RESEARCH ARTICLE

Forecasting COVID-19 daily infected cases in Sri Lanka by Holt-Winters Exponential Smoothing Method

S. S. Wickramasinghe[1]*, K.M.U.B. Konarasinghe[2]

[1] *Department of Microbiology, Faculty of Medicine, University of Ruhuna, Galle, 80000, Sri Lanka*
[2] *Institute of Mathematics and Management, Sri Lanka*

Abstract: The novel coronavirus disease (COVID-19) has spread from China since December 2019 and spread worldwide including Sri Lanka. The aim of this study was to forecast the daily infected cases of COVID-19 in Sri Lanka which in turn help administrators for effective management of the pandemic. The method used in this study was Holt-Winters three parameter with additive or multiplicative models. The daily infected cases in Sri Lanka during the period of 22nd January 2020 to 22nd December 2021 were obtained from the publicly available databases of Epidemiology Unit of Sri Lanka and World Health Organization. The pattern recognition of the daily infected cases was examined by time series plot and Auto Correlation Function (ACF). The model validation was performed by the Anderson Darling test which confirmed the normality of residuals ($p > 0.05$) and ACF that confirmed the independence of residuals of the model. The forecasting ability of the model was assessed by the three measurements of errors; Mean Absolute Percentage Error (MAPE), Mean Absolute Deviation (MAD) and Mean Square Error (MSE). Holt-Winters additive and multiplicative model with α (level) 0.61, β (trend) 0.4 and γ (seasonal) 0.3 at a length of repeating behaviour of 3 days, had the least relative and absolute measurement of errors during the model fitting and verification. İn the multiplicative model, MAPE, MAD and MSE were 0.2847, 0.0187 and 0.0005 respectively. Similarly in the additive model, corresponding values of MAPE, MAD and MSE were 0.0207, 0.0187 and 0.0005. The fits and the forecast of these models followed a similar pattern of the actual daily infected cases concluding that the Holt-Winters model can be used to forecast the COVID-19 outbreak in Sri Lanka.

Keywords: COVID-19, forecasting, infected cases, Holt-Winters

Introduction

Since December 2019, the novel coronavirus disease (COVID-19) has spread from Wuhan city to other cities in China and around the world. On March 11, 2020, the World Health Organization declared the COVID-19 outbreak as a pandemic (WHO, 2020) İn Sri Lanka, the first case of COVID-19 was detected on 27th January 2020. According to the Epidemiology Unit of Sri Lanka, there had been 586,183 confirmed cases and 14,944 deaths due to this virus infection by 29th December 2021.

The application of mathematical models, big data, artificial intelligence, and other similar methods are used to predict the disease burden in a country in which infectious disease data are considered (Abdulmajeed et al., 2020). Forecasting of epidemics and pandemics based on mathematical models played a key role in controlling the spread of previous epidemics such as ebola, influenza, malaria, etc (Kalipe et al., 2018; Jhuo et al., 2019; Forna et al., 2020). In time series analysis, past observations of the same variables at consistent intervals are analyzed and developed models to describe the underlying relationship and extrapolate the time series into the future (Lima et al., 2019). Time series data are extensively used in various fields including tourism, economy, and health specifically in the COVID-19 pandemic and others (Djakaria and Saleh, 2021).

*corresponding author:*subodhawick@med.ruh.ac.lk* iD *https://orcid.org/0000-0002-7628-6645*

Holt-Winters exponential smoothing model can be applied when both trend and seasonality are present, with the two components being either additive or multiplicative. This method calculates dynamic estimates for three components: level (α), trend (β) and seasonality (γ) (Winters, 1960; Holt, 2004) and provides short to medium-range forecasting (Konarasinghe, 2016). In a study done by Djakaria and Saleh (2021) using Covid-19 pandemic data in the area of Gorontalo, Indonesia, from April 10 to October 13, 2020, they have found that Holt-Winters exponential smoothing was best with smoothing parameters of α = 0.1 and γ = delta = 0.5 for trend and seasonality respectively giving the lowest MAPE value of 6.14. Considering COVID-19 data in Jakarta Province in Indonesia, Sulasikin et al. (2020) have compared three different time series approaches, namely Holt's Linear, Holt-Winters' Additive and ARIMA and ARIMA had shown as the best forecasting model.

