Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/120230

ARDL Bound Test Cointegration Modeling For COVID-19 Infected Cases And Deaths

A. Rajarathinam and J. B. Anju*

Department of Statistics, Manonmaniam Sundaranar University, Tirunelveli – 627 012, Tamil Nadu State, India

Received: 22 Jun. 2022, Revised: 22 Aug. 2022, Accepted: 12 Oct. 2022
Published online: 1 May 2023

Abstract: The main purpose of this paper is to investigate the significant long-run and short-run dynamic relationships between the cumulative numbers of COVID-19 infected cases and deaths due to COVID-19 infections as of 31^{st} May 2021, starting from 7^{th} March 2020. Furthermore, the stability of the estimated model parameters is studied. To assess the consistency of the model parameters, the cumulative sum of recursive residuals test and the cumulative sum of recursive residuals squares tests are used. Additionally, cointegration equations such as the Fully Modified Ordinary Least Square, Dynamic Ordinary Least Squares, and Canonical Cointegration Regression are applied to check the long-run elasticities in the concerned relationship.

Keywords: Autoregressive distributed lag model, Bounds cointegration test, Error correction model, Residual diagnostics, Stability tests, Unit root tests

1 Introduction

1.1 Background of the study

The first case of a COVID-19 infections in Tamil Nadu was identified on 7^{th} March 2020. Tamil Nadu ranks fifth in states with the highest number of confirmed cases in India, after Maharashtra, Karnataka, Andhra Pradesh, and Kerala. All 37 districts in Tamil Nadu have been affected by the COVID-19 epidemic, with the capital district of Chennai being the most affected region.

The initial increase in COVID-19 infected cases in Tamil Nadu was considered due to the cluster of cases linked to the Tablighi Jamaat religious congregation in Delhi in early April 2020. Koyambedu in Chennai was identified as another heavily affected place that caused the surge in May 2020.

To understand the disease dynamics and to make appropriate decisions to control the disease, knowledge of the number of COVID-19 infected cases and the number of deaths due to COVID-19 infections and an estimation of the long-run equilibrium relationship between infected cases and deaths are essential for decision-makers.

1.2 Objectives of the present study

The main objectives of the present investigations are to investigate the short-run and long-run cointegration relationships between the cumulative number of new COVID-19 infected cases and the cumulative number of deaths due to COVID-19; to estimate the long-run equilibrium relationship between these using an autoregressive distributed lag model (ARDL) and bounds cointegration tests, and to study the stability of the model parameters.

1.3 ARDL model

The model has p lags of the dependent variable and q lags of the independent variable:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \ldots + \beta_p y_{t-p} + \alpha_0 x_t + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \ldots + \alpha_q x_{t-q} + \mu_t \quad (1)$$

* Corresponding author e-mail: rajarathinam@msuniv.ac.in

$$y_t = \beta_0 + \sum_{i=1}^{p} \beta_i y_{t-i} + \sum_{i=0}^{q} \alpha_i x_{t-i} + \mu_t \quad (2)$$

where μ_t is a random "disturbance" term. Here $\beta_1, \beta_2, \beta_3, ..., \beta_p$ are called long-run dynamics and $\alpha_1, \alpha_2, \alpha_3, ..., \alpha_q$ are short-run coefficients

The model is "autoregressive" in the sense that y_t is "explained" (in part) by lagged values of itself. It also has a "distributed lag" component in the form of successive lags of the "x" explanatory variable. Sometimes, the current value of x_t itself is excluded from the distributed lag part of the model's structure,Soharwardi et al. [1].

2 Materials and Methods

2.1 Materials

The cumulative total number of COVID-19 infections and deaths as of 31th May, 2021, starting on 9th March 2020, were collected from the official website, https://stopcorona.tn.gov.in, maintained by the Health and Family Welfare Department, Government of Tamil Nadu, India. Several statistical methodologies were used to achieve the objectives of the present study. Here, the cumulative number of COVID-19 infected cases is denoted by X (CASES) (independent variable), and the cumulative total number of deaths due to COVID-19 infections is denoted by Y (DEATHS) (dependent variable), which are the study variables. In addition, EViews Ver. 11 software was used to estimate the model parameters, error diagnostics checks and to study the stability of the model.

