Modeling Palestinian COVID-19 Cumulative Confirmed Cases: A Comparative Study

Issam Dawoud
Department of Mathematics, Al-Aqsa University, Gaza, Palestine

ABSTRACT

COVID-19 is still a major pandemic threatening all the world. In Palestine, there were 26,764 COVID-19 cumulative confirmed cases as of 27th August 2020. In this paper, two statistical approaches, autoregressive integrated moving average (ARIMA) and k-th moving averages - ARIMA models are used for modeling the COVID-19 cumulative confirmed cases in Palestine. The data was taken from World Health Organization (WHO) website for one hundred seventy-six (176) days, from March 5, 2020 through August 27, 2020. We identified the best models for the above mentioned approaches that are ARIMA (1,2,4) and 5-th Exponential Weighted Moving Average – ARIMA (2,2,3). Consequently, we recommended to use the 5-th Exponential Weighted Moving Average – ARIMA (2,2,3) model in order to forecast new values of the daily cumulative confirmed cases in Palestine. The forecast values are alarming, and giving the Palestinian government a good picture about the next number of COVID-19 cumulative confirmed cases to review her activities and interventions and to provide some robust structures and measures to avoid these challenges.

© 2020 The Authors. Production and hosting by Elsevier B.V. on behalf of KeAi Communications Co., Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

The SARS-CoV-2 (COVID-19) is a new pandemic which spreads rapidly from person to person and to many countries (Ceylan, 2020). In most countries, the health systems are tenuous and especially in war regions. Now, This pandemic is the world crisis. According to studies, COVID-19 originated from Wuhan, China, in the end of 2019 and has caused an economic crisis whose impact will be felt for the next years (WHO, 2020). On January 30, 2020, the World Health Organization (WHO) declared the outbreak as a Public Health Emergency of International concern (International Health Regulations, 2020 30 January 2020; WHO, 2020). According to the ministry of health in Palestine, the first case was detected on March 5, 2020. 152 deaths are reported on August 27, 2020 with more than 26,764 total number of infected cases. The phase of COVID-19 epidemics can be decomposed as an exponential growth. Since forecasting is an important issue nowadays, there are many different methods have been proposed and developed to get accurate forecasts. The well-known used method is called the Box and Jenkins method using autoregressive integrated moving average (ARIMA) models. Also, many studies in different fields discussed using different forecasting models as ARIMA models and k-th moving averages (k-th SMA), k-th weighted moving average (k-th WMA) and k-th exponential weighted moving average (k-th EWMA) with ARIMA models in order to find the best accurate forecasting models under some conditions, for example, Crane and Crotty.
De (1967), Shami and Snyder (1998), Billah, King, Snyder, & Koehler (2006), Burman and Shumway (2006), Shih and Tsokos (2008), Tsokos (2010), Safi and Dawoud (2013) and Dawoud and Kaciranlar (2017a, 2017b, 2017c). Recently, there are many studies about building models for forecasting the ongoing trend with data-driven analysis and estimating the COVID-19, to mention a few, Li et al. (2020), Fanelli and Piazza (2020), Roda et al. (2020), Wei et al. (2016), Al-qaness et al. (2020), Anastassopoulou et al. (2020), Wang et al. (2020), Ayinde et al. (2020), Ghosal et al. (2020), Ceylan (2020) and Ogundokun et al. (2020). The main purpose of this article is modeling COVID-19 cumulative confirmed cases in Palestine using different existing methods as ARIMA and k-th moving averages – ARIMA models for giving accurate forecast values. This article is organized as follows, in section 2 we present some fundamental definitions. In section 3 we give and describe measures of forecast accuracy. In section 4 we present data description, forecasting techniques and the fitting models for COVID-19 data. Finally, in section 5 some conclusions are given.

2. Fundamental definitions

In this section, we present the definitions of ARIMA, the k-th SMA, k-th WMA and k-th EWMA models.

Definition 1. (Box et al., 1994) (ARIMA Model)

The classical ARIMA$(p, d, q)$ model is defined as

$$\varphi_p(B) (1 - B)^d T_t = \theta_q(B) \varepsilon_t, \quad (2.1)$$

where $B T_t = T_{t-1}$ is the difference filter, $d$ is the degree of differencing of the series, $\varphi_p(B) = (1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_p B^p)$ and $\theta_q(B) = (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q)$. Stationarity and invertibility requires the roots of $\varphi_p(B)$ and $\theta_q(B)$ to lie outside the unit circle, respectively.

