Journal of Statistical and Econometric Methods, Vol.12, No.1, 2023, 1-18
ISSN: 2241-0384 (print), 2241-0376 (online)
https://doi.org/10.47260/jsem/1211
Scientific Press International Limited

Modeling COVID-19 Infected Cases and Deaths Based on Generalized Method of Moments

Rajarathinam, A.[1] and Anju, J.B.[2]

Abstract

This paper investigates the dynamic relationships between the number of COVID-19 infected cases and deaths in all the districts of Karnataka state, India, from July 2020 to December 2021 based on the panel Generalized Method of Moments (GMM). The panel GMM model with the first difference transformation was found suitable for studying the dynamics of the number of deaths due to COVID-19 infections over time. The one-period lag (DEATHS (-1)) has a positive and significant effect on the number of deaths (DEATH). The Wald test confirms the validity of the coefficients' significance and adds explanatory power to the model. The correlation between number of fatalities at time t positively correlated with the number of deaths in the previous period. Also, the number of infected cases positively and significantly influences the number of deaths over time. Granger pairwise causality test reveals the existence of bi-directional causality relationships between the COVID-19 infected cases and deaths.

JEL classification numbers: E18, HO, I1, J64, J88.

Keywords: Arellano-Bond serial correlation test, Cross-sectional dependence test, Cointegration test, Granger causality test, Generalized Method of Moments, Kao cointegration test, Hausman test, Levin-Lin-Chu unit root test, Wald test.

[1] Manonmaniam Sundaranar University.
[2] Manonmaniam Sundaranar University.

Article Info: *Received:* October 7, 2022. *Revised:* October 28, 2022.
Published online: November 4, 2022.

1. Introduction

The usual approach today when facing heteroskedasticity of unknown form is to use the Generalized Method of Moments (GMM), introduced by Hansen and Singleton (1982). It is a method for constructing estimators analogous to the Maximum Likelihood (ML) method. GMM uses assumptions about specific moments of the random variables instead of assumptions about the entire distribution, which makes GMM more robust than ML at the cost of some efficiency. GMM models are used when we are not sure of the distribution of the dependent variable. GMM uses the orthogonality conditions to allow for efficient estimation in the presence of heteroskedasticity of unknown form. These models are used when there is a presence of endogeneity in regression models. It is a combination of ordinary least squares (OLS) and the two stages least square method. GMM models are linear regression models and allow the dependent variable depends on its past realizations, thus making the model *dynamic*.

GMM estimator makes it possible for researchers to eliminate the problems of serial correlation, heteroskedasticity, and endogeneity of some variables. This method uses the lags of dependent variables in the model to consider the dynamic effects. Dynamic relationships are modeled by inserting the lags of dependent variables that appear on the right side of the equation; OLS estimators are not consistent (Dilmaghani and Tehranchian, 2015).

The present paper is aimed to estimate the dynamic relationships between COVID-19 infected cases and the number of deaths due to COVID-19 in all 30 districts of Karnataka state, India, from July 2020 to December 2021, based on the GMM method.

2. Materials and Methods

2.1 Materials

The monthly data on COVID-19 infected cases and deaths dataset was collected from the official Karnataka state government website www.covid19.karnataka.gov.in from July 2020 to December 2021 (eighteen months of data). Various econometric tools related to panel data auto-regression modeling were employed to investigate the dynamic relationships between COVID-19 infected cases and the number of deaths due to COVID-19 in all 30 districts of Karnataka state, India. Several methodologies for panel data regression modeling are discussed in the methods section. EViews Ver. 11. the software was used for the parameter estimation and model fit.

2.2 Methods

2.2.1 Generalized Method of Moments

The GMM model is given by

$$y_{it} = \gamma y_{i,t-1} + x_{it}^{'}\beta + \alpha_i + \nu_{it}, \quad (1)$$

i=1,2, 3,…,N (individuals), t=1,2,3,…,T (time)

The above model is dynamic because the equation for time t includes an element from the previous period, the lagged response $y_{i,t-1}$. The introduction of lags becomes crucial to control the process's dynamics. In the above model, x_{it} the regressor is fixed individual effects, ν_{it} has zero mean and constant variance, and is uncorrelated across time and individuals.

As $y_{i,t-1}$ is correlated α_i because $y_{i,t-1}$ it is a function of α_i, Generalised Least Squares (GLS), and OLS estimators are biased and inconsistent. Within Group (WG) estimators, so-called "fixed effects estimators" are also biased and unreliable because in the transformed model, when using variable deviations from the mean, the independent variable will be endogenous ($\bar{y}_i$ is correlated with $\bar{\nu}_i$). An alternative transformation to remove individual effects α_i is the so-called "first-difference" transformation :

$$\Delta y_{it} = \gamma \Delta y_{i,t-1} + \Delta x_{it}^{'} \beta + \Delta \nu_{it} \tag{2}$$

Again, WG and GLS estimators are inappropriate. Moreover, the model suffers from an endogeneity problem because the dynamic structure of equation (2) $\Delta y_{i,t-1}$ is correlated $\Delta \nu_{it}$. To solve this problem, Anderson and Hsiao (1982) proposed to control endogeneity using $\Delta y_{i,t-2}$ or $\Delta y_{i,t-1}$ as instruments $\Delta y_{i,t-1}$. Lagged levels of the endogenous variable *aw*, three or more periods before, can be used as instruments (Holtz-Eakin et al., 1988), and if the panel includes three or more periods, one should have more available instruments than unknown parameters.

