Schooling Infrastructure and Literacy-Evidence from a National Education Mission in India

Chitra Jogani*

Preliminary Draft-Please do not cite

Nov 11, 2015

Abstract

The Government of India has made constant effort and set targets to achieve a higher literacy rate and bridge the gender gap in literacy. I investigate if investing in schooling infrastructure can help a nation boost its literacy rate using evidence from an education mission called Sarva Sikhsya Abhiyan in India. I find evidence for substantial amount of money being invested into the program and growth of schooling infrastructure over the past decade. Using population census data and implementing regression discontinuity design, I find that there was no significant change in female literacy rates or the gender gap in literacy rates in the rural areas. The zero result obtained is precise and any effect greater than 1 percentage points in the increase in the rural female literacy rates can be ruled out.

Keywords: Schooling Infrastructure; Sarva Sikshya Abhiyan; Regression discontinuity; India; literacy rate

*Graduate student, Department of Economics, University of Illinois–Urbana Champaign. Address: 214 David Kinley Hall, 1407 W. Gregory, Urbana, Illinois 61801. E-mail: jogani2@illinois.edu. I am grateful to Mark Borgschulte, Marieke Kleemans, Benjamin Marx, Adam Osman, Elizabeth Powers, Rebecca Thornton, colleagues and seniors for their helpful comments and guidance. I thank the seminar participants in the applied micro research lunch talk for their useful feedback. I appreciate the efforts of the DISE team including Rohit Hans, Aparna Mookherjee, Arun Mehta and others for helping with the data and answering my queries. Clarifications and informations provided by PratibhaKumari and Anil Kumar from the Census India division are highly appreciated. This is a preliminary draft and any errors however are my own.

I. Introduction

Whether investment in schooling infrastructure leads to rise in years of education, enrollment and educational achievement has been addressed in the literature by several development economists. Deaton (1999) obtained increase in enrollment and test scores due to an increase in educational input or pupil teacher ratio. Duflo (2001) found that there was an average increase of 0.12 to 0.19 years of education for each primary school constructed per 1,000 children. It is also of interest to know how far schooling infrastructure or enrollment can help a nation improve the proportion of its literate population.

It has been established that there are several factors that contribute to the education and learning of children. Going to school may not be enough for a student to achieve any level of learning like being able to read or do math. The ASER report for 2011 for rural India states that there were 38.4% of students in grade one who could not recognize even a letter and 36.5% of students who could not recognize numbers. Banerjee et al (2008) in their experiment in Jaunpur district found similar evidence. Muralidharan (2013) also finds that improving school inputs in the usual manner will have little effect on learning outcomes and suggests changes in pedagogy and governance .

In this paper I explore the progress made in the literacy rates of female and the gender gap in literacy rates in the last decade in India where there was an initiative for building school and infrastructure facilities launched as a nationwide education mission called Sarva Sikshya Abhiyan (SSA) or the Education for all movement. This is an ongoing program that involves not only building schools and improvement of schooling infrastructure but also increasing the number of teachers and training them. The program took into account the necessity for improvement in the basic amenities like textbooks, uniforms, drinking water and separate toilet for girls. In addition, the program also facilitated enrollment of differently abled students, students from the minority or disadvantaged groups.

With the wide disparity between male and female literacy rates in mind, the program has special focus on girls and children from disadvantaged groups or minorities. The Ministry of human resource development has been building Model Schools, providing additional funds under the scheme National Program for Education of Girls at Elementary Level (NPEGEL), establishing Kastruba Gandhi Balika Vidyalay (KGBV) or boarding schools for girls in blocks that are considered as educationally backward (EBB).[1] My analysis examines if the special focus of the program on girls in these blocks led to an improvement in the female literacy rates of the region and helped bridge the gender gap in literacy rates. [2]

[1] Blocks also known as Tehsils, Mandals or sub districts are subdivisions of a district

[2] The literacy rate in India is calculated by taking the percentage of literates above age 6 in the population

Educationally backward blocks received some girl specific programs and since they also overlap with areas that are generally backward and lack schools, they were supposed to receive the entire SSA program with higher intensity.[3] The variables used for classifying the blocks were the corresponding female rural literacy rates and gender gap in total literacy rates based on the Census 2001 data. The classification was based on the twin criteria of rural female literacy rates being below the national average of 46.13% and the gender gap in total literacy being above the national average of 21.59%. Blocks that satisfied the criteria were classified as educationally backward and ones that did not satisfy were classified as non educationally backward (NEBB).[4] I use regression discontinuity as the empirical strategy to see the impact on the literacy variables after a decade. The datasets used in the paper are the decennial Census data for the years 2001 and 2011 to obtain the literacy and demographic indicators, the classification of blocks as EBB/NEBB from the Ministry of Human Resource and Development (MHRD) and the school census from District Information System for Education (DISE). I also use the DISE data to explore the situation of the schooling infrastructure and its growth in the last decade. [5]

Based on regression discontinuity, I find that there was no significant effect in the rural female literacy rates or total female literacy rates of blocks after a decade on being classified as educationally backward. However, the blocks that were educationally backward did show higher percentage increase in the literacy indicators. The estimate I obtain for the outcome variable, increase in rural female literacy rate is statistically zero with a confidence band of 1 percentage points.[6] I find a similar zero result for the decrease in gender gap in rural literacy with a confidence band of 0.7 percentage points. From the infrastructure facilities data, I observe that aspects of the infrastructure situation like number of classrooms, computers, probability of having electricity and other facilities have improved during the program period.

Sarva Sikshya Abhiyan is a huge program that has received lot of attention. Several reports on it have been released, but not much empirical analysis. The closest papers to my study is Litschig, Meller (forthcoming) which finds positive effect of two girl specific program launched in EBB, National Program for Education of Girls at Elementary Level (NPEGEL) and Kasturba Gandhi BalikaVidyalaya (KGBV) on gross and net enrollment. They also find a reduction in the gender differences in enrollment. The other related paper is Glinskaya, Jalan (2003) which is based on the District Primary Education Program in India. The authors in that paper do not find any decrease in gender gap in enrollment or educational achievements. Both these papers have studied the effect on enrollment rates due to such programs. This paper contributes to the literature to see if such programs have an effect on the literacy rates of the nation.

[3] From figure 4 we can also see that the states with larger proportion of EBBs were allocated more funds.

[4] Gender gap in literacy rates is defined as the difference between the male and female literacy rates

[5] While conducting the census, the head of the household or respondent is asked for the number of literate people in the household. In case of doubt, the interviewer is asked to make the individuals read

[6] I refer to the rural FLR, female literacy rates, total literacy rate sometime broadly as literacy indicators

The rest of this paper is organized as follows: Section II describes the program. Section III outlines the empirical analysis. Section IV provides the results obtained. Section V presents a discussion and Section VI concludes the paper.

II. The Education for All Movement

The literacy rate of India has grown to 74.04% (2011 Census) but it still has the largest number of illiterate population in the world. Along with the dismal scenario of the entire population, the picture for children is equally depressing. As of year 2006, 13.4 million children have been out of school [7] and the dropout rate for adolescent girls is as high as 63.5% .[8]Lack of schools makes education an infeasible option keeping children out of school which also contributes to child labor another problem that developing countries are trying to fight. The number of children aged between 5 to 14 involved in child labor in India is reported as 10.12 million. [9]

Universal elementary education has been the aim of India since its independence and it has taken steps towards increasing the literate population and improving the educational status. The first formal statement on universal primary education was the 1968 National Policy on Education and several attempts have been made in that direction. Sarva Shiksha Abhiyan is the result of the same effort in the last decade.

A. Aims and Objectives

Sarva Shiksha Abhiyan is a nationwide program with wide ranging objectives. One of the main aim of the program has been school building and investment in infrastructure, i.e improve infrastructure of existing schools and to build new schools. Improvement in infrastructure and schooling facilities included building and repairing classrooms, toilets, drinking water facilities, boundary wall, playground, furniture, etc. Civil work has been the major component of total allocations and expenditure. The exact proportion varies over years and states but remained closer to 50%. For example, in the year 2006-07, 71% of the total allocation of SSA fund was towards civil works and the final expenditure was 56%. [10]

In addition to infrastructure, the program aims to provide free and compulsory education, increase enrollment and reduce drop out rates of children. It also realizes the importance of providing quality learning and adapting the curriculum to meet the varying needs of children in

[7]Source: Indiastat table based on Lok Sabha Unstarred Question No. 258 dated 21.02.2006
[8]Source: Information from cry.org which is based on MosPI,2012
[9]Source: Information from cry.org which is based on MosPI,2012
[10]Source: State wise and Component wise allocation and expenditure reports, SSA

different locations. As also mentioned earlier, the program has special focus on girls, children from minorities and children with special needs. The dropout rate among girls tend to be higher because they are often required by their families to help with domestic chores, taking care of siblings or are married off at an early age. Other factors which lead to lower enrollment of girls is because of the schools being located far away, lack of female teachers, girls' toilets and safety concerns.

Given the low enrollment or high dropout rate of girls, SSA in its planning has made it explicit that efforts will be made to combat such problems and issues which seem to act as an obstacle to girls education. The program allows extra funds at district level for awareness campaigns on girls education, and enrollment drive campaigns. It also provides free textbooks and trains teachers to be sensitive to girls needs. Girls have been treated as a special focus group in the planning, implementation and evaluation process.

However, the general interventions towards girls under SSA may not be enough and thus specific schemes like the National Program for Education of Girls at Elementary Level (NPEGEL), which is an integral component of SSA but having a separate status and additional funding was launched.[11] The scheme of building residential schools for girls known as Kasturba Gandhi Balika Vidyalaya (KGBV), was also launched as an additional resource.[12] The SSA must also provide the planning for implementation of these schemes.[13] Hence, this program can be seen as a universal drive encompassing schooling infrastructure but catering to gender specific needs at the same time

The department of education and literacy classified blocks as EBB/NEBB to bring into notice the areas that are lagging behind substantially and need greater attention. To have the nation improve as a whole in the frontier of education, it is important to work on these areas intensively and get them at par with rest of the areas or at least the national average. The classification in the last decade was based on the twin criteria of rural female literacy rates being below the national average of 46.13% and the gender gap in total literacy being above the national avergae of 21.59%. Also, some blocks or urban slums were added based on if they were SC/ST concentrated blocks.[14]

B. Planning and Implementation of the Program

SSA is an ongoing program with funds being released every year. The initial release of funds for the scheme was Rs.584 crore (123.75 million dollars) in the year 2001-02 which increased more than ten times to 6,846 crore (1.510.6 billion dollars)by 2004-05 and to 30,793 crore (6.73

[11] Source: Manual for planning and appraisal of SSA, Ministry of human resource and development

[12] KGBV was launched as a separate scheme in 2004 but has been merged into SSA from Apr,2007, SSA framework document

[13] Source: Manual for planning and appraisal of SSA, Ministry of human resource and development

[14] Block of districts, which have at least 5% SC/ST population and SC/ST female literacy rates below 10% ,Source MHRD planning and appraisal

billion dollars) by 2010-11.[15]The total release of funds in the decade from year 2001-2010 has been Rs.1,25,323 crore or 27.345 billion dollars and the audited expenditure has been Rs.120,820 crores (2.63 billion dollars). The expenditures have differed from the allocation of funds, being higher or lower in certain years. The central and state governments have shared the release of funds. The Centre:State ratio was proposed to be 85:15 in 2001-02, 75:25 for the years 2002-07 and gradually change to 50:50.[16] However, the ratio differed for some states.

The financing of SSA was borne by the tax base of the country with an additional levy of educational cess. Initially the decision was to fund the program entirely by the government of India, but later from 2004 onwards due to insufficiency of funds, additional funding was provided by international aid agencies like the World Bank, European Commission.[17]The fund under SSA was used in providing teachers salary, textbooks, civil works, management grants and other expenses.

Implementation of a massive program like SSA requires detailed planning. The planning process seeks more community involvement and thus has planning teams at the district, block and habitation level to accommodate location specific issues and needs. These teams include people from NGOs, teachers, education department, parents among others.[18] The program has a bottom-up approach in implementation, i.e getting feedback from the local level and incorporating it in the planning process. It has established Village Education Committee, School Management Committee, etc representing different groups in the society.[19]

The fund from the centre flows to the state implementation society which is then transferred to the districts. District project offices at the district level have been set up as administrative units. Similarly, the block and cluster resource centres have been set up at the subdistrict levels. [20][21] There have been several institutions involved in monitoring of the program; in the year 2013-2015 this included 38 monitoring institutions which were major reputed universities in different locations.[22] Joint review missions were also held twice a year from 2005 onwards to discuss the progress and sort out any issues that the states could be facing.[23]

[15]Source tables for release and expenditures from Sarva Sikshya Abhiyan

[16]Source: Planning commission study and documents on SSA framework

[17]Source: UNESCO, case study on Sarva Sikshya Abhiyan, 2015

[18]Source: Manual for planning and appraisal of SSA, Ministry of human resource and development

[19] However, the implementation process had some loopholes as cited by some studies. Banerjee et al 2008

[20]Source: UNESCO, case study on Sarva Sikshya Abhiyan, 2015

[21]The funding although at the central level followed some rules but it was decentralised in nature and exact rule used to allocate the funds at central level or the next administrative levels remain to be investigated which is my topic for further research

[22]Source: http://ssa.nic.in/monitoring/monitoring-institutes

[23]Source: UNESCO, case study on Sarva Sikshya Abhiyan, 2015

III. Empirical Methods and Analysis

In this section I provide information about the data used, the identification strategy along with its validation and method of analysis for my outcome variables of interest which are literacy rate indicators and variables related to schooling infrastructure.

A. Data

The datasets used in this study are the population census of India for years 2001 and 2011 for the demographic variables, the classification of blocks as EBB/NEBB from the MHRD and the DISE data for the school related outcomes. Since the program rule of classifying blocks as educationally backward was specified at the subdistrict or block level, I consider the Census data at the block level. This program used the variables of female rural literacy rate and gender gap in total literacy rates for the cutoffs which were based on the Census 2001. So, I consider the blocks in the Census 2001 as my base sample which were 5,463.

I use the primary census abstract tables of the census data for years 2001 and 2011 which has information on population of a region and number of literates at different administrative levels. I use the data at the block/subdistrict level. The tables report values for the demographic variables separately for the rural, urban or total areas in a block. Based on this, I create the rural and total variables separately. For example, the rural female literacy rate of 2001 is calculated by taking the number of literate population in the rural area of the block above age 6 as a percentage of the total rural population in the block in 2001. The total literacy rate is calculated similarly but by taking the total literate population above age 6 as a percentage of the total population in the block.

India currently has 29 states, 7 union territories, 640 districts (Census 2011) but the boundaries or area of geographical regions have not remained the same over the decades let alone their names. However, the boundaries of blocks which are more disaggregated than districts have not been affected much in the decade considered. In India there are blocks or districts of the same name in different states which require careful matching of variables across two time periods. There is no common ID linking the census data across years until now, hence I matched using the combination of state, district and block. In some cases there have been change in names of the administrative divisions or the records having different spellings in different datasets; the major reason for which is that most of the names are in the regional languages and people may spell it differently in English.

To tackle the naming mismatch problem, I followed the strategy of renaming all the divisions as per their name in 2001 for the ones that were not matched and created a unique dataset of demographic variables for the years 2001 and 2011 for blocks. This however reduced my sample to 5,299 blocks for the data on total areas and 5,225 blocks on the data on rural areas.

Since, I want to look at the implementation of the program and its effectiveness across the educationally and non-educationally backward block; I have to match the census data with the list of classification of blocks as EBB/NEBB. I use the data for the classification from the program website. The major difficulty in matching and drop in sample size happened at this stage as this list of classification of blocks prepared by the department of education and literacy had no connection with the names of the areas as mentioned in the census. Thus, I once again followed the strategy of matching these names with the names of my census dataset by renaming the areas consistently with the Census 2001 names for whichever areas possible. This led to my final sample size of 4,218 unique areas which is a combination of state, district and block. Tables 1 and 2 present the summary statistics for the demographic variables from the population census for the years 2001 and 2011 based on the total sample considered.

