



Forecasting Covid-19: SARMA-ARCH approach

Firuz Kamalov¹ · Fadi Thabtah²

Received: 25 April 2021 / Accepted: 3 August 2021 / Published online: 18 August 2021
© IUPESM and Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

Forecasting the number of Covid-19 cases is a crucial tool in public health policy. In this paper, we construct seasonal autoregressive moving average and autoregressive conditional heteroscedasticity models to forecast the spread of the infection in the UAE. While most of the existing literature is dedicated to forecasting the total number of infections, we endeavor to forecast the number of *new* infections which is a significantly more challenging task due to the greater volatility. Our models are based on a careful analysis of correlation plots and residual analysis. In addition, we employ highly accurate population data that leads to more reliable outcomes. The results reveal a high degree of accuracy of the proposed forecasting methods. The constructed models can be used by health officials to better anticipate and plan for new cases of Covid-19.

Keywords Covid-19 · SARMA · AR-ARCH · Forecasting · Autoregression

1 Introduction

The Covid-19 pandemic has affected the social and economic lives of people around the globe. It has created an enormous public health challenge for government agencies. One of the key tools in the fight against the virus is communal quarantine. In order to choose the right time to implement quarantine measures it is important to have accurate forecast of the spread of the virus. Advance knowledge of the number of new cases allows government officials to decide if a lockdown is necessary. In this paper, we consider the task of forecasting the number of new Covid-19 cases in the UAE. The UAE has one of the highest testing rates in the world making its population data extremely reliable. We propose seasonal autoregressive moving average (SARMA) and autoregressive conditional heteroscedasticity (ARCH) models that can predict with relative accuracy the number of new cases 10 days ahead.

The ability to accurately forecast the number of new Covid-19 cases has several benefits. It allows public health agencies to better plan for future new cases. A forecast of a rapid increase in the number of new cases would give government officials time to take the necessary steps to prevent the rapid spread of the virus. Conversely, a forecast of a decrease in the number of new cases would allow officials avoid implementing unnecessarily harsh restrictions. In general, accurate forecasts would help officials make more informed decisions. Thus, it is imperative to develop tools that provide accurate forecasts of the infection rates.

There exists a number of approaches to time series forecasting. The traditional approach is based on the autoregressive moving average (ARMA) model. The ARMA model is built on the assumption that the value of a time series depends linearly on the previous values of the series together with past shocks. The ARMA model is widely used in finance, meteorology, astronomy, and many other fields. Although the ARMA model provides a well-understood and solid theoretical framework its simple structure does not always fit in practice. Consequently, a number of extensions of the ARMA model were developed. Two of the most popular extensions of the ARMA model are the SARMA and ARCH models. In our study, we apply both SARMA and ARCH to build forecasting models for the number of new Covid-19 cases. The results reveal that the two approaches are equally effective in forecasting the number of new cases.

This article is part of the COVID-19 Health Technology: Design, Regulation, Management, Assessment

✉ Firuz Kamalov
firuz@tud.ac.ae
Fadi Thabtah
fadi.fayez@manukau.ac.nz

¹ Department of Electrical Engineering, Canadian University
Dubai, Dubai, UAE

² School of Digital Technologies, Manukau Institute
of Technology, Auckland, New Zealand

The SARMA model takes into account seasonal variations in time series. It is used to model sales, temperature, internet traffic, and other periodic time series. Since Covid-19 is a flu virus there is a possibility that it possesses seasonal characteristics. Therefore, the SARMA model would be a suitable candidate to forecast the spread of the virus. In this paper, we carefully analyze the residual plots to construct an appropriate SARMA model that achieves robust out of sample accuracy.

The ARCH model is designed to reflect a nonconstant conditional variance in time series. It can be used in time series where the variance at time t depends on the value of the series at time $t - 1$. It is often used to model time series with large shocks such as stock price or macroeconomic indicators. There are two main reasons why the ARCH model would be appropriate in the study of Covid-19. First, since the virus can spread with geometric progression if unimpeded the variance in the number of daily cases would be strongly related to the number of cases in the previous days. Second, governmental measures such as lockdown, quarantine, and others can have a sudden and dramatic impact on the variance in the number of daily cases. The ARCH model can be used in order to take into account these factors.

A number of different approaches to forecasting Covid-19 have been proposed in the literature. Mathematical models such as the exponential growth, self-exciting branching process, and the susceptible–infected–resistant (SIR) compartment models have been used for their simplicity. These models have been applied in different countries around the world albeit with mixed results [1–3]. Machine learning techniques have also been a popular tool in forecasting Covid-19. Given the temporal nature of the Covid-19 data, the most commonly used approaches have been based on long short-term memory and convolutional neural networks [4, 5].

One of the main drawbacks of the existing studies in forecasting the spread of Covid-19 is the use of low quality data. The poor quality of data is manifested in two ways. First, the data encompasses a short time frame sometimes as few as 2 weeks [6, 7]. In addition, the data is often taken from the early 2020 and does not match well with the existing situation. The number of cases has increased dramatically since the start of the pandemic so the models developed based on the early data do not apply well in current environment. Second, the data is obtained from countries with low Covid-19 testing rate. The data derived from low test rate countries is unreliable as it does not reflect the full spread of the virus. Models based on low test rate data cannot be used to forecast the true spread of the virus.

In our paper, we focus on forecasting the number of new Covid-19 cases in the UAE. The UAE was chosen due to the reliability of the country Covid-19 data. The UAE has one of the highest Covid-19 testing rates in the world. Indeed,

on Aug 5, 2020 the UAE became the world's first country with populations over one million to hit a 50% testing rate for COVID-19. As a result, the reported data is highly reflective of the true number of cases in the country. Although our models are developed specifically for the UAE, their success shows that similar approaches can be applied to other countries. In addition, we use data from Apr 1, 2020 to Dec 1, 2020 obtained from European Centre for Disease Prevention and Control (ECDC). The time interval is chosen to match the infection rates closer to the full scale pandemic. All the calculations including the SARMA and ARCH models were done in Python. The SARMA and ARCH models were implemented via the statsmodel [8] and arch [9] libraries respectively.