Definition 2. (Shih & Tsokos, 2008) (The k-th SMA method)

The k-th SMA method of a time series $\{T_t\}$ is defined as

$$S_t = \frac{1}{k} \sum_{j=0}^{k-1} T_{t-k+1+j}, \quad (2.2)$$

where $t = k, k+1, \ldots, n$.

Definition 3. (Shih & Tsokos, 2008) (The k-th SMA Back-Shift)

The k-th SMA Back-Shift operator in order to obtain the estimates of the original time series data $\{T_t\}$, that is.

$$\tilde{T}_t = k S_t - T_{t-1} - T_{t-2} - \cdots - T_{t-k+1}, \quad (2.3)$$

Definition 4. (Tsokos, 2010) (The k-th WMA Method)

The k-th WMA method of a time series $\{T_t\}$ is defined as:

$$W_t = \frac{\sum_{j=0}^{k-1} (j + 1) T_{t-k+1+j}}{(1 + k) k/2}, \quad (2.4)$$

where $t = k, k+1, \ldots, n$.

Definition 5. (Tsokos, 2010) (The k-th WMA Back-Shift)

The k-th WMA Back-Shift operator in order to obtain the estimates of the original time series data $\{T_t\}$, that is.

$$\tilde{T}_t = \frac{(1 + k) k/2}{k} \sum_{j=1}^{k} T_{t-k+1+j} - T_{t-k+1}, \quad (2.5)$$

Definition 6. (Tsokos, 2010) (The k-th EWMA Method)

The k-th EWMA method of a time series $\{T_t\}$ is defined as:
\[
E_t = \frac{\sum_{j=0}^{k-1} (1 - \alpha)^{j-1} T_{t-k+1+j}}{\sum_{j=0}^{k-1} (1 - \alpha)^{j}},
\]

(2.6)

where \( t = k, k + 1, \ldots, n \) and the smoothing factor \( \alpha = \frac{2}{k+1} \).

**Definition 7.** (Tsokos, 2010) (The \( k \)-th EWMA Back-Shift)

The \( k \)-th EWMA Back-Shift operator in order to obtain the estimates of the original time series data \( \{T_t\} \), that is,

\[
\hat{T}_t = \sum_{j=0}^{k-1} (1 - \alpha)^{j} \hat{E}_t - (1 - \alpha) T_{t-1} - (1 - \alpha)^2 T_{t-2} - \cdots - (1 - \alpha)^{k-1} T_{t-k-1}.
\]

(2.7)

### 3. Measures of forecast accuracy

Several measures of forecast accuracy have been introduced by many authors. These measures are recommended to use in comparing the forecast methods accuracy which are applied to univariate time series data. For example, Hyndman and Koehler (2006) introduced the Mean Square Error (MSE) as the most used measure of deviation between the actual and the predicted value.

**Definition 9.** The MSE measure is defined as:

\[
MSE = \frac{1}{N} \sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2
\]

(3.1)

where \( Y_i \) is the actual value and, \( \hat{Y}_i \) is the predicted value. MSE is one of the most commonly used measures of forecast accuracy and the \( RMSE = \sqrt{MSE} \).

### 4. Fitting forecasting models for COVID-19 data

This section presents data description, forecasting techniques and the fitting models for COVID-19 data by using the two different approaches, ARIMA \((p,d,q)\) and \( k \)-th moving averages-ARIMA \((p,d,q)\) models. Consider the daily COVID-19 cumulative confirmed cases in Palestine, from 5-3-2020 through 27-8-2020, the forecasting results are presented in the following subsections.

#### 4.1. Data description

We use a daily COVID-19 cumulative confirmed cases dataset of Palestine from 5-3-2020 through 27-8-2020 (data source: WHO website, 2020). The time-series plot of the daily COVID-19 cumulative confirmed cases is in Fig. 1.
**Fig. 1** shows that the Daily COVID-19 cumulative confirmed cases over time is exponentially increasing which the cumulative number of confirmed cases reaches 26,764 cases as of 27, August 2020.

### 4.2. Forecasting technique

In this study, we are fitting the classical ARIMA, a k-th SMA, a k-th WMA and k-th EWMA with ARIMA models for 90% (approximately 158 observations) of the available data which is called the in-sample forecast or the training data and the remaining 10% (approximately 18 observations) is used for the out-of-sample forecast or testing the models. The forecast accuracy measures to these forecasting models are given for the in-sample and the out-of-sample. R-statistical software is used for fitting the above mentioned models.

### 4.3. Fitting the classical ARIMA model for COVID-19 data

After taking the second difference of the original data in order to make it stationarity, different combinations of ARIMA models with \( p + q \leq 5 \) and their corresponding RMSE values are shown in Table 1.