Arellano and Bond (1991) proposed a method that exploits all possible instruments. Using the GMM, they obtained estimators using the moment conditions generated by lagged levels of the dependent variable $\left(y_{i,t-2}, y_{i,t-3}, \ldots\right)$ with $\Delta \nu_{it}$. These estimators are called difference GMM estimators.

Similar to all instrumental variables regressions, GMM estimators are unbiased. Arellano and Bond (1991) compared the performance of different GMM, OLS, and WG estimators. Using simulations, they found that GMM estimators exhibit the smallest bias variance.

2.2.2 Wald test

The Wald test (Wald, 1943) tests which model variables significantly contribute to the effect. The test (also called the Wald chi-squared test) is a way to determine whether the explanatory variables in a model are significant, meaning that they add explanatory power to the model; variables with no explanatory power can be deleted without affecting the model in any meaningful way. The null hypothesis for the test is *some parameter = some value*

2.2.3 Instrument validation

Once difference or system GMM estimators are obtained, the model's validity must be checked. Arellano and Bond (1991) proposed a test to detect serial correlation in disturbances. Note that the presence of serial correlation in the disturbances affects the validity of some instruments: If ν_{it} they are serially correlated of order 1, then $\Delta\nu_{it}$ it is endogenous to $\Delta\nu_{it}$ (by the presence of $\nu_{i,t-1}$ in the difference) and, therefore, $y_{i,t-2}$ would be an invalid instrument. They tested the serial correlation of disturbances using difference $\Delta\nu_{it}$ instead of level ν_{it}. Therefore, to test the serial correlation of order 1 in levels, one must check for the correlation of order 2 in differences. When the null hypothesis of this test (no serial correlation) is not rejected, validation of the instrumented variables is obtained.

2.2.4 Testing for causality (Granger, 1969)

The causal relationship between two stationary series, Xt and Yt, can be assessed based on the following bivariate autoregression:

$$X_t = \varphi_0 + \sum_{k=1}^{p} \varphi_k Y_{t-k} + \sum_{k=1}^{p} \phi_k X_{t-k} + \upsilon_t \quad \text{and} \quad Y_t = \alpha_0 + \sum_{k=1}^{p} a_k Y_{t-k} + \sum_{k=1}^{p} \beta_k X_{t-k} + u_t$$

where p is a suitably chosen positive integer; α_k and β_k, k=0,1,2,3,…,p, are constants; u_t and υ_t are the usual disturbance terms with zero mean and finite variance. The null- hypothesis that X_t does not Granger-cause Y_t is rejected if β_k k>0 in the first equation is jointly significantly different from zero according to a standard joint test (e.g., an F test). Similarly, Y_t Granger causes X_t if the coefficients of ϕ_k and k>0 in the second equation are jointly different from zero. A bidirectional causality (or feedback) relation exists if both β_k and ϕ_k k>0 are different from zero.

3. Results and Discussion

3.1 Summary statistics

District-wise total numbers of infected cases are presented in Table 1 and depicted in Figure 1. The result reveals that all over Karnataka state, 3768297 COVID-19 infected cases were registered during the study period. The highest number of COVID-19 infected cases were reported in Bengaluru Urban (1759033), followed by Mysuru (214833). The lowest number of cases are reported in Haveri (23570).

Table 1: District-wise summary statistics of COVID-19 infected cases

District	Sum	Mean	Std.Err	Maxi.	Mini
Bengaluru Urban	1759033	97724.06	34172.74	456462	4169
Ballari	106102	5894.56	2279.16	38689	25
Dakshina Kannada	126434	7024.11	1810.92	30157	369
Dharwad	77730	4318.33	1625.16	24741	73
Bengaluru Rural	76198	4233.22	1491.26	23182	63
Vijayapura	37025	2056.94	695.27	11306	10
Hassan	134045	7446.94	2745.31	44488	228
Uttara Kannada	65319	3628.83	1465.09	25736	117
Udupi	88427	4912.61	1566.02	26494	161
Chamarajanagar	40227	2234.83	2234.83	15394	48
Bagalkot	37347	2074.83	870.38	15360	3
Tumakuru	147553	8197.39	3409.28	55328	226
Davanagere	72263	4014.61	1879.02	34082	14
Chikkaballapura	51693	2871.83	1042.00	16689	22
Kalaburagi	69511	3861.72	1376.20	19445	23
Ramanagara	27490	1527.22	644.03	11089	8
Koppal	37796	2099.78	794.81	13222	3
Raichur	42132	2340.67	946.31	16100	7
Chitradurga	38949	2163.83	686.71	10076	41
Yadgir	24910	1383.89	676.33	12161	1
Bidar	27298	1516.56	560.70	8310	4
Belagavi	86713	4817.39	1859.06	32858	76
Kodagu	44871	2492.83	820.27	13351	119
Mandya	91789	5099.39	2003.12	29152	93
Kolar	59298	3294.33	1282.73	19609	95
Shivamogga	75775	4209.722	1431.20	24042	87
Gadag	28452	1580.67	625.47	10659	2
Chikkamagaluru	55514	3084.11	1123.72	19188	80
Mysuru	214833	11935.17	4075.50	65511	457
Haveri	23570	1309.44	467.91	7311	2

In Bengaluru, a maximum (456462) number of cases was reported in December 2021, and a minimum (4169) was reported in October 2021. In the case of the Mysuru district, a maximum (of 65511) number of cases was reported in April 2021, and a minimum (457) cases were reported in November 2021.

