Inter-District Disparity of Socio-Economic Development in West Bengal: Does Space Matter?

By

SOHINI MUKHERJEE

Roll No. 001700303006

Registration No. 142300 OF 2017-2018

Examination Roll No. MPEC194006

Session: 2017-19

Dissertation submitted to the Jadavpur University, West Bengal for the Degree of Master of Philosophy in Economics, 2017

Supervised by:

Dr. Sushil Kr. Halder

Professor

Department of Economics

Jadavpur University

Kolkata- 700032

Certified that the thesis entitled, “Inter-District Disparity of Socio-Economic Development in West Bengal: Does Space Matter?”, submitted by me towards the partial fulfilment of the Degree of Master of Philosophy (Arts) in Economics of Jadavpur University, is based upon my own original work and there is no plagiarism. This is also to certify that the work has not been submitted by me for the award of any other degree/diploma of the same Institution where the work is carried out, or to any other Institution. A paper out of this dissertation has also been presented by me at a seminar/conference at Economics Department, Jadavpur University on 26.04.2019, thereby fulfilling the criteria for submission, as per the M.Phil Regulation (2017) of Jadavpur University.

SOHINI MUKHERJEE
ROLL NO. 001700303006
REGISTRATION NO.142300 OF 2017-2018
(Name of the M.Phil Student with
Roll number and Registration number)

On the basis of academic merit and satisfying all the criteria as declared above, the dissertation work of Sohini Mukherjee entitled Inter-District Disparity of Socio-Economic Development in West Bengal: Does Space Matter? is now ready for submission towards the partial fulfilment of the Degree of Master of Philosophy (Arts) in Economics of Jadavpur University.

Head Department of Economics	Supervisor & Convener of RAC	Member of RAC

Dedicated to My Parents

Acknowledgments

I would like to express sincere gratitude to my supervisor, Prof. Sushil K. Haldar, Economics Department, Jadavpur University, Kolkata- 700032 for his full support, expert guidance, understanding and encouragement throughout the M.Phil study.

Without any hesitation and impatience, he continuously gave me proper time to build up better understanding in this subject. Sir always gave me proper space to clarify each and every doubt. Without this support, guidance and sincerity, this Dissertation would never complete. I have learnt a lot from Sir and also I am grateful to him for giving me a chance to work under his guidance. I am truly fortunate to have him as a mentor.

In addition, I give my heartiest thanks to the Head and other faculty members of the Department of Economics and my RAC member, Prof. Tanmoyee Banerjee for providing valued comments on my Dissertation.

Further, I am also thankful to all research scholars, especially Sulekha Hembram and Antara Bhattacharyya for their limitless academic help and support. I have presented few research papers (out of my M.Phil Dissertation) in different seminars (at CESP-JNU, Economics Department-CU, Kalyani University) and I have received valued comments from the participants; I am thankful to them.

Finally, I would also like to thank my parents, sister and friends for their unconditional love and support.

SOHINI MUKHERJEE

May, 2019

CONTENTS

Page No.

Chapter- I

INTRODUCTION

Chapter -II

TRENDS, PATTERNS AND INEQUALITY OF SOCIAL DEVELOPMENT INDEX AND PCDDP OF WEST BENGAL: AN INTER-DISTRICT ANALYSIS

Chapter-III

SOCIAL DEVELOPMENT AND SPACE: A DISTRICT LEVEL ANALYSIS IN WEST BENGAL

LIST OF TABLES

LIST OF FIGURES

LIST OF MAPS

Introduction

1.1 Background of the Study

Economic growth measured by Per Capita Income (PCI) fails to capture some of the important issues like distribution of the resources, health and educational outcome of the people, gender imbalance in respect of income earning, labour force participation etc. Higher economic growth does not necessarily imply higher social development. Now, by the term social development, we generally mean the non-monetary measure of social progress. Therefore, indicators which are used to measure the social development are called social indicators of development. Social indicators may be different types- some indicators are used to measure the health status of community and some are used to measure the education and cultural progress of the society. Gender differential in work force participation, health status and educational attainment belong to socio-cultural progress; gender equality in various aspect of life though broadly belongs to cultural development of a society; therefore, we can include this as social progress.

Human Development Index (HDI) was introduced in 1990 by the United Nations Development Programme (UNDP) and HDI has become widely used to measure socio-economic development of a country. The Social Progress Index (SPI) captures people's quality of life and wellbeing of the society, independent of wealth; about fifty-four indicators in the areas of basic human needs, foundations of well-being and opportunity to progress show the relative performance of nations. The index is published by the non-profit Social Progress Imperative and is based on the writings of Amartya Sen, Douglass North, and Joseph Stiglitz. The SPI measures the well-being of a society by observing social and environmental outcomes directly rather than the economic factors. But there is a complication to measure all these dimensions. More generally, HDI actually shows the performance of well-being of a country against other countries by ranking them in respect of basic capabilities.

However, per capita income cannot fully reflect the development level of a country because it ignores the welfare of a society and human being. Since, higher national wealth of a country does not necessarily imply lower human poverty; similarly HDI ignores the problem of inequality which occurs due to misdistribution of national income. For example, a country may suffer from malnutrition and lack of educational facilities but the status of gross national income may be high enough but might bypass human development.

Though, HDI is an appropriate measure of human development but it does not specify the cultural progress along with gender sensitivity. Culture may vary within or between countries; for example, in India, there exists wide differences of culture between Northern and Southern India. More specifically, the gender differential outcomes in respect of health, survival and education differ across regions in India; this is due to differences of religion, belief and trust. In India, there are different communities and different communities have different culture; the position of female in north-east states (Assam, Manipur, and Arunachal Pradesh etc.) differ from and north-western states (Punjab, Kashmir, Haryana, Rajasthan and Gujarat) in India. So, the cultural dimension varies across societies and again there is a neighbourhood impact of cultural transmission.

It can be argued that more or less same region may form a homogeneous group in respect of socio-cultural development and they are also related to each other by spatial autocorrelation. Social and cultural factor influence the health. There is a neighbourhood effect of health seeking behaviour. Exposure and vulnerability to disease, risk-taking behaviour, the effectiveness of health promotion efforts and access to medical facilities and availability of and quality of health care are influenced by space. Understanding spatial variation of health outcomes like morbidity, survivals and mortality are important for policy intervention.

The origin of the idea of cultural values came from Adam Smith (1759). It includes the social norms, beliefs, ethics, morality and how economy reacts by changing the role of cultural. After that following the idea of Smith, Max Weber (1930) put the cultural dimension in a different way by using religions reaction. Following the same line of thinking, other economists like, Landes (2000), Sen (2002), Borttke (2001) and Grief (1994) have argued that the cultural improvement has been pre-requisite for ensuring long term development. But there is no such history or theoretical background explaining the effects of culture on social development.

As culture is one of the factors of development; in the same manner space and neighbour does matter towards diffusion and spill over of culture and development. If one region is more developed culturally; then this will influence its neighbouring region as well.

For example in case of Kerala, there is higher literacy rate for both male and female and this influences its neighbouring regions like Karnataka, Tamil Nadu and Andhra Pradesh. I am trying to capture such spatial dependence relating to socio-cultural development in the Districts of West Bengal.

In our study, the concept of ‘culture’ is carefully grounded in development outcomes. The concept of culture is complex and implies a dynamic and ever-changing process. Definitions and dimensions of cultural phenomenon belong to inter-cultural and intra-cultural variations and change that are most useful in understanding health and the mechanisms through which cultural phenomena influence health. In this study, we use a new concept like social development index (SDI) which differs from other studies but similar to HDI in respect of methodology.

1.2 Scope of the Study

Numerous studies have undertaken in West Bengal focussing mainly on income disparity across the districts, agricultural production, urbanisation, industrial production etc. But the comprehensive study of West Bengal based on human development is not done in earlier studies. First human development index (HDI) across districts in West Bengal was undertaken by UNDP in 2004 (West Bengal Human Development Report, 2004). It was observed that Burdwan, Hoogli and Haora are found to be better off whereas Uttar Dinajpur, Malda and Murshidabad are found to be backward in respect of HDI. Raychaudhuri and Haldar (2009) have studied the inter-district disparity of income over time among 17 districts of West Bengal; they have observed that the inequality trend in respect of income is consistent with physical infrastructure index but not social infrastructure. They find that both social and physical infrastructure indices are U shaped over time across districts.

But my study is different from others in respect of social development index (SDI) and per capita district domestic product (PCDDP). I use both District Level Household Survey (DLHS) and National Family Health Survey (NFHS) data in my analysis.

The distinct feature of my work is to capture spatial dimension of social and economic variables and their corresponding determinants. None of works is done earlier in the view of spatial dimension in West Bengal. So, my study is very different from other studies in West Bengal.

1.3 Major Research Questions

There are four main research objectives of my Study:

- The first objective is to formulate a new non-monetary measure of social progress, known as Social Development Index (SDI) across 17 districts for five time points corresponding to DLHS (I-IV) and NFHS (I-IV).
- The second objective is to study the distribution, dispersion and inequality of the SDI and income (PCDDP) among 17districts over five time points (corresponding to DLHS and NFHS-IV survey).
- The third objective is to study the relationship between (i) space and SDI; and (ii) space and income (PCDDP). This is captured by Global and Local Moran's Index using spatial weight matrix.
- The fourth objective is to study the determinants of spatial variations of SDI and PCDDP using Spatial Auto-Correlation (SAR) and Spatial Error Model (SER).

1.4 Data and Methodology

1.4.1 Data and Variables

The data on health variables (especially reproductive and child health) are drawn from DLHS (I-IV) and NFHS (IV) surveys:

Health variables like Percentage of woman who have less than three order births(HOB<3), percentage of Institutional birth delivery(SID), percentage of woman who had 3 times ANC visits(PWANC), percentage of children who are fully immunized (PCCI) and percentage of girls who got married after age 18(PWMA18) are drawn from DLHS (I-IV) and NFHS (IV) round surveys. The data on per capita district domestic product (PCDDP) are collected from Statistical Abstract of West Bengal. The net enrolment ratio at the upper primary level is drawn form DISE. The data of female literacy rate and combine literacy rate are estimated from Census 2001 and 2011, Government of India. Various physical infrastructures like, road length (PWD, Zila Parishad and Gram Parishad), net irrigated area, village electrified, number of banks etc. are drawn from Statistical Abstract of West Bengal. For social infrastructure like, number of primary school, number of upper primary school etc. are collected from DISE; health infrastructure like Sub-Centre (SC) and primary health centre (PHC) are drawn from Statistical Abstract of West Bengal.

1.4.2 Methods

I have used the following methodology or econometric techniques and statistical tools in my study. Here, in introductory chapter, I have just mentioned those very briefly; but detailed discussions on these techniques are mentioned in the respective chapters. The brief introductions of the methods used in my Dissertation are mentioned as follows:

A. Construction of SDI

Construction of Social Development Index (SDI) is similar to the methodology used in measuring HDI developed by UNDP. I construct three separate indices like health index (HI), education index (EI) and cultural index (CI) from a set of indicators; considering the fixed range of each indicator which is time and space invariant. Finally, I estimate SDI for each time point for each district using geometric mean of three dimension indices (health, education and cultural). This helps in measuring the progress of SDI over time. Moreover, unlike arithmetic mean of three separate dimensions, the geometric mean is more scientific because, we can estimate the marginal effect of HI, EI and CI. The marginal effect of each dimension index is not only depends on its own rather it also depends on other two dimensions also. Thus, we define Social Development Index (SDI) as:

Social Development Index (SDI)= $[HIxEIxCI]^{\frac{1}{3}}$ ……….(1.1)

The value of SDI varies from 0 to 1. The marginal effect of HI is:

$\frac{\delta SDI}{\delta HI} = \frac{1}{3} * HI^{\frac{-2}{3}} * (EI * CI)^{\frac{1}{3}}$. In the same manner, the marginal effect of EI and CI can be estimated.

B. Generalized Entropy Measure of Inequality:

To measure the inequality of SDI and PCDDP over time across districts, I have used Generalized Entropy (GE) measure of Inequality. GE has some advantages over Gini because GE captures different parts of distribution. Gini is distribution insensitive, this is why I use GE measures of inequality.

$$GE(\alpha) = \frac{1}{n(\alpha^2 - \alpha)} \sum \left[\left(\frac{X_i}{\overline{X}} \right)^{\alpha} - 1 \right] \text{..........(1.2)}$$

In this equation, α=0,1and 2.

C. Principle Component Analysis (PCA):

In order to construct physical and social infrastructure indices, I have used the Principle Component Analysis (PCA). The first PC is considered since it captures maximum proportion of variability in the data set. If we consider 8 factors, then the first PC can be written as:

$$PC_1 = W_{11}.P_1 + W_{12}.P_2 + W_{13}.P_3 + W_{14}.P_4 + W_{15}.P_5 + W_{16}.P_6 + W_{17}P_7 + W_{18}P_8 \text{........(1.3)}$$

Here, W_{ij} stands for factor loadings or weights. It basically represents correlation coefficient between PC and the original factor. For example, W_{11} is the correlation coefficient between P_1 and PC_1; the square of W_{11} means percentage of variation of P_1 captured by first PC, PC_1. In the same way, other factor loadings could be interpreted.

Since the factor loadings represent the correlation coefficients between PC (artificial variable) and original variables (P's), we go for pair-wise statistical testing using the following t test for determining influential factors/variables. This helps to construct the physical and social infrastructural stock index across districts over time. Since, PCs are extracted from correlation matrix, we use that correlation matrix derived from panel data regression. This means that our correlation coefficient matrix is independent over time and space.

$$t = \frac{r\sqrt{(n-2)}}{\sqrt{(1-r^2)}} \text{.......(1.4)}$$

D. Spatial Autocorrelation:

Further, we try to incorporate the role of spatial effects on SDI and PCDDP in the districts of West Bengal. This spatial effect can be measured by Moran's Index. Moran's Index can capture the spatial autocorrelation; in my case I examine whether SDI is space dependent or not. The aggregate or total effect is captured by Global Moran's Index (GMI) and second one is Local Moran's Index (LMI) which captures the spatial autocorrelation at the disaggregate level. Detailed methodological explanations pertaining to GMI and LMI are given in Chapter-III.

E. Spatial Regression:

In order to in depth study, I have used spatial regression technique towards capturing spatial autocorrelation and neighbourhood effect more prominently. I have used two models Spatial Lag Model (SLM) and Spatial Error Model (SEM). A detailed explanation of SLM and SEM is written in Chapter-IV.

1.5 Organization of the Chapters

On the basis of the overall background, I have developed three inter-related Chapters (excluding this one) to carry out my research. In Chapter-I, I have explained the background of this study; the major research objectives; the data and methodology. In Chapter-II, I have developed a new measure of social development which is non-monetary, known as Social Development Index (SDI). I have studied the trend, dispersion and inequality of SDI and per capita district domestic product (PCDDP) over time (1998-99 to 2015-16). In Chapter-III, I have studied the impact of space on SDI and PCDDP over time using Global and Local Moran's Index. In Chapter-IV, I have explored the determinants of SDI and PCDDP using Spatial Panel Data Regression method. Chapter-V deals with Conclusion and Policy relevance.

Chapter-II

Trends, Patterns and Inequality of Social Development Index (SDI) and Per Capita District Domestic Product (PCDDP) of West Bengal Over Time: An Inter-District Analysis

2.1 Introduction

The HDI is a number estimated by using relative distance method in three basic dimensions of human life like income (capturing access to resources), literacy (capturing knowledge dimension) and life expectancy at birth (which captures health dimension) but one can raise the two following questions:

a) What does this number mean?

b) Does this number really capture development of a nation?

Indeed, higher HDI of a country is always good and desirable; this means that the country is well-off on an average in three basic dimensions of human life. But, higher HDI does not necessarily imply lower multidimensional poverty or inequality in basic capabilities; this is because HDI is based on range equalization method. Higher HDI cannot ensure bliss. How far and to what extent the people in a country are deprived in basic dimensions of life is captured by human poverty index or multidimensional poverty index. Moreover, HDI is an average measure of social progress but it is gender insensitive. HDI is not non-monetary measure of social progress; the cultural dimension is missing inn HDI though literacy partially captures cultural aspect of the society. The position of female in social life needs to be incorporated in measuring social development.

If income is very high of a nation it does not indicate that the nation is fully developed in true sense of human capital. For example, in case of West Bengal, the HDI rank is very poor around 0.637 (2017) which indicates 28th position among 32 states including UTs. But in case of GDP per capita, West Bengal is quite better off and rank-wise it takes 6th position. Thus, we can see that though West Bengal is well off in respect income but its human capital (education and health) is not very satisfactory. So, higher income does not lead to have better human capital accumulation.

It is clear that education is an important aspect to measure human capability. To capture education it is necessary to improve literacy rate. In case of West Bengal, the literacy rate is not very high i.e. 77.08% (total literacy) whereas male literacy is 82.67% (Census 2011) and female literacy rate is 71.16% (Census 2011). It means that female education of West Bengal is not satisfactory; at the same time child education is also not very high.

Female education plays a significant role in societal long-run development; educated female may aware of their health and at the same time improves their productiveness. Health in general and reproductive and child health in particular is assumed to play a significant role towards achieving long-run sustainable development path.

2.2 Review of Literature

Sharma (1997) has used the concept of human development to study the pattern of socio-economic development in India. The HDI offers an alternative index income and neoclassical measure of 'consumer utility' by capturing levels of human development within countries and measuring relative socio- economic progress. As HDI is not properly gender-sensitive, it must include women's vital contribution to development. The UNDP has introduced the concept of gender development and gender empowerment index during mid 1990.

Estes (1997) has examined the social development trends in Europe over time, 1970-1994; empirical evidence provides that cluster of countries emerge which are similar to their socio-economic characteristics in Europe.

Vyasulu and Vani(1997) have studied intra-state disparity of development and deprivation in Karnataka by applying human development index and ranks among districts on the basis of HDI. They have found differential development outcomes measured in HDI among the districts in Karnataka and have found that low ranked districts are found to be backward in respect of child health, sanitation and lack of drinking water. The morbidity is high among the backward districts.

Antony, Rao and Balakrishna (2001) have found that health inequality, low standard of living and HDI do vary due to differences of demographic, socio-economic and nutritional status of population.

They have indicated that some indicators like life expectancy at birth, per capita income, and prevalence of contraceptive usage are good indicators to measure development.

Ntibagirirwa (2009) has claimed that the freedom of an individual and rational choice lead to economic growth and development. He observes by an example that, if Africans want to achieve the economic growth and development of the Western country, then they have to adopt their culture, values and belief at first. This is the main argument to develop by the proponents of the modernisation theory. In other words, he put, 'it is the totality of the values, norms, attitudes, beliefs of a society which shapes its social, political and economic organisation and inculcates a general feeling towards development'.

Shoham, Shoham and Malul (2011) have showed the relationship between a country‘s level of literacy and its national culture. This study also focuses on the two cultural variables like power distance (PD) and gender egalitarianism (GE) that leads to social development.

Natoli and Zuhair (2011) have given emphasis on social values, norms and non-monetary approach to development. However, many national index measure monetary measurement to capture social progress but recently, non-monetary approaches to progress measurement have been undertaken by the new economics foundation with its happy planet index (HPI) and the gross national happiness (GNH) initiated by the Centre for Bhutan Studies.

Mishra, Kristle Nathan (2014) tries to construct a multidimensional development and social inclusiveness by using the methodology of human development. The Human Development Index (HDI) is calculated using normalized indicators from three dimensions- health, education, and standard of living (or income) but this paper evaluates three aggregation methods of computing HDI using a set of axioms. The old measure of HDI taking a linear average of the three dimensions does satisfy monotonicity, anonymity, and normalization (or MAN) axioms. The current geometric mean approach additionally satisfies the axiom of uniformity, which shows unbalanced or skewed development across dimensions. They also propose an alternative measure of HDI which is the additive but inverse of the distance from the ideal method.

Wu, Fan and Pan (2014) try to employ a super efficiency model to evaluate the rationality of the HDI rankings of 19 evaluated OECD countries in 2009. Compared to the HDI rankings, the efficiency rankings measured by the super-efficiency model having two advantages: 1) they consider the inputs to generate the indicators for constructing the HDI, and decide the weights of inputs and outputs as endogenous and 2) by inputs provisions of super-efficiency model which evaluate whether the inputs are over-used (or under used) and provide the improvement path of each country's input variables. Empirical result shows that approximately, 75 % of the evaluated countries had rather different results in the efficiency rankings and the HDI rankings.

Sachs, Traub, Kroll, Dalacre and Teksoz (2016) construct new index, SDG index (Sustainable Development Goals) and compare between the ranking of SDG and HDI and find out some countries are lower and perform worse in SDG rank but in HDI they do well in meeting basic human needs. Again SDG Index and SDG Dashboards show that even many high-income countries fall far short of achieving the SDG as sustainable development including three pillars- economic development, social inclusion and environmental sustainability supported by good governance. It may be possible that high-income country (rich) with significant inequality and unsustainable practices.

Kapas (2017) has given the idea of how cultural values affect economic growth by inspiration of Adam Smith and other renowned economists and critically has assessed their perspectives of social capital, trust, respect, individual self-control and obedience. This paper has tried to summarize all the perspectives in detailed and assessment of the literature.

2.3 Major Findings and Research Gap

In HDI, income component is there but in Social Development Index (SDI), we do not include income. Most of the earlier studies have estimated the HDI; but studies on social development are extremely few. I could not find any study that takes into account the cultural variable. I have included those cultural variables which are gender sensitised. So, it is my opportunity to do this work on the basis of cultural aspect sensitised by gender. This is the first attempt of time series study on social development in West Bengal among districts.

2.4 Methodology and Data

2.4.1 Methods

i. Construction of SDI:

From the concept of HDI which is an average of three dimensions of human life i.e. longevity, knowledge and access to resource. It is a minimal measure; here we try to broaden the concept of development by incorporating cultural dimension(with a focus on gender issue) along with education and health. We deliberately exclude income with a view to understand the separate spatial and neighbourhood effect on SDI(Social Development Index) and income. Thus, or proposal SDI is purely a non-income measure of development, broadly social development that comprises three dimensions- health, education and cultural which are assumed to be highly inter-related. Thus, my proposal SDI is a non-income measure of social progress.

The following three sub-indices are used:
(a) Health Index (HI)
(b) Education Index (EI)
(c) Cultural Index (CI)

Health Index (HI) is comprised of the following indicators:

1. Percentage of woman who have less than three order births (HOBL3)
2. Percentage of Institutional birth delivery(SD)
3. Percentage of woman who had 3 times ANC visits(PWANC3)
4. Percentage of children who are fully immunized(PCCI)

Education Index (EI) considers the following indicators:

1. Combined literacy rate.(CLR) which is defined as $CLR = P_f . X_f + P_m . X_m$ where P_f and P_m are the population(excluding 0-6 years of age) proportion of female and male respectively; X_f X_m are the female and male literacy rate (in percentage) respectively.
2. Net enrolment ratio to upper primary school(NERUP)

Cultural Index (CI) consists of the following two indicators:

1. Percentage of girls who got married after age 18(GMA18)
2. Female literacy rate (FLR).

Justification of the variables:

Life expectancy at birth (LEB) is generally considered as the best indicator of health outcome; when life expectancy data were not available, sometimes infant mortality rate (IMR) or infant survival rate (ISR) is used towards comparing health outcomes across regions. But, in India the LEB or IMR is not available for five time points across districts.

