(2021) 51, 41–63

The Review of Regional Studies

The Official Journal of the Southern Regional Science Association

Spatial Analysis of an Education Program and Literacy in India*

Chitra Jogani

Department of Economics, Trinity College, USA

Abstract: This paper explores the inclusion of spatial dependency in measuring the impact of geographically targeted programs. Using an education program in India, which targeted educationally backward districts, I study the influence of the program on the change in the rural female literacy rate and the gender gap in the literacy rate. In the estimation of a non-spatial model, the residuals exhibit spatial dependency, and the data suggests the spatial error model or the spatial Durbin error model (SDEM) as the appropriate specification. According to the SDEM estimates, with a one percentage point increase in the educational backwardness of a district, there was a 0.08 percentage point increase in the rural female literacy rate and a 0.02 percentage point decrease in the gender gap in literacy rate. The results imply a small but insignificant influence of the program received by the neighboring districts on the change in rural female literacy rate of a district. Limited financial flexibility and the lack of incentive to engage in a competition is a possible explanation for the absence of strategic interaction between districts.

Keywords: Spatial Dependency; India; Education; Literacy

JEL Codes: C21, I24, J18

1. INTRODUCTION

A popular strategy used by governments and international organizations for implementing public programs is to target geographic areas in need. Researchers and policy makers are often interested in measuring the impacts of such programs. Given the spatial contiguity of geographic areas, there is a possibility of spatial correlation between areas that receive the program and in the variables used to measure the outcome of the program. In standard program evaluation, limited attention is paid to possible spatial dependency among geographic neighbors (Baylis and Ham, 2015). Disregarding the spatial dependency and estimating a non-spatial model could lead to biased and inconsistent estimates. Using a geographically

*I am grateful to Rebecca Thornton, Richard Akresh, Tatyana Deryugina, Benjamin M. Marx, and Sandy Dall'erba for their comments and feedback on the paper. Remaining errors are my own.

Chitra Jogani is an Assistant Professor of Economics and International Studies at Trinity College, Hartford, CT. E-mail: chitra.jogani@trincoll.edu.

ISSN 1553-0892, 0048-49X (online)
www.srsa.org/rrs

targeted education program in India, I investigate the presence of spatial dependency while measuring the influence of that education program on female literacy.

This paper focuses on an education program in India called Sarva Sikshya Abhiyan (SSA) or the Education for All program, which was launched in 2001. The program was a nationwide initiative targeted to building schools and providing other necessities, such as textbooks and uniforms, to students. Given the low literacy level of females in India (rural female literacy rate was 46.13 percent in 2001) the program incorporated specific schemes for girls: building residential schools and launching enrollment campaigns to encourage admission. The program targeted the schemes for girls and provided larger amounts of funds under SSA to relatively educationally backward districts, districts with a low rural female literacy rate and a high gender gap in literacy rates.[1] This study examines whether there was an increase in the rural female literacy rate and a decrease in the gender gap in literacy rates after the implementation of SSA in districts that were more educationally backward (districts that had higher concentrations of blocks with low rural female literacy rates and a high gender gap in literacy rates).

I assess whether or not there is spatial correlation in the educational backwardness of districts, which determines the intensity of the program received, and in the outcome variables of interest (an increase in the rural female literacy rate and a decrease in the gender gap in literacy rates). Spatial dependency between districts may exist because of geographic, demographic, administrative, or any reason which can be related to distance. Using the the Moran's I statistic for global spatial correlation, I find a positive spatial correlation for both the independent variable and the outcome. The data also shows spatial clusters or "hotspots," regions with high local spatial correlation.

To further validate the presence of spatial correlation, I begin by estimating a non-spatial model for predicting the influence of educational backwardness of a district on the overall literacy rate. The residuals from the estimation are not normally distributed, are heteroskedastic, and are spatially correlated. I use the spatial error model (SER) and the spatial Durbin error model (SDEM) to account for spatial dependency. Both the SER and SDEM incorporate spatial dependency in the error term, but the SDEM also allows for the independent variables of neighboring districts to influence the outcome of a district. Thus, the SDEM may indicate whether or not there is an additional influence on the literacy rate of a district because of the intensity of the program received by the neighbors.

The estimates from the spatial models suggest that with a one point increase in the educational backwardness of a district, the rural female literacy rate increases by 0.076 percentage point and the gender gap in rural literacy rates decreases by 0.02 percentage points. The magnitude of the estimate from the spatial models are similar compared to estimates from the non-spatial model: 0.076 compared to 0.073 for the outcome increase in rural female literacy rate and 0.02 compared to 0.017 for the outcome decrease in rural gender gap in literacy rates. The coefficient capturing the influence of the treatment received by the neighboring districts on the outcome of a district is 0.011 and is insignificant. The results

[1]Blocks are sub-divisions of districts; on average there are ten blocks in a district. Districts with a higher concentrations of educationally backward blocks are referred to as districts that are more educationally backward. The program's geography focused on educationally backward blocks. The percentage of educationally backward blocks in a district is the measure of the intensity of the program.

do not suggest a significant influence of the treatment received by the neighboring districts on the literacy rate of a district; the source of spatial dependency being the dependency in the errors.

The presence of spatial dependence among public schools has been widely studied for programs and schools in the United States. The reason for the existence of the spatial dependence is the strategic interaction and competition between schools and school districts for resources and students (Ghosh, 2010; Ajilore, 2011; Millimet and Rangaprasad, 2007). Such strategic interaction is unlikely in India due to the lack of financial flexibility of individual districts under the SSA program: the resources are distributed by the central and state governments is in line with the educational backwardness of districts. The system of public schools in India and the administration of resources for education differ on several dimensions from that of the U.S., such as an absence of separate boundaries defining school districts. Thus, this paper provides a different context for investigating the presence of spatial dependence, when without an incentive for competition, presence of such spatial dependency may be limited.

2. LITERATURE REVIEW

Spatial dependency and interactions among neighbors has been an important question in the literature and is observed in various contexts, such as in education, government expenditures, environmental policy, politics, and demography. In the context of schooling, there is evidence of spatial dependence in spending by schools in a district depending on the spending by schools of neighboring districts (Ajilore, 2011; Ghosh, 2010). For example, research has found a positive correlation in adopting open enrollment policy between neighboring school districts (Brasington et al., 2016). In addition, previous research has found that teacher salaries tend to be highly influenced by salaries in similar districts (Greenbaum, 2002) and evidence suggests that geographical contiguity fosters local competition and increases efficiency for public and private schools (Millimet and Collier, 2008; Misra et al., 2012; Gonzalez Canche, 2014; McMillen et al., 2007).

The reason for such strategic interaction and spillovers in the U.S. has been the existence of competition between school districts for students, as parents "shop" for schools, and also because housing prices depend on the school quality (Ajilore, 2011). Millimet and Rangaprasad (2007) show that school districts in Illinois compete with nearby districts for students and other important measurable quality criteria such as school spending and pupil-teacher ratio. Similarly, Rincke (2006) show school districts in Michigan compete for students and for non-resident students if neighboring districts did.

Strategic interaction has also been observed for property-tax competition among local governments (Brueckner and Saavedra, 2001), expenditure of state governments (Case et al., 1993), spending decisions of local governments in Portugal (Costa et al., 2015), recreational and cultural services provided in municipalities in Sweden (Lundberg, 2006), and in incumbent behavior of politicians (Besley and Case, 1992).[2] Similar questions of dependency has been studied for stringency of environmental policies (Fredriksson and Millimet, 2002b), pol-

[2]For an overview of strategic interaction among governments refer to (Brueckner, 2003).

lution abatement expenditure (Fredriksson and Millimet, 2002a), and sex ratio of a district (Echávarri and Ezcurra, 2010).

