USING TIME SERIES ANALYSIS TO FORECAST THE PREVALENCE OF THE COVID-19 PANDEMIC(IRAQ AND GULF COOPERATION COUNCIL COUNTRIES (GCC))

Prof. Dr. Obaid Mahmmood Alzawbaee, College of Administration and Economic/Cihan University Sulaimanya - Camp

Ass. Prof. Dr. AbedAli Hamad, University of Anbar

Ass. Lec. Ahmed Obaid Mahmood, Middle Technical University –Instituteof Technical Anbar

Received: 14 March 2020 Revised and Accepted: 8 July 2020

Abstract:

Corona pandemic (COVID - 19) that appeared in China in December 2019 and spread rapidly around the world. Whereas, at the time of writing this paper, its impact was more than (6,800,000) cases, and more than (380,000) deaths and the most affected countries are USA, Brazil, Russia, Spain, and UK. Nevertheless, this paper dealt with the number of cases in Iraq and the six countries of the Gulf Cooperation Council for the period from first of March to 15th of May (Series size 76). The technique of time series analysis was adopted which showed that it is nonstationary and was converted to stationary by taking the differences and a model was built for each country and used in the forecasting. Thus, the study result showed that the models that were built are effective by using it to forecast future cases and the possibility of benefiting from the proposed models to forecast the number of cases for each country by updating the data every period of time and rebuilding the models. In addition, the models clarify that the number of cases is increasing for all countries in the coming few days.

key Words: (COVID – 19), Time series, ARIMA Model, Forecasting, Pandemic.

1- Introduction:

The Corona pandemic (COVID – 19) started in December 2019 at China and within a few months, it spread rapidly around the world. And at the time of writing this paper, its impact was more than (6,800,000) cases, and death cases over (380,000). Meanwhile, the most affected countries are: USA,Brazil, Russia, Spain, UK. As for Iraq and the countries of the Gulf Cooperation Council, they are not considered as main epicenters of the pandemic and will discuss in this paper. In addition, the main aim of this study is to forecastcases size in these countries by building mixed models for time series analysis, autoregressive integrated moving average (ARIMA). {[15], [9], [10]}

Where the paper includes, additionally to the introduction, an explanation and presentation of the data used, and then testing the stationary of the series, after which the models are built. Finally, these models are usedin forecasting after diagnostic checking step. {[1], {[14], [12]}

1. Data description:

The data used in this study is the daily new cases of corona pandemic (COVID – 19) of the period from first of march to the 15th of May 2020, which has been collected from Worldoeter, Coronavirus website, and the figure number(1) represents the time series graph for each of the seven countries (research data).{[15]}

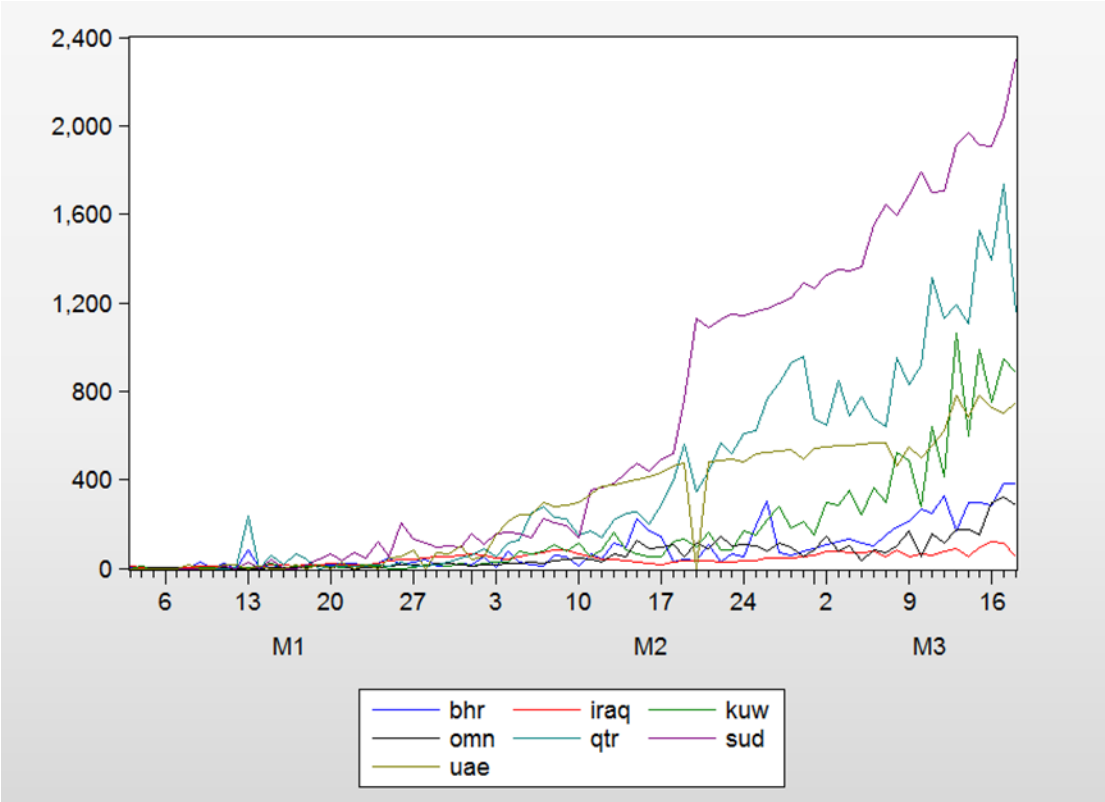


Figure: (1) Time Series Graph

From the above figure cannotice that the descriptive analysis of the data indicates that some series are nonstationary(containing a decreasing trend) clearly, while other series are not clear their stationary and will be confirmed by using Dickey-Fuller test.

3- Methods and Results:

There are many studies that tried to forecast the final size of COVID-19 cases through using many different methods over the past few months. Whereas, the steps of time series analysis are multiple and are in order; Identification which includes knowing the type of the model and determining its rank, Estimation to estimate the model parameters, Diagnostic checking, and the last step is Forecasting. However, all of these steps must be preceded by a series stationary test and the transformation of the nonstationary series into a stationary series. In this paper a new technique which is determine the model has been done ARIMA (p, d, q) where P, q = 0, 1, 2

Meanwhile, the best model will be chosen after finding therank selection criteria Akaike's Information Criterion, (AIC) and Bayesian Information Criterion, (BIC) as well as, find Root Mean Square Error, (RMSE) and Mean Absolute Error, (MAE), and take the model that matches to the minimum of these criteria, then its parameters are estimated and diagnostic checking and then used for forecasting.

