

Panel Data Modelling of COVID-19 Infected Cases

RAJARATHINAM ARUNACHALAM (✉ arrathinam@yahoo.com)

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

TAMILSELVAN PAKKIRISAMY

Manonmaniam Sundaranar University

Research Article

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

Posted Date: June 18th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-635014/v1>

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

[Read Full License](#)

Panel Data Modelling of COVID-19 Infected Cases

Rajarathinam, A., and Tamilselvan, P

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

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

Running title: Modelling of 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 analyses 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 regarding the prediction of future coronavirus cases.

The coronavirus epidemic caused an emergency in South Korea. The first infected case came to light on 20 January 2020, followed by 9583 more cases that were reported by 29 March 2020. This indicates that the number of confirmed cases is increasing rapidly, which can cause a nationwide crisis for the country. The aim of this study is to fill a gap between previous studies and the current rate of spreading of COVID-19 by extracting a relationship between independent variables and the dependent variables. This study statistically analyzed the effect of factors such as sex, region, infection reasons, birth year, and released or diseased date on the reported number of recovered and deceased cases. The results found that sex, region, and reasons for infection affected both recovered and deceased cases, while birth year affected only the deceased cases. In addition, no deceased cases were reported for released cases, while 11.3% of deceased cases were confirmed to be positive after their death. An unknown reason for infection is the main variable that is detected in South Korea, with more than 33% of total infected cases¹.

The novel coronavirus (COVID-19), which has been spreading worldwide since December 2019, has sickened millions of people, locked down major cities and some countries, prompting unprecedented global travel restrictions. Real data-driven modeling is an effort to help evaluate and curb the spread of the novel virus. Lockdowns and the effectiveness of reduction in the contacts in Italy have been measured via our modified model, with the addition of auxiliary and state variables that represent, contacts with infected, conversion rate and latent propagation. The results show a decrease in infected people due to stay-at-home orders and tracing quarantine interventions. The effect of quarantine and centralized medical treatment was also measured through numerical modeling analysis².

It has been reported that the novel coronavirus (COVID-19), which was first known at the end of 2019, has impacted almost every aspect of life. This paper focuses on the incidence of the disease in Italy and Spain—two of the first and most affected European countries. Using two

simple mathematical epidemiological models—the susceptible-infectious-recovered model and the log-linear regression model—we modeled the daily and cumulative incidence of COVID-19 in the two countries during the early stage of the outbreak and computed estimates for basic measures of the infectiousness of the disease, including the basic reproduction number, growth rate, and doubling time. Estimates of the basic reproduction number were found to be larger than 1 in both countries, with values between 2 and 3 for Italy and 2.5 and 4 for Spain. Estimates were also computed for the more dynamic effective reproduction number, which showed that since the first cases were confirmed in the respective countries, the severity has generally been decreasing. The predictive ability of the log-linear regression model was found to give a better fit, and simple estimates of the daily incidence for both countries were computed³.

The novel coronavirus COVID-19 was originally identified in December 2019. Based on the data issued by March 30, 2020 daily report, the epidemic of SARS-CoV-2 so far has caused 693224 cases and resulted in 33106 deaths in more than 200 countries. Referring to the data reported, the World Health Organization declared the outbreak a pandemic. We considered the chain-binomial type of the model, which involves short stages of high infectivity and approximately constant incubation periods. This research paper aims to study and analyze the COVID-19 virus spreading statistics on examples of cases from different countries. High correlation coefficients (91.64%) and determinants (83.98%) between the total volumes of virus spread and recovery are high and indicate the correctness of the Bailey model. Thus, as of March 30, with the results of statistical and mathematical data processing, it is difficult to predict the future spread-reduction variables of the pandemic⁴.

A simple mathematical model was used to predict the characteristics of the epidemic caused by coronavirus in mainland China. Unfortunately, the number of coronavirus victims is expected to be much higher than that predicted on February 10, 2020, since 12289 new cases (not previously included in official counts) were added two days later. Further research should focus on updating the predictions with the use of up-to-date data and using more complicated mathematical models⁵.

The coronavirus outbreak has been the most notable world crisis since the Second World War. The pandemic that originated from Wuhan, China, in late 2019 affected all nations of the world and triggered a global economic crisis whose impact will be felt for years to come. This

necessitates the need to monitor and predict COVID-19 prevalence for adequate control. Linear regression models are prominent tools in predicting the impact of certain factors on the COVID-19 outbreak and taking the necessary measures to respond to this crisis. The data were extracted from the NCDC website and spanned from March 31, 2020 to May 29, 2020. In this study, we adopted the ordinary least squares estimator to measure the impact of travelling history and contacts on the spread of COVID-19 in Nigeria and made a prediction. The model was conducted before and after travel restriction was enforced by the Federal government of Nigeria. The fitted model fit well to the dataset and was free of any violation based on the diagnostic checks conducted. The results show that the government made the right decision to enforce travelling restrictions because we observed that travelling history and contacts made increased the chances of people being infected with COVID-19 by 85% and 88%, respectively. This prediction of COVID-19 shows that the government should ensure that most travelling agencies should have better precautions and preparations in place before reopening⁶.

