Does Affirmative Action Impact Inter-Generational Mobility? Evidence from India

Manisha Jain [*]

September 15, 2022

ABSTRACT

I estimate the causal impact of an affirmative action policy, the implementation of job quotas for a socio-economically disadvantaged group in India, the Other Backward Castes (OBCs), in 1993, on their inter-generational education and occupation mobility. I use data drawn from a nationally representative survey, the Indian Human Development Survey (2011-12), and employ a difference-in-differences strategy motivated by the fact that only OBCs of a school or college-going age in 1993 could potentially be impacted. Older OBCs would have already made human capital investment decisions that would be impossible or too costly to change in response to the policy. I study various measures of absolute education mobility and absolute occupation mobility. I find that the job quotas resulted in an increase in the absolute upward mobility of OBC sons, as measured by (1) an increased probability that their education is greater than their father's, (2) an increase in the years of education of sons born to illiterate fathers, (3) an increase in the probability that a son born to an illiterate father is literate, (4) a higher expected education rank of sons born to below-median-education fathers, and (5) a higher probability that a son born to a below-median-education father ends up in the top half of their education distribution.

1 Introduction

Do affirmative action policies help in improving Inter-Generational Mobility (IGM) of the communities targeted by such policies? I address this question in the context of the implementation of the Mandal Commission reforms, which were a series of reservation policies that introduced a 27% reservation of seats in central government jobs and colleges for the Other Backward Castes (OBCs), a historically disadvantaged group in India, in 1993 and 2006, respectively. In this paper, I focus only on the impact of the job quotas.

Inter-generational mobility, or the degree to which the socio-economic status of a generation persists, is often used as an indicator of the equality of economic opportunities in a society (e.g., Restuccia

[*]Ph.D. Student, Dept. of Economics, University of Wisconsin-Madison

I wish to thank Corina Mommaerts, Jeffrey Smith, Matt Wiswall, Chris Taber, Jesse Gregory, Martin O'Connell, Naoki Aizawa,and Jean-Francois Houde for their suggestions and feedback. All remaining errors are my own.

and Urrutia (2004); Piketty (2000)). Understanding and measuring this degree of persistence, as well as unravelling the respective contributions of the transmission of innate abilities, family background, and economic policy in generating this persistence, has been one of the most controversial issues, not only in actual political conflicts, but also in academic writings by social scientists (e.g.,Lee and Seshadri (2019); Piketty (2000)). The ultimate aim of affirmative action policies is to ensure equal access of opportunities by all members of the society; hence, it is a natural question to study their effect on inter-generational mobility.

There are at least five motivations for studying the impact of affirmative action laws on inter-generational mobility in the Indian context. First, affirmative action policies in India are larger, more aggressive, and more salient than in most other countries (Khanna, 2020). Second, over time, reservation policies have increased in their scope, rather than decreased. Third, these policies are widely discussed in households and often dominate the public discussion regarding education policies in India. Public colleges and jobs, in which these affirmative action policies are implemented, are still considered very attractive and prestigious. Thus, I expect households to be well aware of these highly-publicized policies and take them into consideration when making human capital formation decisions for their children. Fourth, studies on IGM are rare in India, even more so for studies that look at causal impacts. Fifth, studying inter-generational mobility against the background of the social architecture shaped by the caste system is even more interesting as the caste system served as a natural barrier to IGM by restricting individuals to their traditional caste-based occupations and preventing them from marrying outside their caste community.[1] Each caste was traditionally linked to (usually) a single occupation (Cassan et al., 2021), and individuals were fated to follow that occupation by virtue of their birth.

Even today, caste as an institution plays a vital role in all stages of an individual's education and labor market outcomes in India, by determining the quality of school and college that they have access to, how they are treated within that school or college by their teachers and peers, whether they can benefit from affirmative action in access to colleges and government jobs, how they are perceived by potential employers and clients, and the informal caste-based networks they can rely on for job referrals and credit (Munshi, 2019). For example, Figure 1 shows how the education of general category individuals is consistently higher than the education of OBC individuals, especially for those born in earlier birth cohorts. Similarly, Figure 6 portrays that general category individuals are more likely to be employed in professional or skilled occupations, which on average, offer a higher pay than other occupations. Caste also plays an important role in the political sphere, and indeed any discussion about caste, especially caste-based affirmative action policies, is a politically sensitive topic. Changes in caste-based affirmative action policies are often met by large protests and political parties often indulge in vote-bank politics,

[1]This is the view taken by most historians, but some historians like (Chandavarkar, 2003) and (Rudner, 1996) theorize that mobility is possible in a caste-based society, but this mobility means the mobility of the entire group, rather than of an individual.

by promising to target public resources to certain castes, even at the local level. Thus, the caste system makes India an important setting to understand the inter-generational persistence in socio-economic outcomes and also to isolate the impact of group-based affirmative action policies on inter-generational mobility.

Theoretically, the job quotas could influence the education and occupation choice of the OBCs through the following two mechanisms. First, the job quotas could have a direct positive impact on the occupation choices of the OBCs as a result of expanded opportunities in the public sector. However, these opportunities can be availed only by acquiring at least the minimum level of education required for these jobs. At the same time, private sector employers might start perceiving OBCs who fail to secure a government job as inferior candidates, and thereby reduce their probability of employment (Coate and Loury, 1993). Negative stereotypes about the ability of minority candidates would be exacerbated and could lead to more discrimination by employers. Second, the job quotas could alter the human-capital-formation decisions of OBCs due to a change in the perceived returns to education. A higher investment in education could, in turn, lead to more opportunities in the private sector. However, such policies can also have a negative impact by reducing the competition for public sector jobs, and thus, reducing the incentive to invest in human capital formation.

There are two strands of literature that are related to my work: (i) papers that look at the impact of affirmative action, for example, on the education and labour market outcomes (the probability of being employed, the probability of being salaried, etc.) of the targeted minorities and (ii) papers that focus on empirically estimating measures of IGM. To the best of my knowledge, my paper is the first one to attempt to connect these two strands of literature by trying to estimate the causal impact of an affirmative action policy on inter-generational mobility.

Extensive literature exists on affirmative action and its impact on education and labour market related outcomes. Khanna (2020) examines the effect of the same policy that I study, i.e. i.e. the implementation of the job quotas for the OBCs, and finds that OBC students are incentivized to study in school for another 0.8 years on average. The younger the OBC student is, the higher is the expected increase in their educational attainment as they would have more time to adjust their human capital formation decisions in response to changes in future prospects. Prakash (2020) uses exogenous variation in the the updating of state-level job quotas for the Scheduled Castes (SCs) and the Scheduled Tribes (STs), two other disadvantaged sections in India, and concludes that a one percentage point increase in the job quotas for the SCs raises their probability of being salaried by 0.6%. However, no such effect is observed for the STs. I study the same policy and use the same methodology as in Khanna (2020), but my main objective is to look at measures of IGM. One of the most common criticisms leveled against affirmative action policies in India is "elite capturing", and thus, it is important to examine if the benefits of the job quotas are concentrated among members who are born to relatively better-off parents.

On the other hand, the empirical literature of IGM aims to measure the degree to which a parent's socio-economic status determines their child's socio-economic status (Chetty et al., 2014). Because opportunities and socio-economic status are hard to measure, most of the literature on IGM is set in developed countries and considers income to be the most informative measure of socio-economic status (e.g., Solon (1999); Chetty et al. (2014)). However, IGM studies set in developing countries (or even historical periods set in developed countries [2] tend to avoid using income and instead use education (Alesina et al., 2021; Asher et al., 2018; Azam and Bhatt, 2015; Hnatkovska et al., 2013) or occupational choice (Azam, 2013; Hnatkovska et al., 2013; Iversen et al., 2016) as the relevant proxy of socio-economic status due to the following reasons.

First, income in developing countries is subject to significant measurement errors. Availability of long panel data is rare and studies typically rely on cross-sectional data that record an individual's income in the year in which the survey was conducted. This measure of income tends to be a poor estimate of permanent lifetime income as it is affected by transitory income shocks, which are especially prevalent in economies predominated by the agrarian and informal sectors. Analysis of income mobility requires that we compare the outcomes of children to their parents when they are at similar ages in their life-cycle because measuring children's income at early ages can positively bias inter-generational mobility in lifetime income as children with high lifetime incomes have steeper earnings profiles when they are young (Zimmerman, 1992; Grawe, 2006). In contrast, education and occupation choices rarely change in adulthood and thus, do not suffer from this life-cycle bias.[3] Education and occupation also tend to be precisely reported whereas income is often under-reported.

Second, many households in developing countries engage in joint production activities, like cultivation and self-employment in a family-owned business, which sometimes makes it impossible to divide the household income among the family members.[4] Education and occupational choice, in contrast, can directly be attributed to an individual. Third, information on education and occupation of parents might be available even if they are dead. However, education and occupation are not perfect measures of socio-economic status. The emphasis on studying educational mobility in developing countries is driven by the belief that education opens up better opportunities in the future, but the returns to education vary for different social groups. The empirical literature at best compares the trends of the estimates of IGM or explains how the estimates could be different for different social groups, without examining potential mechanisms that could explain such differences. My contribution to the literature

[2] Examples of studies looking at educational inter-generational mobility in historical times in developed countries include Derenoncourt (2019) and Card et al. (2018).

[3] I still try to control for the life-cycle bias by adding age and the square of age of the son and the father as covariates.

[4] One possible way of allocating household income among household members could be to use adult equivalent scales. The survey data set that I use, the IHDS-II, reports the number of hours worked by each household member in joint production activities, and hence, another method could be to divide the household income in accordance with their time contribution. However, the other two issues in using income as a measure of socio-economic status still remain.

is to investigate the role of one possible mechanism (i.e. affirmative action) in explaining the differences in IGM among different social groups.

As mentioned earlier, most of the literature on measurement of IGM is set in developed countries and uses income as the relevant outcome. Interested readers could look at Appendix C for a brief summary of this literature. The main two take-aways from that discussion are the following. First, IGM can be measured in various ways and the results could be different depending on the measure chosen as they each serve to answer related but distinct questions. Thus, it can be helpful to present results using more than one mobility measure. Second, the literature distinguishes between two classes of measures of IGM that capture different normative concepts: *relative mobility* and *absolute mobility* (Chetty et al., 2014). Relative mobility is the degree to which parental outcomes determine children's outcomes, whereas absolute mobility looks at childrens' outcomes for parents of a given outcome to determine if, on average, children are doing better than their parents. Since absolute mobility is arguably of greater interest than relative mobility, the measures of mobility I use in this paper are absolute mobility measures.

The measures of absolute educational mobility that I use are inspired by the empirical literature that measures income mobility. Education could, thus, be measured in terms of levels (i.e. years of education) or rank in the education distribution of a particular cohort. For example, Alesina et al. (2021) studies education IGM in African countries and uses (1) the probability that a child born to illiterate parents is literate as a measure of upward mobility and (2) the probability that a child born to literate parents is illiterate as a measure of downward mobility. On the other hand, Asher et al. (2018) develops a rank-based measure of upward educational mobility, called the bottom half mobility, which is the expected rank of a child born to parents in the bottom half of the distribution. Therefore, to study the effect of affirmative action implemented in the form of job quotas on educational IGM, I use the following measures of absolute upward education mobility: (1) probability that a child acquires higher years of education as compared to their parent, (2) change in years of education of children born to illiterate parents, (3) probability that a child born to an illiterate parent is literate, (4) change in education rank of children born to below-median-education parents, and (5) probability that a child born to a below-median-education parent ends up in the top half of their education distribution. Similarly, to look at absolute downward education mobility, I use the following measures: (1) change in years of education of children born to at-least-high-school-graduate parents (2) probability that a child born to an at-least-high-school-graduate parent does not complete high school, (3) change in education rank of children born to above-median-education parents, and (4) probability that a child born to an above-median-education parents ends up in the bottom half of their education distribution. I restrict my analysis to only son-father pairs due to reasons that are detailed in Section 3.2.1.

Occupations, unlike education, can not be easily ranked as there is some subjectivity in what a "better" occupation is.[5] Thus, I classify the occupations of individuals as (1) professional, (2) skilled, (3) unskilled, and (4) farmers. To study absolute upward occupation mobility, I use the probability that a son who is born to a father employed in an unskilled or farming occupation is employed in a professional or skilled occupation. Similarly, to study absolute downward occupation mobility, I use the probability that a son who is born to a father employed in a professional or skilled occupation is employed as a farmer or in an unskilled job.

I use data drawn from the second round of the Indian Human Development Survey (IHDS), which was conducted in 2011-12, and construct a data set of matched son-father pairs. I then restrict my sample to OBC and general-category individuals who have completed their education at the time of the survey. Next, I exploit the plausibly exogeneous nature of the implementation of the job quotas and employ a difference-in-differences strategy to estimate the causal impact of the job quotas for OBCs on inter-generational education and occupation mobility. I do so by comparing the outcomes of OBC sons belonging to a particular subsample based on father's education or occupation, to their counterparts in the general category, for individuals who are "old" at the time of the policy enactment and those who are "young". Only OBCs of a school or college-going age in 1993 could potentially be impacted by the job quotas as older OBCs would have already made human capital investment decisions that would be impossible or too costly to change in response to the policy. I find that the job quotas resulted in an increase in the absolute upward mobility of OBC sons, as measured by (1) an increased probability that their education is greater than their father's, (2) an increase in the years of education of sons born to illiterate fathers, (3) an increase in the probability that a son born to an illiterate father is literate, (4) a higher expected education rank of sons born to below-median-education fathers, and (5) a higher probability that a son born to a below- median-education father ends up in the top half of their education distribution. On the other hand, I find that the policy resulted in an absolute downward occupation mobility of OBCs, as indicated by a decrease in the probability that a son born to a father employed in an unskilled or farming occupation is employed in a professional or skilled occupation. Thus, OBC sons acquired more education in response to perceived future opportunities, but the downward occupation mobility could perhaps be understood by the exacerbation of negative attitudes of private employers regarding OBCs in response to the quotas.

The rest of the paper proceeds as follows. Section 2 discusses the institutional background of reservation policies in India, and in particular, the implementation of the Mandal Commission reforms. Section 3 provides a description of the data source, the construction of the data set, and the variables used in the econometric analysis. Section 4 details the empirical strategy employed to study the impact of the job quotas on inter-generational mobility. Section 5 reports the estimation results and interprets the

[5]One way to rank occupations could be to use the average income from those occupations. Another way, which was popularly used by American sociologists in the 1960s, is to rank occupations based on occupation prestige. See for example, Nakao and Treas (1994).

main findings. Section 6 concludes and discusses some ways in which the analysis could be extended in the future.

2 Institutional Background

I divide this section into four subsections. In the first subsection, I explain what the caste system is. The second subsection details the history of reservation policies in India and the third subsection discusses the particular reservation policy that I am interested in, i.e. the implementation of the Mandal Commission reforms. The fourth subsection presents some facts on the public sector in India.

2.1 Caste System

The majority of reservation policies in India are based on caste and take the form of quotas in educational institutions, jobs, and legislatures, which means that only the eligible groups for the reservation policies can apply for a specified percentage of seats. Thus, before proceeding to discuss the reservation policies in India, it is crucial to provide a little background on what the caste system is.

The caste system was a hereditary and hierarchical way of organization of society and division of labour in ancient India that was developed between 1500 to 500 BCE (Munshi, 2019). The caste system is characterized by the following main features. Every Hindu individual was (and for the most part still is) born into a *jati*, or caste, which in turn belongs to a *varna*, though there is sometimes some ambiguity about which *jati* belonged to which *varna*. The *varna* system stratified the Hindu society into four hierarchical classes (called *varnas*), with the *Brahmins*, who were priests and scholars, at the top of the social hierarchy. Next were the *Kshatriyas*, or the warriors and rulers. They were followed by the *Vaishyas*, or merchants, and at the bottom of these four caste categories were the *Shudras*, who were usually laborers, peasants, artisans, and servants. Outside of the *varna* system were a large sub-population of *Dalits* or untouchables who were excluded from much of the Indian village life. The other sub-population that was not a part of the *varna* system was the tribals. Thus, the *varna* system defined a broad social structure within which the thousands of castes and subcastes were placed. There are about 2,000-4,000 castes in India, and many more subcastes. It is important to realize that it is the *jatis* that define and form the nuts and bolts of the caste system, rather than the better-known *varna* system (Osborne, 2001). Each *jati* was linked to a particular occupation and tended to be confined to particular regions. Each *jati* was obligated to perform certain services tied to their hereditary occupations and could draw upon the services that other *jatis* were ordained to follow. The resulting *jajmani* system, which was the complex system of obligations among *jatis*, mimicked exchange in a conventional monetary economy (Osborne, 2001).

The caste system was very rigid, with strict restrictions on inter-caste dining and especially inter-caste marriage, to prevent mixing of castes. A notion of purity and impurity emerged, with the *Brahmins* considered the most "pure", and purity declining successively with *Kshatriyas*, *Vaishyas*, *Shudhras*, and the untouchables. A touch from a lower caste was considered to "pollute" the upper caste. The caste system, with all of its taboos and restrictions, was authenticated by religion, and the justification behind this hierarchy was tied to the Hindu belief in rebirth, wherein a person who was born as a *Brahmin* was considered to be able to do so by virtue of doing good deeds in their past life.

Of course, the system was unfair (and inefficient too) as one's occupation, and thus, fate was determined at birth, permitting no occupational or socio-economic mobility and resulting in persistent inequality across generations. This is the main feature that distinguishes caste from class, as the latter is not hereditary, at least in principle. Over time, the caste system lost its influence due to the influx of other cultures and religions, and it was believed to be fairly flexible by the 18^{th} century. Harder boundaries were again set by the British, who wanted a way of classifying a large and diverse Indian population, and they made caste India's defining social feature. They also practised discriminatory policies, such as allowing only Christians and upper-caste Hindus to be eligible for administrative positions in the British Raj.

Today, caste has became linked to religions apart from Hinduism as well. Many communities that converted from Hinduism to Islam, Christianity, or other religions, still maintained their caste and sub-caste affiliations.[6] Caste identities remain strong, and untouchability and prohibition on inter-caste dining is still practised in many remote areas. Field-based ethnographic studies in villages present mixed evidence regarding the hereditary relationship between occupations and caste (Mendelsohn, 1993; Mayer, 1996; Jodhka, 2004), thus, highlighting the need to use large scale survey data to explore if the relationship between caste and occupational segregation has weakened, and if so, how much can be attributed to affirmative action policies. Inter-caste marriages remain low, at around 5.8%[7] according to the 2011 Census and this rate has remained virtually unchanged over the past four decades.

2.2 History of Reservation Policies in India

Caste-based discrimination, especially the practise of untouchability, was banned after independence in 1947 in hope that the caste system would cease to be a prominent part of Indian society. However, the post-independence government decided to continue the British policy of reserving seats, a policy, ironically, that had been opposed pre-independence (Osborne, 2001). The difference, though, was that now, affirmative action policies were introduced to correct inequalities arising from the caste

[6] Thus, non-Hindus could also be eligible for caste-based reservation policies.

[7] Ray et al. (2020) study the patterns of inter-caste marriage and find that the rate is not statistically different for rural and urban areas, and education of the spouses themselves plays no role, but education of the mother of the groom does.

system and explicitly target the historically disadvantaged castes, whereas earlier, the policies favoured upper-caste representation.

Thus, for the purpose of caste-based reservation laws, there are 4 distinct caste categories in society today: (i) Forward Castes or General Category, (ii) Scheduled Castes (SCs), (iii) Scheduled Tribes (STs), and (iv) Other Backward Castes (OBCs). This categorization is different from the original *varna* system of organizing the castes. However, there is some general overlap, for example, the castes comprising the SCs are mostly the castes classified as untouchables in the original *varna* system categorization. Similarly, the *jatis* comprising the STs are mostly the aboriginal tribal communities that existed outside the *varna* system. *Brahmins* are generally a part of the forward castes, while the *jatis* that belongs to the *Kshatriyas*, *Vaishyas*, or *Shudras* are generally a part of either the forward castes, or the OBCs, depending on the region.

The forward castes are considered the most advantaged sections of the society, and the SCs and the STs are considered the most disadvantaged. From a policymaker's perspective, there are at least three factors to consider while forming reservation policies: (i) the category of individuals to target i.e. which castes belong to these four caste categories, (ii) the institutions in which to introduce reservations (schools/colleges/jobs/legislatures and also the the jurisdiction level– central/state/local/private), and (iii) the percentage of seats to reserve. The government comes up with a list of castes and sub-castes to classify as SCs, STs, and OBCs, and by default, all other castes and sub-castes that do not belong to either of these three lists are a part of the Forward Castes[8].

