Panel Regression Models for Paddy (Oryza sativa) Crop Production

Rajarathinam, A[1] and Suba S.S[2].
[1,2]*Department of Statistics, Manonmaniam Sundaranar University, Tirunelveli – 627 012, Tamil Nadu State, India*
rajarathinam@msuniv.ac.in[1] and sssuba@gmail.com[2]

Abstract

The present investigation was carried out to study area production trends of Paddy crop grown in different districts of Tamil Nadu state, India during the period 1998-99 to 2010-2020 based on Panel Regression Model. The statistically most suited Panel Regression model was selected based on Hausman and Wald test. The study variables namely the area under the Paddy crop (AREA) and the production (PRODN) of Paddy crop were found to be stationary at level. Analysis of variance test indicated that district to district crop productions were highly significant. Highest area under the crops and productions were registered in Tiruvarur, Thanjavur etc., Very lowest were registered in Coimbatore and Nilgiris districts. The fixed effect model was found to be suitable to study the trend and this model explains the 87% of variations in Paddy crop production.

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

1. Introduction

Over the last few decades, regression modelling has traditionally been employed in agricultural production prediction and classification. For agricultural planning purposes, decision-makers need simple and reliable estimation techniques for crop production prediction. Multiple regressions, Discriminant analysis, factor analysis, principal component analysis, cluster analysis and logistic regression analysis are the most commonly used statistical techniques for the prediction and classification of agricultural-related production. In agricultural production time series data, the problems of multicollinearity, autocorrelation and extreme values are unavoidable. In such complex situations, regression models may not provide accurate predictions. Regression models need to fulfil regression assumptions such as autocorrelation and multiple collinearity between the independent variables, which causes the estimated regression models to be unfit and the estimated parameter values obtained based on these models to be inefficient. In most agricultural practices, crop production is influenced by a great variety of interrelated factors such as autocorrelation, and it is difficult to describe their relationships using conventional methods (Zaefizadahet al., 2011).

In this study, panel data regression model is used to combat the complicated relations and strong autocorrelation present in the crop production data.

Panel data is a combination of cross-sectional and time series data. Therefore, using a regression suited to panel data has the advantage of distinguishing between fixed and random effects. Fixed effects, effects that are independent of random disturbances, e.g., observations independent of time. Random effects, effects that include random disturbances. Panel data is more informative since it includes more information, but it has to be modeled correctly by taking into account fixed vs. random effects.

Panel data helps us to controls heterogeneity of cross-section units such as individuals, states, firms, countries etc., over time. Panel data estimation considers all cross-section units as heterogeneous. It helps us to get unbiased estimation. There are time invariant and state invariant variables which we observe or not. As compared to pure cross section and time series, panel data estimation is better to identify and measure effects of independent variables on dependent variables what we cannot measure using time series and cross section data. In addition to this “Panel data give more informative data, more variability, less collinearity among the variables, more degree of freedom and more efficiency”. It is also better estimation method to study the duration of economic states and the “dynamics of change”, over time. It is a good estimation method to ‘construct and test complicated behavioral models’, (Baltagi, (2001)).

Based on the above discussion, the present study is aimed to study the trends in food grain production in different states in India over the period 2001-02 to 2020-2021 based on panel regression model.

2. Materials and Methods

2.1. Materials: The present investigation was carried out to study the dynamic relationships between area and production and its trends in paddy crop in different districts of Tamil Nadu. The cross-sectional time series data on paddy crop during the period 1998-99 to 2010-2020) have been collected from Reserve Bank of India - Handbook of Statistics on Indian Economy (rbi.org.in).

2.2. 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.

A simple panel data regression model is specified as

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

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, $\alpha \; and \; \beta$ are the intercept and slope, i stands for the ith cross-sectional unit and t for the tth 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 28 (i=1, 2, 3, 4,…,28), and t=1, 2, 3,…, 22.Detailed discussions of panel data models were given in Hsiao,(2003),Greene, (2008) and Gujarathi, (2017).

2.2.1. Unit Root Test: Unit roots for the panel data can be tested using either the Leuin-Llin-Chu, (2002) test or the Hadri, (2000) 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.