The time series for the countries above are drawn as shown in Figure (1), which shows that most of the series are nonstationary and for the purpose of examining the stationarity of the series, Dickey-Fullerand Modifid for the unit root (ERS) tests were used, as it shown in table number (1). {[1], [2], [4], [5], [3], [6], [11]}

ISSN- 2394-5125 VOL 7, ISSUE 15, 2020

Table:(1) Results of ADF and ERS test for unit root

			UNIT ROOT TEST TABLE (PP)					
	At Level							
		IRAQ	BHR	KUW	OMN	SUD	UAE	QAT
With Constant	t-Statistic	-2.231	-1.120	-1.348	-0.411	2.752	-0.469	-0.452
	Prob.	0.197	0.704	0.603	0.901	1.000	0.891	0.894
		no	no	no	no	no	no	
With Constant & Trend	t-Statistic	-3.275	-4.002	-2.631	-3.814	-0.702	-5.531	-3.330
	Prob.	0.078	0.013	0.268	0.021	0.969	0.000	0.069
		*	**	no	**	no	***	
Without Constant &Trend	t-Statistic	-0.710	0.057	-1.135	0.640	4.486	1.054	0.720
	Prob.	0.406	0.698	0.231	0.852	1.000	0.922	0.869
		no	no	no	no	no	no	
	At First Difference							
		d(IRAQ)	d(BHR)	d(KUW)	d(OMN)	d(SUD)	d(UAE)	
With Constant	t-Statistic	-9.517	-15.560	-25.711	-15.504	-7.399	-20.916	-13.775
	Prob.	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		***	***	***	***	***	***	
With Constant & Trend	t-Statistic	-9.361	-21.424	-31.954	-17.330	-8.214	-25.111	-22.335
	Prob.	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		***	***	***	***	***	***	
Without Constant & Trend	t-Statistic	-9.405	-14.155	-24.333	-15.126	-6.542	-16.001	-12.884
	Prob.	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		***	***	***	***	***	***	***
			UNIT ROOT TEST TABLE (ADF)					
	At Level							
		IRAQ	BHR	KUW	OMN	SUD	UAE	QAT
With Constant	t-Statistic	-2.393	-0.466	4.719	2.035	2.053	-0.388	0.678
	Prob.	0.147	0.891	1.000	1.000	1.000	0.905	0.991
		no	no	no	no	no	no	n0
With Constant & Trend	t-Statistic	-3.390	-4.002	2.549	-0.929	-1.025	-5.523	-1.672
	Prob.	0.061	0.013	1.000	0.947	0.934	0.000	0.754
		*	**	no	no	no	***	n0
Without Constant & Trend	t-Statistic	-0.986	0.417	5.546	2.794	3.799	0.921	1.795
	Prob.	0.288	0.801	1.000	0.999	1.000	0.903	0.982
		no	no	no	no	no	no	n0
	At First Difference							
		d(IRAQ)	d(BHR)	d(KUW)	d(OMN)	d(SUD)	d(UAE)	D(QAT)
With Constant	t-Statistic	-9.164	-12.598	1.096	-15.161	-7.489	-9.293	-13.258

	Prob.	0.000	0.000	0.997	0.000	0.000	0.000	0.000
		***	***	no	***	***	***	***
With Constant & Trend	t-Statistic	-9.063	-12.673	-7.369	-15.477	-8.158	-9.293	-13.450
	Prob.	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		***	***	***	***	***	***	***
Without Constant & Trend	t-Statistic	-9.202	-12.524	1.649	-14.903	-6.571	-14.103	-12.865
	Prob.	0.000	0.000	0.975	0.000	0.000	0.000	0.000
		***	***	no	***	***	***	***
Notes: (*)Significant at the 10%; (**)Significant at the 5%; (***) Significant at the 1%. and (no) Not Significant								

From the table above it is clear that all the time series of all countries are nonstationary and it became stationary after taking the first difference (d = 1).

The series are drawn after taking the differences, and figure number(2) shows that, as it becomes clear that they are stationary.

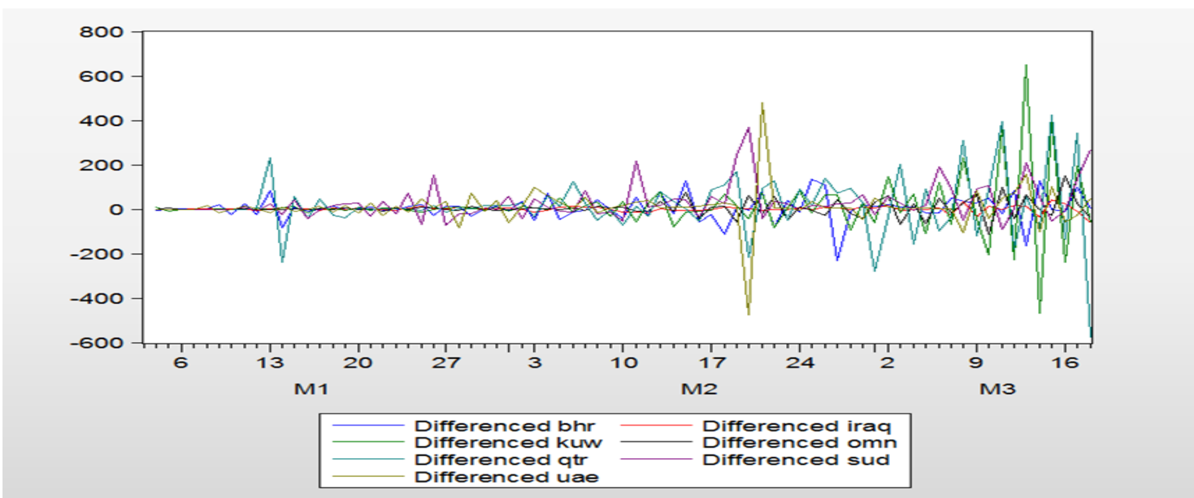


Figure: (2)The series after taking the differences

And after that, each country was studied separatelyand determined the model for each country as ARIMA (p, d, q), where:

$$p = 0, 1, 2$$

$$d = 1$$

$$q = 0, 1, 2$$

All possible models for each country are:

ARIMA (0, 1, 1), (0, 1, 2), (1, 1, 0), (1, 1, 2), (1, 1, 1), (2, 1, 1), (2, 1, 2), (2, 1, 0)

And by using the criteria for choosing a rank, two criteria were taken which are:

Akaike's Information Criterion (AIC)

ISSN- 2394-5125 VOL 7, ISSUE 15, 2020

Bayesian Information Criterion (BIC)

As well as, found:

Root Maen Square Error (RMSE)

Maen Absolute Error (MAE)

The appropriate model was chosen for each country, which corresponds to the minimum value of the above four criteria, and Table number(2) shows the results of Oman, for example.