2.2 Methods

To apply the ARDL model, the study variables should fulfill certain stationarity conditions. That is, the variables should be purely I(0), purely I(1) or I(0)/I(1), Alimi [2]. To test this, three different tests, viz., the Dickey and Fuller [3], Phillips and Perron [4], and Kwiatkowski et al. [5] tests were used. The Akaike information criterion (AIC) was used to select the optimal lag. The Jarque-Bera test [6] is used to test the normality of the residual's. For testing for autocorrelation and serial correlation, the Ljung-Box test [7] and the Breusch-Godfrey test [8],[9], respectively, were used. The Breusch-Pagan-Godfrey heteroscedasticity test [10] was used to test the heteroscedasticity. Model stability was studied based on the cumulative sum of recursive residuals (CUSUM) and cumulative sum of recursive residuals squares (CUSUMSQ) tests [11]. Finally, to test the cointegration (long-run relationship), the bounds test [12] was employed.The long-run elasticities among predetermined variables were analyzed with the Fully Modified Ordinary Least Square (FMOLS) as proposed by Phillips and Hansen [13], and Dynamic Ordinary Least Squares (DOLS) technique as suggested by Stock and Watson [14], and Canonical Cointegration Regression (CCR) were applied to check the long-run elasticities in the concerned relationship. Details of these methods have been omitted in this paper and are available extensively in the literature.

3 Results and Discussions

In this section, the empirical findings, and their interpretations are discussed in sequence.

3.1 Unit root test

The results presented in Tables 1 and 2 reveal that the ADF and PP tests statistics values are significant at a 5% level of significance, and both the study variables, CASES, and DEATH, are found to be stationary without differencing and are hence they are of order I(0).

3.2 Summary statistics

The results presented in Table 3 reveal that the study variables are not normally distributed since the Jarque-Bera statistics values are found to be significant at 1% level of significance. Deaths have more range value than cases. The study variables are positively leptokurtic according to the kurtosis values 28.04 and 25.21.

Table 1: Characteristics of Augmented Dickey-Fuller test at level

Variables	Intercept	Intercept & Trend	None
Cases	-3.91* (0.004)	-4.32* (0.008)	-8.28** (0.000)
Death	-4.52* (0.000)	-4.94* (0.001)	-3.71* (0.000)

** 1% level of significance;*5% level of significance ;Figures in the () represents p -values.

Table 2: Characteristics of Phillips-Perron test at level

Variables	Intercept	Intercept & Trend	None
Cases	-3.98* (0.004)	-4.40* (0.007)	-3.04** (0.003)
Death	-4.52* (0.000)	-4.94* (0.001)	-3.71* (0.000)

** 1% level of significance;*5% level of significance ;Figures in the () represents p -values.

Fig.1 depicts the cumulative number of COVID-19 infected cases in different districts of Tamil Nadu through 31^{st} May 2021. The most significant number of COVID-19 infections was registered in Chennai (504502), and the lowest registration was in Perumbalur (8430). The total number of COVID-19 infected cases in Tamil Nadu as of 31^{st} May 2021 was 2096516.

Fig.2 depicts the cumulative number of deaths due to COVID-19 in different districts of Tamil Nadu through 31^{st} May 2021. The most notable deaths due to COVID-19 infections were registered in Chennai (7091), and the lowest registration was in Perumbalur (57). The total number of deaths due to COVID-19 in Tamil Nadu as of 31^{st} May 2021 was 24232.

Fig.3 depicts the cumulative death rate due to COVID-19 in different districts of Tamil Nadu through 31^{st} May 2021. The highest death rates due to COVID-19 infections were registered in Vellore (1.76%) and Tirupattur (1.55%), and the lowest registration was in Nilgiris (0.49%). The overall cumulative death rate due to COVID-19 in Tamil Nadu as of 31^{st} May 2021 was nearly 1.6%.

Table 3: Summary statistics

Statistics	DEATHS	CASES
Mean	637.68	55171.47
Median	349.50	35475.50
Maximum	7091.00	504502.00
Minimum	2.00	2507.00
Std. Dev.	1145.20	81985.45
Skewness	4.91	4.58
Kurtosis	28.04	25.21
Jarque-Bera	1145.91	913.67
Probability	0.00	0.00
Sum	24232.00	2096516.00

