Journal of New Frontiers in Healthcare and Biological Sciences

ISSN 2719-2385 (Online)

Volume 1, Issue 1, 2020, 1-14

Home page: www.imathm.edu.lk/publications

Modeling COVID -19 Epidemic of USA, UK and Russia

K.M.U.B. Konarasinghe
Institute of Mathematics and Management, Sri Lanka

udaya@imathm.edu.lk

ABSTRACT

A novel coronavirus was found in a seafood wholesale market in Wuhan, China. The World health organization (WHO) officially named this coronavirus as COVID-19. The disease had spread well outside China. An increase in confirmed cases is very high in the USA, UK, and Russia. Therefore, the purpose of the study was to forecast the number of infected cases in the USA, UK, and Russia. The daily confirmed cases of COVID-19 of the three countries for the period of 22nd January 2020 to 28th May 2020 were obtained from the WHO database. The Autoregressive Integrated Moving Average (ARIMA), Autoregressive Distributed Lag Models (ADLM), and Double Exponential Smoothing (DES) were tested. The Anderson–Darling test, Auto-Correlation Function (ACF), and Ljung-Box Q (LBQ)-test were used as the goodness of fit tests in model validation. The best-fitting model was selected by comparing relative and absolute measurements of errors. The ARIMA did not satisfy the model validation criterion for any of the countries, but the ADLM and DES did. It is concluded that the ADLM is the most suitable model for forecasting the USA and the DES is the best model for the UK. However, both models are equally good for Russia.

Keywords: COVID -19, ARIMA, ADLM, DES

1. INTRODUCTION

The corona virus disease (COVID -19) emerged in Wuhan the capital city of Hubei Province, China in December 2019. Early stages of the disease have been linked to a live animal seafood market in Wuhan, pointing to a zoonotic origin of the epidemic (Chowell et al., 2020), but the origin of the virus is questionable. The disease had spread well outside China, reaching more than 190 countries and infected more than 5 million people globally. More than 300,000 people died due to the diseases up to now, the World Health Organization (WHO) reported. The worst situation experienced by the USA followed by the UK, Italy, France, and Spain. The USA found itself grappling with the worst outbreak after Italy and Spain. According to WHO, 13,09,550 confirmed cases in USA, 2,15,260 in UK and 1,98,676 cases in Russia as of May 9, 2020. This rapid spread generating considerable turmoil among the population of many courtiers. At present, Russia and the UK report a high volume of active cases daily.

1.2 Research Problem

Increasing of confirming cases generate outnumbered of health care facilities and other medical supplies. Hospitals, quarantine sites, ventilators, beds, Personal Protective Equipment (PPE) kits, and other medical facilities need to strengthen the health system of those counties to overcome the negative consequences of COVID 19. Forecasting infected cases is an important activity to anticipate the requirement of healthcare resources and to save human lives. It helps them to maintain economic and social stability. Therefore finding suitable models for forecasting is a timely requirement. The study was designed to fill the knowledge gap.

1.3 Objective of the Study

The objective of the study is to forecast number of infected cases of COVID -19 in USA, UK and Russia.

1.4 Significance of the Study

The results of the study can be used for proactive decision making to minimize the risk to human life. The results will provide a platform to anticipate medical and health care resources to combat the epidemic. Government authorities can decide the lockdown procedures and minimize the movements of the general public.They can guide businesses to implement new doorstep delivery systems to the consumers to avoid gatherings in the market. Restricting the movements and quarantine are critical practices to avoid community spread. The results of this study would be facilitated to exercise those practices.

2. LITERATURE REVIEW

This section of the study reviewed research papers focus on forecasting COVID -19 for various destinations. Researchers used mathematical and statistical models for estimation and forecasting pandemic. Fanelli & Piazza (2020) have done a comparative assessment of the evolution of the outbreak in mainland China, Italy, and France. They provide estimates of using susceptible (S), infected (I), recovered (R), dead (D) (SIRD) models. COVID -19 epidemics in China forecasted by Chowell et al., (2020). They have used Generalized Logistic Models for forecasting cumulative cases. They have done short term forecasts for 5, 10, and 15days. Peng et al., (2020) have done a propagation analysis and