Another question that has been widely studied in the context of education is the effectiveness of education programs on various economic and educational outcomes. The studies have generally used experimental or quasi-experimental methods, such as difference-in-differences, to causally identify the impact of various education programs. For instance, researchers have examined the effect of an investment in schooling infrastructure on years of education (Duflo, 2001), enrollment (Barrera-Osorio et al., 2011), and educational achievement (Case and Deaton, 1999). Geographic based targeting is also practiced in the case of education programs such as the Head Start program in the U.S. (Ludwig and Miller, 2007). Additionally, to address the inequality of education opportunities for girls in developing countries, the construction of schools (Andrabi et al., 2013; Kazianga et al., 2013) or scholarship programs (Filmer and Schady, 2008; Kremer et al., 2009) are not uncommon. In (Jogani, 2018), I study the effect of the Sarva Sikshya Abhiyan (SSA) program on the female literacy rates using a regression discontinuity method, a quasi-experimental approach.

Although experimental and quasi-experimental approaches have been accepted as the reliable approach for establishing causality, the importance of incorporating spatial components in the above designs has received recent attention. Kolak and Anselin (2020) discuss the possible violation of the Stable Unit Treatment Value Assumption (SUTVA), which is an important assumption for causality, in the presence of spatial dependence. Baylis and Ham (2015) also discuss the possible failure of SUTVA in the presence of spatial correlation when using a randomized control trial, which would lead to biased causal estimates. Additionally, a common approach to account for spillovers when randomization occurs at the group level, but outcomes are measured at the individual level, is to cluster standard errors at the group-level. Baylis and Ham (2015) show that in the presence of spatial spillovers, correcting for standard errors may not be sufficient. Similarly using geographic neighbors to the treated units as a 'control' group may lead to biased estimates in the presence of spatial spillovers due to the treatment (Hanson and Rohlin, 2013). Hopefully, we continue to make progress and have a merger of the causal and the spatial literature, as the standard spatial literature has paid inadequate attention to causality thus far (Mur and Paelinck, 2009).

3. CONTEXT AND DATA

3.1. Program and Measure of Treatment

The Sarva Sikshya Abhiyan (SSA) or the Education for All program was launched in 2001 to increase access to education in India. This was an effort to achieve universal elementary education, which has been a goal of the government since India's independence in 1947. The program aimed to improve the education infrastructure in the country, which included building and repairing classrooms, building girls' toilets, and drinking water facilities. The program also hired new teachers and designed curriculum to include the interests of children from different backgrounds. In addition to providing the necessary infrastructure, the program aimed to increase enrollment and reduce school dropout rates.

The government of India implemented an educational tax to raise funding for the SSA

program. The total allocation of funds in 2001-2010 was Rs.125,323 crores or 27.345 billion U.S. dollars and the audited expenditure was Rs.120,820 crores (2.63 billion U.S. dollars). The funds from the central government were transferred to the states, which were then transferred to districts.[3] The program followed a bottom-up approach in planning and sought more community involvement with planning teams at the district, block, and habitation level to accommodate location specific issues and needs. These teams included teachers, parents, and employees from NGOs and the education department.

Given the low enrollment or high dropout rate of girls, the program made special efforts to increase the enrollment of girls in schools in India.[4] India has a low level of literacy, and especially a low level of female literacy (the average rural female literacy rate was 46.13 percent according to Census 2001[5]). To identify areas that are falling behind substantially in rural female literacy, the Department of Education and Literacy of India classified blocks into two categories: educationally backward block (EBB) and not educationally backward block (NEBB). Blocks are sub-divisions of districts and on average there are ten blocks in a district.

The classification of a block as an EBB was based on the criteria of rural female literacy rate being below the national average of 46.13 percent and the gender gap in total literacy being above the national average of 21.59 percent. Classification as an EBB entitled a block to receive additional funding to build special facilities, such as residential schools for girls known as Kastruba Gandhi Balika Vidyalay (KGBV), and for conducting campaigns to encourage enrollment of girls under the National Program for Education of Girls at Elementary Level (NPEGEL).

This study takes advantage of the process of classification of EBBs to investigate the effect of expansion in schooling infrastructure on literacy using a regression discontinuity method in (Jogani, 2018). The results indicate a significant expansion of school infrastructure in educationally backward blocks. However, being classified as an EBB did not lead to a significant increase in the rural female literacy rate. But, at the aggregate level of districts, the results suggest that districts with a higher percentage of EBBs experience an increase in the rural female literacy rate and a decrease in the gender gap in literacy rate.[6] To explore spatial correlation in the dependent and independent variables, districts are the relevant spatial unit. The percentage of EBBs in a district is defined as the treatment intensity of the program in that district and I use the data on the literacy rate of districts to study this.

There are various reasons to expect spatial dependency between districts with respect to the intensity of the program received and the literacy rates. First, neighboring a district with a high treatment intensity may have negative (i.e. receive less treatment due to crowding out by the neighboring district) or positive spillovers (i.e. receive attention and more treatment by being near a high treatment intensity district). Second, districts in a state are impacted by the same state policies and exogenous events which would lead to spatial correlation in

[3]The funding transfer from the central level followed a rule, but the exact rule used to allocate the funds to the next administrative levels remains to be investigated.

[4]The dropout rate for adolescent girls in India is as high as 63.5 percent (See https://www.cry.org/old-statistics-on-children-in-india.

[5]See: https://censusindia.gov.in/Census_Data_2001/India_at_glance/literates1.aspx

[6]For further details on the analysis, please refer to (Jogani, 2018).

the errors. Finally, a high treatment intensity district may attract students from neighboring districts and the increase in the literacy rate of the district may affect the literacy rate of neighboring districts through peer or network effects.

Table 1: Summary Statistics for Districts in India in 2001

Variable	Mean
% of EBBs	45.954
	(1.683)
Rural Female Literacy Rate	47.417
	(0.635)
Gender Gap in Literacy Rate	24.158
	(0.324)
No. of Districts	581

% of EBBs: Percentage of Educationally Backward Blocks in a district. The table presents the summary statistics using the Census 2001.

3.2. Data

Several datasets are used in this analysis. First, spatial maps of the administrative districts of India from Census 2001 shared by the Datameet community are used. Second, information on the classification of blocks as educationally backward or non-educationally backward are obtained from the Ministry of Human Resource Development. The classification was based on the rural female literacy rate and the gender gap in total literacy rate for the year 2001 based on the population Census. Finally, for examining the growth in the literacy rate after a decade of the program, the population census of India for 2001 and 2011 are used. The number of districts used for the analysis is 576 out of a total number of districts of 592.[7] Table 1 provides the summary statistics of the treatment variable (percentage of EBBs) and the literacy variables for districts in India in 2001. The table shows that the mean percentage of EBBs in a district is 45.9. Figure 1 shows the frequency distribution of EBBs and the median percentage of EBBs in a district to be 42.9.