Table 1 shows that the lowest RMSE value among all models is for ARIMA (1,2,4) which is equal to 60.66346 for in-sample forecasts. This result indicates that ARIMA (1,2,4) model is more efficient than other mentioned ARIMA models with \( p + q \leq 5 \) for in-sample forecasts.

We use maximum likelihood estimation and show the results obtained from the R statistical software in Table 2. Here we see that \( p_1 = -0.7880, \theta_1 = 0.2011, \theta_2 = -0.9299, \theta_3 = -0.0099 \) and \( \theta_4 = 0.4589 \). We also see that the estimated noise variance is \( \sigma_e^2 = 3680 \).

### 4.4. Fitting the k-th moving averages - ARIMA models for COVID-19 data

Different combinations of k-th moving averages-ARIMA models with \( k \leq 5 \) and their corresponding RMSE values are shown in Table 3.

Table 3 shows that the lowest RMSE value among all models is for 5-th EWMA-ARIMA (2,2,3) which is equal to 59.9422 for in-sample forecasts. This result indicates that 5-th EWMA-ARIMA (2,2,3) model is more efficient than other k-th moving averages-ARIMA models with \( k \leq 5 \) for in-sample forecasts.

We use maximum likelihood estimation and show the results obtained from the R statistical software in Table 4. Here we see that \( p_1 = -0.0086, \varphi_2 = -0.8788, \theta_1 = 0.1584, \theta_2 = 0.9112 \) and \( \theta_3 = 0.3873 \). We also see that the estimated noise variance is \( \sigma_e^2 = 3593 \). Noting the \( P \)-values, the estimates of all autoregressive and moving average coefficients are significantly different from zero statistically.

Table 5 shows the actual and forecasting results for COVID-19 cumulative confirmed cases from August 14, 2020 to August 31, 2020 based on the classical ARIMA (1,2,4) and 5th EWMA-ARIMA (2,2,3) models.

Table 5 shows that the 5-th EWMA-ARIMA forecasts are so near to the actual values of the COVID-19 cumulative confirmed cases than that of the classical ARIMA. Also, the RMSE for ARIMA and 5-th EWMA-ARIMA equal 158.1097 and 138.3910, respectively for the out-of-sample forecasts. This result shows that RMSE of 5-th EWMA-ARIMA is 87.53\% of RMSE for ARIMA. In other words, the RMSE of ARIMA model is 1.14 times RMSE of the 5-th EWMA-ARIMA model. This means 5-th EWMA-ARIMA model for forecasting is much more accurate and efficient than the ARIMA forecasting model.

Table 6 shows the forecasting results for COVID-19 cumulative confirmed cases from 28 to 08-2020 to 30-10-2020 (62 values) based on the classical ARIMA (1,2,4) and 5th EWMA-ARIMA (2,2,3) models.

| Model Order (1,2,0) | (2,2,0) | (3,2,0) | (4,2,0) | (5,2,0) | (0,2,1) |
|---------------------|---------|---------|---------|---------|---------|
| RMSE 74.70927       | 63.1612 | 63.1909 | 62.1342 | 61.8354 | 66.02938|
| Continued           |         |         |         |         |         |
| Model Order (0,2,2) | (0,2,3) | (0,2,4) | (0,2,5) | (1,2,1) | (1,2,2) |
| RMSE 65.71853       | 61.4393 | 61.9999 | 61.1342 | 61.8354 | 65.94331|
| Continued           |         |         |         |         |         |
| Model Order (1,2,3) | (1,2,4) | (2,2,1) | (2,2,2) | (2,2,3) | (3,2,1) |
| RMSE 61.14791       | 60.6634 | 62.4458 | 61.9661 | 61.1100 | 62.21753|
| Continued           |         |         |         |         |         |
| Model Order (3,2,2) | (4,2,1) | (0,2,0) |         |         |         |
| RMSE 61.95796       | 61.7425 | 78.7981 |         |         |         |

Table 1
The RMSE of forecast accuracy for the different-ARIMA models for in-sample.

