Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Forecasting of COVID-19 Reproduction Number by ARIMA Methodology and Quantile Estimation based on Best Fit Distribution by L- moments for Top-10 Affected Countries

Muhammad Waqas (wakas.brw@gmail.com)
Xi'an Jiaotong University Xi'an
Songhua Xu
Second Affiliated Hospital of Xi'an Jiaotong University
Linyun Zhou
Second Affiliated Hospital of Xi'an Jiaotong University

Research Article

Keywords: ARIMA, LMOM, Reproduction Number, Forecasting, Quantile Estimation

Posted Date: May 11th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-471180/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Forecasting of COVID-19 Reproduction Number by ARIMA Methodology and Quantile Estimation based on Best Fit Distribution by L-moments for Top-10 Affected Countries

1. **Author Name: Muhammad Waqas* (1st Author)**

Affiliation: School of Mathematics and Statistics, Xian Jiaotong University, Xian Shaanxi 710049 People's Republic of China, **Email: wakas.brw@gmail.com**

2. **Author Name: Songhua Xu, PhD**

Affiliation: Institute of Medical Artificial Intelligence the Second Affiliated Hospital, Xi'an Jiaotong University Xi'an, Shaanxi, 710049, China, Phone: +86 29 82666758, Email: songhuaxu@126.com

3. **Author Name: Linyun Zhou, MPH**

Affiliation: Institute of Medical Artificial Intelligence the Second Affiliated Hospital, Xi'an Jiaotong University Xi'an, Shaanxi, 710049, China, Email: feiniaolinyun@163.com

Abstract

We utilized the average weekly estimated reproduction number data of COVID-19 from March (2020-2021). By applying ARIMA and L-moments methodology, short-and-long-term forecasting of R0 is made for Govt. officials and public health experts to take before-time policy measures to control the spread of novel coronavirus. This study helps medical staff to measure the expected demand of COVID-19 vaccine doses. We applied various ARIMA models on each country's data and the best selected based on RMSE, AIC, and BIC for point and interval forecasting. Application L-Moments techniques selected GLO, GEV, and GNO distributions and quantile estimation with return period calculations. The forecasting shows that maximum countries mean R0 >1, which is still a serious threat and can lead to heath disaster. The forecasting provided an alarming situation in the coming months for India, France, Turkey, and Spain; health experts should take strict measures because the cases rise due to the high R0 forecast. The USA, Russia, and the UK mean R0 will not suddenly increase; these countries consistent in COVID-19 R0 control. We find that even the significant population differences prevail among selected countries, the R0 is still high in maximum countries, so its a dire need to take strict control actions to minimize the R0 for public safety.

Keywords: ARIMA, LMOM, Reproduction Number, Forecasting, Quantile Estimation

1. Introduction

Start to today, more than 200 countries are fighting the COVID-19 pandemic. COVID-19 originated from Wuhan, China, at the end of 2019, spread across the world, and strongly hit more than 200 countries (WHO 2021). At the end of Jan-2020, WHO announced public health emergency among all member countries, converted to COVID-19 pandemic on March 11, 2020. The pandemic of coronavirus changed the psyche of the world. The virus spread across all continents except Antarctica just within two months.

The number of suspected, infected cases and deaths started to increase in early March 2020 worldwide. Many countries began to follow the successful Chinese model to control the spread of the coronavirus in the provincial and local quarantines, tracing suspected cases, limiting movements and public activities, wearing the face mask, and preventive measures. Because pandemic China was the number 1 country facing the recorded instances that use the quarantine, local control, and preventative measure policy, it now makes it world number 55th in the series of top countries. Those countries that were late in implementing the pandemic control policies and restrictions faced big problems and more deaths. At the start, the COVID-19 hit China, Italy, Iran, the United States, South Korea, and Japan badly. Those countries that adopted policies efficiently and planned properly are now away from pandemic threats, while other countries face the issue of controlling the spread of the coronavirus. Directly these countries reporting the highest number of cases with high reproduction numbers are under discussion. Like the USA, Brazil, India, Russia, France, the UK, Italy, Spain, Turkey, and Germany are on the top 10 list globally (WHO-COVID-19-meter, March 31, 2021). All these countries were having various policies and control programs, varying populations, and different geographical locations. The spread of the virus is based on its reproduction number, which is reducible by using effective lockdown policies. The top-10 countries with the latest population, cases reported, and % population affected is list Table 1 below.

Table 1 Top-10 Countries with the highest spread rate

S. No.	Country	Continent	Population	Confirmed Cases	Pop. Affected
1	USA	North America	331,002,651	30,700,922	9.28%
2	Brazil	South America	212,559,417	12,957,597	6.10%
3	India	Asia	1,380,004,385	12,485,509	0.90%
4	Russia	Asia	145,934,462	4,529,576	3.10%
5	France	Europe	65,273,511	4,883,174	7.48%
6	UK	Europe	67,886,011	4,373,798	6.44%
7	Italy	Europe	60,461,826	3,668,264	6.07%
8	Turkey	Asia	84,339,067	3,487,050	4.13%
9	Spain	Europe	46,754,778	3,300,965	7.06%
10	Germany	Europe	83,783,942	2,895,657	3.46%
04-Apr-21					

Reproduction numbers demarcated as the mean of the number of people infected in a specific case Fraser (2007). Many mathematical models can calculate R0. Among them, two or most widely used SIR and SEIR. R0 differs according to that virus transmissibility, community type, immunity levels, population number, different countries and regions, their policies and time variability, and many other factors for every virus type. Reproduction Number R0 is used to predict at the start level, surge, and Epidemic curve phasing out stage. Average R0, if greater than 1, representing an exponential rise and spread of infection caused by the specific virus in society and a value of R0,

if less than 1, forecasting the decreasing trend and eventually decline of epidemic Hot et al., (2020). Such estimates are beneficial and help experts in planning the epidemic control measures.

The second most important contribution of R0 is to estimate the amount of vaccination required to protect the population from attaining herd immunity. The percentage of the population requiring immunization is calculated by formula $1 - 1/R0(1)$. As the policy of restrictions to stay at home, lockdown on regional, provincial, and national level, and intelligent lockdown for specific areas and many other policies adopted by many countries during this situation. Adequate reproduction number the effectual reproduction number R(t) calculated at day t, i.e., the total number of individuals infected at a specific t time would be infectious if other factors remained unchanged. At the start of COVID-19 Chinese epidemiologists and public health researchers started to work on the count and reproduction number of the coronavirus. As Li Quan et al. (2020) projected, R0 to be 2.2 (2.09–6.02) 95% confidence interval, working on the first 425 patients in Wuhan, China. Other studies estimated R0 to be 1.4–2.5, 2.68 (2.47–2.68, 95% CI), 3.6–3.8, and 6.47 (5.71–7.23, 95% CI). Ying Liu et al. (2020) found that the estimated mean R0 for COVID-19 is around 3.28, with a median of 2.79 and IQR of 1.16 by studying R0 of COVID-19 in 12 studies. Liu (2020) worked on USA data to forecast reproduction numbers by investigating the SIR model. Liu predicted that R0 would show a sudden spike right after the ease of lockdown and go slow over the period. Liao (2020) proposed a time window-based SIR model for real-time R0 forecast for seven countries. The R0 was on start measured by using the China cases record. On March 11, 2020, WHO announced the pandemic, including 110 countries under effect which are now widely spread into 200 countries. But over time, other countries crossed china in a spread of the virus. Currently, the reproduction number of viruses in the top-10 countries discussed, the data-driven from John Hopkins University data website.

