Forecast and analysis of COVID19 Disease 2019 in Iran

Soheil Nosratabadi, Mohamad Bagher Davoodi, Hassan Shafiee alavi, Shaghayegh Zonobi,
Maryam Mirshekari

Abstract— **Since December 2019, coronavirus (COVID-19) pneumonia has occurred in Wuhan, and infection has spread everywhere. Iran is still in its infancy, and there is no indication that the epidemic has reached its peak now, as there is no downward trend in disease and death. Nevertheless, using statistical and data mining methods, future changes can be estimated based on daily mortality. Since the reported statistics in Iran have the highest mortality rate among other countries, we have also been using data from our country. In this study, we intend to predict the daily mortality of coronavirus disease patients in Iran. We have used mortality data from Iran's provinces and estimated implementation based on ARIMA models. The results show that the severe risk of coronavirus reaches its peak point in the coming days of April 4th and is optimistic for between 200 and 300 deaths daily. The results of this model are used for the SIR method with 2 contact rates, to determine the number of the infection and the decay endpoint. Implementation results show that, in the best scenario, the number of infected coronaviruses will reach zero in May-16.**

Index Terms—**COVID19 - forecast mortality – SIR Model – ARIMA Model**

I. Introduction

More than 3,000 cases of COVID-19 have been reported in China since the first cases were reported late last year in Wuhan City, Hubei Province, China. These include all definitive laboratory and clinical diagnoses in Hubei Province, China. Most of the reports were from Hubei and neighboring provinces, but there have been numerous reports from provinces and counties throughout China. Medical data, observation of discharge, infection, non-severe, contaminated, and severe, symptoms of illness and death were collected and analyzed, and there is still an increasing increase in the disease in the coming days. Moreover, the epidemic is still on the rise, though it seems to have slowed slightly in recent weeks. One of the most likely reasons for the decline in people's physical, social connections is the numerous training that encouraged people to stay home.

The World Health Organization on March 11 declared COVID-19[1]. The outbreak of the Coronavirus began in late December last year in Wuhan, China, and now reaches more than 120 countries. Despite these doubts and rumors, however, some modeling and statistical analysis follow scenarios to predict the approximate time of the end of the epidemic in different countries. Miguel and Babak's study suggests that climate change is followed first by people in temperate and warm weather, and then by people in dry weather. SARS-CoV-2 Coronavirus outbreaks have also been concentrated in the Northern Hemisphere, mainly in Asia, the Middle East, Central, South, and Western Europe and the United States, and the virus follows a seasonal pattern and its prevalence favored by dry, dry weather. While in extreme cold and heat conditions as well as moisture reduced[2].

In the 2020 study, Ke and et al. [3]. Used the Chinese experience for Iran and predicted future scenarios for Iran. They have used a generalized logistic growth model to estimate South Korea, and it predicted that within 20 days, 7928 people would infected, and 15% of Italians will be infected with the coronavirus. Unlike South Korea and Japan, Iran and Italy are on the wrong track, with a prevalence rate of 3.6 percent in Iran, which expects Iran to reach 14994 within five days (Mar-15). In the positive scenario, it is predicted that in 10 days (Mar-20), Iran will have 10719 patients, which, unfortunately, has not happened and that the prevalence rate follows the first cynical scenario.

Another study [4] using this Transition matrix model predicts what the peak inflation time (maximum open infections) will be in Hubei province. Also, this model was used to predict patient distribution (severe, non-severe) for better allocation of medical resources, which, under the pessimistic relative scenario, would be the peak inflation time of April 6 - April 14. Moreover, the number of patients will reach 2200-2700.

In a study by xoing and yang[5], the exposed-identified-Recovered (EIR) model used to simulate the disease distribution process. With two approaches before and after the virus intervention, they have made assumptions that warn of the potential dangers of quarantine. They announced that the outbreak would be minimal by the end of March. Quarantine may also result in a 27% decrease in the rate of growth of virus distribution. They announce that if the intervention start date postpones for one day from January 23, the infected peak population will increase to about 6,351. If the delay time is 3 days or 7 days, the increasing number will be 21621 or 65929 people, so the peak number of infected will be 70714 and 115022 people.

