

Estimating Trends in COVID-19 Infected Cases Based on Panel Data Regression Modelling

RAJARATHINAM ARUNACHALAM (✉ arrathinam@yahoo.com)

Manonmaniam Sundaranar University <https://orcid.org/0000-0002-3245-3181>

TAMILSELVAN PAKKIRISAMY

Manonmaniam Sundaranar University <https://orcid.org/0000-0003-4097-8874>

Research Article

Keywords: Panel Regression Model, Least-Squares Dummy Variable, Fixed-Effect Model, Random-Effect Model, Wald Test, Hausman Test

Posted Date: June 30th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-635014/v3>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Estimating Trends in COVID-19 Infected Cases Based on Panel Data Regression Modelling

Rajarithinam, A and Tamilselvan, P

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

Corresponding Author: Dr.A. Rajarithinam, Professor, Department of Statistics, Manonmaniam Sundaranar University, Tirunelveli – 627 008, Tamil Nadu State, India
Tel.+91 8220226545

Running Title: Modelling Trends in COVID-19 Infection Data

Competing Interest: The authors have declared that no competing interests exist.

Abstract

Background and Objective: The novel coronavirus pandemic, known as COVID-19, could not have been more predictable; thus, the world encountered health crises and substantial economic crises. This paper analysed the trends in COVID-19 cases in October 2020 in four southern districts of Tamil Nadu state, India, using a panel regression model. **Materials and Methods:** Panel data on the number of COVID-19-infected cases were collected from daily bulletins published through the official website www.stopcorona.tn.gov.in maintained by the Government of Tamil Nadu state, India. Panel data regression models were employed to study the trends. EViews Ver.11. Software was used to estimate the model and its parameters. **Results:** In all four districts, the COVID-19-infected case data followed a normal distribution. Maximum numbers of COVID-19-infected cases were registered in Kanniyakumari, followed by Tirunelveli, Thoothukudi and Tenkasi districts. The fewest COVID-19 cases were registered in Tenkasi, followed by Tirunelveli, Thoothukudi and Kanniyakumari districts. A random effects model was found to be an appropriate model to study the trend. **Conclusion:** The panel data regression model is found to be more appropriate than traditional models. The Hausman test and Wald test confirmed the selection of the random effects model. The Jarque-Bera normality test ensured the normality of the residuals. In all four districts under study, the number of COVID-19 infections showed a decreasing trend at a rate of 1.68% during October 2020.

Keywords: Panel Regression Model, Least-Squares Dummy Variable, Fixed-Effect Model, Random-Effect Model, Wald Test, Hausman Test

INTRODUCTION

The 2019 COVID-19 pandemic has received much attention, as it has affected most economies worldwide and resulted in innumerable deaths. Because no antiviral drugs or vaccines exist, the number of new coronavirus-affected cases has increased tremendously, and many people have died. The development of various methodologies to analyse these pandemic data has become an especially important research area. The current topic is very much important for investigation since the decision makers can take an appropriate decision to impose the lock down related matters, number of beds needed in hospitals to admit the infected patient, quantity of oxygen needed etc. The present study results would be helpful government, and healthcare communities to initiate appropriate measures to control this outbreak in advance. The trend analysis using sound statistical models are very much important to investigate the pattern in diseases spreading. Hence the current topic is very much important for investigation.

An investigation has been made by estimating hidden models and trends of COVID-19 infected cases in all the 37 districts of Tamil Nadu State, India during the period from 1st August, 2020 to 31st December, 2020 by using different curve fitting tools. The result reported that decreases in trend have been observed in all the district¹. Lockdowns and the effectiveness of reduction in the contacts in Italy have been measured using the modified model. The results showed a decrease in infected people due to stay-at-home orders and tracing quarantine interventions². The prediction, regarding the number of COVID-19 cases in India based on differential equation model. The model shows that the imposition of a countrywide lockdown plays a important role in restraining the spread of the disease³. Log-linear model has been used to estimate the progression of the COVID-19 infection in Tamil Nadu, India. The result indicates that the outbreak is showing decay in the number of infections of the disease which highlights the effectiveness of controlling measure⁴. A study aimed to analyse the spatial distribution of COVID-19 incidence in Brazils municipalities (counties) and investigate its association with socio dimorphic determinants to better understand the social context and the epidemic's spread in the country using the ordinary least squares (OLS), spatial autoregressive model (SAR), conditional autoregressive model (CAR) and the local regression model called multiscale geographically weighted regression (MGWR). The MGWR model fit improved when compared to the OLS, SAR and CAR⁵. A mathematical epidemic model is framed by taking inspiration from the classic Susceptible (S)-Infected (I)-Recovered (R),(SIR) model and develop a compartmental model with ten compartments to study coronavirus dynamics in India with inclusion of factors related to face

mask efficacy contact tracing and testing along with quarantine and isolation⁶. Autoregressive integrated moving average (ARIMA) modeling approach has been used for projecting coronavirus (COVID-19) prevalence patterns in East African countries, mainly Ethiopia, Djibouti, Sudan and Somalia⁷. Dynamic Panel Data Modeling and Surveillance of COVID-19 in Metropolitan area in the United States using the longitudinal trend analysis and Wald test⁸. In ARIMA time series model the variable under study should be stationary. When the variable is not stationary one should convert the original data to the first or second difference. By making the original series into stationary series the estimated trend model is for the transformed series not for the original series. This is one of the main drawbacks of ARIMA time series modelling.

To avoid this problem, in this study Panel Data Regression modeling is used. By combining time series of cross section observations, panel data give “more informative data, more variability, less collinearity among variables, more degrees of freedom with more efficiency”. As Baltagi⁹ stated that compared to pure cross sectional and ARIMA time series, panel data regression estimation is better to identify and measure effects of independent variables on dependent variables, what one cannot measure using time series and cross section data. The functional form of the equation, the slope coefficients, give the percentage change in the dependent variable⁹. By this way, the statistical methodologies used in this paper is more specific, unique which advances a new knowledge in COVID-19 trend analysis.

Based on the above discussion the main purpose of present study is to study the trends in the number of COVID-19 infected cases, i.e., whether the number of cases decreased or increased. Additionally, the impacts of COVID-19 in four different southern districts of Tamil Nadu, namely, Kanniyakumari, Tenkasi, Thoothukudi and Tirunelveli (cross-sections), were investigated by using panel data models. These are the four districts which come under the Manonmaniam Sundaranar University, Tirunelveli, India. Based on the outcome of this study, the University authorities can take an appropriate academic decision whether to go for online class, online examinations, or offline class etc., based on the available COVID-19 infections trends in these four districts. There are nearly ninety affiliated colleges in this University. The result of the present study would be incredibly helpful to take academic related decisions.

MATERIALS AND METHODS

Materials: The COVID-19 dataset was collected from the official Tamil Nadu Government website www.stopcorona.tn.gov.in. In this novel study, only four southern districts of Tamil Nadu, including Kanniyakumari, Tenkasi, Thoothukudi and Tirunelveli, were considered in October 2020.

Methods: Panel data are a type of data that contain observations of multiple phenomena collected over different time periods for the same group of individuals, units, or entities. In short, econometric panel data are multidimensional data collected over a given period.

