Affirmative Action, Education and Gender: Evidence from India.*

Guilhem Cassan†

January 30, 2017

Abstract: This paper studies the impact of affirmative action policies on educational attainment in India. Using plausible exogenous variation, they are shown to have widely differing impact across genders, with males gaining about 2 years of education while females not being affected and maybe losing. This underlines the fact that groups cumulating discrimination (caste and gender) may not be sufficiently well protected by “simple” affirmative action policies.

JEL Classification: I24; O15; H41

Keywords: identity; scheduled caste; quota; affirmative action; gender; India; education; intersectionality.

---

*I am grateful to Gani Aldashev, Denis Cogneau, Christelle Dumas, Esther Duflo, Hemanshu Kumar, Eliana La Ferrara, Sylvie Lambert, Andreas Madestam, Marco Manacorda, Annamaria Milazzo, Barbara Petrongolo, Rohini Somanathan, Vincenzo Verardi and Yanos Zylberberg, as well as participants in the seminar and conferences in which this paper has been presented. Marion Leturcq’s and Lore Vandewalle’s comments on an earlier draft vastly improved this paper. I am indebted to Julien Grenet for his help in the data collection, and to Ashwini Natraj, with whom I started this project. This paper is produced as part of the project “Actors, Markets, and Institutions in Developing Countries: A micro-empirical approach” (AMID), a Marie Curie Initial Training Network (ITN) funded by the European Commission. Funding from the CEPREMAP is gratefully acknowledged. This research used resources of the “Plateforme Technologique de Calcul Intensif (PTCI)” (http://www.ptci.unamur.be) located at the University of Namur, Belgium, which is supported by the F.R.S.-FNRS under the convention No. 2.4520.11. The PTCI is member of the “Consortium des Equipements de Calcul Intensif (CECI)” (http://www.ceci-hpc.be). All remaining errors are mine.

†University of Namur, CRED. Email: guilhem.cassan@unamur.be.
Introduction

Evaluating affirmative action policies is by construction a difficult exercise. As a result, while these policies are often widely debated, the debate is rarely well informed. India, in particular, has implemented the largest affirmative action program in the world, which targets the low castes and in particular the “Scheduled Castes” (SC). Those policies have been controversial since they were systematically introduced after the Independence. In the early 90s, with the enlargement of affirmative action policies to a large new group of relatively well off castes (the “Other Backward Classes”, or OBC) the debate has become particularly intense, as higher castes began to fear competition. The debate has notably focused on quotas in higher education, in which competition with high caste was the strongest. Following this trend, most of the research has dealt with the consequences of quotas in universities (Bertrand et al., 2010; Krishna and Frisvand Robles, 2012) or of quotas for OBC (Khanna, 2013). However, according to the 2011 Census, only 56% of the SC population aged 20 and above was literate (against 66% in the whole population), only 30% had attained an education level higher than primary (against 41%), and only 6% had gone beyond secondary schooling (against 11%). And these vary broadly by gender: only 43% of women aged 20 and above are literate against 66% for the whole population, 21% went above primary school and 3.8% went beyond secondary schooling. Hence, the current focus on affirmative action in higher education entirely neglects the most weaker segment of the Indian population, the Scheduled Castes which in their immense majority never even reach university. Focusing on university level affirmative action programs is not sufficient to get a broad picture of the effect of affirmative action policies targeted to low castes in general and to SC in particular. As a result, the diagnostic drawn several decades ago by Chalam (1990), Chitnis (1972) or Galanter (1984) still holds: we have very little knowledge about the impact of these policies on SC. This paper ambitions to contribute to filling this gap.

I show in this paper that the overall effect of affirmative action policies on educational attainment of SC is quite small, if any. However, this small overall effect hides large heterogenous effects across genders. Indeed, males gain up to 2.5 years of education while females are not affected and may even have suffered from the access to the SC

---

1The “Scheduled Castes” are the castes traditionally the most discriminated against. They represent approximately 17% of the population in 2011.

2And this diagnostic remains true for younger cohorts, as focusing on the 25-29 cohort yields better but still very low figures: 70% literacy (60% for female), 42% (resp. 34%) above primary school and 8% (resp. 6.5%) above secondary school.
status. This result underlines the fact that populations that cumulate disadvantages, that are at the intersection between different discriminated groups, such as women of low castes, may not be sufficiently protected by policies which do not take into account this cumulative discrimination. Strikingly, the gains for the males come from the first years of education: affirmative action policies seem to push families to put their children in the schooling system, in primary school in particular, while there does not seem to be any robust increase in later years of education. Hence, this paper adds two important contributions to the literature on affirmative action in India. First, by focusing only on the higher end of education, impact evaluations may miss important effect of affirmative action policies. Second, by putting the accent on the gender difference of the impact of affirmative action, this paper emphasizes the heterogenous effect of affirmative action policies within the treated population. As such, it relates to the literature on gender discrimination (Jensen, 2012) and in particular to the literature on asymmetric effect of social policies in developing countries (Foster and Rosenzweig, 2003; Rosenzweig and Schultz, 1982; Ashraf et al., 2015) as well as to the “intersectionality” literature in political science, which studies the consequences of the accumulation of discriminations (Hughes, 2011, 2014; Jensenius, 2015).

The research design of this paper relies on a unique natural experiment on the access to the SC status. As a matter of fact, the list of the castes considered as SC were drawn by each Indian state at the Independence. In 1956, the borders of the Indian states have been redrawn, while the lists of SC remained unchanged. This created many within State discrepancies: the SC status of the members of a same caste could vary across space. This situation lasted until 1976, when the list of castes considered as SC were harmonized within each state, allowing 2.4 million individuals (Government of India, ed, 1978) to have access to the SC status. This historical setting allows me to assess the fate of the individuals who had access to the SC status from 1976.

As the treatment status varies across castes, and within caste across time and space, this paper innovates by coding the treatment status of individuals based on their precise

\[\text{In practice, each state had several lists of SC, by regions usually grouping several districts. Those regions are listed in Appendix A.}\]

\[\text{Note that this means that the identification does not rely on policy variation across states, but on variation to exposure to affirmative action policies within a single state. This specific setting allows to safely attribute any change observed to only a change in access to the Scheduled Caste status and not to any other state wide policy.}\]
caste name\(^5\) instead of their declared beneficiary status\(^6\). By using the precise caste name I can use demanding specifications, which identify the effects of the policy within a single caste, something that, to the best of my knowledge, has never been done in the literature. As a matter of fact, the main specifications rely on caste*cohorts and region*cohorts fixed effects, which allow to identify the treatment effect using only the within caste-cohort and within region-cohort variation. In a robustness check, I show that adding a specific trend for the treatment group does not alter the results. Hence, the results presented in this paper are not due to pre-existing trends, and can safely be attributed to the acquisition of the Scheduled Caste status.

In the first section of this paper, I will present the context and the natural experiment exploited in this paper. I will then describe the data as well as the empirical strategy, which will open the way to the presentation of the results, their discussion and various robustness checks (varying treatment across cohorts, differential trends, migration and identity manipulation).

