

Forecasting COVID-19 Confirmed Cases in Ghana: A Model Selection Approach

Twumasi-Ankrah, Sampson; Owusu, Michael; Appiah, Simon Kojo; Pels, Wilhemina Adoma; Arthur, Doris

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Empfohlene Zitierung / Suggested Citation:

Twumasi-Ankrah, S., Owusu, M., Appiah, S. K., Pels, W. A., & Arthur, D. (2021). Forecasting COVID-19 Confirmed Cases in Ghana: A Model Selection Approach. *Path of Science*, 7(2), 4001-4010. <https://doi.org/10.22178/pos.67-2>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
<https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more information see:
<https://creativecommons.org/licenses/by/4.0>

Forecasting COVID-19 Confirmed Cases in Ghana: A Model Selection Approach

Sampson Twumasi-Ankrah¹, Michael Owusu¹, Simon Kojo Appiah¹, Wilhemina Adoma Pels¹, Doris Arthur¹

¹ Kwame Nkrumah University of Science and Technology

PMB, University Post Office, 40080, Kumasi, Ashanti Region, Ghana

DOI: [10.22178/pos.67-2](https://doi.org/10.22178/pos.67-2)

LCC Subject Category: [R5-920](#)

Received 12.01.2021


Accepted 25.02.2021

Published online 28.02.2021

Corresponding Author:

Sampson Twumasi-Ankrah,

sampson.ankrah@yahoo.com.com

© 2021 The Authors. This article is licensed under a [Creative Commons Attribution 4.0 License](#) 

Abstract. This study seeks to determine an appropriate statistical technique for forecasting the cumulated confirm cases of Coronavirus in Ghana. Cumulated daily data spanning from March 12, 2020, to August 04, 2020, was retrieved from the Center for Systems Science and Engineering at Johns Hopkins University. Four statistical forecasting techniques: Autoregressive Integrated Moving Average, Artificial Neural Network, Exponential smoothing and Autoregressive Fractional Integrated Moving Average were fitted to the COVID-19 series. Their respective forecast accuracy measures were compared to select the appropriate technique for forecasting the COVID-19 cases. Our findings revealed that the ARFIMA technique was a suitable statistical model for predicting COVID-19 cases in Ghana. The "best" model for forecasting is ARFIMA (2, 0.49, 4) which passed all the needed diagnostic tests. An unequal weight was estimated to derive a combined model for all four forecasting techniques. A 149-cumulated daily forecast from the "best" model and the combined model revealed that the number of confirmed COVID-19 cases would increase slightly until the end of this year.

Keywords: exponential smoothing; COVID-19; Artificial Neural Network; Forecast.

INTRODUCTION

The occurrence of the COVID-19 pandemic has given the 21st-century generation a feel of the Spanish flu in 1918. The COVID-19 was seen in December 2019, when a cluster of cases with unknown pneumonia with similar clinical manifestations suggesting viral pneumonia appeared in Wuhan City, Hubei Province, China. According to the WHO [1] situation report, SARS-CoV-2 (COVID-19) belong to β -coronavirus, which is a typical RNA-virus and can spread from person-to-person [2]. According to Fernandes [3], the COVID-19 outbreak has caused serious global socio-economic turmoil. Globally, as of August 7 2020, about 21.88 million confirmed cases of COVID-19 had been recorded in a total of about 215 countries with more than 773,926 deaths and 14.6 million recoveries [4, 5]. The Africa sub-region constitutes 5.14% of the global confirmed cases, 3.32 % of deaths and 5.75 % of the recovery's cases [4, 5]. The confirmed cases in Ghana stands at 42,653 with 239 deaths and 40,567 re-

coveries cases as of August 7 2020. Many countries, including Ghana, have responded by implementing self-isolation measures, social distancing, and wearing the mask to prevent further spread [6].

Decision-makers are confronted with considerable uncertainties in deciding how to deal with the pandemic in scarce health resources. In this regard, it is practically essential to construct statistical models that are accurate and realistic enough to help forecast its future behaviour in terms of a possible number of daily cases. This can assist the medical system in better plan the healthcare resources for new patients. These statistical predictive models are useful in forecasting as well as controlling the global epidemic threat.

Some studies have modelled and forecasted the COVID-19 pandemic using the time series analysis methods [7, 8, 9, 10, 11]. Although all countries deal with the same SARS-CoV-2, predicting future outbreaks seems to differ based on cases'

unique pattern. However, there is limited data on statistical methods that best predict SARS-CoV-2 infections in Ghana and other African countries.

This study aims to compare the performance of four different time series methods and determine the appropriate or "best" way that could be used to forecast the confirmed cases of COVID-19 in Ghana. In each time series technique, competing models are constructed, and information criteria are used to select the "within-best" model. The error metric from the out-sample of these forecast techniques is compared to choose the overall "best" forecast model for the COVID-19 cases in Ghana. Therefore, in this study, much attention is giving to how the "best" forecast method is selected.