Table: (2) The all possible models ARIMA (p, 1, q) for OMAN

ARIMA (p, 1, q)	RMSE	AIC	BIC	MAE
ARIMA(1,1,0)	31.929	521.527	7.042	19.219
ARIMA(2,1,0)	31.767	522.764	7.090	18.892
ARIMA(0,1,1)	32.277	523.1533	7.064	19.600
ARIMA(0,1,2)	31.803	522.934	7.092	18.930
ARIMA(1,1,1)	31.848	523.146	7.095	18.864
ARIMA(1,1,2)	32.022	525.963	7.163	18911
ARIMA(2,1,2)	29.743	516.889	7.073	17.892
ARIMA(2,1,1)	29.429	513.297	6.994	17.874

from theresultsof the above table it was found that the modelARIMA (2, 1, 1) is the best among the other models, because it corresponds the minimum values of (AIC, BIC, MAE, RMSE). And table number (3) shows the result summery for the seven countries.

The Optimal Parameters for ARIMA Models for all countries Table: (3)

Country	ARIMA (p, d, q)	RMSE	AIC	BIC	MAE
IRAQ	ARIMA (1, 1, 1)	13.145	390.406	5.325	8.85
QATAR	ARIMA (1, 1, 0)	122.287	722.956	9.728	80.098
AUE	ARIMA (0, 1, 1)	73.873	647.352	8.720	40.499
SAUDE	ARIMA (1, 1,0)	080.85	660.889	8.900	54.351
BAHRAYN	ARIMA (0, 1, 1)	52.384	595.790	8.032	35.710
KWAIT	ARIMA (2, 1, 2)	73.042	651.655	8.870	44.322
OMAN	ARIMA (2, 1, 1)	29.429	513.297	6.994	17.874

From the above table, and after determining the quality and rank of the models for each country, the model parameters for each of the above countries were estimated and were as follows:

IRAQ

ARIMA (1, 1, 1) $Z_t = 0.911 + 0.734Z_{t-1} + 0.998\alpha_{t-1}$ ---------- (1)

ISSN- 2394-5125 VOL 7, ISSUE 15, 2020

QATAR

ARIMA (1, 1, 0) $Z_t = 18.318 – 0.558Z_{t-1} + \alpha_t$ -------- (2)

AUE

ARIMA (0, 1, 1) $Z_t = 10.085 + 0.662\alpha_{t-1} + \alpha_t$ --------- (3)

SAUDE

ARIMA (1, 1, 0) $Z_t = 30.939 + 0.061Z_{t-2} + \alpha_t$ ------ (4)

BAHRAYN

ARIMA (0, 1, 1) $Z_t = 4.819 + 0.531\alpha_{t-1} + \alpha_t$ --------- (5)

KWAIT

ARIMA (2, 1, 2) $Z_t = 11.758 – 0.879Z_{t-1} – 0.954Z_{t-2} – 0.263\alpha_{t-1} + 0.262\ \alpha_{t-2}$ -- (6)

OMAN

ARIMA (2, 1, 1) $Z_t = 3.953 – 0.42Z_{t-1} + 0.605Z_{t-2} – 0.975\alpha_{t-1}$ -------- (7)

After we determined the model for each country and estimated its parameters, in the step of Diagnostic checking the Ljug-Box test for the model of each country has been used. Which depends on the test of significance Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) for the estimated residuals. And its formula is: {[7], [8], [13]}

$$Q = n\,(n+2)\sum_{k=1}^{h}\frac{\hat{\rho}_k^2}{n-k} \sim\chi^2\,(m-p-q) \text{ -------- (8)}$$

And table number (4) shows the results of Ljug-Box test for each model. {[16], [17]}

Table: (4) Results of Ljug-Box test

Model	Ljung-Box Q(18)		
	Statistics	DF	Sig.
IRAQ-Model_1	18.801	16	.279
QTR-Model_2	12.033	17	.798
UAE-Model_3	6.844	17	.985
SUD-Model_4	17.476	17	.423
BHR-Model_5	13.903	17	.674
KUW-Model_6	21.081	14	.100
OMN-Model_7	17.031	15	.317