The next dataset that I use is the dataset from DISE which is the unique source of information on schools in India capturing their enrollment, facilities and location. On linking the combined census and MHRD dataset of 4,218 blocks with the DISE data required me to do matching based on names of the areas as the DISE data has no corresponding code to link with the census data. This reduced my sample size further to 3,991 which is the final dataset I use throughout the paper. From the DISE data I consider only schools which were in rural areas and exclude private unaided schools.

Table 3 compares the average amount of funds a school received and spent, the average number of computers and classrooms in a school for the academic years 2005 and 2011. Since the program involved building schools and investing in infrastructure, it is important to observe if there was an improvement in facilities over the academic years. We see that the kind of facilities in a school have increased over time. For example, the percentage of schools that are electrified has increased by 70 percentage points although being only 38.

B. Identification Strategy: Fuzzy Regression Discontinuity

I am interested in finding the effect of being classified as educationally backward in the SSA program on outcome variables of interest, i. e

$$Outcome_i = \alpha + \beta EBB_i + \varepsilon_i \tag{1}$$

where EBB is a binary variable which equals 1 if the block is EBB and 0 otherwise. The outcome variables of interest in this paper can be broadly classified as variables related to literacy rates of a block which are based on the population census and the variables related to schooling infrastructure of a block which are based on the school census from DISE. The EBBs as mentioned earlier are areas which are educationally backward and were classified based on low literacy rates. This makes the classification and hence the variable EBB as endogenous. Using a simple OLS model would give us biased estimates.

Since the classification of blocks used a criteria, I exploit the fact to use Regression discontinuity (RD) method as the empirical strategy. Regression discontinuity is based on the idea that blocks on both sides of the cutoff are otherwise similar except exposure of the program. I provide evidences for the validity of the assumptions in the next section. This helps me to find an unbiased estimate by comparing observations only around the cutoff, i.e the ones to the right and left of the cutoff. The procedure would provide the local average treatment effect (LATE). To the best of my knowledge the criteria used for classification of blocks as educationally backward is not used for any other classification. Also, this criteria is not used for implementing any other programs or schemes. So, this would isolate the impact only of the SSA or specifically schemes that were implemented heterogeneously across educationally and non educationally backward blocks.

The criteria used two variables; rural female literacy rates and gender gap in total literacy rates based on Census 2001. This would imply two forcing variables for RD. RD estimates obtained using local linear regressions include calculating the optimal bandwidth which is based on RD designs involving one forcing variable and this has been the standard in the literature. Thus, I restrict my sample to only those who satisfy the criteria of having the gender gap above average maintaining the rural female literacy rates (RFRL) for year 2001 as the assignment variable. This procedure gives a sample of 2,626 for the regression discontinuity analysis and is used as a sample for all my RD analysis. [24]

Figure 1 plots the proportion of blocks that were classified as EBB against the RFLR for the whole sample and the sample restricted to only those satisfying the gender gap criteria. From the first panel we see that there were some points at the cutoff but in the second panel with the gender gap restriction we see clear discontinuity in the proportion or probability of blocks being classified at the cutoff of RFLR. However, since the proportion is not 1 or 0 for being on either side of the cutoff and there are some blocks classified as EBB even when the RFRL is above the cutoff or vice versa; this gives a fuzzy regression discontinuity and not a sharp one. One of the reason for this is inclusion of some scheduled caste or scheduled tribes (SC/ST)concentrated blocks or urban slums as an exception to the rule.[25]

[24] I could also do the other way of restricting the sample to ones which satisfy the RFRL threshold and use gender gap in literacy as my forcing variable, but this gives me a smaller sample and hence I do not perform the analysis using this.

[25] Block of districts, which have at least 5% SC/ST population and SC/ST female literacy rates below 10% ,Source MHRD planning and appraisal

C. Estimation of Literacy rate indicators

Fuzzy regression discontinuity has a similar structure to IV 2SLS regression. In the 2SLS setup we would be estimating an equation of the below form:

$$Y_i = \alpha + \beta EBB_i + \delta f(X_i) + \lambda treat f(X_i) + \varepsilon_i \tag{2}$$

EBB=1 if a block is an EBB, x is the rural female literacy rate for 2001, treat=1 if the RFLR for a block in 2001 was below the cutoff. In the fuzzy regression discontinuity framework, the endogenous status variable EBB indicator is instrumented with the dummy variable treat that takes the value one if a block is on the left and 0 if a block is on the right of the RFLR cutoff for year 2001 (given the sample which already satisfies the gender gap criteria), i.e the discontinuity we obtain in the first stage is used as an instrument for obtaining the causal effect of EBB on any outcome variable of interest.

Estimating the above 2SLS equation to provide RD estimates relies on the choice of functional form and also provides global estimates. One of the method suggested in the literature is to find estimates based on local linear regressions that provides a nonparametric method of estimating the causal effect at the cutoff.[26] Restricting the sample to observations close to the cutoff provides unbiased estimates but decreases precision. Therefore, there is tradeoff between bias and precision of the estimates and an optimal bandwidth for the observation needs to be decided.

Hence, for finding the RD estimates I use the procedure by Calonico, Cattaneo, and Titiunik (2014, CCT). The RD estimates are based on local polynomial regressions and I report the bias corrected robust estimates which measure the average treatment effect at the cutoff. The above estimate is calculated using the optimal CCT bandwidth procedure. In addition, I present the estimates based on Nichols, Austin (2014) which uses the bandwidth procedure based on Imbens and Kalyanaraman (2009). The estimates based on Nichols, Austin (2014) are reported for three levels of bandwidth; optimal bandwidth, half of the optimal bandwidth and double of the optimal bandwidth. I also report estimates after including for controls and state fixed effects as robustness checks. Besides, I report estimates based on the linear, quadratic and cubic functional form of the 2SLS regression function but restricting the sample within the CCT optimal bandwidth in the appendix.

I present figures for regression discontinuity for the literacy rates where I restrict my sample within the optimal bandwidth based on the CCT methodology. I also incorporate figures with additional covariates as robustness checks. The numbers of blocks that are educationally backward vary across states and hence I included states as one of the controls. Additionally, differences in the proportion of SC/ST in a block may cause an important difference among them, hence I also

[26] Detailed discussion and review in Lee and Lemieux (2010).

include the percentage of SC/ST population in the year 2001 as controls.

D. Analysis of Schooling Infrastructure

To look at the growth of schooling infrastructure, I provide several figures showing the trend of schooling facilities and number of schools over the academic years 2005-2011. The school census is available annually from 2005 and finding RD estimates for each year could be demanding to observe any effect.[27] Hence, I establish the below framework.

I create a panel dataset of the blocks using the DISE data for the academic years 2005-2011. I aggregate the school level variables to find the values for a block and calculate them as percentages. I consider indicator variables like whether a school has electricity, pre primary section, is a girls school and so on. I aggregate these to derive the percentages of schools in a block that have a facility; for example percentages of schools that are electrified. Similarly, I find the percentage of good classrooms, classrooms needing repair. I also find the average number of visits a representative school of a block can receive from an official belonging to the block or cluster resource centres.

I aim to observe two things; firstly, whether the facilities varied across EBB/NEBB in the seven academic years and secondly, the trend of the facilities over time. I perform this by estimating the below regression:

$$Y_i = \alpha + \beta EBB_i + \gamma EBB_i years + \delta f(X_i) + \lambda treat f(X_i) + \varepsilon_i \tag{3}$$

EBB=1 if a block is an EBB, years is a categorical variable to represent the seven academic years, the third term is the interaction term and x is the rural female literacy rate for 2001. I estimate the above equations for f(x) being a linear, quadratic or cubic function of x. The coefficients of interest are β and γ. Once again since the EBB variable is endogenous I instrument EBB with the variable treat and EBB*years with treat*years.

E. Regression Discontinuity Validation

The RDD analysis is valid under certain assumptions: There must be no manipulation of the treatment variables around the cutoff, the covariates are balanced across the cutoff and the assignment variable is continuous. I provide evidence in support of each of these assumptions being satisfied. The results represented as figures and tables are provided in the appendix.

Firstly, manipulation of the treatment variables is highly unlikely as RFLR and GGLR are aggregate level variables determined after the collection of census data by finding the number of

[27] Results available on request

literates in the total population. The Census in India is conducted and the data is released by the Registrar General and Census Commissioner of India which is a separate organisation. It is thus not possible for blocks to manipulate their aggregate literacy indicators to be on either side of the cutoff. I validate this assumption using the McCrary test (McCrary 2007). Figure A.ii is presented in the appendix and I find that there was no manipulation around the cutoff in the assignment variable ruling out possibility of bunching or manipulation. The estimates obtained under the test imply acceptance of the null hypothesis of no manipulation.

Secondly, the blocks could differ in various demographic indicators like population size, population of males, population of children in the age group 0 to 6, proportion of people from the minorities like SC/ST, and so on. These could influence the socio economic condition of a block and hence the impact of any treatment. Thus, I focus on the above variables to check for the balance of covariates in the sample. Figure A.iii in appendix presents the discontinuity plots for the covariates. The figure has different plots for each of the demographic variables against the assignment variable of rural female literacy rate for 2001. The plots do not show any significant discontinuities around the cutoff for any of the variables. Finally, the assignment variable being the rural female literacy rate and gender gap in total literacy rates, are continuous.

IV. Results

I have divided the results of the paper into four subsections below.

A. Literacy Indicators

I want to find out if being classified as EBB brought a significant increase in the literacy indicators. The classification was based on low female rural literacy rates and gender gap in literacy rates which led to provision of gender specific schemes. The program was implemented in 2000-2001 and thus the literacy rates in 2011 would be after a decade of the implementation of the program. So, there is an entire generation of children who are expected to be affected by it.

First Stage Estimates

Figure 1 shows that the probability of being classified as EBB increased by approximately 70 percentage points at the cutoff, i.e on having a RFLR less than the national average of 46.13% when they already had a gender gap in total literacy rates above the national average of 21.59%. The first stage estimates are represented in table 4. Columns (2) to (4) take into account the controls, state fixed effects. The estimates remain consistent with inclusion of controls or state fixed effects and are around 75% points. The first row represents the RD estimate using the optimal bandwidth based on Imbens and Kalyanaraman (2009). The second row provides estimates using half of this bandwidth and the third row for double the optimal bandwidth for robustness check. The estimates hardly fluctuate with the size of bandwidth. Column (5) and the fourth row presents the bias corrected robust RD estimates obtained using the CCT methodology and the estimates are of the same order. Rest of the tables in this section have the same format.[28]

Reduced Form Estimates

The outcome variables of interest are the increase in rural female literacy rates and the decrease in gender gap in rural literacy rates in the period 2001 to 2011. I use the change in these variables over the past decade to obtain my second stage of the regression. Figure 2 presents the RD graphs for the first outcome variable, which is the percentage increase in the rural female literacy rates from year 2001 to 2011. The CCT bandwidth has been estimated to be around 6.2. The first plot is obtained without including any controls and the consecutive graphs have incorporated additional controls. The figure shows that there is no significant discontinuity in the outcome variable at the cutoff. It is clear that the insignificant result does not change with inclusion of controls, except that the increase in the literacy rates are reduced to much smaller number.

A concern in RD estimation is the test can be low powered, however the confidence interval for the estimates are small implying precise estimation of the zero result. Thus, there exists a possibility of an effect of 1 percentage points which is however very small compared to an increase of 10 percentage points experienced by blocks on the right of the cutoff. Table 5 presents the estimates for the outcome variable increase in the rural female literacy rates and we see that the point estimates are not statistically significant.

Figure 3 presents the similar analysis for the second outcome variable which is the decrease in gender gap in rural literacy rates from year 2001 to 2011. No significant decrease in the gender gap in rural literacy rates is observed. The confidence interval for the estimate is small ruling out concerns of imprecise estimation. As evident from the figure the maximum effect that cannot be ruled out is of 0.7 percentage points which once again is a small effect. The point estimates are represented in Table 6.[29]

B. Funding and Schooling Infrastructure

An important step is to find out if the planned proposal for building schools was converted to reality. Using the DISE data I analyse if number of schools built and schooling infrastructure facilities increased from the launch of the program. Also, it is of interest to see if the funds allocated to the states had any relation to the proportion of educationally backward blocks which would provide some affirmation to the fact that the program prioritized areas with higher EBBs.

[28] I also include the case for excluding states which did not have any block classified as EBB classified in the appendix and the results remain similar.

[29] I also provide the similar analysis for total literacy rate indicators in the appendix and the result remains similar.

Figure 4 plots the allocation and expenditure of SSA funds (in million dollars) to states vis-a-vis the percentage of EBBs.[30] The figure shows the funding allocated and spent under SSA had a positive relation with the proportion of EBBs in a state.[31]

Expansion of the schooling infrastructure and building schools was a major component of the program and we see from Figure 5 that the average number of schools built in a block increased over the years. It was once again important to observe if the number of schools built differed across EBB/NEBB. In this case it is seen that in the past decade on average more schools were built in EBB compared to NEBB. The average number of schools built in the rural region of a block in a year varies between 4 and 6 with few outliers.

The average number of girls schools built in an EBB increased steeply post 2004. There is no clear distinction in the number of boys schools but the number of coeducational schools seems to be declining, although remaining higher for EBBs compared to NEBBs. From the figure it is evident that the KGBVs were built only in the EBBs based on the sample considered. The KGBV schools appear in the data mainly after 2005.[32]Also, the larger number of schools built in EBBs is also reflected in higher number of schools with primary or middle section. [33]

C. Other Schooling facilities

To analyse if other schooling facilities have increased over the years, I plot figure 6 for some of the variables like classrooms, electrification, computers, etc. We would expect the schooling condition to vary based on whether a school was old or new. For example, the focus for a new school will be to build classrooms as it does not have any but that for an old school will be to first repair the existing ones in a bad situation and build additional if necessary. Thus, I have done the analysis to capture the variation in the situation by the year of establishment of the school.

Figure 6 presents the level of schooling facilities over the academic years 2005 and 2011. The figure shows that there was an increase in number of classrooms and classrooms in good condition over the academic years along with probability of having electricity, computers and SSA fund. We can see that the proportion of good classrooms are higher in the newer schools but probably will need more time to catch up with the old schools in the number of computers

[30]The funds value in rupees have been converted into dollars using the exchange rate for the respective year

[31]I have depicted the relation for the funding in year 2008 for clarity but the same relation holds in other years too

[32]I have used the DISE dataset for 2013 to trace back all schools that were established until 2012 and have drawn the figures for all schools built till 2010. I excluded private unaided schools and schools in urban areas. The KGBV schools are recognised based on the residential type of the school.

[33]I have also provided similar figure for the number of schools per thousand children in a block for the last decade and the same figure but for the previous decade in the appendix. The results remain the same when the variable used is number of schools per thousand children. For the years 1990-2000 we see that not much difference was observed between number of schools built in EBBs vs NEBBS

or electricity. This could also be because the old schools have been functioning for longer period of time and could be having more enrollment and capacity. Hence, the different plots tell us that the availability and condition of schooling infrastructure has improved over the academic years and is different between schools that are old and new represented by the year of establishment.

D. IV Estimates for Schooling Infrastructure

The estimates obtained for variables of interest on schooling facilities obtained from equation (3) are reported in Table 7. From the table we can see that the percentages of schools electrified are higher in NEBBs compared to EBBs but the trend over the years have been positive as given by the coefficient for the EBB*years variable. Similar relation is observed for other variables like having a boundary wall, medical checkup, good classrooms, etc. Whereas for some variables like percentages of schools with pre primary section, the EBBs have a higher percentage but it has a declining trend. This could be due to the total number of schools growing at a higher rate in the decade compared to ones with pre primary sections. The declining trend is also observed for the number of pre primary students and teachers which are reported as the average number for a school in the block and not percentages. This implies that the numbers of pre primary students and teachers have fallen in the period.

The estimates also imply the percentage of classrooms needing minor repair is higher in EBB but a definite conclusion cannot be made about the percentage of classrooms needing major repair. There are various dimensions of this. With a new school built, the number of classrooms increase, but those are also the ones which are in good condition or need minor repairs. So, as new schools are built in EBB, the classrooms in those blocks on average will require minor repair rather than major ones. In addition, we see that the trend has been negative implying that the quality of classrooms in an EBB has risen over time.