2.2.2. 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 same irrespective of cross-sectional unit being considered. Accepting this homogeneity assumption (also called pooling assumption), the CCM uses the panel (or pooled) data set, and applies Ordinary Least Squares (OLS) method to estimate unknown parameters of the model. Thus, the CCM is nothing but straightforward application of OLS to a given panel or pooled data to obtain estimates for unknown parameters of the model (Bhaumik, (2017)).

2.2.3. 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.

2.2.4. Fixed-Effect OR Least-Square Dummy Variable Regression Model: Fixed effect regression model indicates that each unit has its own intercept. There will be heterogeneity among the unit due to individual intercepts. Here in fixed effect model the unit intercepts are time-invariant (do not vary over time) even if they might be different among cross section units. However the fixed effect model believes that the coefficients of the independent variables do not vary across cross-section unit or over time.These fixed effects model can be implemented with the dummy variable technique. Therefore, the fixed effects model can be written as

$$Y_{it} = \alpha_1 + \alpha_2 D_{2i} + \alpha_{3i} D_{3i} + \alpha_4 D_{4i} + ... + \alpha_{28} D_{28i} + \beta_2 X_{it} + \nu_{it}$$

where D_{2i} =1 if the observation belongs to Cuddalore district and 0 otherwise, D_{3i} =1 if the observation belongs to Dharmapuri 0 otherwise, D_{4i} =1 if the observation belongs to Dindigul and 0 otherwise, and so on. Here, the district Coimbatore is considered the baseline, or reference, category. Thus, the intercept α_1 represents the intercept value of the Coimbatore district, and the other α coefficients represent how much the intercept values of the other districts differ from the intercept value of the Coimbatore district. Thus, α_2 shows how much the intercept value of the second district, Cuddalore differs from α_1. The sum ($\alpha_1 + \alpha_2$) gives the actual value of the intercept for Cuddalore The intercept values of the other districts can be computed similarly.

2.2.5. Random-Effect (RE) Model: Random effects model is also called error component model (ECM). In this model the cross section units will have random intercept instead of fixed intercept. The rationale behind random effects model is that, unlike the fixed effect model, the variation across entities is assumed to be random and uncorrelated with the predictor or independent variables included in the model, the crucial distinction between the fixed and random effects is whether the unobserved individual effects embodies elements that are correlated with regressors in the model, not whether these effects are stochastic or not (Green, 2008)**.** The RE model assumes that individual-specific effects α_i are random and one should include α_i in the error term. Each cross-section has the same slope parameters and a composite error term. So the model (1) become Random-Effect Model (REM):

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

$$\text{Let } \varepsilon_{it} = \alpha_i + \nu_{it}\,.$$

Here ε_{it}, α_i and ν_i are normally distributed with zero means and constant variances σ_ε^2, σ_α^2 and σ_ν^2, respectively.

Hence: $\mathrm{var}(\varepsilon_{it}) = \sigma_\alpha^2 + \sigma_\nu^2$, and $\mathrm{cov}(\varepsilon_{it}, \varepsilon_{is}) = \sigma_\alpha^2$; therefore, $\rho_\varepsilon = cor(\varepsilon_{it}, \varepsilon_{is}) = \dfrac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\varepsilon^2}$.

Rho is the interclass correlation of the error or the fraction of the variance in the error term due to individual-specific effects. These variable approaches 1 if individual effects dominate the idiosyncratic error (Bhaumik, (2017)).

2.2.6. Hausman test: The Hausman test (Hasman, 1978) is the standard procedure used in empirical panel data analysis to distinguish between the fixed effects and random effects. In the Hausman test the null hypothesis signifies that there is no significant difference in the estimator of fixed effect model and random effect model. If we reject the null hypothesis the fixed effect model will be the appropriate model. Rejecting the null hypothesis shows that there might be correlation between the error term and dependent variable. The test statistic can be calculated is given as follows:

$$H = \left(\widehat{\beta}_{RE} - \widehat{\beta}_{FE}\right)^{|} \left(V(\widehat{\beta}_{RE}) - V(\widehat{\beta}_{FE})\right)\left(\widehat{\beta}_{RE} - \widehat{\beta}_{FE}\right)$$

Here, $\widehat{\beta}_{RE}$ and $\widehat{\beta}_{FE}$ 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 the matrix:

$$\left(V(\widehat{\beta}_{RE}) - V(\widehat{\beta}_{FE})\right)$$

2.2.7. Wald Test: The Wald test (Wald, 1943) 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*.

2.2.8. Breusch-Pagan Lagrange Multiplier Test: The Breusch-Pagan-Godfrey test (Breusch and Pagan, (1980)) 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.

