Assessing the Effectiveness of Affirmative Action in India Using Fertility Rates

Arnav Parikh
April 28, 2021

Advisor: Professor Jessica Reyes

Submitted to the Department of Economics at Amherst College
in partial fulfilment of the requirements for the degree of
Bachelor of Arts with Honors

Abstract

This thesis analyzes the impact of the Central Educational Institutions Amendment Bill (CEIAB), an affirmative action policy that mandated a 27% Other Backward Caste (OBC) quota in all public universities in India. The primary outcomes of interest are a woman's age at marriage and age at first birth. It is hypothesized that these fertility patterns were deferred because access to higher education incentivized OBCs to cross the high school threshold and encouraged them to pursue further studies, which resulted in eminent career aspirations and awareness of more modern perspectives. In the developing world, fertility patterns also proxy for broader socio-economic outcomes and invariably center the analysis around women.

The CEIAB was implemented in 2008 and specifically aimed to increase OBC participation in higher education. Additionally, the CEIAB's impact varied at the regional level since different regions in India had varying levels of OBC quotas prior to the policy. Therefore, I exploit variation across cohorts, castes and regions and find that affirmative action had the effect of differentially increasing the age at marriage and first birth for the targeted group. OBCs in the eastern region, where no affirmative action prior to the policy existed, benefitted the most. The effect of the policy in the southern region, where quotas for OBCs exceeded the centrally mandated 27% before the CEIAB, was negligible. However, the results also suggest that OBCs in the lowest quintile of the wealth distribution were unaffected by the CEIAB. These findings indicate the important role, even with its limited scope, state-sponsored affirmative action plays in helping the historically marginalized backward castes close the gap between the socio-economic outcomes of higher castes in Indian society.

JEL: H4 H75 I24 128 I38

Keywords: Affirmative Action, Fertility, Inequality, Caste, Quotas, Development

Acknowledgements

Firstly, I want to thank Professor Jessica Reyes for her guidance, close engagement and support. This thesis would not have been possible without her encouragement, attention to detail and invaluable feedback. She has been pivotal to this transformative experience.

Thank you to Professor Caroline Theoharides for inspiring me to write this thesis. Her classes and her work have had a profound impact on my college career and my interest in development economics.

A special mention for Professor Nick Horton, who has been paramount to my academic experience at Amherst.

Thank you to Ricardo Diaz for his help throughout the thesis process.

Many thanks to my peers – Sirig and Crystal for motivating me to work and Riddhi, Rafael, Seamus, Charlie and Gahena for their constructive feedback.

Thank you to the Amherst Squash community for playing a major role in my Amherst experience.

Lastly, but most importantly, thank you to Mom, Dad and Ishita. Their sacrifices made it possible for me to have this incredible opportunity to study at Amherst.

TABLE OF CONTENTS

I. Introduction

The rationale behind affirmative action (AA) is that because certain groups in society face systematic discrimination, the gradual process of development and economic growth is not sufficient in closing the gap between marginalized and dominant groups. Hence, AA broadly consists of anti-discrimination measures intended to provide access to preferred positions in society for sections of the population that would be otherwise excluded or underrepresented.

The question of the effectiveness of AA, especially in India, has been long debated. A major reason for the continued debate is the lack of empirical evidence. Unlike traditional redistributive policies, like land reforms or conditional cash transfers, AA does not immediately lead to the redistribution of assets or wealth. Instead, it simply alters the composition of elite positions in society (Deshpande 2013). Therefore, the impact of AA policies on the socio-economic outcomes of its beneficiaries are realized years or decades after their implementation. This time lag has posed a challenge for economists to causally interpret the impact of AA policies and has contributed to the lack of empirical studies.

Motivated by the need for quantifying the long-run impacts of AA, this paper uses the implementation of the Central Educational Institutions Amendment Bill (CEIAB), a policy change that federally mandated a 27% quota in higher education for the Other Backward Caste (OBC) category, as an exogenous change to answer the question: did increased access to higher education through AA close the gap between the socio-economic outcomes of OBCs and the General Population? The answer to this question is important as it could help shape future policy decisions in a country where AA has become a highly controversial issue.

Although India has implemented reservations for Scheduled Castes (SCs), Scheduled Tribes (STs) and OBCs in political representation, public-sector employment and higher education for decades, partisan opinions on its effectiveness remain entrenched in the national discourse. For example, after the extension of quotas in public sector employment for OBCs in 1992, thousands of agitated students threw the country into turmoil demanding the reversal of this apparently retrograde move (Deshpande 2013). Supported by the mainstream media, these detractors raised the questions of why a traditionally divisive element of society was being perpetuated in a modern India and why merit was not the primary determining factor in the labor market. On the other hand, proponents of AA strongly believed that the consequences of past discrimination affect the skill acquisition process today (Fryer and Loury 2005). Lower castes continue to live in areas and communities with inferior public goods and do not have bequeathed social networks to help them in a competitive job market. Thus, supporters of AA believe that equal opportunity in a meritocracy is not enough to ensure caste equality.

Other, more nuanced, doubts around AA in India also persist. Critics of AA argue that the federally mandated 50% quota in the public sector affects the productivity of the entire nation, which not only has a negative impact on the Gross Domestic Product but also drives away crucial foreign investment. Furthermore, in higher education, there is a widespread belief that reservations place minorities in academic environments for which they are ill-prepared for. This 'mismatch hypothesis' theorizes that the lack of preparation leaves disadvantaged groups worse off due to time wasted in college or through a 'discouragement effect' (Bertrand, Hanna and Mullainathan 2010). Additionally, disagreements around who should be eligible for reservations endure. This argument is mainly based on the fact that quotas are filled up by financially affluent minorities and that AA does little to help the 'poorest among the poor' (Chauhan 2008).