Coronavirus disease 2019 (COVID-19) has continued to be a global threat to public health. In fact, unreserved effort is needed to monitor the prevalence of the virus. However, applying an effective prediction of the prevalence is thought to be the fundamental requirement to effectively control the spreading rate. Time series models have extensively been considered convenient methods to predict the prevalence or spreading rate of the disease. This study, therefore, aimed to apply the autoregressive integrated moving average (ARIMA) modeling approach for projecting coronavirus (COVID-19) prevalence patterns in East African countries, mainly Ethiopia, Djibouti, Sudan and Somalia⁷.

COVID-19 is rapidly spreading in South Asian countries, especially in India. India is the fourth most COVID-19-affected country at present, i.e., until July 10, 2020. With limited medical facilities and a high transmission rate, the study of COVID-19 progression and its subsequent trajectory needs to be analyzed in India. Epidemiologic mathematical models have the potential to predict the epidemic peak of COVID-19 under different scenarios. Lockdown is one of the most effective mitigation policies adopted worldwide to control the transmission rate of COVID-19 cases. In this study, we use an improvised five-compartment mathematical model, i.e., Susceptible (S)-Exposed (E)-Infected (I)-Recovered (R)-Death (D) (SEIRD), to investigate the progression of COVID-19 and predict the epidemic peak under the impact of lockdown in India. The aim of this study is to provide a more precise prediction of epidemic peaks and to evaluate the impact of lockdown on epidemic peak shifts in India. For this purpose, we examine

the most recent data (from January 30, 2020 to July 10, 2020, i.e., 160 days) to enhance the accuracy of outcomes obtained from the proposed model. The model predicts that the total number of COVID-19 active cases would be approximately 5.8×10^5 on August 15, 2020 under current circumstances. In addition, our study indicates the existence of underreported cases, i.e., 105 during the post lockdown period in India. Consequently, this study suggests that a nationwide public lockdown would lead to epidemic peak in India. It is expected that the obtained results would be beneficial for determining further COVID-19 mitigation policies not only in India but globally as well⁸.

Aims: The aim of this work was to develop suitable ARIMA models that can be used to forecast daily confirmed/dead cases of COVID-19 in Nigeria. This is subject to developing the model, checking them for suitability, carrying out eight-month forecasts, and making recommendations for the Nigerian health sector. **Study Design:** The study used daily confirmed and dead cases of COVID-19 in Nigeria. **Methodology:** This work covers time series data on the daily confirmed/dead cases of COVID-19 in Nigeria obtained from the Nigerian Centre for Disease Control (NCDC) from 21 March 2020 to 5 May 2020, covering a total of 51 data points. This work aims to develop suitable autoregressive integrated moving average (ARIMA) models that can be used to forecast total daily confirmed/dead cases of COVID-19 in Nigeria. Two adequate subset ARIMA (2, 2, 1) and AR (1) models for the confirmed/dead cases are fitted and discussed **Results:** A forecast of 239 days – from 6th May 2020 to 31 December 2020 – was conducted using the fitted models, and we observed that the COVID19 data have an upward trend and are best forecasted within a short period. **Conclusion:** Critical investigation into the rate of spread of the COVID-19 pandemic has shown that the daily confirmed cases and deaths tend to follow an upward trend. This work aimed to develop suitable ARIMA models that can be used to fit the most appropriate subsets to statistically forecast the actual number of confirmed cases and deaths from COVID-19 recorded in Nigeria for a period of 8 months⁹.

Based on the above information, the present study aimed to assess the trends in the number of cases related to COVID-19, 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, were investigated by using panel data models.

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.

⁹A simple panel data regression model is specified as

$$Y_{it} = \alpha + \beta X_{it} + v_{it}$$

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 nonstochastic 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 $t=1, 2, 3, \dots, 30$. Detailed discussions of panel data models are given in^{10,11}.

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

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

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 α_2 , α_3 , and α_4 are different intercept coefficients that indicate 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 distributed independently of the regressor, and one should include α_i in the error term. Each individual has the same slope parameters and a composite error term $\varepsilon_{it} = \alpha_i + v_{it}$.

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

Here, $\text{var}(\varepsilon_{it}) = \sigma_\alpha^2 + \sigma_v^2$, and $\text{cov}(\varepsilon_{it}, \varepsilon_{is}) = \sigma_\alpha^2$; therefore, $\rho_\varepsilon = \text{cor}(\varepsilon_{it}, \varepsilon_{is}) = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_v^2}$.

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 error⁹.

Hausman test: 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:

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

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	KANNIYAKUMARITENKASI		THOOTHUKUDITIRUNELVELI	
Mean	70.93548	15.61290	51.06452	49.12903
Median	70.00000	10.00000	51.00000	50.00000
Maximum	118.0000	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.