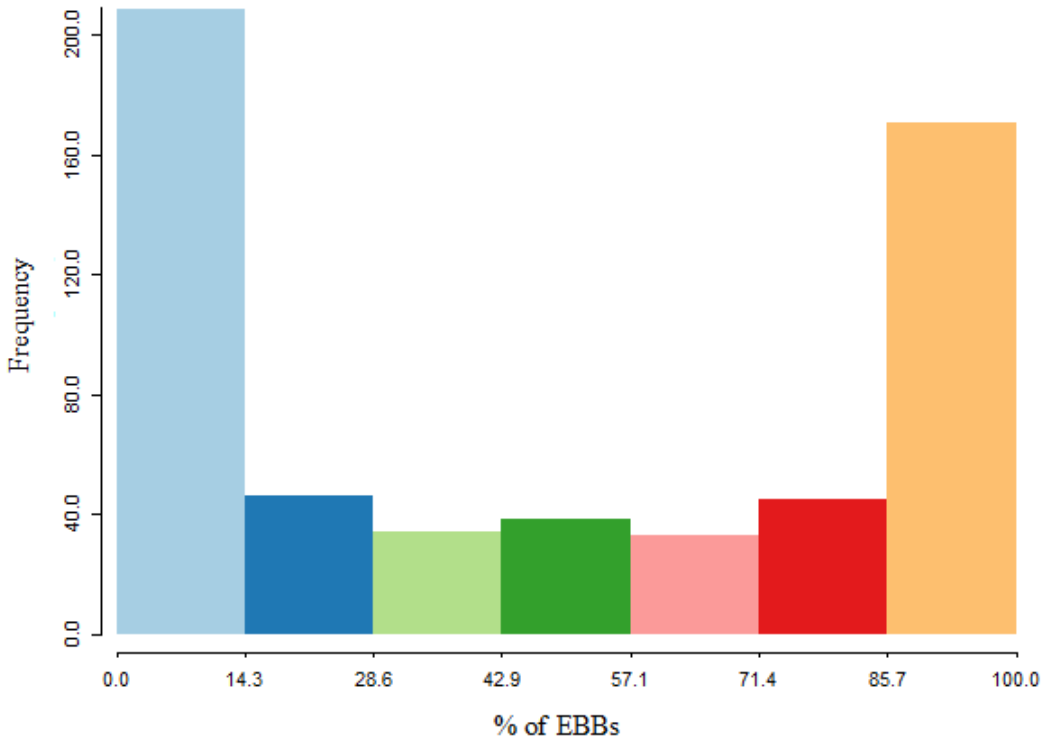
4. EXPLORATORY SPATIAL DATA ANALYSIS

4.1. Quantile Maps

To understand the spatial distribution of educationally backward areas, quantile maps of the districts of India are shown in Figure 2. Districts are shaded according to the percentage of

[7]The total number of districts in 2001 was 592. However, due to lack of a unique identification code for matching the datasets, 576 districts were matched accurately. Total number of blocks: 5,463 (Census 2001).

Figure 1: Frequency Distribution of Educationally Backward Blocks

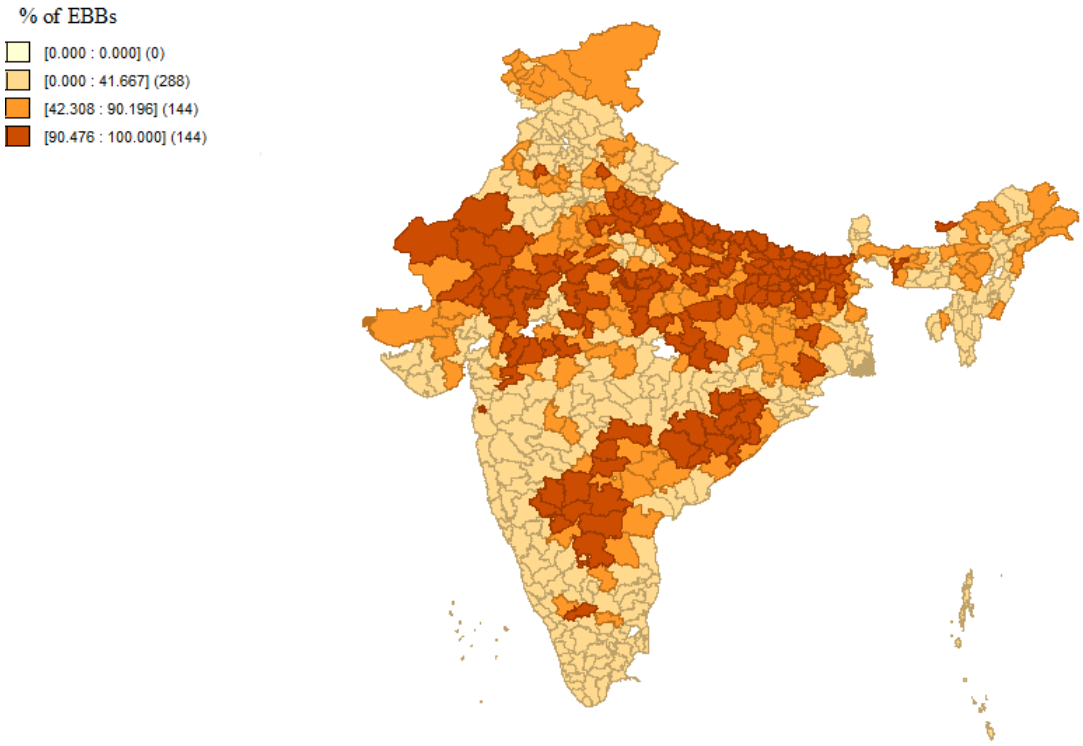


The figure presents the histogram for the percentage of educationally backward blocks (EBBs) in districts of India. There are many districts with no EBBs (0 percent), whereas in some districts all the blocks are EBBs (100 percent). The median percentage of EBBs is 42.9.

EBBs; districts with the darkest shade are ones with 90-100 percent EBBs. The quantile map shows districts with high (low) percentage of EBBs were surrounded by districts with high (low) percentage of EBBs. This indicates spatial correlation among neighboring districts in the percentage of educationally backward blocks and the intensity of treatment.

Figure 3 presents the quantile map of the outcome variable of interest - the increase in the rural female literacy rate (IRFLR) for the period of 2001-2011. The districts with the darkest shade have experienced the highest increase in the rural female literacy rate in the last decade. There exists a similar pattern of spatial correlation between neighboring districts; districts with high (low) IRFLR are surrounded by districts with high (low) IRFLR. Additionally, comparing Figures 2 and 3, there is a positive correlation between districts with a high percentage of EBBs and districts with a high IRFLR, especially in the northeast and west of India - historically regions with low levels of development and literacy. For some districts, such as in the west of India, there is a negative correlation (high percentage of EBBs but a low IRFLR).

Figure 2: Distribution of Educationally Backward Blocks in Districts of India



The figure shows the quantile map for the percentage of educationally backward blocks (EBBs) in districts of India. Districts with a darker shade have a higher concentration of the EBBs and hence are more educationally backward than districts with a lighter shade. There is also a significant concentration of the educationally backward districts in certain states of India, such as Rajasthan and Uttar Pradesh.

4.2. Global Spatial Auto-correlation

One of the common statistics used to understand global and local spatial correlation is the Moran's I statistic. The Global Moran's I can be represented by the following equation:

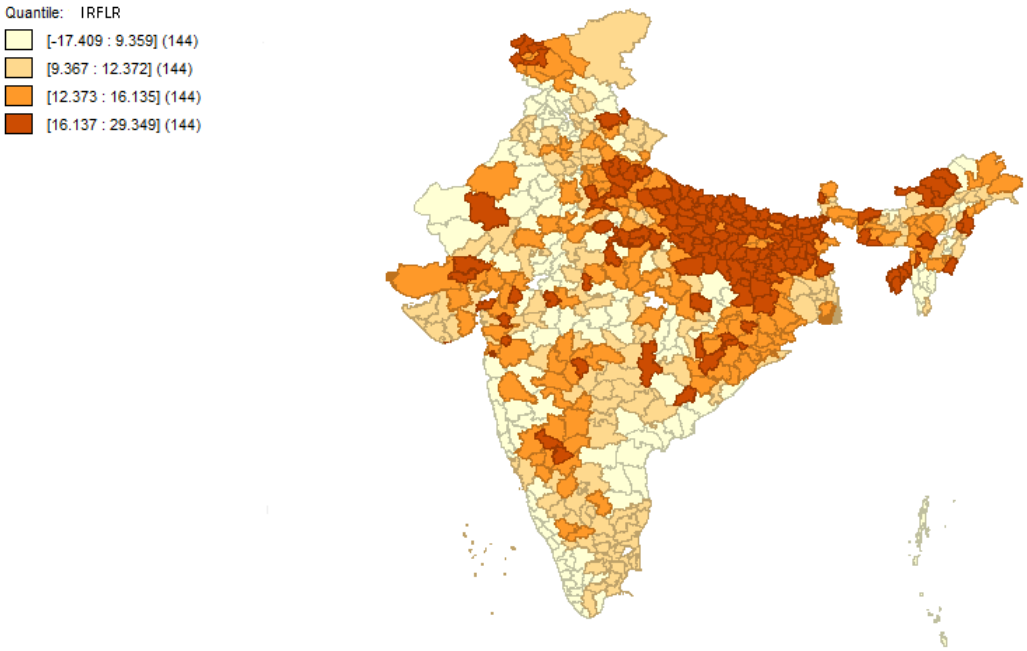
$$I = \frac{N\Sigma_i\Sigma_j W_{ij}(x_i - \overline{x})(x_j - \overline{x})}{\Sigma_i\Sigma_j W_{ij}(x_i - \overline{x})^2} \tag{1}$$

where N is the total number of observations (districts in this case), i and j represent spatial units (or districts), x is the variable of interest for detecting spatial correlation (for example the percentage of educationally backward blocks or the change in literacy rate), $\overline{x}$ is the mean of x, W_{ij} is the optimal weight matrix. The weight matrix is used to provide structure to the nature of correlation between spatial units and to define spatial neighbors.