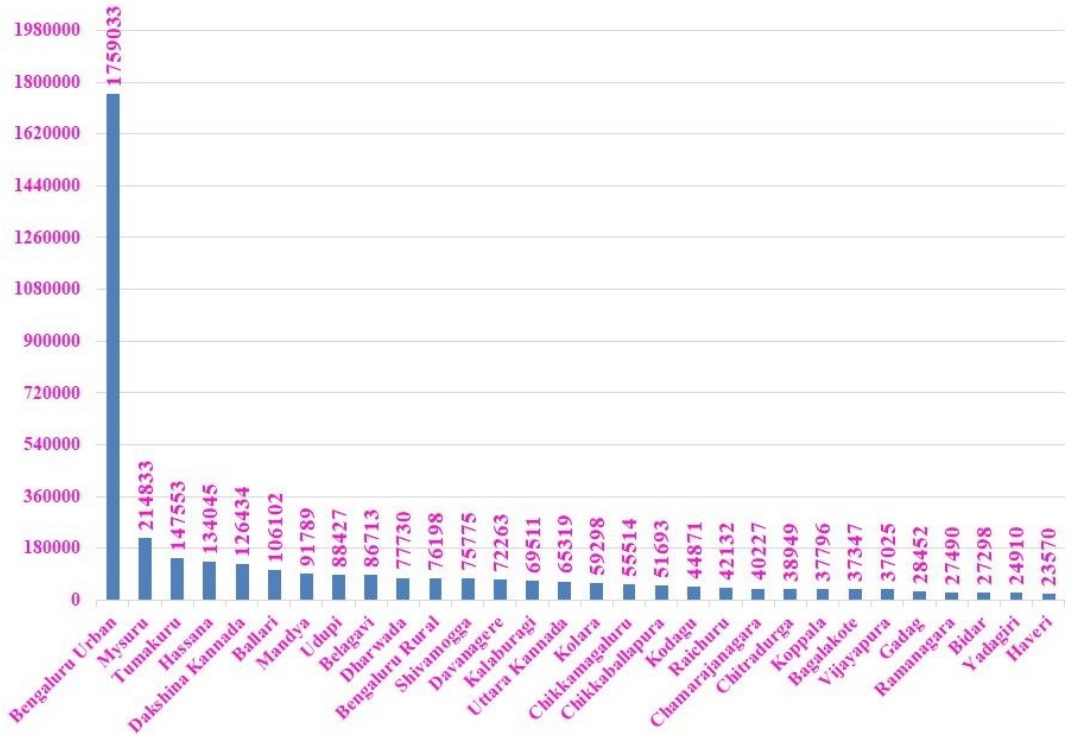

Figure 1: District-wise total number of COVID-19 of infected cases

District-wise total numbers of deaths due to COVID-19 infections are presented in Table 2 and depicted in Figure 2. The result reveals that all over Karnataka state, 36845 deaths due to COVID -19 infections were registered during the study period. The highest number of deaths, 15592, have been reported in Bengaluru Urban, and the lowest of 195 in Yadagiri districts. Next to Bengaluru Urban, 2370 deaths in Mysuru, with a maximum of 538 deaths in May 2021; 1638 deaths in Dakshina Kannada, with a maximum of 276 deaths in May 2021; 1607 deaths in Ballari, with a maximum of 623 deaths in April 2021; 1253 in Dharwad, with a maximum of 262 deaths in April 2021; 1230 in Hassana, with a maximum of 402 deaths in April 2021; 1098 in Tumakuru, with a maximum of 422 deaths in April 2021; 1058 in Shivanmogga, with a maximum of 421 deaths in April 2021; 925 in Belagavi, with a maximum of 209 deaths in May 2021 have been reported. In Bengaluru Rural, 850 total deaths have been notified, and a maximum of 430 deaths was reported in April 2020.

Month-wise total numbers of COVID-19 infected cases are reported in Table 3 and depicted in Figure 3. The highest number of 1058453 infected cases was reported in April 2021, and the lowest of 7086 cases in October 2021. In all the month's maximum number of infected cases have been reported in Bengaluru Urban.
Month-wise total deaths due to COVID-19 infections are reported in Table 4 and depicted in Figure 4. The highest number of 13296 deaths was reported in April 2021, and the lowest of 119 cases was in October 2021. The month's maximum number of deaths have been reported in Bengaluru Urban.

