IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Study on Parametric Model for COVID-19 data

S. Sundarabalan and S. Raguraman

Department of Statistics, Dwaraka Doss Goverdhan Doss Vaishnav College, Chennai – 06.

Abstract

The main objective of this study is to find the best predictive model of COVID-19 cases in Chennai, Tamil Nadu. We have considered daily COVID 19 new cases for the period of April 2021. In parametric models different linear, non-linear, smoothing methods and time series models have been employed. Among the non-linear and ARIMA models, appropriate model was selected based on highest adjusted R^2 values, significance of the model parameters, lowest values of RMSE, MAE and MAPE values. Further the residual of the selected model was tested for randomness and normality.

Key words: COVID 19, Non-linear models, Run test, Durbin-Watson Statistics, parametric model.

1. Introduction

China Country office informed the WHO of unknown cases of pneumonia on 30th December 2019. On investigation, the officers found a link to the Huanan Seafood market of Wuhan. Out of 44 cases, 33 patients were stable, but 11 were severely ill (WHO, 2020).

Since its discovery, approximately 216 countries have been affected by it, with the total number of confirmed cases being 11.4 Million and 534K deaths. India currently has 697K cases and 19693 deaths, placing it at the third-highest position in the world as of 6 July (WHO, 2020). Coronavirus showed similar etiology to the Severe Acute Respiratory Syndrome (SARS) virus of 2006. Hence, it is also known as SARS-2 or COVID-19 (WHO, 2020). It is a contagious disease that spreads via minute respiratory droplets through coughing, sneezing, or close contact. Patients can experience the symptoms in 2 to 14 days of incubation, such as high fever, body pain or weakness, dry cough, breathlessness, pneumonia, kidney failure, and respiratory distress. In some cases, the patients do not show any symptoms of the virus; that is, they are asymptomatic. In both scenarios, testing of the patient is essential for verification, and they should be isolated (He et al., 2020; Singhal et al., 2020). Thus to curb the rate of infection and prevent community spread wearing masks, to avoid crowded spaces and maintaining the protocols of social distancing is essential (WHO, 2020).

In India, the first confirmed case of COVID-19 was of a man with travel history from Wuhan (WHO, 2020). Although the progression of cases was slow initially, the Indian government imposed a lockdown on 25th March to prevent human-to-human transmission and contain the pace of the virus. The main objective of this study is to find the best predictive model of COVID-19 cases in Chennai, Tamil Nadu.

2. Materials and Methods

Confirmed, recovered and death cases of COVID-19 infection are collected for Chennai, Tamil Nadu, as per World Health Organization region classification, from the official website of **www.stopcorono.tn.gov.in** from 01 April 2021 to 31 April 2021.

To study the different non-linear models and ARIMA models have been employed. The performances of the different models have been studied based on model performance measures such as RMSE, MAE and MAPE values. The widely used Levenberg-Marquardt algorithm (Ratkowsky, 1990) was utilized to fit the non-linear models. Different initial parameters values were used to ensure global convergence. The standard SPSS 22 package was used to fit the different non-linear models. The procedures of non-linear and ARIMA models are being discussed in detail in the following sections.

2.1 Parametric Models

In parametric model different non-linear models given in table -1 (Bard, 1974; Seber and Wild, 1989; Ratkowsky, 1990, Draper and Smith, 1998 and Montgomery *et al*., 2003) are employed. Among the non-linear models, the model having highest adjusted R^2 with significant F value is selected, so that it satisfies test for goodness of fit (Montgomery *et al*., 2003). Normality of residuals is examined by using Shaprio-Wilks test (Agostid'no and Stephens, 1986). Furthermore, while dealing with time-series data it may be possible that successive observations may be auto-correlated among themselves (Venugopalan and Shamasundaran, 2003). To overcome all these problems, performing residual analysis is strongly advised. Randomness assumption of the residuals needs to be tested before taking any final decision about the adequacy of the model developed. To carry out the above analysis "Run test" procedure developed in the literature (Ratkowsky, 1990). Further, to test the presence or absence of auto- correlation in the data set Durbin-Watson test procedure (Lewis-Beck, 1993) is utilized. In case of more than one models being the good fit for the data, the best model is selected having lower values of RMSE, MAE and MAPE. The Standard SPSS 22 Package was used to fit the models given in Table 1.