Unlike most of the similar studies in the literature, we attempt to forecast the number of new daily cases instead of the cumulative number of cases. Due to the differences in magnitude, the forecasts of the total number of cases appear more accurate than that of the number of new cases. Our paper makes several contributions to the existing literature as listed below:

1. The use of SARMA model justified by the seasonal patterns of the flu virus. The model is based on the analysis of correlation and residual plots.
2. The use of ARCH model justified by the conditional variance of daily number of cases. The model is carefully built using first order residual correlations.
3. The use of high quality data based on a thoroughly tested population. In addition, the time frame used in the study is chosen to represent the full scale of the pandemic.

The paper is structured as follows. In Section 2, we provide a brief review of the existing literature related to forecasting Covid-19 cases. In Section 3, we present our forecasting models based on SARMA and ARCH processes. We conclude the paper with closing remarks in Section 4.

2 Literature review

There have been a number of attempts to forecast the spread of Covid-19 both on the international [10] and country specific [7, 11, 12] levels. The existing forecasting methods include a range of techniques - both traditional and nontraditional approaches. The results have been mixed. While some authors argue that forecasting Covid-19 has failed [13], others have shown promising results [14]. In [10], the authors propose a simple iteration method based on the average growth rate over the previous m timesteps. The authors apply their method to a several countries to forecast the number of new cases. The study shows that it is necessary to keep the grow rate below 5% to curve the spread of the virus. One of

the early studies done by [6] focuses on the data from Hubei province of China. The authors use a calibrated SIRD model to fit the reported data and produce forecasts. A multi-part forecasting system was proposed by [12]. The authors propose a model based 6 parameters including recovered and infected rates. The model is applied to Covid-19 infection in India and a number of recommendations about curbing the spread of the virus were made. The authors in [15] produce forecasts using models from the exponential smoothing family. The models were constructed continuously over six week and applied to several countries. The results were mixed with forecast at times over-estimating and under-estimating the number of cases.

An alternative approach to the classical ARMA-based forecasting has been pursued by the authors using machine learning (ML) algorithms. A comparative study of ML-based algorithms for Covid-19 forecasting was done in [16]. The authors analyzed a number of evolutionary algorithms such as Genetic Algorithm, Particle Swarm Optimization, and Gray Wolf Optimizer as well as ML algorithms such Multilayer Perceptron (MLP) and adaptive network-based fuzzy inference system (ANFIS). The models were evaluated on the basis of their accuracy for different prediction lead times. The authors employed data from 5 different countries in their study experiments revealed that MLP and ANFIS algorithms produce the best results. The authors in [17] test 3 long short term memory (LSTM) based models to forecast the number of infected individuals for 32 states in India. The tested models include stacked, convolutional, and bi-directional LSTM neural networks. The predictions are made one day and one week ahead. The results show that the bi-directional LSTM produces the optimal results.

One the main issues with many of the existing forecasting models is the quality of data used to build the models. There are two primary concerns with the data. First, the data covers a small time frame. Several studies use data from as little

as 2-3 weeks. In [18], the authors use data from Jan 1 - Feb 18, 2020 while in [6] from Jan 16 - Feb 10, 2020, and in [7] from Jan 20 – Feb 11, 2020. In general, the majority of studies do not use data beyond the end of spring 2020 [17, 19, 20]. In addition, the majority of studies use data from the early stages of the pandemic. Since the number of cases has increased exponentially, the models constructed based on the early data do not apply to the current situation. Second, the data is based on sparsely tested populations. The two most used datasets are based on daily number of cases in China and India. The testing rates in both countries were relatively low during the time frame used in the studies. For instance, by Jun 29, 2020 China had tested only 6.5% of its population and India 0.65% [21]. A low testing rate does not reflect the true spread of infection and lead to biased forecasting models. In comparison, the UAE had tested over half of its population by Aug 4, 2020 [22].

3 Forecasting models

3.1 Data preprocessing

In our study, we employ data for the number of new daily Covid-19 cases in the UAE. The dataset is available for download from the ECDC website <https://opendata.ecdc.europa.eu/covid19>. Although the data is available as far back as Jan 1, 2020 the number of new recorded cases in the UAE is close to zero before Apr 1, 2020. Given the substantial increase in the number of cases in the following months it would be misleading to include the data prior to April 1 in our analysis. Therefore, in order to work on relatively uniform data we focus on data post Apr 1, 2020.

The graph in Fig. 1 shows a rapid increase in the number of new cases during the period of Apr 1 - May 21. The growth can be attributed to two primary causes: communal

Fig. 1 The number of daily new cases of Covid-19 together with a fitted degree 4 polynomial regression

