

Long-Term Spatial Associations and Regional Convergence of Rural Literacy Rate in India



Hazem Mohamed

University of Manouba, TUNISIA

ABSTRACT: Although there are claims that inequality in India has stagnated or even increased in the post-liberalization period, careful analysis of data shows that these views are mostly unsubstantiated. The main aspects of social inequality can be structured, to a large extent, along urban-rural and gender lines. In this paper, we explore a nonparametric approach to highlight the form of the long-term spatial associations of the georeferenced units. Relying on the method of Markov chains, we develop a stationary form for the Moran scatter plot, one of the techniques of the exploratory spatial data analysis. As an illustration, we highlight the forms of spatial autocorrelation and heterogeneity of long-term literacy rates in rural districts in India observed by the two censuses of 1991 and 2001. We use the same data to estimate a spatial β -convergence model. It appears that the growth of literacy rate is higher in rural districts with the lowest literacy and proceeds by spatial diffusion. Beyond his correspondence with the β -convergence model, the proposed nonparametric approach has the advantage of identifying the location of spatial associations

KEYWORDS: Stationary Spatial Associations, Regional Convergence, Exploratory Spatial Data Analysis, Spatial Econometrics, Literacy Rate, Rural Districts, India.

JEL Classification: *C14 ; C31 ; I24.*

I- INTRODUCTION

In India, the mode of development has generated strong territorial inequalities. Like population and female literacy, cultural, health, infrastructure and geographic factors influence the education of rural regional units in particular and the education system in general. As infrastructure has improved in urban areas, rural communities have been largely left on their own, without government assistance or sufficient access to foreign structural aid. This has led to discrepancies in economic development within Indian rural districts, as rundown regions have failed to attract investment, thus falling behind in employment and income levels.

Since the revolution in geographic information systems in the 1980s, spatial databases have multiplied and have made it possible to use specific analysis and modeling methods. The relative position of regional units is not neutral and there are multiple spatial interactions. Hence, the inclusion of autocorrelation and spatial heterogeneity in recent empirical models has become essential. The methods used by spatial analysis are numerous. Exploratory Spatial Data Analysis (ESDA) is a set of techniques designed to describe and visualize spatial distributions, extreme observations and detect patterns of spatial association (Haining, 1990; Bailey and Gatrell, 1995; Anselin, 1998; Le Gallo, 2002). The spatial econometric approach has been employed in several studies alongside convergence models, both unconditional (Rey & Montouri, 1999) and conditional (López-Bazo et al., 1999; Fingleton & López-Bazo, 2006; Ezcurra & Rios, 2015). In the other hand, econometric analysis of convergence processes across countries or regions usually refers to a transition period between an arbitrary chosen starting year and a fictitious steady state (Kosfeld & Lauridsen, 2004). In this sense, the issue of regional convergence has been the subject of a wide range of empirical research and the most widely used model for testing convergence hypotheses is beta-convergence analysis. When a clear convergence pattern does exist, this process can be explained by many variables, especially human capital, trade and regional integration, transports and infrastructures (Hammouda et al., 2009; Guétat & Serranito, 2010). Other factors commonly included in the econometric modelling of convergence are demographic variables, labor market conditions, industrial structure, institutional factors and overall government policy (Davor & all, 2013).

In this article, we explore a non-parametric approach to determine the form of long-term spatial associations of georeferenced units. This method consists in determining a stationary form on the Moran Scatter plot, using the Markovian approach. We

Long-Term Spatial Associations and Regional Convergence of Rural Literacy Rate in India

highlight the correspondence that may exist between this approach and the convergence models in a spatial framework, and its contribution. We carry out an application to illustrate the methodological approach proposed, taking the case of the evolution of the literacy rate in Indian rural districts in India observed from the two censuses of 1991 and 2001.

The structure of this article is as follows. In Section 2, a spatial model is specified to estimate regional convergence of literacy rate in Indian rural districts. In section 3, we highlight the forms of long-term spatial associations of districts by proceeding through the proposed non-parametric approach. In conclusion, we show the contribution of the proposed method with regard to the results of the $c\beta$ -convergence model in a spatial framework.

II- DISPARITIES AND REGIONAL CONVERGENCE OF LITERACY RATE IN INDIAN RURAL DISTRICTS OVER THE PERIOD 1991-2001

1. Regional disparities

Although there are claims that inequality in India has stagnated or even increased in the post-liberalization period, careful analysis of data shows that these views are mostly un-substantiated. In this sense, one of the peculiarities of Indian territory is its federal nature which results in the continuous interactions between the central government and these different states. It is a country characterized by both extreme cultural diversity and strong inequalities (Tawa et al, 2008). In the aftermath of independence, India wanted to prove to the world its ability to be self-sufficient and its ability to cope: to tap into its heritage to fight injustices. The main aspects of social inequality in India can be structured, to a large extent, along urban-rural and gender lines. It is the country that offers a particularly complex subject of study in the social sciences because of its high degree of segmentation (Guilmoto, 2005). From independence, education was perceived by the new leaders as a necessity, both from a political point of view (to ensure national unity and consolidate democracy), and economically (to meet the needs of development: industrialization, tertiarization) and social (family planning, etc.), argue Oliveau and Chasles (2005). Its credibility therefore rests on a school system built around the idea of universal education (Buisson, 2007). The weakness of the educational policy can be explained by an investment strategy which has not really integrated investment in human capital, which explains a persistent weakness of the education system after independence and therefore, a remarkable illiteracy rate especially among women. In this work, we ask the question about the forms of disparities and interdependencies between Indian rural districts and thus we seek to know whether women benefit from improvements in the field of education. However, it seems more interesting to look in detail for the answers through the literacy data in the neighborhoods. By placing ourselves in the context of inequalities and the determinants of growth, we are interested here in a variable that characterizes human capital, the literacy rate in Indian rural districts.

In 2001, figure 1 shows that the highest levels of the literacy¹ rate in the rural districts² are notably recorded in the districts of the states of Kerala³, Mizoram, Karnataka, Maharashtra and New Delhi. In the same map, we notice that the lowest levels are recorded in the states of Orissa, Bihar, Andhra Pradesh and Uttar Pradesh. Note that the lowest rate (26%) is recorded in the Dantewada district of the state of Chhattisgarh. The Indian population is in the process of urbanizing and if the Indian rural population is for the moment greater than 65% the large rural exodus is expected to continue, which would reduce this figure to less than 50% by 2025-2030.

