JOURNAL OF BUSINESS & ECONOMICS
Volume No. 12 (2), pp. 133–143 doi:10.5311/JBE.2020.12.12

RESEARCH ARTICLE

Identifying Optimal ARIMA Type Forecasting Model for COVID-19 Datasets of Pakistan

Mudassar Rashid[1], Shahid Akbar[2], Usman Shakoor *[3], and Sher Akbar[4]

[1,3]*Department of Economics, COMSATS University, Islamabad*
[2]International Institute of Islamic Economics, International Islamic University Islamabad
[4]Department of Statistics, Quaid-i-Azam University, Islamabad

Received: October 18, 2020; Accepted: December 12, 2020.

Abstract: The COVID-19 has adversely affected almost all countries of the world in a few months span and most of the countries are not able to manage the adverse impacts of it. There is a need for study that gives some help in management of this pandemic in future. Keeping this objective, this research evaluated the forecast performance of three ARIMA type models; Simple ARIMA, Sutte-ARIMA, and Wavelet-ARIMA for the COVID-19 data sets of Pakistan from 26 February 2020 to 13 July 2020. The data is separated into two portions: training and testing data. Training data ranges from 26th Feb, 2020 to 6th July, 2020 with 132 data points, for three series confirmed cases, deaths, and recoveries, respectively. On the other hand, testing data comprised of seven observations ranges from 7th July, 2020 to 13th July, 2020 for possible data validation process. The testing data is used to check the forecasting accuracy of the models used in this study. This study has compared three ARIMA type models predictive performances based on our data. Prediction performance for the next 7 days of testing data is applied through 3 different error measurements (MAE, MAPE, RMSE). Based on our findings, it is observed that Sutte-ARIMA performs best as compared to other two ARIMA type models, while Wavelet ARIMA performs worst. This study will help policy makers to make future management in better ways to keep away people from adverse effect of such pandemics.

Keywords: ARIMA, Wavelet-ARIMA, Sutte-ARIMA, COVID-19, Loss Functions
JEL Classification Code: B23, C53

*Corresponding author: usman.shakoor@comsats.edu.pk

1 Introduction

Since last seven months, the world is witnessing a pandemic named COVID-19 which is identified as a transmitted disease with deadly nature (*WHO* , 2020). This infectious disease, which (started its roots from city of Wuhan) was reported in Wuhan, instigated by a new virus named SARS-CoV-2reported in China (Yang et al., 2020). Aerosol transmission, human to human contact, and transmission by touch are the three basic primary ways to spread the virus (Malik et al., 2016; WHO, 2003). The COVID-19 has adversely affected over 209 countries in all regions of the world which shows that how fast it is spreading. Up till now (13thJuly, 2020) the total confirmed cases and deaths are reported as 12,769,098 and 566, 652 around the globe, respectively (WHO, 2020 b). To date, USA with 3,225, 950 confirmed cases followed by Brazil with 1,839,850and India with 878,254 are identified as the most affected nations.

In Pakistan, a first infected person was reported in Karachi on 26th February, 2020and on the same date another confirmed patient of COVID-19 was identified in capital territory of Islamabad. A few days later with increasing number of confirmed cases, government-imposed lockdown to control the spread of the disease. However, due to already weak economic conditions the COVID-19 further much affected the economy as a result the government uplifted the lockdown very soon. Soon after that the number of cases started rise exponentially and up to date Pakistan with confirmed cases 251, 625 stands as the 11th country in the world with the most number of cases (*WHO* , 2020).

To evaluate the impact of uplifting of lockdown in Pakistan, it is essential to make a forecast projection of the COVID-19 dataset. Time series data varies with time and in certain cases it varies in an abrupt way. To investigate these variable pattern of time series, different studies have been conducted to model and forecast the nature of the data. Most recently different models are proposed to model and forecast the COVID-19 datasets: a SIRD model is proposed to forecast spread of COVID-19 in France, Italy and China in a study conducted by Fanelli Piazza (2020), an ARIMA model is applied by (Benvenuto et al., 2020) in their study to forecast COVID-19, a generalized logistic growth model (GLM) is suggested to make forecast projection of COVID-19 in a study (Roosa et al., 2020) for China, a simple exponential curve is fitted to predict COVID-19 outside China by (Koczkodaj et al., 2020).