This is why we have considered those health indicators which are readily available over five time points across districts. Only, reproductive and child health (RCH) outcomes data were available in the DLHS: I-IV and NFHS-IV across districts. This is why the above four RCH indicators were chosen. Higher fertility is detrimental to reproductive health of the mothers; similarly institutional delivery aims to secure health of the new born and her mother. Institutional delivery is considered as reproductive security. Complete immunization of the children and ante-natal care visits of pregnant mothers is expected to reduce IMR and maternal mortality rate (MMR). In order to capture knowledge dimension, we have considered combined literacy rate and net enrolment ratio at the upper primary level (viz. class VIII); lower primary enrolment is found to be high, in some districts it exceeds 100 due to mid-day meal programme! Net enrolment at the upper primary in most of the districts are found to be less than 80, this is due to dropout of children at the upper primary level.

The variables like Girls' marriage before attaining age 18 and Female literacy rate are very crucial in case of West Bengal. As per census 2011, 7.8% of the female married before attaining age 18 in West Bengal compare to all-India average 3.7%. A recent regression analysis on DLHS 4-unit-level data by Sen and Modak (2017) revealed for determining the probability of family bringing in an underage bride in the rest of India but economic factors and welfare schemes in the village played a vital role for this matter, it was not so in West Bengal. But most of the districts in West Bengal are not so developed and also a large amount of the districts are stuck into village. None of the economic factors affect the probability of a woman marrying before 18 in this state, suggesting that poverty is not responsible. It is also a culture basis as in Rajasthan one of the state in India, it is a compulsion for underage marriage both for boys and girls especially for girls. So there is another factor comes which is more dependent on society. Some how due to social pressure, the villagers have to do this type of social action which is very illegal. Mass education can prevent this social problem. A girl with higher education does have lower probability to get marriage before attaining age 18. So, West Bengal govt. takes some policies to prevent this problem and to influence in education give some incentives like Kannyashree. By this policy girls are more motivating to study especially up to class 12 as it is a preliminary level of study. Very small amount of money takes some incentive to family member of the girls as she get money on her own account. The assurance of Rs.25,000 discourages dropping out among girls at the secondary level and prompts them to defer marriage till the legally permissible age. So the girls are taking her own decision for her life in case of marriage purpose. This policy is found to be more successful all through the districts of West Bengal.

SDI Computation

We make the indicators scale free by way of relative distance methods. We compute the range for each indicator in each dimension (category). For the j-th indicator belonging to the r-th dimension (category), the range R^*_{rj} would be:

$\left(X_{\substack{rjMax \\ i,j}} - X_{\substack{rjMin \\ i,j}} \right) = R^*_{rj}$ where, $\left(X_{\substack{rjMax \\ i,j}} \right)$ and $\left(X_{\substack{rjMin \\ i,j}} \right)$ are the maximum and minimum values for the jth indicator in the r-th dimension (r=1, 2 and 3) over time(t) and space(district(i)). This helps to calculate the growth of SDI over time of any district.

Compute, $\left(X_{rji(t)} - X_{\substack{rjMin \\ i,t}} \right)$ and divide each value by R^*_{rj} , we thus obtain the scale-free values of the j-th indicator as:

$\dfrac{\left(X_{rji(t)} - X_{\substack{rjMin \\ i,t}} \right)}{R^*_{rj}}$ for all i (i=1,2……17) and t(1,2….5). We add the scale-free values of the indicators within each dimension for each district. Assume, Y_{ri} as the component index for the r-th dimension for the i-th district, thus we can write:

$$Y_{ri} = \sum \left(X_{rij(t)} - X_{\substack{rjMin \\ i,t}} \right) \div R^*_{rj} \text{............(2.1)}$$

Each j=1,2..Zr, where Z_r is the number of indicators in the r-th dimension. Finally, the SDI is the geometric mean of the previous three normalized indices; it helps us to measure the progress of SDI over time of a particular district (UNDP 2010).
Each indicator has calculated by using that indicator index and by applying simple arithmetic mean we get health index, education index and cultural index respectively.

$$HI = \frac{1}{N}\left(\frac{HOBL3_{i(t)(A)} - HOBL3_{i(MIN)}}{R^{*}_{HOBL3}} + \frac{SD_{i(t)(A)} - SD_{i(MIN)}}{R^{*}_{SD}} + \frac{PCCI_{i(t)(A)} - PCCI_{i(MIN)}}{R^{*}_{PCCI}} + \frac{PWANC3_{i(t)(A)} - PWANC3_{i(MIN)}}{R^{*}_{PWANC3}}\right)$$

where, N=4

$$EI = \frac{1}{N}\left(\frac{CLR_{i(t)(A)} - CLR_{i(MIN)}}{R^{*}_{CLR}} + \frac{NERUP_{i(t)(A)} - NERUP_{i(MIN)}}{R^{*}_{NERUP}}\right)$$

where, N=2

$$CI = \frac{1}{N}\left(\frac{GMA18_{i(t)(A)} - GMA18_{i(MIN)}}{R^{*}_{GMA18}} + \frac{FLR_{i(t)(A)} - FLR_{i(MIN)}}{R^{*}_{FLR}}\right)$$

where, N=2

Finally, by using those dimensions we calculate SDI by using geometrical mean following the methodology of HDI (UNDP, 1990).

Social Development Index (SDI)= $[HIxEIxCI]^{\frac{1}{3}}$ ……….(2.2)

Thus, SDI is a composite index measuring average achievement in three dimensions of social development that actually captures non-monetary measure of social progress. The index is best seen as a measure of people’s ability to live a long and healthy life, to communicate and participate in the life of the community. It purely considers non-monetary measure of social progress and it differs from HDI; because HDI takes into account the access to resource dimension but our SDI does not. The cultural dimension, though it is heavily influenced by education dimension is added here in measuring SDI. Also it focused mainly gender sensitise. So, the cultural variables matter in case of education and marriage. Especially by education girls can earn more knowledge what they choose in their life. Sometimes money or income does not as a barrier of education and to improve own selves by getting healthy and to know how they can improve their physical fitness as before attaining age 18 is not proper time to get married as they learn through education and their culture. So culture matters in a sense of how a girls progress her selves and also they aware of their health also. That's why we are taking cultural dimension which is different from income to study more properly as this study is more focus of female gender i.e. gender biased study.

ii. Generalised Entropy:

First to check the regional disparity or inequality of SDI and its components across the districts over five time points (viz. 1998-99, 2003-04, 2007-8, 2012-13 and 2015-16). The Mean Log Deviation (MLD or GE(0)), Theil Index (GE(1)) and coefficient of Variation (CoV or GE(2)) belong to Entropy(GE) class measure of inequality. The generalized form of Entropy class measure of inequality is shown by the following equation:

$$GE(\alpha) = \frac{1}{n(\alpha^2 - \alpha)} \sum \left[\left(\frac{X_i}{\overline{X}} \right)^{\alpha} - 1 \right] ....(2.3)$$

$\overline{X}$ is the mean of the variable (say, SDI or its components), α be the sensitivity parameters capturing different parts of distribution; *n* stands for number of observations. GE (0) gives more relative importance to the lower tail of the distribution, GE (2) gives relatively greater importance to the upper tail of the distribution, and GE (1) gives equal weights to both the tails (Litchfield 1999, Cowell 2003). If we directly put α=0 and 1, the function (equation 3) cannot be evaluated; this can be solved by using L' Hospital's rule. Therefore, for α=0, 1 and 2, the above equation is reduced to:

$$GE(\alpha)\Big|_{\alpha=0} = -\frac{\sum \ln\left(\frac{X_i}{\overline{X}}\right)}{n} = \frac{\sum \ln\left(\frac{\overline{X}}{X_i}\right).}{n} .....(2.4)$$

$$GE(\alpha)\Big|_{\alpha=1} = \frac{\sum \frac{X_i}{\overline{X}} \ln \frac{X_i}{\overline{X}}}{n} ......(2.5)$$

$$GE(\alpha)\Big|_{\alpha=2} = \frac{1}{2}\frac{Var(X)}{\overline{X}^2} = \frac{1}{2}.CoV^2 ....(2.6)$$

From equation (2.3), (2.4) & (2.5), we can see the lower, middle and higher districts respectively. One can roughly guess about disparity or inequality and distribution of the districts by way of estimating GE(0), GE(1) and GE(2) over time.

2.4.2 Data Sources:

Data pertaining to Health indicators are drawn from DLHS (I-IV) and NFHS-4 surveys. Upper primary enrolment is drawn from DISE Data; data on enrolment for 1998-99 was extrapolated from FLR. Per Capita DDP are drawn from Economic Survey, Govt. of West Bengal.

2.5 Analysis

I. Value of SDI:

The UNDP revised the methodology towards calculating of HDI. The position of HDI of West Bengal compared to all the states in India is 8 (1991). Just like HDI, I have estimated the Social Development Index (SDI) of 17 districts for five time points as shown in Table-2.1. Based on observed values of SDI, I have ranked the districts according to their values as shown in parentheses.

The relative positions of the districts in respect of SDI vary to a large extent over time; however, some districts like Howrah, Hooghly, Nadia, and North 24 Parganas are found to be better off in respect of SDI and their positions change marginally over time. The worst position in social development over time is found in Uttar Dinajpur throughout DLHS-1 to NFHS-4. The other bad performing districts are Malda, Murshidabad and Purulia. In case of Bankura, Birbhum, Coach Behar, they are in a fluctuating trend. Trend and position of districts change in case of NFHS-4; this is because the differences of same frame and sample design adopted in DLHS and NFHS.

Out of 17 districts, few districts have witnessed relative deterioration of ranking in respect of SDI- these are Darjeeling, Medinipur(combined) and South 24 Parganas (except DLHS-4). On the other hand, over time, few have performed well in raising relative ranks in SDI and they are Dakhsin Dinjpur, Coach Behar (except DLHS-2) and North 24 Parganas (except DLHS-4). Thus, we can roughly conclude that there may be some possibility of the existence of low level trap found in few districts like Uttar Dinajpur, Malda, Murshidabad and Purulia. That means that these districts are stuck into a social development trap.

Table: 2.1 Social Development Index (SDI) of 17 Districts over 5 Time Points

Districts	DLHS 1	DLHS 2	DLHS 3	DLHS 4	NFHS 4
Bankura	0.5449(9)	0.5796(6)	0.6366(12)	0.7213(10)	0.7585(5)
Bardhaman	0.5427(10)	0.5790(7)	0.6490(9)	0.6941(14)	0.7392(8)
Birbhum	0.5021(12)	0.5238(13)	0.5852(13)	0.7022(12)	0.7165(10)
D. Dinajpur	0.4882(13)	0.5548(11)	0.6483(10)	0.7491(7)	0.7259(9)
Darjeeling	0.5712(6)	0.5695(10)	0.6559(8)	0.7181(11)	0.7073(12)
Howrah	0.6308(3)	0.6737(1)	0.7110(4)	0.7859(2)	0.7908(3)
Hooghly	0.6304(4)	0.6627(2)	0.7358(1)	0.7702(5)	0.7820(4)
Jalpaiguri	0.5586(7)	0.5755(8)	0.6789(6)	0.7550(6)	0.6952(13)
Koch Bihar	0.5056(11)	0.5501(12)	0.6597(7)	0.7780(4)	0.7580(6)
Maldah	0.4004(15)	0.4424(16)	0.5318(16)	0.6970(13)	0.6149(16)
Mednipure	0.6469(1)	0.6584(4)	0.7084(5)	0.7805(3)	0.7562(7)
Murshidabad	0.4023(14)	0.4652(15)	0.5494(15)	0.6906(15)	0.6729(14)
Nadia	0.6328(2)	0.6622(3)	0.7219(3)	0.7880(1)	0.8043(1)
N24Parganas	0.6028(5)	0.6357(5)	0.7255(2)	0.7250(9)	0.7986(2)
Purulia	0.3711(16)	0.5204(14)	0.5587(14)	0.6884(16)	0.6611(15)
S24 Parganas	0.5480(8)	0.5729(9)	0.6367(11)	0.7255(8)	0.7158(11)
U. Dinajpur	0.3551(17)	0.4011(17)	0.5055(17)	0.6389(17)	0.5755(17)

Source: own estimation. Note: Values in parentheses represent rank

In the first time point (DLHS-1), Medinipur takes the first position but over time the position deteriorates, in NFHS it ranks 7^{th}. In case of Howrah and Hooghly, the position is quiet similar and they are in a good position. The positions of these districts are good because of agricultural development as well as industrial support. The position of credit facilities like cooperative banks as well as commercial banks is prominently noticed among these districts.

II. Dispersion of SDI among Districts over Mean overtime (1998-99 to 2015-16)

In this part, I have tried to explain the difference of SDI among the districts from the mean value of SDI for each time point(1998-1999, 2002-2004, 2007-2008, 2012-2013, 2015-2016). It is found that out of 17districts, mainly four districts are always showing a bad position which means they are lying below the mean value for each time point and these are: Uttar Dinajpur, Malda, Murshidabad and Purulia. At the same time there are some well off districts which lie above the mean consistently over the five time points; these are Howrah, Hooghly, Darjeeling, and North 24 Parganas. But the remaining districts are found to be fluctuating in respect of relative positions of SDI. I have shown this graphically Fig(2.1) to Fig(2.5) using radar diagrams as follows.

Fig. 2.1 Dispersion of SDI of Districts over Mean: 1998-1991(DLHS-1)

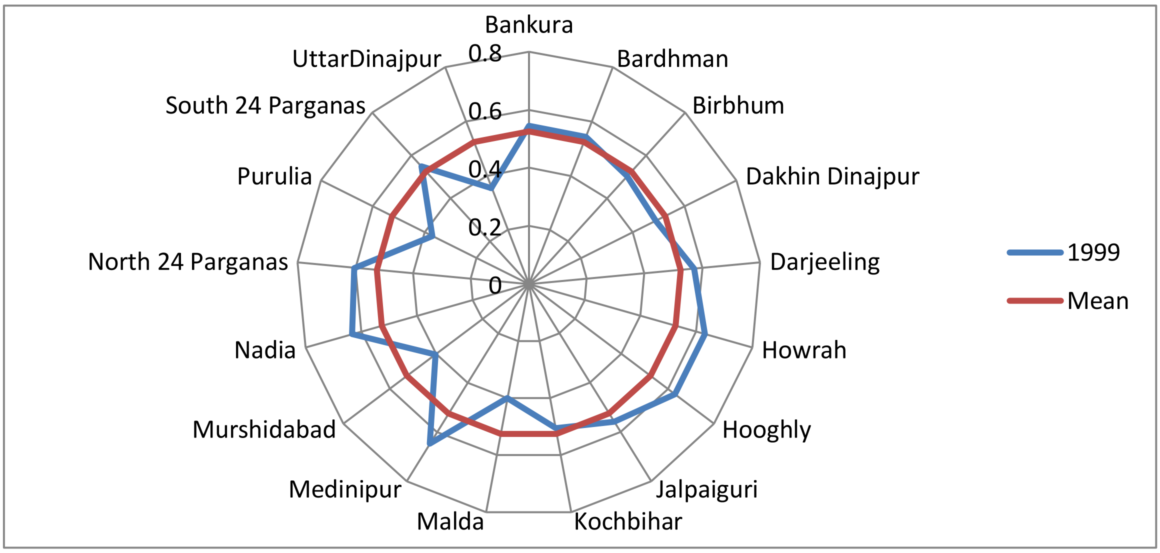


Fig. 2.2 Dispersion of SDI of Districts over Mean: 2002-2004(DLHS-2)

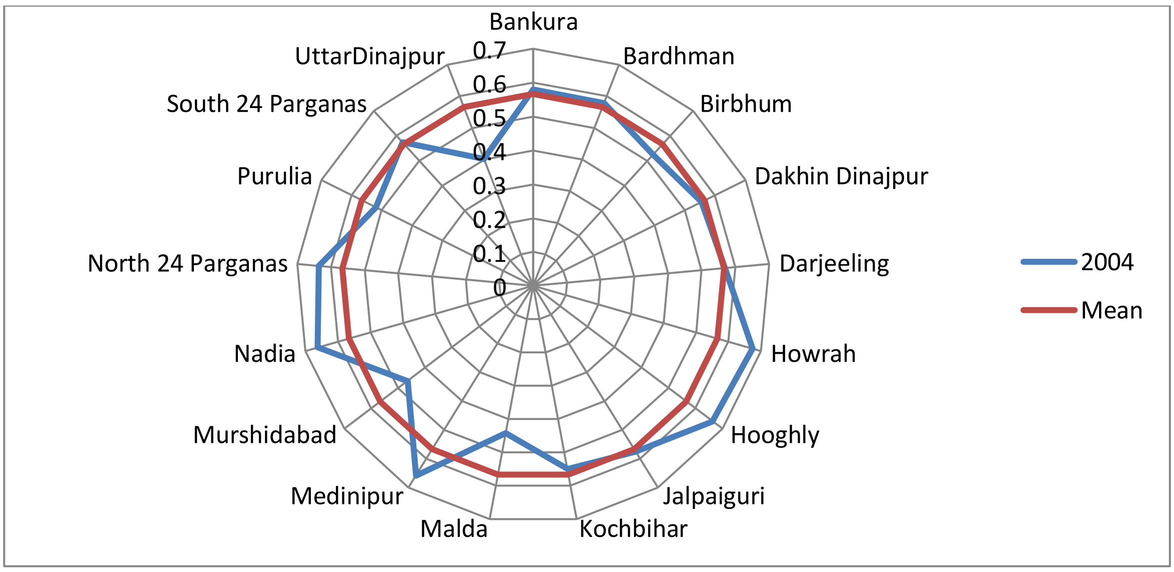


Fig. 2.3 Dispersion of SDI of Districts over Mean: 2007-2008(DLHS-3)

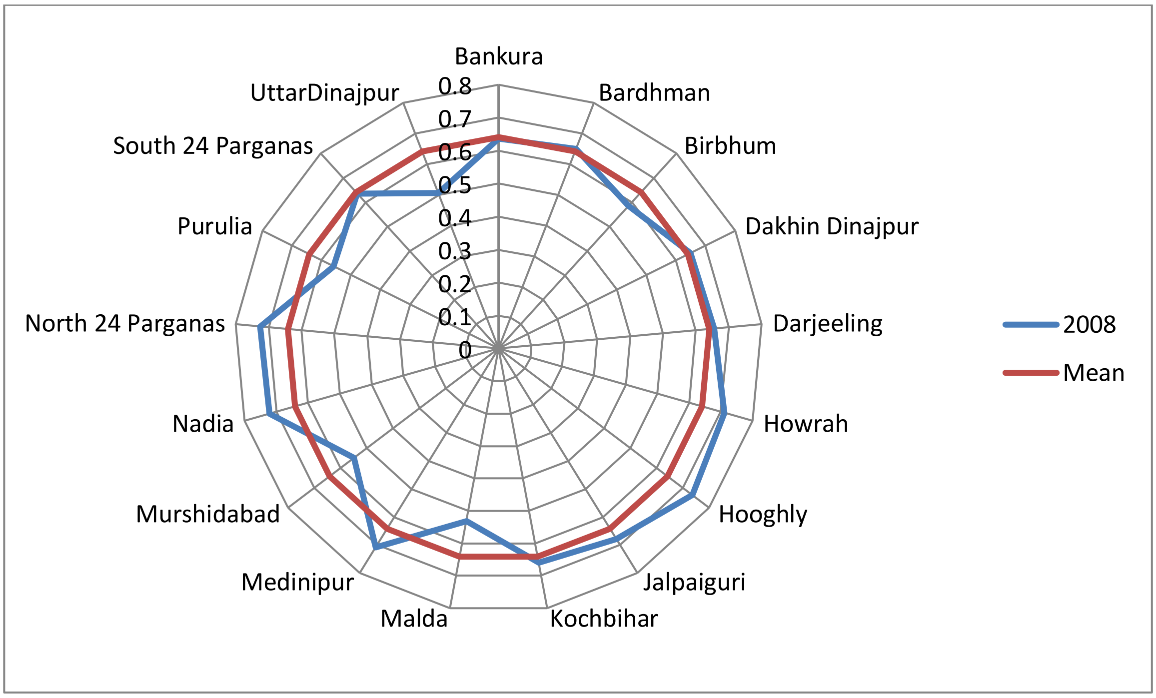


Fig. 2.4 Dispersion of SDI of Districts over Mean: 2012-2013(DLHS-4)

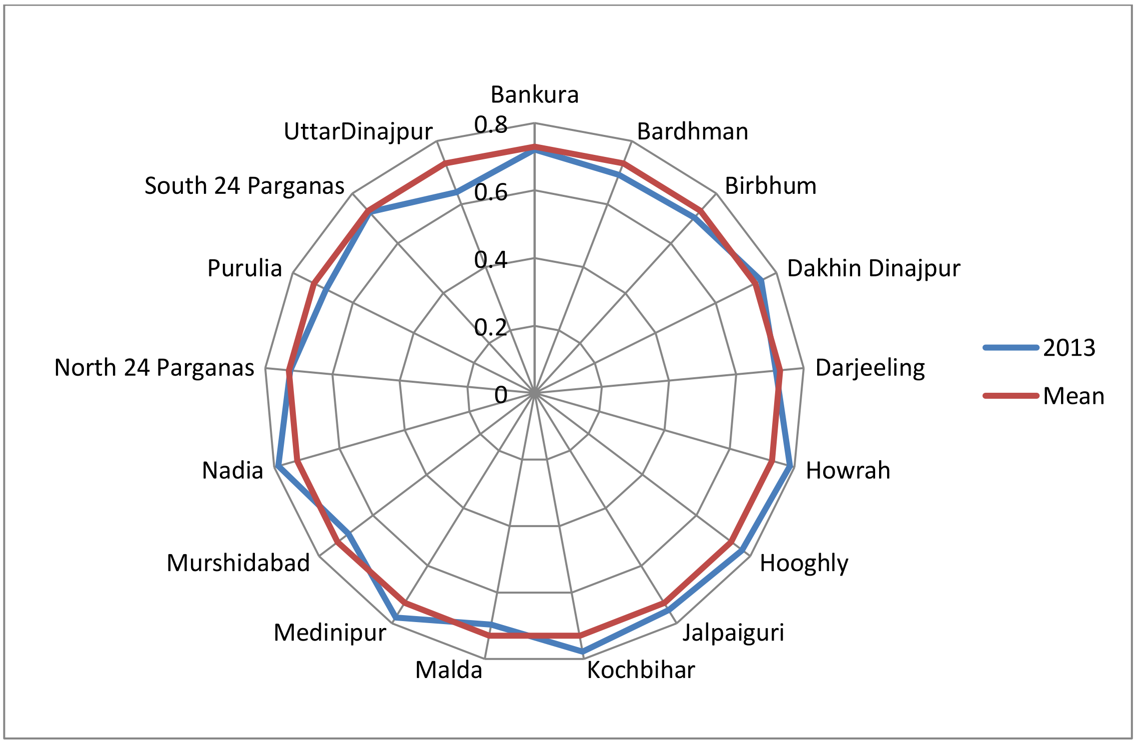
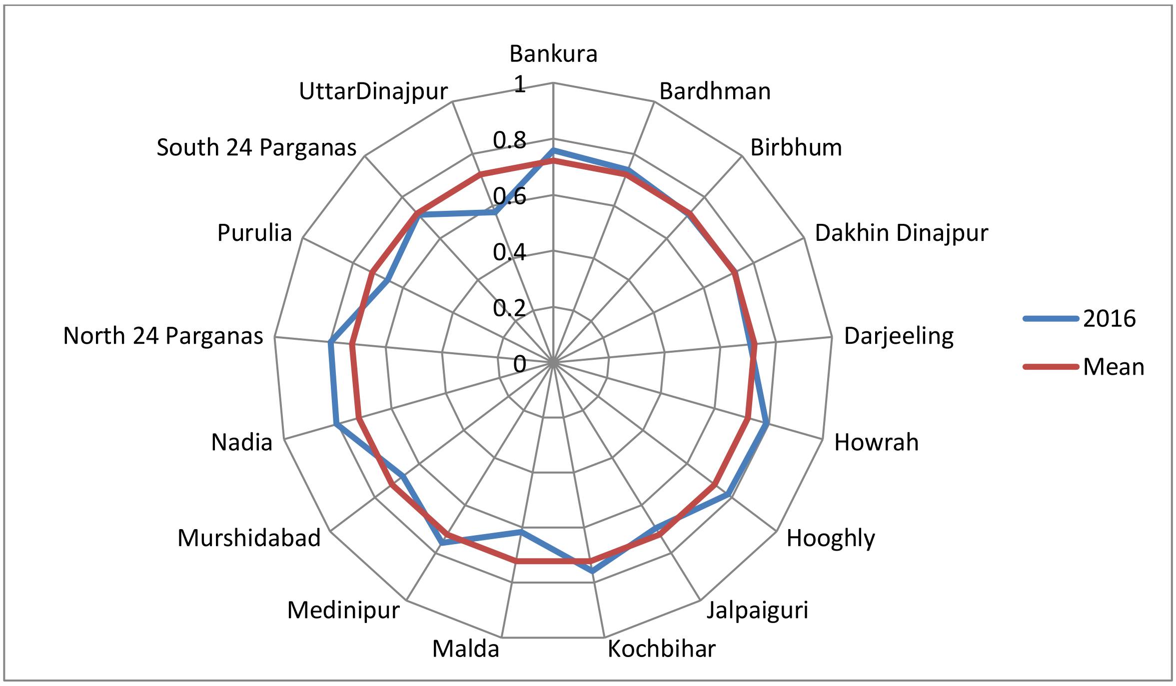