^{9,10}A simple panel data regression model is specified as

$$Y_{it} = \alpha + \beta X_{it} + v_{it} \quad (1)$$

where v_{it} are the estimated residuals from the panel regression analysis. Here, Y is the dependent variable, X is the independent or explanatory variable, α and β are the intercept and slope, i stands for the i^{th} cross-sectional unit and t for the t^{th} month, and X is assumed to be non-stochastic and the error term to follow classical assumptions, namely, $E(v_{it}) = N(0, \sigma^2)$. In this study, i, the number of cross-sections is 4 (i=1, 2, 3, 4), and t=1, 2, 3, ..., 30. Detailed discussions of panel data models are given in¹¹.

Unit Root Test: Unit roots for the panel data can be tested using either the Levin-Lin-Chu¹² test or the Hadri¹³ LM stationarity test. The null hypothesis is that panels contain unit roots, and the alternative hypothesis is that panels are stationary. In the results, if the p value is less than 0.05, then one can reject the null hypothesis and accept the alternative hypothesis. Similarly, the unit root for the first difference can also be tested using a similar method.

The Constant Coefficients Model : The Constant Coefficients Model (CCM) assumes that all coefficients (intercept and slope) remain unchanged across cross-sectional units and over time. In other words, the CCM ignores the space and time dimensions of panel data. Put differently, under the CCM, the cross-sectional units are assumed to be homogeneous such that the values of intercept and slope coefficients are the same irrespective of the cross-sectional unit being considered. Accepting this homogeneity assumption (also called the pooling assumption), the CCM uses the panel (or pooled) data set and applies the ordinary least squares (OLS) method to estimate unknown parameters of the model. Thus, the CCM is nothing but a straightforward application of OLS to a given panel or pooled data to obtain estimates for unknown parameters of the model.

Individual Specific-Effect Model: Here, it is assumed that there is unobserved heterogeneity across individuals and captured by α_i . The main question is whether the individual-specific effects α_i are correlated with the regressor; if they are correlated, a fixed effects model exists. If these factors are not correlated, a random effects model exists.

Fixed-Effect OR Least-Square Dummy Variable Regression Model: One way to consider the individuality of each district or each cross-sectional unit is to let the intercept vary for each district but still assume that the slope coefficients are constant across districts. The model is written as :

$$^{9,10} y_{it} = \beta_{1i} + \beta_2 X_{it} + v_{it} \quad (2)$$

The difference in the intercept may be due to COVID-19 infections pre-cautionary measure followed in each of the four districts.

The model (2) is known as the fixed effect model (FEM), The term “fixed effects” is due to the fact that, although the intercept may differ across districts, each district’s intercept does not vary over time; that is, it is time invariant.

These fixed effects models can be implemented with the dummy variable technique. Therefore, the fixed effects model can be written as

$$^{10} Y_{it} = \alpha_1 + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \alpha_4 D_{4i} + \beta_1 X_{it} + v_{it} \quad (3)$$

where $D_{2i}=1$ if the observation is from Tenkasi District and is 0 otherwise, $D_{3i}=1$ if the observation is from Thoothukudi and is 0 otherwise, and $D_{4i}=1$ if the observation is from Tirunelveli and is 0 otherwise. Here, α_1 represents the intercept of Kanniyakumari, and $\alpha_2, \alpha_3,$ and α_4 are different intercept coefficients, tell by how much the intercepts of Tenkasi, Thoothukudi and Tirunelveli differ from that of Kanniyakumari District. Since the dummies are used to estimate the fixed effects, the model is also known as the least-squares dummy variable (LSDV) model; hence, one can conclude that the restricted panel regression model is invalid and that the LSDV model is valid.

Random-Effect (RE) Model: The RE model assumes that individual-specific effects α_i are random and one should include α_i in the error term. Each cross-sections have the same slope parameters and a composite error term. So the model (1) become random-effect model (REM)

$$y_{it} = x_{it}\beta + (\alpha_i + v_{it}) \quad (4)$$

Let $\varepsilon_{it} = \alpha_i + v_{it}$, here ε_{it} , α_i and v_{it} are normally distributed with zero means and constant variances σ_ε^2 , σ_α^2 and σ_v^2 , respectively. Hence the $\text{var}(\varepsilon_{it}) = \sigma_\alpha^2 + \sigma_v^2$, and $\text{cov}(\varepsilon_{it}, \varepsilon_{is}) = \sigma_\alpha^2$;

therefore,
$$\rho_\varepsilon = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\varepsilon^2}.$$

Rho is the interclass correlation of the error or the fraction of the variance in the error term due to individual-specific effects. These variables approach 1 if individual effects dominate the idiosyncratic error^{9,10}.

Hausman test¹⁴: The null hypothesis of the Hausman test is that the preferred model includes random effects and not fixed effects. This test determines whether the unique error (α_i) is correlated with the regressor, and the null hypothesis is that they are not correlated. The random effects estimator is highly efficient, so it should be used if the Hausman test supports it. The Hausman test statistic can be calculated only for time-varying regressors and is given as follows:

$$H = \left(\hat{\beta}_{RE} - \hat{\beta}_{FE} \right)' \left(V(\hat{\beta}_{RE}) - V(\hat{\beta}_{FE}) \right) \left(\hat{\beta}_{RE} - \hat{\beta}_{FE} \right)$$

Here $\hat{\beta}_{RE}$ and $\hat{\beta}_{FE}$ are the vector of parameter estimates of random effect and fixed effect, respectively. Under the null hypothesis, this statistic has asymptotically the chi-squared distribution with the number of degrees of freedom equal to the rank of matrix $\left(V(\hat{\beta}_{RE}) - V(\hat{\beta}_{FE}) \right)$.

Wald Test: The Wald test can determine which model variables make significant contributions. The **Wald test** (also called the Wald chi-squared test) is a way to determine if explanatory variables in a model are significant, meaning that they add something to the model; variables that add nothing can be deleted without affecting the model in a meaningful way. The test can be used for a multitude of different models, including those with binary variables or continuous variables. The null hypothesis for the test is: *some parameter = some value*.

Breusch-Pagan Lagrange Multiplier Test: The Breusch-Pagan-Godfrey test is a Lagrange multiplier test of the null hypothesis of no heteroskedasticity, i.e., constant variance among residuals.

Ho: The null hypothesis of the test states that there is constant variance among residuals.

RESULTS AND DISCUSSION

The results obtained in this paper based on applying different statistical tools related to panel regression models are discussed in subsequent sections. This is the first kind of work based on COVID-19 infected case data sets; hence, the current findings are not compared with existing results available in the literature.

Summary Statistics: The results presented in Table 1 reveal that the highest number of COVID-19 infected cases was registered in Kanniyakumari (118), followed by Tirunelveli (86), Thoothukudi (77), and Tenkasi (49) districts. The lowest number of COVID-19 cases registered in Tenkasi (3) was followed by Thoothukudi (17), Tirunelveli (15), and Kanniyakumari (25) districts. In all four districts, the number of COVID-19-infected cases follows a normal distribution since the Jarque-Bera statistical values are nonsignificant at the 5% level of significance.

Category wise Statistics: The categorical number of COVID-19-infected cases is given at the bottom of Table 1. The results show that the number of infected cases in category [0 50) is 7 days in Kanniyakumari district, 31 days in Tenkasi, and 15 days each in Thoothukudi and Tirunelveli districts. In the case of [50 100), category 22 days in Kanniyakumari district and 16 days in each of the Thoothukudi and Tirunelveli districts, whereas the number of cases registered in the category of [100 150), is 2 days in Kanniyakumari districts only.

