The Effectiveness of Decentralized Education Programs: A Multi-Dimensional Regression Discontinuity Approach

Gaurav Khanna *
University of Michigan

PRELIMINARY

July 2015

Abstract

This paper develops ideal estimators to causally estimate the effect of two education reforms in India that expanded the access to schooling. These policies differed in their levels of decentralization in decision making. The different administrative levels decided which inputs the funds were allocated for and where the school-building was targeted. Comparing the two policies, I find evidence that the decentralized policy increases literacy rates more than the relatively centralized one. While one of the policies requires a Regression Discontinuity (RD) design, the other uses a multi-dimensional RD (MRD) approach. Since the MRD is new to the literature, I perform a Monte Carlo exercise to determine the ideal estimators, that may be used by researchers in other contexts as well.

JEL: C1, O1, I20, I28, C5, O53,

Keywords: regression discontinuity, Monte-Carlo, education, India, decentralization

*Department of Economics, 611 Tappan Street, Ann Arbor, MI 48109. gkhanna@umich.edu. gauravkhanna.info. I would like to thank Ach Adhvaryu, John Bound, Matias Cattaneo, Jeff Smith, Dean Yang and seminar participants at the University of Michigan for valuable comments. All mistakes are my own.

1 Introduction

The effectiveness of large-scale government policies in health and education depend on the implementation quality of these policies. One crucial aspect in implementation is who the decision-makers are, and on where and how the funds are spent. In a federal system, this then boils down to how decentralized the decision making process is. Furthermore, to causally identify the effects of policies a good identification strategy is required. This paper will examine two large education policies in India, that differed on their levels of decentralized decision making, to study which policy was more effective. In doing so, a relatively new empirical strategy will be developed, to identify the causal impacts of these policies.

In the policies studied in this paper, there is a lot of leeway on how the money is spent, as districts or sub-districts can allocate funds to inputs that they deem necessary. The main difference between the two policies is the level of decentralization. This allows me to compare the effectiveness of the two policies, and determine whether the centralized policy had larger impacts than the decentralized education policy.

Which region received funds depended on an allocation rule that was different for each policy. While one of the allocation rules allows me to use a simple Regression Discontinuity (RD) design, the other requires a Multi-Dimensional Regression Discontinuity (MRD) approach. Since the MRD approach is relatively new to the literature, I conduct a Monte Carlo exercise to determine which estimators are ideal given different contexts. The results from the exercise can be used to inform researchers who face similar policy-designs in other contexts as well.

The first policy I study, is the District Primary Education Program (DPEP), that was implemented over a decade beginning in 1994. Under DPEP, districts were allocated funds to build new schools, hire teachers and upgrade the infrastructure of existing schools. Any district that had a female literacy rate below the national average (based on the 1991 Indian Census), was eligible to become a DPEP district. However, in reality, not all eligible districts received funds, and there were ineligible districts that did. Such a set-up allows for a fuzzy Regression Discontinuity (RD) design, where I compare districts just above and below the female literacy rate cutoff.

In a companion paper (Khanna (2015)), I study the effectiveness of the DPEP, focusing on the general equilibrium (GE) effects in the education and the labor market. I find that the DPEP changed the skill and wage distribution in districts that received it, and the labor market general equilibrium effects substantially altered the returns to education in these regions. I also show how the schooling interventions were not permanent, and as the funding is cut in the early 2000s, many of the effects dissipate over time.

When the DPEP was winding down, the Indian government replaced it with a new policy called the Sarva Shiksha Abhiyan (SSA). Under SSA, there were certain schooling schemes that were targeted to disadvantaged regions of the country called Educationally Backward Blocks (EBBs). A block is a sub-district, and decisions on how and where to spend the money was made at this sub-district level.

EBBs were identified based on two criteria. First, the female literacy rate had to be below the national average (based on the 2001 Census), and second, the gender gap had to be above the national average.[1] Since there are two running variables, I use a multi-dimensional Regression Discontinuity (MRD) approach to identify the impacts of this policy.

Since the MRD approach is new to the literature, I use a Monte Carlo setup to isolate the optimal estimation strategy. Recent work by Reardon and Robinson (2012) and Wong et al. (2012) discuss estimation methods for MRD setups. My Monte Carlo design differs from these papers in a few ways. In many practical contexts, like the ones studied in the paper, assignment can be 'fuzzy.' For example, not all eligible EBBs received funds, whereas some ineligible ones did. In such situations, when there isn't perfect compliance, some of the MRD candidate estimators will suffer from weak instrument problems, for example, if one of the cutoffs is ignored. Furthermore, in many contexts, we may expect the two running variables to be highly correlated. In such situations some estimators perform better than others. My Monte Carlo setup explores which estimators are biased and efficient under different sets of assumptions, allowing future researchers to use the right estimator for whichever context they wish to apply this to.

In this paper, I first identify the optimal estimation strategy for each policy, and then compare the impacts of the policies on literacy rates in India. The MRD estimation strategies I develop may be used by researchers in other contexts. I find that the more decentralized policy had a larger impact on increasing literacy, although the difference between the two effects are not statistically significantly different from each other. There is therefore, suggestive evidence that decentralized policies may be more effective.

1.1 Regression Discontinuities and Multiple Running Variables

Under the DPEP, districts that had a female literacy rate below the national average were eligible to receive funds. This set-up allows for a regression discontinuity (RD) design. The

[1] The gender gap is officially defined to be the difference between the male and female literacy rates.

RD design compares districts just above and below the literacy rate cutoff in order to determine the causal impact of the policy. If the only thing that changes discontinuously at the cutoff is the probability of receiving funds under DPEP, then this comparison estimates the local average treatment effect (LATE) of the policy near the cutoff. In order to estimate the parameters, I use optimal bandwidth selection procedures proposed by Calonico et al. (2014) and Imbens and Kalyanaraman (2012).

Policies under the SSA had a more complex assignment rule which requires a multi-dimensional RD (MRD) approach. Since the MRD approach is new to the literature, I use a Monte Carlo setup to determine the optimal estimation strategy. There are many possible candidate estimators for conclusively estimating the treatment effect, so the Monte Carlo exercise is necessary to figure out which is the best possible estimator. Wong et al. (2012) and Reardon and Robinson (2012) suggest some possible estimators to this effect. However, their set-up of their problem is rather from the one discussed here. First, they concentrate on the sharp RD case rather than the fuzzy. In many practical contexts, like the ones studied in the paper, assignment can be 'fuzzy.' There are at least two concerns that arise out of fuzzy designs. First, if the discontinuity is very fuzzy, then the instrument may not be strong. In such situations, when there isn't perfect compliance, some of the MRD candidate estimators will suffer from weak instrument problems. This is especially relevant in such a set-up, where there are potentially more than one discontinuity to exploit and strengthen the instrument. The second issue is that districts were selected on characteristics not observable to the researcher. Since manipulation and selection is a constant concern in such policies, especially in a developing country context, the fuzzy design needs to account for this possibility.

Furthermore, the correlation in the errors in the two different running variables is extremely high in the current context. Since the gender gap is a function of the female literacy rate, there will exist some amount of mechanical correlation. In other contexts where, for example text scores in two different subjects are the running variables, we may expect high correlation as well. The Monte Carlo exercise will therefore not only study how the extent of manipulation and selection affects the quality of the estimators, but also how the correlation in the running variables results in some estimators being better than others. Despite these changes, the preferred estimators are in many instances similar to the ones discussed by Wong et al. (2012) and Reardon and Robinson (2012).[2] However, these considerations are particularly relevant given that many governmental policies that are based on two cutoffs will suffer from the same issues highlighted here. Hence, a discussion of which is the ideal estimator, is of great policy relevance.

[2] I assume that the treatment effect is the same across each cutoff since there is no prior reason to believe otherwise in this context. For a different treatment effect at each cutoff, see Wong et al. (2012) and Reardon and Robinson (2012).

1.2 Decentralization of Policies

In the policies studied in this paper districts or sub-districts can allocate funds to regions and inputs that they deem necessary. One of the primary differences between the two policies is the level of decentralization. This allows me to explore the effectiveness of a centralized versus a decentralized education policy. There exists a theoretical literature that discusses the costs and benefits of decentralized policies (Bardhan (2002); Bardhan and Mookherjee (2000, 2005); Niehaus et al. (2013)). Better information allows for better targeting, but decentralized policies may have other drawbacks as well (Coady et al. (2004)).

For instance, decentralized policies may have more corruption at each stage of the disbursement process (Ferraz and Finan (2008); Olken (2007); Reinikka and Svensson (2004)). This freedom to allocate funds may lead to leakages, and spending may not be very effective. Furthermore, a capture of the entire decision making process by the local elite can misallocate the funds to a certain section of the populace (Bardhan and Mookherjee (2005)). Despite having good information on who needs the funds most, local elites may divert funds to friends and family (Basurto et al. (2015)). On the other hand, decentralized decisions can utilize more localized knowledge and be better targeted (Olken (2010)). This allows authorities to target areas that need theses funds the most. In general, with any policy of this kind where funds are allocated to lower administrative levels, there are concerns that include the possible wastage of funds (Bandiera et al. (2009); Easterly (2006, 2008)). It is, therefore, an important empirical question as to whether decentralized spending is more effective than centralized spending.

1.3 Schooling Interventions in India

There is a large literature in the US context that explores whether spending on education actually affects educational outcomes [3] While some find positive effects, others find no effects. One crucial take-away from the US literature is that not all schooling inputs have similar impacts, and so it is necessary to understand which inputs matter.[4]

We may, however, expect different results in developing countries, which have more leakage of funds, more informational constraints, lower average levels of education, more spending constraints and other institutional drawbacks. For example, one often cited reason for low educational outcomes are the high levels of teacher absence (Das et al. (2013b); Duflo et al.

[3]See, for example, Card and Payne (2002); Hanushek (1997, 2003, 2006); Hoxby (2001); Krueger (2003).
[4]See Grogger (1996); Hanushek (1986, 2008); Krueger (1999); Loeb and Bound (1996).

(2012)).[5] There would arguably be a larger return to relaxing these constraints via increased spending on education-related programs. Some government programs like the mid-day meal scheme in India, do have a positive impact on caloric intake of students (Afridi (2010)), but no impact on test-scores (Jayaraman et al. (2010)). Chin (2005) shows how 'Operation Blackboard' redistributed teachers from larger to smaller schools, which raised primary school enrollment for girls and the poor.

In the Indian context, micro-interventions that change inputs within schools seem to produce mixed results. For instance, Borkum et al. (2010) study the impact of a school library program in one state, and find no impact on student reading scores. Similarly some studies don't find any positive relationship between teacher training and improved test scores (Kingdon and Teal (2010)) or between teacher salaries and student outcomes (Muralidharan and Sundararaman (2010)). Remedial education does seem to have a positive impact, but a reduction in class sizes does not (Banerjee et al. (2007)). However, other studies do find a positive, yet small, impact of smaller class sizes (Jacob et al. (2008); Muralidharan (2013)). Banerjee et al. (2010) try several interventions with community participation, and find that the only one that effectively improved test scores was a remedial instruction program. And Banerjee et al. (2007) find that computer-aided remedial programs had larger impacts on test-scores than remedial teachers. Linden (2008) further evaluates computer-aided instruction and finds positive impacts if these were implemented after school but negative impacts on if implemented in school hours, where they end up displacing traditional modes of instruction. There is therefore a substantial literature on supply-side interventions in developing countries that focuses on micro-interventions.

Unfortunately, micro-interventions are inherently different from large-scale policies. Micro-interventions are usually implemented by researchers, not governments, and governments have to be concerned with questions of who controls the funds. Duflo (2001) uses school construction in Indonesia and shows increases in education and earnings for cohorts that would have benefited most from these policies. Similar programs can be found across the developing world and can be thought of as reducing the marginal cost of schooling by improving the access to schools (Behrman et al. (1996); Birdsall (1985)).[6] Similarly, the policies that I study should be thought of as mostly affecting the extensive margin, since the Indian government built newer schools in most regions.

In the policy studied in this paper, there is a lot of leeway on how the money is spent,

[5] Possible solutions to the suggested problem include more effective monitoring of absenteeism (Muralidharan (2013)), or performance-linked-pay (Das et al. (2013b); Duflo et al. (2012)).

[6] Some examples are found in Zimbabwe (Aguero and Bharadwaj (2014)), Nigeria (Osili and Long (2008)), Sierra Leone (Cannonier and Mocan (2012)), Uganda (Deininger (2003)), Zambia (Ashraf et al. (2015)), Kenya (Bold et al. (2013)), Tanzania (Sifuna (2007)).

as districts or sub-districts can allocate funds to inputs that they deem necessary. This may lead the funds to be targeted to areas that need them most. On the contrary, this freedom to allocate funds may lead to corruption or leakages, and spending may not be very effective. In an accompanying paper (Khanna (2015)), I extensively study one of the policies to disentangle the general equilibrium effects in the labor and education market, to derive the welfare impacts of the policy.

The persistence of impacts depends on how long these interventions last. Interventions that lower the costs of schooling seem to have positive impacts, at least in the very short run. In the longer-run, however, the evidence is mixed. While Das et al. (2013a) find substantial test scores improvements in the first year of a school grant program, but no impacts in the following years, the Angrist et al. (2006) study of schooling vouchers in Colombia shows impacts on school completion rates and test scores even in the long run. The Das et al. (2013a) results suggest that households reduce their own spending by the amount of the grant. One interpretation of these results is that the demand for education is highly inelastic in the longer run, and supply-side policies may not be as effective as hoped. Another, is that interventions need a constant source of funds to maintain persistent impacts. I assemble a panel dataset of schools that allows me to study the long-run impacts of these interventions as well. I find that once the funding is reduced, teachers leave, but physical infrastructure upgrades last for longer. Over time, many impacts dissipate as the funding runs out. The two policies I study were implemented one after the other - as the effects of the first policy wear out, spending on the newer one is ratcheted up.

2 The Two Programs

I find exogenous variation in two large policies implemented by the Indian government. The government selected areas based on certain criteria that allow me to use different RD approaches. This was also a period of rapid growth and development in the Indian economy, which is why it is necessary to isolate the impact of the policy from other changes. The schemes spanned a decade each and were a decade apart. Subtle differences across the schemes will also be interesting variation that I will study.

The two policies differed on the administrative levels at which the funds were received and spent. There are various administrative levels in India. At the top there is the Central government that can preside over national policy decisions. Next, there are 36 states and union territories. States have their own elected state governments and non-elected bureaucracy, whereas union territories are administered by the Central government. These states can be divided into districts. The exact number of districts change over time as some districts are split into others or merged with others - the 2011 Census lists 640 districts. District officials are non-elected and district boundaries do not necessarily correspond with constituency boundaries for elected officials. These districts can be further divided into sub-districts, which also regularly split or merge over time. In 2011, there were 5564 sub-districts, each with appointed officials for administrative purposes.[7] While the DPEP was administered at the district level, the policies studied here under SSA were at the sub-district level.

