THE UNIVERSITY OF CHICAGO

ESSAYS IN DEVELOPMENT ECONOMICS

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE IRVING B. HARRIS
GRADUATE SCHOOL OF PUBLIC POLICY STUDIES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

BY
DEVIKA LAKHOTE

CHICAGO, ILLINOIS
JUNE 2023

For Akka and Baba

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I thank my friends and family for providing moral support. I thank my committee: Marianne Bertrand, Oeindrila Dube and Luis Martinez for their guidance. I thank the Harris School of Public Policy, the Pearson Institute, the Becker Friedman Institute and the Center for the study of Race, Politics and Culture for financial and administrative support. I thank Claire Fan And Tausif Alam for their contribution to my dissertation. I thank Rathan K. for data collection support. I thank all the participants of workshops at the University of Chicago for their comments. I thank Arun Chandrasekhar, Terence Chau, Emileigh Harrison, and Mythili Vinnakota for their overall support and feedback.

ABSTRACT

A large share of public expenditure goes towards building public infrastructure. Often, the infrastructure is targeted particularly to groups that have previously been under-served. Sometimes, there is no targeting but infrastructure improvements end up benefiting these groups more because of pre-existing inequalities. Indian society is deeply unequal along several such group lines: caste, religion, gender, language, and more. Therefore, group identity can be a pivotal determinant of access to goods, services, and infrastructure. Through this dissertation, I contribute to the big picture question: What is the relationship between group identity, and access to goods and infrastructure? We know from existing causal evidence that group identity plays an important role in determining economic outcomee, improving schooling infrastructure can lead to long-lasting impacts on educational attainment and workforce participation and household water access can have important consequences for health and overall welfare. My thesis contributes to both strands of the literature by evaluating two (previously unevaluated) nationwide infrastructure projects and measuring their effect on under-served groups (religious minorities and women). Many studies in the past have been constrained by a lack of granular nationwide data on many measures in India. There have been recent efforts to alleviate this problem. My thesis also contributes to this effort by creating two new granular nationwide datasets for public use. I collect and digitize data on infrastructure spending and construction in India from the last decade, that has not been used before. I also put together nationwide consumer data at a granular level to study identity-based goods consumption. In Chapter 1, I evaluate an infrastructure policy in India that invests in regions with a high share of religious minorities. I use an event study design to measure the medium-term effect of these investments on schooling infrastructure and schooling outcomes. I find that there is an improvement in classroom infrastructure (as targeted by the policy), but this improvement doesn't lead to an improvement in schooling outcomes. I also use an alternate approach to study the effect of the policy on regions with

close to 25% share of religious minorities using an RD design. In Chapter 2, I study the effect of a rollout of household water access points (pipes, wells, tanks, pumps, etc.) on educational outcomes for children, especially girls. I use an event study design to measure the medium-term effect of receiving access to a substantial volume of water on student enrollment. I find that there is a large increase in student enrollment and retainment over time. I see a slight increase in share of girls enrolling in school but only 3 years after treatment; I discuss reasons as to why this could be the case. In Chapter 3, I ask if religious minority groups make different consumption choices based on the religious identity of their neighbors. I use National Sample Survey data to measure these patterns. I find that Muslims who live amongst Hindu neighbors silence their Islamic identity more than those who live amongst Muslim neighbors. I don't see this difference for Hindus.

CHAPTER 1

EVALUATION OF A PLACE-BASED INFRASTRUCTURE POLICY IN INDIA

1.1 Introduction

Governments in developing countries have always invested in public infrastructure, often in the interest of furthering the development of citizens. It is important to understand who benefits from these investments, if at all, and whether the investments are yielding long-term returns. Furthermore, when governments target specific groups or geographic regions for investments, it calls for inquiry into the effects of such projects on the targeted beneficiaries. The literature has shown that if the government can provide public goods at low cost then a place-based policy can improve upon the underprovision of public goods by the private sector (Kline and Moretti [2014a]). Duflo [2001] shows that construction of schools in certain areas (i.e. place-based policy) in Indonesia increased educational attainment as well as wages in the long-term. Adukia [2017] finds that improving sanitation infrastructure in schools in India improves enrollment of girls. Beyond the positive and lasting effect of infrastructure on development outcomes, there is also a literature on the importance of group identity in determining long-term outcomes (Akerlof and Kranton [2000], Shayo [2020]). For example, Ellison and Pathak [2021] find that adopting a place-based policy in Chicago Public Schools that is race-neutral is not as effective as a race-based (i.e. identity-targeted) policy, when it comes to educational attainment. By evaluating a place-based infrastructure policy, that is also identity-targeted (targeted to benefit religious minorities in India), this paper contributes to two both strands of literature.

In 2006, a committee was formed by the Indian goverment to prepare a report on the "Social, Economic and Educational status of the Muslim community in India". This committee created a report (Sachar Commitee Report, 2006) which concluded, among other things, that areas with a high concentration of Muslims were severely underserved by public infrastructure. In response to the report, the government took several steps to correct the widening development gap between Muslims and other groups. One of the responses was to target areas with a high concentration of Muslims and other minority groups for infrastructure investments. They selected areas with at least 25% share of minorities in the population and a below-average performance on development indicators.

This policy sets up a unique natural experiment that is helpful to answer an important question: what are the effects of targeting a particular group in a particular geographic location for state-funded benefits? I utilize plausibly exogenous variation in the timing and location of the treatment to measure the impact of large infrastructure investments in education, health and water access on the medium-term outcomes of the targeted areas. The variation lends itself to evaluation using two separate approaches: an event study design and a fuzzy regression discontinuity design. I use an event study design to measure the medium-term effect of these investments on schooling infrastructure and schooling outcomes. I find that there is an improvement in classroom infrastructure (as targeted by the policy), but this improvement doesn't lead to an improvement in schooling outcomes. In the fuzzy regression discontinuity design, I utilize the threshold at the 25% mark to study the effect of the policy on those who received infrastructure investments as opposed to the ones who just missed out. I find that there is a significant improvement in education infrastructure, which eventually translates into an increase in enrollment.

In Section 1.2, I describe the setting of the paper and the data collection process. In Section 1.3, I present the event study design, along with its assumptions. I also lay out the fuzzy regression discontinuity design with its set of assumptions. All results are presented in Section 1.4 with discussion of key takeaways, their interpretation as well as verification of important assumptions. Section 1.5 concludes and provides suggestions for future work.

1.2 Setting and Data

In response to the committee report, the Ministry of Minority Affairs (MoMA), a federal government agency in India, implemented an infrastructure policy called the "Multi-Sectoral Development Programme". The focus of the policy was on investing in physical infrastructure for education, health and housing - by filling gaps in existing projects or by starting new projects when required. The investment timeline can be divided into two periods, 2008-2012 and 2013-2017. In the first period, 90 out of the 593 districts in India were targeted for investment. They were selected based on certain eligibility criteria. Rule 1 was that the share of minorities in the district population should be 25% or more (for districts with $>= 500,000$ population, it should be 20% or more). Rule 2 was that the district should rank below the national average on basic development indicators [1]. During this period, annual investments were made at the district level in various physical infrastructure. In the second period, instead of targeting districts, the policy targeted census blocks i.e. a group of 20-40 contiguous villages within a district. The two rules remain the same, except they now apply at the block level. Not only does it target blocks in the 90 districts mentioned above, but also in other districts where eligible blocks exist. Since I focus on the effect of this policy on districts with a high share of religious minorities, I restrict my analysis to comparing districts treated in the first time period to districts that never got treated.

In order to determine the timing of treatment for each district between 2008-2011, I make use of signed receipts that show the investment funds allocated to each project in a given district. These receipts are publicly available on the MOMA website. However, they are stored in the form of scanned PDFs that cannot be easily digitized using OCR. This is because each of them have different formatting, and they have not been converted into digitally readable scans. Shen et al. [2021] face a similar issue in their paper and solve it by creating a PDF-reading algorithm called "Layout Parser". Since the volume of my data is

1. Literacy, work force participation, electricity, access to water, and housing condition

much smaller than theirs, I used a combination of manual data entry and a data extraction algorithm to digitize the relevant information from the receipts. Figure 1.1 shows an example of a receipt document. I extract information from the table in the receipt that outlines the type of infrastructure being built by each project and the amount of funds allocated to it. Beyond its use for this paper, this data contributes to the sparse but growing repository of granular, detailed data availble on India's infrastructure expenditure (Asher and Novosad [2020]).

In order to measure schooling infrastructure and enrolment outcomes, I use the Unified District Information System for Education (U-DISE). This is an annual survey of all schools in India which includes, among other things, information on number of classrooms, physical condition of the classrooms, and number of students enrolled. To measure the share of minorities in each district, I use Census of India 2001.

The spellings of district names are different in each dataset that I use because of variations in phonetic translations from local Indian languages into English. Moreover, there are changes in district boundaries as well as district names over time. Using a custom fuzzy-matching algorithm and manual name changes, I create a uniform district-level dataset that includes information on fund allocation, treatment timing, and information on schooling outcomes. I also account for district boundary changes by merging newly formed districts back to their original boundaries from 2001. Therefore, I conduct my analysis using Census of India 2001 district boundaries.

1.3 Research design

The variation in timing of investment lends itself to an event study design. Using this variation, I define an event as the year when a district first receives an investment. 87 districts were treated from 2008-2011, 349 districts were never treated. 33 districts were treated in 2008, 38 in 2009, 12 in 2010 and 4 in 2011.

$$Y_{it} = \gamma_i + \eta_t + \sum_k \beta_k D_{it}^k + \varepsilon_{it} \tag{1.1}$$

Equation 1.1 is a dynamic two-way fixed-effects model that estimates the effect of infrastructure investment for every year t. Y_{it} is the outcome for district i in year t. The main outcomes of my analysis are number of classrooms, classroom condition and enrollment. γ_i denotes the district fixed-effects and η_t denotes year fixed-effects. D_{it}^k takes value 1 if it has been k years since treatment for district i in year t, and 0 otherwise.

The event study design, with the appropriate assumptions, estimates the Average Treatment effect on Treated (ATT) of the investment policy. To get an unbiased estimate of ATT, I assume that districts treated at time t, in the absence of said treatment, would follow parallel trends to un-treated districts at time t, in all time periods after t. I also assume that once a district is treated in time t, it is considered as treated for the rest of the time periods. The setting of this policy circumvents some of the common threats to the parallel trends assumption. The goal of the policy was to supplement existing projects undertaken by the government by filling in gaps where necessary. They also undertook some projects of their own but the primary spending happened on supplementing existing projects. Therefore, it doesn't explicitly advertise itself as a policy that targets certain locations, but still does that implicitly. It also doesn't advertise as targeting minority groups. This feature allows me to shut down the channel of anticipatory effects such as private investments or migration since people will only realise these locations have improved once they start seeing the results of the policy. Also, the kind of infrastructure they are investing in is incremental, it is unlikely to have big demand effects. This second feature allows me to shut down the political economy effects of such a policy because it is not obvious that the policy is targeting religious groups.

However, according to Callaway and Sant'Anna [2021], even if the parallel trends assumption holds in the staggered treatment setting, the estimates from the DTWFE model may be biased if different cohorts (districts treated in different years) have heterogeneous trajectories of treatment effects. In the case of the MOMA policy, this may be the case if, for example, treatment timing and treatment effects both depend on the districts' pre-existing infrastructure access. Then, due to negative weighting issues, the $\hat{\beta}_k$ estimates from the DTWFE regression no longer provide the ATT. Therefore, to derive interpretable estimates under heterogeneous treatment effects, I estimate the event study using the Callaway Sant'Anna estimator, with never-treated districts as a control group when estimating each cohort-year treatment effect. The parallel trends assumption is modified to only allow comparison to never-treated villages. The parallel trends assumption for comparing treated villages to never-treated villages is that in the absence of treatment, the districts first treated in time t and the "never treated" districts would have followed parallel paths in all post-treatment periods.

The eligibility rules also allow me to use an alternate research design, the fuzzy regression discontinuity design. The treatment depends on Rule 1 and Rule 2 eligbility. I measure Rule 1 eligibility using share of religious minority population in a given district. Conditional on being eligible under Rule 2, districts on the right side of 0.25 are eligible and districts on the left side are not eligible. However, due to the policy change in 2013, some ineligible districts were also targeted for investment. Using a FRD allows for the possibility that some of the eligible districts were not treated and some of the ineligible districts were treated. However, for my main analysis, I exclude the districts that were first treated after 2012. Moreover, there is variation in the yearly investments across districts i.e. despite a sharp cutoff, there is variation in treatment intensity on the right side of the cutoff. This also calls for the use of FRD.

In order to construct the running variable and identify the treated districts, I use population data from Census of India, 2001. This allows me to construct x_{it} i.e. share of religious minority population in a given district. I also use this dataset to predict Rule 2 eligibility.

According to policy documentation, they use Census 2001 to determine eligibility on Rule 1 and Rule 2. Therefore, I use Census 2001 to construct the share of religious minority population and assign Rule 1 eligibility. It matches perfectly with the definition in the policy documentation. Next, I construct development indicators that are used to determine Rule 2 eligibility: literacy rate, work force participation, female work force participation, share of households with access to safe drinking water, with electricity and with pucca walls. If a district falls below the national average on any of these indicators, it is eligible under Rule 2. However, I am not able to perfectly match my assignment with the policy documentation. This could be because of lack of clear definitions of safe drinking water, workforce participation and pucca walls. There could also be a measurement error on the part of MOMA. Since I am using a fuzzy design, I use a probabilistic measure of Rule 2 eligibility.

In the first stage, I instrument the investment amount using Rule 1 eligbility. In the second stage, I use the predicted investment amount to measure the effect of the policy on relevant outcomes.

$\forall x_i \in (c-h,\ c+h)$ and satisfies Rule 2,

$$A_{it} = \alpha_0 + \alpha_1 D_i + \alpha_2(x_i - c) + \alpha_3(D_i \times (x_i - c)) + u_{it} \tag{1.2}$$

$$Y_{it} = \beta_0 + \beta_1 D_i + \beta_2(x_i - c) + \beta_3(D_i \times (x_i - c)) + v_{it} \tag{1.3}$$

$$Y_{it} = \theta_0 + \theta_1 \widehat{A_{it}} + \theta_2(x_i - c) + \theta_3(\widehat{A_{it}} \times (x_i - c)) + w_{it} \tag{1.4}$$

Equation 1.2 predicts amount invested (A_{it}) in district i in year t using a local linear sharp RDD. D_i indicates a dummy for Rule 1 eligbility. x_i is the running variable, share of religious minority population in district i. c is the threshold for eligbility of district i i.e. 0.25. h gives the selected bandwidth range. Equation 1.3 estimates the effect of being on the right side of the threshold on relevant outcomes using a local linear sharp RDD. Finally, equation 1.4 gives the 2SLS estimate using $\widehat{A_{it}}$.

Y_{it} is the outcome of interest at the district-year level. $\hat{\theta_1}$ gives the key estimate i.e. the effect of being on the right side of the threshold, within the chosen bandwidth, on the outcome of interest. Optimal bandwidth h is selected by minimizing mean squared error using `rdrobust` package Calonico et al. [2017]. Note that I only use observations after $t >= 2008$ for this analysis.

To get an unbiased estimate of θ_1, I assume that probability of treatment assignment across the threshold in a narrow bandwidth is as-if random, accounting for linear trends. More formally, for treatment status D and running variable x, threshold c, I assume that there is a jump at the threshold in probability of treatment assignment.

$$\lim_{x \downarrow c} \mathbb{P}(D = 1|X = 1) \neq \lim_{x \uparrow c} \mathbb{P}(D = 1|X = c) \tag{1.5}$$

I assume that conditional regression functions, in the absence of treatment, are continuous across the threshold.

$$\lim_{x \downarrow c} E[Y_i(0)|X_i = x] = E[Y_i(0)|X_i = c] \tag{1.6}$$

1.4 Results

Figures 1.2, 1.3 and 1.4 show the event study estimates for dynamic two-way fixed effects and Callaway Sant'Anna. The y-axis denotes the outcome variable i.e. total number of classrooms in the district. The x-axis denotes 4 years before and after treatment. The panel is balanced in event-time i.e. I restrict analysis to districts for which I can observe outcomes 4 years before and after they were treated. This ensures that each event-time estimate is computed using the same sample of districts.

Figure 1.2a shows that the effect of the policy in the year of treatment is close to 0 but imprecise. Total number of classrooms increases by 1200 in a year after treatment, by 1700 in two years after treatment, and by 2200 three years after treatment. The baseline number of classrooms in never-treated districts in 2007 was 9800. Taken together, this means that infrastructure policy increased the number of classrooms in treated districts by 22% on the base of 9800. Figure 1.2b shows the Callaway Sant'Anna estimates for the average effect by the length of exposure to the treatment. In the first year of exposure, there is no detectable effect of the policy. The number of classrooms increase by 1800 when the treated districts have been exposed to the policy for a year. After two years of exposure, the classrooms increase by 2100 and by 3000 after three years of exposure.

Figure 1.3 measures the effect on number of classrooms that are in good condition. Number of classrooms in good condition increase by 1100 a year after treatment, by 1500 two years after treatment, and by 2200 three years after treatment. The baseline number of classrooms in good condition in the control districts before treatment year was 8900. Thus, classroom quality in treated districts improved by 33% after treatment. The Callaway Sant'Anna estimates are similar, but less precise. There is a detectable improvement in classroom quality only two years after exposure to treatment. The number of classrooms in good condition increase by 2100 two years after exposure, and by 2900 three years after exposure.

So far, I show that there is an improvement in schooling infrastructure as a result of this policy. This is an important finding for two reasons. First, from the literature on corruption in public works projects (Olken and Pande [2012]), we know that funds allocated to certain projects do not always get utilized in the intended manner or the quality of the projects is not up to the mark. Thus, when analyzing the effect of an infrastructure policy like the one in this paper, there is a concern that the funds allocated to improve schooling infrastructure did not, in reality, get used to build additional classrooms or that the classrooms built were of a bad quality, for example. I show that despite plausible corruption there is an increase in schooling infrastructure and that the quality of new infrastructure is good. Second, since this policy targets districts that have been under-served by public infrastructure, the relative benefit of additional classrooms is higher than in a place that already has good infrastructure. Now, I check if the improvement in schooling infrastructure leads to changes in actual schooling outcomes i.e. enrollment.

Figure 1.4a shows that there is no detectable effect of the policy on the number of students enrolled in the year of treatment and in the next year after treatment. Two years after treatment, I see an increase in enrollment by 35,000, but that effect is not detectable three years after treatment. The Callaway Sant'Anna estimates, given by Figure 1.4b, also show that there is no detectable effect of the policy on enrollment, regardless of length of exposure to the treatment. The baseline enrollment in the control district, pre-treatment is 290,000. Thus, I see that despite a large point estimate (12% effect), I cannot reject that there is no effect on enrollment. This could be because the pre-existing classroom infrastructure was falling short for the current students and the addition of new classrooms reduced class size for each classroom, instead of attracting more students to the school (Banerjee et al. [2016]). Exploring the possible mechanism behind this result is beyond the scope of this paper due to current data limitations. However, this is an important avenue for future research to understand the co-inciding effects of increase in infrastructure supply and demand.

In an event study design, I cannot verify the parallel trends assumption since it would require me to observe counterfactual trends that do not exist. But, as a robustness check, I can test if parallel trends hold in years before treatment begins. Although this doesn't guarantee that the parallel trends will hold post-treatment, it provides information on differences in trend before treatment. In Figures 1.2a, 1.3a and 1.4a, I show pre-trends for three time periods before treatment for the DTWFE model and I cannot reject parallel trends in the pre-treatment period. To deal with the issue of "selective treatment timing", Figures 1.2b, 1.3b and 1.4b show pre-tests based on aggregated group-time average treatment effects using the Callaway Sant'Anna method. In all pre-periods but one, for all outcome variables, I don't see a significant difference between treatment and control districts.