¹ UNESCO defines an illiterate person as "someone, who understands the situation, cannot read or write a simple statement about their day-to-day life". UNESCO defines a literate person as "a person who has acquired the essential knowledge and skills which make them capable of engaging in all activities in which literacy is necessary for effective community functioning." The National Literacy Mission (NLM) defines literacy like "acquiring the skills of reading, writing and arithmetic and the ability to apply them in one's life". The Indian census defined literacy as "reading and writing with the ability to understand in any Indian language".

² The average literacy rate in rural districts is 60.4158%. Among 581 districts, 292 are associated with rates below this value.

³ The highest value is recorded in the Kottayam district (95.9%) of this state.

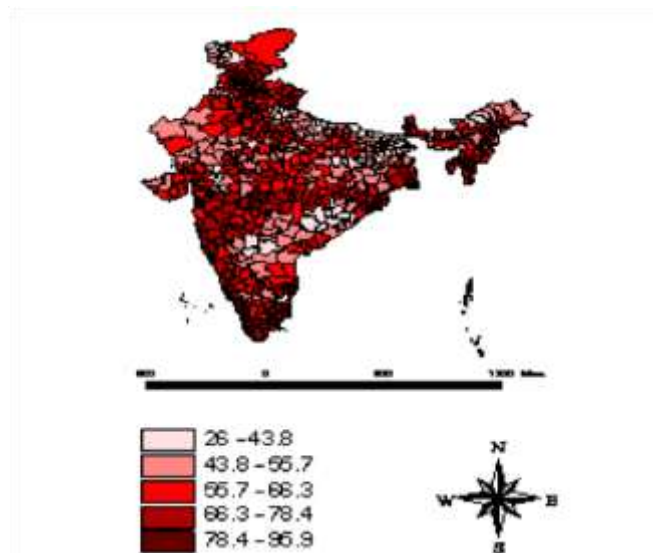


Figure 1: Literacy rate in rural districts (2001)

The “Trend-surface” is a representation in which the exploratory variables consist of polynomials of geographical coordinates of districts. An expression with three dimensions is generally used, the abscissa (x) and ordinate (y) are respectively represented by the geographical coordinates of the observations and the value of the variable represents the side (z). Linear (x and y) and quadratic (x, y, x², y², xy) forms are the most frequently used. Table (1) shows that the coordinates (x and y) are negatively significant (-0.2 and -0.588). This proves that there is a growing trend in the level of literacy from East to West to and from North to South.

Table 1: Linear trend surface

Variable	Coefficient	Standard-Err	T-Statistic	Critical-Pr
<i>cste</i>	90.388	6.846	13.202	0.00000
<i>x</i>	-0.200	0.084	-2.387	0.01729
<i>y</i>	-0.588	0.091	-6.459	0.00000

Table (2) indicates that all variables are statistically significant excepting the variables y and xy. The two variables x and y are negatively significant while both variables x² and y² are positively significant. This means that the model reveals a “bowl-like Shaped” for the trend surface, where the level of literacy is low in the middle and increases by approaching to the borders.

Table 2: Quadratic trend surface

Variable	Coefficient	Standard-Err	T-Statistic	Critical-Pr
<i>cste</i>	492.47	88.065	5.592	0.0000000
<i>x</i>	-9.698	2.088	-4.642	0.0000043
<i>y</i>	-2.212	1.964	-1.125	0.2606763
<i>x²</i>	0.060	0.013	4.574	0.0000059
<i>y²</i>	0.068	0.012	5.306	0.0000002
<i>xy</i>	-0.016	0.023	-0.734	0.4632009

2. Regional convergence

Income growth and convergence also known as the catch-up effect has been studied extensively in the literature with most studies citing the works of Solow (1956). Recent research on economic growth and regional convergence has incorporated the analysis of spatial spillovers, acknowledging that traditional determinants of regional growth are subtly altered when the spatial effect is taken into account (Abreu et al., 2005). In this sense, many studies have shown that geographic location does matter in terms of regional growth performance. Therefore, it is necessary to include the location in growth models, because otherwise the results obtained could be biased and any conclusions misleading (Ferrer, 2017).

Long-Term Spatial Associations and Regional Convergence of Rural Literacy Rate in India

The β -Convergence hypothesis presupposes the presence of a catch-up mechanism which reduces the gap between regional units. To identify it, Barro and Sala-i-Martin (1991) considered that the error terms are independent and estimated the regression coefficients from the ordinary least squares (OLS) method. In the case of georeferenced observations, the econometric estimates neglect a specific characteristic of the spatial data. They cannot be considered as independent because the geographical distribution of growth on a regional scale is rarely random. The use of localized data involves mobilizing an appropriate statistical and econometric framework to detect and treat a geographic interactions between regions.

First, we will test the hypothesis of absolute and conditional β -convergence of Indian rural districts in terms of literacy over the period 1991-2001. Then, we show the importance of the role played by geographic indicators to improve the model. Our analysis covers 563 rural districts in India⁴. The absolute β -convergence hypothesis is usually tested on the following cross-sectional model:

$$\begin{cases} GRRL = \alpha S + \beta \text{LogRL91} + \varepsilon \\ \varepsilon \sim N(0, \sigma^2 I) \end{cases} \quad (1)$$

where $GRRL$ is the vector of the average annual growth rates of the literacy rate in the rural districts over the period 1991-2001, LogRL91 is the vector of the literacy rates expressed in logarithm at the initial date (1991), S is the vector sum of dimension (563,1) and α and β are the parameters to be estimated.