2.1 District Primary Education Project (DPEP)

In 1992, the Indian Parliament updated their National Policy on Education with a renewed focus on primary and upper primary education. The Parliament amended the constitution and transferred education related decisions to local bodies, based on recommendations from the Central Advisory Board of Education. They stressed the decentralization of decision making by helping districts plan and manage both primary and upper-primary education.[8]

The District Primary Education Project (DPEP) began in 1994 in seven states and 42 districts, and was over time expanded to 271 of the 600 districts in the country. The program grew to consist of seven projects, with funding from the World Bank, the European Commission (EC), the U.K. Department for International Development (DFID) and Official Development Assistance (ODA), the Royal Government of the Netherlands, and UNICEF, making it one of the largest donor assisted programs in the world (Jalan and Glinskaya (2013)). The Central government channeled these funds to the states as their 85% contribution, whereas the states funded the remaining 15%.[9]

The project spanned four phases, the last of which were implemented in the mid-2000s. While a portion of the funds were released under DPEP through the mid-2000s, the bulk of the funding ended in 2005 when other policies under the Sarva Shiksha Abhiyan (SSA)

[7] Sub-districts can be split into villages which will have elected Gram Panchayats making decisions at the local level.

[8] Primary is usually grades 1 through 4 or 5, and upper primary is grades 5 or 6 through 8.

[9] India has received aid on various social and infrastructure programs, and in 2005-6 alone it received $4 bn. (Colclough and De (2010)). By 2002 the World Bank alone had committed about $1.62 billion on DPEP, whereas the other donors concentrated on certain states. For example, in the first few years of the program, the EC spent ECU 150m in Madhya Pradesh, the Netherlands spent $25.8m in Gujarat, DFID spend 80m pounds in Andhra Pradesh and West Bengal, whereas UNICEF spent $ 153m in Bihar (GOI (2000)). WorldBank (1997) claims that in 1993, the EC provided a grant of ECU 150 mn, whereas the World Bank approved credits of $265 mn in 1994 and $425mn in 1996. At the time of the transfer to the wider SSA program in 2004, the World Bank's contribution consisted of less than half of the external aid funds, with DFID and the EC being the other major donors. Between 2004 and 2007 alone, about $7.8 bn was spent on the expanded SSA program, including the Government's contributions (Ayyar (2008)).

were growing in strength. In 2006, only 2 states received any money under DPEP, and 2007 onwards none did. When the shift to the newer SSA program happened, the government levied a 2% education cess to fund the expansion to all districts.[10]

The broader program claims to have covered about 271 low literacy districts, and served approximately 51.3 million children and 1.1 million teachers in about 375,000 schools (Jalan and Glinskaya (2013)). It created 63,000 new schools, including more than 50,000 'alternative' or 'community schools',[11] and trained about 1 million teachers and 3 million community members. The first two phases of the program covered 14 large states in the country, and over time it was extended to other states. States had to maintain the level of expenditure that existed before the program was implemented in an attempt to ensure that there was no crowd-out of state funds.[12] However, states did have the ability to re-allocate funds across districts. Within the state, there was major inter-district variation in planning and management as the districts had the flexibility to allocate funds. In the project states, it increased the average allocation of funds for primary school education by between 17-20 % (Jalan and Glinskaya (2013)).[13]

The objectives of the project were two-fold. The first was to improve student access to and retention in primary education by building schools, supporting school and community organizations, constructing new classrooms and improving existing school facilities. The guidelines of the program stipulated that the "*project would be a reconstruction of primary education as a whole in selected districts instead of piecemeal implementation of schemes*" (GOI (1994)). While most funds were directed towards the government schools, there was also a training-drive for teachers of private and government-aided schools.[14]

The second was to improve the access to primary and upper-primary education by establishing district institutions to decentralize planning. Specifically, this was to be done by

[10] The phase-out was fairly rapid. In the 2002-3 financial year, the government spent approximately $ 345 mn on DPEP, whereas in the 2006-7 financial year, they spent only $24 mn on it.

[11] Alternative or community schools are part of the non-formal schooling system. They provide the most basic schooling infrastructure to remote areas and disadvantaged groups with the help of local community participation.

[12] Varghese (1994) claims that states had to maintain their expenditures on education at at least their 1992 values, whereas the WorldBank (1997) guidelines claim the states had to maintain the same growth rate in their education budgets.

[13] In this period, DPEP was the flagship education, despite being restricted to less than half the country. For example, in 2001 alone, the Ministry of Human Resource Development, estimates spending approximately $275 mn on DPEP for the limited number of districts. The second and third largest expenditures were on schemes that covered all districts like the Mid-day Meal Scheme ($232 mn), and Operation Blackboard ($130mn).

[14] The guidelines of the policy also discussed the local community initiatives in promoting enrollment and retention. For example, Village Education Committees and local bodies like the Mother-Teacher Associations were tasked with creating local awareness campaigns and getting more children into schools and preventing them from dropping out of schools.

managing the delivery of education, including teacher support and materials development through Block Resource Centers (BRC) and Cluster Resource Centers (CRC), and strengthening the District Institutes of Education and Training (DIET). This also included targeted interventions for girls, minority groups and children with mild to moderate disabilities, and the expansion of Early Childhood Education (ECE). The program established a DPEP Bureau in the Ministry of Human Resource Development that served as a financial and technical intermediary. They appraised, monitored and supervised the district programs. The programs were developed by each participating district and appraised by the Bureau that also provided implementation support. The programs were evaluated and the poorly performing subprojects are dropped.

Amongst the many World Bank and Government of India briefs,[15] and media reports[16] that refer to the program's success, is a working paper by Jalan and Glinskaya (2013) that uses a difference-in-differences methodology to look at the enrollment rates for students in the 42 Phase 1 districts where the project was implemented. They find that five years after the program started, enrollment and grade progression of minority groups in some specific states improve. Furthermore, grade progression for boys in certain states is higher, but there are little to no impacts on girls.[17] Over the entire period districts were not allowed to receive more than $8 million, which comes to approximately $9.1 per student, lowering private costs by between 20 to 40% (Jalan and Glinskaya (2013)). Their paper does not exploit the discontinuity in treatment, nor does it look at schooling inputs, and instead uses two repeated cross sections of enrollment to look at the short-run impacts of the program. In contrast, my RD approach looks at the longer run effects, fifteen years after the program started. Other descriptive studies examine the outcomes for DPEP districts without comparing them to other districts (Aggarwal (2000); Menon (2001); Pandey (2000)), and hence cannot distinguish between the changes in overall education taking place all across the country driven by a robust economic growth, and the changes specifically attributable to the program.

2.2 Policies under Sarva Shiksha Abhiyan (SSA)

Sarva Shiksha Abhiyan (The Education for All Movement) was launched in 2001, but many of its policies did not take effect till 2004. The goal was to build and open new schools, and upgrade the existing schooling infrastructure, hire and train new teachers, and provide "quality" education. A strong focus on girls' education was implemented under two separate sub-schemes - the National Program for Education of Girls at the Elementary Level

[15] See World Bank Report (2003) and Government of India (2011)

[16] See, for example, "World Bank praises India for DPEP" Economic Times, (Sep 2005)

[17] My results also show larger impacts on boys than girls.

(NPEGEL), and the Kasturba Gandhi Balika Vidyalaya (KGBV) available in certain regions called "Educationally Backward Blocks (EBBs)",[18] but in 2011 it was expanded to the entire country (under the "Right to Education Act"). The funds were largely from the Central Government and the State Government, but also from the World Bank, DFID and UNICEF. The NPEGEL program focused on elementary school education for girls, whereas the KGBV program built residential schooling facilities for girls from minority groups, lower castes and below-poverty-line families.

Figure 1 shows foreign aid earmarked for primary and upper primary education only, and the amount spent on DPEP. Both amounts rise steadily in the 1990s till the peak in 2002-3. When foreign aid is cut in 2003, so is DPEP expenditure, which falls steadily over the next few years. Foreign aid spikes up again a few years later as the next policy - the SSA - is ramped up. Figure 2 shows how much the Central government transferred to the states for social sector spending (health and education). This amount, and the share of total transfers, rises steadily during this period. The figures make it clear that this was a period of a large increase in externally-financed expenditure on education, most of which was concentrated in less than half the districts of the country, allowing for a valuable policy experiment.

EBBs were identified by the following methodology outlined by the Planning Commission and the Ministry of Human Resource Development: *"Initially a list of 3073 educational backward blocks (EBBs) was drawn up in connection with the Sarva Shiksha Abhiyan. This was arrived at on the basis of twin criteria of the female literacy rate (FLR) being below the national average of 53.7% and the gender gap in literacy being above the national average of 21.59%. Both these criteria had been earmarked by the Registrar General of India (RGI). Subsequently this list was expanded to include 406 more blocks, out of which 404 blocks were having a rural FLR of less than 45% irrespective of the gender gap."* The numbers in this paragraph were derived using the publicly available 2001 Census data.

Unlike the DPEP, the NPEGL-KGBV was far more decentralized. While the decision making and resource allocation under the DPEP happened at the state and district levels, the decisions under the NPEGL-KGBV were made by BRCs (Block Resource Centers) and CRCs (Cluster Resource Centers). Decentralization of the policy may have the added benefit of better local management and planning for regional constraints and demands. This extent of decentralization was the main difference between the two policies.

To causally estimate the parameters of the model, I use a Regression Discontinuity (RD) design. Since the MRD approach is new to the literature, I provide Monte-Carlo evidence

[18] A block (also known as a sub-district) is the third level of administrative disaggregation after states and districts.

to identify the optimal estimation strategy in this context. My Monte Carlo results may be useful for other researchers in selecting the appropriate estimators given the correlations in errors in other contexts.

3 The Multi-Dimensional Regression Discontinuity

The NPEGEL-KBGV program was instituted at the sub-district level, where two criteria were required for the selection of these sub-districts - first, the female literacy rate in the sub-districts had to be below the national average, and second the gender gap (the difference between the male and female literacy rates) had to be above the national average. Since many of the policies under this program were targeted to girls, the difference between male and female literacy was used to identify which regions had a higher level of gender inequality. In order to credibly estimate the causal impacts of this policy using the institutional set-up, this paper outlines an estimation strategy akin to the regression-discontinuity approach. While the discussion below will be tailored towards the set-up at hand, the broader lessons about the different methods of estimation can be used in many other contexts that have a fuzzy-multi dimensional RD (MRD) set-up. This is one of the first applications of a fuzzy MRD setup, which also addresses the high levels of correlation between the two running variables.

3.1 Parameters to be Estimated

Since the sub-districts were chosen based on the twin criteria of female literacy and the gender gap, I use a multi-dimensional regression discontinuity (MRD) design, where female literacy F and gender gap G are the running variables. Unlike the set-up studied here, Papay et al. (2014) discuss a context where each cutoff generates a different treatment. The most relevant papers, therefore, for this context are by Reardon and Robinson (2012) and Wong et al. (2012). Both are theoretical papers and discuss ideal estimation strategies for such a design, but do not have any real fuzzy applications studied as examples. This paper, therefore is one of the first applications of a $fuzzy$ multi-dimensional RD design. Since the theoretical literature is also new, this paper will support the estimation strategies with Monte Carlo exercises.

To understand the parameters of interest, let us first define the neighborhood of the cutoff. Remembering that districts are eligible for 'treatment' when female literacy rates are below the national average ($F \leq F_c$), and gender gaps are above the national average $G \geq G_c$, we can define the 'treated' region to be: $C^+ = \{(F,G)|(F \leq F_c \& G = G_c) \cup (F = F_c \& G \geq G_c)\}$. And the 'control' region can similarly be defined as $C^- = \{(F,G)|(F \geq F_c \& G = G_c) \cup (F = F_c \& G \leq G_c)\}$

Similar to the uni-dimensional set-up, I define ϵ-neighborhoods around these regions: $N_\epsilon \equiv \{(F,G) \text{ s.t. } |(F-F_c)(G-G_c)| < \epsilon^2\}$. And the ϵ-neighborhoods in the treated and control groups as $N_\epsilon^+ \equiv N_\epsilon \cap C^+$, and $N_\epsilon^- \equiv N_\epsilon \cap C^-$.

Given this notation, let us define τ_{FG} as the average treatment effect at the cutoff for the two running variables F and G to be:[19]

$$
\begin{aligned}
\tau_{FG} &= \lim_{\epsilon \to 0} E[Y|(F_i,G_i) \,\epsilon\, N_\epsilon^+] - \lim_{\epsilon \to 0} E[Y|(F_i,G_i) \,\epsilon\, N_\epsilon^-] \\
&= E[Y_i(1) - Y_i(0)|(F_i,G_i) \,\epsilon\, N_\epsilon] \qquad (6)
\end{aligned}
$$

When treatment is not sharp, let us define W to be the probability of treatment. If there is a discontinuity in this probability, then we have a relevant first stage:

$$
W_{FG} = \lim_{\epsilon \to 0} E[W|(F_i,G_i) \,\epsilon\, N_\epsilon^+] - \lim_{\epsilon \to 0} E[W|(F_i,G_i) \,\epsilon\, N_\epsilon^-] > 0 \qquad (7)
$$

The fuzzy RD estimator in this case would be:

[19] Using Wong et al. (2012) we can decompose this parameter into the weighted average of the average treatment effect at each frontier estimated separately:

$$
\tau_{FG} = \omega_F \tau_F + \omega_G \tau_G \qquad (1)
$$

where the weights are increasing functions of the probabilities of being near the F or G cutoff:

$$
w_F = \frac{\int_{G \geq G_c} f(F = F_c, G)\, dG}{\int_{G \geq G_c} f(F = F_c, G)\, dG \,+\, \int_{F \geq F_c} f(G = G_c, F)\, dF} \qquad (2)
$$

$$
w_G = \frac{\int_{F \geq F_c} f(G = G_c, F)\, dF}{\int_{F \geq G_c} f(F = F_c, G)\, dG \,+\, \int_{F \geq F_c} f(G = G_c, F)\, dF} \qquad (3)
$$

And the treatment effects at each frontier are given by:

$$
\tau_F = E[Y_i(1) - Y_i(0)|F_i \epsilon C^+] = \frac{\int_{G \geq G_c} (y_1(F,G) - y_0(F,G)) f(F = F_c, G)\, dG}{\int_{G \geq G_c} f(F = F_c, G)\, dG} \qquad (4)
$$

$$
\tau_G = E[Y_i(1) - Y_i(0)|G_i \epsilon C^+] = \frac{\int_{F \geq F_c} (y_1(F,G) - y_0(F,G)) f(G = G_c, F)\, dF}{\int_{F \geq F_c} f(G = G_c, F)\, dF} \qquad (5)
$$

This result of the estimators being a weighted average has been highlighted by Wong et al. (2012).

$$\tau_{FG}^{fuzzy} = \lim_{\epsilon \to 0} \frac{E[Y|(F_i, G_i)\ \epsilon\ N_\epsilon^+] - E[Y|(F_i, G_i)\ \epsilon\ N_\epsilon^-]}{E[W|(F_i, G_i)\ \epsilon\ N_\epsilon^+] - E[W|(F_i, G_i)\ \epsilon\ N_\epsilon^-]}$$
$$= E[Y_i(1) - Y_i(0)|(F_i, G_i)\epsilon N_\epsilon\ ;\ W^+ > W^-] \quad (8)$$

In order to estimate τ_{FG} there a few standard assumptions required. First, there must be a relevant first-stage - there exists a discontinuity in treatment probabilities at the cutoff, such that both a treatment and control group exist in C. Second, the monotonicity assumption should be satisfied, and there are no defiers close to the cutoff. Last, there must be continuity, both in compliance, and in the potential outcomes.