Figures 1.6, 1.7, and 1.8 present results from the fuzzy RD. The variables displayed on the y-axis are number of classrooms, number of classrooms in good condition and number of students enrolled, respectively. The x-axis shows the running variable i.e share of religious minorities in the district. Since the FRD focuses on districts just above and below the threshold, I restrict the x-axis to a narrow bandwidth from -0.015 to 0.025. Panel (a) shows estimates from Equation 1.2 and panel (b) shows estimates from Equation 1.3, along with their 95% confidence band. Panel (a) of Figure 1.6 shows that there is a large first-stage effect of the policy on amount invested in the district. Panel (b) shows that there is a jump at the threshold in total number of classrooms. Taking a ratio of the first-stage and reduced form estimates, the 2SLS estimate shows that there is an increase in number of classrooms for districts just above the threshold. I see a very similar effect on classroom quality in Figure 1.7 which indicates that the improvement in classroom infrastructure is being driven by high quality classrooms. I also see an increase in enrollment as a result of this infrastructure policy as seen in 1.8.

The FRD results are based on two key assumptions as detailed in Section 1.3: the discontinuity of total investments at the threshold and the continuity of outcomes at the threshold in the absence of treatment. Panel (a) of Figures 1.6, 1.7, and 1.8 verify the first assumption by showing that there is an increase in total investments at the threshold. However, since I cannot observe the counterfactual, I cannot directly verify the continuity assumption. But in Figure 1.9, I present reduced form results for outcome variables in the pre-treatment period i.e. 2005-2007. This allows me to check if, before the treatment started, there was already a jump at the threshold. Even though this does not give me the counterfactual, it is the closest I can get to verifying the continuity assumption. Panel (c), in particular, shows that there was already a jump at the treshold in school enrollment. This warrants further examination to disentangle effects of the policy from pre-existing differences between these districts at the threshold. However, the current sample does not have enough within-state variation to be able to delve into this result. As the results stand, they fail to pass the key falsification test of continuity of pre-treatment outcomes.

In addition to the main tests above, I also conduct two other tests: McCrary test to check if there is bunching of the running variable at the threshold and Placebo test to check if the effect is observed at any other "placebo" threshold. Figure 1.10 presents results from the McCrary test of the running variable. I observe no jump in the density of the running variable at the threshold, in 2001 and in 2014. Figures 1.11, 1.12 and 1.13 present placebo thresholds for all outcomes. Out of all the possible thresholds, we see a positive effect on only a handful of them. The density plots in panel (b) show the distribution of z-statistics across thresholds and verified that the distribution is roughly centered around 0. This indicates that there is no threshold other than 0 where we see the same effects. Taken together, the results and the assumption tests lead me to the conclusion that further analysis at a granular level is required to be able to say that there is a definitive effect of this policy at the threshold.

1.5 Conclusion

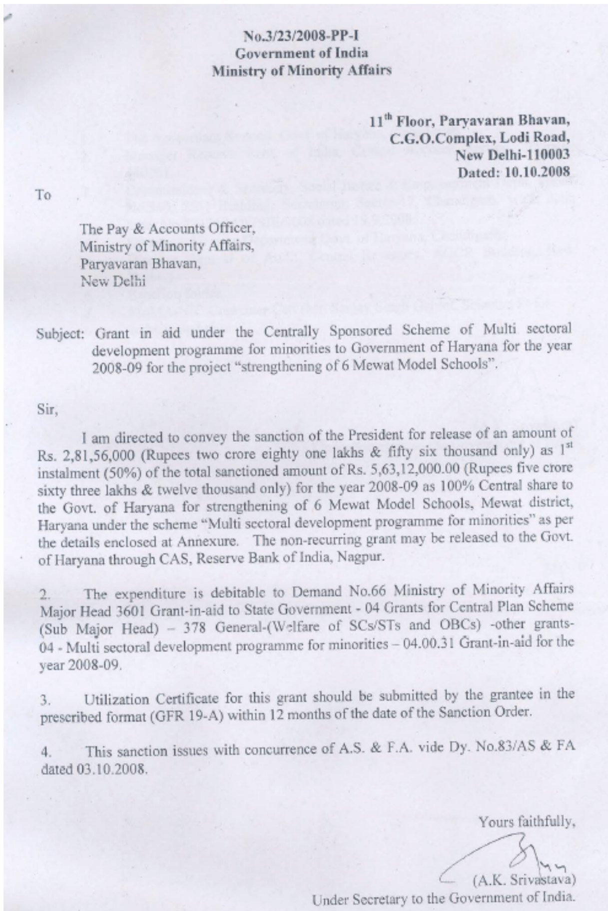
Using an event study design, I show that investing in infrastructure in a given district, in fact, improves infrastructure in that district. If anything, the effect I find is attenuated given that we know that not all the fund allocated to public works projects get utilized as intended. In future work, it will be important to track the investments in these districts and be able to measure the funds being utilized for construction, funds being used for personal use (corruption practices) by presiding officials, funds being misdirected to lower quality infrastructure, and so on. Not only does the investment improve infrastructure but the quality of the infrastructure being built is good. However, I don't see an effect of this policy on enrollment of students. This is an important avenue for future work to examine general equilibrium effects of supply-side interventions. It remains to be seen how the demand-side forces interact with the increase in supply of infrastructure. As we know, an actual improvement in educational outcomes depends a lot on teacher availability, student-teacher ratio, class size, etc.

In future work, I also plan to collect more granular data on the block-level investments as part of this policy and hope to have more definitive results using the fuzzy RDD.

The main takeaway of this paper is that if governments intentionally direct funds to places that have been neglected in the past, it can improve infrastructure in those places significantly. Specifically, if marginalized groups (such as religious minorities) systematically co-locate close to each other and are the receiving end of the neglect, even small improvements in infrastructure could have lasting impacts.

1.6 Figures

Figure 1.1: Examples of fund allocation receipts



No.3/23/2008-PP-I
Government of India
Ministry of Minority Affairs

11th Floor, Paryavaran Bhavan,
C.G.O.Complex, Lodi Road,
New Delhi-110003
Dated: 10.10.2008

To

The Pay & Accounts Officer,
Ministry of Minority Affairs,
Paryavaran Bhavan,
New Delhi

Subject: Grant in aid under the Centrally Sponsored Scheme of Multi sectoral development programme for minorities to Government of Haryana for the year 2008-09 for the project "strengthening of 6 Mewat Model Schools".

Sir,

I am directed to convey the sanction of the President for release of an amount of Rs. 2,81,56,000 (Rupees two crore eighty one lakhs & fifty six thousand only) as 1st instalment (50%) of the total sanctioned amount of Rs. 5,63,12,000.00 (Rupees five crore sixty three lakhs & twelve thousand only) for the year 2008-09 as 100% Central share to the Govt. of Haryana for strengthening of 6 Mewat Model Schools, Mewat district, Haryana under the scheme "Multi sectoral development programme for minorities" as per the details enclosed at Annexure. The non-recurring grant may be released to the Govt. of Haryana through CAS, Reserve Bank of India, Nagpur.

2. The expenditure is debitable to Demand No.66 Ministry of Minority Affairs Major Head 3601 Grant-in-aid to State Government - 04 Grants for Central Plan Scheme (Sub Major Head) – 378 General-(Welfare of SCs/STs and OBCs) -other grants-04 - Multi sectoral development programme for minorities – 04.00.31 Grant-in-aid for the year 2008-09.

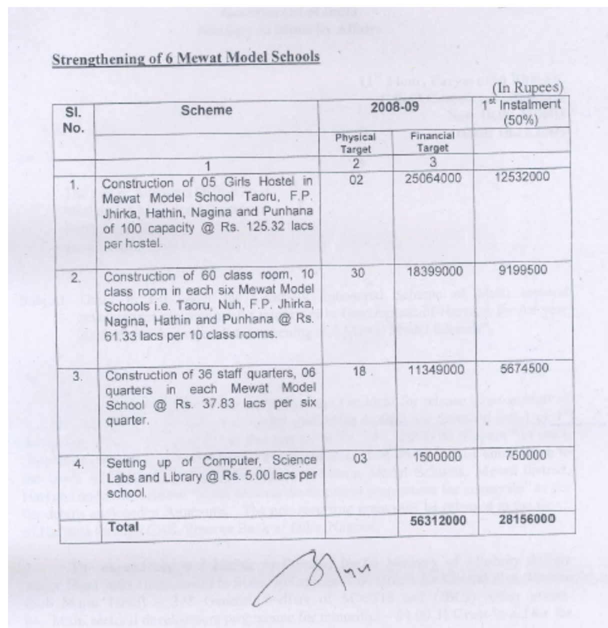
3. Utilization Certificate for this grant should be submitted by the grantee in the prescribed format (GFR 19-A) within 12 months of the date of the Sanction Order.

4. This sanction issues with concurrence of A.S. & F.A. vide Dy. No.83/AS & FA dated 03.10.2008.

Yours faithfully,

(A.K. Srivastava)
Under Secretary to the Government of India.

(a) Image of fund allocation letter



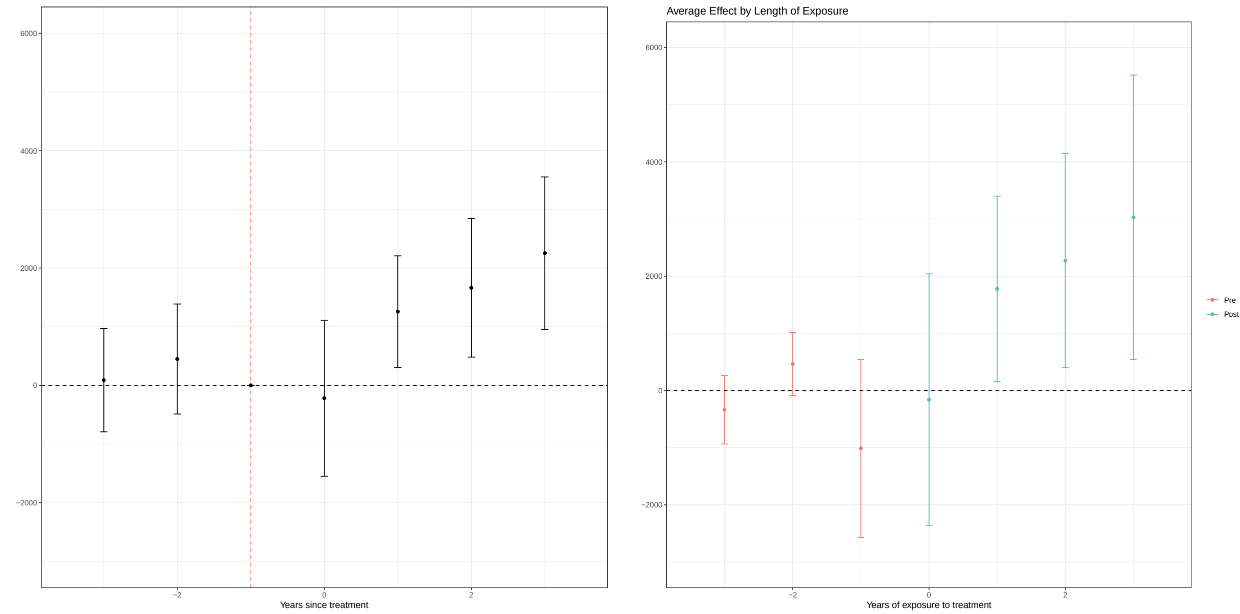
Strengthening of 6 Mewat Model Schools

(In Rupees)

Sl. No.	Scheme	2008-09		1st Instalment (50%)
		Physical Target	Financial Target	
	1	2	3	
1.	Construction of 05 Girls Hostel in Mewat Model School Taoru, F.P. Jhirka, Hathin, Nagina and Punhana of 100 capacity @ Rs. 125.32 lacs per hostel.	02	25064000	12532000
2.	Construction of 60 class room, 10 class room in each six Mewat Model Schools i.e. Taoru, Nuh, F.P. Jhirka, Nagina, Hathin and Punhana @ Rs. 61.33 lacs per 10 class rooms.	30	18399000	9199500
3.	Construction of 36 staff quarters, 06 quarters in each Mewat Model School @ Rs. 37.83 lacs per six quarter.	18	11349000	5674500
4.	Setting up of Computer, Science Labs and Library @ Rs. 5.00 lacs per school.	03	1500000	750000
	Total		**56312000**	**28156000**