METHODS AND MATERIALS

Dataset and Approach of Analysis. We focus on the confirmed cumulative daily COVID-19 cases in Ghana starting from March 12, 2020, to August 4 2020. The data from March 12, 2020, to July 9, 2020, were used as the training data for fitting the model, while the daily confirmed cases from July 10, 2020, to August 4, 2020, were used as test data for the comparison of the forecast performance of the models.

The procedure used to analyze the dataset in this study are indicated as follows:

1. The COVID-19 confirmed daily cases are plotted to observe the trend pattern and other features.
2. Three different unit root tests of stationarity are performed on the time series data.
3. For each forecasting technique employed, competing models are fitted to the cumulative COVID-19 case series; the "best" model is selected using the minimum information criterion.
4. In a situation where the information criteria disagree, we compare the models' forecast accuracy measure suggested by each of the information criteria. The final "best" model is selected based on the minimum forecast accuracy measure.
5. The forecast performance of "best" models from each forecasting technique in step 3 is compared using their error metric.
6. Forecast COVID-19 confirmed cases using the overall "best" forecasting technique in step 5.

Unit Root Tests. The Augmented Dickey-Fuller (ADF), Phillips & Perron (PP) and the Kwiatkowski Phillips Schmidt and Shin (KPSS) tests are the three most commonly used unit root tests of which the ADF and the PP have the same null hypothesis that the given time series data set have a unit root (that is, it is not stationary). The alternative idea is that the data set does not have a unit root (that is, it is fixed). However, the KPSS has its null hypothesis as the data set is stationary with alternative as the series is not stationary

Time Series Models

1. *Autoregressive Integrated Moving Average (ARIMA) Model:*

$$\phi(B)(1 - B)^d Y_t = \theta(B)e_t \quad (1)$$

where $\phi(B)$ is the autoregressive in a backshift form, $(1 - B)^d$ is the differencing order, and $\theta(B)$ is the moving average part of the ARIMA model.

2. *Exponential Smoothing Technique (ETS).* Author [12] extended the simple exponential smoothing to allow the forecasting of data with a trend. This method involves a forecast equation and two smoothing equations (one for the level and one for the direction):

Forecast equation:

$$y(t + h|t) = \ell_t + hb_t \quad (2.1)$$

Level Equation:

$$\ell_t = \alpha y_t + (1 - \alpha)(\ell_{t-1} + b_{t-1}) \quad (2.2)$$

Trend equation:

$$b_t = \beta^* (\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1} \quad (2.3)$$

where ℓ_t denotes an estimate of the level of the series at time t , b_t denotes an estimate of the trend (slope) of the series at time t , α is the smoothing parameter for the level, $0 \leq \alpha \leq 1$, and β^* is the smoothing parameter for the trend, $0 \leq \beta^* \leq 1$.

According to [13], the simple exponential smoothing method is defined by cell (N, N), Holt's linear method by cell (A, N), the damped trend method by cell (Ad, N), Holt-Winters' additive method by cell (A, A), and Holt-Winters' multiplicative method is given by cell (A, M) in Table 1.

Table 1 – A two-way classification of exponential smoothing methods

| Trend Component | Seasonal Component | | |
|-----------------------|--------------------|--------------|--------------------|
| | N (None) | A (Additive) | M (Multiplicative) |
| N (None) | (N, N) | (N, A) | (N, M) |
| A (Additive) | (A, N) | (A, A) | (A, M) |
| Add (Additive damped) | (Ad, N) | (Ad, A) | (Ad, M) |

3. *Autoregressive Fractional Integrated Moving Average (ARFIMA)*. The ARFIMA is considered an extended memory model. In ARFIMA, the idea of assigning $d = 1$ or 2 to make a series stationary has been extended to the class of fractionally integrated ARMA, or ARFIMA models, where we allow $-0.5 < d < 0.5$; when d is negative [14]. Now, d becomes a parameter to be estimated, and a better way to calculate d is using the expression:

$$\pi_j = \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} \tag{3}$$

Neural Network Models. A neural network is a network of "neurons", which are organized in layers. The predictors (or inputs) form the bottom layer, and the forecasts (or outputs) include the top layer (Figure 1). Most external networks contain no hidden layers and are equivalent to linear regressions. In time series, the series' lagged values can be used as inputs to a neural network autoregression or NNAR model. The notation NNAR (p, k) is used to indicate that there are p lagged inputs and k nodes in the hidden layer.

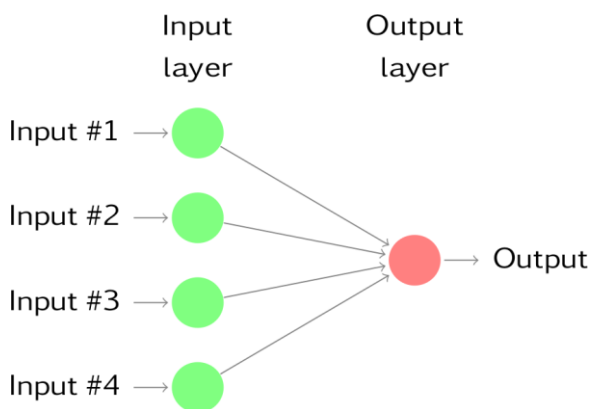


Figure 1

Model Selection Criteria. In this study, three information criteria are utilized; the Akaike Information Criterion (AIC), the Corrected Akaike information criterion (AICc) and the Bayesian Information Criterion (BIC). The AIC is given by (4):

$$AIC = -2 \ln L(\hat{\theta}_k) + 2k \tag{4}$$

where $L(\hat{\theta}_k)$ is the likelihood of the fitted model, and k is a number of unknown parameters free to vary.

