WP-2014-044

Neighborhood and agricultural clusters across states of India

Tirtha Chatterjee and A. Ganesh-Kumar

Indira Gandhi Institute of Development Research, Mumbai
October 2014
http://www.igidr.ac.in/pdf/publication/WP-2014-044.pdf

Neighborhood and agricultural clusters across states of India

Tirtha Chatterjee and A. Ganesh-Kumar
Indira Gandhi Institute of Development Research (IGIDR)
General Arun Kumar Vaidya Marg
Goregaon (E), Mumbai- 400065, INDIA
Email(corresponding author): tirtha@igidr.ac.in

Abstract

In this study we trace how number and members of income clusters have changed in Indian agriculture over the last four and a half decades. Two features which stand out in our results are that not all geographical neighbors belong to the same cluster and clusters include both geographical neighbors and non-neighbors. To identify the factors driving a pair of states to common cluster we then use a logit model and find that smaller is the relative difference between them in terms of mechanization, infrastructural support, deviations from normal rainfall and price differences, higher are the chances that they will be in the same income cluster. Between contiguous and non-contiguous state pairs we find that apart from the common factors, smaller relative differences in irrigation support, rainfall and price differences additionally brings non-contiguous states together.

Keywords: agriculture, rural development, club convergence, income cluster determinants, geography

JEL Code: O13, O18, O47, Q18, R11, R12

Geographic neighborhood and agricultural cluster formation: Evidence from states of India

Tirtha Chatterjee[a] and A. Ganesh-Kumar[b]

ABSTRACT

We study an empirical-occurrence largely overlooked in studies on growth-clusters: (i) Not all geographical-neighbors are cluster-co-members; and (ii) Most clusters include geographical-neighbors and non-neighbors. Using agricultural income-per-capita across Indian-states, we identify the income-clusters over the last forty-five years and find evidence of a similar pattern. Logistic estimations show that cluster-membership is driven by the relative-difference between a state-pair in mechanization, infrastructural-support, cropping pattern, rainfall and prices. Further, the relative-difference between a state-pair in irrigation, rainfall and prices matter for the cluster-membership of non-neighbouring state pairs only. Our findings could have implications for policies aimed at reducing spatial differences across states.

KEY WORDS: agriculture, club convergence, income cluster determinants, geography
JELCODES: O47, Q18, R11, R12

a: Doctoral candidate,Indira Gandhi Institute of Development Research (IGIDR), General A. K. Vaidya Marg, Goregaon (E), Mumbai, India-400065, E-mail: tirtha@igidr.ac.in, Phone no: +91-8655848508
b: Professor, Indira Gandhi Institute of Development Research (IGIDR), General A. K. Vaidya Marg, Goregaon (E), Mumbai, India-400065, E-mail:agk@igidr.ac.in

1. Introduction

The impact of a region's geographical location on its economic performance has always been a source of debate in growth literature. Studies on convergence across regions both at cross-country levels (for example Gallup et al (1999)) and country specific levels (like Bloom et al (1998) on Africa, Corrado et al (2005) in case of Europe) find that geographical location of a region, for example being located in the tropics, being land-locked or far away from the coast or facing high transportation costs in accessing markets etc., significantly influences its economic growth.

On the other hand cross-country studies (like Rodrik et al (2004) and Easterly and Levine (2003)) which explore the relationship between geographical location and income levels do not find any evidence of a significant relationship between the two. For example Rodrik et al (2004) document the supremacy of quality of institutions over geography in explaining income levels around the world. The authors add that 'geography have at best weak direct effects on incomes, although they have a strong indirect effect by influencing the quality of institutions'. Easterly and Levine (2003) conclude that 'We find no evidence that tropics, germs, and crops affect country incomes directly other than through institutions, nor do we find any effect of policies on development once we control for institutions'.

On the same lines, there is divergence in the results of various studies which have used club-convergence tests to identify clubs/clusters of convergence and tried to explain the club/cluster constitution on the basis of geographical factors. Studies like Corrado et al (2005) and Bartkowska and Reidl (2012) found geographical location to be significant driver of cluster constitution while Maasoumi et al (2008) concludes that convergence groups that they identified in China could not be defined on the basis of region or extent of policy preference.

Hence, we can see that this debate on the significance of geographical location on growth remains un-settled. One plausible reason for this could be that all these studies consider the economy as a whole. It is conceivable that geographic location may matter only for some, but not all sectors of the economy. For instance, location may not matter for say the financial or educational services, while it could be a major driver of growth of agriculture.

Another possible reason could be that policies, technologies, infrastructural investments, and institutionsmight cancel out the influence of geographical location. Even in agriculture, although it is widely accepted that growth is dependent on the agro-ecological conditions (i.e. geographical location) of a region,advancement in technology, investments aiding growth in infrastructure and input use and other region level policies and institutions can help an initially un-favourably endowed region, to perform better than it would have in their absence. At the same time, it has been argued, that regions with favourable agro-ecological conditions can benefit more from these infrastructural, policy and institutional supports than those with un-favourable ones. For example, Jones and Sen (2006) point out that although irrigation will be beneficial for agricultural performance in all regions, some regions are bound to benefit more owing to the fact that they have a favourable agro-ecological condition that makes large irrigation projects economically and practically feasible. So could be the case with other key infrastructure as well.

Thus, these (counter) arguments only show that we cannot say a priori whether countries / regions with similar agro-ecological conditions will grow similarly. Yet, as Hausmann (2001) in his article 'Prisoners of Geography' in Washington Post warns 'denying the impact of geography will only lead to misguided policies and wasted efforts'.

The clue to understanding the linkages between geographic location and growth across regions could lie in a certain pattern that one finds in almost all the past studies on growth clusters. Two aspects of the growth clusters identified in the past studies stand out. (i) Not all geographically neighboring countries / regions belong to the same cluster; and (ii) Many of the clusters include both geographical neighbors as well as non-neighbors. This pattern is seen in the results of Corrado et al (2005), Fritsche and Kuzin (2011) and Bartkowska and Reidl (2012) for Europe, Ghosh et al (2013) for India, and Maasoumi et al (2008) and Herririas and Ordonez (2012) for China. Understanding the reasons behind this pattern could possibly help in avoiding the misguided policies and wasted efforts that Hausmann (2001) talks about. The factors which determine this pattern are most likely the interventions that are required to bring non-neighbours (or neighbours) together. And these could potentially be able to reduce the widely documented impact of constraints imposed by un-favourable geographical location of regions on their growth.