3. RESULTS AND DISCUSSIONS

The results obtained in this paper based on applying different statistical tools related to panel regression models are discussed in sequence below.

3.1. Unit root tests

In analyses of time series data, it is important that the study variables are stationary, which means that the means and variances of the variable data are the same. Accordingly, Levin-Lin-Chu unit root tests were carried out to test the stationarity of the study variables, area under the crop (AREA) and the production (PRODN). The results are reported in Tables 1 and 2.

Table 1: Unit root test results for area (AREA) under the Paddy crop

Individual Effect		Individual Effect and Trend		None	
Statistic	**Prob****	**Statistic**	**Prob****	**Statistic**	**Prob****
6.43486	0.0000	6.74762	0.0000	3.54163	0.0002
** Probabilities are computed assuming asymptotic Normality					

Table 2: Unit root test results for the production (PRODN) of paddy crop

Individual Effect		Individual Effect and Trend		None	
Statistic	**Prob****	**Statistic**	**Prob****	**Statistic**	**Prob****
7.05042	0.0000	7.45309	0.0000	5.85905	0.0000
** Probabilities are computed assuming asymptotic Normality					

The test results presented in Tables 1 and 2 reveal the two variables under study, NCASE and DEATH, to be stationary in level, since the Levin, Lin and Chu t-statistics are found to be highly significant ($p<0.0000$). Hence, the variables under study are found to be stationary.

3.2. Summary statistics

The descriptive statistics presented in the Table 3 reveal that district wise area under the Paddy crops are normally distributed in all the districts except Kanchipuram and Tiruvallur districts since Jarque-Bera statistics values were non-significant. Highest area under the crops were registered in Tiruvarur, Thanjavur etc., Very lowest were registered in Coimbatore and Nilgiris districts. The trends in area and production of Paddy crop are depicted in the Fig.1. & Fig.2 respectively.

Table 3: Summary Statistics of area under the Paddy crop

Sr. No.	Name of District	Sum	Mean	Max.	Min.	S.D.	Jarque-Bera Prob.
1.	Coimbatore	115414	5246.09	16875	680	4682.55	0.0697
2.	Cuddalore	2623811	119264.10	139987	102336	11754.78	0.4486
3.	Dharmapuri	596753	27125.14	61608	3889	16546.05	0.1299
4.	Dindigul	346772	15762.36	28437	1522	7160.32	0.3840
5.	Erode	782060	35548.18	68130	695	19277.00	0.8635
6.	Kanchipuram	2326587	105754.00	207871	61881	38303.09	0.0030
7.	Kanyakumari	416029	18910.41	32907	9628	6926.36	0.3969
8.	Karur	300142	13642.82	18398	3672	3378.22	0.0562
9.	Madurai	1156984	52590.18	88338	10407	20807.37	0.5518
10.	Nagapattinam	3505507	159341.20	170840	136039	10061.37	0.1351
11.	Namakkal	273511	12432.32	24167	2188	5954.77	0.4685
12.	Perambalur	541331	24605.95	48734	2613	17761.87	0.2296
13.	Pudukkottai	1910130	86824.09	107199	67238	11054.48	0.6272
14.	Ramanathapuram	2758053	125366.00	136902	114981	5724.71	0.2764
15.	Salem	570461	25930.05	48400	5024	12641.51	0.5779
16.	Sivaganga	1680160	76370.91	89924	63492	8074.29	0.5014
17.	Thanjavur	3717398	168972.60	196816	123293	19728.65	0.6061
18.	Nilgiris	21311	968.68	2611	32	793.84	0.3432
19.	Theni	318444	14474.73	21747	6210	3951.38	0.9031
20.	Tiruvallur	2030891	92313.23	140703	66734	16295.10	0.0041
21.	Tiruvarur	3733186	169690.40	194743	121437	19236.23	0.1810
22.	Thoothukudi	349981	15908.23	22709	4808	4694.22	0.3502
23.	Trichirappalli	1296383	58926.50	79576	27068	13390.18	0.9956
24.	Tirunelveli	1690797	76854.41	98506	29881	17281.57	0.0548
25.	Tiruvannamalai	2375224	107964.70	161709	43872	29809.74	0.9282
26.	Vellore	987726	44896.64	77109	26287	11447.83	0.1536
27.	Viluppuram	3122606	141936.60	182303	75279	29396.23	0.1762
28.	Virudhunagar	579368	26334.91	34953	13826.	5842.06	0.4247