Table 2: District-wise summary statistics of deaths due to COVID-19 infections

District	Sum	Mean	Std.Err	Maxi.	Mini
Bengaluru Urban	15592	866.22	376.87	6809	48
Ballari	1607	89.28	36.38	623	0
Dakshina Kannada	1638	91.00	20.99	276	0
Dharwad	1253	69.61	20.20	262	3
Bengaluru Rural	850	47.22	23.85	430	0
Vijayapura	468	26.00	9.89	168	0
Hassan	1230	68.33	23.09	402	1
Uttara Kannada	756	42.00	19.72	362	0
Udupi	486	27.00	7.75	130	1
Chamarajanagar	510	28.33	13.00	237	0
Bagalakot	329	18.28	7.29	129	0
Tumakuru	1098	61.00	22.88	422	5
Davanagere	574	31.89	12.59	192	0
Chikkaballapura	432	24.00	9.41	165	0
Kalaburagi	807	44.83	17.36	296	0
Ramanagara	318	17.67	8.01	147	0
Koppal	498	27.67	10.60	139	0
Raichur	328	18.22	6.74	103	0
Chitradurga	204	11.33	3.73	65	0
Yadgir	195	10.83	6.19	109	0
Bidar	368	20.44	7.93	124	0
Belagavi	925	51.39	15.21	209	0
Kodagu	325	18.06	7.32	135	0
Mandya	655	36.39	13.27	232	1
Kolar	615	34.17	10.00	131	0
Shivamogga	1058	58.78	24.80	421	0
Gadag	315	17.50	6.88	109	0
Chikkamagaluru	408	22.67	6.98	98	0
Mysuru	2370	131.67	36.93	538	3
Haveri	633	35.17	14.05	218	0

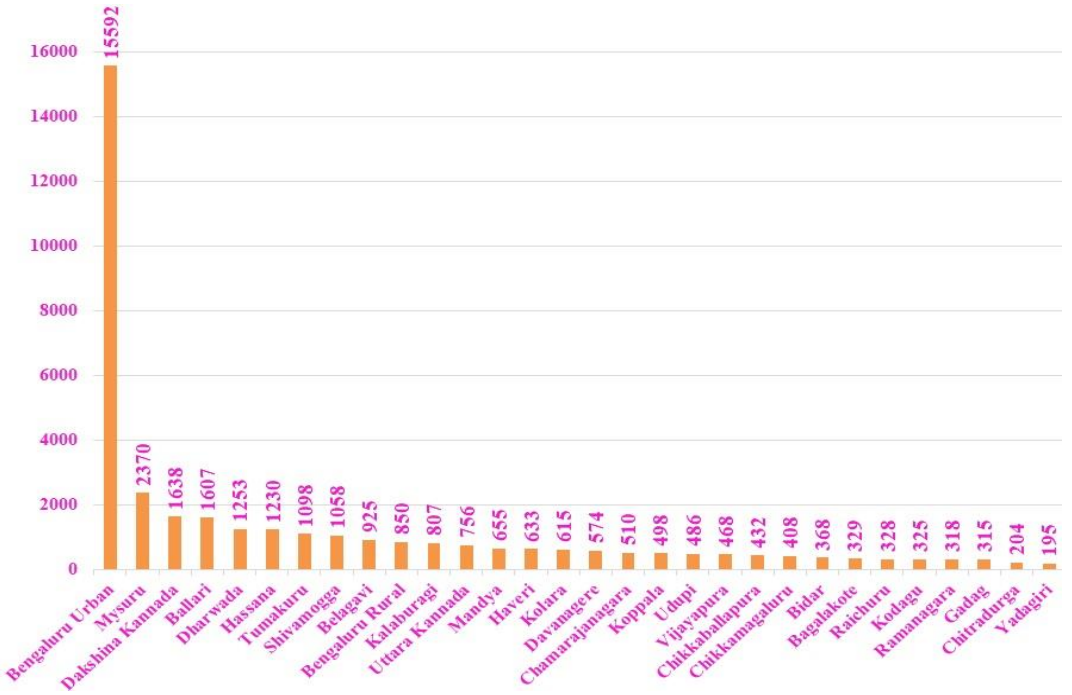

Figure 2: District-wise total number of deaths due to COVID-19 infections

Table 3: Month-wise summary statistics of the number of COVID-19 infections

Month	Sum	Mean	Std.Err	Maxi.	Mini
July-2020	107098	3569.93	1616.47	49954	362
August-2020	213126	7104.2	2241.90	69940	980
September-2020	373121	12437.37	7157.95	219371	1260
October-2020	208293	6943.1	3205.96	98953	544
November-2020	33289	1109.63	582.04	17894	141
December-2020	18791	626.37	327.38	10036	32
January-2021	11473	382.43	214.77	6550	16
February-2021	45404	1513.47	952.36	28980	32
March-2021	519411	17313.7	10440.52	319007	879
April-2021	1058453	35281.77	12394.29	387136	5897
May-2021	224985	7499.50	1650.96	46912	320
June-2021	56121	1870.7	508.03	13104	67
July-2021	42460	1415.33	449.51	10109	32
August-2021	25461	877.97	310.15	8234	12
September-2021	11744	391.47	175.77	5221	4
October-2021	7086	236.2	137.611	4169	2
November-2021	10867	362.23	230.55	7001	1
December-2021	801097	26703.23	14919.78	456462	1941