Table 1: Characteristics of Summary and Category wise Statistics

Statistics	KANNIYAKUMARI	TENKASI	THOOTHUKUDI	TIRUNELVELI
Mean	70.93548	15.61290	51.06452	49.12903
Median	70.00000	10.00000	51.00000	50.00000
Maximum	118.00000	49.00000	77.00000	86.00000
Minimum	25.00000	3.000000	17.00000	15.00000
Std. Dev.	24.61156	11.97686	14.19374	22.86007
Skewness	-0.019244	1.131092	-0.001396	0.023114

Kurtosis	2.154139	3.372580	2.676387	1.508788
Jarque-Bera	0.926077	6.789374	0.135281	2.875059
Probability	0.629368	0.033551	0.934597	0.237514
Sum	2199.000	484.0000	1583.000	1523.000
Sum Sq. Dev.	18171.87	4303.355	6043.871	15677.48

Category	No. of cases			
[0 50)	7	31	15	15 68
[50 100)	22	0	16	16 58
[100 150)	2	0	0	0 2

Fig. 1(a) depicts the total number of COVID-19 cases in all four districts in October – 2020. The highest number of COVID-19-infected cases (2199) was registered in Kanniyakumari district, followed by Thoothukudi (1583), Tirunelveli (1523) and Tenkasi (484). In Tenkasi district, it is exceptionally low due to more awareness among the people, and the district administration might have taken more precautionary measures to prevent COVID-19 infections.

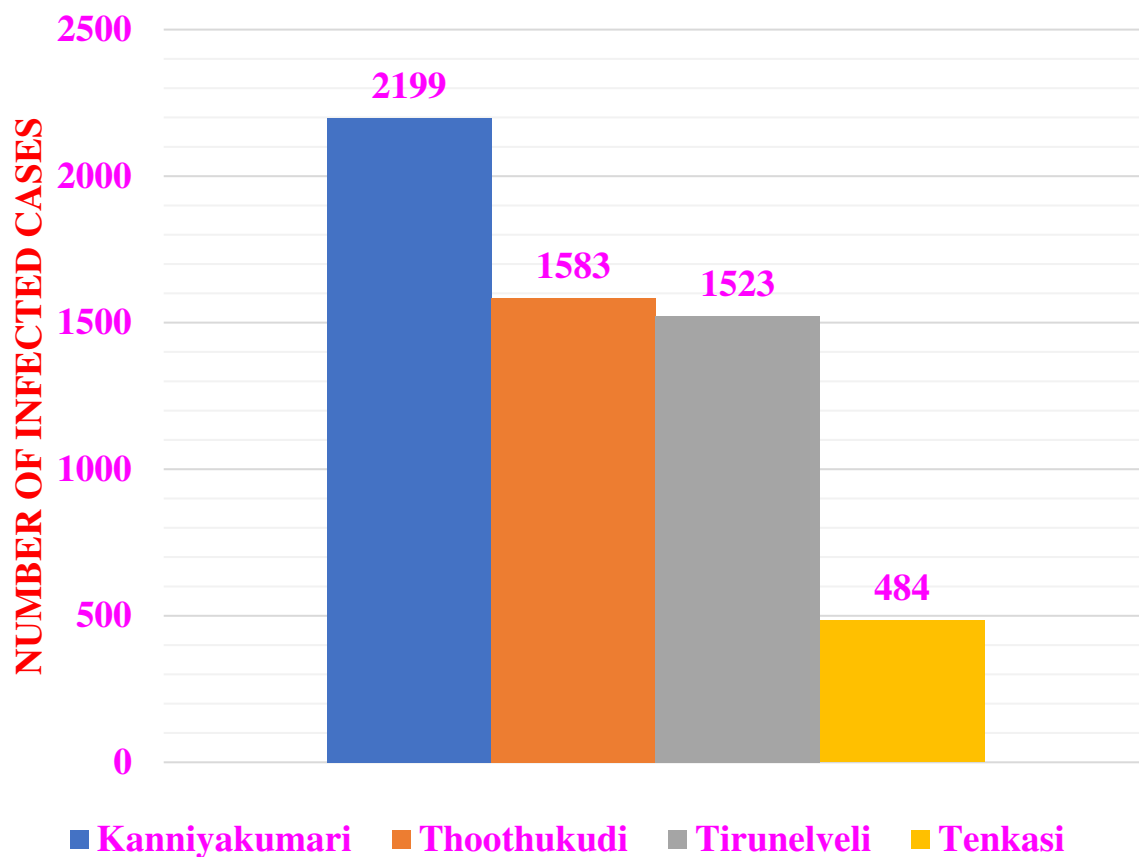


Fig.1(a). Total number of COVID-19 infected cases in all four districts in October 2020

In Kanniyakumari District, Fig. 1(b) depicts that on the 1st and 4th October 2020, the highest numbers of COVID-19 infected cases registered were 118 and 117, respectively, and then they started to decline to 25 cases on 31st October 2020. Step declining trends have been noted in this district.

Rajarathinam *et.al*¹., reported that Holt's linear exponential smoothing with $\alpha=0.1413$ and $\beta=0.1071$ exhibited declining trend in number of infected cases in Kanniyakumari district during the period from 1st August,2020 to 31stDecember 2020. Present result also in agreement with this result.

Yongmei and Gao² by using data-driven modeling reported that evaluation of COVID-19 in Italy decrease due to stay-at-home orders and tracing quarantine interventions.

Roy and Bhattacharya³, also asserted that imposition of a country wide lockdown plays an especially important role in restraining the spread of COVID-19 infections.