The first round of reservation policies were implemented in 1950 and were targeted at the most disadvantaged social groups, the SCs, and the STs. The SCs and the STs occupy a 15% and 7.5% reservation of seats, respectively, in central government colleges and jobs. In 1993 and 2006 respectively, the OBCs also started occupying a 27% reservation in government jobs and colleges, which I will discuss in detail in the next subsection. The OBCs, however, do not enjoy reservations in central, state, and local legislatures, as the SCs and the STs do.

To put these quotas into perspective, it would be helpful to know the share of population of each of these four caste categories. The counting of the population of different castes in India was started by the British in 1872,[9] but this practise was discontinued after the Census of 1931. Caste data in the 1941 Census was collected, but not published. The Post-independence decennial Censuses of India, first conducted in 1951, record if an individual belongs to the SC or the ST caste group, but do not record

[8]These lists are different for different states, thus it is possible, for example, that the same caste could be considered OBC in one state and SC in another state.

[9]In 1872, 3,208 castes were identified. The castes in colonial India weren't divided into the four caste groups (SC/ST/OBC/-General) that they are divided into now, hence there isn't any information available in the colonial censuses on the share of population belonging to these different caste groups. It might be possible to aggregate the 1872 population of all the castes that belong to a particular caste category in 2021 (the list of OBC/SC/ST castes is publicly available) to arrive at these shares. However, I have not carried out this exercise and I couldn't find any other paper that has.

if an individual belongs to the OBC group. According to the 2011 Census, the share of SCs and STs in India's population in 2011 was 16.6% and 8.6%, respectively. This means that the General and the OBC caste group together constituted 74.8% of India's population in 2011. The Mandal Commission that was established to identify OBC castes estimated their popluation share to be 52% in the early 1980s, but there aren't any official figures available since then. While the quotas for the SCs and the STs are somewhat in tune with their population shares, 27% was recommended as the quota for the OBCs by the Mandal Commission, despite their population share being estimated by them to be 52%, owing to a Supreme-court mandated constraint that the total quantum of reservation could not exceed 50%. Nationally representative surveys, like the National Sample Survey (NSS), estimate their number to be around 40%, but the reluctance to officially count the OBCs could stem from the government's fear that the official count could be above 52%, thereby spurring demands for more quotas. The first ever Socio-Economic Caste Census (SECC) was conducted in 2011, which reported the caste of every respondent, but the results of the caste count have not yet been released.[10]

The preceding discussion on the quota shares for different caste groups is at the central level. Besides central government jobs and colleges, one could also apply to colleges owned by state-level governments or get a job in state-level institutions. Different state governments have their own policies regarding reservation in state government jobs and colleges. Each state chooses a different reservation quotas for the SCs, STs, and the OBCs, and these quotas could be different from the central quotas of 17%, 7.5%, and 27%, respectively. [11]. Khanna (2020) provides information on the state-level variation in the quotas applicable to different caste categories in the year 1995, though there is a slight possibility that these numbers may have changed over time.

One concern could be that the list of SCs, STs, and OBCs is in itself determined by factors like political lobbying wherein castes with higher resources and political clout can pressure the government to make them eligible for quotas. Recent anecdotal evidence suggests that it may not be that likely. One caste community, the Jats, led a series of violent protests in February 2016 to seek inclusion in the OBC list of the central government and the state governments of Punjab and Haryana. The state governments ceded and categorized them as OBCs, but the decision was overruled by the court within a couple of days.

Besides caste-based quotas, affirmative action policies are also targeted towards women and low-income general category caste members. In 1992, a constitutional amendment mandated one third of local government head positions to be randomly reserved for women. In 2019, 10% of seats in govern-

[10]This is because of the following two reasons. First, the government has faced logistical difficulties in trying to collate caste information. Around 330 million households were surveyed, and 4.3 million entries were recorded, which clearly exceeds the actual number of castes in India. There are differences in the interpretation of the question "What is your caste?" by different individuals as caste is a complex social structure. Second, the ruling party could be deliberately delaying releasing the results so that they don't have to deal with the resulting political fallout.

[11]Thus, in my analysis in Section 4.1, I control for State-Caste fixed effects.

ment colleges and jobs were reserved for low-income general category caste members. There have also been demands to have quotas in place separately for Muslims. The issue of sub-categorization of OBC quotas has also gained policy attraction, with more than 13 Indian states instituting a quota-within-quota system.

2.3 Mandal Commission Reforms

The identification and implementation of reservation quotas for the OBCs has been a long drawn-out process. Adhering to Article 340 in the Constitution, which obligates the government to promote the interests and welfare of "backward classes", the First Backward Classes Commission was set up on January 29, 1953 under the chairmanship of Kaka Kalelkar. Its objective was to identify sections of society that were socially and educationally backward, in addition to the SCs and the STs, and make recommendations to improve their condition. The commission submitted their report in 1955 and identified 2,399 backward castes, out of which 837 were deemed to be the "most backward". However, the report was not accepted by the government as caste was used as the sole criteria to identify social and educational backwardness, and the government feared severe backlash. The second Backward Classes Commission was established under the leadership of B.P. Mandal (and hence it is popularly called the Mandal Commission) on January 1, 1979, with the same objective. The Mandal commission came out with their report in 1980, wherein they created a list of castes that would be OBCs based on an index of backwardness that captured their socio-economic status at that time and also recommended the percentage of seats to be reserved for them in educational institutions and jobs. The index was based on four social indicators, each carrying a score of three points; three educational indicators, each carrying a score of two points; and four economic indicators, each carrying a score of one point. Thus, the total score could be at maximum 22. These 11 indicators were applied to each caste surveyed by the commission, and castes that scored less than 11 points were deemed backward. 3,473 backward caste (or subcaste) groups were identified.

The contents of the Mandal Commission report were of a politically contentious nature, and the central government did not act upon it for nearly 10 years. They first announced their intent to implement it in 1989, but were immediately faced by (often violent) protests by upper-caste students. In 1991, they again announced that the recommendations will be implemented, but the constitutional validity of the new law was challenged in court. On 16 November 1992, the Supreme Court ruled in favour of the central government and the first stage of the reforms was implemented in 1993. Thus, even though the recommendations were around for a long time, it can be argued that the implementation of these reservation policies was exogeneous from the perspective of the households, though there is some degree of endogeniety in the sense that it was implemented only after a change in the political party in power. Reservation policies take an important position in electoral promises; so, it would be the same households who vote for the parties, keeping in mind the expected policy changes in the future.

2.4 Public Sector in India and Implementation of Job Quotas

The public sector in India still remains a major employer, though the size of the public sector is shrinking over time. The formal/organized sector in India employed around 20% of the workforce in 2018, out of which around 58% comes from the public sector. However, the central government constitutes only 14% of the public sector employment, and the rest comes from state governments. Employment in the public sector is still considered very attractive, as it is seen as a symbol of status and power, provides a competitive salary and perks, and offers job security. This is true especially at the lower end of the skill distribution. For example, the unconditional annual wage premium for government jobs relative to similar formal sector jobs was $465 ($485 for OBCs) in 2000 (Prakash, 2020). Of course, the number of job openings as a percentage of the labour force is low. Competition for these jobs is tough, with one instance in 2018 where 23 million people applied for about 100,000 low-skilled posts advertised by the Railway Recruitment Board, such as porters.

Jobs in the central government of India are classified in four categories (group A, B, C, D)[12] corresponding to the salary, status, qualifications, and the nature of responsibilities attached to the jobs. Group A posts are the most well-paid, carry higher administrative and executive responsibilities and include senior management positions in the ministries/departments and field organisations. Group B jobs constitute middle management. Group A and B posts together made up 11.5% of the total central government jobs and require a higher secondary or a college degree. Group C employees perform some supervisory and operative tasks. These jobs comprise 58% of the jobs and require completing secondary or middle school. Group D employees receive the lowest salary and carry out low-skilled routine tasks such as clerical work. These jobs encompass 30.5% of the jobs and require completing primary school or being literate. Thus, the incentive effect of the job quotas could matter for acquiring each of these minimum education qualifications.

As mentioned before, hiring in the public sector as well as admissions to government universities are subject to caste-based quotas. In practise, the hiring takes place as follows. First, the government announces vacancies for a particular post, the minimum educational qualification and the maximum applicant age required for that post. An applicant can indicate that they intend to apply for a reserved seat, by showing proof of their SC/ST/OBC status.[13] These caste certificates are issued by the sub-district

[12] The quotas are applied to each of the four categories separately.

[13] SC/ST/OBC individuals could also apply as a general category individual by choosing not to show their caste certificate. This could be due to the fear that their peers or superiors in their job could discriminate against them if they reveal they are from a lower caste.

administration. Recruitment is determined based on written competitive exams[14] and the government comes out with a score cutoff for each caste separately. The score cutoffs are usually such that the top 50% of the highest scoring candidates in the general category can be hired, the top 27% OBC candidates can be hired, and so on, or until there is no available candidate that meets a "minimum quality" threshold. The cutoff for the general category is mostly higher than the cutoffs for other caste categories. "Minimum quality" is determined according to the nature of the vacancy, but is ambiguous. Reserved seats that are unfilled could be filled by lowering the cutoff for general category individuals. Thus, in practise, it is possible that more than 50% of the seats go to the general category, or some seats are unfilled even after determining multiple cut-offs. Quotas are also implemented in promotions.

Data on actual shares of different caste groups in government jobs are scant, but some information is available through the replies of government officials to such concerns raised in the parliament or in their response to Right to Information (RTI) requests made by the public. The representation of SCs, STs, and OBCs in central government jobs stood at 17.49%, 8.47% and 21.57%, respectively, as on January 1, 2016. The OBC representation has increased over time; on Jan 1, 2012, the share of OBC employees in central government jobs was 16.55%.[15] Also, to give an idea of the OBC representation before the quota implementation, the Mandal Commission report estimated that their share was 12.55% in 1980 (4.69% in Class I jobs, 10.63% in Class II jobs, and 24.40% in Class III jobs).[16] The actual representation of the OBCs in government jobs falls short of the quota available to them, but this shouldn't be suggestive of the policy being ineffective. The effect of the intention of the government to fill seats such that the OBC share is 27% should be seen more from an Intent-to-Treat framework. The very availability of the quotas could change how much education young children acquire in response to perceived future returns.

3 Data

I partition this section into five subsections. The first subsection discusses the source of the underlying data and the advantages and disadvantages of using it. The second subsection describes the construction of the data set and the sample restriction criteria imposed. The choice of the control caste groups and the treatment cohorts is explained in the third subsection. The fourth subsection describes the variables used in the econometric analysis. The fifth subsection presents summary statistics.

[14] For some positions, there could be more than one round of written exams. There could also be an interview as the final stage of the hiring process. In such cases, the final score of a candidate is determined as a weighted average of their scores in all of the hiring stages. The weights are generally such that more importance is given to written exams.

[15] These aggregate numbers mask heterogeneity in the actual type of jobs secured by different caste groups. For example, lower caste groups are typically better represented in low-skilled jobs than in high-skilled jobs.

[16] The central government earlier used Class I/II/III to classify jobs. Class I jobs were the most sought-after and prestigious jobs, like civil servants, professors, etc. while Class III jobs were unkilled jobs like clerks.

3.1 Data Source

To study the impact of the Mandal Commission reforms on inter-generational education and occupation mobility, I must have, at the very least, data on caste group membership, age, years of education, and occupation for children as well as their parents. The data used in this study come from the second round of the Indian Human Development Survey (IHDS-II), which was conducted by the University of Maryland and the National Council of Applied Economic Research in 2011-12. IHDS is a multi-topic, nationally representative survey [17] and encompasses questions on a wide range of socio-economic and demographic characteristics such as health, education, employment, economic status, marriage, fertility, gender relations, and social capital. Children aged 8-11 completed standard reading, writing and arithmetic tests, whose questions were the same for all the surveyed children. Additional village level data is also available on the status of schools, medical facilities, and other public infrastructure. IHDS-I surveyed 41,554 households in 1,503 villages and 971 urban neighborhoods across India whereas IHDS-II surveyed 42,152 households (in the same villages and urban neighborhoods), with 85 per cent of the households from IHDS-I being resurveyed.

There are at least six advantages of using data drawn from IHDS-II. First, the IHDS collects information on the education and occupation of the father of the majority of male respondents, even if those fathers have died or are not considered part of the same household.[18] Information on deceased fathers will help in understanding mobility of older cohorts. Second, it contains information on the education and occupation of non-resident children and non-resident parents. Most of the literature studying inter-generational mobility in India use data sets like the National Sample Survey (NSS) (for example, in Hnatkovska et al. (2013)) and National Family Health Survey (NFHS), which have information only on the residents of a household. Neglecting the non-residents might lead to negatively biased estimates, as individuals who experience upward mobility are also less likely to co-reside with their parents.[19] Third, the IHDS is a panel data set which allows me to recover the occupational status of some individuals who were retired in 2011-2012 (i.e. in IHDS-II, the data set that I use), but not in 2004-05 (IHDS-I). Fourth, the IHDS identifies the mother's and father's ID of an individual, if their parents are part of the same household. Data sets like NFHS and NSS do not provide an ID for the parents and children can only be indirectly linked to their parents using information from the relationship-to-the-household-head field. This, of course, is possible only in non-ambiguous cases, which might result in some information loss

[17] IHDS covered all the states and union territories of India except Andaman and Nicobar and Lakshadweep islands. These two union territories form less than 0.05% of India's total population according to the 2011 Census of India.

[18] The household file in IHDS specifically asks the occupation and education of the father of a male household head. If the household head is female, they instead ask the occupation and education of her father-in-law.

[19] Asher et al. (2018) find that estimating IGM using only co-resident son-father pairs leads to a bias that increases in the age of the son. Azam and Bhatt (2015) demonstrate that the estimated regression coefficient in a regression of son's educational attainment on father's educational attainment restricting the data based on the co-residence criterion is 17% lower than the estimate based on the full sample. Other papers like Hnatkovska et al. (2013) also acknowledge that using only co-resident son-father pairs leads to a selection bias but do not show the possible extent of this bias.

as many people in India still live in joint-families.[20] Fifth, it contains data on the actual number of years of education rather than levels of schooling, which helps me in avoiding bunching in schooling distribution as a result of the imputation of years of education from the level (Azam and Bhatt, 2015). Sixth, the IHDS provides very detailed information on a variety of household and village level characteristics, which allows for a very rich set of potential covariates. In particular, the IHDS also asks the household their *jati*, besides which of the four caste categories (General/SC/ST/OBC) they belong to at the time of the survey. This feature can be exploited in future work, which I discuss in Section 6.

Before proceeding, it is also important to acknowledge the limitations of IHDS-II for the purpose of this study. IHDS-II has a limited sample size as compared to data sets like the NSS and the NFHS, which are repeated cross-section surveys that have the advantages of a larger sample size and a higher number of rounds. The relatively small sample size of IHDS-II proves to be a limitation in some of my econometric analysis where in I subset the data by some variable indicating the father's education or occupation.

3.2 Construction of the Data Set

3.2.1 Linking Sons to Fathers

I limit my analysis to only son-father pairs due to the following reasons. First, the organization of society in India is patrilineal i.e. the bride moves into the groom's family post marriage. The parents of a woman who enters a household after marriage are, thus, not considered members of the household, and therefore it would be impossible to track the parental information for married women. Married daughters are also not considered a part of the household, and their information is not included in the non-resident file of IHDS. Second, the IHDS only records the education and occupational choice of fathers of male household heads, and not their mothers. Male household heads with a dead mother form a significant portion of my sample, and thus I can't ignore them while possibly trying to link sons with their mothers. Third, I restrict my sample to men because I want to study the impact of affirmative action on occupational mobility, but the female labor force participation in India still remains very low, at around 20% in 2019.

I construct the matched son-father pairs by using the household, individual (resident), and non-resident files of the IHDS as follows. Every resident of a household is assigned an ID and is asked their age, gender, occupation, earnings, relationship to household head, and the ID of their mother and father (if the parents are alive, irrespective of whether the parents are residents or non-residents). The household file also reports the occupation and education of the father of a male household head or

[20] For example, Hnatkovska et al. (2013) are able to identify father's education for less than 15% of adult males surveyed in the NSS. In contrast, I am able to identify father's education for more than 95% of adult males in IHDS-II by using the father's ID and information on the education of the household head variable.

the husband's father of a female household head, even if the father under consideration is dead or no longer a part of that household. The non-resident file contains information on the age, sex, occupation, and education of the non-residents, the ID of the resident to whom they are related, and the type of the relationship (e.g., son/daughter, father/mother, brother/sister, etc.). This enables me to link non-resident fathers and resident sons as well as resident fathers and non-resident sons.

3.2.2 Sample Restrictions

I now outline the major sample restriction criteria that I impose on the son-father data set. First, I restrict the sample to individuals who are more than 21 years old in 2006 for reasons that will become clear in Section 3.3. Second, I exclude individuals who are were enrolled in any educational institution at the time the survey was conducted (i.e. in 2012). This is done to avoid the possible right-censoring of education data by the inclusion of sons who have not fully completed their education. Including such individuals can bias the estimates of inter-generational educational mobility downwards.

Third, I restrict the sample to only General category and OBC members for reasons that will become clear in Section 3.3. Fourth, I drop individuals whose (1) education is missing, or (2) occupation is not defined, or whose (1) father's education is missing, or (2) occupation is not defined. Fifth, there seem to be a few cases in which the father is recorded as less than 10 years older than his son. I drop such cases as this scenario seems to be pretty unlikely to have arisen.

After imposing the above sample restriction criteria, I am left with 30,539 son-father pairs for studying occupation mobility and 36,127 son-father pairs for studying education mobility.

3.3 Control Caste Group and Treatment Cohorts

As explained later in Section 4.1, I use a difference-in-differences framework to estimate the causal impact of the implementation of the job quotas on the inter-generational education and occupation mobility. To identify the causal effects, I leverage variation in caste group affiliation and age. As discussed in Section 2.3, only the OBCs were eligible for the job quotas implemented in 1993. Thus, the OBCs are the obvious choice for the treatment group. This means that there are three potential candidates for the choice of the control group: (i) the general category/forward castes/upper-castes, (ii) the SCs, and (iii) the STs. Following the previous literature (Khanna, 2020), I keep only the general category as the control group. Khanna (2020) demonstrates that the impact of the job quotas for OBCs on their educational attainment was higher for each successive birth cohort affected by the policy change, and hence, it is not unreasonable to rule out dynamic treatment effects of the quotas for the SCs and the STs, especially since their quotas started in the 1950s and have covered more than one generation.

Another concern could be that caste-based reservation in legislatures could also have effects on policy outcomes through changes in the allocation of public goods. Pande (2003) finds that increasing caste-based minority representation in state governments increases transfers to minorities and results in an increased spending on welfare programs and share of seats reserved in state government jobs. Bardhan et al. (2010) find that having a SC/ST village head leads to redistribution of employment program benefits away from non-SC/ST landless households toward SC/ST households. In 1993, the year in which job quotas for OBCs were implemented, gender-based quotas were also introduced in local government bodies, mandating that one third of seats be reserved for women. This new policy, interacted with the pre-existing legislative quotas for the SCs and the STs, could thus change their socio-economic outcomes.

The source of the second difference in the difference-in-differences analysis comes from variation in age. The newly-implemented job quotas for OBCs could only be availed by individuals who were not too old at the time of the policy change. In Section 1, I explained that there are two possible mechanisms by which job quotas could affect IGM : (i) change in human-capital-formation decisions due to a change in the perceived returns to education and (ii) change in the probability of employment in certain occupations, even if there is no additional investment in education, as a direct result of expanded opportunities in the public sector and an indirect result from the change in how private employers perceive them. Government jobs have a maximum age cutoff and require some minimum educational qualifications (discussed in more detail in Section 2.4). It is also uncommon for older individuals to (i) acquire more education after they enter the labour market[21], even though educational institutions do not have a maximum age cutoff and (ii) drastically change their occupational choice, especially when occupation is considered in broader categories. Thus, if OBC individuals are young enough in the years in which the job quotas exist (i.e. of age 21 or younger–the age at which one typically graduates from college), they could have had time to change their education and occupational choice decisions in response to the implementation of the quotas. Individuals above the age of 21 will find it difficult to benefit from the job quotas as they would have already taken many human-capital-formation decisions that might now be too costly to alter. For the same reason, I drop individuals who were younger than 22 in 2006, as 2006 was the year in which quotas for the OBCs in central-government universities were introduced. Thus, I am able to capture the sole effect of the job quotas.