ISSN 2719-2385 (Online)

prediction of the epidemic in Italy, Iran, South Korea, and Hubei Province in China. They have done a comparison between official epidemic data and simulation data. They have compared the simulation results with the real data; analyzed the propagation process and its influencing factors. The authors have used the Gaussian distribution theory to construct a new model of corona virus transmission. Xiong et al., (2020) have used Artificial Intelligence in forecasting the epidemic in China. They have used the latent variables in the auto-encoder and clustering algorithms to group the provinces/cities for investigating the transmission structure. The prediction and analysis of the epidemic in China was done by Jia et al., (2020). Authors have used the Logistic model, Bertalanffy model, and Gompertz models for the purpose. Epidemic trends of SARS 2013 in China were tested and analyzed to prove the validity of existing models. The results of the study revealed that the logistic model outperformed other mathematical models. To develop the artificial intelligence (AI) methods for forecasting and evaluating intervention strategies to curb the spread of Covid-19 in the world was the objective of the study conducted by Xiong et al., (2020). The mean errors of the AI approach were very low. IHME COVID-19 health service utilization forecasting team 2020 estimates of predicted health service utilization and deaths due to COVID-19 in the US. They have used the Gaussian error function for the purpose. Li et al., (2020), establish the dynamics model with People who may be infected by the virus (S), Infected with the virus but without the typical symptoms of infection (E) Infected with the virus and highly infectious but not quarantined (I) Diagnosed and quarantined (Q) Suspected cases of infection or potential victims (D) People who are cured after infection(R) (SEIQDR) and Exponential Smoothing, ARIMA and ARIMAX models for predicting the epidemic in Hubei Province and mainland in China. ARIMAX (0,1,0) and SEIQDR models performed better. The Logistic growth model was used to model epidemic infected cases of 29 provinces in China, Iran, South Korea, Italy, and Europe by Sornette et al., (2020). Gupta & Pal (2020) have analyzed the epidemic outbreak situation in India and assess trends. They have used ARIMA and Exponential smoothing for the purpose, both techniques performed well.

Most of the forecasting epidemic research focuses on China, Italy, France, and some European countries. Very few studies were focused on the USA and Russia. The mathematical models namely: SIRD, Bertalanffy and Gompertz models, and statistical models namely: Exponential smoothing, Logistic, ARIMA, and ARIMAX models were used by many researchers. Besides, some distributions like Gaussian and soft computing techniques namely: Artificial Intelligence (AI) was used by some of the researches for the prediction purpose. SIRD, Logistic, Exponential smoothing, and AI performed better than other techniques. Most of the researchers have paid the least attention to the model validation and verification criterion; hence the reliability of the models is questionable.

3. METHODOLOGY

The daily confirmed cases of COVID-2019 of the USA, UK, and Russia for the period of 22nd January 2020 to 28th May 2020 were obtained from the World health organization (WHO) database. Pattern recognition of a data series paves the path for model developments. It gives an insight into the trends, seasonal variations, cyclical variations, and volatility of the time series. Therefore, time series plots, Auto Correlation Functions

ISSN 2719-2385 (Online)

(ACF), and Partial Auto Correlation Functions (PACF) were used for the purpose. Based on the pattern recognition, the Auto-Regressive Integrated Moving Average (ARIMA) model, Autoregressive Distributed Lag Model (ADLM) and Double Exponential Smoothing (DES) techniques were tested to forecast the pandemic of USA, UK, and Russia. The Anderson–Darling test, ACF, and Ljung-Box Q (LBQ)-test were used to test the validation criterion and fit the model. The forecasting ability of the models was assessed by three measurements of errors; Mean Absolute Percentage Error (MAPE), Mean Square Error (MSE) and Mean Absolute Deviation (MAD) in both model fitting and verification process.