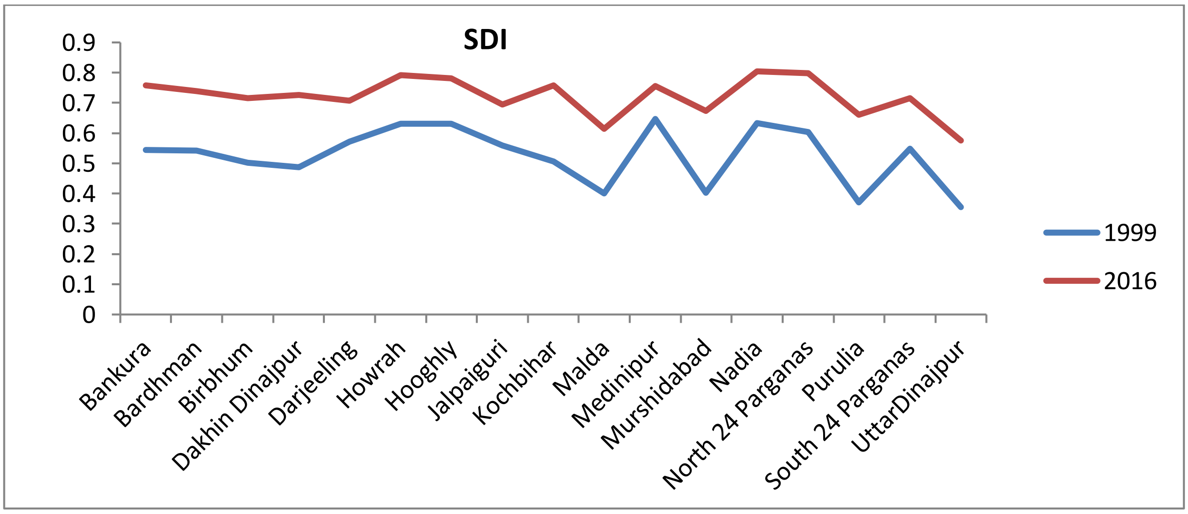
Fig. 2.5 Dispersion of SDI of Districts over Mean: 2015-2016(NFHS-4)



<u>Progress of Districts in respect of SDI between 1998-99 to 2015-16</u>

From the following diagram, we can see that all the districts have improved their position in respect of social development over two time points (1999 & 2016). An important point is noted here that some districts like Malda, Murshidabad, Purulia and Uttar-Dinajpur could not improve remarkably but some districts like Kochbihar and Jalpaiguri have fairly improved over time in SDI.

Fig.2.6 Trends of SDI of DLHS-I (1998-1999) and NFHS-4 (2015-16)



III. Value of PCDDP:

The values of PCDDP are collected from statistical survey for different time points. Here, I estimate the PCDDP at constant (base year 2000) prices over time; values in parentheses represent ranks.

Table:2.2 Per Capita District Domestic Product(PCDDP) over 5 time 5 Points

Districts	DLHS 1	DLHS 2	DLHS 3	DLHS 4	NFHS 4
Bankura	13008 (9)	17409.96(9)	26213.88(10)	50239.92(9)	52966.17(10)
Bardhman	16554.8(1)	23254.99(1)	36655.34(1)	72435.61(2)	76655.97(2)
Birbhum	12977.23 (10)	16747.28(11)	24287.38(12)	44252.39(12)	46642.14(12)
DakhinDinajpur	11743.5(14)	15458.87(15)	22889.61(15)	44194.33(13)	46403.4(14)
Darjeeling	15335.9(2)	22045.2(2)	35463.85(2)	76351.19(1)	80123.98(1)
Howrah	14207.3(4)	20441.95(4)	32911.31(4)	58177.69(6)	62824.33(6)
Hooghly	13274.16(6)	17741.82(8)	26677.15(9)	62011.81(4)	63797.39(4)
Jalpaiguri	13274.16(6)	17741.82(8)	26677.15(9)	62011.81(4)	63797.39(4)
Kochbihar	11920(13)	16014.46(14)	24203.37(13)	44904.64(11)	47588(11)
Malda	12249.55(11)	16134.27(13)	23903.7(14)	43964.72(14)	46472.97(13)
Medinipur	13248.24(7)	20488.6(3)	34969.33(3)	64234.71(3)	69637.74(3)
Murshidabad	12037.46(12)	16298.65(12)	24821.04(11)	43113.64(16)	46197.7(15)
Nadia	14218.3(3)	18551.56(6)	27218.07(7)	50802.19(8)	53491.89(9)
North 24 Parganas	13246.4(8)	19529.23(5)	32094.86(5)	58202.17(5)	62826.81(5)
Purulia	11175.87(16)	14961.87(16)	22533.88(16)	43395.31(15)	45722.4(16)
South 24 Parganas	11661.71(15)	17214.6(10)	28320.37(6)	49744.52(10)	53979.75(7)
UttarDinajpur	10434.99(17)	13233.26(17)	18829.81(17)	35018.77(17)	36669.73(17)

Source: Statistical Survey of West Bengal

From the above Table-2.2, we can see that all most all the districts are growing in respect of PCDDP but some districts have performed well and these are Bankura, Bardhman, Darjeeling, Howrah and Hooghly. It is noticed that few districts like Uttar Dinajpur, Murshidabad, Malda and Purulia could not increase their PCDDP at a faster rate as a result their relative positions are found to be low.

IV. Trend of SDI and PCDDP:

The numerical values of inequality of SDI (and its components) across the districts over five time points are produced in Appendix B (Table-2). The trend of GE(α) of SDI and PCDDP are plotted in Fig.2.7 and 2.8 respectively for five time points and I find a secular decline of inequality of SDI (except NFHS-4) but the trend of inequality of PCDDP is found to be erratic over time. Thus, I can say that the inequality of social development is independent of inequality of income movement; no co-movement is observed between SDI and income (PCDDP).

The inequality of all the components is found to be declining over time; only exception is observed in ante-natal care visits (PWANC3), it increases in DLHS-2 and NFHS-4. Since GE(2) is found to be higher than GE(0) and GE(1), I can say that over time all the districts are moving to the higher values of social development parameters in which inequality is more sensitive compared to middle and lower value of SDI!.

Fig. 2.7: Inequality Trend of SDI

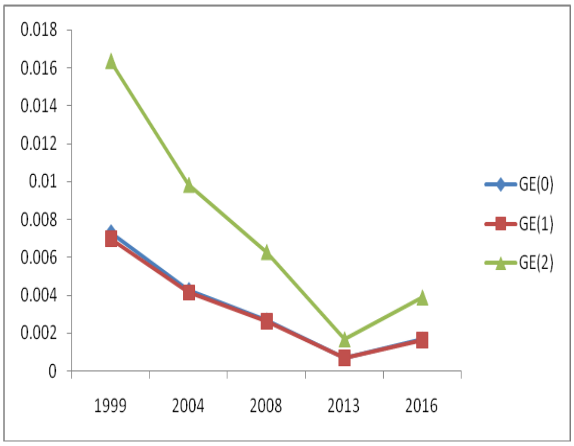
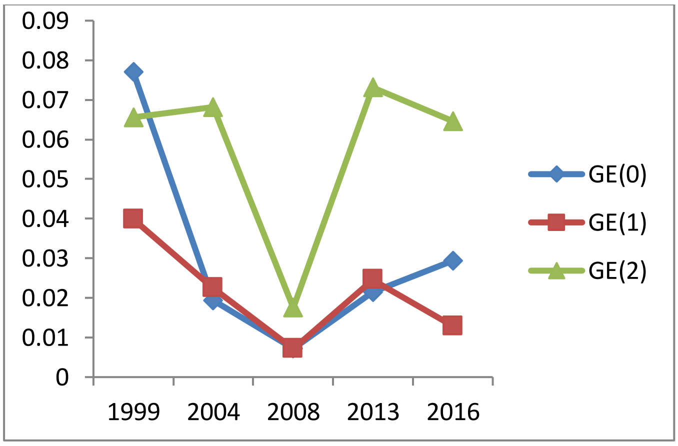


Fig. 2.8: Inequality Trend of PCDDP



V. Components of GE Measures:

A disaggregate analysis of inequality is studied in respect of the components of SDI as shown in Fig.(2.9) to (2.12).

Fig: 2.9: Inequality Trend of HOBL3 and SD

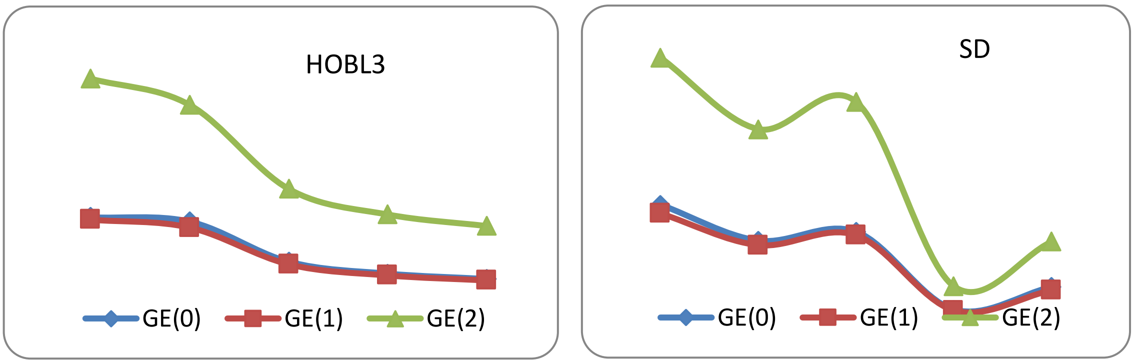


Fig: 2.10: Inequality Trend of PWANC3 and PCCI

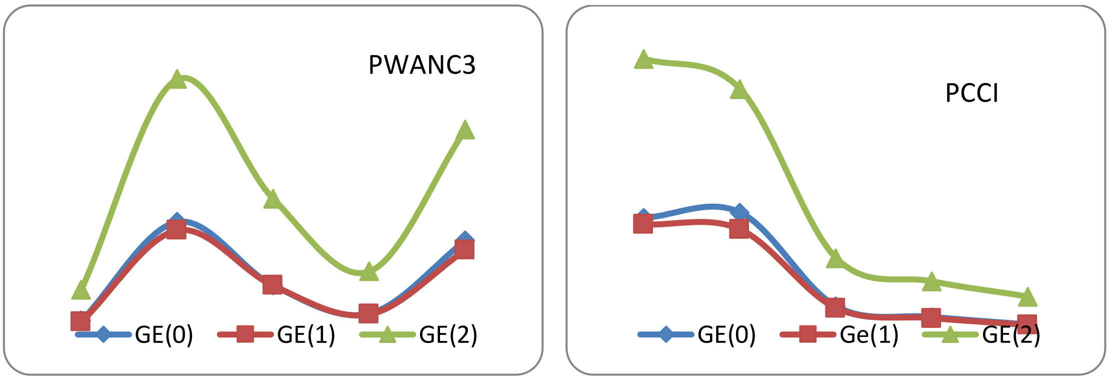


Fig. 2.11: Inequality Trend of CLR and NERUP

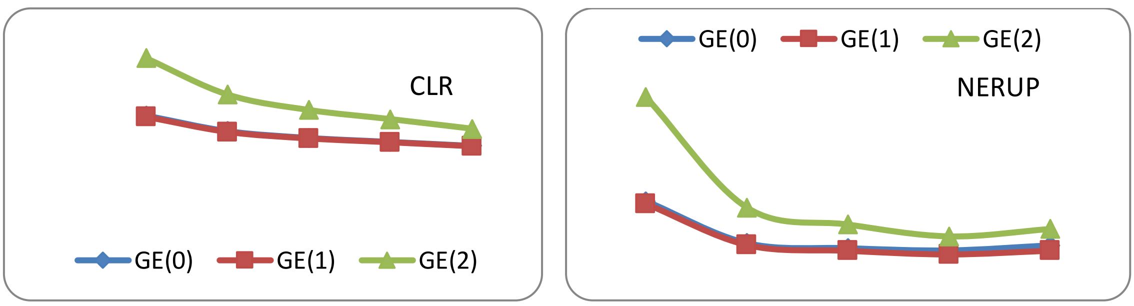
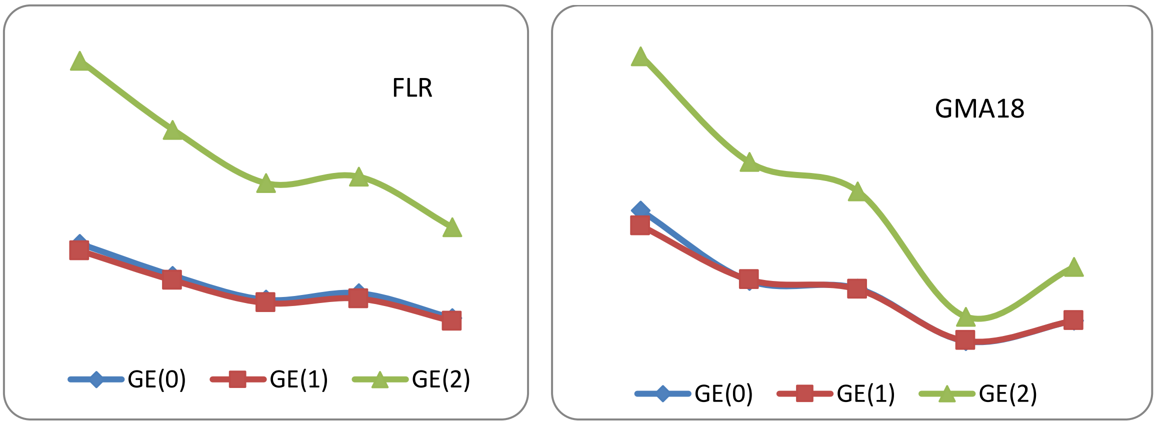


Fig. 2.12: Inequality Trend of FLR and GMA18



The values of GE of different components of SDI is shown in Appendix- B (Table-2).The inequalities of all the components are found to be declining over time; only exception is observed in case of ante-natal care visits (PWANC3), it increases in DLHS-2 and NFHS-4. Since GE(2) is found to be higher than GE(0) and GE(1), we can say that right tail of the distribution is more sensitive, this means that over time all the districts are moving to the higher values of social development parameters.

2.7 Findings

Here, I construct a new measure of social development, known as Social Development Index (SDI). By this index, I can see the development of the districts of West Bengal and also see the relative position of each district and identify that some districts are trapped by low level social development. Districts like Malda, Uttar-Dinajpur, Murshidabad and Purulia are stuck in a low level development trap over all the periods from 1998-99 to 2015-16. I see the dispersion, inequality and their trends of SDI of the districts in West Bengal.

Chapter-III

Social Development and Space: A District Level Analysis in West Bengal

3.1 Introduction

In the previous chapter, we have seen that there exists inequality and also heterogeneity among the districts in respect of SDI. The SDI is a composite index measuring average achievement in three dimensions of social development that actually captures non-monetary measure of social progress. The index is best seen as a measure of people's ability to live a long and healthy life, to communicate and participate in the life of the community. It purely considers non-monetary measure of social progress and it differs from HDI; because HDI takes into account the access to resource dimension but our SDI does not. The cultural dimension, though it is heavily influenced by education dimension is added here in measuring SDI that is special in my research work.

The New Economic Theory developed by Krugman(1991) has emphasised mainly on spatial effect on economic growth. The study of spatial economics is based on location or geography. Spatial economics tells us that any economic activity is driven by spatial forces. Location affects economic activity and many economic activities or market is concentrated geographically. The theory of 'New Economic Geography' has developed from the theory of emergence of large agglomerations which tend to increase returns and also reduce transportation cost. Krugman(1991) has got the idea of new geography and used in international trade theory. He has emphasised the role of distance and space in explaining trade and urban agglomeration.

Spatial econometrics differs from traditional econometrics in a two ways: 1) spatial dependence between the observations and 2) spatial heterogeneity in the relationships. Spatial dependence between the observations by Gauss-Markov assumes the explanatory variables are fixed in repeated sampling i.e. $E(u_i)=0$. Similarly, spatial heterogeneity violates the Gauss-Markov assumption that a single linear relationship with constant variance exits across the sample data observations i.e. $V(u_i)$=constant. Gauss-Markov states that in a linear regression model in which, $E(u_i)=0$ and uncorrelated and lave equal variance and errors do not need to be normal, nor do they need to be independent and identically distributed. Four basic tools are there in spatial econometrics methodology-

1. Specification of spatial effect in econometric models- set up an econometric model where spatial effect happens.
2. The estimation of models that incorporates spatial effects.
3. Specification tests and diagnostic for the presence of spatial effects.
4. Spatial prediction.

3.2 Objectives

My third objective is to study the relationship among space (districts, in my case) in respect of SDI and PCDDP using Global and Local Moran's Index. I use standard spatial weight matrix derived from Rook's Contiguity Matrix.

3.3 Review of Literature

Rondinelli and Ruddle (1997) have shown that location of social and economic activities lies at the core of development strategy. This study confirms the close relationship between location of industry, commerce and public facilities and the distribution and concentration of population. The pattern of population distribution- the spatial arrangement of human settlement- has a pervasive influence on a nation's social, economic and political organization. The locations of public services, physical facilities and productive activities have some spatial impact.

Chamarbagwala (2009) has checked spatial dependence on children's participation in work, idleness and school attendance in Indian districts. He finds significant spatial correlation of education related variables and suggests some policy for implementing the target level of quality and quantity of education.

Ahmed (2011) has checked the spatial educational inequalities in Pakistan. He has investigated the spatial distribution of income inequality, education, growth and development levels for 98 districts between 1998 and 2005. The overall findings suggest that distribution of districts with respect to those indicators exhibit a significant tendency of the presence of spatial autocorrelation.

Rende and Donduran (2013) have studied the impact of space on human development index. They have proposed self-organising maps to explore similarities among countries using the components of the HDI rather than rankings and illustrated clusters of countries and defined as 'neighbourhoods in development'. They Illustrate that countries sharing similar characteristics do change over time and these neighbourhoods do not necessarily overlap with the HDI ranking.

Zhang and Lin (2015) have observed the existence of spatial dependence of homogeneous and heterogeneous populations. Their work analyse the growth and distribution of sectoral and aggregate incomes in the 10-year period 1993-94 to 2002-03 in various regions of Maharashtra and finds notwithstanding its overall high economic development and suffers from actual regional inequality. District level sectoral as well as aggregate per capita income data show marked spatial association and so the spatial spill over and contiguity effect. From central Maharashtra that shocks of development affecting significantly among large number of districts.

3.4 Research Gap

Since Social Development Index (SDI) is a new concept for measuring development of a country state in a true sense and no such study does exist in India, therefore, it is my privilege to undertake such study in the districts of West Bengal. How does space matter towards explaining the variations of SDI in West Bengal? I believe such study might explore some new findings that may be useful to the development planners to formulate development strategy at the disaggregate manner.

Therefore, it is my good opportunity to study the spatial heterogeneity of SDI among the districts of West Bengal over time using newly developed Spatial Econometric methods.

3.5 Methodology and Data

3.5.1 Methods

I. **Spatial Dependence**- Spatial dependence in a collection of sample data means that observations at location *i* depend on other observations at locations *j*, where *j≠i*. It means sample data observed at one point in space to be dependent on values observed at other locations. Spatial dependence is "the propensity for nearby locations to influence each other and to possess similar attributes" (Goodchild, 1992, p.33). A famous geographer named Waldo Tobler(1970), everything is related to everything else, means similar things have tendency to close together and dissimilar things are apart from another. Two things are common: 1) data collection of observations associated with spatial units such as zip-codes, countries, states, census etc. 2) more important reason is to expect spatial dependence is that the spatial dimension of socio-demographic, economic or regional activity may truly be an important aspect of a modelling. The location and distance are important aspects to measure spatial dependence.

$$Y_i = f(Y_j), i = 1,2 ....n; j \neq i ......(3.1)$$

II. **Spatial Heterogeneity**- Spatial heterogeneity refers to variations in respect to space of any socio-economic variable. It refers to an uneven distribution of a particular event or variable within a particular region. Spatial heterogeneity can be of two types, local and stratified. Spatial local heterogeneity refers to the phenomena that the value of an attribute at one site is different from its surrounding and in case of spatial stratified heterogeneity refers for an attribute that within strata variance is less than between strata variance. For example, within the populations that are spread across landscapes, there exist pockets of low and high fertility, mortality, and population movement or migration. Sometimes, events such as births or deaths are clustered. There are a variety of exploratory tools for detecting spatial heterogeneity. Moran’s I is often used as an indicator of spatial association (Anselin, 1995; Anselin, 2005).

III. **Spatial autocorrelation different from autocorrelation-** spatial autocorrelation measures the correlation of a variable with itself through space. It means how much close objects are in comparison with other close objects. Spatial auto-correlation is the same as correlation only the difference is to measure the correlation between variables measure the same attribute but at a different spatial locations. This is the main reason which differentiates spatial autocorrelation than others. Spatial autocorrelation can be defined as a particular relationship between the spatial proximity among observational units and the numeric similarity among their values. Spatial autocorrelation occur means certain amount of information is similar to its neighbouring regions and entire data set possess some amount of redundant information. This feature violates the assumption of independent observations upon which many standard statistical treatments are predicted. Spatial autocorrelation examines the different pairs of sample locations by measuring the distance between the locations. Statistics in geography considers that space and location has an influence on observations. The closer things are, the more likely they will influence each other. Spatial autocorrelation defers from traditional autocorrelation by time variable. Autocorrelation represents of the degree of similarity between a given time series a lagged values either depend on it-self and also time variation. It is actually measures the relationship between current variable and of its past variable. It is basically a time dependent not space dependent. But in case of spatial autocorrelation, it is a apace dependent means change in particular location of an attribute changes with its neighbouring value. So, observations are not independent in nature. Outcome of a variable of one region depends upon outcome of other regions.