The AIC_c is also computed as (5)

$$AIC_c = -2 \ln L(\hat{\theta}_k) + \frac{2kn}{n-k-2} \tag{5}$$

where n is the total number of observation while the BIC is given by (6):

$$BIC = -2 \ln L(\hat{\theta}_k) + k \ln(n) \tag{6}$$

Forecast Accuracy Measures. Three error metrics, namely, root mean square error (RMSE), mean absolute percentage error (MAPE), and mean fundamental error (MAE), were employed to measure the predictive performance of the models in (7)–(9). The RMSE is a measure of the spread of the forecast errors about the actual data points, which informs how far or near the forecasted values of an estimated model are from the real data points. It is computed as (7):

$$RMSE = \sqrt{mean[(e_i)^2]} \tag{7}$$

where $e_i = Y_t - \hat{Y}_t$ is the error.

The MAPE is a measure of the size of the error of a forecast in percentage. It is used to measure the accuracy of a prediction using the formula (8):

$$MAPE_{forecast} = \left(\frac{\frac{1}{N} \sum_{i=1}^n |Y_t - \hat{Y}_t|}{|Y_t|} \right) \times 100 \tag{8}$$

The MAE is a scale-dependent measure that is based on the absolute errors and computed as (9):

$$MAE_{forecast} = \sqrt{mean[(e_i)^2]} \tag{9}$$

RESULTS AND DISCUSSIONS

Firstly, four different univariate time series techniques were employed to model and forecast COVID-19 cases in Ghana. These time series techniques are ARIMA, ETS, ANN and ARFIMA. The various methods' predictive performance was used in selecting the "best" way for forecasting COVID-19 cases in Ghana. For each forecasting technique, appropriate competing models were constructed, and their information criteria

were recorded. The model with the least information criterion was chosen as the 'best' model for forecasting the COVID-19 time series data. The R software precisely predicted, and the ARFRIMA package was used to run the time series models.

Time Series Plot. Generally, in Figure 2, there is a strong upward trend in COVID-19 cases from 2020-03-12 to 2020-08-04. This indicates that the COVID-19 cases series is not stationary.

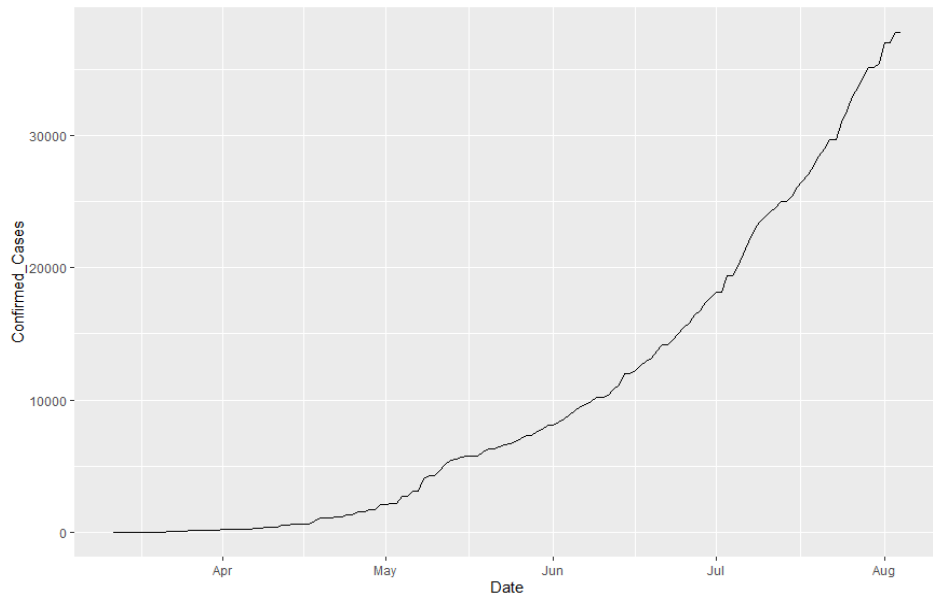


Figure 2 – Time series plot of cumulative confirmed cases of COVID-19 in Ghana

As observed in Figure 2, the strong upward trend of COVID-19 cases in Ghana shows that the series is not stationary. This is confirmed by results of the three-unit root tests ADF, PP and KPSS as presented in Table 2, where the p-values are all greater than 5% level of significance. Thus, there is no enough evidence to reject the null hypothesis that the COVID-19 series of Ghana is non-stationary. Nonetheless, a first difference of the series made it stationary, as confirmed by the ADF and the PP test. Yet, the KPSS test still showed non-stationarity of the series until the second difference.