A data-based predictive model known as Autoregressive integrated moving average (ARIMA) has proved helpful in forecasting short-term forecasts of dengue fever, hemorrhagic fever with renal syndrome, and tuberculosis Darapaneni (2020). ARIMA has proven to be more effective than comparable models such as the support vector machine (SVM) and wavelet neural network. Sharma (2020) uses the ARIMA technique to construct a forecasting model for nonstationary time series and forecasts COVID-19 in India. Khan et al. (2020) found ARIMA methodology consistent on NAR, investigated the number of patients based on ARIM and NAR technique for COVID-19 in India. Auto-Regressive Integrated Moving average (ARIMA) methodology application on John Hopkins Data to forecast the COVID-19 epidemiological tendencies, which are least biased and influential compared to other methodologies Benvenuto et al., (2020). Depending on retrospective data, the modified ARIMA model expert to predict the best possible outcomes. The ARIMA model uses a weighted average of historical values and error values to evaluate model predictions, making it superior to other simple regression and exponential models.

Research demonstrated that the ARIMA methodology used to forecast the occurrence of COVID-19 in the future. The study's findings helped to evaluate the pandemic dynamics and indicated the epidemiological state of the subjected countries. The estimation by applying ARIMA methodology on COVID-19 occurrence patterns in France, Italy, and Spain by Ceylon (2020). L-skewness and L-kurtosis measurements by using the L-moments methodology. Hosking (1990) provides the degree of variation of cases from a distinct bell-shaped distribution on a defined period for surveillance data to measure the inclusive degree of outbreak intensity and differentiate between the consistent and consequently expected seasonal behavior and potential outbreaks. Outbreaks included seasonal rises well perceived by coefficients L-skewness and L-kurtosis Simpson et al.,

(2020); both methodologies are widely applicable in many fields. Still, on COVID-19 data in reproduction number perspective, its application is carried here, leading to the development of estimations tools by using these methods. Financial advisors in the insurance sector also apply expected-loss models to measure the scale of claims in flood situations or health insurance. To properly value the implications of unpredictable circumstances, the insurance sector measures what are known as exceedance probability functions. Such functions produce estimates of the likelihood that casualties from an unknown occurrence, such as influenza, Ebola pandemic, reach any defined amount within a given period or not Fan et al., (2018). In a probabilistic context, it is expressed as the ratio of the losses generated by a pandemic of any given severity level multiplied by the likelihood of a pandemic of the same severity occurring in the coming year.

2. Methodology

2.1. The SIR and SEIR model

The fundamental difference between the SIR and SEIR model is that the first one contains three components as susceptible, infected, and recovered. In comparison, SEIR is derived from SIR, which has four parts: sensitive, exposed, infected and recovered. The aggregative number N = S + I + R showen in Figure 1, consisting of the components, generally grows from susceptible to infectious to recover. The application of the SIR model is on several diseases, particularly measles, rubella, and mumps, which are airborne diseases with lifetime immunity upon recovery. Thus, this model is practically applicable to the projection of infectious illnesses and human-to-human transmissible diseases after COVID-19 researchers use this model to calculate the numbers or expected patients to estimate the reproduction number of novel coronavirus.

Figure 1 Depicting SIR Model for R0

The susceptible, exposed, infectious, and recovered model (SEIR) depicted in Figure 2 is derived from the elementary SIR model. The components of this model represent the following measurements: susceptible individuals (S), those individuals who experience a long incubation duration (E), the total number of infected persons (I), and the sum of recovered patients represented by R. Therefore, the SEIR model differs from the SIR fundamentally by the amalgamation of an expectancy period. Both models provide the same number of reproduction numbers.

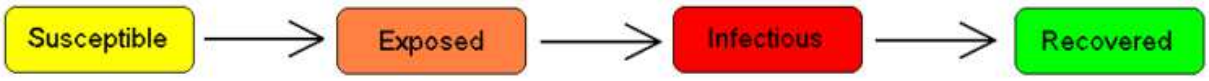

Figure 2 Depicting SIER Model for R0

2.2. Auto-Regressive Moving Average Methodology (ARIMA)

ARIMA (p, d, q) method applies lag at the 1st or 2nd level if the nonstationary problem exists in the data; otherwise, if stationary without lag, than ARMA (p, q) is an alternative method, hence p for Moving Average (MA) and q for Autoregressive (AR) order that is the number of errors lag in

ARIMA model forecast. Subtracting the initial value from the current is the most popular way to make a sequence stationary. Based on the nature of the univariate series, one or more lag is expected. Consequently, the value of d represents the smallest number of differentiation necessary to keep the series stationary, so if without differentiation, the data series is still stationary, then d = 0. The identification method began by measuring the presence of autocorrelation (ACF) and partial autocorrelation (PACF) by plotting the correlogram by Brockwell et al. (2003) and Awazuzu et al. (2008), Identification by using Jarque-Bera test for normality, Unit root test to test stationarity and Ljung-Box Q test. Then estimation of appropriate models, selecting the level of auto-regressive and moving averages based on the ACF and PACF of series. Based on the spikes and curve in the graph of ACF and PACF, the (p, q) identified and the best model is performed. After selecting the best model, forecasting is performed based on parameters (p,d,q) suggested by a suitable model. Diagnostic evaluation of forecasting involves assessing the efficacy of the currently developed model by feasible statistically relevant measures such as Akaike information criterion (AIC), Bayesian criterion (BIC), measurement of mean square error (MSE). The model with the least MSE and best AIC, BIC, is fitted for a forecast. The abovementioned ARIMA methodology is elaborated in the flowchart Figure 3 below.

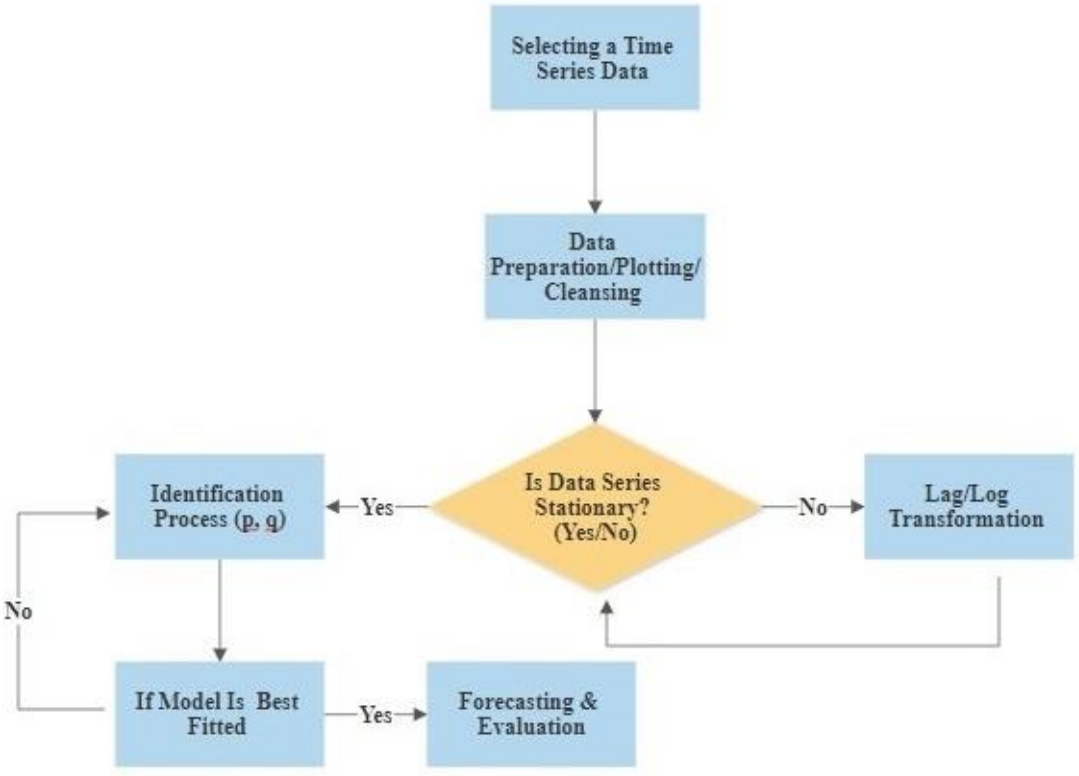

Figure 3 Flow Chart Presenting ARIMA Process

2.3. Linear Moments Technique

PWM is mainly delimited by Greenwood et al. (1979), later called LM, the linear combination of PWM given by Hosking and Wallis (1990, 1997). The PWM of order r drawn by Greenwood (1979)

$$\beta_r = \int_0^1 x(F)F^r dF$$

$x(F)$ Stands for quintile function and $F = F(x)$ cumulative distribution function. The first four L moments and PWM's are defined below,