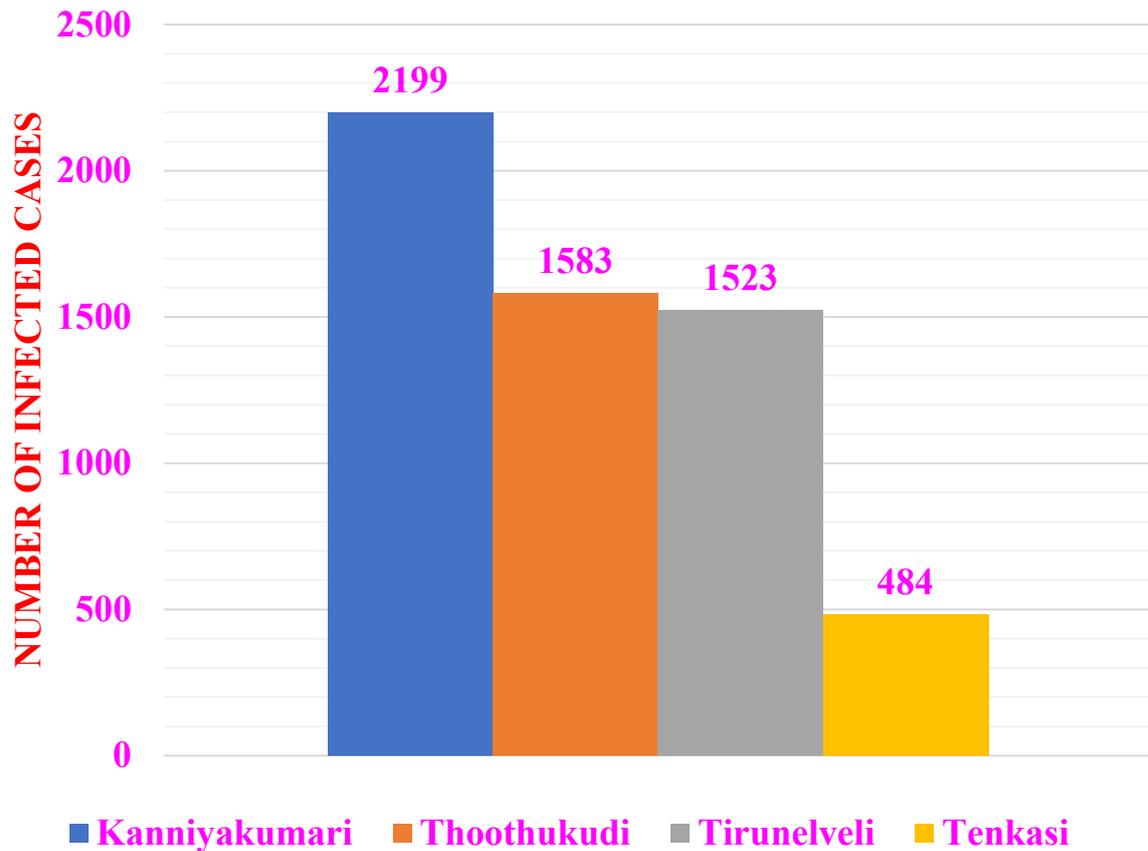


Fig.1(a). Total number of COVID-19 infected cases in all four districts in October 2020

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.

