The Pharma Innovation Journal 2022; SP-11(2): 273-281

www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277- 7695
ISSN (P): 2349-8242
NAAS Rating: 5.23
TPI 2022; SP-11(2): 273-281

www.thepharmajournal.com
Received: 13-12-2021
Accepted: 18-01-2022

**N Uttam Singh**
ICAR Research Complex for NEH Region, Umiam, Meghalaya, India

**A Roy**
ICAR Research Complex for NEH Region, Umiam, Meghalaya, India

**Kishore K Das**
Department of Statistics, Gauhati University, Guwahati, Assam, India

**Pampi Paul**
ICAR Research Complex for NEH Region, Umiam, Meghalaya, India

**C Gowda HR**
ICAR Research Complex for NEH Region, Umiam, Meghalaya, India

**Kamni PB**
ICAR Research Complex for NEH Region, Umiam, Meghalaya, India

**Corresponding Author**
**N Uttam Singh**
ICAR Research Complex for NEH Region, Umiam, Meghalaya, India

# Panel data approach on cropping intensity under some land use dimensions and population density among the States of India

**N Uttam Singh, A Roy, Kishore K Das, Pampi Paul, C Gowda HR and Kamni PB**

**Abstract**
This paper investigated the pattern and behavior of variations of cropping intensity across the 28 states of India over the year from 2005 to 2015 using panel data models. The cross sectional as well as temporal or dynamic changes in gross irrigated area, forest area, fallow land and population density, and their panel effect on cropping intensity were analyzed by construction of various panel models using panel estimators including Pooled OLS, between effect, fixed effects, Random effects in a dynamic mode of estimation in SAS/ETS® software. The variables were chosen in view of their strong linkage with the agricultural production dynamics. The models were compared to conclude the most appropriate one through proper diagnostics and model identification tools including F-test for no fixed effect, Breusch-Pagan Lagrange Multiplier test for random effect, Hausman test for either fix or random effect, Variance Influence Factor for multicollinearity, and Durbin Watson for autocorrelation. Variation between states were much greater than their within variations for all the variables. The most appropriate model was found to be one way fixed effect model. Existence of unobserved heterogeneities was also detected during the process. This unobserved effect might include economic forces, population pressure, livestock pressure, natural factors, demand for industrialization, urbanization, housing and infrastructure and else others. Gross irrigated area had positive and significant effect. Forest area was also found positively associated with cropping intensity. The inverse relationship between fallow land and cropping intensity is highly disturbing factor as present data shows that fallow land increases over the years though the rate is slow.

**Keywords:** Cropping intensity, panel data, dynamic model, within and between variations, fixed and random effects

## Introduction

As per the land use statistics 2012-13, the total geographical area of the country is 328.7 million hectares, of which 139.9 million hectares is the net sown area. The gross cropped area is 194.4 million hectares with a cropping intensity of 138.9%. Land not available for cultivation has witnessed a continuous increase over time while land available for cultivation has declined. For example, net area sown has declined from about 142 million hectares in Triennium Ending (TE) 1983-84 to 140.82 million hectares in TE 2012-13, a reduction of about 1.2 million hectares. Land not available for cultivation has increased from about 40 million hectares in TE 1983-84 to 43.61 million hectares in TE 2012-13. Similarly, area under fallow land has increased from 23.26 million hectares to 25.35 million hectares during the same period (Vijay Paul Sharma, 2011) [14]. In view of declining availability of land for agriculture, increasing cropping intensity is the only answer to the problem of land constraint. The net sown area has remained around 141 million hectares during the last 40 years. Cropping intensity however has increased from 124.17 percent in TE 1983-84 to 139.1 percent in TE 2012-13. The increase in cropping intensity has been primarily driven by improved irrigation facilities (Vijay Paul Sharma, 2011) [14]. Irrigation remains the most dominant component in the overall investment in agriculture. Without proper use of water, it is difficult to get good returns on better high yielding seeds and higher doses of fertilizers. Water will remain a critical input for agriculture in the decades to come until science develops seeds that can thrive in dry climate with very little water. Another factor, worth discussion, is forest area. Out of the direct causes of deforestation, expansion of farming land for settlement of agriculture is one. Mostly all reports indicate shifting agriculture as responsible for about one half of tropical deforestation and some put it up to two-thirds.

Indeed, it is feared that agricultural expansion which is the main cause of deforestation in the tropics might replace forestry in the remaining natural forests (Anon., 2002; Cossalter and Pye-Smith, 2003; Anon., 2005) [1, 6, 2]. Again owing to a burgeoning population, it is estimated that per capita total land availability which was 0.32 ha in 2001 against the world average of 2.19 ha will decrease to 0.23 ha in 2025 and 0.19 ha in 2050. (State of Indian Agriculture 2011-12) [12].

Panel data, sometimes called longitudinal data, is the joining of cross-sectional and time series data where panel of individual states of India are observed over a period of year during which 11 observations per individual state are obtained. Panel data enabled us to control for factors that cannot be considered by simple cross-sectional regression models that ignored the time dimension. These factors, which were unobserved by us, make biased regression coefficients if they were ignored. The biggest advantage of the panel structure is that the multiple observations afford us more flexibility in how we can model the cropping intensity of these states. The aim of the present investigation was to determine the association between Cropping Intensity (*CI)* and each of the four variables Gross Irrigated Area (*GIA), Forest Area (FR), Fallow Land (FL)* and Population Density *(PD)* while controlling for other three at one time. Therefore, throughout this study we considered the panel regression model for state I during year t as follows:

$$CI_{it} = \beta_0 + \beta_1 GIA_{it} + \beta_2 FR_{it} + \beta_3 FL_{it} + \beta_4 PD_{it} + v_{it} + \varepsilon_{it}$$

The $v_{it}$ are individual state specific effects, and the $e_{it}$ are the observation-level regression errors. There are several ways to fit the preceding regression model, and each strategy differs in what it is willing to assume about the explanatory variables, the individual effects, the observation-level errors, and their relationships. A priory, *GIA* is positively related with *CI* and *FR, FL* and *PD* have inverse relationship with *CI* and there is unobserved cross-sectional heterogeneity.