$\lambda_1 = \beta_0$ $(\lambda_1\ location\ measure)$

$\lambda_2 = \beta_1 - \beta_0$ $(\lambda_2\ scale\ measure)$

$\lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0$ (λ_3 skewness measure)

$\lambda_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0$ (λ_4 kurtosis measure)

Where

LCv $\tau_2 = \frac{\lambda_2}{\lambda_1}$ $(coeffecient\ variation\ measure\ based\ on\ L-moments)$

LCs $\tau_3 = \frac{\lambda_3}{\lambda_2}$ $(skewness\ measure\ based\ on\ L-moments)$

LCk $\tau_4 = \frac{\lambda_4}{\lambda_2}$ $(kurtosis\ measure\ based\ on\ L-moments)$

LM is calculated by sample values, to practice unbiased estimators is an additional property b_r of PWM

β_r as:

$$b_r = \frac{1}{n}\sum_{i=r+1}^{n} \frac{(i-1)(i-2)\ldots..(i-r)}{(n-1)(n-2)\ldots..(n-r)} x_{i:n}$$

Connection of both LM and PWM is as

$$l_{r+1} = \sum_{k=0}^{r} p^*_{r,k} b_k \; r = 0,1,2,\ldots\ldots\ldots(n-1)$$

For $r = 0,1,2,3$

$l_1 = b_0$

$l_2 = b_1 - b_0$

$l_3 = 6b_2 - 6b_1 + b_0$

$l_4 = 20b_3 - 30b_2 + 12b_1 - b_0$

After an initial screening of the data, the L-moments procedure follows four steps, like ARIMA (p, d, q) methodology.

2.3.1. Selecting the Appropriate Distribution

L-moments ratio diagram LMRD, Hosing (1997), along with the use of Z-statistic. Application of simulation studies with L-skewness and L-kurtosis of at site situation or at cluster situation,

$$Z^{Dist} = \frac{(\tau_4^{Dist} - t_4^R + B_4)}{\sigma_4}$$

Where

$$B_4 = \frac{\sum_{m=1}^{N\,sim}(\, t_4^{(m)} - t_4^R)}{N}$$

$$\sigma_4 = [\frac{\{\sum_{m=1}^{N\,sim}\left(t_4^{(m)} - t_4^R\right) - N_{sim}B_4^2\}}{(N_{sim} - 1)}]^{\frac{1}{2}}$$

$$t_4^{Dist} = \; L - Cs\; of\; fitted\; distribution$$

$$B_4 = \; cluster\; bias$$

$$\sigma_4 = \; cluster\; standard\; deviation$$

$$N_{sim} = number\; of\; simulations\; to\; be\; run\; cluster \\ /country\; data\; by\; kappa\; distribution$$

Best fitted distribution selection based on the lowest $Z^{Dist} - statistic$ or closest to zero, as the $\left|Z^{Dist}\right| \leq 1.64$ is the criterion for confidence level at 90%; by using these criteria, more than one qualified distribution is drivable.

2.3.2. Estimations by Selected Distributions

For a given country i included in the cluster, the quantile estimates attainable by substituting the index of reproduction number estimates μ_i And quantile function of $q(.)$. F is the probability for quantile estimates,

$$Q_i(F) = \mu_i q(F) \qquad\qquad i = 1,2,3, \ldots\ldots\ldots\ldots N$$

where μ_i $country\; dependent\; scale\; factor$.

$Q_i(F)$ stated to the at country quantile function

$q(F)$ stated to the cluster quantile function

Let $\widehat{\mu_i}$ scale factor at country i, the cluster contains N countries, with country i with sample size n_i and observed data Q_{ij} then

$$q_{ij} = \frac{Q_{ij}}{\widehat{\mu_i}}$$

Where $i = 1,2,3, \ldots\ldots\ldots\ldots N, \quad j = 1,2,3, \ldots\ldots\ldots\ldots \; n_i$

Setting cluster mean is equal to 1 ($l_1 = 1$), by equating LMR's $\lambda, \tau, \tau_3, \tau_4$ mean to LMR of cluster $t^R,\; t_3^R,\; t_4^R$.

$$t_r^R = \frac{\sum_{i=1}^{N} n_i t^{(i)}}{\sum_{i=1}^{N} n_i} \qquad r = 3,4, \ldots \ldots$$

3. Results and discussions

Assumption of stationarity and no autocorrelation, the data is scrutinized by the software SPSS 25, E-view 10, and R-language. Augmented ducky fuller test application to all data series shown unit root and data are absent is stationary for further analysis. Just France (p-value=0.5091), which was acceptable after clearance from other tests. The Ljung Box test, autocorrelation function (ACF), and partial autocorrelation function (PACF) shown USA, Brazil, India, UK, France, and Spain were stationary at the first difference, Russia, and Italy on 2nd difference where the Turkey and Germany on zero lag were stationary. The correlogram of ACF and PACF of the USA, Brazil, India, and Russia is in Figure 4 below.