I find the weight matrix that would best capture the spatial correlation in the data, or the matrix with the highest Moran's I. I calculate the Moran's I for weight matrices defined on the distance based spatial weight, on contiguity, and on the K-nearest neighbors[8]. Results

[8]See the appendix for details on the weight matrices

Figure 3: Quantile Map for the Outcome Variable:
Increase in Rural Female Literacy Rate (IRFLR)



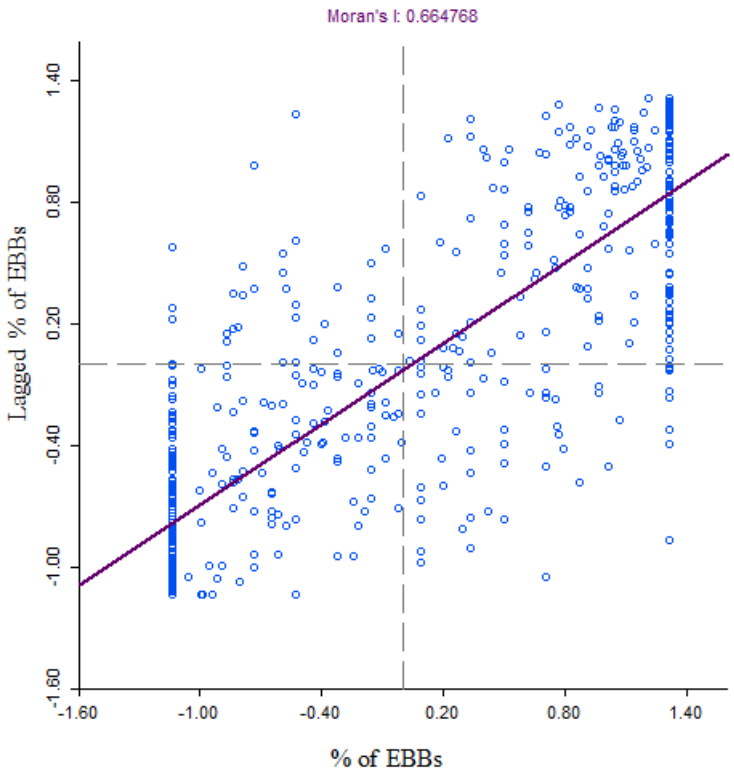
The figure shows the quantile map for the outcome variable - the increase in rural female literacy rates from 2001 to 2011 in the districts of India. Districts with a darker shade experienced a higher increase in the rural female literacy rate in the last decade. There is some overlap of districts with a high percentage of educationally backward blocks and districts that experienced a higher increase in the rural female literacy rate.

show the average Moran's I is the highest for the K5 weight matrix (5 nearest neighbors) for both EBB and IRFLR (the independent variable and outcome). Thus, I use the K5 matrix as the weight matrix of choice to define spatial neighbors for all analyses.

To determine whether or not there is significant spatial correlation, we need to compare the statistic I obtained from equation (1) with the expected value of I under the null hypothesis of no spatial correlation, given by $-1/(N-1)$ (Dall'erba, 2005). The Moran's I ranges from -1 to 1: -1 implies perfect negative spatial correlation, 1 implies perfect positive spatial correlation, and 0 implies no spatial correlation. If I is greater (lower) than $-1/(N-1)$, then the data suggests positive (negative) spatial correlation. In this analysis, $N = 576$, therefore, $-1/(N-1)$=-0.00174.

Figure 4 presents the univariate global Moran's I plot for the intensity of treatment variable - the percentage of educationally backward blocks. The figure captures the spatial correlation in the treatment variable between a district and its neighbors (represented by the spatially lagged variable on the vertical axis). The Moran's I obtained for the relation is 0.66, which is greater than -0.00174 calculated above. Figure 5 presents the univariate global Moran's I scatter plot for the outcome variable IRFLR and the Moran's I statistic obtained is 0.51. The positive Moran's I and the linear relationship suggest a significant positive correlation among spatial units and their neighbors with respect to both the outcome

Figure 4: Global Moran's I plot: Percentage of Educationally Backward Blocks



The figure presents the global Moran's I plot for the treatment variable - the percentage of educationally backward blocks (EBBs). The figure plots the relationship between the treatment variable and its spatial lag (that is the value of the variable for its neighbors). The plot is constructed using standardized values on the axes, such that each unit corresponds to a standard deviation. The positively sloping fitted line through the scatter shows a positive spatial correlation between the districts in the intensity of treatment. The scatter plot is centered on the mean to divide the plot into four quadrants. The top right and bottom left correspond to positive spatial correlation whereas the bottom right and top left correspond to negative spatial correlation.

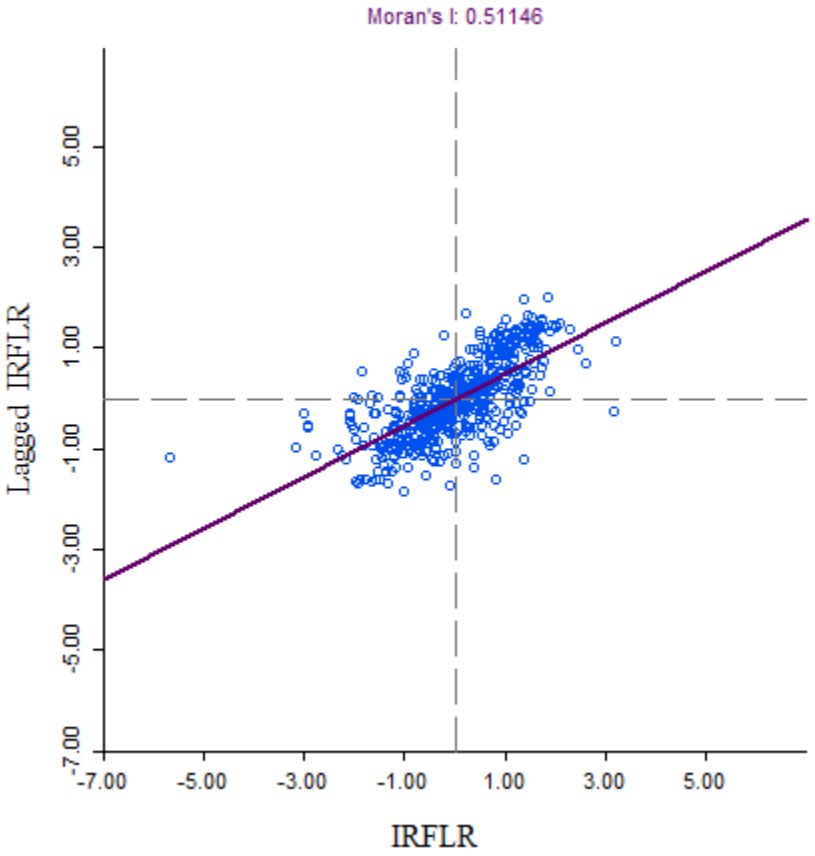
(the increase in rural female literacy rate) and the independent variable (the percentage of EBBs). Additionally, the null hypothesis of such correlation to be random, is rejected at a significance level of 1 percent for both variables.