Table 3: Estimation of the absolute β -convergence model

Estimation results by the OLS			
α	0.1213 (0.0000000)*	VC	5.14%
β_1	-0.0402 (0.0000000)*	DV	16.89 ans
R2	0.662	JB	21.1152 (0.0000055)*
LogL	1751.41	BP	3.5421 (0.0864182)***
AIC	-3895.13	KB	2.7451 (0.0952142)***
SC	-3885.15	WH	17.1245 (0.0975213)***
Specification of a spatial model			
	$W(dist)$	$W(10)$	$W(20)$
IM-Err	0.2252 (0.0000002)*	0.2661 (0.0000012)*	0.2712 (0.0000014)*
LM-Lag	20.1245 (0.0000122)*	17.5241 (0.0006125)*	19.2152 (0.0005212)*
RLM-Lag	2.5212 (0.0000453)*	2.1582 0.00005241)*	2.35216 (0.0000325)*
LM-Err	17.1254 (0.0001254)*	15.1352 (0.0008521)*	16.5213 (0.0001952)*
RLM-Err	1.9856 (0.0000725)*	1.1235 (0.0004525)*	1.1652 (0.0000521)*

Notes: *LogL* designates the value of the log likelihood function. *AIC* is the information criterion of Akaike (1974). *SC* is the information criterion of Schwarz (1978). *JB* is the Residue Normality test of Jarque-Bera (1987). *BP*, *KB* and *WH* are the heteroscedasticity tests of the Breush-Pagan, Koenker-Bass and White error terms respectively. *IM (Err)* is Moran's test (Cliff and Ord, 1981). *LM-Lag* and *RLM-Lag* denote respectively the Lagrange multiplier test applied to the lagged endogenous variable and

⁴ We do not include the districts that underwent redistricting between 1991 and 2001 and whose values cannot be restored between the two censuses.

Long-Term Spatial Associations and Regional Convergence of Rural Literacy Rate in India

its robust version. LM-Err is the Lagrange multiplier test applied to the error terms and RLM-Err indicates its robust version. Critical values are in parentheses. *: significant at 1%, **: significant at 5%, ***: significant at 10%.

The OLS estimation results (Table 3) show the existence of an global convergence of the rural districts since the parameter β (-0.040) is significantly negative. This suggests that the geographic dispersion of literacy rates is decreasing and that the improvement is most marked for the rural districts with the highest illiteracy. Convergence speed (CS)⁵ and the half-life (LT)⁶ are respectively equal to 5.14% and 16.89 years. But the value of Moran's statistic (IMErr) is very high (0.225, 0.2661 and 0.2712)⁷ whatever the spatial weight matrix used and shows that the error terms resulting from the estimation by the OLS of the model of absolute β -convergence are spatially auto-correlated. Model (1) therefore suffers from wrong specification due to the omission of the spatial autocorrelation of the literacy rates recorded in the 563 rural districts. Hence, the specification of an adequate spatial model is essential. In this sense, the two most commonly used spatial models are the spatial lag model and the spatial autocorrelation model of error terms. Simple Lagrange multiplier tests do not take into account the possible presence of a lagged endogenous variable when testing for the presence of spatial autocorrelation of errors. The application of the robust approach⁸ leads us to retain the spatially lag model¹⁰ as the best specification that can describe the process of convergence of rural districts in terms of literacy over the period 1991-2001. This model is defined by the following expression:

$$\begin{cases} GRRL = \alpha S + \rho WGRRL + \beta \text{LogRL91} + \varepsilon \\ \varepsilon \sim N(0, \sigma^2 I) \end{cases} \quad (2)$$

with the same notations as the previous model and where W is the spatial weight matrix. It is a square matrix with as many rows and columns as there are regional points. This tool is essential because not only does it make it possible to solve the estimation problems posed by the multidimensional nature of spatial data, but it also makes it possible to define the topology of the space studied (by making it possible to locate the spatial units in relation to each other) and the relative weight of each of the spatial units that compose it (Ertur and Thiaw, 2005). The autoregressive parameter ρ measures the intensity of the spatial interaction that exists between a rural district and its neighbors.

The results of the estimation of model (2) are given in Table 4. The regional convergence hypothesis is verified since the parameter β (-0.043) is significantly negative. The speed of convergence and the half-life are respectively equal to 5.65% and 15.7 years. Compared to model (1), it is clear that the regional proximity effects introduced in the model clearly accentuate the catching up in terms of literacy rate in rural districts. The model, based on the inverse distance matrix, also shows a positive spatial autocorrelation: the parameter is estimated at 0.35 with a very high critical probability ($p = 0.0004$) and the likelihood ratio test is very significant. (LRT = 15.14 and $p = 0.0000$). This result shows the extent of the similarities that exist between the growth rate of the literacy rate observed in a rural district and the weighted average of the values observed in neighboring rural districts. It reflects the idea that the increase in the literacy rate in rural districts is by contagion, by a certain spatial diffusion. Note however that in this specification the problem of heteroscedasticity of the error terms is not solved (BP = 4.586 and $p = 0.0023$).

Table 4 : Estimation of spatial models

	Spatial convergence model	Lag	Spatial Lag conditional convergence model
α	0.213 (0.0000020)*		0.524 (0.0000510)*
β	-0.043 (0.0000521)*		-0.044 (0.0000322)*
δ	-		-0.152 (0.0008851)*

⁵ $CS = -\ln(1 + T\hat{\beta}) / T$ indicates the speed necessary for an economy to reach its state of equilibrium.

⁶ $LT = -\ln(2) / \ln(1 + \hat{\beta})$ defines the time required for an economy to reach half the distance between it and its state of equilibrium.

⁷ The matrix of inverse distances and those of the 10 and 20 closest neighboring districts are used.

⁸ There are four approaches to specifying spatial models: the classical approach, the hybrid approach, the robust approach and the Hendry approach. The robust approach is the most commonly used. For more details see Florax et al (2003). Specification Searches in Spatial Econometrics: The Relevance of Hendry's Methodology.

⁹ The two statistics RLM-Lag and RLM-Err are significant, hence their comparison is necessary.

Long-Term Spatial Associations and Regional Convergence of Rural Literacy Rate in India

R^2	0.721	0.7512
VC	5.6552%	5.830%
DV	15.7033 ans	15.34 ans
ρ	0.351 (0.0000421)*	0.45216 (0.0000001)*
LogL	1781.18	1790.20
AIC	-3386.37	-3401.29
SC	-3373.37	-3352.51
BP	4.5857 (0.002312)*	5.458 (0.0981232)***
LRT	15.139 (0.0000999)*	16.522 (0.0000011)*

Notes: LogL designates the value of the log likelihood function. AIC is Akaike's information criterion. SC is the Schwarz information criterion. BP is the Breush-Pagan error terms heteroskedasticity test. LRT is the likelihood ratio test. Critical values are in parentheses. *: significant at 1%, **: significant at 5%, ***: significant at 10%.