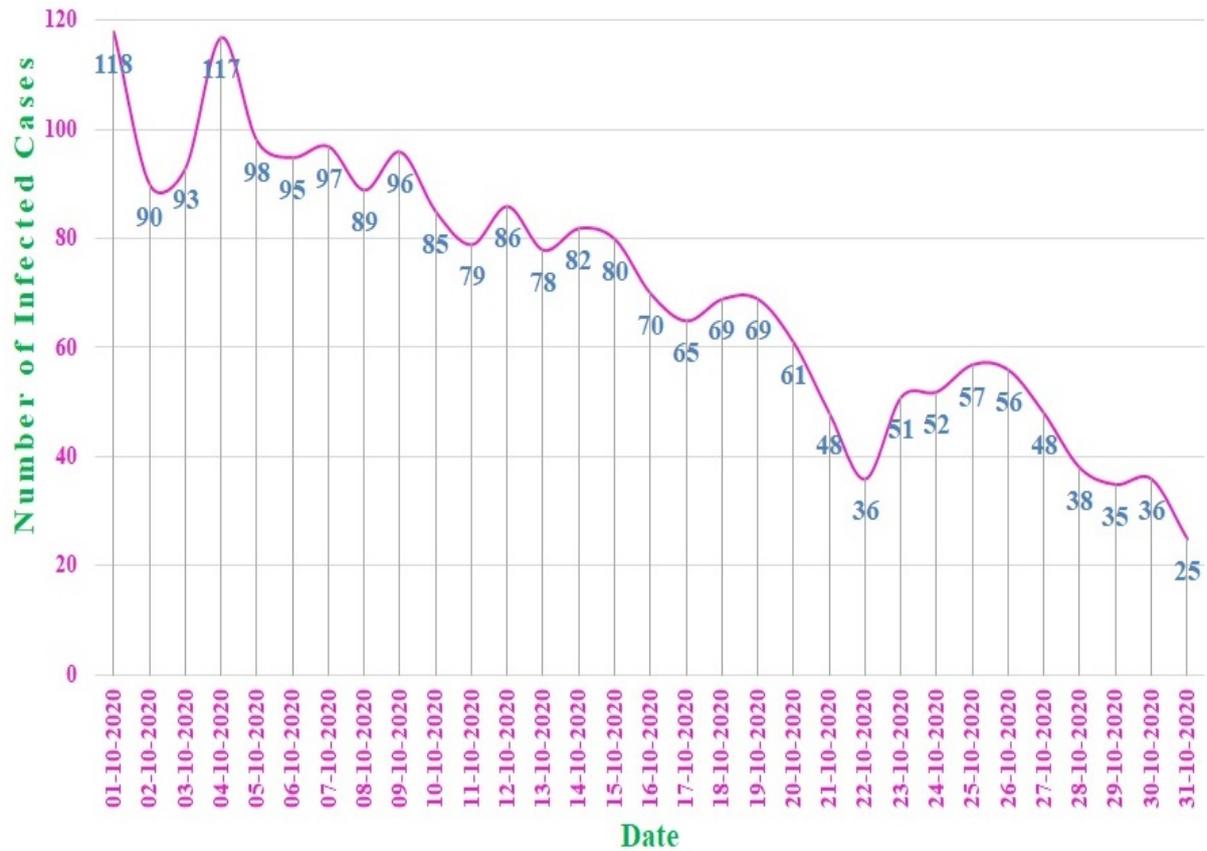


Fig. 1(b): Daily COVID-19 infected cases in Kanniyakumari District

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.

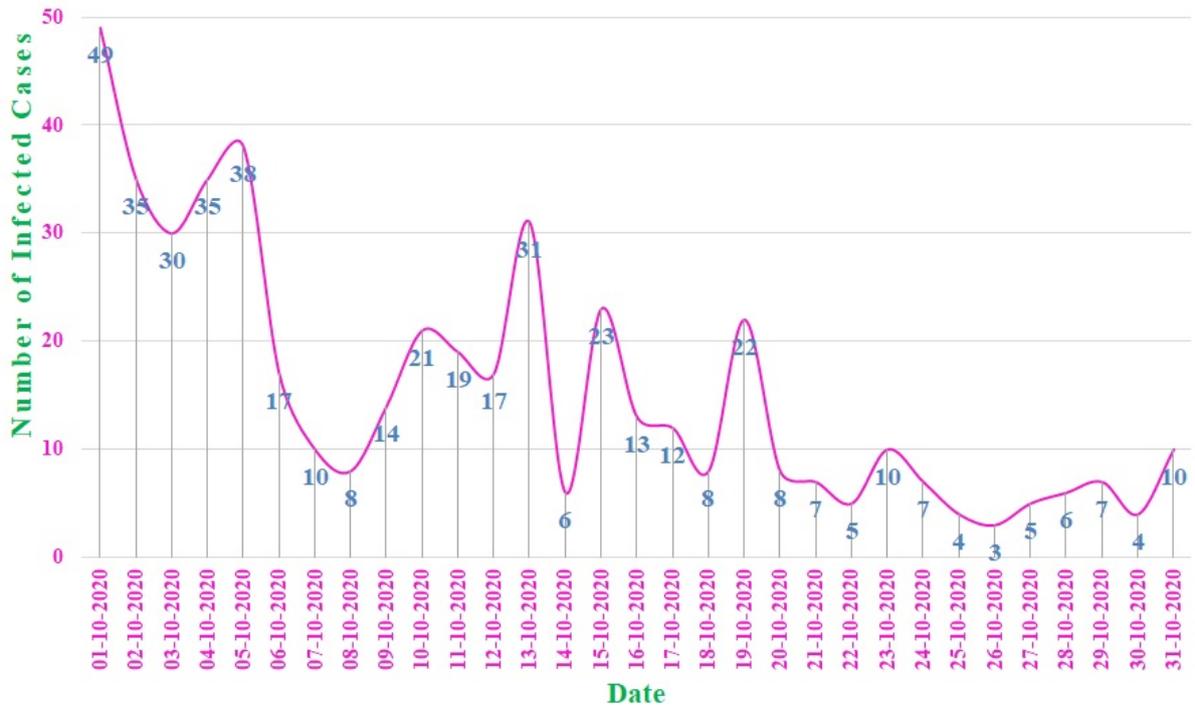


Fig. 1(c): Daily COVID-19 infected cases in Tenkasi District

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.