So, what explains these phenomena? Internationally there are very few studies that have explored the factors that drive cluster formation. Bartkowska and Reidl (2012) and Kim and Rous (2012) use ordered probit and multimomial logit models, respectively, to relate clusters to certain determining factors. Corrado et al (2005) computed cluster correlations between clusters generated on the basis of Hobijn and Franses (2000) and hypothetically created clusters on the basis of various criteria like core-periphery, spatial proximity etc. to identify the reasons driving cluster formation. However, to the best of our knowledge, none of these studies address the issues of why all geographical neighbors are not members of the same cluster and why non-neighbors are members of the same cluster.

This paper addresses this question taking the case of agricultural income per capita in India. Indian agriculture offers a good case to study the issue of (club) convergence given the perennial nature of the huge disparity that exists across states (Bhalla and Singh (1997, 2009), Bhide et al (1998), Chand and Chauhan (1999), NAP (2000), Ghosh (2006), Chand and Raju (2008)). Further, studies such as those by Bhide et al (1998), Ghosh (2006), and Chatterjee (2014) find evidence of beta but not sigma convergence, suggesting the possibility of club convergence amongst the states.[1]

Hence, we first test for clubs/clusters in Indian agriculture over the period 1966-67 to 2010-11 using the methodology proposed Philips and Sul, 2007 (PS from now on).[2]The analysis has been done for three sub-periods to understand the temporal and spatial dynamics of cluster formation in Indian agriculture. As will be seen later both the number and members of clusters have changed in these sub-periods. We find that there were four, one and three clusters respectively in the three sub-periods. As in the case of other studies on club convergence, our results also show that (i) not all geographically neighboring states belong to the same income cluster; and (ii) most of the clusters include both geographical neighbors as well as non-neighbors.

Estimating ordered/multinomial logit/probit methodsas in Bartkowska and Reidl (2012) and Kim and Rous (2012) will not help understand the above patterns in clusters. Such models can help understandwhy a certain state is in a certain income-club/cluster but not answer our question of why neighbouring or non-neighbouring states are (are not) in the same cluster.

For that we consider a setting that differs from others. We use a pair of states as a cross-sectional observation in our analysis (the strategy has been discussed in detail in later sections of the paper). This will help us incorporate both the aspects: (a) the fact that the pair of states are (are not) members of the same income cluster and (b) they are (are not) geographical neighbours simultaneously in our estimation strategy. We then try to understand why a pair of states in India is in the same agricultural income cluster and estimate how relative similarity between them in various plausible control factors pushes them towards the same

[1]The methodological aspects of testing for (club) convergence are discussed later.

[2] To the best of our knowledge, in Indian context, Ghosh et al (2013) is the only other study that used PS methodology on per capita NSDP for the aggregate economy and for agriculture, industry and services sub-sectors from 1968 to 2008 and they found two clusters in agriculture sector.

cluster. Further, we also try to understand if factors driving state pairs which are not geographical neighbors are different from those pairs which are geographical neighbors.

The rest of the paper is organized as follows: The next section briefly describes the data used for the analysis. The methodology adopted for identifying clusters and the results have been discussed in section 3. Section 4discusses the methodology used to identify the factors determining the clusters and reports the results of the analysis while section 5 provides some concluding remarks.

2. Data

The main variable of interest here is the Net State Domestic Product (NSDP) per rural person, which is constructed out of the data on NSDP taken from the EPWRF database and rural population data from Census. The analysis is carried out for 17 states[3], which on an average account for over 95% of Net Domestic Product from agriculture. Data on the newly formed states of Jharkhand, Chhattisgarh and Uttaranchal have been merged with their parent states of Bihar, Madhya Pradesh and Uttar Pradesh, respectively, to maintain uniformity in the panel data set.

To understand the changing patterns of growth of agricultural income, the time period in this study, viz., 1966-67 to 2010-11, has been divided into three sub-periods on the basis of changing policies in agriculture sector: 1st sub-period (1966-1977) the period of Green Revolution; 2nd sub-period (1978-1989) period of falling public investment in agriculture; and 3rd sub-period (1990-2010) period of economic reforms. As seen in Figure-1, public investment in agriculture as a percentage of agricultural GDP was high in the latter half of 1960s and early 1970s (1st sub-period), fell sharply in the 1980s (2nd sub-period) and continue to fall for most of the period of economic reforms (3rd sub-period). Only since the late-2000s, is there a pickup in public investments in agriculture.

Figure-1: Share of Public GFCF in agriculture in GDP from agriculture

Table 1 reports the levels and growth of NSDP over the various periods and brings out some interesting aspects of the inter-state variation over the years. In terms of levels of per capita income, compared to other states Uttar Pradesh and eastern states like Assam and Bihar have always been poorly performing states while Punjab and Haryana in north-west have always

[3] The 17 states are Andhra Pradesh (AP), Assam, Bihar+Jharkhand, Gujarat, Haryana, Himachal Pradesh (HP), Jammu & Kashmir (JK), Karnataka, Kerala, Maharashtra, Madhya Pradesh +Chhattisgarh (MP), Orissa, Punjab, Rajasthan, Tamil Nadu (TN), Uttar Pradesh+ Uttarakhand (UP) and West Bengal (WB)

been better performers. However, as documented in results from studies like Bhalla and Singh (1997) and also from Table 1 we can see that growth rates of states like Punjab and Haryana have come down in the later sub-periods. Growth rates for almost all the states were highest in the first sub-period (1966-2010) which was the initial green revolution period while they were lowest and had the highest coefficient of variation in the second sub-period (1978-89). Third sub-period (1990-2010) saw some revival but growth rates were still lower than the first sub-period. Yet, one can also see some instances of catching up in states like Gujarat and Maharashtra in the west and Andhra Pradesh and Tamil Nadu in the south as they experience a higher growth rate in last sub-period.

Table 1: Agricultural performance across states of India

Data sources for the various explanatory variables were quinquennial livestock Census which is conducted by Department of animal husbandry, dairying and fishing, Government of India for tractors, 'Basic Road Statistics' and 'Statistical abstracts of India' for road density which is defined as of total road length per square kms geographical area in the state, EPWRF database for power/electricity consumed for agricultural purposes, 'Land use statistics, Department of Economics and Statistics, Ministry of Agriculture' for data on share of gross area irrigated in total cropped area, 'Finances of state government' published by RBI for state expenditure in agriculture[4]. Data of share of area under each crop-groups like cereals, fiber, pulses, sugar, oilseeds and rest was obtained from 'Area, Yield, Production of Principle Crops' by Ministry of Agriculture. Average annual data on rainfall (both actual and normal) was collected from various publications of Statistical abstract of India. Data on markets was compiled from 'Bulletin on Food statistics' published by Ministry of Agriculture for various years. Aggregate price index (ratio of NSDP from agriculture in current prices to constant prices) is used as a proxy for market integration and both the data was collected from EPWRF database.

