



# Forecasting COVID-19: Vector Autoregression-Based Model

Khairan Rajab<sup>1</sup> · Firuz Kamalov<sup>2</sup> · Aswani Kumar Cherukuri<sup>3</sup>

Received: 12 April 2021 / Accepted: 28 October 2021 / Published online: 3 January 2022  
© King Fahd University of Petroleum & Minerals 2022

## Abstract

Forecasting the spread of COVID-19 infection is an important aspect of public health management. In this paper, we propose an approach to forecasting the spread of the pandemic based on the vector autoregressive model. Concretely, we combine the time series for the number of new cases and the number of new deaths to obtain a joint forecasting model. We apply the proposed model to forecast the number of new cases and deaths in the UAE, Saudi Arabia, and Kuwait. Test results based on out-of-sample forecast show that the proposed model achieves a high level of accuracy that is superior to many existing methods. Concretely, our model achieves mean absolute percentage error (MAPE) of 0.35%, 2.03%, and 3.75% in predicting the number of daily new cases for the three countries, respectively. Furthermore, interpolating our predictions to forecast the cumulative number of cases, we obtain MAPE of 0.0017%, 0.002%, and 0.024%, respectively. The strong performance of the proposed approach indicates that it could be a valuable tool in managing the pandemic.

**Keywords** COVID-19 · Forecasting · VAR · ARIMA · Autoregression

## 1 Introduction

The COVID-19 pandemic has brought tremendous challenges to the governments and public health authorities around the globe. One of the key aspects to managing the governments' response to the pandemic is forecasting the spread of the infection. Accurate forecast of the expected number of new cases can help the authorities better plan their policies and actions to achieve the optimal outcome. In this paper, we propose a vector autoregressive (VAR) model to forecast the daily number of new cases and deaths. The proposed algorithm produces accurate results and outperforms the existing state-of-the-art forecasting models.

The ability to accurately forecast the spread of the infection allows governments to make smart and informed decisions. If the number of infections is expected to rise sharply, the government may consider imposing a lockdown in order to stop the spread of the virus. On the other hand, if the number of new cases is expected to decline the government may consider easing some of the social and economic restrictions to improve the quality of life. Similarly, accurate forecast allows public health authorities to better manage their limited resources.

Forecasting COVID-19 infection has received a considerable amount of attention among researchers. However, our method differs from the existing approaches in two important ways. First, we consider the number of new cases in conjunction with the number of deaths. We believe that the two time series are related. Consequently, the information about one time series can be used to forecast the other time series. Using the VAR model, we can take into account the cross-correlation between the series and achieve more accurate forecasts. Second, we use data from some of the most extensively tested populations in the world—the UAE, Saudi Arabia, and Kuwait—to train our model. So the data accurately represent the true prevalence of the infection in each country. Additionally, our data cover a 12-month period which is considerably longer than many existing studies. We

This work was supported in part by the Najran University under Grant NU-/SERC/10/597.

✉ Firuz Kamalov  
firuz@cud.ac.ae  
Khairan Rajab  
khairanr@gmail.com  
Aswani Kumar Cherukuri  
cherukuri@acm.org

<sup>1</sup> Najran University, Najran, Kingdom of Saudi Arabia

<sup>2</sup> Canadian University Dubai, Dubai, UAE

<sup>3</sup> Vellore Institute of Technology, Vellore, India



believe that the quality and depth of the data lead to more reliable results.

Although vector autoregression was used to forecast COVID-19 in the past [1], its application has been sparse. Our approach is based on a careful study of the times series plots, correlation plots, and information criteria. Remarkably, our analysis produced the same order model for each country. It indicates that the proposed approach can be adopted to forecast the infection rates in other countries. To test the efficacy of our approach, we measured the model accuracy based on 10-day ahead forecast. The mean absolute percentage error for the three countries is 0.0017%, 0.002%, and 0.024%, respectively.

Our paper is structured as follows. In Sect. 2, we briefly discuss the existing efforts in the literature to forecast the COVID-19 infection. In Sect. 3, we provide the required theoretical background about the VAR model. In Sect. 4, we construct and apply the proposed model to forecast the infection rates in the UAE, Saudi Arabia, and Kuwait. We conclude the paper with a few closing remarks in Sect. 5.

## 2 Literature

The topic of forecasting COVID-19 has recently attracted a significant amount of attention in the literature. A number of different forecasting approaches have been explored. The results vary depending on the model and data used in the study. Despite the considerable volume of research dedicated to the subject, the results have sometimes been criticized as inadequate [2]. One of the issues with the existing attempts in the literature is the size and the quality of the training data. The data are often too little or originate from countries with low testing rates [3]. To address this issue in our paper, we employ a 12-month dataset from rigorously tested countries. Another issue is the vulnerability of the underlying assumptions of the model. Most of the forecasting models are based on certain assumptions about the time series. If the underlying assumptions are not satisfied, then the model is not technically sound.

The majority of the existing forecasting methods can be grouped into three categories: autoregressive integrated moving average (ARIMA) models, mathematical growth models, and machine learning (ML) models. In an ARIMA model, the values of an individual time series are forecasted based on a linear combination of the past values and random shocks. Formally, ARIMA models are denoted  $ARIMA(p, d, q)$ , where  $p$  is the number of time lags of the autoregressive model,  $d$  is the degree of differencing, and  $q$  is the order of past shocks. The authors in [4] find  $ARIMA(0, 1, 0)$  to be the best fit for predicting the trend of daily confirmed COVID-19 in Malaysia. The authors use the data from January 22 to March 31, 2020, for training and April 1 to April 17, 2020,

for testing. The test results show MAPE of 16.01%. In a similar study [5], the authors attempted to estimate the total daily infected cases from the top five countries: US, Brazil, India, Russia, and Spain. The authors obtained different optimal ARIMA models for each country: (4,2,4), (3,1,2), (3,0,0), (4,2,4), and (1,2,1), respectively. Model specifications were estimated using Hannan and Rissanen algorithm [6]. The data for the study were taken from February 15 to June 30, 2020, for training and July 1 to July 18, 2020, for testing. The MAPE for each country is 3.701%, 1.844%, 1.090%, 0.832%, and 2.885%, respectively.

The mathematical growth (contagion) models are based on differential equations that model the spread of infection. In [7], the authors forecast the total number of daily confirmed and death cases in India using several models based on gene expression programming. The data for the study are taken from April 7 to May 5, 2020. The results show root-mean-squared error on training data: The confirmed and death cases are 5.5574 and 90.1863, respectively. The authors do not provide out-of-sample forecasting and testing. In [8], the authors forecast the total number of daily infected individuals in Brazil, UK, and South Korea using discrete-time-evolution model based on a set of four equations. The results show MAPE: Brazil 5.25%, UK 4%, and South Korea 3.75%, respectively.