So, spatial autocorrelation can be formally expressed by the condition,

$$COV(Y_i, Y_j) = E(Y_i, Y_j) - E(Y_i).E(Y_j) \neq 0; i \neq j \text{...........(3.2)}$$

This covariance becomes meaningful from a spatial perspective as this takes **non-zero** value. Spatial autocorrelation can be two types: positive or negative. Positive spatial autocorrelation means high or low values making a cluster between two different locations.

1	1	
1	1	

Here for shaded regions are making a cluster which means these four regions are neighbouring to each other and one region's impact may influence to other 3 regions may be at high values or low values.

Negative spatial autocorrelation means that all regions are not relating to each other and there is dissimilarity between the regions.

1		1
	1	
1		1

The above table describes that the regions are not related to each other and they are not making any cluster. Means that the surrounded neighbouring values have dissimilar connection and they are not forming any cluster.

IV. Spatial Neighbours and weights:

Spatial statistics is related to space and the relation of space is directly associated with mathematics like area, distance, length etc. a spatial weights matrix is a representation of the spatial structure of the data. Spatial weights matrix imposes a structure of data to show how spatial data are interact to each other. Spatial weights are measured in a $(N \times N)$ ("N" is the number of features in the data set). There is one row for every feature and one column for every feature. The every cell value gives row and column combination representing the weights that qualifies the spatial relationship between those row and column features. The matrix is given below:

$$W = \begin{pmatrix} w_{11}w_{12}.........w_{1n} \\ w_{21}w_{22}........w_{2n} \\ .... \\ .... \\ w_{n1}w_{n2}..........w_{nn} \end{pmatrix}........(3.3)$$

The values of w_{ij} of the weights for each pair of location assigned by some pre-set rules which define the relationship among regions. The diagonal values of the weight matrix **W** is 0 i.e. w_{ij}=0. Row standardization is the distribution of the features is potentially biased due to sampling design or to an imposed aggregation scheme.

Row standardization- spatial weights are standardized by row. Each weight is divided by its row sum. It means the elements of the row-standardized weight matrix equal to:

$$w_{ij}^{s} = \frac{w_{ij}}{\Sigma j \, w_{ij}}$$

The sum of each row has to be 1 means,

That is,

$$\sum_{i=1}^{n} w_{ij} = 1, i = 1,2.......n...........(3.4)$$

This ensures that all weights lay between 0 and 1 and facilities the interpretation of operation with the weight matrix as an averaging of neighbouring values. This process is called row-standardization of W; it means row-normalised weights, w_{ij} can be interpreted as a fraction of all spatial influence on unit *i* to unit *j*. This weight assigns to measure how two regions are intensively related to each other.

There are two strategies for creating weights to quantify the relationships among data features: binary or variable weights. In case of binary weights there is a fixed distance with their K nearest neighbours and it named as contiguity like if a feature has either a neighbour then its value is 1 or it is not then, value be 0. For weighted strategies, there are varying amount of impact and it is called influence and weights are computed to reflect that variation. The availability of polygon or lattice data permits the construction of contiguity based spatial weight matrix. A typical specification of the contiguity relationship in the spatial weight matrix is,

$$w_{if} = \begin{cases} 1 \ \ if & i \ and \ j \ are \ contiguous \\ 0 \ \ if & i \ and \ j \ are \ not \ contiguous \end{cases}$$

Binary contiguity mainly of 3 types:

A. Rook contiguity- two regions are neighbours if the two regions share a common border (either left and right or up and down). It is basically calculated by distance. Larger the border, lager is their distance.
B. Bishops contiguity- two regions are neighbouring to each other if they meet at a "point". The two spaces meet in a graph by vertex. That means two regions have common vertex.
C. Queen contiguity- this is the union of Rook and Bishop Contiguity. If the two or more regions are sharing their border and vertex at a one stage then it is called Queen Contiguity.

These 3 types are examined by a table.

1	2	3
4	5	6
7	8	9

In this table there are 9 regions. If we examine for region 5 then,

- Rook Contiguity to be region 2, 4, 6 & 8 which are red in colour. Because for 'region 5' these four regions have common border.
- For Bishop Contiguity the regions are 1, 3, 7 & 9 as in respect of 'region 5' these regions are related by vertexes which are green colour.
- Lastly for Queen Contiguity the regions are 1, 2, 3, 4, 6, 7, 8 & 9 as these eight regions are related with 'region 5' by sharing common border and vertex.

V. Measures of Spatial Autocorrelation:

Spatial autocorrelation is multi-dimensional and useful to make specific patterns in complicated data sets.

1. **Moran's I**
2. **Geary's C**
3. **Getis-Ord G statistic**

Here, we are focussing only on Moran's I. Moran's I index is very popular to capture spatial autocorrelation. To capture the impact of one region to another or spatial autocorrelation, we employ Moran's Index. This index value captures both z-score and p-value. P-values are numerical approximations of the area under the curve for a known distribution, limited by test statistics. It means this index can be tested statistically.

The null hypothesis for the test is that one data is randomly distributed i.e.

H_0 : there is no spatial clustering with the associated area

But the alternative hypothesis is that the data is more spatially connected and forming a cluster i.e.

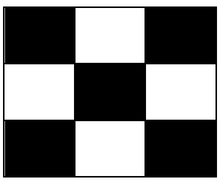
H_1 = there is spatial clustering with the associated area.

This statistic gives two possible outcomes:

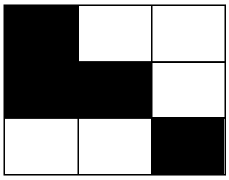
a) **A positive z-value**: data is spatially clustered and high value and low value are forming a cluster.
b) **A negative z-value**: data is clustered in such a way that, high value clustered with high values and low value clustered with low values.

Moran's index is similar to correlation coefficient but not equivalent. The value of this index varies from -1 to +1. The patterns are as follows:

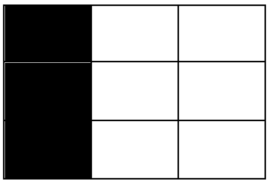
- -1 is perfectly dissimilar; values are not forming any cluster. It is the case of perfect dispersion.



- 0 is no autocorrelation means the regions are random in nature. There is no systematic pattern.



- +1 indicates perfect clustering of similar values; it means the regions are having the tendency to form a group.



The Moran's I index can be divided into two ways.

a) Global Moran's I (GMI)and
b) Local Moran's I (LMI)

a) GMI: GMI is mainly calculated the spatial effect on an aggregated level. The value of GMI can be positive or negative and also statistically measurable. Positive but statistically significant means that the dispersion between high values and low values are relatively less and they are forming a group or cluster. So, the high valued and low valued regions are not forming polarization and they are more connected to each other. But, if the value of GMI is negative and statistically significant that means the high valued and low valued regions are dispersed and they are not forming any group or cluster. It means they are belonging to two polarizations, having separate regions:

The mathematical statistics pertaining to GMI is given below:

$$I = \frac{N}{W} \cdot \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2} \quad \text{........(3.5)}$$

Where, X_i and X_j are the two regions and $\bar{x}$ is the means of the two regions; w_{ij} is the elements of spatial weight matrix that measures the spatial distance or connectivity between i and j and i$\neq j$. In the absence of autocorrelation and regardless of the specific weight matrix, its expectation is,

$$E(I) = -\frac{1}{N-1} ......(3.6)$$

This tends to zero as sample size increases. If the coefficient of GMI is larger than -1/(n-1) then it indicates positive spatial autocorrelation and if coefficient of GMI is smaller than -1/(n-1) then it indicates negative spatial autocorrelation.

The Z_i-score for the statistics is computed as:

$$z_i = \frac{I - E(I)}{\sqrt{V(I)}} ......(3.7)$$

The variance of GMI is:

$$V(I) = E(I^2) - [E(I)]^2$$

Where, Variance of GMI can be calculated as:

$$V(MI) = \frac{NS_4 - S_3 S_5}{(N-1)(N-2)(N-3).W^2} - (E[MI])^2 ....(3.8)$$

Where,

$$S_1 = \frac{1}{2}\sum_i \sum_j (w_{ij} + w_{ji})^2$$

$$S_2 = \sum_i \left(\sum_j w_{ij} + \sum_j w_{ji} \right)^2$$

$$S_3 = \frac{N^{-1}\sum_i (X_i - \overline{X})^4}{\left[N^{-1}\sum_i (X_i - \overline{X})^2 \right]^2}$$

$$S_4 = (N^2 - 3N + 3)S_1 - NS_2 + 3W^2$$

$$S_5 = (N^2 - N)S_1 - 2NS_2 + 6W^2$$

b) LMI: Local Moran's Index (LMI) captures the same effect but at the disaggregated level. It means one region getting impacted by or may have an impact on other regions. It also gives three types of result i.e. positive, negative and insignificant. Positive local Moran's index means one region is influenced by other region and they are forming a group and it has neighbourhood impact and it is also statistically measurable. Negative Local Moran's Index means one region is independent to its neighbouring regions. These regions have enough endowment to become self-sufficient if it appears significant statistically. The insignificant values have no spatial meaning.

The statistics is:

$$MI_i^L = \frac{X_i - \overline{X}}{S_i^2} \sum_{j=1, i \neq j} w_{ij} \left(X_j - \overline{X} \right) \text{......(3.9)}$$

Where,

$$S_i^2 = \frac{\sum_{j=1, i \neq j}^{N} w_{ij}}{N - 1} - \overline{X}^2$$

The expectation of LMI is:

$$E\left(I_i\right) = -\frac{\sum_{j=1, i \neq j}^{N} w_{ij}}{N - 1} \text{.......(3.10)}$$

The Zi-score for the statistics is computed as:

$$z_i = \frac{I - E(I)}{\sqrt{V(I)}} \text{.....(3.11)}$$

The variance of LMI is:

$$V(I) = E(I^2) - [E(I)]^2$$

$$Var(MI_i^L) = \frac{\left(N - B_2\right) \sum_{j=1, i \neq j}^{N} w_{ij}^2}{N - 1} - \frac{\left(2B_2 - N\right) \sum_{k=1, k \neq i}^{N} \sum_{k=1, k \neq i}^{N} w_{ik} w_{ik}}{\left(N - 1\right)\left(N - 2\right)} - \left[E(MI_i^L)\right]^2 \text{.....(3.12)}$$

Where, $$B_2 = \frac{N \sum_{i=1, i \neq j}^{N} \left(X_i - \overline{X}\right)^4}{\left[\sum_{i=1, i \neq j}^{N} \left(X_i - \overline{X}\right)^2\right]^2}$$

3.5.2 Data Sources:

Data pertaining to Health indicators are drawn from DLHS (I-IV) and NFHS-4 surveys. Upper primary enrolment is drawn from DISE Data; data on enrolment for 1998-99 was extrapolated from FLR. Per Capita DDP are drawn from Economic Survey, Govt. of West Bengal.

3.6 Analysis

3.6.1 Results

I. Global Moran's Index of SDI:

The result of SDI will more prominent if we use the spatial econometrics. If we don't see the impact of space, then we cannot identify is there any relation in between the districts or not. To capture the spatial autocorrelation we use Moran's Index by the following Table 3.1:

Table:3.1 Results of Global Moran's(GMI) Index of SDI

Time Points	SDI		
	GMI	Z-value	P-value
DLHS-1	0.317^{***}	3.626	0.000
DLHS-2	0.466^{***}	4.160	0.000
DLHS-3	0.370^{***}	3.566	0.000
DLHS-4	0.203	1.116	0.132
NFHS-4	0.365^{***}	3.557	0.000

Note: *** denotes 1% level of significance and ** denotes 5% level of significance

From the above Table 3.1, we can see that the index values are positive and statistically significant except DLHS-4. This means that space does matter towards variation of SDI and heterogeneity is less between high and low value of SDI. Now a positive value of Global Moran's Index (GMI) means spatial autocorrelation does exist. This is also supported by the graphical representation of SDI for each of five time points. The figures are as follows:

Fig. 3.1 Moran's of SDI of Districts: DLHS-1

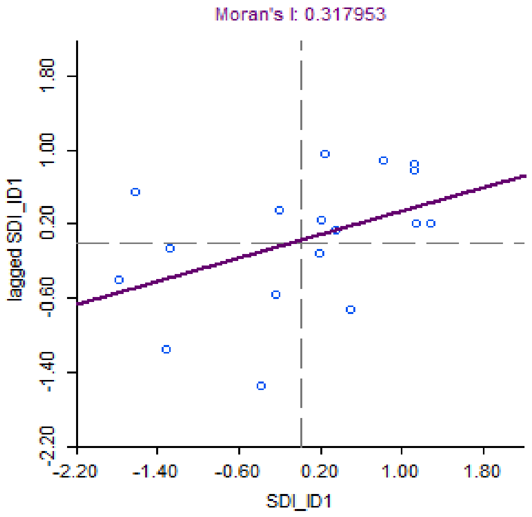


Fig.3.2 Moran's of SDI of Districts: DLHS-2

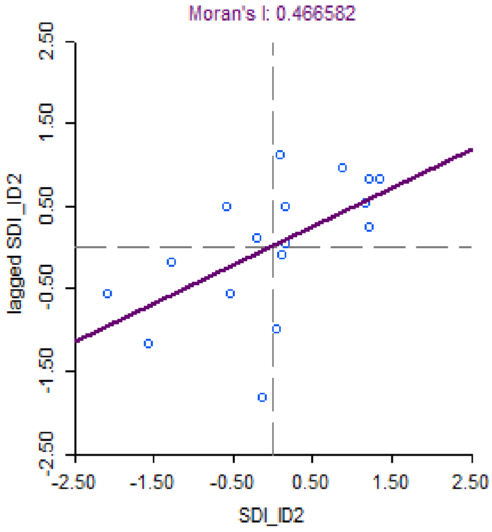


Fig. 3.3 Moran's of SDI of Districts: DLHS-3

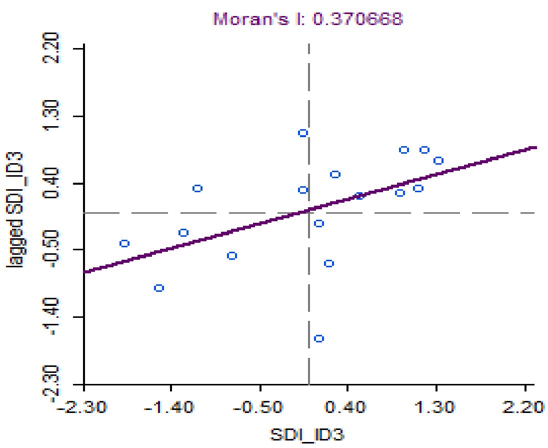


Fig.3.4 Moran's of SDI of Districts: DLHS-4

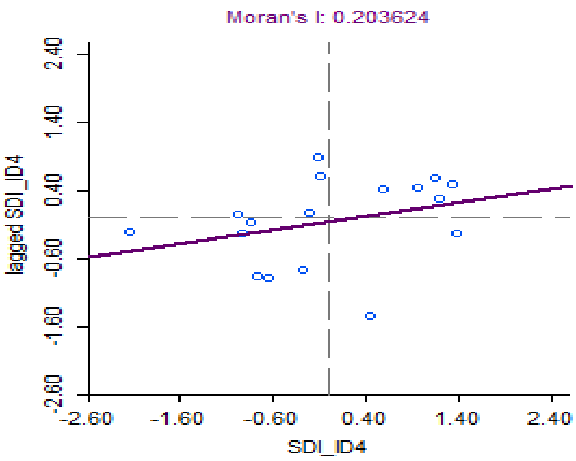
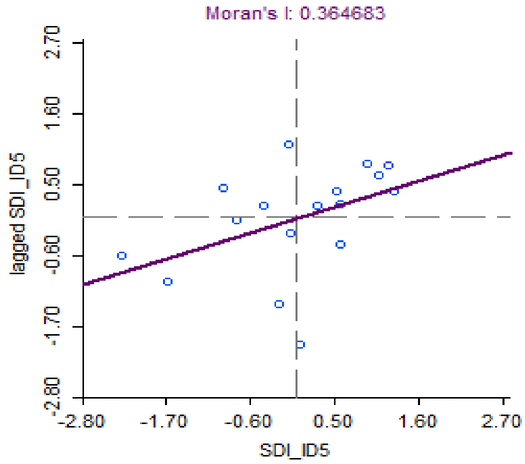


Fig. 3.5 Moran's of SDI of Districts:

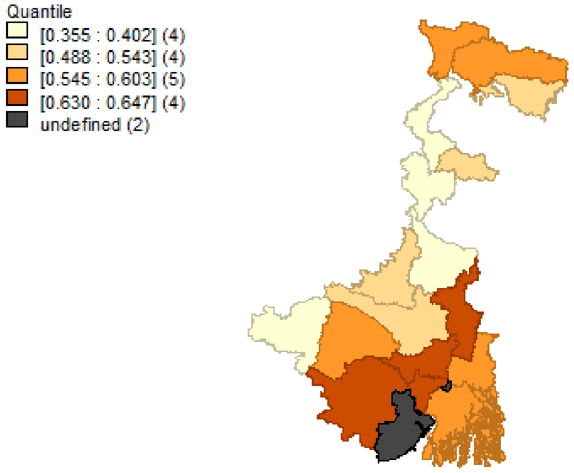
NFHS-4



The above five figures Fig(3.1) to Fig(3.5) show that there is a positive relation between the districts and the districts having high value of SDI with high values of SDI with neighbours; similarly districts with low values of SDI are clustered with districts having low values of SDI. The values of GMI is also statsitically significant which means that the districts are forming group or cluster and the disparsion of high valued districts and low valued districts are relatively less. Anselin (1996) has demonstrated that the slope of the regression line through these points expresses the global Moran's I value which, for SDI for 5 time points are given in Table-3.1 with their respective significant level. This suggests a clustered spatial pattern in distribution of districts in respect of SDI. The GMI is found to be insignificant in DLHS-4 that means that districts are forming a group with high-high and low-low rather they are forming a group with high-low and low-high. This is an expectional case and the divergence between the high valued districts and low valued districs are in a broader sense widened which proves the existence of polarization. This is happened in case of DLHS-4 (2012-13). The upper right quadrant of the Moran scatter plot shows those districts with above average SDI and share above average SDI with neighboring districts (high-high). Also, the lower left quadrant shows districts with below average SDI values and neighbors also with below average SDI values (low-low). The lower right quadrant displays districts with above average SDI surrounded by districts with below average values of SDI (high-low), and the upper left quadrant contains the reverse (low-high). Now, in case of SDI, this spatial dependency can also be shown in terms of Mapping:

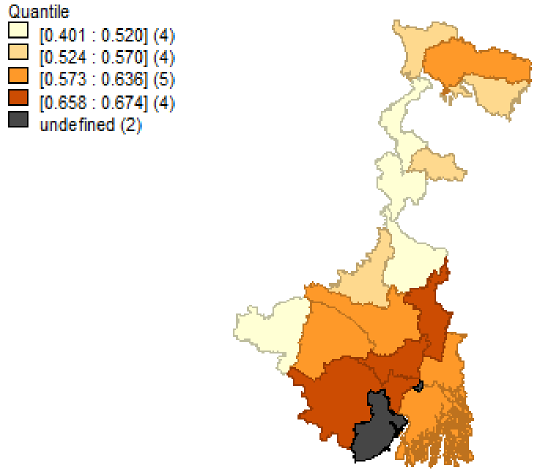
Map. 3.1 GMI of SDI of Districts:

DLHS-1



Map.3.2 GMI of SDI of Districts:

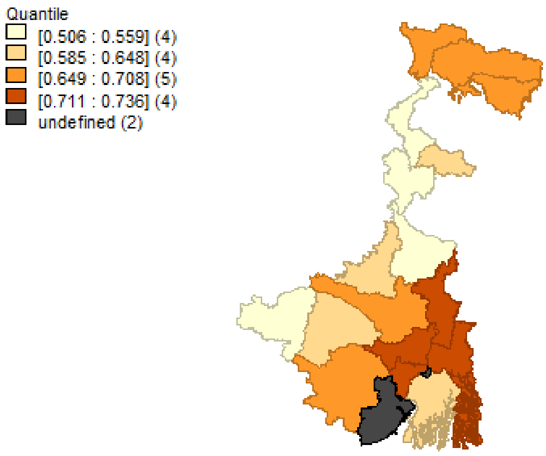
DLHS-2



In these Maps of West Bengal for different time points Map(3.1) and Map(3.2), we see that districts like Howrah, Hooghly, Medinipur (Combined), Nadia form a group or making a cluster with higher values of SDI. The districts like Malda, Murshidabad, Uttar Dinajpur and Purulia are again forming a cluster with lower values of SDI. In these two time points, the districts are relatively in a same position.

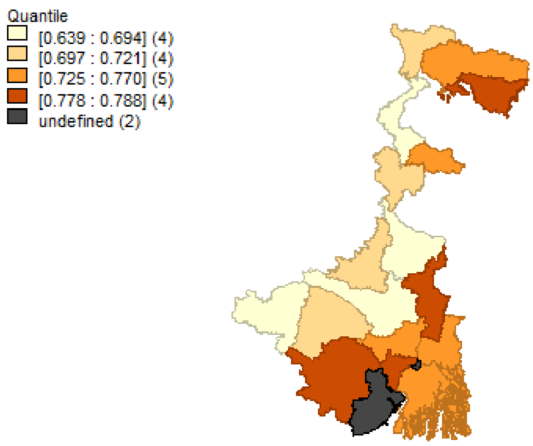
Map. 3.3 GMI of SDI of Districts:

DLHS-3



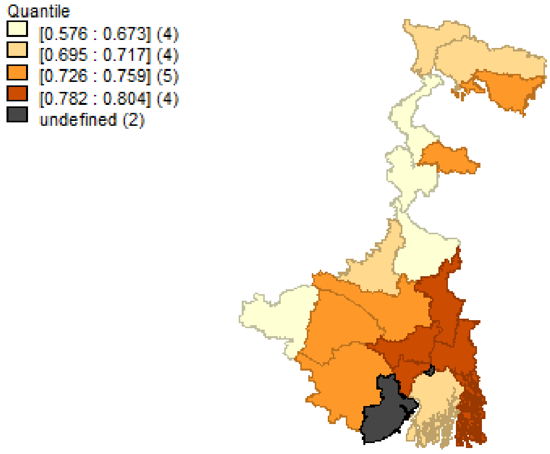
Map.3.4 GMI of SDI of Districts:

DLHS-4



Map. 3.5 GMI of SDI of Districts:

NFHS-4



In Map(3.3), the position of the high valued district like Medinipur is deteriorated but the position of the other three districts(Howrah, Hooghly and Nadia) remain same and North 24 Parganas joins into this group that means this district is performing better in respect of SDI. In case of the worst performing districts, they are also remaining their own position. Another group of districts is formed in North Bengal which are relatively better in respect of SDI, these are Darjeeling, Jalpaiguri, Coach Behar. In Map(3.4), the positions of all the districts are found to be changing; we did not find any clear clustering of districts. For this cause the value of GMI at this point is found insignificant; that means the districts are randomly behaving. But in the final time point Map(3.5), which is nothing but NFHS-4, the position of all the districts are quite good and the districts are retained to their earlier positions (like DLHS-1,2). Some districts like Howrah, Hooghly, Nadia, North 24 Parganas are better performing districts in respect of SDI, more prominently for Dakhshin Dinajpur is improved very much as the colour is changed with deeper shade, whereas the bad performing districts appear to be Uttar Dinajpur, Malda, Murshidabad and Purulia forming a low level social development trap. Another cluster of districts is found which lie on average SDI value; these are Medinipur (combined), Bankura and Bardhaman. These districts are remaining to their positions throughout all the time points.