1 Context

1.1 Affirmative action in India

While the first affirmative action policies for Scheduled Castes were implemented under British rule, it is not before the Independence that a systematic positive discrimination policy was implemented. Also called “reservations”, it has 3 main dimensions: political representation, education and public employment. Electoral quotas are indeed implemented for SC in all federal and state assemblies, as well as in panchayat elections (since the mid 90s), while public recruitment is subject to a quota policy, with a share of new openings to be reserved for low caste groups. Finally, in education, affirmative action consists in various policies. There are quotas in higher education institutions, free secondary schooling, as well as access to scholarships, specific schools and hostels and free mid day meals\(^7\). Positive discrimination might thus affect schooling through various channels. By reducing the cost of education, it favors longer studies in the cost-benefit arbitrage of the household, while quotas in higher education will help the pursuit of

\(^5\)Their “jati”, that is the endogamous group which is the relevant social category in every day life, as opposed to their caste category, such as “Scheduled Caste”, which is an administrative category.

\(^6\)Indeed, most of the micro level literature on affirmative action in India uses the the household’s declaration of its SC status as a basis for identifying the “treatment” group (Khanna, 2013; Hnatkovska et al., 2012, 2013; Prakash, 2009)

\(^7\)See Galanter (1984), chapter 3.B for a presentation of the main educational policies targeting low castes apart from quotas in higher education.
studies after secondary schooling. Moreover, the quotas in public employment increases the returns to education. Hence, this paper does not evaluate the effect of affirmative action in the educational sector but the effect of the package of affirmative action policies on educational attainment.

1.2 The definition of the Scheduled Castes

How are the beneficiaries of these affirmative action policies determined? As underlined by Galanter (1984), one of the main difficulties of the creation of a list of Scheduled Castes is the criteria of inclusion in that list. Indeed, while “untouchability” is the prerequisite to be considered a SC, defining “untouchability” is actually not straightforward since its meaning varies across the subcontinent. Hence, the Constitution of 1950 avoids to define a clear concept of untouchability and only provides a procedure of designation that each State is to follow. This allowed for the possibility of inconsistencies across States as well as within States, as certain States decided to give the SC status to certain castes only in certain areas. Despite those inconsistencies, the lists were revised only four times since the Independence. With an increase of 2.4 million SC over an original population of 80 million SC, the Scheduled Castes and Scheduled Tribes (Amendment) Act of 1976 was the most dramatic change in the list of SC in India.

1.3 The Scheduled Castes and Scheduled Tribes Orders (Amendment) Act of 1976

In 1956, India reorganized the borders of its States along linguistic lines. But, as the borders of the States were redefined, the State-wise SC lists remained unchanged. This led to a large increase of within-state discrepancies in the SC lists, since there were now systematically several SC lists per state. Hence, from 1956, the number of castes considered as SC in one part of a State and not in another part of the same State vastly increased. It is only in 1976 that the SC lists were harmonized within states by the SC

---

8 “castes, races or tribes or parts of or groups within castes, races and tribes which shall for purposes of this Constitution be deemed to be Scheduled Castes in relation to that State.”

9 Bayly (1999) gives the example of the Khatik caste, considered as SC in Punjab, but classed as a “forward” caste in Uttar Pradesh, a neighboring State at the time of the establishment of the lists.

10 The first change, in 1951 was a matter of correcting small anomalies. The second change, in 1956 mainly affected Rajasthan and Uttar Pradesh, and also allowed all Sikh untouchable castes to claim SC status, while the fourth change of 1990 allowed the Buddhists to have access to the SC status in all the States. The most important change took place in 1976 and is described in more details in the paper.

11 And in 1960, the states of Maharashtra and Gujarat were created from the former state of Bombay.
and ST (Amendment) Act of 1976, also called the “Area Restriction Removal Act”\textsuperscript{12}. This Act removed almost all intra-State restrictions\textsuperscript{13}. This removal of restrictions led to an increase of 2.4 million of the SC population\textsuperscript{14}. The list of these areas of restriction (which I’ll call “regions” in the paper) is in Appendix A. In practice, the paper uses 31 regions.

Hence, individuals from the same caste could be considered as SC in one part of a State but not in another until 1976, when anyone whose caste was on the SC list somewhere in a state would be considered SC in that state. This situation thus creates a natural experiment setting in which the members of a same caste faced different access to affirmative action status within the same state. This unique historical event creates a plausible exogenous variation in the SC status across individuals, allowing me to assess the causal impact of the SC status on educational attainment.

2 Data and Empirical Strategy

2.1 Data and descriptive statistics

The National Family and Health Survey of 1998-99 (NFHS2) is, to my knowledge, the only dataset offering both the precise caste name (the “jati”) of respondents, their district of residence and a sufficient sample size to perform the type of analysis done in this paper\textsuperscript{15}.

\textsuperscript{12}The reason for the list not to be adjusted to the new borders was the slowness of the administration: “It has been mentioned in the last report that the President has issued the SC and ST Lists (Modification) Order, 1956, specifying the SC and ST in the re-organized States. As these lists had to be issued urgently for the re-organized States, it was not possible to prepare comprehensive and consolidated lists and therefore, the SC and ST had to be specified in these lists territory-wise within each re-organized State” (Government of India, ed, 1958). But not only did the administration fail to change the lists on time, it failed to do so for a period of twenty years. The yearly reports of the Commissioner on SC and ST are particularly telling in this aspect, as many of its yearly occurrences refer to the fact that “[...]the question of preparation of comprehensive lists of SC and ST for the reorganized States [...] remained pending [...]” (Government of India, ed, 1960).

\textsuperscript{13}According the Galanter (1984) their number dropped from 1,126 to 64.

\textsuperscript{14}The states affected by the reform were Andhra Pradesh, Bihar, Gujarat, Karnataka, Kerala, Madhya Pradesh, Maharashtra, Rajasthan, Uttar Pradesh, Tamil Nadu, West Bengal and Himachal Pradesh. The analysis excludes Himachal Pradesh, as when this Union Territory was granted the State status in 1966, large portions of the state of Punjab were also transferred to it: hence, between 1956 and 1966, the population living in the contemporary borders of Himachal Pradesh were exposed to different State policies, preventing me from identifying the sole effect of the access to the SC status. Results are unaffected by the inclusion of Himachal Pradesh. Note that if the changes in border mainly affected the South of India, the Area Restriction Removal Act also affected northern states such as Bihar and Uttar Pradesh, who had within states variations in their SC list before 1956.

\textsuperscript{15}For example, the 2004-5 round of the NFHS does not contain information on the district of residence of the respondents.
The jati names present in the NFHS2 are present only in a raw form. As a consequence, an effort has been made to clean the names of the jatis of Scheduled Castes in a systematic manner. The procedure followed was to send the list of the raw caste names with the list of castes and synonyms of Kitts (1885)\textsuperscript{16} as well as the a second caste list which (supposedly) had been used for the coding of the first round of the NFHS\textsuperscript{17} to a data entry team based in India for them to match names from the NFHS2 to the two lists. I then refined this matching using the state wise list of Scheduled Caste synonyms available in the Census of 1971 and 1981. This methodology has led to the identification of a total of 179 SC jati, defined at the state level\textsuperscript{18}. Using the cleaned jati names, the 1971 and 1981 Census lists of Scheduled Castes and the district of residence of households, I am able to identify the households that were granted the SC status in 1976\textsuperscript{19}. Throughout this paper, I will call this population “New SC”, as opposed to the “Old SC” who were already on the list in 1976. This methodology differs from most studies of caste using nationally representative surveys. Indeed, while these studies generally rely on the declarations of the respondent on their Scheduled Caste status, I use the jati name of the respondent to attribute the SC status, which allows to have within jati estimates of the access to the SC status.