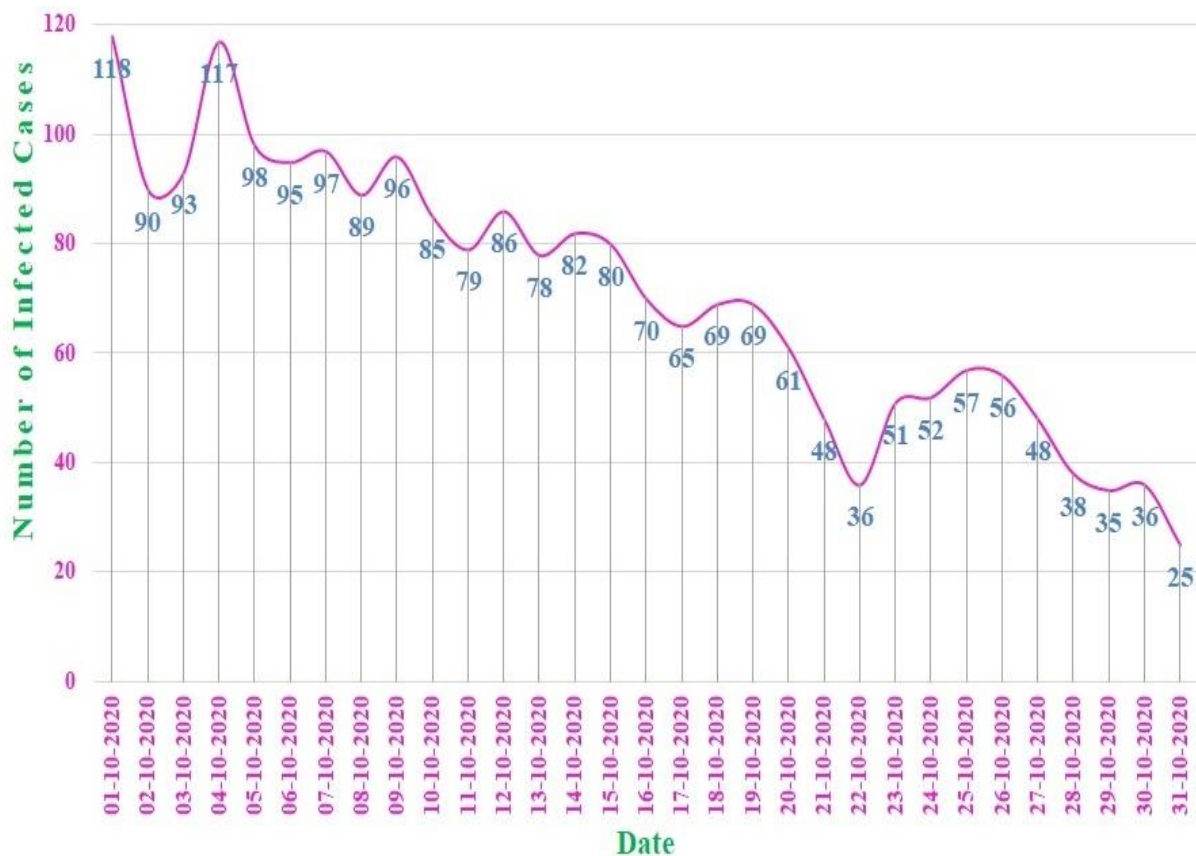


Fig. 1(b): Daily COVID-19 infected cases in Kanniyakumari District

In the case of Tenkasi District, Fig. 1(c) shows that the highest number of COVID-19 infected cases registered on 1st October 2020 decreased to 4 cases on 30th October and then increased to 10 cases at the end of the month. Additionally, in 14 days, only single-digit numbers of infected cases were registered. Upward and downward trends were noted in this district.

Rajarathinam *et.al*., reported that Auto-Regressive Integrated Moving Average Model ARIMA (0,1,1) exhibited Upward, downward followed by step declining trend in number of infected cases in Tenkasi district during the period from 1st August,2020 to 31st December 2020. Present result also in agreement with this result.

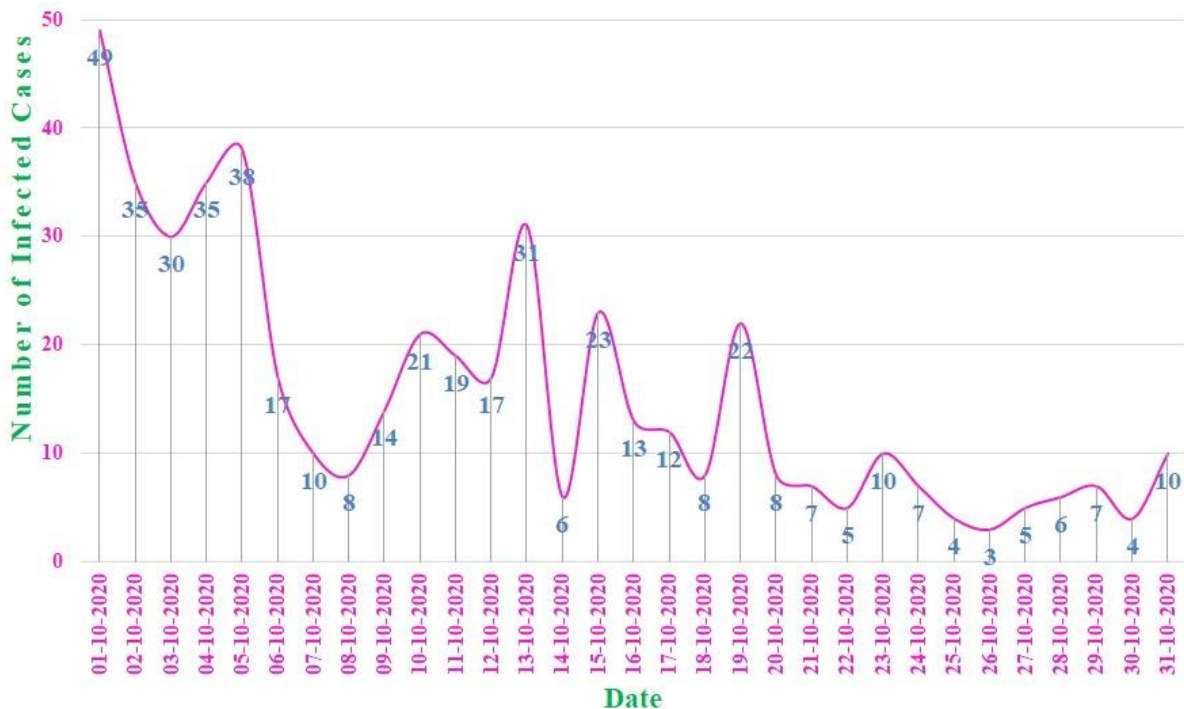


Fig. 1(c): Daily COVID-19 infected cases in Tenkasi District

Fig. 1(d) of Thoothukudi District shows that the highest number of 77 cases registered on 9th October decreased to 42 cases at the end of the month. Very peculiar upward and downward trends have been observed in this district.

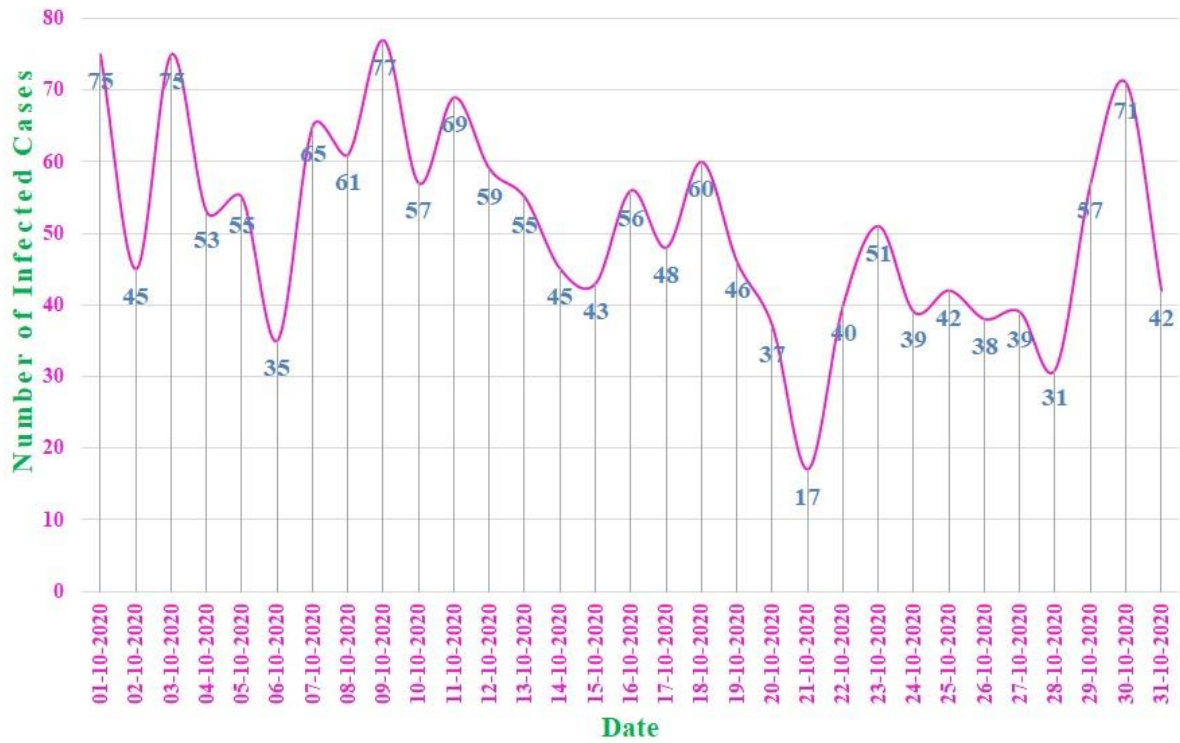


Fig. 1(d): Daily COVID-19 infected cases in Thoothukudi District

In the case of Tirunelveli District, Fig. 1(e) depicts a stepwise declining trend, with the highest number of 86 infected cases on 1st October directly declining to 15 at the end of the month.