4.3. Local Spatial Auto-correlation

The global Moran's I statistic informs us about the overall spatial auto-correlation in a data. However, due to spatial heterogeneity the degree of correlation may vary across space. The local Moran's I statistic can help to determine the spatial correlation in different areas across space.

The local Moran's I statistic for each region i can be obtained using the following equation (Anselin, 1995):

Figure 5: Global Moran's I plot: Increase in Rural Female Literacy Rate (IRFLR)



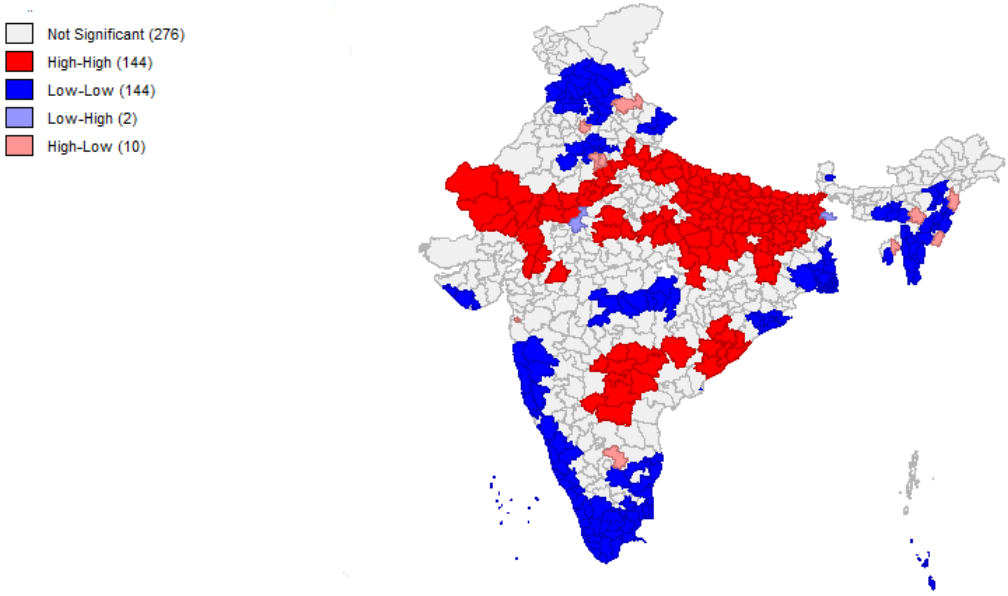
The figure presents the global Moran's I plot for the outcome variable - the increase in the rural female literacy rate from 2001 to 2011. The figure plots the relationship between the treatment variable and its spatial lag (that is the value of the variable for its neighbors). The plot is constructed using standardized values on the axes, such that each unit corresponds to a standard deviation and the scatter plot is centered on the mean. The positively sloping fitted line through the scatter shows a positive spatial correlation between the districts in the outcome variable. The scatter plot is centered on the mean to divide the plot into four quadrants. The top right and bottom left correspond to positive spatial correlation whereas the bottom right and top left correspond to negative spatial correlation.

$$I_i = \frac{N(x_i - \overline{x})\Sigma_j W_{ij}(x_j - \overline{x})}{\Sigma_i (x_i - \overline{x})^2} \tag{2}$$

where I_i is the local Moran's statistic for spatial unit i, and j represents the neighboring units, N is the total number of observations (districts in this case), x is the variable of interest for detecting spatial correlation (for example the percentage of educationally backward blocks or change in literacy rate), $\overline{x}$ is the mean of x, W_{ij} is the optimal weight matrix.

Figures 6 and 7 present the local indicator of spatial correlation (LISA) cluster map for the treatment variable and for the outcome variable. The districts with high (low) EBBs are labeled as high (low) in Figure 6 and districts with high (low) IRFLR are labeled as high (low) in Figure 7. The plots are constructed using standardized values on the axes,

Figure 6: LISA cluster map: Percentage of Educationally Backward Blocks

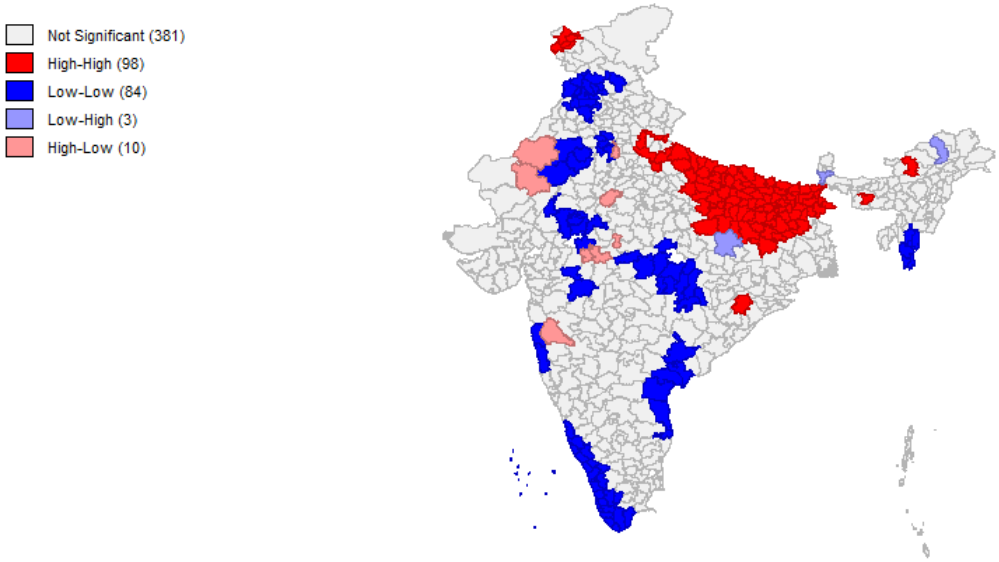


The figure presents the local indicator of spatial correlation (LISA) cluster map for the treatment variable - the percentage of educationally backward blocks (EBBs). High: Districts with high percentage of EBBs, Low: Districts with low percentage of EBBs. The map shows the different spatial clusters or areas where there is higher degree of spatial auto-correlation. For example, a High-High region implies districts with a high percentage of EBBs are surrounded by districts with a high percentage of EBBs (positive spatial correlation). Similarly, the Low-Low region implies a cluster of districts with a low concentration of EBBs (positive spatial correlation). A High-Low region implies negative spatial correlation as High districts are surrounded by Low districts. Not significant implies no significant local spatial correlation. The significance level used to detect local spatial auto-correlation is less than 5 percent.

such that each unit corresponds to a standard deviation. The scatter plot is centered on the mean to divide the plot into four quadrants. The top right and bottom left correspond to positive spatial correlation, whereas the bottom right and top left correspond to negative spatial correlation.

The LISA cluster map identifies the spatial clusters or hotspots: regions with high spatial correlation. The red regions, labeled as the High-High regions, show districts with high values of the treatment or outcome variable are surrounded by similar districts, implying positive local spatial correlation. The blue regions, labeled as the Low-Low regions, show districts with low values of the treatment or outcome variable are surrounded by similar districts, also implying positive local spatial correlation. A High-Low region implies negative spatial correlation as High districts are surrounded by Low districts. The other regions do not show a significant presence of local spatial correlation between the districts at a significance level of less than 5 percent. Figure 6 shows spatial clusters of districts with high percentage of EBBs in northeast India. The northeast region has historically remained among the under developed regions in India.

Figure 7: LISA cluster map: Increase in Rural Female Literacy Rate (IRFLR)



The figure presents the local indicator of spatial correlation (LISA) cluster map for the outcome variable - the increase in rural female literacy rate (IRFLR). The map shows the different spatial clusters, or hotspots, where there is a higher degree of local spatial auto-correlation. The districts with high (low) IRFLR are labeled as high (low). A High-High region implies districts with a high IRFLR are surrounded by districts with a high IRFLR (positive spatial correlation). Similarly, the Low-Low region implies a cluster of districts with a low concentration of IRFLR (positive spatial correlation). A High-Low region implies negative spatial correlation as High districts are surrounded by Low districts. Not significant implies no significant local spatial correlation. The significance level used to detect local spatial auto-correlation is less than 5 percent.