3.4 Variables

Since I am interested in studying education and occupation IGM, I first describe the education and occupation related variables. I then proceed to describe other variables that serve as the main independent variables or as covariates.

[21] E.g., according to data in in IHDS-II, in 2012, less than 0.5% of the individuals above the age of 25 were enrolled in any kind of educational institute.

3.4.1 Years of Education

IHDS-II reports the number of completed years of education of residents in the individual file, non-residents in the non-resident file, and the father of a male household head (or husband's father of female household heads) in the household file. I fill in the father's education first by using information from the resident and the non-resident files, before proceeding to fill in missing information from the household file for male household heads and spouses of female household heads.[22] The years of education range from 0 to 16, with years 1-12 indicating the 12 years of primary and secondary schooling, 13-15 representing years of college education,[23] and 16 signifying more than 15 years of education.[24] Figure 1 plots the average value of the years of education acquired by different birth cohorts for the OBCs and the general category separately. Even though I do not use the full sample to analyse education IGM, it is reassuring to see (roughly) parallel pre-treatment trends for this variable.

I also construct a binary variable that takes the value one if a son acquires higher years of education as compared to their father, and zero otherwise. Figure 2 depicts the percentage of individuals acquiring education greater than their father's education for different birth cohorts and caste groups. The trend for OBC and general category sons seems broadly parallel before 1971.

3.4.2 Education Categories

I focus on the subsamples at the extreme end of the education distribution and construct two binary variables that represent the highest level of education attained by an individual: (1) illiterate (0 years of schooling), and (2) at least high school (12 or more years of education) I use this categorization to set the data based on father's education category.

To study drop out rates of sons, I create two binary variables from the years of education variable: (1) illiterate, and (2) completed at least high school.

3.4.3 Education Ranks

Following Asher et al. (2018), from the years of education variable, I create the rank of an individual in their education distribution. I do this before I drop the SCs and STs from my sample, as the rank should indicate the position of an individual in the national education distribution, and thus would be more suitable for comparing different social groups. First, I pool data into 10-year cohorts and calculate

[22] There are some misreporting issues in the education of the father. In practise, the education of the household head's father can be recorded from three sources: (1) the resident file (if the father is co-resident); (2) from the household head's response to the question on his father's education; and (3) from his wife's response to the husband's father's education question in the women's survey file. Asher et al. (2018) show that the average correlation between father's education measured across these three sources in IHDS-II is 0.9, but the misreporting errors are uncorrelated with household characteristics.

[23] Most undergraduate college education in India takes the form of three-year programs, but some fields like engineering have a four-year undergraduate program.

[24] Attained while pursuing a master's degree, for example.

the cohort-specific rank of a son in the education distribution of all sons born in that cohort. Since there are 16 possible values of the years of education variable, the rank of individuals in one birth cohort would also have at most 16 values, with the same rank assigned to all those who have the same educational attainment.[25] However, the values of rank would be more than 16 for the "young" and "old" individuals, as both these subsamples would contain more than one 10-year cohort. Second, I calculate the cohort-specific rank of a father in the education distribution of all fathers with sons born in that cohort. Figure 6 depicts the average value of the education rank acquired by sons born in different years for the OBCs and the general category separately. Again, the pre-treatment trends seem broadly parallel in the entire sample. General category individuals do relatively better than OBCs, but as can be seen from Figure 6, the gap between the two castes seems to narrow after 1980.

I also create binary variables that takes the value zero if the father's rank is below (above) the median (i.e. father's rank is < 50 (> 50)) and the value one if it is above (below) the median. Since education is coarsely measured and there are lots of older individuals reporting a bottom-coded level of education (for e.g., over 50% of fathers in the 1960-69 birth cohort are illiterate), I am unable to create finer categories like quartiles.[26] I do the same for the rank of the son variable.

3.4.4 Occupation Categories

I ascertain the main work activity (i.e. the work activity in which an individual spends most of his labor hours) and the associated occupation codes of that activity for fathers and sons using the method delineated in Appendix A. Occupations in IHDS-II are classified at the two-digit level following the 1968 National Classification of Occupations (NCO) codes. NCO codes are aligned with the International Standard Classification of Occupations (ISCO) with appropriate adjustments suitable for the Indian economy. Table A1 lists the two-digit NCO-1968 codes.

Of course, the two-digit occupation codes in the IHDS-II will only be available for those who are employed in the labor market. Some individuals might be unemployed or not participate in the labor force, for example, due to retirement, health reasons, engagement in household work, etc. As the IHDS is a panel data set, I am able to recover the occupation codes for some individuals who were employed in 2004-05 (i.e. when the first round was conducted) but not in 2011-12 (i.e. when the second round was conducted). Ultimately, for the analysis of occupational mobility, I drop observations in which the two-digit occupation code is still not available for the son and/or the father.

Next, I aggregate the two-digit occupation codes into four categories (professional, skilled, unskilled, and farmer) by combining similar occupations. I use the classification schemes following Azam (2013) and present the two-digit occupation codes of these categories in Table A2. Figure 6 shows the percent-

[25] Ranks are assigned using the midpoint method.

[26] If I create quartiles of father's rank, I observe that there are no old individuals with father's rank in the first quartile.

age of sons employed in a professional or skilled occupation by caste and birth year. These trends do not seem to be very parallel for the overall sample.

3.4.5 Independent Variables

The two most important independent variables in my analysis are caste group affiliation and birth cohort. The construction of the caste group affiliation follows from the discussion in Section 3.3. *OBC* is a binary variable that takes the value one if an individual is OBC, and zero if he belongs to the general category. *Young* is a binary variable that takes the value one if an individual was 21 or younger in 1993 (i.e. the time when the job quotas for OBCs were introduced), and zero otherwise. As will be seen in Section 4.1, the coefficient on the interaction term of these two variables will be the coefficient of interest. I also include the following socioeconomic variables as co-variates: age of son, square of age of son, age of father[27], square of age of father, relationship of son to household head, marital status of son, indicator for whether the household is below poverty line, religion, and place of residence (urban/rural).

3.5 Summary Statistics

Column (1) of Table 1 details the mean and standard deviation of various education, occupation, father's education, father's occupation, and other socio-economic characteristics of sons belonging to my sample consisting of OBC and general category individuals born before 1985. Columns (2), (3), (5), and (6) carry out the same exercise, but for old general-category sons (i.e. those born before 1971), Old OBC sons, young general-category sons, and young OBC sons, respectively. Column (4) tests whether the population mean of characteristics of old general-category sons is equal to those of old OBC sons. This is done to check if the levels of the outcome variables or other characteristics of the treatment and control group are similar before treatment. As can be seen from Column (4), almost all the differences are significantly different from zero. Even though I did not do a formal one-sided test, the average education and occupation outcomes of general category members are better than the OBCs. I discuss what this means for my identification strategy to ascertain the impact of job quotas on mobility in Section 4.3.

4 Methodology

I partition this section into three subsections. First, I discuss the difference-in-differences framework that I use to ascertain the impact of job quotas on inter-generational mobility. In the second subsection,

[27] Age of father is missing for fathers whose information was obtained from education/occupation of father of household head variable. I impute the age of such fathers as the mean age of fathers with sons as the same age as the son of that father.

I discuss multiple hypothesis testing. The third subsection discusses some concerns in the empirical strategy.

4.1 Impact of Job Quotas: Difference-in-Differences Framework

I now turn to the key question of this paper: how do the education and occupation choices of children relative to their parents change as a result of the job quotas? I exploit the quasi-experimental nature of the implementation of the Mandal Commission reforms and use a difference-in-differences approach to compare the variation in the outcome variables of individuals affected by the program, relative to those who are not. The following two sources of variation identify an individual's exposure to the Mandal Commission reforms: his (1) caste group affiliation and (2) birth cohort. Recall that in section 3.3, I had specified that the treatment group is the OBCs and the control group is the General category. The treatment cohort is those who were 21 or younger at the time the job quotas were implemented. Thus, only young OBCs could be affected by the policy change, while the rest of the individuals should be unaffected.

I employ various specifications in a difference-in-differences framework to see how absolute mobility changes as a result of the job quotas. I set the data based on father's education or occupation and then estimate the following regression:[28]

$$y_{iga} = \beta_0 + \beta_1 obc_g + \beta_2 young_a + \beta_3 obc_g \times young_a + X'_{iga}\boldsymbol{\gamma} + \varepsilon_{iga} \tag{1}$$

where y_{icgt} is the outcome of interest of son i in caste group g and of age a; obc_g is a dummy that takes the value one if caste group is OBC and zero otherwise; $young_a$ is a dummy that takes the value one if age is less than 22 in 1993 and zero otherwise; X'_{iga} are control variables added incrementally, and ϵ_{iga} is the error term.

The control variables could include (1) state-cohort fixed effects to take into account time-varying state-level policies that could affect education or occupation choice decisions, (2) state-caste fixed effects to take into account the state-level quotas for OBCs for employment in state-government jobs, and (3) co-variates (age of son, square of age of son, age of father, square of age of father, relationship to household head, marital status, amount of land owned, religion, and place of residence (urban/rural). My preferred specification is including all three categories of control variables listed above.[29]

Standard errors are clustered at the state caste-group level. Thus, there are 62 clusters, two for each

[28] I set my data in all the regressions except the one in which I consider the binary variable that a son has higher education than his father as the dependent variable.

[29] As a robustness check, I try alternative specifications to see how the magnitude and the significance of my coefficients changes if I do not include some controls. I find that my results are not very sensitive to the controls I include. These results are available from the author on request.

of the 31 states (or union territories). An alternative way of clustering could be at the state level, as in Khanna (2020). The basic idea behind clustering standard errors is the assumption that errors are correlated within a cluster but are independent across clusters. Failing to account for this correlation could lead to misleadingly small standard errors, and consequently low p-values. This discussion is extremely relevant in difference-in-differences studies (Bertrand et al., 2004). Had treatment been assigned at the individual level, there would be no need for clustering, but clustering is needed when cluster of units are assigned to the treatment rather than individual units (Abadie et al., 2017). Cameron and Miller (2015) suggest that there is no formal test to determine the suitable level of clustering, and the consensus is to be conservative and use bigger clusters till the point that there is concern about having too few clusters. Depending on the context, "few" might mean less than 20 to less than 50 clusters. Hence, I cluster my standard errors at the state caste-group level rather than the caste group level or the state level. Thus, by clustering at the state caste-group level, I am assuming that there exists a correlation between belonging to a certain caste in a certain state and the values of outcome variables. This seems like a reasonable assumption to me because every state has their own caste-specific policies that could influence education or occupation choice of individuals.

The coefficient of interest is β_3, and is identified under the assumption that the average outcomes of OBC sons (in a sample restricted by some indicator of the father's education or occupation) and their counterparts in the general category would follow parallel paths over birth cohorts in the absence of the job quotas (i.e. other factors that can cause inter-generational mobility to change over time (e.g. trade liberalization) would affect different castes belonging to the same age cohort in the same manner).

I now discuss the exact y_{iga} variable used and the method of sub-sectioning the data for studying education and occupation mobility. As will be evident shortly, restricting the sample based on some indicator of father's education or occupation allows us to interpret β_3 as changes in absolute mobility.[30]

4.1.1 Effect on Education IGM when Education is Measured in Levels

I use three different methods to study the effect of job quotas on absolute education IGM. First, I look at the effect of the policy on the probability that a son acquires higher years of education as compared to their father. I do this analysis using my entire sample. However, the β_3 coefficient in this case gives us the impact of the policy on average and could mask important heterogeneity. For example, if we observe β_3 to be positive (which would be an indication of upward absolute education mobility) , could it be possible that sons of highly-educated fathers are acquiring more education than their (highly educated fathers) and sons of lower-educated fathers are not? To study this, I use subsamples constructed using the highest education level attained by fathers.

[30] An alternative way could be to interact obc_g, $young_a$, and an indicator for father's education or occupation category. However, I follow the prior literature, such as Alesina et al. (2021) and Asher et al. (2018), that seems to prefer the subsectioning method.

Second, I use the two subsamples based on father's education categories defined in Section 3.4.2 (the first one consisting of illiterate fathers and the second one consisting of fathers who have completed at least high school) and for each of the 2 subsamples, I estimate Equation (1) separately with y_{icgt} being the son's years of education variable as defined in Section 3.4.1.

Third, I analyse inter-generational education transitions. I focus on the two most important transitions. First, I look at the effect of the policy on the probability that the son is literate conditional on his father being illiterate. A positive β_3 in this case would signal absolute upward education mobility. Second, I look at the effect of the policy on the probability that the son does not complete high school conditional on his father being at least a high school graduate. A positive β_3 in this case would signal absolute downward education mobility.

4.1.2 Effect on Education IGM when Education is Measured in Ranks

Next, I look at education in terms of ranks instead of levels or years of education. First, I study how the average gain in a son's standing in their education distribution changes as a result of the policy when their fathers are restricted to (1) having below-median education (refer to Section 3.4.3 for how this indicator is constructed) and (2) having above-median education. Thus, in these two equations, I use son's education rank as the y_{iga} variable. I also analyse inter-generational education rank transitions i.e. I use the the dummy variable for the son's education rank being above the median as the y_{iga} variable for the former sub-sample and the dummy variable for the son's education rank being below the median as the y_{iga} variable in the latter sub-sample.

In conclusion, to study the effect of affirmative action implemented in the form of job quotas on educational IGM, I use the following measures to check for absolute upward education mobility: (1) probability that a son acquires higher years of education as compared to their father, (2) change in years of education to children born to illiterate fathers, (3) probability that a son born to an illiterate father is literate, (4) change in education rank of children born to below-median-education fathers, and (5) probability that a son born to a below-median-education father ends up in the top half of their education distribution. Similarly, to look at absolute downward education mobility, I use the following measures: (1) change in years of education to children born to at-least-high-school-graduate fathers (2) probability that a son born to an at-least-high-school-graduate father does not complete high school, (3) change in education rank of children born to above-median-education fathers, and (4) probability that a son born to a above-median-education father ends up in the bottom half of their education distribution.

4.1.3 Effect on Occupation IGM

I now turn to inter-generational occupation mobility. I use occupation transition probabilities to study occupation IGM. The conditional probability of an occupation transition from the father's generation to the son's generation is obtained in a manner similar to the education-transition probabilities. Thus, in Equation (1), y_{iga} now represents the occupation category instead of the education category and the sample is partitioned into subsamples using father's occupation categories. Occupation categories are defined in Section 3.4.4, but I combine (1) professional and skilled and (2) unskilled and farmer occupations together instead of producing the effect of job quotas using a 4×4 occupation transition matrix. I do this because it is not immediately obvious how farming and unskilled jobs rank against each other (for education, more years of education would be considered better, and thus there is no ambiguity in the interpretation of a full educational mobility matrix).

To study absolute upward occupation mobility, I use the probability that a son who is born to a father employed in an unskilled or farming occupation is employed in a professional or skilled occupation. Similarly, to study absolute downward occupation mobility, I use the probability that a son who is born to a father employed in an professional or skilled occupation is employed as a farmer or in an unskilled job.

4.2 Multiple Hypothesis Testing

Because I estimate Equation (1) for many different outcome variables even for the same sub-sample, there is a multiple inference concern that the probability that I reject at least one null hypothesis (i.e. $\beta_3 = 0$ in Equation (1) is greater than the significance level used for each individual hypothesis test. This probability increases with the number of hypotheses being simultaneously tested, and thus the p-value of the coefficients in each hypothesis test must be adjusted to account for this fact. There are various ways of adjusting the p-values in a multiple hypothesis test, the most common being (1) controlling the false discovery rate (FDR), or the proportion of rejections that are "false discoveries" (type I errors), (2) controlling the Family Wise Error Rate (FWER), or the probability of (incorrectly) rejecting at least one true null hypotheses belonging to a family of hypotheses i.e. the probability of making at least one false discovery within the family, and (3) the Bonferroni correction, which uses α/n to determine critical values, where α is the level of significance and n is the number of hypotheses to be tested.

In my analysis, I account for the multiple testing concern by controlling the FWER. The main advantage of using this method is that outcomes are not assumed to be independent of each other. For example, in my case, since I am considering the impact of treatment on the years of education as well as the probability of a son being literate, it is very likely that these two outcome variables are correlated. Both the Bonferroni correction and the FDR[31] assume independent outcomes.

To control the FWER, I first define mutually exclusive families of hypotheses that include all of my regressions. Hypotheses tests from different subsamples aren't considered together, as in Jones et al.

[31] At least, Anderson (2008)'s code for FDR sharpened q-values assumes this.

(2019). Thus, I have the following four family of hypotheses- (1) Subsample: sons born to illiterate fathers, Dependent variables: (i) years of education of son and (ii) binary variable indicating whether son is literate; (2) Subsample: sons born to at-least-high-school-graduate fathers, Dependent variables: (i) years of education of son and (ii) binary variable indicating whether son is not a high school graduate; (3) Subsample: sons born to below-median-education-rank fathers, Dependent variables: (i) education rank of son and (ii) binary variable indicating whether son is in the top half of their education distribution; and (4) Subsample: sons born to above-median-education-rank fathers, Dependent variables: (i) education rank of son and (ii) binary variable indicating whether son is in the bottom half of their education distribution.

I use 10,000 bootstraps of the free step-down procedure of calculating family-wise adjusted p-values listed in Westfall and Young (1993) when testing multiple hypotheses. The adjusted p-values would have increased as a result of adding more outcome variables to each family of hypotheses, but since I have only two regression models per family, I do not expect the adjusted p-values to be dramatically different from the unadjusted p-values.

4.3 Possible Concerns in the Difference-in-Differences Analysis

In an ideal world, we would use well designed experiments to uncover causal effects of policy interventions. However, in the real world, randomized experiments are costly or just infeasible in many substantive domains of interest, and economists use empirical design-based tools that widen the range of possible variations that can be used to uncover causal impacts. Difference-in-differences is "probably the most widely applicable design-based estimator" (Angrist and Pischke, 2010), but its credibility rests on some identification assumptions. I now list the main identification concerns in the use of a difference-in-differences estimator to explain the effect of job quotas on the IGM of the OBCs.

First, it is possible that the parallel trends assumption (control units provide a good counterfactual of the trend that the treated units would have followed if they had not been treated) is not satisfied i.e. the change in the outcomes of the OBC children that would have occured absent the quota policy differs from the change in children belonging to the general category. This may arise if the selected control group is not really a good control group. Because this parallel trends assumption is ultimately untestable, it is, in practise, tested using pre-trends. Pre-trends is neither neither implied nor is implied by the identifying assumption, and the test is a suggestive one. The logic here is that if we have sufficient pre-treatment periods and parallel trends in the outcome variable are exhibited for the two groups in the absence of treatment, then it is more plausible to establish the parallel trends identifying assumption.

Researchers test parallel pre-trends using three methods. First, by plotting the raw treatment and control time series of the outcome variable and eyeballing if the two trends are parallel in the pre-treatment period. If we see that the levels of the two groups are similar to begin with, then it makes the parallel trends assumption more credible (Kahn-Lang and Lang, 2020). Kahn-Lang and Lang (2020) also advise that any difference-in-differences paper finding different levels should justify why this should be the case. In my case, I would expect that the levels of education or occupation patterns of the OBCs and the general category caste group would differ, due to the centuries-old caste-based discrimination ingrained into the social-cultural milieu. The general category has been a historically advantaged group and the affirmative action policy seeks to correct this imbalance. Thus, this mechanism would not affect the trends so long as affirmative action policies are not introduced. It is clear from Figures 4-12 that the levels of y_{iga}, for the particular subsample considered, are different for the general category and OBC sons who were born before 1971. The pre-trends do not seem all that parallel in many of these figures.

Second, one could plot the difference in the outcome variable between the two groups, relative to the difference in a base year (typically the time period immediately before the policy implementation). This is called an event study graph and researchers then test whether the regression coefficients on the treatment group in the pre-period (relative to the base year) are individually or jointly statistically significant. Figures B1-B11 plot these event studies. To plot these figures, I aggregate individuals into 5 year birth cohorts so that I have sufficient observations for individuals belonging to a particular caste and birth-cohort group and can thus carry out somewhat meaningful regressions. The base birth cohort is individuals aged 25-30 in 1993 i.e. the first pre-treatment period that I would have for these 5-year birth cohort categories. In these regressions, I do not include any controls, state-caste fixed effects, or state-cohort fixed effects. The β coefficients for each 5-year birth cohort category, thus, shows the difference in the outcome variable between the two castes, relative to the difference for individuals aged 25-30 in 1993. As can be seen in these figures, the pre-treatment coefficients are individually insignificant (except in Figure B10). However, it is a bit concerning that the confidence intervals are very wide and the coefficients insignificant even after the policy implementation.