3. Identifying clusters

Traditionally, convergence across regions has been analyzed through sigma and beta convergence measures. However, as discussed in studies like Azariadis and Drazen (1990), Quah (1993), Durlauf and Quah (1999), and Islam (2003), there are various problems associated with using these measures of convergence. Often beta convergence is seen along

[4]Expenditure on agriculture and allied activities include expenditure on crop husbandry, soil and water conservation, animal husbandry, dairy development, fisheries, forestry and wild life, plantations, food storage and warehousing, agriculture research and development, food and nutrition, community development and other agricultural programs(Both revenue and capital expenditure have been included)

with sigma divergence. Besides, there is also the issue of multiple steady states, which gives inconsistent results. This inconsistency has been resolved in the literature by exploring the possibility of multiple club/cluster convergence. Studies like Hobijn and Franses (2000), Busetti et al (2006) and Philips and Sul (2007) among many others have suggested methods to test for cluster/club convergence. These tests have been widely used in various studies like Maasoumi et al (2008) for China, Corrado et al (2005) and Bartkowska and Reidl (2012) for Europe among many others.

In order to analyze the transitional behavior of per capita income among the states in India, we apply the test developed by PS (the method has been described in detail in the Appendix). Briefly, this method first tests whether all the states converge with each other at the end of the time period and then if not, it proposes a cluster identification algorithm. The states are first sorted in descending order. Then starting from the 1st state, the first convergence cluster is identified by testing each state sequentially for membership of the group. The states that are not members of this group are then tested again sequentially to see if they form a second cluster, and so on till all the clusters are identified.

Accordingly, we tested for convergence for the three sub-periods and the t-values turned out to be -121.45, -35.75 and -92.17 for the three sub-periods, respectively, all of which are significant at <5%. That is, the null hypothesis that all the states converge with each other is rejected for all the three sub-periods. Hence, we proceed with the cluster identification algorithm.

Results of the clusters so identified for the three sub-periods are given in Table 2. The results show that in the first sub-period there were four converging clusters with different number of members and no divergent states. In the 2nd sub-period, there is only one cluster with 11 states and 6 diverging states, while in the 3rd sub-period there are 3 clusters with different number of members and 2 divergent states.

We can see a clear pattern from Table 2 that some states have been in the same cluster always. For example, Andhra Pradesh (AP) and Gujarat have been together always, though the rank of the income cluster (representing the level of income) that they were in has been changing over the sub-periods. Similarly, Karnataka and Himachal Pradesh (HP) have been together, as also Assam and Orissa, Madhya Pradesh (MP) and Uttar Pradesh (UP), and Punjab and Haryana. Besides these pairs of states, the five states of Jammu and Kashmir (JK), Kerala, Tamil Nadu (TN), Rajasthan and West Bengal (WB) as a group also show a similar pattern. Bihar is the only state that has always been the member of the last club with the lowest income levels and growth rates.

Table-2: Members of the clusters in the three sub-periods

To see if these clusters themselves are converging or diverging relative transition parameters are estimated using h_{it} from Equation A-3 in Appendix. Plots of these transition parameters for the three sub-periods are given in Figure 2. It can be seen that the clubs and divergent states indeed follow a different pattern from one another and none of them tend to converge with one another.

Figure-2: Relative transition parameters of clubs and divergent states

Two interesting features of these clusters come out from a geographical perspective. From Figure 3 it can be seen that in each sub-period several clusters include both contiguous and non-contiguous states. For example, in 1^{st} sub-period cluster-2 includes Gujarat, Karnataka, HP and AP, of which only Karnataka and AP are neighboring states. A second feature of these clusters is that not all neighboring states belong to the same cluster. For instance, Punjab-Haryana are neighbors and form the first income cluster in 1^{st} sub-period, their neighbors Jammu & Kashmir, HP, Rajasthan and UP are not members of this cluster. These two features are seen in several clusters in the three sub-periods. The next section explores the factors that determine cluster formation, and in particular examines the reasons behind this spatial pattern in the income clusters.

Figure-3: Spatial pattern of clusters

4. Factors driving states to common clusters

To identify the factors which drive the states to common income clusters we estimate a logistic model where the dependent and explanatory variables have been constructed in such a way that each cross-sectional observation refers to a particular pair of states. That is, first the 17 states were arranged in alphabetical order and every possible state-pair combination was taken to constitute one cross-sectional observation for a particular sub-period.

The dependent variable is a binary variable which takes the value one if both states in a state pair are members of the same income cluster in a particular sub-period as reported in Table 2 above; and zero otherwise. Thus, out of the time series data for a particular sub-period we come out with a cross-sectional data that tells whether a pair of state has been in the same cluster over that period. Therefore, there are 136 ($^{17}C_2$) observations for each sub-period and total number of observations over the three sub-periods is 408.It is worth stressing here that the dependent variable does not differentiate among high, low or middle level of clusters. For example in 1st sub-period, both the state pairs Punjab-Haryana (highest income cluster 1) and Bihar-Uttar Pradesh (lowest income cluster 4) are given a value of one.

On the same line as the dependent variable, explanatory variables were constructed in such a manner that they concurrently represent the state-pairs and the sub-periods. For each explanatory variable, the absolute difference between the two states in a state pair was constructed to capture the relative gap between the two states. Absolute differences were used because the order of the states in a state pair is alphabetic, and hence the sign of the difference has no particular interpretation. Absolute differences were computed for the initial and final time period of each sub-period, which can be written mathematically as: $|x_{i0} - x_{j0}|$ and $|x_{iT} - x_{jT}|$ respectively where 'i' and 'j' are the states and 0 and T are the initial and last years of each sub-period. Additionally, ratio of the difference in the final values to difference in the initial values of the explanatory variable was also constructed as a proxy for the relative growth of the two states in a state pair. This can be written as: $|x_{iT} - x_{jT}|/|x_{i0} - x_{j0}|$. When this ratio takes a value greater (lesser) than one then it implies that the two states are moving away from (closer to) each other in the explanatory variables.

Binary logistic regressions with state fixed effects are then estimated to identify the determinants of cluster formation. To study the factors that drive neighbors (non-neighbors) into different (same) income cluster, we divide the state-pairs into two sub-groups on the basis of state-level contiguity and estimate separate logistic models. Out of the total 408 observations, 96 are contiguous state-pairs and 312 non-contiguous state-pairs observations. The sub-group regressions help us understand how factors driving contiguous states to a common income cluster are different from those of non-contiguous ones.

Combined sample results: Columns 1-3 in Table 3 gives the results of the logistic regression with state fixed effects. It was found that power/electricity consumption and irrigation are highly correlated and hence they could not be used together in the same regression model. So in columns 1 and 2 we have controlled for irrigation only while in column 3 only power/electricity consumption has been controlled.

