Cointegrated Movement of Covid Cases with Cured and Deaths in Kerala

Unnikrishnan T[1], Prasanth C.B[2*] & Ramya. M. S[3].

[1]Assistant Professor, Department of Statistics, Shri C Achuthamenon Govt. College, Thrissur. Kerala,t.unnikrishnan@gmail.com

[2]Assistant Professor, Department of Statistics, *Sree Kerala Varma College,* Thrissur. Kerala,cbpwarrier@gmail.com

[3] Research Scholar, Department of Statistics, Annamalai University, msramyaotp@gmail.com

Abstract: *Epidemiologists and policymakers turn to mathematical models to navigate the pandemic and help take critical decisions. These models, which extrapolate from existing data to predict the progression of an infectious disease outbreak, have come to play an integral role in infectious disease epidemiology.Such modeling helps one understand the trajectory of a disease over time, how fast it is increasing and what might determine that increase. Herebyanalyzing aCo integrated Movement of Covid Cases with cured and deaths in Kerala. Vector Auto Regressive model were also tried to fit the model.*

Key Words: Covid, Time Series, Cointegration, Engle-Granger, Vector Auto Regression

INTRODUCTION

Coronavirus disease (COVID-19) is an infectious disease caused by a newly discovered coronavirus. Most people who fall sick with COVID-19 will experience mild to moderate symptoms and recover without special treatment.The virus that causes COVID-19 is mainly transmitted through droplets generated when an infected person coughs, sneezes, or exhales. These droplets are too heavy to hang in the air, and quickly fall on floors or surfaces. It can be infected by breathing in the virus if you are within close proximity of someone who has COVID-19, or by touching a contaminated surface and then your eyes, nose or mouth.

Materials and Methods

A novel concept of cointegration techniques to forecast number of covid cases based on lagged data of other series was tried. The data were collected from the official website of health department. Cointegration was developed by the Nobel laurate Granger (1986). The Engle - Granger two-step approach is a residual based cointegration procedure which uses the ordinary least squares in the estimation process. In general, economic, financial and accounting variables are non-stationary. However, there may be a long-run relationship between variables which is stationary. Hence, we need another measure of the degree of association between variables that takes into account the possibility of a series not being jointly stationary in the short run, but which has a long-run equilibrium implying cointegration. The simple idea behind cointegration

* Corresponding author –mail id : *cbpwarrier@gmail.com*

is that sometimes lack of stationarity of a multi-dimensional process is caused by common stochastic trends which can be eliminated by taking suitable linear combinations of the process thereby making the linear combination stationary. It describes the long-run relationship between variables and results from those variables having a common stochastic trend over time.

Table 1.District wise Population in Kerala and Literacy percentage.

SlNo.	District	Population	Sex Ratio	Literacy	Density
1	Malappuram	4,112,920	1098	93.57%	1157
2	Thiruvananthapuram	3,301,427	1087	93.02%	1508
3	Ernakulam	3,282,388	1027	95.89%	1072
4	Thrissur	3,121,200	1108	95.08%	1031
5	Kozhikode	3,086,293	1098	95.08%	1316
6	Palakkad	2,809,934	1067	89.31%	627
7	Kollam	2,635,375	1113	94.09%	1061
8	Kannur	2,523,003	1136	95.10%	852
9	Alappuzha	2,127,789	1100	95.72%	1504
10	Kottayam	1,974,551	1039	97.21%	895
11	Kasaragod	1,307,375	1080	90.09%	657
12	Pathanamthitta	1,197,412	1132	96.55%	452
13	Idukki	1,108,974	1006	91.99%	255
14	Wayanad	817,420	1035	89.03%	384

(Source :https://www.census2011.co.in/census/state/districtlist/kerala.html)

Table 1.describes the District wise population and literacy rates in Kerala along with the population density. Fromthe table it is clear that the districts Thiruvananthapuram, Alappuzha, Calicut and Malappuram are highly dense and have the high chance to spread the cases. Since literacy rates are high, the state can understand the uptodate details very fact and can do changes needed. The Self cleanliness in Keralitesare world famous and is also a factor to help in safeguard ourselves. Moreover the food habits, timely and ayurvedic food habits, various rituals followed in Kerala based on ayurvedic concepts and moreover highly active health and police departments are also doing well in controlling spread.