3.2 Monte Carlo Exercises

Since there are multiple possible candidate estimators for conclusively estimating the treatment effect, a Monte Carlo exercise is necessary to figure out which is the best possible estimator. The Monte Carlo exercise will therefore not only study how the extent of manipulation and selection affects the quality of the estimators, but also how the correlation in the running variables results in some estimators being better than others.[20] However, these considerations are particularly relevant given that many governmental policies that are based on two cutoffs will suffer from the same issues highlighted here. Hence, a discussion of which is the ideal estimator, is of great policy relevance.

3.2.1 Monte Carlo Framework

In order to describe the Monte-Carlo exercise I set up a framework to determine which parameters affect which estimators. Female literacy F and male literacy M can be characterized by the following equations:

$$F = \gamma_F + \epsilon_F \quad (9)$$

$$M = \gamma_M + \epsilon_M\ , \quad (10)$$

[20] I assume that the treatment effect is the same across each cutoff since there is no prior reason to believe otherwise in this context. For a different treatment effect at each cutoff, see Wong et al. (2012) and Reardon and Robinson (2012).

where, γ_k is the mean value of literacy, and ϵ_k is an error term. The gender gap is just the difference between male and female literacy:

$$G = M - F = \gamma_M - \gamma_F + \epsilon_M - \epsilon_F \equiv \gamma_G + \epsilon_G \tag{11}$$

One way to understand how manipulation and selection of districts may work given the policy rules is to define the two following 'fuzzy' running variables:

$$F_z = F + \epsilon_{F_z} \tag{12}$$

$$M_z = M + \epsilon_{M_z} \tag{13}$$

An alternative is to use the traditional approach in univariate fuzzy designs and assign the 'error' to the treatment status rather than the running variable. However, such an approach would not let us distinguish which of the errors along which of the running variables matter for which estimators.

Given these running variables, the actual treatment status is defined as follows:

$$D = \begin{cases} 1, & \text{if } F_z < F_c \text{ and } G_z > G_c \\ 0, & \text{otherwise} \end{cases} \tag{14}$$

Finally, the outcome of interest depends on the treatment effect τ:

$$Y = \gamma_Y + \tau D + \epsilon_Y \tag{15}$$

Given this set-up, let us define the error structure amongst the different variables:

$$\epsilon \sim N(\mathbf{0}, \mathbf{\Omega}) \tag{16}$$

where the elements of the variance-covariance matrix $\mathbf{\Omega}$ capture the correlations between the errors in the above equations. These correlations capture various aspects of the identification problem. While σ_{kk} are all normalized to 1 for $k = \{F, M, F_z, G_z, Y\}$, the correlations $\sigma_{k,\,-k}$

capture how the errors depend on each other.

The two running variables F and G have a mechanical correlation of the following form $\sigma_{FG} = \sigma_{MF} - \sigma_{FF}$. It is reasonable to expect male and female literacy within a district are correlated $\sigma_{MF} > 0$. Therefore, the amount and direction of correlation between the two running variables depends on the value of σ_{MF} relative to the variance of the error term in F (which is normalized to 1). The 'fuzzy' version of the running variable should have a high level of correlation with the normal running variable, else we would have a weak-instrument like problem in the analysis. A correlation between the running variables and the outcome of interest σ_{YF} and σ_{YG} is what biases the OLS estimates of the treatment effects. Furthermore, a correlation between the 'fuzzy' running variables and the outcome of interest will arise from selection and manipulation of certain districts into getting the treatment when they weren't supposed to, and vice versa, will be captured by correlations between ϵ_Y and ϵ_{F_z} *or* ϵ_{G_z}. These are represented by σ_{YF_z} and σ_{YG_z}. Covariates may not be able to control for this manipulation as policy makers will observe things about the district that researchers do not.

3.2.2 Candidate Estimators

A few candidate estimators are considered here so as to identify the ideal estimators given the setup. Six approaches are investigated under various assumptions to find the ideal estimator. Two of the basic estimators are the OLS and the 'control-function approach,' which flexibly controls for the running variable.

The other approaches are all instrumental variable (IV) designs. First, is the dual-index frontier approach, also known as 'response surface approach.' This approach is analogous to the control-function-approach in the single-variate RD case. Instead of controlling for a flexible polynomial of one running variable, this framework will parametrically control for flexible polynomials of both running variables. Second, is the uni-variate centering or 'binding-score' approach which collapses the dual-index problem into a single running variable by picking the normalized distance of the running variable to the closest cutoff. Third, is the 'added fuzziness' approach which disregards the existence of one of the running variables, and calculates the discontinuity at the other running variable's cutoff using the usual single-dimensional RD methods. This can be done separately for each running variable. The last candidate is the 'conditional sub-sample' approach where the sample is restricted to be above one cutoff, and the discontinuity is calculated at the other cutoff.

The OLS approaches merely regress the outcome Y on actual treatment D, whereas the control-function approach extends this to flexibly control for the running variables in the following manner:

$$Y_b = \alpha + \tau \hat{D}_b + h(F_b, G_b) + \epsilon_b \tag{17}$$

In order to operationalize the frontier approach it is necessary to define what the predicted treatment variable is in the first place. Analogous to the single-dimensional RD, the predicted treatment variable can be defined as

$$D^{pred} = \begin{cases} 1, & \text{if } F < F_c \text{ and } G > G_c \text{ ;} \\ 0, & \text{otherwise,} \end{cases} \tag{18}$$

The 2SLS (two-staged least squares) estimation would have the following functional form:

$$Y_b = \alpha_0 + \tau \hat{D}_b + h(F_b, G_b) + \epsilon_{ob} \tag{19}$$

$$D_b = \alpha_1 + \delta_1 D_b^{pred} + h(F_b, G_b) + \epsilon_{1b} \text{ ,} \tag{20}$$

where $h(F_b, G_b)$ is a flexible polynomial in the two running variables, estimated at the block b level.

The uni-variate centering or binding-score approach, on the other hand, would require a different definition of the running variable. The new running variable for the single-index problem SI_b and is normalized given the distribution of F and G so as to make them in comparable units. This procedure selects the running variable differently for each district depending on which cutoff it's closest to:[21]

$$SI = \min\{\frac{F - F_c}{\sigma_F}, \frac{G - G_c}{\sigma_G}\} \tag{21}$$

The big advantage of the single-index approach is that optimal bandwidth selection for a single-running variable is already known in the RD literature (Calonico et al. (2014); Imbens and Kalyanaraman (2012)).

[21] Normalizing the running variable is not required when they are in the same units, as in the case of geographic RDs on latitude and longitude.

The parametric form of the 2SLS estimation would then have the following functional form:

$$Y_b = \alpha_0 + \tau \hat{D}_b + h(SI_b) + \epsilon_{ob} \tag{22}$$

$$D_b = \alpha_2 + \delta_2 D_b^{pred} + h(SI_b) + \epsilon_{2b} \tag{23}$$

The third approach is the added fuzziness design which ignores one of the running variables altogether and simply estimates a fuzzy discontinuity at the other variable's cutoff. For example, one can define the new predicted treatment probability to be:

$$D_F^{pred} = \begin{cases} 1, & \text{if } F < F_c \text{ ;} \\ 0, & \text{otherwise,} \end{cases} \tag{24}$$

This ignores the gender gap altogether. The parametric 2SLS regression form of this therefore only controls for the flexible polynomial in female-literacy:

$$Y_b = \alpha_0 + \tau \hat{D}_b + h(F) + \epsilon_{ob} \tag{25}$$

$$D_b = \alpha_3 + \delta_3 D_{Fb}^{pred} + h(F) + \epsilon_{3b} \tag{26}$$

The last candidate tested in the Monte-Carlo setup is the 'conditional subsample' approach. To study the discontinuity at the female-literacy F cutoff, for example, one must restrict the sample to being above the gender-gap cutoff, and then re-do the analysis in the 'added fuzziness' approach.

3.2.3 Results of the Monte Carlo Exercise

In Table 1 I summarize the results of this exercise. Which estimators have lower biases or higher precision than others depends on the variance-covariance matrix of the errors, and the number of observations in the regressions. Since any 2SLS estimator is biased but consistent, they may do worse than the OLS estimator in small samples. The other consideration is that the correlation between the errors may bias the various estimators.

In Figures 15 and 16 the results are presented for varying correlations in the errors with 1000 simulations of 5000 observations with a treatment effect of 10. One can see that the OLS estimates perform poorly across the board. The coefficients are driven in a large part by the correlation in the errors, especially in the correlation between the running variables and the outcome.

The non-IV 'control-function' approach performs fairly well even when there is correlation between the running variables and the outcome variables (Figures 16a and 15d). However, it does poorly the moment the actual selection of districts is systematic and probably manipulated by policy makers. If there is correlation between the fuzzy-version of the running-variable that the policy makers use for selection of districts, and any other variable, the estimates are severely biased.

The conditional-subsample approach does poorly if there is high correlation between the running variable and the outcome of interest. Since the conditional-subsample restricts the sample to the portion where one of the running variables is above a certain cutoff, a high correlation between the running variable and outcome variable leads to a problem akin to the 'selection on the dependent variable' issue. For example, in Figure 16a one can see that conditioning on being above the gender-gap cutoff leads to a bias in the estimator as the correlation between male literacy (gender gap) and the outcome is varied. Whereas, this is not an issue for this same estimator when the correlation between female literacy and the outcome is varied (Figure 15d). This is because the selection of the sub-sample is based on the gender-gap and not the female literacy running variable. Furthermore, not controlling for the other running variable can lead to an omitted variables bias when the errors across the running variables are highly correlated (Figure 15a).

The other issue with the 'conditional subsample' approach is the possibility of a weak-instruments problem, which would then bias the IV estimate towards the OLS estimate. In Figure 17 I show the F-statistics for the excluded instruments in the first-stage. The bias can be seen in Figure 16d where a change in the correlation between the running variable and fuzzy version of the running variable increases the amount of noise and weakens the instrument (Figure 17). This bias would not have arisen if the gender-gap instrument was also being used. In the other panels, however, it can be seen that the conditional-subsample approach does not perform too badly.

In the added-fuzziness approach, the gender gap variable was ignored completely, and female literacy was treated as the only running-variable. Ignoring one running variable may lead to a weak-instruments problem if the discontinuity at the cutoff along the other running-variable is not strong. The first-stage F-statistic for the variables of interest are shown in Figure 17, and the added-fuzziness estimator has a particularly weak first-stage when the correlation in the errors between female-literacy and the fuzzy-version of the variable is low. For example, Figure 15c shows how this approach may perform poorly when the running variable of interest (in this case female-literacy) is driven by the manipulation in actual treatment assignment (i.e. the correlation between the actual running variable, and the running variable used by policymakers is low). In the absence of correlation with errors of the fuzzy-running variables - the absence of systematic manipulation in the selection of districts - the added-fuzziness performs fairly well.

From the Figures it's clear that while the 2SLS frontier approach does well across any kind of error-covariance matrix, all the 2SLS estimators are biased but consistent. In small samples, therefore, the OLS and the non-IV control-function approach may perform better. In Figure 18 the sensitivity to the number of observations is tested. In Figure 18b, for example, it is clear that in very small samples, the OLS and non-IV control-function approach actually do better than the 2SLS flexible frontier or the added-fuzziness approach. In larger samples (of about 300 observations or more in this setup), the IV approaches start doing better as their estimators are consistent, whereas the non-IV estimators are inconsistent given the selection of districts on characteristics that are correlated with the outcome variable. In Figure 18a it can also be seen that the conditional-subsample approach performs particularly badly in smaller samples (of about less than 1000 observations in this example). This is because conditioning on being above a certain cutoff disregards a major chunk of the sample and exacerbates the small-sample biases.

In terms of bias, the response-surface approach, and the normalized single-index approach performs well across the board. The additional benefit of the single-index approach, not captured by this exercise, is the fact that the optimal bandwidth calculation for a single variable has already been solved in the RD literature. Whereas, the benefit of the response-surface approach is that the parametric assumptions help reduce the need for computationally intensive optimal bandwidth calculations.

When it comes to precision, the 2SLS frontier approach once again performs well. In Figure 18d it can be seen that the 95% confidence intervals around the frontier-approach estimator are wide for smaller samples, but still a lot tighter than the intervals around the added-fuzzy estimator seen in Figure 18c. Even comparing the confidence intervals between the single-index approach and the response surface approach in Figures 16d and 16c one can see that the level of precision is higher in the response surface approach. This has to do with the fact that the added-fuzzy estimator is noisier given that it disregards a major chunk of the data when trying to explain the changes in the outcome variable. However, the precision of the 2SLS frontier estimator, (and the other 2SLS estimators), depends on the noise to signal ratio of the instrument. As one can see in Figure 16d as the correlation in the actual running variable and the fuzzy running variable is higher, it adds to the precision of the 2SLS estimator. Since the sample in this application is sizable, the 2SLS frontier approach is the best candidate in obtaining the treatment effect. The strength of the instrument can also be seen in the F-statistic of the first-stage (Figure 17).