Similarly, a comparison between ARIMA and Sutte-ARIMA on the basis of short-term forecast for COVID-19 dataset of Spain is made and it is witnessed that Sutte-ARIMA achieves best forecast (A. S. Ahmar & del Val, 2020). Further, Wavelet-ARIMA and ARIMA models are compared on the basis of prediction gain for COVID-19 dataset of Italy, France, USA, UK, Spain and it is detected that Wavelet-ARIMA produces good forecast (Singh et al., 2020). Majority of the studies have used ARIMA type models to forecast the COVID-19 dataset. In view of the existing literature, ARIMA type models are used to model and predict the COVID-19 data set of Pakistan and to identify the optimal forecast model having minimum prediction error in terms of Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE) and Root Mean Squares Error (RMSE).

2 Material and Methods

2.1 Data

In order to carry out this study; data for total recoveries, confirmed cases, and deaths of COVID-19 is obtained from February 26, 2020 to July 13, 2020 from National Institute of Health (NIH)-Islamabad daily based reports (NIH, 2020). NIH works under the Ministry of Institute of Health Services of Pakistan and reports daily data of COVID-19 since February 26, 2020. The reported data comprises total confirmed cases, deaths, and recovered cases of each province of Pakistan. Our investigation is based on aggregate number of recoveries, confirmed cases, and deaths.

2.2 Forecasting Models

To analyze the COVID-19 data, this research study considers Autoregressive Integrated Moving Average (ARIMA), Sutte-ARIMA, and Wavelet-ARIMA models.

2.2.1 ARIMA Model

This model is generally denoted by ARIMA (p,d,q) where 'p', 'q', and 'd' shows the lag order of Autoregressive (AR), Moving Average (MA), and order of integration respectively. Mathematically, for series X:

$$X_t = \theta_1 X_{t-1} + \theta_2 X_{t-2} + \theta_p X_{t-p} + \varepsilon_t - \phi_1 \varepsilon_{t-1} - \cdots - \phi_q \varepsilon_q \cdots\cdots \varepsilon_t \sim WN(0, \delta^2) \quad (1)$$

$$X_t - \theta_1 X_{t-1} - \theta_2 X_{t-2} - \cdots - \theta_p X_{t-p} = \varepsilon_t - \phi_1 \varepsilon_{t-1} - \cdots - \phi_q \varepsilon_{t-q}$$

$$X_t - \theta_1 B X_t - \theta_2 B^2 X_t - \cdots - \theta_p B^p X_t = \varepsilon_t - \phi_1 \varepsilon_t - \cdots - \phi_q B^q \varepsilon_t$$

$$\theta_p(B)^p X_t = \phi_q(B)^q \varepsilon_t \quad (2)$$

$$\theta_p(B) = (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_p(B)^p), \phi_q(B) = (1 - \phi_1 B - \cdots - \phi_q(B)^q)$$

With 1st difference the ARIMA is:

$$\theta_p(1 - B)^d X_t = \phi_q(B)\varepsilon_t, \varepsilon_t \sim WN(0, \delta^2) \quad (3)$$

2.2.2 Sutte ARIMA Model

This forecasting model uses α-Sutte indicator to forecast the series under consideration on the basis of previous data of the series, (Ahmar et al., 2018). Last four observations (X$_($$t-1$),X$_($$t-2$),X$_($$t-3$),X$_($$t-4$)) provides support to forecast the data, [14]. The following equations are used to perform -Sutte indicator,

$$X_t = \frac{\alpha\left(\frac{\frac{\Delta x}{\alpha+\delta}}{2}\right) + \beta\left(\frac{\frac{\Delta y}{\beta+\alpha}}{2}\right) + \gamma\left(\frac{\frac{\Delta z}{\gamma+\beta}}{2}\right)}{3} \quad (4)$$