5. SPATIAL ECONOMETRIC METHODOLOGY

5.1. Determining the Spatial Model

I am interested in estimating the influence of the SSA program on the increase in the rural female literacy rate and the decrease in the gender gap in total literacy rates of districts. A non-spatial model can be represented by the following equation:

$$Y_i = \alpha + \beta T_i + \gamma X_i + \varepsilon_i \tag{3}$$

where T_i is the intensity of treatment measured by the percentage of educationally backward blocks (EBBs) in a district, X_i are other demographic characteristics of the districts (used as controls), and Y_i are the outcome variables: the increase in rural female literacy rate (IRFLR) and the decrease in rural gender gap in literacy rates (DGGRLR). The coefficient of interest is β.

However, using a non-spatial model would ignore any form of spatial dependency. We can incorporate spatial dependency by using the following groups of models: the spatial lag

Table 2: Lagrange Multiplier Test for Spatial Model Selection

Statistic	Δ Rural Female Literacy Rate	Δ Gender Gap in Literacy
LMError	290.8 ***	351.4***
LMLag	247.6***	328.86***
RLMError	44.3***	22.9**
RLMLag	1.2	0.35
SARMA	291.9***	351.8***

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Δ Rural Female Literacy Rate: Increase in rural female literacy rate, Δ Gender Gap in Literacy: Decrease in rural gender gap in literacy rates, LM: Lagrange multiplier, RLM: Robust Lagrange multiplier. The table presents the statistics from the Lagrange multiplier test for the spatial error and lag models. The Lagrange multiplier statistics for the error and lag models are significant but the statistic from the Robust Lagrange multiplier test is significant only for the error model. Thus, the suggested model belongs to the category of spatial error model.

and the spatial error model.[9] Using matrix notation, the spatial lag model and the spatial error model can be represented by equations 4 and 5 below:

$$Y = \rho WY + X\beta + \varepsilon \tag{4}$$

$$Y = X\beta + \varepsilon, \varepsilon = \lambda W\varepsilon + \mu \tag{5}$$

where $\mu \backsim N(0, \sigma^2 I)$.

Estimating a non-spatial model when a spatial lag model is appropriate will lead to biased estimates (an illustration is provided in the appendix). Similarly, ignoring the correlation in errors will result in inconsistent estimates. Some other common spatial models are the spatial lag of X (SLX) model and the spatial durbin error model (SDEM) (Vega and Elhorst, 2015). These models are a variant of the spatial lag and spatial error models and are represented by the following equations:

$$Y = \rho WX + X\beta + \varepsilon \tag{6}$$

$$Y = \rho WX + X\beta + \varepsilon, \varepsilon = \lambda W\varepsilon + \mu \tag{7}$$

Compared to the spatial lag model, which included spatial dependency in Y as shown by equation (3.4), the SLX model includes spatial dependency in X. In the SLX model, the outcome Y of a district depends on the covariates of the district and also the covariates of the neighbors. The SDEM model is a combination of the SLX model and the spatial error model - it includes spatial dependency in X and spatial dependency in the error term.

[9]The models in the spatial literature can also be compared with those in the time series literature, such as the spatial lag model can be compared with the AR(1) model which is represented as $Y_t = \rho Y_{t-1} + \varepsilon_t$. For review of spatial models refer to (Anselin, 2002; LeSage, 2014; LeSage and Pace, 2009).

Table 3: Likelihood Ratio Test to Determine the Optimal Spatial Model

Outcome	SDEM	ERROR	OLS
Δ Rural Female Literacy Rate	-1552.3	-1553.8	-1660.9
Δ Gender Gap in Literacy	-1181.9	-1184.4	-1308.9

SDEM: Spatial Durbin Error Model, ERROR: Spatial Error model, OLS: Ordinary least squares, Δ Rural Female Literacy Rate: Increase in rural female literacy rate, Δ Gender Gap in Literacy: Decrease in rural gender gap in literacy rates. The likelihood ratio in the above table is highest for the SDEM, closely followed by the Error model. Thus, I choose SDEM as the optimal model.

5.2. Estimation Results

I begin by estimating the OLS model in equation (3), where the treatment variable is the percentage of blocks that are educationally backward in a district. I include the following variables as controls in all estimations: the percentage of minority population and the percentage of female population in the district in 2001 (before the implementation of the project). Areas with high minority population overlap with economically backward areas. The residuals from the estimation for the outcomes (IRFLR and DGGRLR) were found to be spatially correlated.[10] To account for the spatial correlation, I work with the spatial models described previously and use the K5 weight matrix to define spatial neighbors.

To choose between the spatial lag and error model, I use the Robust Lagrange Multiplier test (Anselin et al., 1996). Table 2 presents the statistics from the Lagrange multiplier test for the spatial error and lag models. The Lagrange multiplier statistics are significant for both error and lag models, but the statistic from the Robust Lagrange multiplier test is significant only for the error model. As suggested by the Robust Lagrange multiplier test, the optimal model belongs to the category of spatial error model. In addition to the spatial error model, I estimate the spatial Durbin error model (SDEM) which incorporates correlation among errors along with the dependency in the treatment variable as shown by equation (7).

Table 3 provides the log likelihood results, which compares the performance of the different spatial models. According to the likelihood ratio test, the difference in the spatial error model and SDEM for both outcomes is small, but they provide a better fit than the OLS model (as indicated by a lower absolute value of the ratio).

Table 4 presents the estimates from the SDEM, the spatial error model, and the OLS model. The SDEM estimates imply a 0.076 percentage point increase in the rural female literacy rate with a one percentage point increase in the intensity of treatment. However, the coefficient estimate for WX (or the Weighted percentage of EBBs), which captures the influence of the treatment received by the neighbors is insignificant. Thus, the results do not suggest a significant association between the intensity of the treatment or the program

[10]The Moran's I statistic obtained was 0.42 and 0.47 for IRFLR and DGGRLR and the statistic was significant at a p-value of 0.001 with 999 permutations

Table 4: Estimation Results for Increase in Rural Female Literacy Rates

	OLS	SDEM	Error
Variables	Estimate	Estimate	Estimate
% of EBBs	0.073***	0.076***	0.074***
	(0.004)	(0.006)	(0.005)
Weighted % of EBBs		-0.011	
		(0.01)	
No. of Districts	576	576	576

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. SDEM: Spatial Durbin Error Model, ERROR: Spatial Error model, OLS: Ordinary least squares, % of EBBs: Percentage of Educationally Backward Blocks in a district, Weighted % of EBBs: Weighted matrix of Percentage of Educationally Backward Blocks in the neighboring districts. The above table compares the estimates from the different models. The SDEM estimates are slightly greater than the OLS estimates (0.076 compared to 0.074). The SDEM estimate implies a 0.076 percentage point increase in the rural female literacy rate for a one point increase in the intensity of treatment or the educational backwardness of a district.

received by the neighboring districts and the literacy rate of a district.

Table 5 provides the estimates from the various models for the second dependent variable - the decrease in the rural gender gap in literacy rates (DGGRLR). The estimates imply a 0.02 percentage point decrease in the gender gap in rural literacy rates with a one point increase in the intensity of treatment. The coefficient estimate for WX (or the Weighted percentage of EBBs) is insignificant.

Finally, to capture any heterogeneity in the association, I divide the sample of districts based on the intensity of treatment into two groups: low and high. Districts for which the treatment variable, the percentage of blocks that are educationally backward, is lower (higher) than the median belong to the group low (high). Table 6 presents the SDEM estimates for the low and high districts separately. The table shows that the influence of educational backwardness of a district is larger for districts in the high group compared to districts in the low group. However, the influence of the educational backwardness of the neighboring districts, measured by the Weighted percentage of EBBs, remains insignificant even for the districts in the high group, implying no influence of the treatment received by the neighbors.