(b) Image of fund allocation letter

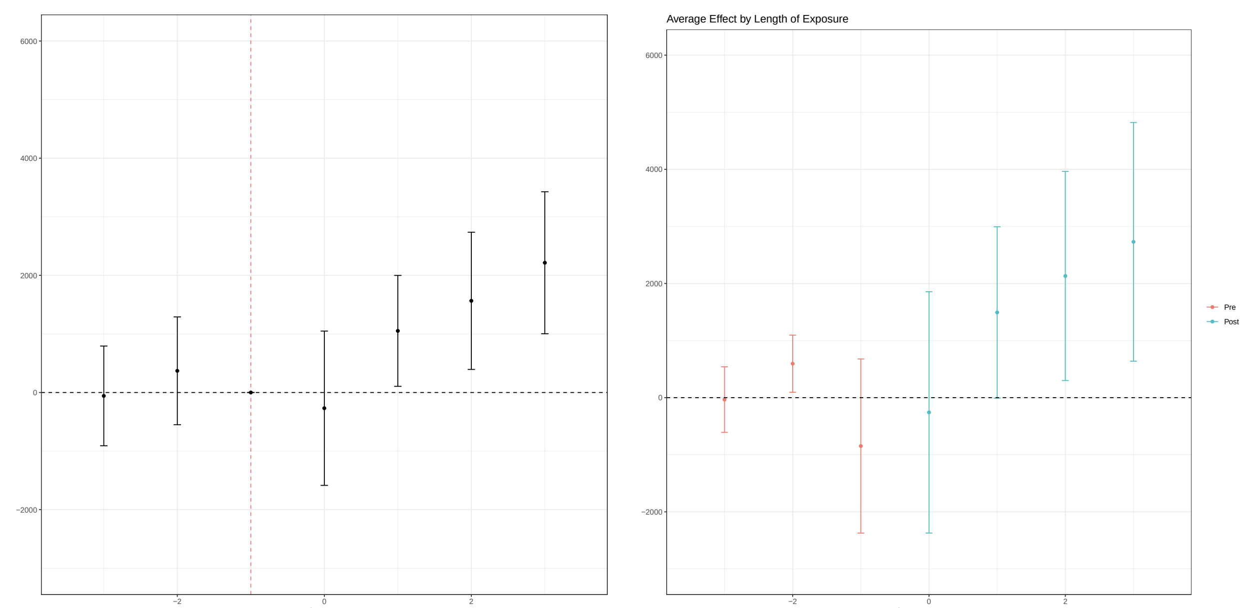
Figure 1.2: Number of classrooms



(a) Dynamic two-way fixed effects

(b) Callaway Sant'Anna

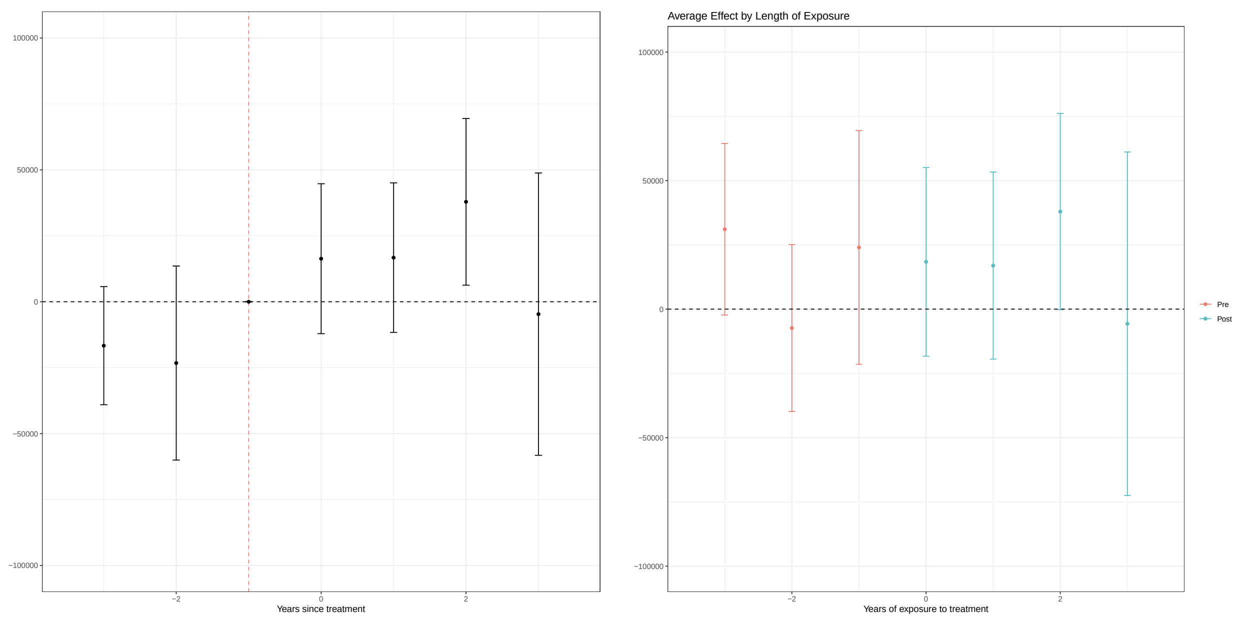
Figure 1.3: Classroom quality



(a) Dynamic two-way fixed effects

(b) Callaway Sant'Anna

Figure 1.4: Enrollment



(a) Dynamic two-way fixed effects

(b) Callaway Sant'Anna

Figure 1.5: Variation in investments by category

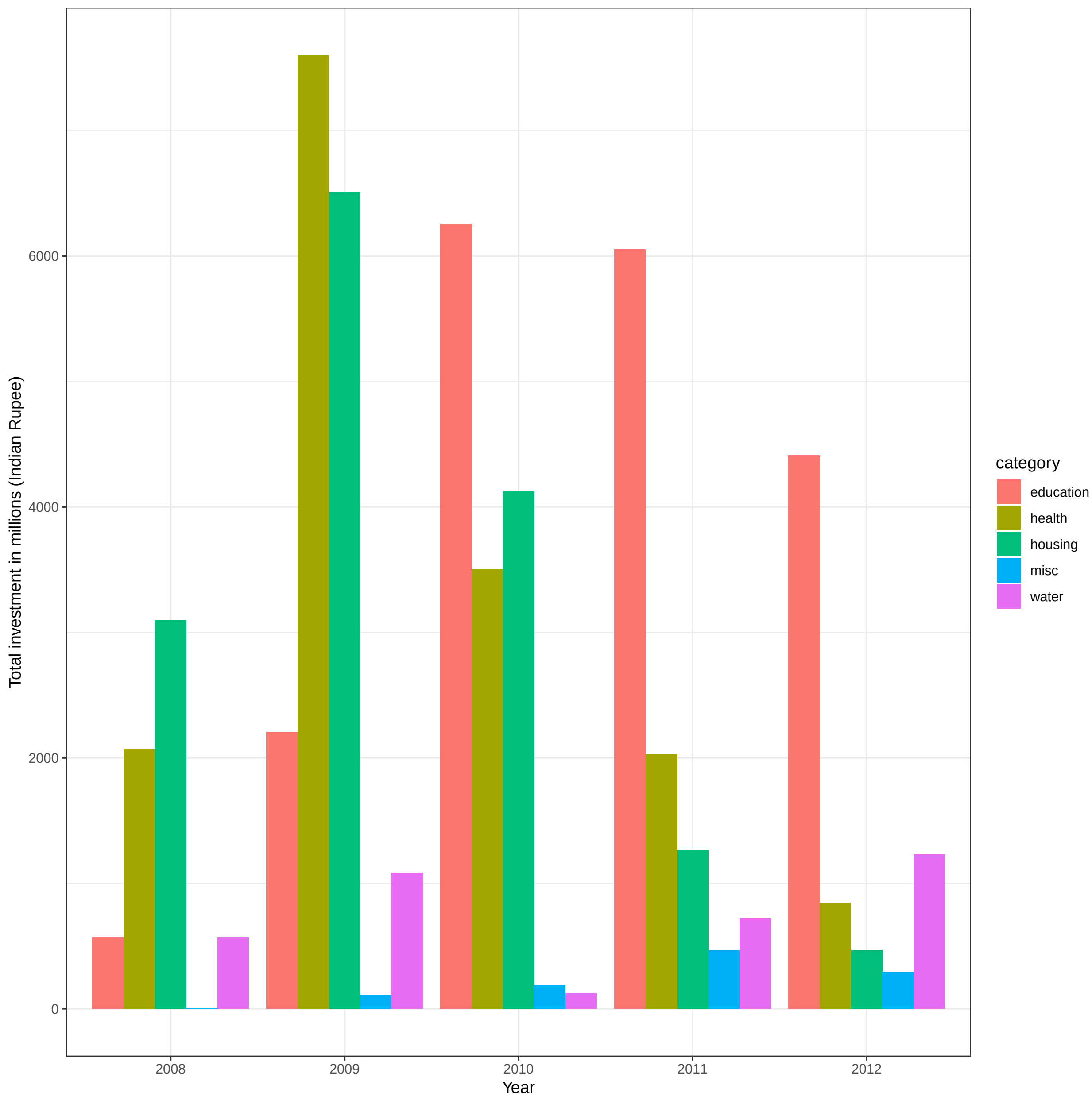
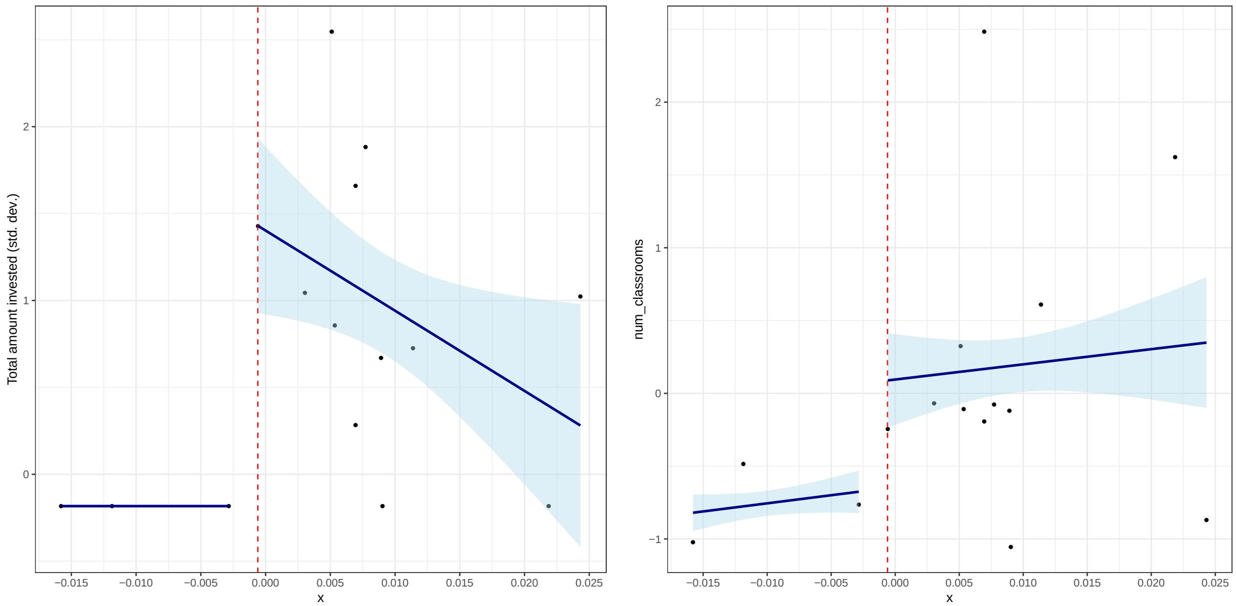


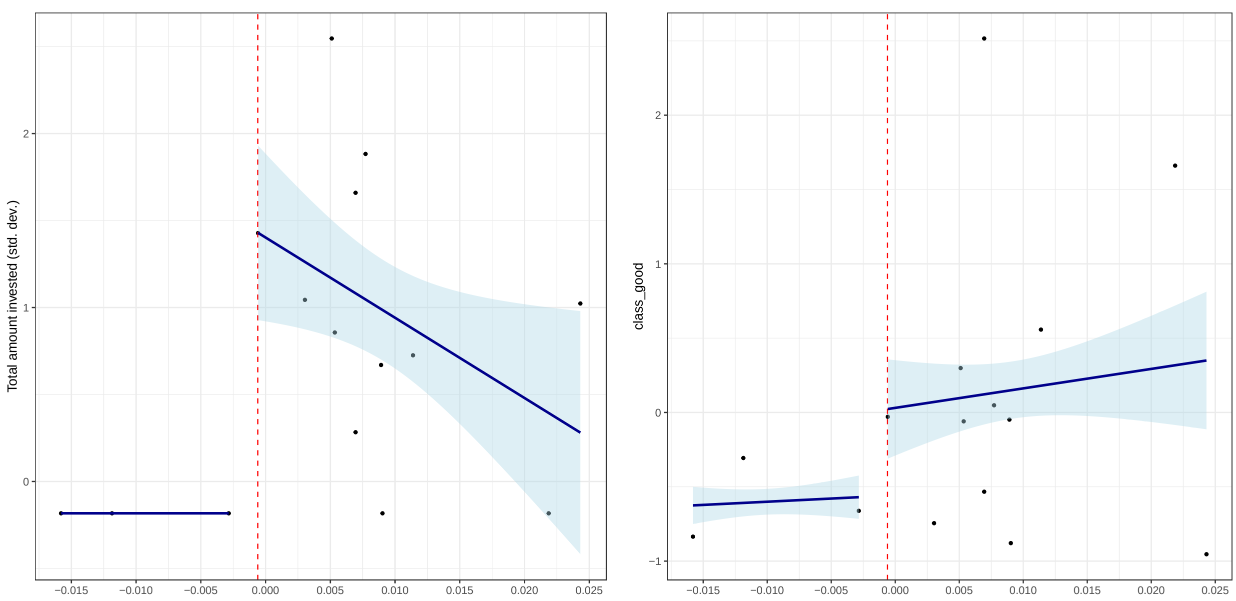
Figure 1.6: Number of classrooms



(a) First stage

(b) Reduced form

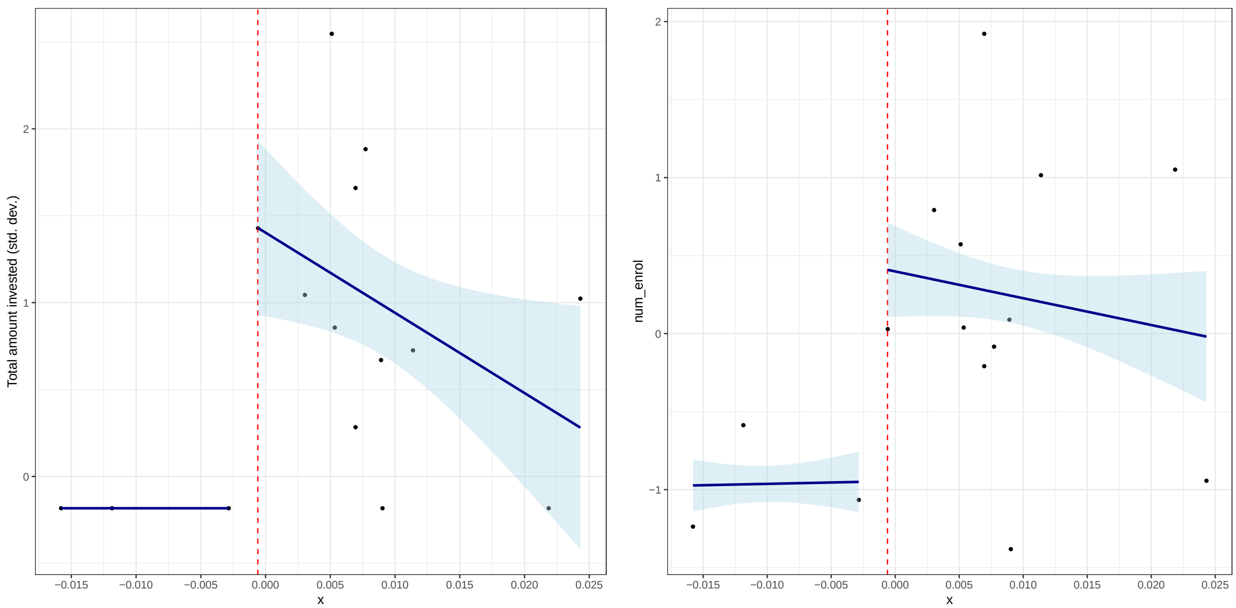
Figure 1.7: Classroom quality



(a) First stage

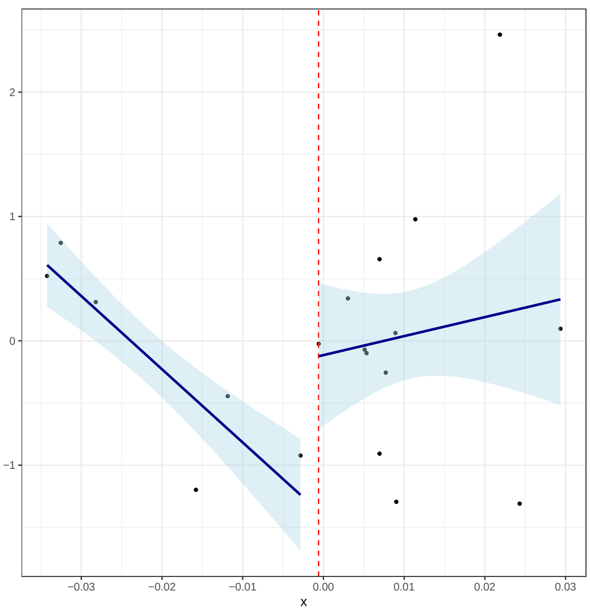
(b) Reduced form

Figure 1.8: Enrollment

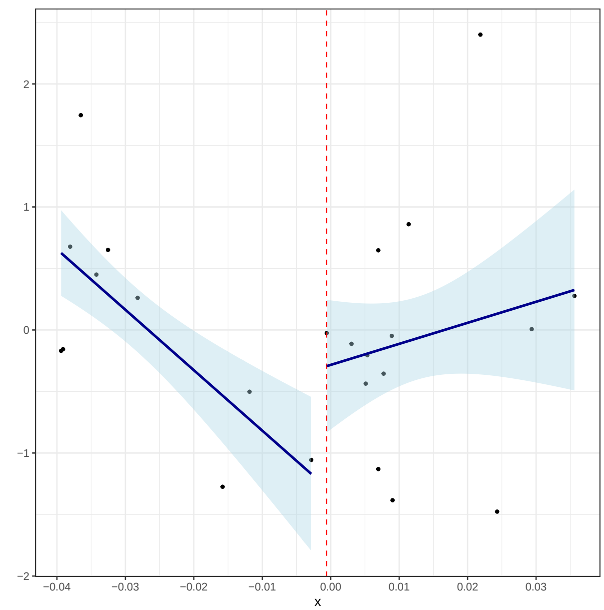


(a) First stage (b) Reduced form

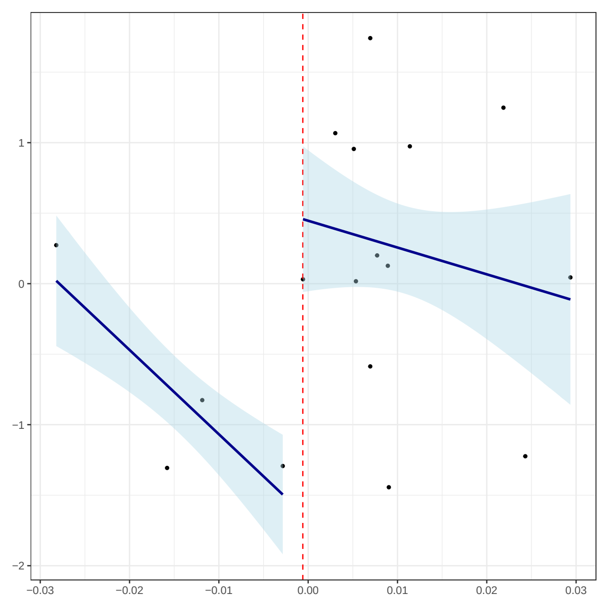
Figure 1.9: Outcomes in pre-treatment period



(a) Number of classrooms

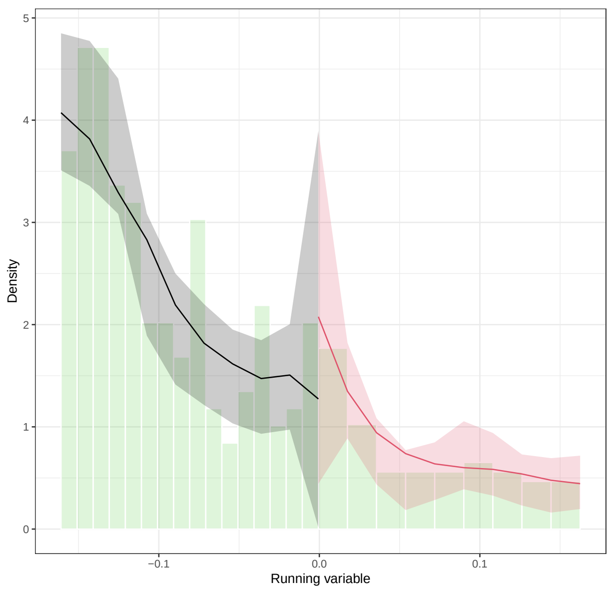


(b) Classroom quality

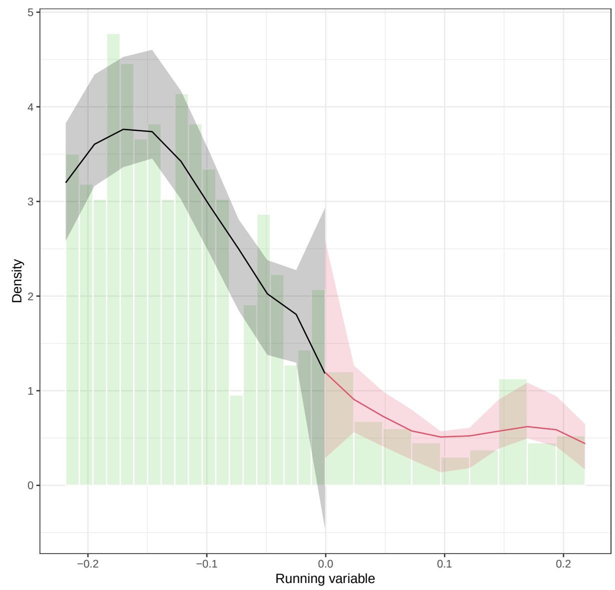


(c) Enrollment

Figure 1.10: McCrary Test (density of running variable)

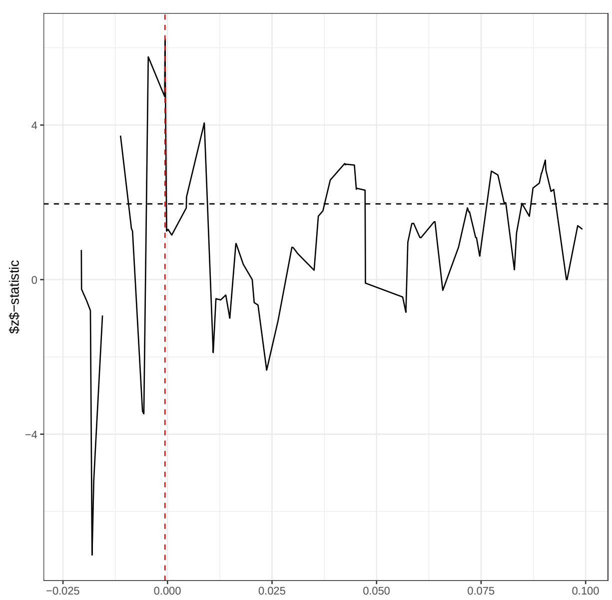


(a) Census 2001 (policy occured in 2008)

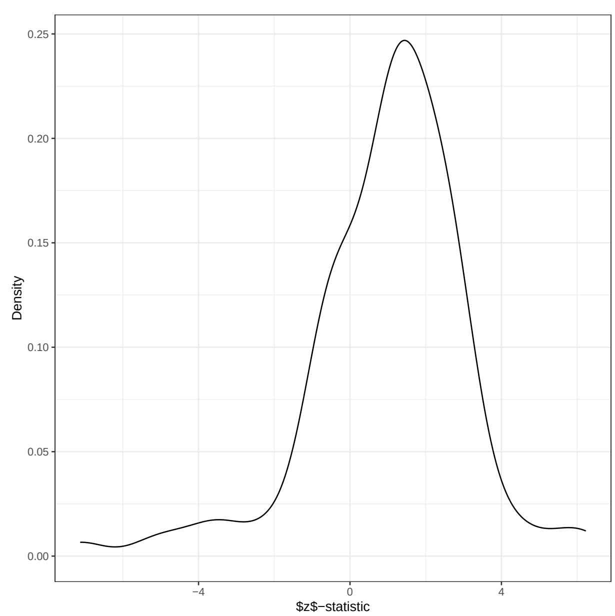


(b) NSS 2014 (policy occured in 2008)