Machine learning models have been used for forecasting in a variety of applications including finance [9,10], energy [11], education [12], temperature [13], and many others. A number of authors have employed ML methods such as regularized linear regression (LASSO) and recurrent neural networks (RNN) to forecast the spread of the infection [14]. In [15], the authors compared RNN, long short-term memory (LSTM), BiLSTM, gated recurrent units (GRU), and variational autoencoders (VAE) to forecast the total number of daily cases in Italy, Spain, Italy, China, the USA, and Australia. The study used data from January 22 to June 1, 2020, for training and June 1 to June 17, 2020, for testing. The VAE model achieved the best MAPE values: 5.90%, 2.19%, 1.88%, 0.128%, 0.236%, and 2.04%, respectively. Similar studies using LASSO, SVM, logistic regression, and others have also been carried out in [16,17].

Vector autoregression is used to model the joint dynamic behavior of a collection of time series. It was used in [18] to forecast mortality rates, where mortality rates of each age depend on the historical values of itself and the neighboring ages. The VAR model is commonly used in spatiotemporal settings. For instance, in [19] wind power forecast at multiple plants at different locations was done within a single framework using LASSO vector autoregression. The predicted output from each plant in the model is based on its own past values and the past values of the other plants included in the model. Similarly, LASSO vector autoregression was applied for wind power prediction in [20]. Vector autore-

gression has also been applied in a number of other fields including finance [21], tourism [22], and commodity prices [23].

The application of vector autoregression model in the context of COVID-19 has been limited. For instance, in [1] the VAR model was studied together with linear regression and multilayer perceptron. The authors in [24] used VAR model to forecast the infection, hospitalization, and ICU bed numbers in Italy. None of the two studies include any measures of accuracy to evaluate the models.

### 3 Vector Autoregressive Model

The VAR process is traditionally used to model together two or more related time series. In case of COVID-19, the number of new cases is related to the number of deaths. As the number of new cases increases, so does the number of deaths. Therefore, information about the former can help predict the latter. The VAR process allows to incorporate both the number of new cases and deaths into a single model, producing a more powerful forecasting paradigm. In addition, the VAR process requires minimal assumptions about the nature of the time series. As will be shown in Sect. 4.3, modeling the number of new cases in conjunction with the number of deaths is more effective than modeling the series individually.

A good introduction to vector autoregression can be found

in [25]. Let  $\mathbf{x}_t = \begin{bmatrix} x_{t,1} \\ x_{t,2} \\ \vdots \\ x_{t,k} \end{bmatrix}$  be a vector-valued time series

consisting of  $k$  individual time series. Assume that  $\mathbf{x}_t$  is stationary, i.e., the cross-covariance function  $\text{Cov}(x_{t,i}, x_{s,j})$  depends only on  $s - t$ . Then, the VAR( $p$ ) model is given by the following equation:

$$\mathbf{x}_t = \Phi_1 \mathbf{x}_{t-1} + \Phi_2 \mathbf{x}_{t-2} + \dots + \Phi_p \mathbf{x}_{t-p} + \mathbf{w}_t, \quad (1)$$

where  $\Phi_j$  are matrices of coefficients and  $\mathbf{w}_t$  is the vector Gaussian white noise with  $\text{Cov}(\mathbf{w}_t, \mathbf{w}_s) = 0$  for  $s \neq t$ . In our paper, we examine a vector of two time series, so the corresponding VAR( $p$ ) model is given by the following equation:

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \Phi_1 \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \dots + \Phi_p \begin{bmatrix} x_{t-p} \\ y_{t-p} \end{bmatrix} + \begin{bmatrix} w_{t,1} \\ w_{t,2} \end{bmatrix},$$

where  $x_t$  and  $y_t$  are the numbers of new cases and deaths at time  $t$ , respectively. The coefficients of matrix  $\Phi_j = \begin{bmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{21} & \Phi_{22} \end{bmatrix}$  are estimated based on the maximum likelihood estimation. In other words, the matrix coefficients are

calculated to maximize the likelihood of obtaining the sample. In our paper, we use the `statsmodels` package in Python [26] to implement the VAR model. The order  $p$  of the VAR model is chosen based on a combination of different factors including the time series plots, the correlation plots, and information criteria. In addition, residual analysis is employed to confirm the model assumptions about normality and independence.

## 4 Model Construction and Forecasting

### 4.1 Datasets

The data used in this study consist of the daily number of new cases and deaths. The data are collected for three countries: the UAE, Saudi Arabia, and Kuwait. The countries were chosen due to rigorous COVID-19 testing conducted within the populations [3]. The data range over a 12-month period starting from March 20, 2020, to March 20, 2021. We used the data until March 10, 2021, for training and the data from March 11 to March 20, 2021, for testing. The data are sourced from OurWorldInData.org [27]. The time series plots for the number of new cases and deaths for each country are presented in Fig. 1.

### 4.2 Data Preprocessing

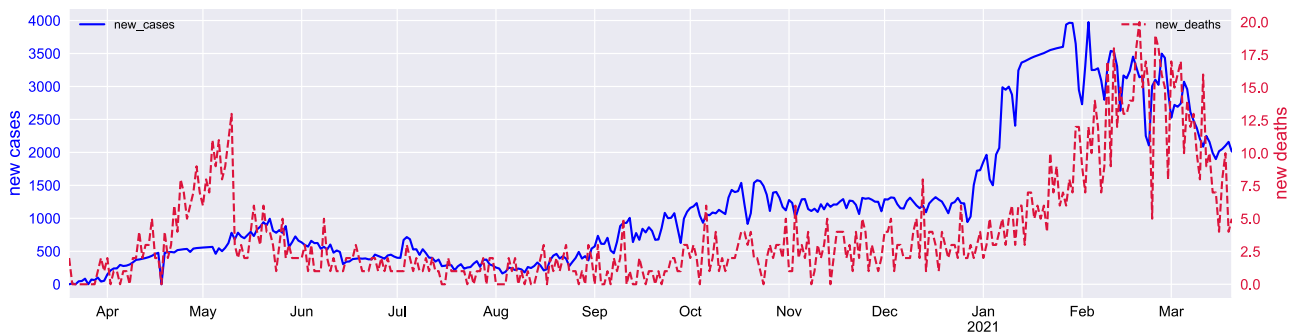
According to the European Centre for Disease Prevention and Control, “the daily number of cases is frequently subject to retrospective corrections, delays in reporting and/or clustered reporting of data for several days.” Therefore, daily variations in the number of cases are unreliable for effective forecasting. To obtain a more balanced and well-founded time series, we employ a 7-day rolling mean. As shown in Fig. 2, using the rolling mean produces a more reliable time series. We can also see from the plots that the time series for the number of cases and number of deaths are correlated with an approximately 30-day lag. For instance, in the case of the UAE, the peak for the number of new cases occurs in the end of January, while the peak for the number of new deaths occurs in the end of February.