5.3. Discussion and Policy Implications

The above results suggest spatial correlation in the errors, but the literacy rates of a district were not significantly influenced by the intensity of the program received or the literacy rates of the neighbors. A possible explanation of the insignificance is the absence of strategic

Table 5: Estimation Results for Decrease in Rural Gender Gap in Literacy Rates

	OLS	SDEM	Error
Variables	Estimate	Estimate	Estimate
% of EBBs	0.02***	0.0175***	0.017***
	(0.002)	(.003)	(0.003)
Weighted % of EBBs		-0.003	
		(0.006)	
No. of Districts	576	576	576

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. SDEM: Spatial Durbin Error Model, Error: Spatial Error model, OLS: Ordinary least squares, % of EBBs: Percentage of Educationally Backward Blocks in a district, Weighted % of EBBs: Weighted matrix of Percentage of Educationally Backward Blocks in the neighboring districts. The above table compares the estimates from the different models. The SDEM estimates are similar to the OLS estimates for decrease in rural gender gap in literacy rates and it implies a 0.02 percentage point decrease in the rural gender gap in literacy rates for a one point increase in the intensity of treatment or the educational backwardness of a district.

Table 6: Effect by Different Spatial Regimes

	Δ Rural Female Literacy Rate	Δ Gender Gap in Literacy
Variables	Estimate	Estimate
% of EBBs (High)	0.09***	0.03***
	(0.02)	(0.009)
% of EBBs (Low)	0.06***	0.01***
	(0.01)	(0.007)
Weighted % of EBBs (High)	-0.01	-0.003
	(0.01)	(0.007)
Weighted % of EBBs (Low)	-0.01	-0.004*
	(0.02)	(0.009)

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Δ Rural Female Literacy Rate: Increase in rural female literacy rate, Δ Gender Gap in Literacy: Decrease in rural gender gap in literacy rates. The above table presents the SDEM estimates for the two sub samples, low and high. The districts in the low (high) are districts with percentage of educationally backward blocks lower (higher) than the median value. The percentage of EBBs and the Weighted percentage of EBBs have the same meaning as in the previous tables, except that they are estimated for the low and high groups separately. The table shows that the SDEM estimate is larger for the high districts than the districts in the low group (0.09 compared to 0.06). However, the association of the neighboring districts, which is measured by the Weighted percentage of EBBs term remains insignificant even for the high districts.

interaction and financial flexibility of districts in regard to resources under the SSA program. As evidenced by the literature, the underlying reason for spatial dependence was the strategic competition and the ability of jurisdictions to influence spending. When a district does not have guaranteed funding and has to compete for resources and students, the actions of neighboring districts affect a district's own decision (Ghosh, 2010). Such competition and interdependence can only exist in a financially flexible environment, allowing districts to change their spending or level of resources (Millimet and Collier, 2008). Strategic interaction was improbable in the SSA program's context for the following reasons.

First, the central and state governments distributed resources for SSA to districts in proportion to the percentage of EBBs (Jogani, 2018). The classification of a block as an EBB was based on the rural female literacy rate of the block, which is measured by the census division of India and is a difficult variable to manipulate (Jogani, 2018). Therefore, the districts had less influence on the resources received under the program as it was unlikely for them to be able to manipulate the percentage of EBBs in the district. With a constrained revenue source from the government, a district could not compete for resources or students with neighboring districts.

Second, unlike in the U.S., there are no separate boundaries or regions classified as school districts in India. The competition observed in the U.S. for students or resources across school districts and the relation with house prices is lacking. A more common phenomenon with regards to "shopping" for schools is moving to cities for better education opportunities or to private schools. Furthermore, the SSA program was majorly for building and improving infrastructure of public schools in rural areas, and migration to neighboring districts for the purpose of education has been unobserved among families in rural areas.

The program studied in this paper has an important policy implication for understanding spatial dependence in case of geographically targeted programs. Strategic interactions and spatial dependence among geographic areas is less probable amidst resource constraints and limited flexibility in undertaking decisions.

6. CONCLUSION

Several nations and states direct public programs to underdeveloped geographic areas. There may exist spatial correlation between adjacent geographic areas, and not accounting for the dependency while measuring the impact of a program may lead to biased and inconsistent estimates. In this paper, I investigate and account for spatial dependency as I estimate the influence of an education program in India. The program targeted districts with low rural female literacy rates and high gender gap in total literacy rates. I measure the influence of the program on the increase in the rural female literacy rate and the decrease in the gender gap in rural literacy rates.

The data suggests presence of both global and local spatial correlation in the intensity of the program received and the literacy rates. The source of spatial dependency was the spatial correlation in errors, and the intensity of the program received or the literacy rate of neighboring districts did not significantly influence the outcomes of a district. In the presence of spatial correlation in the errors, using a spatially blind model would lead to inconsistent

estimates. Therefore, it is recommended to be wary of spatial dependency and to use a spatial model in addition to incorporating any other controls or sources of endogeneity.

Given that the schooling institutions and the implementation of public programs in India differs from those of the U.S., this paper has an important policy implication. The paper conjectures that the administration structure and incentive environment may determine whether or not there will be strategic interaction and spatial dependence. When the financial environment is constrained and there is little autonomy in acquiring resources, spatial spillovers may be limited.

REFERENCES

Ajilore, Olugbenga. (2011) "The Impact of Ethnic Heterogeneity on Education Spending: A Spatial Econometric Analysis of United States School Districts," *Review of Urban & Regional Development Studies*, 23(1), 66–76.

Andrabi, Tahir, Jishnu Das, and Asim Ijaz Khwaja. (2013) "Students Today, Teachers Tomorrow: Identifying Constraints on the Provision of Education," *Journal of Public Economics*, 100, 1–14.

Anselin, Luc. (1995) "Local Indicators of Spatial Association—LISA," *Geographical Analysis*, 27(2), 93–115.

Anselin, Luc. (2002) "Under the Hood: Issues in the Specification and Interpretation of Spatial Regression Models," *Agricultural Economics*, 27(3), 247–267.

Anselin, Luc. (2003) "Spatial Econometrics," In *A Companion to Theoretical Econometrics*. Blackwell: Malden, MA, USA, pp. 310–330.

Anselin, Luc. (2005) *Exploring Spatial Data with GeoDaTM : A Workbook*. Center for Spatially Integrated Social Science, Urbana, IL.

Anselin, Luc, Anil K. Bera, Raymond Florax, and Mann J. Yoon. (1996) "Simple Diagnostic Tests for Spatial Dependence," *Regional Science and Urban Economics*, 26(1), 77–104.

Barrera-Osorio, Felipe, David S. Blakeslee, Matthew Hoover, Leigh L. Linden, and Dhushyanth Raju. (2011) "Expanding Educational Opportunities in Remote Parts of the World: Evidence from a RCT of a Public Private Partnership in Pakistan," In *Third IZA Workshop*. Institute for the Study of Labor: Mexico City, Mexico.

Baylis, Kathy and Andres Ham. (2015) "How Important is Spatial Correlation in Randomized Controlled Trials?," In *2015 AAEA & WAEA Joint Annual Meeting*. Agricultural and Applied Economics Association: San Francisco, California.

Besley, Timothy and Anne Case. (1992) "Incumbent Behavior: Vote Seeking, Tax Setting and Yardstick Competition," National Bureau of Economic Research: Cambridge, MA.

Brasington, David, Alfonso Flores-Lagunes, and Ledia Guci. (2016) "A Spatial Model of School District Open Enrollment Choice," *Regional Science and Urban Economics*, 56, 1–18.

Brueckner, Jan K.. (2003) "Strategic Interaction Among Governments: An Overview of Empirical Studies," *International Regional Science Review*, 26(2), 175–188.