## Data

Secondary data of land used statistics from ministry of agriculture; GOI from the period 2005 to 2015 was used as panel data set. Population statistics was retrieved from census of India 2011. Non-census years' population was generated with the help of decadal growth rate of corresponding states as per 2011census. The variables State and Year were identification variables that represented the state and year respectively. The dependent variable *CI* records percentage of total cropped area over net sawn area. The explanatory variables are *GIA, FR, FL* and *PD. PD* is the ratio of population to the total geographical area. Data for each state on the preceding four variables were available without any missing observation. Thus, there were twenty-eight cross-sectional units and 11 time periods. In all, therefore, we had 308 observations. As always with panel data, it was vital that we kept track of which variables in our data varied within state (the Time variant variables) and which were constant (the Time invariant variables). In the data used for this study, all the variables were time variant.

## Methodology

The panel data was analyzed following the procedure adopted by various authors, Katchova, Ani L. (2013) [9], Thomas Plümper *et al*. (2005) [13], SAS Institute Inc. 2010 [11]. SAS/ETS® 9.22 User's Guide, etc. and the steps followed are presented below.

### Variation types

In panel data analysis we find three variation types *viz* overall, between and within variations. These variations are inspected to have an idea of variation across the individuals over time (Gelman Andrew, 2005) [7]. Let *N* = number of individuals and *T* = number of time periods. Then,

Overall variance $s_O^2 = \frac{1}{NT-1}\sum_i\sum_t(x_{it} - \bar{x})^2$

Between variance $s_B^2 = \frac{1}{N-1}\sum_i(\bar{x}_i - \bar{x})^2$

Within variance $s_W^2 = \frac{1}{NT-N}\sum_i\sum_t(x_{it} - \bar{x}_i)^2 = \frac{1}{NT-1}\sum_i\sum_t(x_{it} - \bar{x}_i + \bar{x})^2$

To test hypothesis that population means from the individuals are equal (and any differences are due to natural variation), F test is used where $F = \frac{s_B^2}{s_W^2}$. F is distributed with *N-1* and *N(T-1)* degrees of freedom.

### Pooled model

Assumedthattheinterceptandslopecoefficientsareconstantacrostimeandspaceand the error term captures differences over time and individuals, the usual assumptions for cross-sectional analysis, and then we have the general regression model as $y_{it} = \alpha + x'_{it}\beta + e_{it}$. The symbols have the usual meanings.

### Within or fixed effects model

We assume that there is unobserved heterogeneity across individuals captured by $\alpha_i$. Fixed effect (FE) model assumes the individual-specific effects are correlated with the regressors. Each individual has a different individual-specific effects or intercept term $\alpha_i$ and the same slope parameters.

$$y_{it} = \alpha_i + x'_{it}\beta + e_{it}$$

In other words, the individual-specific effects are the leftover variation in the dependent variable that cannot be explained by the regressors. Time dummies can also be included in the regressors*x*. Then we have got one-way and two-way FE models.

### Random effects model

The random effect (RE) model assumes that the individual-specific effects are distributed independently of the regressors. All coefficients (the intercept as well as slope coefficients) vary over individuals. We include $\alpha_i$in the error term. Each individual has the same slope parameters and a composite error term $\varepsilon_{it} = \alpha i + eit$.

$$y_{it} = x'_{it}\beta + (\alpha_i + e_{it})$$

### Between estimator

The between estimator only uses the between variation (across individuals). It uses the time averages of all variables. This is an OLS estimation of the time-averaged dependent variable on the time-averaged regressors for each individual.

$$\bar{y}_i = \alpha + \bar{x}'_{it}\beta + (\alpha_i - \alpha + \bar{e}_{it})$$

The number of observations is *N*. The time variation is not considered and the data are collapsed with one observation per individual.

**Dynamic panel estimator**

Arellano *et al*., 1991 [3] and Baltagi *et al*., 1991 [4] have shown that dynamic panel models are linear regression models that are generalized in two ways. First, they include individual effects, yielding a two-tiered error structure: individual-level errors and overall residual errors. Second, they allow the dependent variable to depend on its value from the previous time period, thus making the model dynamic. The first generalization adds individual (cross-sectional) effects to linear regression. Formally, for individual $i=1....N$ at time $t=1...T_i$

$$y_{it} = \beta_0 + \beta_x X_{it} + v_i + \epsilon_{it}$$

The model contains a set of explanatory variables $X_{it}$. The $v_i$ are individual effects, and the $\varepsilon_{it}$are observation-level regression errors.

The second generalization of linear regression results in an autoregressive (AR) model. Ignoring individual effects for now, consider a regression model that includes a "lagged" version of the dependent variable:

$$y_{it} = \beta_0 + \phi y_{i,t-1} + \beta_x X_{it} + \epsilon_{it}$$

The model is dynamic because the equation for time t includes an element from the previous time period, the lagged response $y_{i,t-1}$. In the dynamic AR model, the dependent variable depends on its value from the previous time period in a way that is not explained by the regressors $X_{it}$. One way to interpret $\phi$ is that when $|\phi|<1$, $\phi$ is the correlation between $y_{it}$ and $y_{i,t-1}$.