To help with the planning and monitoring block and cluster resource centres (BRC and CRC) were set up. The officials from this centre were to visit the schools or sites of construction. From the table we can see that the number of visits by an official from CRC are higher in EBBs but not the visits by the BRCs. Clusters are smaller units than blocks and this could be causing some of the difference. However, the visits have been declining. It is also evident that a school in EBB has received less school development grant (SDG) on average but the number has grown over the years. But, SDG is received and reported only by an existing school and not by schools that are in the process of construction. I do not have presently any data to illustrate the nature and amount of funds received by a new school.

V. Discussion

The RD estimates obtained for literacy indicators can be interpreted as an effect of classification of blocks as EBB or of schemes that were implemented in blocks based on their EBB status like NPEGEL and KGBV. To be conservative I would like to take the above estimates as the ITT estimates of SSA. This is because to find the effect of the entire program, knowing the classification of blocks as EBB is not enough. The first stage would need an outcome variable like the amount of funds allocated for each block under SSA and whether that varied differentially across the threshold. Moreover, it would require a detailed study on the fund allocation of other rural development schemes and crowding out or involvement of funds from these schemes, sources and data for which I currently do not have.

A possible issue in the nature of results obtained can be the choice of outcome variable. Literacy rate may not be a good measure. A drawback of this measure is that this captures all the literate population above age 7, and to derive more meaningful estimates a cohort based analysis would be useful. But, the census of India does not provide data for population by different age groups at the block level which is a limitation of data.[34] In addition, it is plausible that it may take more time for such infrastructure investments to cause a change in the percentage of literate population or be reflected in the literacy rate indicators. However, in this paper the literacy rates used are for the blocks which are smaller regions, so an increase in literate population in the era hopefully would have been captured.

A major concern to obtain treatment on the treated estimate would be if such a large scale program was implemented effectively and efficiently. Most importantly did the schools receive fund as promised and reported, were the funds received in a timely manner and finally what was the rule followed for allocation of money to the blocks or schools after disbursement from the Central and State governments.

Following the question of if the schools received the funds, the next question is do the schools spend the money they receive and what do they spend it on. I use the DISE data to plot Figure 7. The SSA funds are disbursed in the form of three grants; school development grant (SDG), school maintenace grant (SMG) and teaching learning material grant (TLM).[35]The DISE data reports each of these funds received and spent by schools over the academic years 2005-2011. I aggregate this information at the block level to obtain the grant received and the expenditure by all the schools in a block. Finally, I calculate the difference in the amount of funds received and spent.

For comparison purposes I have calculated the funds received by all the schools in a block per

[34]Other surveys are not at the sub district level and besides the surveys do not provide sub district identifiers.

[35] Data on funds under NPEGEL and KGBV scheme is unavailable at the block level currently. Also, these are the funds received for open and operational schools

thousand children in thousand dollars.[36]From figure 7 we can see the difference is positive, i.e the expenditure has always been lower than the amount received suggesting unspent balance. To understand if the situation differed across a EBB and NEBB on average, I split my analysis across EBB and NEBB. We can see there is more unspent balance in the NEBB and the unspent balance of the SMG has grown dramatically over the years. Data on the components of expenditure is unavailable at the block level but from the allocations and expenditure report at state level, the major components of expenditure were civil works and teacher salaries.[37]

VI. Conclusion

In this paper, I study if schooling infrastructure can contribute to the literacy of a region and help decrease the gender gap in literacy. To answer this I use a policy experiment in India which took the form of a nationwide education mission. It is important to study the effect on literacy rates as literacy rates reflect the number of literates in the nation and for example; a low literacy rate in the largest democracy implies a large illiterate population. Literacy rates are also important variables for policy making and knowing what factors can improve it is necessary.

Using regression discontinuity I find that incorporating the needs of girls by specific interventions did not cause effect in either the total or rural female literacy rates. Similar result was found for the gender gap in literacy. Being conservative I have proposed them to be intent to treat estimates but ITT can be considered to be the relevant estimate as this tells us about the impact that can be expected in reality from large scale programs like the SSA.

In addition, I have provided analysis or evidence on inputs made in the schooling infrastructure which is a good proxy for funds or the ground level reality and to show outcomes of the program. The results obtained depict the improvement of infrastructure and schooling facilities in the EBBs. To, the best of my knowledge this is the first paper dealing with a panel dataset for seven academic years of around a million school in India.

Some interesting future works would be to study the allocation of funds and crowding in or out of funds from other programs. Looking at measures of learning for children like test scores as outcome variables will be helpful and can substantiate the current findings of the paper. It is also possible that some aspects of investment in schooling facilities could be more effective than

[36] For the population of children I use the data from Census 2001 which reports population of children aged 0 to 6 in rural region of the respective block. Population of children in the age group 0 to 6 may not be the best measure, but I use this as a proxy for population of children that would go to schools. I also divide the funds variable measured in rupees by the exchange rate for the respective academic year in which the fund was received by the school as reported and captured in the DISE data. I finally convert the variable to be measured as funds in thousand dollars (per thousand children).

[37]ASER in their PAISA survey conducted in 2012 found evidences on schools spending their money but rarely taking up repair work

others. For example providing free textbooks could be more effective than building drinking water facilities and having a detailed component level analysis of the program could help understand inputs that work better for education and literacy.

References

- ASER, *Annual Status of Education Report*, 2011
- ASER, Paisa survey. *Do schools get their money*, 2012
- Banerjee, Abhijit V., et al. *Can information campaigns spark local participation and improve outcomes? A study of primary education in Uttar Pradesh, India.A Study of Primary Education in Uttar Pradesh, India* (July 1, 2006). World Bank Policy Research Working Paper 3967 (2006).
- Banerjee, Abhijit, et al. *Pitfalls of participatory programs: Evidence from a randomized evaluation in education in India. No. w14311.* National Bureau of Economic Research, 2008.
- Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik. *Robust Nonparametric Confidence Intervals for RegressionDiscontinuity Designs.* Econometrica 82.6 (2014): 2295-2326.
- Case, Anne, and Angus Deaton. *School quality and educational outcomes in South Africa.* No. 993. 1998.
- Duflo, Esther. *Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment.* No. w7860. National Bureau of Economic Research, 2000.
- Jalan, Jyotsna, and Elena Glinskaya. *Improving Primary School Education in India: An Impact Assessment of DPEP-Phase 1.* World Bank, Washington, DC (2003).
- Imbens, Guido, and KarthikKalyanaraman. *Optimal bandwidth choice for the regression discontinuity estimator.*The Review of Economic Studies (2011): rdr043.
- Lavy, Victor. *School supply constraints and children's educational outcomes in rural Ghana.* Journal of Development Economics 51.2 (1996): 291-314.
- Lee, David S., and Thomas Lemieux. *Regression discontinuity designs in economics.* No. w14723. National Bureau of Economic Research, 2009.
- Meller, Marian, and Stephan Litschig. *Adapting the Supply of Education to the Needs of Girls: Evidence from a Policy Experiment in Rural India.* No. 805. 2014.
- MHRD, Manual for planning and appraisal, Apr 2004
- Muralidharan, Karthik. *Priorities for primary education policy in Indias 12th Five-year plan.* India Policy Forum. Vol. 9. 2013.
- Nichols, Austin. "rd: Stata module for regression discontinuity estimation." Statistical Software Components (2014).

- Planning Commision,Programme Evaluation Organisation.*Evaluation Report on SarvaShikshaAbhiyan*. May,2010

- SSA framework documents,Planning, Appraisal and Funds Flow under SarvaShikshaAbhiyan (SSA)*http://ssa.nic.in/ssa-framework/planning-appraisal-and-funds-flow-under-sarva-shiksha-abhiyan-ssa*

- UNESCO, case study on SarvaSikshyaAbhiyan, 2015 *http://unesdoc.unesco.org/images/0023/002330/233008*

Tables

Table 1: Summary Statistics for Total Sample, 2001

	EBB		NEBB	
Variable	Mean	Std. Dev	Mean	Std. Dev.
%Scheduled Caste	16.272	8.25	17.844	10.813
%Males	51.196	1.285	51.015	1.422
%0 to 6 child	17.843	3.08	14.343	2.475
%Scheduled tribe	15.02	23.78	11.21	20.934
Rural Female Literacy Rate'01	33.433	8.04	57.936	9.228
Gender Gap in Tot Literacy Rate'01	28.076	5.711	20.244	6.735

Note: The above table summarises some of the key demographic variables based on Census 2001

Table 2: Summary Statistics for Total Sample, 2011

	EBB		NEBB	
Variable	Mean	Std. Dev	Mean	Std. Dev.
%Scheduled Caste	16.687	8.609	18.915	11.285
%Males	51.045	1.297	50.932	1.779
%0 to 6 child	15.121	3.433	11.817	2.342
%ST	15.636	24.245	11.747	21.012
Rural Female Literacy Rate'11	48.341	7.367	67.161	8.827
Gender Gap in Tot Literacy Rate'11	22.005	4.803	15.161	5.763

Note: The above table summarises some of the key demographic variables based on Census 2011

Table 3: Variable Averages for a school, 2005-2011

	Aca Year 2005	Aca Year 2011
No of visits by CRC	4.87	5.75
No of visits by BRC	1.61	1.93
SSA fund received	4,025	12,755
SSA fund spent	3,570	11,587
No of classrooms	3.2	3.8
%Schools having electricity	22	38
No of computers	0.27	0.83

Note: The table summarises some of the variables for schooling infrastructure using the DISE data for the two academic years of 2005 and 2011.

Table 4: First Stage Estimates-EBB on RFLR

Bandwidths	(1)	(2)	(3)	(4)	(5)
IK Optimal	-0.754***	-0.729***	-0.756***	-0.731***	
	(0.0759)	(0.0834)	(0.0754)	(0.0841)	
Half of IK Optimal	-0.733***	-0.675***	-0.732***	-0.668***	
	(0.111)	(0.139)	(0.112)	(0.150)	
Double of IK Optimal	-0.768***	-0.761***	-0.768***	-0.762***	
	(0.0515)	(0.0534)	(0.0512)	(0.0531)	
CCT					-0.778***
					(0.0480)
Control	X	State	Others	State+Others	X

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The table represents the first stage estimates where the dependent variable is proportion of EBB and the independent variable is the rural female literacy rates in 2001 of the blocks. Column (1) is the estimate obtained without taking any controls, column (2) is by including state fixed effects for 31 States/UTs, column (3) is by including controls. The controls used are percentages of Scheduled caste and Scheduled tribe in the respective rural areas based on Census 2001 data. Column (4) is taking all the controls in column (3) and state fixed effects. The above estimates are obtained using the user written code of Austin Nicolas. The optimal bandwidth is calculated based on Imbens and Kalyanaraman (2009) which for this regression has been obtained around 2.6. The second row provides estimates using half of this bandwidth and the third row for double the optimal bandwidth. The fourth row is the RD estimate obtained from the user written code by Calonico, Cattaneo and Titiunik (2014) with the corresponding optimal bandwidth which was 6.1

Table 5: RD estimates for effect of EBB classification on increase in Rural female literacy rates

Bandwidths	(1)	(2)	(3)	(4)
IK Optimal	0.406	0.0849	0.364	
	(1.010)	(0.672)	(1.002)	
Half of IK Optimal	0.470	0.855	0.412	
	(1.463)	(1.003)	(1.441)	
Double of IK Optimal	0.157	0.236	0.122	
	(0.675)	(0.446)	(0.672)	
CCT				0.521
				(1.228)
Control	X	State	Others	

Standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Note: The outcome considered is the increase in rural female literacy rates of the block from 2001 to 2011. Column (1) is the estimate obtained without taking any controls, column (2) is by including state fixed effects for 31 states/UTs, column (3) is by including controls.The controls used are percentages of Scheduled caste and Scheduled tribe in the respective rural areas based on Census 2001 data.Column (4) is taking all the controls in column (3) and state fixed effects. The above estimates are obtained using the user written code of Austin Nicolas. The optimal bandwidth is calculated based on Imbens and Kalyanaraman (2009) which for this regression has been obtained as 6. The second row provides estimates using half of this bandwidth and the third row for double the optimal bandwidth. The fourth row provides the bias corrected robust estimates obtained from the user written code by Calonico, Cattaneo and Titiunik (2014) with the corresponding optimal bandwidth which was 6.2

Table 6: RD estimates for effect of EBB classification on decrease in gender gap in rural literacy rates

Bandwidths	(1)	(2)	(3)	(4)	(5)
IK Optimal	0.823	0.299	0.804	0.305	
	(0.570)	(0.392)	(0.564)	(0.392)	
Half of IK Optimal	0.716	0.514	0.672	0.523	
	(0.824)	(0.581)	(0.810)	(0.586)	
Double of IK Optimal	0.599	0.324	0.577	0.344	
	(0.376)	(0.259)	(0.374)	(0.259)	
CCT					0.815
					(0.610)
Control	X	State	Others	Sta+Others	

Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Note: Gender gap in rural literacy rate for a block=Male rural literacy rate-Female rural literacy rate of the block. The outcome considered is the decrease in gender gap in rural literacy rate of the block from 2001 to 2011 and hence is constructed by subtracting the gender gap in rural literacy rate of 2011 from gender gap in rural literacy rate of 2001. Column (1) is the estimate obtained without taking any controls, column (2) is by including state fixed effects for 31 states/UTs, column (3) is by including controls for percentage of rural SC population, percentage of rural ST population. Column (4) is taking all the controls in column (3) and state fixed effects. All the control variables are of the 2001 census. The above estimates are obtained using the user written code of Austin Nicolas. The optimal bandwidth is calculated based on Imbens and Kalyanaraman (2009) which for this regression has been obtained around 5. The second row provides estimates using half of this bandwidth and the third row for double the optimal bandwidth. The fourth row provides the bias corrected robust estimates obtained from the user written code by Calonico, Cattaneo and Titiunik (2014) with the corresponding optimal bandwidth which was 6.2

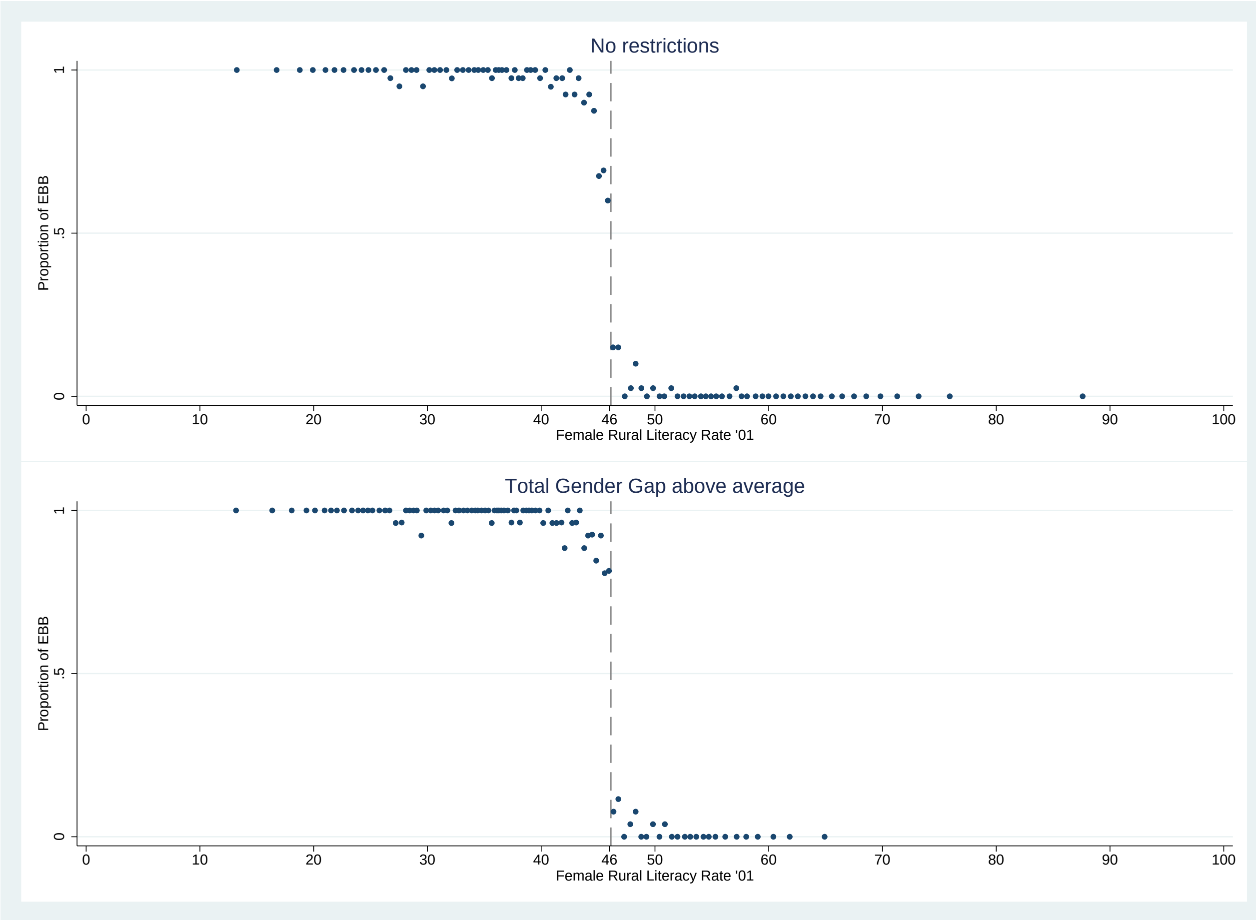
Table 7: Estimates for the schooling facilities and its trend