Given the qualities of the response-surface 2SLS frontier approach, and the normalized single-index approach, these are therefore the estimators of preference. The response surface approach can be represented by:

$$Y_b = \alpha_0 + \tau \hat{D}_b + h(F_b, G_b) + \epsilon_{ob} \tag{19}$$

$$D_b = \alpha_1 + \delta_1 D_b^{pred} + h(F_b, G_b) + \epsilon_{1b} \ , \tag{20}$$

where actual Educationally Backward Block (EBB) assignment in subdistrict (block) b is instrumented with predicted assignment based on the criteria detailed above, and $h(.)$ is a quadratic of all the terms that were used to determine EBB assignment.

The normalized, single index approach, on the other hand, creates a single running variable, represented by:

$$SI = \min\{\frac{F - F_c}{\sigma_F}, \frac{G - G_c}{\sigma_G}\} \tag{21}$$

Then, using the single-index SI, standard RD estimation procedures can be performed.

In Table 1 I summarize the results of this exercise.

4 Data

In order to study the policies in a comprehensive manner, I put together a number of large datasets, that have not been used in this manner before. To conduct this analysis, I required data on school inputs at the school level, household level data on years of education, and Census data on literacy rates. Since there is no dataset that contains all these details, various sources of data needed to be combined. Some of the datasets are representative at the district level and others at the sub-district level.

Table 1: Summary of Monte Carlo Results

Estimator	**Description**	**Pros and Cons**
OLS	Regress outcome on treatment	If (and only if) the error term in the outcome variable is uncorrelated with the other errors, then unbiased, whereas the 2SLS estimators are only consistent. Needs to be scaled up by probability of treatment.
OLS- Control Function	OLS but control for F and G	Does poorly when there is systematic selection into treatment that is unobservable to the researcher. Needs to be scaled up by probability of treatment.
Conditional Subsample	Condition on being above one cutoff, and do single-RD on other cutoff	Optimal bandwidth selection possible for one running variable. Does poorly when high correlation between the outcome and the variable that conditioning happens on. Instrument can be weak.
Added fuzziness	Ignore one running variable	Univariate optimal bandwidth selection possible. Weak instruments problem, especially when assignment is fuzzy on the running variable used for cutoff.
2SLS Frontier	IV-2SLS, and parametrically control for surface	Biased, but consistent. No non-parametric optimal bandwidth selection. Precision depends on the fuzziness in treatment assignment. No weak-IV problem.
Normalized single-index	Normalize running variables, and RD on closest to cutoff	Univariate optimal bandwidth selection possible. Not as precise as 2SLS frontier, and needs larger sample.

Data for inputs into schools comes from the District Information System for Education (DISE), which was established to collect data at the school level in order to inform policy makers in the Indian government about the bottlenecks in the education sector. While a limited number of their variables are available freely at an aggregated level, the bulk of their interesting data is obtainable only at a school-by-school basis on their website. I therefore accessed the data on a school-by-school basis and compiled it for each school separately. For the analysis, a 10% random sample of the data was collected, stratified by year and district. DISE claims to cover all the schools in the country (about 1.45 million schools in 2014) each year between 2005 and 2014, and consists of detailed information on number of schools, when they were built, whether they are public or privately owned, number of teachers by levels of education, and various infrastructural features.[22]

Table 3 summarizes the various variables of interest in the year 2005, twelve years after the DPEP started. The top panel classifies schools by ownership (government vs private), and when they were built (before 1993, the first year of the earliest program and after). 27% of all schools existing in 2005 were built post 1993, and while 20% were government schools, the remaining 7% were private schools. There is large variation in the amount of infrastructural provisioning - while 84% of the schools had drinking water, only 31% had constant electric supply. DISE data also collects information on Block and Cluster Resource Centers which are often used for training teachers, and have computerized facilities which are used to access other teaching materials. On average, schools were about 13 kms away from the closest Block Resource Center (BRC), and were visited by a BRC official about 1.5 times a year. The data also has information on various sources of funding, one of which is the Teacher's Learning Material (TLM) grant that all schools are eligible for irrespective of other programs. On average, schools got about Rs. 1517 ($38) in total from the TLM grant every year.

In order to study educational outcomes, household surveys and Census data was used. While household surveys are representative at the district level, only Census data has detailed information at the sub-district (block) level. Since the variation under the second program (NPEGEL-KGBV) is at the sub-district level, it is necessary to use Census data. Census data, unfortunately only has a limited number of outcome variables, including literacy by gender and rural-urban status. A panel of blocks can be created using the 1991, 2001 and 2011 Census years, all of which include block-level statistics. The panel is particularly challenging because of splits and merges in various districts, so I used detailed information on administrative areas and shape-files to compile the panel.

[22]The DISE data was initially meant to cover only in DPEP districts, but was expanded to cover the rest of the country in the early 2000s. The data is collected by head teachers, and verifies by cluster resource coordinators and block educational officers. Cross verification is done by head teachers of one school for another, and by Department of Education officials.

Of the available Census years, the 2001 data was used by the Planning Commission to identify Educationally Backward Blocks (EBBs). Since the NPEGEL-KGBV policy was implemented in the early 2000s, the only available "post-policy" year is 2011. It is not possible to use household-surveys since there is no survey in India that is representative at the block level. Household surveys collect only information on a small proportion of blocks, and the Multi-dimensional RD empirical strategy necessitates having a large density of blocks near the cutoffs of the literacy distribution.

Household surveys, can be used to study the impacts of the DPEP policy which varied at the district level. The National Sample Survey (NSS) is a nationally-representative survey used by many researchers studying India. Merging the household data to the schooling data is challenging due to changes in administrative boundaries over time. I use Census data and shape-files on changes in administrative boundaries to merge the datasets. I show household-level results for a subset of districts which had the DISE data, and also separately for all districts in the country. The 2009 round was used to study the longer-term impacts of the DPEP policy, and summary statistics for this round are presented in Table 2. In 2009, only about 60% of the population had finished primary school, and on average people had about 6 years of education and earned about Rs 1466 ($30) a month. While earlier rounds of the dataset are available, the 2009 round is the first large-sample round after the end of the DPEP program, and has the added advantage of allowing enough time for students affected by the policy to become a part of the labor market.

5 Estimation

To causally estimate the parameters of the model, I use a Regression Discontinuity (RD) design. Since the MRD approach is new to the literature, I provide Monte Carlo evidence to identify the optimal estimation strategy in this context. My Monte Carlo results may be useful for other researchers in selecting the appropriate estimators given the expected unobserved correlations in their contexts.

5.1 One-Dimensional Regression Discontinuity

In order to target the DPEP program to districts that were worst off in terms of educational outcomes, a selection criterion was used. Districts that had a female literacy below the national average (based on the previous 1991 Census) were eligible for the program, but not all those districts were selected. Furthermore, for some states, which have no low-literacy districts, a few districts were selected at the state's discretion. This selection procedure allows for a fuzzy regression discontinuity design using the 1991 female-literacy as a running variable. The fuzzy design allows for slippage on both sides, since not all low-literacy districts were selected, and for states with no low-literacy districts, some high-literacy districts were selected.

Since around the cutoff we should not expect any discontinuity in the distribution of individual labor-market abilities or individual-specific costs of going to school, unrelated to the policy, the RD estimator is unbiased. Furthermore, at the cutoff, we should expect no discontinuity in pre-policy labor market characteristics and regional outputs, that would otherwise bias the estimated parameters.

The first stage is presented in Figure 4. It is clear that the more literate amongst the eligible districts (i.e. amongst the districts with lower than average female literacy) were selected for the program, leading to a discontinuity at the cutoff. There is also visible slippage at both ends, with not all eligible districts being selected, and not all selected districts being eligible. Since it is clear that policy makers selected the most literate of the low-literacy districts, there is a high likelihood of political manipulation that is correlated with a whole host of unobserved characteristics in these regions. Given such a set-up, regression specifications that do not allow for these differences in unobserved characteristics will be biased. An RD specification can, therefore, provide a causal estimate of the impact of this program. This will be the Local Average Treatment Effect (LATE) for districts near the cutoff value.

Estimating causal impacts requires that the cutoff not be manipulated, which is likely in this case since the cutoff chosen was the national average of the female literacy rate from the previous 1991 Census. Furthermore, the McCrary (2008) tests indicate that the density of districts and of households around the cutoff is smooth (Figure 5), since the p-value of the change in density at the cutoff is 0.71. In the process of the analysis, other placebo tests will be discussed that solidify the RD assumptions that there were no other discontinuities at the same cutoff.[23]

While RD results will be represented graphically, the coefficients of interest will also be calculated using a two-stage least squares procedure where the optimal bandwidth will be calculated using two different methods - the Calonico et al. (2014) method, and the Imbens and Kalyanaraman (2012) method. The Imbens and Kalyanaraman (2012) method uses a data-driven bandwidth selection algorithm to identify the optimal bandwidth for a local linear regression given a squared-loss function, and the Calonico et al. (2014) method also

[23] Cattaneo et al. (2015) offers an alternative test for manipulation at the cutoff that does not rely on the selection of binning parameters. The p-value of a discontinuity in the density using their method is 0.97, indicating almost no likelihood of discontinuity in densities.

performs a bias-correction to the coefficient of interest and the standard errors calculated. Results using both the optimal bandwidth procedures are presented, and are robust to using more parametric approaches like local-linear and quadratic control-function approaches as suggested by Hahn et al. (2001) and Imbens and Lemieux (2008).

5.2 Estimation: Multi-Dimensional Regression Discontinuity

As described in Section 2.2, the NPEGEL-KBGV program was instituted at the sub-district level, where two criteria were required for the selection of these sub-districts - first, the female literacy rate in these sub-districts had to be below the national average, and second, the gender gap (the difference between the male and female literacy rates) had to be above the national average. Since many of the policies under this program were targeted to girls, the difference between male and female literacy was used to identify which regions had a higher level of gender inequality.

In Section 3, I discuss the estimation strategy of the MRD approach, where I detail the parameters estimated, the candidate estimators, and perform Monte Carlo exercises to identify the best estimators for a context where assignment is fuzzy and there is high correlation between the two different running variables.

Table 5 shows the prediction success rate of EBBs, where "slippage" in assisgnment, necessitating the need for the fuzzy-design. There is however no evidence of discontinuities in baseline variables from the 2001 Census (literacy, sex ratio, agricultural workers, cultivators, population, etc.), showing that there is balance pre-treatment.

Figures 7 and 8 show the scatter plots for the sub-districts that were selected as Educationally Backward Blocks (EBBs) across the gender gap and female literacy axes. The probability of being an EBB appears higher for the high-gender gap, low-female literacy districts. Furthermore, an additional criterion of having a rural female literacy rate below the national average was also used, and this was mostly adhered to as well (Figure 8).

Figures 9 to 14 show the first-stage graphs at the different cutoffs conditional on certain sub-samples of the data. In Figure 9 the sample is restricted to having a rural female literacy rate above the national average and total female literacy rate below the national average, and we see the expected discontinuity at the gender gap cutoff. In Figure 10 the sample is restricted to having a female literacy above the national average, and as expected there is no discontinuity at the gender gap cutoff. This is because the female literacy is too high and none of these blocks should have received the program in the first place.

In Figure 11 there is again no discontinuity because the sample is restricted to having a female rural literacy below 45% and therefore all these sub-districts were expected receive the program anyway. Restricting the sample to having a high gender gap and high rural female literacy, one can see the expected discontinuity at the female literacy cutoff in Figure 12. However, restricting it to the sample that has a low gender gap, we see no visible discontinuity in Figure 13, since none of these blocks should have received the program. Finally, there is again the lack of a discontinuity for the sample that has low rural female literacy, since all those blocks receive the program (Figure 14). While these figures clearly show discontinuities that are consistent with the assignment rule, they also show clear evidence of slippage, necessitating the use of a fuzzy design.

6 Which Policy Was More Effective?

The effectiveness of a policy may depend on the level of decentralization. Local bodies have better information on what the funds should be spent on, whereas more centralized bodies may have better knowledge of which regions the money should be allocated to. The DPEP policy was largely administered at the district level, and the NPEGEL-KGBV at the sub-district level. I am, therefore, able to compare the impacts between the two policies and study the extent to which decentralization matters.

The household-level analysis can also be split up by the age-groups that should and should not have been directly affected by the DPEP program. There is a sharp drop in schooling-enrollment rates at the age of 19, because by that age students have already usually finished schooling. Since the household survey was conducted 16 years after the start of the DPEP program, anybody above the age of 35 should not be directly affected by DPEP. Those under the age of 35 in treated districts, however, should be directly affected.[24]

Tables 4 and 6 show the impacts on illiteracy rates for regions near the cutoff. Table 4 shows the proportion of persons who are 'not literate'. The table splits up the results by those who should be directly affected (under the age of 35), and those who should not be directly affected (above 35) by DPEP. The fall in the rate of illiteracy at the cutoff lies between 2.5 and 6.3 percentage points.

Table 6 shows the impacts of the decentralized NPEGEL-KGBV policy by using the multi-dimensional RD method. Seven years after the program is implemented, the male literacy

[24]Since the parameters are estimated using a RD design, they will all be LATEs in the neighborhood of the cutoff. For example, the estimated returns to education are for those students who were induced into getting more schooling and lived in the districts near the RD cutoff.

rate is 5.5 percentage points higher, and female literacy rates are 7.3 percentage points higher, leading to a fall in the gender gap of about 2.6 percentage points. Since the NPEGEL-KGBV is targeted towards girls, the larger impact on female literacy rates is expected. There is however a substantial impact on male literacy rates as well. While the overall literacy rate rises by 6.3 percentage points, this is mostly due to a rise in rural literacy rates which witnessed a rise of 7.8 percentage points. This concentration of the effects in rural areas is expected because of the program prioritizing these regions. While this is suggestive evidence that the decentralized NPEGEL-KGBV policy may have been more effective, the impacts across the two policies are not statistically different.

7 Conclusion

Evidence from other contexts suggests mixed results on the effects of decentralization of government programs. Some studies suggest that corruption and leakages at each stage of the disbursement process may undermine the effectiveness of the policy (Ferraz and Finan (2008); Olken (2007); Reinikka and Svensson (2004)). On the other hand, decentralized policies may utilize local knowledge to better target the policies to areas that need them more (Olken (2010)). Aside from the direct evidence on decentralized policies, there are concerns of the wastage of funds on any government program as well (Bandiera et al. (2009); Easterly (2006, 2008)).

This paper develops ideal estimators to causally examine the effect of decentralization on the implementation quality of government policy. The two education policies studied in this paper largely differed on the administrative levels on where the funds were controlled and where the school-building was targeted. Comparing the two policies, therefore, allows me to see which type of policy is more effective. One of the policies requires a multi-dimensional RD approach, for which I develop estimators that may be used in other contexts as well. Comparing the two policies, I find evidence that the decentralized policy is more effective.