Recent economic studies have attempted to settle some of these arguments using empirical analysis. Analyzing data from the Indian Railways, the world's largest employers subject to AA, *Weisskopf & Deshpande (2014)* find no evidence to support the claim that an increasing proportion of AA negatively impacts productivity or productivity growth. In fact, they find some suggestive evidence of AA having a favorable impact on overall productivity. Plausible reasons for this finding may include higher motivation levels for marginalized groups and a greater diversity of perspectives in the workplace. *Bertrand et al. (2010)* find that despite poor exam scores, lower caste entrants in higher education obtain a positive return to admission, contrary to the aforementioned 'mismatch hypothesis.' However, even though AA improved the wages of its beneficiaries, the returns to enrollment were greater for higher castes compared to lower castes. This corroborates the theories that limited social networks affect the labor market outcomes of lower castes and that a meritocracy is not enough to ensure caste equality.

This paper adds to the existing literature by using fertility patterns, a unique outcome variable, to assess the long-term effects of AA in India. Fertility patterns are the chosen outcome variables because, according to *the Demographic Transition Theory*, they are highly correlated with education, wealth, health and labor market outcomes, making them an effective proxy for broader socio-economic outcomes (Amonker and Brinker 2007). The use of fertility patterns also centers the analysis on women, contrary to the existing literature that has predominantly focused on men. Furthermore, as previous studies have assessed the impact of AA on enrolment and wages, fertility rates help determine if the effects of AA have spilled over to the overall quality of life. However, it must be noted that the emphasis on fertility patterns excludes men from the analysis, which is an inevitable shortcoming of this paper.

To assess the effectiveness of the CEIAB, I use a cross section of women born between 1980 and 1995 and surveyed in the National Family Health Survey (NFHS) 3 & 4, collected in 2006 and 2015 respectively. Since the CEIAB was implemented in 2008 and specifically aimed to increase OBC participation in higher education, I exploit variation across caste and cohorts. OBC women born between 1990 and 1995, who were 13-18 years old in 2008, are considered treated. Older OBC women are either partially treated or untreated. Women from the General Population are the ideal comparison group since they were exposed to all the cultural changes in India that affected fertility patterns during the period of interest, but never benefitted from AA. The third dimension of variation I exploit is across regions since the expected impact of the CEIAB varied according to the regional level of quotas extended to OBCs prior to 2008. The strength of the triple difference approach is that it not only controls for the effects of urbanization, modernization and liberalization on fertility patterns, but it also accounts for the impact of other contemporaneous national-level policies. However, despite the fact that no policies that directly benefitted OBCs in the eastern region[1] were found, the effects of region-specific anti-poverty measures is a threat to identification that remains unaccounted for. The inability to account for age-reporting issues, seen by the clustering of birth years around round years like 1975 and 1980 and several inconsistent ages at marriage, is another weakness of this study.

My findings suggest that the CEIAB had the effect of differentially increasing the age at marriage and age at first birth of OBC women. Treated OBCs in the eastern region, where no reservations for OBCs existed prior to 2008, experienced an 8 percentage point decrease in the probability of marriage at age 19, which completely offset the pre-policy gap in the share of women married at age 19 between OBCs and the General Population. To further substantiate the causal interpretation of the CEIAB, no evidence of a differential effect was found for OBCs in the southern region where quotas exceeded the centrally mandated 27% before 2008. These results

[1] The eastern region is where the CEIAB is expected to have the greatest impact; explained in detail in the Conceptual Framework.

provide suggestive empirical evidence for the effectiveness of state-sponsored AA in closing the gap between the socio-economic outcomes of lower and upper castes in India.

II. Caste Based Affirmative Action in India

To understand the rationale behind AA policies in India, a short dive into the history of the caste system is required. The traditional caste system divided Hindu society into mutually exclusive groups; the 2,500-year old *varna* system consisted of *Brahmins* (priests, teachers), *Kshatriyas (*warriors), *Vaishyas* (tradesmen, moneylenders), *Shudras* (laborers) and *Ati-Shudras* (untouchables, who did the most menial jobs). Since these hereditary groups were largely based on specific occupations, and because inter-caste marriages and social mobility were rare, the caste system contributed to persisting inequality in India.

In an attempt to adopt an inclusive model of a pluralist, participant and federal political system, the founding fathers of independent India guaranteed AA to the historically marginalized castes (Chhetri 2012). AA was constitutionally guaranteed with the Constitution of India itself stating, "The State shall promote with special care the educational and economic interests of the weaker sections of society . . . and shall protect them from social injustice and exploitation."

The government identifies four broad caste categories, specifically for the implementation of social sector policies. Scheduled Castes (SC) and Scheduled Tribes (ST)[2] belong to the historically marginalized castes of Indian society and have benefitted from AA since the 1950s. The Other Backward Castes (OBC), the third caste category that benefits from reservations, consists of castes and communities that have been identified as economically and socially backward. The General Population category remains outside the purview of AA policies.

---

[2] Scheduled Tribes consist of India's indigenous and tribal people while Scheduled Castes are castes that were predominantly categorized as untouchables in the aforementioned *varna* system

Since this paper focuses on an AA policy that only affected the OBCs, a closer look at this caste category is necessary. The OBC category was created due to the findings of the Mandal Commission. The Mandal Commission, established in 1978 with a mandate to identify the socio-economically backward castes in Indian society, identified 3,743 backward *jatis*[3] which made up 52 percent of the Indian population. The implementation of the Mandal Commission in 1992 extended reservations in public sector employment to these newly identified backward classes. Based on a 1963 Supreme Court verdict that limited quotas to 50 percent, and because 22.5 percent of seats were already guaranteed for SCs and STs, an OBC reservation of 27 percent in public sector employment was rolled out throughout the country (Jayal 2015). Initially, higher education remained beyond the scope of AA policies for OBCs because it did not achieve the necessary legislative support. However, in 2008, the CEIAB was passed in the Indian parliament which federally mandated the same 27% quota for OBCs in public institutes of higher education[4]. This specific extension of reservations for OBCs took the aggregate quotas for SCs, STs and OBCs to 49.5% in all public universities, to go along with the already implemented 49.5% quota for these three caste categories in public sector employment and political representation.