Figure 1.11: Placebo test: Number of classrooms

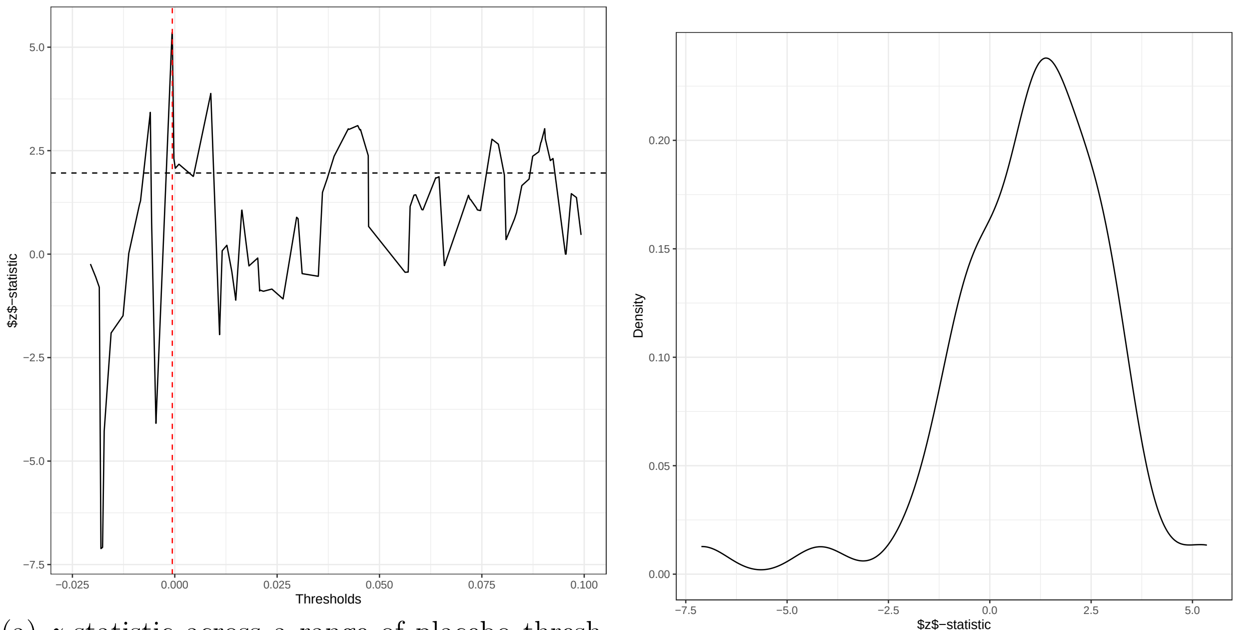


(a) z-statistic across a range of placebo thresholds



(b) Density plot of z-statistics

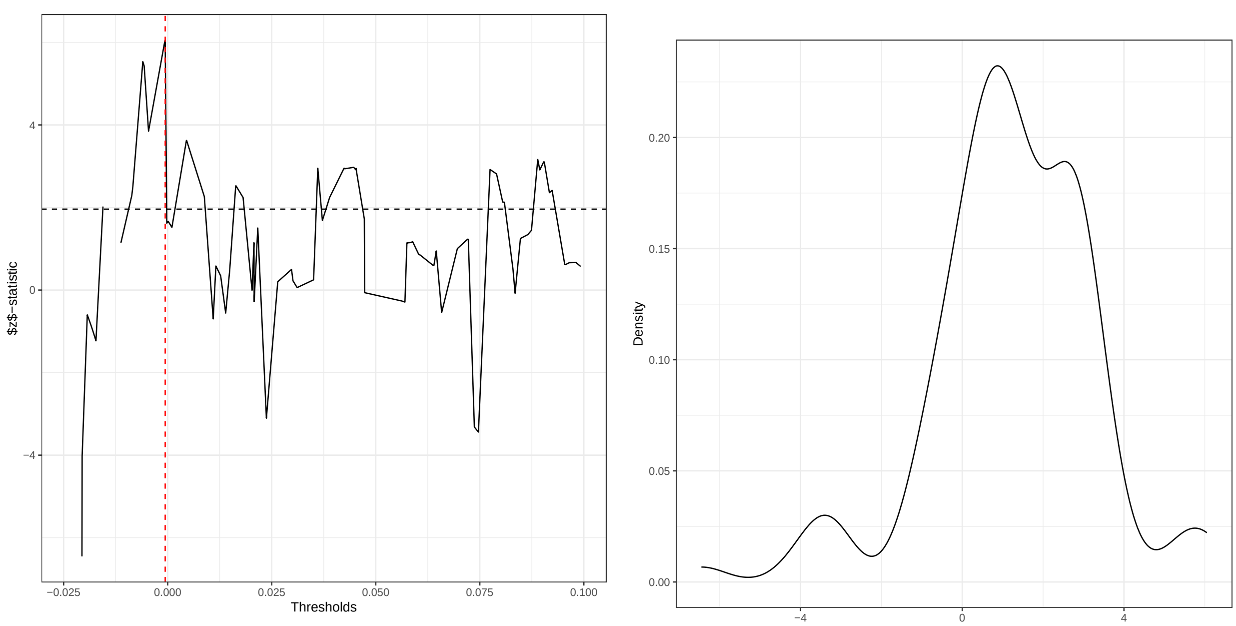
Figure 1.12: Placebo test: Classroom quality



(a) z-statistic across a range of placebo thresholds

(b) Density plot of z-statistics

Figure 1.13: Placebo test: Enrollment



(a) z-statistic across a range of placebo thresholds

(b) Density plot of z-statistics

CHAPTER 2

EFFECTS OF IMPROVED HOUSEHOLD WATER ACCESS (BY CLAIRE FAN AND DEVIKA LAKHOTE)

2.1 Introduction

As aquifers are en route to depletion and climate change is expected to exacerbate water scarcity in many parts of the world, applied microeconomists have directed a sizable amount of attention to the welfare consequences of access to and allocation of water for agriculture (e.g. Duflo and Pande [2007], Hornbeck and Keskin [2014], Sekhri [2014], Blakeslee et al. [2020], Ryan and Sudarshan [2022]). But equally jeopardized, and arguably even more important than irrigation, is access to adequate and clean water for drinking and household needs, which 2 billion people currently lack (World Bank [2021]).

The consequences of changes to household water access (hereon referred to as, HWA) are borne disproportionately by women. In India (the situation is similar or worse in much of sub-Saharan Africa), in cases where water is not piped into the home, it is women and girls who fetch it from outside 80% of the time, spending on average 35 minutes per day, equivalent to the loss of 27 days over a year (UNICEF). Accordingly, drinking water is a much more salient issue to women than to men (Chattopadhyay and Duflo [2004]). In fact, decreased HWA may be costly to women even if the cause is water competition from irrigation itself (Karim et al. [2012]). Agriculture and industry stake large politically and financially backed claims to water supplies, and policymakers must make increasingly constrained water allocation decisions between places and industries. Thus, it is crucial that our accounting of the costs and benefits of various kinds of water use reflect those borne by both men and women, those with and without political and intra-household bargaining power.

Existing causal evidence has shown that HWA can have important consequences for health and overall welfare. It has begun to establish the relevance of girls' schooling, women's time use, and market participation as further outcomes(Devoto et al. [2012], Ashraf et al. [2021]). However, these studies have largely been confined to local urban RCTs or historical settings, with uncertain generalizability to the people across continents, especially in rural areas, facing water access challenges today.[1]

Our analysis focuses on the breadth of rural households in India - where the scarcity of water infrastructure continues to be a pressing concern. Scarcity of water infrastructure, in turn, affects household access to clean, potable water. This has direct effects on health. Several demand-side solutions have been created to address the symptoms of low HWA - such as deworming, bottled water, water purifying kits, etc. (Ahuja et al. [2010]). In this paper, we focus our attention on the supply-side problem - lack of water access points - that the government of India is trying to address at a large-scale through the nationwide roll-out of piped drinking water, tubewells, handpumps and open wells.

We use village-level HWA data from the entirety of rural India to provide large-scale, spatially granular causal evidence on the familiar first-order questions of the effects of HWA infrastructure on girls' schooling. We do this by leveraging timing of treatment in an event study design. We find that improved HWA has positive and significant effects on children's enrollment in school as well as year-end exam attendance. However, contrary to our hypothesis, we see a very small differential effect on girls' outcomes. We discuss reasons as to why that may be the case.

Section 2.2 describes the setting of our study, the data construction process, and outcomes of interest. Section 2.3 describes the event study design we employ to estimate the main result of our paper. Section 2.4 discusses the results and our interpretation of the results. Section 2.5 concludes and provides direction for future work.

1. The exception is Gamper-Rabindran et al. [2010], which looks at the effect of piped water rollout across Brazil on infant mortality.

2.2 Setting and Data

We have acquired India's Integrated Management Information System (IMIS) database, which contains the universe of government-funded rural water access points, including pumps, wells, and piped water services, nation-wide. For each of these water access points, the database contains its location at the sub-village level, dates of commissioning and construction, water source, population covered, liters per capita per day (LPCS) provided, contamination status, and more. IMIS is likely unparalleled among all countries, including developed ones, for providing a comprehensive database of local water access systems in a country (Wescoat et al., 2016). However, it has not (to our knowledge) previously been used at scale in social science research, likely because the database structure makes it costly to download large portions of the data. We suspect that a chief reason for the relative paucity of large-scale empirical evidence on supply-side effects of HWA is the difficulty of obtaining large-scale access points data. By taking on the one-time cost, we hope to open up unique research opportunities beyond this project as well.

Figure 2.1 shows an example of the webpage that stores the IMIS data. The scraping procedure involved two main steps: building a point-and-click algorithm to click through the webpages and arrive at the page for each access point, and extracting relevant information from that page. First, the algorithm clicks through several administrative levels (state, district, sub-district, gram panchayat, village) to get to a list of habitations. A habitation is defined as the catchment area within a village where an access point is located. A habitation can have one or more access points. Each habitation page has a unique URL that we store so that we can access each habitation page independently. Then, the algorithm opens up each habitation page and stores a list of unique access point URLs. Each access point has its own webpage with information the construction date, funds required for construction, volume of water to be supplied by the access point, total target population for the access point, and so on. Using a combination of scraping techniques, we extract this information for each access point and save it in an accessible format for further analysis.

Since the 1970s, India has been gradually expanding rural HWA, and this intertemporal variation is captured by IMIS. Each water access point provides certain liters per capita per day (LPCD) of water to population covered by that access point. As a summary measure of HWA, for each village i and year t with set of access points S_{it}, we calculate:

$$\mathrm{HWA}_{it} = \frac{\sum_{s \in S_{it}} (\mathrm{p}_s \times \mathrm{LPCD}_s)}{\mathrm{P}_i} \tag{2.1}$$

The unit of measurement for HWA is LPCD. p_s gives the population covered by access point s. This is an estimate provided by the government as part of the database. We take it at face value, i.e. we cannot determine if this is the actual number of people in the village who use this access point. Instead, we think of it as the intended coverage of the access point. $LPCD_s$ is the liters per capita per day that can be provided by access point s. Again, this is an estimate provided by the government based on water availability, engineering calculations, etc. and we take it as given. We calculate the total liters per day provided to village i by access point s by multiplying p_s with $LPCD_s$. We calculate this quantity for every access point available to the village and sum over all access points. Note that a single village can have multiple access points that cover different parts of the village as well as different sub-populations. To get our final HWA_{it}, we divide the total liters per day provided to the village by the number of people who live in the village. This gives us an average of the volume of water each person in the village has access to, daily.

To measure outcomes, we use schooling outcomes data from the annual survey of schools in India (U-DISE)[2]. The survey is conducted at the school-level, we aggregate it to the village-level and match it to our water access points dataset using a fuzzy matching algorithm. We use student enrollment as our primary outcome, and in addition, we also look at share of girls' in total enrollment and year-end exam attendance.

2. Data provided by Prof. Marianne Betrand

Therefore, our final data for analysis is unique at the village-year level and contains information on annual HWA and schooling outcomes for each village in our sample. Table 2.1 present summary statistics of our sample. It contains 150,060 villages between 2005-2017 across 27 states in India. An average village in the sample has a population of 1792, has about 368 households, 396 students enrolled in school and 177 LPCD of water supply.

2.3 Research Design

Since this is a nationwide policy roll-out with a sizeable amount of funding being directed to it, we want to estimate the effect of this rollout on certain outcomes of interest, the main one being student enrollment in schools. We hypothesize that an increase in HWA through this policy will, in turn, increase the number of students enrolling in school through multiple channels that we shall discuss later in the paper.

In order to do this, we construct a measure of household water access (HWA) that captures the liters per capita per day (LPCD) of water that is intended to be supplied to a given village through the policy roll-out. In an ideal setting, a random assignment of access points to each village would allow us to get an unbiased estimate the effect of HWA on enrollment. However, that is not a feasible strategy in our setting because the roll-out across the country is not random: allocation of federal funding between states is determined partly by existing access points, while funding allocation within states does not always follow any set criteria (e.g. in Odisha).

In lieu of an ideal experiment, our strategy is to use an event study design. We construct a binary and absorbing treatment status, defined as 1 when the combination of all HWA infrastructure recorded for a village has total water-giving capacity of at least 40 LPCD, which the Indian government stated as its first-order goal when launching the Accelerated Rural Water Supply Programme in 1972. The government considered 40 LPCD to be the minimum amount of water needed to complete basic daily chores in addition to drinking and eating. For context, the WHO names 50 LPCD as the water needed "to ensure that most basic needs are met while keeping public health risks at a low level" (UN Water), so from the perspective of development policymakers and organizations, the marginal benefits of reaching 40 LPCD should be sizable. Further, Figure 2.2b shows that this treatment definition induces substantial variation in treatment status and timing across villages. The 75th percentile village has already reached the 40 LPCD goal at the beginning of our analysis period. The median village reaches the goal in 2011. The 25th percentile village does not reach the goal by the end of the analysis period in 2017.

We can estimate the event study using a dynamic two-way fixed effects (DTWFE) model,

$$Y_{it} = \gamma_i + \eta_t + \sum_k \beta_k D_{it}^k + \varepsilon_{it} \tag{2.2}$$

In the above specification, Y_{it} is the outcome of interest for village i in year t. γ_i and η_t are the village and year fixed-effects, respectively. D_{it}^k is dummy that takes value 1 if, village i first reached 40 LPCD in year $t-k$, and 0 otherwise. The parameters of interest, β_k, are meant to be interpreted as the ATT k years after first receiving 40 LPCD. The identifying assumption here is that villages reaching 40 LPCD at different times (or not at all) would have parallel trends in Y in the counterfactual where no village is ever treated. More precisely, define $G_i = g$ if village i belongs to the cohort of villages first treated in year g, and denote $Y_{i,t}(g)$ as the potential outcome for i in year t if $G_i = g$. Then, using notation from Roth et al. [2023], we are assuming that for any village i and years t, $t' \neq t$,

$$E[Y_{i,t}(\infty) - Y_{i,t'}(\infty)|G_i = g] = E[Y_{i,t}(\infty) - Y_{i,t'}(\infty)|G_i = g'] \tag{2.3}$$

However, according to Callaway and Sant'Anna [2021], even if the parallel trends assumption holds in the staggered treatment setting, the estimates from the DTWFE model may be biased if different cohorts (villages treated in different years) have heterogeneous trajectories of treatment effects. In the case of our sample, this may be the case if, for example, treatment timing and treatment effects both depend on villages' pre-existing water access. Then, due to negative weighting issues, the $\hat{\beta_k}$ estimates from the DTWFE regression no longer provide the ATT. Therefore, to derive interpretable estimates under heterogeneous treatment effects, we estimate the event studies using the Callaway Sant'Anna estimator, with a combination of never-treated villages and not-yet-treated villages as control groups when estimating each cohort-year treatment effect. The parallel trends assumption is modified to only allow comparison to never-treated and not-yet-treated villages. If C denotes the never-treated group, then the parallel trends assumption for comparing treated villages to never-treated villages is as follows,

$$E[Y_t(0) - Y_{t-1}(0)|G = g] = E[Y_t(0) - Y_{t-1}(0)|C = 1] \tag{2.4}$$

To compare to the not-yet-treated villages, the parallel trends assumption is as follows,
For all $g = 2, \ldots, \tau, s, t = 2, \ldots, \tau$ with $t \geq g$ and $s \geq t$,

$$E[Y_t(0) - Y_{t-1}(0)|G = g] = E[Y_t(0) - Y_{t-1}(0)|D_s = 0, G \neq g] \tag{2.5}$$

The Callaway Sant'Anna estimates are our preferred estimates and our results present an aggregate of those cohort-wise treatment effects.

A common threat to the parallel trends assumption is events other than the treatment coninciding with the treatment. We examine treatment timing in Figure 2.2a. 30% of the villages do not reach the 40 LPCD goal by the end of the analysis period in 2017. 30% of the villages have already reached the goal by 2005. The remaining 40% of the villages reach the goal, almost evenly between 2005 and 2017. There is no spike in any given year. The lack of bunching of treatment in any particular year reduces the likelihood of treatment coinciding with other policies that could affect our outcomes.

2.4 Results

Improving household water access is one of the Sustainable Development Goals. It is commonly associated with improved health of all household members consuming water, as well as savings of time and physical energy for those who would otherwise collect water from farther away, which can now be devoted to other productive tasks such as education and labor force participation. Accordingly, developing countries around the world are devoting large sums of money into household water infrastructure. Moreover, given that women and girls bear the brunt of the burden of collecting water from outside the home, a common thread of discourse among policymakers and development organizations is that HWA is critical to promoting not only general development, but also gender equity outcomes, specifically.

Our main outcome of interest is enrollment in school and taking year-end exams, for girls and boys. Since we expect improved HWA to reduce the time that girls would spend on fetching water, we hypothesize that they would spend that time on things that are more useful for them. We can think of change in school enrollment as one of the ways to measure time substitution between water-fetching and schooling. We provide evidence from 60% of rural India about whether rural HWA infrastructure actually improves development outcomes. Under the assumption of parallel trends between villages that reach 40 LPCD at different times, our event study results indicate that improving HWA does have a persistent positive effect on school enrollment, suggesting that reduced water collection duties could be freeing up time for children to attend school.

In Figures 2.3, 2.4 and 2.5, we present estimates for our main outcomes on the y-axis: total enrollment, total exam-takers, and share of girls enrolled. The x-axis shows leads and lags from the year of treatment year. We use a panel balanced in event-time i.e. we use the same sample and measure the effect of treatment upto 4 years before and after treatment.

Panel (a) of Figure 2.3 shows the estimates from the dynamic two-way fixed effects model. There is an increase in enrollment by 33 in the year of treatment. A year after treatment, enrollment increases by 13, by 15 two years after treatment, by 22 three years after treatment, and by 25 four years after treatment. The Callaway Sant'Anna group-time averages presented in panel (b) of Figure 2.3 are smaller and less precise but trend similarly. There is a increase in enrollment of 23 in the year of treatment, and a consistent increase of approximately 10 every year after treatment. The baseline enrollment in a control village pre-treatment is 350. This implies that improved water infrastructure has a positive, significant and lasting effect of 6.5% on the number of students enrolling in school.

To measure actual attendance beyond enrollment, we look at the effects on the numbers of students taking end-of-year exams from grades 5-8. This allows us to measure if the enrolled students continue to stay in school and progress. Panel (a) of Figure 2.4 shows that the number of students who take end-of-year exams increases significantly in year of treatment by 2 students. It continues to increase by 2.5 a year or two after treatment, by 4 students three years after treatment, and by 6 students four years after treatment. In panel (b), we see muted effects on exam-takers. There is an increase of 1 in the year of treatment and of 2.5 a year after treatment. However, the effects are imprecise in the following years. The increase immediately after treatment is an increase of 11% on the base of 21.

In panels (a) and (b) of Figure 2.5, we see that the share of female students enrolled in the school does not change immediately after treatment. But there is a slight increase of 0.002 3-4 years after treatment. The baseline share of girls in total students enrolled is 0.49. Thus, the change in share is only an increase of 0.4%. Given that the baseline share is already close to equal and there is a lack of a large effect, we infer that enrollment is perhaps not the margin on which HWA has gendered effects. We interpret these results cautiously, however, as nominal enrollment is not the same as daily school attendance, studying, and learning.

If taken at face value, our results indicate that HWA investments do not generate the educational returns commonly expected by policymakers and development organizations, either overall or for girls specifically. A few explanations for this underwhelming conclusion are possible and deserve further study. One is that water quantity may not be so beneficial if water quality is not ensured. Another possibility is that HWA infrastructure does not remain functional for long, due to seasonal or permanent water depletion, or lack of maintenance. Although infrastructure construction is primarily funded by central and state governments, operation and maintenance are up to local panchayats, which may face budget constraints or expertise limitations, or may not prioritize the maintenance of infrastructure that do not benefit men as much as women. Moreover, without regular monitoring of existing HWA infrastructure, it is possible that state governments looking to fulfill nominal water availability targets could lack incentives to invest in the infrastructure's potential lifespan or ease of maintenance. Although India's IMIS database is already unparalleled for granularity and depth of rural HWA information, these two classes of possibilities point to the potential need for even more data and continuous monitoring (e.g. of water quality and continued water levels and operational status) if we are to fully realize the benefits of HWA infrastructure.

A third possible explanation for our findings so far is that at the village level, rollout of HWA coincides with improved water access in the village's schools, and students from neighboring villages not yet receiving HWA enroll at the treated villages' schools instead. If neighboring villages receiving HWA at different times differ in levels of development (but not in time trends absent treatment, per the event study assumption), this could explain both the positive effect on enrollment numbers and the attenuated effect on learning outcomes. To explore this possibility, in the future, we will compare HWA rollout with the rollout of water access in schools, as recorded in the DISE panel data.

2.5 Conclusion and next steps

We conclude that, given our assumptions, we see a positive, significant and persistent effect of HWA on schooling outcomes for children. We do not see any difference by gender.