Per capita tractor ownership, which is a proxy for both mechanization and asset ownership, is a significant driver. Lower the difference between the states in per capita tractor ownership, higher is the probability of the states to form a part of the same cluster. It is worth stressing here that the model only captures the probability of the two states being in the same income cluster be it high/medium /low-income cluster. This probability is influenced by the relative gap in the explanatory variable (for example per capita tractor ownership), and not so much the level of tractor ownership per se, which could be high /medium / low in the two states. The model by no means implies any relationship between the level of the explanatory variables and the levels of income across the states. This aspect needs to be borne in mind while interpreting the results vis-à-vis the other explanatory variables.

Among variables indicating status of key infrastructure in the states, we find that relative differences in growth in total road density, irrigation measured in terms of share of gross area irrigated and power consumed for agricultural purposes per total cropped area in the state are significant drivers in the sense that smaller is the difference between any two states in these infrastructural status, higher are the chances that the two states are in the same income clusters. Impact of state support has been controlled in terms of expenditure of state on agriculture and it is also a significant driver. Share of area under all the crops grown in the states was used as a substitute of cropping pattern and among all the crop-groups, share of area under fiber was a significant factor. Again smaller is the difference in the absolute deviation of actual rainfall from normal levels, higher is the probability that states will be in the same cluster.

Recent literature stresses on the importance of market integration in growth of income and following that line, we created an aggregate price index (deflator[5]) for agriculture for the states and it was found to be again a significant driver i.e. smaller is the difference between the states in terms of price index, higher are the chances that they are in the same income cluster. Market per cropped area and deflator were highly correlated so we have controlled for market and deflator interchangeably in our models.

To summarize, smaller is the difference between the two states in state pair on policy variables like state expenditure on agriculture, mechanization substituted by per capita tractor ownership, infrastructure like markets, roads, irrigation and electricity consumption, price index substituted by deflator and deviation of actual rainfall from their normal levels, higher is the probability they will be in the same cluster.

Sub-sample results: Column 4 in table 3 gives the results of a similar regression when performed on contiguous states. Tractors again are a significant driver of states to common income clusters and so are markets, state expenditure on agriculture, growth of total road density and share of cropped area covered with fiber in the final year. Columns 5 and 6 show that the factors which drive non-contiguous states to common income clusters are tractors, irrigation, state expenditure, cropping pattern (share of fiber)and percentage deviation of

[5]Deflator = NSDP per rural person from agriculture in current prices/NSDP per rural person from agriculture in constant 2004-05 prices

actual rainfall from the normal levels are also significant. Here, it is the initial road density and not its growth which is significant driver of states to common clusters. Unlike contiguous state-pairs price deflators are significant drivers in this sub-sample (column 6).

Comparing the results for non-contiguous states with the contiguous states, we find that factors like tractor ownership, expenditure by state governments, markets, roads and cropping pattern are common to both sets of models. The factors which differentiate non-contiguous and contiguous states are deviation of actual rainfall from normal levels, irrigation and price differentials between state pairs.

Table 3: Results of the logistic regression on samples and sub-samples

5. Conclusion

The debate on the inter-play of geographical location and economic performance across regions, unsettled as it is, has largely overlooked an empirical regularity that past studies on growth clusters have thrown up, viz., (i) Not all geographically neighboring countries / regions belong to the same cluster; and (ii) Many of the clusters include both geographical neighbors as well as non-neighbors. This paper sought to study these phenomena using the case of agricultural income per capita across 17 major states of India over the period 1966-2010.For that we first divided the time period of our analysis into three sub-periods: first sub-period (Green Revolution period 1966-1977), second sub-period (period of falling public investment in agriculture 1978-89) and third sub-period (period of economic reforms 1990-2010).

Using the methodology proposed in Philips and Sul (2007), we identified the income clubs/clusters in terms of Net State Domestic Product (NSDP) from agriculture per rural person. We found that both the number and constitution of the clubs have changed over the sub-periods. All the states converged to four clusters in sub-period one. Sub-period two had only one cluster with 11 states, while all other states diverged from one another. In sub-period three we find three clusters and two divergent states. Conforming to the pattern in other studies on growth clusters, our results also showed that (a) clusters comprise of neighboring and non-neighboring states, and (b) not all neighboring states are in the same cluster.

To understand this pattern we adopted a different setting whereinwe used a pair of states as a cross-sectional observation in our analysis.This enabled us to define a new binary variable that takes the value one if both states in a particular state pair are in the same income cluster and zero if not and estimate a logistic model with state fixed effects for the full sample and for two sub-samples, viz., for the neighboring and non-neighboring states. The explanatory variables in these logistic regressions were constructed for a pair of states as the absolute difference between them in the levels of the variable. Further, the widening / narrowing of the gap over time between these states is also captured in the form of ratio of the difference in the variable in the final year to that in the initial year.

The results show that smaller is the relative difference between two states in a state pair in tractor ownership, infrastructure variables like irrigation, power consumption, roads and number of markets, state expenditure on agriculture, price-differentials and rainfall, higher is their probability of being in the same cluster. The results for the two sub-samples for the neighboring and non-neighboring states show that factors like tractor ownership, expenditure by state governments, markets, roads and cropping pattern are common to both of them. However, factors such as deviation of actual rainfall from normal levels, irrigation and price differentials between state pairs are the only ones that drive non-neighboring states into the same income cluster.

The empirical evidence presented here highlights the importance of mechanization, infrastructure, market integration and state support in growth in agriculture. Therefore, economic policy measures targeting improvement and expansion of infrastructural support, especially irrigation and efficient water management to reduce the dependence on vagaries of monsoon, mechanization in agriculture and institutional reforms that accelerate market integration and reduce price differentials across states can have a significant impact in promoting long run agriculture growth and convergence across Indian states.

Some of the limitations of the present study have to be kept in mind while drawing conclusions. Choice of the period of analysis has been largely been based on availability of data on various variables. To the extent possible the length of the sub-periods has been based on major technological / policy changes that could have affected Indian agriculture. Nevertheless, it is possible that the years chosen as the beginning and end years of the sub-periods have influenced the results on cluster constitution. Further, to the extent the explanatory variables are themselves defined for the beginning / end years of the sub-periods, they could have also affected the results of the logistic regressions reported here. And finally, due to multi-collinearity, the impact of irrigation and power consumption could not be estimated simultaneously though we find both of them to have significant impact independently. Despite these limitations, it is hoped that the results reported here do offer an alternative perspective on the inter-play of geographical location and economic performance across regions.