Tables 1 and 2 provide the summary statistics for the variables used throughout the paper. I restrict the sample to individuals born between 1951 (below school age in 1956, year of the redefinition of borders) and aged 17 and above at the time of the survey. It can be seen that the educational attainment in the sample is lower for females than for males, and that overall, “Old SC” have a higher level of education than “New SC”. It can also be noted that the declared religions are somewhat different across groups. All the results shown are robust to restricting the sample to only Hindus\textsuperscript{20}.

\textsuperscript{16}This source was used at it had been the base of the coding of caste names in the first round of NFHS (1992-3).

\textsuperscript{17}Indeed, the first round of the NFHS (1992-3) also offers jati and district identifiers. This project initially planned to use both first and second NFHS rounds. However, the first round turned out to be unusable as the documentation on the codes of caste (which is based on Kitts (1885)) does not correspond to the codes present in the data. It has also come to my knowledge that a different list of caste codes had been used for the first round of NFHS, partially overlapping Kitts (1885), and which I managed to acquire. However, this second list appeared to be unknown to the central organization of NFHS and also to not correspond perfectly to the codes present in the raw data either. As a consequence, the first round of the NFHS could not be used.

\textsuperscript{18}There are two main reasons to define the jatis at the state level. First, the same jati name might not mean the same thing in different areas of India. Second, since the list of synonyms of SC present in the 1971 and 1981 Census is state based, the coding of the synonyms was also state based.

\textsuperscript{19}In order to do so, I simply match the 1999 districts of the NFHS2 data to their 1971 counterparts.
### Table 1: Descriptive Statistics: Male population.

<table>
<thead>
<tr>
<th>Education variables</th>
<th>New SC Mean</th>
<th>New SC Std. Dev.</th>
<th>Old SC Mean</th>
<th>Old SC Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Years of Schooling</td>
<td>5.01</td>
<td>4.82</td>
<td>5.97</td>
<td>4.84</td>
</tr>
<tr>
<td>Literacy</td>
<td>0.61</td>
<td>0.49</td>
<td>0.68</td>
<td>0.47</td>
</tr>
<tr>
<td>Schooling</td>
<td>0.64</td>
<td>0.48</td>
<td>0.72</td>
<td>0.45</td>
</tr>
<tr>
<td>Primary Completion</td>
<td>0.51</td>
<td>0.50</td>
<td>0.61</td>
<td>0.49</td>
</tr>
<tr>
<td>Incomplete Secondary</td>
<td>0.45</td>
<td>0.50</td>
<td>0.53</td>
<td>0.50</td>
</tr>
<tr>
<td>Secondary Completion</td>
<td>0.20</td>
<td>0.40</td>
<td>0.26</td>
<td>0.44</td>
</tr>
<tr>
<td>Higher Education Completion</td>
<td>0.13</td>
<td>0.34</td>
<td>0.16</td>
<td>0.37</td>
</tr>
</tbody>
</table>

**Control variables**

| Urban                               | 0.24        | 0.43             | 0.30        | 0.46             |
| Hindu                               | 0.99        | 0.11             | 0.91        | 0.29             |
| Christian                           | 0.01        | 0.09             | 0.03        | 0.17             |
| Sikh                                | 0.00        | 0.00             | 0.00        | 0.05             |
| Buddhist/Neo Buddhist               | 0.00        | 0.00             | 0.06        | 0.23             |
| No Religion                         | 0.00        | 0.04             | 0.00        | 0.01             |
| Declares to be SC                   | 0.68        | 0.47             | 0.82        | 0.38             |

N = 680

### Table 2: Descriptive Statistics: Female population.

<table>
<thead>
<tr>
<th>Education variables</th>
<th>New SC Mean</th>
<th>New SC Std. Dev.</th>
<th>Old SC Mean</th>
<th>Old SC Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Years of Schooling</td>
<td>2.11</td>
<td>4.00</td>
<td>2.87</td>
<td>4.25</td>
</tr>
<tr>
<td>Literacy</td>
<td>0.26</td>
<td>0.44</td>
<td>0.35</td>
<td>0.48</td>
</tr>
<tr>
<td>Schooling</td>
<td>0.28</td>
<td>0.45</td>
<td>0.39</td>
<td>0.49</td>
</tr>
<tr>
<td>Primary Completion</td>
<td>0.21</td>
<td>0.41</td>
<td>0.30</td>
<td>0.46</td>
</tr>
<tr>
<td>Incomplete Secondary</td>
<td>0.19</td>
<td>0.39</td>
<td>0.25</td>
<td>0.43</td>
</tr>
<tr>
<td>Secondary Completion</td>
<td>0.08</td>
<td>0.28</td>
<td>0.11</td>
<td>0.31</td>
</tr>
<tr>
<td>Higher Education Completion</td>
<td>0.05</td>
<td>0.23</td>
<td>0.07</td>
<td>0.25</td>
</tr>
</tbody>
</table>

**Individual and household variables**

| Urban                               | 0.20        | 0.40             | 0.29        | 0.45             |
| Hindu                               | 0.98        | 0.12             | 0.90        | 0.29             |
| Christian                           | 0.01        | 0.11             | 0.03        | 0.17             |
| Sikh                                | 0.00        | 0.00             | 0.00        | 0.05             |
| Buddhist/Neo Buddhist               | 0.00        | 0.00             | 0.06        | 0.24             |
| No Religion                         | 0.00        | 0.04             | 0.00        | 0.00             |
| Declares to be SC                   | 0.67        | 0.47             | 0.82        | 0.38             |

N = 653

8
2.2 Identification Strategy

As the SC are a different population from the general population, much poorer in particular, comparing the “new” SC to the general population does not allow to identify the specific effect of the access to the SC status on educational attainment from other social policies. Hence, the natural counterfactual are the SC already on the lists in 1976. Due the specific setting of the natural experiment, I am able to compare within the same caste, within the same state, the fate of those who had access to the SC status in 1976 against those who had access to it earlier. This calls for a difference in differences specification, the first difference comparing cohorts too old to benefit from the access to the SC status in 1977\textsuperscript{21} in terms of education to those that were young enough. Throughout the paper, I will consider as “young enough” the cohorts aged either 6 or 11 years old and below in 1977. The second difference would then compare those who had access to the SC status prior to 1977 to those that did not.

I will thus run regressions of the type:

\[
Edu_{idt} = constant + \beta NewSC_{id} + \delta NewSC_{id} \times posttreatment_t + \gamma posttreatment_t + \lambda X_{idt} + \epsilon_{idt}
\]

(1)

Where \( Edu_{idt} \) is a measure of educational attainment of individual \( i \) born in year \( t \) and residing in region \( d \), \( NewSC_{id} \) a dummy indicating whether individual \( i \) residing in region \( d \) is member of a caste added to the SC list in 1976, \( posttreatment_t \) a dummy taking value 1 if individual \( i \) is in a treated cohort and \( X_{idt} \) a set of control variables.

Note that since the treatment is attributed at the caste and region level, this indicates the need to two way cluster the standard errors along those two dimensions. As there are 31 regions, I check the “robustness” of the standard errors by two-way clustering and bootstrapping at the region cluster level using the method proposed by Cuskey (2015). All tables present simple two-way clustered standard errors as well as 95% confidence intervals calculated with Cuskey (2015), which do not alter the statistical significance of the results.

\textsuperscript{20} Using the Indian Administrative Atlas of 2001, which follows district changes over time.

\textsuperscript{21} Results available on request.