Finally, our event study analysis thus far provides causal treatment effect estimates under the assumption of parallel trends between villages which do and do not reach 40 LPCD, and between villages reaching the target at different times. It is quite possible that this assumption does not in fact hold.

Our highest-priority next step, therefore, is to complement our current approach with an IV approach using exogenous geology and state policy timing variation. We plan to leverage variation in state-level water policies interacted with village-level geology to predict rollout within the state. Specifically, under the following assumptions:

- Exclusion restriction: Geology affects Y only through the timing of D.
- Parallel trends with staggered instrumented treatment: Potential ΔY between any two years, if a village were predicted to be never treated in our sample time frame, does not depend on when it is actually predicted to be treated.

we can estimate with 2SLS:

$$Y_{ist} = \alpha_i + \delta_t + \beta D_{ist} + \gamma \kappa_{st} + \varepsilon_{ist}$$

$$D_{ist} = \alpha_i + \delta_t + \theta(\kappa_{st} \times G_{is}) + \gamma \kappa_{st} + \varepsilon_{ist}$$

where κ_{st} equal 1 if state s had articulated a 40 lpcd target by year t, and zero otherwise. The instrument is the interaction between state policy and geology[3].

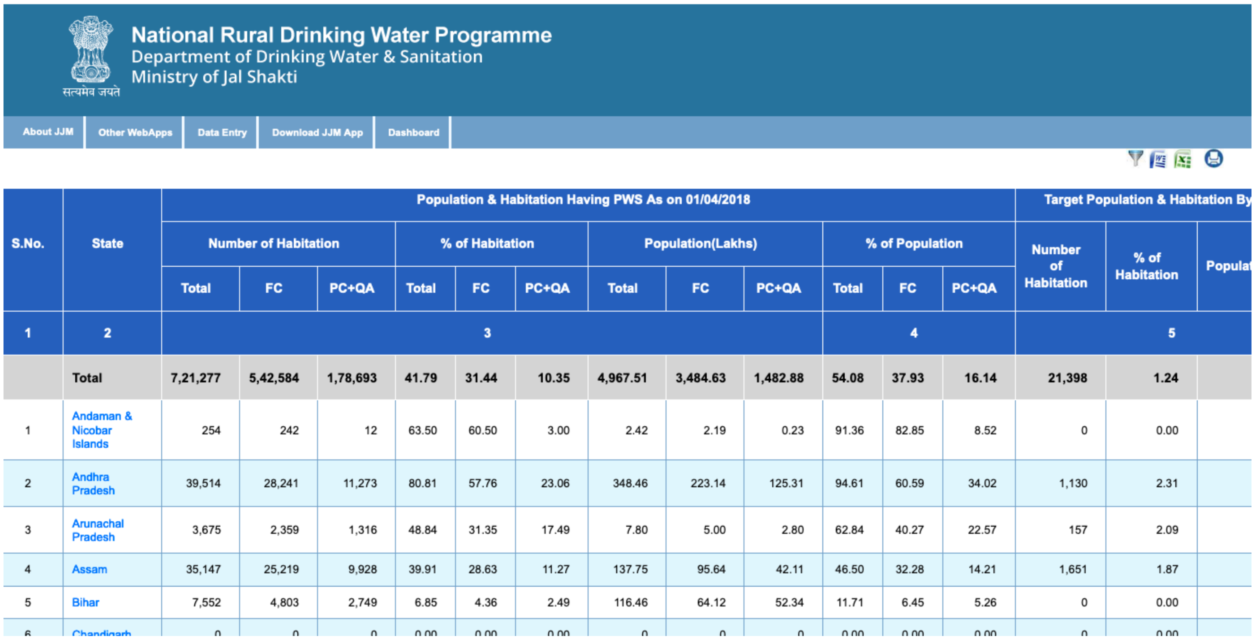
The key takeaway from our analysis is that getting access to a substantial amount of water everyday has immediate and large effects on student enrollment and retainment. This could be because they are spending less time fetching/boiling/collecting water and have more time for school. It could also be that their overall health improves and allows them to

3. We do not need to include geology as its own term because it is absorbed in the unit fixed effect. However, we do need to include the state policy status as its own term. A state having the target could be correlated with other development policies that affect Y not through the geology and HWA channel.

stay in school consistently. In future work, we hope to provide precise estimates on these mechanisms due to current limitations of the health and time-use panels in India. In terms of a differential impact on girls, we do see a slight increase in girls enrollment compared to boys, 3 years after treatment. This indicates that there is some benefit for girls but it is perhaps attenuated by disproportionately positive impacts for boys through other channels (e.g. health). All of these mechanisms deserve further examination and provide direction for future work.

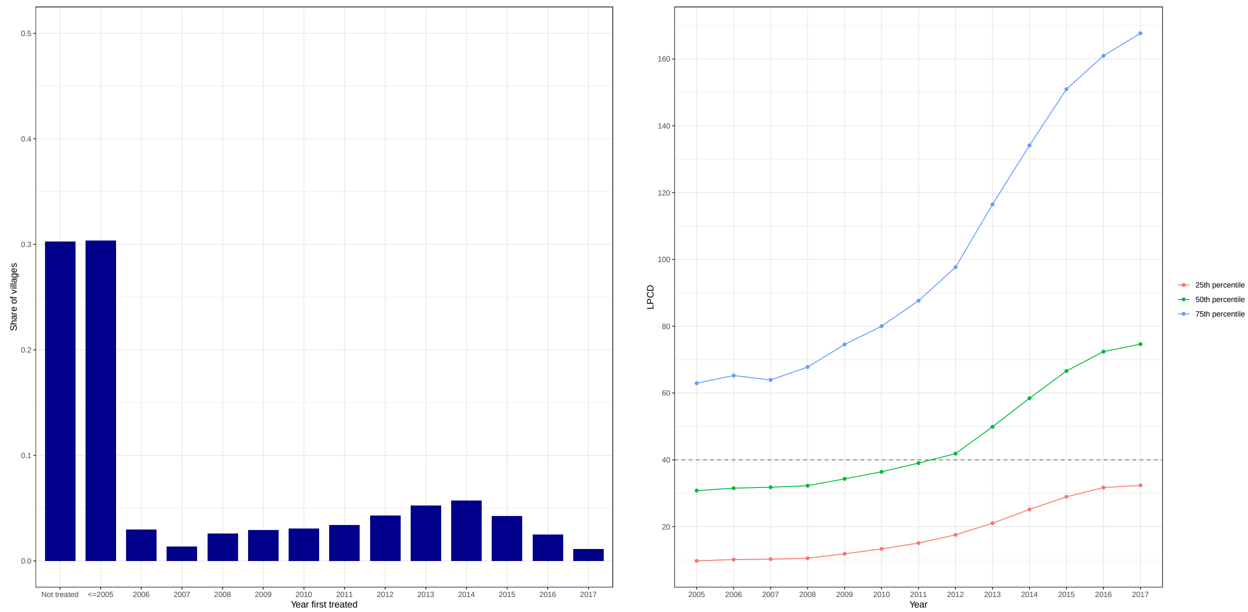
2.6 Figures

Figure 2.1: Screenshot of IMIS platform



S.No.	State	Population & Habitation Having PWS As on 01/04/2018												Target Population & Habitation By		
		Number of Habitation			% of Habitation			Population(Lakhs)			% of Population			Number of Habitation	% of Habitation	Populat
		Total	FC	PC+QA	Total	FC	PC+QA	Total	FC	PC+QA	Total	FC	PC+QA			
1	2	3									4			5		
	Total	7,21,277	5,42,584	1,78,693	41.79	31.44	10.35	4,967.51	3,484.63	1,482.88	54.08	37.93	16.14	21,398	1.24	
1	Andaman & Nicobar Islands	254	242	12	63.50	60.50	3.00	2.42	2.19	0.23	91.36	82.85	8.52	0	0.00	
2	Andhra Pradesh	39,514	28,241	11,273	80.81	57.76	23.06	348.46	223.14	125.31	94.61	60.59	34.02	1,130	2.31	
3	Arunachal Pradesh	3,675	2,359	1,316	48.84	31.35	17.49	7.80	5.00	2.80	62.84	40.27	22.57	157	2.09	
4	Assam	35,147	25,219	9,928	39.91	28.63	11.27	137.75	95.64	42.11	46.50	32.28	14.21	1,651	1.87	
5	Bihar	7,552	4,803	2,749	6.85	4.36	2.49	116.46	64.12	52.34	11.71	6.45	5.26	0	0.00	
6	Chandigarh	0	0	0	0.00	0.00	0.00	0	0	0	0.00	0.00	0.00	0	0.00	

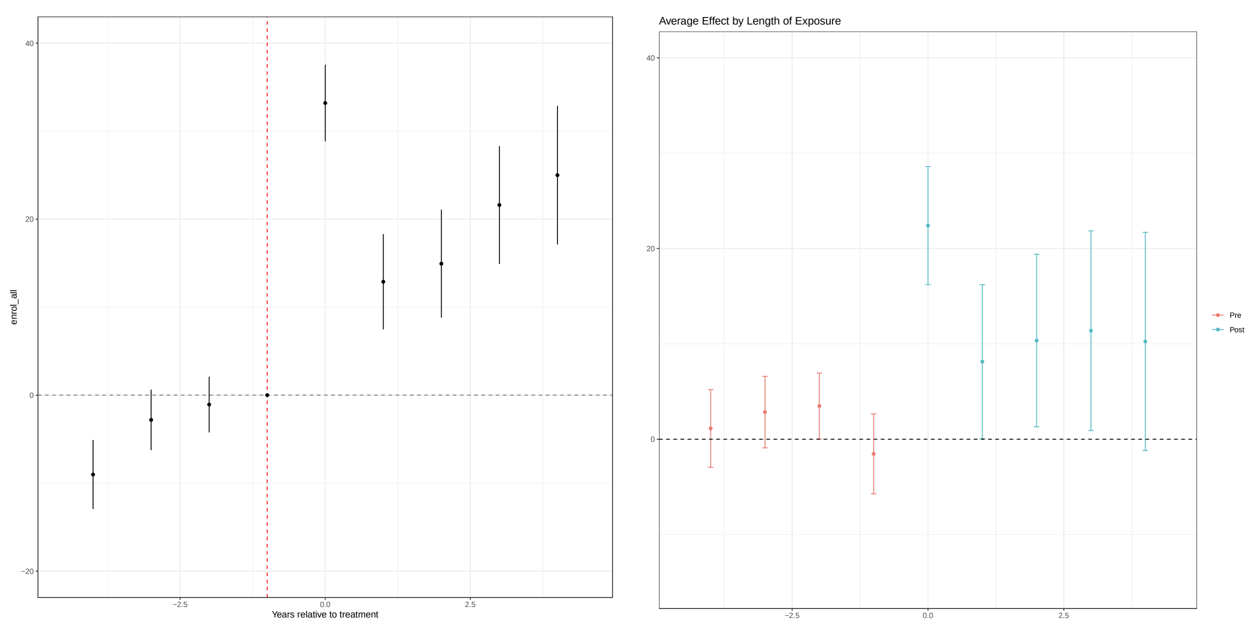
Figure 2.2: Treatment timing across years and villages



(a) Timing of treatment event across years

(b) Variation in treatment across years

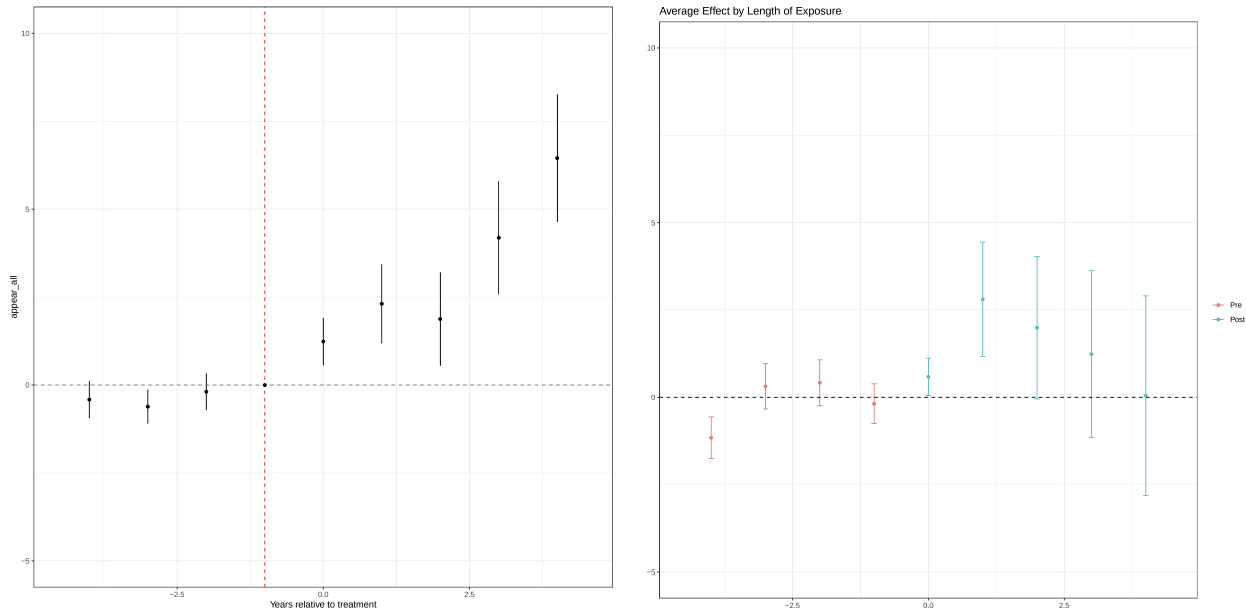
Figure 2.3: Effect of HWA on total enrollment



(a) Dynamic two-way fixed effects

(b) Callaway Sant'Anna

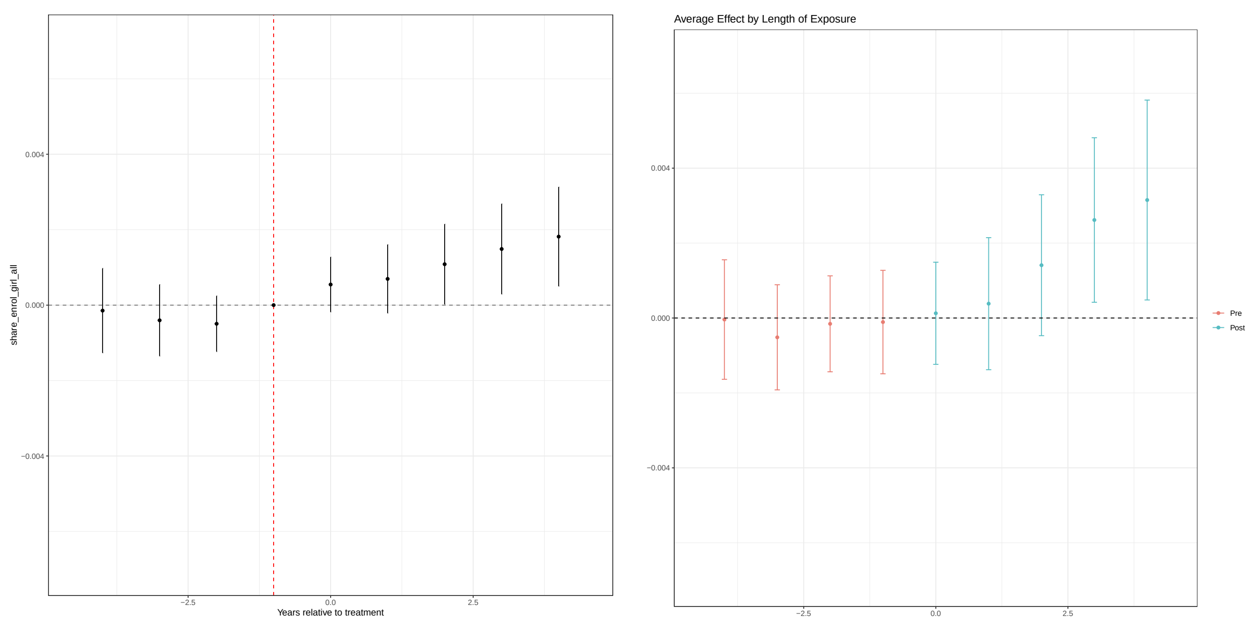
Figure 2.4: Effect of HWA on total exam-takers in grades 5-8



(a) Dynamic two-way fixed effects

(b) Callaway Sant'Anna

Figure 2.5: Effect of HWA on share of girls' in total enrollment



(a) Dynamic two-way fixed effects

(b) Callaway Sant'Anna

2.7 Tables

Table 2.1: Summary statistics

Statistic	N	Mean	St. Dev.	Min	Median	Max
Population	150,060	1,791.882	2,782.451	1	1,020	223,929
Number of HHs	150,060	367.594	601.959	1	210	42,848
Enrolment	150,060	396.248	824.219	0.000	182.667	52,479.580
LPCD	150,060	176.697	7,415.409	0.000	50.800	2,511,407.000

CHAPTER 3

PATTERNS OF IDENTITY-BASED CONSUMPTION

3.1 Introduction

Social identity plays an important role in determining one's economic outcomes Akerlof and Kranton [2000], Shayo [2020]. For example, women face discrimination in the workplace of various kinds because of their gender identity; members of dominant identity groups, in general, enjoy access to influential networks that can help them further improve their economic well-being. Over the course of history, some identity groups have become marginalized as a result of dominance of some other identity groups. Members belonging to both these groups followed norms in order to build trust within the group. Then, members of an in-group could help each other improve their economic outcomes, especially in a society that lacked institutional safety nets (e.g. social security, unemployment insurance). However, given the socio-economic asymmetry between dominant and marginalized groups, the benefits provided by the two will naturally vary. But both of these identity groups exist in the same society and interact with each other in different settings e.g. school, work, place of residence, public space, etc. Not only does one receive benefits from one's in-group but could also incur costs from the out-group (e.g. discrimination, exclusion). Therefore, in order to understand how one's in-group can benefit or disadvantage them, it is important to study how dominant and marginalized identity groups co-exist in a given society and what costs do they impose on each other, if any.

In this paper, I study how one's social identity changes with the social identity of one's neighbours. Accounting for the hierarchical nature of social identities, I ask if, compared to living among their own group, marginalized identity groups strengthen (or weaken) their identity when living among dominant identity groups (and vice versa). Atkin et al. [2021] show that social identities can be fungible and they can be measured using consumption expenditure. Specifically, they find that individuals choose to identify more or less with their group based on how salient the group identity is for them, what the status of the group is and how costly it is for them to identify with it. I build on these findings in 2 ways, (1) I show that marginalized group members who live among dominant group members identify less with their own group i.e. they follow in-group norms less and out-group norms more. I find that the same is not true for dominant group members who live among the marginalized group. (2) I propose an “amenities-insurance tradeoff” framework to explain and further test these patterns.

I use the setting of urban India, where hierarchy based on religious identity can be widely observed Jaffrelot and Gayer [2012]. Hinduism is the most prevalent religious identity (75%) and Muslims are the largest religious minority (14%)[1]. In the modern history of India, the Islamic religious identity has been politically and socially marginalized (see Sachar Committee Report, 2006). Meanwhile, the dominant-caste Hindu identity has been politically and socially powerful.

Individuals carry various identities as part of their daily life (race, gender, sexuality, religion, ethnicity, school/college/profession). The identities a person holds can put them at an advantage compared to others but it can also disadvantage them. In most cases, it is not possible to completely hide these identities. However, it is often possible to choose how strong one wants to make them. This can be done by following group norms: wearing certain types of clothes, eating certain types of foods, engaging in certain kinds of activities, speaking in a particular way, and so on. These actions are visible to others and can communicate to them whether you belong to a certain group or not. Potentially, one can choose which group norms to follow in order to weaken or strengthen parts of their identity.