Forecasting the number of infected cases especially in a situation of an ongoing pandemic is necessary for better preparedness to overcome the problems that may encounter. However, there was the least number of studies found in the literature which have been carried out to forecast COVID-19 daily infected cases by applying Holt-Winters model in Sri Lanka. One such study carried out by Attanayake et al., (2020), they have considered the number of daily new infections and cumulative number of infections in COVID-19 from the first day of appearance of cases to 2nd July 2020 in four selected countries, Sri Lanka, Italy, the United States, and China (Hebei province). Four different models namely, Gompertz, logistic, Weibull, and exponential growth curves were fitted on the cumulative number of infections across countries. Results revealed that the most appropriate growth curve for Sri Lanka was logistic and the Gompertz, Weibull, and Gompertz curves were best for Italy, United States, and China (Hebei) respectively (Attanayake et al., 2020). Therfore, the aim of this study was to determine the best model and to forecast the daily infected cases of COVID-19 in Sri Lanka by using additive and multiplicative Holt-Winters model. This would be the first study that this model has been used in forecasting of the daily infected cases of COVID-19 in Sri Lanka.

Studies Based on Applications of Holt's Winters Model

In a study performed by Papastefanopoulos et al. (2020), six different time series modeling approaches in ten different countries for COVID-19 outbreak detection were investigated. İt was concluded that Holt-Winters additive model is best for data with trend and seasonality that do not increase over time and results in curved forecast that showed that shows seasonal changes in the data.

Time series analysis using daily infected cases in the Philippines, it was found that Holt-Winters additive and multiplicative with (α = 0.6, γ = 0.2 and δ = 0.4) had the least relative and absolute measurement of errors during the model fitting and verifications. It was concluded that the forecasting ability of Holt's Winters additive and multiplicative models was extremely high and both additive and multiplicative models equally well in forecasting daily infected cases in the Philippines (Konarasinghe, 2021a).

In Indonesia, time series forecasting for the spread of COVID-19 was done using hybrid ARIMA and Holt-Winters methods which proved the best method for predicting the number of COVID-19 spread rates in November and December 2020 (Pramudya and Marina, 2021). In a study done by Djakaria and Saleh (2021) using Covid-19 pandemic data in the area of Gorontalo, Indonesia, from April 10 to October 13, 2020, they have found that Holt-Winters exponential smoothing was best with smoothing parameters of α = 0.1 and β = γ = 0.5 for trend and seasonality respectively giving the lowest MAPE value of 6.14.

Materials and Methods

A retrospective analysis was performed using daily infected cases of COVID-19 in Sri Lanka. Data were obtained from 22nd January 2020 to 22nd December 2021 from publicly available databases of Epidemiology Unit of Sri Lanka (www.epid.gov.lk) and World Health Organization. Time series analysis was performed using Minitab statistical software (18th version).

To select the mathematical model, pattern recognition of the daily infected cases examined by time series plot and Auto Correlation Function (ACF). Based on the pattern, the Holt-Winters three parameter model with additive or multiplicative was used to forecast daily infected cases in Sri Lanka. Log transformed data from 28th September 2021 to 12th December 2021 were used to analyse further. Outlier detection was done using a boxplot followed by adjustment of outliers by moving average order 3. Validation and verification of the model was performed.

Model fitting or validation was carried out by using 2/3 (n=50) of the data set. Weights were assigned into 3 parameters, level, trend and season with the seasonal length of 3 days and these were changed until the assumptions were satisfied with the lowest error level.