References

Azariadis, C., Drazen, A. (1990), Threshold externalities in economic development, *Quarterly Journal of Economics,* 105 (2): 501–526

Bartkowska, M., Reidl, A. (2012), Regional club convergence in Europe: Identification and conditioning factors, *Economic Modelling,* 29: 22-31

Bloom, D.E., Sachs, J.D., Collier, P., Udry, C. (1998), Geography, Demography, and Economic Growth in Africa, *Brookings Papers on Economic Activity*, 2: 207-295

Bhalla, G.S., Singh, G. (1997), Recent Developments in Indian Agriculture: A State Level Analysis, *Economic and Political Weekly,* 32 (13): A2-A18

Bhalla, G.S., Singh, G. (2009), Economic Liberalization and Indian Agriculture: A State-wise Analysis, *Economic and Political Weekly,* 44 (52): 34-44

Bhide, S, Kalirajan, K.P., Shand, R.T. (1998), India's Agricultural Dynamics: Weak Link in Development, *Economic and Political Weekly,* 33 (39): A118-A127

Busetti F., Forni L., Harvey A., Venditti F. (2006), Inflation convergence and divergence within the European monetary union, *ECB Working Paper Series.*574, European Central Bank

Corrado, L., Martin, R., Weeks, M. (2005), Identifying and Interpreting Regional Convergence Clusters across Europe, *The Economic Journal*, 115(502): C133-C160

Chand, R., Chauhan, S. (1999), Are disparities in Indian agriculture growing? *NCAP Policy Brief No. 8*

Chand, R., Raju, S.S. (2008), Instability in Indian agriculture during different periods of technology and policy, *.NPP discussion paper*-01

Chatterjee, T. (2014), Spatial convergence and growth in Indian agriculture: 1967-2010, *Indira Gandhi Institute of Development and Research, Working paper series,* WP-035

Durlauf, S.N., Quah, D.T. (1999), *The new empirics of economic growth.* In: Taylor JB, Woodford M (Eds) Handbook of macroeconomics, Elsevier Science, North-Holland, 1A, 235–308

Easterly, W. and Levine, R. (2003), Tropics, germs, and crops: how endowments influence economic development, *Journal of Monetary Economics*, 50: 3-39

Fritsche, U. and Kuzin, V. (2011), Analysing convergence in Europe using the non-linear single factor model, *Empirical Economics*, 41: 343-369

Gallup, J.L., Sachs, J.D., and Mellinger, A.D. (1999), Geography and Economic Development, *International Regional Science Review*, 22(2): 179-232

Ghosh, M. (2006), Regional convergence in Indian agriculture, *Indian Journal of Agricultural Economics,* 61 (4): 610–629.

Ghosh, M., Ghoshray, A., Malki, I. (2013), Regional divergence and club convergence in India, *Economic Modelling*, 30: 733-742

Government of India, *National Agricultural Policy, 2000*, accessed from http://agricoop.nic.in/agpolicy02.htm, last accessed on 10th December, 2014.

Hausmann, R. (2001), Prisoners of Geography, Foreign Policy, No.122, pp-44-53, published by: Washington Post, Newsweek Interactive, LLC

Herririas, M.J. and Ordonez, J. (2012), New evidence on the role of regional clusters and convergence in China (1952–2008), *China Economic Review*, 23: 1120-1133.

Hobijn, B., Franses, P.H. (2000), Asymptotically perfect and relative convergence of productivity, *Journal of Applied Econometrics*, 15: 59–81

Islam, N. (2003), What have we learnt from the convergence debate, *Journal of Economic Surveys*, 17(3): 309–362

Jones, R., Sen, K. (2006), It is where you are that matters: the spatial determinants of rural poverty in India, *Agricultural Economics,* 34: 229–242

Kim, Y.S., Rous, J.J. (2012), House Price Convergence: Evidence from U.S. State and Metropolitan Area Panels, *Journal of Housing Economics,* 21(2): 169-186.

Landes, D. S. (1998), The Wealth and Poverty of Nations: Why Some Are So Rich and Others So Poor, W.W. Norton.

Maasoumi, E., Wang, L. (2008), Economic Reform, Growth and Convergence in China, *The Econometrics Journal, 11: 128–154.*

Phillips, P.C.B., Sul, D. (2007), Transition modelling and econometric convergence tests, *Econometrica,* 75: 1771–1855

Phillips P.C.B., Sul, D. (2009), Economic transition and growth, *Journal of Applied Econometrics,* 24 (7): 1153–1185

Quah, D. T. (1993), Galton's Fallacy and Tests of the Convergence Hypothesis*,* *Scandinavian Journal of Economics,* 95 (4): 427-443

Rodrick, D., Subramanian, A., Trebbi, F. (2004), The Primacy of Institutions over Geography and Integration in Economic Development, *Journal of Economic Growth*, 9(2): 131-165

Appendix

Philips and Sul (2007) test is based on an innovative decomposition of the variable of interest. Usually panel data for the variable of interest (y_{it}) is decomposed in the following way:

$$logy_{it} = \varphi_i \mu_t + \varepsilon_{it} \qquad (A-1)$$

Where, φ_i represents the unit characteristic component, μ_t, the common factor and ε_{it}, the error term. Here, a time varying representation oflog y_{it} can be derived from equation A-1 and can be written as follows:

$$logy_{it} = (\varphi_i + (\varepsilon_{it}/\mu_t))\mu_t = \delta_{it}\mu_t \qquad (A-2)$$

Where both the error term and the unit specific component are condensed into δ_{it} which then represents the idiosyncratic part that varies over time.δ_{it}, here represents the share of the common growth path μ_t that the i-th economy goes through. This transition coefficient (δ_{it}) is modeled in such a way that the common growth path is eliminated. A relative transition coefficient, h_{it} is constructed from δ_{it} as follows:

$$h_{it} = \log(y_{it}) \Big/ N^{-1} \sum_{i=1}^{N} \log(y_{it}) = \delta_{it} \Big/ N^{-1} \sum_{i=1}^{N} \delta_{it} \qquad (A-3)$$

Hence, h_{it} represents the transition path of i-th economy relative to the cross-section average that is it measures the individual behavior in relation to other economies. In case of convergence, i.e. when all economies move towards the same transition path, $h_{it} \rightarrow 1$, for all i as $t \rightarrow \infty$. Then the cross sectional variance of h_{it}, denoted by $V_t^2 = N^{-1} \sum_i (h_{it} - 1)^2$, converges to 0. In order to specify the null hypothesis of convergence, PS model δ_{it} in a semi-parametric form:

$$\delta_{it} = \delta_i + \ (\sigma_i \varepsilon_{it}/L(t)t^{\alpha}) \qquad (A-4)$$