Third, if there are a limited number of pre-treatment time periods, or if there is not sufficient power because of a small number of observations in each time period, one could replicate the model using two periods prior to the treatment, in which case, the second period becomes the placebo treatment period. The pre-trends assumption is likely to hold if the coefficient on the interaction term between the treatment group and the placebo treatment year is insignificant. In Section 5.4, I test parallel pre-trends using the third method. I prefer this over the other two methods, because after subsampling based on father's occupation or education (as discussed in Section 4.1, I do not have enough observations in the two groups in each time period. The third method also makes more sense as I control for other characteristics that could affect the outcome variable and be correlated with caste, such as state-caste fixed effects to account for different state-level caste-based policies, in the placebo regressions. I do not do so in the first or the second method.

Kahn-Lang and Lang (2020) suggest that if the estimates of the treatment effect change after adding a linear (or other) group-specific trends, then we must proceed with caution when interpreting the results as causal. However, adding even a linear group-specific time trend can dramatically decrease the power to detect a true significant effect due to the resulting loss of degrees of freedom. For example, if the causal effect of treatment increases over time (as I expect would be in my analysis, since OBC children who are younger would have more time to change their human capital formation decisions in response to the job quotas), the treatment effect in the difference-in-differences analysis would be an average of these heterogeneous treatment effects. If we add a linear time trend, then the coefficient on the treatment could have the wrong sign (Kahn-Lang and Lang, 2020). Thus, I neither include caste-specific time trends in my main analysis to mitigate the possibility that the parallel trends assumption is not satisfied, nor use that to check robustness.

The second concern is the assumption that there are no spillover effects of the policy change on caste groups other than the OBCs. A widely-held belief regarding reservations in India is that they harm the non-targeted social groups i.e. the forward castes by effectively decreasing the percentage of seats available for them in government jobs and higher educational institutions. Indeed, any proposed increase in reserved seats is met by large-scale protests and strikes. For e.g., the implementation of the Mandal Commission Reforms, the specific affirmative action policy that I consider, was met by violent protests in the 1990s in which some upper-caste students even self-immolated themselves. Thus, the upper-caste students might feel discouraged by the resultant increase in competition and work less hard, or alternatively, they might increase their effort to be able to secure these seats. When spillover effects are suspected, identifying a suitable control group in a non-experimental research design is, therefore, further complicated by having to identify a subject group (a subset of the ineligible groups of the treatment policy) for which spillover effects are possible and a control group that won't be affected by the treatment at all. The other two caste groups in India, the SCs and the STs, should not be affected by the implementation of the job quotas for the OBCs as the percentage of seats reserved for them did not change. However, as mentioned in Section 3.3, they also experienced a policy change that affected their access to public resources at the same time as the implementation of the job quotas for the OBCs, and thus, can't be suitable control groups. The suspected spillover effects are offset to some degree by political promises that the total number of government jobs would be increased so that the absolute number of jobs available to the upper-caste members is unchanged (Khanna, 2020), but the difference-in-differences model will prove to be inadequate if the spillover effects still remain large. As a more recent example, the government implemented a 10% quota in educational institutions for economically weaker sections of the general category in 2019. To ensure that the pre-2019 reservation remains unaffected, the government plans to increase the total seats in educational institutions by as much as 25%.

The third assumption is that there are no anticipatory effects of the policy. Given the long and tumultuous journey of the OBC quotas, it is unlikely that households actually expected the quotas to be actually implemented. As mentioned in Section 2.3, the initial impetus for OBC quotas came in the 1950s, but no action was taken by the government for decades. Even when the government first announced their intention to act upon the recommendations of the Mandal commission, for a couple of years the fate of the policy pretty much hung in the balance due to the widespread anti-reservation protests. Finally, whether the quotas would be imposed or not was ultimately decided by the Supreme Court. Indian courts are notorious for cases being dragged on for several years, even decades, and thus, it was certainly a surprise that the verdict was announced within one year. Hence, I do not expect that the prospect of the policy being adopted changed between 1989 and 1993.[32]

5 Results

I divide this section into three subsections. I present results on the effect of the implementation of the Mandal Commission Reforms in 1993 on absolute education mobility when education is measured as levels. In the next subsection, I discuss the effect on absolute education mobility when education is measured as ranks. The third subsection contains results on absolute occupation mobility. Finally, the fourth subsection conducts a robustness check by assigning a placebo treatment year.

5.1 Education Mobility when Education is Measured in Levels

Table 2 presents the impact of the OBC job quotas on education measured in years. Table 2 is read as follows. Each column of the table represents a different OLS regression defined by the dependent variable and the sample used. I include state-age fixed effects, state-caste fixed effects, and controls (defined in Section 3.4.5) in all these regressions.

Column (1) of Table 2 estimates Equation (1) using the probability that a son acquires higher education than his father as the dependent variable and the entire sample. For representational ease, I omit reporting the coefficients on the fixed effects or the covariates. I report only the coefficient on the interaction term between obc_g and $young_a$ i.e. β_3. I also include its standard error, unadjusted p-value, p-value adjusted using the FWER method, and the p-value using the Bonferroni correction method. β_3 in Column (1) is significant and positive, which means that the job quotas lead to an increase in this probability by 6.39 percentage points, indicating absolute upward education mobility.

However, as mentioned in Section 4.1.1, this number could mask important heterogeneity. Thus, in Columns (2) and (3), I restrict my sample to sons born to illiterate fathers. Column (2) uses the years of

[32] A robustness check could be to assume that cohorts younger than 22 in 1989 rather than 1993 would be affected by the policy. If the results from this treatment year do not change much in magnitude as compared to the results using 1993 as the treatment year, then it would suggest that the policy had no anticipatory effects.

education attained by the son as the dependent variable, while Column (3) uses the probability of such son being literate. Again, I find positive and significant coefficients, which indicate absolute upward education mobility. The quotas increased the average education of these sons by 0.619 years, and the probability that they are literate by 5.05 percentage points. These two results, therefore, also indicate upward absolute education mobility.

Column (4) shows the effect of the policy on the years of education acquired by sons, conditional on their fathers having at least high school education. Column (5) uses the same subsample as Column (4), but uses the probability that the son is not a high school graduate as the dependent variable. Had the β_3 coefficient in these two columns been negative and significant, we would have interpreted that as evidence of absolute downward education mobility. Reassuringly, these coefficients are insignificant.

5.2 Education Mobility when Education is Measured As Ranks

Table 3 provides the impact of the job quotas on the measures of absolute education mobility when education is measured as rank. Because the ranks are defined in the (cohort-specific) national education distribution of sons, this transition probability can be interpreted as a measure of absolute upward mobility and is a simple measure of success.

Table 3 is obtained by estimating Equation (1) for different dependent variables and subsamples. Similar to Table 2, I include state-age fixed effects, state-caste fixed effects, and controls in all these regressions and present information only for the β_3 coefficient. Column (1) of Table 2 estimates Equation (1) using the rank of a son as the dependent variable and the subsample as sons with below-median-education fathers. β_3 in this case is positive and significant, and thus, the job quotas resulted in an increase in the average education rank of OBC sons born to below-median-education fathers by 3.5. In Column (2), the dependent variable is a binary variable that is equal to one if the son's education is above the median, and zero otherwise. The subsample considered is still below-median-education fathers. Since the coefficient on the interaction term between obc_g and $young_a$ is significant, again, this signals that absolute upward mobility of the OBCs increased in response to the implementation of the job quotas for them. The quotas increased the probability that a son born to a below-median-education father ends up in the top half of their education distribution by 5.64 percentage points.

In Columns (3) and (4), I try to ascertain if the quotas affected the absolute downward mobility of OBCs. In these specifications, I restrict the subsample to sons with fathers who have education above the median level of education. Column (3) uses the rank of son as the dependent variable, while in Column (4), the dependent variable is a binary variable that is equal to one if the son's education is below median, and zero otherwise. The β_3 coefficient in these two columns are insignificant, indicating that the policy did not result in an absolute downward mobility of the OBCs.

5.3 Occupation Mobility

Table 4 contains results of the effect of the policy on inter-generational occupation transitions. Column (1) of table 4 restricts the sample to sons born to fathers engaged in farming or other unskilled occupations. The dependent variable used in the estimation of Equation 1 is a binary variable that takes the value one if the son's occupation is professional or skilled, and zero otherwise. A significant and positive coefficient on the interaction between obc_g and $young_a$ would have indicated that the quotas resulted in absolute upward occupation mobility of OBCs. However, the coefficient is negative and significant, perhaps due to the mechanism outlined early on in Section 1. The quotas could have exacerbated the negative stereotypes private employers have about OBCs, and even though their employment opportunities in the public sector increased, their employment opportunities in the private sector could have decreased. Since the private sector forms a considerable part of the formal sector in India, the overall result could be a decline in the probability of being employed in the skilled or professional occupations.

Column (2) of this table restricts the sample to sons born to fathers employed in skilled or professional occupations. The dependent variable used in the estimation of Equation (1) is a binary variable that takes the value one if the son's occupation is farming or unskilled, and zero otherwise. The β_3 coefficient is small and statistically insignificant, which means that the quotas did not change the probability that an OBC son born to skilled fathers ends up in an unskilled occupation.

Hence, overall, the job quotas for OBCs resulted in their absolute downward education mobility.

5.4 Robustness: Placebo Treatment Effect

The large time span of my data set allows me to conduct a placebo test by reassigning the year of treatment from 1993 to a year before 1993. Thus, I use two older generations: those aged 25 to 40 years old in 1993 and those aged 40 to 60 years old in 1993. Both these generations would have been too old for the implementation of the job quotas to have had any direct effect on them. If my results in Tables 2 - 4 are truly capturing the causal effect of the OBC job quotas, then I should not find any significant results with this new definition of "young" (i.e. those 25-40 years old in 1993) and"old" (i.e. those 40-60 years old in 1993) birth cohorts. Tables 5- 7 repeat the analysis in Tables 2 - 4 respectively and show that I find insignificant coefficients (except in one regression) on the interaction between OBC and this new "young" variable. This test lends support to my difference-in-differences estimator and give more credibility to interpreting the main results as causal.

6 Conclusion and Future Work

Centuries-old caste-based discriminatory practises and attitudes have given rise to heterogeneity in the existing socio-economic status in India, with only some who are able to climb the socio-economic ladder, while the rest either remain at the same position in society or find their position declining (Ramaiah, 1992). Affirmative action policies, enacted as reservation quotas in jobs, higher educational institutes, and legislatures, try to correct this imbalance. Using a difference-in-differences strategy on a matched son-father pair sample constructed from IHDS-II, I have evaluated the impact of the implementation of job quotas for the OBCs on their absolute education and occupation inter-generational mobility. I find that the quotas resulted in an increase in the absolute upward education mobility of OBC sons (measured in terms of levels as well as ranks), but also decreased their absolute occupation mobility. Thus, OBC sons acquired more education in response to perceived future opportunities, but the downward occupation mobility could perhaps be understood by the exacerbation of negative attitudes of private employers regarding OBCs in response to the quotas.

In future work, I plan to use *jati* information of households in the following two ways. First, I want to see if there was an effect of job quotas on individuals who are not classified as OBC, but could still be affected through the caste-network channel. Each state has its own list of *jatis* which are considered OBC. Recent genetic evidence suggests that inter-caste breeding has been rare for the past 2,000 years (Munshi, 2019), a fact that is reflected in survey data as well, with over 95% of marriages in IHDS-II being within couples belonging to the same caste. Castes are usually spatially segregated within villages, which results in the prevalence of caste-based social networks. These caste clusters within a village are connected to other villages, even across states by matrimonial ties. The resultant caste networks are of enormous size and scope and help in job referrals, credit for consumption smoothing, and providing an easier access to capital, know-how, and connections for business activities (Munshi, 2019). The caste-based networks could have also helped the castes who are not classified as OBCs in their own state, but are classified as OBCs in another state. The major challenge in undertaking this analysis is harmonizing the *jati* names. IHDS reports *jati* names verbatim and the total number of unique *jati* names without any data cleaning is around a 1,000, when the total number of castes in India is between 2,000-3,000. Data cleaning is complicated as (1) caste itself has an ambiguous meaning[33] and (2) there are many synonyms and spellings of the same *jati*. Spellings of the same *jati* could be identified using fuzzy string matching algorithms such as fingerprint and nearest-neighbour matching. To identify synonyms of a *jati*, Cassan et al. (2021) suggests referring to Singh (1996), which lists all *jati* names and their synonyms at the state level. Kitts (1885) also contains this information, though Singh (1996) is a more contemporary source.

[33]For example, many people report their caste or *jati* as Hindu/Muslim/Christian, etc. IHDS also asks the subcaste of individuals, and in majority of these cases, it makes more sense to use the reported subcaste information as caste.

Second, I want to test whether the job quotas affected the probability that an individual is employed in his traditional, *jati*-based occupation, especially if his father is or was employed in the traditional occupation. This analysis would be interesting because one of the central premises of the caste system was that each *jati* was linked to a particular occupation, and the *jati* of an individual was effectively determined at birth due to strict marital endogamy and the hereditary nature of the caste system. Traditional occupations of different *jatis* can be identified from the the colonial Census of 1911 or Kitts (1885). The traditional occupations must be converted to the 1968-NCO codes that are used in IHDS-II.

References

Abadie, A., Athey, S., Imbens, G. W., and Wooldridge, J. (2017). When should you adjust standard errors for clustering? Technical report, National Bureau of Economic Research.

Alesina, A., Hohmann, S., Michalopoulos, S., and Papaioannou, E. (2021). Intergenerational mobility in africa. *Econometrica*, 89(1):1–35.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American statistical Association*, 103(484):1481–1495.

Angrist, J. D. and Pischke, J.-S. (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of economic perspectives*, 24(2):3–30.

Asher, S., Novosad, P., and Rafkin, C. (2018). Intergenerational mobility in india: Estimates from new methods and administrative data. *World Bank Working Paper.*

Azam, M. (2013). Intergenerational occupational mobility in india. *Available at SSRN 2287215.*

Azam, M. and Bhatt, V. (2015). Like father, like son? intergenerational educational mobility in india. *Demography*, 52(6):1929–1959.

Bardhan, P. K., Mookherjee, D., and Torrado, M. P. (2010). Impact of political reservations in west bengal local governments on anti-poverty targeting. *Journal of Globalization and development*, 1(1).

Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly journal of economics*, 119(1):249–275.

Cameron, A. C. and Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of human resources*, 50(2):317–372.

Card, D., Domnisoru, C., and Taylor, L. (2018). The intergenerational transmission of human capital: Evidence from the golden age of upward mobility. Technical report, National Bureau of Economic Research.

Cassan, G., Keniston, D., and Kleineberg, T. (2021). A division of laborers: Identity and efficiency in india. Technical report, National Bureau of Economic Research.

Chandavarkar, R. (2003). *The Origins of Industrial Capitalism in India: Business strategies and the working classes in Bombay, 1900-1940*, volume 51. Cambridge University Press.

Chetty, R., Grusky, D., Hell, M., Hendren, N., Manduca, R., and Narang, J. (2017). The fading american dream: Trends in absolute income mobility since 1940. *Science*, 356(6336):398–406.

Chetty, R., Hendren, N., Kline, P., and Saez, E. (2014). Where is the land of opportunity? the geography of intergenerational mobility in the united states. *The Quarterly Journal of Economics*, 129(4):1553–1623.

Coate, S. and Loury, G. C. (1993). Will affirmative-action policies eliminate negative stereotypes? *The American Economic Review*, pages 1220–1240.

Derenoncourt, E. (2019). Can you move to opportunity? evidence from the great migration. *Unpublished Manuscript*.

Grawe, N. D. (2006). Lifecycle bias in estimates of intergenerational earnings persistence. *Labour economics*, 13(5):551–570.

Hertz, T., Jayasundera, T., Piraino, P., Selcuk, S., Smith, N., and Verashchagina, A. (2008). The inheritance of educational inequality: International comparisons and fifty-year trends. *The BE Journal of Economic Analysis & Policy*, 7(2).

Hnatkovska, V., Lahiri, A., and Paul, S. B. (2013). Breaking the caste barrier intergenerational mobility in india. *Journal of Human Resources*, 48(2):435–473.

Iversen, V., Krishna, A., Sen, K., et al. (2016). Rags to riches? intergenerational occupational mobility in india. Technical report, GDI, The University of Manchester.

Jodhka, S. S. (2004). Sikhism and the caste question: Dalits and their politics in contemporary punjab. *Contributions to Indian sociology*, 38(1-2):165–192.

Jones, D., Molitor, D., and Reif, J. (2019). What do workplace wellness programs do? evidence from the illinois workplace wellness study. *The Quarterly Journal of Economics*, 134(4):1747–1791.

Kahn-Lang, A. and Lang, K. (2020). The promise and pitfalls of differences-in-differences: Reflections on 16 and pregnant and other applications. *Journal of Business & Economic Statistics*, 38(3):613–620.

Khanna, G. (2020). Does affirmative action incentivize schooling? evidence from india. *Review of Economics and Statistics*, 102(2):219–233.

Kitts, E. J. (1885). *A compendium of the castes and tribes found in India.* Bombay: Education Society's Press Byculla.

Kundu, A., Sen, K., et al. (2021). Multigenerational mobility in india.

Lee, S. Y. and Seshadri, A. (2019). On the intergenerational transmission of economic status. *Journal of Political Economy*, 127(2):855–921.

Mayer, A. (1996). Caste in an indian village: Change and continuity 1954–1992. *Caste today*, pages 32–64.

Mendelsohn, O. (1993). The transformation of authority in rural india. *Modern Asian Studies*, 27(4):805–842.

Munshi, K. (2019). Caste and the indian economy. *Journal of Economic Literature*, 57(4):781–834.

Nakao, K. and Treas, J. (1994). Updating occupational prestige and socioeconomic scores: How the new measures measure up. *Sociological methodology*, pages 1–72.

Osborne, E. (2001). Culture, development, and government: Reservations in india. *Economic Development and Cultural Change*, 49(3):659–685.

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

Piketty, T. (2000). Theories of persistent inequality and intergenerational mobility. *Handbook of income distribution*, 1:429–476.

Prakash, N. (2020). The impact of employment quotas on the economic lives of disadvantaged minorities in india. *Journal of Economic Behavior & Organization*, 180:494–509.

Ramaiah, A. (1992). Identifying other backward classes. *Economic and Political Weekly*, pages 1203–1207.

Ray, T., Chaudhuri, A. R., and Sahai, K. (2020). Whose education matters? an analysis of inter caste marriages in india. *Journal of Economic Behavior & Organization*, 176:619–633.

Restuccia, D. and Urrutia, C. (2004). Intergenerational persistence of earnings: The role of early and college education. *American Economic Review*, 94(5):1354–1378.

Rudner, D. (1996). David west rudner. caste and capitalism in colonial india: The nattukottai chettiars. berkeley and los angeles: University of california press. 1994. pp. xix, 341.

Singh, K. S. (1996). *Communities, segments, synonyms, surnames and titles*, volume 8. Oxford University Press.

Solon, G. (1999). Intergenerational mobility in the labor market. In *Handbook of labor economics*, volume 3, pages 1761–1800. Elsevier.

Westfall, P. H. and Young, S. S. (1993). *Resampling-based multiple testing: Examples and methods for p-value adjustment*, volume 279. John Wiley & Sons.

Williams, H. L. (2015). Intergenerational mobility: Theory, measurement, and empirics.

Zimmerman, D. J. (1992). Regression toward mediocrity in economic stature. *The American Economic Review*, pages 409–429.