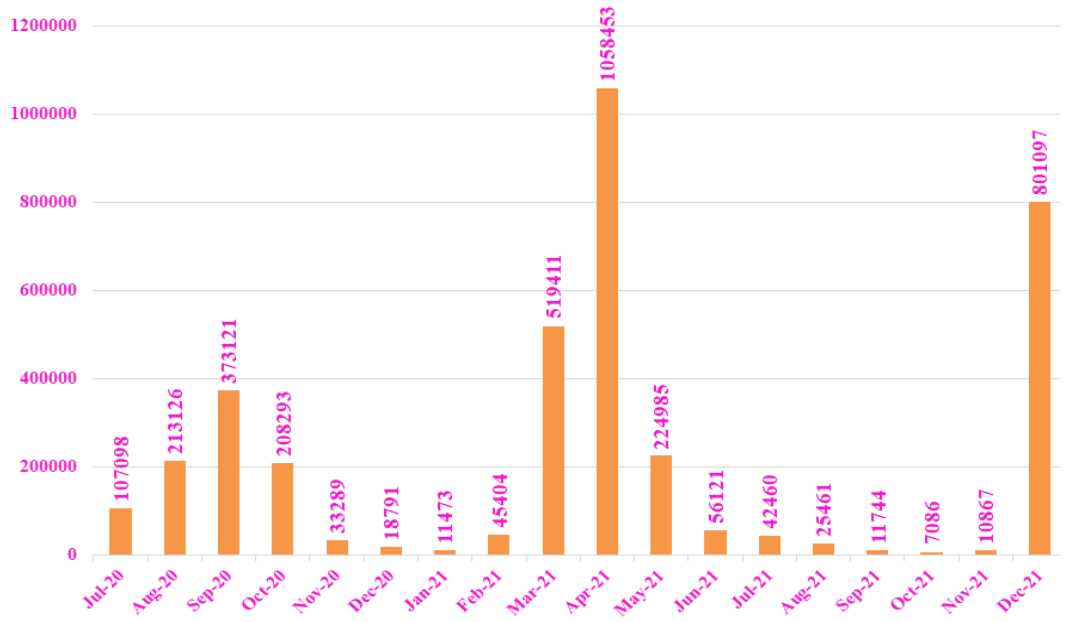

Figure 3: Month-wise total numbers of COVID-19 infected cases

Table 4: Month-wise summary statistics of the number of deaths due to COVID-19 infections

District	Sum	Mean	Std.Err	Maxi.	Mini
July-2020	2063	68.77	30.48	932	3
August-2020	3284	109.47	29.97	903	12
September-2020	3027	100.9	30.87	931	10
October-2020	2174	72.47	29.01	887	8
November-2020	298	9.93	5.62	171	0
December-2020	122	4.07	2.33	70	0
January-2021	111	3.70	2.85	86	0
February-2021	234	7.80	4.52	136	0
March-2021	2938	97.93	57.25	1745	0
April-2021	13296	443.2	220.94	6809	65
May-2021	5486	182.87	65.79	2004	5
June-2021	1433	47.77	10.12	228	0
July-2021	731	24.37	5.96	142	1
August-2021	451	15.55	5.80	155	0
September-2021	274	9.13	4.06	121	0
October-2021	119	3.97	1.58	48	0
November-2021	122	4.07	2.03	62	0
December-2021	678	22.6	6.41	191	0

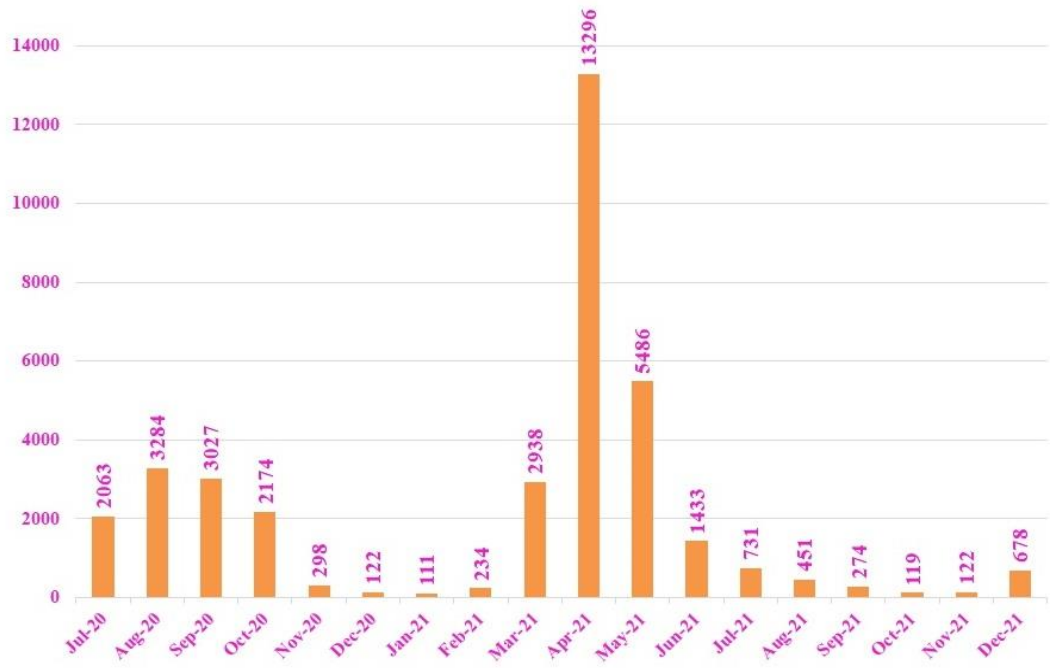

Figure 4: Month-wise total numbers of deaths due to COVID-19 infections

3.2 Investigating cross-sectional dependence

Cross-sectional dependence is one of the essential diagnostics a researcher should investigate before performing panel data analysis. In this context, the Breusch and Pagan (1980) LM test, Pesaran (2004) scaled LM test, Pesaran (2004) CD test, and Baltagi et al. (2012) biased-corrected scaled LM test statistics values have been calculated and presented in Table 5.