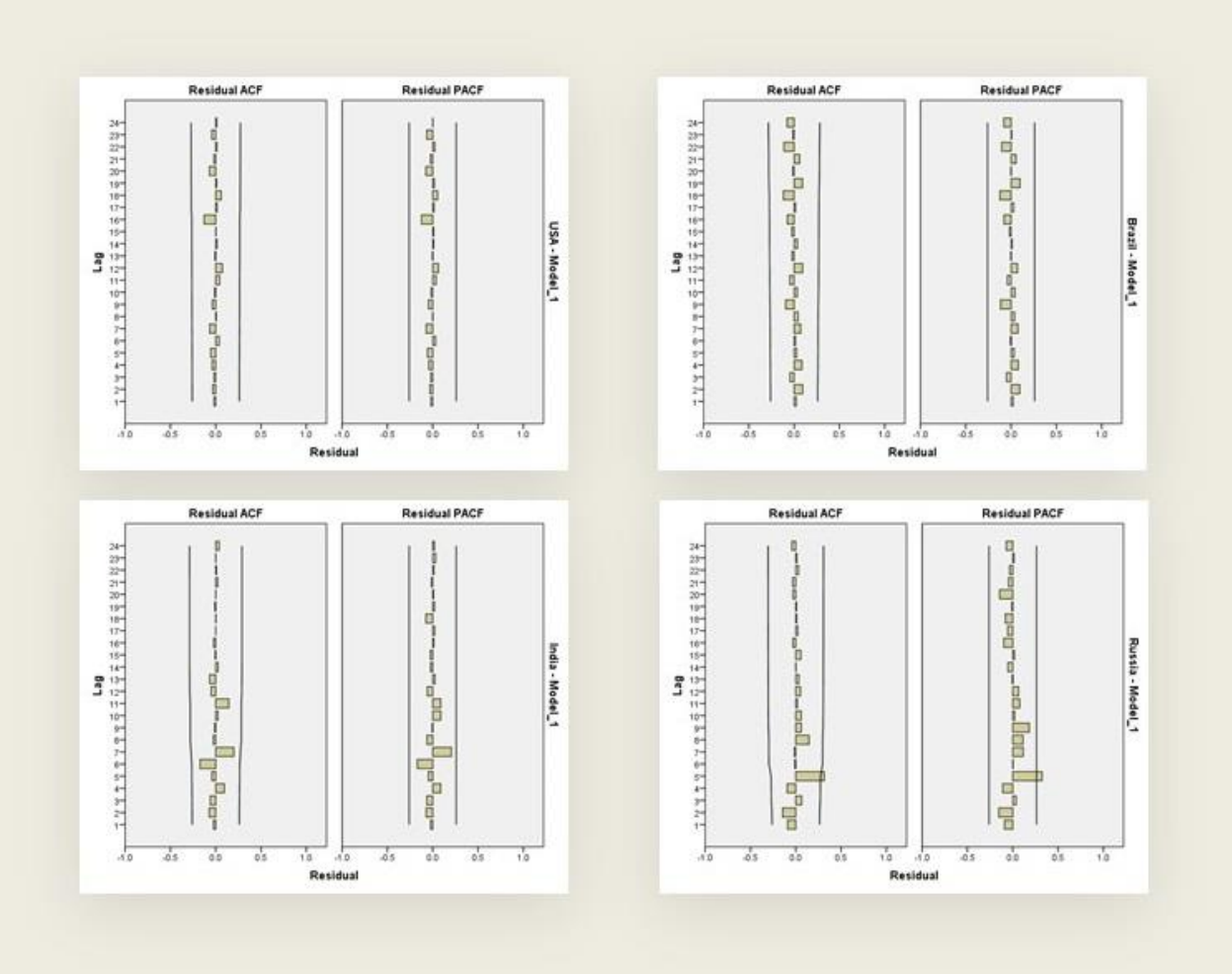

Figure 4 Correlogram of ACF and PACF for countries

3.1. Auto-Regressive Integrated Moving Average Methodology

For a comprehensive data set of top-10 countries in coronavirus cases, deaths and higher reproduction numbers (WHO, COVID-19 meter) investigated. The data of estimated reproduction number by methodology Fraser (2007), which is updated routinely from last 58 weeks starting from February-2020 to March 31, 2021, from John Hopkins University, is utilized to forecast each country in upcoming months ARIMA methodology. Selection of p & q based on the judgment from ACF and PACF. If the ACF shows a sharp cutoff or at 1st difference and autocorrelations, not positive means the model consists of MA process. The PACF charts go slow down for the AR process, indicating how many AR levels be utilized. As the variety exists in trends, spread, and control policies, and the number of COVID-19 cases in each county, different autoregressive (AR), lags, and moving averages (MA) are out-turned by each country presented in Table 2. USA best performed for at lag one without MA and AR. The UK also seems to be the same as the USA , but ARIMA (1,1,0) and ARMIA (0, 1,1) perfumed very well. The selection of the best model measured on minimum root mean square error (RMSE), Akaike information criterion (Akaike AIC, 1974), and Bayesian information criterion (BIC). Russia among three suggested models ARIMA (1,2,1),

ARIMA(2,2,1) the ARIMA (2,2,2) performed very well. Italy ARIMA (1,2,1) and ARIMA (2,2,0) performed best. Turkey AR(1) and Germany simply following MA(1) as the best model. The auto Arima function in R-3.6.3 further verified the proposed models for each country, finally approved for forecasting reproduction number.

Table 2 Identification of ARIMA(p,d,q) and Model Fitting RMSE, AIC, BIC

Country	ARIMA(p,d,q)	Sigma^2	ME	RMSE	MAE	AIC	BIC
USA	(1,1,0)	1.053	0.434	1.017	0.497	170.600	174.720
	(0,1,0)	0.309	-0.066	0.551	0.309	96.920	98.970
	(1,1,1)	0.320	-0.066	1.017	0.497	100.920	107.050
Brazil	(0,1,0)	1.299	-0.188	1.138	0.669	183.780	190.020
	(1,1,0)	0.563	0.087	0.700	0.421	110.690	112.560
	(1,1,1)	0.582	-0.128	0.756	0.732	136.980	145.670
India	**(0,1,1)**	0.124	0.079	0.320	0.185	46.440	50.530
	(1,1,0)	0.125	-0.044	0.390	0.221	47.390	62.340
	(1,1,1)	0.121	-0.060	0.345	0.221	47.820	58.910
Russia	(1,2,1)	0.822	0.037	0.907	0.533	157.290	160.370
	(2,2,1)	0.809	-0.059	0.345	0.221	27.890	31.720
	(2,2,2)	0.066	0.045	0.244	0.136	17.220	27.330
UK	(1,1,0)	0.502	-0.128	0.702	0.429	126.570	130.650
	(0,1,0)	0.506	-0.117	0.705	0.440	125.000	127.040
	(0,1,1)	0.503	-0.127	0.703	0.432	126.640	130.370
France	(1,1,0)	2.432	-0.055	1.532	1.016	215.840	219.930
	(0,1,0)	3.690	-0.029	1.004	1.213	238.180	240.230
	(0,1,1)	2.058	-0.180	1.409	0.950	207.090	211.180
Italy	(1,2,1)	0.280	0.080	0.511	0.402	92.580	98.650
	(2,2,2)	0.263	0.068	0.486	0.388	91.250	101.380
	(2,2,0)	0.258	0.070	0.490	0.386	88.160	94.240
Spain	(0,1,0)	4.856	0.028	2.184	1.230	253.840	255.880
	(0,1,1)	2.671	-0.209	1.605	0.990	222.650	226.740
	(1,1,0)	4.029	0.022	1.972	1.972	244.380	248.870
Turkey	(0,0,0)	20.560	-1.916	4.494	2.386	342.940	347.060
	(1,0,0)	19.720	-0.086	4.363	2.224	341.580	347.760
	(1,0,1)	19.420	-0.316	4.291	2.211	341.810	350.050
Germany	(0,0,0)	1.314	-3.759	1.136	0.634	184.430	187.560
	(1,0,1)	1.174	-0.055	1.055	0.584	179.140	187.380
	(0,0,1)	1.169	-0.033	1.062	0.571	177.850	184.030

The selected models-based forecasts up to the next four months are displayed. The US mean reproduction number approximately 1, where Russia is the only country with an estimated

reproduction number <1 and Spain is >2. All countries forecast depicted in Table 3 below, with point estimation of the reproduction number for coming months with confidence interval 95% lower and upper limit. France tends to be increasing in cases shortly (April-June), with an increase in reproduction is forecasted. Russia and UK reproduction numbers are forecasted to be decreasing in June and July. Turkey and Italy are also predicting a rise in cases in later months. Overall, the estimated reproduction number forecasted > 1 in these countries except Russia. India and Germany predicted (>1), especially both can face a sudden spike in cases. The countries depicting higher reproduction numbers in May-July 2021 and the reproduction number of more than 1.5 or 2 are high to cause severe threats to the countries' health systems. Those countries can take precautionary measures to control the spread by using such useful evidence-based predictions.