3.1 Autoregressive Distributed Lag Model (ADLM)

Distributed lags arise when any cause-effect occurs after a period of time between one event and another (lag) in time. The effect doesn't feel all at once at a single point in time but is distributed over a period of time (Konarasinghe, 2015). Chen (2010) defines that the regressors may include lagged values of the dependent variable and current and lagged values of one or more explanatory variables. Further, he said that this model allows determining what the effects are of a change in a policy variable.
In usual notations the ADLM is;

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_n Y_{t-n} + \varepsilon_t \quad (1)$$

Or

$$Y_t = \alpha + \sum_{i=1}^{n} \beta_n Y_{t-n} + \varepsilon_t \quad (2)$$

Where Y_t, Y_{t-1} and $\ln Y_{t-n}$ are predictors and ε_t are white noise,

$\varepsilon_t \sim N(0, \sigma^2)$.

3.2 Double Exponential Smoothing Models

Double exponential smoothing provides short-term forecasts. This technique works well when a trend is present, but it also serves as a general smoothing method (Konarasinghe, 2016). This method is found using two dynamic estimates, α and β; with values between 0 and 1 (Konarasinghe, 2016). They represent level and trend respectively. Formulae of double smoothing technique (Holt' method) are;

$$L_t = \alpha Y_t + (1-\alpha)(L_{t-1} + T_{t-1}) \quad (3\text{-}1)$$

$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1} \quad (3\text{-}2)$$

$$\hat{Y}_t = L_{t-1} + T_{t-1} \quad (3\text{-}3)$$

$$F_{t+m} = L_t + mT_t \quad (3\text{-}4)$$

ISSN 2719-2385 (Online)

Where,

L_t : is the level at the end of period t , α is the weight of level, T_t = is the estimated trend at the end of period t, β is the weight of trend, m = is the forecast horizon.

4. RESULTS

The analysis contains three main parts:

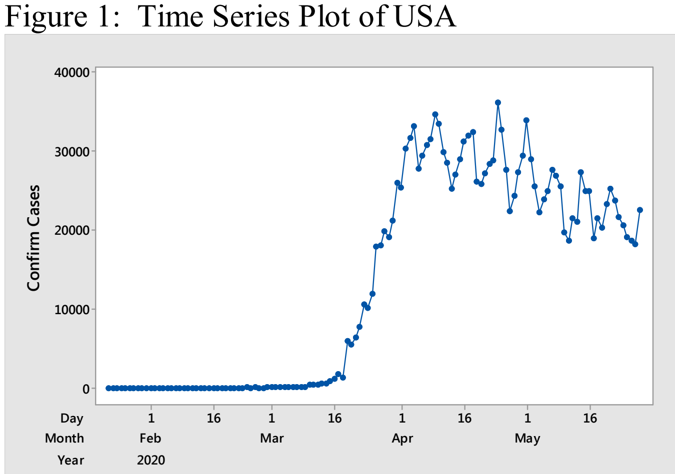
1. Pattern recognition and forecasting pandemic of the USA
2. Pattern recognition and forecasting pandemic of the UK
3. Pattern recognition and forecasting pandemic of the Russia

Log transformed data were used for the analysis. At first pattern recognition of the data series of USA, UK, and Russia was done, and then ARIMA, ADLM, and Exponential Smoothing techniques were tested for forecasting.

4.1 Pattern Recognition and Forecasting Pandemic of USA

Time series plot of confirmed cases of the USA was obtained for the period of 22nd January 2020 to 28th May 2020 (Figure 1). The first confirmed case reported from the USA on 22nd January 2020. The number of cases was low up to 18th March 2020 and shows a rapid growth afterward.

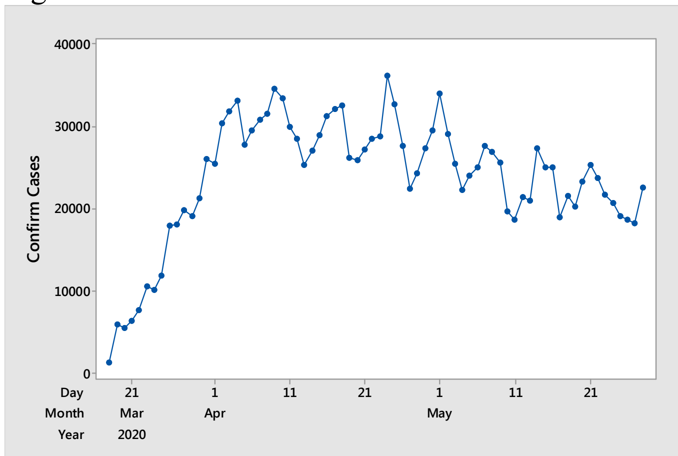
Figure 1: Time Series Plot of USA



Hence, the data set for the period of 18th March to 28th May 2020 was used to forecast the USA.