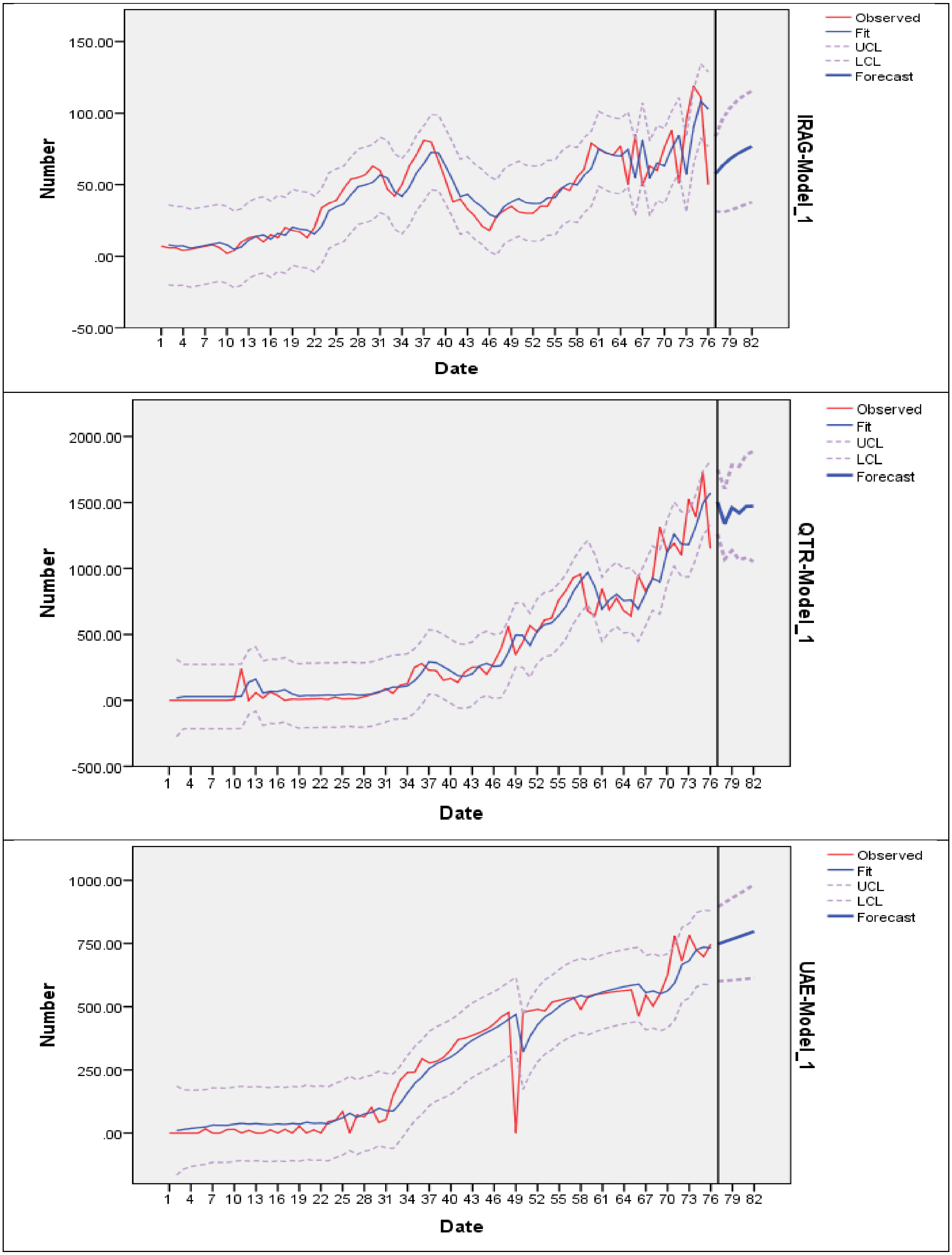
The results of the above table show that all the models are efficient. Also Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) for the estimated residuals were drawn, and its shows that there is no significant as its shown in the appendix (A).{[16], [17], [18]}

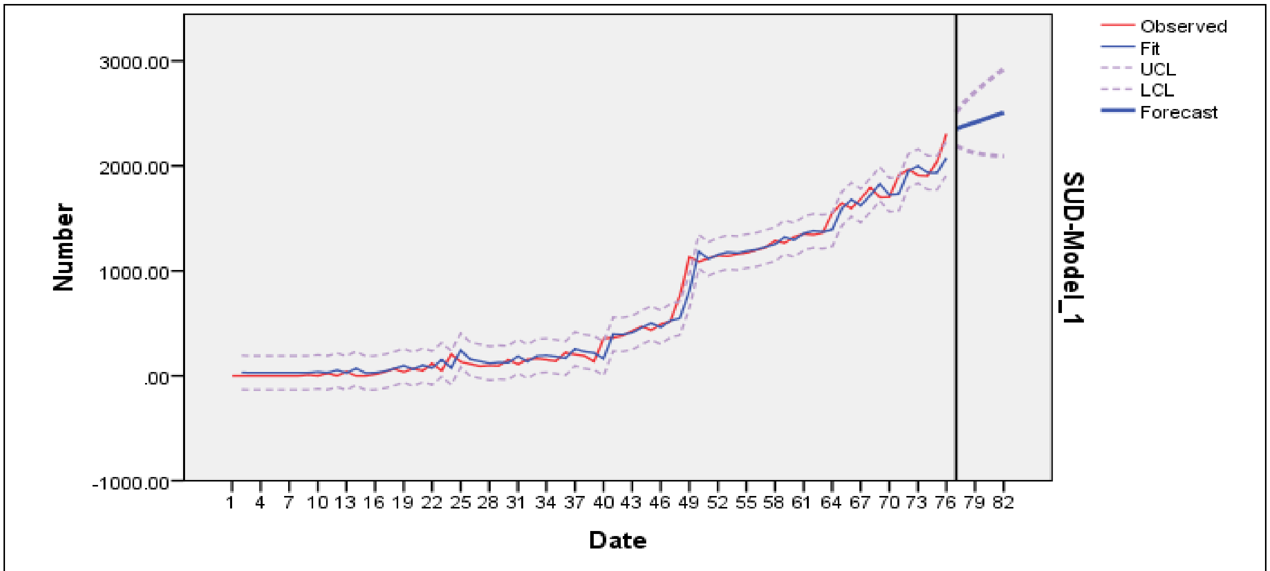
The final step which is, forecasting step by using all the models, one by one, where it forecast for the period of (6) days for each country as its shown in the following table number (5)