III. Previous Studies on Affirmative Action in India

The literature on AA in India is limited but growing. In addition to the already mentioned studies by *Bertrand et al. (2010)* and *Weisskopf & Deshpande (2014), Desai & Kulkarni (2008)* investigate if educational inequalities between SCs and STs and upper caste Hindus have decreased due to AA. They use successive rounds of the National Sample Survey (NSS) between 1983 and

[3] Over time, as the economy grew more complex, and as new castes emerged through fissure and fusion, the *varna* system was transformed into the *jati* system which is essentially a system of regional caste groupings.
[4] In India, higher education has been historically managed by the public sector, and due to concerns regarding quality and difference in cost, even after a recent spur in growth of private institutions, public institutions still remain the most preferable option for students (Sen 2019).

2000 to calculate transition probabilities across six levels of education. Their findings suggest a declining gap between the beneficiaries of AA[5] in the odds of completing primary school but they find little improvement at the college level. They also find no evidence for AA policies disproportionately benefitting high-income lower caste students, contrary to public opinion in India. Also studying the impact of AA on educational attainment, *Cassan (2019)* uses plausibly exogenous variation of the harmonization of *jati* lists to be included within the SC caste category in each state to show that AA increased educational attainment for lower castes. *Cassan (2019)* finds substantial increases in literacy and secondary schooling for beneficiaries. However, the benefits of AA were not distributed evenly across genders; low caste females remained unaffected.

Studying the 'mismatch hypothesis' *Bagde et al. (2016)* analyze data from 214 engineering colleges from one state in India. Using student exam scores from high-school completion, entrance exams and final year exams, *Bagde et al. (2016)* find AA increases participation of SCs and STs, particularly at high quality institutions. Furthermore, they find no evidence of adverse impacts of AA on lower castes. Contrastingly, *Robles and Krishna (2012)* use a dataset on the 2008 graduating class from an elite engineering institution and find evidence of a 'mismatch'. SC and ST students started off with lower grades, never caught up with the rest of the students, and eventually graduated with lower scores. However, this study does conclude that targeted SC and ST students are poorer than displaced students and that, irrespective of lower grades, admission to a prestigious college via AA will undoubtedly alter their life trajectories.

Since SCs and STs have had access to AA since the 1950s, most of the AA literature has focused on these two caste categories. To my knowledge, only *Basant & Sen (2019)* and *Khanna (2020)* have studied the recent effects of AA on the OBCs. *Khanna (2020)* leverages variation

[5] There was no evidence for a declining gap for Muslims, a minority group that does not benefit from affirmative action.

across cohorts and eligibility and finds that the 27% OBC quota in public-sector employment incentivized OBCs to remain in school longer. His results suggest that as a result of AA in employment, the average OBC student increased schooling by 0.8 years. On the other hand, *Basant & Sen (2019)* look at the effects of the CEIAB on OBC participation rates. They find that post 2008, the CEIAB improved OBC participation significantly in states where no AA for OBCs before the CEIAB existed and in states where OBC participation was low to begin with. A further supply-side analysis to investigate if the new demand for higher education increased the number of higher educational institutions was also conducted. They find evidence of supply-side constraints and argue that the expansion of the higher education system could be a better policy option to address low OBC participation.

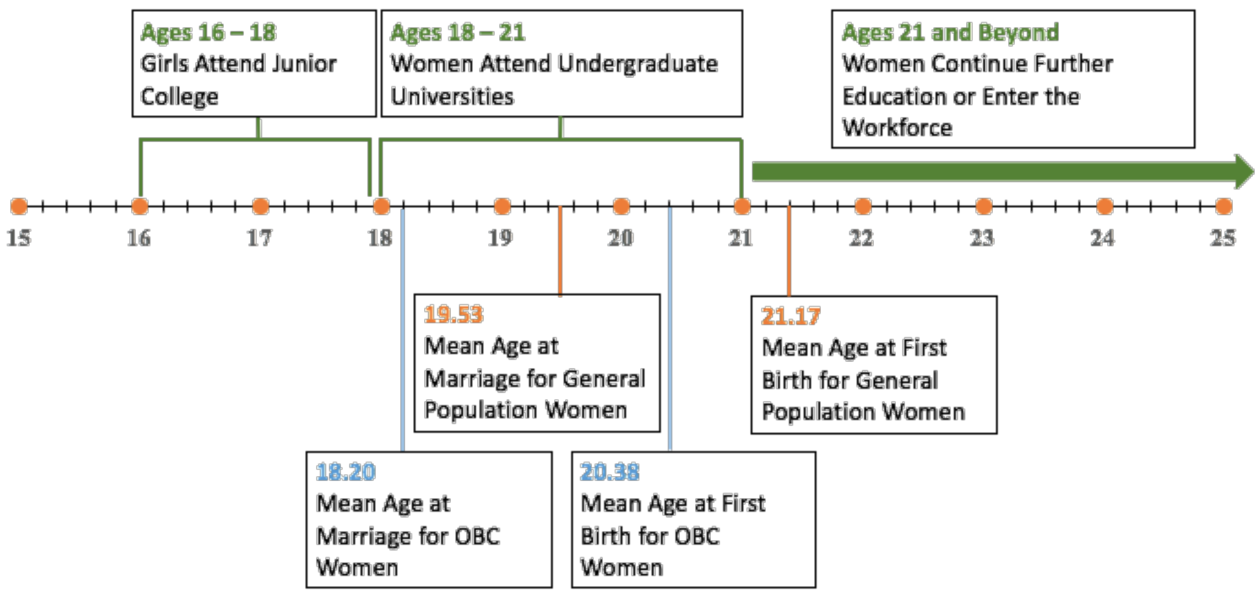
This paper expands on the methodology used by *Basant and Sen (2019)* to investigate if the CEIAB had longer term impacts on the outcomes of OBCs. It also builds on the modest literature on OBC affirmative action. In the context of the wider literature of AA in India, this paper contributes by attempting to casually interpret the effects of AA on the broader socio-economic outcomes of the targeted group. To my knowledge, it is the first paper that uses fertility patterns as a primary outcome in the AA literature.