	EBB			EBB*Years		
	(1)	(2)	(3)	(1)	(2)	(3)
Variables	Linear	Quadratic	Cubic	Linear	Quadratic	Cubic
% of Schools having electricity	-30.74***	-28.32***	-33.21***	5.190***	5.060***	5.095***
	(2.134)	(2.321)	(3.055)	(0.246)	(0.167)	(0.164)
% of Schools with Pre Primary section	11.22***	11.52***	12.05***	-2.125***	-1.963***	-1.914***
	(1.116)	(1.473)	(1.890)	(0.122)	(0.104)	(0.0988)
% of Schools with medical checkup facility	-0.758	-3.225	-0.622	1.562***	1.371***	1.334***
	(2.662)	(2.359)	(2.756)	(0.315)	(0.180)	(0.157)
% of Schools with boundary wall	-8.925***	-9.154***	-9.307***	1.704***	1.917***	1.915***
	(1.585)	(1.606)	(2.156)	(0.186)	(0.115)	(0.115)
% of girls school	1.448***	1.265***	1.709***	-0.0112	0.0221	0.0238
	(0.260)	(0.229)	(0.287)	(0.0366)	(0.0188)	(0.0192)
% of boys school	0.613***	0.822***	1.027***	-0.0118	-0.00745	-0.00290
	(0.164)	(0.172)	(0.246)	(0.0193)	(0.0117)	(0.0138)
% of good classrooms	-3.806***	-3.046**	-2.867*	1.355***	1.418***	1.499***
	(1.303)	(1.414)	(1.551)	(0.142)	(0.0903)	(0.0796)
% of classrooms needing minor repair	3.228***	4.373***	3.742***	-1.280***	-1.327***	-1.335***
	(0.880)	(0.777)	(0.890)	(0.100)	(0.0596)	(0.0531)
% of classrooms needing major repair	0.612	1.217**	0.0875	-0.180***	-0.263***	-0.241***
	(0.545)	(0.520)	(0.778)	(0.0603)	(0.0401)	(0.0456)
No. of girls toilet	0.830***	0.863***	0.884***	-0.200***	-0.204***	-0.206***
	(0.0242)	(0.0328)	(0.0404)	(0.00247)	(0.00205)	(0.00183)
No. of books in library	-131.6***	-152.9***	-151.0***	17.18***	17.43***	16.68***
	(21.87)	(21.85)	(22.53)	(2.561)	(1.665)	(1.148)
No. of computers	-0.664***	-0.583***	-0.614***	0.131***	0.115***	0.102***
	(0.0779)	(0.0849)	(0.0928)	(0.0129)	(0.00876)	(0.00642)
No. of Pre Primary Section students	4.568***	4.863***	4.882***	-0.578***	-0.509***	-0.518***
	(0.848)	(0.862)	(1.223)	(0.124)	(0.0847)	(0.0990)
No. of Pre Primary Section teachers	0.105***	0.0957***	0.0835**	-0.0108**	-0.0104**	-0.00750**
	(0.0323)	(0.0346)	(0.0421)	(0.00488)	(0.00422)	(0.00371)
No. of inspections	0.168	0.348***	0.336**	-0.00708	-0.0236**	-0.0216***
	(0.133)	(0.126)	(0.146)	(0.0144)	(0.00927)	(0.00818)
No. of visits by CRC	0.811***	1.783***	2.747***	-0.128***	-0.137***	-0.143***
	(0.283)	(0.378)	(0.450)	(0.0328)	(0.0291)	(0.0247)
No. of visits by BRC	-0.633***	-0.754***	-0.731***	0.000687	0.00231	0.00363
	(0.107)	(0.168)	(0.188)	(0.0111)	(0.0118)	(0.00948)
School development grant received	-1,814***	-1,825***	-1,557*	563.8***	593.3***	591.1***
	(537.7)	(550.9)	(805.8)	(51.67)	(28.48)	(30.08)
TLM fund received	-17.30	53.93	16.79	33.68***	39.05***	39.75***
	(81.24)	(76.51)	(99.07)	(9.318)	(5.517)	(5.284)

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

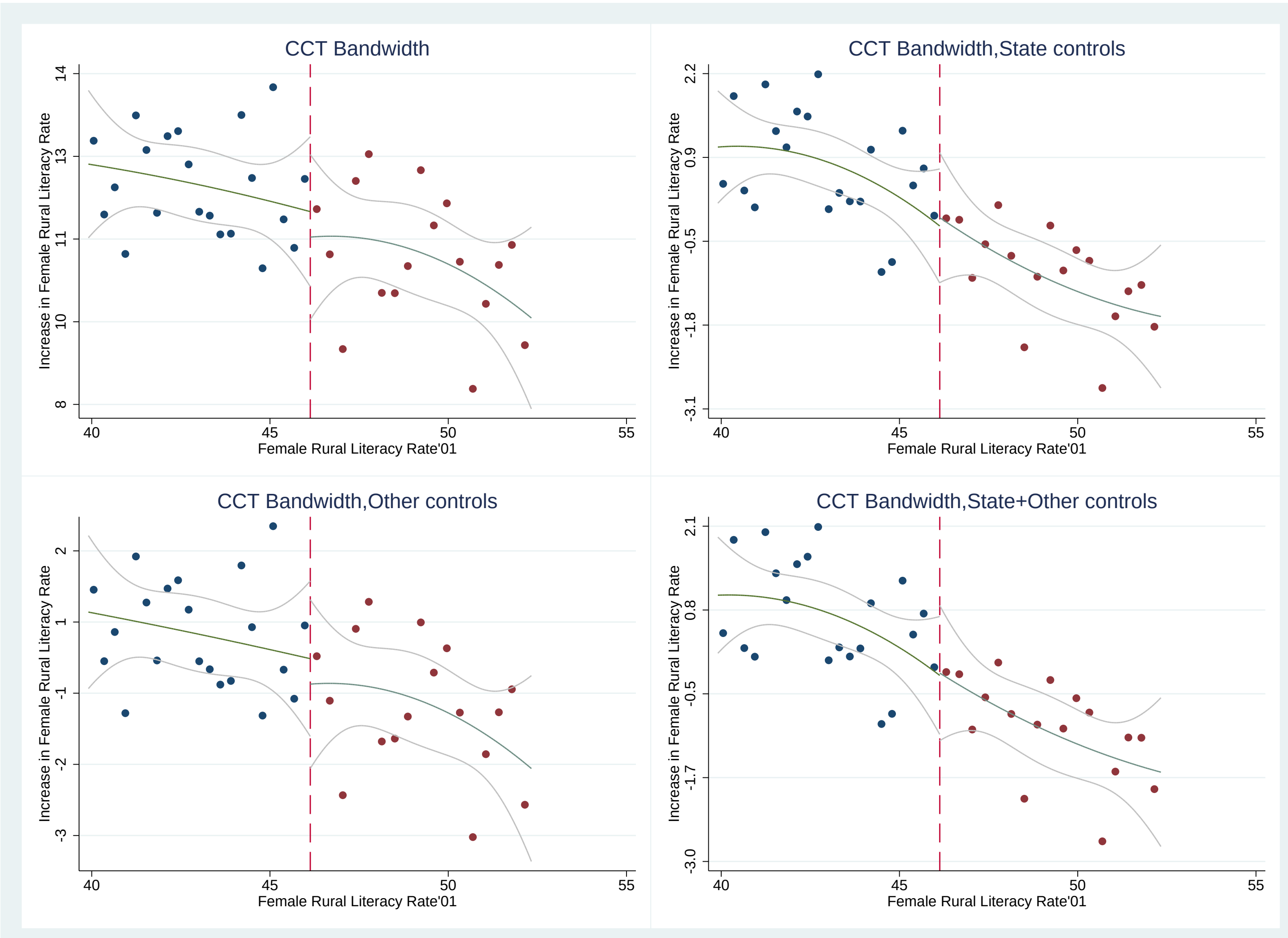
Figures

Figure 1: Regression Discontinuity for first stage



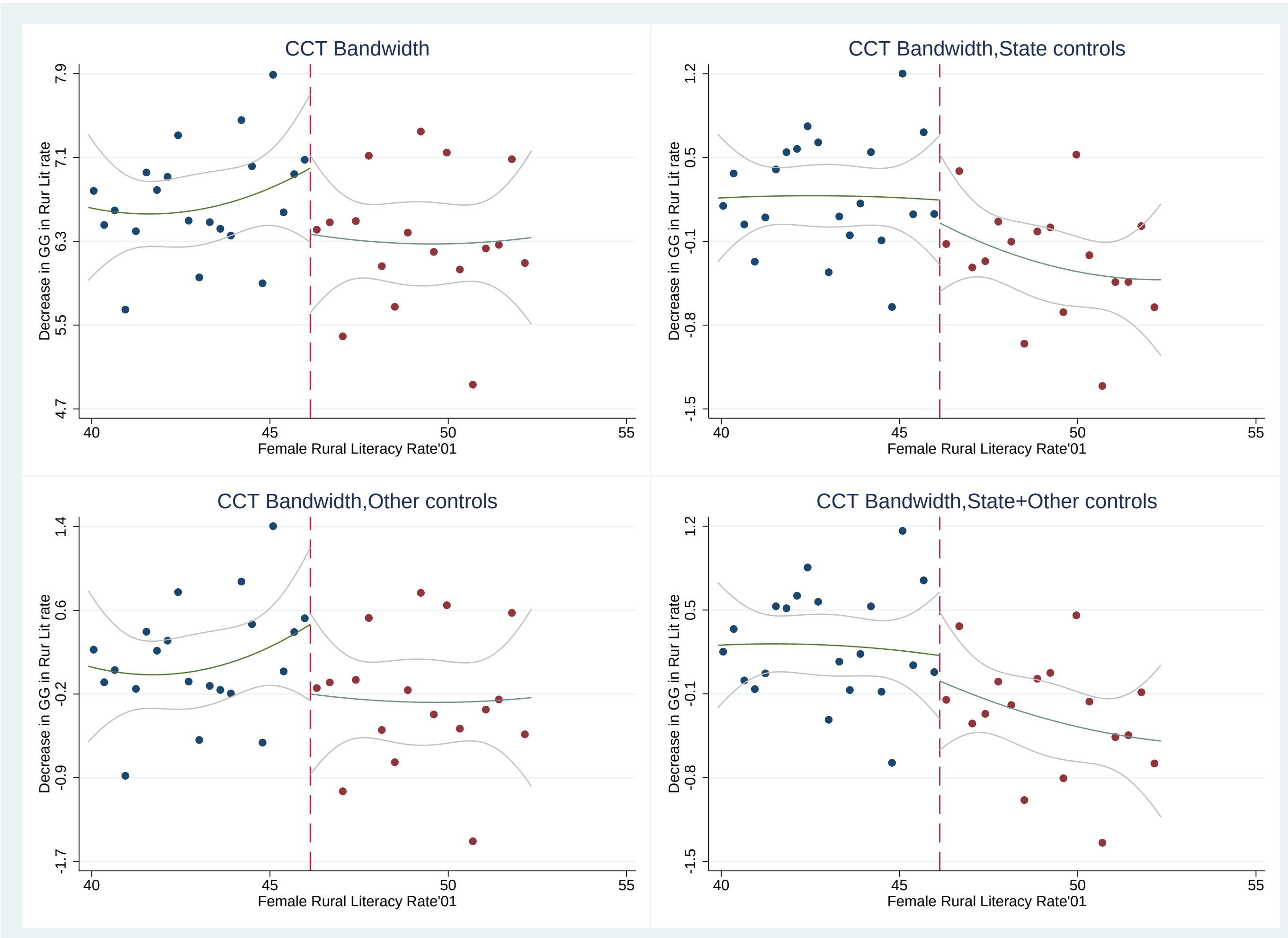
The figure shows that the probability of being classified as EBB increased by 70 percentage points at the cutoff.

Figure 2: Regression Discontinuity for reduced form outcomes-Increase in Rural Female literacy rates



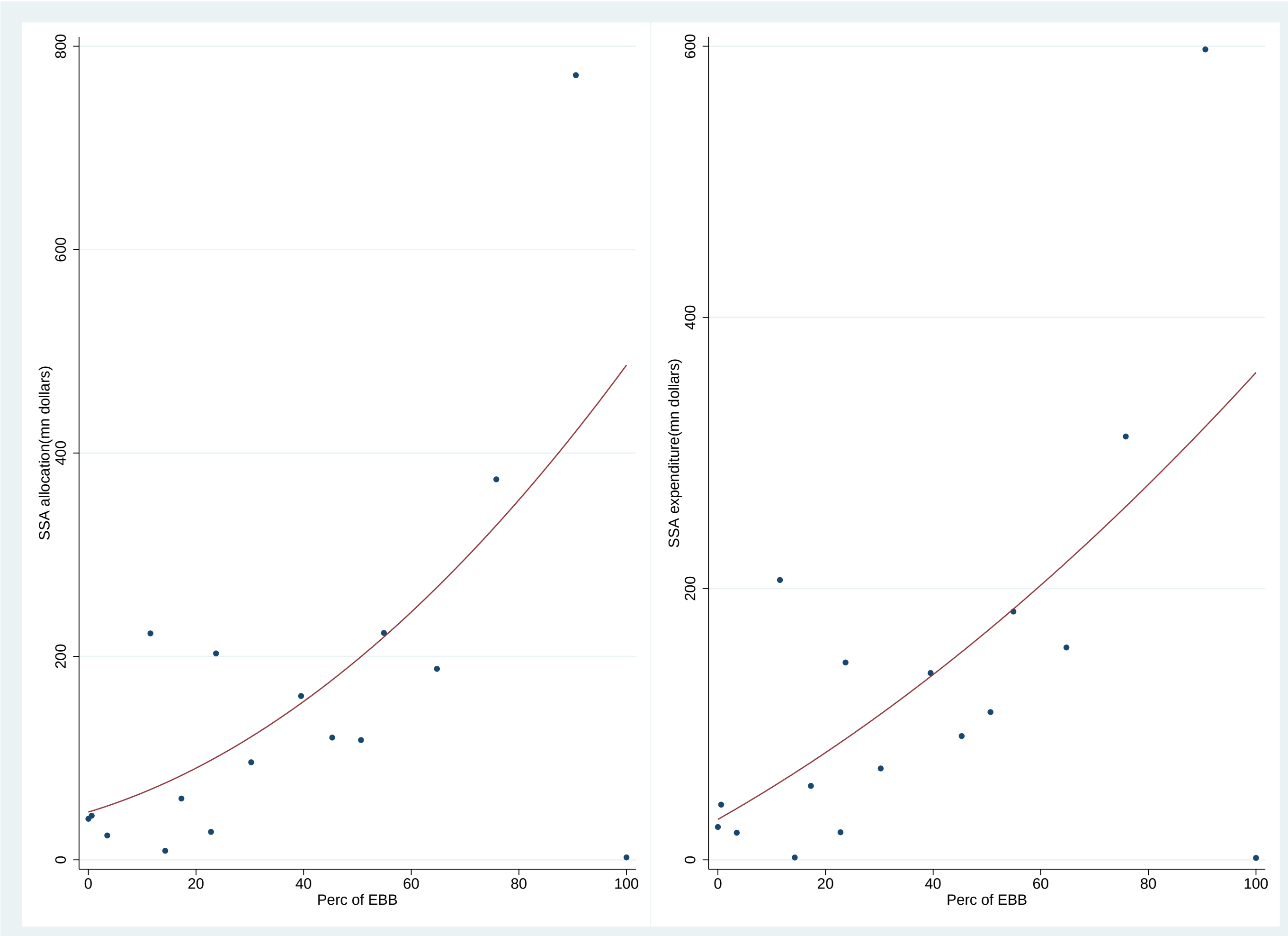
The figure plots the increase in rural female literacy rates over the decade for the optimal bandwidth calculated based on Calonico,et al (2014). The graphs show that there was no significant discontinuity at the cutoff with or without controls. The controls used are percentages of Scheduled caste and Scheduled tribe in the respective rural areas based on Census 2001 data.