\textsuperscript{21} Year of implementation of the 1976 change.
3 Results

3.1 Main specifications

The identification strategy of this paper relies on within caste variation in the exposure to the SC status. The specifications use caste-5 years cohort FE and region-5 years cohort FE. That is, the estimation is solely identified on within region-cohort and within caste-cohort variation. However, the first cohort to be exposed to the treatment is not well defined a priori. Indeed, treated individuals are the ones that would still be at school at the time of the reform. Since I do not have that information (I observe only the schooling choices after the reform has been implemented, which is endogenous) I resort to a rule of thumb. I assume that the first treatment cohort would be the cohort that reaches the median age of schooling in 1977, that is cohort 1966 for males and cohort 1971 for females. I then repeat the regression twice, first with 1966 as the cut off years and second with 1971.

Tables 3 presents the results of those specifications. Access to the SC status does not seem to affect the level of education much. However, once the population is divided by gender, very different effects appear. Indeed, it can be seen that the SC status leads to an increase of 1 to 2.1 years of schooling for males, spread across pre secondary schooling levels. For women, access to the SC status leads to a decrease of 0.7 to 1 year of schooling, driven by a decrease in primary school completion. Note that this is a relative decline: the context is one of overall increase of schooling, so the treated group does not see its absolute level of education decrease. Note also that the treatment effect seems larger for males when the treatment year is 1966 instead of 1971, which points to the fact that male cohorts were treated earlier than 1971, while the contrary is true for females. I will test more formally that insight in the next subsection.

Hence, the overall null effect of the access to the SC status on educational attainment hides asymmetrical effects by gender which cancel each other out in the aggregate. Indeed, while males do benefit from the SC status, it seems that females do not, and may actually lose from the status.

3.2 Varying treatment year

We have seen that male’s treatment effect seemed to be higher when the treated cohort was assumed to be 1971 instead of 1966, and that the contrary was true for female. This

\[22\]\footnote{11 years old in 1977 for a median of 5 years of schooling for the cohorts born up to 1971.}

\[23\]\footnote{6 years old in 1977, for a median of 0 years of schooling for the cohorts born up to 1971.}
Table 3: Educational attainment. OLS regressions.

<table>
<thead>
<tr>
<th></th>
<th>Years of Schooling</th>
<th>Schooling</th>
<th>Primary</th>
<th>Literate</th>
<th>Some Secondary</th>
<th>Secondary</th>
<th>Higher Education</th>
</tr>
</thead>
<tbody>
<tr>
<td>New SC* post 1966</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full Sample</td>
<td>0.831*</td>
<td>0.006</td>
<td>0.077</td>
<td>0.056</td>
<td>0.113***</td>
<td>0.056</td>
<td>0.024</td>
</tr>
<tr>
<td></td>
<td>[0.450]</td>
<td>[0.057]</td>
<td>[0.050]</td>
<td>[0.050]</td>
<td>[0.040]</td>
<td>[0.041]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.09, 1.48]</td>
<td>[-0.1, 0.11]</td>
<td>[-0.01, 0.16]</td>
<td>[-0.02, 0.13]</td>
<td>[0.04, 0.18]</td>
<td>[-0.01, 0.12]</td>
<td>[-0.04, 0.09]</td>
</tr>
<tr>
<td></td>
<td>0.061</td>
<td>-0.012</td>
<td>0.025</td>
<td>0.016</td>
<td>0.034</td>
<td>0.006</td>
<td>-0.032</td>
</tr>
<tr>
<td></td>
<td>[0.526]</td>
<td>[0.065]</td>
<td>[0.053]</td>
<td>[0.063]</td>
<td>[0.038]</td>
<td>[0.055]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[-0.85, 0.83]</td>
<td>[-0.12, 0.08]</td>
<td>[-0.06, 0.11]</td>
<td>[-0.08, 0.11]</td>
<td>[-0.3, 0.09]</td>
<td>[-0.08, 0.09]</td>
<td>[-0.11, 0.45]</td>
</tr>
<tr>
<td></td>
<td>N 14609</td>
<td>14609</td>
<td>14609</td>
<td>14609</td>
<td>14609</td>
<td>14609</td>
<td>14609</td>
</tr>
<tr>
<td>Male Sample</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New SC* post 1966</td>
<td>2.164***</td>
<td>0.13</td>
<td>0.261***</td>
<td>0.206**</td>
<td>0.227***</td>
<td>0.086</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td>[0.663]</td>
<td>[0.085]</td>
<td>[0.075]</td>
<td>[0.090]</td>
<td>[0.057]</td>
<td>[0.068]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[1.09, 3.09]</td>
<td>[-0.005, 0.26]</td>
<td>[0.13, 0.39]</td>
<td>[0.06, 0.35]</td>
<td>[0.12, 0.33]</td>
<td>[-0.02, 0.19]</td>
<td>[-0.07, 0.10]</td>
</tr>
<tr>
<td></td>
<td>0.988*</td>
<td>0.077</td>
<td>0.146**</td>
<td>0.158**</td>
<td>0.119**</td>
<td>0.002</td>
<td>-0.08</td>
</tr>
<tr>
<td></td>
<td>[0.518]</td>
<td>[0.054]</td>
<td>[0.060]</td>
<td>[0.063]</td>
<td>[0.050]</td>
<td>[0.097]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.07, 1.24]</td>
<td>[-0.02, 0.17]</td>
<td>[0.05, 0.24]</td>
<td>[0.05, 0.26]</td>
<td>[0.05, 0.2]</td>
<td>[-0.15, 0.14]</td>
<td>[-0.20, 0.03]</td>
</tr>
<tr>
<td></td>
<td>N 7320</td>
<td>7320</td>
<td>7320</td>
<td>7320</td>
<td>7320</td>
<td>7320</td>
<td>7320</td>
</tr>
<tr>
<td>Female Sample</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New SC* post 1966</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.684</td>
<td>-0.113</td>
<td>-0.111**</td>
<td>-0.071</td>
<td>-0.035</td>
<td>0</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td>[0.443]</td>
<td>[0.083]</td>
<td>[0.050]</td>
<td>[0.060]</td>
<td>[0.034]</td>
<td>[0.035]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[-1.44, -0.02]</td>
<td>[0.25, 0.04]</td>
<td>[-0.19, -0.03]</td>
<td>[-0.17, 0.03]</td>
<td>[-0.09, 0.02]</td>
<td>[-0.05, 0.05]</td>
<td>[-0.05, 0.05]</td>
</tr>
<tr>
<td></td>
<td>-1.003</td>
<td>-0.09</td>
<td>-0.114**</td>
<td>-0.081</td>
<td>-0.095*</td>
<td>-0.016</td>
<td>-0.043</td>
</tr>
<tr>
<td></td>
<td>[0.619]</td>
<td>[0.077]</td>
<td>[0.053]</td>
<td>[0.066]</td>
<td>[0.051]</td>
<td>[0.050]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[-2.10, -0.05]</td>
<td>[-0.21, 0.04]</td>
<td>[-0.2, -0.03]</td>
<td>[-0.19, 0.02]</td>
<td>[-0.18, -0.01]</td>
<td>[-0.1, 0.06]</td>
<td>[-0.11, 0.03]</td>
</tr>
<tr>
<td></td>
<td>N 7289</td>
<td>7289</td>
<td>7289</td>
<td>7289</td>
<td>7289</td>
<td>7289</td>
<td>7289</td>
</tr>
</tbody>
</table>