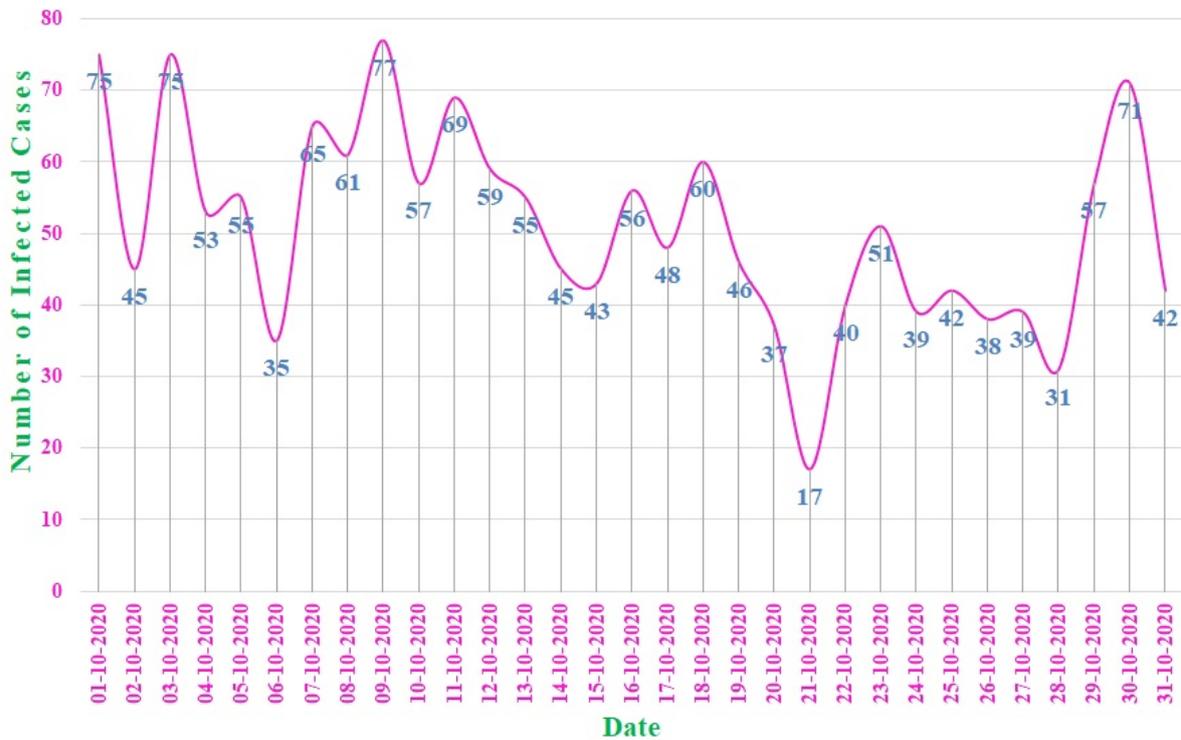


Fig. 1(d): Daily COVID-19 infected cases in Thoothukudi District

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.

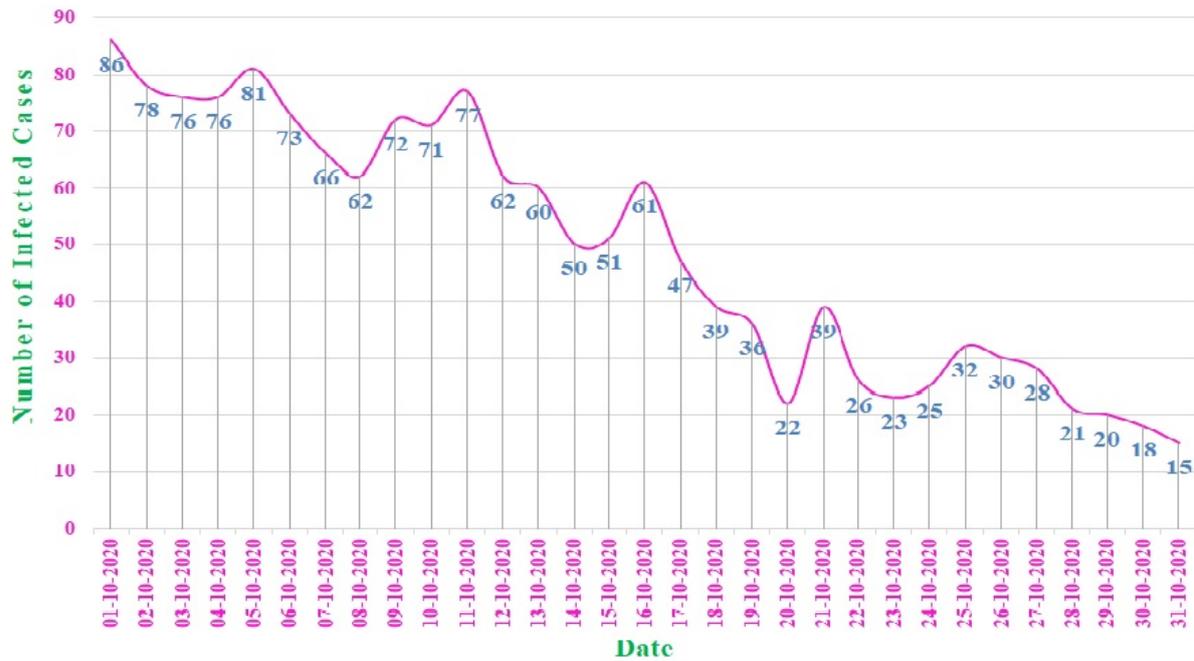


Fig. 1(e): Daily COVID-19 infected cases in Tirunelveli District

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 curve of Tenkasi District is at a lower level, followed by those of Thoothukudi, Tirunelveli and Kanniyakumari. From this trend pattern, it is very clear that the highest number of infected cases is registered in Kanniyakumari District and the least number of registered cases in Tenkasi District.