The plots in Fig. 2 show that the time series are not stationary. Let  $x_t$  be the value of the times series at time  $t$ . To obtain a more stationary time series, we take the first difference of the time series values:

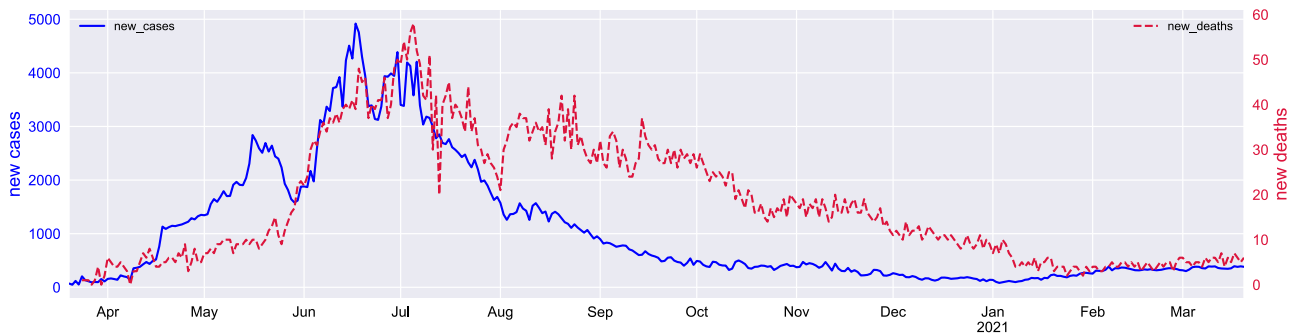
$$\nabla x_t = x_t - x_{t-1}. \quad (2)$$

As shown in Fig. 3, the resulting series is more stationary with a constant mean around zero. There remain spikes in the variance especially in the end of the period (Fig. 3a) which can be attributed to the stochastic realization of the time

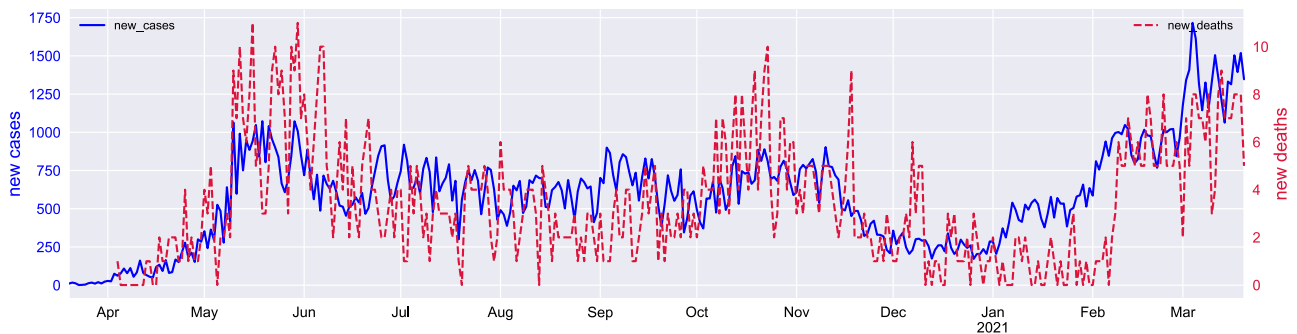




(a) The number of new cases and deaths in the UAE.



(b) The number of new cases and deaths in Saudi Arabia.



(c) The number of new cases and deaths in Kuwait.

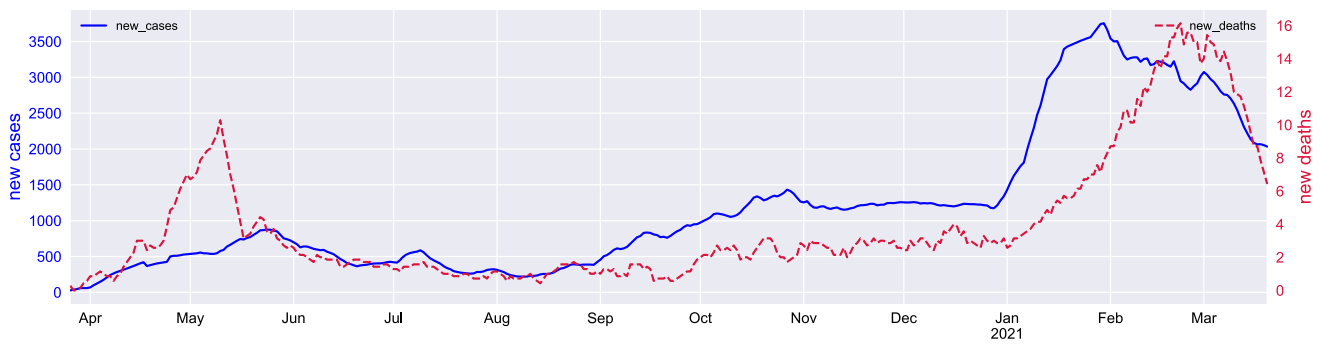
**Fig. 1** The original time series data for the three countries

series. Nevertheless, the transformed series appears largely stable to proceed with our analysis.

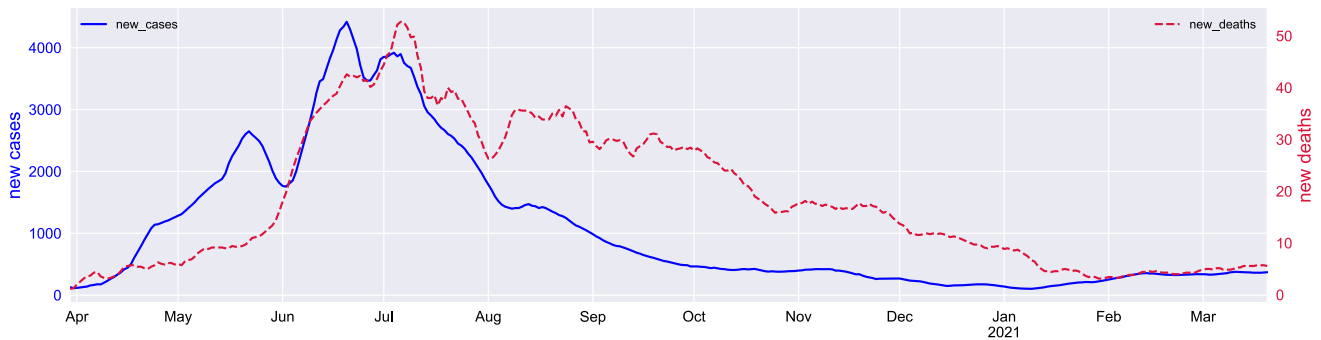
### 4.3 Building the Forecasting Model

The key to building an effective VAR model is determining the correct order. To identify the order of the model we consider 3 factors: time series plots, correlation plots, and information criteria. We illustrate our approach to building a model in the case of the UAE. The models for Saudi Arabia and Kuwait are constructed in similar fashion. Visual examination of the time series in Fig. 1a shows that the time series for the new cases and the time series for new deaths are cor-