Where, α= X$_{t-3}$, β=X_{t-2}, γ=X_{t-3}, δ=X_{t-4}, Δx=α-δ,Δy=β-α,Δz=γ-β.
Then Sutte ARIMA forecast is defined as the average of ARIMA and -Sutte indicator, hence adding equation (2) and equation (4) after further simplifying both of the equations, so

$$(1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_p(B)^p)X_t, \phi_q(B) = (1 - \phi_1 B - \cdots - \phi_q(B)^q)\varepsilon_t$$

$$X_t - \theta_1 X_{t-1} - \theta_2 X_{t-2} - \cdots - \theta_p X_{t-p} = \varepsilon_t - \phi_1 \varepsilon_{t-1} - \cdots - \phi_q \varepsilon_{t-q}$$

$$X_t = \theta_1 X_{t-1} + \theta_2 X_{t-2} + \cdots + \theta_p X_{t-p} + \varepsilon_t - \phi_1 \varepsilon_{t-1} - \cdots - \phi_q \varepsilon_{t-q}$$

$$X_t = \theta_1 \gamma + \theta_2 \beta + \theta_3 \alpha + \theta_4 \delta + \cdots + \theta_p X_{t-p} + \varepsilon_t - \phi_1 \varepsilon_{t-1} - \cdots - \phi_q \varepsilon_{t-q}$$

and,

$$X_t = \alpha(\frac{\Delta x}{\frac{3\alpha+3\delta}{2}}) + \beta(\frac{\Delta y}{\frac{3\beta+3\alpha}{2}}) + \gamma(\frac{\Delta z}{\frac{3\gamma+3\beta}{2}})$$

$$X_t = \alpha(\frac{2\Delta x}{3\alpha + 3\delta}) + \beta(\frac{2\Delta y}{3\beta + 3\alpha}) + \gamma(\frac{2\Delta z}{3\gamma + 3\beta})$$

Now,

$$2X_t = \theta_1 \gamma + \theta_2 \beta + \theta_3 \alpha + \theta_4 \delta + \cdots + \theta_p X_{t-p} + \varepsilon_t - \phi_1 \varepsilon_{t-1} - \cdots - \phi_q \varepsilon_{t-q} + \alpha(\frac{2\Delta x}{3\alpha + 3\delta}) + \beta(\frac{2\Delta y}{3\beta + 3\alpha}) + \gamma(\frac{2\Delta z}{3\gamma + 3\beta})$$

$$X_t = (\frac{\theta_1}{2} + \frac{2\Delta z}{3\gamma + 3\beta})\gamma + (\frac{\theta_2}{2} + \frac{2\Delta y}{3\beta + 3\alpha})\beta + (\frac{\theta_3}{2} + \frac{2\Delta x}{3\alpha + 3\delta})\alpha + \frac{\theta_4 \delta}{2} + \frac{\theta_p X_{t-p}}{2} + \cdots + \frac{\theta_4 \delta}{2} + \frac{\varepsilon_t}{2} - \frac{\phi_1 \varepsilon_{t-1}}{2} - \cdots - \frac{\phi_q \varepsilon_{t-q}}{2} \quad (5)$$

The last equation is called Sutte-ARIMA model.

2.2.3 Wavelet-ARIMA Model

Wavelet is basically a mathematical approach to transform the series under investigation into a dissimilar domain in order to analyze and process the series (Dong et al., 2001).It is a localized function having mean zero and capable to analyze transient and non-periodic signals (Davidson et al., 1997; Parmar & Bhardwaj, 2013; Saâdaoui & Rabbouch, 2014; Salzberger et al., 2020) and is considered to be suitable for analyzing the nonstationary time series data. Wavelets are based on the "mother wavelet" which are defined as,

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi(\frac{t-b}{a}); a, b \in R; a \neq 0 \quad (6)$$

Here parameters 'a' and 'b' are called scaling and translation parameters measuring the degree of compression and time location of the wavelet respectively. Since the data under consideration is discrete with defined length in nature so the following time series, with decomposition into detail components and approximation, is used.