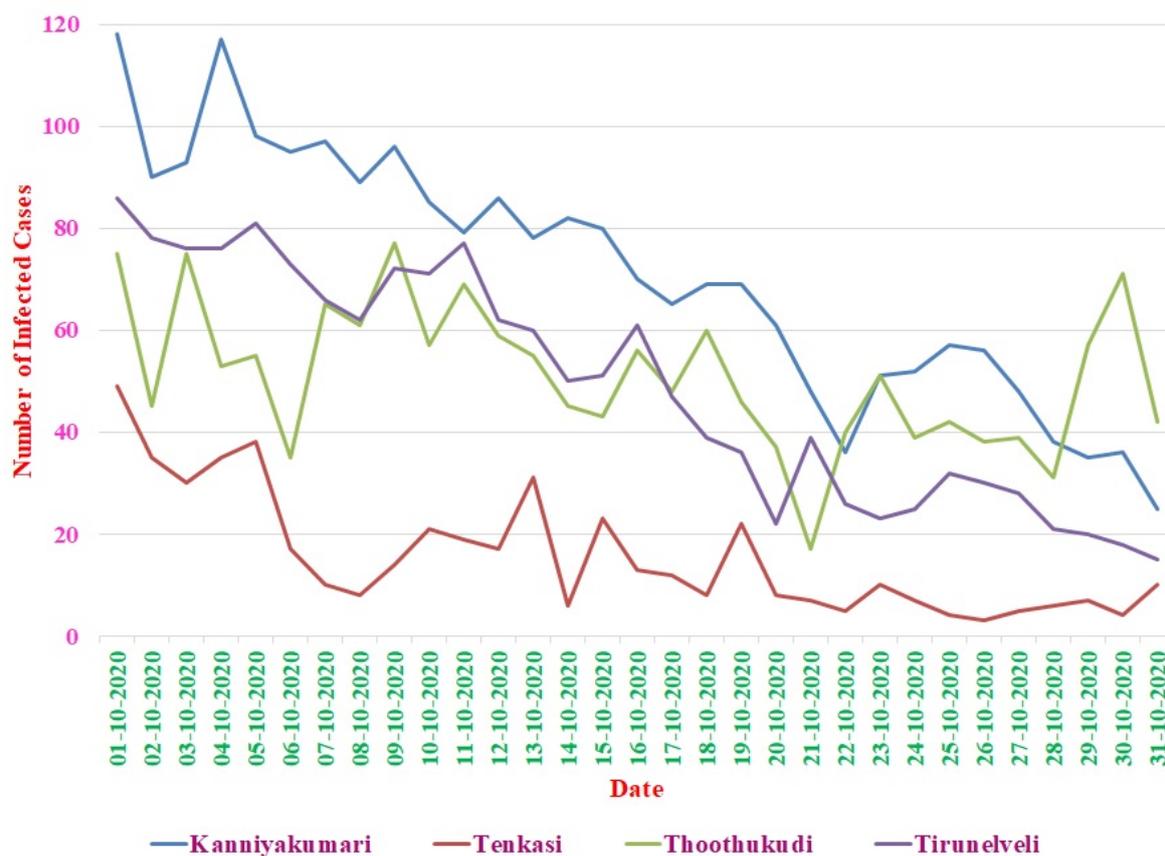


Fig. 2. Comparisons of COVID-19 infection trends in all four districts

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

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.

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.

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 very 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.