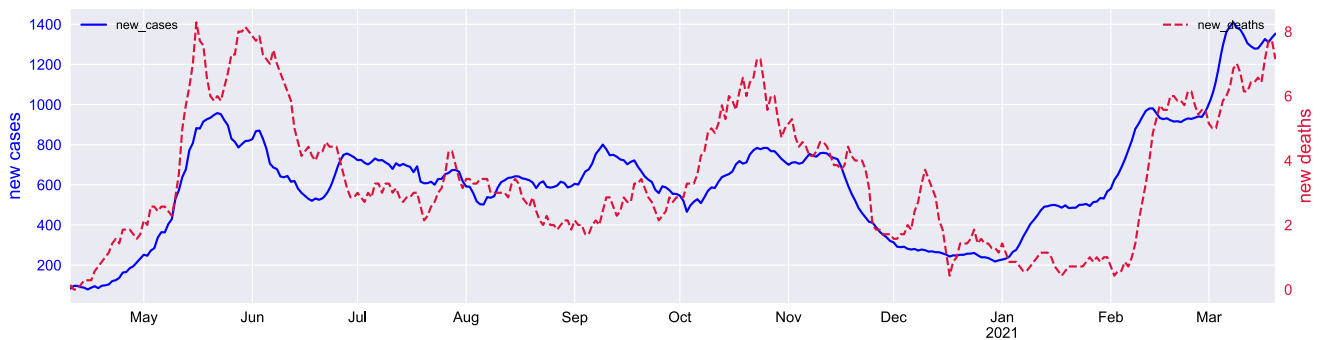
related with an approximately 30-day lag. Next, we consider the auto and cross-correlation plots. As shown in Fig. 4, there exists nontrivial correlation in both time series. Concretely, the autocorrelation plot for the new cases (upper left) contains nonzero values up to lag 28 while for the new deaths (lower right) it is up to lag 17. The cross-correlation function is not even, i.e.,  $\rho_{xy}(s, t) \neq \rho_{yx}(s, t)$ . Therefore, the off-diagonal cross-correlation plots are different. Since the number of new cases leads the number of new deaths, we consider the cross-correlation plot on the upper right. The cross-correlation plot shows nontrivial correlation which indicates that the number of new cases has a lagged correlation with the number of new deaths. So the use of the VAR model is statistically justified.



(a) The rolling mean of the data for the UAE.



(b) The rolling mean of the data for Saudi Arabia.



(c) The rolling mean of the data for Kuwait.

**Fig. 2** The 7-day rolling mean of the number of new cases and deaths

In addition, the nonzero value at lag 29 is consistent with our visual observations in Fig. 1a.

As the final step in identifying the order of the model, we study information criteria metrics. In particular, we calculate the Akaike information criterion (AIC) of the VAR model for the first 33 orders. The AIC is given by the following equation:

$$\text{AIC} = 2k - 2\ln(\hat{L}), \quad (3)$$

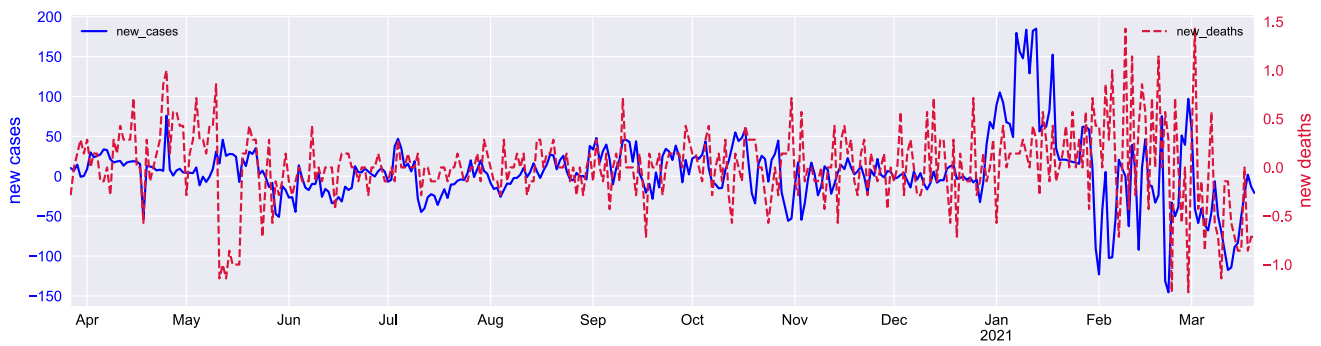
where  $k$  is the number of parameters in the model and  $\hat{L}$  is the maximum sample likelihood for a model. The model with the lowest AIC is considered optimal. As shown in Table 1, the

minimum AIC is achieved at lag 29. This is consistent with our earlier observations. Recall that the plots in Fig. 1a show an approximately 30-day lag between the two time series. In addition, the cross-correlation plot had a maximum nonzero value at lag 29. We conclude that the optimal order for our VAR model is  $p = 29$ .

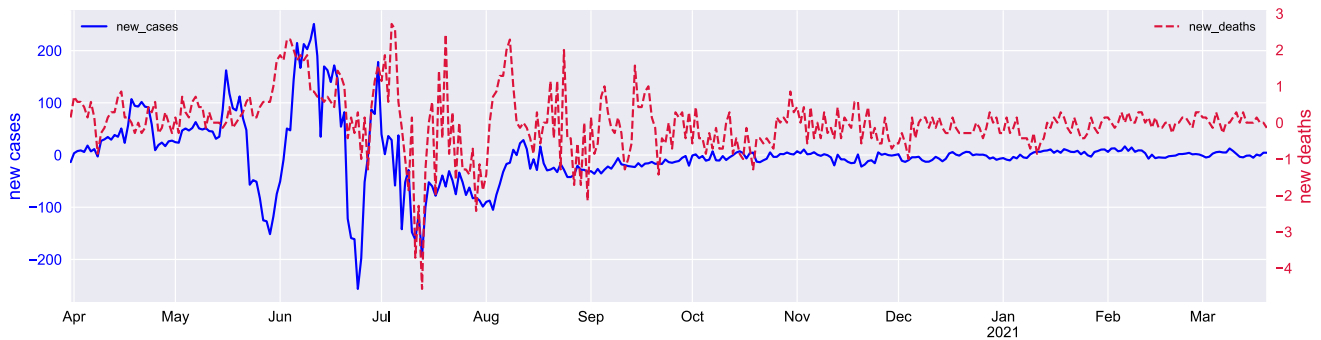
We split the data into training and testing subsets. The training subset encompasses the period March 20, 2020–March 10, 2021, while the testing subset includes the period March 11–March 20, 2021. We fit the VAR(29) model on the training. Then, we use the fitted model to forecast the new daily number of cases and deaths for the testing period. As shown in Fig. 5a, the forecasted numbers of new cases match