ISSN 2719-2385 (Online)

Figure 2: Time Series Plot of Growth of USA



The behavior of the daily confirmed cases for the selected period was further examined. Figure 2 is the time series plot of confirmed cases for the period of 18th March to 28th May 2020. There is a rapid growth from 18th March to 4th April 2020. After 4th April 2020, it shows a slight decline with a high irregular fluctuation. The ACF and Partial Auto Correlation Function (PACF) of the series are shown in Figures 3 and 4;

Figure 3: ACF of Daily Cases

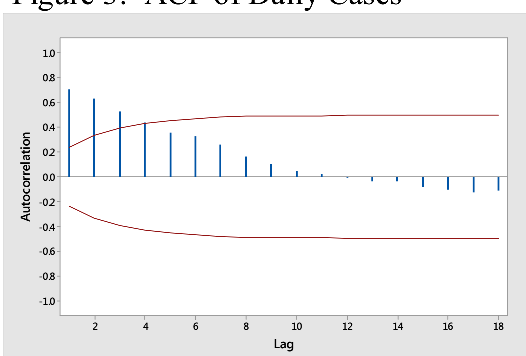
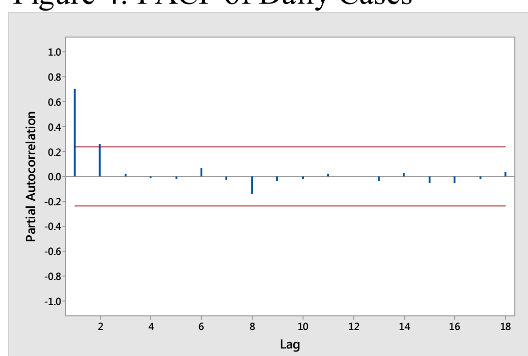


Figure 4: PACF of Daily Cases



The ACF has an exponential decline with three significant lags. The PACF has two significant spikes. The series confirm the stationary criteria. Hence, the ARIMA was tested, but it was not significant. The DES also was not successful. Finally, the ADLM was tested with three lags. The P-value of ANOVA is less than the significance level (α =0.05). It clearly showed that there is a linear relationship between the variables Y_{t-1}, Y_{t-2}, and Y_{t-3}, with Y_t. The next step was to test the individual regression coefficients. The results are in table 1.

ISSN 2719-2385 (Online)

Table 1: Summary Table for Regression Coefficients with all Lags

Term	Coefficients	SE Coefficients	T-Value	P-Value
Constant	1.872	0.440	4.25	0.000
Y_{t-1}	0.923	0.136	6.78	0.000
Y_{t-2}	-0.187	0.140	-1.33	0.189
Y_{t-3}	0.081	0.074	1.08	0.285

The P values correspond to Y_{t-2} and Y_{t-3} are (P= 0.189 and 0.285) are greater than the significance level (α =0.05). Model summary of the best fitting model is given in Table 2;

Table 2: Summaries for Regression Coefficients of Significant Lags

Term	Coefficients	SE Coefficients	T-Value	P-Value
Constant	1.712	0.386	4.43	0.000
Y_{t-1}	0.8328	0.0384	21.71	0.000

The Anderson Darling test confirmed the normality of residuals. The ACF of the residuals and LBQ test confirmed the independence of residuals. Table 3 is the model summary of ADLM. The R-Sq (adj) 89.70%, both relative and absolute measurements of errors are very low under the fitting and verifications.

Table 3: Summary of Model Fittings and Verifications of ADLM

Model	Model Fitting		Model Verification	
$\ln Y_t = 1.712 + 0.8328 \ln Y_{t-1}$	R-Sq(adj)	89.70%		
	MAPE	0.921313	MAPE	1.04308
	MAD	0.092503	MAD	0.103743
	MSE	0.013113	MSE	0.014509
	Normality	P = 0.295		
	Independence of Residuals	Yes		

The actual vs. fit and actual vs. forecast are in Figure 5 and 6. It is clear that the actual and fits are closer to each other and the fits follow the pattern of the actual series.