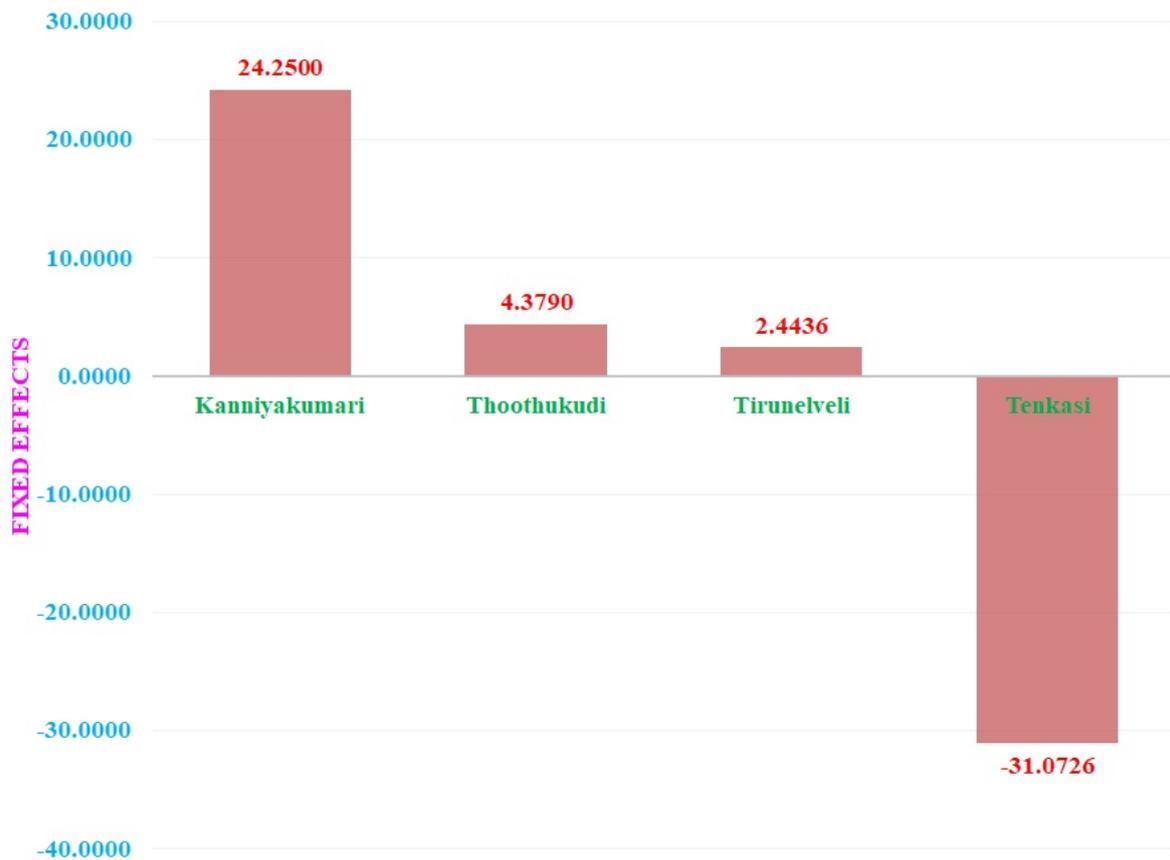


Fig.3. Fixed effect in different districts

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

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%. 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.

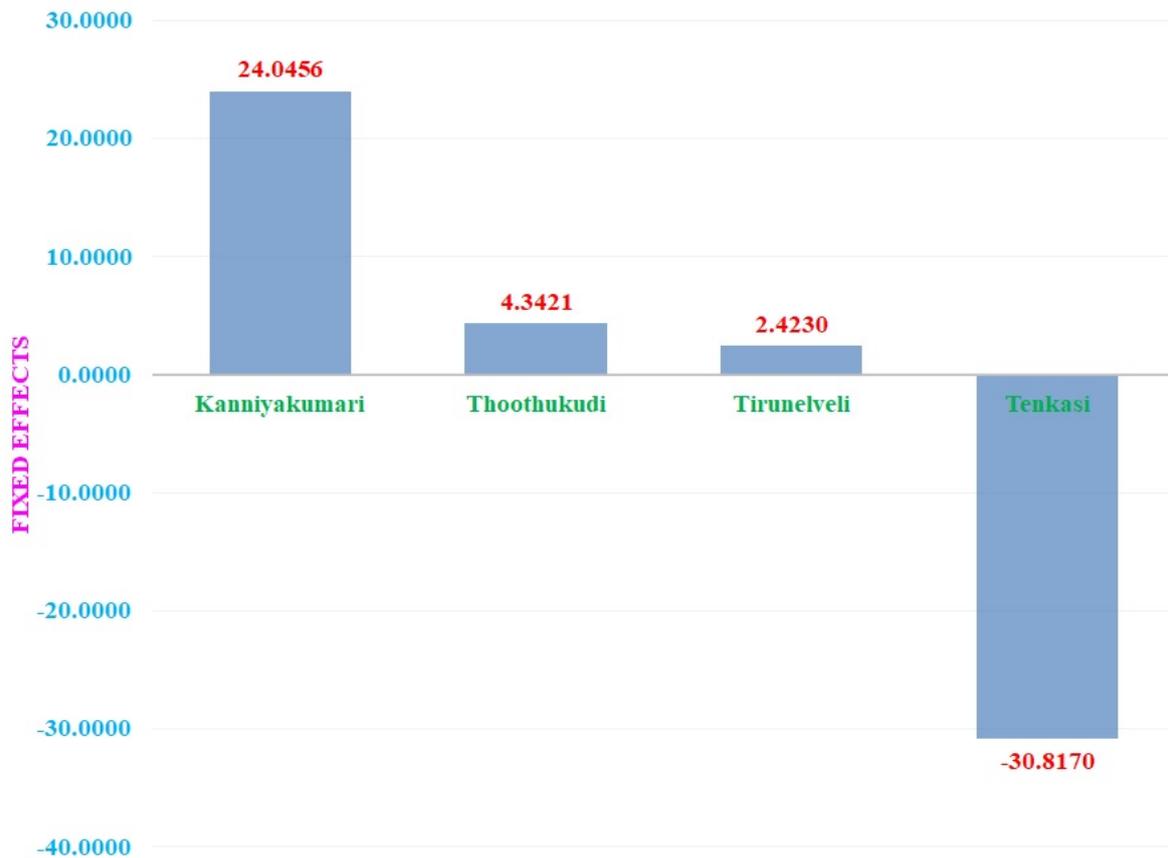


Fig.4. Fixed effect in different districts

Breusch-Pagan Lagrange-Multiplier Test (Heteroskedasticity Test): It is well known that the presence of heteroskedasticity in the disturbances of an otherwise properly specified linear model leads to consistent but inefficient parameter estimates and inconsistent covariance matrix estimates. As a result, faulty inferences will be drawn when testing statistical hypotheses in the presence of heteroskedasticity¹⁴.