I. Dawoud Infectious Disease Modelling 5 (2020) 748–754
5. Conclusion

In previous studies, many statistical methods and time series models were used to forecast epidemic cases. In this article, we have applied the ARIMA and the k-th moving averages-ARIMA models on the data of the cumulative confirmed cases of COVID-19 in Palestine. We have examined the model selection sensitivity based on forecast accuracy criterion, RMSE, for the above mentioned models. The main finding in general is that some of k-th moving averages ARIMA models give better results for in-sample and out-of-sample forecasts than the classical ARIMA models. In particular, the best model is 5-th EWMA-
Table 6
Forecast values for daily COVID 19 cumulative confirmed cases from 01 to 09–2020 to 30-10-2020 (64 values) based on ARIMA (1,2,4), 5th EWMA-ARIMA (2,2,3).

| Date     | ARIMA | 5th EWMA-ARIMA | Date     | ARIMA | 5th EWMA-ARIMA |
|----------|-------|----------------|----------|-------|----------------|
| 28–08–2020 | 27,293 | 26,897 | 29–08–2020 | 42,453 | 41,797 |
| 29–08–2020 | 27,767 | 27,531 | 30–08–2020 | 42,927 | 42,258 |
| 30–08–2020 | 28,240 | 27,994 | 01–09–2020 | 43,401 | 42,718 |
| 31–08–2020 | 28,715 | 28,463 | 02–09–2020 | 43,874 | 43,177 |
| 01–09–2020 | 29,188 | 28,937 | 03–09–2020 | 44,348 | 43,638 |
| 02–09–2020 | 29,662 | 29,351 | 04–09–2020 | 44,822 | 44,098 |
| 03–09–2020 | 30,136 | 29,831 | 05–09–2020 | 45,296 | 44,558 |
| 04–09–2020 | 30,609 | 30,291 | 06–09–2020 | 45,769 | 45,018 |
| 05–09–2020 | 31,083 | 30,755 | 07–09–2020 | 46,243 | 45,479 |
| 06–09–2020 | 31,557 | 31,218 | 08–09–2020 | 46,717 | 45,939 |
| 07–09–2020 | 32,031 | 31,660 | 09–09–2020 | 47,191 | 46,399 |
| 08–09–2020 | 32,504 | 32,131 | 10–09–2020 | 47,664 | 46,859 |
| 09–09–2020 | 32,978 | 32,594 | 11–09–2020 | 48,138 | 47,320 |
| 10–09–2020 | 33,452 | 33,055 | 12–09–2020 | 48,612 | 47,780 |
| 11–09–2020 | 33,926 | 33,514 | 13–09–2020 | 49,085 | 48,240 |
| 12–09–2020 | 34,399 | 33,972 | 14–09–2020 | 49,559 | 48,700 |
| 13–09–2020 | 34,873 | 34,434 | 15–09–2020 | 50,033 | 49,160 |
| 14–09–2020 | 35,347 | 34,895 | 16–09–2020 | 50,507 | 49,621 |
| 15–09–2020 | 35,821 | 35,354 | 17–09–2020 | 50,980 | 50,080 |
| 16–09–2020 | 36,294 | 35,814 | 18–09–2020 | 51,454 | 50,541 |
| 17–09–2020 | 36,768 | 36,275 | 19–09–2020 | 51,928 | 51,001 |
| 18–09–2020 | 37,242 | 36,736 | 20–09–2020 | 52,402 | 51,461 |
| 19–09–2020 | 37,716 | 37,195 | 21–09–2020 | 52,875 | 51,921 |
| 20–09–2020 | 38,189 | 37,655 | 22–09–2020 | 53,349 | 52,381 |
| 21–09–2020 | 38,663 | 38,116 | 23–09–2020 | 53,823 | 52,842 |
| 22–09–2020 | 39,137 | 38,576 | 24–09–2020 | 54,297 | 53,302 |
| 23–09–2020 | 39,611 | 39,036 | 25–09–2020 | 54,770 | 53,762 |
| 24–09–2020 | 40,084 | 39,496 | 26–09–2020 | 55,244 | 54,222 |
| 25–09–2020 | 40,558 | 39,957 | 27–09–2020 | 55,718 | 54,683 |
| 26–09–2020 | 41,032 | 40,417 | 28–09–2020 | 56,192 | 55,143 |
| 27–09–2020 | 41,506 | 40,877 | 29–09–2020 | 56,665 | 55,603 |
| 28–09–2020 | 41,979 | 41,336 | 30–09–2020 | 57,139 | 56,063 |

ARIMA (2,2,3) among all models. It is recommended for practitioners to use the k-th moving averages – ARIMA models for modeling any phenomenon and getting accurate forecast values.

Declaration of competing interest

The author declares that he has no conflict of interest.

References

Al-qaness, M. A. A., Ewees, A. A., Fan, H., & ElAiz, M. A. (2020). Optimization method for forecasting confirmed cases of COVID-19 in China. Journal of Clinical Medicine, 9, 674. https://doi.org/10.3390/jcm9030674

Anastassopoulou, C., Russo, L., Tsakris, A., & Siettos, C. (2020). Data-based analysis, modelling and forecasting of the COVID-19 outbreak. PLoS One, 15, Article e0230405. https://doi.org/10.1371/journal.pone.0230405

Ayinde, K., Lukman, A. F., Rauf, I. R., Alabi, O. O., Okon, C. E., & Ayinde, O. E. (2020). Modeling Nigerian covid-19 cases: A comparative analysis of models and estimators. Chaos, Solitons & Fractals, 138. Article 109911.