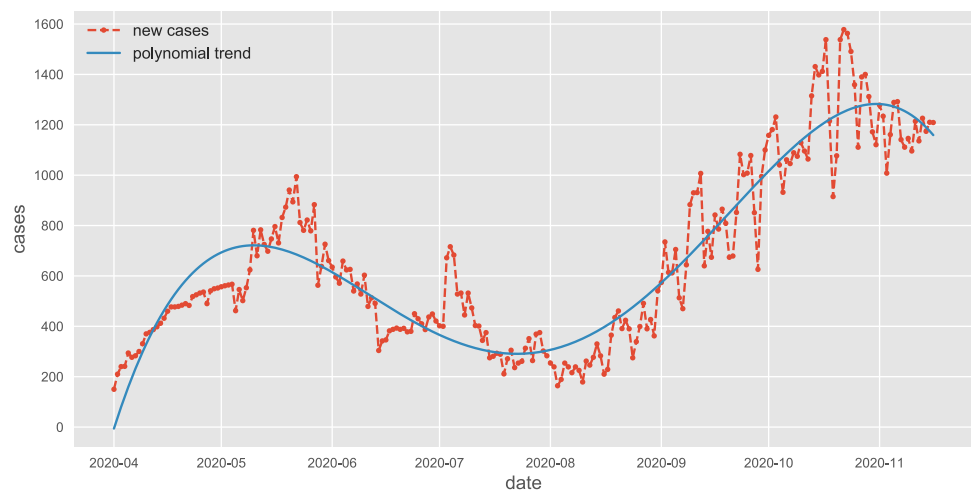
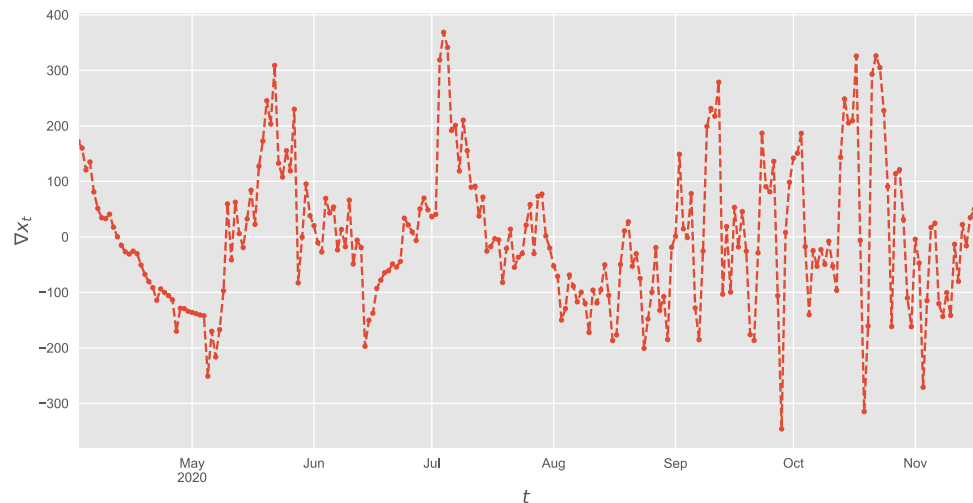


Fig. 2 The detrended series of the number of daily new cases obtained by subtracting the regressed polynomial from the original series



spread of the virus and increase in the number of tests. After achieving a relative peak during the third week of May the number of new cases starts to decline significantly until the beginning of August. The decrease in the number of cases is explained by the quarantine measures together with the summer months that kept people from outdoor gatherings. However, the number of new cases started to increase again from the beginning of August. The softening of full lockdown together with the opening of schools contributed to the rise in cases. The number of new cases continues to grow through November, considerably surpassing the relative peak of May.

As shown in Fig. 1, the time series is clearly not stationary. The mean of the series is not constant. It is steadily increasing from the start of August. The variance of the series is also not constant. The series is growing more volatile with time. The series can be smoothed by either applying the differencing operation or detrending the series by subtracting the estimated trend. We choose the latter approach

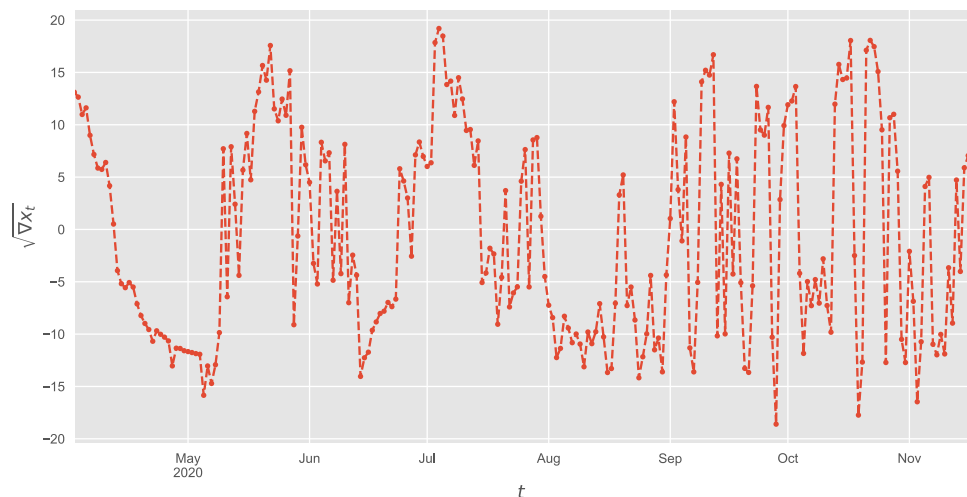
to detrend the series. The graph of the time series shows 3 critical points which indicates that a degree 4 polynomial would fit the series. As can be seen from the figure, the regressed polynomial captures the trend and fits naturally to the time series.

We improve the stationarity properties of the series by subtracting the regressed polynomial from the original series. Let x_t be the original series and \hat{x}_t be the regressed polynomial. We define the detrended series by

$$\nabla x_t = x_t - \hat{x}_t. \quad (1)$$

As can be seen from Fig. 2, the detrended series has a constant mean around zero. The variance is not constant and further transformations such as second difference, square root, log, and others can be applied to smooth the series. However, additional transformations do not result in a significant improvement in stationarity of the series. For instance, as shown in Fig. 3, the square root transformation of the detrended series achieves only a slightly better variance.