Table 2 – Unit root Tests on COVID-19 Cases in Ghana

| Test | Order of Differencing | p-value | Conclusion |
|------|--------------------------------|---------|------------------------------|
| ADF | I (0) (original data) | 0.99 | The series is not stationary |
| | I (1) (first differenced data) | 0.01 | The series is stationary |

| Test | Order of Differencing | p-value | Conclusion |
|------|---------------------------------|---------|------------------------------|
| PP | I (0) (original data) | 0.99 | The series is not stationary |
| | I (1) (first differenced data) | 0.01 | The series is stationary |
| KPSS | I (0) (original data) | 0.01 | The series is not stationary |
| | I (1) (first differenced data) | 0.01 | The series is not stationary |
| | I (2) (second differenced data) | 0.1 | The series is stationary |

Model Selection

In statistical model building, the standard practice fits several candidate models to a dataset to choose the "best" model, thus using the minimum information criterion.

Modelling with ARIMA Model. With the "differencing" information acquired at the test of stationarity in Table 2, the ADF and PP tests suggest a different order of "1" whiles the KPSS test means a differencing order of "2". Hence, two sets of

competing models are built based on the differencing order, and their respective information criteria are computed. Their performance metric will then suggest the model be chosen for the ARIMA technique.

Table 3 presents results with differencing order of "1", and all the three information criteria (AIC, AICc and BIC) suggest ARIMA (1, 1, 2) as the "best" model.

Table 3 – Competing models and their Information criterion values for the first difference

| Model | AIC | AICc | BIC |
|------------|----------|----------|----------|
| (1, 1, 0) | 1665.084 | 1665.188 | 1670.643 |
| (2, 1, 0) | 1624.472 | 1624.681 | 1632.809 |
| (3, 1, 0) | 1611.746 | 1612.097 | 1622.863 |
| (0, 1, 1) | 1691.335 | 1691.438 | 1696.893 |
| (0, 1, 2) | 1669.936 | 1670.145 | 1678.274 |
| (0, 1, 3) | 1657.831 | 1658.182 | 1668.948 |
| (1, 1, 1) | 1604.478 | 1604.687 | 1612.816 |
| (1, 1, 2)* | 1593.076 | 1593.427 | 1604.193 |
| (1, 1, 3) | 1595.186 | 1595.717 | 1609.081 |
| (2, 1, 1) | 1594.779 | 1595.13 | 1605.895 |
| (2, 1, 2) | 1595.235 | 1595.766 | 1609.13 |
| (2, 1, 3) | 1595.997 | 1596.747 | 1612.672 |
| (3, 1, 1) | 1595.925 | 1596.456 | 1609.82 |
| (3, 1, 2) | 1598.154 | 1598.904 | 1614.829 |
| (3, 1, 3) | 1597.541 | 1598.55 | 1616.994 |

* The "best" model, boldface=minimum information criterion

Table 4 presents results with differencing order of "2", as suggested by the KPSS test. All three information criteria (AIC, AICc and BIC) indicate ARIMA (0, 2, 2) as the "best" model.

Table 4 – Competing models and their Information criterion values for the second difference

| Model | AIC | AICc | BIC |
|------------|----------|----------|----------|
| (1, 2, 0) | 1613.13 | 1613.234 | 1618.671 |
| (2, 2, 0) | 1596.538 | 1596.748 | 1604.85 |
| (3, 2, 0) | 1576.038 | 1576.392 | 1587.121 |
| (0, 2, 1) | 1587.659 | 1587.764 | 1593.201 |
| (0, 2, 2)* | 1575.96 | 1576.171 | 1584.273 |
| (0, 2, 3) | 1577.953 | 1578.307 | 1589.036 |
| (1, 2, 1) | 1577.769 | 1577.979 | 1586.081 |
| (1, 2, 2) | 1577.953 | 1578.307 | 1589.036 |
| (1, 2, 3) | 1578.802 | 1579.338 | 1592.656 |
| (2, 2, 1) | 1578.869 | 1579.223 | 1589.951 |
| (2, 2, 2) | 1580.092 | 1580.628 | 1593.946 |
| (2, 2, 3) | 1580.901 | 1581.658 | 1597.525 |
| (3, 2, 1) | 1576.631 | 1577.167 | 1590.484 |
| (3, 2, 2) | 1578.394 | 1579.151 | 1595.018 |
| (3, 2, 3) | 1580.212 | 1581.231 | 1599.607 |

*The "best" model, boldface=minimum information criterion

To select the appropriate model for the ARIMA method for COVID-19 cases in Ghana, the two models' forecast values with a different order of difference were then compared. Their accuracy measures were computed using the 3-error metrics (RSME, MAE and MAPE). From Table 5, it is evident that ARIMA (1, 1, 2) is the "best" model since it had the minimum error metric.

Table 5 – Forecast Accuracy Measures for ARIMA

| Order of Differencing | RMSE | MAE | MAPE |
|-----------------------|---------|---------|---------|
| I (1) | 3406.81 | 3223.66 | 10.3764 |
| I (2) | 3993.27 | 3728.77 | 11.8768 |

boldface = minimum error metric

Modelling with Exponential Smoothing. The appropriate exponential smoothing technique for the COVID-19 series is Holt's linear trend method. This is because the COVID-19 series exhibited a strong upward trend. Technically, from the set of six competing models, the "best" practice was ETS (A, A, N); that is, additive error, trend and no seasonality. In other words, the appropriate technique is Holt's linear trend method with additive errors.