*Significant at the 10%, **significant at the 5%, ***significant at the 1%. Standard errors two way clustered at the jati and region level in parenthesis. 95% confidence interval from two way clustering at the jati and region level, bootstrapped at the region level in brackets. Controls include regions FE, state-cohort FE, jati FE, gender FE and religion of household head FE. Population born after 1950 and aged 17 and above at the time of the survey.
has a logical interpretation: given the education differential between boys and girls, much less women were still at school at age 11 in 1977 compared to males. As a result, the cohort of treatment should vary across gender. To test more formally this intuition, I will allow the treatment effect to vary both for cohorts born in 1966 and after and for cohorts born in 1971 and after, running the following regression:

\[
Edu_{idt} = \text{constant} + \beta_{\text{NewSC}_{id}} + \delta_{1}\text{NewSC}_{id} \ast \text{post1966}_t + \gamma_{1}\text{post1966}_t
+ \delta_{2}\text{NewSC}_{id} \ast \text{post1971}_t + \gamma_{2}\text{post1971}_t + \lambda X_{idt} + \epsilon_{idt}
\] (2)

This will allow me to test if indeed males are treated in earlier cohorts that females. Table 6 presents the results of such a specification. It can be seen that while the female result loses significance, the size of the coefficient is larger for 1971 than for 1966, suggesting that if anything, women from later cohorts are more affected than older cohort. For males, it is clear that the affected cohort start after from 1966 and not 1971. This reinforces our conclusion that the results are indeed driven by the access to the SC status: if the access to the SC status is to have an impact on school choices, it can only affect the cohorts still at school at the time of its implementation. As a result, since females leave school much earlier than males, cohorts of females born between 1966 and 1971 should be much less affected than their male counterpart.

Why such large differences across genders? Women in India are generally at a disadvantage in terms of schooling. For example, in the sample, women have on average 2.8 years of schooling as opposed to 5.8 for their males counterpart. One of the main reasons usually put forward to explain this difference are the lower returns to schooling (Kingdon, 1998), or lower perceived returns to schooling (Dreze and Sen, 2002) for females. Indeed, in the NFHS2 for example, women’s participation in the labour force is much smaller than males’, with 63% of women “not classified by occupations”\(^{24}\) as opposed to 14% of males. As a result, women’s returns to education are often perceived as nonexistent. Moreover, as marriage practices are generally patrilocal, the perception of the investment in the daughter’s skills is often seen as a waste from the point of view of the parents. Dreze and Sen (2002) write for example: “[...] the gender division of labour [...] tends to reduce the perceived benefits of female education. [...] It is in the light of these social expectations about the adult life of women that female education appears

\(^{24}\)Which essentially means being a housewife.
Table 4: Educational attainment: varying treatment year. OLS regressions.

<table>
<thead>
<tr>
<th></th>
<th>Years of Schooling</th>
<th>Schooling</th>
<th>Primary</th>
<th>Literate</th>
<th>Some Secondary</th>
<th>Secondary</th>
<th>Higher Education</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Full Sample</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New SC* post 1966</td>
<td>1.473*** 0.028</td>
<td>0.110*</td>
<td>0.082</td>
<td>0.165***</td>
<td>0.097**</td>
<td>0.088**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.399] [0.664]</td>
<td>[0.069]</td>
<td>[0.062]</td>
<td>[0.083]</td>
<td>[0.051]</td>
<td>[0.049]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[-0.06, 0.21]</td>
<td>[-0.06, 0.21]</td>
<td>[-0.07, 0.27]</td>
<td>[-0.01, 0.2]</td>
<td>[0.01, 0.17]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>New SC* post 1971</td>
<td>-0.873 -0.03</td>
<td>-0.045</td>
<td>-0.036</td>
<td>-0.07</td>
<td>-0.056</td>
<td>-0.087</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[-0.593] [-0.082]</td>
<td>[-0.067]</td>
<td>[-0.096]</td>
<td>[-0.051]</td>
<td>[-0.075]</td>
<td>[-0.059]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[-1.48, 0.9]</td>
<td>[-1.51, 0.06]</td>
<td>[-1.61, 0.12]</td>
<td>[-1.19, 0.06]</td>
<td>[-1.19, 0.02]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>14609</td>
<td>14609</td>
<td>14609</td>
<td>14609</td>
<td>14609</td>
<td>14609</td>
<td></td>
</tr>
</tbody>
</table>

|                          | Male Sample       |           |         |          |                |           |                  |
| New SC* post 1966        | 2.521*** 0.13     | 0.272**   | 0.169   | 0.245*** | 0.140*         | 0.107***  |
|                          | [0.745] [0.132]   | [0.109]   | [0.160] | [0.069]  | [0.075]        | [0.050]   |
|                          | [-1.20, 3.78]     | [-0.09, 0.46] | [-0.09, 0.41] | [-0.11, 0.38] | [-0.00, 0.29] |
| New SC* post 1971        | -0.54 0           | -0.017    | 0.057   | -0.027   | -0.082         | -0.144*   |
|                          | [-0.640] [0.101]  | [-0.092] | [0.133] | [0.065]  | [0.124]        | [0.084]   |
|                          | [-1.76, 0.66]     | [-0.16, 0.12] | [-0.15, 0.26] | [-0.14, 0.09] | [-0.3, 0.12] |
| N                        | 7320              | 7320      | 7320    | 7320     | 7320           | 7320      |

|                          | Female Sample     |           |         |          |                |           |                  |
| New SC* post 1966        | 0.214 -0.09       | -0.044    | -0.016  | 0.083    | 0.026          | 0.065     |
|                          | [0.562] [0.090]   | [0.072]   | [0.090] | [0.065]  | [0.055]        | [0.053]   |
|                          | [-0.57, 1.02]     | [-0.16, 0.07] | [-0.18, 0.14] | [-0.01, 0.18] | [-0.06, 0.12] |
| New SC* post 1971        | -1.145 -0.03      | -0.085    | -0.07   | -0.150*  | -0.034         | -0.086    |
|                          | [-0.831] [-0.085] | [-0.075] | [-0.098] | [0.082]  | [0.075]        | [0.069]   |
|                          | [-2.58, 0.14]     | [-0.21, 0.03] | [-0.23, 0.1] | [-0.28, -0.02] | [-0.16, 0.09] |
| N                        | 7289              | 7289      | 7289    | 7289     | 7289           | 7289      |

*Significant at the 10%, **significant at the 5%, ***significant at the 1%. Standard errors two way clustered at the jati and region level in parenthesis. 95% confidence interval from two way clustering at the jati and region level, bootstrapped at the region level in brackets. Controls include regions FE, state-cohort FE, jati FE, gender FE and religion of household head FE. Population born after 1950 and aged 17 and above at the time of the survey.
to many parents to be of somewhat uncertain value, if not quite ‘pointless’.2526

The NFHS2 survey contains a question on the reason why a person did not attend school. Table 5 presents the responses to this question by gender. The main reasons for males are first and foremost the cost, and then that the child is required for household work or work on farm or family business. For females, the picture is strikingly different: the main reasons for which they do not attend school is because it is not considered necessary, closely followed by the fact that they are required for household work, and that schooling costs too much. Hence both cost and opportunity costs appear to be a major reason for which children are not sent to school, with the additional issue that often, education is not considered necessary for girls, which confirms the qualitative evidence discussed earlier.

Table 5: Reason not going to school by gender.