Anderson Darling test (probability test) was used to confirm the normality of residuals (at level of significance at 0.05) ($p > 0.05$) and ACF was used to confirm the independence of residuals of the model. Same procedure was followed for both multiplicative and additive types of Holt-Winters.

Model verification was performed by using remaining 1/3 (n=26) of the data set to assess the suitability of model by comparing forecasts and actual data. The forecasting ability of the model was assessed by the three measurements of errors; Mean Absolute Percentage Error (MAPE), Mean Square Error (MSE), and Absolute Deviation (MAD) in both model fitting and verification processes. The model with the lowest errors was selected as the best model for forecasting.

Results and Discussion

Pattern recognition

Time series plot and Auto Correlation Function (ACF) of daily infected cases for the period of January 2020 to December 2021 was used to observe the pattern. There were many growths including a few exponential growths of daily cases within the precise period. The cases were weakly (non) stationery and trend and seasonality or repeating behaviour were observed (Fig 1 and 2).
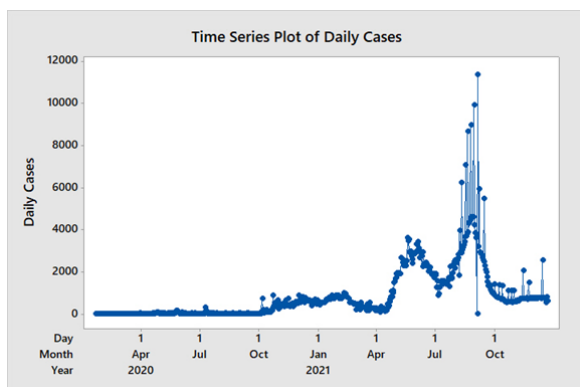

Figure 1: Time series plot of daily cases of COVID-19 in Sri Lanka from January 2020 to December 2021

Data for the period of 28th September 2021 to 12th December 2021 were used for further analysis to forecast daily infected cases of COVID-19 in Sri Lanka. A declining trend with irregular fluctuations with a seasonal behaviour or repeating behaviour was observed (Fig. 3 and 4). Outlier detection was done using a boxplot followed by adjustment of outliers by moving average order 3.
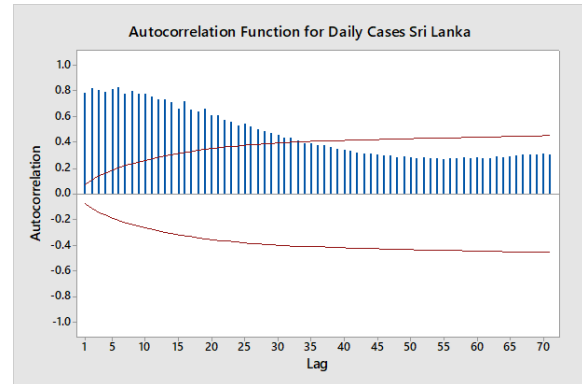

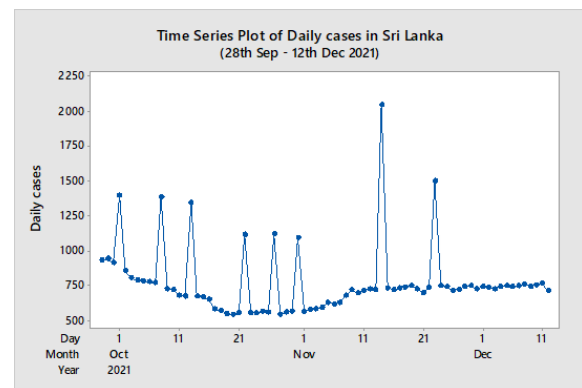Figure 2: Autocorrelation function for daily cases Sri Lanka