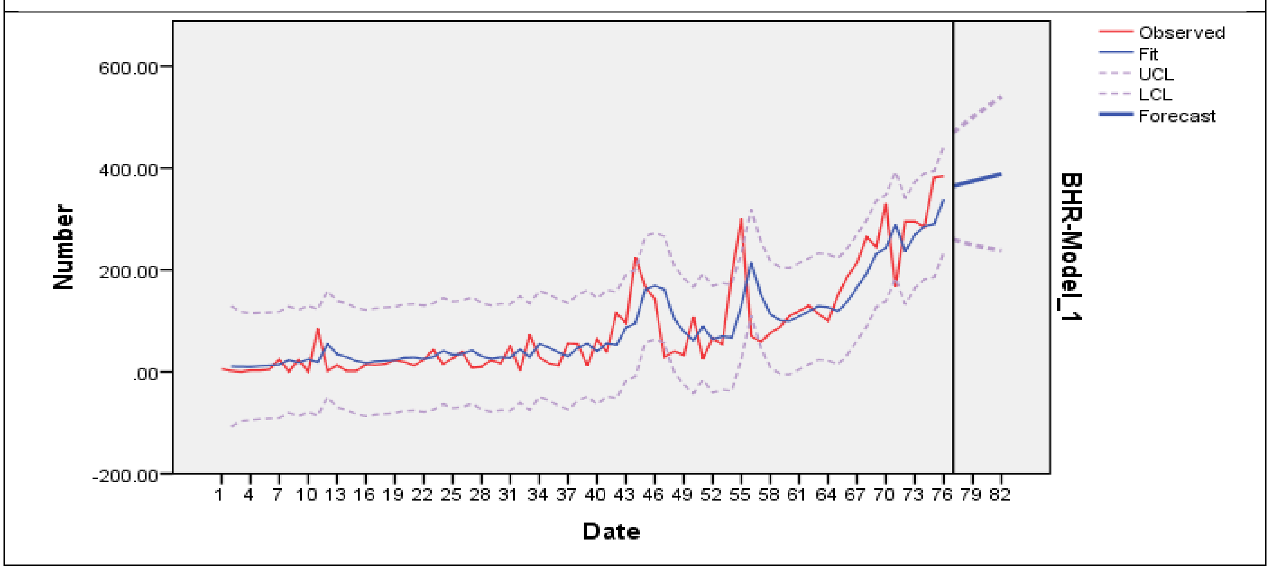
TABLE: (5)values of forecasting

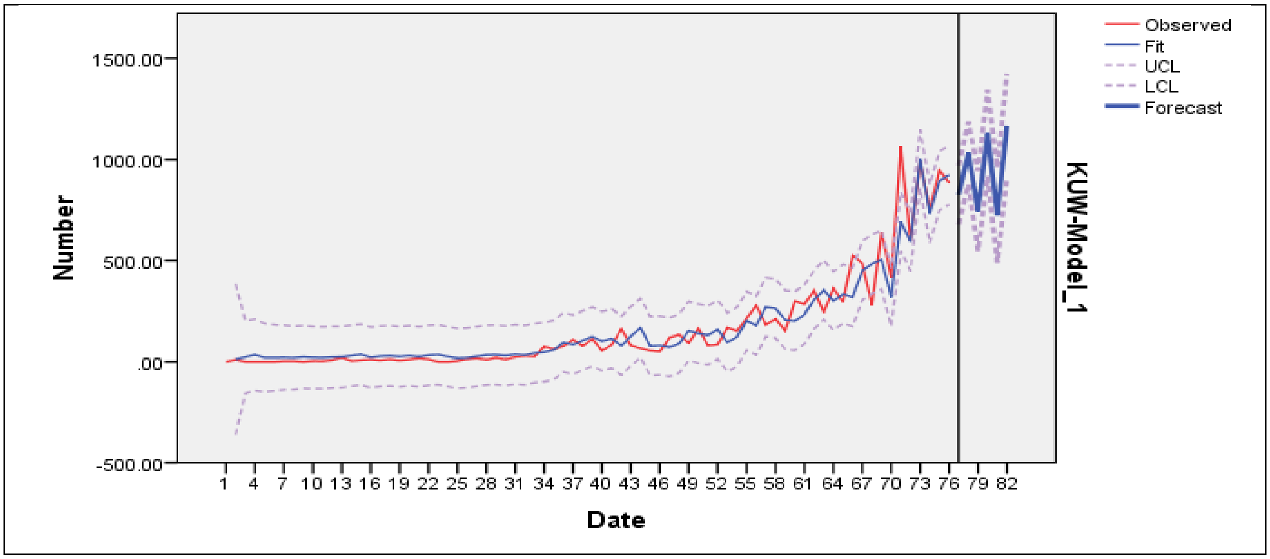
Model		77	78	79	80	81	82
IRAQ-Model_1	Forecast	64.14	70.62	75.17	77.04	79.23	80.04
	UCL	91.01	104.25	111.90	116.41	120.38	123.09
	LCL	37.27	36.99	38.43	37.66	38.09	36.99
Model		77	78	79	80	81	82
QTR-Model_2	Forecast	1505.23	1337.20	1459.52	1419.79	1470.50	1470.74
	UCL	1748.94	1603.64	1783.09	1772.85	1860.16	1888.89
	LCL	1261.52	1070.76	1135.94	1066.73	1080.84	1052.60
Model		77	78	79	80	81	82
UAE-Model_3	Forecast	747.80	757.89	767.97	778.06	788.14	798.23
	UCL	895.01	913.26	931.10	948.59	965.77	982.68
	LCL	600.59	602.51	604.84	607.52	610.51	613.78
Model		77	78	79	80	81	82
SUD-Model_4	Forecast	2352.44	2384.27	2415.26	2446.21	2477.14	2508.08
	UCL	2513.58	2619.22	2706.21	2784.00	2856.04	2924.04
	LCL	2191.31	2149.32	2124.32	2108.41	2098.25	2092.12
Model		77	78	79	80	81	82
BHR-Model_5	Forecast	364.96	369.78	374.60	379.41	384.23	389.05
	UCL	469.35	485.10	499.90	513.96	527.42	540.39
	LCL	260.56	254.45	249.29	244.87	241.04	237.71
Model		77	78	79	80	81	82
KUW-Model_6	Forecast	824.50	1032.29	744.70	1131.81	724.03	1165.81
	UCL	969.73	1187.17	943.47	1345.65	958.94	1423.70
	LCL	679.28	877.42	545.93	917.96	489.12	907.93
Model		77	78	79	80	81	82
OMN-Model_7	Forecast	301.74	311.50	298.87	322.85	308.41	326.36
	UCL	360.35	378.51	374.07	411.85	401.55	431.10
	LCL	243.13	244.49	223.67	233.85	215.26	221.62

In addition, the original and forecasting series for each country were drown, as shown in below figure number (3).