Figure 3: Regression Discontinuity for reduced form outcomes-Decrease in Gender gap in rural literacy rates



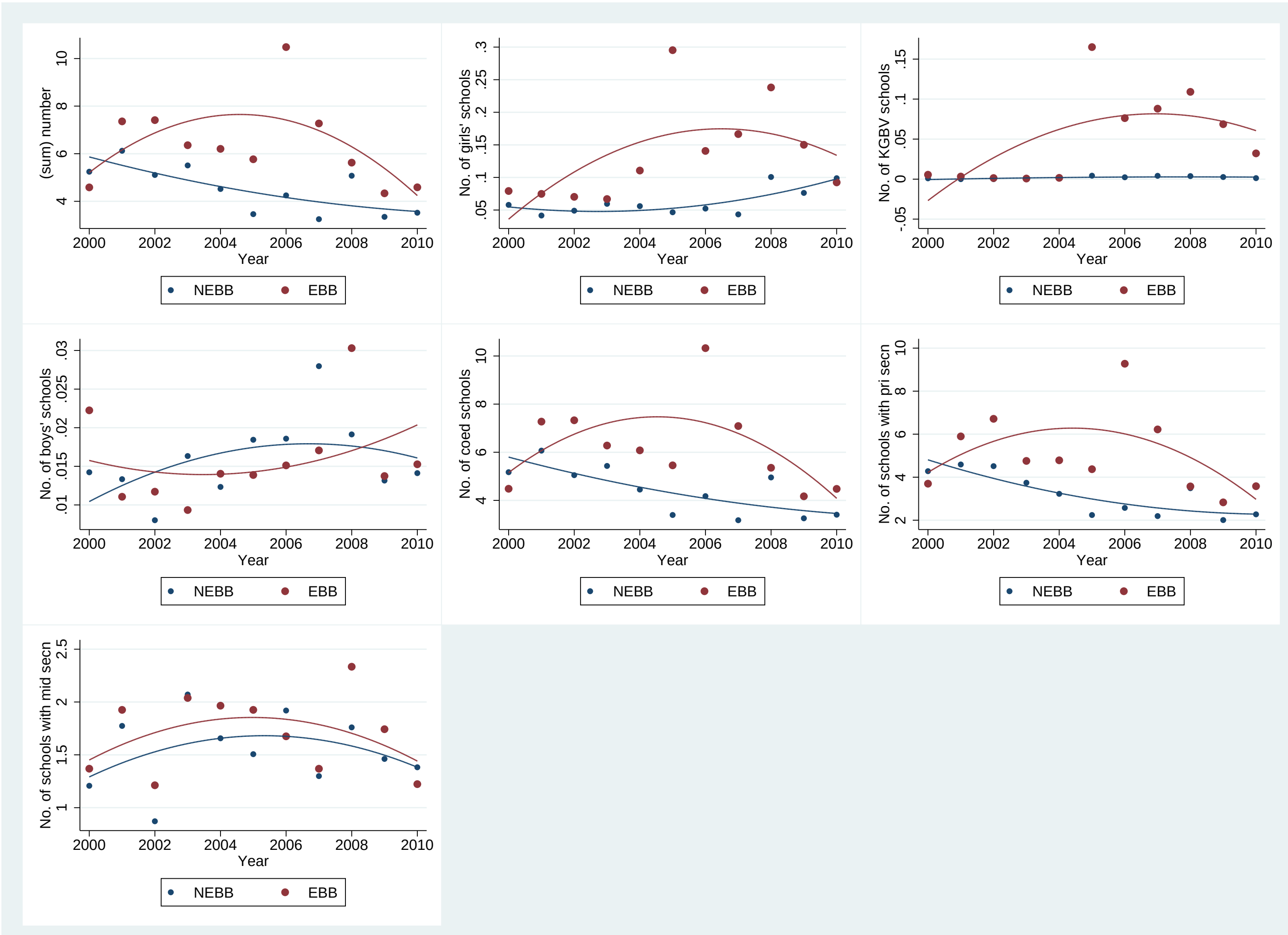
The figure plots the decrease in gender gap in rural literacy rates over the decade for the optimal bandwidth calculated based on Calonico,et al (2014). The graphs show that there was no significant discontinuity at the cutoff with or without controls. The controls used are percentages of Scheduled caste and Scheduled tribe in the respective rural areas based on Census 2001 data.

Figure 4: SSA funds allocated and spent



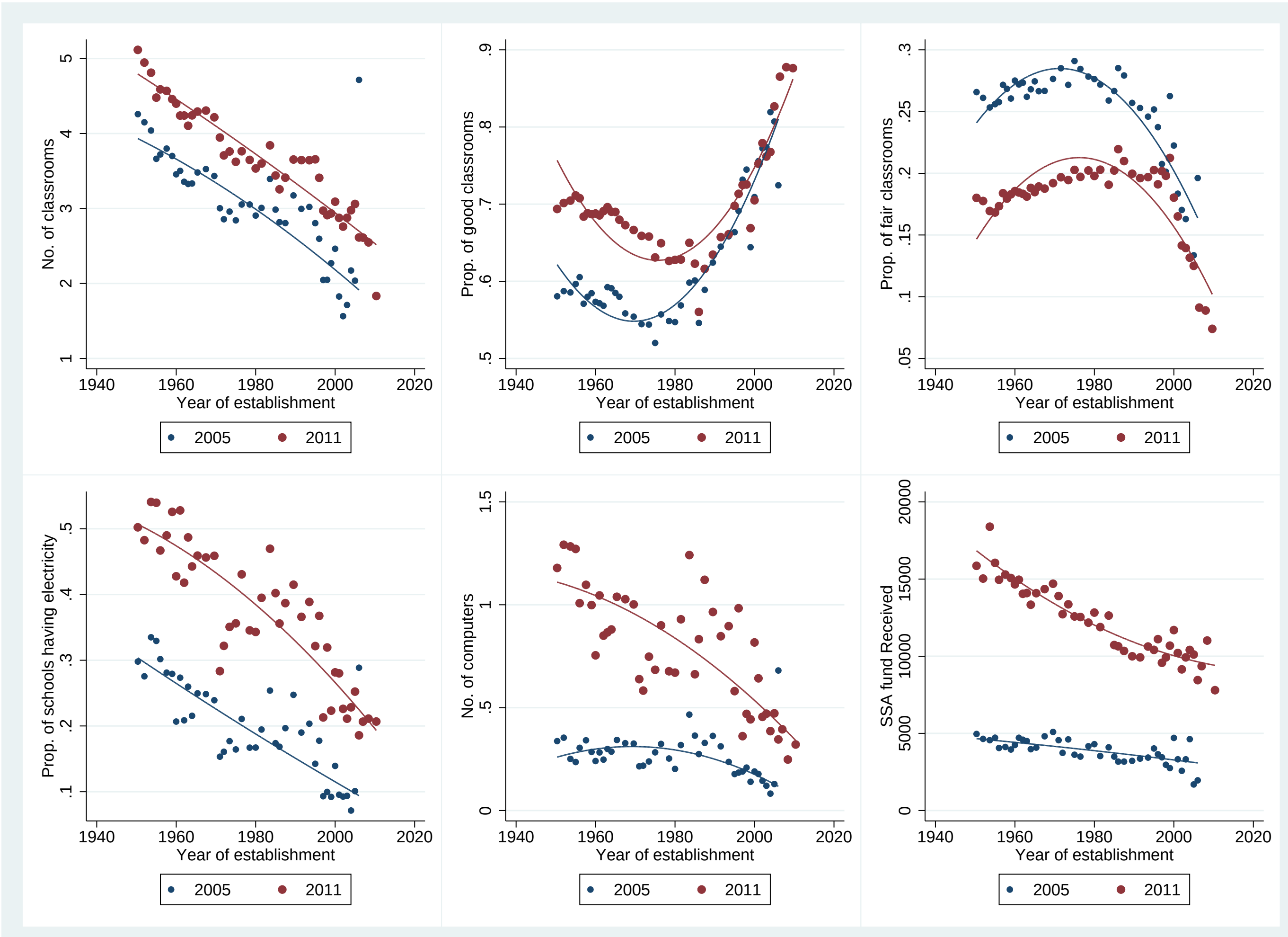
The figure shows that there was a positive relation between the proportion of EBB in a state with the SSA fund allocation and expenditure

Figure 5: Average number of Schools built in the last decade



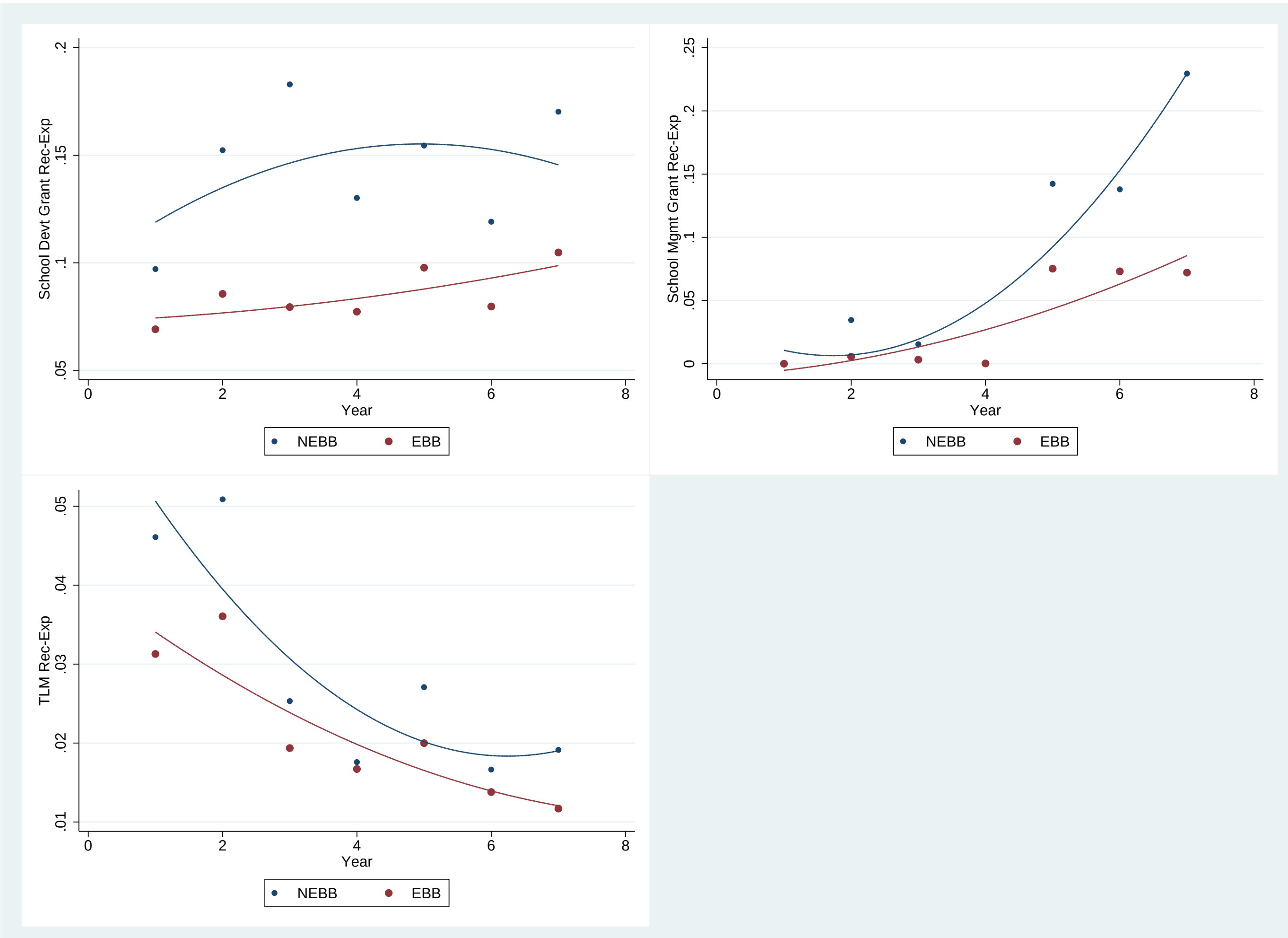
The figure plots the average number of schools built in an EBB/NEBB in the last decade. The growth of different kinds of schools is also depicted above. We see that in the past decade on average more schools were built in EBB compared to NEBB. KGBV schools have been built only in the EBBs and the number of girls schools have been on the rise too.

Figure 6: Situation of Infrastructure by year of establishment of school



The figure shows that there was an increase in number of classrooms and classrooms in good condition over the academic years alongwith probability of having electricity,computers and SSA fund. We can see that the proportion of good classrooms are higher in the newer schools but probably will need more time to catch up with the old schools in the number of computers or electricity. This could also be because the old schools have been functioning for longer period of time and could be having more enrollment and capacity.

Figure 7: Difference in the received and expenditure amounts of the various components of SSA funds

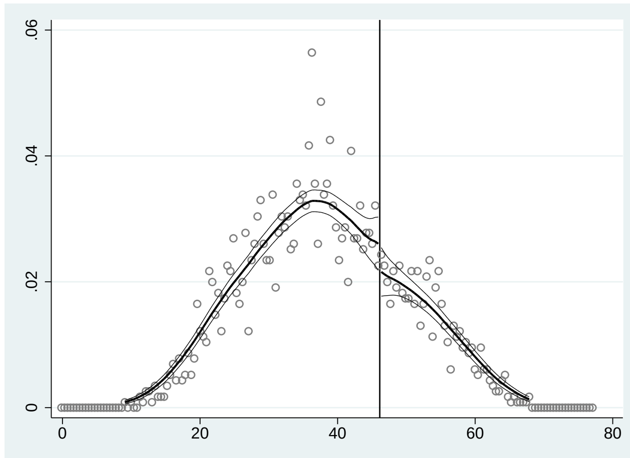


Rec-received, Exp-Expenditure. The figure plots the difference in received and expenditure amounts of the school development grant (SDG), school maintenace grant (SMG), teaching learning material grant (TLM) over the academic years 2005-2011. The variables are measured in 1000 dollars and per thousand children in the rural region of the block.The years are labelled from 1 to 7. The graphs provide a comparison between the EBB vis-a-vis NEBB. It shows that the expenditure of funds was lower than the funds received and there is more unspent balance in the NEBB. The unspent balance of the SMG has grown dramatically over the years.