IV. Conceptual Framework

A. The Effect of Affirmative Action on the OBC Life Trajectory

Stratification in Indian society is reflected in inequalities in educational attainment across caste, religion and ethnic boundaries (Desai and Kulkarni 2009). In India, OBC participation in higher education remains disproportionately lower than the share of OBCs in the total population (Basant and Sen 2019). Thus, the CEIAB attempted to increase OBC participation in higher

Figure 1: A Typical OBC Woman's Trajectory in Adolescence



education via a 27% quota in public universities. As a result of the policy, OBCs were more likely to cross the high school threshold and apply to two-year junior college programs. The policy also encouraged OBCs to continue their education by applying to four-year undergraduate universities. In the long-run, the policy hoped that access to higher education would lead to better labor market outcomes, increased income, improved access to healthcare and a better overall quality of life, which would work towards reducing the gap in socio-economic outcomes between OBCs and the General Population. The top half Figure 1 depicts the hypothesized trajectory of an OBC woman as a result of AA and the bottom half shows the mean age at marriage and first birth for OBCs and the General Population reported in the NFHS–4.

B. Differences in Fertility Rates by Caste

Prior research has shown that fertility rates in India tend to be higher among lower caste women (Gandotra 1998). The NFHS–3 reports that STs had the highest total fertility rate (3.12), followed by SCs (2.92), OBCs (2.75) and the General Population (2.35). Higher fertility rates among lower castes is indicative of the inverse relationship between socio-economic outcomes and fertility rates in India. This is consistent with the *Demographic Transition Theory* which suggests that improved standards of living, educational attainment, access to technology and better health outcomes promote a decline in the level of fertility (Amonker & Brinker 2007).

C. Affirmative Action and Fertility Patterns

The inverse relationship between fertility rates and broader socio-economic outcomes makes fertility a great proxy for investigating if the CEIAB's main aim of reducing the socio-economic differences between OBCs and the General Population was achieved.

In an ideal setting, the eventual fertility of the women that benefitted from the CEIAB would be used to study the effectiveness of the policy. However, as only a short period of time had elapsed from when the bill was passed to when the women were surveyed, the eventual fertility of eligible women was not observed. Nevertheless, because it is expected that the CEIAB would encourage OBC women to enroll in higher education, have eminent career aspirations and make them more aware of more modern perspectives, it is hypothesized that eligible OBC women would defer their age at marriage and age at first birth[6]. These fertility patterns are undoubtedly different than total fertility rates. Yet, age at marriage and first birth can be used to assess the medium-term impact of the CEIAB since they are a determinant of fertility rates and are also heavily influenced by several demographic, social and economic phenomena (Bloom and Reddy 1986).

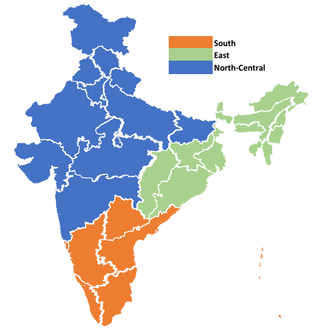
D. Policy Variation Across Regions

To assess effectiveness of the CEIAB, it is imperative to consider state-wise levels of AA before 2008. Following *Basant & Sen (2019)*, I divide the states of India into three broad regions: *South, North-Central, East*[7]*,* based on their respective levels of AA prior to the central mandate[8].

[6] Read 'Age Patterns of Women of Women at Marriage, Cohabitation and First Birth in India' by David Bloom and P.H. Reddy to understand the factors affecting age at marriage and first birth in India.
[7] *Basant and Sen (2019)* have an East and North-East region. However, because both these regions had no AA prior to 2008, I combine them into just one Eastern region.
[8] A table for a detailed list of states included within each region can be found in the Appendix.

Figure 2: Regional Variation of the Policy

Region	Level of AA Prior to 2008	Hypothesized Impact
South	Greater than 27%	No Impact
North-Central	Between 10% and 20%	Small Impact
East	Less than 5%	Immediate Impact



States in the southern region started implementing reservations in higher education for OBCs as early as the 1970s (Bayly 1999). By 2008, these states already had robust AA policies, in some cases greater than the centrally mandated 27%. Hence, I hypothesize that the southern region should have no impact as a result of the policy. In the north-central region, reservations began in the late 1980s and 1990s and were implemented in a fragmented manner (Basant and Sen 2019). By the early 2000s, there was about a 10% to 20% quota for OBCs in this region[9]. So the 27% quota should have had a small impact in the north-central region. Lastly, negligible, if any, quotas existed prior to the national mandate in the eastern region. Therefore, the strongest and an immediate impact of the policy is expected in the eastern region.

V. Data

The data used in this paper comes from the NFHS-3 & NFHS-4, collected in 2005-6 and 2015-16 respectively. Both rounds of the survey are cross-sectional individual level data, representative at the national and state level, and publically available through the Demographic and Health Surveys (DHS) program. Data is collected for women aged between 15 and 49 from selected households and include individual characteristics including caste, age, educational attainment, health outcomes, employment, fertility patterns and so forth.

[9] Extensive research was conducted, however no information of the exact level of AA policies in each state of the north-central region was found in the public domain. Therefore, there is a possibility of variation in the levels of AA before 2008 for states within the north-central region.