Also, even if the spatial autocorrelation can serve as a proxy for the explanatory variables omitted in the model, it seems important, given the information available at the geographic level considered, to explain one: the growth rate of the fertility of women (GFER). The conditional convergence model is then defined by:

$$\left\{ \begin{array}{l} GRRL = \alpha S + \beta \text{LogRL91} + \delta \text{GFER} + \varepsilon \\ \varepsilon \sim N(0, \sigma^2 I) \end{array} \right. \quad (3)$$

India illustrates the paradox of global demographic: the population is increasing rapidly; although the birth control is largely widespread and some states have a fertility rate below the level of replacement generations (Véron, 2006). During one century, India has seen a multiplication by 4.5 of its population and it is expected that it will increase by 50 percent in the next fifty years. The United Nations project that India could exceed 1.5 billion in 2050.

Population growth is explained by the structure of population relatively young by age and by an average fertility rate of 3.1 children per woman. The relative decline in Indian fertility in the past twenty-five years is mainly attributable to the reduction in number of the births of three or more children, indicating that the model of the family with two children begins to spread throughout the country (Spoorenberg, 2001). The Indian population is now significantly urbanizing even if the rural population still above 65 percent. Kerala, one of the most developed state in the south of the country, has made an exception by the precocity and speed of its demographic transition: the crude birth rate declined from 44 per thousand in the 1950's to 17 per thousand in 2002 and the fertility which was 5.6 children per woman in the 1950s, is now below that of France: 1.8 against 1.9 (Veron, 2006). Districts with a fertility level below the national Indian average 3.31 children per woman, are more widely observed in the southern part of the country (particularly, in the states of Kerala and Karnataka, the district Chennai in Tamil Nadu state). In these states, couples use contraception for spacing and limiting births and therefore the fertility rate revolves around the replacement level (two children per women). Between the years 1974 and 1980, the government conducted a campaign of forced sterilization: more than seven million fathers of two or more children were vasectomized by force between April and January 1977 and thousands of women had to undergo a ligation horn (Oyarzun, 2010). The poor's are particularly affected by this practice. When they apply for job or request a credit, they are explained that they would have more chance of accepting their applications if they had a certificate of sterilization (Oyarzun, 2010). The coastal regions often stand out from the inside (ancient Christianization and Islamization, fishermen caste). Trades with the rest of the world, especially during the colonization, were the most intense and the progress of education was much faster (Guilmoto, 2007). The highest fertility rates are recorded mainly in the districts located in the northern half of the country (except in Jammu Kashmir state, particularly in the Leh district): districts in the states of Rajasthan, Uttar Pradesh Meghalaya, Madhya Pradesh, Bihar and especially in the Jaisalmer's district (5.79) in the state of Rajasthan, where the major campaigns of birth control, beginning in the late 1960s⁹, had little influence on demographic behavior in India.

The social and cultural diversity that characterize the Indian population explain that the introduction of new policies on birth (spacing and birth control) can not be done uniformly by the states. For example, Hindus hardly conceive a family without at least

¹⁰ In an attempt to curb population growth, India has held numerous political birth since the mid-twentieth century, starting with the famous slogan "Hum do Hamare Do " ("We're two, we do have two") in 1960 (Oyarzun, 2010).

Long-Term Spatial Associations and Regional Convergence of Rural Literacy Rate in India

two sons, and Muslims are opposed to the family planning. Besides, their beliefs are rejected and insulted to further reduce their minority. The values of fertility are almost homogenous in some states like Uttar Pradesh (it has the largest number of districts, 70 districts in 2001), Bihar, Kerala, and Tamil Nadu. Conversely, states as Jammu and Kashmir, Orissa, Andhra Pradesh, and Maharashtra are particularly heterogeneous. Generally, the distribution of fertility shows that 17 districts have below 50 percent of the national average (3.30), 136 have a level between 50 percent and 75 percent, 159 between 75 percent and 100 percent. Finally, 279 districts have levels above the national rate. Note that the highest fertility rate (1.75) is 4.36 times higher than the lowest one (0.40). Conversely, modernization" explained by the progress of education and urbanization, can directly affect the fertility of the regions. This can create new forms of regional disparities since the access to education; the improvement of infrastructures and the degree of urbanization are related to the economic development of regions.

In fact, in addition to the economic repercussions linked to the level of education (employment and income opportunities), the education of women contributes to changing attitudes towards the sociological and psychological conditions that perpetuate poverty (UNDP, 1997). It is a factor that strongly contributes to emancipation and women's empowerment (Roser, 2003). In the majority of Indian regions, a high level of education of women has a direct impact on fertility. In fact, the heterogeneity of Indian states and their individual autonomy translates into a diversity of statistics. In 1956, the scalar curriculum reform was enacted to promote gender equity (Roser, 2003). Literate women are better able to make responsible decisions in matters that affect themselves, their families, and children, including health practices, household income, education and household well-being. their children 15. Thus, the research results demonstrate that a child's ability to develop properly is closely related to the educational level of its mother (UNESCO, 2006).

The robust approach to the specification of spatial models shows that the conditional convergence model with spatial lag is the most adequate. This model is defined by:

$$\begin{cases} GRRL = \alpha S + \rho GRRL + \beta \text{LogRL91} + \delta GRFER + \varepsilon \\ \varepsilon \sim N(0, \sigma^2 I) \end{cases} \quad (4)$$

Taking into account the fertility rate as an explanatory variable, Table 4 (2nd column) shows that the problem of heteroskedasticity is eliminated: the Breush-Pagan test is no longer significant with a confidence level of 95%. The regional convergence hypothesis is verified since the parameter β (-0.043) is significantly negative. The speed of convergence and the half-life are respectively equal to 5.83% and 15.34 years, which proves that the introduction of the fertility variable in the spatial-lagged convergence model accelerates the catching-up of the districts. As for the previous model, a positive spatial autocorrelation is observed: the parameter is estimated at 0.452. The parameter associated with the fertility growth rate is significantly negative with a confidence level of 99%. Also, the increase in literacy rates is favored in the districts experiencing weak growth in fertility rates over the period 1991-2001. Reducing unwanted pregnancies, especially among adolescent girls, increases education and employment opportunities. Generally, limiting the number of children allows parents to invest more in the education of children.

