predict a smoothly decline trend in the occurrence of death cases as shown in Figure 1C.

The forecasting in this study was based on the primary time series methods. This means that it is affected by the outlier data, not considering the unknown noise. Therefore, the models have better performance for the short term, but the findings should be explained with thriftiness[8,9]. However, the application and interpretation of these models are simple and is an immediate tools for monitoring systems[6,7].

The government of the Islamic Republic of Iran advised to close the educational centers and locked down activities and other confronting approaches from the earliest days of the outbreak on 24th February.

It is noted that Iranian people celebrate their own new year starting on 21th March 2020. They follow the calendar which is based on solar and is different from Christian's calendar. In their new year, people visiting family and friends traditionally and this results in the growing number of contacts between people which eventually can increase the number of new cases and casualties with the spread of COVID-19. It is anticipated that these patterns may be repeated during or after the Ramadan (the holy month in Islam) due to crowding people for praying in mosques and holy shrines. Therefore, the government should consider preventing measures to control the spread of the viruses under these conditions.

The predicted number of new confirmed, death, and recovery cases indicated somewhat is decreasing. The goodness-of-fit criteria were suitable for these events. However, the confirmed cases can rise remarkably, unless necessary preventive measures are kept in place.

In conclusion, the proposed models in this work can act as a predictive tool for public health planning for better understanding of the dynamics of COVID-19 in a resource-constrained context with minimal data entry. Updating these data can be highly useful for an accurate predictions.

Conflict of interest statement

The authors declare that there is no conflict of interest.

Acknowledgements

This study was conducted using existing COVID-19 data on the web official site of the World Health Organization and did not impose additional costs. The authors would like to thank for the support received from Mashhad University of Medical Sciences (Identified No: IR.MUMS.REC.1399.140).

Authors' contributions

All authors contributed equally in conceptualizing the article, retrieving related literature and drafting the final manuscript.

References

- [1] WHO. *Coronavirus 2020*. [Online]. Available from: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/> [Accessed on 2 May 2020].
- [2] Eubank S, Eckstrand I, Lewis B, Venkatramanan S, Marathe M, Barrett C. Commentary on Ferguson, et al. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. *Bull Mathem Biol* 2020; **82**: 1-7.
- [3] Anderson RM, Heesterbeek H, Klinkenberg D, Hollingsworth TD. How will country-based mitigation measures influence the course of the COVID-19 epidemic? *Lancet* 2020; **395**(10228): 931-934.
- [4] Deb S, Majumdar M. A time series method to analyze incidence pattern and estimate reproduction number of COVID-19. [Online]. Available from: <https://arxiv.org/abs/2003.10655>. [Accessed on 2 May 2020].
- [5] Benvenuto D, Giovanetti M, Vassallo L, Angeletti S, Ciccozzi M. Application of the ARIMA model on the COVID-2019 epidemic dataset. *Data Brief* 2020; **29**: 105340.
- [6] Anne R. ARIMA modelling of predicting COVID-19 infections. medRxiv. 2020. doi: doi.org/10.1101/2020.04.18.20070631.
- [7] Ding G, Li X, Shen Y, Fan J. Brief analysis of the ARIMA model on the COVID-19 in Italy. medRxiv 2020. doi: doi.org/10.1101/2020.04.08.20058636.
- [8] Beckett S. *Introduction to time series using Stata*. TX: Stata Press College Station; 2013.
- [9] Rahmanian V, Bokaie S, Rahmanian K, Hosseini S, Firouzeh AT. Analysis of temporal trends of human brucellosis between 2013 and 2018 in Yazd Province, Iran to predict future trends in incidence: A time-series study using ARIMA model. *Asian Pac J Trop Med* 2020; **13**(6): 272-277
- [10] Esmailzadeh N, Bahonar A, Foroushani AR, Nasehi M, Shakeri MT. Which type of univariate forecasting methods is appropriate for prediction of tuberculosis cases in Razavi Khorasan Province? A need for surveillance and biosurveillance systems. *J Archiv Milit Med* 2019; **7**(3): e96229.