Figure 5: Actual Vs Fits

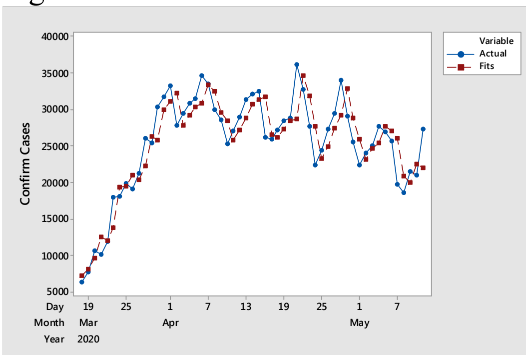
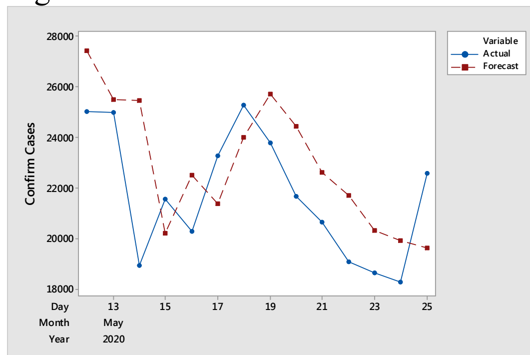


Figure 6: Actual Vs Forecast

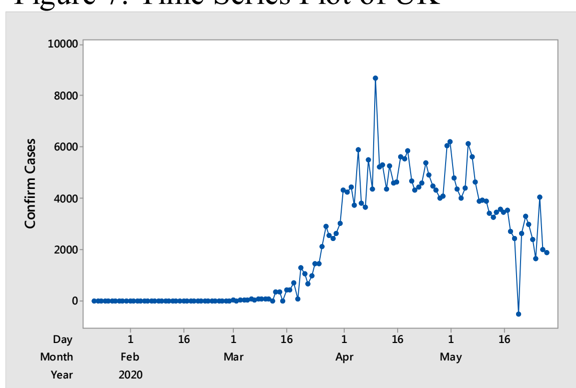


The ADLM is the most suitable model to forecast COVID -19 of the USA.

4.2 Pattern Recognition and Forecasting Pandemic of UK

The pattern recognition of confirmed cases of the UK was examined. The time series plot of confirmed cases was obtained for the period of 22nd January 2020 to 28th May 2020. Figure 7 is the time series plot of confirmed cases.

Figure 7: Time Series Plot of UK



The first confirmed case was reported on 31st January 2020. The number of cases was low up to 16th March and shows a minor growth up to 20th March 2020 and rapid growth afterward. Hence, the data set for the period of 20th March to 28th May 2020 used to forecast confirm cases of the UK.

ISSN 2719-2385 (Online)

Figure: 8 Time Series Plot of Growth of UK

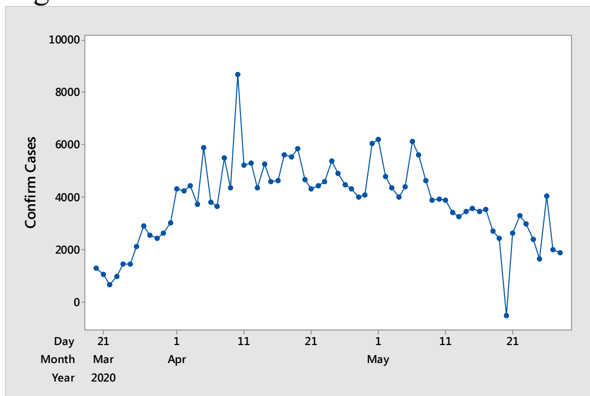


Figure 8 is the time series plot of daily confirmed cases for the period of 20th March to 28th May 2020. It is clear, that there is a rapid growth of confirmed cases from 20th March to 10th April 2020. After 10th April 2020, it shows a slight decline with a high irregular fluctuation. The ACF and Partial Auto Correlation Function (PACF) of the series are shown in Figures 9 and 10.