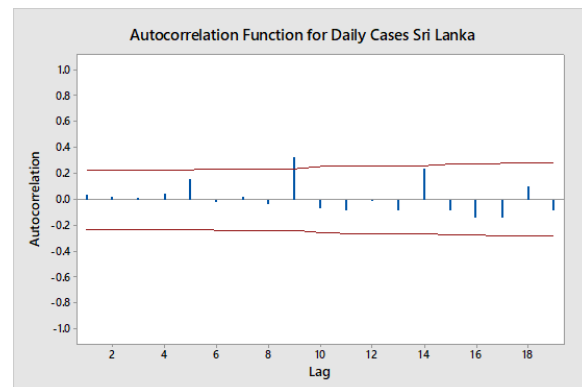Figure 3: Time series plot of COVID-19 daily cases in Sri Lanka from 28th September 2021 to 12th December 2021

Figure 4: Correlogram of autocorrelation function for COVID-19 daily cases in Sri Lanka from 28th September 2021 to 12th December 2021.

Model fitting or validation

Model fitting or validation was carried out by using 2/3 (n=50) of the data set. Weights were assigned into 3 parameters, level, trend and season with the seasonal length of 3 days and these were changed until the assumptions were satisfied with the lowest error level. Anderson Darling test (probability test) was used to confirm the normality of residuals at 0.05 of level of significance and ACF was used to confirm the independence of residuals of the model (Fig 5 and 6). Same procedure was followed for both multiplicative and additive types of Holt-Winters.
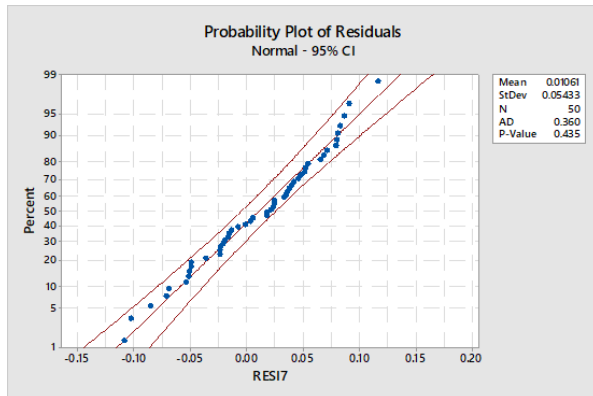

Figure 5: Probability test (Anderson Darling test) confirmed the normality of residuals as p > 0.05
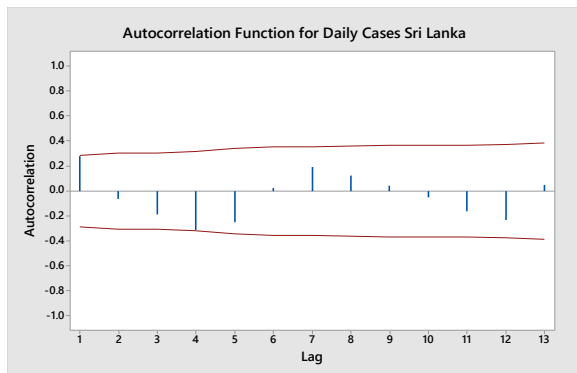

Figure 6: Correlogram of ACF - confirmed the independence or randomness of residuals as all the spikes are within CI lines.

Fits and actual values of daily cases in Sri Lanka by multiplicative method and additive method are shown below (Fig. 7 and Fig. 8).
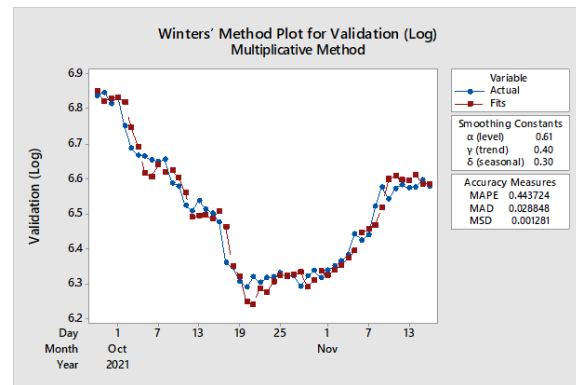