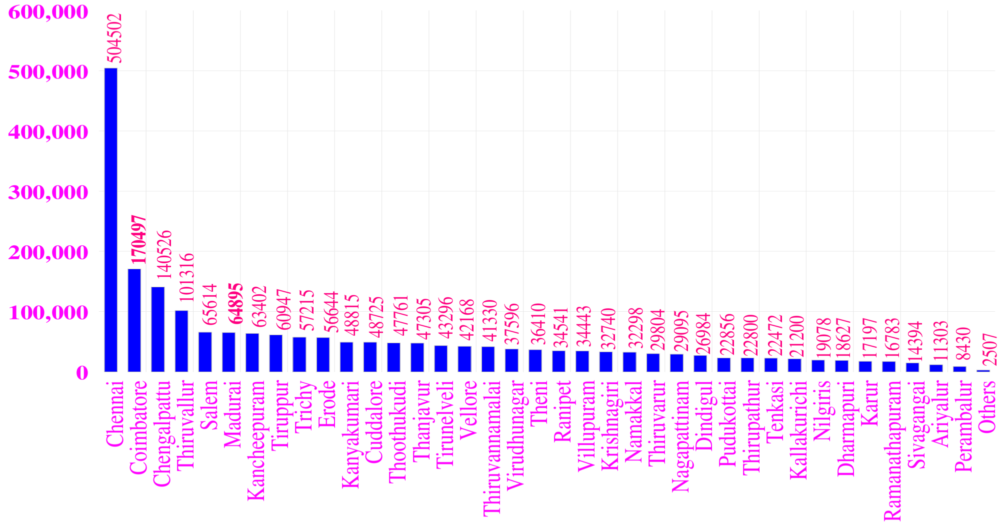

Fig. 1: Cumulative number of COVID-19 infected cases

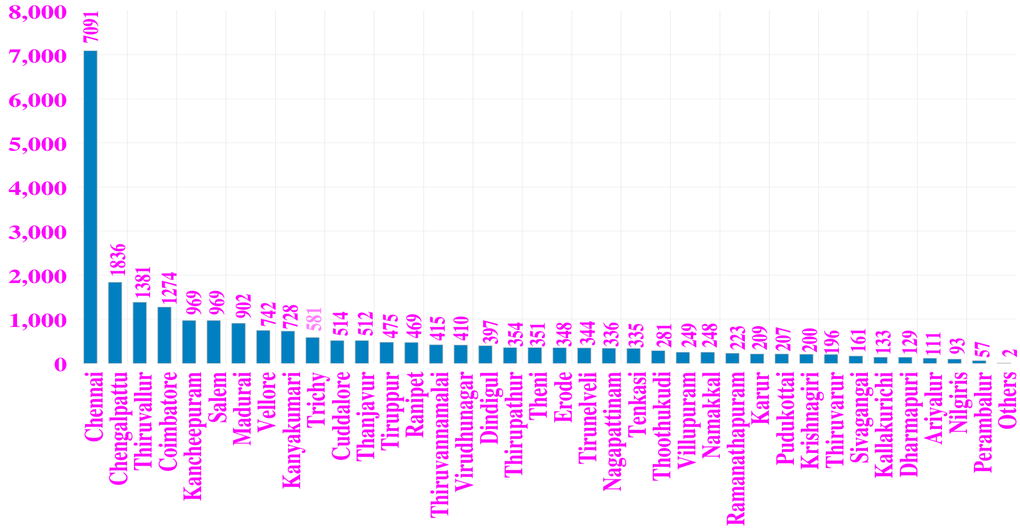

Fig. 2: Cumulative number of deaths due COVID-19 infections

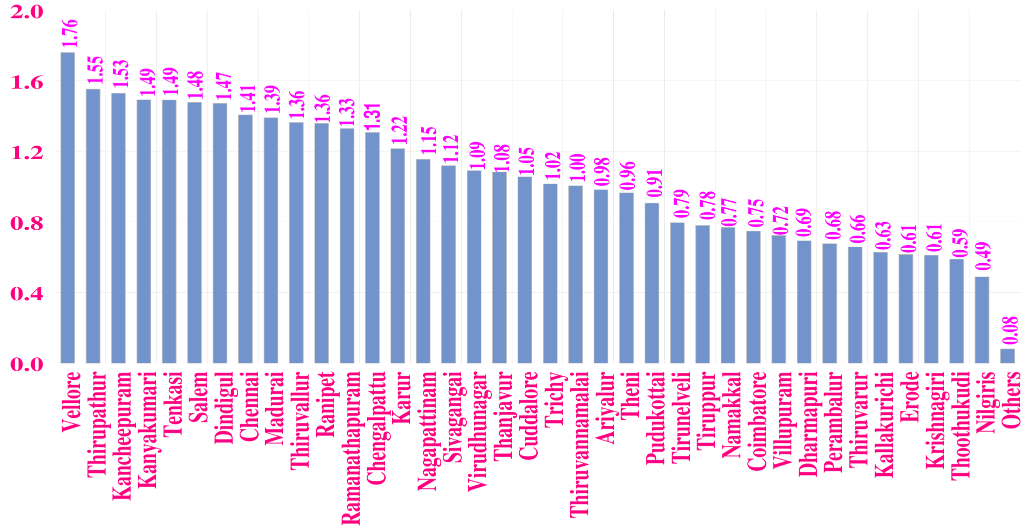

Fig. 3: Death rate due to COVID-19 infections

3.3 Model selection

To choose the optimal lag values for p and q, the Akaike information criterion (AIC) was calculated for the different values of p and q. The lower the AIC values are, the better the lag values for p and q. Fig.4 depicts that the AIC value is extremely low for the lags p=1 and q=0. Accordingly, the ARDL(1,0) is found an appropriate model among the 20 models investigated with different lag values.