<table>
<thead>
<tr>
<th>main reason never went to school</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>school too far away</td>
<td>3.38</td>
<td>2.44</td>
</tr>
<tr>
<td>transport not available</td>
<td>0.91</td>
<td>0.76</td>
</tr>
<tr>
<td>education not considered necessary</td>
<td>23.69</td>
<td>11.94</td>
</tr>
<tr>
<td>required for household work</td>
<td>20</td>
<td>12.76</td>
</tr>
<tr>
<td>required for work on farm/family business</td>
<td>3.71</td>
<td>12.47</td>
</tr>
<tr>
<td>required for outside work for payment</td>
<td>2.6</td>
<td>8.03</td>
</tr>
<tr>
<td>cost too much</td>
<td>19.84</td>
<td>25.56</td>
</tr>
<tr>
<td>no proper school facilities for girls</td>
<td>3.45</td>
<td></td>
</tr>
<tr>
<td>required for care of siblings</td>
<td>2.49</td>
<td>0.72</td>
</tr>
<tr>
<td>not interested in studies</td>
<td>6.85</td>
<td>12.47</td>
</tr>
<tr>
<td>other</td>
<td>5</td>
<td>4.3</td>
</tr>
<tr>
<td>don’t know</td>
<td>7.65</td>
<td>8.55</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>N</td>
<td>4 496</td>
<td>2 093</td>
</tr>
</tbody>
</table>

In this context of lower perceived returns from education for girls from the point of view of the parental household, a gender neutral increase in returns to education and decrease in costs of education might have different effects by gender. Indeed, education of the children is costly, both in terms of direct cost and opportunity costs. As a result, the perception by the parents that returns might be gender-asymmetric either because

25Dreze and Sen (2002) also underline a second important reason for the low enrolment of women: the low quality of schooling infrastructures. Indeed while, say, the absence of a nearby school could be thought to be gender neutral, in practice, girls are more affected, as parents tend to be more reluctant to send their daughters to far away schools than their sons. Finally, an other reason put forward by Dreze and Sen (2002) for the lower education of girls is that even if education can be an asset on the marriage market, it can only be so if the girl’s education remains lower than that of her potential husband.

26Note that this tendency is in no way specific to the lower castes in India. Actually, lower castes tend to be less gender biased than higher castes (Srinivas (1966), Chakravarti (1993), Mencher (1988), Kapadia (1997), Field et al. (2010), Cassan and Van de Walle (2016)).
they do not believe that returns to schooling exist for girls, or because they benefit less
from the returns to schooling of girls than that of boys is sufficient a mechanism to drive
those type of behaviour. In addition, under imperfect access to credit and budgetary
constraints, a household may not have the resources to increase the education of both
genders, and favour the ones with highest perceived returns to education. In the most
extreme cases, this may even lead to a reallocation of resources from female to male child,
leading to an actual decrease of education of females. Banerjee and Duflo (2011) discuss
a similar type of reallocation of schooling investment across children. Hence, a plausible
interpretation of the asymmetric results by gender of the access to the affirmative action
program are gender norms.

4 Robustness check

4.1 Differential trend

A concern with any difference in difference estimation are differential trends. A first
check is to examine visually the data. I run the following regression:

\[
Edu_{idt} = \text{constant} + \beta \text{NewSC}_{id} + \delta_1 \text{NewSC}_{id} \ast C_t + \gamma_1 C_t
+ \lambda S_{it} + \epsilon_{idt}
\] (3)

In which \(C_t\) is a vector of 5 years cohorts and \(S_{it}\) a vector of state fixed effects. Figure
1 plots the coefficients on \(\text{NewSC}_{id} \ast C_t\) of these regressions. One can see that there
does not seem to be any major pre-trend driving the results, while both the increase for
males and the decrease for females are clear for treated cohorts.

Let’s now control more formally for pre-trend by adding a “new” SC specific trend
to the specification. I run the following regressions:

\[
Edu_{idt} = \text{constant} + \beta \text{NewSC}_{id} + \delta \text{NewSC}_{id} \ast \text{posttreatment}_t + \gamma \text{posttreatment}_t
\rho \text{NewSC}_{id} \ast \text{trend}_t + \pi \text{trend}_t + \lambda X_{idt} + \epsilon_{idt}
\] (4)

The notation is the same as earlier and \(\text{trend}_t\) stands for a linear time trend. As can
be seen in Table 6, the results on males become larger but less precisely estimated.
For females however, the negative results, already only weakly significant, seem to both
decrease in magnitude and lose significance. As a consequence, the negative result on
female can be interpreted as evidence that women did not benefit from the SC status
Figure 1: Cohort specific effects. Coefficients of the interaction terms “New SC”* years cohorts dummies (controlling for state FE). Population born after 1950 and aged 18 and above at time of survey.

for educational attainment, with only weak evidence that they could have lost from it.
Table 6: Educational attainment: varying treatment year and including differential trend. OLS regressions.

<table>
<thead>
<tr>
<th>Years of Schooling</th>
<th>Schooling</th>
<th>Primary</th>
<th>Literate</th>
<th>Some Secondary</th>
<th>Secondary</th>
<th>Higher Education</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Male Sample</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New SC* post 1966</td>
<td>3.018**</td>
<td>0.355**</td>
<td>0.308*</td>
<td>0.305*</td>
<td>0.224</td>
<td>0.139</td>
</tr>
<tr>
<td></td>
<td>[1.380]</td>
<td>[0.164]</td>
<td>[0.181]</td>
<td>[0.167]</td>
<td>[0.153]</td>
<td>[0.089]</td>
</tr>
<tr>
<td></td>
<td>[0.79, 5.16]</td>
<td>[0.1, 0.59]</td>
<td>[0.05, 0.57]</td>
<td>[0.04, 0.57]</td>
<td>[-0.02, 0.48]</td>
<td>[-0.06, 0.35]</td>
</tr>
<tr>
<td>New SC* post 1971</td>
<td>-0.267</td>
<td>0.171</td>
<td>0.16</td>
<td>-0.054</td>
<td>-0.092</td>
<td>-0.206**</td>
</tr>
<tr>
<td></td>
<td>[-1.79, 1.38]</td>
<td>[-0.13, 0.34]</td>
<td>[-0.18, 0.16]</td>
<td>[-0.08, 0.42]</td>
<td>[-0.21, 0.09]</td>
<td>[-0.32, 0.14]</td>
</tr>
<tr>
<td>New SC* post 1966</td>
<td>3.010**</td>
<td>0.366**</td>
<td>0.309*</td>
<td>0.315*</td>
<td>0.221</td>
<td>0.134</td>
</tr>
<tr>
<td></td>
<td>[1.395]</td>
<td>[0.166]</td>
<td>[0.182]</td>
<td>[0.164]</td>
<td>[0.156]</td>
<td>[0.087]</td>
</tr>
<tr>
<td></td>
<td>[0.77, 5.16]</td>
<td>[0.1, 0.6]</td>
<td>[0.05, 0.57]</td>
<td>[0.07, 0.57]</td>
<td>[-0.02, 0.47]</td>
<td>[-0.06, 0.33]</td>
</tr>
<tr>
<td>New SC* post 1971</td>
<td>-0.134</td>
<td>0.188*</td>
<td>0.014</td>
<td>0.174</td>
<td>-0.045</td>
<td>-0.086</td>
</tr>
<tr>
<td></td>
<td>[-1.74, 1.53]</td>
<td>[0.03, 0.34]</td>
<td>[-0.16, 0.17]</td>
<td>[-0.07, 0.42]</td>
<td>[-0.2, 0.11]</td>
<td>[-0.32, 0.15]</td>
</tr>
<tr>
<td>N</td>
<td>7320</td>
<td>7320</td>
<td>7320</td>
<td>7320</td>
<td>7320</td>
<td>7320</td>
</tr>
</tbody>
</table>