A Appendix

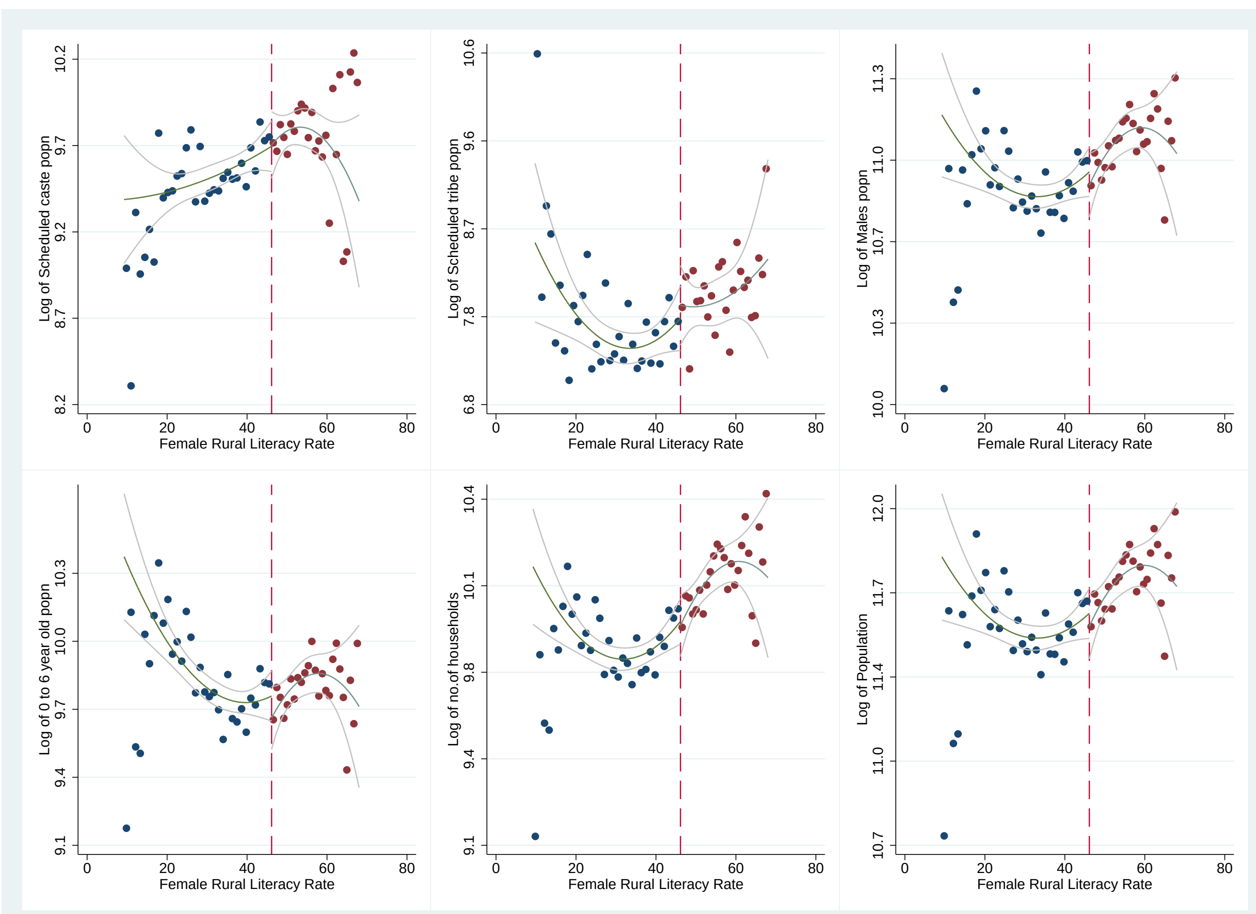
A. Additional figures and tables

Figure A.ii: Testing for Manipulation in the assignment variables



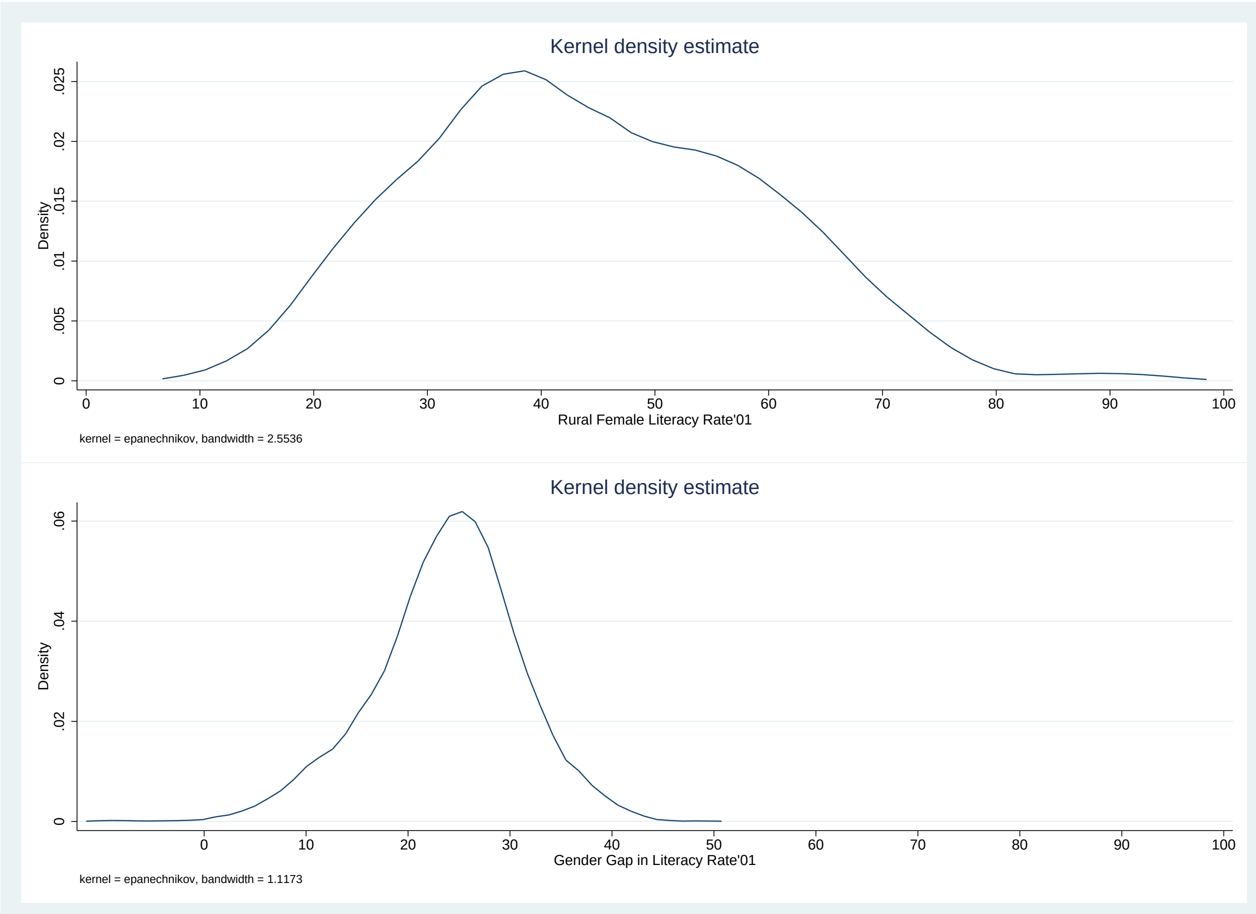
The figure shows that there was no discontinuity at the cutoff ruling out chances of manipulation.

Figure A.iii: Covariates Balanced Test



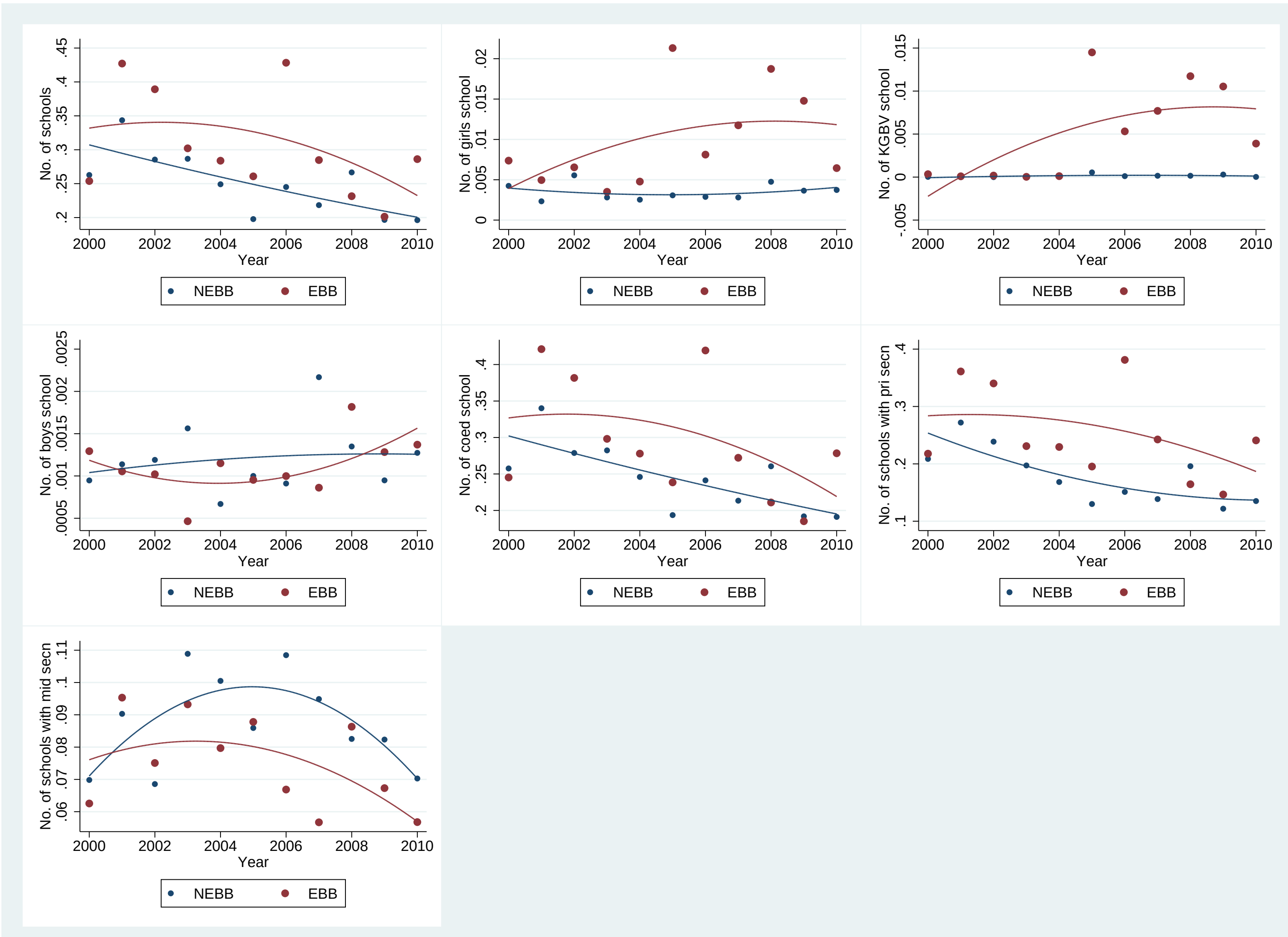
The figure shows that there was no discontinuity at the cutoff for any of the covariates.

Figure A.iv: Density Plots for the assignment variables



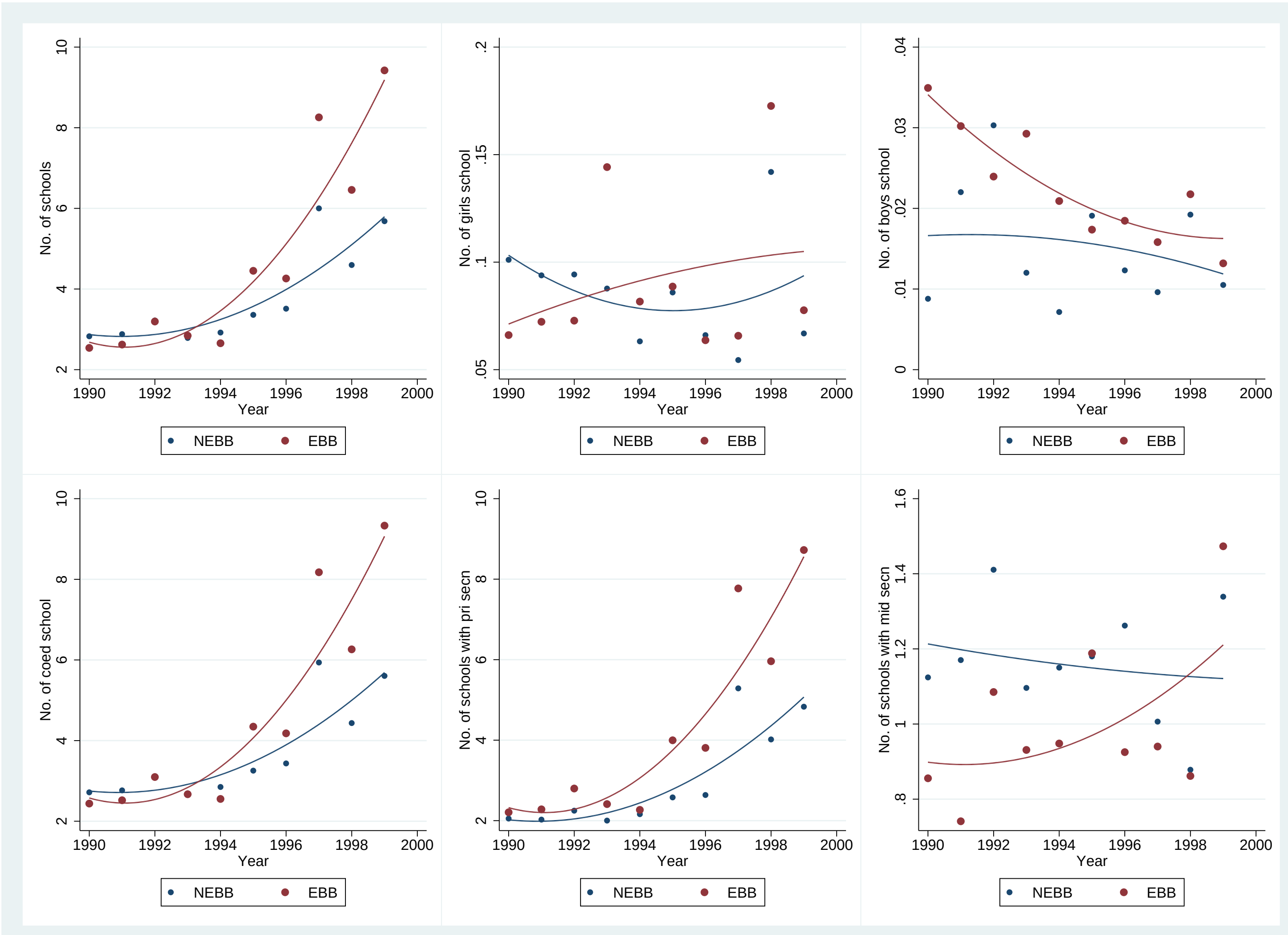
The figure shows that there was no discontinuity at the cutoff for any of the covariates.