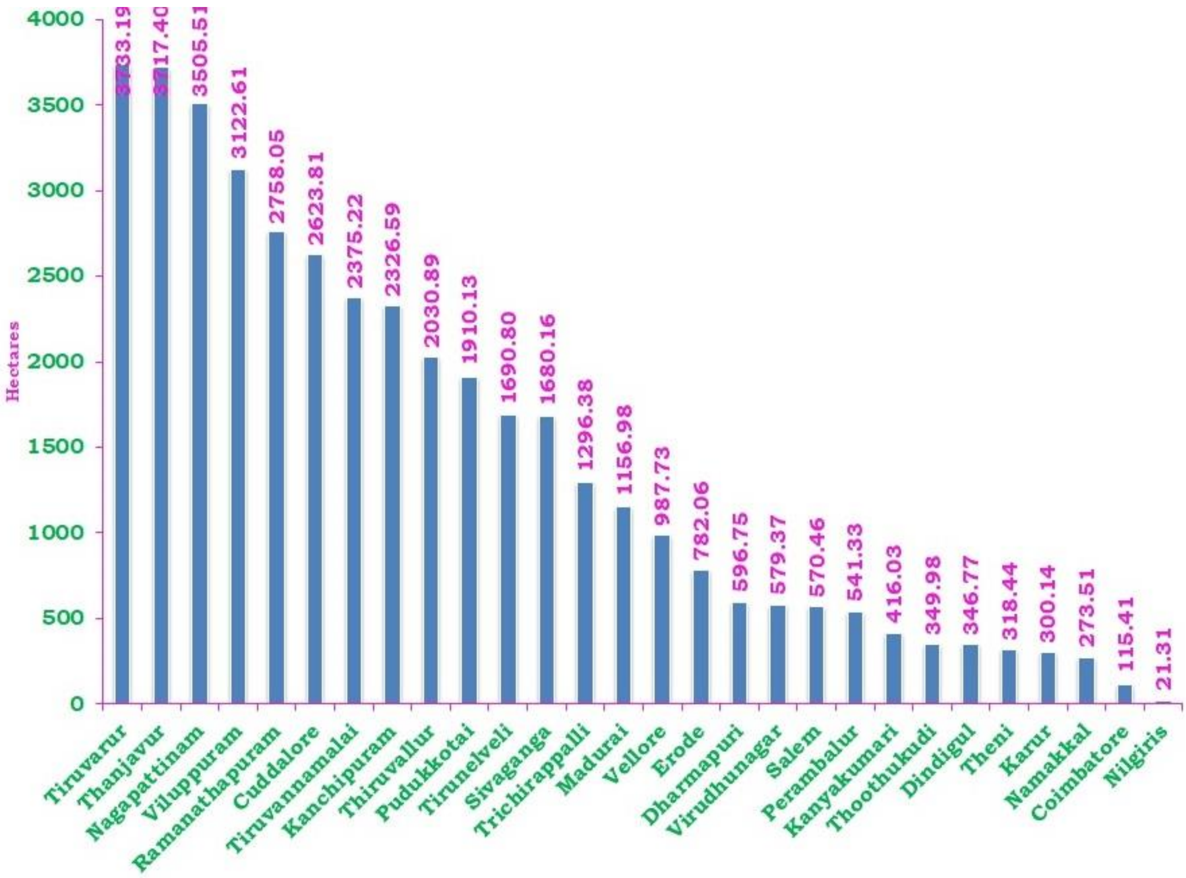

Fig.1: Trends in Area under the Paddy crop

Table 4: Summary statistics for Paddy crop production

Sr. No.	Name of District	Sum	Mean	Max.	Min.	S.D.	Jarque-Bera Prob.
1.	Coimbatore	429818	19537.18	60494	2326	16443.40	0.0859
2.	Cuddalore	8771665	398712.00	641958	218345	118217.90	0.9159
3.	Dharmapuri	2245912	102086.90	227438	17025	58567.46	0.3357
4.	Dindigul	1462684	66485.64	118998	4506	29903.17	0.7592
5.	Erode	3407341	154879.10	306471	1568	79742.41	0.6008
6.	Kanchipuram	8619116	391778.00	711233	278087	114038.30	0.0086
7.	Kanyakumari	1765023	80228.32	143642	33034	29248.21	0.6233
8.	Karur	1089394	49517.91	69898	10792	14312.75	0.1752
9.	Madurai	4474860	203402.70	331342	34120	87674.52	0.5475
10.	Nagapattinam	8788465	399475.70	630607	136637	169431.40	0.3525
11.	Namakkal	1159115	52687.05	105619	7133	26345.87	0.5688
12.	Perambalur	1706031	77546.86	164567	9439	47849.37	0.4684
13.	Pudukkottai	5069817	230446.20	368902	73659	79133.99	0.7713
14.	Ramanathapuram	4133352	187879.60	471001	244	143349.80	0.4918
15.	Salem	2350945	106861.10	204300	18415	51924.89	0.7386
16.	Sivaganga	3246373	147562.40	310176	41283	76099.80	0.7291
17.	Thanjavur	12171223	553237.40	831483	318325	147642.60	0.4881

18.	Nilgiris	74483	3385.59	9192	150	2651.01	0.3592
19.	Theni	1418410	64473.18	80292	36668	14752.21	0.2835
20.	Tiruvallur	7341596	333708.90	484861	195536	82330.56	0.6577
21.	Tiruvarur	10413521	473341.90	852925	112183	231344.00	0.4561
22.	Thoothukudi	1484443	67474.68	108163	15683	22266.58	0.6769
23.	Trichirappalli	5093080	231503.60	309928	76591	60909.63	0.1738
24.	Tirunelveli	7213127	327869.40	488445	105193	87059.67	0.3692
25.	Tiruvannamalai	8267609	375800.40	704247	115453	138544.10	0.8328
26.	Vellore	3772874	171494.30	263941	89863	46508.35	0.9286
27.	Viluppuram	11514713	523396.00	832585	234750	156740.30	0.8697
28.	Virudhunagar	1897842	86265.55	158214	15883	30326.26	0.7807