A study by Lixiang and et al. also found that virological transmission follows a Gaussian distribution. The Gaussian distribution used to demonstrate the virus's ability to spread. They used a formula to determine the transmission capacity to estimate the rate of disease distribution in Hubei. They expect corona infection to reach 20,000 by the end of March and then be substantially controlled by the beginning of April. It was found that Iran was infected on January 13[th][6].

In a different study, different nonlinear and Arima methods are done for coronavirus decay and outbreak rates in the world. In this study, we use the ARIMA and SIR models to investigate the decay rate and the infected number within the 60-day period. In the following, we evaluate the Iranian model with two different methods and its results with the ARIMA and SIR methods and finally provide its general conclusions.

II. PROCEDURE AND METHOD

The implemented method consists of two different parts, Arima and SIR, and it consists of 5 steps (figure 1). In the first step, we estimate the daily mortality rate in the algorithm, and in the second step, we estimate the future mortality. And based on the mortality rate, we reach the number of infections in the peak. And we'll give that value to the SIR model. Eventually, we will find the amount of recovered, infected, and recoveries associated with the day.

The purpose of the Arima model is to find the daily mortality and find its peak, and we convert this mortality rate to the contaminated rate, so we estimate the total peak population in the middle scenario. The goal of the SIR model is to come with two strict modes of call rate, the last day of virus decay, and the infected and the susceptible.
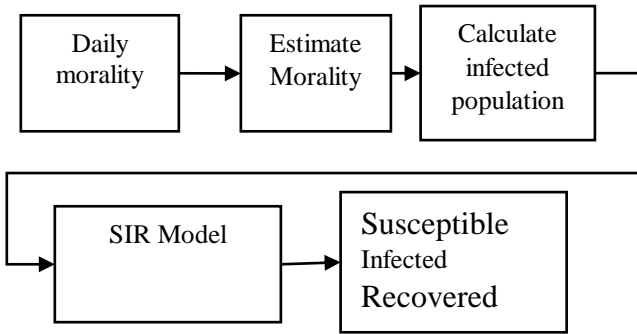

Figure 1. Hybrid Arima and SIR Model

A. Method (1)

One of the popular methods of time series modeling mainly uses for time series prediction. The methodology of this method base on the stochastic analysis of time series based on their information.

In ARMA models[7], the future value of a variable is a linear function of past observations and the magnitude of a random error. The general form of the process of producing a time series is given by Equation (1):

$$y_t = \theta_0 + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \theta_q \varepsilon_{t-q} \quad (1)$$

In equation (1), the y_t time-series values, φ, and θ are the moving average and the autoregressive model, respectively, ε is a stochastic process with the mean zero and the variance σ^2. It is worth noting that one of the requirements for applying ARMA models is the reliability of the time series studied, which is rarely the case for economic variables. Hence, ARMA models have extended to be used even when time series are unstable. In this case, ARMA models are called ARIMA, and their degree of dependence is the degree of dependence of the time series studied. The process of applying this method involves four stages of diagnosis, estimation, diagnostic control, and prediction. After estimating and fixing it as the final model, the model is predicted.

ARIMA finally recognizes a random step with a deviation for mortality; that is, the motion of death is random but follows a targeted pattern over time. This process, which usually follows a random deviation step, means that the targeted patterns are critical in the short term and that the following

28-day forecast can be examined as follows: Ideally, ARIMAX is used to predict an ARIMA model amidst explanatory variables in mind. However, when the goal of predicting a time series is purely base on previous values, ARIMA is a standard model.

B. Actual Specification Tests

Unlike the stipulation error that we consider a given model to compare the results, in the model stupid error, we do not already have the correct model in mind. Usually, we come across two or more competing models or theories that each require investigation. One way to choose between competing models is to combine them into one more substantial model, and then decide on the meaningfulness of the models above.

The most general method base on good fit and simplicity of the model, and it defines that the model and the variable with the lowest Akaike Information Criterion (AIC) values are the most appropriate model and the most important variable. Furthermore, equation (2) describes this model.

$$AIC = n \log \left\{ \frac{RSS}{n} \right\} + 2P \quad (2)$$

n is the number of model parameters, L is the function likelihood of the model probability.