Table 3 Forecasting with point estimation and 95% confidence interval

Weekly Forecast for COVID-19 Reproduction Number				
Country	**Weeks**	**Predicted Value**	**95% (Low, High)**	**Trend**
USA	65th	1.02	(0.601, 4.103)	R0≈1
	69th	1.05	(0.251, 4.290)	
	73rd	1.30	(0.422, 4.465)	
Brazil	65th	1.10	(0.644, 3.696)	R0> 1
	69th	1.23	(0.101, 4.153)	
	73rd	1.15	(0.100, 4.480)	
India	65th	1.20	(0.196, 2.648)	R0> 1
	69th	1.22	(0.272, 2.724)	
	73rd	1.30	(0.343, 2.796)	
Russia	65th	0.83	(0.464, 2.694)	R0< 1
	69th	0.79	(0.190, 2.529)	
	73rd	0.73	(0.270, 2.090)	
UK	65th	1.19	(0.517, 4.317)	R0> 1
	69th	1,23	(0.457, 4.84)	
	73rd	0.90	(0.285, 4.311)	
France	65th	1.48	(0.272, 4.649)	R0> 1
	69th	1.59	(0.411, 4.690)	
	73rd	1.68	(0.445, 4.81)	
Italy	65th	1.09	(0.487, 2.968)	R0> 1
	69th	1.22	(0.803, 3.545)	
	73rd	1.34	(0.826, 3.590)	
Spain	65th	2.82	(0.915, 5.481)	R0> 2
	69th	2.83	(0.993, 5.564)	
	73rd	2.85	(0.997, 5.574)	
Turkey	65th	1.21	(0.490, 4.373)	R0> 1
	69th	1.44	(0.501, 4.566)	
	73rd	1.53	(0.537, 4.746)	

Germany	65th	1.87	(0.348, 2.990)	R0> 1
	69th	1.66	(0.352, 2.987)	
	73rd	1.49	(0.302, 2.986)	

3.2. Linear Moments Methodology

Linear Moments technique on these 10-countries data is processed, and comparable results are drawn. All countries considered in one cluster follow alike or a unique distribution in their reproduction number trend or number following different distributions selected from the family of five extreme value distributions. So, each country's weekly means reproduction number analyzed as well as cross-checked by using a cluster. Assumed as one cluster because of their previous outcomes in coronavirus cases, the performance of the reproduction number after fulfilling the basic assumptions of mean and variance consistency over time. Extreme value distributions included in linear moment methodology (LMM) are five, generalized logistic distribution (GLO), generalized normal distribution (GNO), generalized Pareto distribution (GPA), generalized extreme value distribution (GEV), and Pearson type3 (PE3) distribution, each having three parameters. Top-10 countries discordancy measure value compared with threshold value 2.76 for 10-countries considered as a cluster. But all countries discordancy value is under the limit, mostly <1 considered to be the part of one cluster. A separate analysis by treating each country as a special and unique identity was performed to check the more comprehensives of results at each country.

$Sr.No.$	$Country$	n	l_1	t	t_3	t_4	Best Dist. 1st,2nd,3rd
1	**USA**	58	1.368	0.254	0.592	0.543	GLO, GEV, GPA
2	**Brazil**	58	1.712	0.279	0.436	0.391	GLO, GEV, GNO
3	**India**	58	1.204	0.185	0.361	0.311	GLO, GEV, GNO
4	**Russia**	58	1.326	0.249	0.685	0.508	GLO, GEV, GPA
5	**UK**	58	1.461	0.312	0.512	0.411	GLO, GEV, GNO
6	**France**	58	2.334	0.333	0.303	0.189	GNO, GEV, PE3
7	**Italy**	58	1.363	0.306	0.496	0.458	GLO, GEV, PE3
8	**Spain**	58	2.632	0.267	0.302	0.275	GLO, GEV, GNO
9	**Turkey**	58	2.437	0.578	0.77	0.639	GEV, GPA, GLO
10	**Germany**	58	1.621	0.283	0.429	0.388	GLO, GEV, GNO

Table 4 Showing Length of series, average R0, L-cv, L-Kts, and best performing distributions

Z-fit and L-moments ratio diagram results are depicted separately in (Appendix Figure 1). The lowest Z-statistic criteria and the Figure 6, linear moment ratio diagram (LMRD) both are applicable to choose the most appropriate distribution Hosking (1990). The detailed analysis for each country provided the suitable distribution is generalized logistic distribution (GLO) as bet fit to perform the quantile estimates of reproduction number of COVID-19, see Table 4 and estimate the return periods, which also facilitate the probability value of exceedance. For further validation analysis, based on the lowest Z-statistic threshold, the generalized extreme value (GEV) and generalized normal distributions are selected, the Pearson type 3 (PE3) and generalized Pareto (GPA) excluded because of this minor appearance and higher Z-statistic.

Dist.	Parameters			Quantile Estimates									
	b, a, k			Q_{10}	Q_{20}	Q_{50}	Q_{100}	Q_{200}	Q_{500}	Q_{800}	Q_{900}	Q_{950}	Q_{980}
GNO	0.7281	0.3519	-1.1144	2.598	0.728	0.539	1.013	0.931	0.641	0.534	1.120	0.931	0.609
GEV	0.6501	0.2447	-0.4681	1.710	0.580	1.782	1.075	1.217	1.496	0.915	0.506	0.508	0.567
GLO	0.7537	0.2092	-0.5086	0.923	0.422	2.037	0.393	0.909	0.805	0 .671	1.032	0.652	0.803

Figure 5 Quantile estimation over various return periods

After choosing the best fit is to figure out the estimation of the quantiles for multiple return periods, shown in Figure 6 the growth curves by the suitably identified distributions. Return periods can be describable as the return time or the mean recurrence interval of an event, such as the COVID-19 reproduction number. Return period is defined in the form of probability as any period T can be called as $\frac{1}{P}$ With the probability of exceedance P. The occurrence of exceedance probability is the chance of happening of an incident in a defined period, i.e., $P = \frac{1}{T}$ Probability of occurrence. For illustration, in the situation of COVID-19 reproduction number, of 20 weeks ($\frac{1}{20} = 0.05$) can be explained as the chance of exceedance, where $(1 - \frac{1}{20} = 0.95)$ is the probability of non-exceedance.

The quantile estimates for a given cluster of 1, 2, 5, 10, 20, and 50 return periods are presented in Figure 5. Because for each country, separate analysis and top-10 countries considering cluster provided the same distribution as best fit, using finalized distributions that calculated the cluster quantile estimates. Quantile estimate for individual i^{th}country for a specific return period of reproduction number can be drawn and forecasted. The USA, which has a mean reproduction number of 1.368, can be described by multiplying the cluster quantile estimate for selected distribution. As the table values $\hat{q}_{GNO}(10)$=2.598, 1.368*2.598=3.554 predicted reproduction number once incoming ten weeks (for specific return period) with probability 0.99 of non-exceedance and exceedance probability 0.01. For France table value $\hat{q}_{GLO}(10)$=0.923, where the mean value of France reproduction number 2.334, so the predicted reproduction number one incoming ten weeks with the probability of exceedance 0.01 is 2.334*0.923=2.154. Spain $\hat{q}_{GLO}(50)$=2.037 and with mean estimated reproduction number 2.632 will 5.361 reproduction number with once in the coming year with 0.95 probability of non-exceedance if all other lockdowns, spread control policies followed. Let for India, $\hat{q}_{GNO}(10)$ =2.598 with mean reproduction number for India is 1.204, from which we can predict that R0 3.127 incoming 10 weeks with the probability of exceedance 95%. All other countries can construct the relative index similarly for the next weeks and months. Suppose a homogeneous cluster satisfies the criteria for all countries within the cluster represented by a single probability distribution holding distribution parameters jointly, so after the rescaling of country data by their at-country mean. In that case, this rescaled dimensionless probability distribution is described as a regional growth curve. The Monte Carlo simulation technique constructed on simulations to measure estimated quantiles and growth curves is applied, introduced by Hosking (1997). Monte Carlo simulations, over 10000 reiterations provided the results presndted in Table 5, of root mean square error (RMSE), relative bias (RB), relative absolute bias (RAB). Along with lower bound 0.05 and upper bound 0.95 for growth

curves, the recommended most suitable and appropriate distributions are GLO and GEV. For further modeling of COVID-19 reproduction number estimation, these distributions are suggested. These results are helpful a lot in counter and control the spread of the virus in minimum time to reduce the loss of humans' health, time, and growth.