Observed
Fit
UCL
LCL
Forecast
Number
Date
IRAG-Model_1
QTR-Model_1
UAE-Model_1


Observed
Fit
UCL
LCL
Forecast
3000.00
2000.00
1000.00
.00
-1000.00
Number
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49 52 55 58 61 64 67 70 73 76 79 82
Date
SUD-Model_1


Observed
Fit
UCL
LCL
Forecast
600.00
400.00
200.00
.00
-200.00
Number
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49 52 55 58 61 64 67 70 73 76 79 82
Date
BHR-Model_1



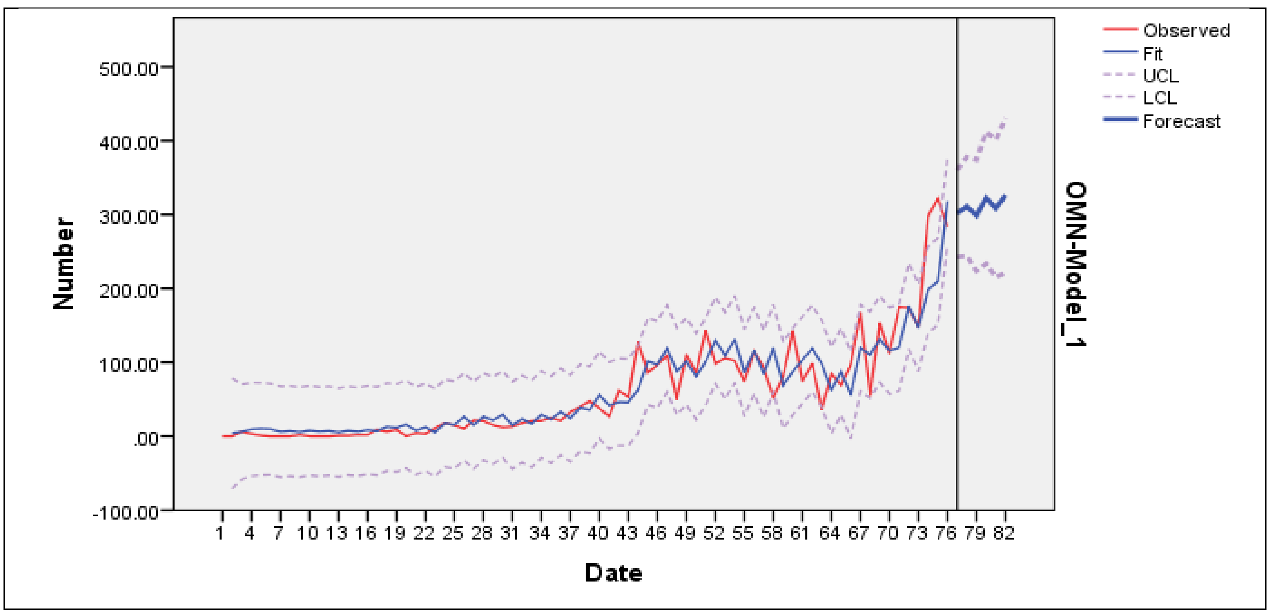


Figure: (3)values of the original and forecasting series

By looking at the figures above and observing the original series behavior thatforecasted for each country under the search,can notice the strength and efficiency of the models that were built on the one side, and the continuing number of cases increasing relatively for the next period.

4- Conclusions

Showed by the analysis that all the time series for all countries are nonstationary series and were converted to stationary series by taking the first differences.

1- The appropriate model for each country waschosen, due to,it corresponds the minimum values of (AIC, BIC, MAE, RMSE).
2- The tests have proved the efficiencyof the selected models, which were used for forecasting.
3- The models that has been builtare effective by adopting to forecast the future cases.

4- The possibility of benefiting from the proposed models to forecast the numbers of cases at each country by updating the data for each period of time and rebuilding the models in the same manner.
5- From the models, it is clear that the number of cases is increasing for all countries in the coming few days
6- It is possible by updating the models to reach inflection point, to know the expected end of the cases.
7- It is necessary to continuingin preventive measures, and social spacing to reduce the cases.

References:

[1]Alzawbaee, O. M. M., AL-Takriti, M. S. (2013), A New Suggested Technique to Determine the Best Models of Time Series. Damascus University Journal for Basic Sciences, Volume (29) - First Issue.

[2]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*, 105340.

[3]Cao, J., Jiang, X., & Zhao, B. (2020). Mathematical modeling and epidemic prediction of COVID-19 and its significance to epidemic prevention and control measures. *Journal of Biomedical Research & Innovation*, *1*(1), 1-19.

[4]Dickey D. A., Fuller W. A. (1981), Likelihood ratio statistics for autoregressive time series with a unit root, Econometrica, 49 (4), pp. 1057-1072. DOI: 10.2307/1912517.

[5]Ding, G., Li, X., Shen, Y., & Fan, J. (2020). Brief Analysis of the ARIMA model on the COVID-19 in Italy. *medRxiv*.

[6]Doornik J. A., Hansen H. (1994). An Omnibus Test for Univariate and Multivariate Normality, Working Paper, Nuffield College, Oxford University.

[7]Elliott G., Rothenberg T., Stock J. (1996), Efficient Tests for and Autoregressive Unit root, Econometrica, 64 (4), pp. 813-836. DOI: 10.2307/2171846.

[8]Engle R. F. (1982), Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, Econometrica: Journal of the Econometric Society, 50 (4), pp. 987-1007. DOI: 10.2307/1912773.

[9]Fanelli D., Piazza F. (2020), Analysis and forecast of COVID-19 spreading in China, Italy and France, Chaos, Solitons & Fractals, 134, 109761. https://doi.org/10.1016/j.chaos.2020.109761

[10]Gujarati D. N., Porter D. C. (2009), Basic Econometrics, 5th Edition, McGraw Hill Inc., New York.

[11]He Z., Tao H. (2018), Epidemiology and ARIMA model of positive-rate of influenza viruses among children in Wuhan, China: A nine-year retrospective study, International Journal of Infectious Diseases, 74, pp. 61-70. doi: 10.1016/j.ijid.2018.07.003.

[12]Hyndman R. J., Athanasopoulos G. (2018), Forecasting: principles and practice, OTexts, Melbourne.

[13]Ljung G. M., Box G. E. (1978), On a measure of lack of fit in time series models, Biometrika, 65 (2), pp. 297-303. DOI: 10.2307/2335207.