Schwarz proposed the Bayesian information criterion [9], and it is defined that the model and the variable have the lower Bayesian information criterion (BIC) value, the more appropriate model, and the more important variable. This criterion for finite samples, due to the large number of negative points considered for complexity, often selects elementary models. The BIC is a measure of good fit evaluation and also calculates the penalty for increasing the number of parameters. Therefore, models that typically have a smaller AIC and BIC are preferred.

$$BIC = n \ln(\text{RMSE}) + P \ln(n) \quad (3)$$

C. Method (2)

Since we are in the medical community with a Compartmental model of people, infectious, susceptible, and recovered, these people enter the cycle of epidemiological models, and this model called the SIR. We are looking for a decay point in the Iranian Coronavirus epidemic. In this study, we have studied the mathematical epidemiology and modeling of coronavirus disease to determine the epidemic in Iran from the perspective of mathematical epidemiology.

Three states are likely to occur for SIR values. If the value is higher than one, the risk of SIR epidemics or the increase in the number of people affected will be higher, and the population is at greater risk. Therefore, we should try to keep the SIR value less than one or at least equal to one so that it does not threaten the epidemic area of our disease. Different studies used to estimate SIR in this field. We used the equation (4,5,6) of Kermack and McKendrick (1927).

$$\frac{dS}{dt} = - \frac{\beta SI}{N} \quad (4)$$

$$\frac{dI}{dt} = \frac{BSI}{N} - \gamma I \qquad (5)$$

$$\frac{dR}{dt} = \gamma I \qquad (6)$$

Where β is the contact rate and γ is the average rate of recovered and N is the total population. S number of susceptible and I contagious and R recovered.

III. RESULT

In figure 2, the coronavirus mortality figures showed while most of the deaths reported in cities with the coronavirus in Qom, Tehran, Mazandaran, Alborz, Esfahan, and Gilan.
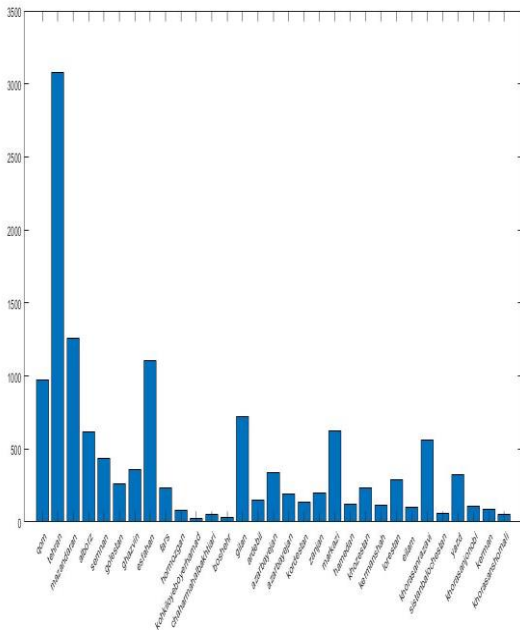

Figure 2. Number of people infected with coronavirus to
03-15-2020 in different cities

In Figure 3, the rate of increase in patients is increasing more than expected, with a mortality rate of 6.5% of total confirms in 27 days, and a recovery rate of 32% of total confirms.
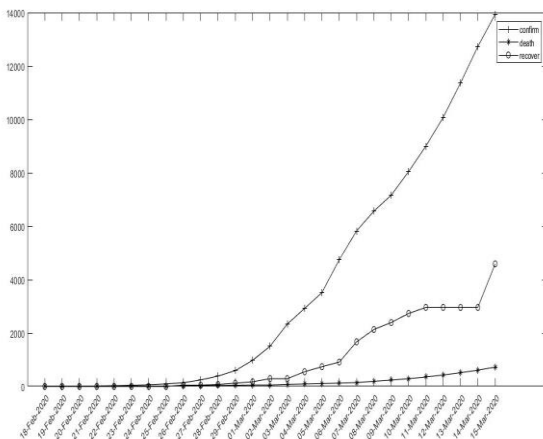

Figure 3. Coronavirus confirm, death, recover in Iran

IV. DAILY MORTALITY ESTIMATES USING THE ARIMA MODEL

Table 1 calculates the Autoregressive and Moving Average values based on the P-value, assuming values less than 0.05 are assumed. With this state showing better Autoregressive value, optimistically, this could be an Autoregressive of future deaths for Iranians on holiday, and in a pessimistic state, it could return. It is possible to have a daily death toll of 500. In Figure 4, the model is estimated by ARIMA from 2020-01-26 to 2020-03-22.