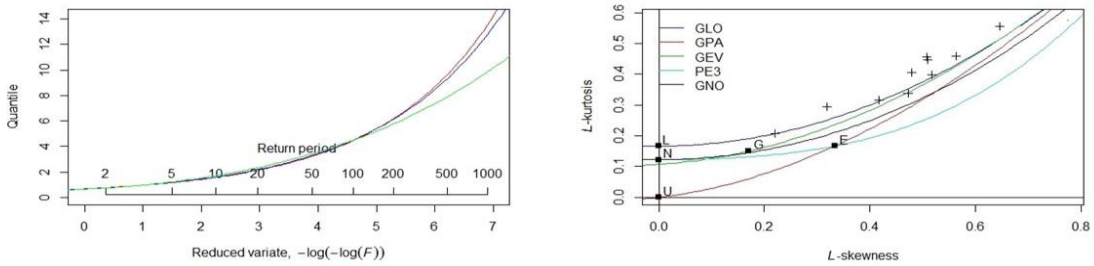

Figure 6 LMRD and return period with GLO, GEV and GNO

Table 5 Error bounds, root mean square error, and 95% CI LEB & UEB

Dist.	*F*	*1*	*2*	*5*	*10*	*20*	*50*	*100*	*500*	*1000*
	R(B)	0.321	0.195	0.125	0.080	0.048	0.027	0.016	0.014	0.030
	R(AB)	0.582	0.417	0.311	0.228	0.157	0.093	0.046	0.075	0.172
GNO	**R(RMSE)**	0.884	0.579	0.409	0.289	0.195	0.115	0.057	0.090	0.207
	LEB(0.05)	0.580	0.628	0.676	0.728	0.786	0.854	0.914	0.884	0.764
	UEB(0.95)	3.349	2.397	1.913	1.608	1.390	1.218	1.086	1.165	1.406
	R(B)	0.297	0.216	0.163	0.119	0.080	0.044	0.010	-0.020	-0.039
	R(AB)	0.572	0.425	0.323	0.239	0.164	0.094	0.040	0.075	0.168
GLO	**R(RMSE)**	0.862	0.595	0.432	0.309	0.207	0.118	0.050	0.088	0.194
	LEB(0.05)	0.567	0.640	0.701	0.759	0.817	0.876	0.918	0.852	0.714
	UEB(0.95)	3.287	2.444	1.977	1.663	1.425	1.230	1.076	1.125	1.310
	R(B)	0.309	0.216	0.156	0.111	0.073	0.040	0.011	-0.012	-0.023
	R(AB)	0.578	0.426	0.321	0.237	0.162	0.094	0.042	0.075	0.168
GEV	**R(RMSE)**	0.875	0.596	0.429	0.305	0.204	0.117	0.053	0.089	0.194
	LEB(0.05)	0.572	0.640	0.698	0.754	0.812	0.872	0.913	0.856	0.726
	UEB(0.95)	3.319	2.446	1.968	1.650	1.415	1.226	1.079	1.137	1.333

4. Conclusion

Those countries that faced a significant spike on start and took measures controlled the spread very nicely and contained reproduction numbers to a minimum. A few countries, such as China, S. Korea, and Australia, have dealt with the coronavirus

The USA, with diverse geography facing a varying situation. California and Florida are on the top among the affected states in the USA and produce many cases. Northeast states of the USA have shown a decline where transmission of the virus accelerates in other parts of the United States, particularly in the South and West. The policy was opted to stay at home, public places clubs and bars distancing, social gathering avoiding and time to time gathering to prevent a rush to reduce the spread. But health officials nominating and advising that without implementing the health benchmarks on forbidden, the public meetings and social events spread can't be reduced. The main things involved in all this virus spread prevention are the timings of people bounding at home, duration of staying at home, economic activities in specific areas, local community behavior, and local culture also incorporated in the spreading and controlling the virus. In the USA in the absence of clear federal policy in different states created the situation of chaos. There was a clear difference between the strategy of the northeast and other States. Overall, in the USA, strategies involving outperforming and causing a surge and mess in the control of the epidemic are politicization in mitigation work, mask-wearing, for example, Democratic-leaning persons cooperating more than Republicans school of thought. (Survey: by PEW Research Center) other relative factors are Salience, Climate, and Seroprevalence. Because of number 3rd in the world's largest population, the reproduction number > 1, which is estimated to be in June 1.05, is still a severe threat to the increase in cases, especially in upcoming months.

India received the first case of coronavirus at the end of January 2020 in Kerala State. India responded promptly. A health emergency imposed, capacity building of health staff, emergency control rooms designed, tracing of the cases started on early stages, isolation, quarantine, extensive care units were functional to manage the patients and reduce the disaster which was helpful to reduce the losses. World Health Organization played an influential role in mitigating outbreaks, health officials' experience, and capacity development and training them by webinars. Even having such a compelling policy, cases are rising again. The most affected states are Maharashtra, Tamil Nadu, Delhi, and Gujrat. Because of the world's 2nd largest population, India is still a serious threat like the USA, depicting R0>1, a reproduction number that is 1.204, which is predicted to rise to 2.042 in June 2021, not controlled effectively.

The case with France opted for lockdown in March, then softened it, and again opted for national lockdown in November. But the matter is the cases, and the death toll is rising in France till Match 2021 and estimated to surpass reproduction number 1.69 in July. It is now facing severe threats based on the increase in the number of cases, rather than having a not much bigger population, depicting a higher reproduction number with an increasing trend. Russia also opted for lockdown in March 2020, but after two months its opened its social activities gradually, now it is on number 4th in the world in coronavirus spread and cases. However, it is showing control in the circumstances in upcoming months prediction based on reproduction number. Turkey portrayed the highest reproduction number in November 2020, which is still accumulating its cases. Spain has an estimated average number of>2, making it seriously threatening that the issues can arise in the coming months, with a mean R0 surpassing 2.85 in July 2021.

L-moments methodology for quantile estimation on weekly mean estimated COVID-19 reproduction number provided analogous estimates to ARIMA forecasting. Consisting on

following four steps of the method, the discordancy measure in which all countries satisfied the threshold criteria. The discordancy measure value for critical countries like France, Spain, and the UK was also within limits. The heterogeneity H measure of the 10-countries provided the best results. The value of H less than or equal to 1 shown the cluster homogeneity perfectly acceptable to fit the distribution and country-wise analysis by considering each country as single-identity has done. Because this methodology depicted matching results, selected outcomes are presented here. L-skewness and L-kurtosis out turned the Z-statistic and LMRD the best-suited fit as delivered by the whole cluster. Among five distributions, GLO, GNO, GEV, GPA, and PE3, three distributions using the Goodness-of-fit measure selected as the best fit for all countries in a cluster to predict the COVID-19 reproduction number for the Year 2021 and so on. Based on the growth curve, RMSE, relative bias, and relative absolute bias, the GLO and GEV distributions are nominated for other quantile estimations. They recommended the GLO distribution as a priority, GEV distribution on a second, and GNO distribution on a third.