Table 1: List of parametric models

Model No.	Model	Name of the Model
I.	y=a*e**(b*x) + e	Exponential
II.	y= a*exp(-exp(b-c*x)) + e	Gompertz
III.	y=a*b**(1/x)*x**c + e	Modified Hoerl
IV.	y= a/(1+b*exp(-c*x)) + e	Logistic
V.	y=(a*b+c*x**d)/(b+x**d) + e	MMF
VI.	y=a+b*cos(c*x+d) + e	Sinusoidal
VII.	y=(a+b*x)/(1+c*x+d*(x**2)) + e	Rational Function
VIII	y=1/(a+b*x**c)+ e	Harris Model
IX	y=a*(b-exp(-c*x))+ e	Exponential Association

In Table 1, *y* is the exchange rates; x is the time point; a, b, c and d are the parameters and e is the error term. In addition, a represents the carrying capacity, c is the intrinsic growth rate, b represents different functions of the initial value y(0) and d is the added parameter in the sinusoidal and rational function models.

Model performance measures:

The following model performance measures are used to select the appropriate models.

Root Mean Square Error (RMSE) = $\left[\sum_{i=1}^{n}\frac{\left(Y_i - \hat{Y}_i\right)}{n}\right]^{1/2}$

Mean Absolute Error (MAE) = $\sum_{i=1}^{n}\left|\frac{\left(Y_i - \hat{Y}_i\right)}{n}\right|$

Mean Absolute Percentage Error (MAPE) = $\frac{\left|\left(Y_i - \hat{Y}_i\right)\right| / Y_i}{n} \times 100\%$

where n and p are number of observations and number of parameters, respectively in the model. The lower the values of these statistics, the better are the fitted model.

As pointed out by Kvalseth (1985), before taking any final decision about the appropriateness of the fitted model, it is paramount importance to investigate the basic assumptions regarding the error term, viz., randomness and normality. Randomness assumptions of the residuals are to be tested before taking any final decision about the adequacy of the model developed. To carry out this analysis “Run test” procedure is employed.

3. Results and Discussions

Different non-linear and ARIMA models have been employed to study trends in COVID 19. The characteristic fitted non-linear, smoothing methods and ARIMA models are presented in Table 2 and 3. Among the non-linear, smoothing methods and ARIMA models, appropriate model was selected based on highest adjusted R^2 values, significance of the model parameters, lowest values of RMSE, MAE and MAPE values.

3.1 Parametric models for COVID 19

The result presented in table 2 for the petrol rate revealed that among the non- linear models the Harris, Logistic, Exponential and Sinusoidal model was fitted having the adjusted R^2 (64 %) and highly significant R^2 values. The estimated parameter values were found to be within the 95% confidence interval, indicating that the estimated parameter values were significant at 5% level of significance. For all these models the model performance measures viz. RMSE, MAE and MAPE values were almost same. The p-values of S-W test as well as Run test values were found to be significant, indicating that the residuals due to these models were not independently distributed. Also the D-W Statistic values were found to be less than two, indicating that the residuals due to these models were correlated. Since none of the non-linear models were found suitable to fit the trends in euro exchange rates, different smoothing techniques and ARIMA time series models have been employed and the results are presented in Table 3.

The results presented in table 3 revealed that among four smoothing methods and the ARIMA time series models the ARIMA (1,1,1) had the lowest values of RMSE **(1.9017)**, MAE **(1.4018)** and MAPE **(2.0561)**. Hence among the parametric models the ARIMA (1,1,1) was found suitable to fit the trends in COVID 19.

Table 2: Characteristics of fitted parametric models for COVID 19.

		Models			
		Harris Model	**Logistic Model**	**Exp. Association Model**	**Sinusoidal Fit**
Parameters	**A**	0.020* (0.002)	82.688* (3.185)	28.485* (4.091)	48.396* (86.5179)
	B	-0.001* (0.001)	0.451* (0.044)	2.993* (0.270)	29.726* (87.912)
	C	0.320* (0.126)	0.014* (0.003)	0.009* (0.003)	0.008* (0.014)
	D	-	-	-	-1.264* (2.072)
Goodness of Fit	**R^2%**	65**	65**	65**	65**
	Adj. R^2 %	64	64	64	64
	S-W test (p-value)	0.000	0.000	0.000	0.001
	Run Test (p-value)	0.000	0.000	0.000	0.000
	D-W Statistic	0.1758	0.1753	0.1750	0.1753
	RMSE	4.5427	4.5297	4.5342	4.5288
	MAE	3.4598	3.5372	3.5265	3.5633
	MAPE	4.8862	5.0043	4.9882	5.0399

*Significant at 5% level ** Significant at 1% level; RMSE: Root Mean Square Error, MAE: Mean Absolute Error, Values in brackets ( ) indicates standard errors.

Table 3: Parametric models for COVID 19.