Fig. 3 The square root transformation of the detrended series



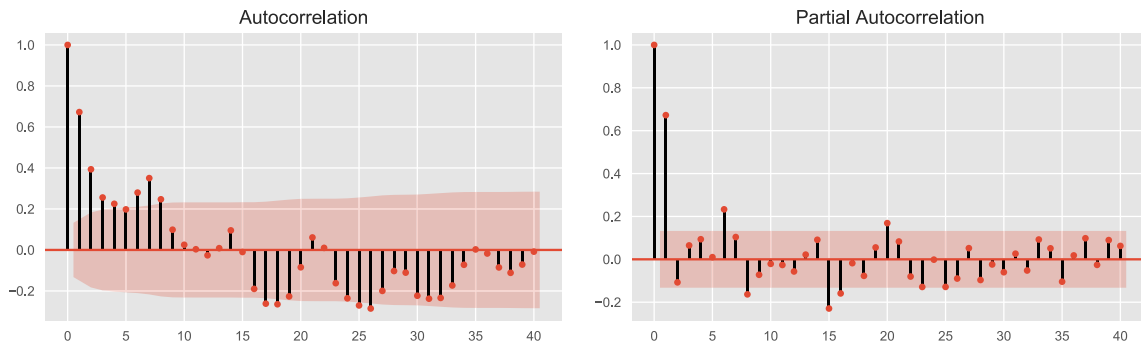


Fig. 4 The ACF/PACF plots of the detrended series ∇x_t

Therefore, we will employ the regression detrended series (Fig. 2) through the rest of the paper.

3.2 SARMA model

The SARMA model allows us to account for the previous values of the time series as well as seasonal patterns. The SARMA model of order $(p, q) \times (P, Q)_s$ for a time series x_t is given by the equation

$$\Phi(B^s)\phi(B)x_t = \Theta(B^s)\theta(B)w_t, \tag{2}$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are seasonal and regular AR polynomials, $\Theta(\cdot)$ and $\theta(\cdot)$ are seasonal and regular MA polynomials, and B is the backshift operator. To determine the order of SARMA model the autocorrelation and partial autocorrelation function (ACF/PACF) plots can be considered. The ACF/PACF plots are presented in Fig. 4. The ACF plot of

the detrended series ∇x_t shows a steady decay of values starting with $h = 1$ which indicates an AR(1) process. There is a spike at lag $h = 7$. Since the spike is singular it can indicate seasonal MA(1) process with $s = 7$. The smaller spikes at $h = 6$ and $h = 8$ support the theory of seasonal MA process with $s = 7$. The PACF plot has a single large spike at $h = 1$ which supports our earlier hypothesis that there is AR(1) process. The spikes at $h = 6$ and $h = 8$ are potentially connected to the corresponding spikes in the ACF plot. Given the above analysis we conclude that SARMA(1, 0) \times (0, 1)_s model fits the detrended series ∇x_t .

To confirm the goodness of fit of SARMA(1, 0) \times (0, 1)_s we consider the residuals of the fitted model. The diagnostic plots of the residuals are presented in Fig. 5. The histogram of the residuals shows an approximately normal distribution. The Normal Q-Q plot supports the evidence from the histogram albeit with some deviations at the tail quantiles. The residuals of the fitted model are approximately

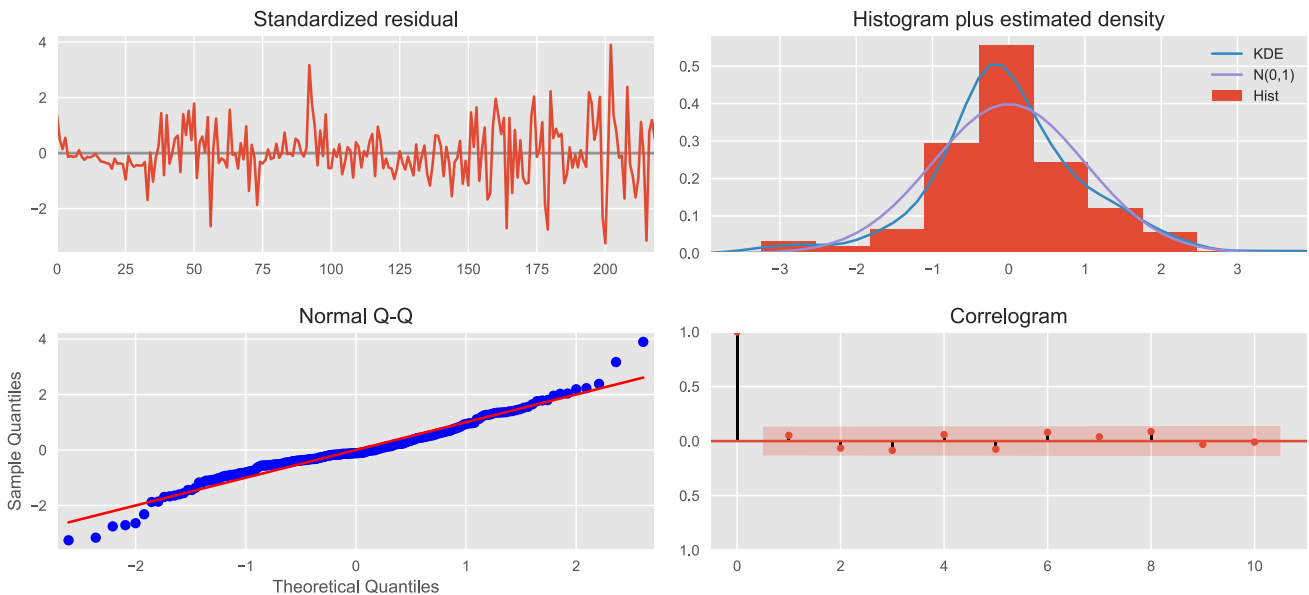


Fig. 5 The residual plots of the fitted SARMA(1, 0) \times (0, 1)_s model

normally distributed with mean zero. The correlogram shows absence of correlations between the lagged residuals. The correlogram has no statistically significant values. The analysis of the residuals leads us to conclude that the SARMA(1, 0) \times (0, 1)_s model fits the series.