We obtain a dynamic panel model by adding cross-sectional effects ($v_i$) to the AR model:

$$y_{it} = \beta_0 + \phi y_{i,t-1} + \beta_x X_{it} + v_i + \epsilon_{it}$$

**Result and Discussion**

**Overall, between and within variations:** Before proceeding to model construction, let's have a look at the summary of the panel data used. Here, the interest is on variation. Overall variation, between variation (variation across individual states) and within variation (variation over time) along with their mean, maximum and minimum values of all the variables (*CI*, *GIA*, *FR*, *FL* and *PD*) are given in table 1. The dependent variable cropping intensity had more between variations (80.49) than within variation (8.25). Overall variation of *CI* across the individual states over the eleven years was quite less than that of the between variation as the within variation was comparatively very low. This indicated the existence of state specific identifiable factors that causes the variation in a large scale. Though the rate of growth of cropping intensity was slow, there was positive growth over the study period (GOI, 2015). The average cropping intensity of an individual state was between 100 (Manipur and Mizoram) to 189.27 (Punjab) across the states but varied by 8.25 for each individual state over time. Smaller within variation indicated that for every state, on an average the variable *CI* did not vary much during this period from 2005 to 2015. The year 2005 had the lowest (135.72) and the year 2015 had the highest (141.39) cropping intensity when we looked at time wise. As expected, rest of the explanatory variables, *GIA*, *FR*, *FL* and *PD* had more between variations than within variation as shown in table 1. This means the variables varied very slowly over the 11 years with a large variability among the states. If the between variance were smaller than the within variance, then the means would be really close to each other and we would fail to reject the claim that "they are all equal". Another prerequisite task was diagnostic for multicollinearity among the explanatory variable. Multicollinearity were diagnosed with the help of Variance Influence Factor (VIF) presented in the last column of table 1 and the result shows VIF values are less than 2.5 (a rule of thumb) and we could assume no correlation between predictors or no multicollinearity. VIF of an explanatory variable is given by $1/(1-R_i^2)$ when the $j^{th}$ explanatory variable is regressed upon other *j-1* explanatory variables. VIF measure how much the variance of the estimated regression coefficients are inflated as compared to when the predictor variables are not linearly related (Kutner *et al*., 2004) [10].

**Table 1:** Overall, between and within variations for the variables, cropping intensity, gross irrigated area, forest area, fallow land and population density used in the panel.

| Variable | Variation | Mean | Std. deviation | Minimum | Maximum | VIF |
|---|---|---|---|---|---|---|
| CI | Overall | 138.09 | 25.14 | 100.00 | 196.24 | |
| | Between | | 80.49 | 131.75 | 146.21 | |
| | Within | | 8.25 | 100.00 | 189.99 | |
| GIA | Overall | 3158.82 | 4245.21 | 9.59 | 20964.77 | 1.60 |
| | Between | | 14232.44 | 2786.32 | 3570.56 | |
| | Within | | 476.26 | 12.36 | 19665.49 | |
| FR | Overall | 2507.98 | 2269.26 | 37.623 | 8702.55 | 1.39 |
| | Between | | 7617.91 | 2356.03 | 2561.67 | |
| | Within | | 223.89 | 40.08 | 8693.76 | |
| FL | Overall | 895.3801 | 1179.61 | 0.2 | 4987.22 | 1.42 |
| | Between | | 3907.55 | 710.90 | 1067.35 | |
| | Within | | 230.86 | 0.20 | 4125.65 | |
| PD | Overall | 357.99 | 285.83 | 14.48 | 1217.94 | 1.54 |
| | Between | | 960.54 | 331.83 | 386.19 | |
| | Within | | 24.65 | 16.26 | 1088.29 | |

Next prerequisite was test for autocorrelation as we had been dealing with time series. The Durbin-Watson (DW) test statistic was used to tests the null hypothesis that "the residuals from an ordinary least-squares regression are not autocorrelated against the alternative that the residuals follow an autoregressive of order one i.e. AR (1) process". The DW

statistic ranges in value from 0 to 4. A value near 2 indicates non-autocorrelation; a value toward 0 indicates positive autocorrelation; a value toward 4 indicates negative autocorrelation (Bhargava *et al*., 1982) [5]. Hence in this case 1st order autocorrelation was very high (0.89) and DW statistic was 0.218 which indicated the existence of autocorrelation of response variable *CI* at least at first order as shown in table 2. Therefore, we introduced a lagged *CI* i.e. *CI_l* as an additional regressor in the regression model and resulted Durbin-Watson statistic (1.91) showed very weak autocorrelation (0.045) of the residuals as presented in the same table 2.

**Table 2:** Durbin-Watson (DW) test statistic for static and dynamic panels.

| Panel Type | Regression | DW | AR(1) |
|---|---|---|---|
| Static | CI = GIA + FR + FL + PD | 0.218 | 0.890 |
| Dynamic | CI = CI_l + GIA + FR + FL + PD | 1.910 | 0.045 |

This situation invited requirement of dynamic panel models (using lagged dependent variable) rather than static panel models. Dynamic panel estimators allow the dependent variable to depend on its own value from the previous time period, thus making the model dynamic by using a lagged dependent variable as a regressor, whereby capturing the autoregressive nature of the underlying process. We wanted to determine the association between *CI* and the independent variables *GIA*, *FR*, *FL* and *PD*. We suspected that *CI* differed by state in ways other than by *GIA*, *FR*, *FL* and *PD* and, because the data are limited i.e., we suspected that there must be unmeasured factors that might influence both present and future *CI*. Therefore, we posited a dynamic panel model that included state effects and a lagged dependent variable $CI_{i,t-1}$for state i in year *t*. The first order of business for estimation of dynamic panel models is to produce a version of our data that contains first-order lags of *CI*.

$$CI_{it} = \beta_0 + \phi CI_{i,t-1} + \beta_1 GIA_{it} + \beta_2 FR_{it} + \beta_3 FL_{it} + \beta_4 PD_{it} + v_{it} + \varepsilon_{it}$$