The NFHS-4 provides a great opportunity to compare younger treated women to older untreated women. With the CEIAB implemented in 2008, the fact that the NFHS-4 was collected in 2015 allows a cohort of treated women to be surveyed a few years after treatment. This delay is crucial as it allows the medium-term fertility patterns of treated women to be observed.

However, because of the cross sectional nature of the NFHS-4, answers given by older women about their age at marriage and first birth could be inaccurate; several people in India do not know their own age and have given inconsistent and implausible answers to questions in the survey[10]. Furthermore, answers for health outcomes, contraceptive usage and employment

Table 1: Descriptive Statistics

	Entire Sample	ST & SCs	OBCs	General Population
Panel A: Individual Level Means (NFHS - 4)				
Total Sample	699,686	251,946	273,700	141,428
Percentage of Sample	100.00%	36.01%	39.12%	20.21%
Literacy Rate	71.70%	66.09%	70.53%	83.37%
Years of Education	6.73	5.86	6.59	8.61
	(5.19)	(4.98)	(5.18)	(5.10)
Age at Marriage	18.70	18.02	18.20	19.53
	(4.37)	(4.47)	(4.21)	(4.29)
Age at First Birth	20.56	20.12	20.38	21.17
	(3.84)	(3.93)	(3.62)	(3.91)
Number of Children	1.88	1.98	1.90	1.71
	(1.82)	(1.91)	(1.82)	(1.64)
Ideal Children	2.27	2.45	2.22	2.09
	(1.01)	(1.13)	(0.93)	(0.85)
Panel B: Individual Level Means (NFHS - 3)				
Total Sample	124,385	20,397	39,035	42,676
Percentage of Sample	100.00%	16.40%	31.38%	34.31%
Literacy Rate	67.81%	57.20%	63.35%	79.71%
Years of Education	6.12	4.64	5.46	7.89
	(5.27)	(4.78)	(5.12)	(5.35)
Age at Marriage	17.47	16.36	16.66	18.55
	(4.35)	(4.21)	(4.21)	(4.31)
Age at First Birth	19.82	18.95	19.41	20.62
	(3.86)	(3.75)	(3.58)	(3.97)
Number of Children	2.06	2.27	2.17	1.83
	(2.04)	(2.21)	(2.08)	(1.83)
Ideal Children	2.30	2.27	2.28	2.08
	(0.93)	(1.08)	(0.86)	(0.80)

Notes: The SCs and STs are grouped into one category. For the Ideal Children outcome, women that reported more than 10 ideal children were excluded. Standard deviations are in paranthesis. The literacy was calculated as the percentage of women that have reported attending school. This is admittedly not the best measure for literacy. Literacy rates do not have standard deviations because of the way they were calculated. Panel A uses the NFHS - 4 and Panel B uses the NFHS - 3. All birth years were included.

[10] For example, 10,832 out of 529,872 married women in the NFHS-4 reported inconsistent ages at marriage or that they did not know when they were married.

outcomes could also vary by age, making it harder to compare younger and older women[11]. Due to these reasons, data from NFHS-3 serve as an excellent control group in the analysis since it is useful for comparing women of the same ages that were observed almost 10 years apart[12].

Table 1 provides descriptive statistics for the data from the NFHS-3 & 4. Each panel displays the individual level means for the entire sample as well as separately for the SCs & STs, OBCs and the General Population. Across both datasets, it is evident that the General Population has the highest literacy rate and years of education. The General Population also tends to get married later and have fewer children. In the NFHS-4, on average, OBCs attend school for 6.59 years and get married at 18.20 years old. On average, OBCs attend school for 2.02 less years compared to the General Population, get married earlier by 1.33 years and have 0.19 more children. The SCs & STs, that have been combined for the purposes of this table, have the worst outcomes with the least years of education, earliest start to fertility patterns and the most children. The hypothesized pattern of fewer years of education leading to the earlier onset of fertility patterns and a greater total number of children is clearly observed in Table 2.

VI. Empirical Strategy

A. Difference in Difference Estimation

A woman's exposure to the policy is jointly determined by her caste category and birth year (Duflo 2001).

A majority of the students in India apply to junior college when they are 15-16 years old. Following two years at junior college, students apply to undergraduate universities when they are 17 -18 years old. If a student elects to continue to study, applications to graduate level schools are

[11] Variation in responses by age is not a problem for the primary outcomes of interest because age at marriage and first birth are both age-invariant. That is, a married mother observed at age 25 would report the same age at marriage and first birth compared to if she was observed at age 40.

[12] No woman observed in the NFHS-3 was exposed to the policy since the survey was conducted in 2006.

generally sent out in the early 20s. Therefore, women born after 1985, who were between the ages 13 and 22 in 2008, are deemed eligible for the policy. Among this eligible cohort, women who were 13-18 years old in 2008, and born between 1990 and 1995, are likely to directly benefit from the policy in the application process. This age-group of women could have also potentially changed their educational path to take advantage of the policy. Hence, they are considered the fully treated cohort. The ability to respond successfully and change their education trajectory is constrained for OBC women born between 1986 and 1989, who were already between the ages of 19 and 22 in 2008 (Shrestha 2015). Therefore, this age-group of women is the partially treated cohort. On the other hand, women born in or before 1985, who were 23 years old or older in 2008 are ineligible to benefit from the policy and are considered untreated. It is possible that grade repetition and delayed school entry could lead to a few of the untreated women being exposed to the policy (Duflo 2001). However, in the Indian context, given the universality of marriage and the familial expectations for women in their 20s, the possibility is negligible (Bloom and Reddy 1986).

The second dimension of variation is the caste category. Since the CEIAB was only aimed at increasing OBC participation in higher education, SCs, STs and the General Population were unaffected. Since the General Population category was exposed to urbanization, liberalization and other unobservable cultural changes in India during the period of interest, but was never eligible for AA, the General Population category is the preferred control group[13]. As STs and SCs have had access to AA since the 1950s, they are excluded from the main analysis.