Figures

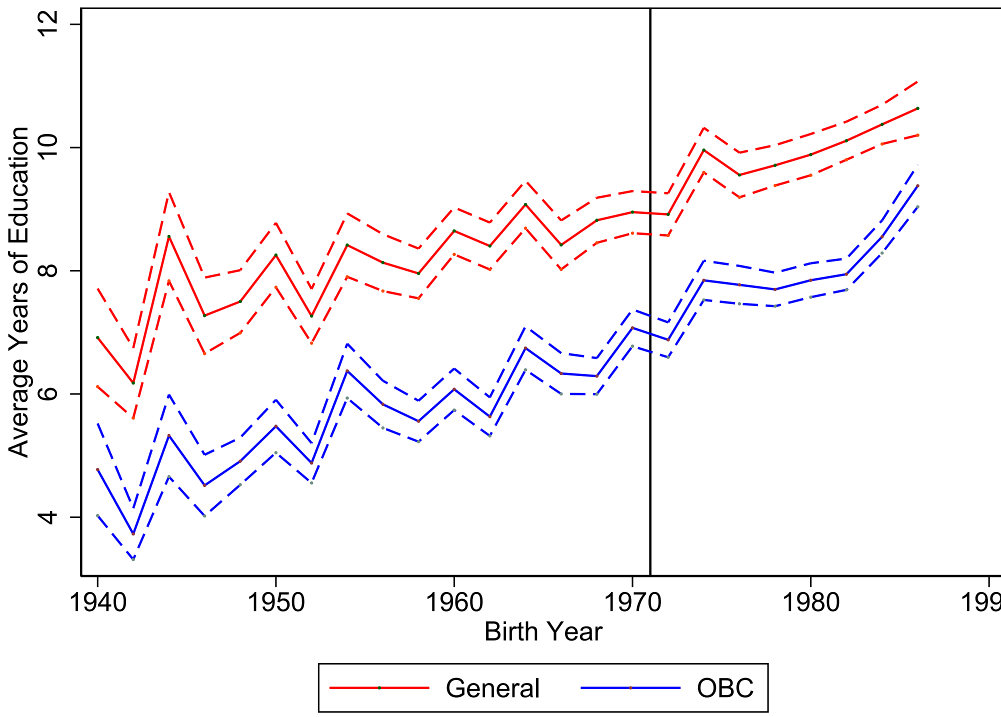

Figure 1: Years of Education

This figure plots the average value of the years of education acquired by different birth cohorts. See Section 3.4.1 for variable definition. The solid red line denotes average education for the general caste category, and the blue line denotes the average education for the OBCs. Dashed lines represent the 95% confidence intervals. The solid black line at birth year= 1971 indicates the earliest birth cohorts treated by the implementation of the job quotas.

,

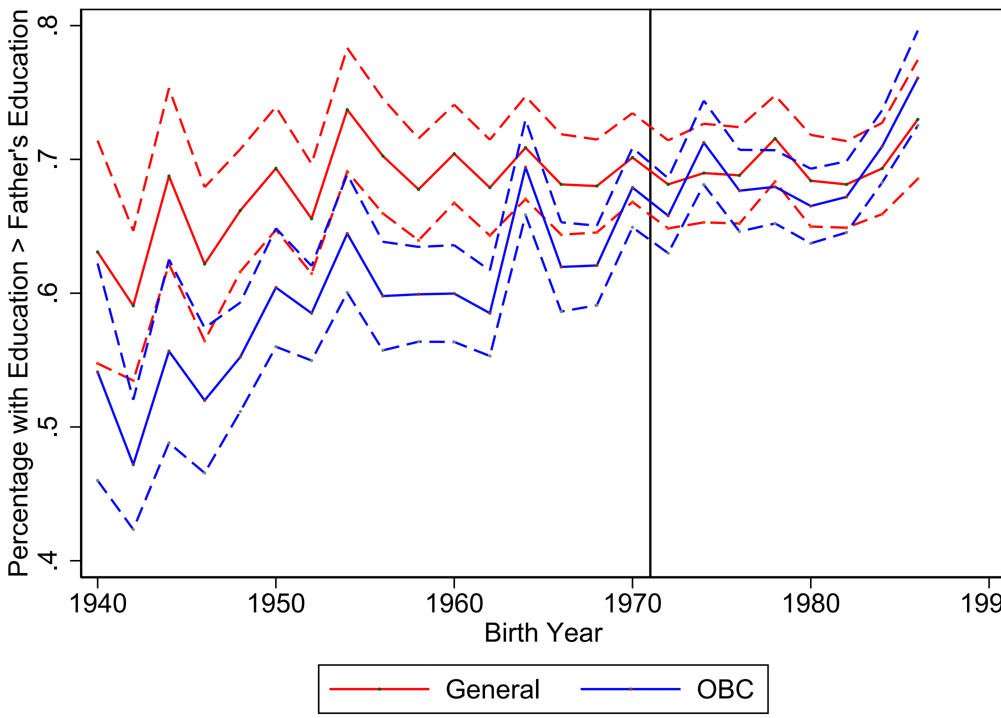

Figure 2: Percentage of Individuals with Education Greater Than Their Father's Education

This figure plots the percentage of individuals acquiring education greater than their father's education for different birth cohorts. See Section 3.4.1 for variable definition and construction. The solid red line denotes the percentage for the general caste category, and the blue line denotes the percentage for the OBCs. Dashed lines represent the 95% confidence intervals. The solid black line at birth year= 1971 indicates the earliest birth cohorts treated by the implementation of the job quotas.

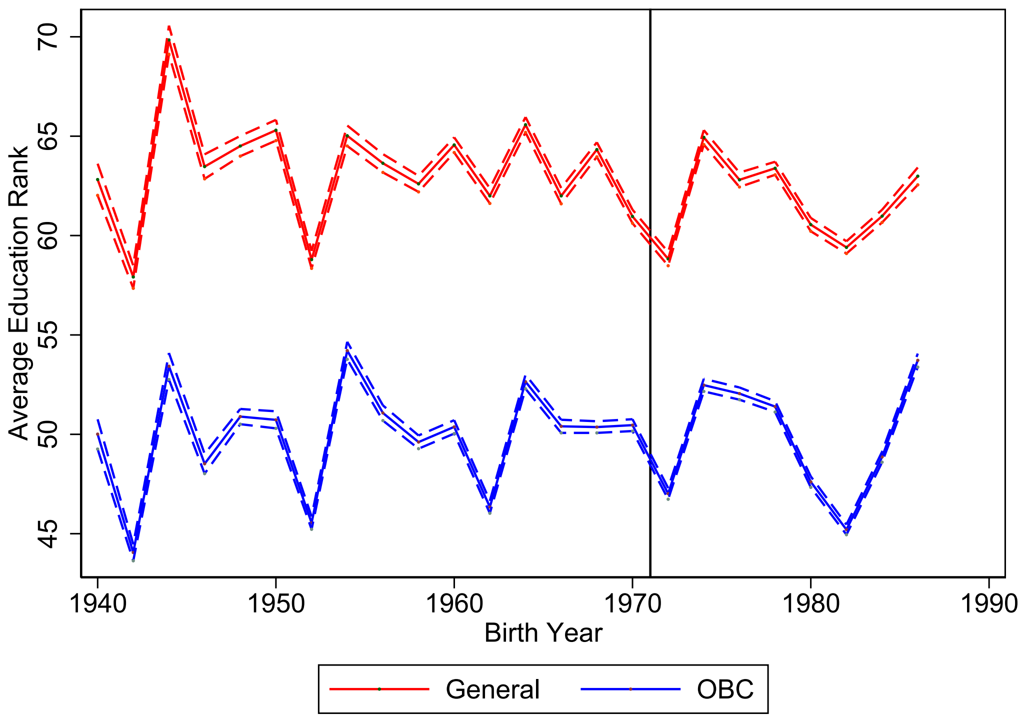

Figure 3: Education Rank

This figure plots the average value of the education rank acquired by different birth cohorts. See Section 3.4.3 for variable definition and construction. The solid red line denotes average education for the general caste category, and the blue line denotes the average education for the OBCs. Dashed lines represent the 95% confidence intervals. The solid black line at birth year= 1971 indicates the earliest birth cohorts treated by the implementation of the job quotas.

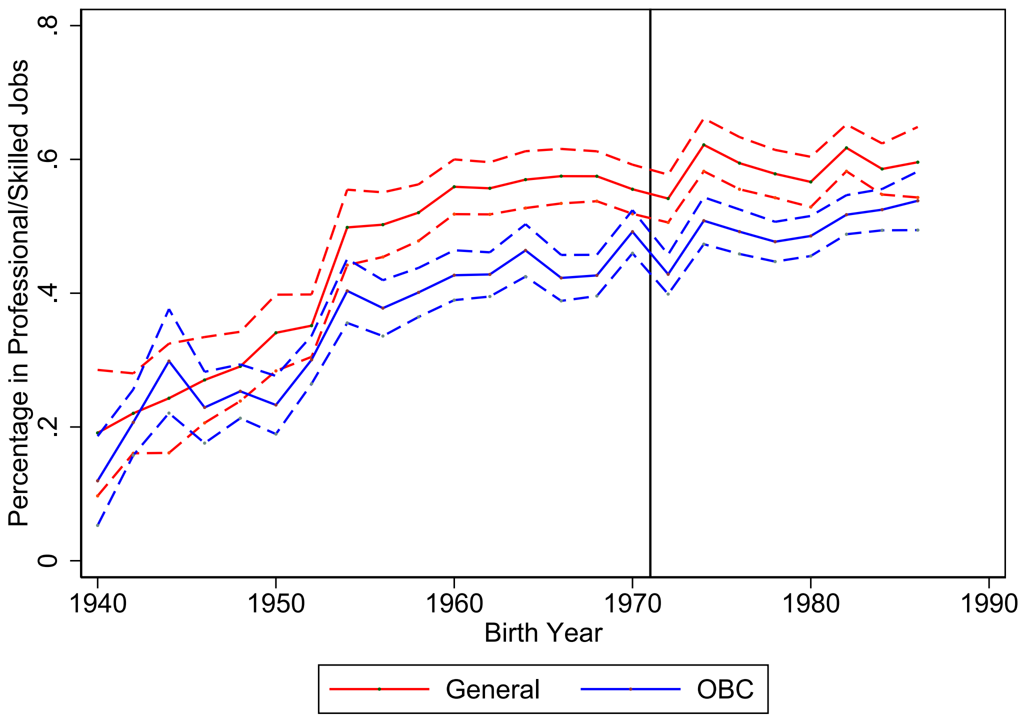

Figure 4: Percentage of Individuals in Professional/Skilled Jobs over Time

This figure plots the percentage of individuals employed in professional or skilled occupations for different birth cohorts. See Section **??** for variable definition and construction. The solid red line denotes this percentage for the general caste category, and the blue line denotes this percentage for the OBCs. Dashed lines represent the 95% confidence intervals. The solid black line at birth year= 1971 indicates the earliest birth cohorts treated by the implementation of the job quotas.

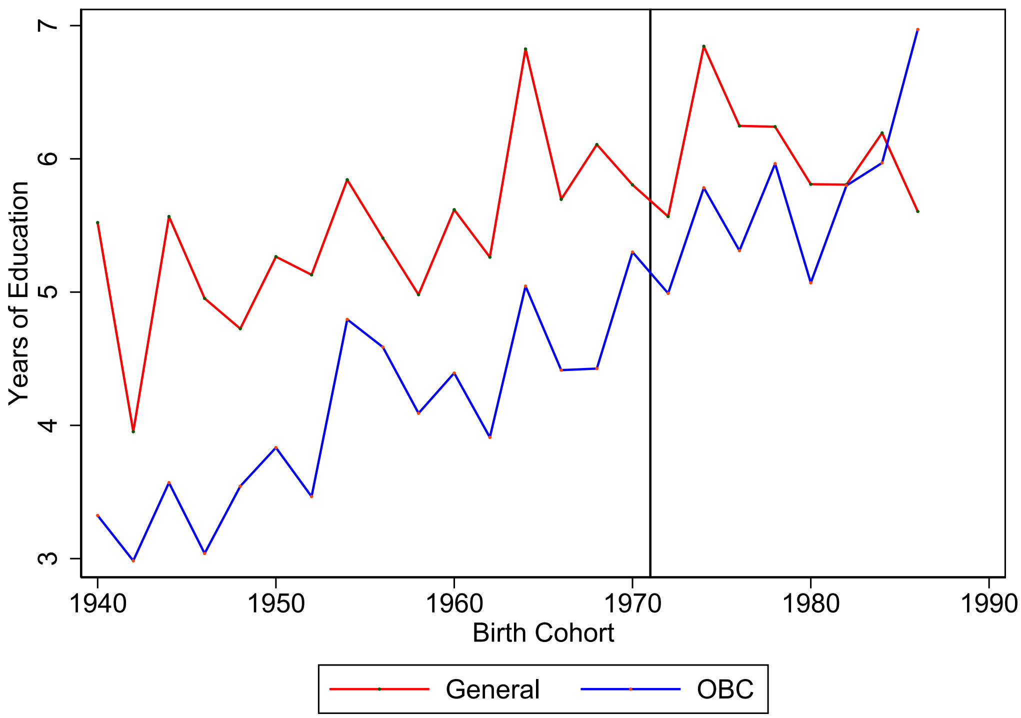

Figure 5: Years of Education for Sons Born to Illiterate Fathers

This figure plots the average value of the years of education acquired by sons born in different birth cohorts to illiterate fathers. See Section 3.4.1 and 3.4.2 for variable and sample definition. The solid red line denotes average education for the general caste category, and the blue line denotes the average education for the OBCs. The solid black line at birth year= 1971 indicates the earliest birth cohorts treated by the implementation of the job quotas.

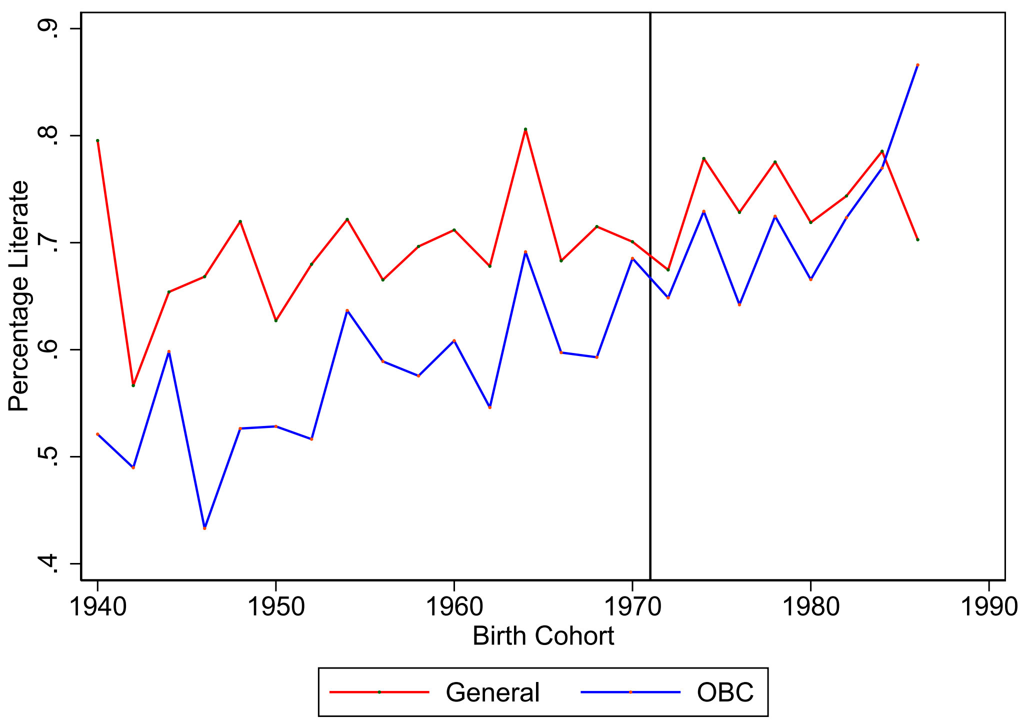

Figure 6: Percentage of Sons Born to Illiterate Fathers who are Literate

This figure plots the percentage of literate sons born in different birth cohorts to illiterate fathers. See Section 3.4.2 for variable and sample definition. The solid red line denotes this percentage for the general caste category, and the blue line denotes this percentage for the OBCs. The solid black line at birth year= 1971 indicates the earliest birth cohorts treated by the implementation of the job quotas.

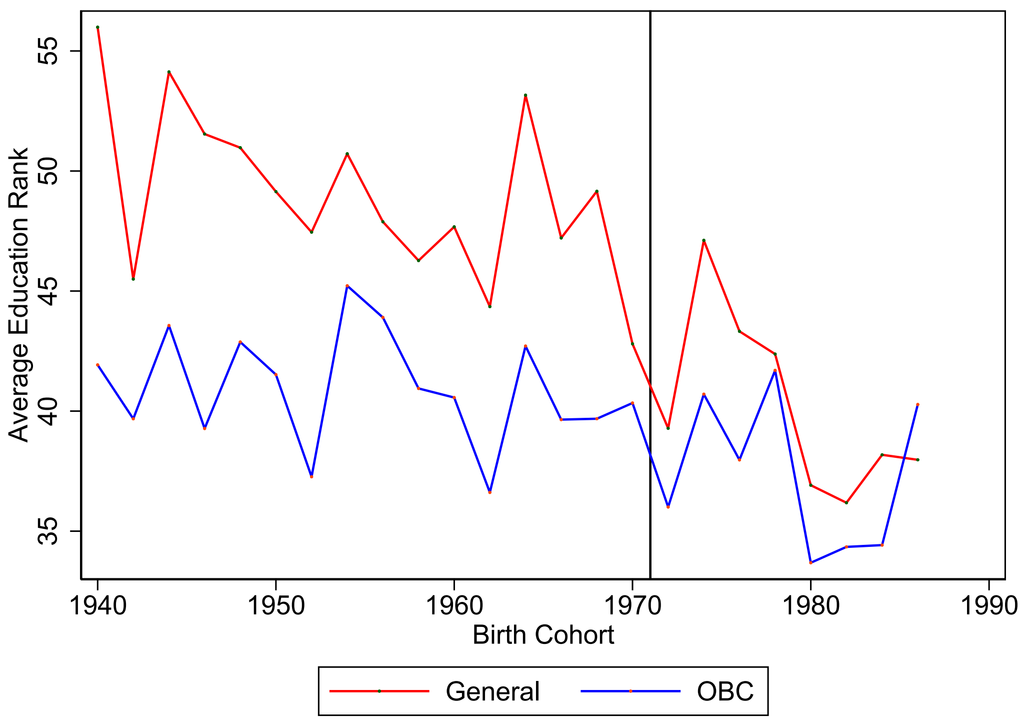

Figure 7: Average Education Rank for Sons Born to Below-Median-Education Fathers

This figure plots the average value of the education rank acquired by sons born in different birth cohorts to fathers who were in the bottom half of the education distribution. See Section 3.4.3 for variable definition and sub-sample construction. The solid red line denotes average education rank for the general caste category, and the blue line denotes the average education rank for the OBCs. Dashed lines represent the 95% confidence intervals. The solid black line at birth year= 1971 indicates the earliest birth cohorts treated by the implementation of the job quotas.

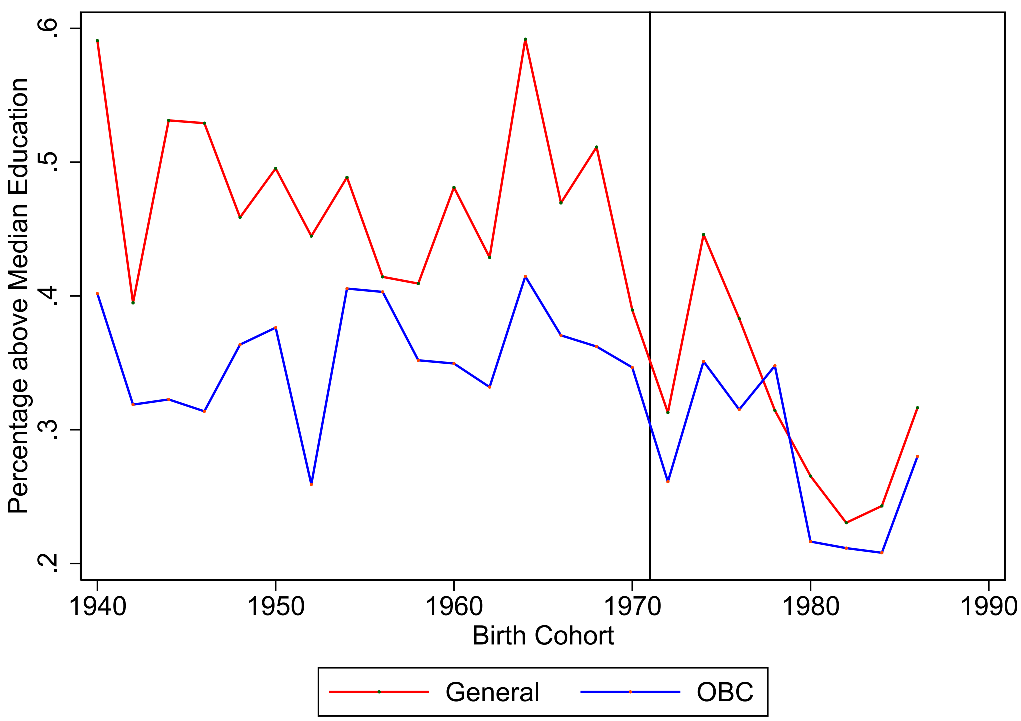

Figure 8: Percentage of Sons born to Below-Median-Education Fathers who are Above Median

This figure plots the percentage of above-median-education sons born in different birth cohorts to fathers who were in the bottom half of the education distribution. See Section 3.4.3 for variable definition and sub-sample construction. The solid red line denotes this percentage for the general caste category, and the blue line denotes this percenage for the OBCs. The solid black line at birth year= 1971 indicates the earliest birth cohorts treated by the implementation of the job quotas.