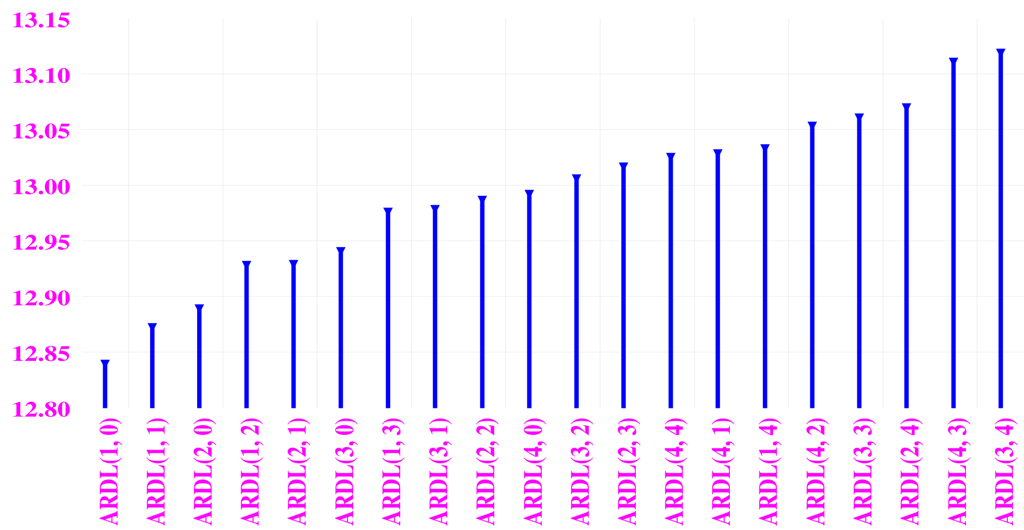

Fig. 4: Selection of the appropriate model based on the AIC

3.4 The ARDL(1,0) model

The ARDL(p=1, q=0) model is employed to study the short-run relationship between the cumulative number of COVID-19 infected cases, and the cumulative deaths due to COVID-19 infections. The findings are reported in Table 4. The results reveal that the overall goodness of fit of the model, as shown by the coefficient of determination, R^2 = 98%, is extremely high and highly significant, implying that the model explains almost 98% of the variation in the dependent variable and the rest is explained by the error term. The value of the D-W statistic is nearly equal to two, which confirms that there are no spurious results. Here the coefficient of the variable DEATH at lag 1 and the coefficients of the variable CASES are highly significant at 1% level of significance. The slope value is negative and significant at a 5% level of significance The estimated ARDL(1,0) model is

DEATHS = -0.14517** DEATHS(-1) + 0.01455** CASES - 72.8137*

Table 4: Characteristics of estimated ADRL model

Variable	Coefficient	Std. Error	t-Statistic	Prob.*
DEATHS(-1)	-0.145130	0.023528	-6.168271	0.0000
CASES	0.014546	0.000329	44.27287	0.0000
C	-72.81371	31.02838	-2.346681	0.0249
R-squared	0.984153	Mean dependent var		651.9189
Adjusted R-squared	0.983220	S.D. dependent var		1157.588
S.E. of regression	149.9492	Akaike info criterion		12.93608
Sum squared resid	764482.4	Schwarz criterion		13.06669
Log-likelihood	-236.3174	Hannan-Quinn criteria.		12.98212
F-statistic	1055.733	Durbin-Watson stat		1.595202
Prob(F-statistic)	0.000000			
*Note: p-values and any subsequent tests do not account for model selection				

3.5 Test for normality of the residuals

Fig.5 illustrates that the errors are normally distributed, as the Jarque-Bera test statistic's value, 1.4956 is non-significant (p=0.4734) at a 5% level of significance. The calculated values of skewness and kurtosis of the residuals are -0.4906 and 2.9144, respectively.

To ensure the consistency of the ARDL(1,0) model, the following residual diagnostic tests are carried out.

3.6 Ljung-Box test for autocorrelation

The results of the Ljung-Box test [7] indicate that the p-values of the Q statistics are non-significant at a 5% significance level and strongly suggest the absence of autocorrelation in the model error. If there is an autocorrelation of residuals, estimated parameters will not be consistent, due to the lagged dependent variable appearing as an exogenous variable in the model.