Since the MRD approach is new to the literature, I use a Monte Carlo setup to isolate the optimal estimation strategy. Unlike recent work by Reardon and Robinson (2012) and Wong et al. (2012), I concentrate on an application that may be of interest to applied economists or policy makers, where there is imperfect assignment, possible selection of regions by policy makers, and high correlation between the running variables. The exercise discusses how correlations in the different errors may affect the bias and precision of the different estimators. This can then allow researchers in other contexts as well to use these estimators.

References

Afridi, F. (2010). Child Welfare Programs and Child Nutrition: Evidence from a Mandated School Meal Program in India. *Journal of Development Economics*, 92(2):152–165.

Aggarwal, Y. (2000). Monitoring and Evaluation Under DPEP: Measurement of Social Impact. *NIEPA*, New Delhi.

Aguero, J. M. and Bharadwaj, P. (2014). Do the More Educated Know More about Health? Evidence from Schooling and HIV Knowledge in Zimbabwe. *Economic Development and Cultural Change*, 62(3):489–517.

Angrist, J., Bettinger, E., and Kremer, M. (2006). Long-Term Consequences of Secondary School Vouchers: Evidence from an Administrative Records in Colombia. *American Economic Review*, 96(3):847–72.

Ashraf, N., Bau, N., Nunn, N., and Voena, A. (2015). Bride Price and the Returns to Education for Women. *Mimeo*, Harvard.

Ayyar, R. (2008). Country Agency Relationship in Development Cooperation: An Indian Experience. *Global Monitoring Report*, UNESCO.

Bandiera, O., Prat, A., and Valletti, T. (2009). Active and Passive Waste in Government Spending: Evidence from a Policy Experiment. *American Economic Review*, 99(4):1278–1308.

Banerjee, A., Banerji, R., Duflo, E., Glennerster, R., and Khemani, S. (2010). Pitfalls of Participatory Programs: Evidence From a Randomized Evaluation in Education in India. *American Economic Journal: Economic Policy*, 2(1):1–30.

Banerjee, A., Cole, S., Duflo, E., and Linden, L. (2007). Remedying Education: Evidence from Two Randomized Experiments in India. *Quarterly Journal of Economics*, 22(3):1235–1264.

Bardhan, P. (2002). Decentralization of Governance and Development. *The Journal of Economic Perspectives.*

Bardhan, P. and Mookherjee, D. (2000). Capture and Governance at Local and National Levels. *American Economic Review.*

Bardhan, P. and Mookherjee, D. (2005). Decentralizing Antipoverty Program Delivery in Developing Countries. *Journal of Public Economics.*

Basurto, P., Dupas, P., and Robinson, J. (2015). Decentralization and Efficiency of Subsidy Targeting: Evidence from Chiefs in Rural Malawi. *Mimeo*.

Behrman, J. R., Birdsall, N., and Kaplan, R. (1996). The Quality of Schooling and Labor Market Outcomes. *Nancy Birdsall and Richard H. Sabot (eds.) Opportunity Foregone: Education in Brazil. Johns Hopkins University Press.*, 8.

Birdsall, N. (1985). Public Inputs and Child Schooling in Brazil. *Journal of Development Economics*, 18(1):67–86.

Bold, T., Kimenyi, M., and Sandefur, J. (2013). Public and Private Provision of Education in Kenya. *Journal of African Economies*, 22(2):39–56.

Borkum, E., He, F., and Linden, L. (2010). School Libraries and Language Skills in Indian Primary Schools: A Randomized Evaluation of the Akshara Library Program. *Mimeo*, Columbia.

Calonico, S., Cattaneo, M., and Titiunik, R. (2014). Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica*, 82(6):2295–2326.

Cannonier, C. and Mocan, N. (2012). Empowering Women Through Education: Evidence from Sierra Leone,. *NBER Working Papers 18016*.

Card, D. and Payne, A. A. (2002). School finance reform, the distribution of school spending, and the distribution of student test scores. *Journal of Public Economics*, 83:49–82.

Cattaneo, M., Jansson, M., and Ma, X. (2015). Simple Local Regression Distribution Estimators with an Application to Manipulation Testing. *Mimeo*, Michigan.

Chin, A. (2005). Can Redistributing Teachers Across Schools Raise Educational Attainment? Evidence from Operation Blackboard in India. *Journal of Development Economics*, 78(2):384–405.

Coady, D., Grosh, M., and Hoddinott, J. (2004). Targeting of Transfers in Developing Countries : Review of Lessons and Experience. *Washington, DC: World Bank*.

Colclough, C. and De, A. (2010). The Impact of Aid on Education Policy in India. *Research Consortium on Educational Outcomes and Poverty*, 30(5):497–507.

Das, J., Dercon, S., Habyarimana, J., Krishnan, P., Muralidharan, K., and Sundararaman, V. (2013a). School Inputs, Household Substitution, and Test Scores. *American Economic Journal: Applied Economics*, 5(2):29–57.

Das, J., Holla, A., Kremer, M., Mohpal, A., and Muralidharan, K. (2013b). The Fiscal Costs of Weak Governance: Evidence from Teacher Absence in India. *Mimeo*, UC San Diego.

Deininger, K. (2003). Does Cost of Schooling Affect Enrollment by the Poor? Universal Primary Education in Uganda. *Economics of Education Review*, 22(3):291–305.

Duflo, E. (2001). Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment. *American Economic Review*, 1(4):795–813.

Duflo, E., Hanna, R., and Ryan, S. (2012). Incentives Work: Getting Teachers to Come to School. *American Economic Review*, 102(4):1241–78.

Easterly, W. R. (2006). *The White Man's Burden: Why the West's Efforts to Aid the Rest Have Done So Much Ill and So Little Good*. Oxford: Oxford University Press.

Easterly, W. R., editor (2008). *Reinventing Foreign Aid*. Cambridge, MA. MIT Press.

Ferraz, C. and Finan, F. (2008). Exposing Corrupt Politicians: The Effects of Brazil's Publicly Released Audits on Electoral Outcomes. *Quarterly Journal of Economics*, 123(2):703–745.

GOI (1994). *DPEP Principles*, volume New Delhi. Government of India Report.

GOI (2000). *DPEP Moves on ... Towards Universal Primary Education*, volume New Delhi. Government of India Report.

Grogger, J. (1996). School Expenditures and Post-Schooling Earnings: Evidence from High School and Beyond. *The Review of Economics and Statistics*, 78(4):653–664.

Hahn, J., Todd, P., and van der Klaauw, W. (2001). Identification and Estimation of Treatment Effects with a Regression Discontinuity Design. *Econometrica*, 69(1):201–209.

Hanushek, E. A. (1986). The Economics of Schooling: Production and Efficiency in Public Schools. *Journal of Economic Literature*, 24(3):1141–1147.

Hanushek, E. A. (1997). Assessing the Effects of School Resources on Student Performance: An Update. *Educational Evaluation and Policy Analysis*, 19(2):141–164.

Hanushek, E. A. (2003). The Failure of Input-Based Schooling Policies. *The Economic Journal*, 113:64–98.

Hanushek, E. A. (2006). School Resources. *Handbook of the Economics of Education*, 2:865–908.

Hanushek, E. A. (2008). Education Production Functions. *The New Palgrave Dictionary of Economics*, Basingstoke.

Hoxby, C. M. (2001). All School Finance Equalizations Are Not Created Equal. *Quarterly Journal of Economics*, 116(4):1189–1231.

Imbens, G. W. and Kalyanaraman, K. (2012). Optimal Bandwidth Choice for the Regression Discontinuity Estimator. *Review of Economic Studies*, 79(3):933–959.

Imbens, G. W. and Lemieux, T. (2008). Regression Discontinuity Designs: A Guide to Practice. *Journal of Econometrics*, 142(2):615–635.

Jacob, V., Kochar, A., and Reddy, S. (2008). School Size and Schooling Inequalities. *Mimeo*, Stanford.

Jalan, J. and Glinskaya, E. (2013). Small Bang for Big Bucks? An Evaluation of a Primary School Intervention in India. *Centre for Training and Research in Public Finance And Policy*, Calcutta.

Jayaraman, R., Simroth, D., and De Vericourt, F. (2010). The Impact of School Lunches on Primary School Enrollment: Evidence from India's Mid-day Meal Scheme. *CESifo Working Papers*, Munich.

Khanna, G. (2015). Large-Scale Education Reform in General Equilibrium: Regression Discontinuity Evidence from India. *mimeo*.

Kingdon, G. and Teal, F. (2010). Teacher Unions, Teacher Pay and Student Performance in India: A Pupil Fixed Effects Approach. *Journal of Development Economics*, 91(2):278–288.

Krueger, A. B. (1999). Experimental Estimates of Education Production Functions. *The Quarterly Journal of Economics*, 114(2):497–532.

Krueger, A. B. (2003). Economic Considerations and Class Size. *The Economic Journal*, 113:34–63.

Linden, L. (2008). Complement or Substitute? The Effect of Technology on Student Achievement in India. *InfoDev*, Columbia University, MIT Jameel Poverty Action Lab, IZA.

Loeb, S. and Bound, J. (1996). The Effect of Measured School Inputs on Academic Achievement: Evidence from the 1920s, 1930s and 1940s Birth Cohorts. *The Review of Economics and Statistics*, 78(4):653–664.

McCrary, J. (2008). Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test. *Journal of Econometrics*, 142(2):698–714.

Menon, P. (2001). Content Analysis of training Modules for Village Education Committees: A Study of seven DPEP states (Phase I). *Mimeo*, NIEPA, New Delhi.

Muralidharan, K. (2013). *Priorities for Primary Education Policy in Indias 12th Five-year Plan*. NCAER-Brookings India Policy Forum.

Muralidharan, K. and Sundararaman, V. (2010). The Impact of Diagnostic Feedback to Teachers on Student Learning: Experimental Evidence from India. *Economic Journal*, 120(546):F187–F203.

Niehaus, P., Attanassova, A., Bertrand, M., and Mullainathan, S. (2013). Targeting with Agents. *American Economic Journal: Economic Policy*.

Olken, B. (2007). Monitoring Corruption: Evidence from a Field Experiment in Indonesia. *Journal of Political Economy*, 115(2):200–249.

Olken, B. (2010). Direct Democracy and Local Public Goods: Evidence from a Field Experiment in Indonesia. *American Political Science Review*, 104(2):243–267.

Osili, U. O. and Long, B. T. (2008). Does Female Schooling Reduce Fertility? Evidence from Nigeria. *Journal of Development Economics*, 87(1):57–75.

Pandey, R. S. (2000). Going to Scale with Education Reform: India's district Primary Education Program, 1995-99. *Washington DC: The World Bank*, 1(4).

Papay, J. P., Murnane, R. J., and Willett, J. B. (2014). Hih-School Exit Examinations and the Schooling Decisions of Teenagers: Evidence from Regression-Discontinuity Approaches. *Journal of Research on Educational Effectiveness*, 7(1):1–27.

Reardon, S. F. and Robinson, J. P. (2012). Regression Discontinuity Designs with Multiple Rating-Score Variables. *Journal of Research on Educational Effectiveness*, 5(1):83–104.

Reinikka, R. and Svensson, J. (2004). Local Capture: Evidence from a Central Government Transfer Program in Uganda. *Quartely Journal of Economics*, 119(2):679–705.

Sifuna, D. N. (2007). The Challenge of Increasing Access and Improving Quality: An Analysis of Universal Primary Education Interventions in Kenya and Tanzania Since the 1970s. *International Review of Education*, 53(5).

Varghese, N. (1994). District Primary Education Programme: the Logic and the Logistics. *Journal of Educational Planning and Administration*, 2(1):69–106.

Wong, V. C., Steiner, P. M., and Cook, T. D. (2012). Analyzing Regression-Discontinuity Designs With Multiple Assignment Variables: A Comparative Study of Four Estimation Methods. *Journal of Educational and Behavioral Statistics*, 38(2):107–141.

WorldBank (1997). *Primary Education in India.* The World Bank, Washington DC.

8 Tables and Figures

8.1 Tables

Table 2: Summary Statistics: Household Level

	Non DPEP Mean	Non DPEP SD	DPEP Mean	DPEP SD	All Mean	All SD
Finished Primary School	0.6	0.5	0.5	0.5	0.6	0.5
Years of Education	6	4.9	5.1	4.9	5.7	4.9
Male	0.5	0.5	0.5	0.5	0.5	0.5
Age	29.3	18.9	28.2	18.7	28.9	18.8
Wage Earnings (2005 Rs.)	1,602.60	2,012.50	1,198.40	1,504.50	1,466.10	1,866.30

Source: National Sample Survey (2009). Age in years. Wages in 2005 Indian Rupees.

Table 3: Summary Statistics: School Level (2005)

	Mean	SD
Fraction of Schools:		
Built post 1993	0.277	0.447
Gov schools built post 1993	0.200	0.400
Pvt school built post 1993	0.075	0.263
Built between 1973-93	0.227	0.419
Gov schools built 1973-93	0.170	0.376
Pvt Schools built 1973-93	0.055	0.228
Fraction of Schools Having:		
Girl's Toilet	0.400	0.490
Electricity	0.312	0.463
Playground	0.549	0.498
Medical Checkups	0.541	0.498
Ramps	0.182	0.386
Boundary Wall	0.506	0.500
Drinking Water	0.846	0.361
A Pre-primary section	0.213	0.410
Block and Cluster Resource Centers:		
Visits by BRC Official	1.485	2.543
Distance to BRC (km.)	13.462	15.936
Visits by CRC Official	4.496	5.612
Distance to CRC (km.)	4.438	8.689
Teacher Learning Materials Grant:		
Amount Received (Rs.)	1517.100	8010.138
Amount Spent (Rs.)	1332.604	7611.869

Source: DISE (2005) - District Information System for Education.