As shown in Fig. 6, the SARMA model fits well to the training data. The fitted series is consistent with the original detrended series. It correctly predicts the direction of change. The model underestimates large swings in the value of the series. On the other hand, it produces accurate predictions for smaller values of the series. Since the series is stochastic a perfect prediction is never possible. Figure 6 also contains the plot of forecasted values of the series from day 220 to day 229, where day 1 corresponds to Apr 1, 2020. The forecasted values are, as expected, close to the mean. At the beginning of the forecast period the predicted values significantly overestimate the true values of the series. Interestingly, the forecast values become more accurate during the last five days. In general, the forecast values are relatively close to the actual values of the series.

We employ Mean Absolute Error (MAE) to measure the performance of the proposed model. The MAE is given by the following equation

$$\frac{\sum |x_i - \hat{x}_i|}{n},$$

where x_i is the true value, \hat{x}_i is the predicted value, and n is the number of observations. The MAE for fitted values and forecasted values are 66.83 and 67.96 respectively. The model MAE on the train set is close to the MAE on the test set which indicates that the model did not overfit. We can benchmark the performance of the model against the naive

approach where all the values are forecasted to be the same as on the last day before the beginning of the forecast period. The MAE for the naive approach is 82.62 which is significantly greater than the model test MAE. We conclude that the model achieves nontrivial results.

Finally, we use the detrended series to reconstruct the original series for the number of daily new cases in the UAE. The forecasted values of the original series are given by the equation

$$\hat{x}_t = x_{t-1} - \nabla \hat{x}_t,$$

where $\nabla \hat{x}_t$ is the forecasted value of the detrended series. As shown in Fig. 7, the fitted values of the series are consistent with the original values of the series. The model correctly predicts the direction of change in almost all cases. The accuracy of the prediction is encouraging based on a visual comparison of the forecasted and actual values of the time series in Fig. 7. The model underestimates some of the extreme values of the series but it performs well on moderate values. The model also performs well on the forecasted values. As can be seen from Fig. 7, the forecasted and actual values of the series are very close. Indeed, the accuracy of the forecasts is maintained even at the end of the period.

The MAEs for the fitted and forecasted values of x_t are 116.7 and 63.1 respectively. To put it in perspective, the average number of new cases during the forecast period was 1174.6/day. In other words, the model predicts the number of new cases with average accuracy of 5.4% during the forecast period. The model performance during the forecast period can be viewed as performance on unseen data. It is indicative of the model performance in the future. Thus, we can conclude that the 10-day ahead accuracy of the model is around 5.4% for future predictions.