Model	Model Performance Measures		
	RMSE	MAE	MAPE
Linear trend	3.8152	2.5519	3.0379
Quadratic trend	5.5740	4.5510	6.0225
Exponential trend	5.7876	4.5980	6.0892
S-curve trend	6.9069	5.4855	7.8637
Simple Exp.	1.9179	1.4553	2.1799
Brown's Linear exp.	2.1657	1.5387	2.2652
Holt's Linear exp.	1.9991	1.6692	2.2018
Brown's quadratic exp.	2.2499	1.6744	2.3985
ARIMA(0,1,0)	1.9079	1.4647	2.1933
ARIMA(0,1,1)	1.9155	1.4521	2.1765
ARIMA(1,1,0)	1.9140	1.4658	2.0948
ARIMA(1,1,1)	**1.9017**	**1.4018**	**2.0561**

Model	Test for Randomness				
	RUNS	RUNM	AUTO	MEAN	VAR
Linear trend	**	***	***	OK	***
Quadratic trend	**	***	***	*	**
Exponential trend	**	***	***	OK	***
S-curve trend	***	***	***	***	OK
Simple Exp.	OK	OK	OK	OK	OK
Brown's Linear exp.	OK	OK	OK	OK	OK
Holt's Linear exp.	OK	OK	OK	OK	OK
Brown's quadratic exp.	**	OK	OK	OK	OK
ARIMA(0,1,0)	OK	OK	OK	OK	OK
ARIMA(0,1,1)	OK	OK	OK	OK	OK
ARIMA(1,1,0)	OK	OK	OK	OK	OK
ARIMA(1,1,1)	**OK**	**OK**	**OK**	**OK**	**OK**

RMSE: Root Mean Squared Error;

RUNS: Test for Excessive Runs up and Down;

RUNM: Test for Excessive Runs Above and Below Median;

AUTO: Box-Pierce Test for Excessive Autocorrelation;

MEAN: Test for difference in Mean 1st Half to 2nd Half;

VAR: Test for Difference in Variance 1st Half to 2nd Half;

OK: non - significant

4. Conclusion

The present study was carried using India COVID 19 data. In parametric models different linear, non-linear, smoothing methods and time series models have been employed. Among the non-linear and ARIMA models, appropriate model was selected based on highest adjusted R^2 values, significance of the model parameters, lowest values of RMSE, MAE and MAPE values. The study revealed that non-linear models, among four smoothing methods and the ARIMA time series models the ARIMA (1,1,1) had the lowest values of RMSE (1.9017), MAE (1.4018) and MAPE (2.0561). Hence among the parametric models the ARIMA (1,1,1) was found suitable to fit the trends in COVID 19 in India.

References

1. Agostid'no, R.B., Stephens, M.A., (1986). *Goodness of fit technologies*, Marcel Dekker, New York.
2. Altman, N. S. (1990). Kernel smoothing of data with correlated errors. *Journal of the American Statistical Association*, ***85***, 749–759.
3. Bard, Y., (1974). *Nonlinear Parametric Estimation*, Academic Press, New York.
4. Box, G. E. P. and Jenkins, G. M., *Time Series Analysis: Forecasting and Control*, Second Edition, Holden Day, 1976.
5. Drapper, N. R. and Smith, H. (1998). *Applied Regression Analysis.* 3rd Edition, John Wiley & Sons, USA.
6. Ljung, G. M. and Box, G. E. P. (1978). On a Measure of a lack of Fit in Time Series Models, *Biometrika*. 65(2): 297-303.
7. Montgomery, D. C., Peck, E. A. and Vining, G. G.(2003). *Introduction to Linear Regression Analysis*, john Wiley & sons, Inc.
8. Ratkowsky, D.A., (1990). *Handbook of Non-linear Regression Models*. Marcel Dekker,

New York.

9. Sananse, S.L., Maidapwad, S.L., (2009). On estimation of growth rates using linear models. *Int. J. Agric. Statistics Sci*., 5: 463-469.

10. Seber, G. A. F., Wild, C. J., (1989). *Non-Linear Regression*. John Wiley and Sons, Newyork.

11. Sharma VK, & Nigam U, Modeling and Forecasting for Covid-19 growth curve in India. *medRxiv*, 2020.

12. Venugopalan, R., Shamasundaran, K.S., (2003). Nonlinear Regression: A realistic modeling approach in Horticultural crops research". *Jour. Ind. Soc. Ag. Statistics* 56(1): 1-6.

13. WHO. 2020. Pneumonia of Unknown Cause – China. URL https://www.who.int/csr/don/05-january-2020-pneumonia-of-unkown-cause-china/en/

14. WHO. 2020. Rolling Updates on Coronavirus Disease (COVID-19). URL https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen

15. WHO. 2020. Coronavirus Disease 2019 (COVID-19) Situation Report – 67. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports

16. He G, Sun W, Fang P, Huang J, Gamber M, Cai J, & Wu J (n.d.). The clinical feature of silent infections of novel coronavirus infection (COVID-19) in Wenzhou. *Journal of Medical Virology, n/a*, 2020.