Figure A.v: Average number of Schools built per 1000 children in the last decade



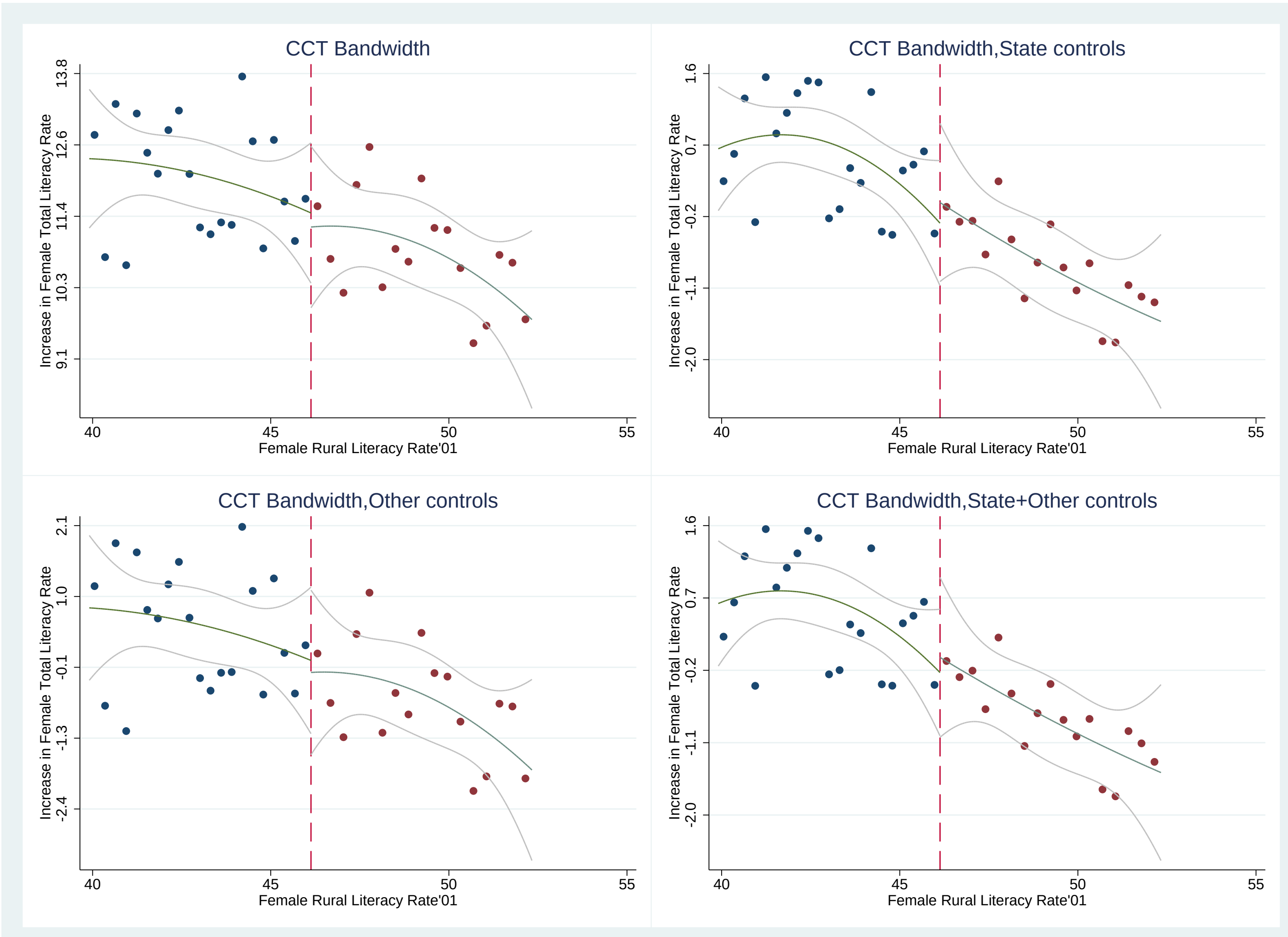
The figure plots the average number of schools per 1000 children built in an EBB/NEBB in the last decade. The growth of different kinds of schools is also depicted above. We see that the number of schools built in a EBB remains higher than in a NEBB.

Figure A.vi: Average number of Schools built per 1000 children in the years 1990-2000



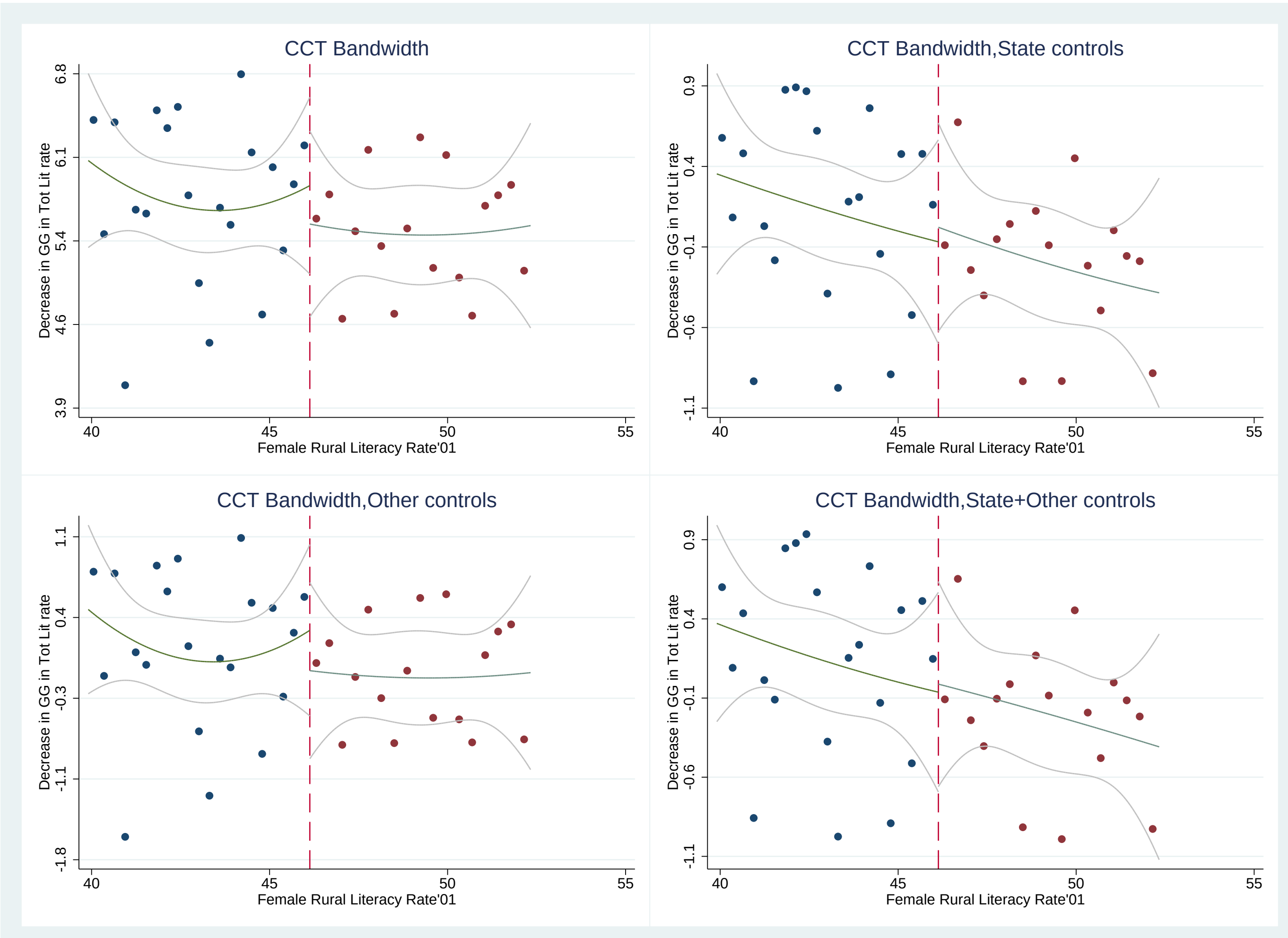
The figure plots the average number of schools built in an EBB/NEBB in the decade of 1990-2000. The growth of different kinds of schools is also depicted above. We see the number of schools built in a EBB vis--vis NEBB is not very different in this decade compared to the last decade.

Figure A.vii: Regression Discontinuity for reduced form outcomes-Increase in Total Female literacy rates



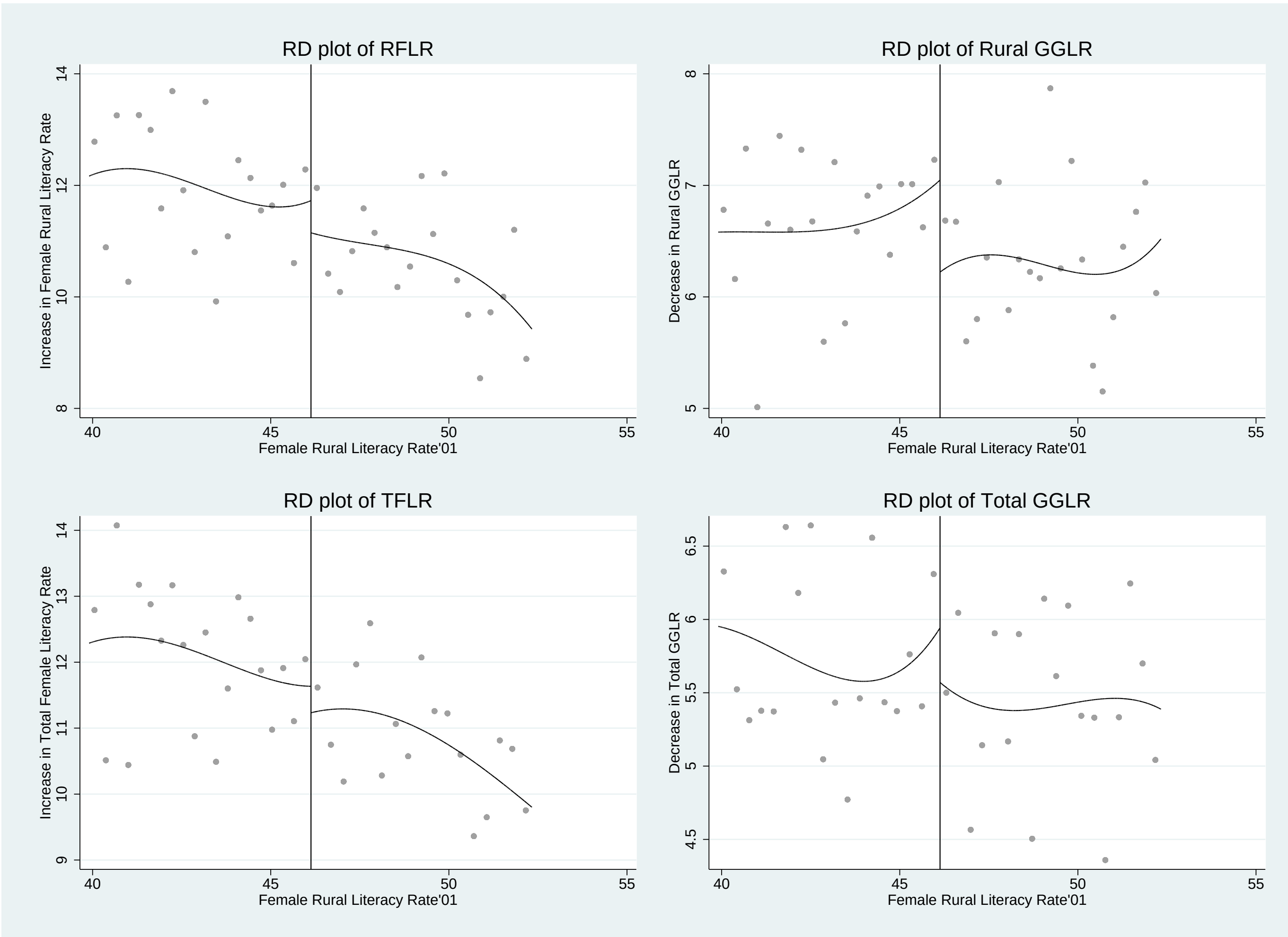
The figure plots the increase in total female literacy rates over the decade for the optimal bandwidth calculated based on Calonico,et al (2014). The graphs show that there was no significant discontinuity at the cutoff with or without controls. The controls used are percentages of Scheduled caste and Scheduled tribe in the respective rural areas based on Census 2001 data.

Figure A.viii: Regression Discontinuity for reduced form outcomes-Decrease in Gender gap in total literacy rates



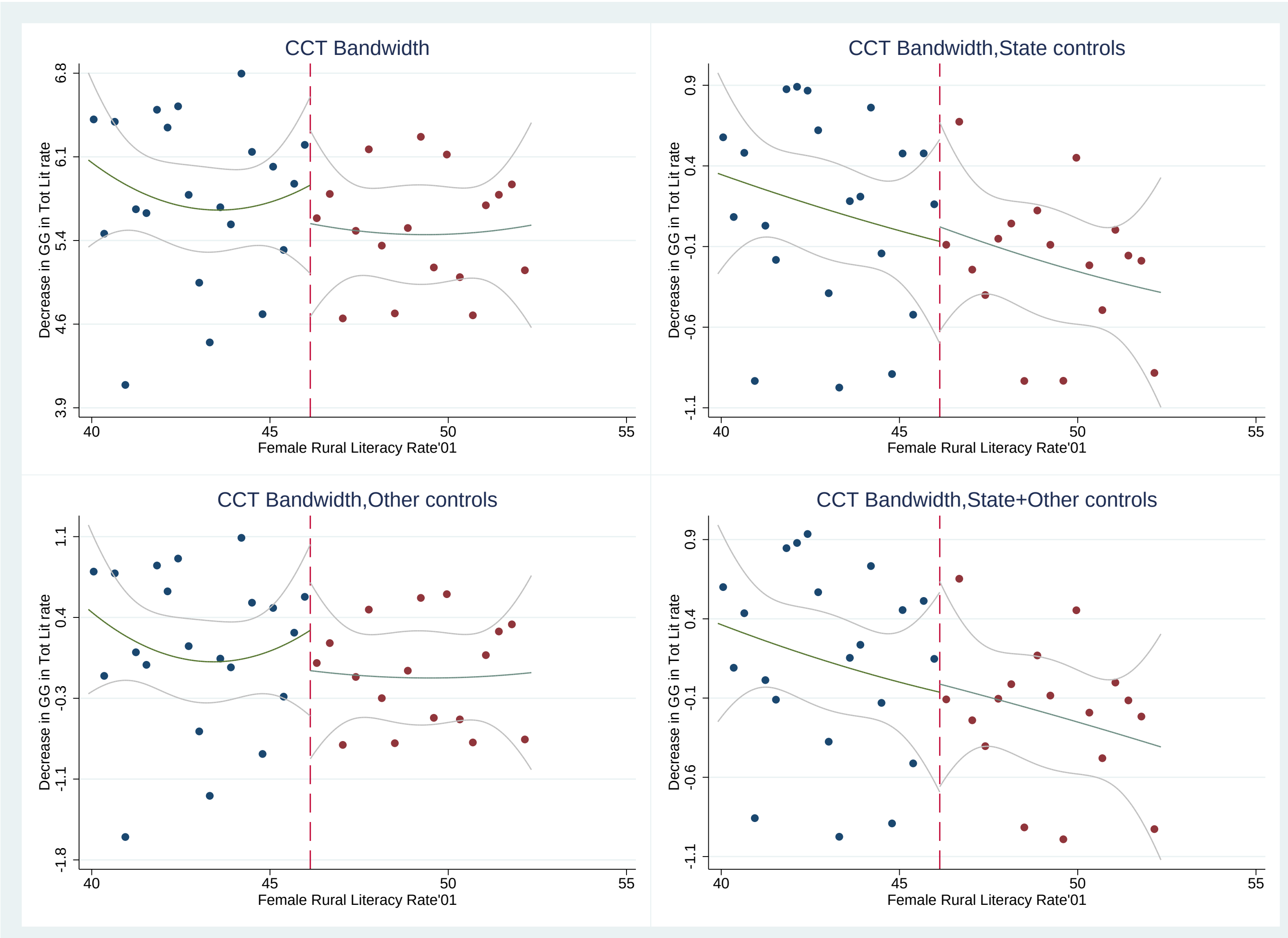
The figure plots the decrease in gender gap in rural literacy rates over the decade for the optimal bandwidth calculated based on Calonico,et al (2014). The graphs show that there was no significant discontinuity at the cutoff with or without controls. The controls used are percentages of Scheduled caste and Scheduled tribe in the respective rural areas based on Census 2001 data.