Figure: 9 ACF of Daily Cases

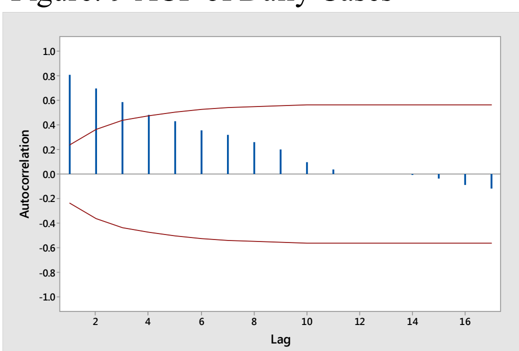
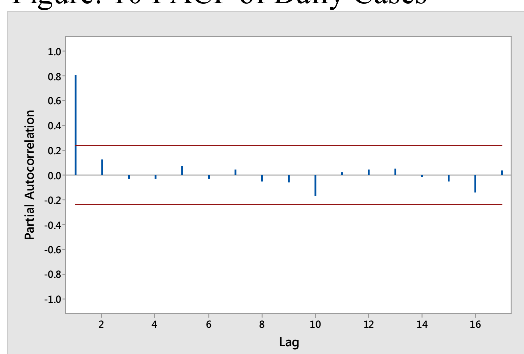


Figure: 10 PACF of Daily Cases



The ACF has an exponential decline, with three significant lags. The PACF has a single significant spike. The series does not confirm the stationary criteria. Hence, ADLM and ARIMA cannot be tested. The DES tested with log transformations for different α and γ values. Model Summary in table 4 shows the outputs at various levels.

Table 4: Summary of Model Fittings and Verifications of DES

Model	Model Fitting		Model Verification	
α (level) 0.85 γ (trend) 0.20	MAPE	2.48429	MAPE	3.56146
	MAD	0.20056	MAD	0.285752
	MSE	0.06733	MSE	0.122425
	Normality	P= 0.085		

ISSN 2719-2385 (Online)

	Independence of Residuals	Yes		

Among all DES, $\alpha = 0.85$ and $\gamma = 0.20$ had the least relative and absolute measurement of errors during the model fitting and verifications. The residuals were normally distributed and independent.

Figure 11: Actual vs Fits and Forecast of UK

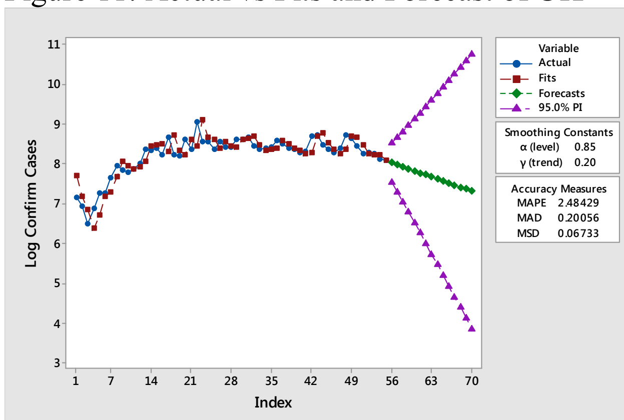


Figure 11 is the actual vs. fits and forecast of daily confirmed cases. The fits and the forecast followed a similar pattern of actual confirm cases. The deviation between actual values fits, and the forecast is less. Therefore, DES, $\alpha = 0.85$ and $\gamma = 0.20$ is the suitable model for forecasting the infected cases of the UK.

4.3 Pattern Recognition and Forecasting Pandemic of Russia

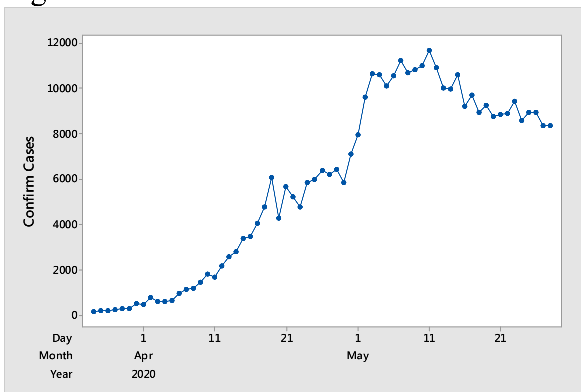
The pattern recognition of daily confirmed cases was examined. Time series plot of confirmed cases for the period of 22nd January 2020 to 28th May 2020 (Figure 12). There is a growth pattern of daily confirmed cases.