These recommended models, forecasting by ARIMA methodology and distributions by L-moments with quantile estimates and retune periods of mean reproduction number, addS to the development of new control applications, amendment in policies, and more detailed insights further planning and control of the virus. Both methodologies are comparable to each other, supporting both techniques and forecasting's authenticity and accuracy. Although L-moments got on edge to ARIMA forecasting because of the power weighted months built-in functions and short to long term return periods with the probability of exceedance. L-moments also have an edge because of their more weightage to lower values and smaller weightage to more significant values to bring the order in data, converting this technique to a more robust and most minor error in more significant period estimations. But both methodologies apply to any country's estimated reproduction number for forecasting if facing a similar scenario. The application of both methods covered the short- and long-term reproduction number forecasting's accuracy. So, every country should take precautionary measures accordingly, strict action against violations to control the spread, minimize R0 to zero and reduce the damages to public health, education, economies, and growth of the society. These recommendations play a vital role in policy development to reduce R0 and add inns public health matters to take precautionary measures before time.

Acknowledgment

I am grateful to the Institute of Medical Artificial Intelligence, The Second Affiliated Hospital, Xi'an Jiaotong University, Xi'an, China, for funding support. I am also grateful to the University of WAH for supporting my research and logistic support.

Declaration of competing interest

No competing interests are in this study.

References:

1. Benvenuto, D., et al., The global spread of 2019-nCoV: a molecular evolutionary analysis. Pathog Glob Health. 114:64–7, (2020). DOI: 10.1080/20477724.2020.1725339
2. Benvenuto, D., et al., Application of the ARIMA model on COVID19 epidemic dataset. Data in Brief. Volume 29, 105340, (2020).
3. Brockwell et al., Time-series: Theory and methods, (2003). ISBN 978-1-4419-0320-4

4. Ceylan Z. Estimation of COVID-19 prevalence in Italy, Spain, and France. The Science of the total environment, 729, 138817, (2020).
5. Dickey, D.A. and W.A. Fuller, Distribution of the estimators for autoregressive time series with a unit root. Journal of the American statistical association. 74(366a): p. 427-431. (1979)
6. Farhan Mohammad Khan, Rajiv Gupta. ARIMA and NAR-based prediction model for time series analysis of COVID-19 cases in India, Journal of Safety Science and Resilience, Volume 1, Issue 1, Pages 12-18, (2020). ISSN 2666-4496
7. Fraser, C. Estimating individual and household reproduction numbers in an emerging epidemic. PLOS ONE 2 (8), (2007).
8. Fan Y, Jamisonb T & Lawrence H., Pandemic risk: how large are the expected losses?. Bull World Health Organ 96:129–134, (2018) DOI: http://dx.doi.org/10.2471/BLT.17.199588
9. H. Akaike, "A new look at the statistical model identification," in *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716-723, December 1974, DOI: 10.1109/TAC.1974.1100705
10. Hotz, T., et al., (2020). Monitoring the spread of COVID-19 by estimating reproduction numbers over time.
11. Hosking J.R.M, Wallis J.R, (1997) Regional frequency Analysis Book, An approach based on L moments, Cambridge University Press.
12. Hosking, J.R.M., "L-moments: Analysis and estimation of distributions using linear combinations of order statistics," Journal of Royal Statistical Society, Series B; 52, 105-124, (1990).
13. Johns Hopkins University Center for Systems Science and Engineering (2019) https://github.com/CSSEGISandData/COVID-19
14. Johns Hopkins University Center for Systems Science and Engineering (2019). https://stochastik-tu-ilmenau.github.io/COVID-19/#ref2
15. Liao, Z., Lan, P., Liao, Z. et al. TW-SIR: time-window based SIR for COVID-19 forecasts. Sci Rep 10, 22454 (2020). https://doi.org/10.1038/s41598-020-80007-8
16. Liu, M., Thomassen, R. & Yao, S., Forecasting the spread of COVID-19 under different reopening strategies. Sci Rep 10, 20367. (2020). https://doi.org/10.1038/s41598-020-77292-8
17. Liao X, Wang B, Kang Y., Novel coronavirus infection during the 2019-2020 epidemic: preparing intensive care units-the experience in Sichuan Province, China. Intensive Care Med. 46:357–60, (2020). DOI: 10.1007/s00134-020-05954-2
18. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. N Engl J Med. 382:1199–207, (2020). DOI: 10.1056/NEJMoa2001316
19. Liu, T.; Hu, J.; Kang, M.; Lin, L.; Zhong, H.; Xiao, J.; He, G.; Song, T.; Huang, Q.; Rong, Z.; et al. Transmissiondynamics of 2019 novel coronavirus (2019-nCoV). bioRxive, (2020). doi:10.1101/2020.01.25.919787
20. Liu Y, Gayle AA, Wilder-Smith A, Rocklöv J., The reproductive number of COVID-19 is higher compared to SARS coronavirus. J Travel Med. 27:taaa02, (2020). DOI: 10.1093/jtm/taaa021
21. Moss, R., Wood, J., Brown, D., Shearer, F. M., Black, A. J., Glass, K. McVernon, J. Coronavirus disease model to inform transmission-reducing measures and health system

preparedness, Australia. Emerging Infectious Diseases, 26(12), 2844-2853, (2020). https://dx.doi.org/10.3201/eid2612.202530.
22. N. Darapaneni, D. Reddy, A. R. Paduri, P. Acharya and H. S. Nithin, "Forecasting of COVID-19 in India using ARIMA model," 2020 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), New York, NY, USA, 2020, pp. 0894-0899, DOI: 10.1109/UEMCON51285.2020.9298045.
23. Read, J.M., Bridgen, J.R.E., Cummings, D.A.T., Ho, A., Jewell, C.P. Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions. MedRxiv, Version 2, 01/28/2020
24. Read JM, Bridgen JRE, Cummings DAT, Ho A, Jewell CP. Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions. medRxiv [Preprint] (2020). DOI: 10.1101/2020.01.23.20018549
25. R. R. Sharma, M. Kumar, S. Maheshwari, and K. P. Ray, "EVDHM-ARIMA-based time series forecasting model and its application for COVID-19 cases," in *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1-10, 2021, Art no. 6502210, DOI: 10.1109/TIM.2020.3041833
26. Simpson, R.B., Zhou, B. & Naumova, E.N., Seasonal synchronization of foodborne outbreaks in the United States, 1996–2017. Sci Rep 10, 17500, (2020). https://doi.org/10.1038/s41598-020-74435-9
27. Singh, R. K., et., al., Prediction of the COVID-19 Pandemic for the top 15 affected countries: advanced autoregressive integrated moving average (ARIMA) model. JMIR public health and surveillance, 6(2), e19115, (2020).
28. Tang B, Wang X, Li Q, Bragazzi NL, Tang S, Xiao Y, et al. Estimation of the transmission risk of the 2019-nCoV and its implication for public health interventions. J Clin Med.9:462, (2020). DOI: 10.3390/jcm9020462
29. Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modeling study. Lancet. 395:689–97, (2020). DOI: 10.1016/S0140-6736(20)30260-9
30. WHO (2020). Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19)
31.