The first case of the COVID-19 pandemic in Kerala (which was also the first in all of India) was confirmed in Thrissur on 30 January 2020. The number of active cases initially peaked at 266 on 6 April before declining. For the first time in over 45 days, there were no new cases on 1 May. However, following the return of Keralites from other countries and states, more cases were reported in mid-May, with the biggest single-day spike (240 cases) on 4 July. As of 5 July, there have been 5429 confirmed cases with 3174 (58.46%) recoveries and 25 (0.46%) deaths in the state. Kerala has one of the lowest mortality rates of 0.46% among all states in India.

Table 2.Districtwisecovid cases in Kerala on 10th July 2020.

District	Total Cases	Recoveries	Death	Active
Alappuzha	451	212	1	237
Ernakulam	388	194	2	191
Idukki	175	84	0	91
Kannur	651	373	4	274
Kasaragod	549	415	0	134
Kollam	444	276	2	166
Kottayam	294	172	0	122
Kozhikode	370	224	2	144
Malappuram	856	416	5	435
Palakkad	748	510	1	236
Pathanamthitta	406	219	1	185
Thiruvananthapuram	505	197	5	303
Thrissur	557	339	3	215
Wayanad	140	77	1	62
Total	**6534**	**3708**	**27**	**2795**

(Source :https://covidindia.org/kerala)

Table 2.shows the District wise number of cases in Kerala upto 10th July 2020. This shows that the number of cases in Kerala is mostly aliens who returned from other countries and different States of India. In such a situation it is also clear that the disease can't attack with out consistent contacts with the patients.

David F. Hendry (2010) detailed about Granger 's co integration. Meuriot, V. (2015) discussed the concept of Granger 's co integration and Granger et.al, (1993) explained Modelling Nonlinear Economic Relationships. Palash Ghoshet.al., (2020) done a State-wise Analysis and Prediction about COVID-19 in India.QasimBukhari and Yusuf Jameel(2020) discussed about "Will Coronavirus Pandemic Diminish by Summer?".Acosta et.al, (2020) explained a Correlation Analysis of Hot Weather and Number of Recovery of the Pandemic Corona virus in The Philippines. Qifang Bi et.al, (2020) studied Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China. Shinde*et al.*(2020)explained a Forecasting Model for Coronavirus Disease.Malhotra and Kashyap,(2020) detailed a Progression of COVID-19 in Indian States. Shayak& Sharma (2020) explained a Retarded logistic equation as a universal dynamic model for the spread of COVID.Shaikhet.al., (2020) detailed a mathematical model of COVID-19 using fractional derivative.

Engle-Granger's Test for cointegration between two variables

Granger and Weiss (1983) and Engle and Granger (1987) have shown that even though a given set of series may be non stationary, there may exist various linear combinations of the individual series that are stationary. The desire to estimate models that combine both short run and long run properties and that at the same time maintain stationarity in all the variables, has prompted a reconsideration of the problem of regression using variables measured in their levels. Engle and Granger (1987) give the formal definition of cointegration of two variables. According to that definition, cointegration between two variables occurs when two series are each integrated of order *b* (I(*b*)), but some linear combination of the two series results in a third series which is integrated of order a (I(*a*)), *a*<*b*. In this case, the two series following I(*b*) are said to be cointegrated. Stationarity of the series can be tested by Augmented Dickey-Fuller Tests before cointegration.

Augmented Dickey Fuller Test for Stationarity

A non-stationary variable has a definite positive or negative trend and so mean, variance, and covariance are changing over time, so that standard t tests in regression are no longer valid. Dickey Fuller Test (DFT) assumes the model $Y_t = \rho Y_{t-1} + e_t$, where e_t is assumed to define a sequence of independently and identically distributed (i.i.d.) random variables with expected value zero and variance σ^2. The process in equation is stationary when ρ is less than one in absolute value; i.e., $-1 < \rho < 1$. The AR (1) process has a unit root if and only if ρ is one. In such a situation, the AR (1) process is nonstationary. But if the errors are dependent, it will be a violation of DF test. So to assure that the time series is non- stationary, Augmented Dickey-Fuller unit root test is used for the determination of the order of differencing and stationarity of the independent variable.

The long run relationship between number of Covid cases confirmed and Number of Covid cases cured, number of Covid cases confirmed and Number of Covid cases deaths, number of Covid cases cured and Number of Covid cases deathsfrom the daily data over a period from 30.01.2020 to 15.06.2020 was analyzed using cointegration technique.