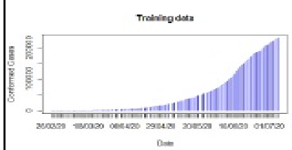
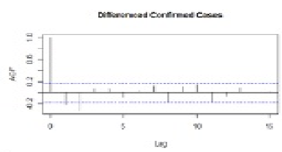
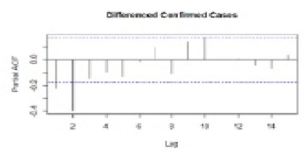
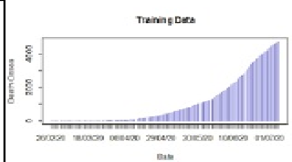
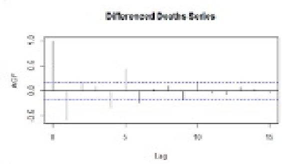
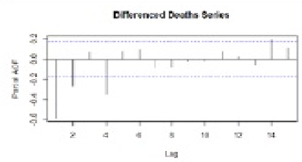
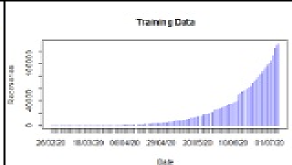
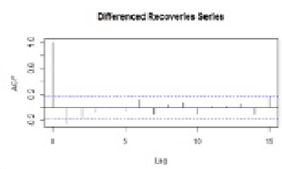
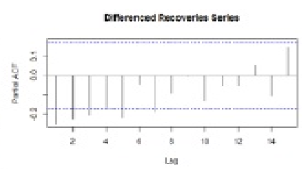
$$f(t) = \sum_{k=-\infty}^{=1+2^{j-k}} \alpha_{j,k} \varphi_{j,k}(t) + \sum_{j=1}^{j} \sum_{k=-\infty}^{=1+2^{j-k}} \beta_{j,k} \psi_{j,k}(t) \quad (7)$$

The Wavelet decomposition and ARIMA model both deals with non-linear and linear tendencies of the data features in order to model and forecast the data. For our procedure, we have used Haar filter to estimate Wavelet-ARIMA model in correspondence with study of Nguyen & Nabney (2010) and Jin & Kim (2015). They pointed out that Haar filter performs well over other filters like Symlet filter, Daubechies filter and Coiet filter. The reason is that Haar filter uses current and preceding values within a given series while other filters use succeeding values. Through this characteristic, Haar filter is considered to be a more appropriate to prediction as it does not need estimated values outside the original series while the other filters use the values from the beginning of the data series.

3 Results

This study started empirical calculations for three datasets (i.e. confirmed, Deaths and Recovered cases) with implication of classical time series ARIMA (p,d,q) model using a "forecast" package in R software (Hyndman et al., 2020). Box-Jenkins methodology ((Benvenuto et al., 2020) is implemented to identify the parameters of the ARIMA model with the help of ACF and PACF plots for possible parameters identifications, Table 1. In order to check for stationarity of the series under investigation Augmented Dicky Fuller (ADF) test (Dickey & Fuller, 1981) is used. An ARIMA model with minimum values of AIC and BIC (Akaike, 1978; Schwarz et al., 1978) is picked and identified as the best fitted model. The fitted ARIMA models for three datasets are; ARIMA (1,2,3) for confirmed cases, ARIMA (2,2,0) for death cases and ARIMA (3,2,1) for recovered cases series.

Table 1. Training datasets with their corresponding PACF and ACF plots for Confirmed, Deaths and Recovered cases of COVID-19 in Pakistan

COVID-19 cases	Training Period	ACF Plot	PACF Plots
Confirmed Cases			
Death Cases			
Recovered Cases			

Finally, for comparison purpose of predictive performance, forecasts for next 7 days are obtained for each and every series. Plots for the residuals error using ARIMA models are calculated and presented in figure 1.