a

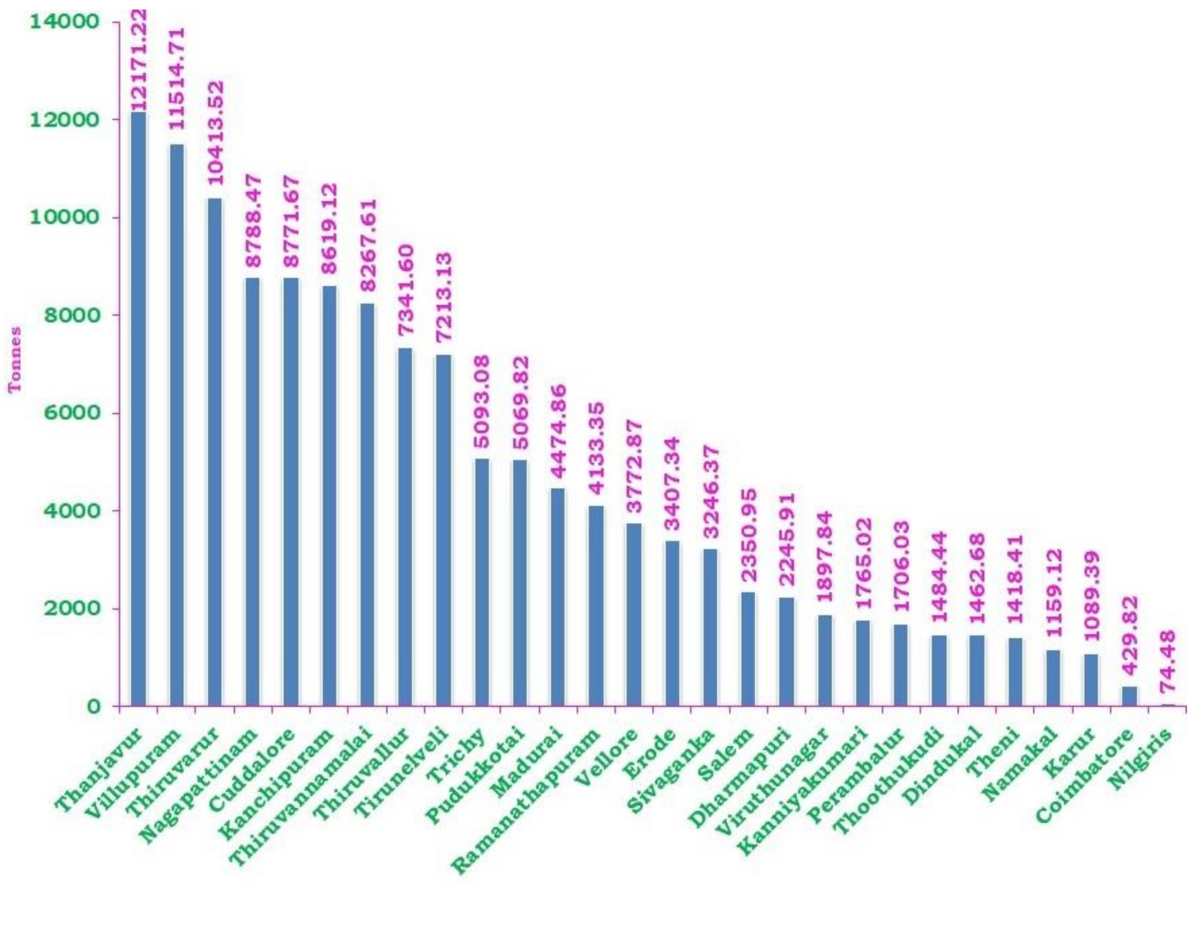

Fig.2. Trends in Paddy crop production

3.3. Variations between districts

To determine the variations across districts, ANOVA tests were carried out individually for each of the study variables, AREA and PRODN, and the results are presented in Tables 5 and 6.

Table 5: Results of test for equality of means of number of COVID-19 infections.

Method	df	Value	Probability
Anova F-test	(27, 588)	251.3787	0.0000
Welch F-test*	(27, 206)	911.7723	0.0000

*Test allows for unequal cell variances

Analysis of Variance

Source of Variation	df	Sum of Sq.	Mean Sq.
Between	27	1.72E+12	6.38E+10
Within	588	1.49E+11	2.54E+08
Total	615	1.87E+12	3.04E+09

Table 6: Results of test for equality of means of number of deaths due to COVID-19.

Method	df	Value	Probability
Anova F-test	(27, 588)	64.51287	0.0000
Welch F-test*	(27, 204.955)	147.5820	0.0000

*Test allows for unequal cell variances

Analysis of Variance

Source of Variation	df	Sum of Sq.	Mean Sq.
Between	27	1.58E+13	5.87E+11
Within	588	5.35E+12	9.09E+09
Total	615	2.12E+13	3.45E+10

The results reveal that since the ANOVA tests are highly significant ($p<0.0000$) for both study variables, highly significant variations occur between the districts.

3.4. Pooled OLS regression or constant coefficients model

The panel least squares method is employed with production (PRODN) as the dependent variable and the area (AREA) under the paddy crop as the independent variable. The regression results based on EViews, Version 11, are presented in Table 7.

Table 7: Results of pooled OLS Regression or Constant Coefficients Model

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	17985.55	5580.043	3.223191	0.0013
AREA	2.948231	0.065388	45.08822	0.0000