Figure 12: Time Series Plot of Russia

ISSN 2719-2385 (Online)

The first confirmed case reported from Russia on 31st January 2020. There is a rapid growth of daily cases after 25th March. Hence, the data set for the period of 25th March to 28th May 2020 used to forecast Russia. Figure 15 is the time series plot for the period of 25th March to 28th May 2020.

Figure 13: Time Series Plot of Growth of Russia



According to Figure 13, there is a growth of daily cases between 25th March and 19th April 2020. Afterward, there is an irregular fluctuation up to 29th April 2020. It is clear, that there was a rapid growth between 29th April and 3rd May 2020. Afterward, there was an increasing trend with irregular fluctuation up to 11th May and a decline with an irregular pattern. The ACF and PACF obtain from the data set to examine the relationship within the confirm cases shows as Figures 14 and 15.

Figure 14: ACF of Daily Cases

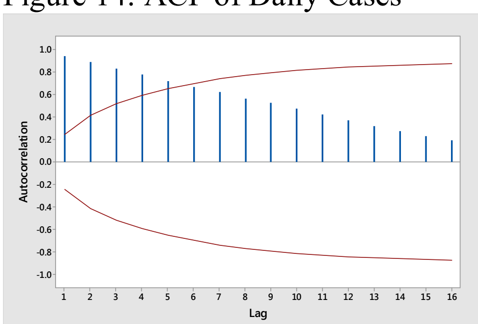
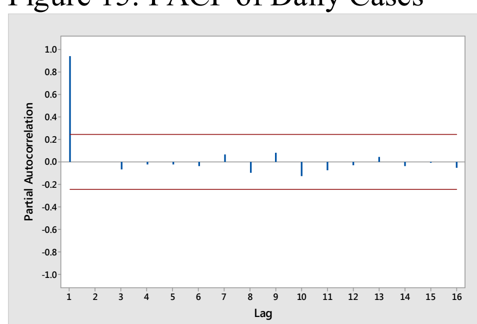


Figure 15: PACF of Daily Cases



The ACF has an exponential decline, with five significant lags; the PACF has a single significant spike. The series does not confirm the stationary criteria. Hence, ARIMA and ADLM cannot test. The DES technique was tested on data with log transformations for different α and γ values. The model Summary in Table 5 shows the outputs at various levels.

ISSN 2719-2385 (Online)

Table 5: Summary of Model Fittings and Verifications of DES

Model	Model Fitting		Model Verification	
α (level) 0.56 γ (trend) 0.32	MAPE	2.02213	MAPE	1.00644
	MAD	0.14034	MAD	0.091605
	MSE	0.03794	MSE	0.011664
	Normality	P = 0.062		
	Independence of Residuals	Yes		

Among all DES, α = 0.56 and γ = 0.32, has the least relative and absolute measurement of errors during fitting and verifications. The residuals are normally distributed and independent.

Figure 16: Actual vs Fits and Forecast of Russia

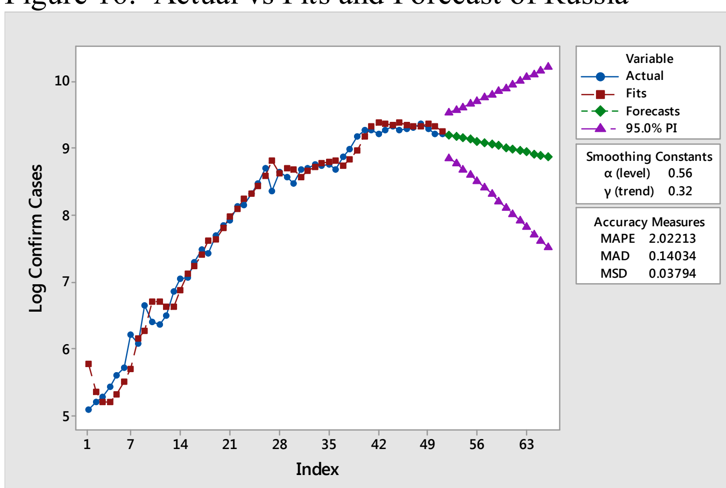


Figure 16 shows the actual vs. fits and forecast of the confirmed cases of Russia. The fits and the forecast follow a similar pattern with a small deviation. Therefore, DES, α = 0.56 and γ = 0.32 is the suitable model for forecasting the infected cases of Russia.