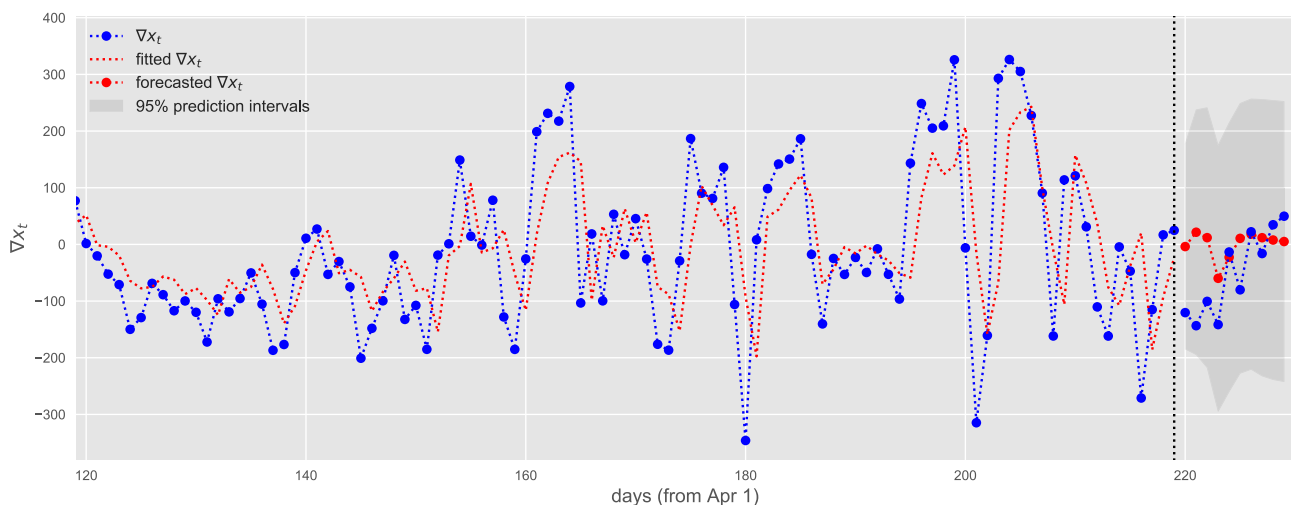


Fig. 6 The original and fitted detrended series

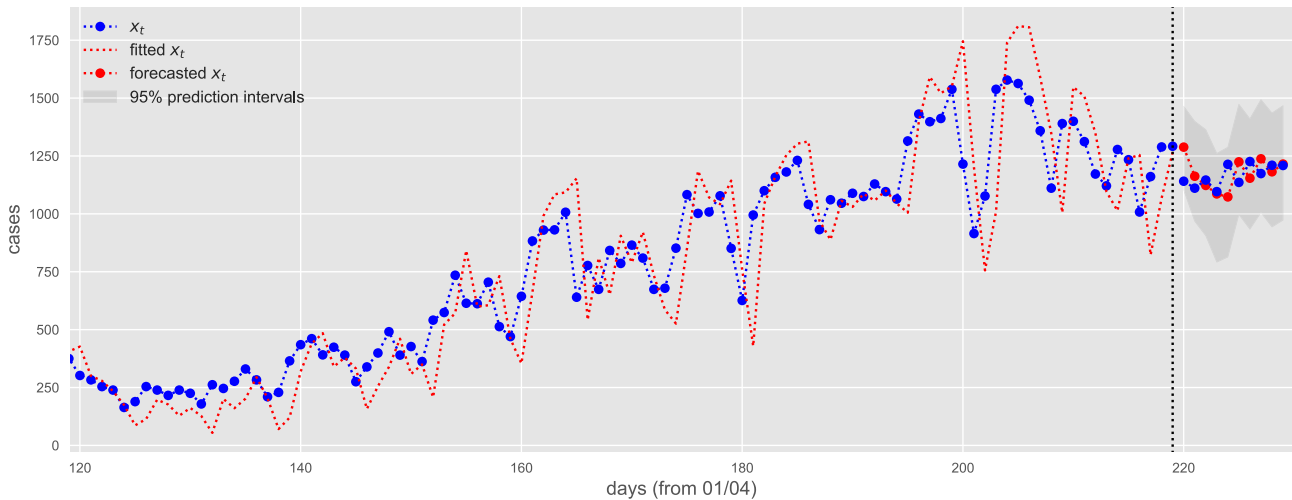


Fig. 7 The number of daily new Covid-19 cases: the original and fitted series

3.3 ARCH model

The ARCH model is traditionally employed to account for unexpected swings in the variance of a time series. It is a sensible model to assume in the case of daily Covid-19 infections. For instance, a quarantine or a lockdown can have a significant sudden impact on the number of new cases. Thus, the ARCH model is appropriate for the use in forecasting the number of new Covid-19 cases. The AR-ARCH model is given by the following equations

$$\begin{aligned}
 \phi(B)x_t &= r_t \\
 r_t &= \sigma_t \varepsilon_t, \text{ where } \varepsilon_t \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1) \\
 \sigma_t^2 &= \alpha_0 + \sum_{j=1}^p \alpha_j r_{t-j}^2
 \end{aligned}
 \tag{3}$$

The PACF plot of the detrended series in Fig. 4 shows several nonzero values. Some of the nonzero PACF values occur at such high lags that can be reasonably dismissed. Therefore, we only consider PACF values for lags $h \leq 10$. There

are three such nonzero PACF values at lags $h = 1, 6, 8$. We use the highest of the nonzero lags, i.e. $h = 8$, to construct the initial AR model to fit the series.

We study the residual plots of the fitted AR(8) model to determine the goodness of fit. As shown in Fig. 8, the ACF/PACF plots have only a few nontrivial values. The values do not appear high enough to be truly significant. Therefore, we conclude that the AR(8) model fits relatively well to the detrended series. Since the original series contained an upward trend it is possible that the series volatility may be dependent on the level of the series. To test this hypothesis we analyze the correlations of squared residuals.

As shown in Fig. 9, the plot of squared residuals contains a nontrivial PACF value at lag $h = 1$. It indicates a possible impact of the level of the series on the series volatility. Since the original series showed increased volatility at higher levels, it is reasonable to assume the presence of ARCH(1) process. Although the PACF plot also contains nontrivial values at other lags, we only use $h = 1$ to keep the model simple and

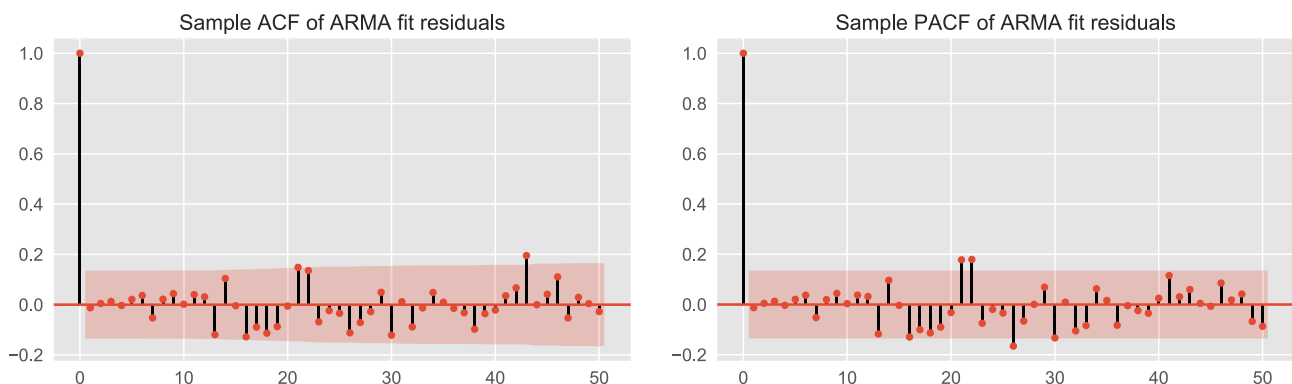


Fig. 8 The ACF/PACF plots of the residuals of the fitted AR(8) model

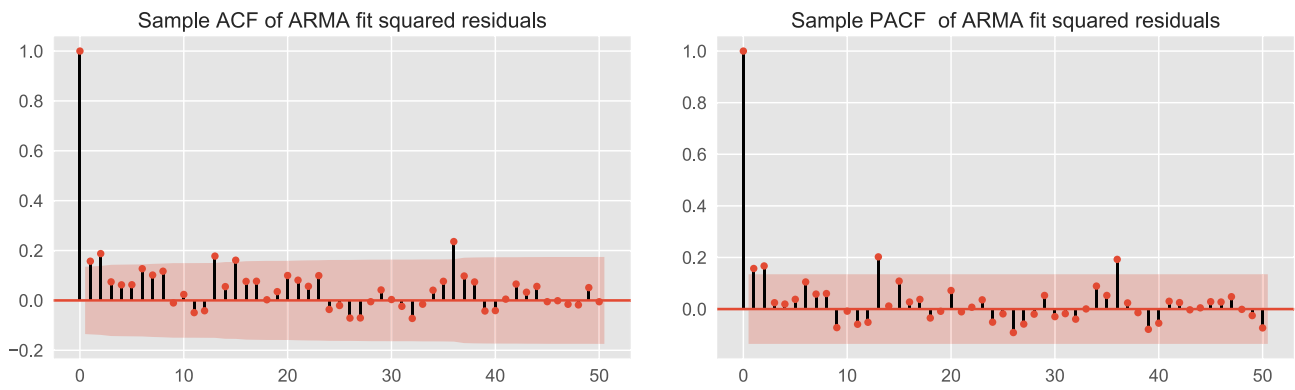


Fig. 9 The ACF/PACF plots of the squared residuals of the fitted AR(8) model

efficient. Thus, we conclude that AR(8)-ARCH(1) model will be a good fit to the detrended series.