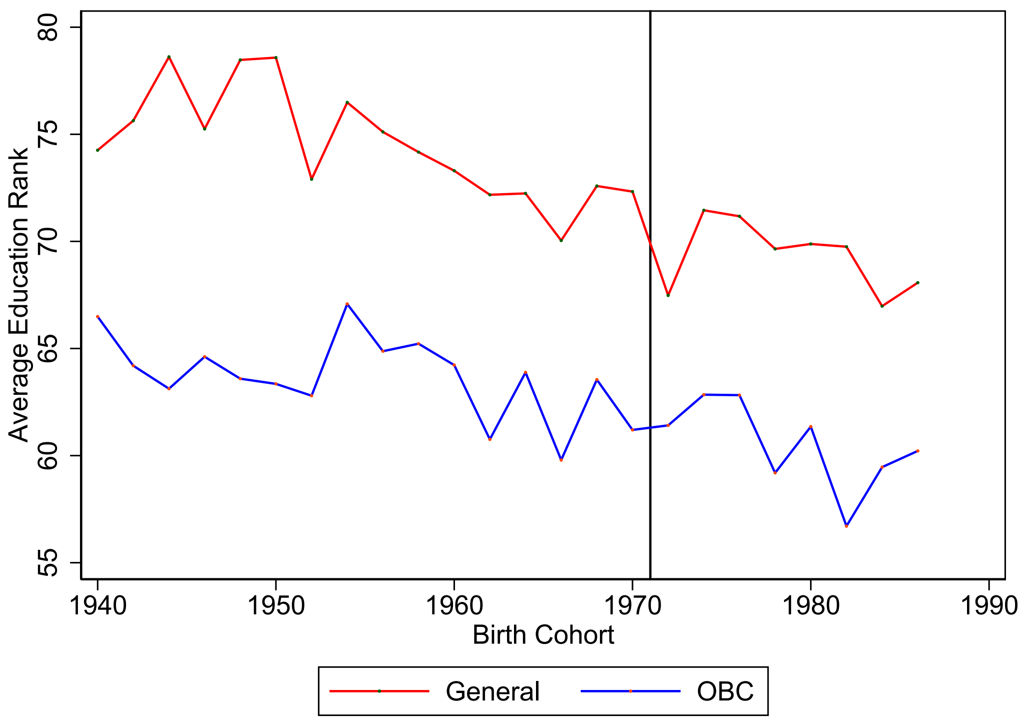

Figure 9: Average Education Rank for Sons Born to Above-Median-Education Fathers

This figure plots the average value of the education rank acquired by sons born in different birth cohorts to fathers who were in the top half of the education distribution. See Section 3.4.3 for variable definition and sub-sample construction. The solid red line denotes average education rank for the general caste category, and the blue line denotes the average education rank for the OBCs. The solid black line at birth year= 1971 indicates the earliest birth cohorts treated by the implementation of the job quotas.

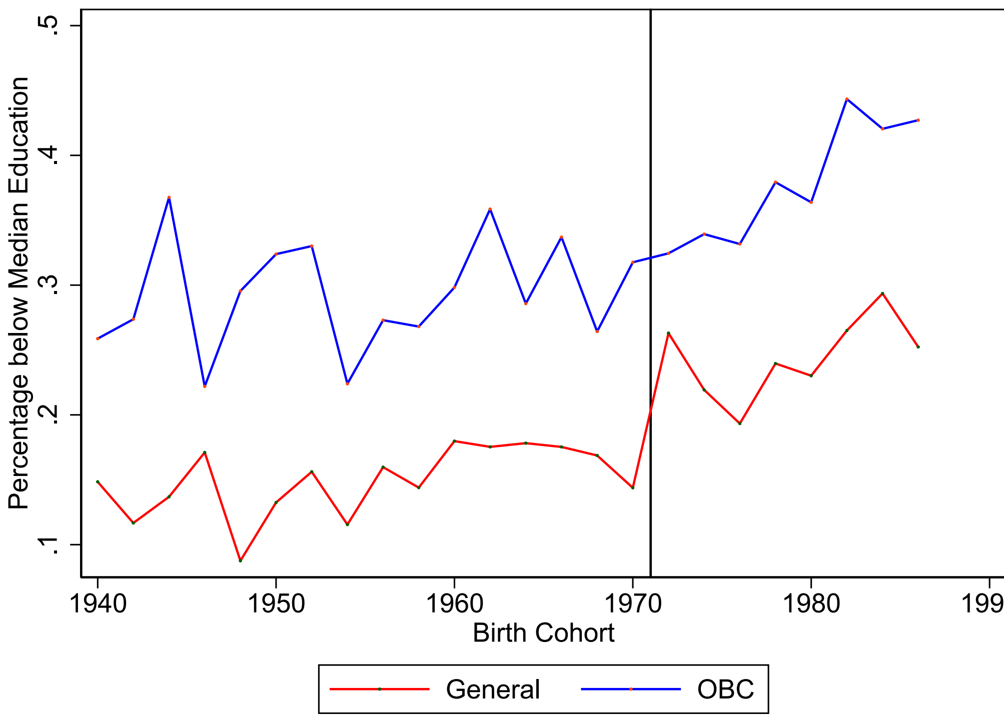

Figure 10: Percentage of Sons born to Above-Median-Education Fathers who are Below Median

This figure plots the percentage of below-median-education sons born in different birth cohorts to fathers who were in the top half of the education distribution. See Section 3.4.3 for variable definition and sub-sample construction. The solid red line denotes the percentage for the general caste category, and the blue line denotes the percentage for the OBCs. The solid black line at birth year= 1971 indicates the earliest birth cohorts treated by the implementation of the job quotas.

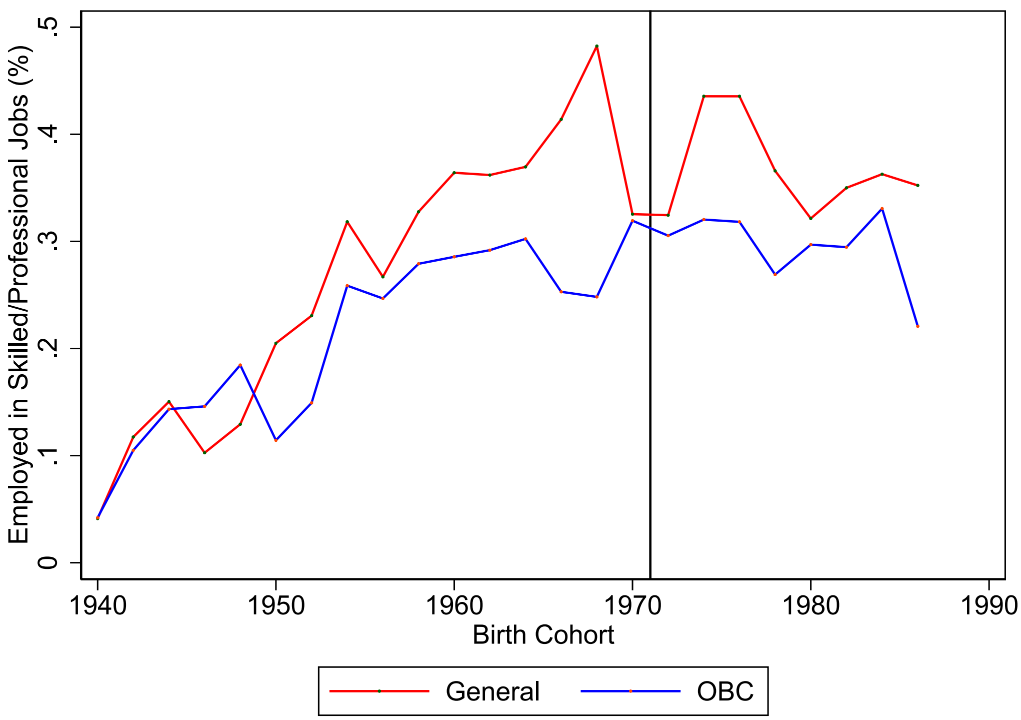

Figure 11: Percentage of Sons in Professional/Skilled Jobs born to Unskilled/Farmer Fathers

This figure plots the percentage of sons employed in professional or skilled occupations born in different birth cohorts to fathers employed in unskilled professions or farming. See Section **??** for variable definition and sub-sample construction. The solid red line denotes the percentage for the general caste category, and the blue line denotes the percentage for the OBCs. The solid black line at birth year= 1971 indicates the earliest birth cohorts treated by the implementation of the job quotas.

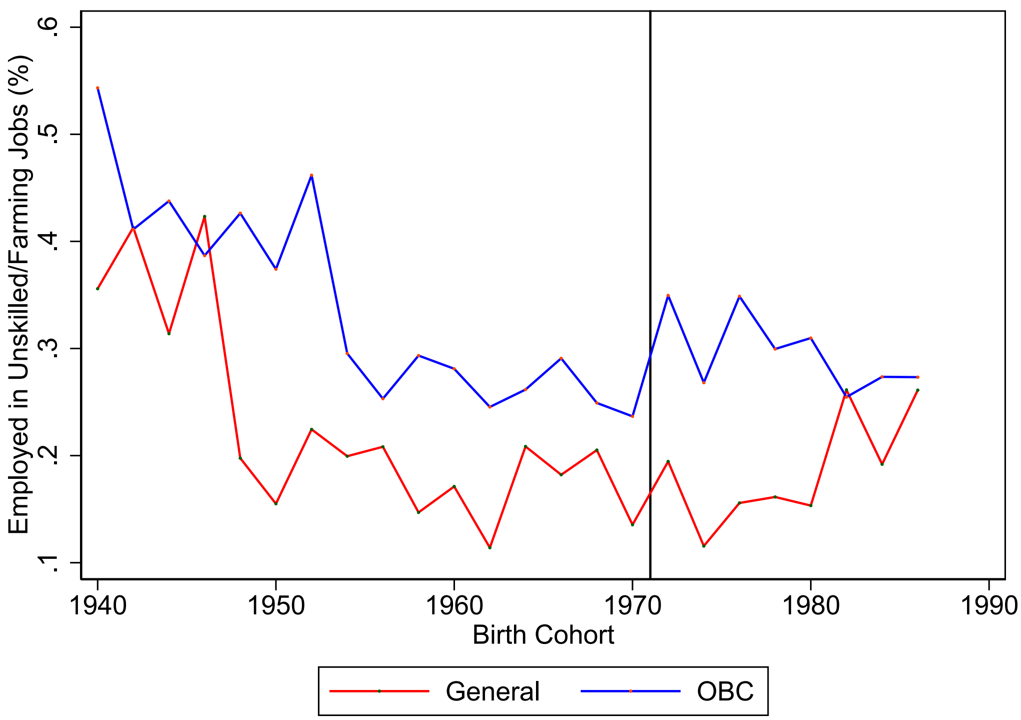

Figure 12: Percentage of Sons in Unskilled/Farming born to Professional/Skilled Fathers

This figure plots the percentage of sons employed in unskilled professions or farming born in different birth cohorts to fathers employed in professional or skilled occupations. See Section **??** for variable definition and sub-sample construction. The solid red line denotes the percentage for the general caste category, and the blue line denotes the percentage for the OBCs. The solid black line at birth year= 1971 indicates the earliest birth cohorts treated by the implementation of the job quotas.

Tables

Table 1: **Summary Statistics By Caste and Age Group**

Variable	All mean/sd	Old General mean/sd	Old OBC mean/sd	Difference b/t	Young Genneral mean/sd	Young OBC mean/sd
Age						
Son's Age	46.56	55.80	55.09	0.71***	33.39	33.21
	(13.86)	(10.88)	(10.43)	(4.56)	(4.17)	(4.14)
Father's Age	65.69	76.75	76.11	0.64	63.60	63.03
	(9.34)	(7.46)	(7.91)	(1.65)	(7.83)	(7.89)
Caste Group						
General	0.42	1.00	0.00	.	1.00	0.00
	(0.49)			.		
OBC	0.58	0.00	1.00	.	0.00	1.00
	(0.49)			.		
Education						
Son: Years of Edu	7.54	8.11	5.72	2.39***	9.90	7.91
	(4.98)	(5.03)	(4.76)	(33.27)	(4.54)	(4.61)
Father: Years of Edu	3.77	4.13	2.21	1.93***	6.14	3.98
	(4.51)	(4.61)	(3.49)	(31.68)	(5.12)	(4.44)
Son:Illiterate (%)	0.19	0.17	0.29	-0.12***	0.08	0.15
	(0.39)	(0.37)	(0.45)	(-20.08)	(0.28)	(0.35)
Father:Illiterate (%)	0.48	0.44	0.63	-0.18***	0.30	0.45
	(0.50)	(0.50)	(0.48)	(-25.56)	(0.46)	(0.50)
Son: Highest Education = Pre-Primary (%)	0.09	0.08	0.13	-0.05***	0.04	0.07
	(0.28)	(0.27)	(0.33)	(-11.00)	(0.20)	(0.25)
Father: Highest Education = Pre-Primary (%) 0.13	0.12	0.14	-0.01*	0.10	0.13	
	(0.33)	(0.33)	(0.34)	(-2.49)	(0.31)	(0.34)
Son: Highest Education = Primary (%)	0.08	0.07	0.10	-0.03***	0.04	0.08
	(0.27)	(0.26)	(0.30)	(-6.99)	(0.20)	(0.27)
Father : Highest Education = Primary (%)	0.10	0.11	0.09	0.03***	0.08	0.10
	(0.30)	(0.32)	(0.28)	(6.20)	(0.28)	(0.31)
Son: Highest Education = Middle (%)	0.25	0.23	0.23	-0.01	0.24	0.32
	(0.43)	(0.42)	(0.42)	(-1.41)	(0.43)	(0.47)
Father: Highest Education = Middle (%)	0.13	0.13	0.09	0.05***	0.19	0.16
	(0.34)	(0.34)	(0.28)	(9.90)	(0.40)	(0.36)
Son: Highest Education = Secondary (%)	0.16	0.19	0.13	0.06***	0.18	0.16
	(0.37)	(0.39)	(0.34)	(11.18)	(0.39)	(0.37)
Father: Highest Education = Secondary (%)	0.09	0.11	0.04	0.07***	0.16	0.09
	(0.28)	(0.31)	(0.19)	(17.25)	(0.37)	(0.29)
Son: Highest Education = Higher Secondary (%) 0.10	0.09	0.06	0.04***	0.16	0.11	
	(0.30)	(0.29)	(0.23)	(10.00)	(0.37)	(0.32)
Father: Highest Education = Higher Secondary (%)	0.03	0.03	0.01	0.02***	0.06	0.03
	(0.17)	(0.18)	(0.10)	(11.13)	(0.24)	(0.18)
Son: Highest Education = College (%)	0.13	0.17	0.07	0.11***	0.25	0.12
	(0.34)	(0.38)	(0.25)	(22.21)	(0.43)	(0.32)
Father: Highest Education = College (%)	0.04	0.05	0.01	0.03***	0.10	0.03
	(0.20)	(0.21)	(0.12)	(12.40)	(0.30)	(0.18)
Son: Education Rank 54.92	62.84	49.55	13.29***	61.72	49.21	
	(28.07)	(27.35)	(26.93)	(33.45)	(27.58)	(27.68)
Father: Education Rank	53.72	59.91	48.25	11.66***	61.35	49.46

Continued on next page

Table 1 – *Continued from previous page*

Variable	All mean/sd	Old General mean/sd	Old OBC mean/sd	Difference b/t	Young Genneral mean/sd	Young OBC mean/sd
	(26.99)	(27.59)	(24.16)	(30.49)	(28.17)	(26.66)
Occupation						
Son: Occupation =Professional (%) 0.10	0.12	0.06	0.05***	0.15	0.09	
	(0.30)	(0.32)	(0.24)	(12.53)	(0.36)	(0.29)
Father: Occupation =Professional (%) 0.06	0.09	0.04	0.05***	0.07	0.04	
	(0.23)	(0.28)	(0.18)	(14.26)	(0.26)	(0.20)
Son: Occupation = Skilled (%)	0.32	0.29	0.27	0.01*	0.40	0.38
	(0.47)	(0.45)	(0.44)	(2.27)	(0.49)	(0.48)
Father: Occupation = Skilled (%)	0.26	0.25	0.26	-0.01	0.27	0.27
	(0.44)	(0.43)	(0.44)	(-1.51)	(0.44)	(0.44)
Son: Occupation = Unskilled (%)	0.12	0.08	0.14	-0.06***	0.08	0.14
	(0.32)	(0.27)	(0.35)	(-14.10)	(0.28)	(0.35)
Father: Occupation = Unskilled (%)	0.17	0.15	0.22	-0.07***	0.09	0.17
	(0.37)	(0.35)	(0.41)	(-13.18)	(0.29)	(0.37)
Son: Occupation = Farmer (%)	0.36	0.36	0.40	-0.04***	0.30	0.34
	(0.48)	(0.48)	(0.49)	(-5.70)	(0.46)	(0.47)
Father: Occupation = Farmer (%)	0.44	0.47	0.46	0.01	0.39	0.40
	(0.50)	(0.50)	(0.50)	(1.96)	(0.49)	(0.49)
Son: Employed in Govt. Job (%)	0.15	0.21	0.14	0.07***	0.13	0.09
	(0.35)	(0.40)	(0.34)	(8.86)	(0.34)	(0.29)
Father: Employed in Govt. Job (%)	0.01	0.00	0.00	0.00	0.05	0.03
	(0.11)	(0.03)	(0.02)	(0.43)	(0.21)	(0.17)
Son: Salaried (%)	0.23	0.27	0.17	0.10***	0.31	0.22
	(0.42)	(0.44)	(0.38)	(15.31)	(0.46)	(0.41)
Father: Salaried (%)	0.08	0.09	0.04	0.06***	0.16	0.09
	(0.27)	(0.29)	(0.19)	(14.67)	(0.36)	(0.29)
Other Socio-Economic Variables						
Married (%)	0.90	0.93	0.94	-0.01	0.84	0.87
	(0.30)	(0.25)	(0.24)	(-1.41)	(0.37)	(0.34)
Hindu (%)	0.80	0.76	0.84	-0.08***	0.75	0.82
	(0.40)	(0.42)	(0.37)	(-13.00)	(0.43)	(0.38)
Below Poverty Line (%)	0.13	0.08	0.14	-0.06***	0.11	0.18
	(0.33)	(0.27)	(0.34)	(-12.98)	(0.32)	(0.39)
Urban Residence (%) 2011	0.40	0.44	0.35	0.09***	0.45	0.38
	(0.49)	(0.50)	(0.48)	(12.47)	(0.50)	(0.49)
N	31688	8306	10716	19022	5137	7529

Notes: The number of observations for different variables could be different than what is stated in the last row of the Table.
For example, a lot of observations have missing information regarding whether the father was/is a govt employee or salaried.
The average values or the percentages depicted in this table should thus be interpreted with some caution.

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

Table 2
Effect of Access to Job Quotas on Educational IGM when Education is measured in levels

This table presents estimation results of the following OLS regression:

$$y_{iga} = \beta_0 + \beta_1 obc_g + \beta_2 young_a + \beta_3 obc_g \times young_a + X'_{iga}\gamma + \varepsilon_{iga}$$

where y_{icgt} is the dependent variable of son i in social group g and of age a ; obc_g is a dummy that takes the value one if caste group is OBC and zero otherwise; $young_a$ is a dummy that takes the value one if age is 21 or younger in 1993 and zero otherwise; X'_{iga} are control variables (state-cohort fixed effects, state-caste fixed effects, and covariates (age of son, square of age of son, age of father, square of age of father, relationship to household head, marital status, below poverty, religion, and place of residence (urban/rural)); and ϵ_{iga} is the error term. Standard errors are clustered at the state caste group level. The sample is restricted to sons who are (i) 21+ in 2006 and (ii) are either OBC or Forward Castes. Additional sample restrictions are indicated in the table. Dependent variables are indicated in column headings and described in detail in Sections 3.4.1 and 3.4.2.