## Dynamic panel models

When we included a lagged dependent variable as one of the regressors, the panel model became dynamic. Here the lagged dependent variable was *CI_l* which was the 1 lagged series of the original response series *CI*. We first fitted the dynamic panel model to the data by OLS and it just takes data and applied OLS on it like traditional regression ignoring the fact that it was panel data. That is, 11 observations for each state were stacking the one on top of the other, thus giving in all 308 observations for each of the variables in the model. The result of pooled OLS is shown in tables 3 and 4. Pooled estimation did not consider state affiliation, the estimates did not account for the state effects. Ignoring that deficiency for the time being, we interpreted the coefficient on *CI_l* (0. 0.966) as the correlation between *CI* for the current year and *CI* for the previous year, after controlling for other regressors *GIA*, *FR*, *FL* and *PD*. The autocorrelation was statistically significant, indicating that we had unmeasured factors affecting *CI* and that the effects of these factors linger for several years. Other coefficients were probably inaccurate because we omitted state effects - effects that could be shown to be correlated with lagged *CI* from the analysis. Besides, Pooled OLS was reasonable when it satisfied the assumption that the intercept values of the states were the same i.e., state specific effects were same. Pooling was admissible if there were no fixed effects or random effects present in the data.It also assumed that the slope coefficients of *GIA*, *FR*, *FL* and *PD* were all identical for all the twenty-eight states. Obviously, these were very restricted assumptions. There fore, despite its simplicity the pooled regression might distort the true picture of the relationship between *CI* and the four explanatory variables across the twenty-eight states. To detect the state specific and/or time specific effect we used fixed effect estimator letting the intercept varied for each state but still assumed that the slope coefficients were constant across states. The term "fixed effects" was due to the fact that, although the intercept might differ across individual states, each individual's intercept did not vary over time; i.e., time invariant. Because the effects were fixed, we could treat them as regression coefficients in a standard regression model, to be estimated along with the coefficients of *GIA*, *FR*, *FL* and *PD*. This was done by the dummy variable technique. The difference in the intercept might bedue the managerial style of managerial philosophy.

**Table 3:** Comparing model fit statistic of estimators of dynamic panel models.

| Model | F/Breusch Pagan | $R^2$ | Hausman | Var_CS | Var_TS | Var_ERR/MSE |
|---|---|---|---|---|---|---|
| Pooled | | 0.95 | | | | 34.34 |
| Fix2 | 2.30*** | 0.95 | | | | 34.31 |
| Fix1 | 2.82*** | 0.95 | | | | 34.20 |
| Fix1TM | 1.74* | 0.94 | | | | 38.81 |
| Ran2FB | 4.90** | 0.93 | 24.73*** | 1.42 | 0.00 | 32.61 |
| Ran2WK | 4.90** | 0.77 | 16.22*** | 20.20 | 0.00 | 32.61 |
| Ran2WH | 4.90** | 0.95 | 25.82*** | 0.00 | 0.54 | 34.91 |
| Ran2NL | 4.90** | 0.60 | 6.47 | 97.51 | 0.99 | 28.06 |
| Ran1FB | 4.20** | 0.93 | 24.73*** | 1.42 | 0.00 | 32.61 |
| Ran1WK | 4.20** | 0.77 | 16.22*** | 20.20 | 0.00 | 32.61 |
| Ran1WH | 4.20** | 0.95 | 25.82*** | 0.00 | 0.54 | 34.91 |
| Ran1NL | 4.90** | 0.60 | 6.47 | 97.51 | 0.99 | 28.06 |
| Btwgrps | | 0.99 | | | | 1.56 |
| Btwtime | | 0.95 | | | | 0.42 |

***, ** and * refer to significant at 1%, 5% and 10% respectively.

**Table 4:** Comparing parameters estimated by estimators of dynamic panel models.

| Model | Intercept | CI_l | GIA | FR | FL | PD |
|---|---|---|---|---|---|---|
| Pooled | 5.043<br>(2.287)** | 0.966<br>(0.016)*** | 0.000<br>(0.000) | 0.000<br>(0.000) | -0.001<br>(0.000) | 0.000<br>(0.001) |
| Fix2 | 26.233<br>(23.782) | 0.622<br>(0.047)*** | 0.003<br>(0.001)*** | 0.002<br>(0.002) | -0.005<br>(0.002)** | 0.014<br>(0.022) |
| Fix1 | 20.859<br>(16.498) | 0.603<br>(0.046)*** | 0.003<br>(0.001)*** | 0.002<br>(0.002) | -0.006<br>(0.002)*** | 0.023<br>(0.016) |
| Fix1TM | 4.065<br>(2.715) | 0.954<br>(0.017)*** | 0.000<br>(0.000) | 0.000<br>(0.000) | -0.001<br>(0.000) | 0.000<br>(0.002) |
| Ran2FB | 6.266<br>(2.706)** | 0.957<br>(0.019)*** | 0.000<br>(0.000) | 0.000<br>(0.000) | -0.001<br>(0.000) | 0.001<br>(0.002) |
| Ran2WK | 15.801<br>(4.954)*** | 0.883<br>(0.034)*** | 0.001<br>(0.000)** | 0.000<br>(0.000) | -0.002<br>(0.001)** | 0.002<br>(0.004) |
| Ran2WH | 4.988<br>(2.282)** | 0.967<br>(0.016)*** | 0.000<br>(0.000) | 0.000<br>(0.000) | -0.001<br>(0.000) | 0.000<br>(0.001) |
| Ran2NL | 24.650<br>(7.073)*** | 0.801<br>(0.044)*** | 0.001<br>(0.000)** | 0.001<br>(0.001) | -0.004<br>(0.001)*** | 0.004<br>(0.008) |
| Ran1FB | 6.266<br>(2.706)** | 0.957<br>(0.019)*** | 0.000<br>(0.000) | 0.000<br>(0.000) | -0.001<br>(0.000) | 0.001<br>(0.002) |
| Ran1WK | 15.801<br>(4.954)*** | 0.883<br>(0.034)*** | 0.001<br>(0.000)** | 0.000<br>(0.000) | -0.002<br>(0.001)** | 0.002<br>(0.004) |
| Ran1WH | 4.988<br>(2.282)** | 0.967<br>(0.016)*** | 0.000<br>(0.000) | 0.000<br>(0.000) | -0.001<br>(0.000) | 0.000<br>(0.001) |
| Ran1NL | 24.650<br>(7.073) | 0.801<br>(0.044) | 0.001<br>(0.000) | 0.001<br>(0.001) | -0.004<br>(0.001) | 0.004<br>(0.008) |
| Btwgrps | 2.078<br>(1.693) | 0.988<br>(0.012)*** | 0.000<br>(0.000) | 0.000<br>(0.000) | 0.000<br>(0.000) | 0.000<br>(0.001) |
| Btwtime | 173.404<br>(58.569) | -0.218<br>(0.300) | 0.006<br>(0.006) | -0.015<br>(0.014) | -0.005<br>(0.011) | 0.056<br>(0.038) |

Figures under parenthesis are the standard error.
***, ** and * refer to significant at 1%, 5% and 10% respectively.