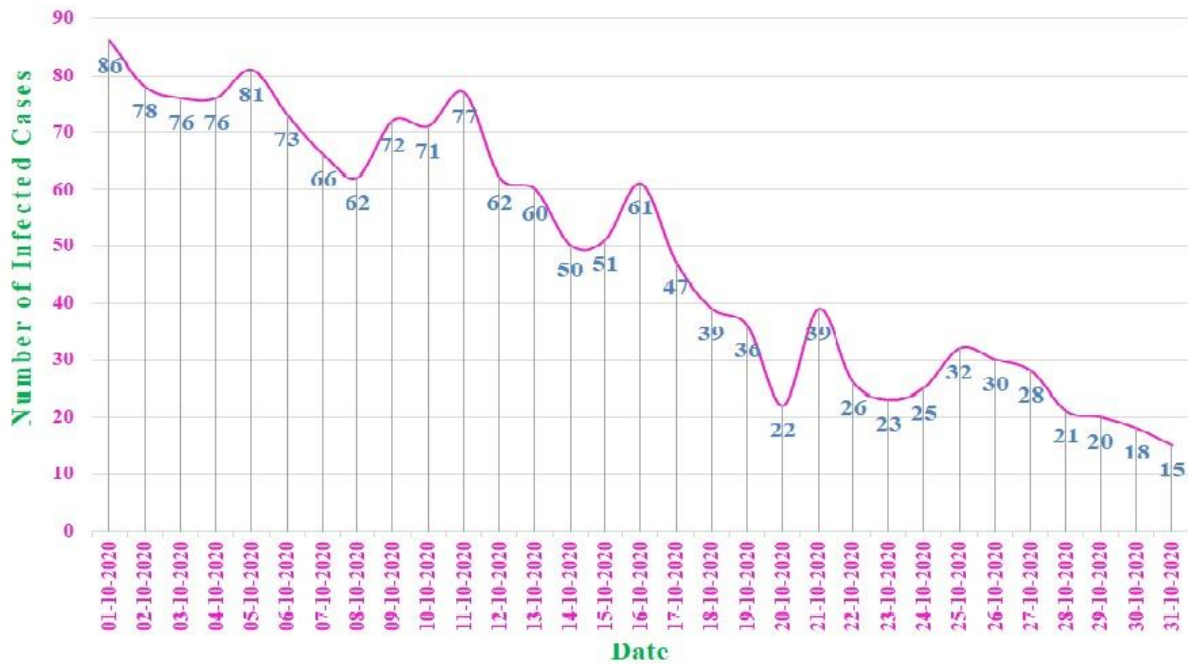


Fig. 1(e): Daily COVID-19 infected cases in Tirunelveli District

The pattern of COVID-19-infected cases of Kanniyakumari and Tirunelveli followed a similar declining trend, whereas the same upward and downward trend was noted in the cases of Tenkasi and Thoothukudi.

Fig.2 depicts the dates-wise comparison of trends exhibited in the four districts, in which the trend cure of Tenkasi District is at a lower level, followed by those of Thoothukudi, Tirunelveli and Kanniyakumari. From this trend pattern, it is noticeably clear that the highest number of infected cases is registered in Kanniyakumari District and the least number of registered cases in Tenkasi District.

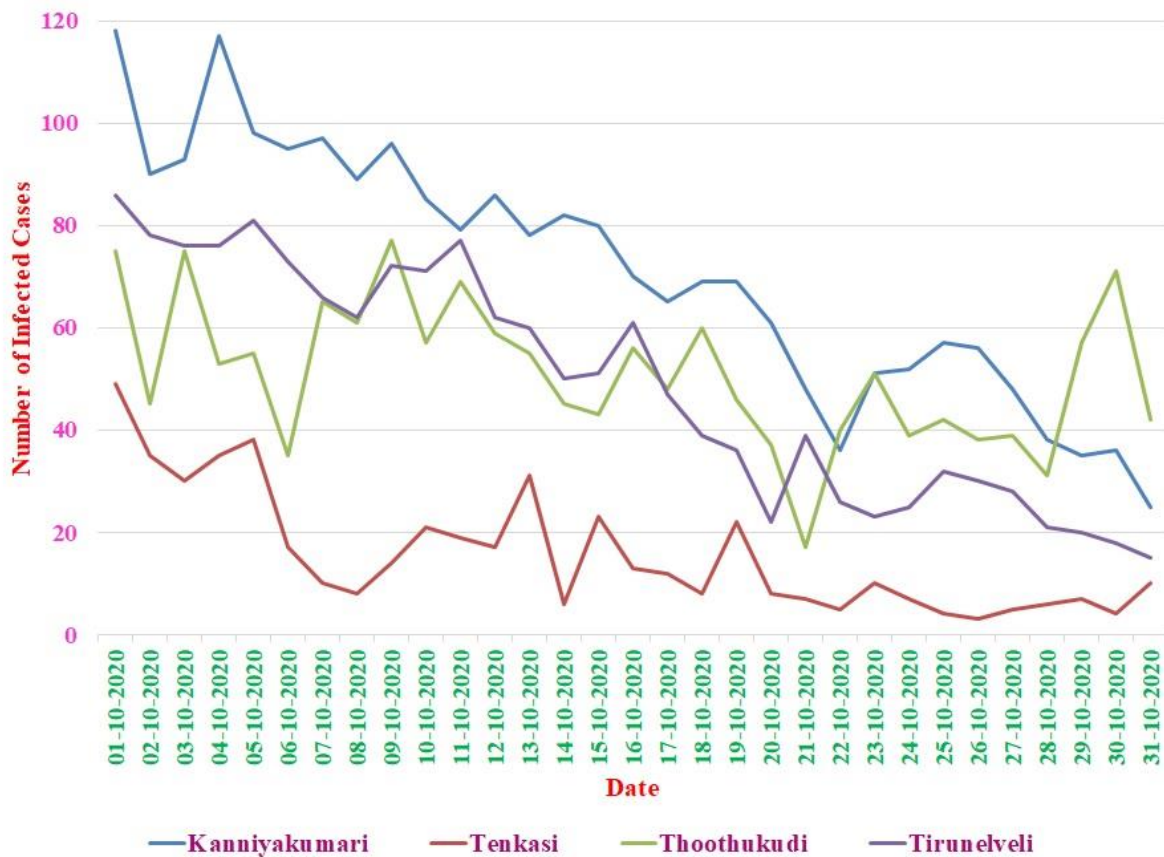


Fig. 2. Comparisons of COVID-19 infection trends in all four districts

The results presented in Table 2 reveal that the ANOVA F-test value 44,29316 and Welch F-test value (64.31809) with a p-value=0.0000 indicate that both tests are highly significant at the 1% level of significance, indicating that the number of COVID-19-infected cases is different from one district to another district (cross-section).

Table 2: Analysis of variance test for equality of means for NCASE.