Figures

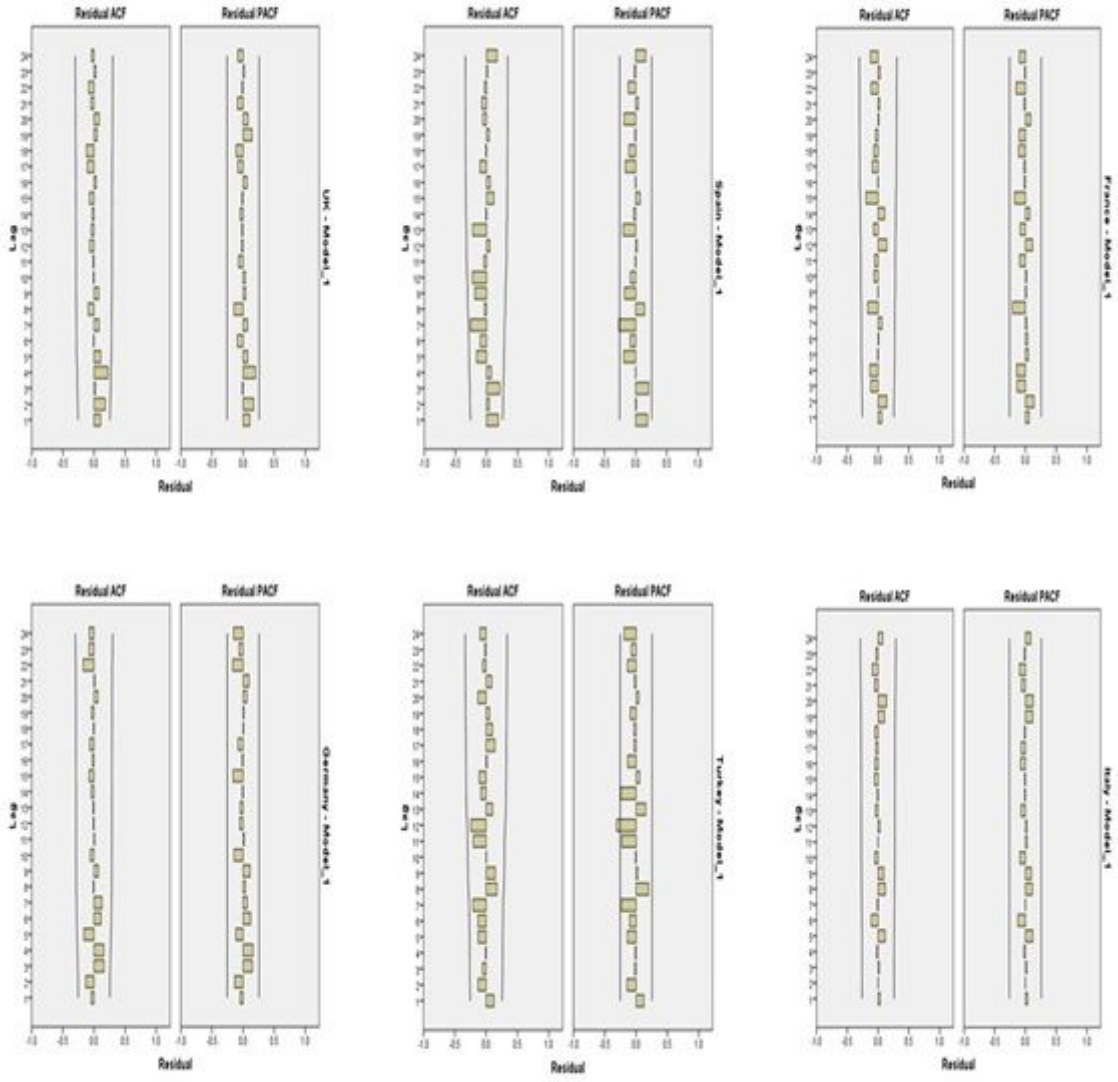

Figure 1

ACF and PACF of Six Countries

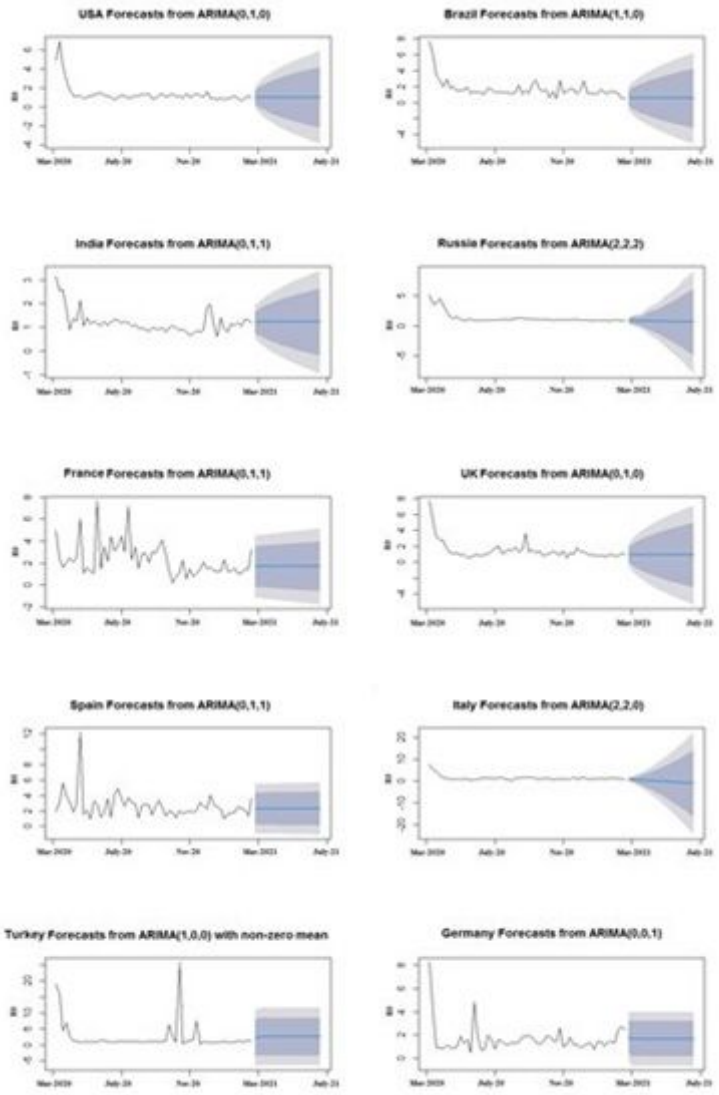

Figure 2

ARIMA Forecasting Selected Models

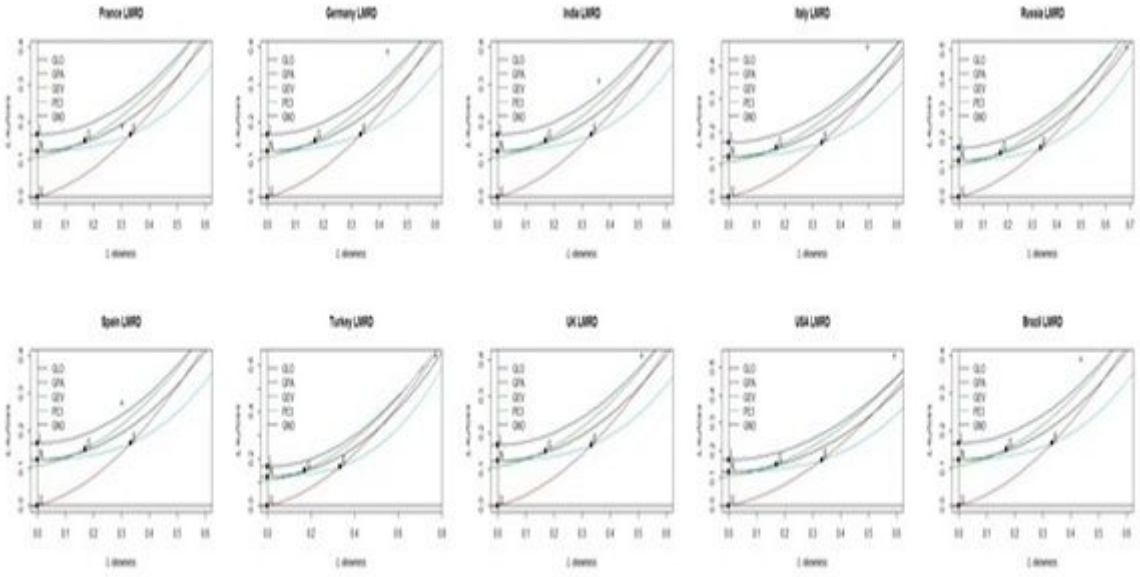

Figure 3

LMRD for Best Fit Distribution