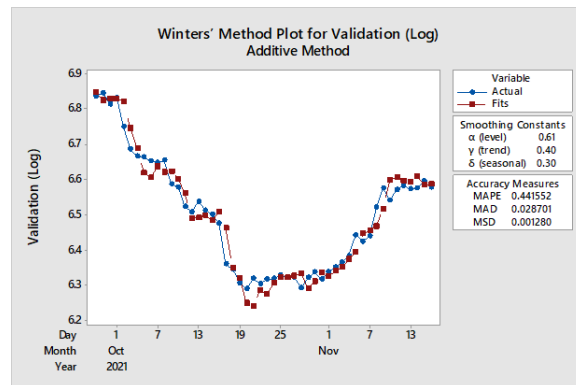Fig. 7: Fits and actual values of Daily cases in SL (A) Multiplicative method

Figure 8: Fits and actual values of Daily cases in SL (Additive method)

Model verification and calculation of measurement of errors

Model verification was performed by using remaining 1/3 (n=26) of the data set to assess the suitability of model by comparing forecasts and actual data. The forecasting ability of the model was assessed by the three measurements of errors; Mean Absolute Percentage Error (MAPE), Mean Square Error (MSE), and Absolute Deviation (MAD) in both model fitting and verification processes. The model with the lowest errors was selected as the best model for forecasting.

Therefore, Holt-Winters model with additive or multiplicative models were tested with log-transformed data for different α, β and γ values and the seasonal length of 3 days. Model summary in table 1 shows the performance of Holt-Winters additive and multiplicative models with the seasonal length of 3.

Table 1: Model Summary of Holt-Winters method

		Model fitting		Model verification	
Multiplicative (3) α (level) 0.61 β (trend) 0.40 γ (seasonal) 0.30	MAPE	0.4437	MAPE	0.2847
	MAD	0.0288	MAD	0.0187
	MSE	0.0012	MSE	0.0005
	Normality	P=0.829		
	Independence of Residuals	Yes		
		Model fitting		Model verification	
Additive (3) α (level) 0.61 β (trend) 0.40 γ (seasonal) 0.30	MAPE	0.4415	MAPE	0.0207
	MAD	0.0287	MAD	0.0187
	MSE	0.0012	MSE	0.0005
	Normality	P=0.812		
	Independence of Residuals	Yes		

Accordingly, both multiplicative and additive models with α (level) 0.61, β (trend) 0.4 and γ (seasonal) 0.3 had the least relative and absolute measurement of errors during the model fitting and verifications. İn the multiplicative model, MAPE, MAD and MSE were 0.2847, 0.0187 and 0.0005 respectively. Similarly, in the additive model, MAPE, MAD and MSE were 0.0207, 0.0187 and 0.0005 respectively. The residuals were normally distributed and independent.

The fits and the forecast of Holt-Winters additive and multiplicative models followed a similar pattern of the actual daily infected cases in Sri Lanka (Fig 9 and 10). The deviation between the actual values of fits and the forecast was less. Therefore, Holt-Winters both additive and multiplicative models with α 0.61, β 0.4 and γ 0.3 are the suitable models at a length of repeating behaviour of 3 days to forecast daily infected cases in Sri Lanka.
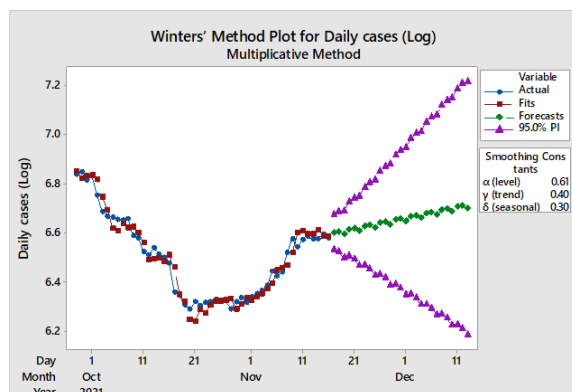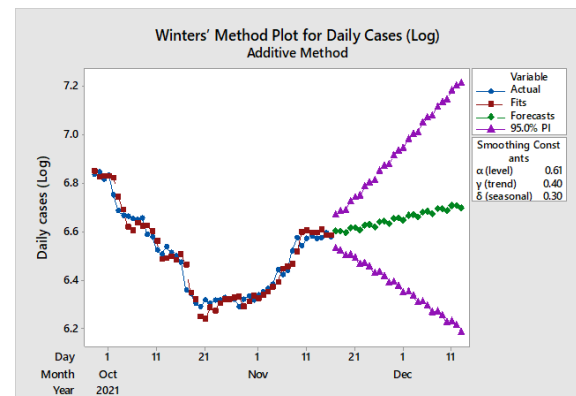