However, individuals do not choose the “strength” of their identity in a vacuum. They are influenced by their economic, social and political conditions. For example, an individual

1. Both religious groups are further divided on the basis of caste. I will be accounting for these divisions in future analysis.

might have to make their identity stronger or weaker in order to work at a job where they are the minority. Anecdotally, Dalits (historically marginalized castes in India) who live in dominant-caste apartment buildings are expected to celebrate festivals they may not subscribe to and are also expected to refrain from publicly celebrating festivals of their own community. Similarly, workplace attire requirements may not be friendly to Muslim women who might want to wear a hijab to work. Many predominantly Brahmin schools don't allow children to bring meat in their lunches[2]. To send one's kid to such a school, they will have to weaken the meat-eating part of their child's identity. In this paper, I will be focusing on one such setting, place of residence.

The key hypothesis here is that there is a tradeoff between where one lives and how strong their identity can be, especially for marginalized groups. In order to live among the dominant group, marginalized group members may have to weaken their identity. This could be for various reasons: to avoid discrimination, to assimilate into the community, ex-ante preferences, and so on. Regardless, weakening one's identity can be costly, since a strong group identity helps access in-group networks that fill in for lack of institutional safety nets and access to dominant group networks. Moreover, there can be intrisic loss from weakening one's identity.

Thus, dominance or marginalization of a group identity can affect economic outcomes through many important ways. In my proposed framework, I focus on two ways: through housing location and through access to informal insurance networks.

Where one lives is an important determinant of one's well-being Ludwig et al. [2013]. Within a monetary budget constraint, individuals choose where to live based on several non-monetary costs and benefits i.e. access to amenities like schools and hospitals, commute time to work, and so on. However, a salient fact is that neighbourhoods are segregated on religion and caste in Indian cities (see Figure 3.2). Thus, the social identity of one's neighbours will

2. Meat is taboo among many Brahmins, a historically dominant caste group in India

enter the cost-benefit calculation. We also know that Hindu-majority neighbourhoods have better access to amenities and public goods Adukia et al. [2019]. Especially for a religious minority household i.e. Muslims household, there might be a cost to incur from living in a Hindu-majority neighbourhood Thorat et al. [2015]. For one, they may lose the community networks they would have in a neighbourhood where they are a majority. This cost is more salient for those who rely on networks for social insurance in absence of institutional support. Secondly, they may also feel unsafe or uncomfortable displaying their religious identity in a Hindu-majority neighbourhood. If not that, they may find the need to assimilate into the Hindu community by silencing their Islamic identity. These can all be characterized as costs.

In this paper, I find the following: Muslims who live in Hindu-majority neighbourhoods are less likely to consume goods associated with Islamic identity and more likely to consume goods associated with Hindu identity, relative to their counterparts who live in Hindu-minority neighbourhoods. Section 3.2 proposes a theoretical framework. Section 3.3 describes the empirical strategy. Section 3.4 describes main results as well as supporting evidence. Section 3.5 concludes.

3.2 Theoretical framework

3.2.1 Defining identity strength

Similar to Atkin et al. [2021], I define identity strength in terms of distance from group norm. Let there be a good x. Average consumption of x by group A is given by $\bar{x}_A$. Thus, $\bar{x}_A$ is the group norm. For a given household i, consumption of x is given by x_i. Distance from group A's norm is given by $x_i - \bar{x}_A$. Thus, household i "strengthens" its group A identity as its distance from group A norm reduces.

In the empirical section, distance from norm is currently defined as a binary based on observed group norms. However, this can be refined to match the definition above.

Individuals can be motivated to strengthen their identity for multiple reasons: (a) it gives them a sense of belonging, (b) it helps them build trust with the in-group, (c) it can help with access to in-group networks (through trust), (d) it signals their identity to the out-group (if consumption is observable), (e) it is costly to deviate because of fear of being punished by in-group and/or out-group members.

3.2.2 Defining marginalization

Historical as well as modern-day processes have rendered some social identities dominant and some others marginalized. Dominant identity groups are disproportionately represented in elected office, thus leading to over-provision of needs of dominant group voters (and under-provision of needs of marginalized voters).

Amenities like schools, hospitals, public transport are under-provided in areas with higher population of marginalized groups. Similarly, under-provision of institutional safety nets disproportionately affects marginalized groups (since they are economically worse off).

3.2.3 Identity and in-group networks

In the absence of institutional safety nets, individuals rely on informal insurance from their in-group. In-group networks can provide loans, gifts and in-kind transfers to their members. There are other benefits such as informal childcare, information about jobs, and supply of group-specific goods (e.g. religious food, clothing). Since under-provision of institutional safety nets disproportionately affects marginalized group members, they are the ones more dependent on in-group networks for goods provision. In order to access these in-group networks, one must build trust within the in-group. By strengthening group identity, one can gain trust of the in-group.

3.2.4 Identity and physical location

Being physically proximate to the in-group can also be important to acess some of the in-group provisions. Furthermore, discrimination in dominant-group neighbourhoods can contribute to physical proximity of marginalized group members. As mentioned above, these neighbourhoods with high populations of marginalized groups members have under-provision of amenities. In such a case, the in-group can also provide for amenities using coordination and in-group trust.

However, the in-group can't provide for all amenities. It can be difficult to provide for schools, hospitals, public transport without institutional support. Thus, living among the marginalized in-group has its limitations in terms of provision of goods and services.

One can choose to move to a neighbourhood with better amenities to alleviate these limitations. However, neighbourhoods with better public goods provision are more likely to be occupied by dominant group and assimilation or acceptance into a dominant out-group network is difficult and costly. Further, moving away from one's in-group can weaken access to the safety nets provided by the in-group.

3.2.5 Framework

→ Define identity strength as "distance" from group norms i.e. if one follows norms of group A more, their identity A is stronger (they are closer to the group A) and vice versa.

→ Social insurance (lending/borrowing/information/advice/jobs) is decreasing in distance, for marginalized group identity.

→ Public goods provision is increasing in distance (by way of housing location), for marginalized group identity.

3.2.6 Endogenous relationships

To distill the framework into a simple endogenous relationship, I use distance, social insurance and public goods as three variables that affect each other. Moreover, how they have affect each other varies by residence among in-group vs. out-group.

$\rightarrow$ Distance $\leftrightarrow$ Social Insurance and Public Goods

$\rightarrow$ *Living among dominant out-group:*

High Distance $\leftrightarrow$ Low Social Insurance and High Public Goods

$\rightarrow$ *Living among marginalized in-group:*

Low Distance $\leftrightarrow$ High Social Insurance and Low Public Goods

In the following sections, I show evidence on the relationship between distance and place of residence. Studying the relationship between distance, social insurance and public goods is beyong the scope of this analysis, at the moment.

3.3 Empirical strategy

The empirical section of this paper is set in urban India. I use the Indian National Sample Survey (NSS) data from 1983-2005. It is a household-level sample survey conducted across the country every 5 years and contains detailed information on demographics and consumption expenditure. I use data on religious identity, consumption of food/drink items and consumption of clothing items.

The NSS follows a multi-stage stratified sampling strategy. For the urban sector, the sampling design is as follows:

1. Within each district of a state, all urban areas formed a single stratum. The exception to this is towns with a population of 1 million or more that form their own stratum.

2. Each urban area (i.e. town/city) is divided into 'blocks' based on a geographical mapping survey. These blocks consist of approximately 80-100 households. They are demarcated by natural boundaries of the neighbourhood.

3. Sub-strata are formed such that each sub-stratum has approximately equal number of blocks.

4. From each sub-stratum, a number of blocks were randomly sampled. This is the first stage of sampling.

5. Next, from each of the sampled blocks, a number of households were randomly sampled. This is the second and final stage of sampling.

6. In my analysis, the mean number of households sampled from each sampled block is 10.25.

First, I construct a measure at the neighbourhood-level (80-100 households) of the share of Hindus in each neighbourhood. I use this to define Hindu-majority and Hinud-minority neighbourhoods (referred to as "blocks" in the survey and hereafter in this paper). Figures 3.1 and 3.2 show the distribution of Hindu-majority blocks in my sample. The median block has 90% Hindus. However, there is enough variation in the share of Hindus that I can utilize. Second, I interact this binary variable with the religious identity of each household in the data. This gives me 6 distinct groups: 1) Muslims living in Hindu-majority blocks, 2) Muslims living in Hindu-minority blocks, 3) Hindus living in Hindu-majority blocks, 4) Hindus living in Hindu-minority blocks, 4) Other religious minorities living in Hindu-majority blocks, 5) Other religious minorities living in Hindu-minority blocks.

It is useful to note here that the analysis in this paper is the most granular level of analysis at this scale that I have come across in the Indian setting. The population census data is at a ward-level, which is upwards of 100 households. I face a limitation of only 10 households per sampled block. But, given that it is a survey that spans across the entire nation over decades, it is offering a unique insight into how religious group identity operates at the smallest unit measurable.

Based on the theoretical framework set up in Section 3.2, I test if distance from in-group identity as well as distance to out-group varies by residence location of a marginalized group member. I also test if it varies for a dominant group member. In order to do this, first, I compare consumption patterns of Muslims living in Hindu-majority blocks to those of Muslims living in Hindu-minority blocks. I use food and clothing to measure group norms of both religions. Based on the theoretical predictions, I expect Muslims who live in Hindu-majority blocks to have a weaker Islamic identity (i.e. larger distance from group norms) compared to those who live in Hindu-minority blocks.

I use the following specification,

$$
\begin{aligned}
y_i = {} & \beta_1 \text{Muslim in Hindu-majority block}_i + \beta_2 \text{Hindu in Hindu-majority block}_i \\
& + \beta_3 \text{Hindu in Hindu-minority block}_i) + \beta_4 \text{Others in Hindu-majority block}_i \\
& + \beta_5 \text{Others in Hindu-minority block}_i \\
& + \text{Region FE} + \text{Survey round FE} + \varepsilon_i
\end{aligned}
$$

y_i denotes whether the household consumes a given item of food/drink/clothing or not. The omitted category is Muslims in Hindu-minority blocks (notice that Hindu-minority doesn't necessarily mean Muslim-majority since there are other groups in the analysis). Out of the 5 co-efficients, I am most interested in β_1. This is because Muslims are the largest and most marginalized religious minority in India. Moreover, Islamic norms and Hindu norms are well-defined and widely known. Focusing on Muslims in relation to Hindus allows me to observe one marginalized and one dominant identity group. There is scope to explore the relationship between different religious groups in India as well. But for the ease of reading, I only report β_1 in my results. In order to account for regional and temporal variation, I use region[3] and survey round fixed-effects. Moreover, I also add controls for demographic characteristics that could have otherwise confounded my results.

3.4 Results

3.4.1 Norms in Hinduism and Islam

Table 3.1 shows the share of households in each religious group that consume a given item. In terms of meats, only 1.5% of Hindus in the sample consume beef, whereas 28.5% of them consume mutton. Less than a percent of them consume pork. 10.2% of them consume alcohol. In terms of clothing, 71% of them wear saris, 22.5% of them wear dhotis and 1.8% of them wear headgear. Compared to this, 50.5% of Muslims wear saris, 9.9% of Muslims wear Dhotis and 7.6% of them wear headgear. 25.7% of them consume beef, 45.8% of them consume mutton. Only 0.3% consume pork and 3% consume alcohol. 88.5% of Hindus and 87.9% of Muslims consume rice. 75.4% of Hindus and 74.2% of Muslims consume wheat. It will be useful to keep these base means in mind when thinking about norms of consumption and taboo in each religion.

3.4.2 Muslims in Hindu-majority blocks follow more Hindu norms and less Islamic norms

Panel A of Table 3.2 shows four dependent variables I use to measure Hindu norms. Consuming beef is taboo amongst dominant-caste Hindus. Consuming mutton (and other types of meat) is common practice in some Hindu castes and not others. Dhoti and Sari are attire typically worn by Hindus (although they have become more widely accepted attire over time).

3. A region is a geographic unit below state but above district. This is the unit at which I have NSS data across all rounds.

Muslims living in Hindu-majority blocks are less likely to spend on beef(-15.1pp) and more likely to spend on mutton instead(+7.2pp), relative to Muslims living in Hindu-minority blocks. This can indicate that they follow Hindu norms more (by consuming less beef) and perhaps substitute into a more acceptable form of meat (by consuming more mutton).

Panel B of Table 3.2 shows three dependent variables that I use to measure Islamic norms. Alcohol and pork are taboo to consume. Wearing headgear (prayer cap for men, hijab for women) is a common practice. However, notice that Muslims living in Hindu-majority blocks are more likely to spend on alcohol(+1.7pp) compared to their counterparts living in Hindu-minority blocks. This can indicate that Muslims living in Hindu-majority blocks are less likely to follow Islamic norms, especially ones that are not aligned with Hindus norms. We can see this by looking at consumption of pork among the Muslim groups – we see no difference. Pork is taboo in Islam but is also not widely consumed by Hindus. Perhaps abstaining from it aligns with Hindu norms already[4].

Moving on to clothing, we see that Muslims in Hindu-majority blocks are more likely to spend on dhotis(+1.5pp) and saris(+8.7pp) and less likely to spend on headgear(-2.5pp). This further indicates that they spend on following Hindus norms more and Islamic norms less. Figures 3.3 and 3.4 show the same patterns but with raw data i.e. no added controls.

Panel C of Table 3.2 shows rice and wheat, that aren't typically associated with either religious identity. We see no difference among the two Muslim groups in consumption of the neutral goods.

3.4.3 Hindus living in Muslim-majority blocks don't exhibit these patterns.

Table 3.3 shows the same dependent variables. But now, I am comparing Hindus who live in Muslim-majority blocks to Hindus who live in Muslim-minority blocks. Overall, you do

4. I run the same analysis for Christians (for whom consumption of pork is a norm) and find that Christians in Hindu-majority blocks are less likely to consume pork compared to their counterparts in Hindu-minority blocks (Table 3.7)

not see significant differences in their consumption patterns. This suggests that Hindus do not follow Islamic norms more or Hindu norms less when they live amongst Muslims. The result provides evidence that the tradeoff may be faced more acutely by marginalized identity groups.

3.4.4 Demographic differences

Panel A Table 3.4 shows the demographic differences between Hindu-majority blocks and Hindu-minority blocks. We see that households in Hindu-majority blocks have fewer members (-0.6 members), have a lower share of females (-1pp), and are more likely to be literate (+2pp). Among literates, they are less likely to be below high school education (-7.6pp) and are more likely to have completed high school or more (+9.7pp). They also spend more per month per capita (+INR 40). In Panel B, I compare Muslims in Hindu-majority blocks to Muslims in Hindu-minority blocks – we see the same patterns. I control for these factors in my main results. However, it is still useful to note these differences because they may be useful when thinking of explanations for the phenomena I observe in the results.

3.4.5 Supporting evidence

One of the main findings has to do with consumption of meat. It could be argued that Muslims in Hindu-majority blocks don't have easy access to beef and hence consume less of it. To check for this, I identify households in the data that are engaged in butcher, meat trade and related occupations. I restrict the analysis to these households in Panel A of Table 3.5. I see the same patterns of beef and mutton consumption as the main result. By restricting the sample to butcher households, I am easing the supply-side constraint assuming that a butcher would have access to any kind of meat if they wanted to have it. This provides suggestive evidence that there is more to the store of weakening identity than simply supply-side constraints. Panel B of Table 3.5 restricts the sample to blocks where butchers live and

I still find the same patterns.

Table 3.8 restricts the sample only to Muslim households in a block and compares the difference across varying shares of Hindus in the block. We see that Muslims follow Hindu-norms more and Islamic norms as the share of Hindus in the block increases. This is consistent with my main result and provides additional evidence for it. Table 3.9 shows the results over time but I don't see any clear trend in the norm-following.

3.5 Conclusion

In this paper, I find that Muslims who live in Hindu-majority neighbourhoods are less likely to consume goods associated with Islamic identity and more likely to consume goods associated with Hindu identity, relative to their counterparts who live in Hindu-minority neighbourhoods. I explain this finding using an "amenities-insurance" tradeoff framework. I argue that there are benefits to living among the dominant out-group (Hindus): such as access to amenities. However, in order to live amongst the dominant group identity, the marginalized group members (Muslims) weaken their own religious identity. This leads me to believe that there are costs to living among the dominant out-group as well: (1) instrinsic loss of identity, (2) loss of in-group networks. It is important to understand how marginalized group members make these tradeoffs if we want to further our understanding of the relationship between identity and economic welfare.