Result and Discussion

In the present study, ACF of all the series showed a non-stationary pattern as it remained significant over half a dozen or more lags rather than quickly declining to zero. To assure the time series was non- stationary, augmented Dickey-Fuller unit root test was additionally used for testing stationarity of the variable using the computer software gretl 1.9. Results of the ADF test were supporting the results of ACF. According to this test it was seen that the entire variables were found to be non-stationary as all the probabilities were above 0.05. The differenced series were also nonstationary. But the second differenced series were found to be stationary. Hence they are found to be Integrated.

Hence in the present study the Engle and Granger (1987) two step procedures for modeling the relationship between co-integrated variables has been employed. The co-integration between the two variables was tested by conducting the ADF test on residuals obtained from running the OLS regression, called the co-integrating regression. Co-integration theory suggests that if two non-stationary time-series are co-integrated, residuals of the linear combination of these two non-stationery series are stationary. Therefore, co-integrated series

indicate stable long-run relationship between them. Evidence of co-integration between the variables indicated that there was a stable long-run relationship between them.

Table 3.gives the parameter estimates for the prediction model for forecasting number of covid cases expected based on number of cured and death cases 1 day before. Table 4.gives the parameter estimates for the prediction model for forecasting number of covid cases expected based on previous days cases using Vector Auto Regression (VAR) modeling.

Table 3. Prediction model for Confirmed from Cured+Death data

Regressor	Coefficient	Std.error	t-ratio	p-value
Constant	31.7017	33.1272	0.9570	0.3407
Recovered	1.78813	0.0273072	65.48	2.38e-089 ***

Since the residuals were stationary at lag 1, the variables could be cointegrated and the cointegration model could be defined by the relation

$$C_t = 1.78813 * S_{t-1} + 31.7017 \quad (*)$$

Where $S_t$denotes the number of cases cured and death and C_t the Confirmed patients at the t^{th}day. Since adjusted R^2 is 0.974982, 97.49% of the variations in Confirmed was due to the variation in Cured and death cases only and can be explained through the above model.

Vector Auto regression model to predict number of covid cases from past data.

Table 4. Parameters of the VAR model for forecasting number of covid cases

	coefficient	std. error	t-ratio	p-value
Constant	5.08998	2.26036	2.252	0.0264
Confirmed_Ker_1	0.0349745	0.00762192	4.589	1.23e-05
Recovered_Ker_1	0.994619	0.0153602	64.75	4.64e-087
Time	−0.309660	0.0995302	−3.111	0.0024

The model can be written as,

$$C_t = 0.0349745 * C_{t-1} + 0.994619 * S_{t-1} - 0.30966t + 5.08998 \quad (**)$$

Where C_t the Confirmed patients and S_t denotes the number of recovered cases adding number cured and death on the t^{th} day.

The R^2 for the model was found to be 0.9993 indicating 99.93% variations can be explained through this model. The actual estimated cases were given in fig 1. Since the actual and expected values go side by side the model is very good for prediction purposes.