Table 6 – Competing Models with respective information criterion

| Models | AIC | AICc | BIC |
|----------|----------|----------|----------|
| A, A, N* | 1837.047 | 1837.574 | 1850.985 |
| A, N, N | 1963.621 | 1963.828 | 1971.983 |
| M, Md, N | 1850.231 | 1850.974 | 1866.955 |
| M, N, N | 1949.955 | 1950.162 | 1958.318 |
| M, Ad, N | 1927.998 | 1928.741 | 1944.723 |
| A, Ad, N | 1843.119 | 1843.862 | 1859.844 |

*The "best" model, boldface = minimum information criterion

Modelling with Artificial Neural Network. Several competing artificial neural networks were constructed after setting seed, and NNAR (3, 1, 2) model was considered the "best" since it had the minimal forecast accuracy measure in Table 7.

Table 7 – Competing Models with their respective performance error metric

| Model | RMSE | MAE | MPE | MAPE |
|-------------|----------|----------|-----------|-----------|
| NNAR (1, 1) | 5234.329 | 3841.024 | 10.798237 | 11.167796 |
| NNAR (2, 1) | 3782.61 | 2720.209 | 6.3341544 | 7.934181 |
| NNAR (3, 1) | 4076.732 | 2934.931 | 7.0995651 | 8.548622 |
| NNAR (1, 2) | 4302.927 | 3222.375 | 6.1057301 | 9.576065 |
| NNAR (2, 2) | 3039.127 | 2541.629 | 0.8910087 | 7.977517 |

| Model | RMSE | MAE | MPE | MAPE |
|--------------|----------|----------|-----------|----------|
| NNAR (3, 2)* | 2828.959 | 2518.455 | -5.016486 | 7.564587 |
| NNAR (1, 3) | 4326.255 | 3242.918 | 6.1598855 | 9.638857 |
| NNAR (2, 3) | 3705.861 | 2998.043 | 2.7467558 | 9.246132 |
| NNAR (3, 3) | 2959.644 | 2653.104 | -4.652074 | 8.984779 |

*The "best" model, boldface = minimum error metric ARFIMA Model

An optimal difference integer (d) was estimated to be 0.49; nine competing models were constructed. Information criteria suggested two models. Thus AIC suggested ARFIMA (2, 0.49, 4), while BIC suggested ARMA (2, 0.49, 0) as presented in Table 8.

Table 8 – Competing Models with respective information criterion

| Models | AIC | BIC |
|---------------------|----------|----------|
| ARFIMA (2, 0.49, 0) | 1296.327 | 1310.265 |
| ARFIMA (3, 0.49, 0) | 1297.811 | 1314.536 |
| ARFIMA (4, 0.49, 0) | 1299.81 | 1319.322 |
| ARFIMA (0, 0.49, 1) | 1638.439 | 1649.589 |
| ARFIMA (0, 0.49, 2) | 1541.114 | 1555.052 |
| ARFIMA (0, 0.49, 3) | 1501.865 | 1518.59 |
| ARFIMA (0, 0.49, 4) | 1454.009 | 1473.522 |
| ARFIMA (0, 0.49, 5) | 1434.665 | 1456.965 |
| ARFIMA (2, 0.49, 1) | 1297.88 | 1314.605 |
| ARFIMA (2, 0.49, 2) | 1299.67 | 1319.183 |
| ARFIMA (2, 0.49, 3) | 1298.966 | 1321.266 |
| ARFIMA (2, 0.49, 4) | 1294.283 | 1319.37 |
| ARFIMA (3, 0.49, 1) | 1299.811 | 1319.323 |
| ARFIMA (3, 0.49, 2) | 1300.885 | 1323.185 |

boldface = minimum information criterion

We estimated the two ARFIMA models' forecast performance suggested by the information criteria (AIC and BIC), and the minimal performance metric was used to select the "best" ARFIMA. The results presented in Table 9, ARFIMA (2, 0.49, 4), were chosen as the "best" model for the ARFIMA.

Table 9 – Comparison of Performance Metric for the "best" ARFIMA Model

| Models | RMSE | MAE | MAPE |
|---------------------|----------|----------|----------|
| ARFIMA (2, 0.49, 4) | 1932.628 | 1637.928 | 5.137979 |
| ARFIMA (2, 0.49, 0) | 2320.084 | 1754.166 | 5.247669 |

boldface = minimum error metric

We compare the forecast performance of the "best" models from the four different time series modelling techniques using the 3-performance

metrics computed from the "test" data. The time series technique with the minimum performance metric is selected as the "best" method. From Table 10, it is obvious that the ARFIMA (2, 0.49, 4) has the least error metric values among the other three forecasting techniques. Hence, it is concluded that the ARFIMA (2,0.49,4) is the 'best' model for forecasting COVID-19 confirmed cases in Ghana.