Billah, B., King, M. L., Snyder, R. D., & Koehler, A. B. (2006). Exponential smoothing model selection for forecasting. International Journal of Forecasting, 22, 239–247.

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (1994). Time series analysis forecasting and control (3rd ed.). New Jersey: Prentice Hall.

Burman, P., & Shumway, R. H. (2006). Generalized exponential predictors for time series forecasting. Journal of the American Statistical Association, 101, 1508–1606.

Ceylan, Z. (2020). Estimation of COVID-19 prevalence in Italy, Spain, and France. The Science of the Total Environment, 729. Article 138817.

Crane, D., & Crotty, J. (1967). A two stage forecasting model: Exponential smoothing and multiple regression. Management Science, 13, 501–507.

Dawoud, I., & Kaciranlar, S. (2017a). An optimal k of kth MA-ARIMA models under a class of ARIMA model. Communications in Statistics - Theory and Methods, 46(12), 5754–5765.

Dawoud, I., & Kaciranlar, S. (2017b). An optimal k of kth MA-ARIMA models under AR(p) models. Communications in Statistics - Simulation and Computation, 46(4), 2842–2864.

Fanelli, D., & Piazza, F. (2020). Analysis and forecast of COVID-19 spreading in China, Italy and France. Chaos, Solitons & Fractals, 134, 1e12. https://doi.org/10.1016/j.chaos.2020.109751

Ghosal, S., Sengupta, S., Majumder, M., & Sinha, B. (2020). Linear Regression Analysis to predict the number of deaths in India due to SARS-CoV-2 at 6 weeks from day 0 (100 cases – March 14th 2020). Diabetes & Metabolic Syndrome. Clinical Research Reviews, 14. Article 311e315.

Hyndman, R. J., & Koehler, A. B. (2006). Another Look at measures of forecast accuracy. International Journal of Forecasting, 22, 679–688.
International Health Regulations. (2020). Statement on the second meeting of the Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV). World Health Organization, 30 January 2020.

Li, Q., Feng, W., & Quan, Y. H. (2020). Trend and forecasting of the COVID-19 outbreak in China. Journal of Information Security, 80. https://doi.org/10.1016/j.jis.2020.02.014, 469e496.

Ogundokun, R. O., Lukman, A. F., Kibria, B. M. G., Awotunde, J. B., & Aladeitan, B. B. (2020). Predictive modelling of COVID-19 confirmed cases in Nigeria. Journal of Infectious Disease Modelling, 5, 543–548. https://doi.org/10.1016/j.idm.2020.08.003

Roda, W. C., Varughese, M. B., Han, D., & Li, M. Y. (2020). Why is it difficult to accurately predict the COVID-19 epidemic? Infection Disease Modelling, 5, 271e281.

Safi, S., & Dawoud, I. (2013). Comparative study on forecasting accuracy among moving average models with simulation and PALTEL stock market data in Palestine. American Journal of Theoretical and Applied Statistics, 2, 202–209.

Shami, R., & Snyder, R. (1998). Exponential smoothing methods of forecasting and general ARMA time series representations. Monash Econometrics and Business Statistics Working Paper, 3, 98.

Shih, S., & Tsokos, C. (2008). A weighted moving average process for forecasting. Journal of Modern Applied Statistical Methods, 7, 187–197.

Tsokos, C. (2010). K-th moving, weighted and exponential moving average for time series forecasting models. European Journal of Pure and Applied Mathematics, 3, 406–416.

Wang, L., Li, J., Guo, S., Xie, N., Yao, L., Day, S. W., … Sun, D. (2020). Real-time estimation and prediction of mortality caused by COVID-19 with patient information based algorithm. Journal of Science Teacher Education. https://doi.org/10.1016/j.jstotenv.2020.138394

Wei, W., Jiang, J., Liang, H., Gao, L., Liang, B., Huang, J., Zang, N., Liao, Y., Yu, J., Lai, J., Qin, F., Su, J., Ye, L., & Chen, H. (2016). Application of a combined model with autoregressive integrated moving average (ARIMA) and generalized regression neural network (GRNN) in forecasting hepatitis incidence in Heng County, China. PLoS One, 11, Article e0156768. https://doi.org/10.1371/journal.pone.0156768

WHO. (2020). Novel coronavirus-China. WHO. Retrieved 9 April 2020. https://covid19.who.int/table, (2020).