Method	df	Value	Probability
--------	----	-------	-------------

ANOVA F-test	(3, 120)	44.29316	0.0000
Welch F-test*	(3, 64.3377)	64.31809	0.0000

Based on Rajarathinam et.al¹ results the above ANOVA test results are ascertained because, Kanniyakumari district follow Holt's linear exponential smoothing trend whereas in cases of Tenkasi, Thoothukudi and Tirunelveli exhibited trends are ARIM(0,1,1) , ARIMA(2,0,0), ARIMA(2,2,2), respectively. These results reveal that the pattern of COVID-19 infection is differ from one district to another.

Takele⁷, also employed ARIMA modeling approach for projecting coronavirus (COVID-19) prevalence patterns in East Africa Countries, mainly Ethiopia, Djibouti, Sudan and Somalia. ARIMA (1,2,1), ARIMA (3,1,1) with drift, ARIMA (0,2,2) and ARIMA (2,2,1) models were chosen as the optimal models for predicting coronavirus cases of Ethiopia, Djibouti, Somalia, and Sudan respectively. By means of those models, a forecast of four month a heads future scenario COVID-19 prevalence (July until October 30, 2020) has made. This study models are at par with the study models of Rajarathinam et.al¹.

Unit Root Test: Before estimating the panel data regression model, it is necessary to determine the stationarity of the variable under study. The unit root test result presented in Table 3 reveals that since the Levin, Lin & Chu t statistics value -3.76884 is significant at the 1% level of significance since the p-value is 0.0001 and hence the study variable, NCASE, is stationary at the level, the variable is I(0).

Table 3: Characteristics of the unit root test.

Method	Statistic	Prob.**
Levin, Lin & Chu t*	-3.76884	0.0001

** Probabilities are computed assuming asymptotic normality.

Takele⁷ employed Augmented Dickey-Fuller (ADF) test were applied to check the stationary of the daily time series data of Nobel Coronavirus (COVID-19) of Ethiopia, Sudan, Djibouti, and Somalia from 13 March 2020 to June 30, 2020. In our study the data type is time series of panel type and hence Levin, Lin & Chu t – test is employed.

Constant Coefficient Model(Panel OLS).The CCM es method is employed considering the number of new cases (NCASE) as the dependent variable and X, time, as the independent variables; the results are presented in Table 4. The result reveals that the intercepts and slopes are highly significant at the 1% level of significance. The slope is negative, which indicates that the number of COVID-19-infected cases decreased by 1.68% in October 2020.The model is highly significant at the 1% level of significance with an incredibly low R² value of 29%. Additionally, the estimated Durbin-Watson value of 0.242273 is quite low, which suggests the presence of autocorrelation in the data.

Table 4: Characteristics of the fitted panel least-squares method

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	73.48710	4.257903	17.25899	0.0000
X	-1.675101	0.232288	-7.211321	0.0000
Root MSE	22.94833	R-squared		0.298863
Mean dependent var.	46.68548	Adjusted R-squared		0.293116
S.D. dependent var.	27.51743	S.E. of regression		23.13566
Akaike info criterion	9.136625	Sum squared residuals		65301.58
Schwarz criterion	9.182114	Log likelihood		-564.4708
Hannan-Quinn criter.	9.155104	F-statistic		52.00315
Durbin-Watson stat.	0.242273	Prob. (F-statistic)		0.000000

The estimated model assumes that the slope coefficients of time variables X are all identical in all four states. Therefore, despite its simplicity, the CCM may distort the true relationship between the dependent variable—the number of cases (NCASE)—and time, the independent variable X, across the four districts.

Raymundo *et al.*⁵, reported that the MGWR model was the most suited model in comparison of OLS (CCM) model in the study of spatial analysis of COVID-19 incidence and the sociodemographic context in Brazil.

Fixed-Effect OR Least-Square Dummy Variable Regression Model: The results presented in Table 5 reveal that the fixed effect model explains 82% of the variation in the dependent variable. The model is highly significant at the 1% level of significance. The

dummy variables were also highly significant at the 1% level of significance. The root mean square error is 11.4868 with the S.E. of regression is 11.7256.

$$\text{NCASE} = 97.7370 - 1.6751 * X - 55.3225 * (D2) - 19.8709 * (D3) - 21.8064 * (D4)$$

Based on the statistical significance at the 1% level of significance of the estimated coefficients and the substantial increase in the R² value to 82% (significant at the 1% level of significance), one can conclude that the fixed effects model or the LSDSV regression model performs better than the panel least-squares regression model (CCM).

Table 5: Characteristics of the fixed effects or LSDSV regression model

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	97.7371	2.8254	34.5913	0.0000
C(2)	-1.6751	0.1177	-14.2285	0.0000
C(3)	-55.322	2.9783	-18.5751	0.0000
C(4)	-19.8709	2.9783	-6.67187	0.0000
C(5)	-21.8064	2.9783	-7.32173	0.0000
Root MSE	11.4868	R-squared		0.8243
Mean dependent var.	46.6854	Adjusted R-squared		0.8184
S.D. dependent var.	27.5174	S.E. of regression		11.7256
Akaike info criterion	7.80092	Sum squared resid.		16361.43
Schwarz criterion	7.91464	Log likelihood		-478.6572
Hannan-Quinn criter.	7.84711	F-statistic		139.6006
Durbin-Watson stat.	0.96695	Prob (F-statistic)		0.0000

The cross-sectional fixed effects (as deviations from the common intercept) in the context of the fixed effect model are calculated and presented in Table 6. In Kanniyakumari district, 24.2500 is positive and high in comparison to that in the other three districts. This may be due to extremely high infection rates. In the case of Tenkasi District, it is -31.07258, which is extremely exceptionally low. This is because in this district, an incredibly low rate of infections is noted. In Thoothukudi District, the effect of 4.379032 is exceptionally low in comparison to that of Kanniyakumar1 District. In case of Tirunelveli district, it is of 2.443548. This fixed effect value is exceptionally low in comparison to that of Kanniyakumari and Thoothukudi districts.

Table6 : Cross-Section Fixed Effects Values

Sr.No.	CROSSID	Effect
1	1	24.25000
2	2	-31.07258
3	3	4.379032
4	4	2.443548

The diagrammatic representation of fixed effects in four different districts is depicted in Fig.

3. Based on this result, it is concluded that the fixed effect model is better than CCM.



Fig.3. Fixed effect in different districts

To confirm the presence of the fixed effect, the redundant fixed effect test was carried out, and the results are presented in Table 7. The test results reveal that the Cross-section F and Chi-square statistics values are significant at the 1% level of significance, indicating that the presence of fixed effects is different from one district to another.

Table 7 : Results of Redundant Fixed Effects Test

Effects Test	Statistic	d.f.	Prob.
Cross-section F	118.650551	(3,119)	0.0000
Cross-section Chi-square	171.627101	3	0.0000

Wald Test: To compare the fixed effect model with CCM, the Wald test was carried out. The null hypothesis of the Wald test is $H_0=C(3)=C(4)=C(5)=0$, i.e., All three dummy variable values are zero (there is no fixed effect). The result presented in Table 8 reveals that since the F and Chi-square statistics values are significant at the 1% level of significance, the null hypothesis $H_0=C(3)=C(4)=C(5)=0$ is rejected, which indicates that the values of the dummy variables are not equal to zero, which confirms fixed effects, or the LSDV regression model is an appropriate model in comparison to CCM.

Table 8: Characteristics of the Wald test

Test Statistic	Value	df	Probability
F-statistic	118.6506	(3, 119)	0.0000
Chi-square	355.9517	3	0.0000

Oehmke, *et al.*⁸, also used Wald test in dynamic panel data modeling to study the surveillance of COVID-19 in different metropolitan area in the United States. The test result reveal that the lack of persistence effect is most likely to occur when the number of new cases per day is at constant speed.