Table 10 – Comparison of Forecasting Techniques: ARIMA, E.T.S., NNAR and ARFIMA

| Models | RMSE | MAE | MAPE |
|---------------------|----------|----------|----------|
| ARIMA (1, 1, 2) | 3406.808 | 3223.663 | 10.37643 |
| ETS (A, A, N) | 3988.566 | 3724.23 | 11.86228 |
| NNAR (3, 2) | 2828.959 | 2518.455 | 8.564587 |
| ARFIMA (2, 0.49, 4) | 1932.628 | 1637.928 | 5.137979 |

boldface = minimum error metre diagnostic Checking

In Figure 3, the diagnostic checks on the residuals of the chosen model [ARFIMA (2, 0.49, 4)] is presented. This is done to see if it does not violate any of the assumptions underlying the model. From Figure 3, we observed the following:

1. There is no apparent trend in the plot of the standardized residuals over the days.
2. A plot of the ACF of the residuals confirms that none of their lags is statistically significant, implying that the residuals are not correlated.
3. The box plot shows that most of the errors are normally distributed except for a few at almost the midpoint that potential cases of outliers are observed.

The test of autocorrelations provides an essential diagnostic tool. Therefore, the Box-Ljung test was used to check for autocorrelation under the hypotheses:

H_0 : residuals are not auto-correlated versus

H_1 : residuals are auto-correlated.

From the results presented in Table 11, the null hypothesis of residuals not being auto-correlated is not rejected since, at a significance of 5%, the p-value (0.9786) is more generous.

Table 11 – Box-Ljung test

| Variable | Test Statistic | P-value |
|-----------------------|----------------|---------|
| Residuals of COVID-19 | 3.1127 | 0.9786 |

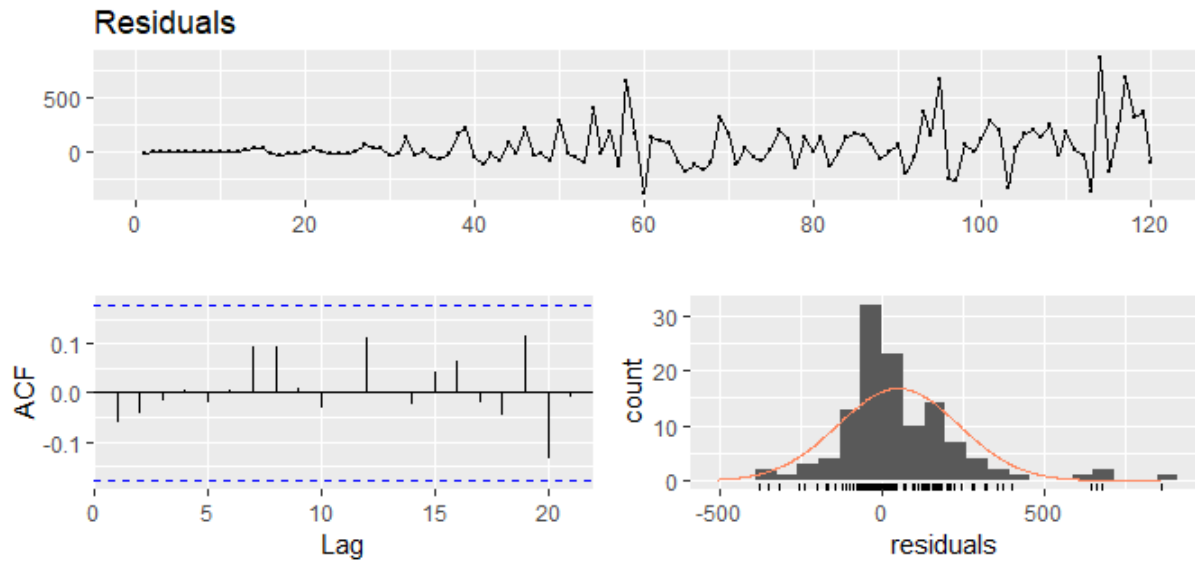


Figure 3 – Diagnostic checking on the residuals of ARFIMA (2, 0.49, 4)

In Figure 4, the forecast of cumulated confirmed COVID-19 cases from four forecast techniques (i.e., respective "best" models) starting from August (starting from 05/08/2020) to the end of

December is presented. The NNAR forecast technique gives the lowest forecast value, while the ETS technique provides the highest forecast.