Two-way fixed effect estimator (designated as Fix2 in tables 3 and 4) concerning both state effect and time effect shows a significant F test (F-value is 2.30) shown in table 3. The F-test for fixed effects was testing the hypothesis that there were no fixed effects and we easily rejected the null of poolability. So, there were state effects, or time effects, or both as we run two-way fixed effect estimator. OLS would not give reasonable results as we discussed earlier. We see a large R-square (0.95) in Fix2, so there is a reasonable fit. Further to investigated the true effect of individual states ignoring the time effect we ran one-way fixed effect estimator designated by Fix1 in tables 3 and 4. We see a good R-square value (0.95) little less than Fix2 and we still reject poolability based on the F-test (F-value is 2.82) in all accepted levels of significance while number of significant individual effects increased from 3 in two-way fixed effect to 13 in one-way fixed effect which shows existence of distinct state specific unobserved heterogeneity. Hence Fix1 is still a good fit with MSE=34.20 which is less than that of Fix2.

Just as we used the dummy variables to account for individual effect, we can allow for time effect. For sucha situation we introduce time dummies, one for each year. To check the existence of time effects separately while controlling state specific effect we ran another one-way time fixed effect estimator designated by Fix1TM in tables 3 and 4. As suggested by F-test (F-value is 1.74, p-value is 0.071) we may reject the null hypothesis at 10% level of significance that there is no time effect. It can be seen from the result that few of the time dummies are significant. The less probability of time effect must be due to the short time period (11 years) of the panel. Still we can consider the time effect as accountable. Existence of significant state specific and time fixed effects are clearly shown by Fix2, Fix1 and Fix1TM.

Now we investigate existence of random effect and its significant effect on cropping intensity using random effect estimators. We ran both two way and one-way random effect estimators in four different methods of variance components namely Fuller and Battese, Wansbeek and Kapteyn, Wallace and Hussain and Nerlove. These are designated by Ran2FB, Ran2WK, Ran2WH and Ran2NL, in two way and Ran1FB, Ran1WK, Ran1WH and Ran1NL in one way respectively in tables 3 and 4. Breusch-Pagan test was found significant in all the cases of random effect estimators (table 3) and it suggested that existence of significant random effects in both two way and one-way cases rejecting the null hypothesis that “there are no random effects”. The Breusch-Pagan test helps us decide between a random effect regression and a simple OLS regression. Now it is seen that both fix effect and random effect estimators are reasonable fit. To determine the model between fixed and random, Hausman Test was performed and it suggested fixed effect model in all the cases (except Nerlove) because we rejected null hypothesis either in one way or two-way effect (table 3). So, Hausman test showed significant differences between the coefficients for the fixed effects and random effects model. In other words, null hypothesis of the Hausman test is that individual state effect is uncorrelated with each explanatory variable, and thus the problem with the random-effects strategy is that state-level effects are correlated with one or more explanatory variables. Now the dilemma is selection between one-way and two-way fixed effect models. In such case we went subjectively because the interest was heavier on cross-section side rather than the time. We made an excuse to exclude time effect as the rejection line of one-way time effect model (Fix1TM) was at 10% level of significance as p- value was 0.071 which might be due to short time of only 11 years.

Moreover, the MSEs of Fix2 and Fix1TM are more than that of Fix1 (table 3) and none of the coefficients of the regressors are significant in Fix1TM. Hence, we stick to one-way analysis model where only states are considered ignoring time.

Further we go for between estimators, between individual states as well as between time periods and they are designated as Btwgrps and Btwtime respectively in tables 3 and 4. Btwgrps is an OLS estimation of the time-averaged dependent variable on the time-averaged regressors for each individual. The time variation is not considered and the data are collapsed with one observation per individual. Since we lose much information by collapsing the data into averages, we should not rely on the between estimator for anything other than to help diagnose correlated individual effects. By comparing the between estimates to the within estimates, we can determine where the bias in GLS (random effect model) occurs. Of course, we can detect bias in GLS by directly comparing it to the within estimator but using the between estimator makes the bias more obvious. To illustrate, the GLS estimate of the coefficient of *GIA*is 0.0002 (Ran1fb), the within estimate is 0.003 (Fix1), and the between estimate is 0.00004 (Btwgrps). The bias is much more evident when we compare 0.003 to 0.00004, knowing that under GLS these should estimate the same quantity. This again intensifies our doubt of using random estimators. If we confine ourselves to comparing only state averages, we obtain the between estimator. Btwtime is another between estimator when we take it time wise instead of state wise. The parameter estimates of various estimators are given in table 4. Out of the estimators listed in table 4, Fix1tm, Btwgrps and Btwtime may be kept aside as we found weak significant F-test in Fix1tm shown in table 4 and; Btwgrps and Btwtime could be used for diagnostic purpose rather than model construction as discussed above. Hence, our best subjective model will be one-way fixed effect model (Fix1) and is given by,

CI = 20.86 + 0.60 CI_l + 0.003 GIA + 0.002 FR - 0.006 FL + 0.023 PD + 14.99 Andhra Pradesh + 15.22 Arunachal Pradesh + 23.26 Assam + 6.46 Chhattisgarh + 19.01 Goa + 3.21 Gujarat +23.83 Haryana + 42.14 Himachal Pradesh + 32.96 Jammu Kashmir + 24.03 Jharkhand + 14.09 Karnataka + 8.62 Kerala – 4.43 Madhya Pradesh + 13.46 Maharashtra + 11.93 Manipur + 22.33 Meghalaya + 15.09 Mizoram + 25.47 Nagaland + 12.45 Odisha + 19.91 Punjab + 20.65 Rajasthan + 44.94 Sikkim + 14.45 Tamil Nadu + 20.62 Tripura + 28.50 Uttarakhand – 26.36 Uttar Pradesh + 11.85 West Bengal.