3.6 Figures

Figure 3.1: Share of Hindus in sample blocks

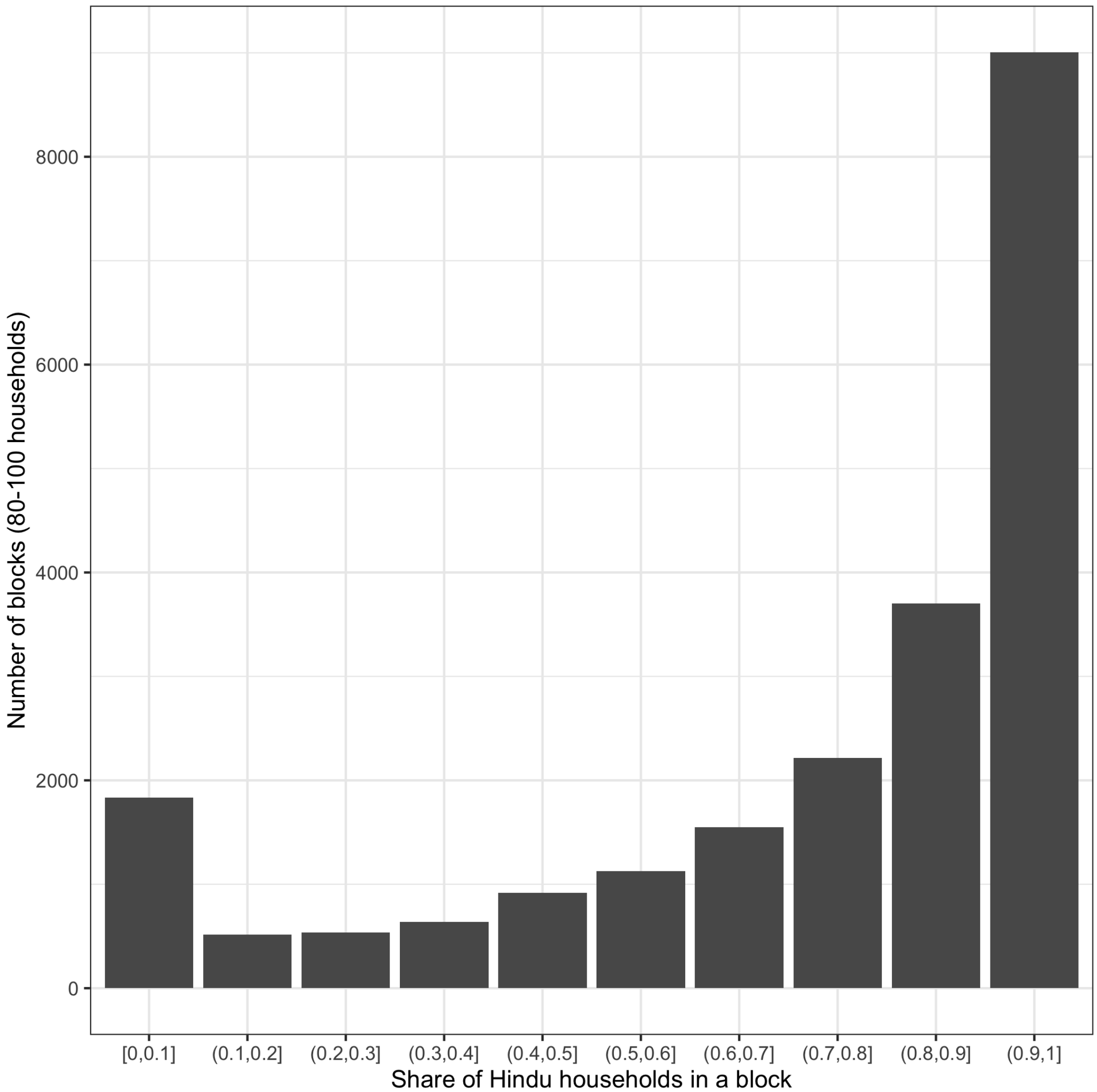


Figure 3.2: Cumulative distribution of blocks by share of Hindus

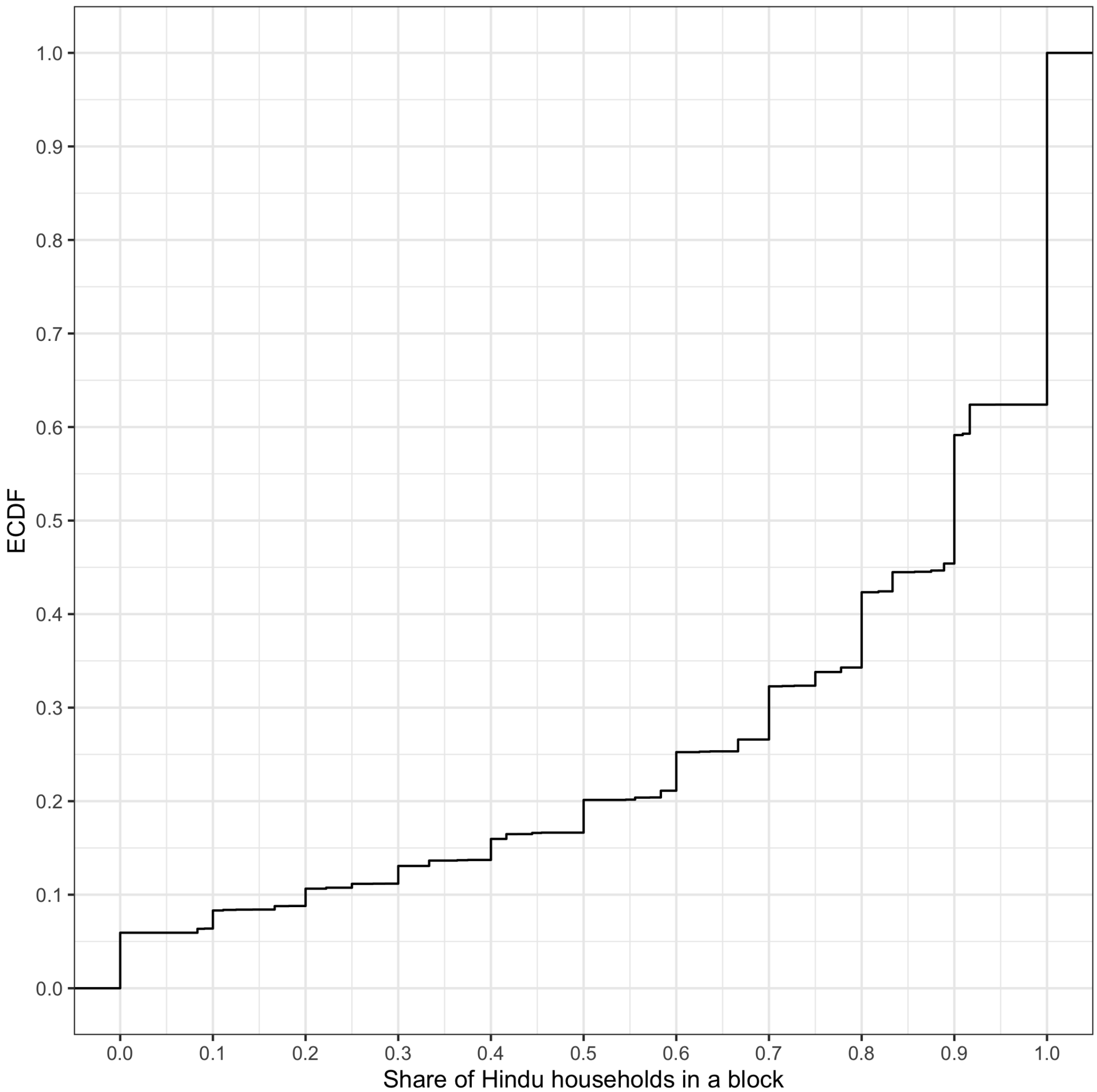


Figure 3.3: Hindu norms

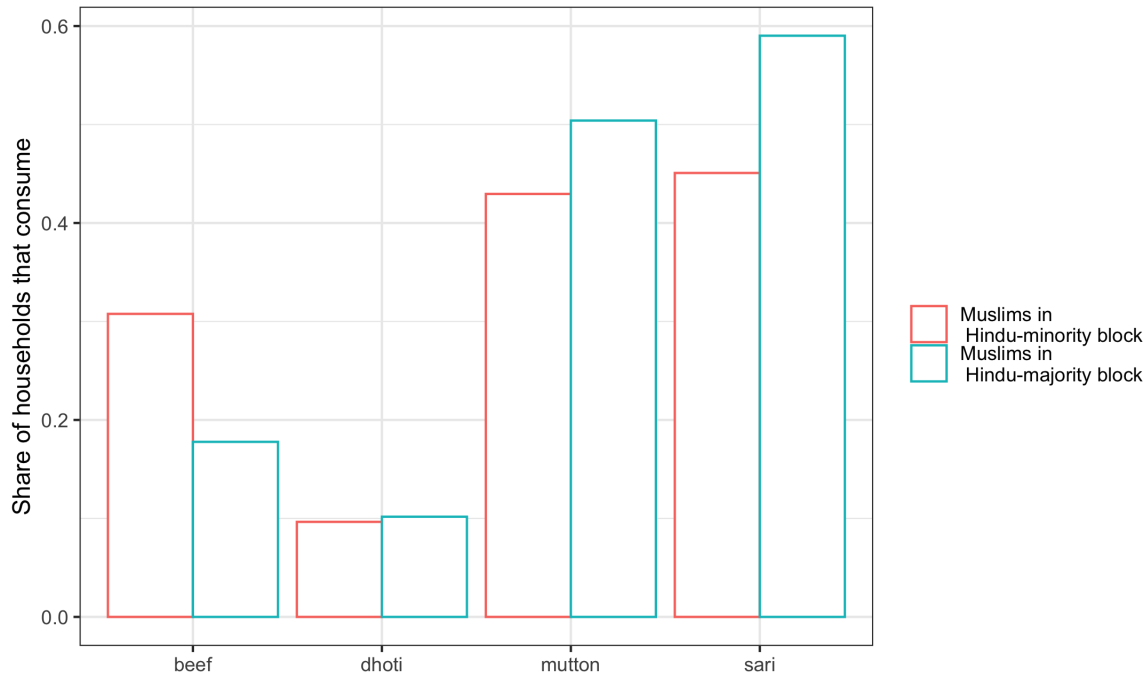


Figure 3.4: Islamic norms

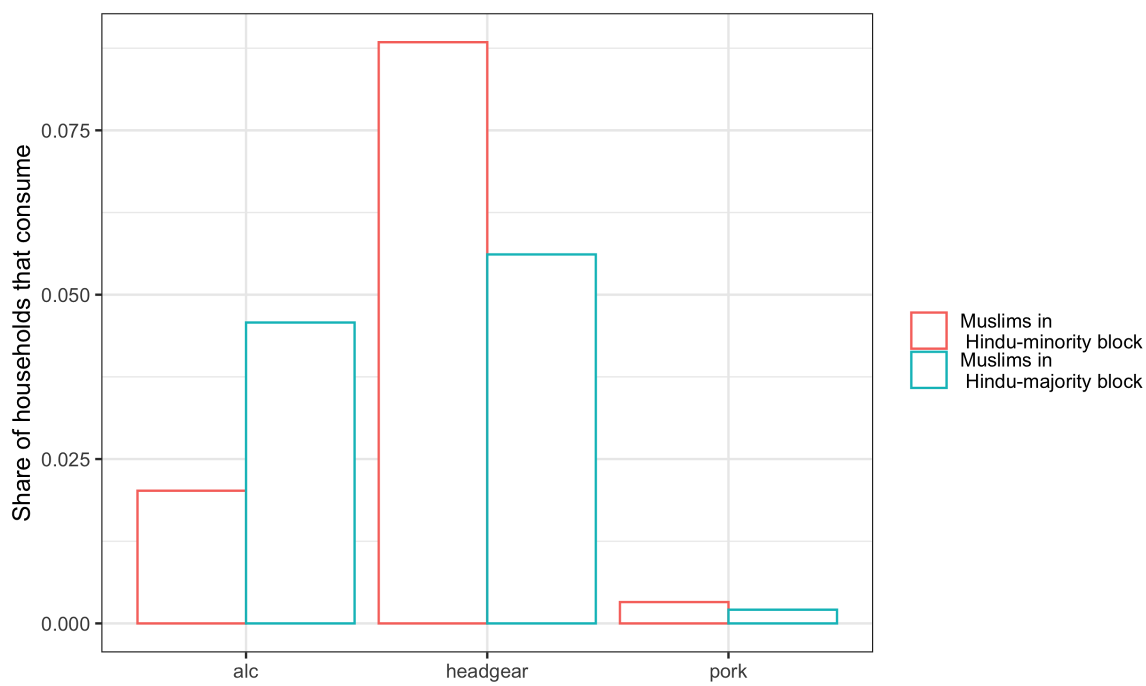


Figure 3.5: Time trend

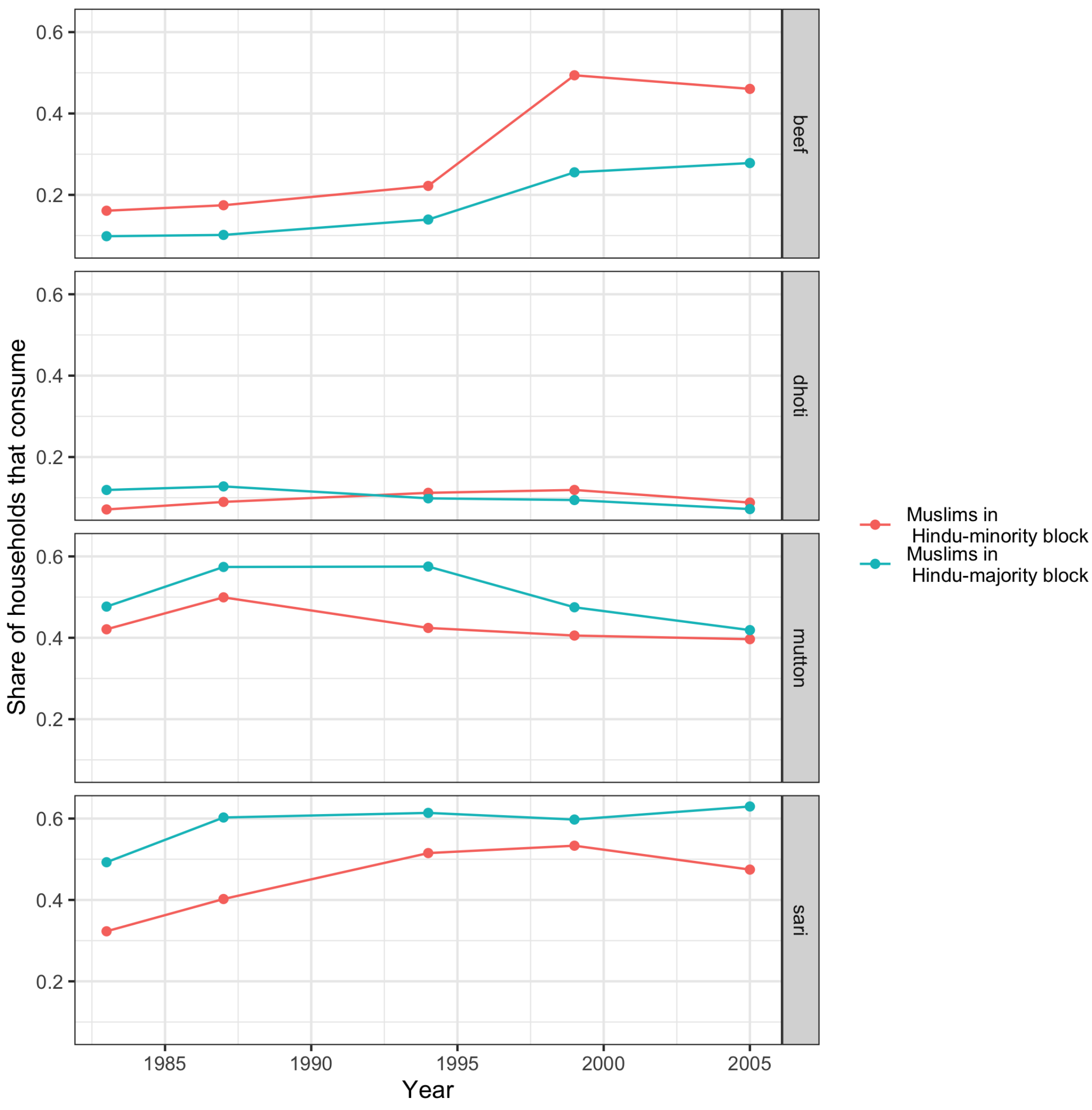
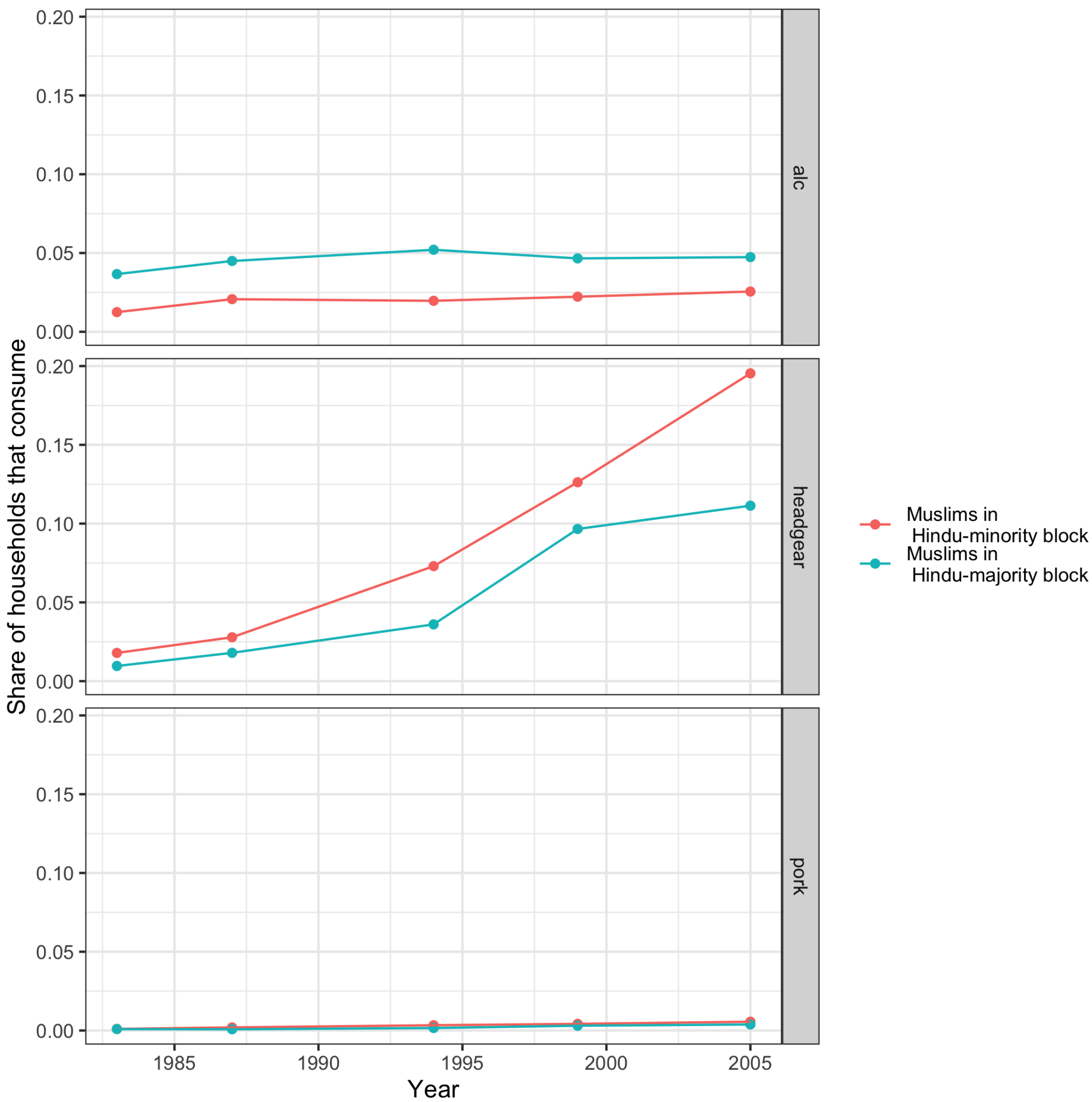


Figure 3.6: Time trend



3.7 Tables

Table 3.1: Consumption averages

	Hindu	Muslim	Other
Pork	0.009	0.003	0.262
	(0.092)	(0.053)	(0.44)
Alcohol	0.102	0.03	0.147
	(0.302)	(0.171)	(0.354)
Headgear	0.018	0.076	0.097
	(0.133)	(0.265)	(0.296)
Beef	0.015	0.257	0.275
	(0.121)	(0.437)	(0.446)
Mutton	0.285	0.458	0.227
	(0.451)	(0.498)	(0.419)
Sari	0.71	0.505	0.401
	(0.454)	(0.5)	(0.49)
Dhoti	0.225	0.099	0.103
	(0.418)	(0.298)	(0.305)
Rice	0.885	0.879	0.894
	(0.319)	(0.326)	(0.307)
Wheat	0.754	0.742	0.596
	(0.431)	(0.438)	(0.491)

Notes: Each row displays the share of Hindus, Muslims and Other religious groups that consume each good. Standard deviations are reported in parentheses.

Table 3.2: Muslims follow different norms based on who they live amongst

Panel A: Hindu norms

	Dependent variable:			
	beef	mutton	dhoti	sari
	(1)	(2)	(3)	(4)
Muslim in Hindu-majority block	-0.151^{***}	0.072^{***}	0.015^{***}	0.087^{***}
	(0.010)	(0.010)	(0.005)	(0.010)
Muslim in Hindu-minority block mean	0.308	0.429	0.097	0.451
Observations	184,512	184,512	184,512	184,512

Panel B: Islamic norms

	Dependent variable:		
	alc	pork	headgear
	(1)	(2)	(3)
Muslim in Hindu-majority block	0.017^{***}	-0.0004	-0.025^{***}
	(0.003)	(0.001)	(0.006)
Muslim in Hindu-minority block mean	0.02	0.003	0.088
Observations	184,512	184,512	184,512

Panel C: Neutral goods

	Dependent variable:	
	rice	wheat
	(1)	(2)
Muslim in Hindu-majority block	0.002	0.006
	(0.005)	(0.006)
Muslim in Hindu-minority block mean	0.88	0.719
Observations	184,512	184,512

Notes: Dependent variables are dummies for if household spends on given item. All columns include region fixed-effects (geographic unit below state, above district), survey round fixed-effects, monthly per capita expenditure of household, dummy for literacy, dummy for high school completion, household size, and share of female members. Co-efficients on Hindus and other religious minorities in the blocks not displayed for brevity. Standard errors (clustered at the block-level) are reported in parentheses.