As can be seen from Fig. 10, the fitted and actual values of detrended series match very closely to one another. The fitted values are consistent with the actual values. The fitted values correctly predict the direction of the actual values. In case of large values, the fitted values underestimate the actual values. However, in most cases the fitted values are close to the actual values. The predicted values during the forecast period (test set) are also close to the actual values. The MAE for forecasted values is 63.1 which is lower than the test MAE of the SARMA model. For comparison, recall that the MAE for the naive strategy is 82.62 which shows that the proposed model fits the series well.

Finally, the estimates of the detrended series obtained above can be used to construct the forecasted values of the original series. The fitted and forecasted values for the original series are presented in Fig. 11. As can be seen from the figure, the fitted values closely approximate the actual values

of the series. Although the model underestimates the original series at large values it is generally consistent with the actual values. Since the series contains a stochastic component a perfect prediction would be impossible. Similarly, the forecast values accurately predict the actual values. Remarkably, the forecasted values remain accurate even at the end of the forecast period. The MAE of the forecasted values is 64.84 which is equivalent to 5.6% of the mean number of daily new cases during the same period. We also note that the actual values of the series are well within the confidence interval bounds.

The proposed models can be used by the public health officials in the UAE and other countries with similar characteristics to forecast the spread of Covid-19. Each model incorporates a specific forecasting philosophy and can be applied as needed. The SARMA model is appropriate in case when the spread of Covid-19 is modeled as a seasonal infection similar to flu, while the ARCH model is appropriate for modeling time series shocks such as lockdowns. As shown

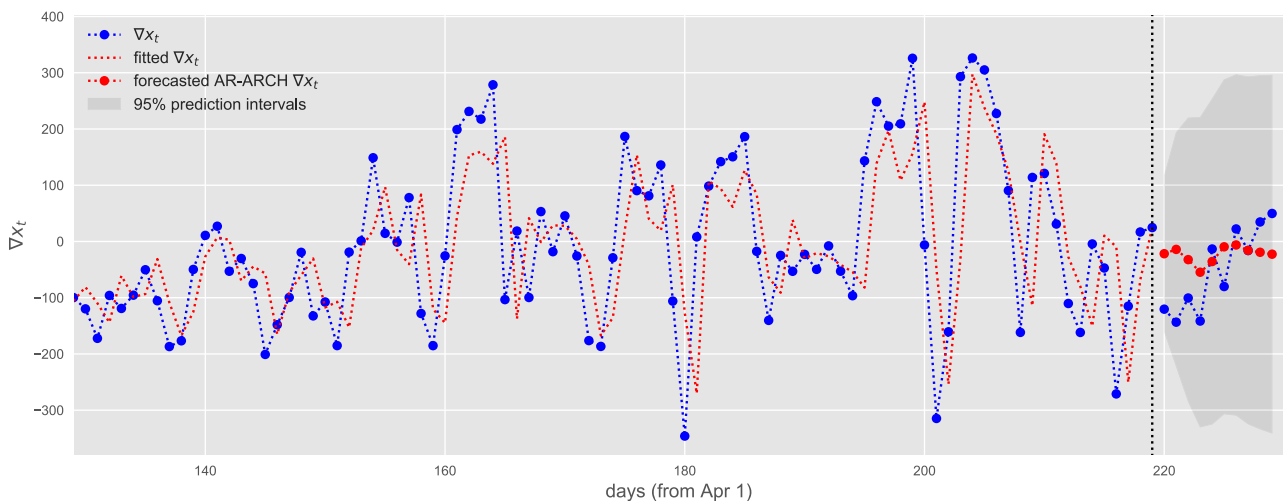


Fig. 10 The AR-ARCH fitted and actual values of detrended series

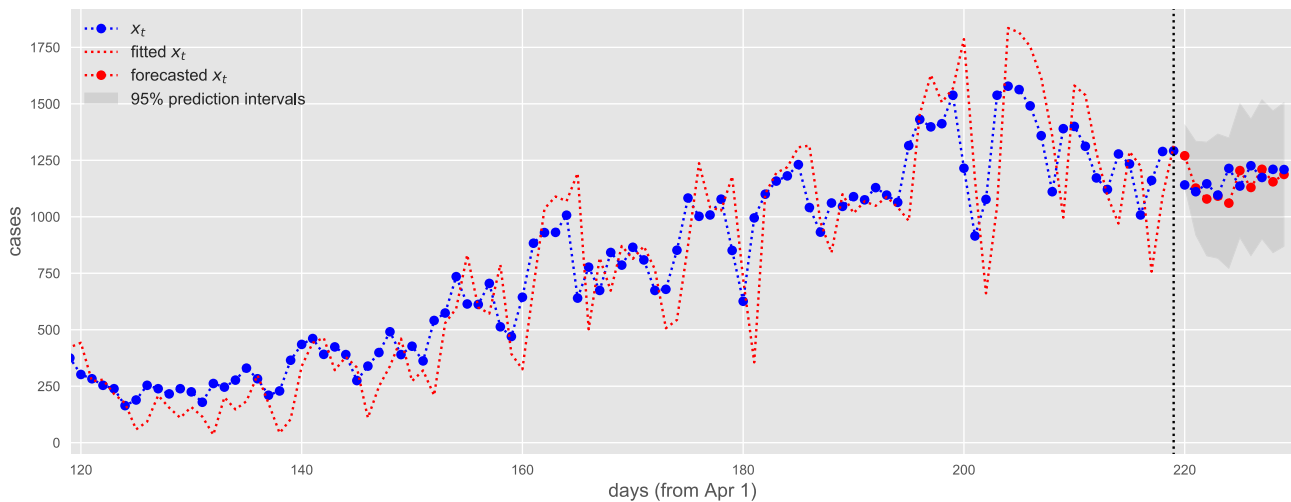


Fig. 11 The number of daily new cases: the actual and AR-ARCH fitted values of the series

above, the parameters of the SARMA and ARCH models can be estimated based on the historic data and the correlation plots. The models can be subsequently deployed to forecast the number of new cases several days or weeks ahead. Accurate forecasts provided by these models would help officials make more informed decisions and help manage the pandemic.