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)

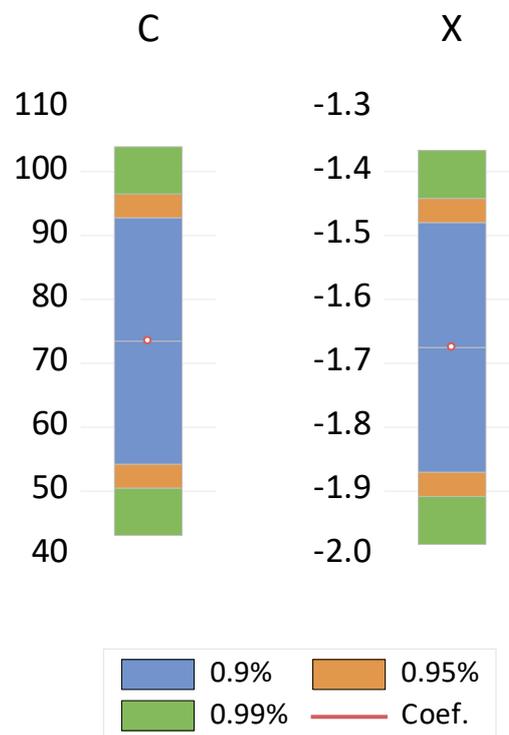


Fig.5. Confidence interval

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.

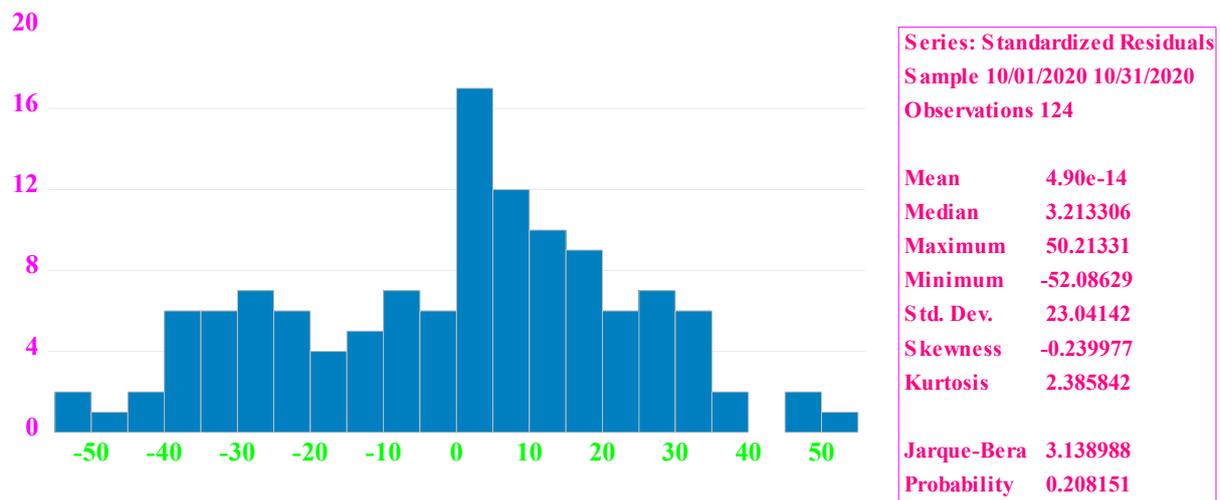


Fig. 6. Characteristics of residuals based on the random effects model.

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.

ACKNOWLEDGEMENT

Both authors expressed their sincere thanks and gratitude to the Chief Editor and Referee of this paper for providing suitable suggestions to improve the quality and presentation of the earlier version of this paper.

REFERENCES

1. Al-Rousan, N., Al-Najjar, H., 2020. Data analysis of coronavirus COVID-19 epidemic in South Korea based on recovered and death cases. *Journal of Medical Virology*, 92, 1603-1608, <https://doi.org/10.1002/jmv.25850>.
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. Chu, J., 2021. A statistical analysis of the novel coronavirus (COVID-19) in Italy and Spain. *PLoS ONE*, 16(3), <https://doi.org/10.1371>.
4. Gondauri, D., E. Mikautadze, and M. Batiashvili, 2020. Research on COVID-19 Virus Spreading Statistics based on the Examples of the Cases from Different Countries. *Electron Journal of General Medicine*, 17, 1-4, <https://doi.org/10.29333/ejgm/7869>.
5. Nesteruk, I., 2020, Statistics-Based predictions of coronavirus Epidemic Spreading in Mainland China. *Innov. Biosyst. Bioeng*, 4(1) :13-18, <https://doi.org/10.20535/ibb.2020.4.1.195074>
6. Ogundokun, R.O., A.F. Lukman, G.B.M. Kibri, J.B. Awotunde, B.B. Aladeitan, B.B., 2020, Predictive modelling of COVID-19 confirmed cases in Nigeria. *Infectious Disease Modelling*, 5:543-548, <https://doi.org/10.1016/j.idm.2020.08.003>.
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. Tiwari, V., N. Deyal, and N. S. Bisht, Mathematical Modeling Based Study and Prediction of COVID-19 Epidemic Dissemination Under the Impact of Lockdown in India. *Front. Phys.* 2020, 8:586899, <https://doi:10.3389/fphy.2020.586899>.
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.
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. Halbert, W., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48 (4), 817-838, <https://doi.org/10.2307/1912934>.