Figure A.ix: Regression Discontinuity plots for Literacy rates



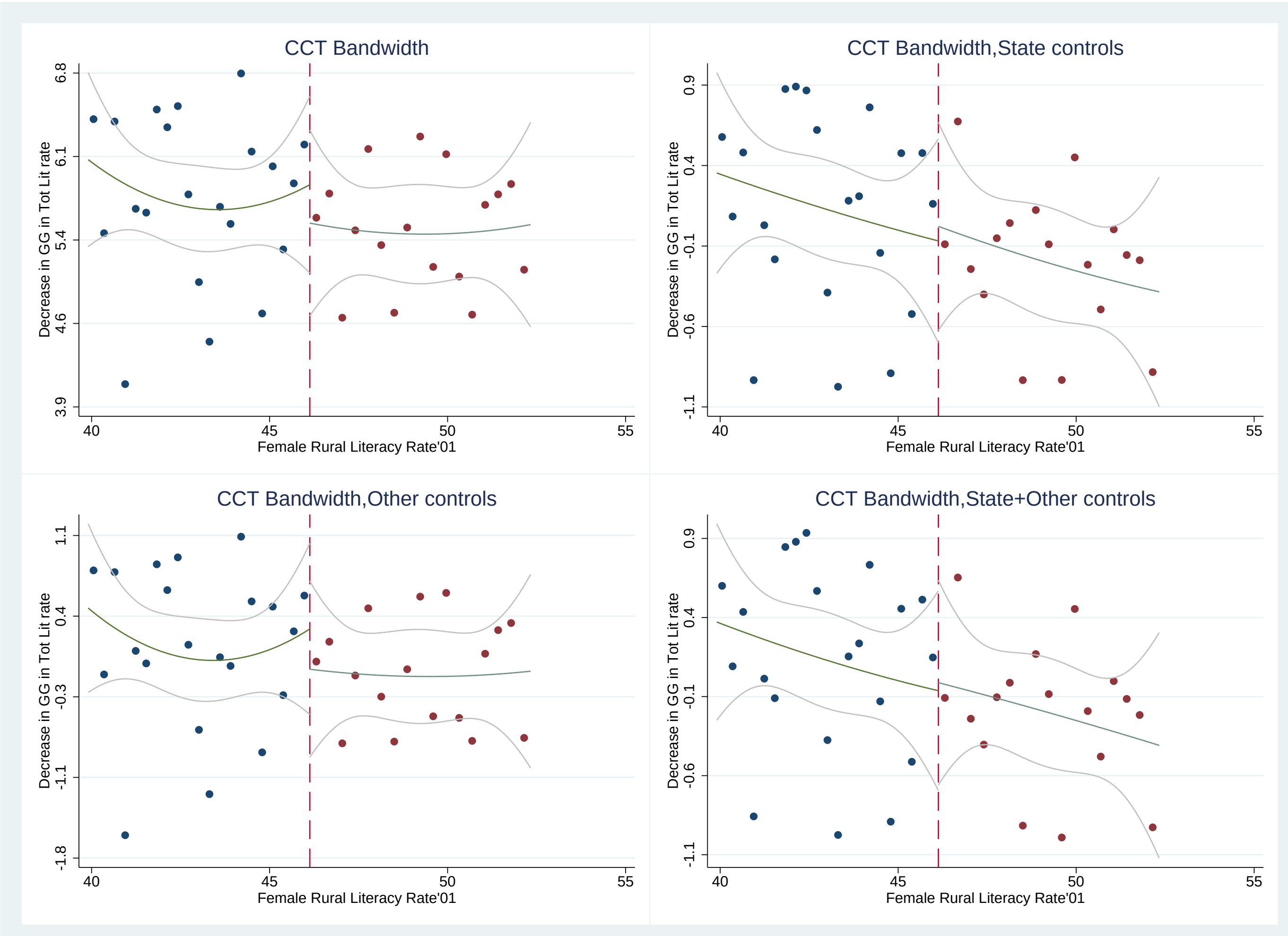
The figures are data driven regression discontinuity plots based on the methodology of Calonico, Cattaneo and Titiunik (2014)

Figure A.x: Regression Discontinuity for reduced form outcomes-Decrease in Gender gap in total literacy rates



The figure plots the decrease in gender gap in rural literacy rates over the decade for the optimal bandwidth calculated based on Calonico,et al (2014). The graphs show that there was no significant discontinuity at the cutoff with or without controls. The controls used are percentages of Scheduled caste and Scheduled tribe in the respective rural areas based on Census 2001 data.

Figure A.xi: Regression Discontinuity for reduced form outcomes-Decrease in Gender gap in total literacy rates



The figure plots the decrease in gender gap in rural literacy rates over the decade for the optimal bandwidth calculated based on Calonico,et al (2014). The graphs show that there was no significant discontinuity at the cutoff with or without controls. The controls used are percentages of Scheduled caste and Scheduled tribe in the respective rural areas based on Census 2001 data.

Table 7: IV 2SLS estimtates for different functional forms

VARIABLES	(1) Linear	(2) Quadratic	(3) Cubic
Increase in Total Female Literacy rate	0.197	0.120	0.900
	(0.768)	(1.203)	(1.658)
Increase in Rural Female Literacy rate	0.231	0.667	0.639
	(0.856)	(1.344)	(1.827)
Decrease in Total Gender gap in Literacy	0.260	0.358	0.572
	(0.474)	(0.743)	(1.012)
Decrease in Rural Gender gap in Literacy	0.637	0.866	1.039
	(0.440)	(0.691)	(0.941)
Bandwidth	CCT	CCT	CCT

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The above estimates provide the IV 2SLS estimates for four outcome variables. Observations in the optimal CCT bandwidth were considered for obtaining the estimates. The estimates were obtained without taking any controls.

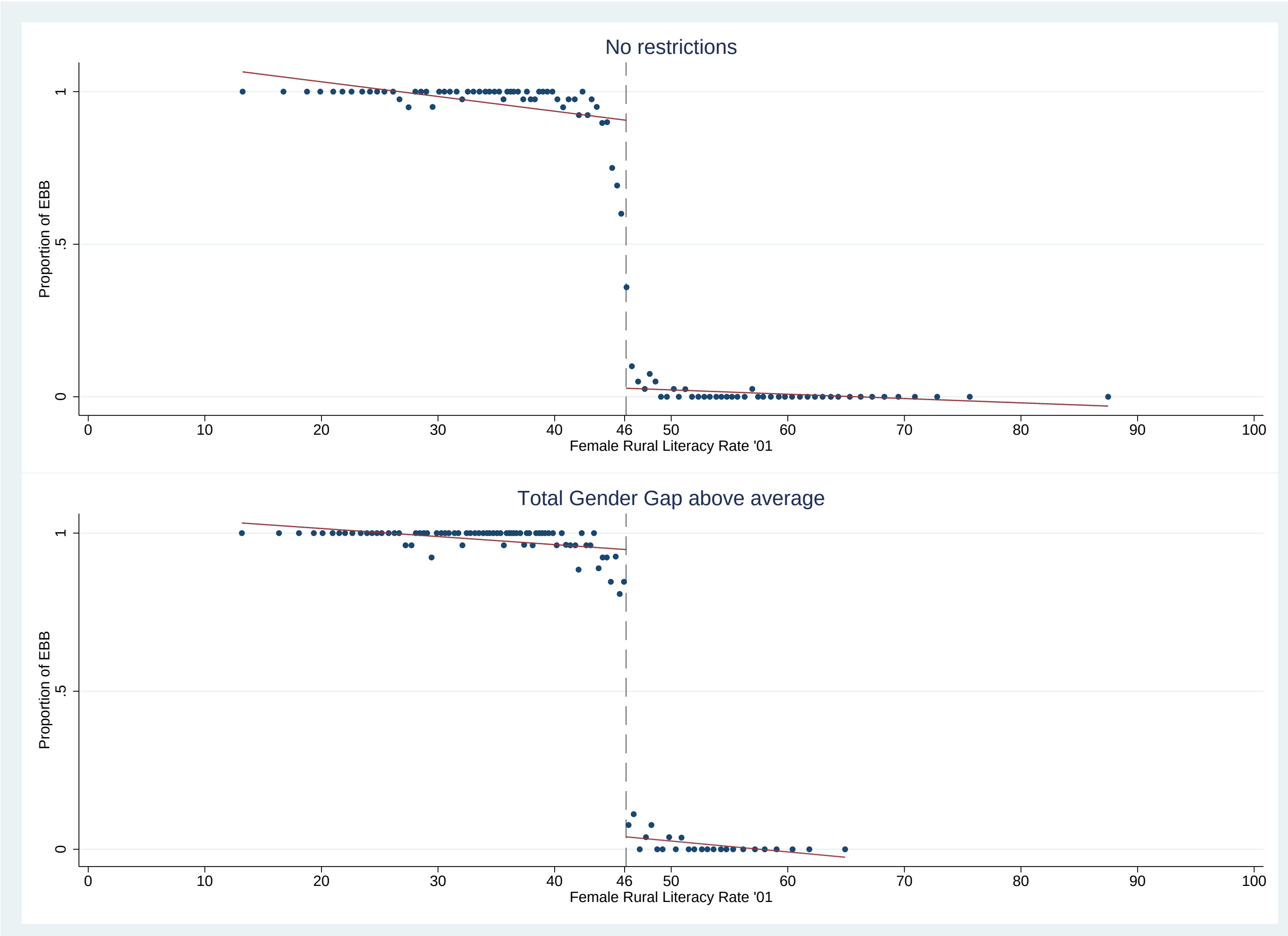
B. Analysis based on exclusion of states with no EBBs

The states and Union territories of Delhi, Chandigarh, Goa, Andaman and Nicobar Islands, Sikkim, Lakshadweep, Puducherry, Daman and Diu, Dadra and Nagar Haveli had no blocks classified as educationally backward. I replicate the entire analysis for a sample excluding these states forms my current sample. My previous sample of 31 states and UTs however did not have Delhi, Chandigarh, Puducherry or Dadra and Nagar Haveli. They were lost during the matching process, but the above states and UTs are small and had very few blocks. For this analysis I also drop the observations for state of Arunachal Pradesh from my previous sample which has only two observations. This would reduce the number of states to be included as fixed effects.

Thus, on dropping rest of the states my sample for this analysis reduces to 3,964. This also implies there were very few observations belonging to these states in my previous sample, however for completeness I replicate the analysis done.

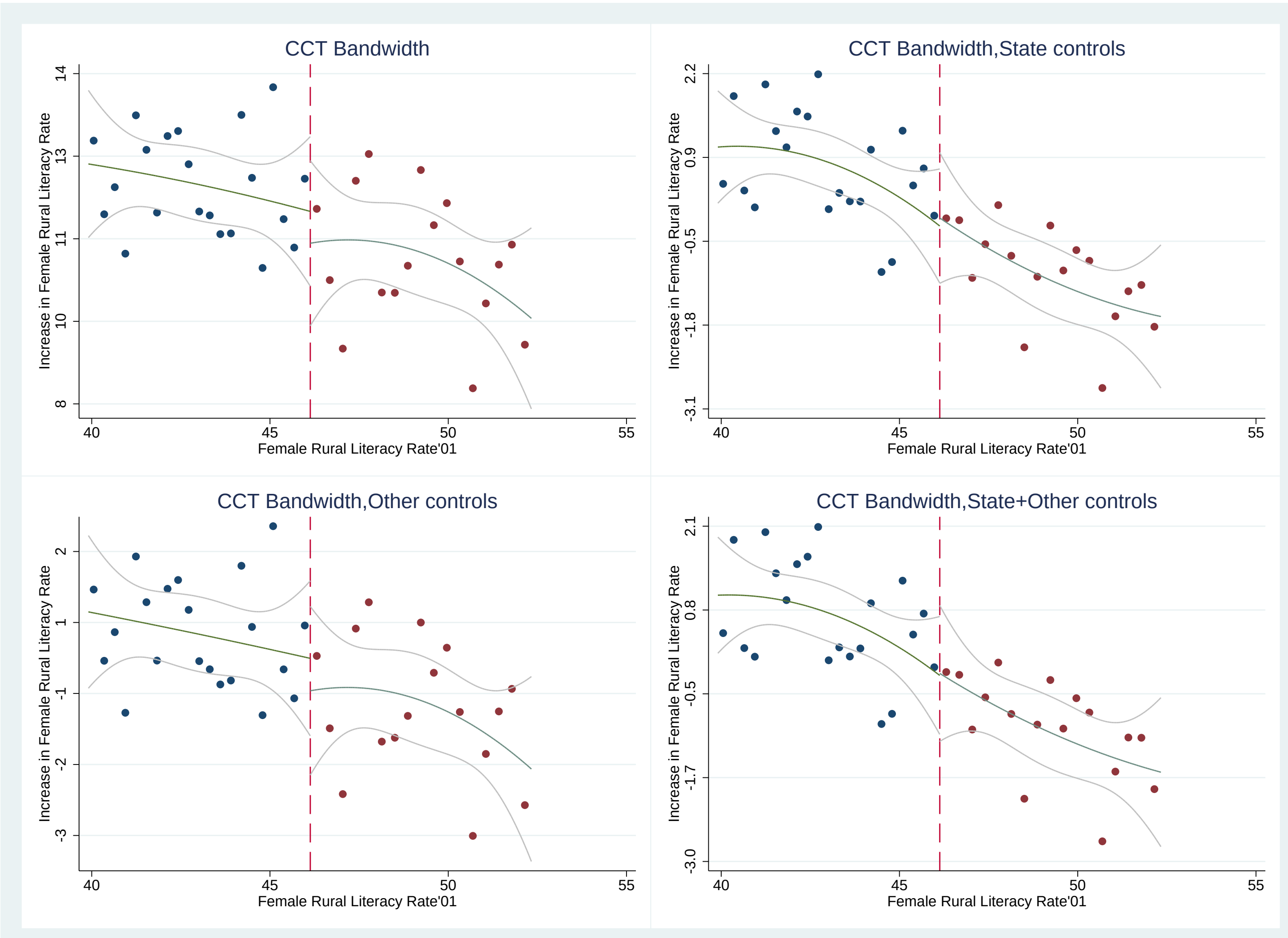
Below I present the main figures from the RD results:

Figure A.xii: Regression Discontinuity for first stage



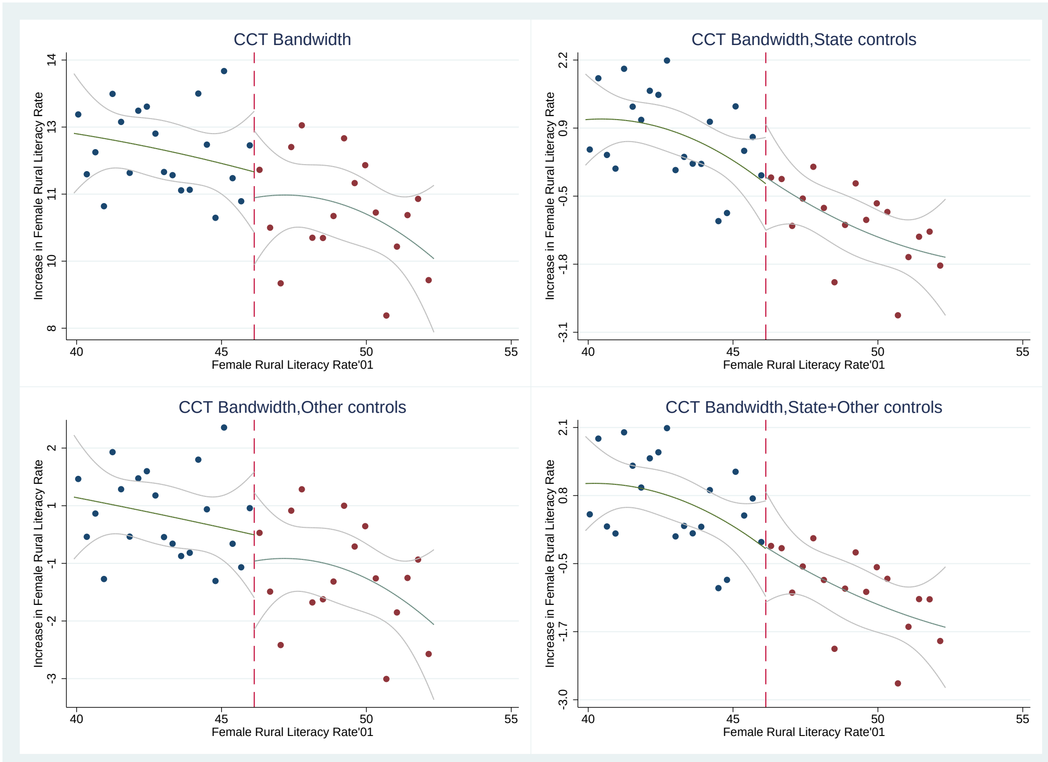
The figure shows that the probability of being classified as EBB increased by 70 percentage points at the cutoff.

Figure A.xiii: Regression Discontinuity for reduced form outcomes-Increase in Rural Female literacy rates



The figure plots the increase in rural female literacy rates over the decade for the optimal bandwidth calculated based on Calonico,et al (2014). The graphs show that there was no significant discontinuity at the cutoff with or without controls. The controls used are percentages of Scheduled caste and Scheduled tribe in the respective rural areas based on Census 2001 data.

Figure A.xiv: Regression Discontinuity for reduced form outcomes-Decrease in Gender gap in rural literacy rates



The figure plots the decrease in gender gap in rural literacy rates over the decade for the optimal bandwidth calculated based on Calonico,et al (2014). The graphs show that there was no significant discontinuity at the cutoff with or without controls. The controls used are percentages of Scheduled caste and Scheduled tribe in the respective rural areas based on Census 2001 data.