Root MSE	89320.98	R-squared		0.768035
Mean dependent var	210037.1	Adjusted R-squared		0.767657
S.D. dependent var	185607.2	S.E. of regression		89466.34
Akaike info criterion	25.64435	Sum squared resid		4.91E+12
Schwarz criterion	25.65872	Log likelihood		-7896.461
Hannan-Quinn criter.	25.64994	F-statistic		2032.948
Durbin-Watson stat	0.965869	Prob(F-statistic)		0.000000

The results reveal that the intercept and slopes are very highly significant, and the model F-statistic is also highly significant, with an exceedingly high R^2 of 77%. This assures the production of paddy crops where significantly influenced by the area under the paddy crop. Every unit increases in area the production would be increased by 3 %.

3.5. FE least squares dummy variable (LSDV) model

The results presented in Table 8 reveal that the FE model is highly significant, with a high R2 of 87%. The slope coefficient for the area under the paddy crop was found to be highly significant, which shows that the area under the paddy crop played an important role in variation in paddy production The dummy variables for Dharmapuri, Namakkal, Ramanathapuram, Salem, Thanjavur, Nilgiris, Tiruvarur, Thoothukudi, Vellore and Virudhunagar were found to be very highly significant, suggesting that perhaps these district variations were heterogeneous; therefore, the pooled regression model values might not be informative. Additionally, the values of the slope coefficients in Table 6 were also different, again casting some doubt in the results in Table 5. Additionally, if the Durbin-Watson d value was nearer to 2, there was no autocorrelation in the FE model. It seems that the FE model have found better than the pooled regression model.