| **Female Sample**  |           |         |          |                |           |                 |
| New SC* post 1966  | 0.202     | -0.022  | -0.38    | -0.029         | 0.095     | -0.008          | 0.021           |
|                    | [0.686]   | [0.115] | [0.097]  | [0.116]        | [0.065]   | [0.055]         | [0.052]         |
|                    | [-0.94, 1.41] | [-0.24, 0.20] | [-0.19, 0.13] | [-0.25, 0.18] | [-0.03, 0.21] | [-0.11, 0.09] | [-0.06, 0.10] |
| New SC* post 1971  | -0.908    | 0.022   | -0.072   | -0.082         | -0.099    | -0.055          | -0.100*         |
|                    | [0.900]   | [0.098] | [0.089]  | [0.116]        | [0.097]   | [0.067]         | [0.060]         |
|                    | [-2.70, 0.71] | [-0.15, 0.19] | [-0.24, 0.09] | [-0.28, 0.13] | [-0.27, 0.60] | [-0.18, 0.07] | [-0.20, 0.01] |
| New SC* post 1966  | 0.396     | -0.027  | -0.24    | -0.023         | 0.119*    | 0.003           | 0.042           |
|                    | [0.709]   | [0.113] | [0.096]  | [0.117]        | [0.067]   | [0.063]         | [0.060]         |
|                    | [-0.75, 1.55] | [-0.24, 0.19] | [-0.17, 0.13] | [-0.23, 0.19] | [0.01, 0.23] | [-0.1, 0.11] | [-0.06, 0.01] |
| New SC* post 1971  | -0.975    | 0.026   | -0.067   | -0.078         | -0.119    | -0.055          | -0.107          |
|                    | [0.916]   | [0.097] | [0.089]  | [0.117]        | [0.100]   | [0.073]         | [0.066]         |
|                    | [-2.75, 0.65] | [-0.13, 0.18] | [-0.23, 0.09] | [-0.27, 0.13] | [-0.29, 0.04] | [-0.2, 0.08] | [-0.22, 0.00] |
| N                  | 7289      | 7289    | 7289     | 7289           | 7289      | 7289            |

*Significant at the 10%, **significant at the 5%, ***significant at the 1%. Standard errors two way clustered at the jati and region level in parenthesis. 95% confidence interval from two way clustering at the jati and region level, bootstrapped at the region level in brackets. Controls include regions FE, state-cohort FE, jati FE, gender FE and religion of household head FE. Population born after 1950 and aged 17 and above at the time of the survey.
4.2 Selective migration and identity manipulation

This section will deal with two additional concerns: migration and identity manipulation. Empirically, I attribute the “New SC” status based on the district of residence of respondents at the time of the survey. Therefore, if prior to 1976, some households have migrated in order to benefit from the SC status, I would wrongly code them as “Old SC”, which may bias the results. However, migration is relatively low in India (Munshi and Rosenzweig, 2009) and particularly so in the 1970’s and earlier. Indeed, from the Migration Volumes of the 1981 Census, one can estimate that only 6.1% of the males and 11.1% of females had changed district before 1977. A similar argument can be made about identity manipulation (Cassan, 2015): people may well have been incited to manipulate their caste identity in order to be considered as eligible to the SC status before 1976. However, as for migration, I do not believe that this behaviour may bias the results. Indeed, first and foremost, as the data is collected in 1999, 25 years after the harmonisation of the SC lists, there are actually no incentive to manipulate caste identity at the time of survey. Second, one may worry that individuals who manipulated their caste identity prior to 1976 permanently changed their jati identity. This second pattern would result in a miscoding of the treatment status very similar to the one due to migration: individuals that should have been coded as “New SC” had they not migrated or manipulated their identity are coded as “Old SC”. To my knowledge, there is no estimates of the amount of identity manipulation that the SC status may lead to. However, Cassan (2015) shows that in the context of colonial Punjab, a caste based policy with arguably stronger impact on day to day life led to an identity manipulation movement of up to 7.5% of the population, which I’ll use as an upper bound on caste identity manipulation.

It is to be noted that the households choosing to migrate/manipulate would probably be the ones that would have benefited the most from the access to the SC status had they not migrated. This means that, if anything, migration/identity manipulation is likely to bias the estimates for males downwards and for females upwards, as I would consider those households as “Old SC”. As the NFHS2 data does not offer useful migration information, it is not possible to directly verify if the migrating/manipulating households indeed are different in any form from the other households. Despite the weak probability that migration or manipulation could be an issue for my estimation strategy,

\(^{27}\)Note that this is an overestimation of the type of migration that concerns us, since a bias might emerge if those migrants moved in a way that would change their SC status, which is not systematically the case when crossing a district border since regions are larger than districts.

\(^{28}\)And obviously no information about identity manipulation.
I will show in this subsection that the results hold even with very strong assumptions on the profiles of migrants/manipulators.

As a robustness check, I reallocate the individuals who may contribute the most to the results from control to treatment. I identify the individuals that are the most likely to contribute to bias the results if they had been migrants/manipulators. That is, for male, I identify “old SC” individuals born prior the first treatment cohort with an above median occupation, and the ones born after the first treatment cohort with a below median education (and symmetrically for female). Among these, I randomly draw a number corresponding to 33%, 66% and 100% of the total migrating/manipulating population\textsuperscript{29}, reallocate them to the “New SC” status and bootstrap 1000 times the main estimations at each of those steps. Figure 2 plots the coefficients of those estimations. Figure 2 plots the coefficient on New SC*post1966 (for males) and New SC*post1971 (for females). It can be seen that the results are very robust, since even under these strong and counter intuitive assumptions about the education profile of migrants, the male result remains positive even if 100% of migrants/manipulators had an above the median education, and remain significant up to 66% of the total population of migrants/manipulators.

5 Conclusion

This paper studies the impact of affirmative action policies for SC in India. It shows that they have an ambiguous effect on schooling attainment. Indeed, the overall null effect actually hides differences across genders, with male benefiting largely from those policies while female to the contrary seem not to be affected, or may even suffer from it. This suggests that individuals at the intersection of different discriminated groups may not be sufficiently well protected by public policies. More generally, this paper underlines the need to focus on the role of accumulation of discrimination, the question of “intersectionality”, which so far has been neglected in the economics literature.

\textsuperscript{29}I take the largest population between the estimates of migration and manipulation. That is migrants/manipulators that would account for 7.5% of the total male “New SC” and 11.1% of the total female “New SC”.

19
Figure 2: Reallocation of potential migrants and robustness of the coefficients.

(a) Male sample, coefficient on New SC*post1966

(b) Female sample, coefficient on New SC*post1971
References


**Srinivas**, *Caste in Modern India and Other Essays*, University of California Press, 1966.

### Appendices

#### A Areas of restriction

The following list presents state by state the different “regions” for which separate SC list were drawn until 1976. This information has been extracted from the 1971 Census’ Social and Cultural Tables, publicly available, which also contain the whole SC lists by region. It can be seen that certain “regions” are included in larger ones. In that case, in the empirical exercises of the paper, I took the most conservative approach, that is: each region is given a specific fixed effect, while for the computation of the clustered standard errors, those regions are included in larger ones (except is that would lead to create only one region for the state). In addition, in certain cases, only taluks are listed, not entire districts. In that case, I took the most conservative coding approach, which is to code the whole district as part of the region. In total, the paper uses 31 regions.