Table 1. Estimation Results ARIMA(1-1-1)

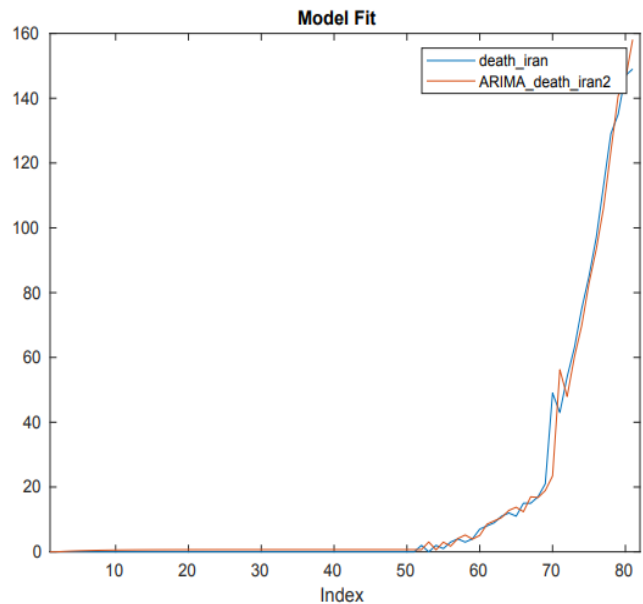| Parameter	P-Value
AR{1}	9.734e-185
MA{1}	0.77556

Figure 4. Plot the fit of model ARIMA_1_1_1 time series
Daily Death Iran.

In Table 2, comparing ARIMA (2-2-1), it can be seen that the value of p_value in Autoregressive or AR {1} is less than 0.05, but AR {2} violates the Autoregressive model. Therefore in this study, we propose the ARIMA model (1-1-1). The estimation results of this model calculation shown in Figure 5.

Table 2. Estimation Results (2-2-1)

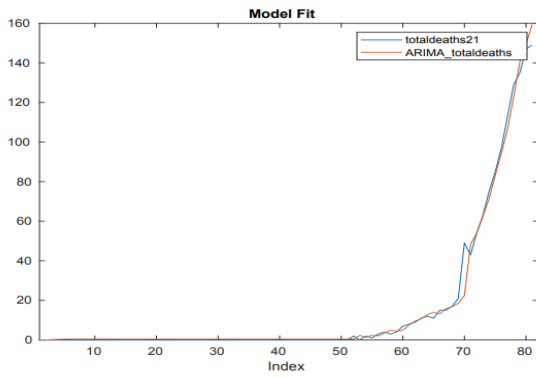| Parameter	P-Value
AR{1}	0.012253
AR{2}	0.28288
MA{1}	0.018296

Figure 5. Plot the fit of model ARIMA_2_2_1 time series Daily Death Iran.

By comparing the AIC and the BIC (Table 3), we find that the minimum value for the ARIMA model is (1-1-1), so we will follow the prediction process of this study with this model. Model BIC (2-2-1) increases due to excessive parameter increase, and this has resulted in a higher BIC penalty.

Table 3. Goodness of Fit

	ARIMA(2-2-1)	ARIMA(1-1-1)
AIC	365.3521	342.1327
BIC	374.9315	351.7918

In Figure 6, we performed the ARIMA (1,1,1) model from 2020-02-23 to 2020-03-22 and estimated Iran's daily mortality. We eliminate the first 15 days; we did in the model above because of zero mortality in the early onset of coronavirus in Iran.