Brueckner, Jan K. and Luz A. Saavedra. (2001) "Do Local Governments Engage in Strategic Property—Tax Competition?," *National Tax Journal*, 54(2), 203–229.

Case, Anne and Angus Deaton. (1999) "School Inputs and Educational Outcomes in South Africa," *The Quarterly Journal of Economics*, 114(3), 1047–1084.

Case, Anne C., Harvey S. Rosen, and James R. Hines. (1993) "Budget Spillovers and Fiscal Policy Interdependence: Evidence from the States," *Journal of Public Economics*, 52(3), 285–307.

Costa, Hélia, Linda Gonçalves Veiga, and Miguel Portela. (2015) "Interactions in Local Governments' Spending Decisions: Evidence from Portugal," *Regional Studies*, 49(9), 1441–1456.

Dall'erba, Sandy. (2005) "Distribution of Regional Income and Regional Funds in Europe 1989–1999: An Exploratory Spatial Data Analysis," *The Annals of Regional Science*, 39(1), 121–148.

Duflo, Esther. (2001) "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment," *The American Economic Review*, 91(4), 795–813.

Echávarri, Rebeca A. and Roberto Ezcurra. (2010) "Education and Gender Bias in the Sex Ratio at Birth: Evidence from India," *Demography*, 47(1), 249–268.

Filmer, Deon and Norbert Schady. (2008) "Getting Girls into School: Evidence from a Scholarship Program in Cambodia," *Economic Development and Cultural Change*, 56(3), 581–617.

Fredriksson, Per G. and Daniel L. Millimet. (2002a) "Is there a 'California Effect' in US Environmental Policymaking?," *Regional Science and Urban Economics*, 32(6), 737–764.

Fredriksson, Per G. and Daniel L. Millimet. (2002b) "Strategic Interaction and the Determination of Environmental Policy across U.S. States," *Journal of Urban Economics*, 51(1), 101–122.

Ghosh, Soma. (2010) "Strategic Interaction among Public School Districts: Evidence on Spatial Interdependence in School Inputs," *Economics of Education Review*, pp. 440–450.

Gonzalez Canche, Manuel Sacramento. (2014) "Localized Competition in the Non-resident Student Market," *Economics of Education Review*, 43, 21–35.

Greenbaum, Robert T. (2002) "A Spatial Study of Teachers' Salaries in Pennsylvania School Districts," *Journal of Labor Research*, 23(1), 69–86.

Hanson, Andrew and Shawn Rohlin. (2013) "Do Spatially Targeted Redevelopment Programs Spillover?," *Regional Science and Urban Economics*, 43(1), 86–100.

Jogani, Chitra. (2018) Does More Schooling Infrastructure Affect Literacy? SSRN Working Paper 3426016.

Kazianga, Harounan, Dan Levy, Leigh L. Linden, and Matt Sloan. (2013) "The Effects of "Girl-Friendly" Schools: Evidence from the BRIGHT School Construction Program in Burkina Faso," *American Economic Journal: Applied Economics*, 5(3), 41–62.

Kolak, Marynia and Luc Anselin. (2020) "A Spatial Perspective on the Econometrics of Program Evaluation," *International Regional Science Review*, 43(1), 128–153. Publisher: SAGE Publications Inc.

Kremer, Michael, Edward Miguel, and Rebecca Thornton. (2009) "Incentives to Learn," *Review of Economics and Statistics*, 91(3), 437–456.

LeSage, James and R. Kelley Pace. (2009) *Introduction to Spatial Econometrics*. CRC Press, Taylor and Francis Group.

LeSage, James P. (2014) "Spatial Econometric Panel Data Model Specification: A Bayesian

Approach," *Spatial Statistics*, 9, 122–145.

Ludwig, Jens and Douglas L. Miller. (2007) "Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design," *The Quarterly Journal of Economics*, 122(1), 159–208.

Lundberg, Johan. (2006) "Spatial Interaction Model of Spillovers from Locally Provided Public Services," *Regional Studies*, 40(6), 631–644.

McMillen, Daniel P., Larry D. Singell, and Glen R. Waddell. (2007) "Spatial Competition and the Price of College," *Economic Inquiry*, 45(4), 817–833.

Millimet, Daniel L. and Trevor Collier. (2008) "Efficiency in Public Schools: Does Competition Matter?," *Journal of Econometrics*, 145(1), 134–157.

Millimet, Daniel L. and Vasudha Rangaprasad. (2007) "Strategic Competition amongst Public Schools," *Regional Science and Urban Economics*, 37(2), 199–219.

Misra, Kaustav, Paul W. Grimes, and Kevin E. Rogers. (2012) "Does Competition Improve Public School Efficiency? A Spatial Analysis," *Economics of Education Review*, 31(6), 1177–1190.

Mur, Jesús and Jean HP Paelinck. (2009) "Some issues of the concept of causality in spatial econometrics models," In *International Conference of the Spatial Econometrics Association*. SEA: Barcelona, Spain, pp. 9–10.

Rincke, Johannes. (2006) "Competition in the Public School Sector: Evidence on Strategic Interaction among US School Districts," *Journal of Urban Economics*, 59(3), 352–369.

Vega, Solmaria Halleck and J. Paul Elhorst. (2015) "The SLX Model," *Journal of Regional Science*, 55(3), 339–363.

A. APPENDIX

A.1. Weight Matrix

Spatial correlation implies units i and j are correlated, i.e $Cov(X_i, X_j \neq 0)$. However, with N number of observations, the number of correlations to estimate would be $N(N-1)/2$, which can be a very large number. The number of correlations to be estimated can also be reduced once we know the neighbors that interact with each other. To do this, I define a spatial weight matrix which imposes a structure on the nature of correlation between the spatial units. A spatial weight matrix usually relies on the distance between neighbors and is thus exogenous.

Distance Based Spatial weights: One of the common method is to use the great circle distance (or arc distance, which is calculated using the latitudes and longitudes of the spatial units), where $W_{ij} = 1$ if the distance between i and j is below a user defined threshold (Dall'erba, 2005; Anselin, 2005). For example, spatial units i and j are defined as neighbors if the distance between them is below the threshold of 250 miles. However, the distance has to be above the minimum distance required for every spatial unit to have at least one neighbor. I find the minimum distance required for every spatial unit to have at least one neighbor for the data set as 270 miles. I find distance based spatial weight matrices for the minimum distance (270 miles). I also find distance based spatial weight matrices for 300 miles and 325 miles to check for robustness, the results remain similar on using the different distances.

Higher order contiguity: The units i and j are said to be contiguous of the order K if the maximum number of borders to cross to reach j from i is K (Anselin, 2005). The contiguous relations can be of various kinds which lead to different weight matrices, such as matrices Queen, Rook, and Bishop. For example, in the below table, we can define the following weight matrices based on different method of selection of neighbors:

X	Y	X
Y	Z	Y
X	Y	X

Rook: The spatial units labeled as Y are considered neighbors.
Queen: The spatial units labeled as X and Y are considered neighbors.
Bishop: The spatial units labeled as X are considered neighbors.

K-Nearest Neighbors: The value of K is chosen a-priori which is used to define the K nearest neighbors of j, the definition is based on the distance between the centroids of the spatial units. $W_{ij} = 1$ if the centroid of area i is one of the K nearest neighbor from j, $W_{ij} = 0$ otherwise (Anselin, 2005).

I use the above weight matrices to find the weight matrix which captures the nature of spatial correlation best for the given data set, or has the highest value for the local Moran's

I statistic.[11]

Simple illustration of why estimates may be biased: Consider the spatial lag model as shown by equation (4). (4) can be rewritten as below:

$$Y = (I - \rho W)^{-1} X\beta + (I - \rho W)^{-1}\varepsilon \tag{8}$$

$$\implies \frac{\partial Y}{\partial X} = (I - \rho W)^{-1}\beta \neq \beta$$

Hence the estimates are biased.

[11] The understanding and description of the methodology in this section is derived from (Anselin, 2003, 2005; Dall'erba, 2005).