Table 5: Characteristics of Residual Cross-Sectional dependence test

Test	Statistic	d.f.	Prob.
Breusch-Pagan LM	2812.204	435	0.0000
Pesaran scaled LM	80.59476		0.0000
Pesaran CD	44.24982		0.0000

Evidence from the above table suggested a rejection of the null hypothesis of no cross-sectional dependence, i.e., the existence of cross-sectional dependence among the regressors at a 1 % significance level for the Breusch-Pagan LM and Pesaran Scaled LM tests. This means that there is a certain level of dependence among districts of Karnataka state, thereby confirming the appropriateness of the panel data modeling for the number of COVID-19 infected cases and deaths in the 30 different districts (cross-section) of Karnataka state.

3.3 Unit root tests

Before estimating the model, it is necessary to conduct stationary tests for the variables. Spurious regression may occur if the variables are non-stationary (Dilmaghani and Tehranchian, 2015). Therefore, Levin-Lin-Chu unit root tests (Levin et al. 2002) were performed to test the stationarity of the study variables, viz., the number of COVID-19-infected patients (CASES) and deaths (DEATHS) due to COVID-19 and the results are reported in Table 6. The test results reveal that the two variables under study, CASE and DEATH, are stationary since the values of the Levin, Lin, and Chu t-statistics are highly significant ($p<0.0000$). Hence, the variables under study are stationary.

Table 6: Unit root test results for the variable Cases and Deaths

Method	Cases		Deaths	
	Statistic	Prob**	Statistic	Prob**
Levin, Lin & Chu t*	-8.7557	0.0000	-10.9169	0.0000

** Probabilities are computed assuming asymptotic normality

3.4 Panel Cointegration tests

To avoid spurious regression, one should test the cointegration between the independent and dependent variables (Dilmaghani and Tehranchian,2015). For this purpose, the Pedronic cointegration test (Pedroni, 2000, 2004); (with no deterministic trend; deterministic intercept and trend; and no deterministic intercept or trend) has been carried out in addition to the Kao cointegration test (Kao and Chiang,2000) to assess the long-run equilibrium relationships between the variables in the model. The test results are presented in Table 7 through Table 10.
The test results reveal that out of eleven tests for the null hypothesis of No Cointegration is rejected since most of the test statistics p values are <0.0.000, indicating that there exists a cointegration, i.e., long-term relationship between the number of deaths due to COVID-19 and the number of COVID-19 infected cases.
The Kao cointegration test results in Table 10 reveal a strong long-term relationship and cointegration between the study variables since the null hypothesis of no cointegration is rejected.

Table 7: Characteristics of Pedroni cointegration test (No deterministic trend)

Name of the test	Statistic	Prob.	Weigted Statistic	Prob.
Panel v-Statistic	31.48084	0.0000	5.122892	0.0000
Panel rho-Statistic	-9.530230	0.0000	-7.768760	0.0000
Panel PP-Statistic	-3.262046	0.0006	-7.765797	0.0000
Panel ADF-Statistic	-4.783584	0.0000	-9.051099	0.0000
Group rho-Statistic	-5.497392	0.0000		
Group PP-Statistic	-4.716377	0.0000		
Group ADF-Statistic	-4.825069	0.0000		

Table 8: Characteristics of Pedroni cointegration test (Deterministic intercept and trend)

Name of the test	Statistic	Prob.	Weigted Statistic	Prob.
Panel v-Statistic	14.47767	0.0000	-1.347181	0.9110
Panel rho-Statistic	-4.567400	0.0000	-3.851955	0.0001
Panel PP-Statistic	-1.525791	0.0635	-7.193887	0.0000
Panel ADF-Statistic	-2.525812	0.0058	-8.135059	0.0000
Group rho-Statistic	-1.748681	0.0402		
Group PP-Statistic	-3.487189	0.0002		
Group ADF-Statistic	-3.547681	0.0002		

Table 9: Characteristics of Pedroni cointegration test (No deterministic intercept or trend)

Name of the test	Statistic	Prob.	Weigted Statistic	Prob.
Panel v-Statistic	53.91298	0.0000	12.53456	0.0000
Panel rho-Statistic	-15.42834	0.0000	-13.12926	0.0000
Panel PP-Statistic	-4.878379	0.0000	-9.039844	0.0000
Panel ADF-Statistic	-4.835779	0.0000	-9.834622	0.0000
Group rho-Statistic	-8.778302	0.0000		
Group PP-Statistic	-7.285127	0.0000		
Group ADF-Statistic	-7.087520	0.0000		

Table 10: Characteristics of Kao cointegration test (No deterministic trend)

Test Name	t-Statistic	Prob.
ADF	-5.889986	0.0000

Therefore, the long-run equilibrium relationship between the variables and the absence of spurious regression in the model is confirmed.

3.5 Pooled OLS regression or Constant Coefficients Model

The panel least squares method is employed with the number of deaths due to COVID-19 as the dependent variable and the number of COVID-19-infected patients as the independent variable. The regression results based on EViews, Version 11, are presented in Table 11. The results reveal that the slope is highly significant, and the model F-statistic is also highly significant, with a very high R^2 of 47%. This shows that the number of COVID-19 infected cases is related to a considerable variation in the number of deaths due to COVID-19.