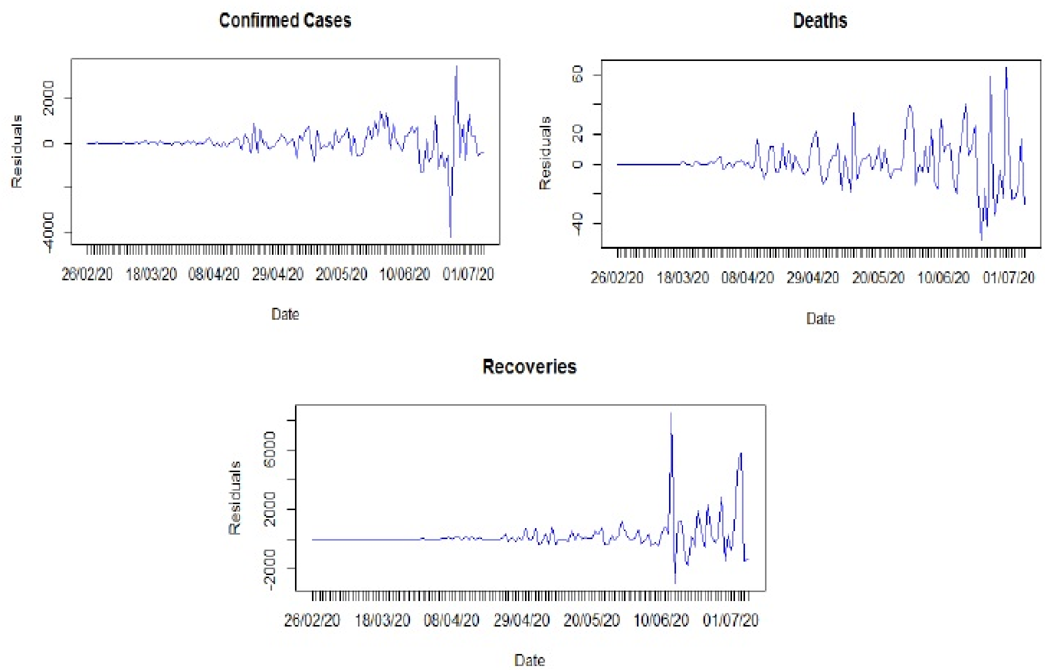


Figure 1. Plots for the residuals of ARIMA model applied on Confirmed, Deaths and Recovered cases

An interesting scenario is to see the oscillating and non-stationary residuals plots for all datasets generated through ARIMA model. These oscillations are captured through application of wavelet transformation, for decomposition of time series datasets in a linear combination manner. Thus we used wavelet transformation along with ARIMA model for comparative analysis using "TSPred" package in R software (Rebecca Eduardo, 2018). Forecasts for next 7 days are also obtained using Wavelet ARIMA model. Moreover, we have also used Sutte-ARIMA model to make our study more comprehensive. For this purpose we used "suttee Forecast R" and "Rcmndr Plugin. Suttee Forecast" packages using R version (3.5.1) (A. S. Ahmar, 2017a,b). Forecasts for 7 days ahead are obtained using Sutte-forecast algorithm for possible comparative analysis.

Table 2 shows the forecast of confirmed cases in Pakistan for the next 7 days using three different models and error measures. Results shows that Sutte-ARIMA outperformed other ARIMA type models, having evidence from all three accuracy measures (MAE, MAPE and RMSE) with lowest forecast errors.

Table 2: Comparison for the forecast error of next 7 days on confirmed cases ofCOVID-19 in Pakistan.

Date	Actual	ARIMA	Sutte-ARIMA	Wavelet-ARIMA
7/7/2020	234509	235515	235332.25	235047.2
7/8/2020	237489	239321	238462.9	238314.8
7/9/2020	240848	243066	241789.95	241582.2
7/10/2020	243599	246769	245323.15	244849.8
7/11/2020	246351	250441	248544.4	248117.2
7/12/2020	248856	254090	251706.35	251384.8
7/13/2020	251625	257722	254631.15	243254.2
	MAE	3378.07	**1787.59**	2287.83
	MAPE	0.0137	**0.0073**	0.0094
	RMSE	3791.52	**1980.57**	3436.35

Moreover, Sutte-forecast also have best predictive performance shown in table 3 applied on the deaths series with lowest error measures calculated. Table 4 shows best predictive performance of the Sutte-forecast model with lowest MAE=1266.10, MAPE= 0.0086 and RMSE=1408.99 respectively, as compared to other ARIMA and Wavelet-ARIMA models.