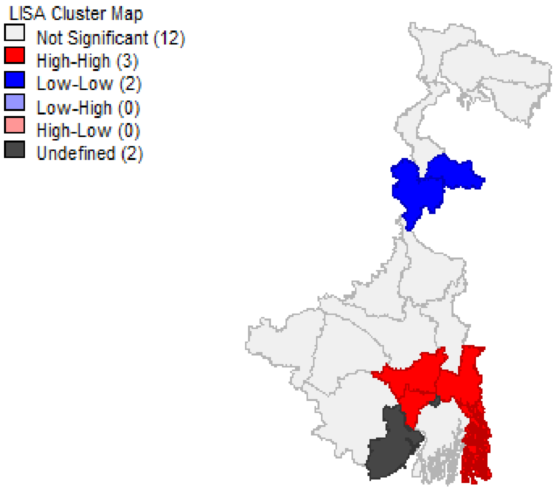
II. Local Moran's Index (LMI) of SDI:

The result of Global Moran's Index (GMI) remains incomplete if we do not incorporate the results of Local Moran's Index (LMI). Spatial outliers can be detected using Local Moran‘s Index (LMI). The LMI indicates the location of local cluster and spatial outliers (Anselin 1988). Like GMI, the LMI provide the information about the presence of both positive and negative spatial autocorrelation. The sum of the values of all observations is proportional to Global Moran‘s Index (GMI). If we plot the index values in a scatter plot, we can observe the location of local cluster and spatial outliers. A positive value of index indicates that the development of that region and other regions are positively related. They are related to each other whereas a negative index value indicates that the region is an outlier. In the context of spatial autocorrelation, the localized phenomena of interest are those areas on the map that contribute strongly to the overall trend (which is usually positive autocorrelation). Methods enable an analyst to identify localized regions in a map; where data values are strongly positively or negatively associated with each other and these are collectively known as Local Indicators of Spatial Association (or LISA). We can also see that the districts having positive values, they form a club but the negative value clearly manifests an outlier.

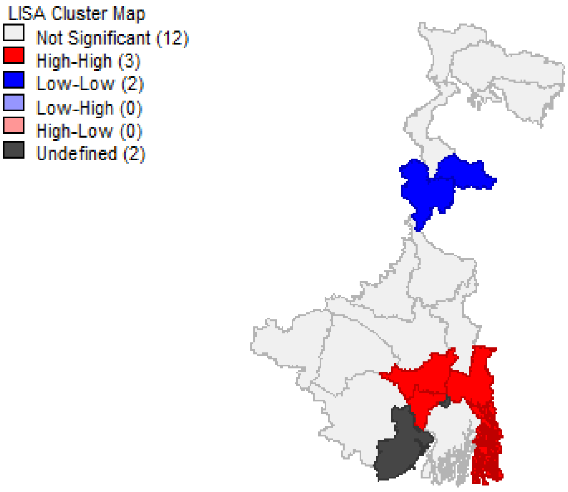
We observe that few districts like Darjeeling, Dakshin Dinajpur, Purulia have come out as outliers. It is to be mentioned here that districts with low values of SDI (like Uttar Dinajpiur, Malda and Murshidabad) that persist over long period of time do share international border with Bangladesh. But, we could not capture the spatial weights relating to international border in our analysis. The pictorial representation of LMI for each 5 time points which are following:

From the Maps (3.6 and 3.7), we can clearly see that the districts are forming groups or clusters according to high-high values and low-low values. The blue region shows that regions forming low-low clustering and the red regions represent high-high clustering. The grey region shows the values which are undefined. Here, we use the district Medinipur in a combine (East and West) has been an outlier shown as grey colour. Again, we do not take the value of Kolkata as it is an extremely outlier. The white portion shows the regions which are not significant at all; that means the districts are random in nature and no prominent neighbourhood effect is obtained. There is a proper neighbourhood effect with high-high and low-low regions.

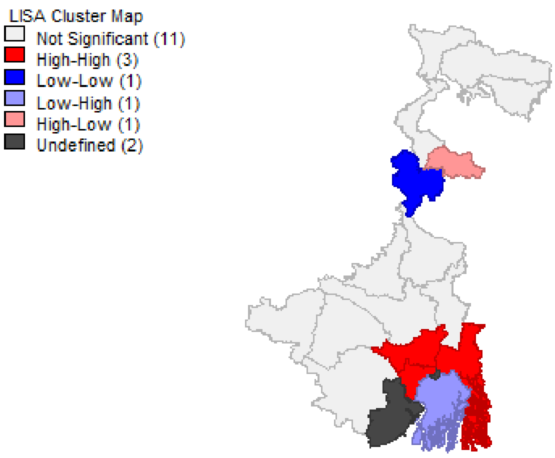
Map. 3.6 LMI of SDI of Districts: DLHS-1



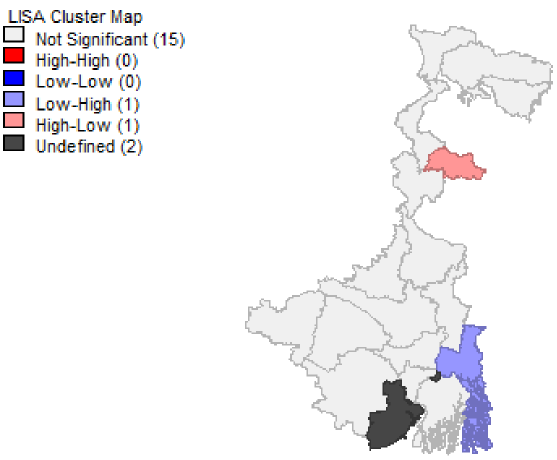
Map.3.7 LMI of SDI of Districts: DLHS-2



Map. 3.8 LMI of SDI of Districts: DLHS-3



Map.3.9 LMI of SDI of Districts: DLHS-4

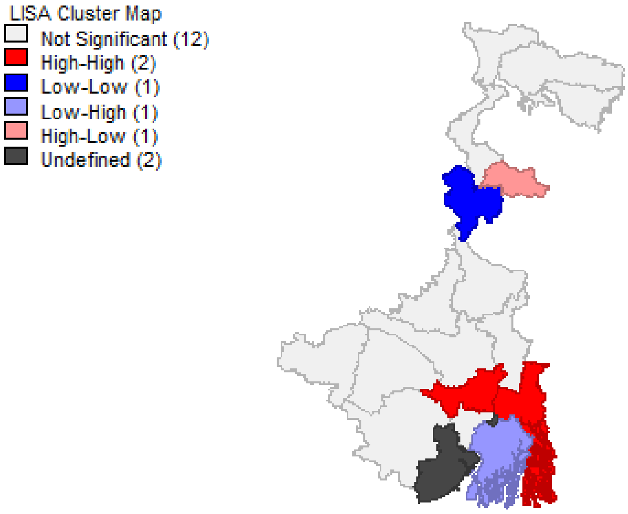


In the last two Maps (3.8), the results are slightly different; it means that in the two time points, the regions are not strictly following to high-high and low-low regions. The faded blue region shows (in the two Maps) that there is a relationship between low valued regions and high valued regions i.e. low-high in case of South 24-Parganas. The faded red region shows that there is a relationship between high valued region and low valued region making a cluster i.e. high-low in case of Dakhsin Dinajpur as the improvement of this district is very significant so far in terms of SDI but it shares its border with lower district like Malda. So it becomes an outlier.

Most of the districts are found to be insignificant which means they are random in nature. In Map 3.9, there does not exist any true cluster occurring as high-high and low-low. Therefore, this point is nothing but an exceptional case.

Map. 3.10 LMI of SDI of Districts:

NFHS-4



In the last time point (NFHS-4) i.e. Map(3.10), it is more or less identical to DLHS-3. High valued region making a cluster with high valued, low valued region making a cluster with low valued, high valued region making a cluster with low valued and lastly low valued region making a cluster with high valued and other districts are found to be insignificant that is they are random in nature.

What we find is that the SDI is space dependent. Any development of a district affects the other neighbourhood districts. This proves the existence of spatial autocorrelation and neighbourhood impact. This work shows a significant spatial model with spatial interdependences. The spatial model remains incomplete if we don't go for spatial regression.

Only the significant locations matches with significant maps and the high-high and low-low clusters are spatial clusters and the high-low and low-high clusters are spatial outliers. The strongly coloured regions are therefore those that contribute significantly to a positive global spatial autocorrelation outcome, while faded colours contribute significantly to a negative autocorrelation outcome.

III. Global Moran's Index of PCDDP:

In order to understand the spatial impact of income (PCDDP), we use both GMI and LMI. To capture the spatial autocorrelation we use Moran's Index by the following Table 3.2:

Table: 3.2 Results of Global Moran's(GMI) Index of Income(PCDDP)

Time Points	PCNDDP		
	GMI	**Z-value**	**P-value**
DLHS-1	-0.0295	1.237	0.108
DLHS-2	-0.0520	0.953	0.170
DLHS-3	-0.0618**	2.959	0.002
DLHS-4	-0.1031	1.116	0.132
NFHS-4	-0.0905	1.226	0.110

*Note: *** denotes 1% level of significance and ** denotes 5% level of significance*

From the above Table 3.2, we can see that the index values are negative and statistically insignificant except DLHS-3. This means that space does not matter towards variation of PCDDP and heterogeneity is more between high and low value of PCDDP. Now a negative but insignificant value of Global Moran's Index (GMI) means spatial autocorrelation does not exist. The districts are scattered in respect of income; no uniform pattern is noticed in respect of space. This is also supported by the graphical representation of PCDDP for each five time points which proves that for income, space is somehow independent. The scatter plots are shown as follows:

The above five graphs Fig(3.6) to Fig(3.11) show that there is a negative relation in between the districts and the districts having very low values of PCDDP with low valued neighbors of PCDDP. The values of GMI is also statsitically insignificant which means that the districts are not forming a group or cluster and the dispersion of high valued districts and low valued districts are relatively greater in this case. The GMI is found to be insignificant means that districts are forming a group with high-low and high-low rather they are forming a group with high-high and low-low. The upper right quadrant of the Moran scatter plot shows those districts with above average PCDDP and share above average PCDDP with neighboring districts (high-high); the lower left quadrant shows districts with below average PCDDP values and neighbors also with below PCDDP values (low-low). The lower right quadrant displays districts with above average PCDDP surrounded by districts with below average values (high-low), and the upper left quadrant contains the reverse (low-high).

Fig. 3.6 Moran's of PCDDP of Districts: DLHS-1

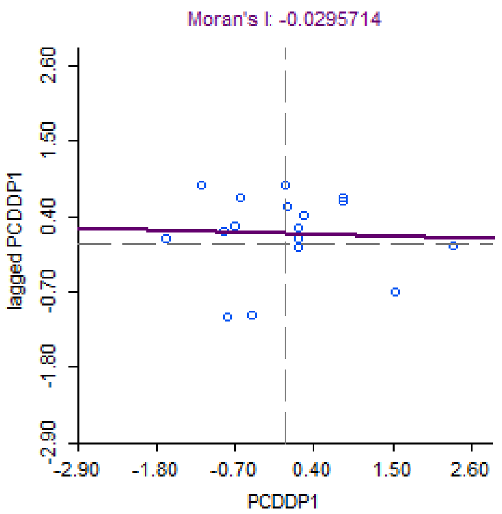


Fig.3.7 Moran's of PCDDP of Districts: DLHS-2

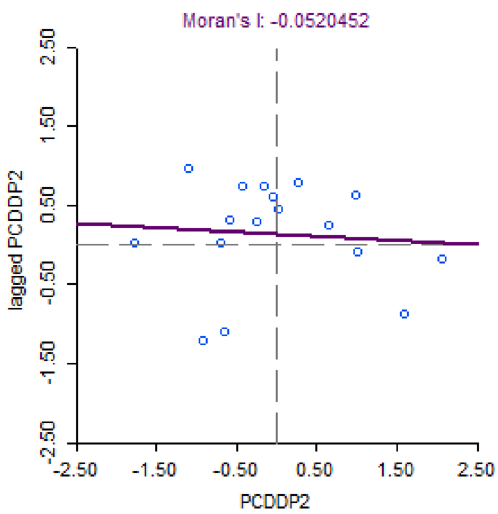


Fig. 3.8 Moran's of PCDDP of Districts: DLHS-3

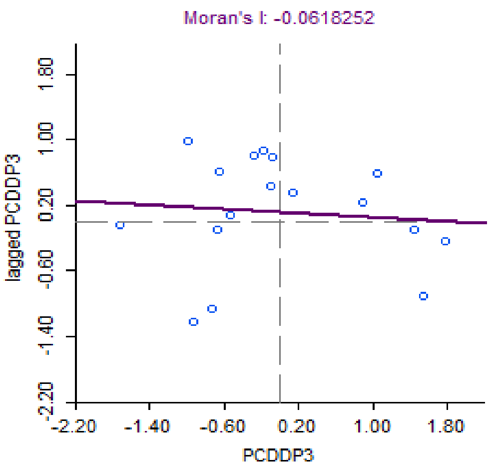


Fig.3.9 Moran's of PCDDP of Districts: DLHS-4

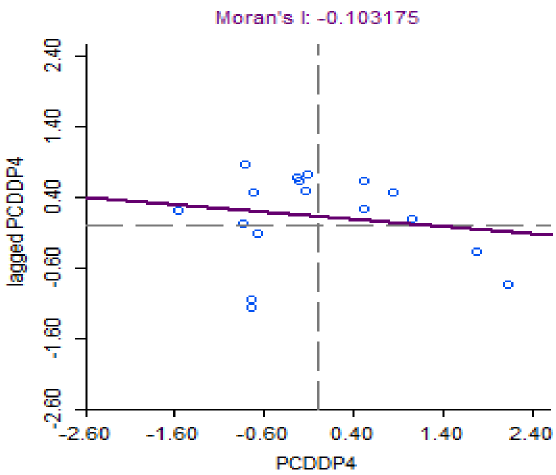
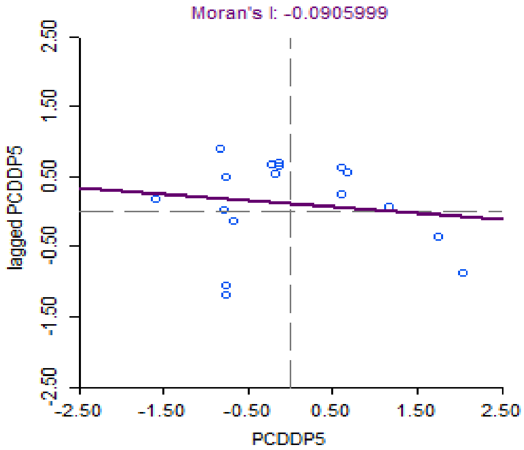


Fig. 3.10 Moran's of PCDDP of Districts:

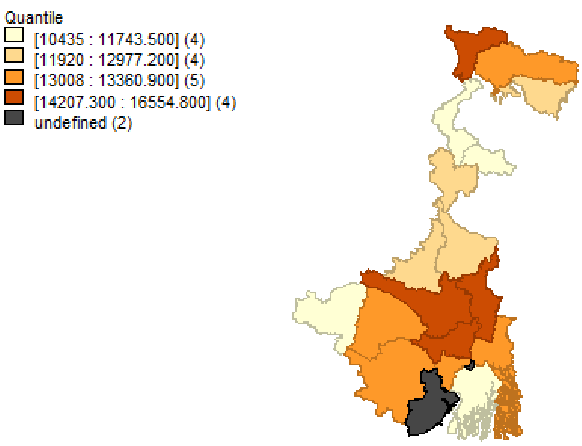
NFHS-4



This is an expectional case and the divergence between the high valued districts and low valued districs are in a broader sense which proves that there is some polarization occures. This is happening throughout all the time points from DLHS-1 to NFHS-4. The most intersting result is coming out from SDI which is actually an opposite of PCDDP. Therefore, we can draw conclusion that in case of income, districts are space independent. We can not see a proper clustering or trend of significant level in case of income as it may happen the clusters between high-low or low-high. Not excat or proper relation came out for income(PCDDP). This spatial independency can also be shown in terms of Maps as shown below (Map 3.11-3.15):

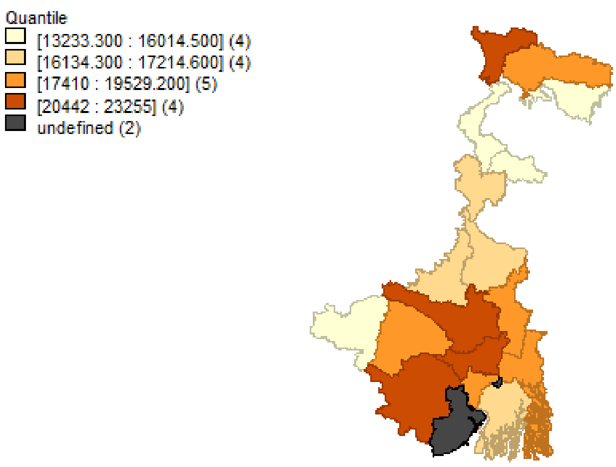
Map. 3.11 GMI of PCDDP of Districts:

DLHS-1

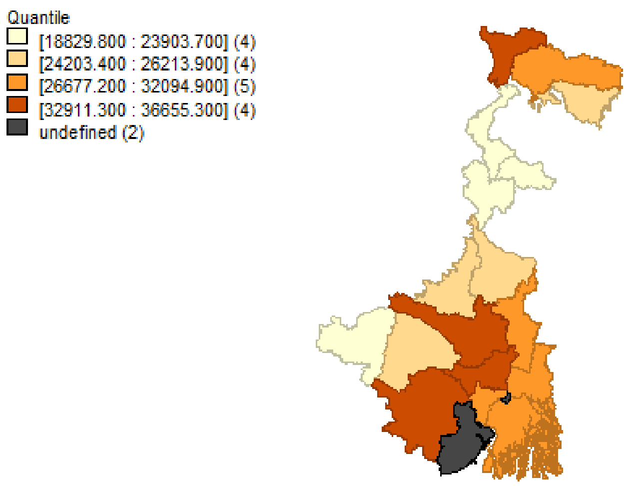


Map.3.12 GMI of PCDDP of Districts:

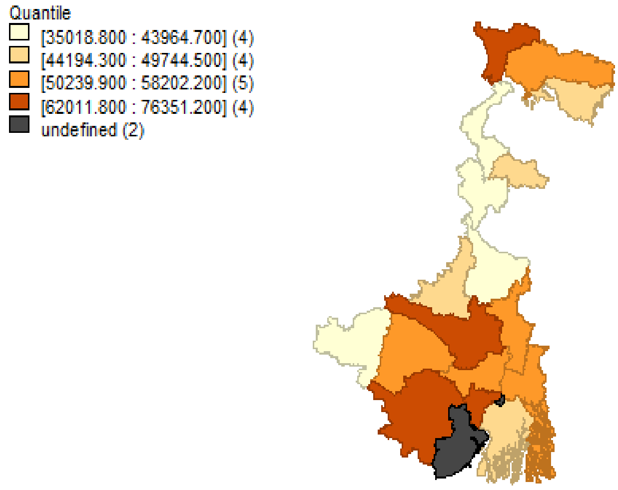
DLHS-2



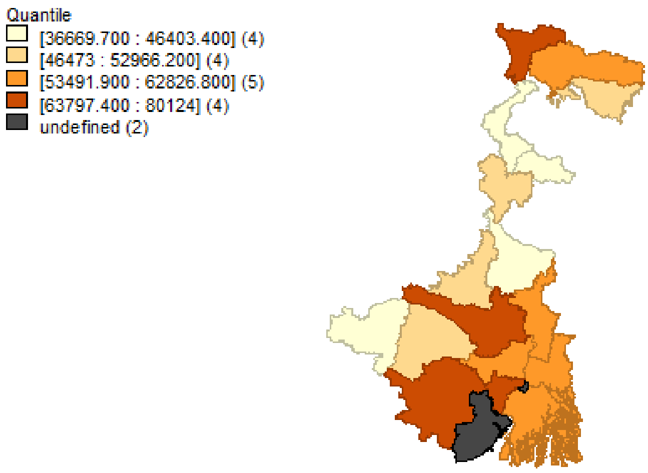
Map. 3.13 GMI of PCDDP of Districts: DLHS-3



Map.3.14 GMI of PCDDP of Districts: DLHS-4



Map. 3.15 GMI of PCDDP of Districts: NFHS-4



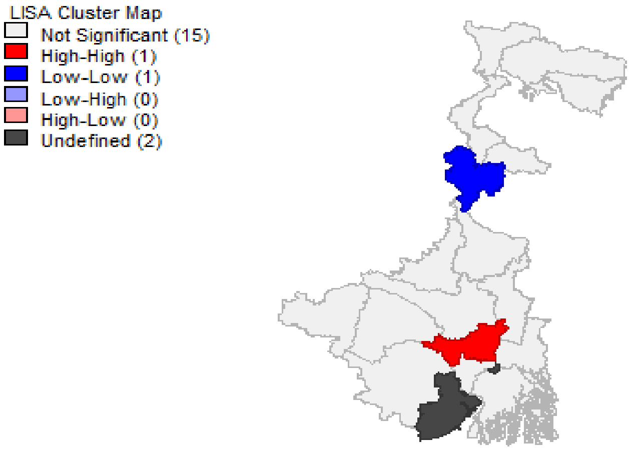
From the above Maps, we find that the districts like Bankura, Medinipur, Hooghly, North 24 Parganas are making a cluster but the districts are forming a group not in a sustained and continuous manner from DLHS-1 to NFHS-4. The other districts like Bankura, Murshidabad, Malda are forming a group with lower valued of domestic product. This group is continuing though second time point DLHS-2. In case of Bankura, the domestic product of this district is showing a very fluctuating trend, sometimes it gets increase and sometime it falls at a higher rate. Throughout DLHS-1 to NFHS-4, the districts like Uttar Dinajpur, Malda, Murshidabad and Purulia are stuck in a low level domestic product. Again for some districts like Darjeeling, Bardhaman, Hooghly, North 24 Parganas are standing at a higher position continuously. The scatter plot gives us that the overall income dependency has not occur in between districts as the value of Global Moran's I is negative for each of five time points, except DLHS-3. That means in the long run, there does not exist any specific clustering among the districts in West Bengal. Since, GMI appears to be insignificant in case of income (PCDDP), there is no necessity of estimating the LMI, and however, I have further explored the existence of clustering and outliers of PCDDP at the local level.

IV. Local Moran's Index of PCDDP:

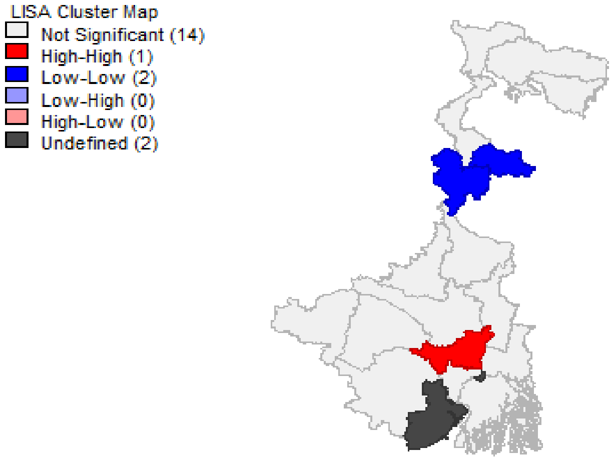
From the Map(3.16), we can see that, two districts are forming a group i.e. Malda and Hooghly. The deep blue region shows that region is forming a group with low values means low-low region and the deep red region shows a cluster with high values means high-high type. The grey region shows that the region is undefined. Here we are combining the two region of Medinipur that's why it shows grey and for Kolkata, it shows grey because of we do not incorporate the value of Kolkata as it performs well always. The white region shows that the regions are insignificant that means they are not making any cluster which proves the variable (PCDDP) is not space dependent. In case of Map(3.17), this situation is quite different as there is only one cluster with Malda and Dakhin Dinajpur, this means the low valued region clustered with low value i.e. low-low cluster occurs. But, the others are same as previous time point.