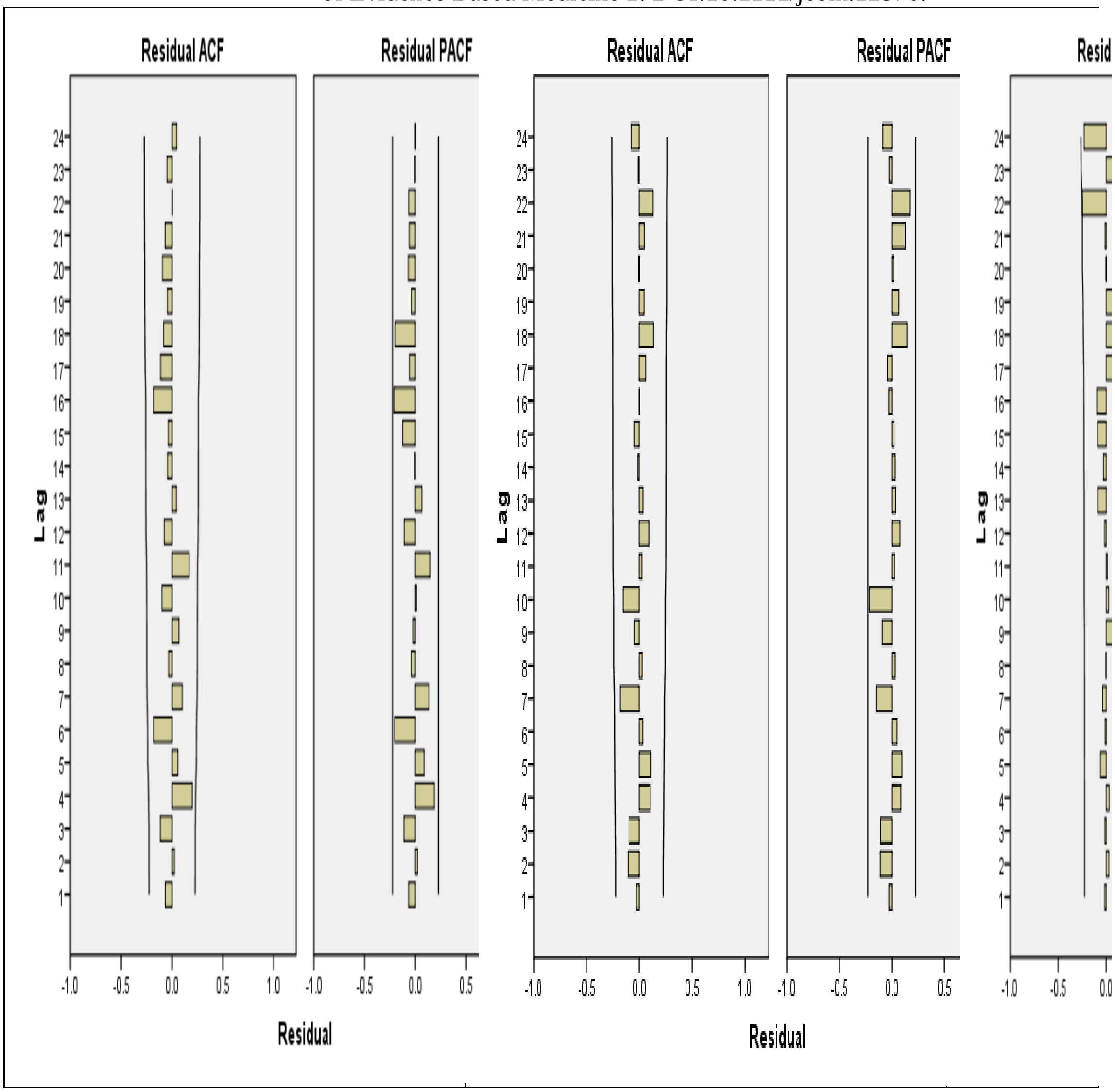
[14]Perone, G. (2020). An ARIMA model to forecast the spread and the final size of COVID- 2019 epidemic in Italy. *medRxiv*.

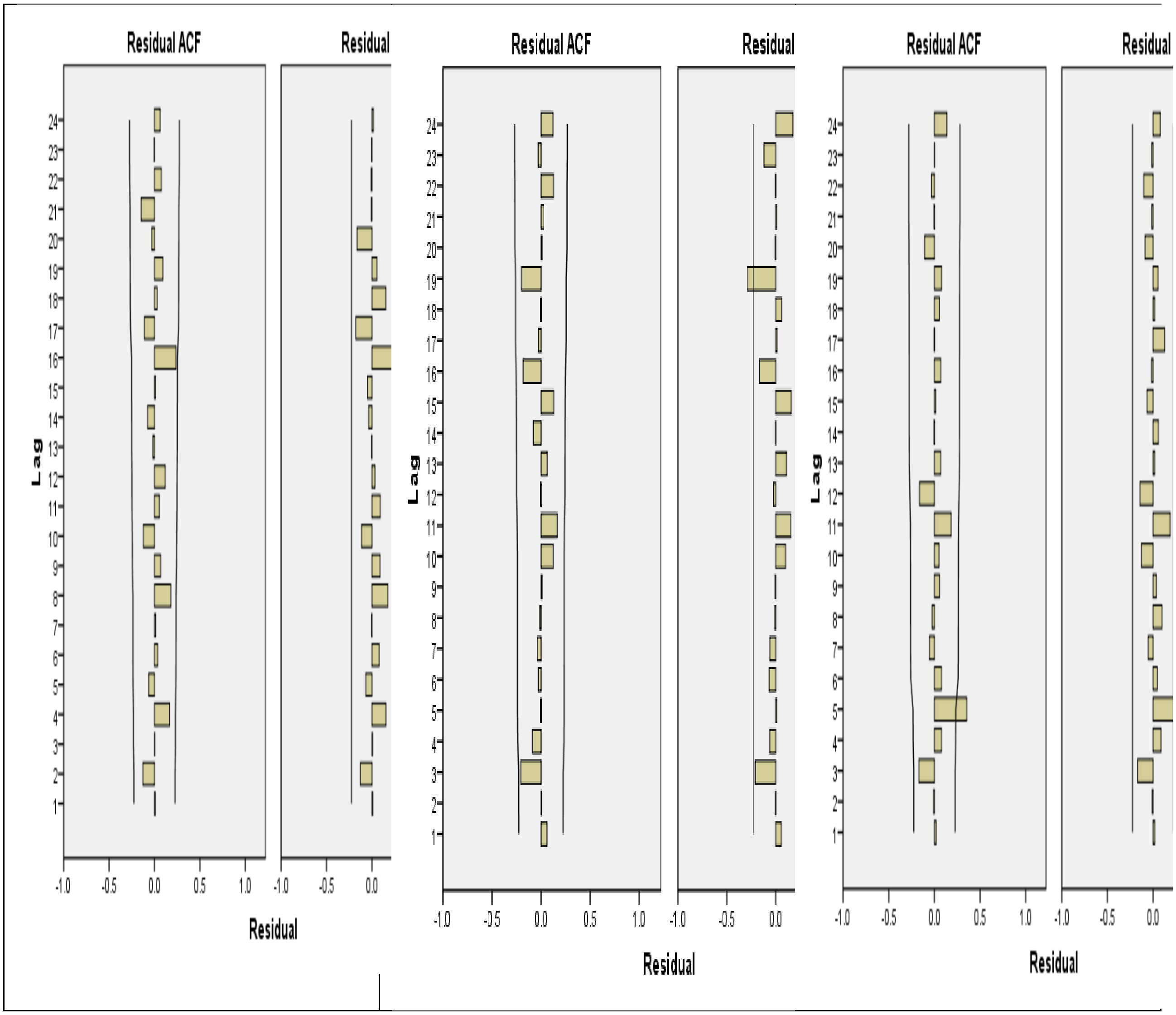
[15]Worldoeter, Coronavirushttps://www.worldometers.info/coronavirus/.