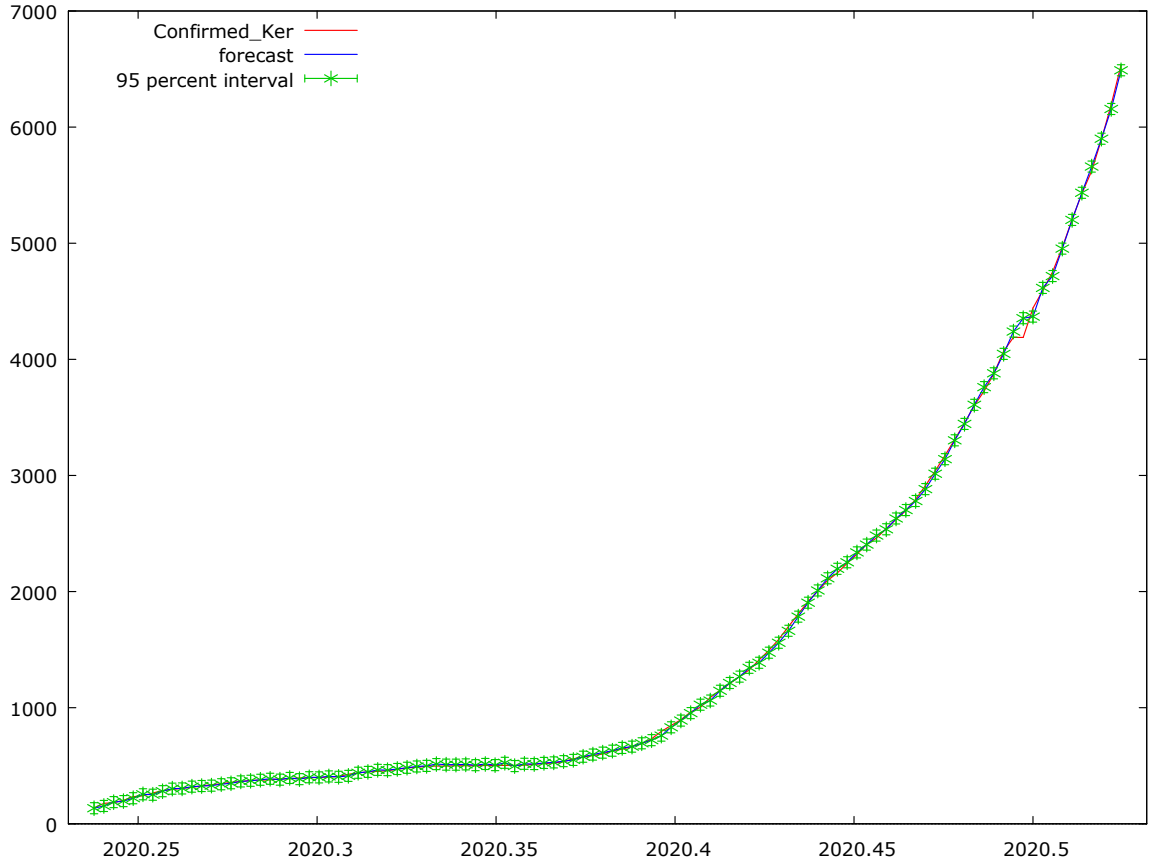

Figure 1. Actual and Estimated number of covid cases using VAR modeling.

Conclusion

The cointegration model and Vector auto correction model can predict next day's number of cases with high accuracy. It is really significant in this present situation. Since the R^2 value of VAR model is high, and it incorporates a time trend with the data, it is best suited for prediction purposes.

Reference

1. Acosta, Dianne and Alquizar, Maria Lavelle and Junio, CJ Alexes and Buladaco, Mark Van. (2020), Correlational Analysis of Hot Weather and Number of Recovery of the Pandemic Coronavirus in The Philippines (May 9, 2020). Available at SSRN: https://ssrn.com/abstract=3597254 or http://dx.doi.org/10.2139/ssrn.3597254

2. Bi, Q., Wu, Y. and Mei, S. (2020), Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study, https://www.thelancet.com/action/showPdf?pii=S1473-3099%2820%2930287-5
3. Bukhari Q, Jameel Y. (2020) Will Coronavirus Pandemic Diminish by Summer? SSRN Electron J. doi:10.2139/ssrn.3556998
4. David F. Hendry (2010) Granger 's and co integration, Oxford University, UK on Nobel Prize Talk about Clive Granger to the American Economic Association Meeting in Philadelphia.http://www.nuffield.ox.ac.uk/Users/Hendry/Papers/DFHonCliveJFEC09.pdf
5. Ghosh, P., Ghosh, R. and Chakraborty, B. (2020), COVID-19 in India: State-wise Analysis and Prediction medRxiv 2020.04.24.20077792; doi: https://doi.org/10.1101/2020.04.24.20077792
6. Granger, C. W. J., and Ter¨asvirta, T. (1993). Modelling Nonlinear Economic Relationships. Oxford: Oxford University Press.
7. Malhotra,B. and Kashyap,V.(2020) "Progression of COVID-19 in Indian States - Forecasting Endpoints using SIR and Logistic Growth Models,", pp. 2014–2015, https://doi.org/10.1101/2020.05.15.20103028.
8. Meuriot, V. (2015). The concept of co integration: the decisive meeting between Hendry and Granger (1975). *Cahiers d'économiePolitique*, 68(1), 91-118. doi:10.3917/cep.068.0091.
9. Shaikh, A.S., Shaikh, I.N. and Nisar, K.S. (2020), A mathematical model of COVID-19 using fractional derivative. Outbreak in India with dynamics of transmission and control preprints.org
10. Shayak, B.SharmaandMohit (2020). Retarded logistic equation as a universal dynamic model for the spread of COVID-19. 10.1101/2020.06.09.20126573.
11. Shinde, G.R., Kalamkar, A.B., Mahalle, P.N. *et al.* Forecasting Models for Coronavirus Disease (COVID-19): A Survey of the State-of-the-Art. *SN COMPUT. SCI.* **1,** 197 (2020). https://doi.org/10.1007/s42979-020-00209-9