For Map(3.18), again this situation is different from previous two time points as there are not only one low-low cluster in between Malda and Dakhn Dinajpur rather two outliers are also coming out in this time point. But, the two outlier regions are also in a scattered way i.e. for Purulia and Howrah they are making a cluster but separately and also with low valued region cluster with high valued region means low-high cluster occurs which is nothing but two outliers (as shown in coloured by faded blue).

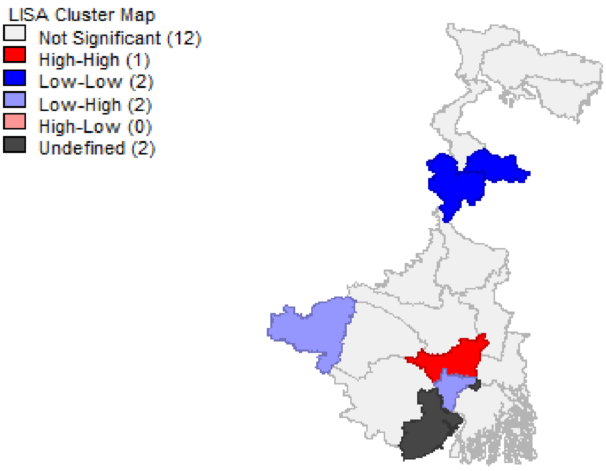
Map. 3.16 LMI of PCDDP of Districts: DLHS-1



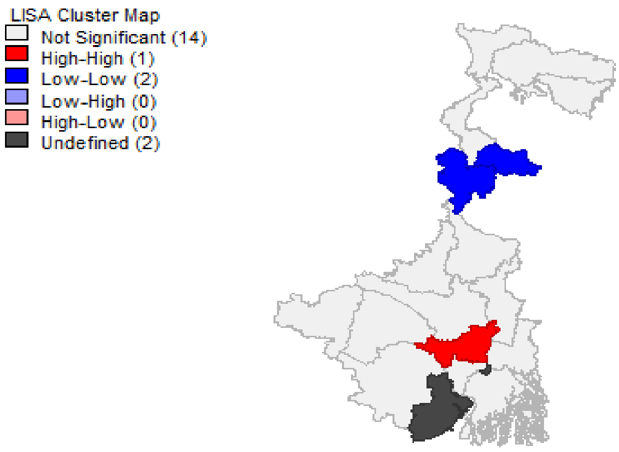
Map.3.17 LMI of PCDDP of Districts: DLHS-2



Map. 3.18 LMI of PCDDP of Districts: DLHS-3

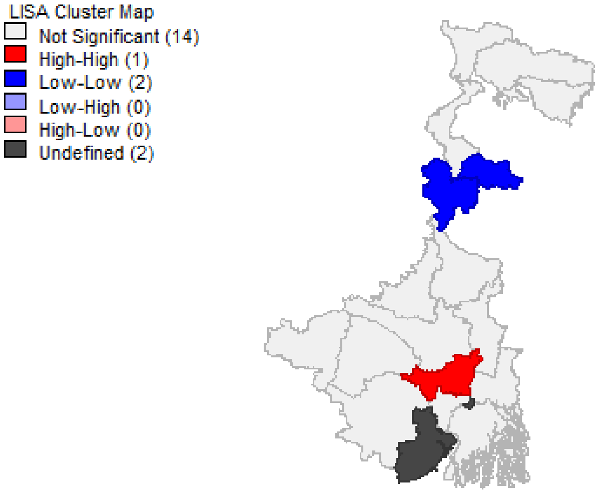


Map.3.19 LMI of PCDDP of Districts: DLHS-4



Map. 3.20 LMI of PCDDP of Districts:

NFHS-4



In case of Map 3.19, this picture is very similar to second time point (DLHS-2: 2002-2004). Here only one cluster comes out by two districts Malda and Dakhsin Dinajpur which is again a low-low cluster and the other one is clustering with itself i.e. Hooghly. This situation persists to the last time point (NFHS-IV: 2015-16). Therefore, to sum up, we can see that for each time point there does not exist any clear cluster that persists over long time period. Some districts are making a cluster but they are not sustained over time; therefore, we find that income is space independent.

3.7 Findings:

The main objective of this Chapter is to find out spatial impact of SDI and PCDDP. We can see that SDI is dependent on space and for income it is independent. The overall findings that emerge from empirical analysis is that socio-cultural development as captured by SDI is affected by space and there exists neighbourhood effect whereas income (PCDDP), generally does not follow such event. Due to variations of geographical constraints and agro-climatic conditions, there is a variation in income. As a result, we could not find any systematic clustering and neighbourhood effect in respect of income (PCDDP). However, I strongly believe more research is needed in this direction.

Chapter-IV

Determinants of Social Development Index (SDI) and Per Capita District Domestic Product (PCDDP) in West Bengal: A Spatial Panel Data Regression Approach

4.1 Introduction

Spatial analysis remains incomplete if we do not run spatial regression. It is complementary analysis to Moran's Index. How far the development takes place over the spaces is captured by spatial regression model. How far the benefit of development of a specific geographical region percolates to its just neighbour and neighbours of neighbours can be captured by this analysis. The spill-over effect is more appropriately captured by spatial regression. This chapter deals with the problem of inference in models with spatial data. Basic and traditional regression is not similar to this regression as it allows the restriction part of independent and identical distribution of error term. Spatial regression deals with dependent variables as a measure spatial lag and also errors of that model.

Spatial regression models are used as Local Indicator for Spatial Autocorrelation, LISA (Anselin 1994, 1998 and 1999a, Haining 2003, Bailey and Gatrell 1995). Haining (2003) has introduced two spatial regression models: the "Spatial Lag Model (SLM)" and the "Spatial Error Model (SEM)". Spatial regression analyses is mainly conducted with data aggregated to geographic areas such as counties, regions etc. It is easy to find spatially auto-correlated residuals.

4.2 Objectives

My last and final objective is to study the determinants of spatial variation of SDI and PCDDP using Spatial Auto-Correlation (SAR) and Spatial Error Model (SEM) of regression. The SAR and SEM are more appropriate and scientific to capture the spatial neighbourhood impact.

4.3 Review of Literature

Chi and Zhu (2008) have analysed spatial autocorrelation for demographic transition. They have given a practical guide to spatial demographic analysis, with a focus on the use of spatial regression models. They also have reviewed spatial regression models including spatial lag models, spatial error models, and spatial autoregressive moving average models and used these models for analysing the data and have finally suggested opportunities and directions for future research on spatial demographic theories and practices.

Hession and Moore (2010) have used spatial regression analysis and checked for spatial autocorrelation in the study of climate change for Kenya. They have checked how the topographic variables such as elevation and slope strongly influence rainfall during the ‘long rains’ and ‘short rains’, which are vital for Kenyan agriculture.

Higazi, Abdel-Hady and Al-Oulfi (2013) have analysed the violation of the assumption of iid by two derived models that put contiguity of observations into consideration. The dependent variable is the percentage of individuals classified as poor; explanatory Spatial Data Analysis (ESDA) is performed to examine the existence of spatial clustering and spatial autocorrelation between neighbouring counties. Spatial analysis on poverty reflects many economic and living conditions (unemployment, illiteracy rate, average GDP, education drop-outs, access to sanitation facilities, dependency ratios, health care . . . etc) and have found that the Spatial Error Model (SEM) is better than the Spatial Lag Model (SLM).

4.4 Methodology and Data

4.4.1. Traditional Regression Model:

A. **Classical Model:** The linear-cross-sectional model

$Y = \beta X + u ....(4.1)$, where u is the classical error term

B. **Panel Data Linear Regression:**

Panel data are a type of longitudinal data, or data collected at different points in time. Panel data are repeated cross-sections over time, in essence there will be space as well as time dimensions and gives efficient estimation (Baltagi 2005). Main motivation of using panel data regression is to check unobserved heterogeneity in between the regions.

Panel data may useful for two or more observations (small t) on many units (large N) examples:

- Panel surveys of households and individuals.
- Data on organizations and firms at different time points.
- Aggregated regional data over time.

The regression equation as,

$$Y_{it} = \alpha_i + \beta X_{it} + U_{it} ......(4.2)$$

α_i is constant but different for each i.

and $U_{it} = \eta_i + \theta_t + \varepsilon_{it}$...........(4.3)

where, η_i is for number of regions, θ_t is for time period and ε_{it} is white noise.

Panel Data has two parts:

a) Fixed Effect.
b) Random Effect.

a) Fixed Effect (FE):

FE explore the relationship between predictor and outcome variables within an entity (Country, person, company, etc.). FE removes the effect of those time-invariant characteristics so we can assess the net effect of the predictors on the outcome variable. Each entity is different therefore the entity's error term and the constant which captures individual characteristics should not be correlated with the others. Technically, time-invariant characteristicsof the individuals are perfectly collinear with the person [or entity] dummies. Fixed effects capture the individual heterogeneity.

The estimation:

i) The least squares dummy variable estimator

ii) The fixed effects estimator

Introduce dummy variable to capture the difference in α_i's as, $\alpha_i, \alpha_2, ..., \alpha_N$ equation (4.2).

b) Random Effect:

The rationale behind random effects model is that, unlike the fixed effects model, the variation across entities is assumed to be random and uncorrelated with the predictor or independent variables included in the model:

"…the crucial distinction between fixed and random effects is whether the unobserved individual effect embodies elements that are correlated with the regressors in the model, not whether these effects are stochastic or not" [Green, 2008, p.183].

The random effect model is:

$$Y_{it} = \alpha + \beta X_{it} + (U_{it} + v_i)\text{.......}(4.4)$$

Where $w_{it} = (U_{it} + v_i)$ is the random effect of the model and v_i is the random part and it changes one part to another part.

There is an autocorrelation occurs,

$$\begin{aligned}
&\operatorname{cov}(w_{it}, w_{is}) = E(w_{it}, w_{is}) \\
&= E[(U_{it} + v_i) * (U_{is} + v_i) \\
&= E(U_{it} * U_{is}) + E(U_{it} * v_i) + E(U_{if} * v_i) + E(U_i)^2 \\
&= \sigma_v^2
\end{aligned}$$

If this autocorrelation problem is looked over, there will be a problem of inefficiency.

But, which model is to be chosen? Fixed Effect or Random Effect?

To decide between fixed or random effect we can run a Hausman test where the null hypothesis is that the preferred model is random effect vs. the alternative the fixed effects (Green, 2008, chapter 9). It basically tests whether the unique errors (U_i) are correlated with the regressors, the null hypothesis is they are not.

4.4.2. Spatial Regression Model:

C. **Spatial Autoregressive Model (SAR):** This model says that levels of the *dependent* variable *y* depend on the levels of *y* in neighbouring regions. The SAR model accounts for the presence of spatial autocorrelation in the dependent variable, but assumes spatial independence of the error terms It is thus a formulation of the idea of a spatial spill over. The formal model is:

$$Y = \rho WY + \beta X + u ....(4.5)$$

With u assumed to be classical error. Note that, *Wy* makes sense since the diagonal elements of W are zero, which implies that we do not have the circular specification that y_j on the left is influenced by the same y_j on the right. Clearly we would not want to run OLS on this model, since the presence of y on both the left and right sides means that we have a correlation-between-errors-and-regressors problem, and the resulting estimates will be biased and inconsistent. But we can easily obtain the reduced form as,

$$Y = \rho WY + X\beta + u$$
$$or, (I - \rho W)Y = \beta X + u$$
$$or, Y = (I - \rho W)^{-1} X\beta + (I - \rho W)^{-1} u ......(4.6)$$

Two inferences can be drawn:

- Firstly, the new error term $u^* = (I - \rho W)^{-1} u$ is no longer homoscedastic.
- Secondly, and probably more fundamentally, the model is no longer linear-in parameters because of the new unknown parameter ρ.
- $u \sim N(0, \sigma^2 I)$
- If $(I - \rho W) = A$ then, the equation will be, $AY = X\beta + u$
- The model is usually estimated by maximum likelihood. the log-likelihood function is,

$$\ln L(\beta, \rho, \sigma) = -(R/2)\ln \pi - (R/2)\ln \sigma^2 + \ln\|A\| - (1/2\sigma^2)(AY - X\beta)'(AY - X\beta)......(4.7)$$

where $\|A\|$ is the determinant of A. Anselin (1988) suggests a way to do the estimation. Focussing first on β, it is straightforward to show that the ML estimator is given by,

$$b = (X'X)^{-1} X'AY$$
$$or, b = (X'X)^{-1} X'Y - \rho(X'X)^{-1} X'WY$$
$$or, b = b_0 - \rho b_L$$

Where, $b_0 = (X'X)^{-1} X'Y$ and $b_L = (X'X)^{-1} X'WY$. It shows that b_0 is the coefficient vector from the OLS regression of Y on X, while b_L is from the OLS regression of WY on X. So, if ρ is known, we could compute the ML estimate of β:

Next, write the residuals of these two OLS regressions as,

$$e_0 = Y - Xb_0$$
$$e_L = WY - Xb_L$$

It can be shown that the ML estimate of σ^2 is,

$$s^2 = (1/R)(e_0 - \rho e_L)^{'}(e_0 - \rho e_l)$$

Therefore, once again we could estimate σ^2 if λ were known. We can now use all this to write down a version of the log-likelihood function in terms of ρ only: the result is the *concentrated* log-likelihood, ln L^* which can be seen to be,

$$\ln L^* = C - (R/2)\ln\left[(1/R)(e_0 - \rho e_L)^{'}(e_0 - \rho e_L)\right] + \ln\|A\| \text{........}(4.8)$$

where C does not involve any unknown parameters. We can now maximize ln L^* with respect to λ and obtain the ML estimate of this parameter, and work backwards.

In detail, the estimation steps are:

1) Regress Y on X : this gives b_0: Compute the residual $e_0 = Y - Xb_0$.
2) Regress WY on X : this gives b_L: Compute the residual $e_L = WY - Xb_L$.
3) Find the ρ that maximizes the concentrated log-likelihood function. Call it $\hat{\rho}$.
4) Given $\hat{\rho}$, compute $b = b_0 - \hat{\rho} b_L$ and $s^2 = (1/R)(e_0 - \hat{\rho} e_L)'(e_0 - \hat{\rho} e_L)$.

Note that steps (1) and (2) are simply OLS linear estimation problems; and that step (3) is a one-parameter nonlinear optimization problem.

One problem with this is that, due to the stepwise nature of the estimation process, we do not get estimates of the (joint) covariance matrix of all the estimated parameters. However, because they are maximum-likelihood estimates, we know that they are asymptotically efficient; meaning that for large samples the covariance matrix attains the Cramer-Rao lower bound, given by,

$$-\mathrm{E}(\delta^2 \ln L / \delta\theta\delta\theta')^{-1} \text{.......}(4.9)$$

where, $\theta = (\beta, \rho, \sigma^2)$: This in turn can be estimated by the numerical Hessian of the log likelihood.

D. **Spatial Error Model (SEM):** In this model, the spatial influence comes only through the error terms. The SEM allows for spatial dependence ofthe error terms. we have,

$$Y = \beta X + u$$
$$here, u = \lambda Wu + v......(4.10)$$

Where, $v =$ classical error term.

With v assumed to be normal with $E(v) = 0$; $E(v'v) = \delta^2 I$ (i.e. completely classical).
Solving the error specification for u we find,

$$(I - \lambda W)u = v$$
$$or, u = (I - \lambda W)^{-1} v......(4.11)$$

So putting eq.(4.10) on eq.(4.11), we get eq.(4.12),

$$Y = \beta X + (I - \lambda W)^{-1} v.....(4.12)$$

This is conceptually simpler than the SAR case because the only problems are heteroskedasticity and non-linearity in ρ.
If, $B = (I - \lambda W)$ then the log-likelihood is,

$$\ln L = -(R/2)\ln \pi - (R/2)\ln \sigma^2 + \ln\|B\| - (1/2\sigma^2)(Y - \beta X)' B' B(Y - \beta X)....(4.13)$$

Here, as previously described, Bv is heteroskedastic. We could estimate β using GLS, and an estimate of δ^2 is similar to the SAR case. The concentrated log-likelihood is,

$$\ln L^* = C - (R/2)\ln[(1/R)e' B' Be] + \ln\|I - \lambda W\|......(4.14)$$

where, $e = y - Xb_{GLS}$. The problem of b_{GLS} is that it itself depends on ρ (unlike the SAR case). Anselin suggests an iterative procedure, essentially as follows:

1) Regress y on X: Call the coefficient estimate b_{OLS} and compute the residual vector: $e = Y - Xb_{OLS}$.
2) Use this e in the concentrated log-likelihood, and optimize to find $\hat{\lambda}$.

3) Use $\hat{\lambda}$ to compute the GLS estimator b_{GLS} and then a new residual vector $e = Y - Xb_{GLS}$

4) First time or if the residuals have not converged: go back to step 2 and re-estimate λ : Otherwise: go to step 5.

5) At this point we have a converged estimate of λ (say, $\hat{\lambda}$) and the associated residual vector e, and a GLS estimator of β. We can now estimate σ^2 by: $(1/R)e'B'Be$.

E. **Spatial Durbin Model**: The spatial Durbin model is,

$$Y = X\beta + WX\theta + X \text{.....(4.15)}$$

which just adds average-neighbour values of the independent variables to the specification. Example: the level of crime in region j depends on the intensity of policing in j as well as on the intensity in neighbouring jurisdictions. Apart from potential problems of multi-collinearity (recall that row-wise, X and WX are for different regions because the diagonal elements of W are zero), this model poses no problems.

Two models SAR and SEM models were considered as alternatives to OLS models because of their ability to incorporate spatial autocorrelation by adding two terms: ρWy in the SAR model and λW*u* in the SEM model. Without these terms, the OLS estimation is to be overestimated and there is a problem of biasedness to capture the presence of spatial autocorrelation, resulting in biased estimates of the OLS coefficients and values such as the coefficient of determination (R^2).

4.4.3. Data Sources:

Data pertaining to Health indicators are drawn from DLHS (I-IV) and NFHS-4 surveys. Upper primary enrolment is drawn from DISE Data; data on enrolment for 1998-99 were extrapolated from FLR. Per Capita DDP are drawn from Economic Survey, Govt. of West Bengal.

4.4.3.1. Explanatory Variables and the application of Principle Component Analysis (PCA):

In order to explain the variations of SDI, we consider supply side factors only; I only consider publicly provided physical and social infrastructures. We broadly categorize two types of infrastructure- Physical and Social.

The explanatory variables included in regression are as follows:

1. Road length per area of the districts(a_1),
2. Net irrigated area out of total area of the districts (a_2),
3. Percentage of village electrified out of the total villages of the districts (a_3),
4. Number of bank facilities out of total 1000 population (a_4).

These explanatory variables are all normalised of their specific area and these variables are all physical infrastructure variables.

Again for social infrastructure variables are also normalised by district population and area and these are:

1. Primary schools (a_5) per cohort percentage out of total population,
2. Upper primary with secondary school (a_6) per cohort percentage out of total population,
3. Primary health centre, Block primary health centre out of 80,000 population of the districts (a_7)
4. Sub centre can access out of 80,000 populations of the districts. (a_8)

In case of physical infrastructure:

- Roads maintained by PWD, Zila-Parishad and Gram-Panchyayet are taken into account and these values are normalised by total area of the districts times 100 for getting percentage of these values. So it becomes km/km^2.
- For irrigation, we are normalising the value as a irrigated area divided by total area of the districts. Then we get net irrigated area.
- Third one is Electricity, it is normalised by total number of village whose have an access of electricity. So, it is village electrified.
- Last one is for number of banks access per 1000 population.

In case of social infrastructure:

- The primary schools are normalised by cohort population i.e. among the total population of the districts which percentage belongs to the age group of primary school. Among the total population only 9% are going to primary school age group.

Then by cohort population, the number of primary schools are normalised times 10000 populations.

- Same is for upper primary schools with higher secondary. Here, the cohort population is 12% of total population time 1000 populations.
- In case of primary health centre it is normalised by total population times 80000 population.
- Like as previous case sub centre also normalised by total population times 80000 population.

Out of these variables, we are trying to make two separate indices- one is physical infrastructure index(PII) and other one is social infrastructure index(SII).

We use the methodology of PCA (Principle Component Analysis) through which we can endogenously determine the weights of the factors. Following steps are adopted to find the best explanatory variable through PCA:

- Infrastructural parameters are included simultaneously in PCA and all infrastructural parameters are assumed to be highly collinear.
- We calculate the correlation matrix of six infrastructure parameters from the district level panel data.
- Only the first PC is taken into account because it absorbs the maximum variance of the data and it consists of six original variables but with differential weights.
- There is no need to consider the 2nd PC since it is orthogonal to the first one. Since all the infrastructure parameters are highly correlated, therefore, the 2nd PC becomes redundant (Nunnally and Bernstein 1994).
- The first principal component (or artificial variable) captures the maximum variability of the data and it is written as:

If we incorporate all the factors (indicators) simultaneously, then the first PC becomes:

$$PC_1 = W_{11}.P_1 + W_{12}.P_2 + W_{13}.P_3 + W_{14}.P_4 + W_{15}.P_5 + W_{16}.P_6 + W_{17}P_7 + W_{18}P_8 \text{........}(4.16)$$

From the above equation, W_{1i} is the factor loading of P_i where i=1,2……..8. Here I have taken 8 variables and among them four belongs to physical infrastructure and remaining four belongs to social infrastructure. Now, in order to get a clear picture of social and physical infrastructure separately on SDI, I consider the following methods:

Since the factor loadings represents the correlation coefficients between PC and original variables (P's), we go for pair-wise statistical testing using the following *t* test:

$$t = \frac{r\sqrt{(n-2)}}{\sqrt{(1-r^2)}} ....(4.17)$$

- It is revealed that all the eight explanatory variables are found to be statistically significant (as reported in the analysis), but in case of a_1 and a_3 carries negatively significant. Therefore, we retain all eight parameters in our analysis.
- In final stage, we calculate two indices as PII (Physical Infrastructure Index) and SII (Social Infrastructure Index) across 17 districts in five time points from panel data correlation matrix. So, the index becomes:

$$PII = W_{11}^{P} * a_1 + W_{12}^{P} * a_2 + W_{13}^{P} * a_3 + W_{14}^{P} * a_4 ....(4.18)$$

$$SII = W_{11}^{S} * a_5 + W_{12}^{S} * a_6 + W_{13}^{S} * a_7 + W_{14}^{S} * a_8 ....(4.19)$$

From those two equations W_{1i} are the respective factor loadings i=1,..4 and these values are constant over 5 time points for all districts in West Bengal; only a's do change across districts over time.

4.5 Empirical Findings

4.5.1. Traditional Regression:

Model-1: Panel Data Regression for Socio Economic Factors:

The regression equation of the panel data linear model is following:

For the physical infrastructure model,

$$SDI_{it} = \alpha_1 + \beta_1(ROAD_{it}) + \delta_1(IRRIGATION_{it}) + \phi_1(ELECTRICITY_{it}) + \varphi_1(BANK_{it}) + \varepsilon_{1it} .....(4.20)$$

For the social infrastructure model,

$$SDI_{it} = \alpha_2 + \beta_2(P.SCHOOL_{it}) + \delta_2(UP.SCHOOL_{it}) + \phi_2(PHC_{it}) + \varphi_2(SC_{it}) + \varepsilon_{2it} .....(4.21)$$

The above two regression models suffer from the problem of collinearity since the factors in each category of infrastructure are found to be highly collinear. This is why we construct, two indices namely PII and SII as outlined earlier.