III- MORAN'S SCATTER PLOT: A NEW STATIONARY REPRESENTATION

1- The dynamics of spatial associations: new approach

The Moran index is the most robust statistic for testing the presence of global spatial autocorrelation even though it does not give any indication of the potential form of spatial autocorrelation. Unlike the other tests, it is not affected by a misspecification of the spatial weight matrix. Moran's statistic (I) is defined by:

$$I_M = \frac{N}{K} \frac{\sum_i \sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2} \quad (6)$$

where, for 2001 as for 1991, x_i denotes the value of the literacy rate in the rural district, \bar{x} is the general average, N is the number of rural districts, w_{ij} measures the intensity of the spatial interaction which exists between the two districts i and j and $K \sum w_{ij}$ is the sum of the interaction coefficients. With the matricial form, the Moran index is defined by:

$$I_M = \frac{ZW^{(sd)}Z'}{ZZ'} \quad (7)$$

where Z is the vector of rural literacy rates defined by their deviation from the national average and $W^{(sd)}$ is the spatial weight matrix in standardized form. Weight matrices are usually standardized in rows to facilitate interpretation and comparison of spatial parameters.

The Moran diagram (Figure 2) is a graphical representation showing on the abscissa the standardized literacy rate of a district and on the ordinate its spatial lag (also standardized). This diagram has four different quadrants. Each quadrant corresponds to a local spatial association between a district and its neighbors. The EE quadrant includes districts that have high literacy rate values

Long-Term Spatial Associations and Regional Convergence of Rural Literacy Rate in India

surrounded by districts that also have high values. The FF quadrant represents districts with low values surrounded by districts with also low values. The EF quadrant groups districts with high values surrounded by districts with low values. Finally, the FE quadrant contains districts with low values surrounded by districts with high values. Quadrants EE and FF refer to positive forms of spatial autocorrelation. In the EF and FE quadrants are grouped the atypical values where the spatial autocorrelation is negative.

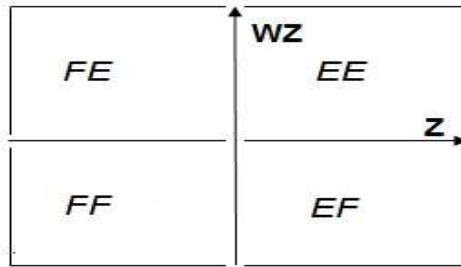


Figure 2: Moran Diagram

At each period, we can construct a diagram that can reveal the form of the spatial associations through the distribution of regional units in the four quadrants. In order to reveal the dynamics of these spatial associations, we propose a stationary form for the Moran diagram. To do this, we use the transition matrix constructed from the distributions of the initial year and the final year by proceeding by the Markov chain method. Next, we determine the form of the resulting long-term spatial associations: global convergence, polarization (local convergence into two spatial regimes) or stratification (local convergence with three or more spatial regimes).

Recall that for a set of m spatial associations, a Markov chain is said to be irreducible if all the groupings form a single equivalence class, each association is accessible from any other association. Therefore, the chain has a unique stationary distribution

$\Pi = (\Pi_1 \dots \Pi_i \dots \Pi_m)$ as a solution of the following system:

$$\begin{cases} \Pi A = \Pi \\ \sum_{i=1}^m \Pi_i = 1 \end{cases} \quad (8)$$

where Π_i denotes the proportion of regional units in the grouping i and A is the matrix of transitions defined by:

$$A = \begin{pmatrix} e_{11} & \dots & e_{1i} & \dots & e_{1m} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ e_{i1} & \dots & e_{ii} & \dots & e_{im} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ e_{m1} & \dots & e_{mi} & \dots & e_{mm} \end{pmatrix} \quad (9)$$

where e_{ij} is the percentage of regional units that transited from state i at period t to state j at period $t + 1$. The elements of the main diagonal (e_{ii}) indicate the percentage of regional units that have kept their initial state between t and $t + 1$.

To determine the stationary distribution, consider:

$$(\Pi_1 \dots \Pi_i \dots \Pi_m) \begin{pmatrix} e_{11} & \dots & e_{1i} & \dots & e_{1m} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ e_{i1} & \dots & e_{ii} & \dots & e_{im} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ e_{m1} & \dots & e_{mi} & \dots & e_{mm} \end{pmatrix} = (\Pi_1 \dots \Pi_i \dots \Pi_m) \quad (10)$$

$$\sum_{i=1}^m \Pi_i = 1$$

or,

$$(\Pi_1 \dots \Pi_i \dots \Pi_m) \begin{pmatrix} e_{11} & \dots & e_{1i} & \dots & e_{1m} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ e_{i1} & \vdots & e_{ii} & \dots & e_{im} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ e_{m1} & \dots & e_{mi} & \dots & e_{mm} \end{pmatrix} - \begin{pmatrix} 1 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & 1 & \\ & & & & 1 \end{pmatrix} = (0 \dots 0 \dots 0) \quad (11)$$

and $\sum_{i=1}^m \Pi_i = 1$

which give,

$$(\Pi_1 \dots \Pi_i \dots \Pi_m) \begin{pmatrix} e_{11} - 1 & \dots & e_{1i} & \dots & e_{1m} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ e_{i1} & \vdots & e_{ii} - 1 & \dots & e_{im} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ e_{m1} & \dots & e_{mi} & \dots & e_{mm} - 1 \end{pmatrix} = (0 \dots 0 \dots 0) \quad (12)$$

and $\sum_{i=1}^m \Pi_i = 1$

or,

$$(\Pi_1 \dots \Pi_i \dots \Pi_m) \begin{pmatrix} e_{11} - 1 & \dots & e_{1i} & \dots & e_{1m} & 1 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ e_{i1} & \vdots & e_{ii} - 1 & \dots & e_{im} & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ e_{m1} & \dots & e_{mi} & \dots & e_{mm} - 1 & 1 \end{pmatrix} = (0 \dots 0 \dots 0 \ 1) \quad (13)$$

and $\sum_{i=1}^m \Pi_i = 1$

hence, the following expression:

$$\begin{cases} \Pi M = V \\ \sum_{i=1}^m \Pi_i = 1 \end{cases} \Leftrightarrow \begin{cases} \Pi M M' = V M' \\ \sum_{i=1}^m \Pi_i = 1 \end{cases} \quad (14)$$

Where,

$$M = \begin{pmatrix} e_{11} - 1 & \dots & e_{1i} & \dots & e_{1m} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ e_{i1} & \vdots & e_{ii} - 1 & \dots & e_{im} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ e_{m1} & \dots & e_{mi} & \dots & e_{mm} - 1 \end{pmatrix} \quad \text{et } V = (0 \dots 0 \dots 1) \quad (15)$$

Finally, the long-term stationary distribution is defined by:

$$\Pi = V M' (M M')^{-1}, \sum_{i=1}^m \Pi_i = 1 \quad (16)$$

For the Moran diagram, the stationary distribution is defined

$$\text{by: } \Pi^{(s)} = \begin{pmatrix} \Pi_{EE}^{(s)} & \Pi_{EF}^{(s)} & \Pi_{FE}^{(s)} & \Pi_{FF}^{(s)} \end{pmatrix} \tag{17}$$

The stationary Moran diagram is represented by figure (3).