Random-Effect Model: Finally, the random-effect model is estimated, and the results are presented in Table 9. The results reveal that the model is highly significant at the 1% level of significance with an extremely high R^2 value of 62% with an S.E. of regression 1.7257., Root MSE, 11.6307. As in the case of the fixed effect model, the random-effect model coefficients, intercept, and slope are highly significant at the 1% level of significance.

The value of slope is -1.675101, which is highly significant, indicating that the COVID-19 infection new cases are decreasing at the rate of 1.68%.

Baskar *et al.*⁴, reported that the progression of the COVID-19 infection in Tamil Nadu, India is showing decay in the number of infections of the disease which highlights the effectiveness of controlling measure.

The rho value is 0.7915, which indicates that the individual effects of cross-sections are 0.8%.

Table 9: Characteristics of the fitted random effects model.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	73.48710	11.62358	6.322241	0.0000
X	-1.675101	0.117728	-14.22852	0.0000
Effect specification				
			S.D.	Rho
Cross-section random			22.84301	0.7915
Idiosyncratic random			11.72566	0.2085
Weighted statistics				
Root MSE	11.63071	R-squared		0.623980
Mean dependent var.	4.285949	Adjusted R-squared		0.620898
S.D. dependent var.	19.04403	S.E. of regression		11.72566
Sum squared resid.	16773.90	F-statistic		202.4507
Durbin-Watson stat.	0.943181	Prob (F-statistic)		0.000000

The cross-sectional random effects in the context of the random effect model are calculated and presented in Table 10. The results reveal that in Kanniyakumari district, 24.04562 is positive and high in comparison to that in the other three districts. This may be due to extremely high infection rates. In the case of Tenkasi District, it is -30.81070, which is extremely exceptionally low. This is because of the incredibly low rate of infections in this district. In Thoothukudi District, the effect is 4.342125, which is exceptionally low in comparison to that of Kanniyakumar1 District. In case of Tirunelveli district, it is of 2.422954. This random effect value is exceptionally low in comparison to that of Kanniyakumari and Thoothukudi districts.

Table 10: Cross-Section Random Effects Values

Sr.No.	CROSSID	Effect
1	1	24.04562
2	2	-30.81070
3	3	4.342125
4	4	2.422954

The diagrammatic representation of random effects in four different districts is depicted in Fig. 4. Based on this result, the presence of random effects in all four different districts is confirmed.

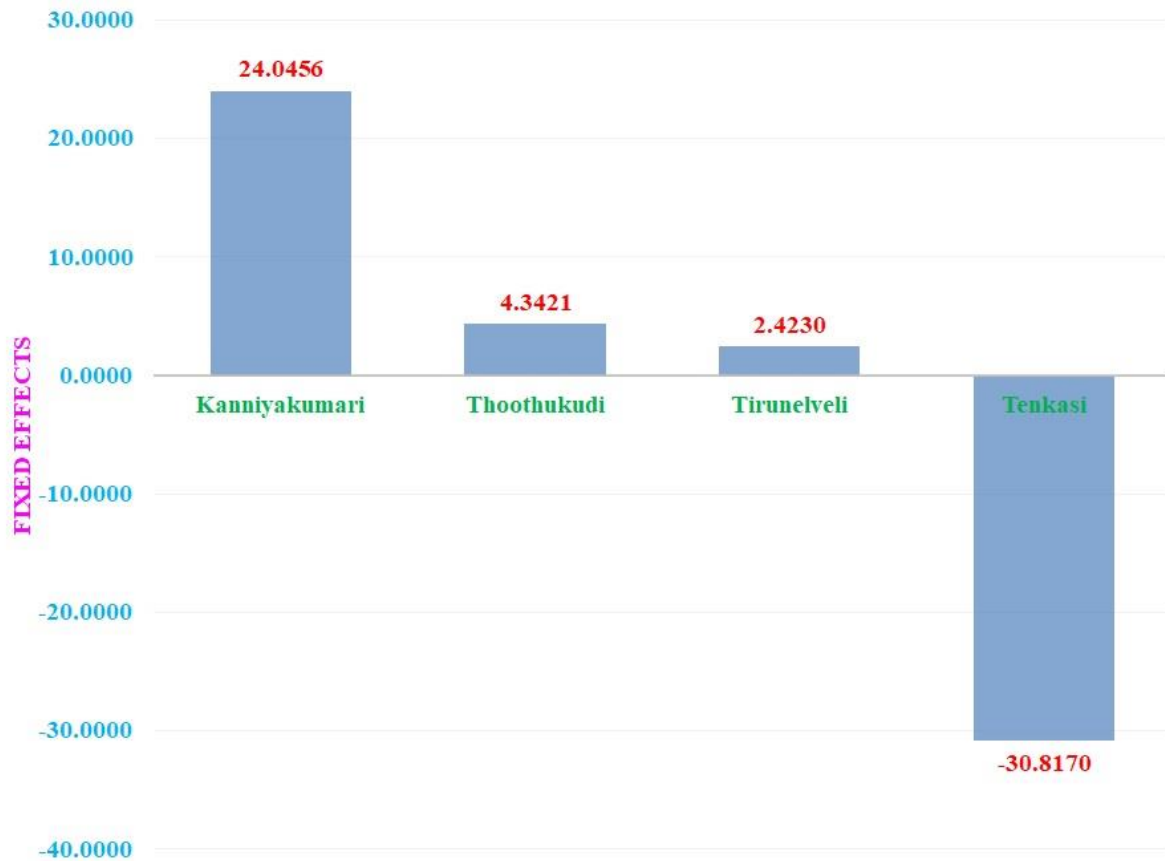


Fig.4. Fixed effect in different districts

Breusch-Pagan Lagrange-Multiplier Test (Heteroskedasticity Test):

The result presented in Table 11 indicates that the Breusch-Pagan LM test statistic value of 30.35083 and Pesaran scaled LM test statistic value of 7.029479 are highly significant at the 1% level of significance since both statistical p-values are equal to 0.0000, indicating that the null hypothesis of the test, “H₀: There is constant variance among residuals”, is rejected. Hence, the above random effect model has the problem of heteroscedasticity.

Table 11: Characteristics of the residual cross-section dependence test

Test	Statistic	d.f.	Prob.
Breusch-Pagan LM	30.35083	6	0.0000
Pesaran scaled LM	7.029479		0.0000

Hausman Test: The Hausman test result presented in Table 12 reveals that the Chi-Sq. A statistical value of 0.0000 with 1 degree of freedom is nonsignificant at a 5% significance level, and the null hypothesis “ H_0 : Random Effect Model” is accepted. So, among the three models viz. CCM, fixed effect and random effect models have emerged as appropriate models.

Table 12: Characteristics of the Hausman Test.