3.7 Breusch-Godfrey serial correlation LM test

Usually, when an analysis involves time series data, the possibility of autocorrelation is high. Therefore, it is necessary to test the residuals for autocorrelation using the Breusch-Godfrey LM test. The results presented in Table 5 reveal that the F-statistic value is non-significant at a 5% level of significance, and hence the null-hypothesis of no serial correlation is accepted, and therefore there is no serial correlation.

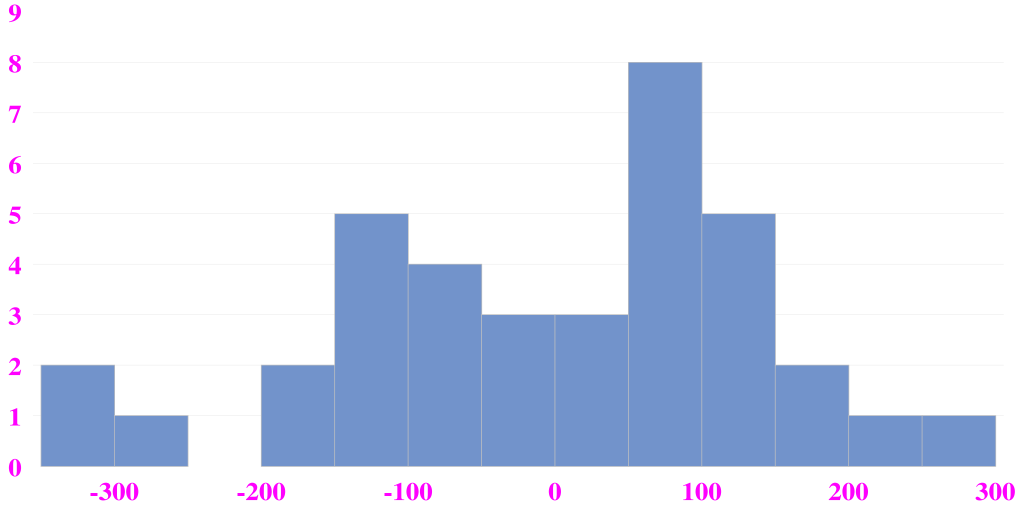

Fig. 5: Test for normality of residuals

Table 5: Characteristics of Autocorrelation's residuals

Lag	AC	PAC	Q-Stat	Prob
1	-0.030	-0.030	0.0358	0.850
2	0.223	0.223	2.0949	0.351
3	-0.017	-0.005	2.1065	0.551
4	0.091	0.043	2.4685	0.650
5	-0.161	-0.161	3.6321	0.603
6	-0.193	-0.247	5.3735	0.497
7	-0.132	-0.091	6.2100	0.515
8	-0.010	0.082	6.2147	0.623
9	-0.030	0.065	6.2596	0.714
10	-0.049	-0.046	6.3888	0.782
11	-0.015	-0.092	6.4007	0.845
12	-0.012	-0.108	6.4092	0.894
13	-0.080	-0.128	6.7958	0.912
14	-0.064	-0.043	7.0550	0.933
15	-0.048	0.005	7.2079	0.952
16	-0.065	-0.065	7.4984	0.962
AC-Autocorrelation; PAC-Partially Autocorrelation				

3.8 Breusch-Pagan-Godfrey heteroscedasticity test

To ensure consistency, the study further employed the Breusch-Pagan-Godfrey heteroscedasticity test, and the test results presented in Table 6 reveal that the F-statistics value is non-significant at a 5% level of significance, the null- hypothesis of no heteroscedasticity is accepted. Hence it shows that the error variance is constant, which is the desirable quality of the fitted model.

3.9 Fit of the model

The estimated plot of the identified ARDL(1,0) model is depicted in Fig.6 which shows that the model's fit is appropriate enough to explain the cumulative total deaths.

Table 6: Characteristics of the Breusch-Godfrey serial correlation LM test of the residuals

F-statistic	0.842467	Prob. F(2,32)	0.4400
Obs*R-squared	1.850755	Prob. Chi-Square(2)	0.3964