Where $\delta_i$is fixed, σ_i is an idiosyncratic scale parameter, ε_{it} is iid (0, 1), L (t) is a slowly varying function (such that L (t) $\rightarrow\infty$ when t $\rightarrow\infty$) and α is the decay rate. The null hypothesis of convergence can be written as: $H_0: \delta_i = \delta$ and $\alpha \geq 0$ and it is tested against the alternative $H_A: \delta_i \neq \delta$ for all i or $\alpha<0$. Under this null hypothesis of convergence various transitional patterns of economies i and j are possible, including temporary divergence, which refers to periods where $\delta_i \neq \delta_j$. Therefore, this method can detect convergence even in the case of transitional divergence, where other methods such as stationary tests for e.g. Hobijn and Franses, 2000 fail (Bartkowska and Reidl, 2012). PS show that under convergence the cross-sectional variance of h_{it} has the limiting form,

$$V_t^2 \sim A/L(t)^2 t^{2\alpha} \qquad \text{as } t \rightarrow \infty \text{ for some } A>0 \qquad (A-5)$$

The following regression based convergence test can be derived from above:

$$log(V_1^2/V_t^2) - 2logL(t) = a + blogt + u_t \qquad (A-6)$$

for t= [rT],[rT]+1……T where r is the initiating sample fraction and general r Є (0,1) and L(t) is a slowly varying function. The log t procedure depends on the choice of the slowly varying function L (t) and 'r'. Rejection rates of the log t test depend on combinations of r, α, N and T. The authors suggest that since value of α is not known so for purposes of empirical application, value of r must be chosen in such a way so that firstly, size will be accurate when α is close to zero, secondly, for which size is not too conservative when α is larger and thirdly for which power is not substantially reduced by the effective sample size reduction. Their Monte Carlo simulation results indicate that r=0.3 seems to be a preferable choice when sample size is below T=50 to secure size accuracy in the test for small α. Simulation tests involving varying the L(t) function showed that the test is conservative when L(t) = log t is used in the regressions. Hence, in this study with a sample size of less than T=50, suggestions by authors have been strictly followed and r=0.3 and L (t) =log t have been used for the analysis. Using$\hat{\beta} = 2\hat{\alpha}$ for the log t regression, a one sided t-test robust to heteroscedasticity and autocorrelation (HAC) is applied to test the inequality of the null hypothesis $\alpha \geq 0$.The null hypothesis of convergence is rejected if $t_{\hat{b}} < -1.65$ (at 5 per cent level of significance). PS state that in case null of convergence is rejected, many possibilities exist for example possible existence of convergence clusters around separate points of equilibria or steady state growth path or cases where there may be possibility of coexistence of convergent clusters and divergent members in the full panel.

If the null hypothesis of convergence is rejected for the overall sample, PS suggests applying the following step-wise cluster mechanism to subgroups:

1. The cross-sectional units are ordered by final observation in descending order as according to them, when there is evidence of multiple club convergence as $T \to \infty$, this is usually most apparent in the final time series observations.
2. Core group is identified- The first k units of the panel are taken such that $2 \leq k < N$ and log t regressions (as discussed above) are run and $t_{\hat{b}}$ is calculated for the k selected units each time adding further units one by one. These log t regressions are run as long as $t_{\hat{b}}$ keeps on increasing and is larger than -1.65. Once, a smaller value of $t_{\hat{b}}$ is obtained, it can be concluded that the core group with $k^{*}=k-1$ members is formed. If that $t_{\hat{b}}>-1.65$ does not hold for the first two units, the first unit is dropped and the log t regression is run again for the second and the third unit and so on till $t_{\hat{b}}>-1.65$ is reached for the two units and then again the same process is repeated. If there are no two units for which $t_{\hat{b}}>-1.65$, it can be concluded that there are no convergence clubs in the panel.
3. New club members are added to the core group- One additional unit is added to the core group and the log t regression is run. This is repeated for all the units outside the core group. Units where $t_{\hat{b}}>c$ where c is critical value ($c \geq 0$) are selected and added to the core group. The authors suggest using conservative critical value as it reduces the chances of including a false member into the group. There is a problem with this approach i.e. chances are that more number of converging groups are identified than what exist. To correct this problem Philips and Sul (2009) suggest a merging procedure[6]. Then the log t regression is run for the whole group. If$t_{\hat{b}}>-1.65$, it can be concluded that all these members are part of the same convergence club otherwise, the critical value is increased for the club membership selection and the process is repeated till $t_{\hat{b}}>-1.65$ for the entire group. Then it can be concluded that these units form a convergence club. If there are no units apart from the core group that result in $t_{\hat{b}}>-1.65$, it can be concluded that the convergence club consists only of the core group.
4. Recursive and stopping rule- A second group is formed consisting of all units outside of the convergence club, i.e. where $t_{\hat{b}}<c$. The log t test is run for the entire group to check whether $t_{\hat{b}}>-1.65$ and the group converges. If not steps 1 to 3 are repeated on

[6]Philips and Sul (2009) suggest a test for merging between the groups formed according to the clustering algorithm (steps 1 to 4) where the log t test is run on the first two groups. If the t-statistic is larger than -1.65 (5 per cent level of significance), both the groups are merged to form a group. The test is repeated after adding the next group until the t-statistic indicates that the convergence hypothesis is rejected and the step is repeated until it is concluded that the remaining groups do not merge with each other.

this group to determine whether the panel includes a smaller subgroup that forms a convergence club. If there is no k in step 2 for which $t_{\hat{b}}>-1.65$, it can be concluded that the remaining units diverge.

As suggested in PS, the log of NSDP per rural person was filtered following the Hodrick Prescott methodology (using 100 as the smoothing parameter as the per capita income data are annual) to remove the business cycle impact from the data.