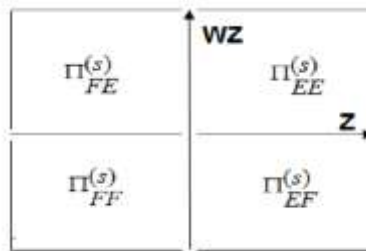


Figure 3: Stationary spatial associations

2- Application to literacy in rural Indian districts

We apply this nonparametric method to the same data used in the previous section. Using the inverse distance-based spatial weight matrix, Moran's statistics are equal to 0.2975 for 1991 and 0.2250 for 2001. This result shows the magnitude of spatial dependencies between rural Indian districts in terms of literacy rate. The very high significance of this test may indicate the existence of extreme points which may affect the value of the global spatial autocorrelation upwards or downwards, if it exists, between the districts. By means of the BoxMap (Figure 4), it appears that there are 9 observations for 1991, 0 for 2001, which can affect the spatial autocorrelation upwards. Thus, the Moran's statistic value cannot be strongly influenced by 9 particular districts, given the high number of districts considered (563), and therefore the global spatial autocorrelation should be considered as the dominant characteristic.

Table (5) shows that 67.99% of the districts fall under a spatial association of similar values of the literacy rate in 1991 and 72.72% of the districts in 2001 (quadrants EE and FF). On the

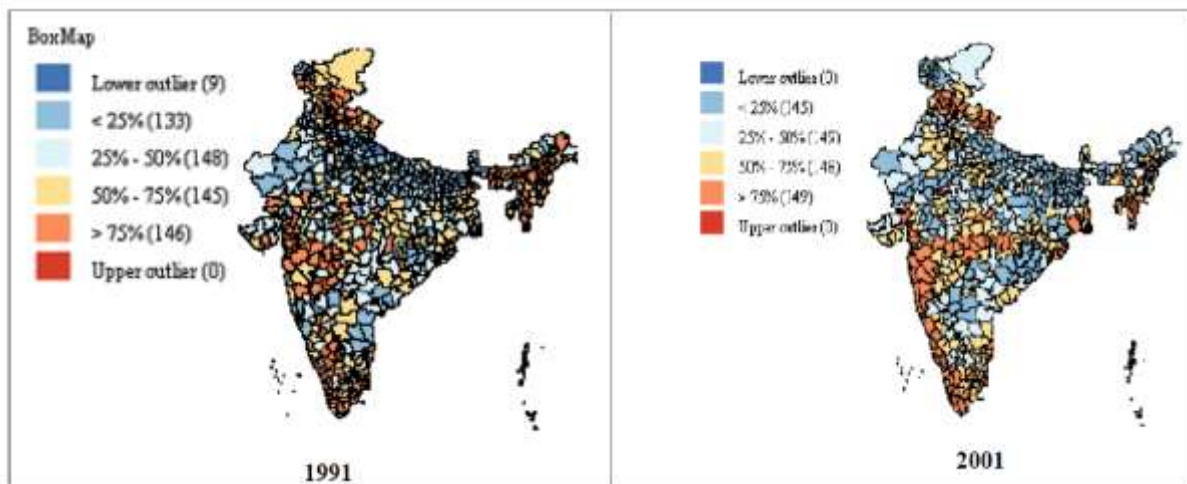


Figure 4: BoxMap: detection of extreme observations for 1991 and 2001

other hand, 32.01% in 1991 and 27.25% in 2001 of these districts are characterized by a spatial association of dissimilar values (quadrants EF and FE). However, it should be noted that the percentage of districts in the EE quadrant increased significantly between 1991 and 2001 to the detriment of the share of the other three quadrants and in particular of the FE and FF quadrants.

Table 5: Moran diagram: spatial associations of Indian rural districts in 1991 and 2001

Year	Quadrant EE	Quadrant EF	Quadrant FE	Quadrant FF
1991	(31.91%)	(15.90%)	(16.11%)	(36.08%)
2001	(39.02%)	(14.10%)	(13.18%)	(33.70%)

From there, the study of the dynamics of spatial associations can be carried out. In theory, each district in a quadrant can transit to a different quadrant. In this sense, we will say that the transition can be *horizontal* (LH-HH, HH-LH, LL-HL, HL-LL) where only the concerned district makes a movement towards a different quadrant, *vertical* (LL-LH, LH-LL, HL-HH, HH-HL) where only neighboring

Long-Term Spatial Associations and Regional Convergence of Rural Literacy Rate in India

districts change quadrants, *diagonal* (LL-HH, HH-LL, LHHL, HL-LH) where a parallel movement of the district and its neighbors is observed. A homogeneous transition can also be carried out by a district and its neighbors: they remain in the same quadrant during the same period (HH-HH, HL-HL, LH-LH, LL-LL). The dynamics of rural Indian districts over the period 1991-2001 in terms of literacy are summarized in Figure 5.