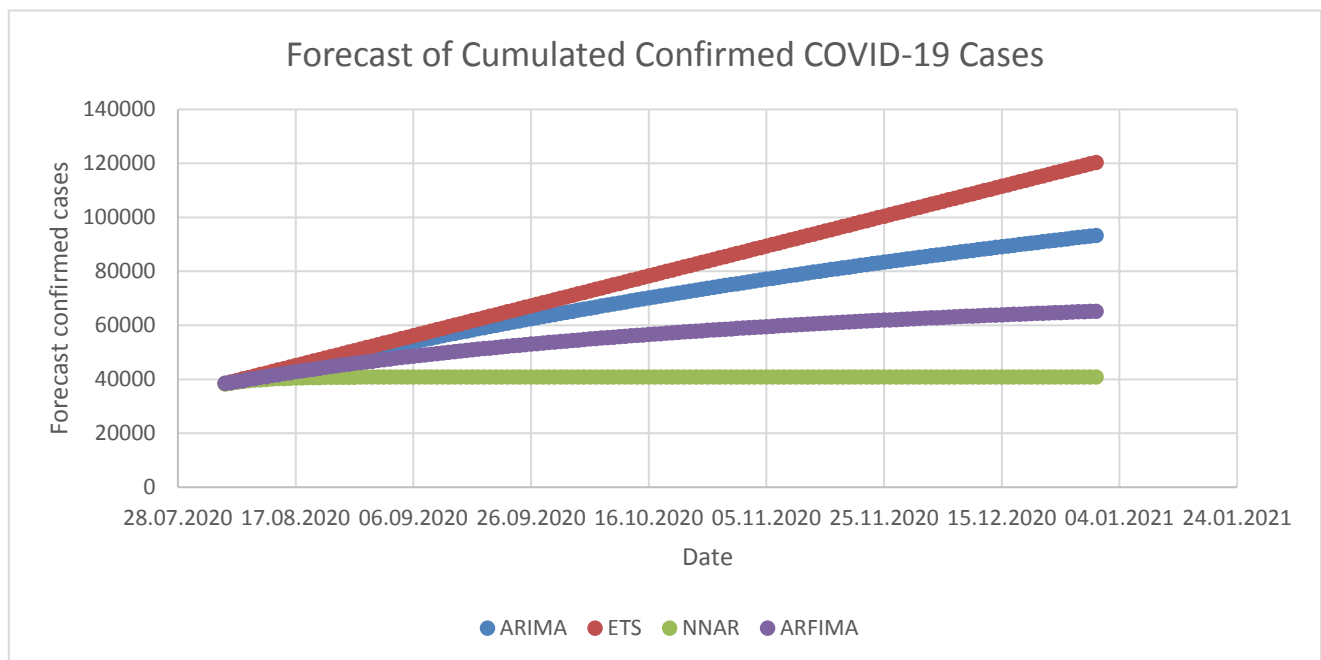


Figure 4 – Forecast plot of cumulated confirmed COVID-19 cases from four forecast techniques

The forecast of the overall "best" model that is ARFIMA (2, 0.49, 4), is slightly above NNAR. Therefore, in Table 13, we combined the forecast values of the "best" models from the four respective forecast methods. An unequal weight is estimated from the MAPE in Table 10. The MAPE of

the overall "best" forecast techniques (ARFIMA (2, 0.49, 4)) is subtracted from the other forecast techniques to get the difference (d) in Table 12.

The forecast values from the overall "best" forecast techniques (ARFIMA (2, 0.49, 4)) are similar to that of the combined model.

Table 12 – Unequal weight estimation for the combined model

| Models | MAPE | d=MAPEi-MAPEmin | | Weight |
|---------------------|----------|-----------------|------------|--------|
| ARIMA (1,1,2) | 10.37643 | 5.238451 | 0.07285927 | 0.05 |
| ETS (A, A, N) | 11.86228 | 6.724301 | 0.03466064 | 0.03 |
| NNAR (3,2) | 8.564587 | 3.426608 | 0.1802692 | 0.14 |
| ARFIMA (2, 0.49, 4) | 5.137979 | 0 | 1 | 0.78 |
| | | | 1.28778911 | |

Table 13 – Forecast cumulated confirmed COVID-19 cases by weeks from the overall best model and combined model

| Month | Weeks | ARFIMA (2, 0.49, 4) | Combined Model |
|-----------|--------|---------------------|----------------|
| August | Week 1 | 39005 | 38993 |
| | Week 2 | 41082 | 41011 |
| | Week 3 | 43476 | 43260 |
| | Week 4 | 45634 | 45257 |
| September | Week 1 | 47575 | 47058 |
| | Week 2 | 49347 | 48716 |
| | Week 3 | 50969 | 50250 |
| | Week 4 | 52464 | 51678 |
| October | Week 1 | 53846 | 53013 |
| | Week 2 | 55127 | 54265 |
| | Week 3 | 56317 | 55442 |
| | Week 4 | 57425 | 56549 |
| | Week 5 | 58458 | 57594 |
| November | Week 1 | 59423 | 58581 |
| | Week 2 | 60323 | 59515 |
| | Week 3 | 61166 | 60398 |
| | Week 4 | 61953 | 61236 |
| December | Week 1 | 62690 | 62030 |
| | Week 2 | 63378 | 62783 |
| | Week 3 | 64023 | 63498 |
| | Week 4 | 64625 | 64177 |
| | Week 5 | 65111 | 64732 |

Some studies have modelled the COVID-19 pandemic using the time series analysis methods [7, 8, 9, 10]. In this study, four competing forecasting techniques (ARIMA, ETS, NNAR, and ARFIMA) were fitted to the COVID-19 confirmed cases so that the appropriate or "best" forecasting techniques would be used to forecast the COVID-19 issues in Ghana.

Although researchers like [9, 10, 11] have used some of these techniques to model and forecast COVID-19 cases in other countries, the selection of the appropriate method was not exhaustive. Here, several competing models were constructed for each forecasting technique and the "best" model was selected to represent that technique. Eventually, the out-sample forecast performance of these respective "best" techniques are compared, and the one with the minimum error metric was selected as the overall "best" forecasting technique.