Table 4: Fraction of people that are 'not literate'

All Districts Not-literate (1/0)	Below 35 years	Above 35 years	Below 35 years	Above 35 years
RD Estimate	-0.0307	0.0407	-0.0631	0.0192
	(0.0132)**	(0.0213)*	(0.00966)***	(0.0147)
Observations	67,764	31,700	159,945	81,040
Fuzzy Conventional p-value	0.0200	0.0556	6.50e-11	0.193
Fuzzy CCT Correct p-value	0.0706	0.0666	0.000992	0
Bandwidth selection procedure	CCT	CCT	I and K	I and K

All Districts - Earners Not-literate (1/0)	Below 35 years	Above 35 years	Below 35 years	Above 35 years
RD Estimate	-0.0833	0.0701	-0.0723	0.0703
	(0.0406)**	(0.0491)	(0.0310)**	(0.0376)*
Observations	7,951	6,711	14,791	10,876
Fuzzy Conventional p-value	0.0403	0.154	0.0198	0.0614
Fuzzy CCT Correct p-value	0.0184	0.0759	0.773	0.216
Bandwidth selection procedure	CCT	CCT	I and K	I and K

DISE Districts Not-literate (1/0)	Below 35 years	Above 35 years	Below 35 years	Above 35 years
RD Estimate	-0.0323	0.0457	-0.0246	0.0467
	(0.0172)*	(0.0237)*	(0.0138)*	(0.0210)**
Observations	63,443	37,631	88,259	44,039
Fuzzy Conventional p-value	0.0611	0.0540	0.0742	0.0263
Fuzzy CCT Correct p-value	0.0231	0.0352	0.00852	0.0469
Bandwidth selection procedure	CCT	CCT	I and K	I and K

DISE Districts - Earners Not-literate (1/0)	Below 35 years	Above 35 years	Below 35 years	Above 35 years
RD Estimate	-0.0755	0.0764	-0.0423	0.0401
	(0.0486)	(0.0487)	(0.0309)	(0.0327)
Observations	7,113	7,215	12,271	13,351
Fuzzy Conventional p-value	0.121	0.116	0.170	0.220
Fuzzy CCT Correct p-value	0.0440	0.0396	0.136	0.0284
Bandwidth selection procedure	CCT	CCT	I and K	I and K

National Sample Survey 2009-10. 'All Districts' include districts that are in the Household Survey. 'DISE districts' include a sub-sample of districts that have school-level DISE data.
Bandwidths: 'CCT' is the Calonico, Cattaneo and Tituinik (2014) method. 'I and K' is the Imbens and Kalyaranam (2011) method. 'CCT corrected p-value' is the bias-corrected p-values using the method in Calonico, Cattaneo and Tituinik (2014).

Table 5: Predicting Educationally Backward Blocks (EBBs)

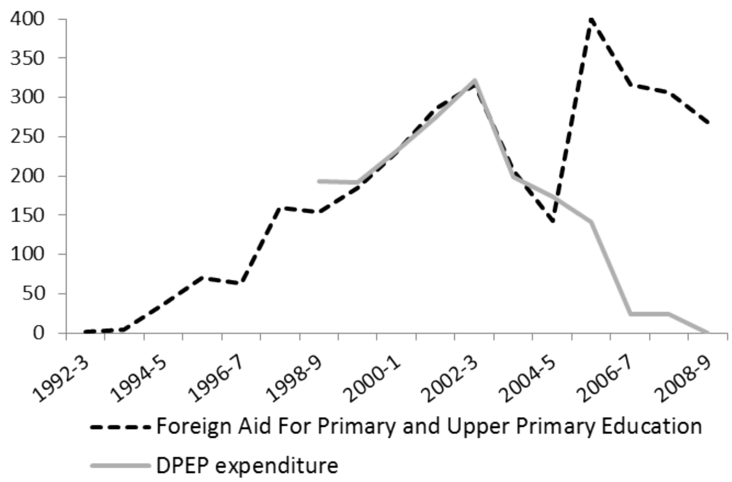
		Predicted Non-EBB	Predicted EBB	Total
Actual Non-EBB	No. of blocks	2,020	65	2,085
	Row percent	97	3	100
	Column Percent	96.65	2.54	44.83
Actual EBB	No. of blocks	70	2,496	2,566
	Row percent	2.73	97.27	100
	Column Percent	3.35	97.46	55.17
Total	No. of blocks	2,090	2,561	4,651
	Row percent	44.94	55.06	100
	Column Percent	100	100	100

Table 6: Literacy Rates by gender and region in the 2011 Census

	Male	Female	Total	Rural	Gender Gap
Coefficient	5.557	7.331	6.262	7.786	-2.592
SE cluster:					
Sub-district	(2.115)	(2.711)	(2.379)	(2.601)	(1.350)
District	(4.089)	(5.123)	(4.550)	(4.485)	(2.485)
Observations	4,621	4,621	4,621	4,606	4,606
R-squared	0.867	0.877	0.869	0.841	0.872

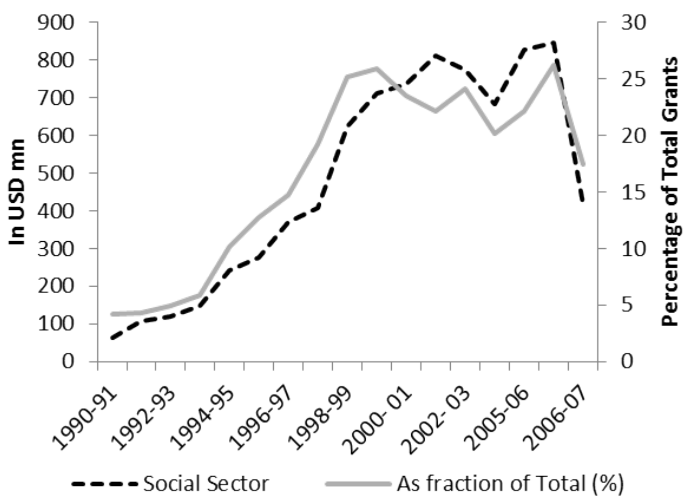
8.2 Figures

Figure 1: Foreign Aid and DPEP Expenditure (in USD mn)



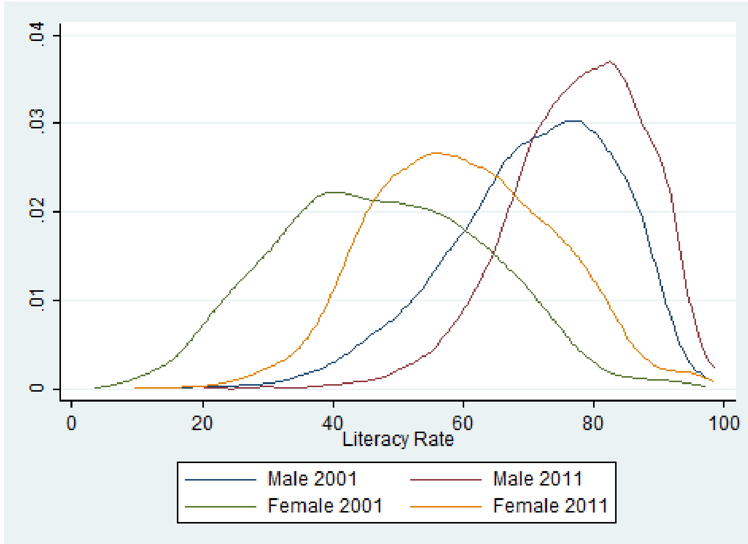
Foreign aid for expenditure on primary and upper primary education, and funds disbursed for DPEP. 1999 Indian rupees converted to USD using the 1999 exchange rate of Rs. 40 to $1. Sources: Foreign Aid from Colclough and De (2010). DPEP expenditures data compiled by author from Ministry of Human Resources and Development Reports, National Institute of Educational Planning and Administration, Lok Sabha Unstarred Question Numbers: 1807- 07.03.2006; 552- 24.02.2009; 55 - 26.02.2008; 267- 22.03.2005; 1320- 10.12.2003; 2018- 4.3.2003, and Rajya Sabha Unstarred Question No. 2855- 19.04.2002.

Figure 2: Social Sector (Health and Education) Grants/Loans from Central to State Governments



Central government grants and loans to State governments for spending in the social sector (health and education), and as a proportion of total grants/loans. 1999 Indian rupees converted to USD using the 1999 exchange rate of Rs. 40 to $1. Source : External Assistance Brochure of CAA&A, Department of Economic Affairs, Ministry of Finance, Government of India.

Figure 3: Literacy Rates Across Census Years



Census 2001 and 2011

Figure 4: First Stage of DPEP

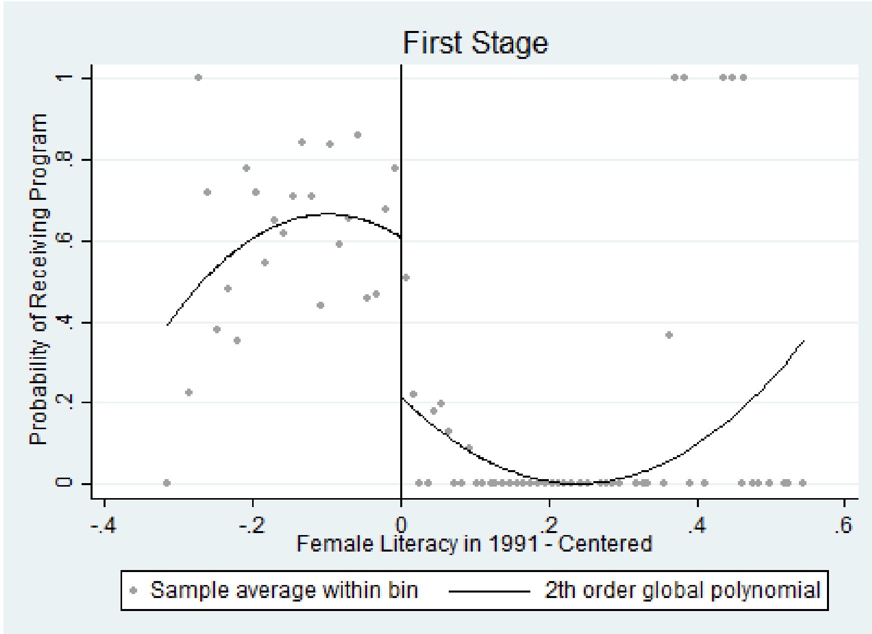
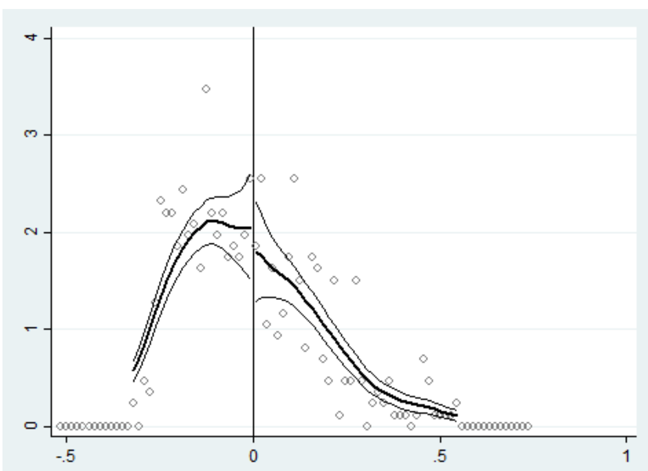
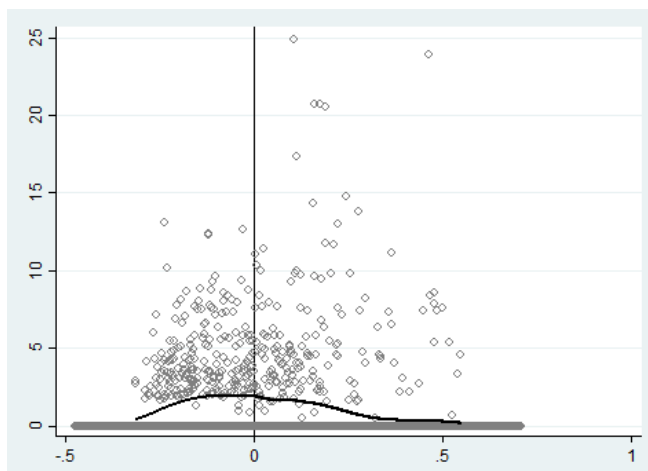