**Table 5:** Unobserved heterogeneity of each states in explaining cropping intensity.

| State | Estimates | State | Estimates | State | Estimates |
|---|---|---|---|---|---|
| Andhra Pradesh | 14.99 (13.81) | Jharkhand | 24.03** (10.35) | Odisha | 12.45 (15.59) |
| Arunachal Pradesh | 15.22 (20.54) | Karnataka | 14.09 (11.92) | Punjab | 19.91* (10.31) |
| Assam | 23.26** (10.89) | Kerala | 8.62 (6.00) | Rajasthan | 20.65 (16.61) |
| Chhattisgarh | 6.46 (17.90) | Madhya Pradesh | -4.43 (19.80) | Sikkim | 44.94*** (14.94) |
| Goa | 19.01* (10.30) | Maharashtra | 13.46 (12.73) | Tamil Nadu | 14.45* (8.72) |
| Gujarat | 3.21 (12.20) | Manipur | 11.93 (14.27) | Tripura | 20.62* (10.91) |
| Haryana | 23.83** (9.24) | Meghalaya | 22.33 (13.89) | Uttarakhand | 28.50* (14.80) |
| Himachal Pradesh | 42.14*** (14.39) | Mizoram | 15.09 (15.03) | Uttar Pradesh | -26.36 (16.05) |
| Jammu Kashmir | 32.96** (15.38) | Nagaland | 25.47* (14.07) | West Bengal | 11.85*** (3.72) |

Figures under parenthesis are the standard errors.
***, ** and * refer to significant at 1%, 5% and 10% respectively.

Individual state specific effects are listed in table 5. These are output from one-way fixed effect model. One state out of 28 states was omitted for reference. Bihar, whose average of cropping intensity (138.76) was the closest to the overall average (138.09) was kept for reference and compared with each state. State specific unobserved heterogeneities were captured to be 20.86 which was the intercept of Bihar. Result shows that out of the 28 states Madhya Pradesh and Uttar Pradesh had unobserved negative effects on cropping intensity. This means that these states' unobserved heterogeneity is less than that of Bihar. Rest of the states showed positive estimate and this means that their unobserved heterogeneity was positively associated with cropping intensity and more than that of Bihar. However, only 13 states Assam, Goa, Haryana, Himachal Pradesh, Jammu and Kashmir, Jharkhand, Nagaland, Punjab, Sikkim, Tamil Nadu, Tripura, Uttarakhand and West Bengal showed significantly positive effect. It is seen from the suggested model that *GIA* has positive and significant effect and one unit increase in *GIA* would increase 60% in cropping intensity while *FR*, *FL* and *PD* are controlled, indicating that irrigation has a big role in increasing cropping intensity. Interestingly *FR* has positive association with cropping intensity and increments in both the entities are desirable though the estimate is not significant. On the other hand, *FL* has significant inverse relationship with cropping intensity. This is a disturbing factor on cropping intensity and land use statistics of India shows that *FL* increases over the years though the rate is slow. Again, *PD* which is positively associated with cropping intensity but threatening part is how long this positive association will prolong as the rate of country's population increment is significant. Expected overall cropping intensity varies from 99.776 to 190.771 while that of actual varies from 100 to 196.244. The interstate variability of the unobserved factors is discussed in table 5. Sikkim had the highest unobserved heterogeneity which was 44.94 more than that of Bihar, followed by Himachal Pradesh (42.14) while Uttar Pradesh had the lowest unobserved heterogeneity which was 26.36 less than that of Bihar, which was followed by Madhya Pradesh (-4.43). The performances of 8 states of north eastern amongst the Indian states are notable. All the eight states' unobserved heterogeneity were more than the reference state, Bihar and positively associated with cropping intensity. Among these 8 states, Assam, Nagaland, Sikkim and Tripura witnessed significantly higher amount of omitted heterogeneity than Bihar. Sikkim, at the national level also, had the highest unobserved heterogeneity which was 44.94 more than that of Bihar, followed by Nagaland (25.47) while Manipur had the lowest unobserved heterogeneity which was just 11.93 more than that of Bihar, which was followed by Mizoram (15.09).