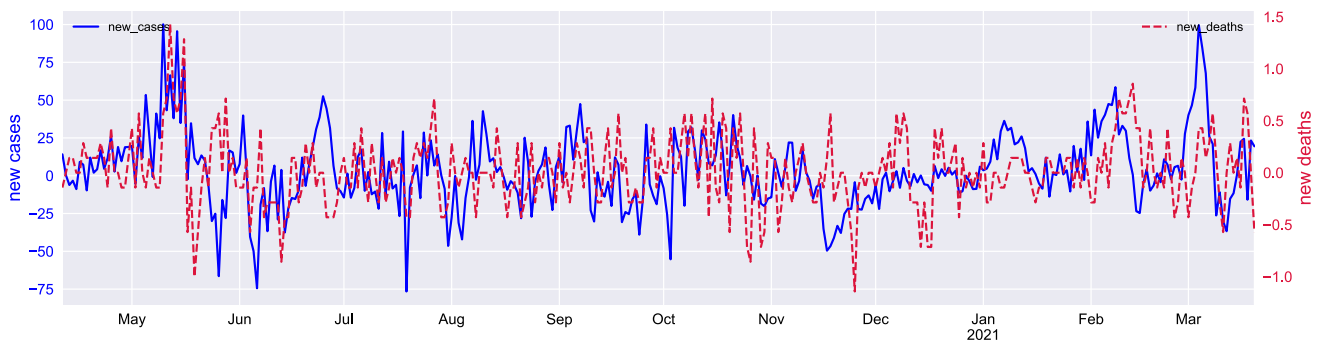




(a) The first difference of the rolling mean series for the UAE.



(b) The first difference of the rolling mean series for Saudi Arabia.



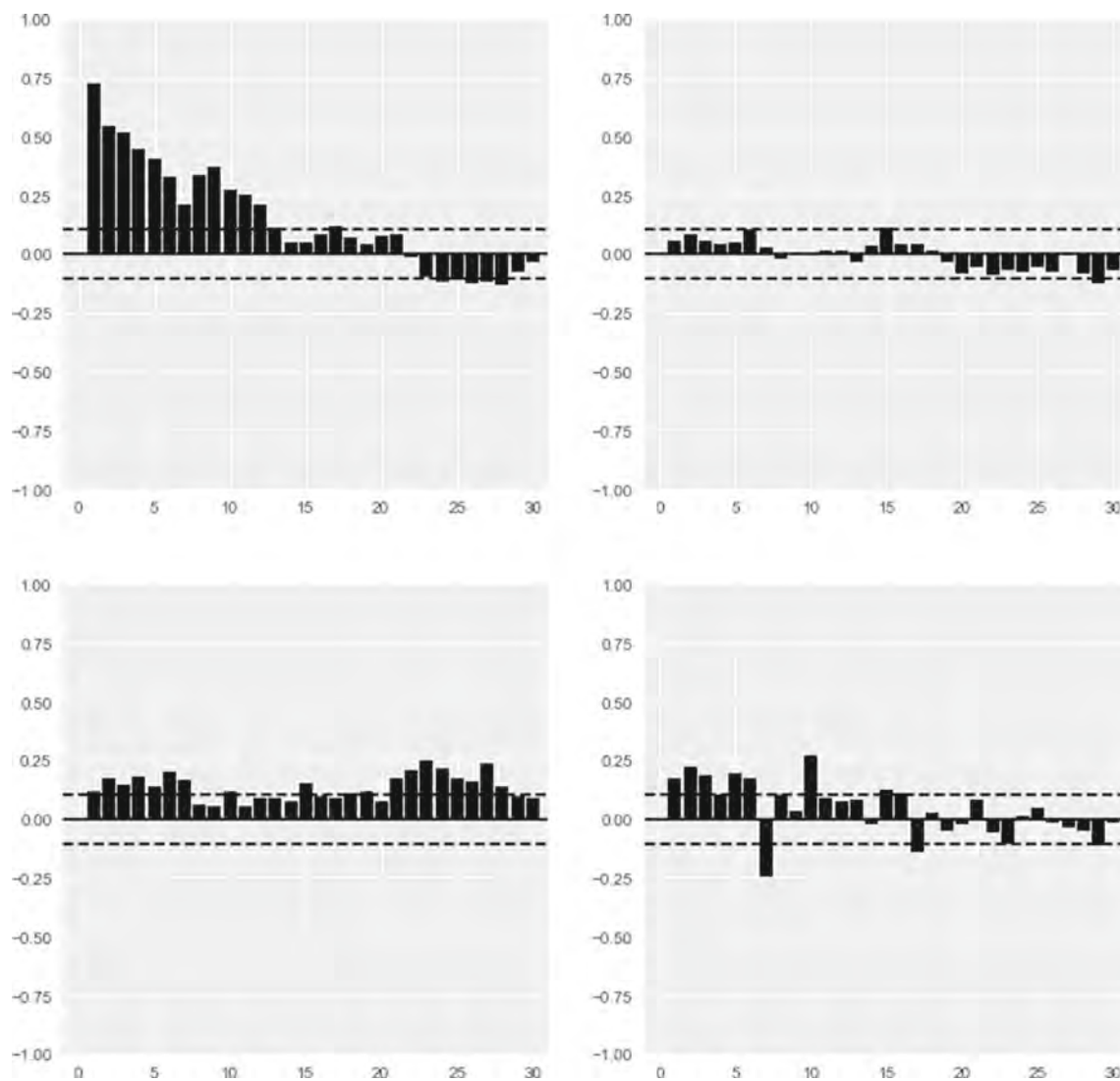
(c) The first difference of the rolling mean series for Kuwait.

**Fig. 3** Taking the first difference of the series helps achieve stationarity

almost perfectly with the actual numbers. Similarly, it can be seen from Fig. 5b, the forecasted numbers of new deaths match closely with the actual numbers.