Table 3.3: Hindus follow similar norms regardless of who they live amongst

Panel A: Hindu norms

	Dependent variable:			
	beef	mutton	dhoti	sari
	(1)	(2)	(3)	(4)
Hindu in Muslim-majority block	0.005	0.017**	0.017**	−0.009
	(0.003)	(0.008)	(0.008)	(0.007)
Hindu in Muslim-minority block mean	0.014	0.285	0.224	0.709
Observations	184,512	184,512	184,512	184,512

Panel A: Islamic norms

	Dependent variable:		
	alc	pork	headgear
	(1)	(2)	(3)
Hindu in Muslim-majority block	−0.012**	0.003	−0.002
	(0.005)	(0.002)	(0.003)
Hindu in Muslim-minority block mean	0.102	0.009	0.018
Observations	184,512	184,512	184,512

Panel C: Neutral goods

	Dependent variable:	
	rice	wheat
	(1)	(2)
Hindu in Muslim-majority block	0.002	0.003
	(0.005)	(0.007)
Hindu in Muslim-minority block mean	0.885	0.754
Observations	184,512	184,512

Notes: Dependent variables are dummies for if household spends on given item. All columns include region fixed-effects (a geographic unit below state, above district), survey round fixed-effects, monthly per capita expenditure of household, dummy for literacy, dummy for high school completion, household size, and share of female members. Co-efficients on Hindus and other religious minorities in the blocks not displayed for brevity. Standard errors (clustered at the block-level) are reported in parentheses.

Table 3.4: Demographic differences

Panel A: Demographics in Hindu-majority vs. Hindu-minority blocks

	Dependent variable:					
	HH size	Share female	Literate	> High school	>= High school	MPCE
	(1)	(2)	(3)	(4)	(5)	(6)
Hindu-majority block	-0.661^{***}	-0.010^{***}	0.020^{***}	-0.076^{***}	0.097^{***}	39.973^{***}
	(0.014)	(0.001)	(0.002)	(0.003)	(0.003)	(6.176)
Observations	225,249	225,242	184,585	184,585	184,585	225,932

Panel B: Demographics of Muslims in Hindu-majority vs. Hindu-minority blocks

	Dependent variable:					
	HH size	Share female	Literate	> High school	>= High school	MPCE
	(1)	(2)	(3)	(4)	(5)	(6)
Muslim in Hindu-majority block	-0.496^{***}	-0.008^{***}	0.020^{***}	-0.004	0.024^{***}	33.588^{***}
	(0.029)	(0.003)	(0.003)	(0.006)	(0.006)	(12.492)
Observations	225,124	225,117	184,512	184,512	184,512	225,807

Notes: Dependent variables are (1) HH size = number of members in the household, (2) Share female = share of females in the household, (3) Literate = dummy for at least one literate member, (4) > High school = No member who studied upto high school, (5) <= High school = At least one membet who studied upto high school or more, (6) MPCE = Monthly per capita expenditure.

Table 3.5: Easing supply-side constraint of access to meat

Panel A: Sample restricted to butchers

	Dependent variable:								
	alc	pork	headgear	beef	mutton	dhoti	sari	rice	wheat
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Muslim in Hindu-majority block	0.019	−0.007	−0.070**	−0.176***	0.162***	0.006	−0.029	0.005	0.003
	(0.028)	(0.012)	(0.029)	(0.049)	(0.053)	(0.034)	(0.043)	(0.017)	(0.041)
Muslim in Hindu-minority block mean	0.042	0.002	0.107	0.392	0.411	0.116	0.564	0.903	0.74
Observations	1,089	1,089	1,089	1,089	1,089	1,089	1,089	1,089	1,089

Panel B: Sample restricted to blocks where butchers live

	Dependent variable:								
	alc	pork	headgear	beef	mutton	dhoti	sari	rice	wheat
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Muslim in Hindu-majority block	0.023	−0.004	−0.065***	−0.096***	0.056*	0.015	0.044	0.011	0.003
	(0.015)	(0.006)	(0.018)	(0.034)	(0.030)	(0.018)	(0.029)	(0.012)	(0.023)
Muslim in Hindu-minority block mean	0.028	0.002	0.095	0.32	0.413	0.105	0.527	0.889	0.719
Observations	8,955	8,955	8,955	8,955	8,955	8,955	8,955	8,955	8,955

Notes: Dependent variables are dummies for if household spends on given item. All columns include region fixed-effects (geographic unit below state, above district), survey round fixed-effects, monthly per capita expenditure of household, dummy for literacy, dummy for high school completion, household size, and share of female members. Co-efficients on Hindus and other religious minorities in the blocks not displayed for brevity. Standard errors (clustered at the block-level) are reported in parentheses. In Panel B, co-efficients on Hindus and other religious minorities in the blocks not displayed for brevity. Standard errors (clustered at the block-level) are reported in parentheses.

Table 3.6: Muslims living amongst Christians

	Dependent variable:						
	alc	pork	headgear	beef	mutton	dhoti	sari
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Muslim in Christian-majority block mean	-0.021	-0.140^{***}	-0.026	-0.024	-0.128^{***}	0.033	0.235^{***}
	(0.013)	(0.030)	(0.017)	(0.049)	(0.034)	(0.029)	(0.032)
Muslim in Christian-minority block mean	0.03	0.002	0.076	0.256	0.46	0.098	0.505
Observations	184,512	184,512	184,512	184,512	184,512	184,512	184,512

Notes: Dependent variables are dummies for if household spends on given item. All columns include region fixed-effects (geographic unit below state, above district), survey round fixed-effects, monthly per capita expenditure of household, dummy for literacy, dummy for high school completion, household size, and share of female members. Co-efficients on Hindus and other religious minorities in the blocks not displayed for brevity. Standard errors (clustered at the block-level) are reported in parentheses.

Table 3.7: Christians following different norms based on who they live amongst

	Dependent variable:						
	alc	pork	headgear	beef	mutton	dhoti	sari
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Christian in Hindu-majority block mean	-0.015	-0.184^{***}	-0.014^{**}	-0.245^{***}	0.197^{***}	-0.005	0.091^{***}
	(0.012)	(0.015)	(0.006)	(0.017)	(0.015)	(0.011)	(0.017)
Christian in Hindu-minority block mean	0.129	0.639	0.03	0.586	0.1	0.084	0.237
Observations	184,512	184,512	184,512	184,512	184,512	184,512	184,512

Notes: Dependent variables are dummies for if household spends on given item. All columns include region fixed-effects (geographic unit below state, above district), survey round fixed-effects, monthly per capita expenditure of household, dummy for literacy, dummy for high school completion, household size, and share of female members. Co-efficients on Hindus and other religious minorities in the blocks not displayed for brevity. Standard errors (clustered at the block-level) are reported in parentheses.

Table 3.8: Muslims follow different norms based on share of Hindus in their block

	Dependent variable:								
	alc	pork	headgear	beef	mutton	sari	dhoti	rice	wheat
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Muslim in 20 − 40% Hindu block	0.004	-0.002^{**}	−0.011	-0.072^{***}	0.018	0.014	−0.006	-0.023^{***}	−0.009
	(0.004)	(0.001)	(0.011)	(0.016)	(0.013)	(0.012)	(0.008)	(0.008)	(0.010)
Muslim in 40 − 60% Hindu block	0.012^{***}	−0.001	−0.013	-0.100^{***}	0.027^{**}	0.010	0.007	−0.003	−0.012
	(0.004)	(0.001)	(0.011)	(0.015)	(0.013)	(0.012)	(0.007)	(0.007)	(0.010)
Muslim in 60 − 80% Hindu block	0.013^{***}	−0.001	-0.020^{**}	-0.147^{***}	0.050^{***}	0.028^{**}	0.008	−0.010	−0.002
	(0.004)	(0.001)	(0.009)	(0.014)	(0.013)	(0.011)	(0.007)	(0.007)	(0.009)
Muslim in 80 − 100% Hindu block	0.022^{***}	−0.001	-0.023^{***}	-0.179^{***}	0.056^{***}	0.060^{***}	0.020^{***}	−0.008	−0.005
	(0.004)	(0.001)	(0.009)	(0.013)	(0.013)	(0.011)	(0.007)	(0.007)	(0.009)
Muslim in 0 − 20% Hindu block mean	0.015	0.004	0.098	0.333	0.427	0.088	0.399	0.885	0.69
Observations	25,912	25,912	25,912	25,912	25,912	25,912	25,912	25,912	25,912

Notes: Dependent variables are dummies for if household spends on given item. All columns include region fixed-effects (geographic unit below state, above district), survey round fixed-effects, monthly per capita expenditure of household, dummy for literacy, dummy for high school completion, household size, and share of female members. Co-efficients on Hindus and other religious minorities in the blocks not displayed for brevity. Standard errors (clustered at the block-level) are reported in parentheses.

Table 3.9: Muslims and norm following over time

	Dependent variable:								
	alc	pork	headgear	beef	mutton	sari	dhoti	rice	wheat
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Muslim in Hindu-majority block (1987)	0.013**	−0.001	0.010	−0.056***	0.069***	0.050***	0.033***	0.004	0.010
	(0.006)	(0.001)	(0.008)	(0.016)	(0.016)	(0.015)	(0.010)	(0.009)	(0.013)
Muslim in Hindu-minority block (1994)	−0.010**	0.0004	0.023**	0.021	−0.083***	−0.010	−0.025**	0.005	0.005
	(0.005)	(0.002)	(0.011)	(0.019)	(0.018)	(0.016)	(0.012)	(0.010)	(0.014)
Muslim in Hindu-majority block (1994)	0.010	0.0001	0.012	−0.009	−0.043**	0.009	−0.011	0.014	0.015
	(0.007)	(0.001)	(0.009)	(0.017)	(0.017)	(0.015)	(0.010)	(0.009)	(0.014)
Muslim in Hindu-minority block (1999)	−0.017***	0.001	0.060***	0.312***	−0.286***	0.004	−0.018*	0.022**	0.006
	(0.006)	(0.002)	(0.013)	(0.021)	(0.018)	(0.017)	(0.010)	(0.010)	(0.016)
Muslim in Hindu-majority block (1999)	−0.005	0.001	0.056***	0.115***	−0.252***	0.004	−0.015	0.020*	−0.012
	(0.007)	(0.002)	(0.013)	(0.020)	(0.019)	(0.017)	(0.011)	(0.011)	(0.015)
Muslim in Hindu-minority block (2005)	−0.013**	0.002	0.117***	0.258***	−0.307***	−0.024	−0.045***	0.026**	0.009
	(0.006)	(0.002)	(0.017)	(0.021)	(0.019)	(0.018)	(0.011)	(0.010)	(0.015)
Muslim in Hindu-majority block (2005)	−0.006	0.002	0.067***	0.123***	−0.315***	0.002	−0.043***	0.015	−0.010
	(0.007)	(0.002)	(0.013)	(0.020)	(0.019)	(0.016)	(0.011)	(0.011)	(0.015)
Muslim in Hindu-minority block (1987) mean	0.021	0.002	0.028	0.175	0.499	0.09	0.402	0.919	0.695
Observations	25,912	25,912	25,912	25,912	25,912	25,912	25,912	25,912	25,912

Notes: Dependent variables are dummies for if household spends on given item. All columns include region fixed-effects (geographic unit below state, above district), survey round fixed-effects, monthly per capita expenditure of household, dummy for literacy, dummy for high school completion, household size, and share of female members. Co-efficients on Hindus and other religious minorities in the blocks not displayed for brevity. Standard errors (clustered at the block-level) are reported in parentheses.

REFERENCES

Anjali Adukia. Sanitation and education. *American Economic Journal: Applied Economics*, 9(2):23–59, 2017.

Anjali Adukia, Sam Asher, Paul Novosad, and Brandon Tan. Residential segregation in urban india. *Available here*, 2019.

Amrita Ahuja, Michael Kremer, and Alix Peterson Zwane. Providing safe water: Evidence from randomized evaluations. *Annu. Rev. Resour. Econ.*, 2(1):237–256, 2010.

George A Akerlof and Rachel E Kranton. Economics and identity. *The quarterly journal of economics*, 115(3):715–753, 2000.

Bhimrao Ramji Ambedkar. *Castes in India: Their mechanism, genesis and development*, volume 1. Ssoft Group, INDIA, 2016.

Sam Asher and Paul Novosad. Rural roads and local economic development. *American economic review*, 110(3):797–823, 2020.

Nava Ashraf, Edward Glaeser, Abraham Holland, and Bryce Millett Steinberg. Water, health and wealth: The impact of piped water outages on disease prevalence and financial transactions in zambia. *Economica*, 88(351):755–781, 2021.

David Atkin, Eve Colson-Sihra, and Moses Shayo. How do we choose our identity? a revealed preference approach using food consumption. *Journal of Political Economy*, 129 (4):1193–1251, 2021.

Abhijit Banerjee, Lakshmi Iyer, and Rohini Somanathan. History, social divisions, and public goods in rural india. *Journal of the European Economic Association*, 3(2-3):639–647, 2005.

Abhijit Banerjee, Rukmini Banerji, James Berry, Esther Duflo, Harini Kannan, Shobhini Mukherji, Marc Shotland, and Michael Walton. Mainstreaming an effective intervention: Evidence from randomized evaluations of "teaching at the right level" in india. Technical report, National Bureau of Economic Research, 2016.

Sharon Barnhardt, Erica Field, and Rohini Pande. Moving to opportunity or isolation? network effects of a randomized housing lottery in urban india. *American Economic Journal: Applied Economics*, 9(1):1–32, 2017.

Eli Berman. Sect, subsidy, and sacrifice: an economist's view of ultra-orthodox jews. *The Quarterly Journal of Economics*, 115(3):905–953, 2000.

David Blakeslee, Ram Fishman, and Veena Srinivasan. Way down in the hole: Adaptation to long-term water loss in rural india. *American Economic Review*, 110(1):200–224, 2020.

Matias Busso, Jesse Gregory, and Patrick Kline. Assessing the incidence and efficiency of a prominent place based policy. *American Economic Review*, 103(2):897–947, 2013.

Brantly Callaway and Pedro HC Sant'Anna. Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230, 2021.

Sebastian Calonico, Matias D Cattaneo, Max H Farrell, and Rocio Titiunik. rdrobust: Software for regression-discontinuity designs. *The Stata Journal*, 17(2):372–404, 2017.

Arun G Chandrasekhar, Melanie Morten, and Alessandra Peter. Network-based hiring: Local benefits; global costs. Technical report, National Bureau of Economic Research, 2020.

Raghabendra Chattopadhyay and Esther Duflo. Impact of reservation in panchayati raj: Evidence from a nationwide randomised experiment. *Economic and political Weekly*, pages 979–986, 2004.

Daniel L Chen. Club goods and group identity: Evidence from islamic resurgence during the indonesian financial crisis. *Journal of political Economy*, 118(2):300–354, 2010.

Raj Chetty, Nathaniel Hendren, and Lawrence F Katz. The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment. *American Economic Review*, 106(4):855–902, 2016.

David M Cutler, Edward L Glaeser, and Jacob L Vigdor. When are ghettos bad? lessons from immigrant segregation in the united states. *Journal of Urban Economics*, 63(3): 759–774, 2008.

Florencia Devoto, Esther Duflo, Pascaline Dupas, William Parienté, and Vincent Pons. Happiness on tap: Piped water adoption in urban morocco. *American Economic Journal: Economic Policy*, 4(4):68–99, 2012.

Esther Duflo. Schooling and labor market consequences of school construction in indonesia: Evidence from an unusual policy experiment. *American economic review*, 91(4):795–813, 2001.

Esther Duflo and Rohini Pande. Dams. *The Quarterly Journal of Economics*, 122(2):601–646, 2007.

Glenn Ellison and Parag A Pathak. The efficiency of race-neutral alternatives to race-based affirmative action: Evidence from chicago's exam schools. *American Economic Review*, 111(3):943–75, 2021.

Shanti Gamper-Rabindran, Shakeeb Khan, and Christopher Timmins. The impact of piped water provision on infant mortality in brazil: A quantile panel data approach. *Journal of Development Economics*, 92(2):188–200, 2010.

Saad Gulzar, Nicholas Haas, and Benjamin Pasquale. Does political affirmative action work, and for whom? theory and evidence on india's scheduled areas. *American Political Science Review*, 114(4):1230–1246, 2020.

Karla Hoff and Priyanka Pandey. Discrimination, social identity, and durable inequalities. *American Economic Review*, 96(2):206–211, 2006.

Richard Hornbeck and Pinar Keskin. The historically evolving impact of the ogallala aquifer: Agricultural adaptation to groundwater and drought. *American Economic Journal: Applied Economics*, 6(1):190–219, 2014.

Christophe Jaffrelot and Laurent Gayer. *Muslims in Indian cities: Trajectories of marginalisation*. Columbia University Press, 2012.

Seema Jayachandran. The roots of gender inequality in developing countries. *economics*, 7 (1):63–88, 2015.

KM Rabiul Karim, Maria Emmelin, Bernadette P Resurreccion, and Sarah Wamala. Water development projects and marital violence: Experiences from rural bangladesh. *Health Care for Women International*, 33(3):200–216, 2012.

Patrick Kline and Enrico Moretti. Local economic development, agglomeration economies, and the big push: 100 years of evidence from the tennessee valley authority. *The Quarterly journal of economics*, 129(1):275–331, 2014a.

Patrick Kline and Enrico Moretti. People, places, and public policy: Some simple welfare economics of local economic development programs. *Annu. Rev. Econ.*, 6(1):629–662, 2014b.

Tanu Kumar. Welfare programs and local political participation: the effects of affordable housing in mumbai. 2019.

Jens Ludwig, Greg J Duncan, Lisa A Gennetian, Lawrence F Katz, Ronald C Kessler, Jeffrey R Kling, and Lisa Sanbonmatsu. Long-term neighborhood effects on low-income families: Evidence from moving to opportunity. *American economic review*, 103(3):226–31, 2013.

Kaivan Munshi and Mark Rosenzweig. The efficacy of parochial politics: caste, commitment, and competence in indian local governments. Technical report, National Bureau of Economic Research, 2008.

Benjamin A Olken and Rohini Pande. Corruption in developing countries. *Annu. Rev. Econ.*, 4(1):479–509, 2012.

Rohini Pande. Can mandated political representation increase policy influence for disadvantaged minorities? theory and evidence from india. *American Economic Review*, 93(4): 1132–1151, 2003.

Jonathan Roth, Pedro HC Sant'Anna, Alyssa Bilinski, and John Poe. What's trending in difference-in-differences? a synthesis of the recent econometrics literature. *Journal of Econometrics*, 2023.

Nicholas Ryan and Anant Sudarshan. Rationing the commons. *Journal of Political Economy*, 130(1):210–257, 2022.

Sheetal Sekhri. Wells, water, and welfare: the impact of access to groundwater on rural poverty and conflict. *American Economic Journal: Applied Economics*, 6(3):76–102, 2014.

Moses Shayo. Social identity and economic policy. *Annual Review of Economics*, 2020.

Zejiang Shen, Ruochen Zhang, Melissa Dell, Benjamin Charles Germain Lee, Jacob Carlson, and Weining Li. Layoutparser: A unified toolkit for deep learning based document image analysis. In *Document Analysis and Recognition–ICDAR 2021: 16th International Conference, Lausanne, Switzerland, September 5–10, 2021, Proceedings, Part I 16*, pages 131–146. Springer, 2021.

Raphael Susewind. Muslims in indian cities: Degrees of segregation and the elusive ghetto. *Environment and Planning A*, 49(6):1286–1307, 2017.

Sukhadeo Thorat, Anuradha Banerjee, Vinod K Mishra, and Firdaus Rizvi. Urban rental housing market. *Economic & Political Weekly*, 27:47–53, 2015.

Winnie Van Dijk. The socio-economic consequences of housing assistance. *University of Chicago Kenneth C. Griffin Department of Economics job market paper, 0–46 i–xi*, 36, 2019.

Joel Waldfogel. Preference externalities: An empirical study of who benefits whom in differentiated product markets. Technical report, National Bureau of Economic Research, 1999.

Ann M Weber, Beniamino Cislaghi, Valerie Meausoone, Safa Abdalla, Iván Mejía-Guevara, Pooja Loftus, Emma Hallgren, Ilana Seff, Lindsay Stark, Cesar G Victora, et al. Gender norms and health: insights from global survey data. *The Lancet*, 393(10189):2455–2468, 2019.