Therefore, the ARFIMA technique was selected as the overall "best" forecasting technique for the COVID-19 cases in Ghana in this study. To the best of our knowledge, this study is the first to construct time series models and specifically selecting ARFIMA techniques as the appropriate forecast technique for Ghana's COVID-19 cases.

CONCLUSION

COVID-19 pandemic has been spreading rapidly across different parts of the world, and Ghana has not been spared. This pandemic continues to cause more havoc, most especially in the economic development of the country. Hence prediction of cases is vital for stakeholders of the public and private sectors of Ghana. Therefore, this research sought to identify an appropriate statistical technique for forecasting the cumulative daily cases of Coronavirus in Ghana. Thus, four com-

peting forecasting techniques were compared to choose the proper method. Four competing forecasting techniques (ARIMA, ETS, NNAR, and ARFIMA) were applied to the COVID-19 series, from 2020-03-12 to 2020-08-04. Our findings revealed that the ARFIMA technique is the appropriate statistical technique for forecasting COVID-19 cases in Ghana. The "best" model for forecasting is ARFIMA (2, 0.49, 4) which passed

all the needed diagnostic tests. A 149-daily forecast from the "best" model revealed that the number of cases of COVID-19 will still be on the rise.

Conflict of Interest

The authors do not have any conflict of interest

REFERENCES

1. World Health Organization. (2020). *WHO Director-General's remarks at the media briefing on 2019-nCoV on February 11 2020*. Retrieved from <https://www.who.int/director-general/speeches/detail/who-director-general-s-remarks-at-the-media-briefing-on-2019-ncov-on-11-february-2020>
2. Chan, J. F.-W., Yip, C. C.-Y., To, K. K.-W., Tang, T. H.-C., Wong, S. C.-Y., Leung, K.-H., ... Yuen, K.-Y. (2020). Improved Molecular Diagnosis of COVID-19 by the Novel, Highly Sensitive and Specific COVID-19-RdRp/Hel Real-Time Reverse Transcription-PCR Assay Validated In Vitro and with Clinical Specimens. *Journal of Clinical Microbiology*, 58(5). doi: 10.1128/jcm.00310-20
3. Fernandes, N. (2020). Economic Effects of Coronavirus Outbreak (COVID-19) on the World Economy. *SSRN Electronic Journal*. doi: 10.2139/ssrn.3557504
4. Worldometer. (2020). *COVID-19 coronavirus pandemic*. Retrieved from <https://www.worldometers.info/coronavirus>
5. Johns Hopkins University. (2020). *COVID-19 Dashboard by the Center for Systems Science and Engineering at Johns Hopkins University*. Retrieved from <https://coronavirus.jhu.edu/map.html>
6. McCloskey, B., Zumla, A., Ippolito, G., Blumberg, L., Arbon, P., Cicero, A., ... Borodina, M. (2020). Mass gathering events and reducing further global spread of COVID-19: a political and public health dilemma. *The Lancet*, 395(10230), 1096–1099. doi: 10.1016/s0140-6736(20)30681-4
7. Papastefanopoulos, V., Linardatos, P., & Kotsiantis, S. (2020). COVID-19: A Comparison of Time Series Methods to Forecast Percentage of Active Cases per Population. *Applied Sciences*, 10(11), 3880. doi: 10.3390/app10113880
8. Maleki, M., Mahmoudi, M. R., Wraith, D., & Pho, K.-H. (2020). Time series modelling to forecast the confirmed and recovered cases of COVID-19. *Travel Medicine and Infectious Disease*, 37, 101742. doi: 10.1016/j.tmaid.2020.101742
9. Khan, F. M., & Gupta, R. (2020). ARIMA and NAR based prediction model for time series analysis of COVID-19 cases in India. *Journal of Safety Science and Resilience*, 1(1), 12–18. doi: 10.1016/j.jnlssr.2020.06.007
10. Yonar, H., Yonar, A., Tekindal, M., & Tekindal, M. (2020). Modeling and Forecasting for the number of cases of the COVID-19 pandemic with the Curve Estimation Models, the Box-Jenkins and Exponential Smoothing Methods. *Eurasian Journal of Medicine and Oncology*, 4(2), 160–165. doi: 10.14744/ejmo.2020.28273
11. Balah, B., & Djeddou, M. (2020). *Forecasting COVID-19 new cases in Algeria using Autoregressive fractionally integrated moving average Models (ARFIMA)*. doi: 10.1101/2020.05.03.20089615
12. Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, 20(1), 5–10. doi: 10.1016/j.ijforecast.2003.09.015
13. Hyndman, R., Koehler, A., Ord, K., & Snyder, R. (2008). *Forecasting with Exponential Smoothing*. *Springer Series in Statistics*. doi: 10.1007/978-3-540-71918-2

14. Shumway, R. H., & Stoffer, D. S. (2016). *Time Series Analysis and Its Applications* (4th ed.). Retrieved from <https://www.stat.pitt.edu/stoffer/tsa4/tsa4.pdf>