To make the comparison between the forecasted and actual values more precise, we calculate the **root-mean-squared error (RMSE)**, the mean absolute error (MAE), and the mean absolute percentage error (MAPE) of the model forecast. Note that RMSE and MAE are conditional on the size of the population and should be used with caution for comparison between countries with significantly different numbers of cases. While our model is designed to forecast the number of *new* cases, it can be easily used to forecast the *total* number of cases. To obtain the forecast for the total number of cases,

we simply add the forecast for the new number of cases to the current total number of cases. The results presented in Table 2 show that the proposed model produces highly accurate forecasts. The RMSE, MAE, and MAPE columns in Table 2 correspond to the number of new cases, while the  $MAPE_T$  column corresponds to the number of total cases. Observe that the MAPE relative to the number of new cases is 0.35%, while the MAPE relative to the total number of cases is 0.0017%. The model is less accurate in predicting the number of deaths. The lower accuracy on the number of deaths can be attributed to two primary factors: insufficient data and the complex nature of death. On the other hand, the



**Fig. 4** Correlation plots for the UAE. The upper left and lower right plots represent autocorrelations for the number of cases and deaths, respectively. The off-diagonal plots are cross-correlations. The dashed horizontal lines represent the confidence limits

MAE for new deaths is 0.86 which means that the forecast is accurate within 1 count.

Indeed, the results are substantially better than for many current models in the literature. As shown in Table 3, the proposed model outperforms existing approaches from a range of fields including ARIMA, mathematical modeling, and ML. In particular, the MAPE for predicting the total number of cases using our method is 0.0017% which is substantially smaller than the MAPE values presented in Table 3. Although comparison with Table 3 is not ideal due to the differences in studies, it does provide a useful benchmark for our proposed model.

To further validate the use of the VAR model, we compare its performance against the basic AR model. To this end, we fitted and tested the AR(30) model on the data. We obtain  $MAPE_T$  of 0.0063% which is more than thrice the

$MAPE_T$  for the VAR model. We conclude that the vector-based approach to forecasting COVID-19 infection is more effective than single-valued approach.

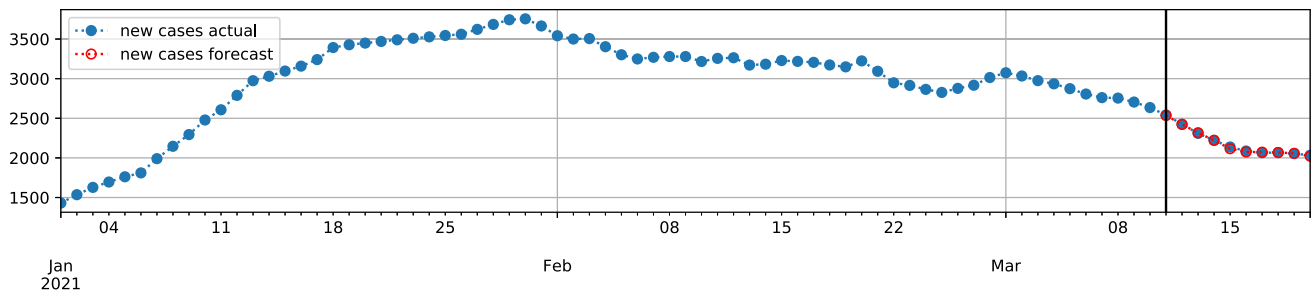
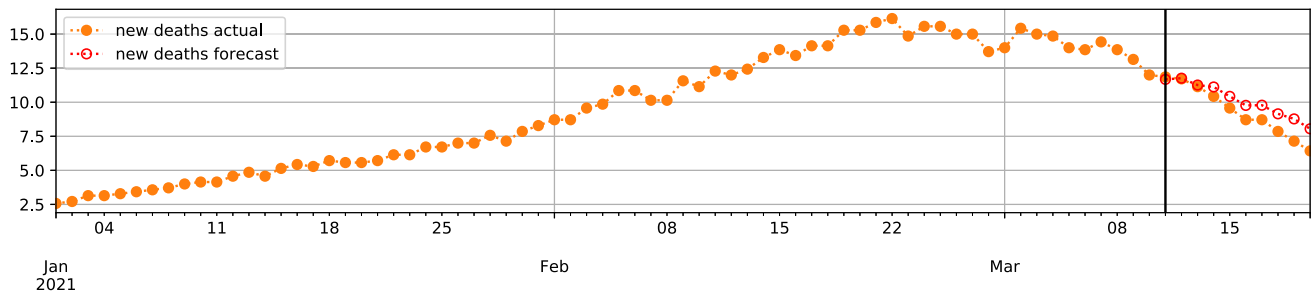
#### 4.4 Saudi Arabia and Kuwait

To demonstrate the effectiveness of the proposed approach, we apply it to the case of Saudi Arabia and Kuwait. To construct the forecasting models for the two countries, we follow the same steps as in the case of the UAE. We consider the time series plots, correlation plots, and AIC. Our analysis yields VAR(29) and VAR(28) as the optimal models for Saudi Arabia and Kuwait, respectively. After fitting the models on the training data, we forecast the number of new cases and deaths for the test period March 11–March 20, 2021. The results are illustrated in Figs. 6 and 7. As shown in Fig. 6, the fore-



**Table 1** AIC values for different orders  $p$  of the VAR model

p	AIC	p	AIC	p	AIC	p	AIC	p	AIC	p	AIC
0	5.693	6	4.785	12	4.337	18	4.106	24	3.990	30	3.968
1	4.911	7	4.624	13	4.292	19	4.113	25	3.990	31	3.998
2	4.877	8	4.436	14	4.198	20	4.088	26	3.986	32	4.008
3	4.824	9	4.442	15	4.081	21	4.019	27	3.999	33	4.017
4	4.841	10	4.310	16	4.089	22	3.977	28	3.997	34	
5	4.826	11	4.325	17	4.102	23	3.978	29	3.951	35	

**(a)** New cases in the UAE.**(b)** New deaths in the UAE.**Fig. 5** The actual and forecast values for the number of new cases and deaths in the UAE**Table 2** 10-day-ahead forecast accuracy for the UAE

	RMSE	MAE	MAPE	MAPE <sub>T</sub>
New cases	9.65	7.45	0.35%	0.0017%
New deaths	1.02	0.86	10.75%	0.0612%

casted number of cases in Saudi Arabia matches very closely with the actual number of cases. Furthermore, the forecasted number of deaths is nearly identical with the actual numbers. Similarly, as shown in Fig. 7, the forecasted values in Kuwait are not too far from the actual values.