FIGURES

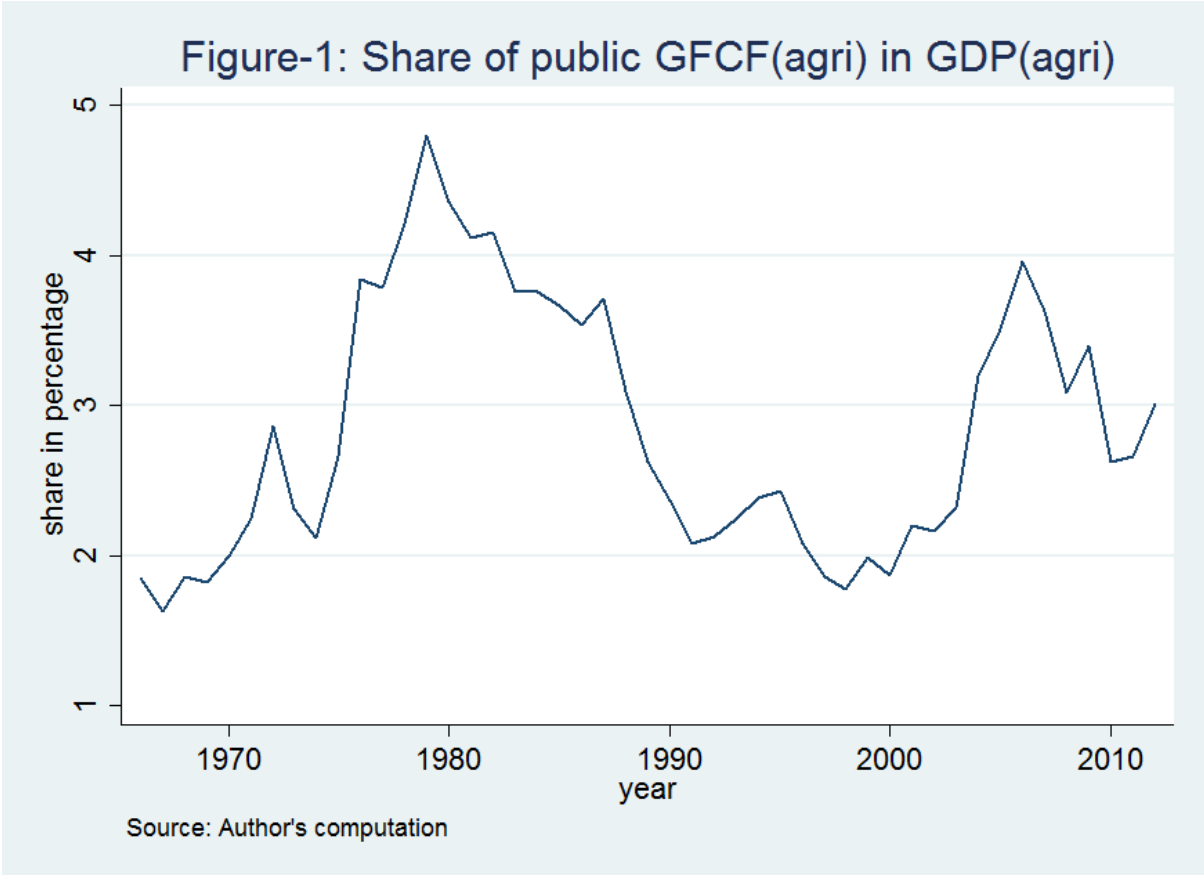
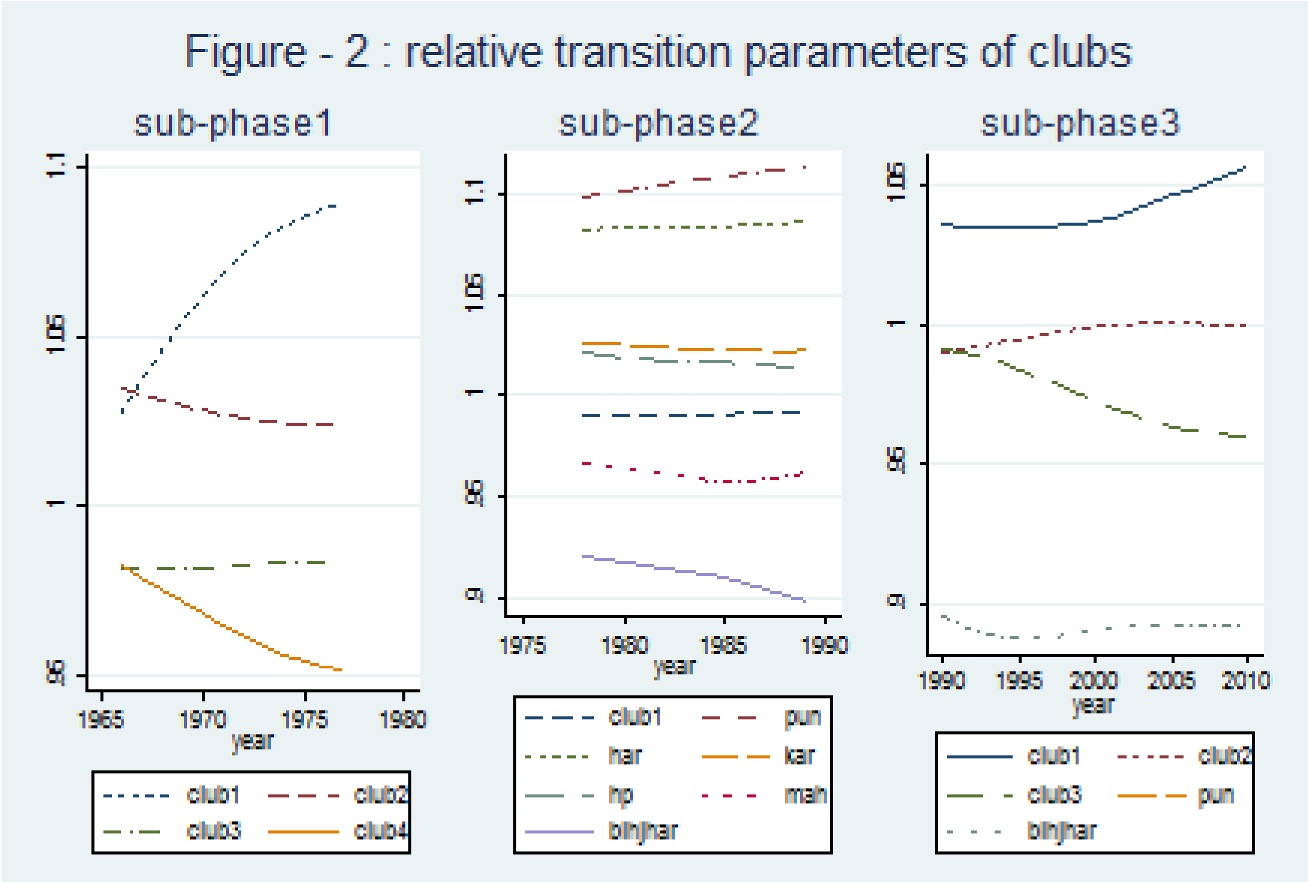
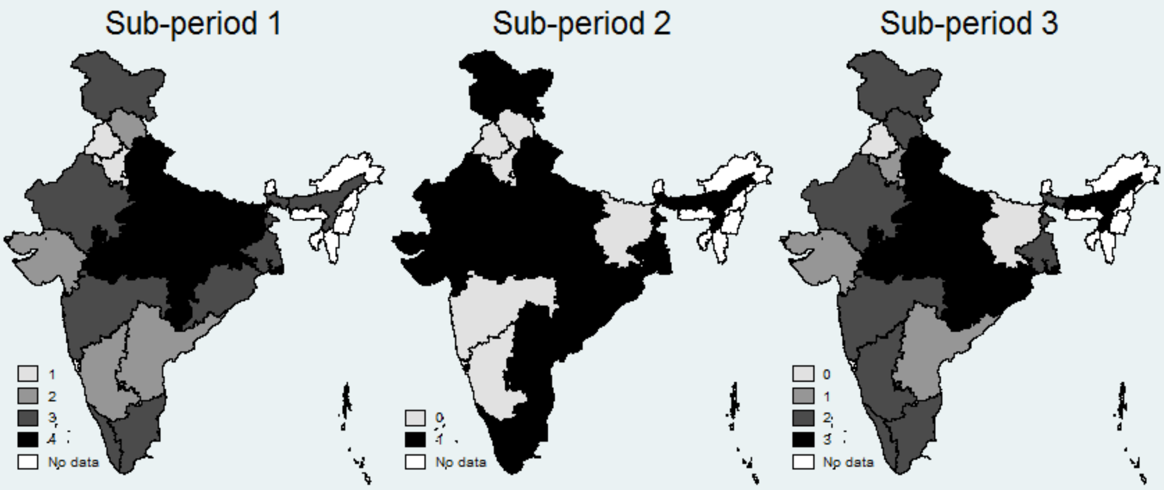