	(1) Son's Edu > Father's	(2) Years of Edu	(3) Literate	(4) Years of Edu	(5) Not High School Grad
$obc \times young$	0.0639***	0.619**	0.0505**	0.0679	-0.0672
(s.e)	(0.0143)	(0.249)	(0.0220)	(0.409)	(0.0539)
p-value	3.49e-05	0.0164	0.0257	0.869	0.218
p-wyoung		0.087	.095	.972	.615
Observations	30,229	14,760	14,760	1,648	1,648
R-squared	0.114	0.224	0.201	0.346	0.317
Sample	All	Illiterate Fathers	Illiterate Fathers	At Least High Grad Fathers	At Least High Grad Fathers

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

Table 3
Effect of Access to Quotas on Educational IGM when Education is measured in Ranks

This table presents estimation results of the following OLS regression:

$$y_{iga} = \beta_0 + \beta_1 obc_g + \beta_2 young_a + \beta_3 obc_g \times young_a + X'_{iga}\boldsymbol{\gamma} + \varepsilon_{iga}$$

where y_{icgt} is the dependent variable of son i in social group g and of age a has an above-median education rank; obc_g is a dummy that takes the value one if caste group is OBC and zero otherwise; $young_a$ is a dummy that takes the value one if age is less than 22 in 1993 and zero otherwise; X'_{iga} are control variables added incrementally (state-cohort fixed effects,state-caste fixed effects, caste time trends, and covariates (age of son, square of age of son, age of father, square of age of father, relationship to household head, marital status, amount of land owned, religion, and place of residence (urban/rural)); and ϵ_{iga} is the error term. Standard errors are clustered at the caste group level. The sample is restricted to sons who are (i) 21+ in 2006 and (ii) are either OBC or Forward Castes. Additional sample restrictions are indicated in the table. Dependent variables are indicated in column headings and described in detail in Section 3.4.3.

	(1) Edu Rank	(2) Above Median Edu	(3) Edu Rank	(4) Below Median Edu
$obc \times young$	3.503**	0.0564**	0.432	0.00238
(s.e)	(1.404)	(0.0252)	(0.476)	(0.0128)
p-value	0.0159	0.0297	0.367	0.853
p-wyoung	0.054	.069	.916	.934
Observations	15,396	15,396	14,543	14,543
R-squared	0.210	0.191	0.227	0.189
Sample	Below-Median-Edu Fathers	Below-Median-Edu 50 Fathers	Above-Median-Edu Fathers	Above-Median-Edu Fathers

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

Table 4
Effect of Access to Quotas on Occupation IGM

This table presents estimation results of the following OLS regression:

$$y_{iga} = \beta_0 + \beta_1 obc_g + \beta_2 young_a + \beta_3 obc_g \times young_a + X'_{iga}\boldsymbol{\gamma} + \varepsilon_{iga}$$

where y_{icgt} is the dependent variable of son i in social group g and of age a; obc_g is a dummy that takes the value one if caste group is OBC and zero otherwise; $young_a$ is a dummy that takes the value one if age is less than 22 in 1993 and zero otherwise; X'_{iga} are control variables added incrementally (state-cohort fixed effects,state-caste fixed effects, caste time trends, and co-variates (age of son, square of age of son, age of father, square of age of father, relationship to household head, marital status, amount of land owned, religion, and place of residence (urban/rural)); and ϵ_{iga} is the error term. Standard errors are clustered at the caste group level. The sample is restricted to sons who are (i) 21+ in 2006 and (ii) are either OBC or Forward Castes. Additional sample restrictions are indicated in the table. Dependent variables are indicated in column headings and described in detail in Section 3.4.4.

	(1) Skilled/Professional	(2) Unskilled/Farmer
$obc \times young$	-0.0540***	-0.0192
(s.e)	(0.0145)	(0.0190)
p-value	0.000462	0.318
Observations	16,671	8,153
R-squared	0.262	0.275
Sample	Unskilled/Farmer Fathers	Skilled/Professional Fathers

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

Table 5

Effect of Access to Job Quotas on Educational IGM when Education is measured in Levels: Placebo Treatment Year

This table presents estimation results of the following OLS regression:

$$y_{iga} = \beta_0 + \beta_1 obc_g + \beta_2 young_a + \beta_3 obc_g \times young_a + X'_{iga}\boldsymbol{\gamma} + \varepsilon_{iga}$$

where y_{icgt} is the dependent variable of son i in social group g and of age a ; obc_g is a dummy that takes the value one if caste group is OBC and zero otherwise; $young_a$ is a dummy that takes the value one if age between 25 and 40 in 1993 and zero if age is between 41 and 60 in 1993; X'_{iga} are control variables (state-cohort fixed effects, state-caste fixed effects, and covariates (age of son, square of age of son, age of father, square of age of father, relationship to household head, marital status, below poverty, religion, and place of residence (urban/rural)); and ϵ_{iga} is the error term. Standard errors are clustered at the state caste group level. The sample is restricted to sons who are (i) 21+ in 2006 and (ii) are either OBC or Forward Castes. Additional sample restrictions are indicated in the table. Dependent Variables are indicated in column headings and described in detail in Sections 3.4.1 and 3.4.2.

	(1) Son's Edu > Father's	(2) Years of Edu	(3) Literate	(4) Years of Edu	(5) Not High School Grad
$obc \times young$	0.0210	0.397	0.0516	1.729	-0.437
(s.e.)	(0.0372)	(0.438)	(0.0436)	(2.547)	(0.555)
p-value	0.575	0.370	0.242	0.502	0.436
Observations	15,518	8,638	8,638	423	423
R-squared	0.133	0.234	0.201	0.438	0.394
Sample	All	Illiterate Fathers	Illiterate Fathers	At Least High Grad Fathers	At Least High Grad Fathers

Robust standard errors in parentheses

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

Table 6
Effect of Access to Quotas on Educational IGM when Education is measured in Ranks:Placebo Treatment Year

This table presents estimation results of the following OLS regression:

$$y_{iga} = \beta_0 + \beta_1 obc_g + \beta_2 young_a + \beta_3 obc_g \times young_a + X'_{iga}\boldsymbol{\gamma} + \varepsilon_{iga}$$

where y_{icgt} is the dependent variable of son i in social group g and of age a has an above-median education rank; obc_g is a dummy that takes the value one if caste group is OBC and zero otherwise; $young_a$ is a dummy that takes the value one if age between 25 and 40 in 1993 and zero if age is between 41 and 60 in 1993; X'_{iga} are control variables added incrementally (state-cohort fixed effects,state-caste fixed effects, caste time trends, and covariates (age of son, square of age of son, age of father, square of age of father, relationship to household head, marital status, amount of land owned, religion, and place of residence (urban/rural)); and ϵ_{iga} is the error term. Standard errors are clustered at the caste group level. The sample is restricted to sons who are (i) 21+ in 2006 and (ii) are either OBC or Forward Castes. Additional sample restrictions are indicated in the table. Dependent variables are indicated in column headings and described in detail in Section 3.4.3.

	(1) Edu Rank	(2) Above Median Edu	(3) Edu Rank	(4) Below Median Edu
$obc \times young$	3.604	0.0988**	1.361	-0.0346
(s.e.)	(2.517)	(0.0401)	(1.940)	(0.0327)
p-value	0.158	0.0173	0.486	0.294
Observations	8,638	8,638	6,711	6,711
R-squared	0.228	0.205	0.274	0.207
Sample	Below-Median-Edu Fathers	Below-Median-Edu 50 Fathers	Above-Median-Edu Fathers	Above-Median-Edu Fathers

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

Table 7
Effect of Access to Quotas on Occupation IGM:Placebo Treatment Year

This table presents estimation results of the following OLS regression:

$$y_{iga} = \beta_0 + \beta_1 obc_g + \beta_2 young_a + \beta_3 obc_g \times young_a + X'_{iga}\boldsymbol{\gamma} + \varepsilon_{iga}$$

where y_{icgt} is the dependent variable of son i in social group g and of age a; obc_g is a dummy that takes the value one if caste group is OBC and zero otherwise; $young_a$ is a dummy that takes the value one if age between 25 and 40 in 1993 and zero if age is between 41 and 60 in 1993; X'_{iga} are control variables added incrementally (state-cohort fixed effects,state-caste fixed effects, caste time trends, and co-variates (age of son, square of age of son, age of father, square of age of father, relationship to household head, marital status, amount of land owned, religion, and place of residence (urban/rural)); and ϵ_{iga} is the error term. Standard errors are clustered at the caste group level. The sample is restricted to sons who are (i) 21+ in 2006 and (ii) are either OBC or Forward Castes. Additional sample restrictions are indicated in the table. Dependent variables are indicated in column headings and described in detail in Section 3.4.4.

	(1) Skilled/Professional	(2) Unskilled/Farmer
$obc \times young$	0.192	-0.0838
(s.e.)	(0.227)	(0.0599)
p-value	0.402	0.167
Observations	8,963	3,701
R-squared	0.321	0.306
Sample	Unskilled/Farmer Fathers	Skilled/Professional Fathers

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

Appendix A Data Appendix

This section details the process by which I fill in the two-digit occupation codes of sons and fathers. I discuss how I construct occupation codes for resident sons, followed by non-resident sons, and finally, I describe how I construct the occupation code for fathers.

A.1 Occupation Code of Resident Sons

The IHDS contains information on the sources of household income from joint production activities (income from cultivation, animal husbandry, and business[34]), each resident household member's participation in each of these activities, and their level of participation (the number of hours worked in a day and the number of days worked in a year[35]). It also asks every resident household member if they engaged in any agricultural wage labour, non-agricultural wage labour, or salaried work and if yes, for how many hours they worked in a year. As many individuals in my data set worked in more than one work activity in the year of the survey, I assign their main work activity (cultivation, working in business 1, working in business 2, working in business 3, salaried/agricultural wage labourer/non-agricultural wage labourer, or animal husbandry) as the activity in which they work the highest hours per year, provided that number is greater than 240 hours/year.[36]

The IHDS also provides two-digit occupation codes for salaried work/agricultural wage labour/non-agricultural wage labour (the variable called "WS4" contained in the individual file) and each of the possibly three businesses of the household (variables "NF1B", "NF21B", and "NF41B", respectively, contained in the household file). Thus, if the main work activity is classified as working in business 1 (2) (3), then I fill in the occupation code for this resident by the occupation code reported in the variable "NF1B" ("NF21B") ("NF41B"). If the main work activity is classified as work/agricultural wage labour/non-agricultural wage labour, I disaggregate workers using the variable "WS4". Lastly, I impute the occupation code as 61 if the main work activity of that resident is cultivation or animal husbandry.

I fill in some of the missing information of the occupation codes [37] from the variable "RO7" that reports the "primary occupation status" [38] of each resident. If the residents report their primary occupation as cultivation or allied agriculture, I classify their (originally missing) occupation code as 61

[34]The IHDS reports information on a maximum of 3 businesses owned by a household.

[35]Information on the number of hours/days worked in animal husbandry is not available; instead, the IHDS asks if the member takes care of the animals "usually", "sometimes", or "never".

[36]I assign animal husbandry as the main work activity if the maximum hours worked per year in all other work activities is less than 240 and the household member "usually" takes care of animals.

[37]This information could be missing, for example, if the main work activity is missing in cases where the individual did not report working hours and/ or days, is unemployed, retired, does household work, or is a student.

[38]The primary occupation status of each resident takes one of the following values: Cultivation, Allied Agriculture, Agricultural Wage Labour, Non-Agricultural Wage labour, Artisan/Independent Work, Small Business, Organized Business, Professional, Salaried, Retired, Housework, Student, Unemployed, Too young/unfit, and Others

(cultivators) and if they report their primary occupation as agricultural labour, I classify their (originally missing) occupation code as 63.

I do not use this variable as the first step in deducing the main work activity as there are many cases in which the number of hours reported in a different work activity is higher than the one stated as the "primary occupation". Azam (2013) and Kundu et al. (2021) also do not rely on the "primary occupation status" of the residents, but instead use the indicators provided in the IHDS for different work activities (cultivation, animal care, business, salaried, agricultural wage labour, non-agricultural wage labour), each of which is reported as "none" (i.e. 0 hours worked in that activity), missing hours, less than 240 hours/year, part time (greater than 240 hours/year, but not full time), or full time (greater than 250 days/year and more than 2000 hours/year) [39]. There are a few cases in which a resident reports having two full time work activities, and there are many cases in which an individual reports having no full time activity but multiple part-time activities. Thus, Azam (2013) and Kundu et al. (2021) end up using a main-work-activity-classification procedure that relies on a subjective order in which they consider these indicators (for example, they first try to categorize residents as salaried/wage labourers if they spend more than 240 hours/year in salaried/wage work, ignoring the possibility that they could indeed be classified as full time in some other work activity. They then try to classify residents as working in a business, and so on. The advantage of using the highest number of hours worked to deduce the main work activity is that I do not have to use an arbitrary order to go through these indicators and I can also infer the particular business in which the resident spends the maximum of his working time, so as to fill in the occupation code of that specific business, if needed.

A.2 Occupation Code of Non-Resident Sons

Filling in the two-digit occupation codes of non-resident sons is a more straight-forward process as the variable "NR11" in the non-resident file of IHDS-II directly contains their occupation code.

A.3 Occupation Code of Fathers

I fill in the two-digit occupation codes of fathers in the following way. More than 50% of sons in my sample are also household heads or husbands of household heads. The household file of IHDS-II contains a variable called "ID18A", which reports the two-digit occupation code of the father of a male household head or the husband's father of a female household head, even if they are dead or are not considered a part of the same household. Thus, I can directly obtain information on father's occupation choice for male household heads and husbands of female household heads from this variable. For sons

[39] Since the number of hours worked in animal care is not reported, the indicator for animal work takes the following values: none, less than 240 hours (i.e. work "sometimes" in animal care), and part time (i.e. work "usually" in animal care)

who are not household heads and have resident fathers, I fill in the father's occupation code in a manner similar to the one I use for resident sons i.e. I infer their main work activity using information on hours worked by the father in different work activities, fill in the occupation codes using the "WS4", "NF1B", "NF21B", or "NF41B" variables corresponding to the father's row in the individual file of the IHDS, wherever appropriate, and replace some missing information in the father's occupation code using information reported in the "RO7" variable corresponding to the father's row. For sons who are not household heads and have non-resident fathers, I use information contained in the "NR11" variable in the non-resident file that corresponds to the individual who is the father of the son under consideration.

Table A1: **NCO-1968 Two-Digit Occupation Codes**

Code	Occupation
	PROFESSIONAL, TECHNICAL, AND RELATED WORKERS
00	Physical Scientists
01	Physical Science Technicians
02	Architects, Engineers, Technologists, and Surveyors
03	Engineering Technicians
04	Aircraft and Ships Officers
05	Life Scientists
06	Life Science Technicians
07	Physicians and Surgeons (Allopathic Dental and Veterinary Surgeons)
08	Nursing and Other Medical and Health Technicians
09	Scientific, Medical, and Technical Persons, Other
10	Mathematicians, Statisticians, and Related Workers
11	Economists and Related Workers
12	Accountants, Auditors, and Related Workers
13	Social Scientists and Related Workers
14	Jurists
15	Teachers
16	Poets, Authors, Journalists, and Related Workers
17	Sculptors, Painters, Photographers, and Related Creative Artists
18	Composers and Performing Artists
19	Professional Workers, n.e.c.
	ADMINISTRATIVE, EXECUTIVE, AND MANAGERIAL WORKERS
20	Elected and Legislative Officials
21	Administrative and Executive Officials, Government and Local Bodies
22	Working Proprietors, Directors and Managers, Wholesale and Retail Trade
23	Directors and Managers, Financial Institutions
24	Working Proprietors, Directors and Managers Mining, Construction, Manufacturing, and Related Concerns
25	Working Proprietors, Directors, Managers and Related Executives, Transport, Storage, and Communication
26	Working Proprietors, Directors and Managers, Other Service
29	Administrative, Executive and Managerial Workers, n.e.c.
	CLERICAL AND RELATED WORKERS
30	Clerical and Other Supervisors
31	Village Officials
32	Stenographers, Typists, and Card and Tape Punching Operators
33	Book-keepers, Cashiers, and Related Workers
34	Computing Machine Operators
35	Clerical and Related Workers, n.e.c.
36	Transport and Communication Supervisors
37	Transport Conductors and Guards
38	Mail Distributors and Related Workers
39	Telephone and Telegraph Operators
	SALES WORKERS
40	Merchants and Shopkeepers, Wholesale and Retail Trade
41	Manufacturers, Agents
42	Technical Salesmen and Commercial Travellers
43	Salesmen, Shop Assistants, and Related Workers
44	Insurance, Real Estate, Securities, and Business Service Salesmen and Auctioneers
45	Money Lenders and Pawn Brokers
49	Sales Workers, n.e.c.

Continued on next page

Table A1 – *Continued from previous page*

Code	Occupation
	SERVICE WORKERS
50	Hotel and Restaurant Keepers
51	House Keepers, Matron, and Stewards (Domestic and Institutional)
52	Cooks, Waiters, Bartenders, and Related Worker (Domestic and Institutional)
53	Maids and Other House Keeping Service Workers n.e.c.
54	Building Caretakers, Sweepers, Cleaners, and Related Workers
55	Launderers, Dry-cleaners, and Pressers
56	Hair Dressers, Barbers, Beauticians, and Related Workers
57	Protective Service Workers
59	Service Workers, n.e.c.
	FARMERS, FISHERMEN, HUNTERS, LOGGERS, AND RELATED WORKERS
60	Farm Plantation, Dairy and Other Managers and Supervisors
61	Cultivators
62	Farmers other than Cultivators
63	Agricultural Labourers
64	Plantation Labourers and Related Workers
65	Other Farm Workers
66	Forestry Workers
67	Hunters and Related Workers
68	Fishermen and Related Workers
	PRODUCTION AND RELATED WORKERS, TRANSPORT EQUIPMENT OPERATORS AND LABOURERS
71	Miners, Quarrymen, Well Drillers, and Related Workers
72	Metal Processors
73	Wood Preparation Workers and Paper Makers
74	Chemical Processors and Related Workers
75	Spinners, Weavers, Knitters, Dyers, and Related Workers
76	Tanners, Fellmongers, and Pelt Dressers
77	Food and Beverage Processors
78	Tobacco Preparers and Tobacco Product Makers
79	Tailors, Dress Makers, Sewers, Upholsterers, and Related Workers
80	Shoe Makers and Leather Goods Makers
81	Carpenters, Cabinet and Related Wood Workers
82	Stone Cutters and Carvers
83	Blacksmiths, Tool Makers, and Machine Tool Operators
84	Machinery Fitters, Machine Assemblers, and Precision Instrument Makers (except Electrical)
85	Electrical Fitters and Related Electrical and Electronic Workers
86	Broadcasting Station and Sound Equipment Operators and Cinema Projectionists
87	Plumbers, Welders, Sheet Metal, and Structural Metal Preparers and Erectors
88	Jewellery and Precious Metal Workers and Metal Engravers (Except Printing)
89	Glass Formers, Potters, and Related Workers
90	Rubber and Plastic Product Makers
91	Paper and Paper Board Products Makers
92	Printing and Related Workers
93	Painters
94	Production and Related Workers, n.e.c.
95	Bricklayers and Other Constructions Workers
96	Stationery Engines and Related Equipment Operators, Oilers and Greasers
97	Material Handling and Related Equipment Operators, Loaders and Unloaders
98	Transport Equipment Operators
99	Labourers, n.e.c.