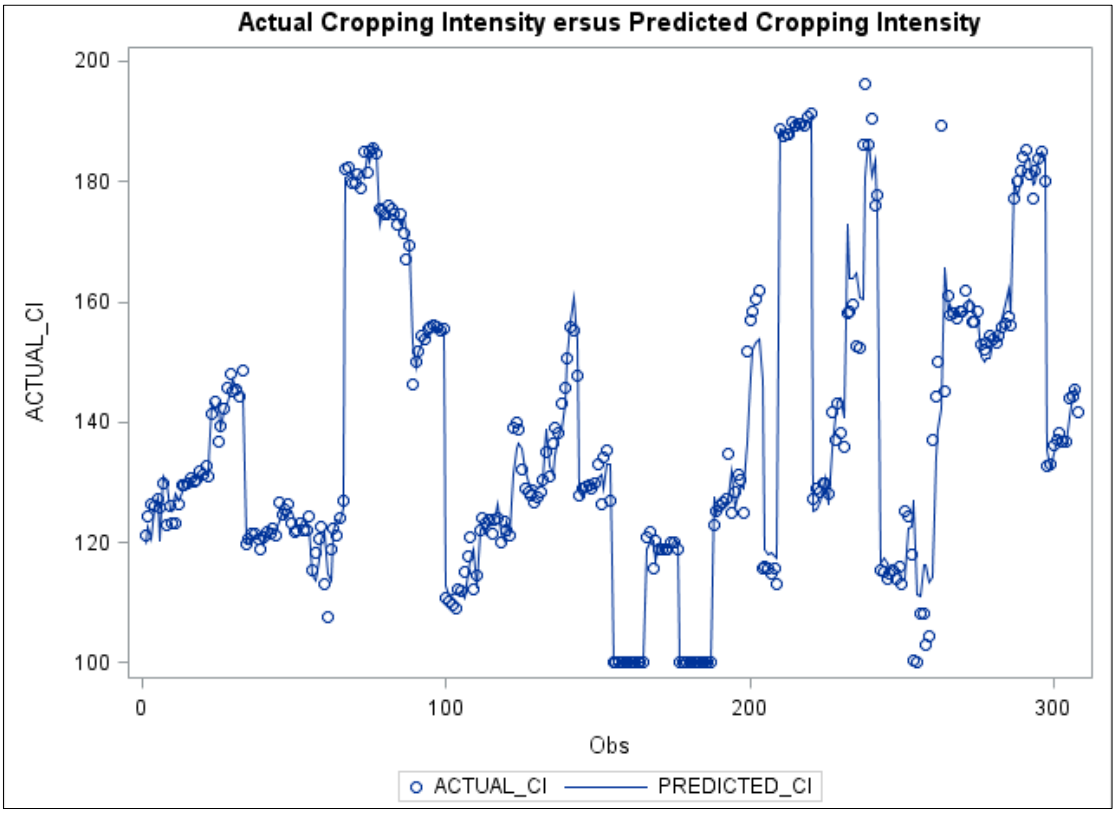


**Fig 1:** Plot of actual cropping intensity versus predicted cropping intensity given by one-way dynamic fixed effect model.

**Appendix 1:** Comparing model fit of various estimators of static (without lagged dependent variable) panel models.

| Model | F/ Breusch Pagan | $R^2$ | Hausman | Var_CS | Var_TS | Var_ERR |
|---|---|---|---|---|---|---|
| Pooled | | 0.29 | | | | 455.79 |
| FD2 | | 0.18 | | | | 47.41 |
| FD1 | | 0.17 | | | | 24.78 |
| FDT | | 0.38 | | | | 557.40 |
| Fix2 | 58.82*** | 0.92 | | | | 56.54 |
| Fix1 | 82.49*** | 0.92 | | | | 55.16 |
| Fix1TM | 0.09 | 0.29 | | | | 469.83 |
| Ran2FB | 1168.30*** | 0.18 | 7.66 | 477.20 | 0.00 | 56.54 |
| Ran2WK | 1168.30*** | 0.18 | 6.97 | 552.60 | 0.00 | 56.54 |
| Ran2WH | 1168.30*** | 0.18 | 7.78* | 475.63 | 0.00 | 57.90 |
| Ran2NL | 1168.30*** | 0.17 | 5.23 | 683.83 | 0.98 | 48.83 |
| Ran1FB | 1163.57*** | 0.20 | 1.09 | 4652.26 | | 55.17 |
| Ran1WK | 1163.57*** | 0.19 | 8.41* | 681.33 | | 55.17 |
| Ran1WH | 1163.57*** | 0.18 | 10.90** | 475.69 | | 57.30 |
| Ran1NL | 1163.57*** | 0.22 | 6.99 | 777.28 | | 49.44 |
| Parks | | 0.41 | | | | 0.21 |
| Dasilva | m=2 | 0.13 | | 456.164 | 0.00 | 401.80 |
| Btwgrps | | 0.30 | | | | 480.91 |
| Btwtime | | 0.94 | | | | 0.38 |

***, ** and * refer to significant at 1%, 5% and 10% respectively.

**Appendix 2:** Comparing parameters estimated by various estimators of static (without lagged dependent variable) panel models.

| Model | Intercept | GIA | FR | FL | PD |
|---|---|---|---|---|---|
| Pooled | 136.64<br>(2.77)*** | 0.003<br>(0.0004)*** | -0.001<br>(0.001) | -0.009<br>(0.001)*** | -0.009<br>(0.005)* |
| FD2 | | 0.008<br>(0.001)*** | -0.001<br>(0.002) | -0.005<br>(0.002)*** | -0.081<br>(0.059) |
| FD1 | | 0.007<br>(0.001)*** | -0.002<br>(0.002) | -0.004<br>(0.002)** | -0.019<br>(0.043) |
| FDT | | 0.002<br>(0.000)*** | -0.001<br>(0.001) | -0.011<br>(0.001)*** | 0.008<br>(0.005) |
| Fix2 | 104.792<br>(29.551)*** | 0.006<br>(0.001)*** | 0.003<br>(0.003) | -0.008<br>(0.003)*** | 0.012<br>(0.028) |
| Fix1 | 80.827<br>(20.125)*** | 0.006<br>(0.001)*** | 0.002<br>(0.003) | -0.008<br>(0.003)*** | 0.034<br>(0.020)* |