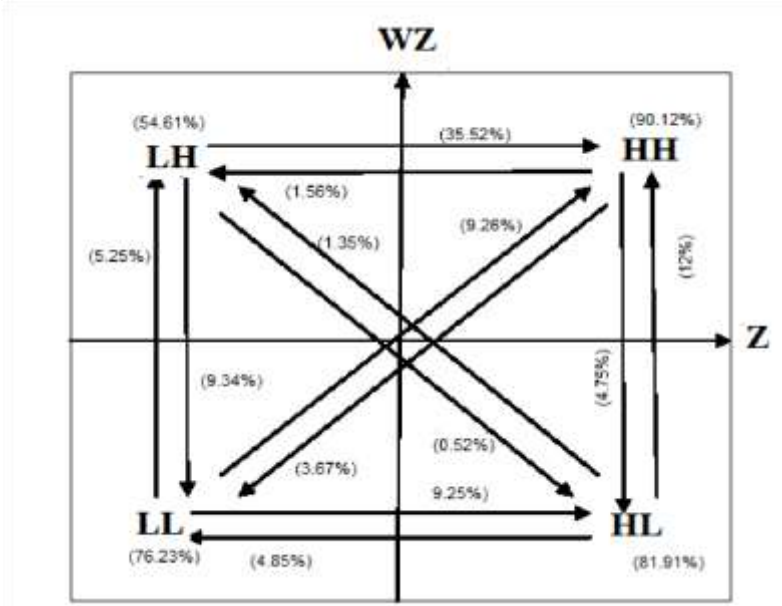


Figure 5: Dynamics of Indian rural districts in the Moran diagram between 1991 and 2001

By referring to this dynamic, we construct the transition matrix (W(T)) between 1991 and 2001 defined by:

$$W(T) = \begin{matrix} & \begin{matrix} \text{HH} & \text{HL} & \text{LH} & \text{LL} \end{matrix} \\ \begin{matrix} \text{HH} \\ \text{HL} \\ \text{LH} \\ \text{LL} \end{matrix} & \begin{pmatrix} 0.90125 & 0.04752 & 0.01562 & 0.03673 \\ 0.12 & 0.81912 & 0.01235 & 0.04852 \\ 0.35521 & 0.00521 & 0.54615 & 0.09341 \\ 0.09261 & 0.09251 & 0.05254 & 0.76233 \end{pmatrix} \end{matrix}$$

The first diagonal of this matrix shows the percentage of rural districts that have remained in the same quadrant. Thus, 90.1% of districts that were in the EE quadrant in 1991 were also in 2001. On the other hand, only 54.6% of districts with low literacy rates in 1991 surrounded by districts with high literacy rates remained in the FE quadrant. It therefore appears a high mobility of the districts and their neighbors in this quadrant. The EE quadrant received 56.8% of the movements made by the districts of the other three spatial associations. In addition, 14.52% of the transitions made by the districts of the three quadrants EE, FE and FF are attributed to the EF quadrant.

To reveal the shape of the long-term spatial associations in the four quadrants, we rely on the matrix of transitions defined above and we proceed by the Markovian method. Stationary distribution is the solution of the following system:

$$\Pi^{(s)} W(T) = \Pi^{(s)}$$

Or,

$$\Pi^{(s)} = V(W(T) - I)^{-1} [(W(T) - I)(W(T) - I)^{-1}]^{-1}$$

where V takes the same expression as in equation (15) and I is the identity matrix. Using the resolution method that we have developed, the stationary distribution of rural districts in the four quadrants of Moran is defined by: (s) (66.25% 9.52%

$$\Pi^{(s)} = (66.25\% \quad 9.52\% \quad 7.85\% \quad 16.38\%)$$

while the initial distribution was: $\Pi^{(I)}$ (31.91% 15.90% 16.11% 36.08%)

The stationary Moran diagram is given in figure 4.

Long-Term Spatial Associations and Regional Convergence of Rural Literacy Rate in India

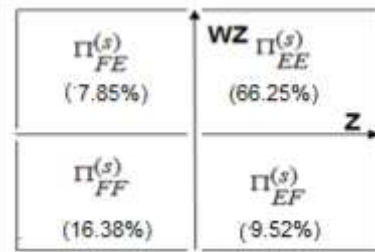


Figure 6: Stationary Moran Scatter plot

The stationary distribution shows that the percentage of rural districts that have high literacy rates surrounded by rural districts that also have high literacy rates (quadrant EE) has significantly increased at the expense of other spatial associations (quadrants EF, FE and FF, particularly the FF quadrant) while between 1991 and 2001 the decrease is particularly marked in the two quadrants FE and FF. This confirms the results resulting from the estimation of the convergence model integrating the effects of regional proximity. Beyond this correspondence between the two approaches, the new non-parametric approach has the advantage of identifying the districts in each spatial association.

IV-CONCLUSION

In this paper, we have developed a non-parametric approach to determine the shape of long-term spatial associations of georeferenced units. This method consists in giving a stationary form to the Moran Scatter plot, relying on the Markovian approach. From there, we have shown the correspondence that exists between this approach and the models of regional convergence. As an application, data on literacy rates according to the 563 Indian rural districts observed after the two censuses of 1991 and 2001 are used.

In the first part of the work, we estimated a β -convergence model of the literacy rate of rural districts over the period 1991-2001. The results obtained tend to confirm the hypothesis of β convergence. The spatially lagged model has been shown to be the most adequate to describe the convergence process and reveals that the growth in literacy rates proceeds through a certain diffusion between areas, explained by their spatial proximity. The fertility rate was considered as a conditional explanatory variable in the convergence model. It thus appears that the convergence of literacy rates is favored for the districts which have experienced weak growth in fertility rates over the period 1991-2001.

In a second part of the work, the construction of the Moran Scatter plot for 1991 as for 2001 made it possible to reveal the geography of the level of literacy of the rural districts during this decade. Despite the disparities, the results obtained show that there is a strong global and local spatial autocorrelation expressed by the existence of a strong spatial association of similar values. The construction of the long-term Moran diagram shows that the percentage of rural districts that have high literacy rates surrounded by districts that also have high literacy rates (quadrant EE) has significantly increased at the expense of other spatial associations, particularly the FF quadrant where the literacy rate of the district and its neighbors is low. This conclusion is consistent with the results found by estimating the β convergence model in a spatial framework.

It appears that the long-term Moran Scatter plot and the lagged endogenous variable regional convergence model both incorporate regional proximity effects between rural Indian districts and lead to the same results. The new non-parametric approach brings the advantage of identifying the location of long-term spatial associations.