Before going to the regression analysis by both these two equations (4.20) and (4.21) then we may face the problem of multi-collinearity as these explanatory or supple side variables are highly co-related to each other. So, to avoid this problem we create a synthetic index using PCA as outlined above.

The first principal component in case of physical infrastructure is reported as:

$$PC_{phys} = 0.563*\mathrm{P}_1+0.861*\mathrm{P}_2+0.724*\mathrm{P}_4 \text{..................} (4.22)$$

Where, P_1 = Road length, P_2 = Net Irrigated Area, P_3 = Village Electrified, P_4 = Bank Facilities. P_3 is excluded because it's factor loadings appears to be insignificant.

Similarly, first principal component of social infrastructure is shown by eqn. (4.23)

$$PC_{social} = 0.942*\mathrm{S}_1+0.813*\mathrm{S}_2+0.852*\mathrm{S}_3+0.943*\mathrm{S}_4 \text{............} (4.23)$$

And, S_1 =Number of Primary School, S_2 =Number of Upper Primary School, S_3 =Number of Primary Health Centre, S_4 = Number of Sub-Centre.

Using equation (4.22) and (4.23), we estimate physical infrastructure index (PII) and social infrastructure index (SII) across districts over time; while estimating principal component (PCs), we consider correlation matrix derived from panel data, therefore, the factor loadings are time and space invariant. It is observed that all the factor loadings are not statistically significant like village electrified as it gives insignificant result. So, we omit this variable as it is less important for physical infrastructure. And all the social infrastructure variables are statistically significant and also positive. So, all the variables of social infrastructure are taken into account.

Therefore, the main regression for the panel data is:

$$SDI_{it} = \alpha_i + \beta(PII_{it}) + \delta(SII_{it}) + U_{it} ....(4.24)$$

Where, $U_{it} = \eta_i + \theta_t + \varepsilon_{it}$ and ε_{it} is white noise.

The empirical findings from the above panel data regression is shown in Table-4.1.

Table:4.1 Model-1: Panel Data Regression for Socio Economic Factors.

Explanatory variables	Coefficients	Z	P>|z|
PII	0.0001838***	7.27	0.000
	(.0000255)		
SII	0.0003871***	3.33	0.001
	(.0001232)		
Constant	0.4875226***	24.93	0.000
	(.0195531)		
Within R squared	0.5399		
Between R squared	0.0383		
Overall R squared	0.0907		
F-test	8.10***		
	FE	RE	
PII	0.0001838	0.0001299	
SII	0.0003871	0.0004264	
Hausman value	23.11***		
No. of observation	85		

Note: 1. PII and SII are two independent variables. Within parenthesis is the standard error of the respective coefficients. 2.*,**,*** indicate significance at 10%, 5% and 1% levels respectively.

Here, we use panel least squares regression to investigate the existence of any relation between PII and SII on SDI. We divide the whole time period into 5 time points. From the above table, it clearly shows that there exists an impact on SDI. Here PII and SII are two explanatory variables which affect on SDI at a very positive and significant level. The F-statistics also shows that the regression is significant at a very high level. This is a result of Fixed Effect model as from Hausman Test we can conclude that fixed effect model is appropriate in this case as the value of the probability is high from Hausman test which can show from the Table(4.1).

From the above results, I find that FE model is appropriate which means that district specific characters do matter towards variations of SDI. This can be captured by LSDV Model but if we include district dummy, there will be a problem of degrees of freedom!

Model-2: Panel Data Regression for Economic Factor (PCDDP):

Our SDI is a non-monetary measure of social progress. Is it affected by income (PCDDP)? In order to understand the impact of income (PCDDP) on SDI, I run panel regression as shown in Eqn. (4.25).

$$SDI_{it} = \alpha_i + \mu PCDDP_{it} + \varepsilon_{it} .....(4.25)$$

Table:4.2 Model-2: Panel Data Regression for Economic Factors.

Explanatory variables	Coefficients	Z	P>|z|
PCDDP	0.000404***	14.08	0.000
	(.000287)		
Constant	0.5014449***	30.67	0.000
	(.0163501)		
Within R squared	0.7404		
Between R squared	0.3959		
Overall R squared	0.5921		
Wald Chi(2)	193.13***		
	FE	RE	
PCDDP	0.00004	0.000404	
Hausman	2.26***		
No. of observation	85		

Note: 1. PCDDP independent variables. Within parenthesis is the standard error of the respective coefficients. 2.*,**,*** indicate significance at 10%, 5% and 1% levels respectively.

Here, similarly we use panel least squares regression to investigate the existence of any relation between PCDDP on SDI. Again, we divide the whole time period into 5 time points. From the above Table(4.2), it clearly shows that there exists an impact on SDI. Here only one explanatory variable PCDDP which effect on SDI a very positive and significant level. Here also the F-statistics also shows that the regression is significant at a very high level. Also here is the result of Random effect model is appropriate as the value of Hausman test is high with higher probability as per Table(4.2).

4.5.2. Spatial Regression

From the above estimation we can see that SDI is space dependent. All districts are making clusters according to their similar districts. Again some dissimilar districts are also making clusters as they are nothing but an outlier. But in case of global Moran's I and local Moran's I we can interpret that districts have neighbourhood impact on an aggregate basis or individually. But the social indicators which are explanatory variables also influence on SDI and the explanatory variables are different in terms of physically and socially. They are named as PII (Physical Infrastructure Index) and SII (Social Infrastructure Index) as discussed above. In a panel data framework, it is obvious that PII and SII in the i^{th} region depend on those of neighbouring regions. If the data do not appear to follow normal distributions or the relations among the variables are not linear, we could consider transforming the variables. However, the transformation may not reduce spatial dependence if it exists (Bailey and Gatrell 1995). Alternatively additional variables such as higher-order terms and interaction terms can be incorporated. Baltagi and Pirotte (2010) showed that if the spatial dimension is neglected, the usual panel data estimators can be greatly affectedand considered various estimators using panel data seemingly unrelated regressions (SUR) with spatial error correlation. Thus we propose to use spatial panel models to take into account regional interdependences. We use two econometric models to study the regional impact on SDI. First one is spatial lag model and second one is spatial error model In particular, we review spatial autocorrelation, spatial heterogeneity, spatial weight matrix based on spatial neighbourhood structures, and discuss the modifiable areal unit problem. These concepts and issues are essential in spatial regression model. We assume that SDI in the i-th region at time t named as, SDI_{it} is generated according to the following linear spatial lag model (SAR) and spatial error model (SEM).

a. **Model-3: Spatial Lag Model of SDI:**

The spatial equation of this model is,

$$SDI_{it} = \rho \sum_{j=1}^{N} w_{ij} SDI_{jt} + \beta PII_{it} + \delta SII_{it} + \varepsilon_{it} .....(4.26)$$

Where i=1,2…..17 ; t=1,2….5 and $\varepsilon_{it} = \theta_i + u_{it}$

where ρ is the spatial autoregressive parameter or spatial lag parameter, β, the $(k \times 1)$ of the parameters, PII_{it} and SII_{it} , a $(1 \times k)$ vector of the explanatory variables. α_i an individual random effect associated to the ith region, IID(0, σ_α^2) , u_{it}, the remainder term IID(0,σ_u^2) and w_{if}, a generic element of spatial matrix W_N. More precisely, W_N is an $(N \times N)$ known spatial weights matrix which has zero diagonal elements. It characterizes the spatial regional interdependences.

Following equation(4.26), the N-dimensional vector of observations on the dependent variable SDI_t for time t is given by:

$$SDI_t = \rho W_N SDI_t + \beta PII_t + \delta SII_t + \varepsilon_t$$
$$or, SDI_t = (I_N - \rho W_N)^{-1} PII_t \beta + (I_N - \rho W_N)^{-1} SII_t \delta + (I_N - \rho W_N)^{-1} \varepsilon_t .....(4.27)$$

Where I_N is an identity matrix of order N, $PII_t = (PII_{1t}^{'},........, PII_{Nt}^{'})'$ is an $(N \times k)$ matrix and $\varepsilon_t = (\varepsilon_{1t}^{'},.............,\varepsilon_{Nt}^{'})'$ is an $(N \times 1)$ vector.

So finally,

$$(I_N - \rho W_N)^{-1} = (I_N + \rho W_N + \rho^2 W_N^2 + \rho^3 W_N^3 + ...\infty)......(4.28)$$

The first term of equation(4.28) corresponds to no neighbor, the second term to immediate neighbors (first order), the third to neighbors of neighbors (second-order), and so on. More prominently as power of ρ is increasing then we can appropriately capture this neighborhood effect as power of ρ is one then one region directly effects to its neighboring region and power of ρ is 2 the that region indirectly effects its neighbors of neighbor region. So as the power of ρ is increasing then we can capture this spatial variation. Thus in the spatial lag model, the interpretation of parameters becomes richer and more complicated. A change in the explanatory variable for a single region can potentially affect the dependent variable in all other regions. The spatial lag model expands the information set to include information from neighboring regions. Several researchers have noted that this kind of specification requires a special interpretation of the parameters (Anselin). A significant spatial lag term may indicate strong spatial dependence. A significant spatial error term indicates spatial autocorrelation in errors, which may be due to key explanatory variables that are not included in the model (Pirotte and Madre 2011).

The following table is witnessed for spatial lag model:

Table:4.3 Model-3: Panel Data Spatial Lag Regression.

Explanatory variables	Coefficients	Z	P>|z|
PII	0.0000591^{***} (.0000284)	2.07	0.039
SII	0.0003041^{***} (.0001273)	2.39	0.017
Constant	0.4846917^{***} (.0372055)	13.03	0.000
Rho(ρ)	0.002396^{***} (.0008443)	2.84	0.005
Log likelihood	77.966814		
	Z chi2(1)	P>|z|	
Wald test (ρ=0)	8.054	0.005	
Likelihood ratio test (ρ=0)	7.695	0.006	
Lagrange multiplier test (ρ=0)	8.76	0.004	
Range	-2.84 to 1		
No. of observation	85		

Note:1. PII and SII are two independent variables. Within parenthesis is the standard error of the respective coefficients.2.*,**,*** indicate significance at 10%, 5% and 1% levels respectively.

From this table we can see that both the explanatory variables PII and SII are spatially dependent at a very high significant level and it is also positively related with SDI. It means that the explanatory variables are space dependent and also moves on same direction as SDImoves. The constant parameter is in this model name as rho (ρ) also positive and significant at a very high level and the value of rho is at a very acceptable range.

b. Model-4: Spatial Lag Model of PCDDP:

The equation of spatial lag model is,

$$SDI_{it} = \rho \sum_{j=1}^{N} w_{ij} SDI_{jt} + \beta PCDDP_{it} + \varepsilon_{it} ......(4.29)$$

Where i=1,2…..17 ; t=1,2….5

and $\varepsilon_{it} = \theta_i + u_{it}$

where ρ is the spatial autoregressive parameter, β, the $(k \times 1)$ of the parameters, $PCDDP_{it}$ a $(1 \times k)$ vector of the explanatory variables. α_i an individual random effect associated to the i^{th} region, IID(0, σ_α^2) , u_{it}, the remainder term IID(0,σ_u^2) and w_{if}, a generic element of spatial matrix W_N. More precisely, W_N is an $(N \times N)$ known spatial weights matrix which has zero diagonal elements. It characterizes the spatial regional interdependences.

Following equation(38), the N-dimensional vector of observations on the dependent variable SDI_t for time t is given by:

$$SDI_t = \rho W_N SDI_t + \beta PCDDP_t + \varepsilon_t$$

$$or, SDI_t = (I_N - \rho W_N)^{-1} PCDDP_t \beta + (I_N - \rho W_N)^{-1} \varepsilon_t ....(4.30)$$

Where I_N is an identity matrix of order N, $PCDDP_t = (PCDDP_{1t}^{'},........,PCDDP_{Nt}^{'})'$ is an $(N \times k)$ matrix and $\varepsilon_t = (\varepsilon_{1t}^{'},.............,\varepsilon_{Nt}^{'})'$ is an $(N \times 1)$ vector.

So finally,

$$(I_N - \rho W_N)^{-1} = (I_N + \rho W_N + \rho^2 W_N^2 + \rho^3 W_N^3 + ... + \infty)......(4.31)$$

The first term of equation(4.31) corresponds to no neighbor, the second term to immediate neighbors (first order), the third to neighbors of neighbors (second-order), and so on. More specifically, the second value of this equation gives that one region shares its border with just its nearest neighbor mean that region is directly affected by its neighboring region; again for third case one region shares its border indirectly with its neighbor’s of neighbor region. Thus in the spatial lag model, the interpretation of parameters becomes richer and more complicated. As an increase in the power of ρ, more appropriate to capture neighbouhood impact. A change in the explanatory variable for a single region can potentially affect the dependent variable in all other regions. The similar explanation is also true as before. A significant spatial lag term may indicate strong spatial dependence.

The following table shows the results of spatial lag model:

Table:4.4 Model-4: Panel Data Spatial Lag Regression.

Explanatory variables	Coefficients	Z	P>|z|
PCDDP	0.000413*** (.0000360)	11.47	0.000
Constant	0.4271758*** (0.233177)	18.32	0.000
Rho(ρ)	0.001854*** (.0005197)	3.57	0.000
Log likelihood	112.73595		
	Z chi2(1)	P>|z|	
Wald test (ρ=0)	12.727	0.000	
Likelihood ratio test (ρ=0)	11.860	0.001	
Lagrange multiplier test (ρ=0)	12.828	0.001	
Range	-2.84 to 1		
No. of observation	85		

Note:1. PII and SII are two independent variables. Within parenthesis is the standard error of the respective coefficients.2.*,**,*** indicate significance at 10%, 5% and 1% levels respectively.

From this table we can see that the variable, PCDDP is spatially dependent at a very high level of significance. It means that the explanatory variables are space dependent and also moves on same direction as SDI moves. The results suggest that not only infrastructural factors rather income also necessary for social development. The constant parameter, rho (ρ) also appears positive and significant at a very high level.

c. Model-5 : Spatial Error Model of SDI:

The equation of spatial error model is,

$$SDI_{it} = \beta PII_{jt} + \delta SII_{jt} + u_{it} ....(4.32)$$

Where, i=1,2…….17 and t=1….5

and $u_{it} = \lambda w_{ij} u_{jt} + \varepsilon_{it}$

Here, SDI is the dependent variable with $(N \times 1)$ vector, PII and SII are the two explanatory variables with $(N \times k)$ vector, w is spatial weight matrix and λ is a coefficient on the spatially correlated errors or spatial error parameter and the parameters β and δ reflects the influence of the explanatory variables on variation in the dependent variable with $(k \times 1)$ vector.

So following the above equation (4.32), N-dimensional vector of observations on the dependent variable SDI_t for time t is given by:

$$SDI_{it} = \beta PII_{it} + \delta SII_{it} + \lambda W_N u_{jt} + \varepsilon_{it} ......(4.33)$$

The following table shows for spatial lag model:

Table:4.5 Model-5: Panel Data Spatial Error Regression.

Explanatory variables	Coefficients	Z	P>|z|
PII	0.0000626*** (.0000294)	2.13	0.033
SII	0.0002827*** (.0001258)	2.25	0.025
Constant	0.4868184*** (.0372055)	13.38	0.000
Lamda(λ)	0.002659*** (.0009167)	2.90	0.004
Log likelihood	77.929399		
	Z chi2(1)	P>|z|	
Wald test (λ=0)	8.413	0.004	
Likelihood ratio test (λ=0)	7.620	0.006	
Lagrange multiplier test (λ=0)	2.982	0.084	
Range	-2.84 to 1		
No. of observation	85		

Note:1. PII and SII are two independent variables. Within parenthesis is the standard error of the respective coefficients.2.*,**,*** indicate significance at 10%, 5% and 1% levels respectively.

A number of other asymptotic approaches exist for testing whether spatial correlation is present in the residuals from a least-squares regression model. Some of those are the likelihood ratio test, the Wald test and lagrange multiplier test, all of which are based on maximum likelihood estimation of the SEM model. The likelihood ratio test is based on the difference between the log likelihood from the SEM model and log likelihood from a least squares regression. This quantity represents a statistic that is distributed $\chi^2(1)$. Finally the Lagrange Multiplier (LM) test which is based on the least-squares residuals and calculations involve the spatial weight matrix W. The LM statistic takes the from: (Anselin, 1998,page 104), where e denote least square residuals.

We can infer from the regression table is that the least squares residuals exhibit spatial correlation. We can say from this a small marginal probability that indicate significance at the 99% level of confidence.

From this table we can see that both the explanatory variables PII and SII are spatially dependent at a very high significant level and it is also positively related with SDI. It means that the explanatory variables are space dependent and also moves on same direction as SDI moves. The constant parameter is in this model name as rho (ρ) also positive and significant at a very high level and the value of rho is at a very acceptable range.

d. Model-6 : Spatial Error Model of PCDDP:

The equation of spatial error model is,

$$SDI_{it} = \beta PCDDP_{jt} + u_{it} \text{....}(4.34)$$

Where, i=1,2…….17 and t=1….5

and $u_{it} = \lambda w_{ij} u_{jt} + \varepsilon_{it}$

Here, SDI is the dependent variable with $(N \times 1)$ vector, PCDDP are the two explanatory variables with $(N \times k)$ vector, w is spatial weight matrix and λ is a coefficient on the spatially correlated errors and the parameters β and δ reflects the influence of the explanatory variables on variation in the dependent variable with $(k \times 1)$ vector.

So following the above equation (4.34), N-dimensional vector of observations on the dependent variable SDI_t for time t is given by:

$$SDI_{it} = \beta PCDDP_{it} + \lambda W_N u_{jt} + \varepsilon_{it} ....(4.35)$$

A number of other asymptotic approaches exist for testing whether spatial correlation is present in the residuals from a least-squares regression model. Some of those are the likelihood ratio test, the Wald test and lagrange multiplier test, all of which are based on maximum likelihood estimation of the SEM model. The likelihood ratio test is based on the difference between the log likelihood from the SEM model and log likelihood from a least squares regression. This quantity represents a statistic that is distributed $\chi^2(1)$. Finally the Lagrange Multiplier (LM) test which is based on the least-squares residuals and calculations involve the spatial weight matrix W. The LM statistic takes the from: (Anselin, 1998,page 104), where e denote least square residuals.

We can infer from the regression table is that the least squares residuals exhibit spatial correlation. We can say from this a small marginal probability that indicate significance at the 99% level of confidence.

The following table is witnessed for spatial lag model:

Table:4.6 Model-6: Panel Data Spatial Error Regression.

Explanatory variables	Coefficients	Z	P>|z|
PCDDP	0.000412*** (.000361)	11.42	0.00
Constant	0.4295535*** (.0231705)	18.54	0.000
Lamda(λ)	0.002339*** (.0006645)	3.51	0.000
Log likelihood	112.50071		
	Z chi2(1)	P>|z|	
Wald test (λ=0)	12.337	0.000	
Likelihood ratio test (λ=0)	11.389	0.001	
Lagrange multiplier test (λ=0)	8.268	0.004	
Range	-2.84 to 1		
No. of observation	85		

Note:1. PII and SII are two independent variables. Within parenthesis is the standard error of the respective coefficients.2.*,**,*** indicate significance at 10%, 5% and 1% levels respectively.

From this table we can see that both the explanatory variables PCDDP is spatially dependent at a very high significant level and it is also positively related with SDI. It means that the explanatory variables are space dependent and also moves on same direction as SDI moves. The constant parameter is in this model name as rho (ρ) also positive and significant at a very high level and the value of rho is at a very acceptable range.

4.6 Findings:

Both the two spatial regression models like SLM and SEM clearly prove that space does matter towards variations of SDI. The neighbouring impact of SDI is captured by both the models. The SDI is not only determined by the physical as well as social infrastructure variables but also is influenced by the level of income (PCDDP).

Chapter-V

CONLUSIONS AND POLICY IMPLICATIONS

I construct a new measure of social development, known as Social Development Index (SDI) across 17 districts over 5 time points using DLHS (I-IV) and NFHS (IV) data. I find that majority of the districts have performed well in raising SDI but few districts like Malda, Uttar Dinajpur, Murshidabad and Purulia could not perform well in raising their SDI over time. It is observed that there exist spatial variations of SDI as well as income (PCDDP). However, the inequality of SDI is found to be declining, giving the notion of sigma convergence except the last time point (2015-16 corresponding to NFHS-IV). This may be the due to differences of sample selection of DLHS and NFHS data. We did not perform the beta convergence but since over time the GE measure of inequality is falling, therefore, it indirectly support the notion of beta convergence in SDI. The SDI is created from those parameters which have asymptotic upper values, this is another point for ensuring beta convergence. The inequality of districts manifesting higher value of SDI is more pronounced compared to lower and middle of the distributions of SDI, this is because GE(2) >GE(1) and GE(0). Income inequality across the districts is found to be erratic, no systematic pattern is noticed, therefore, we can not draw any definite conclusions about sigma convergence in respect of income (PCDDP).

The SDI is influenced by space and neighbour and it is proved using GMI and LMI. Moreover, the spatial econometric methods clearly prove that SDI is significantly influenced by neighing districts. I have explored in details about the impact of publicly provided infrastructure on SDI and income (PCDDP) in panel data regression; find that both physical and social infrastructure play a significant role towards variations of SDI. It is interesting to note that SDI is purely non-monetary measure of social progress but it is highly influenced by monetary variable like income (PCDDP). This means that districts having higher income (PCDDP) generally, on an average, manifest higher value of SDI. But, I could not run both way causality; even the problem of endogeneity is not captured in my econometric models.

The chronic and persistence low level social development trap is found to exist among the four districts (out of 17) like Malda, Uttar Dinajpur, Murshidabad and Purulia. This is an interesting observation and it requires an in-depth study. Why these four districts do fail to improve their SDI over time? I could not study specifically about these districts; however, the demographic characteristics of Malda, Uttar Dinajpur and Murshidabad are more or less identical. The concentration of Muslim community is found to be higher than the West Bengal average. Moreover, these four districts do share common international border with Bangladesh. But, the other district like Purulia belong to western part of WB sharing border with Jharkhand and Bihar(the poorer states in India); moreover, a considerable share of population in Purulia is Scheduled Tribe(ST), who are not only poor in respect of income but they are poor in respect of socio-economic status also.

In West Bengal, there is high degree of occupational dependence on agriculture, especially in terms of agricultural labour. Generally, economic status of the agricultural labour is poor; various development programmes have already been initiated in order to earadicate poverty but the results are not satisfactory. I suggest to strengthen the ongoing affirmative actions plan in one hand and to specifically allocate more development grants among the four districts (viz. Malda, Uttar Dinajpur, Murshidabad and Purulia). Publicly provided infrastructure is not equally distributed, therefore, some equitable distribution of physical and social infrastructure is highly needed to ensure social justice and equity.

Again in case of human development, the health status of women and children assume a special importance. Majority of women is suffering from anaemia and also for underweight. According to human resources, medical and para-medical facilities are one kind of important manpower for physical infrastructure. On the other hand, education is both a constituent and instrumental component of human development which has a significant effect on life expectancy, infant mortality, nutritional status and environmental awareness. The Government of West Bengal has implemented the District Primary Education Programme (DPEP) since 1997 and it has carried out Sarva Shiksha Abhiyan (SSA) to universalise elementary education (UEE) since 2001-02 in all the districts of the state. The recent project especially for women, taken by Government of West Bengal from 2013 is ‘Kanyashree’. This programme is meant for improving both health and education of girls of Bengal by supporting them financially. This project is spreading world-wide and the main agenda of this project is reaching the poorest and most vulnerable through inclusive services and participation. Indeed, it is a welcome move by the Government but we have to wait to get the desired results of this project.