**Andhra Pradesh**

*Region A:* throughout the state:
Region B: districts of Srikakulam, Visakhapatnam, East Godavari, West Godavari, Krishna, Gunter, Ongole, Nellore, Chittoor, Cuddapah, Anantapur and Kurnool;
Region C: districts of Mahbubnagar, Hyderabad, Medak, Nizamabad, Adilabad, Karimnagar, Warangal, Khamman and Nalgonda.

Assam
No regional caste list. This state is not used in the paper.

Bihar
Region A: throughout the state;
Region B: Patna and Tirhut divisions and Monghyr, Bhagalpur, Palamay and Purnea districts; Region C: districts of Patna, Shahabad, Gaya and Palamau.

Gujarat
Region A: throughout the state except Rajkot division and Kutch district;
Region B: Dangs districts and Umbergaon taluka of Surat district;
Region C: Rajkot division;
Region D: Kutch district.

Haryana
Region A: throughout the state;
Region B: throughout the state except in Mahendragarh and Jind districts;
Region C: districts of Mahendragarh and Jind.

Himachal Pradesh
As Himachal Pradesh is carved out of Punjab in 1966, it is not used in the paper.
Region A: districts of Chamba (excluding the towns of Dalhousie MC, Dalhousie CB and Bakloh CB), Mandi, Bilaspur, Mahasu, Sirmaur and Kinnaur;
Region B: districts of Kangra, Kulu, Lahul & Spiti, Simla and the towns of Dalhousie MC, Dalhousie CB and Bakloh CB in Chamba district.

Jammu & Kashmir
No regional caste list. This state is not used in the paper.

Kerala
Region A: Throughout the state;
Region B: throughout the state except in Kasaragod taluka of Malabar district;
Region C: throughout the state except Malabar district (excluding Kasaragod taluka);
Region D: throughout the state except Malabar district;
Region E: Malabar district;
Region F: Malabar district (excluding Kasaragod taluka);
Region G: Kasaragod taluka of Malabar district.

Madhya Pradesh
Region A: districts of Bind, Gird, Morena, Shivpuri, Goona, Rajgarh, Shajapur, Ujjain, Ratlam, Mandsaur, Bhilsa, Indore, Dewas, Dhar, Jhabua and Nimar;
Region B: districts of Chhindwara, Seoni, Betul, Jabalpur, Sagar, Damoh, Mandla, Hoshangabad, Narsimhapur, Nimar, Balaghat, Raipur, Bilaspur, Durg, Rastar, Surguja and Raigarh;
Region C: Bilaspur district;
Region D: Sagar and Damoh districts;
Region E: Damoh district;
Region F: districts of Bilaspur, Durg, Raipur, Bastar, Surguja and Raigarh;
Region G: Sagar district;
Region H: Balaghat district;
Region I: districts of Balaghat, Bilaspur, Durg, Raipur, Surguja, Bastar and Raigarh;
Region J: districts of Balaghat, Betul, Bilaspur, Durg, Nimar, Raipur, Bastar, Surguja, Chhindwara, Sagar and Raigarh, Hoshangabad and Seoni Malwa thesils of Hoshangabad district;
Region K: districts of Sagar and Damoh, Hoshangabad and Seoni Malwa thesils of Hoshangabad district;
Region L: districts of Chhindwara, Seoni, Betul, Jabalpur, Narsimhapur, Sagar, Damoh, Mandla, Nimar, Balaghat, Raipur, Durg, Bastar, Surguja, Raigarh and Hoshangabad (except Harda and Sohagpur thesils);
Region M: districts of Chhindwara, Seoni, Betul, Jabalpur, Narsimhapur, Sagar, Damoh, Mandla, Nimar, Balaghat, Raipur, Bilaspur, Durg, Bastar, Surguja, Raigarh and Hoshangabad (except Harda and Sohagpur thesils);
Region N: Sohagpur thesil of Hoshangabad district;
Region O: districts of Datia, Tikamgarh, Chhatarpur, Panna, Stana, Rewa, Sidhi and Shahdol;
Region P: districts of Raisen and Sehore.

Maharashtra
Region A: throughout the state except Buldana, Akola, Amravati, Yeotmal, Wardha, Nagpur, Bhandara, Chanda, Aurangabad, Parbhani, Nanded, Bhir, Osmanabad and Rajura districts;
Region B: districts of Greater Bombay, Dhulia, Jalgaon, Nasik, Ahmednagar, Poona, Satara, Sangli, Kolhapur, Sholapur, Thana, Kolaba and Ratnagiri;
Region C: districts of Buldana, Akola, Amravati, Yeotmal, Wardha, Nagpur, Bhandara and Chanda;
Region D: districts of Akola, Amravati and Buldana;
Region E: Bhandara district;
Region F: districts of Bhandara and Buldana;
Region G: districts of Amravati, Bhandara and Buldana;
Region H: districts of Aurangabad, Parbhani, Nanded, Rajura, Bhir and Osmanabad.

Manipur
No regional caste list. This state is not used in the paper.

Meghalaya
No regional caste list. This state is not used in the paper.

Karnataka
Region A: throughout the state except Coorg, Belgaum, Bijapur, Dharwar, Kanara, South Kanara, Gulbarga, Raichur and Bidar districts, and Kollegal talukla of Mysore district;
Region B: districts of Belgaum, Bijapur, Dharwar and Kanara;
Region D: Kanara district;
Region E: districts of Gulbarga, Bidar and Raichar;
Region F: district of South Kanara and Kollegal taluka of Mysore district;
Region G: Kollegal taluka of Mysore district;
Region H: district of South Kanara;
Region I: Coorg district.

Orissa
Region A: throughout the state;
Region B: Sambalpur district.

Punjab
Even though there were sub regions, those were not removed in 1976, so Punjab is not used in this paper.
Region A: throughout the state;
Region B: throughout the state except Patiala, Bhatinda, Kapurthala and Sangrur districts;
Region C: districts of Patiala, Bhatinda, Kapurthala and Sangrur districts.

Rajasthan
Region A: throughout the state except Ajmer district, Abu Road taluka of Sirohi district and Sunel Tappa taluka of Jhalawar district;
Region B: Ajmer district;
Region C: Abu Road taluka of Sirohi district;
Region D: Sunel Tappa taluka of Jhalawar district.

Tamil Nadu
Region A: throughout the state;
Region B: throughout the state except Kanyakumari district and Shencottah taluka of Tirunelveli district;
Region C: Nilgiri district;
Region D: districts of Coimbatore and Salem;
Region E: Kanyakumari district and Shencottah taluka of Tirunelveli district;
Region F: Tanjore district.

Tripura
No regional caste list. This state is not used in the paper.

Uttar Pradesh
Region A: throughout the state; Region B: throughout the state excluding Agra, Meerut and Rohilkhand divisions;
Region C: Bundelkhand division and the portion of Mirzapur district south of Kaimur range.

West Bengal
Region A: throughout the state;
Region B: throughout the state except in Purulia district and the territories transferred from Purulia district of Bihar;
Region C: in Purulia district and the territories transferred from Purulia district of Bihar;
Region D: the territories transferred from Purulia district of Bihar.

Arunachal Pradesh
No regional caste list. This state is not used in the paper.

Chandigarh
No regional caste list. This state is not used in the paper.

Dadra and Nagar Haveli
No regional caste list. This state is not used in the paper.

Delhi
No regional caste list. This state is not used in the paper.

Goa, Daman and Diu
No regional caste list. This state is not used in the paper.

Mizoram
No regional caste list. This state is not used in the paper.

Pondicherry
No regional caste list. This state is not used in the paper.