| | | | | | |
|---|---|---|---|---|---|
| Fix1TM | 137.125<br>(4.839) | 0.003<br>(0.000) | -0.001<br>(0.001) | -0.009<br>(0.001) | 0.009<br>(0.005) |
| Ran2FB | 122.934<br>(7.395)*** | 0.004<br>(0.001)*** | 0.001<br>(0.002) | -0.009<br>(0.002)*** | 0.019<br>(0.013) |
| Ran2WK | 121.943<br>(7.715)*** | 0.004<br>(0.001)*** | 0.002<br>(0.002) | -0.009<br>(0.002)*** | 0.020<br>(0.014) |
| Ran2WH | 123.117<br>(7.338)*** | 0.004<br>(0.001)*** | 0.001<br>(0.002) | -0.009<br>(0.002)*** | 0.019<br>(0.013) |
| Ran2NL | 121.209<br>(8.714)*** | 0.005<br>(0.001)*** | 0.002<br>(0.002) | -0.009<br>(0.002)*** | 0.018<br>(0.016) |
| Ran1FB | 112.106<br>(14.834)*** | 0.005<br>(0.001)*** | 0.002<br>(0.003) | -0.009<br>(0.003)*** | 0.031<br>(0.018)* |
| Ran1WK | 120.394<br>(8.249)*** | 0.004<br>(0.001)*** | 0.002<br>(0.002) | -0.009<br>(0.002)*** | 0.021<br>(0.015) |
| Ran1WH | 123.045<br>(7.360)*** | 0.004<br>(0.001)*** | 0.001<br>(0.002) | -0.009<br>(0.002)*** | 0.019<br>(0.013) |
| Ran1NL | 118.874<br>(8.831)*** | 0.005<br>(0.001)*** | 0.002<br>(0.002) | -0.009<br>(0.002)*** | 0.023<br>(0.015) |
| Parks | 132.292<br>(7.511)*** | 0.003<br>(0.001)*** | -0.000<br>(0.001) | -0.006<br>(0.002)*** | 0.004<br>(0.013) |
| Dasilva | 110.12<br>(4.182)*** | 0.006<br>(0.000)*** | -0.001<br>(0.000) | -0.004<br>(0.000)*** | 0.042<br>(0.001)*** |
| Btwgrps | 136.933<br>(9.476)*** | 0.003<br>(0.001)* | -0.001<br>(0.002) | -0.009<br>(0.004)** | 0.009<br>(0.018) |
| Btwtime | 133.232<br>(18.541)*** | 0.004<br>(0.005) | -0.008<br>(0.009) | -0.010<br>(0.008) | 0.059<br>(0.036) |

Figures under parenthesis are the standard error.
***, ** and * refer to significant at 1%, 5% and 10% respectively.

## Conclusion

As expected, variation between states were much greater than their within variations for all the variables. The identification of appropriate model was done on the basis of (i) F test, that suggested fixed effect rather than pooled OLS, (ii) Breusch-Pagan test, that suggested random effect rather than pooled OLS, (iii) Hausman test, that suggested fixed effect rather than random effect, (iv) between effect estimator, that suggested fixed effect rather than random effect and (v) time wise fixed effect estimator, that suggested one-way estimator rather than two-way as time effects are weakly significant. Ultimately the most appropriate model for the panel data used was identified to be "one way fixed (within) effect model". State specific unobserved heterogeneities, which could not be covered by the explanatory variable in the model, were detected and found significant effect. This unobserved effect might include economic forces, population pressure, livestock pressure, natural factors, demand for industrialization, urbanization, housing and infrastructure and else others. In the model, variable *GIA* had positive and significant effect indicating the requirement of irrigated area to be increased to improve cropping intensity. *FR* was found positively associated with cropping intensity. This is an interesting association as increments in both entities are desirable. The reason might be due to increase in water holding capacity of land as forest area increases and in turn a positive effect on cropping intensity. On the other hand, variable *FL* had negative effect. The inverse relationship between fallow land and cropping intensity is highly disturbing factor as present data shows that fallow land increases over the years though the rate is slow. Fallow land varies from 25.670 to 26.182 million hectares over the period from 2004 to 2013 at national level. Again, the country's population is a major concern for not only in agriculture but many other developmental programmes as the growth rate of population is faster than the rates of cropping intensity as well as growth of other agricultural commodities. General panel data analysis or in specific sense static panel data analysis of the study dataset are also estimated as presented in Appendix 1 and 2. These static versions of panel models were found performed less in terms of model fit statistic and parameter estimates as compared to the dynamic models discussed in the study.

## References


1. Anonymous. Forest certification and biodiversity: opposites or complements? Discussion paper prepared for the GEF, International Tropical Timber Organization, Yokohama Japan, 2002.
2. Anonymous. Ecosystems and Human well-being: synthesis. Millennium ecosystem Assessment. Island Press, Washington DC, 2005.
3. Arellano M, Bond S. Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations, Review of Economic Studies. 1991;58:277-297.
4. Baltagi BH, Li Q. A Joint Test for Serial Correlation and Random Individual Effects, Statistics and Probability Letters. 1991;11:277-280.
5. Bhargava Alok, Franzini L, Narendranathan W. Serial Correlation and the Fixed Effects Model, Review of Economic Studies. 1982;49(4):533-549.
6. Cossalter C, Pye-Smith C. Fast-wood forestry, myths and realities. CIFOR, Bogor Indonesia, 2003.
7. Gelman Andrew. Analysis of Variance - Why It Is More Important than Ever? The Annals of Statistics. 2005;33:1-53.
8. Government of India. Directorate of Economics and Statistics, Department of Agriculture, Cooperation and Farmers Welfare, Ministry of Agriculture and Farmers Welfare, 2015.
9. Katchova, Ani L. Panel Data Models Example [PDF document and YouTube Video]. Retrieved from Econometrics Academy, 2013. Website: https://sites.google.com/site/econometricsacademy/

10. Kutner MH, Nachtsheim CJ, Neter J. *Applied Linear Regression Models*, 4th Edition (McGraw-Hill Irwin, 2004.
11. SAS Institute Inc. SAS/ETS® 9.22 User's Guide. Cary, NC: SAS Institute Inc, 2010.
12. State of Indian Agriculture Government of India. Ministry of Agriculture. Department of Agriculture and Cooperation. Directorate of Economics and Statistics. New Delhi, 2012-13.
13. Thomas Plümper, Vera Troeger E, Philip Manow. Panel data analysis in comparative politics: Linking method to theory. European Journal of Political Research. 2005;44:327-354.
14. Vijay Paul Sharma. India's Agricultural Development under the New Economic Regime: Policy Perspective and Strategy for the 12th Five Year Plan. Indian Institute of Management, Ahmedabad. W.P. No. 2011-11-01, 2011.