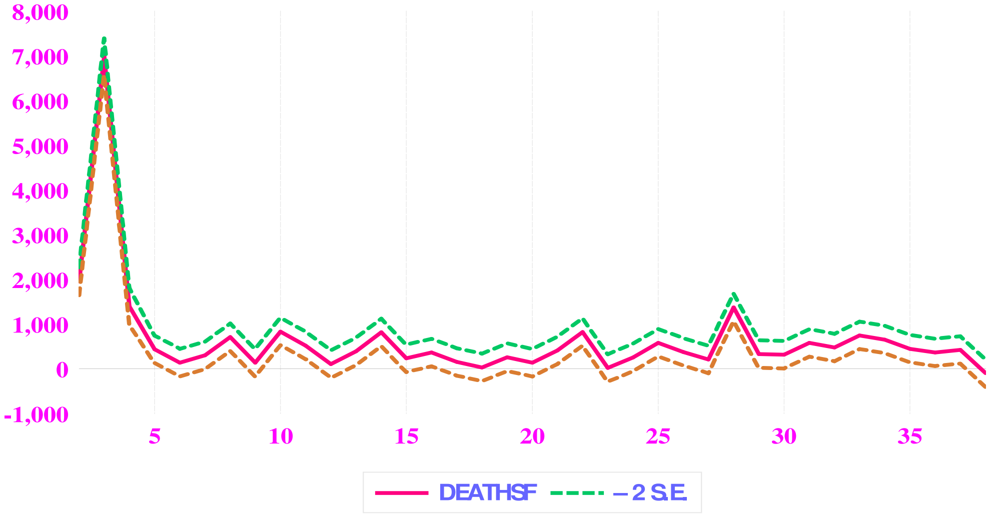

Fig. 6: Model Fit

3.10 Model stability

To check the robustness of the results obtained, structural stability tests of the parameters of the long-run results are performed by the cumulative sum of recursive residuals (CUSUM) and cumulative sum of recursive residuals squares (CUSUMSQ) tests (Brown et al. [11]). This exact procedure has been utilized by Pesaran and Pesaran [15] and Mohsen et al. [16] to test the stability of long-run coefficients. A graphical representations of the CUSUM and CUSUMSQ statistics are depicted in Fig.7 and 8, respectively. The plots of the CUSUM and CUSUMSQ are within the boundaries (indicated by the dotted red lines) of the 5% level of significance, and these statistics confirm the model's stability.

3.11 Bounds test for cointegration

The bounds test Pesaran et al. [12] is employed to test the cointegration (long-run relationship) between the study variables CASES and DEATHS and are presented in Table 7. The test results reveal that there exists a cointegration relationship between CASES and DEATHS, as the bounds test statistic is greater than the upper bound (F-statistics = 1037.636 > 5.58), and it is highly significant at a 1% level of significance. Hence the null-hypothesis of "No Levels Relationships" is rejected, which implies the possibility of estimating a log-run cointegration relationship between the study variables.

Table 7: Characteristics of F-bounds test

Test Statistic	Value	Signif.	Lower Bound I(0)	Upper Bound I(1)	Conclusion
F-statistic	1037.636	10%	3.02	3.51	Cointegration
k	1	5%	3.62	4.16	
		2.5%	4.18	4.79	
		1%	4.94	5.58	