Figure-1: Share of public GFCF(agri) in GDP(agri)
share in percentage
1
2
3
4
5
1970
1980
1990
2000
2010
year
Source: Author's computation



source : author ' s computation

Figure - 3 : Spatial pattern of Clusters



Same colour scheme indicate same income clusters and 0 implies divergent states

Source : author ' s computation

TABLES

Table 1: Agricultural performance across states of India

	Per capita income levels (in Rs.)				Compound growth rate (in %)			
State	1966	1977	1989	2010	1966-2010	1966-77	1978-89	1990-2010
Andhra Pradesh	4936	5126	6421	10652	1.72	0.32	1.89	2.44
Assam	4023	4592	5086	5039	0.5	1.11	0.85	-0.04
Bihar+Jharkhand	755	2483	2113	2784	2.94	10.43	-1.34	1.32
Gujarat	2220	6241	6491	10953	3.61	9.00	0.33	2.52
Haryana	3188	9420	11487	14966	3.5	9.45	1.67	1.27
Himachal Pradesh	2714	5627	7068	7135	2.17	6.26	1.92	0.04
Jammu & Kashmir	2502	5362	4840	6673	2.2	6.56	-0.85	1.54
Karnataka	2922	6122	6389	9159	2.57	6.36	0.36	1.73
Kerala	1875	3696	4109	6859	2.92	5.82	0.89	2.47
Maharashtra	1507	3825	4605	7774	3.71	8.07	1.56	2.52
MP+Chhattisgarh	2238	3491	5229	6145	2.27	3.78	3.42	0.77
Orissa	1668	3527	6066	4945	2.44	6.44	4.62	-0.97
Punjab	2195	10322	14871	17950	4.78	13.77	3.09	0.9
Rajasthan	1855	4742	4889	7687	3.21	8.14	0.25	2.18
Tamil Nadu	2467	4613	4173	6913	2.32	5.35	-0.83	2.43
UP+Uttarakhand	3091	4138	4390	4874	1.02	2.46	0.49	0.50
West Bengal	1374	3794	4820	7024	3.69	8.83	2.02	1.81
India	2449	4615	5009	6793	2.29	5.42	0.69	1.46
CV	41.3	40.04	49.22	46.77	39.26	51.75	78.37	49.87

Source: author's computation

Table 2: Members of the clusters in the three sub-periods

club	no of states	members	t-stat	per capita income in Rs. (final year)	Average annual growth rate (%)
Sub-period 1					
one	2	Punjab, Haryana	-0.3333	9871	17.33
two	4	Gujarat, Karnataka, HP and AP	2.2039	5779	10.56
three	8	JK, Assam, Rajasthan, TN, Kerala, WB, Orissa, Maharashtra	1.7193	4269	9.98
four	3	UP, MP and Bihar	7.9422	3371	8.66
Sub- period 2					
one	11	AP, Gujarat, MP, JK, Assam, Orissa, Rajasthan, WB, UP, TN, Kerala	-0.1062	5138	3.43
divergent states	6	Punjab, Haryana, Karnataka, HP, Maharashtra and Bihar			
Sub- period 3					
one	3	Haryana, AP, Gujarat	2.1818	12191	3.53
two	8	Karnataka, Maharashtra, Kerala, HP, WB, TN, JK, Rajasthan	2.4131	7403	2.59
three	4	MP, Orissa, UP, Assam	6.8639	5251	0.89
divergent states	2	Punjab, Bihar			

source: authors computation

Table 3: Results of the logistic regression on samples and sub-samples

	all state-pairs			contiguous state-pair	non contiguous state-pair	
Variables	[1]	[2]	[3]	[4]	[5]	[6]
per capita tractor	-5.549***	-5.269***	-7.742***	-8.341@	-5.692***	-6.296***
market per cropped area(final)		-11.358***	-11.122***	-21.701**	-8.776*	
Irrigation	-2.270***	-1.774**			-2.255***	-2.639***
power per cropped area			-1.356@			
per cap agriexp (final)	-0.137***	-0.147***	-0.115***	-0.587***	-0.091*	-0.082*
growth in total road density	-2.48e-4*	-2.51e-4@	-2.14e-4**	-2.37e-4***		
total road density					-0.350**	-0.265@
share of fibre(final)	-8.438***	-8.686***	-6.364**	-13.840***	-6.254**	-5.632*
Deflator	-5.731*					-9.648***
abs. deviation of rainfall	-0.027**	-0.030**	-0.025**		-0.030**	-0.026*
State effect: Bihar+Jharkhand	-2.382***	-2.587***	-2.264***			
State effect : Haryana	-2.512**	-2.539**				
State effect :J&K	2.275***	2.234***	2.182***		2.365***	2.533***
State effect :Karnataka	-1.419**	-1.486**				
State effect :Kerala	1.048**	1.118**	1.441***		1.215**	1.234**
State effect :Rajasthan			1.824*			
constant	1.544***	1.703***	1.038***	1.768**	1.598***	1.513***
STATISTICS						
No. of observations	408	408	408	96	312	312
AIC	409.735	406.69	422.456	101.249	334.95	331.014
BIC	461.882	458.836	470.591	116.635	372.38	368.444
log-likelihood	-191.868	-190.345	-199.228	-44.625	-157.475	-155.507
pseudo-R-sq.	0.249	0.255	0.219	0.218	0.204	0.214

@: p<0.15, *:p<0.10, **:p<0.05, ***:p<0.01. Final year values were used for market and state expenditure as data prior to year 1976 and 1972 respectively is not available. Latest data for markets is available for year 2008.