Table 11: Characteristics of pooled OLS regression

Variable	Coefficient	Std. Error	t-Statistic	Prob.
CASES	0.007109	0.000328	21.69516	0.0000
C	18.62529	10.59094	1.758606	0.0792
Root MSE	239.8618	R-squared		0.466630
Mean dependent var	68.23148	Adjusted R-squared		0.465638
S.D. dependent var	328.7375	S.E. of regression		240.3072
Akaike info criterion	13.80541	Sum squared resid		31068195
Schwarz criterion	13.82131	Log-likelihood		-3725.461
Hannan-Quinn criteria.	13.81163	F-statistic		470.6800
Durbin-Watson stat	1.427732	Prob(F-statistic)		0.000000

The major problem with this model is that it does not distinguish between the various districts, nor does it tell us whether the response of total COVID-19 deaths to the explanatory variable over time is the same for all districts. Consequently, the error term may correlate with the model's regressor. If so, the estimated coefficients in the above model may be biased and inconsistent.

3.6 Statistic panel model

The statistic panel models viz, fixed and random effect models have been estimated and presented in Tables 12 and 13, respectively. To decide whether to consider the fixed effect or the random effect models, the Hasman test has been carried out, and the results are presented in Table 14. Since the Hasman test statistics' p-value value is non-significant, indicating that the random effect model is appropriate.

The random effect model presented in Table 13 reveals both the lag variables (exogenous variables) explain 58 % of variations in deaths. Furthermore, both the exogenous variables, viz., DEATHS (-1) and CASES(-1), are highly significant and show a positive influence on the endogenous variable (DEATH).

Since the estimated Durbin-Watson stat value is 2.09, the errors due to the estimated model are uncorrelated, which is preparable.

Table 12: Characteristics of estimated fixed effect model

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DEATHS(-1)	0.376192	0.033038	11.38656	0.0000
CASES	0.006508	0.000362	17.97833	0.0000
C	-5.195117	10.47488	-0.495960	0.6202
Root MSE	215.7034	R-squared		0.586890
Mean dependent var	68.20000	Adjusted R-squared		0.560099
S.D. dependent var	335.9313	S.E. of regression		222.8066
Akaike info criterion	13.71118	Sum squared resid		23729258
Schwarz criterion	13.97686	Log-likelihood		-3464.350
Hannan-Quinn criteria.	13.81534	F-statistic		21.90574
Durbin-Watson stat	2.141517	Prob(F-statistic)		0.000000

Table 13: Characteristics of estimated random effect model

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.110829	10.22960	-0.108590	0.9136
DEATHS(-1)	0.352382	0.030118	11.70007	0.0000
CASES	0.006174	0.000314	19.66744	0.0000
Effects Specification			**S.D.**	**Rho**
Cross-section random			0.000000	0.0000
Idiosyncratic random			222.8066	1.0000
Weighted Statistics				
Root MSE	216.6220	R-squared		0.583364
Mean dependent var	68.20000	Adjusted R-squared		0.581721
S.D. dependent var	335.9313	S.E. of regression		217.2619
Sum squared resid	23931793	F-statistic		354.9453
Durbin-Watson stat	2.092160	Prob(F-statistic)		0.000000
Unweighted Statistics				
R-squared	0.583364	Mean dependent var		68.20000
Sum squared resid	23931793	Durbin-Watson stat		2.092160

Table 14: Characteristics of estimated Hausman test

Test Summary	Chi-Sq. Statistic	Chi-Sq. d.f.	Prob.
Cross-section random	4.061228	2	0.1313

3.7 GMM model

To estimate the models through GMM, it is necessary to use instrumental variables. Instrumental variables should be chosen based on their abilities to estimate and identify the conditions. If adding a new instrumental variable positively affects the estimation quality, the variable will be used as the instrumental variable (Ghiasi et al. 2019).

The panel GMM method with the first difference transformation has been calculated and presented in Table 15. R^2 is not used as a statistical standard for determining the model's goodness of fit, but the J-statistics assess the validity of the instrument variable used in the model. The probability value of the J-test is estimated to be 0.4512. Thus, H_0 (i.e., the validity of the instruments defined in the model) is accepted; hence, the instrumental variable used in the model is valid.

Table 15: Characteristics of estimated GMM model

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DEATHS(-1)	0.336319	0.001514	222.1124	0.0000
CASES	0.011738	3.85E-05	304.7321	0.0000
Effects Specification				
Cross-section fixed (first differences)				
Root MSE	312.5533	Mean dependent var		-6.170000
S.D. dependent var	439.2945	S.E. of regression		318.9984
Sum squared resid	29306879	J-statistic		18.06828
Instrument rank	30	Prob(J-statistic)		0.451163

The one-period lag (DEATHS(-1)) has a positive and highly significant effect on the number of deaths (DEATH). The correlation between number of deaths at time t positively correlated with the number of deaths in the previous period. This result reflects the dynamics of the number of deaths over time. Also, the number of infected cases positively and highly significantly affects the number of deaths.

3.8 Wald test

According to the Wald test (Wald, 1943) results presented in Table 16, the hypothesis that all coefficients are zero is rejected at a 1% significance level; thus, the validity of the significance of coefficients is confirmed, meaning that they add explanatory power to the model.