Given that the number of deaths has reached 1556 on 21-23, we estimate that the deaths will approach 4,000 in the next 15 days. Although official holidays have begun in Iran with the start of the New Year, deaths continue to increase exponentially. We estimate that Iran will be optimistic, with an estimated 150-200 deaths in the next 15 days (April-06) and a pessimistic death rate of 200 to 300. Because Iran is facing a New year holiday, and this is good news for the people quarantined, the mortality trend will have a downward trend at the start of the new Iranian year. This severe and probable risk informs us of future risks and prognosis.
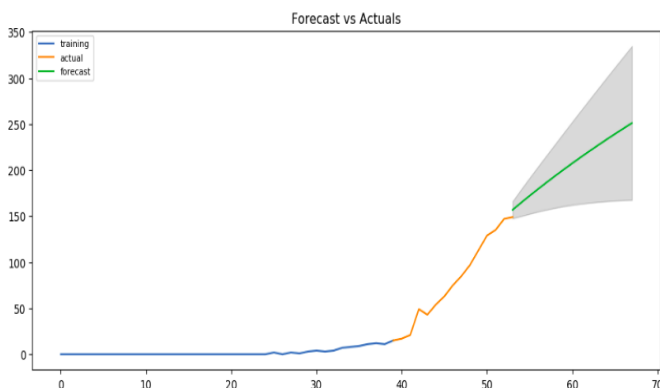

Figure 6. Estimation of the ARIMA model for daily mortality in Iran

V. SIR MODEL

To calculate the number of infected people in the median of Arima, we need to calculate the forecast mortality rate based on the current mortality rate. At present, the number of deaths is about 2517, which represents a rate of 7.1% mortality, which means that 35350 people are infected. However, in the median of the Arima method (green line), the number of morality on April- 06 is 4389. The rate of mortality in the will reach 12.42% (equation 7), accordingly in 4389 deaths, 61225 people infected (equation 8), and twice as many as 122450 will be infected.

$$\frac{M_f * M_{rc}}{M_c} = M_{rf} \qquad (7)$$

That is M_f of Arima is total deaths and M_{rc} is the current mortality rate, M_c is current mortality and M_{rf} estimate the death rate.

$$\frac{M_{rf} * I_c}{M_{rc}} = I_f \qquad (8)$$

That's I_c is total case infected and M_{rc} is the current mortality rate, I_f It is infected in the peak of the Arima model.

The simulations performed in three different scenarios, as illustrated in Figure 7. The scenario is considered based on three SIR parameters with different values of contact rate of 0.65, 0.8, and 0.2 and the default population of 122450. Also, according to March 29, infectious 2500 and recovered 35,000 were considered. In Fig. A, if the population density is minimized, the recoveries will increase rapidly (green and yellow lines), and all will recover in the 50th (May 20). In the worst scenario, if the population density is maximized (black line), 80% of the population will be recovered in the next 60 days. In figure B, if the population density is 1/3 as of March 29, the infected rate will then decay (Blackline).

If the density increases, the outbreak takes its maximum on April 6, and after 40 days (May-06), the number of infected will reach zero. The model's results show that after 1, the number of patients will reach zero as well. In Figure C, if people's contact rate is 1/3 as of March 28, then on the best-scenario of April 19, susceptible people will also take zero (Yellow line). In the worst scenario, if the population increases as their density increases, as a result, a 0/3 of the population will still be susceptible to the coronavirus.
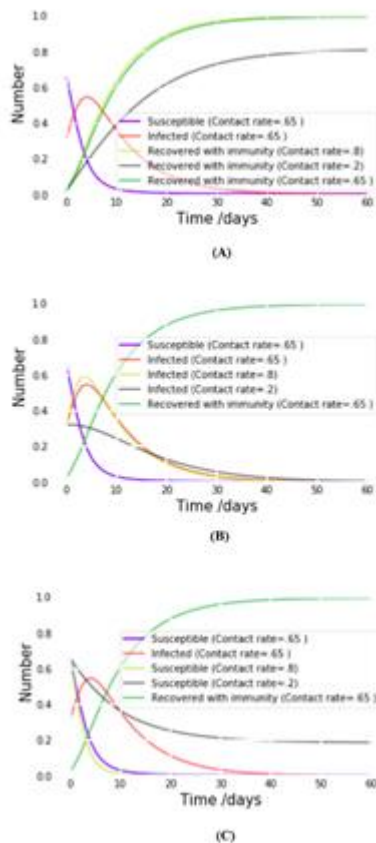
Figure 7. Calculate SIR with different contact rates (0/2, 0/65, 0/8). (A).Different scenarios for computing recovered with immunity- (B). Different scenarios for computing infected-(C).Different scenarios for computing susceptible.