Figure 5: McCrary Tests

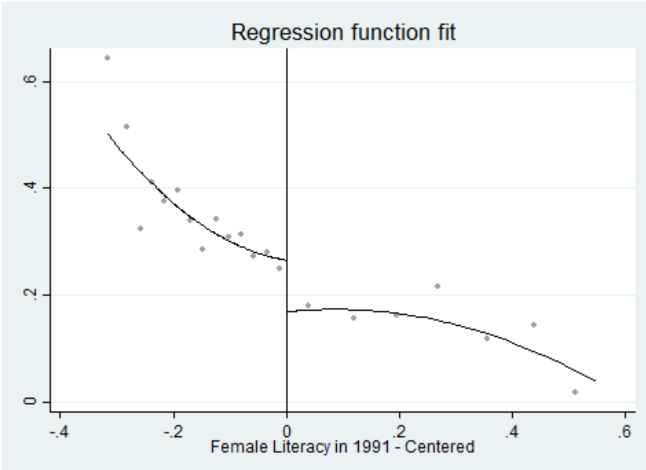


District Level McCrary Test

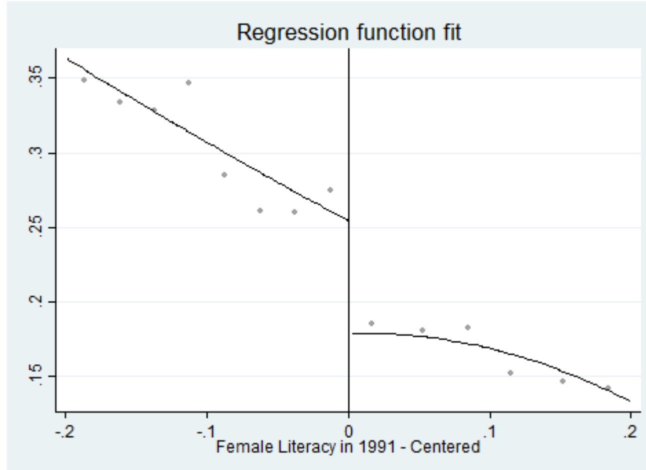


Household Level McCrary Test

Figure 6: Fraction of Schools Built Post 1993

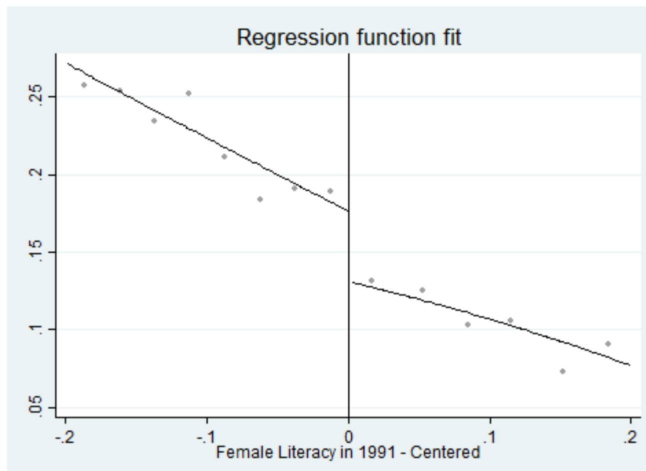


All Schools Built Post 1993

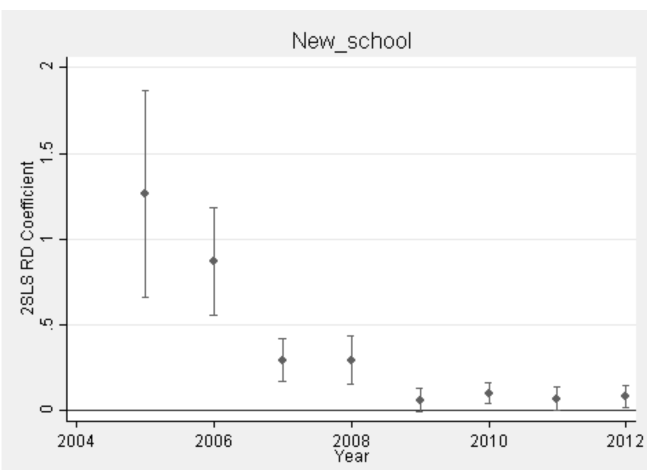


All Schools - smaller bandwidth

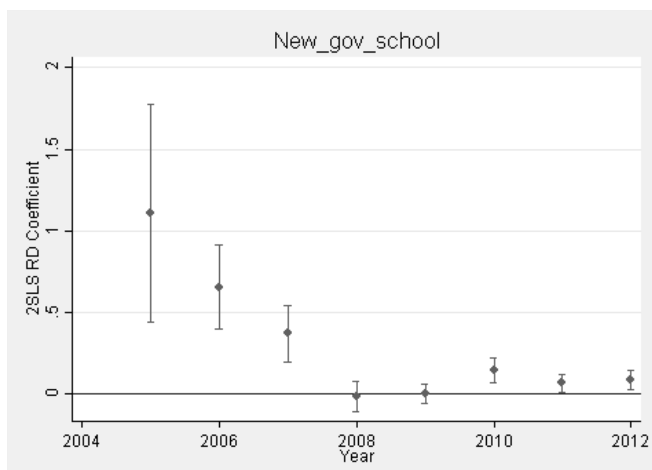
Government Schools Built post 1993



Government schools - smaller bandwidth



All Schools - RD Coefficient Over Time



Government Schools - RD Coefficient Over Time

Source: DISE (District Information System for Education) data. Top panels show RD graphs for a 10% subsample of the 2005 data. Bottom panels show RD coefficients calculated using Calonico, Cattaneo and Tituinik (2014) procedure.

8.3 Multi-Dimensional Regression Discontinuity: First Stage

Figure 7: Identifying Educationally Backward Blocks (EBBs)

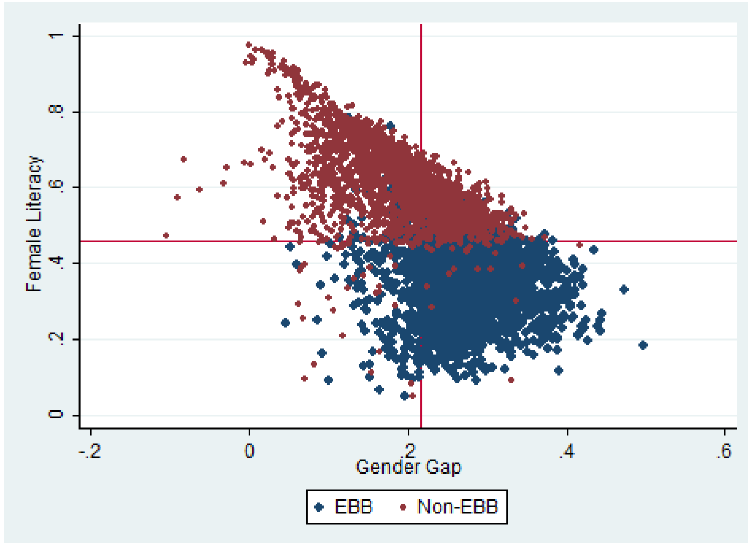


Figure 8: For sub-districts with rural female literacy below 45%

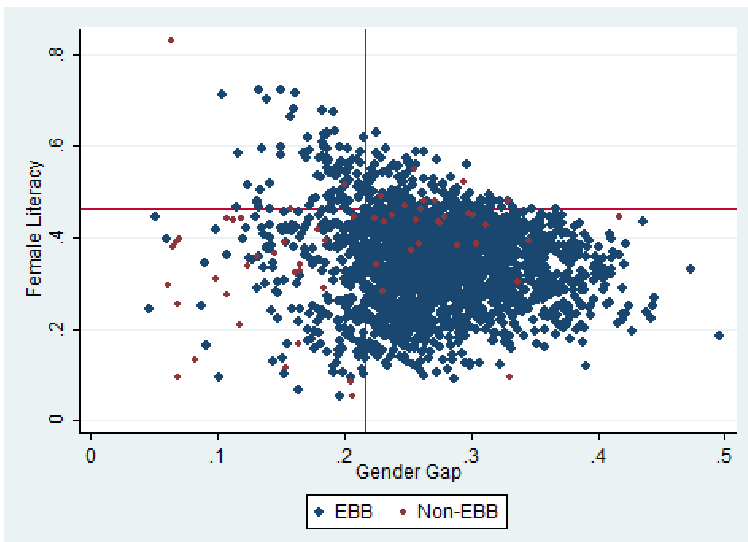


Figure 9: Discontinuity at Gender Gap cutoff for sub-districts with rural literacy above 45% but female literacy below 46.13%

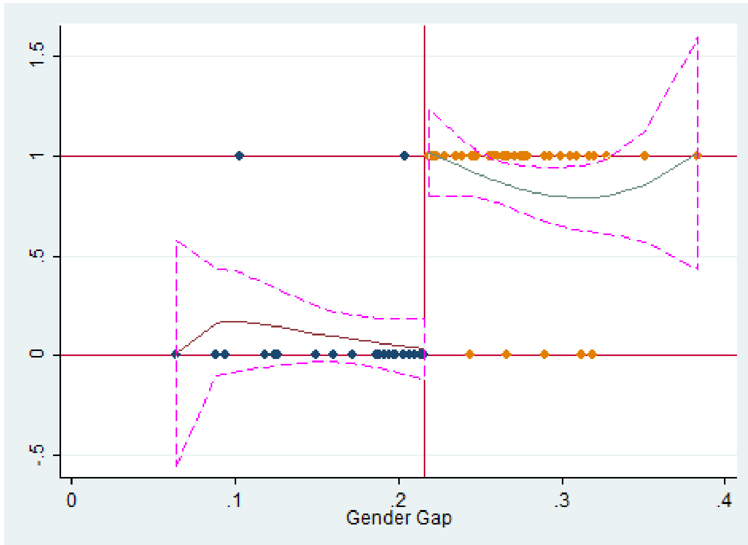


Figure 10: No discontinuity at Gender Gap cutoff for sub-districts with rural literacy above 45% but female literacy above 46.13%

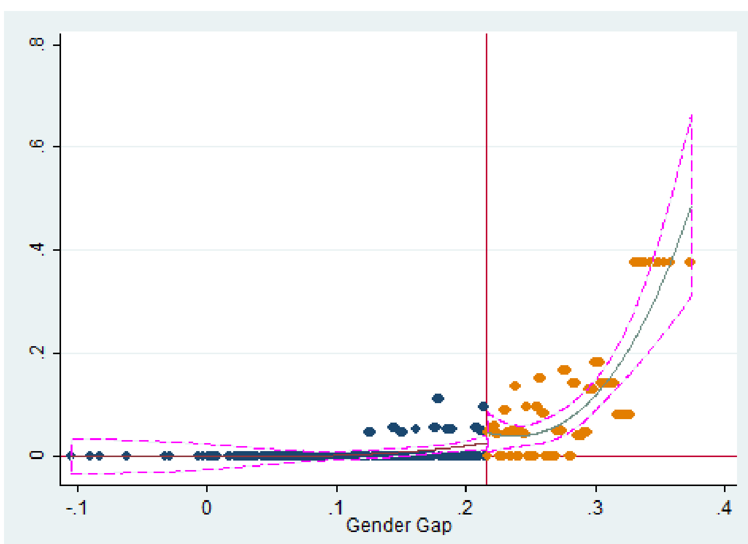


Figure 11: No discontinuity at Gender Gap cutoff for sub-districts with rural literacy below 45% but female literacy below 46.13%

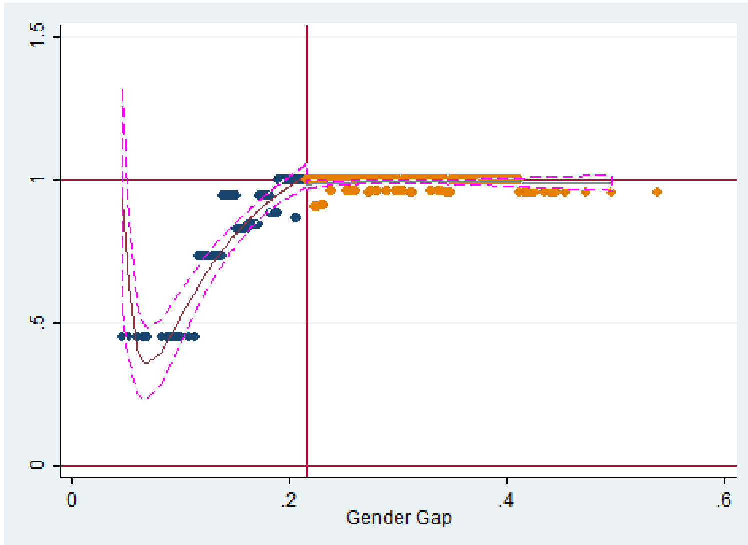


Figure 12: Discontinuity at Female Literacy cutoff for sub-districts with rural literacy above 45% but gender gap above 21%

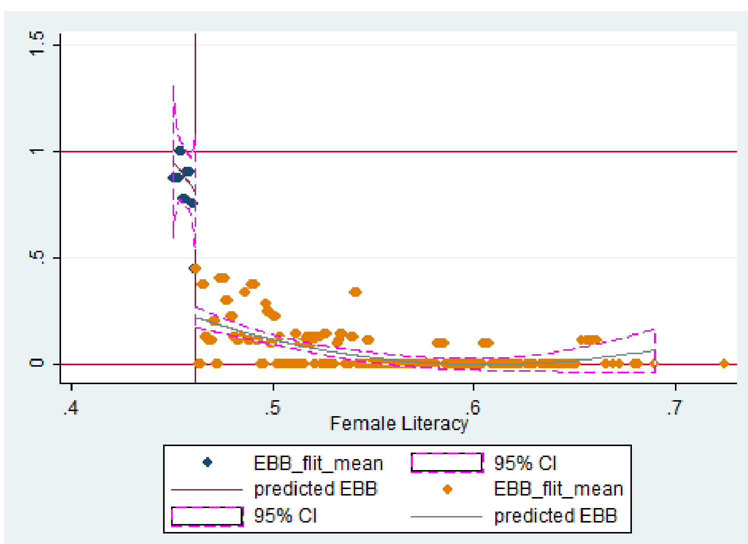


Figure 13: No discontinuity at Female Literacy cutoff for sub-districts with rural literacy above 45% but gender gap below 21%

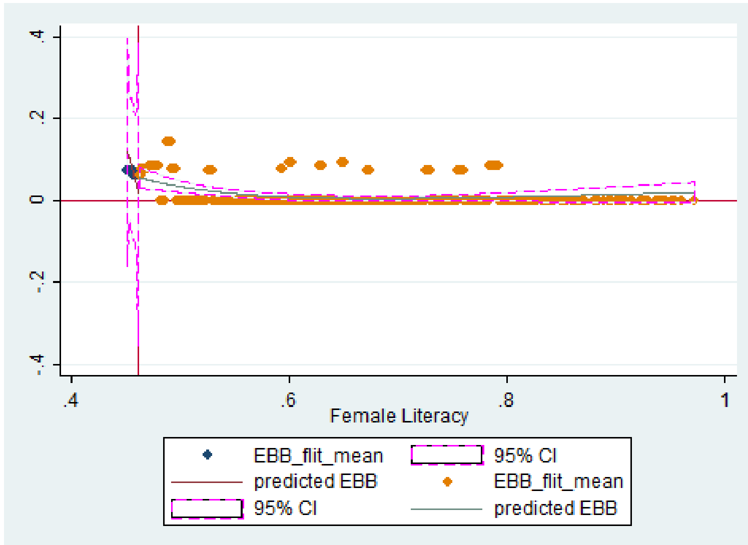
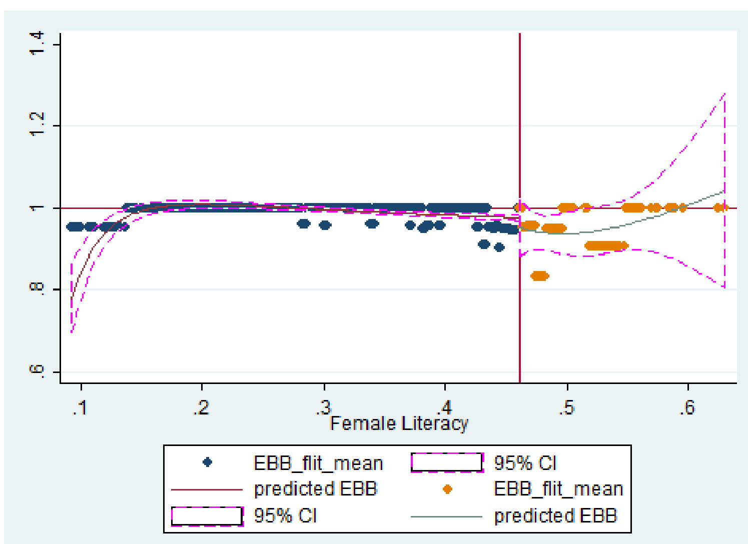
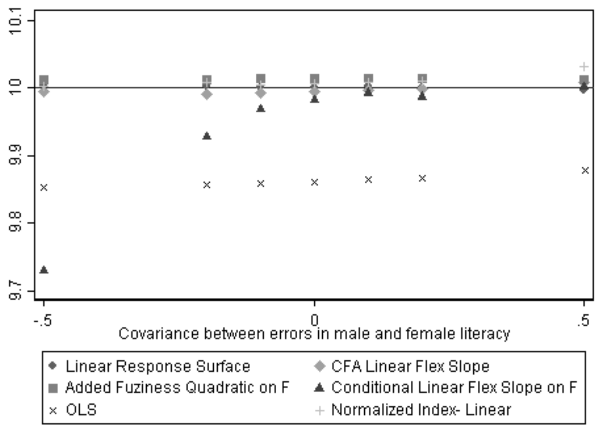