Given this background, the effect on human capital is identified via a difference in difference estimation, comparing fertility patterns between the eligible and ineligible cohorts,

[13] OBCs have slightly worse socio economic outcomes compared to the General Population but are similar to them in terms of social hierarchy in Indian society. Ashwini Deshpande writes that the stigmatization on account of their untouchability confers a particular disadvantage to SCs and STs that goes beyond the economic and social marginalization faced by the OBCs.

Experiment A

Birthyear	Caste Category		Age In		
	General Population	OBC	2006	2008	2015
1980			26	28	**35**
1981			25	27	**34**
1982	Untreated	Untreated	24	26	**33**
1983			23	25	**32**
1984			22	24	**31**
1985			21	23	**30**
1986			20	22	**29**
1987	Untreated	Partially Treated	19	21	**28**
1988			18	20	**27**
1989			17	19	**26**
1990			16	18	**25**
1991			15	17	**24**
1992	Untreated	Treated	14	16	**23**
1993			13	15	**22**
1994			12	14	**21**
1995			11	13	**20**

Experiment B

Birthyear	Caste Category		Age In		
	General Population	OBC	2006	2008	2015
1980			**26**	28	35
1981			**25**	27	34
1982	Untreated	Untreated	**24**	26	33
1983			**23**	25	32
1984			**22**	24	31
1985			**21**	23	30
1990			16	18	**25**
1991			15	17	**24**
1992	Untreated	Treated	14	16	**23**
1993			13	15	**22**
1994			12	14	**21**
1995			11	13	**20**

Notes: Bolded ages signify the age at which women in the corresponding birth year were observed
Ages shaded in grey were not included in the eventual sample used for the regressions. This was done so that women of the same ages when observed, between 21 - 25, were included from in both cohorts

within the OBC and General Population caste categories (Shrestha 2015). The solitary difference of the age at marriage and first birth between the eligible and ineligible OBC cohorts could be correlated with several age-varying unobserved variables, especially since the time period of the study corresponds with significant growth and liberalization in the Indian economy. Therefore, also subtracting the cohort difference in fertility patterns for the General Population, from the difference between eligible and ineligible OBC cohorts, would net out all age-varying characteristics as well as age-invariant caste characteristics that could directly affect fertility patterns (Shrestha 2015).

This empirical strategy of comparing the treated cohorts to partially treated and untreated cohorts that are all observed in 2015 is henceforth referred to as Experiment A. The treated cohort in Experiment B essentially remains the same. However, in Experiment B, the untreated cohort was observed in 2006. This allows me to compare treated women and untreated women that were the same ages, between 21 and 25, when observed.

B. Baseline Regression

The difference in difference strategy can be generalized to a regression framework:

$$FP_{icst} = \beta_1 OBC_{ic} + \beta_2 Treat_{it} + \beta_3 OBC_{ic}\, X\, Treat_{it} + \alpha_s + C_i + \epsilon_{icst} \quad (1)$$

where FP_{icst} is a woman's fertility pattern (either age at marriage or age at first birth) for woman i of caste c living in state s and born in year t. OBC_{ic} is a dummy variable indicating if woman i belongs to the OBC caste category and $Treat_{it}$ is a dummy variable indicating if a woman is born in the previously defined treated cohort. The coefficient of interest is β_3, the coefficient on the interaction term between $OBC_{ic}\, X\, Treat_{it}$, representing the differential fertility outcomes of treated OBC women. A similar regression framework, adding $Partial\, Treat_{it}$ and $OBC_{ic}\, X\, Partial\, Treat_{it}$ can be written for Experiment A.

In Equation 1, C_i refers to a vector of control variables. Fertility patterns in India are diverse, differing by religion, caste, geographic location, urban-rural residence, educational background and income level, and changing over time (Bloom and Reddy 1986). Therefore, I control for the aforementioned factors[14]. It is essential to control for religion because of the prevalence of child marriages amongst Hindus and the tendency of Muslim women to have larger families[15]. To account for the effect of wealth, I include a categorical variable Wealth Index with five levels (poorest, poorer, middle, richer and richest). This categorical variable was constructed using Principal Components Analysis after a score was assigned to each woman according to the number and kinds of consumer goods she owned. A dummy variable with levels for 'rural' and 'urban' is also included. Finally, a linear birth year trend is included because rapid growth, development and urbanization in India led to increasing ages at marriage and first birth during the late 20th Century. The first fully treated birth year, 1990, is selected as the base birth year[16] so that the trend can be interpreted as years after (or before) 1990.

State fixed effects are also included to control for unobserved influences on fertility patterns that vary across states but not with time, as represented by α_s. State fixed effects are essential since state borders in India are drawn along linguistic and cultural lines, with each state having its own unique language and traditions. As outcomes may have been correlated within states, clustering standard errors at the state level ensures that the precision of treatment effects are not overestimated (Abeberese, Kumler and Linden 2014).

[14] A woman's years of education is not controlled for because, unlike the variables that I do control for, education is an outcome of the life trajectory of a woman.
[15] A representative sample of women from nine major religions in India (Hinduism, Islam, Christianity, Sikhism, Buddhism, Jainism, Judaism, Zoroastrianism and atheism) has been collected in both datasets.
[16] 1990 is entered as value a zero in the linear trend.

The identification assumption is that without the implementation of the CEIAB, the changes in the fertility patterns of OBCs and the General Population over time would not have been systematically different. A perennial threat to this parallel trend assumption is omitted time-varying caste-specific effects. In the Indian context, the fact that India made substantial progress in reducing child marriage, which was more prevalent amongst OBCs during the time period of interest, could lead to the convergence of the age at marriage between the two castes. A differential increase in the fertility patterns of OBC women violates the parallel trends assumption. Furthermore, since OBCs are an economically backward caste, they are more likely to live in poorer states. Because India is a federal republic and liberalization, urbanization and modernization in each state is affecting fertility patterns differentially over time, age varying state-specific factors could further confound the estimates of Equation 1. Therefore, I also include both caste and state-specific linear trends in Equation 1.