Table 8: Results of FE or LSDV regression model.

Coefficient	Estimates	Std. Error	t-Statistic	Prob.
C(1)	-2551.651	14491.43	-0.176080	0.8603
C(2)	4.210532	0.175582	23.98037	0.0000
C(3)	-100901.8	28619.73	-3.525601	0.0005
C(4)	-9572.692	20810.18	-0.460000	0.6457
C(5)	2669.353	20535.71	0.129986	0.8966
C(6)	7754.034	21133.24	0.366912	0.7138
C(7)	-50950.75	27013.65	-1.886111	0.0598
C(8)	3157.088	20592.77	0.153311	0.8782
C(9)	-5373.961	20505.60	-0.262073	0.7934
C(10)	-15478.26	22077.33	-0.701093	0.4835
C(11)	-268884.0	33916.88	-7.927734	0.0000
C(12)	2892.024	20491.41	0.141133	0.8878
C(13)	-23505.64	20733.08	-1.133726	0.2574
C(14)	-132577.7	24969.45	-5.309598	0.0000
C(15)	-337426.4	29379.15	-11.48524	0.0000
C(16)	233.5037	20772.47	0.011241	0.9910
C(17)	-171448.1	23963.78	-7.154468	0.0000
C(18)	-155675.6	35280.66	-4.412492	0.0000
C(19)	1858.577	20466.31	0.090812	0.9277
C(20)	6078.532	20516.62	0.296274	0.7671
C(21)	-52427.23	25534.53	-2.053189	0.0405
C(22)	-238593.4	35383.43	-6.743083	0.0000
C(23)	3044.235	20538.03	0.148224	0.8822

C(24)	-14056.62	22519.83	-0.624189	0.5327
C(25)	6823.118	24008.13	0.284200	0.7764
C(26)	-76236.87	27268.82	-2.795753	0.0053
C(27)	-14992.80	21604.96	-0.693951	0.4880
C(28)	-71681.04	31532.97	-2.273210	0.0234
C(29)	-22066.78	20785.01	-1.061668	0.2888
Root MSE	66217.38	R-squared		0.872515
Mean dependent var	210037.1	Adjusted R-squared		0.866434
S.D. dependent var	185607.2	S.E. of regression		67833.36
Akaike info criterion	25.13343	Sum squared resid		2.70E+12
Schwarz criterion	25.34167	Log likelihood		-7712.096
Hannan-Quinn criter.	25.21440	F-statistic		143.4805
Durbin-Watson stat	1.584903	Prob(F-statistic)		0.000000

3.6. Wald test

To determine whether the FE model or pooled OLS regression model is more suitable, we adopt the Wald test. Here, the null hypothesis is that the pooled OLS regression model is appropriate (all dummy variables equal zero), and the alternative hypothesis is that the FE model is appropriate (all dummy variables do not equal zero). Accordingly, the Wald test is carried out and presented in Table 9.

Table 9: Characteristics of Wald test

Test Statistic	Value	df	Probability
F-statistic	18.46420	(26, 587)	0.0000
Chi-square	480.0693	26	0.0000

The Wald test F-statistic is found to be highly significant ($p<0.0000$), indicating that the FE or LSDV regression model is more appropriate than the panel pooled regression model. Not all dummy variables are not equal to zero.

3.6. RE model

The RE model is employed, by keeping the paddy crop production as the dependent variable and the area under the paddy crop as the independent variable, and the test results are presented in Table 10.