Table 16: Characteristics of Wald test

Test Statistic	Value	df	Probability
F-statistic	2.21E+08	(2, 478)	0.0000
Chi-square	4.42E+08	2	0.0000

3.9 Arellano-Bond serial correlation test

To ensure the absence of serial correlation of first-order difference in residuals, the first and second-order serial auto-correlation test proposed by Arellano and Bond (1991) is calculated and presented in Table 17. The test results reveal that the null hypothesis of the absence of serial auto-correlation, which should be greater than 5 % in the first and second orders, is accepted. Hence, the residual due to the model is free from auto-correlation.

Table 17: Characteristics of the Arellano-Bond serial correlation test

Test order	m-Statistic	rho	SE(rho)	Prob.
AR(1)	-1.041994	-18859787.70	18099716.65	0.2974
AR(2)	1.005126	4130237.36	4109172.35	0.3148

Since the null hypothesis of this test (no serial correlation) is accepted, validation of the instrumented variables is obtained. Therefore, the method of estimation is suitable for the model.

3.10 Test of causality

The pairwise Granger causality test was employed to assess whether causal relationships exist among the variables and determine the direction of the causality. The results are presented in Table 18. The test results reveal that the null hypothesis of no causality between the independent and dependent variables running in either direction is rejected; hence, bi-directional causality exists between the study variables.

Table 18: Characteristics of pairwise Granger causality test

Null Hypothesis:	Obs	F-Statistic	Prob.
CASES does not Granger Cause DEATHS	480	417.952	2E-105
DEATHS does not Granger Cause CASES		26.2602	2.E-11

4. Conclusions

The result reveals that all over Karnataka state, 3768297 COVID-19 infected cases have been registered from July 2020 to December 2021. The highest number of infected cases were reported in Bengaluru Urban (1759033), followed by Mysuru (214833). The lowest number of cases are reported in Haveri (23570). Due to COVID-19 infections,36845 deaths were registered all over the state during the study period. The highest number of deaths has been reported in Bengaluru Urban (15592) and the lowest of 195 in Yadagiri districts. The panel GMM model with the first difference transformation was found suitable for investigating the number of death dynamics over time. The one-period lag has a positive and significant effect on the number of deaths. The Wald test confirms the validity of the coefficients' significance and adds explanatory power to the model. The correlation between number of fatalities at time t positively correlated with the number of deaths in the previous period. Also, the number of infected cases positively and significantly affects the number of deaths. Bi-directional causality exists between the COVID-19 infected cases and deaths.

References

[1] Anderson, T.W. and Hsiao, C. (1982). Formulation and estimation of dynamic models using panel data.Journal of Econometrics, 18, pp.47-82.

[2] Arellano, M. and Bond,S.(1991).Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. Review of Economic Studies, 58, pp.277-297, 1991. http://dx.doi.org/10.2307/2297968

[3] Baltagi, B.H., Feng,Q., and Kao,C. (2012).A Lagrange Multiplier test for cross-sectional dependence in a fixed effects panel data model. Journal of Econometrics, 170(1), pp.164-177.

[4] Breusch, T. and Pagan,A. (1980).The Lagrange Multiplier Test and Its Applications to Model Specification in Econometrics. Review of Economic Studies, 47, pp.239-254.

[5] Dilmaghani, A.K. and Tehranchian, A.M. (2015). The impact of monetary policies on the exchange rate: A GMM approach. Iran.Econ.Rev., 19(2), pp.177-191.

[6] Ghiasi, M, Sarlak, A. and Ghafari,H.(2019).Effect of health capital on the economic growth process of Iranian provinces using generalized method of moments approach. J.Research & Health, 9(1), pp.21-28.

[7] Granger, C.W.J. (1969). Investigating causal relations by econometric models and cross-spectral methods. Econometrica, 37(3), pp.424-438.

[8] Hansen, L.P. and Singleton, K.J. (1982). Generalized instrumental variables estimation of nonlinear rational expectations models. Econometrica, 50(5), pp.1269-1286.

[9] Holtz-Eakin, D., Newey, W. and Rosen, H.S. (1988). Estimating Vector Autoregressions with Panel Data. Econometrica, 56(6), pp.1371-1395.

[10] Kao, C. and Chiang, M.H. (2000). On the Estimation and Inference of a Cointegrated Regression in Panel Data, in B.Baltagi (ed.), Nonstationary Panels, Panel Cointegration, and Dynamic Panels. Advances in Econometrics, 15, JAE Press, pp. 179-222.

[11] Levin, A., Lin, C.F. and Chu, C.S. (2002). Unit root test in panel data: Asymptotic and Finite-Sample properties. Journal of Econometrics, 108(1), pp.1-24.

[12] Pedroni, P. (2000). Fully Modified OLS for Heterogeneous Panels, in B.Baltagi (ed.), in B.Baltagi (ed.), Nonstationary Panels, Panel Cointegration, and Dynamic Panels. Advances in Econometrics, 15, JAE Press, pp. 93-130.

[13] Pedroni, P. (2004). Panel Co-integration: Asymptotic and finite sample properties of pooled time series tests with an application to PPP hypothesis: new results.Econometric Theory, 20, pp.597-625.

[14] Pesaran, M.H. (2004). General Diagnostic Test for Cross Section Dependence in Panels. University of Cambridge, Faculty of Economics, Cambridge Working Paper in Economics No.0435.

[15] Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. Transactions of the American Mathematical Society, Vol.54, pp. 426-482.