Figure.9: Forecasts by Multiplicative method

Figure.10: Forecasts by Additive method

These findings may be beneficial for the Government for better preparedness for COVID-19 pandemic with a view of controlling the spread by better decision making and adhereing to a proper action plan based on these results.

There are number of studies which were carried out to predict the number of COVID-19 cases in other countries using the Holt's Winters additive and multiplicative model. Similar to our study, this same model was used to forecast daily infected cases in Philippines and it has been shown that the model with α = 0.6, γ = 0.2 and δ = 0.4, had the least relative and absolute measurement of errors during the model fitting and verifications. Hence, it concluded that both additive and multiplicative models equally well in forecasting daily infected cases in the Philippines (Konarasinghe,   2021a).   According   to Papastefanopoulos et al. (2020), Holt-Winters additive model showed the best performance compared to the other models in forecasting the active cases per

population in Turkey. Similarly, in Indonesia, Irandi et al. (2021), proposed exponential smoothing Holt-Winters with the additive model with seasonal addition considering trend and seasonal factors. They had mentioned that this model provided sharp and well performance for forecasting daily new cases of COVID-19 in Central Java province with MAPE less than 10%. In the study conducted by Pramudya and Marina (2021), a hybrid model that combines ARIMA and Holt-Winters models was adopted to positive confirmed cases, recovered cases and total deaths. Based on the results of the research, they have concluded that the Hybrid ARIMA and Holt-Winters method was the best method for predicting the number of COVID-19 spread rates in Indonesia as it produced the smallest RMSE value.

In contrast to the above studies, there were other studies in which different models were caried out to predict the disease burden of COVID-19. In the study carried out by Konarasinghe (2021b), to forecast daily infected cases in Argentina, it revealed that Sama Circular Model (SCM) was the most suitable one and the Holt's Winters method had not been useful in it as it could not fulfill the model validation criterion. To predict the recovered and death cases of COVID-19 in India, Panda (2020) had employed two methods, namely ARIMA and Holt-Winters exponential smoothing models. İt was shown that proposed ARIMA model was the most accurate one in predicting both recovered and death cases compared to Holt-Winters model. İn a study done by Prajapati et al. (2021), the hybrid combination of ARIMA and NARNN (Nonlinear Auto-Regression Neural Network) had given the best result with a reduced RMSE among the selected forecasting methods such as Prophet, Holt Winters, and the ARIMA model in encapsulating the linear as well as non-linear patterns of the epidemical datasets.

Conclusions

Holt-Winters Exponential Smoothing three-parameter additive or multiplicative models can be used to forecast the daily infected cases in Sri Lanka. Out of these 2, the best fitted model would be the Holt-Winters three-parameter additive one. The results will be helpful for the health authorities for planning and implementing preventive measures against COVID-19 in Sri Lanka.

The limitation of this proposed model is the ability to forecast for a shorter time period. Also, when there are other factors that contribute to the spread of a disease such as availability of tests, interventions, lockdown

policies and non-adherence to preventive measures, it is difficult to forecast using mathmatical models as it may give rise to unrealistic values.

Acknowledgement

Conflict of Interest

The authors declare that they have no competing personal or financial interests that could have influenced the work stated in this paper.