Additionally, it is also plausible that the effect of the program could be increasing with time, as more colleges adjust their admission policies to allow for AA and more OBC women become aware of the benefits of higher education. Younger treated women have a greater chance of planning ahead, changing their educational trajectories, and incorporating a collegiate education into their long-term plans. Therefore, the coefficient of interest could be biased downward. Hence, I also incorporate the triple interaction $OBC_{ic}\ X\ Birth\ Year_{it}\ X\ Treat_{it}$ which relaxes the assumption that treated OBC women and untreated OBC women are trending the same over time and allows the fertility patterns of treated OBC women to have a different slope compared to untreated OBC women.

C. *Accounting for Sample Selection Bias Using Logistic Regression*

Sample selection bias is a major issue in Equation 1. Sample selection bias arises because of the way the variables of age at first marriage and age at first birth are coded. If a woman is married, her age at marriage is recorded. However, if she is not married, her response is recorded as a missing value. Since all missing values are dropped in the regression framework, Equation 1 effectively excludes unmarried women from the analysis. Given the universality of marriage in India and the fact that the mean age of marriage is about 19 years old for women, the exclusion of unmarried women is negligible for the untreated cohort in Experiment A. However, because I hypothesize a deferment of the age at marriage for college-educated treated women, omitting unmarried treated women between the ages of 20 and 25 causes sample selection bias. Treatment, in effect, leads to exclusion from the sample in both experiments.

To deal with the issue of sample selection bias, I estimate the probability of a woman being married at or before a certain age. The ages 15 to 24 are chosen because, in India, most women get married during this 10-year period[17]. As a woman's first birth almost always takes place after marriage in India, choosing ages 15 to 24 is appropriate for age at first birth as well.

$$Pr(Married\ Before\ Age\ a) = \beta_1 OBC_{ic}^{a} + \beta_2 Treat_{it}^{a} + \beta_3 OBC_{ic}^{a}\ X\ Treat_{it}^{a} + \alpha_s + C_i + \epsilon_{icst} \quad (2)$$

The right hand side of Equation 2 is exactly the same as Equation 1. However, the outcome variable is now the probability of being married at or before age a. This empirical strategy, using logistic regression, attempts to solve the issue of sample selection bias as both married and unmarried women are needed to estimate the probability of being married by age a. Once again, $Partial\ Treat_{it}$ and $OBC_{ic}\ X\ Partial\ Treat_{it}$ are included for Experiment A. Since the same

[17] Only 6% of women in the NFHS-4 were married after age 24.

threats to identification for Equation 1 are applicable for Equation 2, the additional control terms are also added. Equation 2 is also used to estimate the probability of giving birth by age a.

For the sample that I have chosen[18], the marital status for all women at ages 15 through 20 is known since the youngest woman in both experiments, when observed, is 20-years old[19]. However, while estimating the probability of getting married at or before age 21, unmarried 20-year olds are omitted from the analysis because their marital status at age 21 is unknown and unobservable[20]. Similarly, while estimating the probability of being married at or before age 22, unmarried 20 and 21-year old women are excluded, and so on. Hence, all married women and only unmarried women that have reached the target age a by the time the survey was conducted can be included in each iterative specification for ages 21 through 24. This strategy of iterating the same specification for different ages is also useful because it helps to understand the ages at which the differential effect is the greatest, which can help to isolate the effects of the policy from the impacts of the reduction of child marriage.

D. *Regional Variation and a Triple Difference Estimation*

As explained above, regional differences in the levels of AA prior to the CEIAB is another source of variation that can be exploited. This lends to a triple difference strategy by comparing the double difference in the eastern region with the same double difference in the southern region[21].

[18] Refer to the tables that visually represent Experiment A and Experiment B on page 15.
[19] For clarity, take the example of the youngest woman who was 20 years old when she was surveyed. If she got married when she was 18, her marital status at age 15 through 17 is unmarried and her marital status at ages 18 though 20 is married. On the other hand, if she was unmarried when observed, then she was unmarried at all ages between ages 15 and 20. Similarly, the marital status for all women at the ages 15 through 20 is known.
[20] Let us take the example of an unmarried 20-year old. As she is observed at age 20, we do not know if she will remain unmarried or get unmarried at age 21 and at all subsequent ages. Therefore, the marital status of the unmarried 20-year old is unobservable for ages 21 through 24.
[21] The reason I leave the north-central region out of the triple difference strategy will be explained subsequently.

The triple difference framework is estimated by:

$$FP_{icrst} = \beta_1 OBC_{ic} + \beta_2 East_{irs} + \beta_3 Treat_{it} + \beta_4 OBC_{ic} \: X \: East_{irs} + \beta_5 OBC_{ic} \: X \: Treat_{it}$$
$$\beta_6 East_{irs} \: X \: Treat_{it} + \beta_7 OBC_{ic} \: X \: East_{irs} \: X \: Treat_{it} + \alpha_s + C_i + \epsilon_{icrst} \quad (3)$$

where all the variables are the same as before. The new variable $East_{irs}$ is an indicator for residing in the eastern region. The main parameter of interest is β_7, the triple difference estimate. β_1 through β_6 are the estimates of the double interaction terms and linear terms respectively. The triple difference is also extended to the logistic regression framework.

VII. Results

A. Difference in Difference Estimation

Table 2 shows the mean age at marriage and mean age at first birth for treated and untreated cohorts from the OBC and General Population caste categories. Panel A is comprised of women observed in 2015 and compares women born between 1990 and 1995 (treated, aged 13-18 in 2008) to women born between 1980 and 1985 (untreated, aged 23-28 in 2008). Panel B uses essentially the same treatment group but a control group of women born between 1981 and 1985 who were observed in 2006.