Figure 14: No discontinuity at Female Literacy cutoff for sub-districts with rural literacy below 45% but gender gap above 21%

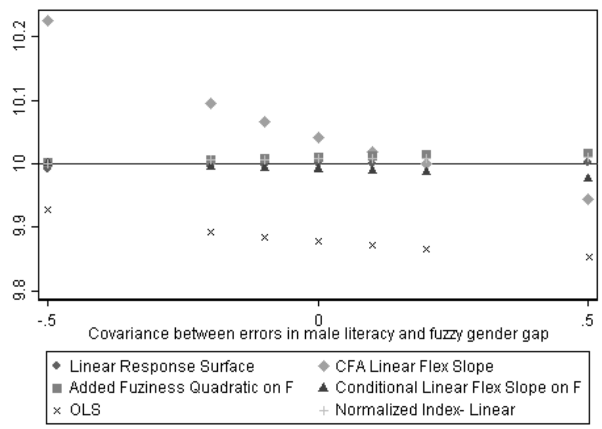


8.4 Multi-Dimensional Regression Discontinuity: Monte Carlo Exercises

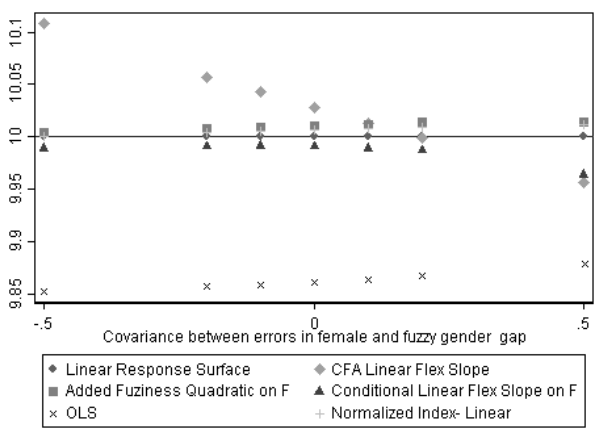
Figure 15: Candidate Estimators by Error Covariances



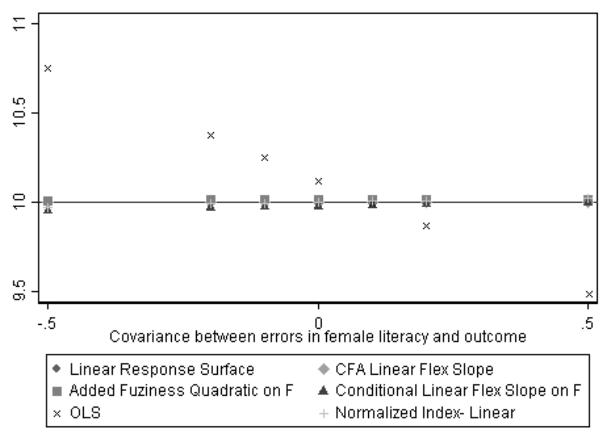
(a)



(b)



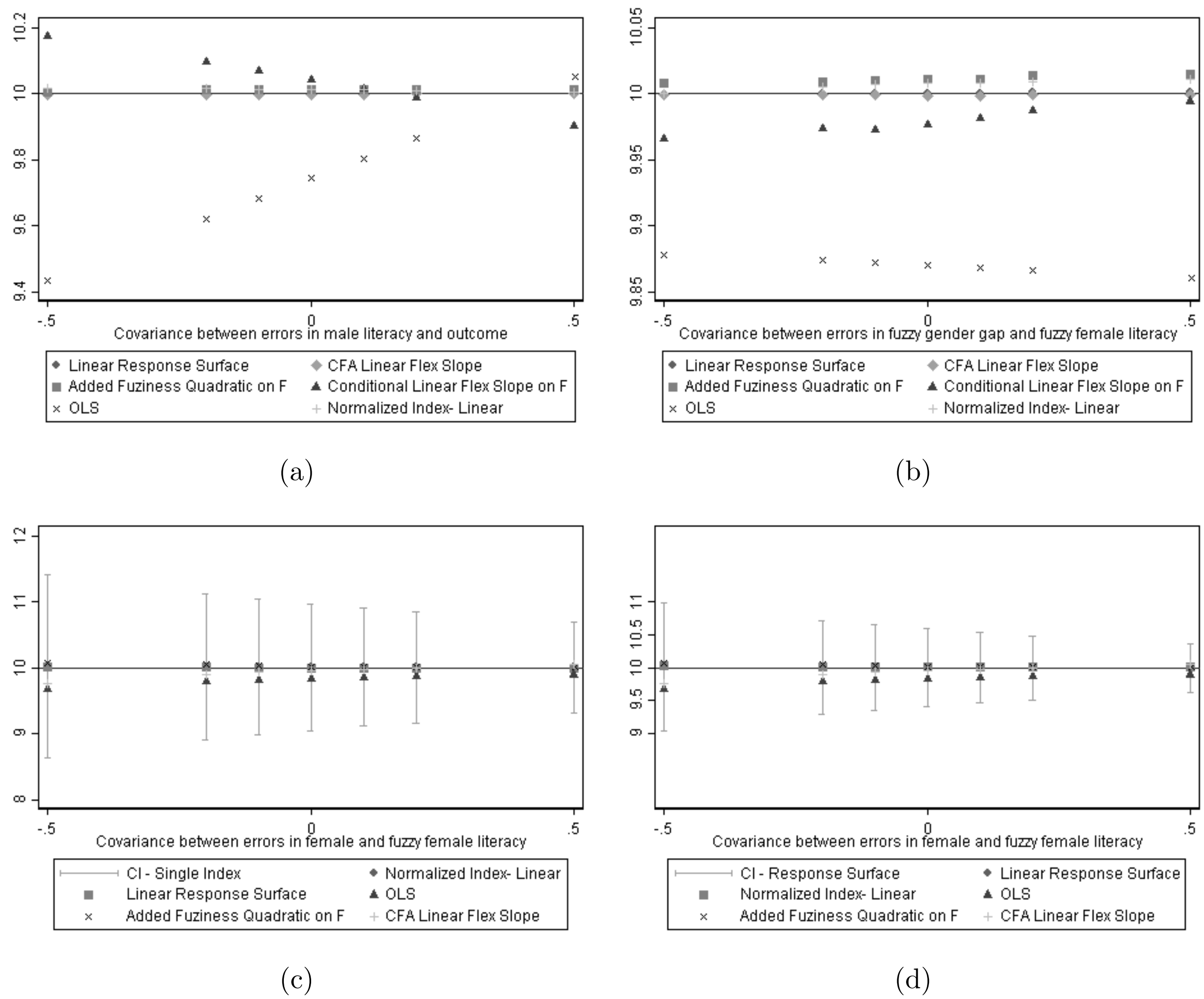
(c)



(d)

These graphs show the estimated coefficient using the different candidate estimators where the true value of the coefficient is 10. Along the horizontal axes, the covariance between the errors of the various variables of interest is varied in order to identify the source of the bias in the various estimators.

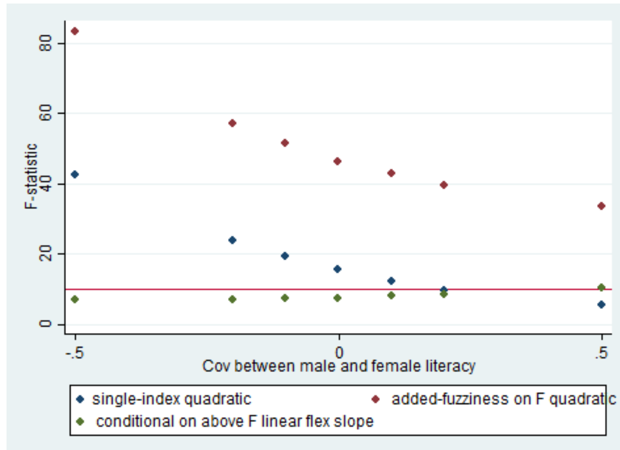
Figure 16: Candidate Estimators by Error Covariances



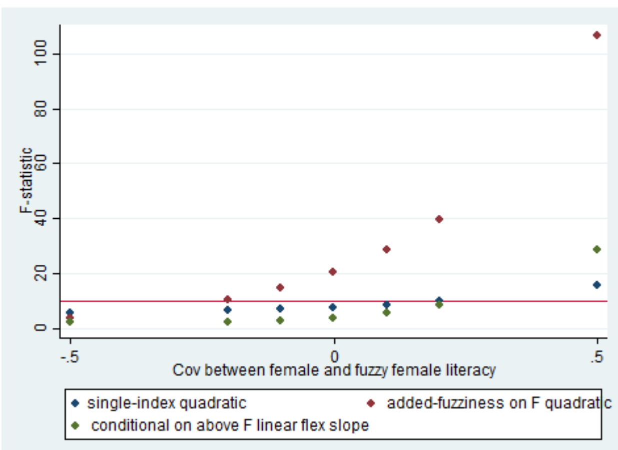
These graphs show the estimated coefficient using the different candidate estimators where the true value of the coefficient is 10. Along the horizontal axes, the covariance between the errors of the various variables of interest is varied in order to identify the source of the bias in the various estimators. Figures 16c and 16c have the same axes

Figure 17: Strength of the First Stage: F statistics

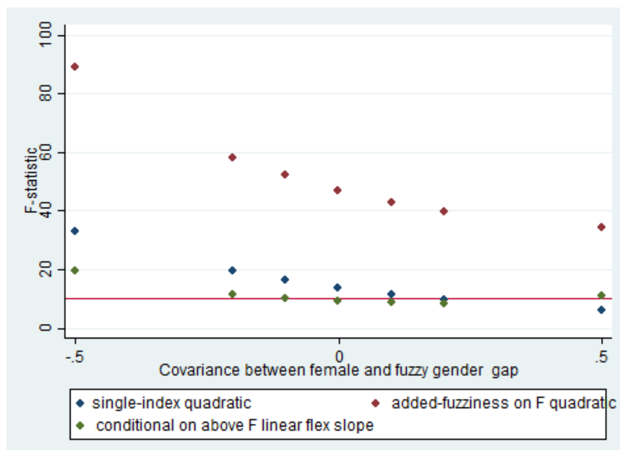
(a)



(b)



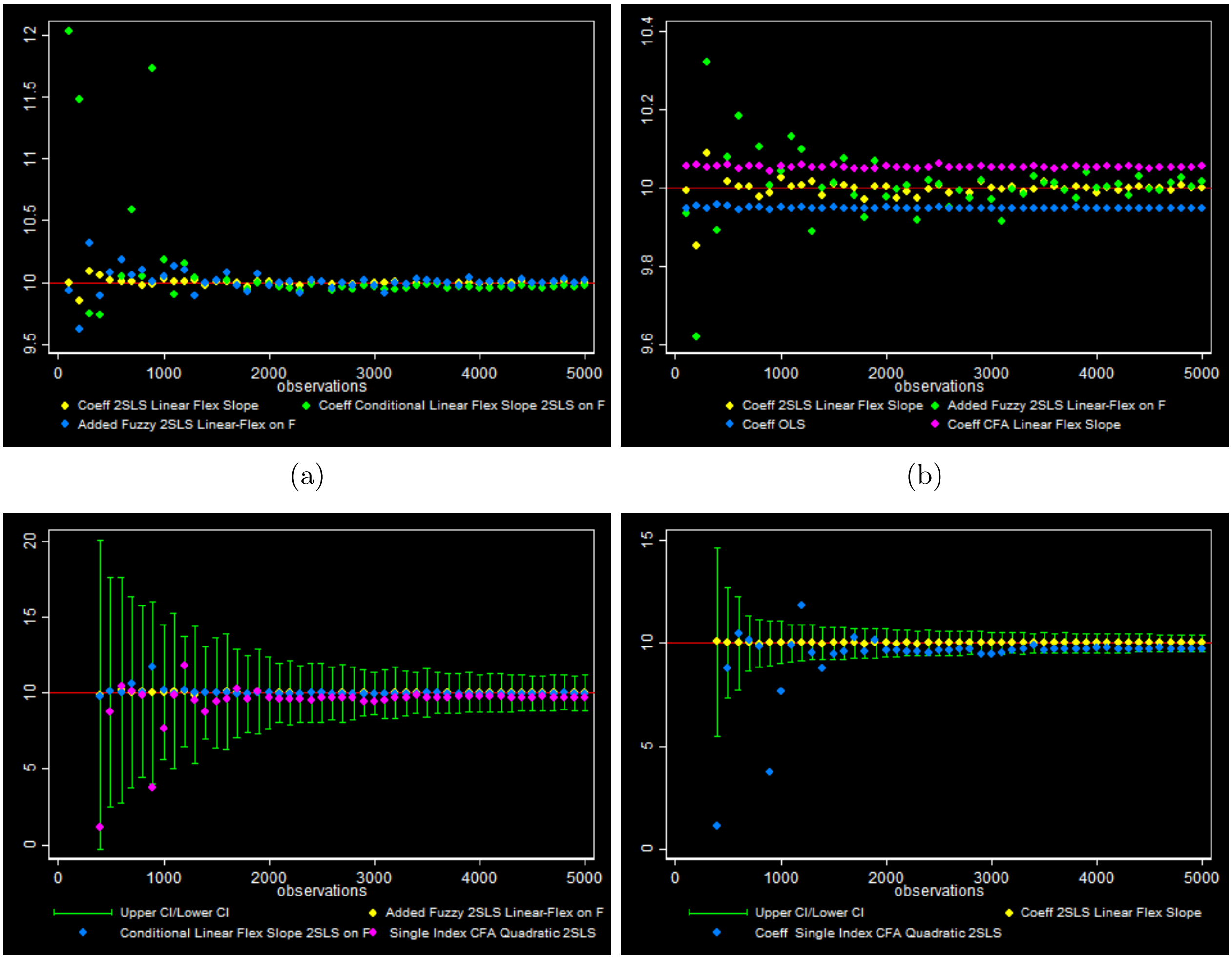
(c)



(d)

These graphs show F - statistics in the first-stage, using the different candidate estimators. A horizontal line at the rule-of-thumb F-stat value of 10 is shown. Along the horizontal axes, the covariance between the errors of the various variables of interest is varied in order to identify the source of the bias in the various estimators.

Figure 18: Candidate Estimators by Observations



(a) (b) (c) (d)

These graphs show the estimated coefficient using the different candidate estimators where the true value of the coefficient is 10. Along the horizontal axes, the number of observations is varied in order to identify the source of the bias and the strength of these estimators in small samples.