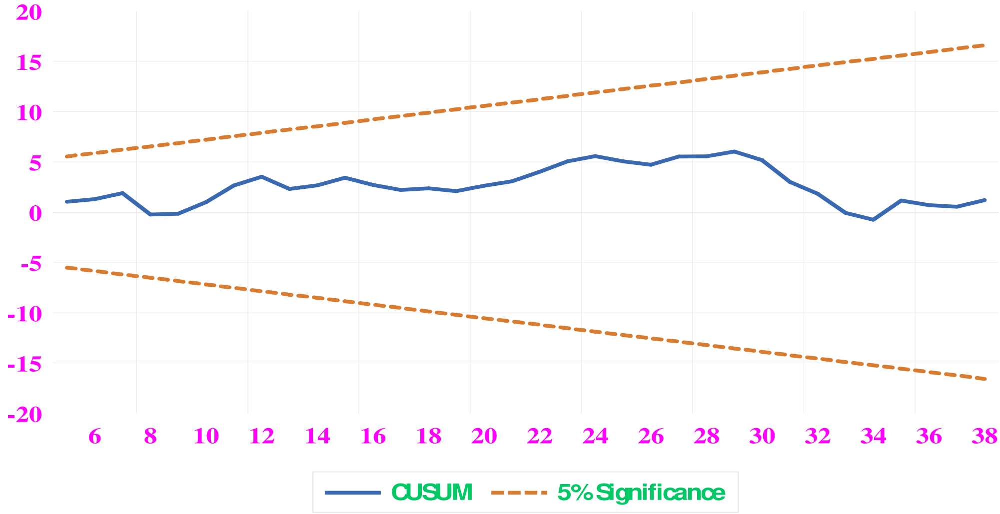

Fig. 7: CUSUM stability test

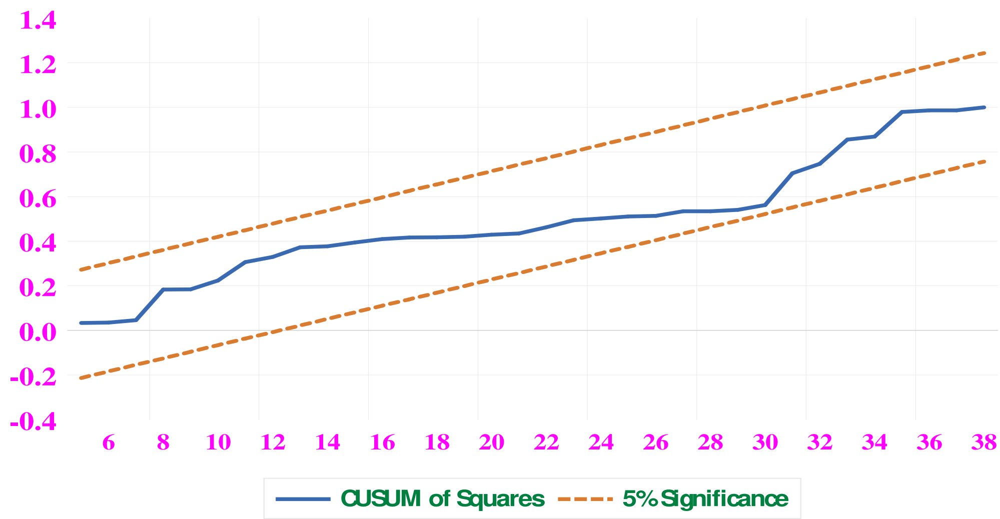

Fig. 8: CUSUMSQ stability test

The conditional error correction regression model is presented in Table 8. All the estimated parameters are highly significant at a 1% significance level. Here the variable ECM(-1) is called the error correction model, and its coefficient value should be negative and significant, which is one of the desirable qualities of the model. ECM(-1) corresponds to the lagged error term equilibrium equation. The coefficient expresses the degree to which the variable DEATH will be recalled towards the long-term target. It is negative (-1.145130) and significant at a 1% level of significance, thus reflecting a relatively quick long-term target adjustment.

The results presented in Table 9 are the estimates of the long-run variables, and the Error Correction (EC) equation is given at the end of the table.

Table 8: Characteristics of conditional error correction regression

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-72.81371	31.02838	-2.346681	0.0249
ECM(-1)	-1.145130	0.023528	-48.66997	0.0000
CASES	0.014546	0.000329	44.27287	0.0000

Table 9: Characteristics of levels equation

Variable	Coefficient	Std. Error	t-Statistic	Prob.
CASES	0.012703	0.000302	42.03245	0.0000
C	-63.58554	27.46597	-2.315066	0.0268

The estimated error correction equation is EC = DEATHS - (0.012703*CASES – 63.5855) and the estimated cointegration equation is D(DEATHS) = -1.1451**(DEATHS(-1) - (0.0127**CASES - 63.5855*)).

The results presented in Table 10 show that the error correction model estimates the speed of adjustment to equilibrium in a cointegration relationship. Here, the error correction term derived as the Levels Equation earlier is included among the regressors and is denoted as CointEq. The coefficient associated with this regressor is typically the speed of adjustment to equilibrium in every period. Here the coefficient of CointEq is negative and highly significant, which are the desirable qualities of the model. Thus, both the variables under study are moving in an opposite positive direction.

Table 10: Characteristics of ARDL ECM regression

Variable	Coefficient	Std. Error	t-Statistic	Prob.
CointEq(-1)*	-1.145130	0.019946	-57.41097	0.0000
R-squared	0.989196	**Mean dependent var**		-2.945946
Adjusted R-squared	0.989196	**S.D. dependent var**		1401.955
S.E. of regression	145.7245	**Akaike info criterion**		12.82797
Sum squared resid	764482.4	**Schwarz criterion**		12.87151
Log-likelihood	-236.3174	**Hannan-Quinn criteria.**		12.84332
Durbin-Watson stat	1.595202			

3.12 Long-Run Elasticities

As cointegration exists among the study variables, long-run elasticities are estimated with the FMOLS, DOLS and CCR equations by considering the number of deaths due to COVID-19 as the regress and the number of COVID-19 infected cases as the regressor. The results are reported in Table 11. Among the three different models, the DOLS model has the highest adjusted values and coefficient of determination (R^2=98%) compared to other models. Furthermore, it shows that 1% increase in COVID-19 infected cases, the death rate would be increased by 1.3%.