Continued on next page

Table A1 – *Continued from previous page*

Code	Occupation
	WORKERS NOT CLASSIFIED BY OCCUPATIONS
X0	New Workers Seeking Employment
X1	Workers Reporting Occupations Unidentifiable
X9	Workers Not Reporting Any Occupation
AA	Housewife/Household work
BB	Student/Too Young to Work
CC	Retired/Too Old to Work
DD	Disabled/Unfit to Work
EE	Out of Labour Force n.e.c

Table A2
Occupation Categories in Azam (2013)

	Category	NCO-1968 Occupation Code
1	White collar	0-29
2	Skilled	30-45, 49, 71-98
3	Unskilled	50-59, 63-68, 99
4	Farmers	60-62

Appendix B Pre-Trends using Event Study

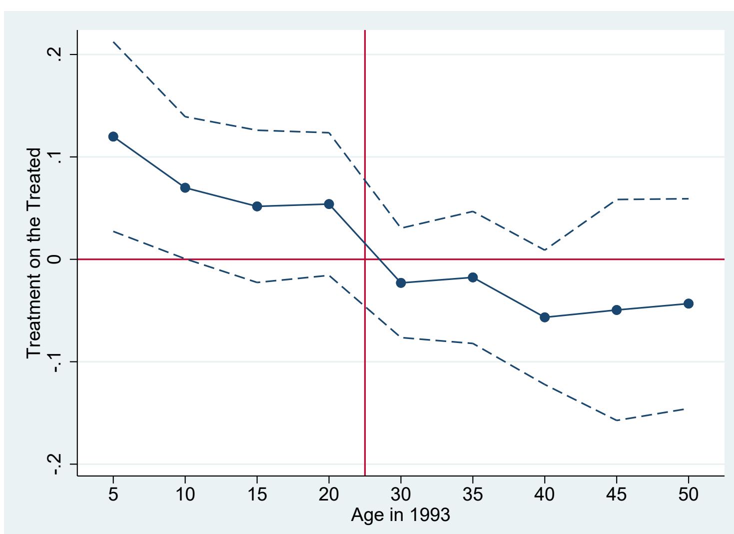

Figure B1: Event Study:Son's Education is Greater than Father's Education

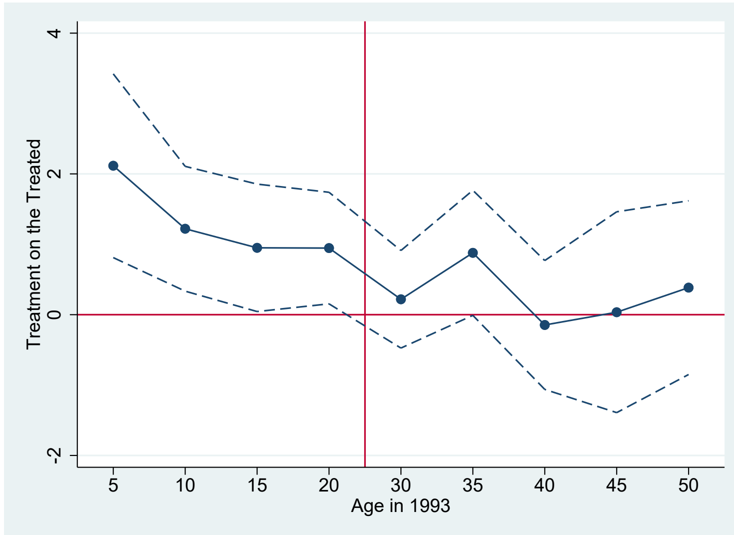

Figure B2: Event Study: Years of Education of Sons Born to Illiterate Fathers

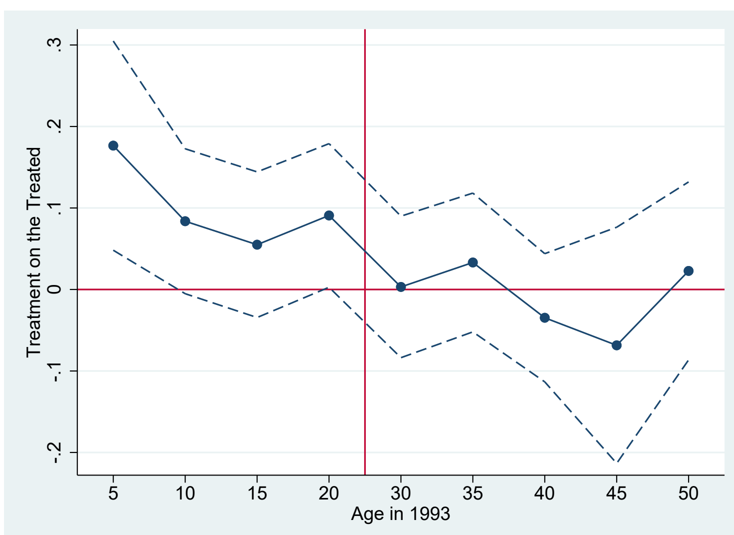

Figure B3: Event Study: Probability of Son Born to Illiterate Father being Literate

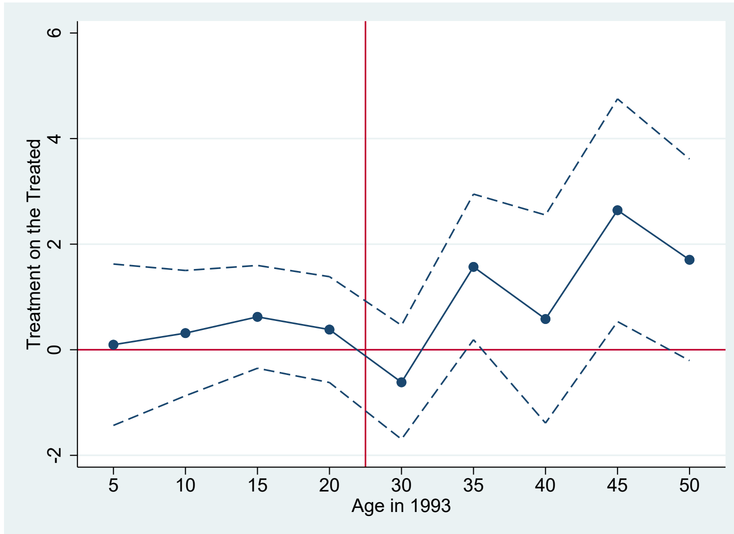

Figure B4: Event Study: Years of Education of Sons Born to At-Least-High-School-Graduate Fathers

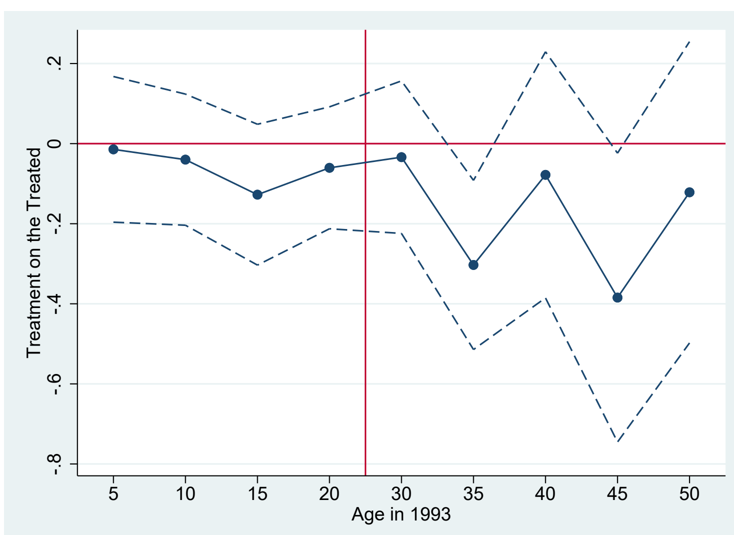

Figure B5: Event Study: Probability of Sons Born to At-Least-High-School-Graduate Fathers not being High School Graduates

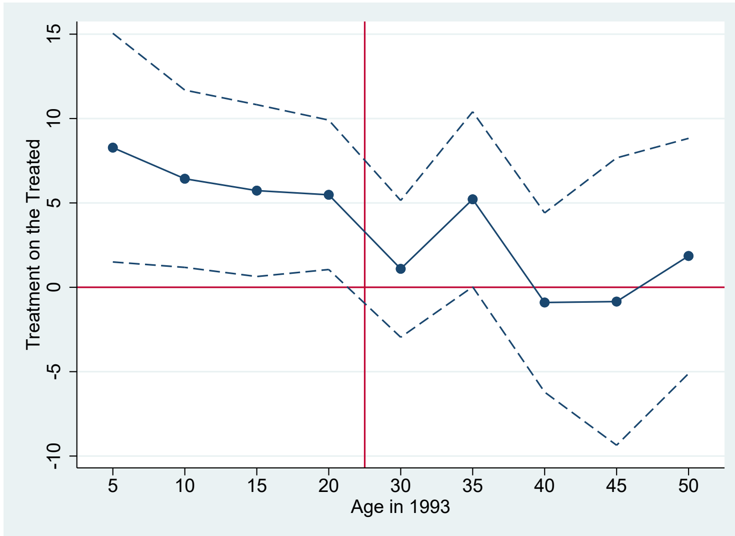

Figure B6: Event Study:Education Rank of Sons Born to Below-Median-Education-Fathers

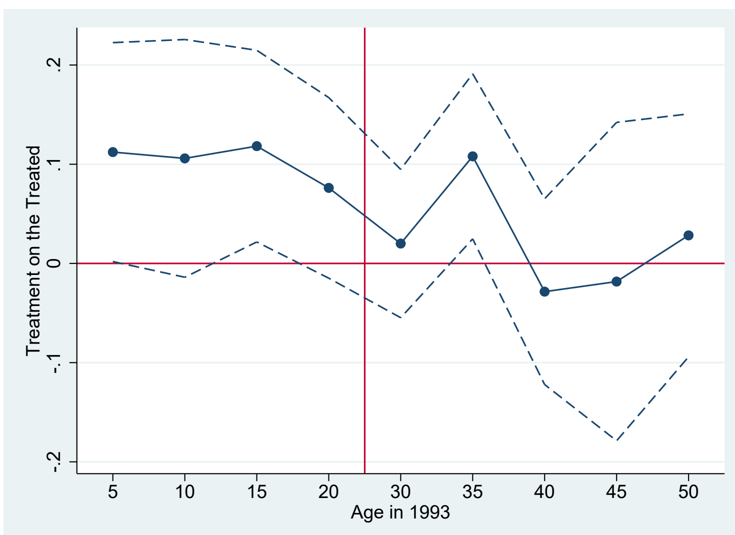

Figure B7: Event Study:Probability of Sons Born to Below-Median-Education-Fathers being in the top half of their education distribution

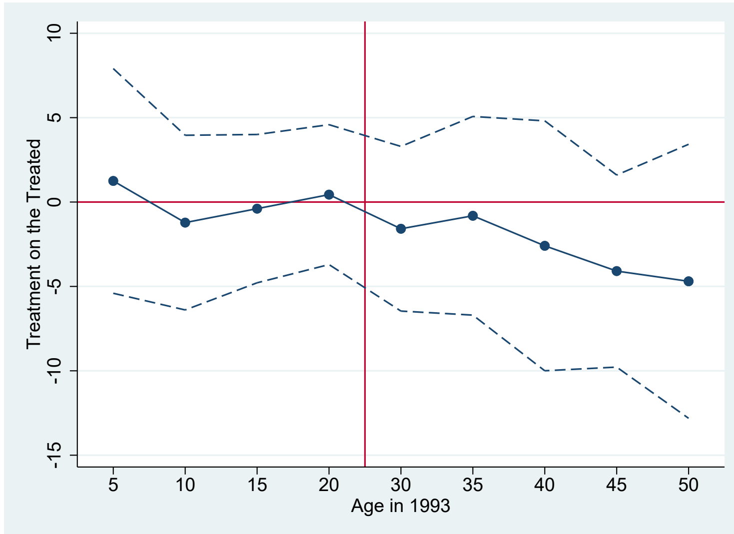

Figure B8: Event Study:Education Rank of Sons Born to Above-Median-Education-Fathers

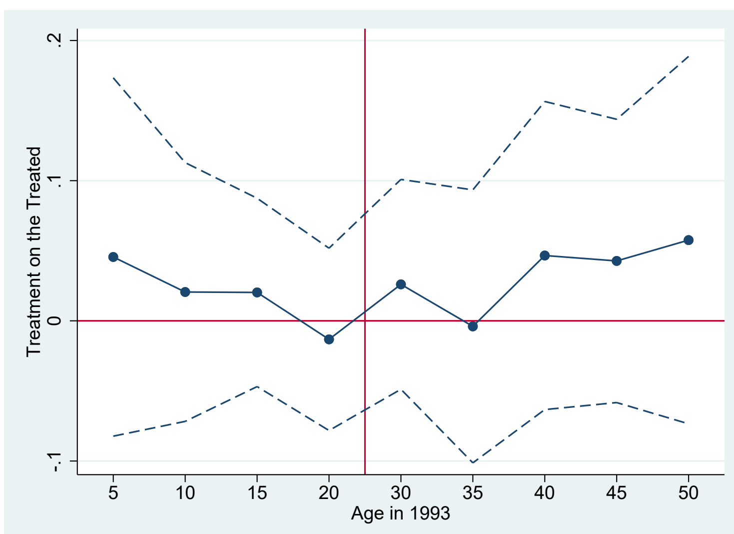

Figure B9: Event Study:Probability of Sons Born to Above-Median-Education-Fathers being in the bottom half of their education distribution

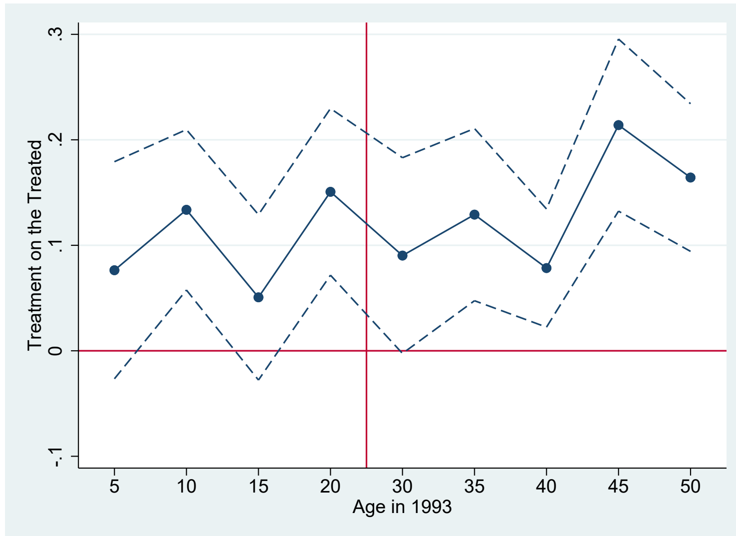

Figure B10: Event Study:Probability of Sons Born to Unskilled/Farmer Fathers being employed in Professional/Skilled Occupations

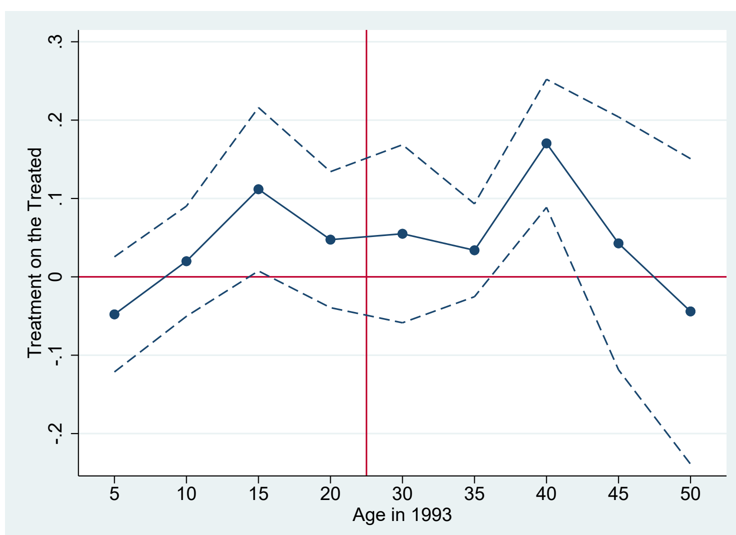

Figure B11: Event Study:Probability of Sons Born to Professional/Skilled Fathers being employed in Unskilled/Farmer Occupations

Appendix C Measures of Income IGM

Most of the empirical literature on measurement of IGM is set in developed countries and uses permanent income as a proxy for socio-economic opportunities. In this section, I discuss the evolution of these measures. These measures can also be divided into two classes that capture different normative concepts: *relative mobility* and *absolute mobility* (Chetty et al., 2014).

C.1 Relative Mobility

The first generation of statistical IGM measures (Solon, 1999) use a single linear parameter to empirically estimate relative mobility by using the following basic regression framework:

$$y_s = \beta_0 + \beta_1 y_f + \varepsilon \tag{2}$$

where y is a measure of the log of permanent income, subscript s refers to the child, and subscript f refers to the parent. The regression coefficient β_1 is called the Inter-Generational Regression Coefficient (IGRC) in general, and the Inter-Generational Elasticity (IGE) specifically when y is the log of income. The parameter $(1-\beta_1)$ is a measure of relative mobility as it measures the difference in the outcomes of children having different parental backgrounds. Thus,a higher β_1 implies a higher degree of persistence (or a lower degree of mobility) because children's outcomes are more dependent on parents' socio-economic status. A related measure of relative mobility is the Inter-Generational Correlation (IGC), which is obtained by dividing the IGRC by the relative standard deviation of the child's and the parents' income distribution i.e.

$$IGC = IGRC(\frac{\sigma_0}{\sigma_1}) \tag{3}$$

Thus, the IGC is a normalized measure of relative mobility that takes into account the cross-sectional dispersion in income in both the generations (Williams, 2015). Even though the IGC and the IGRC are closely linked, the results are often different depending on the measure used (Hertz et al., 2008).

The second generation of IGM measures builds on the discussion in Solon (1999) and was popularized by papers such as Chetty et al. (2014). This measure asks how the child's position in the income distribution depends on their parents' position, by estimating a regression of the following form:

$$R_s = \beta_0 + \beta_1 R_f + \varepsilon \tag{4}$$

where R_s denotes the percentile rank of the child in the income distribution of children and R_f denotes the parents' percentile rank in the parent's income distribution. β_1 is termed the rank-rank slope coefficient, with a higher β_1 representing a lower degree of relative IGM. The rank-rank slope is scaleinvariant like the IGC.

IGRC and IGC are easy to calculate and offer a relatively straightforward interpretation. However, these measures have two limitations: (i) they assume that the underlying Conditional Expectation Function (CEF) of the child's otcome given parental outcome is linear and (ii) they are not suitable for subgroup analysis. Chetty et al. (2014) demonstrates the non-linearity of the income CEF in USA. Chetty et al. (2014) also finds that the rank-rank income CEF is very robust to alternate specifications. This means that the change in opportunities will be different for children at the top and bottom of the parental distribution. Many studies (Azam and Bhatt (2015); Hnatkovska et al. (2013), among others) also conduct a subgroup analysis by estimating Equation 2 separately for each subgroup. However, this effectively means that children are compared with reference to better-off children within their subgroup rather than with reference to the entire sample, which could result in a subgroup having a lower IGRC but worse outcomes at each point in the parental distribution i.e. have a lower upward mobility (Asher et al., 2018).

C.2 Absolute Mobility

Absolute mobility looks at the childrens' outcomes for parents of a given income in absolute terms. There are various methods of measuring absolute mobility. The first measure is obtained by estimating Equation (2) (or (4)) and plugging in a given value of parental outcome (or rank). For example, Chetty et al. (2014) looks at the mean rank of children whose parents are at the 25^{th} percentile of their income distribution i.e. $E(y_1|y_0 = 0.25) = \beta_0 + 0.25\beta_1$. A value greater (smaller) (equal to) than 0.25 implies upward (downward) (no) mobility.

A second measure is to look at the probability that the child's socio-economic outcome is different or higher than their parent's (e.g., as in Chetty et al. (2017)). Ultimately, it is useful to present results from a variety of mobility measures as they each serve to answer related but distinct questions.

Because education is not a continuous variable like income, practically, it makes mores sense to condition parental education rank below a certain rank than at a given rank. In the absolute education and occupation mobility literature, another measure is to use transitional probabilities or construct a mobility matrix. The aim of the transitional probabilities or the mobility matrix approach is to characterize the joint distribution of the child's education (or occupation) and the parents' education (or occupation) using a small set of measures to present mobility estimates in a concise and parsimonious manner.[40] Transitional probabilities analyse the probability of the next generation moving to a certain socio-economic category when the preceding generation belongs to a given socio-economic category and are used in papers like Derenoncourt (2019); Hnatkovska et al. (2013). The categories can be con-

[40] Thus, even though an individual can choose to acquire education ranging from 0 to more than 15 years (or choose from about a 100 different occupations), it makes practical sense to aggregate these choices to education (or occupation) levels.

structed based on either levels or ranks.