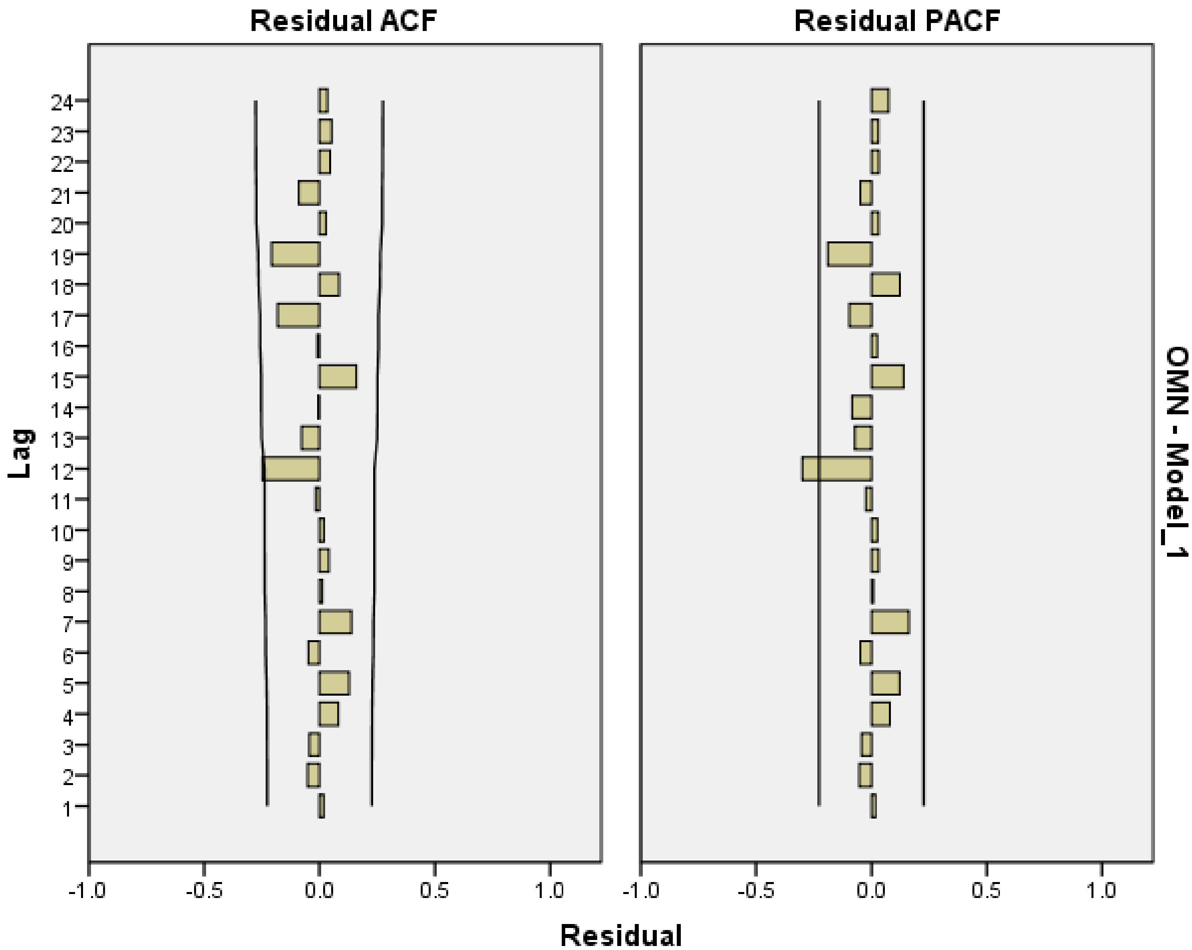
[16]Wu J. T., Leung K., Leung G. M. (2020), Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study, The Lancet, 395 (10225), pp. 689-697. doi:10.1016/S0140- 6736(20)30260-9.

[17]Zhao S., Lin Q., Ran J., Musa S. S., Yang G., Wan W., Lou Y., Gao, D., Yang L., He D., Wang M. H. (2020), Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak, International Journal of Infectious Diseases, 92, pp. 214-217. doi:10.1016/j.ijid.2020.01.050.

[18]Zhou T., Liu Q., Yang Z., Liao J., Yang K., Bai W., Xin L., Zhang, W. (2020), Preliminary prediction of the basic reproduction number of the Wuhan novel coronavirus 2019-nCoV, Journal of Evidence Based Medicine 1. DOI:10.1111/jebm.12376.




Residual ACF
Residual
Lag
Residual
Residual ACF
Residual
Lag
Residual
Residual ACF
Residual
Lag
Residual


Residual ACF
Residual PACF
Lag
24
23
22
21
20
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
-1.0
-0.5
0.0
0.5
1.0
Residual
OMN - Model_1