VI. CONCLUSION

What experts know is that respiratory infections such as influenza, cough, and cold have seasonal effects that can predict and control the spread of this virus and disease. It is also clear that specific environmental conditions, such as cold weather, humidity, and how people behave during the winter, affect the epidemic and infectious process and enhance viral transmission.

In this study, we investigated the ARIMA growth model with two different models (1-1-1) and (2-2-1). We introduced the growth model (1-1-1) as the best model. The ARIMA model expected to take its peak of 150 to 200 coroner deaths on April 06, with the most optimistic, and with the most pessimistic estimates, it will reach between 200 and 300 daily. Therefore, medical advice and home quarantine should be more carefully observed, and this alarm is considered a risk for the coming days in Iran.

In this study, to determine the daily mortality trend of Iran, the interval between Feb-18 and 15-march studied. This trend is also increasing daily in major cities (Tehran, Esfahan, Mazandaran, Gilan, Qom) and may face higher risks in the coming days, with rising mortality rates.

Calculating the SIR with different contact rates shows that if we reduce the contact rate by 1/3 on March 28, then we will have a downward trend in Iran up to 40 days later. If the total population was infected on April 5 (highest contact rate), then recovery would be faster, and as a result, the probabilities would take zero on April 19.

REFERENCES

[1] Sohrabi, Catrin, et al. "World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19)." *International Journal of Surgery* (2020).

[2] Araujo, Miguel B., and Babak Naimi. "Spread of SARS-CoV-2 Coronavirus likely to be constrained by climate." *medRxiv* (2020).

[3] Wu, Ke, et al. "Generalized logistic growth modeling of the COVID-19 outbreak in 29 provinces in China and in the rest of the world." *arXiv preprint arXiv:2003.05681* (2020).

[4] Wu, Ke, Junhua Zheng, and Jian Chen. "Utilize State Transition Matrix Model to Predict the Novel Corona Virus Infection Peak and Patient Distribution." *Available at SSRN 3539658* (2020).

[5] Xiong, Hao, and Huili Yan. "Simulating the infected population and spread trend of 2019-nCov under different policy by EIR model." *Available at SSRN 3537083* (2020).

[6] Li, Lixiang, et al. "Propagation analysis and prediction of the COVID-19." *arXiv preprint arXiv:2003.06846* (2020).

[7] Vecchia, A. V. "Maximum likelihood estimation for periodic autoregressive moving average models." *Technometrics* 27.4 (1985): 375-384.

[8] Sakamoto, Yosiyuki, Makio Ishiguro, and Genshiro Kitagawa. "Akaike information criterion statistics." *Dordrecht, The Netherlands: D. Reidel* 81 (1986).

[9] Bhat, Harish S., and Nitesh Kumar. "On the derivation of the Bayesian Information Criterion." *School of Natural Sciences, University of California* (2010).

[10] Kermack, William Ogilvy, and Anderson G. McKendrick. "Contributions to the mathematical theory of epidemics. II.—The problem of endemicity." *Proceedings of the Royal Society of London. Series A, containing papers of a mathematical and physical character* 138.834 (1932): 55-83.

Soheil Nosratabadi, Islamic azad university safashahr branch, fars.

Mohamad Bagher Davoodi Data Scientist at Company RADCO (sydney).

Hassan Shafiee alavi, Jahad Daneshgahi Academic Institute Of Shiraz.

Shaghayegh Zonobi, Jahad Daneshgahi Academic Institute Of Shiraz .

Maryam Mirshekari, Data Analyst at Atiye Shahr Hoshmand Iranian (Cpay).