References

Abdulmajeed K, Adeleke M, Popoola L (2020). Online forecasting of COVID-19 cases in Nigeria using limited data. Data in Brief 30: 105683

Attanayake A, Perera S, Jayasinghe S (2020). Phenomenological Modelling of COVID-19 epidemics in Sri Lanka, Italy and Hebei Province of China. MedRxiv: 2020.2005. 2004.20091132

Djakaria I, Saleh S (2021). Covid-19 forecast using Holt-Winters exponential smoothing Journal of Physics: Conference Series. IOP Publishing, pp 012033

Forna A, Nouvellet P, Dorigatti I, Donnelly CA (2020). Case fatality ratio estimates for the 2013–2016 West African Ebola epidemic: application of Boosted Regression Trees for imputation. Clinical Infectious Diseases 70: 2476-2483

Holt CC (2004). Forecasting seasonals and trends by exponentially weighted moving averages. International journal of forecasting 20: 5-10

Irandi DF, Rohmawati AA, Gunawan PH (2021). Forecasting Number of New Cases Daily COVID-19 in Central Java Province Using Exponential Smoothing Holt-Winters. Indonesia Journal on Computing (Indo-JC) 6: 23-32

Jhuo S-L, Hsieh M-T, Weng T-C, Chen M-J, Yang C-M, Yeh C-H (2019). Trend prediction of influenza and the associated pneumonia in taiwan using machine learning 2019 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS). IEEE, pp 1-2

Kalipe G, Gautham V, Behera RK (2018). Predicting malarial outbreak using machine learning and deep learning approach: a review and analysis 2018 International Conference on Information Technology (ICIT). IEEE, pp 33-38

Konarasinghe K (2016). Forecasting tourist arrivals to Sri Lanka: Post-war period. International Journal of Novel Research in Physics Chemistry & Mathematics 3: 57-63

Konarasinghe K (2021a). Forecasting COVID-19 Outbreak in the Philippines and Indonesia. Journal of New Frontiers in Healthcare and Biological Sciences 2: 1-19

Konarasinghe K (2021b). Model Base Analysis of COVID-19 Daily Infected Cases in Argentina. Journal of New Frontiers in Healthcare and Biological Sciences 1: 1-14

Lima S, Gonçalves AM, Costa M (2019). Time series forecasting using Holt-Winters exponential smoothing: An application to economic data AIP conference proceedings. AIP Publishing

Panda M (2020). Application of ARIMA and Holt-Winters forecasting model to predict the spreading of COVID-19 for India and its states. medRxiv: 2020.2007. 2014.20153908

Papastefanopoulos V, Linardatos P, Kotsiantis S (2020). COVID-19: a comparison of time series methods to forecast percentage of active cases per population. Applied sciences 10: 3880

Prajapati S, Swaraj A, Lalwani R, Narwal A, Verma K (2021). Comparison of traditional and hybrid time series models for forecasting COVID-19 cases. https://arxivorg/ftp/arxiv/papers/2105/210503266pdf, Access on 1122023

Pramudya DA, Marina N (2021). Time Series Forecasting for the Spread of Covid-19 in Indonesia Using Hybrid ARIMA and Holt-Winters Methods. International Research Journal of Advanced Engineering and Science 6: 201-205

Sulasikin A, Nugraha Y, Kanggrawan J, Suherman AL (2020). Forecasting for a data-driven policy using time series methods in handling COVID-19 pandemic in Jakarta 2020 IEEE International Smart Cities Conference (ISC2). IEEE, pp 1-6

WHO (2020). WHO-China joint mission on coronavirus disease 2019 (COVID-19). Available at https://wwwwhoint/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-reportpdf Accessed on 01122023 2019: 16-24

Winters PR (1960). Forecasting sales by exponentially weighted moving averages. Management science 6: 324-342

www.epid.gov.lk/web/index.php?lang=en, accessed on 30.12.2021.