5. CONCLUSION AND RECOMMANDATION

It is concluded that the ADLM and DES are the most suitable models in forecasting pandemic in the USA, UK, and Russia. The behavior of pandemic is in a decline in the USA, UK, and Russia.

Due to the absence of antiviral drugs for COVID-19, the effective implementation of immunization practices such as; consume suitable food and beverages (Natural Food), avoid bad habits like smoking and alcohol consumption, use warm water, steaming, etc. are important. Besides non-pharmaceutical practices, such as personal protection, social distancing and restrict movements will be critical to bringing the epidemic situation under control. It is recommended to continue further studies as patterns of daily confirmed cases are under rapid change.

REFERENCE

Chen, Y. (2010).Autoregressive Distributed Lag (ADL) Model. http://mail.tku.edu.tw/chenyiyi/ADL.pdf

Chowell, G., Roosa, K., Lee, Y., Luo, R., Kirpich, A., Rothenberg, R., Hyman, J.M., & Yan, P. (2020). Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020. *Infectious Disease Modelling* 5, 256-263. https://doi.org/10.1016/j.idm.2020.02.002

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

Gupta, R., & Pal, S.K. (2020). Trend Analysis and Forecasting of COVID-19 outbreak in India. https://doi.org/10.1101/2020.03.26.20044511

IHME COVID-19 health service utilization forecasting team. (2020). Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator days and deaths by US state in the next 4 months. *www.medrxiv.org* . tracking ID MEDRXIV/2020/043752.

Jia,L., Li,K., Jiang,Y., Guo,X.,& Zhao,T.(2020). Prediction and analysis of Corona virus Disease 2019.

Konarasinghe, K.M.U.B.(2015). *Lin- Log Model for Forecasting International Tourism Income in Sri Lanka: Post-War Period*. Conference Proceedings of the 12th International Research Conference on Business Management (ICBM) – 2015. University of Sri Jayawardanapura. Sri Lanka. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2699689.

Konarasinghe, K.M.U.B. (2016).*Forecasting Tourist Arrivals to Sri Lanka: Post-War Period.* International Journal of Novel Research in Physics Chemistry & Mathematics. 3(1),57-63.

Li, Y., Wang, B., Peng, R., Zhou, C., Zhan, Y., Liu, Z., Jiang, X., Zhao, B. (2020). Mathematical Modeling and Epidemic Prediction of COVID-19 and Its Significance to Epidemic Prevention and Control Measures. *Ann Infect Dis Epidemiol*. 5(1), 1052.

ISSN 2719-2385 (Online)

Peng,H., Li,L.,Yang,Z., Dang,Z., Meng,C.,Huang,J., Meng, H., Wang,D.,Chen,G., Zhang,J., and Shao,Y.(2020). Propagation analysis and prediction of the COVID-19 *Infectious Disease Modelling* 5, 282- 292. https://doi.org/10.1016/j.idm.2020.03.002

Sornette, D., Wu, K., Darcet, D., Wang, Q. (2020). Generalized logistic growth modeling of the COVID-19 outbreak in 29 provinces in China and in the rest of the world.

Xiong, M., Hu, Z., Ge, Q., Li, S., & Jin, L. (2020). Artificial Intelligence Forecasting of Covid-19 in China. *The University of Texas Health Science Center.* Houston. USA.

Xiong, M., Hu,Z., Ge,Q., Li,S., Boerwincle, E., and Jin, L.(2020). Forecasting and evaluating intervention of Covid-19 in the World. *The University of Texas Health Science Center.* Houston. USA.

ISSN 2719-2385 (Online)