Test Summary	Chi-Sq. Statistic	Chi-Sq.D.F.	Prob.
Cross-Section Random	0.0000	1	0.9568

Fig. 5 depicts and confirms that the coefficients of intercept and slope lie in the 99% confidence interval (CI)

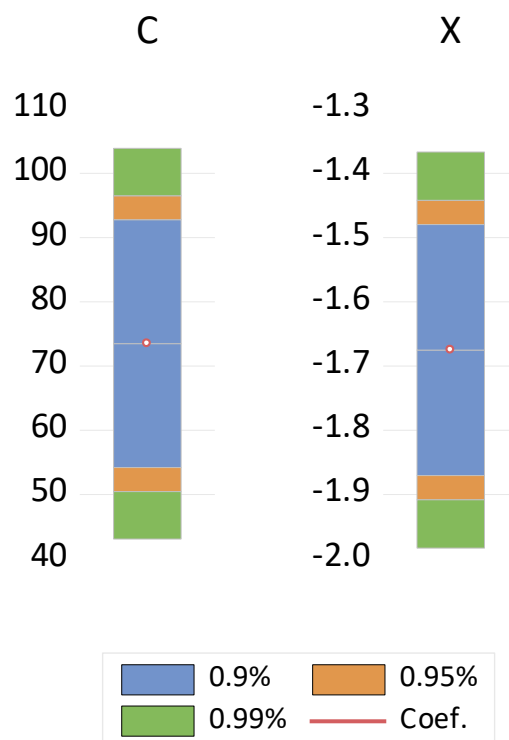


Fig.5. Confidence interval

Additionally, Fig. 6 depicts that the residuals are normally distributed because the Jarque-Bera test value is 3.1389, which is found to be nonsignificant at the 5% level of significance.

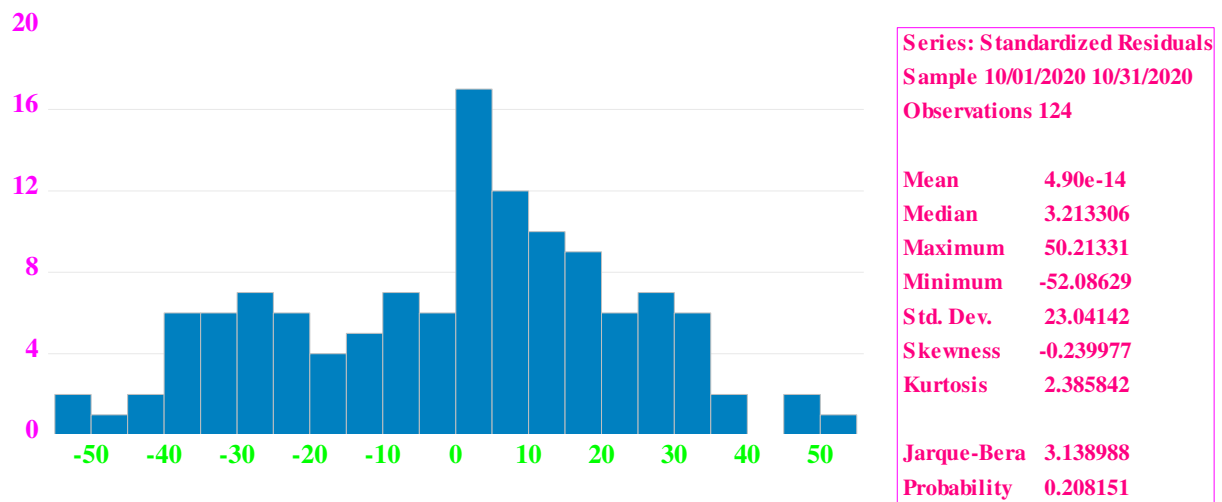


Fig. 6. Characteristics of residuals based on the random effects model.

This model in future can be extended by taking several other parameters related to age, and inclusion of population with the case of respiratory ailments, etc.

CONCLUSION

In this study, the panel data regression model was found to be suitable for assessing the trends in COVID-19-infected cases. The random effects model was found to be suitable to study the trend. The study results reveal that the highest number of new cases was registered in Kanniyakumari, followed by Thoothukudi and Tirunelveli and Tenkasi districts. The lowest number of cases was observed in Tenkasi District. This difference may be due to the precautionary measures taken by the district administration. In general, the trends in the number of COVID-19 infected cases were found to decrease in all four districts by 1.68% in October 2020.

Significance statement: This study revealed that COVID-19-infected cases showed a decreasing trend. The study would be incredibly useful to administrators and decision-makers to take precautionary measures to stop COVID-19 infections. By knowing the COVID-19 infection trends, the academic personal could take appropriate decisions.

REFERENCES

1. Rajarathinam,A., and P.Tamilselvan, and M.Ramji, 2021.Estimating hidden models and trends in COVID-19 infected cases, *Strad Research*, 8(5), 128-133, <http://doi.org/10.37896/sr8.5/012>.
2. Yongmei,D., L.Gao,2020.An evaluation of COVID-19 in Italy: A data-driven modeling analysis. *Infectious Disease Modelling*, 5, 495 - 501, <https://doi.org/10.1016/j.idm.2020.06.007>.
3. Roy,S., and Bhattacharya,K.R., 2020,Spread of COVID-19 in India: A Mathematical Model, *Journal of Science and Technology*, 5(3), 41-47, <https://doi.org/10.46243/jst.2020.v5.i3>.
4. Bhaskar, A., C.Ponnuraja, R.Srinivasan,S.Padmanaban, 2020. Distribution and growth rate of COVID-19 outbreak in Tamil Nadu: A log-linear regression approach. *Indian J Public Health*, 2020,64:S188-S191, https://doi.org/10.4103/ijph.IJPH_502_20.
5. Raymundo,C.E.,M.C.Oliveira, TdA. Eleuterio, S.R.Andre´, M.G. da Silva,ERdS,Queiroz, 2021.Spatial analysis of COVID-19 incidence and the sociodemographic context in Brazil. *PLoS ONE* 16(3): e0247794. <https://doi.org/10.1371/journal>.
6. Bandekar, S.R., M.Ghosh, 2021. Mathematical modeling of COVID-19 in India and its states with optimal control. *Model Earth Syst Environ*, <https://doi.org/10.1007/s40808-021-01202-8>.
7. Takele, R.,2020. Stochastic modelling for predicting COVID-19 prevalence in East Africa Countries. *Infectious Disease Modelling*, 5, 598-607,<https://doi.org/10.1016/j.idm.2020.08.005>.
8. Oehmke, T. B., L.A.Post, C.B.Moss,T.Z.Issa, M.J.Boctor,S.B.Welch, and J.F. Oehmke,2021. Dynamic Panel Data Modeling and Surveillance of COVID-19 in Metropolitan Areas in the United States: Longitudinal Trend Analysis, *Journal of medical Internet research*, 23(2), <https://doi.org/10.2196/26081>.
9. Baltagi, B.H., 2021. *Econometric Analysis of Panel Data*.6th Edition, Wiley, New York, NY, <https://link.springer.com/book/10.1007%2F978-3-030-53953-5>.
10. Gujarati, D.N., D. C. Porter, G.Sangeetha,2017.*Basic Econometrics*, 5th edition, McGraw Hill Education, New York, NY.<https://www.amazon.com/Basic-Econometrics-Damodar-N-Gujarati/dp/0073375772>

11. Hsiao, C., 2010. Analysis of Panel Data. 2nd Edition, Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9780511754203>.
12. Levin, A., C.F. Lin, and C.S.J.Chu, 2002. Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties. Journal of Econometrics, 108,1-24, [https://doi.org/10.1016/S0304-4076\(01\)00098-7](https://doi.org/10.1016/S0304-4076(01)00098-7).
13. Hadri, K., 2000. Testing for stationarity in heterogeneous panel data. Econometrics Journal, 3(2),148-161, <https://doi.org/10.1111/1368-423X.00043>.
14. Patrick, R.H., 2020. Handbook of Financial Econometrics, Mathematics, Statistics, and Machine Learning, World Scientific Publishing Co. Pte. Ltd., 1075-1108, https://doi.org/10.1142/9789811202391_0028.