REFERENCES

- 1) Abreu M., de Groot H.I.F., Florax R.J.G.M. (2005). Space and growth : a survey of empirical evidence and methods, *Région et Développement*, 2005-21, 12-43.
- 2) Anselin L. (1990). Spatial Dependence and Spatial Structural Instability in Applied Regression Analysis, *Journal of Regional Science*, 30, 185-207.
- 3) Anselin L., Bera A. (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. Chapter 7 (pp. 237-289) in Aman Ullah and David Giles (eds.). *Handbook of Applied Economic Statistics*, NewYork: Marcel Dekker.
- 4) Arbia G., Basile R., Salvatore M. (2003). Measuring spatial effects in parametric and non parametric modelling of regional growth and convergence. Paper prepared for the UNU/WIDER Project Meeting on Spatial Inequality in Development Helsinki, 29 May 2003.

Long-Term Spatial Associations and Regional Convergence of Rural Literacy Rate in India

- 5) Bailey T., Gatrell A. (1995). *Interactive Spatial Data Analysis*, Longman, Harlow.
- 6) Barro R., Sala-i-Martin X. (1991). Convergence across states and regions. *Brooking Papers on Economic Activity*, 1, 107-182.
- 7) Basu K., Foster J. E. (1998). On Measuring Literacy', *The Economic Journal*, vol 108, No, 451, pp. 1733-1749.
- 8) Baumont C., Ertur C., Le Gallo J. (2006). Clubs de convergence et effets de débordement géographiques : une analyse spatiale sur données régionales européennes, 1980-1995, *Economie et Prévision*, n° 173, pp 111-134.
- 9) Becker G. S. (1981). *A Treatise on the Family*. Cambridge, MA, Harvard University Press.
- 10) Buisson A. (2007). Alphabétisation et éducation au Kérala, Inde, Thèse.
- 11) Cliff A., Ord K. (1981). *Spatial Processes, Models and Applications*. Pion, London.
- 12) Davor, M., Andrea, G., Zeljko, L (2013). Regional convergence in the European Union, new Member States and Croatia, *South East European Journal of Economics and Business*, Volume 8(1), 7-19.
- 13) Ertur C., Thiaw K. (2005). Analyse spatiale des disparités régionales dans l'Europe élargie, *Région et Développement*, 21, 65-92.
- 14) Davor M., Andrea G. N., Zeljko L. (2013). Regional Convergence in the European Union, new Member States and Croatia. *South East European Journal of Economics and Business*.
- 15) Ertur C., Le Gallo J. (2009). Regional growth and convergence heterogeneous reaction versus interaction in spatial econometric approaches, Chapter 19, p. 374-388 in Capello R., Nijkamp P. (Eds.), *Regional Dynamics and Growth*, Edward Elgar, 2009.
- 16) Ferrer, E (2017). Regional convergence and productive structure in Iberian regions: A spatial approach, *Revista Portuguesa de Estudos Regionais*, n° 47.
- 17) Florax R., Folmer H., Rey S. (2003). Specification Searches in Spatial Econometrics: The Relevance of Hendry's Methodology. *Regional Science and Urban Economics*, 33, 557-579.
- 18) Guilmoto C. (1997). La géographie de la fécondité en Inde (1981-1991). *Espace, Populations. Sociétés*, 145-159.
- 19) Guétat, I. and F. Serranito (2010), "Convergence et Rattrapage Technologique : Un Test par les Séries Temporelles dans le Cas de Pays de la Région MENA », *Revue d'Economie du Développement*, 2 : 5-45.
- 20) Guilmoto C. (2005). L'Inde et son milliard. Démographie en début du siècle, Document de recherche n° 8, Série Santé de la reproduction, fécondité et développement.
- 21) Haining R. (1990). *Spatial Data Analysis in the Social and Environmental Sciences*, Cambridge University Press, Cambridge.
- 22) Hammouda, H., S. Karingi, A. Njuguna and M. Sadni-Jallab (2009) "Why Doesn't Regional Integration Improve Income Convergence in Africa?", *African Development Review*, 21(2): 291-330.
- 23) Jayati. G, Pal. P (2007). Inequality in India: A survey of recent trends. United Nations, department of Economics and Social Affairs, N 45.
- 24) Kosfeld, R. and J. Lauridsen (2009) "Dynamic Spatial Modelling of Regional Convergence Spatial Econometrics: Methods and Applications, pp. 245-61, Studies in Empirical Economics. New York: Springer, Physica.
- 25) Le Gallo J. (2002). Econometrie spatiale : L'autocorrélation spatiale dans les modèles de régression linéaire, *Economie et prévision*, 155, 139-157.
- 26) Mankiw N., Romer D., Weil D. (1992). A contribution to the empirics of economic growth. *Quarterly journal of economics*, 107, 407-437.
- 27) Oliveau S., Chasles V. (2005). Sur les chemins de l'école : accessibilité et alphabétisation dans le monde rural sud-indien, *Espace, Population, Sociétés*, 3, 453-464.
- 28) Oyarzun E. (2010). L'Inde, deuxième berceau du monde. Journée mondiale de la population : égalité entre les genres cette année, <http://inde.aujourdhuilemonde.com/l%E2%80%99indedeuxieme-berceau-du-monde>.
- 29) Rey S., Montouri B. (1999). US regional income convergence: a spatial econometric perspective, *Regional Studies*, Vol 32, pp 143-156.
- 30) Roser C. (2003). L'impact des politiques de scolarisation des filles, Mauritanie, Tunisie, Inde, Bangladesh et Sénégal, Institut National de Planification de l'Education, UNESCO.
- 31) Spoorenberg T. (2010). La transition de la fécondité en Inde entre 1977 et 2004. Analyse des probabilités d'agrandissement, *Population*, 2, vol. 65.
- 32) Upton G., Fingleton B. (1985). *Spatial Data Analysis by Example*, vol. 1, John Wiley, New York.
- 33) Véron J. (2006). Stabiliser la population de l'Inde : plus facile à dire qu'à faire, *Population et Sociétés*, Bulletin mensuel d'information de l'institut national d'études démographiques.

Long-Term Spatial Associations and Regional Convergence of Rural Literacy Rate in India

- 34) Zachary. S, Harper. A. (2013). Spatial Econometric Analysis of Regional Income Convergence: The Case of North Carolina and Virginia. *Research in Business and Economics Journal*.
- 35) Zakir H., Mousumi D., Sriparna G. (2011). Contraceptive use among illiterate women in India: does proximate illiteracy matter? Institute of Economic Growth, Delhi, Presidency University, MPRA Paper N 3090.

Figure 7: Map of Indian States