4 Conclusion

Forecasting the number of daily new cases is an important tool in tackling the Covid-19 pandemic. It allows public health officials to plan for the future and make more informed policy decisions. Therefore, it is vital to develop effective forecasting methods. In this paper, we proposed two approaches to forecast the number of new Covid-19 cases: SARMA and AR-ARCH. Both methods produced accurate forecasting results as demonstrated in the above numerical experiments. The mean absolute percentage errors for SARMA and ARCH are 5.4% and 5.6% respectively indicating that the proposed models are a good fit to the times series. Although our study is based on the UAE country data, it can be extended to other countries. We hope that the constructed models can be used by health officials to better anticipate and plan for the new cases of Covid-19.

Funding The research is not funded by any Agency.

Data availability Not applicable.

Code availability Not Applicable.

Declarations

Conflicts of interest The authors declare there is no conflict of interest.

References

- Bertozi AL, Franco E, Mohler G, Short MB, Sledge D. The challenges of modeling and forecasting the spread of COVID-19. *Proceedings of the National Academy of Sciences*. 2020;117(29):16732–8.
- Ghostine R, Gharamti M, Hassrouny S, Hoteit I. An extended SEIR model with vaccination for forecasting the Covid-19 pandemic in Saudi Arabia using an ensemble Kalman filter. *Mathematics*. 2021;9(6):636.
- Moein S, Nickaeen N, Roointan A, Borhani N, Heidary Z, Javanmard SH, Gheisari Y. Inefficiency of SIR models in forecasting COVID-19 epidemic: a case study of Isfahan. *Sci Rep*. 2021;11(1):1–9.
- La Gatta V, Moscato V, Postiglione M, Sperli G. An epidemiological neural network exploiting dynamic graph structured data applied to the covid-19 outbreak. *IEEE Transactions on Big Data*. 2020;7(1):45–55.
- Mohimont L, Chemchem A, Alin F, Krajecki M, Steffene LA. Convolutional neural networks and temporal CNNs for COVID-19 forecasting in France. *Applied Intelligence*. 2021;1–26.
- Anastassopoulou C, Russo L, Tsakris A, Siettos C. Data-based analysis, modelling and forecasting of the COVID-19 outbreak. *PLoS One*. 2020;15(3):e0230405.
- Li Q, Feng W, Quan YH. Trend and forecasting of the COVID-19 outbreak in China. *J Infect*. 2020;80(4):469–96.
- Seabold S, Perktold J. Econometric and statistical modeling with Python skipper seabold 1 1. In *Proc 9th Python Sci Conf*. 2010;57:61.
- Sheppard K. *bashtage/arch: Release 4.18 (Version v4.18)*. Zenodo. 2021;593254. <https://doi.org/10.5281/zenodo>.
- Perc M, Goriek Miksi N, Slavinec M, Stoer A. Forecasting Covid-19. *Front Phys*. 2020;8:127.
- Bastos SB, Cajueiro DO. Modeling and forecasting the Covid-19 pandemic in Brazil. *arXiv preprint*. 2020. arXiv:2003.14288.

12. Sarkar K, Khajanchi S, Nieto JJ. Modeling and forecasting the COVID-19 pandemic in India. *Chaos, Solitons Fractals*. 2020;139:110049
13. Ioannidis JP, Cripps S, Tanner MA. Forecasting for COVID-19 has failed. *Int J Forecast*. 2020.
14. Kapoor A, Ben X, Liu L, Perozzi B, Barnes M, Blais M, O'Banion S. Examining COVID-19 Forecasting using Spatio-Temporal Graph Neural Networks. *arXiv preprint*. 2020. arXiv:2007.03113
15. Petropoulos F, Makridakis S. Forecasting the novel coronavirus COVID-19. *PLoS One*. 2020;15(3):e0231236
16. Ardabili SF, Mosavi A, Ghamisi P, Ferdinand F, Varkonyi-Koczy AR, Reuter U, Atkinson PM. Covid-19 outbreak prediction with machine learning. Available at SSRN. 2020;3580188.
17. Arora P, Kumar H, Panigrahi BK. Prediction and analysis of COVID-19 positive cases using deep learning models: A descriptive case study of India. *Chaos, Solitons Fractals*. 2020;139:110017.
18. Al-Qaness MA, Ewees AA, Fan H, Abd El Aziz M. Optimization method for forecasting confirmed cases of COVID-19 in China. *J Clin Med*. 2020;9(3):674.
19. Chimmula VKR, Zhang L. Time series forecasting of COVID-19 transmission in Canada using LSTM networks. *Chaos, Solitons Fractals*. 2020;135:109864.
20. Shahid F, Zameer A, Muneeb M. Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM. *Chaos, Solitons Fractals*. 2020;140:110212.
21. Statista. Cumulative COVID-19 tests India 2020-2021. 2020. <https://www.statista.com/statistics/1113465/india-coronavirus-covid-19-tests-cumulative/>.
22. Chaudhary SSBR. COVID-19: UAE among the world's first countries to test 50 per cent of its population. *Health Gulf News*. 2020, August 5. <https://gulfnews.com/uae/health/covid-19-uae-among-the-worlds-first-countries-to-test-50-per-cent-of-its-population-1.72997437>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.