Table 10: Characteristics of Fitted RE model

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-27256.39	14097.69	-1.933394	0.0536
AREA	3.642751	0.134417	27.10045	0.0000
Effects Specification				
			S.D.	**Rho**
Cross-section random			56647.82	0.4109
Idiosyncratic random			67833.36	0.5891
Weighted Statistics				

Root MSE	69048.21	R-squared	0.535030
Mean dependent var	51955.77	Adjusted R-squared	0.534273
S.D. dependent var	101342.8	S.E. of regression	69160.57
Sum squared resid	2.94E+12	F-statistic	706.5166
Durbin-Watson stat	1.507303	Prob(F-statistic)	0.000000
Unweighted Statistics			
R-squared	0.725413	Mean dependent var	210037.1
Sum squared resid	5.82E+12	Durbin-Watson stat	0.760925

The results presented in Table 10 reveal that the RE model explains only 54% of the variation in the Paddy crop production in relation to the area under the Paddy crop. The rho value is 0.4109, which indicates that the individual effects of the cross-sections are 0.4%.

3.7. Hausman test

The results presented in Table 11 reveal that the Hausman test statistic is significant and that the null hypothesis is rejected, indicating that the FE model is an appropriate model. A remarkably high R^2 value of 80% is noted in the Hausman test. This finding supports the rejection of the null hypothesis that the random effect model is appropriate. Additionally, in the last row of Table 9, the FE and RE coefficient values of the regressor variable are found to be highly statistically significant.

Table 11: Hausman test results (Test cross-section random effects).

Test Summary		**Chi-Sq.tatistic**	**Chi-Sq. d.f.**	**Prob.**
Cross-section random		25.261800	1	0.0000
Cross-section random effects test comparisons:				
Variable	**Fixed**	**Random**	**Var(Diff.)**	**Prob.**
AREA	4.210532	3.642751	0.012761	0.0000
Variable	**Coefficient**	**Std. Error**	**t-Statistic**	**Prob.**
C	-64242.34	11759.67	-5.462936	0.0000
AREA	4.210532	0.175582	23.98037	0.0000
Effects Specification				
Root MSE	66217.38	R-squared		0.872515
Mean dependent var	210037.1	Adjusted R-squared		0.866434
S.D. dependent var	185607.2	S.E. of regression		67833.36
Akaike info criterion	25.13343	Sum squared resid		2.70E+12
Schwarz criterion	25.34167	Log likelihood		-7712.096
Hannan-Quinn criter.	25.21440	F-statistic		143.4805
Durbin-Watson stat	1.584903	Prob(F-statistic)		0.000000

4. CONCLUSION

The pooled regression model was not found suitable to study the relationship between the area under the paddy crop and it's production. The panel regression fixed effect model was emerged as an appropriate model and it explained 87 % variations in paddy crop production.

REFERENCES

[1] *Baltagi, B.H., "Econometric Analysis of Panel Data", 6th Edition, Wiley, New York, NY, https://link.springer.com/book/10.1007%2F978-3-030-53953-5,* ***(2021)***.

[2] *Bhamik, S.K., "Principles of Econometrics", 2nd edition, Oxford University Press,* ***(2017)***.

[3] *Breusch, T., and A. Pagan, "the Lagrange Multiplier Test and its Application to Model Specification in Econometrics", Review of Economic Studies, 47,* ***(1980)****, pp.239–253.*

[4] *Greene, W.H., "Econometric Analysis", 2008, Pearson Prentice Hall, Upper Saddle River, NJ,* ***(2008)***.

[5] *Gujarati, D.N., D.C.Porter and G.Sangeetha, "Basic Econometrics", 5th edition. McGraw Hill Education, New York, NY,* ***(2017)***.

[6] *Hadri, K., "Testing for Units Roots in Heterogeneous Panel Data". Econometrics Journal, 3,* ***(2000)****, pp.148-161.*

[7] *Hausman, J. A., "Specification Tests in Econometrics", Econometrica, 46,* ***(1978)****, pp.1251–1272.*

[8] *Hsiao, C., "Analysis of Panel Data. Cambridge University Press", Cambridge,* ***(2003)***.

[9] *Levin, A., C.F. Lin, and C.S.J.Chu, "Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties", Journal of Econometrics, 108,* ***(2002)****, pp.1-24.*

[10] *Wald,A, "Test of statistical hypotheses concerning several parameters when the number of observations is large", Transactions of the American Mathematical Society, 54,* ***(1943)****, pp.426-482, http://www.jstor.org/stable/1990256.*

[11] *Zaefizadah M., Khayatnezhad M., and Ghlaomin M., "Comparison of Multiple Linear Regression (MLR) and Artificial Neural Network (ANN) in Predicting the Yield Using its Components in the Hulless Barley", J. Agriculture and Environmental Sciences, 10(1),* ***(2011)****, pp.60-64.*