Limitations:

1. The main limitation of this study is that the collection of data from two sources-DLHS and NFHS. I acknowledge that the two surveys are completely different in respect of sample design and methodology.
2. My second limitation of this research project is that I am not able to take those variables like IMR, Total Fertility Rate(TFR), Life Expectancy (LE) for capturing health dimension because of lack of availability of data at the district level.
3. Here, I could not take other explanatory variables like urbanization, industrialization, concentration of backward community, concentration of Muslim Community, agricultural situation etc. in the regression of SDI.
4. The problem of endogeneity between SDI and PCDDP is not addressed in regression of SDI. Moreover, convergence study is weak; I could not examine beta and club convergence. I could not incorporate the influence of international border to study GMI and LMI. I strongly believe that there is further scope of research towards existence of low SDI among those four districts.

Bibliography

Alkire, S., Robles, G.(2017): Global Multidimensional Poverty Index, OPHI, Oxford International Development, OUP.

Anand, S. and Sen, A. Human Development Index: Methodology and Measurement. paper for Human Development Report 1993. New York: UNDP, 1992.

Anselin, Luc. (1988). Spatial Econometrics, Methods and Models. Dordrecht: Kluwer Academic.

Anselin, Luc (1988a). "Model Validation in Spatial Econometrics: A Review and Evaluation of Alternative Approaches*", International Regional Science Review* 11, 279-316.

Anselin, L. (1992), *Space Stat*: *A program for the analysis of spatial data*. National Centre for Geographic Information and Analysis, *University of California, Santa Barbara*.

Anselin, Luc (1994). "Exploratory Spatial Data Analysis and Geographic Information Systems."PP. 45-54 in New Tools for Spatial Analysis, edited by M. Painho. Luxembourg: Euro-Stat.

Anselin, L. (1995). "Local Indicators of Spatial Association—LISA", Geographical Analysis 27:93-115.

Anselin, Luc (1996), "The Moran Scatter Plot as an ESDA Tool to Assess Local Instability in Spatial Association." Pp. 111- 125 in Manfred Fischer, Henk J. Scholten and David Unwin (eds.) Spatial Analytical Perspectives on GIS. London: Taylor & Frances.

Anselin, Luc and Anil Bera. (1998). "Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics." PP. 237-289 in AmanUllah and David Giles (eds.) Handbook of Applied economic Statistics. New York: Marcel Dekker.

Anselin, L. (1999). "The Future of Spatial Analysis in the Social Sciences". Geographic Information Sciences 5(2): 67-76.

Anselin, Luc (1999a). "Interactive Techniques and Exploratory Spatial Data Analysis." Pp. 251-264 in Geographic Information Systems: Principles, Techniques, Management and Applications, edited by P.A. Longley, M.F. Goodchild, D.J.(1992)

Anselin, Luc. (2003). An Introduction to Spatial Autocorrelation Analysis with GeoDa. Spatial Analysis Laboratory, University of Illinois.

Antony, G et. al(2001):" Suitability of HDI for Assessing Health and Nutritional Status", Economic and Political Weekly, Vol. 36, No. 31 (Aug. 4-10, 2001), pp. 2976-2979.

Ahmed (2011); ‘Does Economic Geography Matter for Pakistan? A Spatial Exploratory Analysis of Income and Education Inequalities. The Pakistan Development Review, Vol. 50, No.4, Paper and Proceedings Parts I and II, pp. 929-952.

Bailey, T.C. and A.C. Gatrell.(1995). "Interactive Spatial Data Analysis". New York: John Wiley

Baltagi, Badi (2005), Econometric Analysis of Panel Data- — 3rd ed. John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England

Baltagi, B. and D. Li (2006), ‘Prediction in the panel data model with spatial correlation: The case of liquor’. Spatial Economic Analysis 1, 175–185.

Baltagi, Badi H. and Pirotte, Alain, Seemingly Unrelated Regressions with Spatial Error Components (September 1, 2010). Center for Policy Research Working Paper No. 125. Available at SSRN: https://ssrn.com/abstract=1805888 or http://dx.doi.org/10.2139/ssrn.1805888.

Bhattacharya, G. and Haldar, S.K.(2014): “Trend, Differential and Determinants of Deprivation of Reproductive and Child Health in the Districts of West Bengal”, Journal of Health Management, 16(1), March.

Boettke, P.j.-Coyne, C.J.-Leeson, P.T. (2001); ‘Institutional Stickness and the New Development Economics. American Journal of Economics and Sociology 67, 2:331-358.

Census, 2011 and 2011.

Chattopadhyay, S.(2011): Inter-regional Poverty Comparisons: Case of West Bengal, Journal of Quantitative Economics, 2011, 9(2), 104-122.

Chamarbagwala, R. (2009): ‘Social Interactions, Spatial Dependence, and Children's Activities: Evidence from India’, The Journal of Developing Areas, Vol. 42, No. 2 (Spring, 2009), pp. 157-178, College of Business, Tennessee State University.

Chi, G. and Zhu, J. (2008): “Spatial Regression Models for Demographic Analysis”, Population Research and Policy Review, Vol. 27, No. 1, Spatial Demography. Part II(February 2008), pp. 17-42, Springer in cooperation with the Southern Demographic Association.

Clark, C. (1965): ‘The Fundamental Problems of Economic Growth’, WeltwirtschaftlichesArchiv, Bd. 94 (1965), pp. 1-9, Springer.

Cowell, F.(1977). Measuring Inequality. Philip Allan. Oxford, UK. models. Moscow-Izhevsk: RHD, 12, 13.

………….. (2003). Theil, inequality and the structure of income distribution.

District Level Household Survey (DLHS), (I-IV): IIPS, Mumbai

District Information System for Education (DISE)

Estes, R. (1997): Social Development Trends in Europe, 1970-1994: Development Prospects for the New Europe, Social Indicators Research, Vol. 42, No. 1 (1997), pp. 1-19, Springer.

Krugman, P. R. (1991), What's New About The New Economic Geography?, Oxford Review of Economic Policy, VOL. 14, NO. 2.

Fujita, M., Krugman, P. R., &Venables, A. J. (2001). The spatial economy: Cities, regions, and international trade. MIT press.

Green. (2008), Panel Data Analysis Fixed and Random Effects using Stata (v. 4.2), Princeton University, https://www.princeton.edu/~otorres/Panel101.pdf, p.183

Greif, A. (1994): Cultural Beliefs and the Organisation of Society: A Historical and Theoretical Reflection on Collectivist and Individualist Societies. Journal of Political Economy 102. 5:912-950.

Haining, R. (2003) Spatial Data Analysis: Theory and Practice. Cambridge University Press, Cambridge. http://dx.doi.org/10.1017/CBO9780511754944.

Hession, S.L and Moore, N (2010), A spatial regression analysis of the influence of topography on monthly rainfall in East Africa, INTERNATIONAL JOURNAL OF CLIMATOLOGY Int. J. Climatol. 31: 1440–1456 (2011).

Higazi, S. et. al (2013):Application of Spatial Regression Models to Income Poverty Ratios in Middle Delta Contiguous Counties in Egypt", Vol. IX No.1 2013 pp. 93-110.

Kapas, J (2017): "How Cultural Values Affect Economic Growth: A Critical Assessment of the Literature", JEL: O43,Z19,

Khasnabis, R (2009); 'The Economy of West Bengal', Vol. 43, No. 52 (Dec. 27, 2008 - Jan. 2, 2009), pp.103-115, Economic and Political Weekly.

Krugman, P. (2010), Prepared for presentation to the Association of American Geographers, https://www.princeton.edu/~pkrugman/aag.pdf

Krugman, P. (1998), 'WHAT'S NEW ABOUT THE NEW ECONOMIC GEOGRAPHY?' OXFORD REVIEW OF ECONOMIC POLICY, VOL. 14, NO. 2.

Landes, D. (2000), Culture Makes Almost All the Difference, In: Harrison, L.E.-Huntington, S.p. (eds), Culture Matters. New York, NY: Basic Books.

Litchfield, J.A. (1999) 'Inequality: Methods and Tools', Text for World Bank's Web Site on Inequality, Poverty.

Mishra, S. et.al (2014); Measuring HDI – The Old, the New and the Elegant: Implications for multidimensional development and social inclusiveness, working paper 63, Asia research centrehttp://www.lse.ac.uk/asiaResearchCentre/_files/ARCWP63-MishraNathan.pdf

National Family Household Survey (IV): IIPS, Mumbai

Natoli, R. and Zuhair, S. (2011): "Measuring Progress: A Comparison of the GDP, HDI, GS and the RIE", Social Indicators Research, Vol. 103, No. 1 (August 2011), pp. 33-56, Springer.

Ntibagirirwa, S. (2009). 'Cultural Values, Economic Growth and Development', Journal of Business Ethics, Vol. 84, Supplement 3: Global and Contextual Valuesfor Business in a Changing World (2009), pp. 297-311, Springer.

Nunnally, J.C. and Bernstein, I.H. (1994) The Assessment of Reliability. Psychometric Theory, 3, 248-292.

Ord, J.K. and A. Getis. (1995). "Local Spatial Autocorrelation Statistics: Distributional Issues and an Application". Geographical Analysis 27: 286-306.

Pirotte, A and Madre, J.L.(2011); 'Car Traffic Elasticities: A Spatial Panel Data Analysis of French Regions', University of Bath, Journal of Transport Eonomivs and Policy, Vol. 45, No. 3 (September, 2011), pp. 341-365.

Rende, S and Donduran, M (2013).Neighborhoods in Development: Human Development Index and Self-organizing Maps, Social Indicators Research, Vol. 110, No. 2 (2013), pp. 721-734, Springer.

Rondinelli, D. and Ruddle, K. (1997): Integrating spatial development, Ekistics, Vol. 43, No. 257, PLANNING FOR RURAL AREAS (APRIL 1977), pp. 184-194, Athens Centre of Ekistics.

RoyChaudhuri, A &Haldar, S. (2009): An Investigation into the Inter-District Disparity in West Bengal, 1991-2005, Economic & Political Weekly, June 27, 2009 Vol. xliv No.s 26 & 27.

Sachs, J. et. al (2016): 'Sustainable Development Solutions Network', Sustainable Development Solutions Network (2016), https://www.jstor.org/stable/resrep15881.8.

Schmutzler, A. (1999), 'THE NEW ECONOMIC GEOGRAPHY', JOURNAL OF ECONOMIC SURVEYS Vol. 13, No. 4.

Sharma,S. (1997). "Making the Human Development Index (HDI) Gender-Sensitive", Source: Gender and Development, Vol. 5, No. 1 (Feb., 1997), pp. 60-61, Taylor & Francis.

Shoham, A et all (2011): "Society's Level of Literacy: A Cross Cultural Study", Issues in Informing Science and Information Technology, Vol.8,2011.

Sen, A. (2002); How Does Cultural Matter? In: Vijayendra, R.- Walton, M.(eds), Cultural and Public Action, Stanford: Stanford University Press. pp.37-58.

Sen, A and Modak, P (2017), An analysis of the determinants of child marriage in West Bengal and the rest of India, Working paper, Department of Economics, University of Calcutta. Forthcoming.

Sen, A. and Dutta, A. (2018): 'West Bengal's Successful KanyashreePrakalpa Programme Needs More Push From State and Beneficiaries', Vol. 53, Issue No. 17, 28 Apr, 2018, Economic and Political Weekly.

Smith, A. (1759),'The Theory of Moral Sentiments.' London: A. Millar. Available at: http://www.econlib.org/library/Smith/smMS0.html

Tobler W., (1970) "A computer movie simulating urban growth in the Detroit region". Economic Geography, 46(Supplement): 234–240.

United Nation Development Programme (UNDP), 1990

United Nation Development Programme (UNDP), 2010

Vyasulu, V and Vani B.P. (1997)."Development and Deprivation in Karnataka: A District-Level Study", Economic and Political Weekly, Vol. 32, No. 46 (Nov. 15-21, 1997), pp. 2970-2975.

Weber, M (1930), 'The Protestant Ethic and the Spirit of Capitalism.Lnondon, New York: Routledge, 2001.

West Bengal Human Development Report, (2004), West Bengal.

Wu, P. et.al (2014). 'Does Human Development Index Provide Rational Development Rankings? Evidence from Efficiency Rankings in Super Efficiency Model", Social Indicators Research, Vol. 116, No. 2 (April 2014), pp. 647-658, Springer.

Zhang and Lin (2015); 'On Moran's I Coefficient under Heterogeneity', www.elsevier.com/locate/csda

Appendix- A: Data Set of All Variables over Five Time Points (1998-99 to 2015-16)

Districts	Time	HOBL3	SD	PWANC3	PCCI	CLR	NER(UP)	FLR	GMA18
Bankura	DLHS1	68.4	66.3	81	67.3	59.98	31.5177	49.8	50
	DLHS2	76	66.4	74.8	67.4	65.47	45.1	52.88	46
	DLHS3	85.6	61.4	65.8	91.7	68.23	58	56.975	50.3
	DLHS4	88.3	88.9	86.1	88.9	70.3	69.2	62.11	60
	DLHS5	85.8	85.6	89.5	96.2	74.85536	79.95	65.2	61
Bardhman	DLHS1	67.3	52	80.8	51.8	67.71	26.81789	61.93	45.3
	DLHS2	72.9	62.2	63.6	60.2	72	41.8	64.24	41
	DLHS3	84.3	58	70.7	63.8	74.4	49	67.32	60.6
	DLHS4	79.9	71.6	73.9	70.7	76.2	58.4	68.418	65.7
	DLHS5	85.9	82.6	83.3	82.3	79.51763	74.64	66.6	58.8
Birbhum	DLHS1	63.9	68.5	77.2	34.9	57.63	31.38037	52.21	40.7
	DLHS2	65.7	52.8	63	43.2	64.26	42.2	55.789	40.2
	DLHS3	80.1	48.7	58	65.8	67.94	56	60.561	41.7
	DLHS4	81.8	78.3	79.7	67	70.7	70.1	63.324	64.8
	DLHS5	91.7	86.3	78.1	91.4	75.85619	76.88	62.1	48.7
Dakhin Dinajpur	DLHS1	66.7	41.6	77.2	40.5	58.44	33.28054	55.12	34.5
	DLHS2	69.5	50	68.2	60.2	66.36	42.5	58.687	42.4
	DLHS3	82.9	40.3	77.6	88.9	70.04	61	63.443	51.3
	DLHS4	86.5	74.3	85.5	86.6	72.8	76.7	67.126	67.9
	DLHS5	87.4	79.5	69.6	83.2	79.67176	76.85	67.3	54.9
Darjeeling	DLHS1	68.1	60.4	72.2	60.8	67.66	14.29037	63.92	75
	DLHS2	67.9	47.5	51.3	57	74.14	16.4	66.743	79
	DLHS3	86.7	72.4	70.4	86.2	77.26	21	70.507	74.9
	DLHS4	82.9	86.1	87.2	86.5	79.6	25.3	75.198	89.5
	DLHS5	88.1	94.5	65.9	84.2	85.11382	23.79	78	78.1
Howrah	DLHS1	63.3	71.9	83.7	56.1	74.18	35.37142	70.93	62.2
	DLHS2	74.5	65.7	77.9	58.3	77.9	41.6	73.48	74.5
	DLHS3	83.4	65.7	79.8	76.4	79.1	51	76.88	67.8
	DLHS4	83.6	87.4	93.4	82.4	80	61.3	79.018	79.4
	DLHS5	86.8	86.6	86.6	73.8	83.75341	71.3	78.4	74.4
Hooghly	DLHS1	77.1	68.4	87	67.8	72.61	26.79126	67.72	66.4
	DLHS2	76.4	80.3	78.1	73.6	77.11	34.3	70.312	65.1
	DLHS3	89.9	80.1	82.8	92.9	79.79	45	73.768	73.9
	DLHS4	91.4	88.8	77.9	83.5	81.8	59.5	76.336	75
	DLHS5	90.3	91.3	76.6	84.4	85.11835	69.47	76.3	68.1
Jalpaiguri	DLHS1	61.3	35.5	71.6	62	57.56	45.83477	52.9	64
	DLHS2	64.5	44.4	61.6	69.5	66.02	40.7	56.899	62
	DLHS3	76.6	48.4	72.7	78.4	70.18	55	62.231	82.5
	DLHS4	79.4	75.2	76.2	81.8	73.3	74.9	65.418	83.1
	DLHS5	83.1	84	80.7	81.7	80.41238	45.29	64.2	65.5
Kochbihar	DLHS1	58.1	30.3	81	49.8	60.15	29.59672	57.04	47.9
	DLHS2	67.1	40.1	47	53.4	68.37	49.4	60.475	48.3

Districts	Time	HOBL3	SD	PWANC3	PCCI	CLR	NER(UP)	FLR	GMA18
	DLHS3	78.3	46.4	59	78	71.13	63	76.505	54.2
	DLHS4	92.4	79.6	93.3	86.5	73.2	83.8	67.814	68.5
	DLHS5	84	81.2	74.4	76.6	81.38599	94.80984	66.8	58.2
Malda	DLHS1	49	29.7	74.2	38.9	45.89	16.98425	41.67	43.3
	DLHS2	54.7	31.7	62.2	46.9	53.72	34.3	46.257	34.1
	DLHS3	68.1	28.6	59.6	82.2	58.28	47	52.373	43.3
	DLHS4	79.4	72	86.3	80.4	61.7	81.6	44.566	74.2
	DLHS5	72.5	55	52.6	69.5	67.5766	71.01	64.2	43.2
Medinipur	DLHS1	65.2	54.8	78.8	46	73.5	68.31937	64.63	60
	DLHS2	76.5	62.5	73.9	53.1	77.46	71.90289	68.02	46.8
	DLHS3	91.4	42.9	61.7	84.9	80.34	76.68091	72.54	56.3
	DLHS4	89.05	78.5	85.5	77.9	82.5	79.19465	74.918	67.15
	DLHS5	90.25	75.9	81.65	92.4	84.90124	77.59	73.4	51.7
Murshidabad	DLHS1	52.6	39.7	74.6	39.4	49.63	23.11298	48.33	20.9
	DLHS2	61.9	39.2	39.4	27.9	58.06	39.4	52.758	45.2
	DLHS3	71.9	41.6	63	62.5	62.94	51	58.662	38.6
	DLHS4	69.8	58.1	83.6	67.6	66.6	84.1	64.294	60.9
	DLHS5	74.4	63.8	72.1	78.9	72.95619	76.6	66.1	46.5
Nadia	DLHS1	73.3	77.5	87.9	68.9	62.02	49.02787	60.06	58.5
	DLHS2	81.3	76.2	67.6	71.9	68.77	59	63.336	59
	DLHS3	87.6	69.9	76.4	86	72.33	78.3	67.704	57.1
	DLHS4	87.1	88	84.4	91.5	75	83.24	72.068	68.8
	DLHS5	88.2	93.1	91.6	93.2	80.43516	93.62639	73.7	56.9
North 24 Parganas	DLHS1	67.7	65	82	65.6	74.71	25.88548	72.13	52
	DLHS2	76.2	55.1	64.5	62	79.9	36.6	74.593	62.1
	DLHS3	85.9	62.8	86.6	81.7	82.3	47	77.877	71.1
	DLHS4	86.1	79.6	65.5	56.1	84.1	55.7	81.364	70.3
	DLHS5	90.1	86.9	79.3	88.7	88.61454	72.65	82.9	63.5
Purulia	DLHS1	51.5	35.3	82.2	38	51.91	10.87236	37.15	25.6
	DLHS2	61.1	60	69.8	65.5	58.27	39.3	41.161	48.8
	DLHS3	77.2	38.9	65.1	84.3	61.83	52	46.509	45.7
	DLHS4	75.4	78.1	88.2	83.6	64.5	70.4	49.552	69.3
	DLHS5	76.8	72.9	68.6	87.4	69.4165	75.28	48.1	56.3
South 24 Parganas	DLHS1	53.4	48	84.5	59.4	65.18	31.24064	59.73	51.4
	DLHS2	69.4	39	65.4	54.4	71.9	41.1	63.231	53.3
	DLHS3	76.9	36.4	69.1	73.8	75.1	52	67.899	58.8
	DLHS4	80.5	67.7	78.7	70.5	77.5	65.5	72.68	70.9
	DLHS5	78.5	52.2	75.6	94.8	83.25341	71.55	74.6	51.2
UttarDinajpur	DLHS1	46.8	23.6	60.8	28.5	43.91	13.40209	37.16	40.9
	DLHS2	45.2	26.9	42.7	27.6	51.26	28.7	41.663	48.8
	DLHS3	59	27.6	53.3	54.5	55.74	42	47.667	60.9
	DLHS4	62.7	54.5	67	76.8	59.1	73.6	51.742	68.6
	DLHS5	66	47	43.1	66	64.41784	58.66	51.1	60.3

Appendix-B: Value of GE(0), GE(1) and GE(2) of SDI for 5 Time Points:

Variables	**GE**	**DLHS 1**	**DLHS 2**	**DLHS 3**	**DLHS 4**	**NFHS 4**
SDI	GE(0)	0.007332	0.004293	0.002694	0.000703	0.001685
	GE(1)	0.006966	0.004137	0.002629	0.000698	0.001637
	GE(2)	0.016374	0.009838	0.0063	0.001697	0.003906
HOBL3	GE(0)	0.004229	0.004049	0.002401	0.001912	0.001691
	GE(1)	0.004129	0.003806	0.002293	0.001837	0.001632
	GE(2)	0.009927	0.008847	0.00539	0.004341	0.003868
SD	GE(0)	0.025305	0.017949	0.019765	0.003843	0.008482
	GE(1)	0.023521	0.017019	0.019053	0.00366	0.007781
	GE(2)	0.055394	0.040737	0.046338	0.008599	0.017723
PWANC3	GE(0)	0.001561	0.007943	0.003848	0.002045	0.006735
	GE(1)	0.001504	0.007403	0.003842	0.001989	0.006124
	GE(2)	0.003563	0.017128	0.009439	0.004754	0.013881
PCCI	GE(0)	0.014372	0.014997	0.004408	0.003211	0.002348
	GE(1)	0.013637	0.013085	0.004201	0.003067	0.002293
	GE(2)	0.032309	0.028924	0.009873	0.007219	0.005502
CLR	GE(0)	0.005429	0.003471	0.002617	0.002117	0.00162
	GE(1)	0.005274	0.003375	0.002556	0.002073	0.001581
	GE(2)	0.012648	0.008078	0.006135	0.004986	0.003787
NER(UP)	GE(0)	0.044982	0.017923	0.014701	0.01327	0.016493
	GE(1)	0.043223	0.016708	0.012999	0.010673	0.013046
	GE(2)	0.11184	0.040766	0.029872	0.02228	0.027027
FLR	GE(0)	0.008242	0.006445	0.005086	0.005454	0.004056
	GE(1)	0.007846	0.00618	0.0049	0.005146	0.003882
	GE(2)	0.018537	0.014659	0.011647	0.012008	0.009173
GMA18	GE(0)	0.020818	0.010827	0.009741	0.002227	0.005156
	GE(1)	0.01865	0.010974	0.009614	0.002277	0.005162
	GE(2)	0.042796	0.027709	0.023576	0.005726	0.01277
PCDDP	GE(0)	0.0769	0.0192	0.0071	0.0213	0.0298
	GE(1)	0.0398	0.0225	0.0071	0.0245	0.0128
	GE(2)	0.0655	0.081	0.0174	0.0731	0.0646