Table 3: Comparison for the forecast error of next 7 days on Deaths cases of COVID-19 in Pakistan

Date	Actual	ARIMA	Sutte-ARIMA	Wavelet-ARIMA
7/7/2020	4839	4837.87	4835.39	4844.25
7/8/2020	4922	4905.98	4909.46	4915.75
7/9/2020	4983	4972.8	4982.67	4987.25
7/10/2020	5058	5042.77	5050	5058.75
7/11/2020	5123	5110.69	5121.12	5130.25
7/12/2020	5197	5179.3	5184.88	5201.75
7/13/2020	5270	5247.96	5258.4	5012.25
	MAE	13.52	**7.16**	40.89
	MAPE	0.0026	**0.0014**	0.0078
	RMSE	14.85	**8.62**	97.54

Table 4: Comparison for the forecast error of next 7 days on Recovered cases of COVID-19 in Pakistan

Date	Actual	ARIMA	Sutte-ARIMA	Wavelet-ARIMA
7/7/2020	134957	135704	136789.4	135655.8
7/8/2020	140965	140906	139599.05	138933.2
7/9/2020	145311	145942	145340	142210.8
7/10/2020	149092	150704	150323.9	145488.2
7/11/2020	153134	155431	154658.05	148765.8
7/12/2020	156700	160205	158725.6	152043.2
7/13/2020	161917	164997	162770.8	143120.2
	MAE	1704.23	**1266.1**	5322.34
	MAPE	0.0111	**0.0086**	0.0342
	RMSE	2090.87	**1408.99**	7757.96

4 Discussion and Conclusion

4.1 Discussion

Our findings further strengthen the existing literature (A. S. Ahmar & del Val, 2020; Benvenuto et al., 2020; Singh et al., 2020) in favor of using ARIMA type models to forecast a pandemic series. However, among ARIMA types of models, this study founded that Sutte-ARIMA model is the best performing model having minimum loss function values as compared to other two ARIMA type models considered. This indicates that -Sutte indicator (A. Ahmar et al., 2018), as an estimator, gives precise weights to the values to be forecasted and makes Sutte-ARIMA model better than ARIMA and Wavelet-ARIMA models. While, Wavelet-ARIMA model with complicated mathematical structure, as this model depends on different filters, and maximum values of loss functions stands as the worst performer forecasting model. However, simple ARIMA model is the mediocre performer forecasting one with respect to achieved loss function values.

4.2 Conclusion

In this study, we have compared the predictive performance for three types of ARIMA models: Wavelet-ARIMA,ARIMA, and Sutte-ARIMA using earlier data for COVID-19 dataset of Pakistan. We used training data of 132 observations while prediction performance of each model is checked over 7 days ahead data points. The comparative analysis revealed that Sutte-ARIMA model performed best as compared to ARIMA and Wavelet-ARIMA modeling techniques with minimum loss functions recorded for all the three datasets of COVID-19. For confirmed cases dataset Sutte-ARIMA performed 53% better than ARIMA model while, 77% more accurate than Wavelet-ARIMA model. Moreover, for the number of Deaths series Sutte-ARIMA gave similar results as reported earlier for the confirmed cases with 53.8% and 17.9% better than ARIMA and Wavelet-ARIMA model respectively. However, for recovered cases Sutte-ARIMA gave 77.3% and 15.7% better predictive performance over ARIMA and Wavelet-ARIMA model. Therefore, our findings suggest improved performance of Sutte-ARIMA model compared to ARIMA and Wavelet-ARIMA model and can be used for predicting COVID-19 datasets over the world. The expectation of deaths due to COVID-19 through this approach will assist government officials to take more precautionary measures afore any devastating condition.

4.3 Future Research Direction

This study shows the forecasting performance of ARIMA type models with respect to Wave-I of pandemic (COVID-19). In future research the comparison of same ARIMA type forecasting models with respect to Wave-II are required to more comprehensively evaluate their performances in comparison with Wave-I.

References

Ahmar, A., Rahman, A., & Mulbar, U. (2018). α-sutte indicator: a new method for time series forecasting. In *J. phys. conf. ser* (Vol. 1040, p. 012018).