As shown in Tables 2 and 4, the proposed forecasting approach achieves a high degree of accuracy in forecasting the number of new cases and deaths. The 10-day-ahead forecast of the number of new cases in the UAE is accurate within 0.35%. The results attained by the proposed VAR model are significantly better than the results in the existing literature (Table 3). The success of the proposed approach lies in the

simplified assumptions that underlie the VAR model together with the high quality of the time series data. Given the performance of the proposed model, it can be similarly applied to other countries and help health officials combat the pandemic.

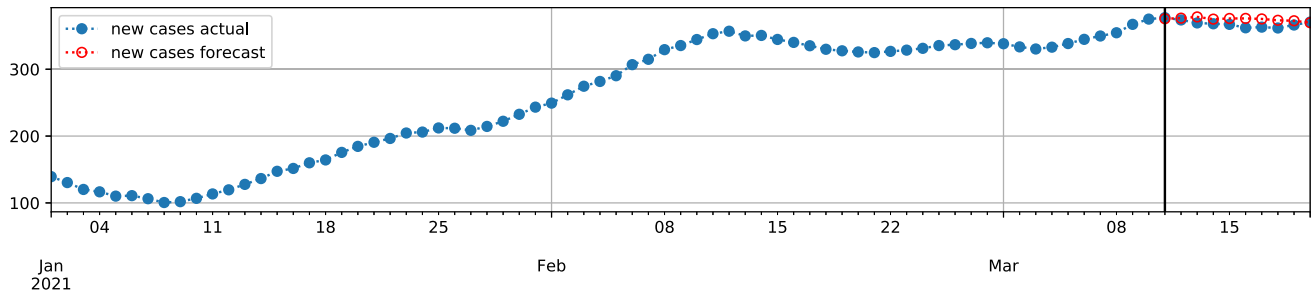
## 5 Conclusion

In this paper, we investigated the use of vector autoregression to forecast the spread of COVID-19 infection. In particular, we applied the VAR model to jointly forecast the number of new cases and deaths the UAE, Saudi Arabia, and Kuwait. The results show a high level of accuracy of the proposed model. Indeed, the MAPE of the model is substantially lower than that of the existing models in the literature. The success of the proposed approach shows that it can also be used for forecasting in other countries.

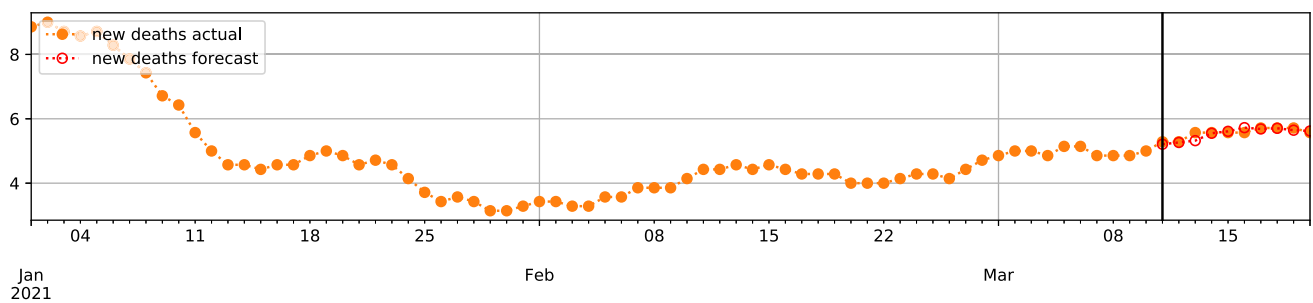


**Table 3** Accuracy results of the existing methods for forecasting the number of cases

Source	Singh et al. [4]	Singh et al. [4]	Curado et al. [8]	Zeroual et al. [15]
Method	ARIMA	ARIMA	Mathematical model	Machine learning
Results	MAPE: 16.01%	MAPE <sub>T</sub> : 0.8–3.7%	MAPE <sub>T</sub> : 3–5%	MAPE <sub>T</sub> : 0.128–5.90%

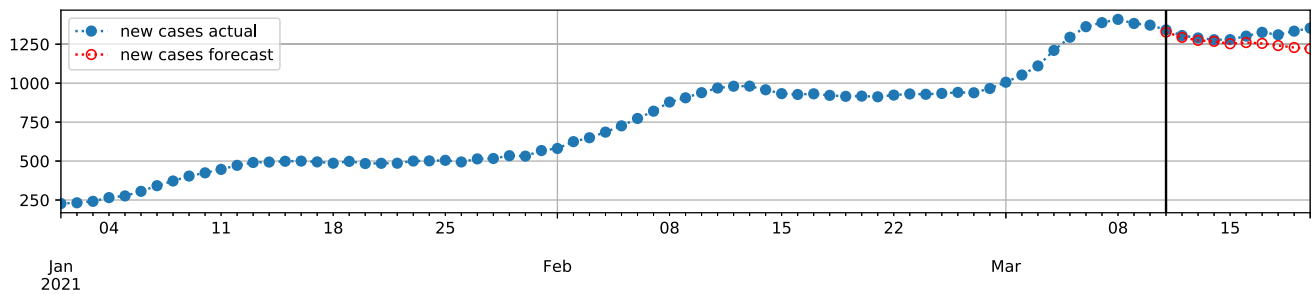


(a) New cases in Saudi Arabia.

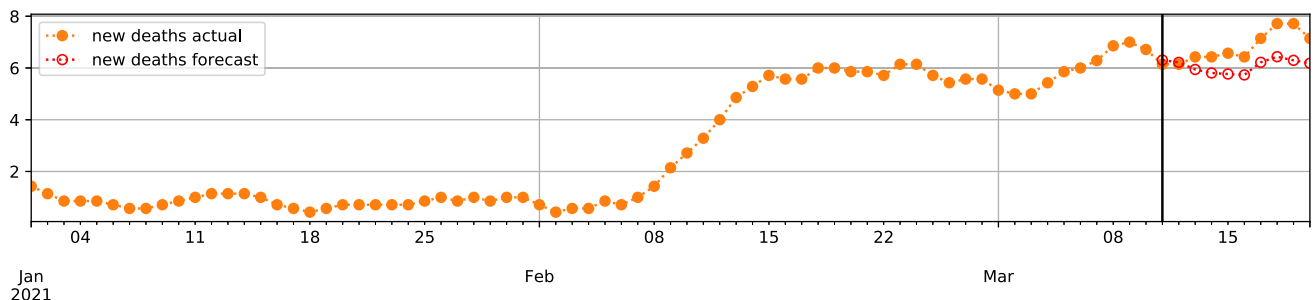


(b) New deaths in Saudi Arabia.

**Fig. 6** The actual and forecast values for the number of new cases and deaths in Saudi Arabia



(a) New cases in Kuwait.



(b) New deaths in Kuwait.