4 Conclusion

The most significant number of new COVID-19 infections was registered in Chennai (504502), and the lowest registration was in Perumbalur (8430). The total number of COVID-19 infected cases in Tamil Nadu as of 31st May 2021 was 2096516. The most notable deaths due to COVID-19 infections were registered in Chennai (7091), and the lowest registration was in Perumbalur (57). The total number of deaths due to COVID-19 in Tamil Nadu as of 31st May 2021 was 24232. The highest death rates due to COVID-19 were registered in Vellore (1.76%) and Thrupathur (1.55%), and the lowest registration was in Nilgiris (0.49%). The overall cumulative death rate due to COVID-19 in Tamil Nadu as of 31st May 2021 was nearly 1.6%. The ARDL(p=1, q=0) model is highly significant, and the value of the coefficient of determination, R^2 = 98%, implies that the model explains almost 98% of the variation in the dependent variable and that the rest is explained by the

Table 11: Characteristics of long-run elasticities for the dependent variable (DEATH)

Variables	FMOLS	DOLS	CCE
CASES	0.013484** (0.000377) [35.79617]	0.012945** (0.001148) [11.26439]	0.013312** (0.000490) [27.18775]
Constant	-107.7520* (37.37328) [-2.883128]	-74.53464 (54.55462) [-1.366239]	-98.13312* (41.28413) [-2.377131]
R^2	97 %	98 %	97 %
Adj.R^2	97 %	98 %	96 %
S.E. of Regression	216.2666	153.5710	218.2074

error term. The value of the D-W statistic is nearly equal to two, which confirms that there are no spurious results. The bounds test results reveal a log-run relationship between the study variables. The error correction term is negative and highly significant, reflecting a relatively quick adjustment to the long-term target.

Acknowledgements

Both the authors highly acknowledge the Department of Scince and Technology for providing the Fund for Improvement in S&T Infrastructure (FIST-2018) (No.SR/FST/MS-11/2018/19 dated 20-12-2018) to the Department of Statistics, Manonmaniam Sundaranar University, Tirunelveli-627 012, Tamil Nadu state, India.

References

[1] M.A.Soharwardi, R.E.A.Khan, S.Mushtaq, Long run and the short run relationship between financial development and income inequality in Pakistan, *Journal of ISOSS*, **42**,105-112 (2018).
[2] R.S. Alimi, ARDL Bounds Testing Approach to Cointegration: A Re-Examination of Augmented Fisher Hypothesis in an Open Economy, *Asian Journal of Economic Modelling*, **2**, 103-114 (2014).
[3] D.A. Dickey and W.A. Fuller, Distribution of the Estimators for Autoregressive Time Series with a Unit Root, *Journal of the American Statistical Association*, **74**, 427-431 (1979).
[4] P.C.B. Phillips and P. Perron, Testing for a Unit Root in Time Series Regression, *Biometrika*, **75**, 335-346 (1988).
[5] D. Kwiatkowski,P.C.B. Phillips, P.Schmidt and Y. Shin, Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root, *Journal of Econometrics*, **54**, 159-178 (1992).
[6] C.M. Jarque and A.K.Bera, Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals, *Economics Letters*, **6**, 255-259 (1980).
[7] G.M. Ljung and G.E.P.Box, The Likelihood Function of Stationary Autoregressive-Moving Average Models, *Biometrika*, **66**, 265-270 (1979).
[8] T.S. Breusch, Testing for Autocorrelation in Dynamic Linear Models, *Australian Economic Papers*, **17**, 334-355 (1978).
[9] L.G. Godfrey, Testing Against General Autoregressive and Moving Average Error Models when the Regressors Include Lagged Dependent Variables, *Econometrica*, **46**, 1293-1301 (1978).
[10] T.S. Breusch and A.R.Pagan, A Simple Test for Heteroscedasticity and Random Coefficient Variation, *Econometrica*, **47**, 1287-1294 (1979).
[11] R.L. Brown,J. Durbin and J.M. Evans, Techniques for Testing the Constancy of Regression Relationships over Time, *Journal of the Royal Statistical Society. Series B (Methodological)*, **37**, 149-192 (1975).
[12] M.H. Pesaran, Y. Shin and R.J. Smith, Bounds Testing Approaches to the Analysis of Level Relationships, *Journal of Applied Economics*, **16**, 289-326 (2001).
[13] P.C.Phillips and B.E.Hansen,Statistical inference in instrumental variables regression with I(1) processes, *Rev.Econ.Stud.*, **57(1)**, 99-125 (1990).
[14] J.H.Stock and M.W.Watson,A simple estimator of cointegrating vectors in higher order integrated systems, *Econometrica*, **61(4)**, 783-820 (1993).

[15] M.H. Pesaran and B. Pesaran, Working with Microfit 4.0: *Interactive Econometric Analysis*, Oxford University Press, Oxford, UK (1997).

[16] Mohsen, Bahmani-Oskooee and R.W.Ng, Long run demand for money in Hong Kong : An application of the ARDL model, *International Journal of Business and Economics*, **1(2)**, 147-155 (2002).