Ahmar, A. S. (2017a). Rcmdrplugin. sutteforecastr:'rcmdr'plugin for alpha-sutte indicator. Journal of Physics: Conference.

Ahmar, A. S. (2017b). Sutteforecastr: Forecasting data using alpha-sutte indicator. OSF Preprints.

Ahmar, A. S., & del Val, E. B. (2020). Suttearima: Short-term forecasting method, a case: Covid-19 and stock market in spain. *Science of The Total Environment*, 138883.

Akaike, H. (1978). A new look at the statistical model identification. *Automatica*, *19*(6), 465–471.

Benvenuto, D., Giovanetti, M., Vassallo, L., Angeletti, S., & Ciccozzi, M. (2020). Application of the arima model on the covid-2019 epidemic dataset. *Data in brief*, *29*, 105340.

Davidson, R., Labys, W. C., & Lesourd, J.-B. (1997). Walvelet analysis of commodity price behavior. *Computational Economics*, *11*(1-2), 103–128.

Dickey, D. A., & Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica: journal of the Econometric Society*, 1057–1072.

Dong, Z., Guo, X., Zheng, J., & Xu, L. (2001). Calculation of noise resistance by use of the discrete wavelets transform. *Electrochemistry communications*, *3*(10), 561–565.

Hyndman, R. J., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M., … Razbash, S. (2020). Package 'forecast'. *Online] https://cran. r-project. org/web/packages/forecast/forecast. pdf*.

Jin, J., & Kim, J. (2015). Forecasting natural gas prices using wavelets, time series, and artificial neural networks. *PloS one*, *10*(11), e0142064.

Koczkodaj, W., Mansournia, M., Pedrycz, W., Wolny-Dominiak, A., Zabrodskii, P., Strzaška, D., … Mazurek, J. (2020). 1,000,000 cases of covid-19 outside of china: The date predicted by a simple heuristic. *Global Epidemiology*, *2*, 100023.

Malik, M., Elkholy, A., Khan, W., Hassounah, S., Abubakar, A., Minh, N. T., & Mala, P. (2016). Middle east respiratory syndrome coronavirus: current knowledge and future considerations. *EMHJ-Eastern Mediterranean Health Journal*, *22*(7), 533–542.

Nguyen, H. T., & Nabney, I. T. (2010). Short-term electricity demand and gas price forecasts using wavelet transforms and adaptive models. *Energy*, *35*(9), 3674–3685.

Parmar, K. S., & Bhardwaj, R. (2013). Wavelet and statistical analysis of river water quality parameters. *Applied Mathematics and Computation*, *219*(20), 10172–10182.

Roosa, K., Lee, Y., Luo, R., Kirpich, A., Rothenberg, R., Hyman, J., … Chowell, G. (2020). Real-time forecasts of the covid-19 epidemic in china from february 5th to february 24th, 2020. *Infectious Disease Modelling*, *5*, 256–263.

Saâdaoui, F., & Rabbouch, H. (2014). A wavelet-based multiscale vector-ann model to predict comovement of econophysical systems. *Expert systems with applications*, *41*(13), 6017–6028.

Salzberger, B., Glück, T., & Ehrenstein, B. (2020). *Successful containment of covid-19: the who-report on the covid-19 outbreak in china.* Springer.

Schwarz, G., et al. (1978). Estimating the dimension of a model. *The annals of statistics*, *6*(2), 461–464.

Singh, S., Parmar, K. S., Kumar, J., & Makkhan, S. J. S. (2020). Development of new hybrid model of discrete wavelet decomposition and autoregressive integrated moving average (arima) models in application to one month forecast the casualties cases of covid-19. *Chaos, Solitons & Fractals*, *135*, 109866.

WHO. (2003). *Consensus document on the epidemiology of severe acute respiratory syndrome (sars) 2003. geneva: Who; 2003.*

WHO . (2020). .https://www.who.int/. (Accessed: 2020)

Yang, Y., Peng, F., Wang, R., Guan, K., Jiang, T., Xu, G., … Chang, C. (2020). The deadly coronaviruses: The 2003 sars pandemic and the 2020 novel coronavirus epidemic in china. *Journal of autoimmunity*, *109*, 102434.