**Fig. 7** The actual and forecast values for the number of new cases and deaths in Kuwait

**Table 4** 10-day-ahead forecast accuracy for Saudi Arabia and Kuwait

	Saudi				Kuwait			
	RMSE	MAE	MAPE	MAPE <sub>T</sub>	RMSE	MAE	MAPE	MAPE <sub>T</sub>
New cases	8.59	7.41	2.03%	0.002%	64.49	49.78	3.75%	0.024%
New deaths	0.10	0.07	1.29%	0.001%	0.85	0.74	10.55%	0.063%

Despite the high accuracy of the model, there remains room for improvement. Although taking the first difference stabilized the mean of the series around zero, additional transformations can be applied to improve the variance. The model may benefit by expanding and including quantities such as the percentage of tested population, the number of recoveries, and others. Alternative VAR models such as LASSO VAR and VARCH may also be investigated in future research. The proposed approach has a great potential to be a valuable tool in managing the government response against COVID-19.

## References

1. Sujath, R.; Chatterjee, J.M.; Hassanien, A.E.: A machine learning forecasting model for COVID-19 pandemic in India. *Stoch. Env. Res. Risk Assess.* **34**, 959–972 (2020)
2. Ioannidis, J.P.; Cripps, S.; Tanner, M.A.: Forecasting for COVID-19 has failed. *Int. J. Forecast.* (2020)
3. COVID-19 tests by country. (2021, March 22). Retrieved April 01, 2021, from <https://www.statista.com/statistics/1028731/covid19-tests-select-countries-worldwide/>
4. Singh, S.; Sundram, B.M.; Rajendran, K.; Law, K.B.; Aris, T.; Ibrahim, H.; Gill, B.S.: Forecasting daily confirmed COVID-19 cases in Malaysia using ARIMA models. *J. Infect. Dev. Ctries.* **14**(09), 971–976 (2020)
5. Sahai, A.K.; Rath, N.; Sood, V.; Singh, M.P.: ARIMA modelling & forecasting of COVID-19 in top five affected countries. *Diabetes Metabol. Syndr. Clin. Res. Rev.* **14**(5), 1419–1427 (2020)
6. Hannan, E.J.; Rissanen, J.: Recursive estimation of mixed autoregressive-moving average order. *Biometrika* **69**(1), 81–94 (1982)
7. Salgotra, R.; Gandomi, M.; Gandomi, A.H.: Time series analysis and forecast of the COVID-19 pandemic in India using genetic programming. *Chaos Solitons Fractals* **138**, 109945 (2020)
8. Curado, E.M.; Curado, M.R.: A discrete-time-evolution model to forecast progress of COVID-19 outbreak. *PLoS ONE* **15**(10), e0241472 (2020)
9. Kamalov, F.: Forecasting significant stock price changes using neural networks. *Neural Comput. Appl.* **32**(23), 17655–17667 (2020)
10. Kamalov, F.; Gurrib, I.; Rajab, K.: Financial forecasting with machine learning: price vs return. *J. Comput. Sci.* **17**(3), 251–264 (2021)
11. Fathi, S.; Srinivasan, R.; Fenner, A.; Fathi, S.: Machine learning applications in urban building energy performance forecasting: a systematic review. *Renew. Sustain. Energy Rev.* **133**, 110287 (2020)
12. Kamalov, F.; Sulieman, H.; Santandreu Calonge, D.: Machine learning based approach to exam cheating detection. *PLoS ONE* **16**(8), e0254340 (2021)
13. Praveen, B.; Talukdar, S.; Mahato, S.; Mondal, J.; Sharma, P.; Islam, A.R.M.T.; Rahman, A.: Analyzing trend and forecasting of rain-fall changes in India using non-parametrical and machine learning approaches. *Sci. Rep.* **10**(1), 1–21 (2020)
14. Kamalov, F.; Cherukuri, A.; Sulieman, H.; Thabtah, F.; Hossain, A.: Machine learning applications for COVID-19: A state-of-the-art review (2021). arXiv preprint [arXiv:2101.07824](https://arxiv.org/abs/2101.07824).
15. Zeroual, A.; Harrou, F.; Dairi, A.; Sun, Y.: Deep learning methods for forecasting COVID-19 time-series data: a comparative study. *Chaos Solitons Fractals* **140**, 110121 (2020)
16. Rustam, F.; Reshi, A.A.; Mehmood, A.; Ullah, S.; On, B.W.; Aslam, W.; Choi, G.S.: COVID-19 future forecasting using supervised machine learning models. *IEEE Access* **8**, 101489–101499 (2020)
17. Wang, P.; Zheng, X.; Li, J.; Zhu, B.: Prediction of epidemic trends in COVID-19 with logistic model and machine learning technics. *Chaos Solitons Fractals* **139**, 110058 (2020)
18. Li, H.; Lu, Y.: Coherent forecasting of mortality rates: A sparse vector-autoregression approach. *ASTIN Bull. J. IAA* **47**(2), 563–600 (2017)
19. Cavalcante, L.; Bessa, R.J.; Reis, M.; Browell, J.: LASSO vector autoregression structures for very short-term wind power forecasting. *Wind Energy* **20**(4), 657–675 (2017)
20. Messner, J.W.; Pinson, P.: Online adaptive lasso estimation in vector autoregressive models for high dimensional wind power forecasting. *Int. J. Forecast.* **35**(4), 1485–1498 (2019)
21. Antonakakis, N.; Gabauer, D.; Gupta, R.: International monetary policy spillovers: evidence from a time-varying parameter vector autoregression. *Int. Rev. Financ. Anal.* **65**, 101382 (2019)
22. Liu, Y.Y.; Tseng, F.M.; Tseng, Y.H.: Big data analytics for forecasting tourism destination arrivals with the applied vector autoregression model. *Technol. Forecast. Soc. Chang.* **130**, 123–134 (2018)
23. Vu, T.K.; Nakata, H.: Oil price fluctuations and the small open economies of Southeast Asia: an analysis using vector autoregression with block exogeneity. *J. Asian Econ.* **54**, 1–21 (2018)
24. Monllor, P.; Su, Z.; Gabrielli, L.; Taltavull de La Paz, P.: COVID-19 infection process in Italy and Spain: are data talking? Evidence From ARMA and vector autoregression models. *Front. Public Health* **8**, 784 (2020)
25. Hamilton, J.D.: *Time Series Analysis*. Princeton University Press, Princeton (2020)
26. Seabold, S.; Perktold, J.: Statsmodels: econometric and statistical modeling with python. In: *Proceedings of the 9th Python in Science Conference*, vol. 57, p. 61 (2010).
27. Coronavirus source data. (n.d.). Retrieved April 01, 2021, from <https://ourworldindata.org/coronavirus-source-data>