The positive difference in difference in Columns 3 and 6 of Panel B indicates that age at marriage and first birth increased more for OBCs compared to the General Population. While comparing treated and untreated women, OBC women increased their age at marriage over time by 0.73 more years and increased their age at first birth over time by 0.52 more years compared to the same difference for women from the General Population. These differences in differences are significantly different from 0 at the 1% level.

The average age at marriage and first birth was lower for OBC women compared to women from the General Population within all cohorts. Furthermore, in Panel B, average age at marriage

Table 2: Means of Age at Marriage and First Birth by Cohort and Caste

	Age At First Marriage			Age At First Birth		
	Caste Category			Caste Category		
	General Population (1)	OBC (2)	Difference (3)	General Population (4)	OBC (5)	Difference (6)
Panel A: Women Observed in NFHS-4						
Untreated: Born 1980-1985, Ages 23-28 in 2008, observed in 2015	19.87	18.27	-1.60***	21.50	20.52	-0.98***
	(0.03)	(0.02)	(0.04)	(0.03)	(0.02)	(0.03)
Treated: Born 1990-1995 Ages 13-18 in 2008, observed in 2015	18.97	18.32	-0.65***	19.81	19.65	-0.16***
	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)
Difference	-0.90***	0.05***	0.95***	-1.70***	-0.87***	0.82***
	(0.04)	(0.03)	(0.05)	(0.03)	(0.03)	(0.04)
Panel B: Women Observed in NFHS-3 & NFHS-4						
Untreated: Born 1981-1985, Ages 21-25 in 2006, observed in 2006	18.13	16.70	-1.42***	19.44	18.74	-0.70***
	(0.04)	(0.05)	(0.06)	(0.04)	(0.04)	(0.05)
Treated: Born 1990-1994, Ages 21-25 in 2015, observed in 2015	19.13	18.44	-0.69***	19.96	19.78	-0.18***
	(0.02)	(0.02)	(0.03)	(0.02)	(0.01)	(0.02)
Difference	1.00***	1.73***	0.73***	0.52***	1.04***	0.52***
	(0.05)	(0.05)	(0.07)	(0.04)	(0.04)	(0.06)

Notes: The sample is made out of women that were married when observed. Panel A consists of women observed in the NFHS - 4. The treated cohort of Panel B is observed in the NFHS - 4 while the untreated cohort in Panel B is observed in the NFHS - 3. Standard errors are clustered at the state level and are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

and first birth increased over time, irrespective of caste. These observations corroborate the initial hypotheses that fertility patterns begin earlier for women from lower castes and that ages at marriage and first birth have been increasing for all women, in a modernizing India.

In this context of increasing ages at marriage, it is noteworthy that the differential increase in the age at marriage for OBC women accounts for $(0.73 / 1.73) * 100 = 42.20\%$ of the overall increase between untreated and treated OBCs. Similarly, the differential increase constitutes half of the overall increase in age at first birth between untreated and treated OBCs. Therefore, the difference in difference estimation in Panel B can be interpreted as the causal effect of the policy, under the assumption that in the absence of the policy, the changes in the age at marriage and age at first birth over time would not have been systematically different for the OBC and General Population caste categories.

Although there is a positive difference in difference estimation, the difference between treated and untreated women in Columns 1, 4 and 5 of Panel A implies that average age at marriage and first birth decreased over time. For example, the average age at first birth for treated OBC women fell by 0.87 years from the average age at first birth for the untreated OBC women. The contrast in the trends over time between Panel A and Panel B is most likely driven by sample selection bias. As explained above, treated women may have deferred their fertility patterns and, as a result, have unknown and unobserved ages at marriage and first birth in 2015. Their exclusion downwardly biases these averages since women who likely never took advantage of the policy and married / gave birth early are observed. Although the issue of sample selection exists in Panel B, its effect is smaller as Experiment B compares women of the same ages when observed.

Even though the results of Table 2 are imprecisely estimated, the differences in differences provide some suggestive evidence of the positive differential impact of the policy on the fertility patterns of OBC women. The following sections will extend this basic difference in difference identification strategy to lead to more persuading results.

B. Baseline Regression

Tables 3A & 3B show the regression results of Equation 1. The main coefficient of interest in all the specifications is *OBC X Treat*. If the AA policy had an effect, I would expect to see a positive statistically significant coefficient on *OBC X Treat*, reflecting a differential increase in the age at marriage and first birth for treated OBC women. It is important to keep in mind that women born between 1990 and 1995, who were aged 13-18 in 2008, are considered treated whereas women born between 1980 and 1985, who were aged 23-28 in 2008, are untreated. Furthermore, in Table 3A, women born between 1986 and 1989 are partially treated: while they could have benefitted from AA, it is less likely that they changed their educational trajectory since they were

Table 3A: Baseline Regression for Experiment A

	Age At Marriage			Age at First Birth		
	Base	Controls	State Effects	Base	Controls	State Effects
	(1)	(2)	(3)	(4)	(5)	(6)
OBC	-1.60 ***	-0.98 ***	-0.72 ***	-0.98 ***	-0.63 ***	-0.51 ***
	(0.04)	(0.03)	(0.14)	(0.03)	(0.03)	(0.11)
Partial Treat	0.02 ***	0.49 ***	0.44 ***	-0.35 ***	0.37 ***	0.36 ***
	(0.04)	(0.05)	(0.09)	(0.04)	(0.04)	(0.07)
Treat	-0.90 ***	-0.02	0.01	-1.70 ***	-0.14 **	-0.10
	(0.04)	(0.06)	(0.13)	(0.03)	(0.06)	(0.12)
OBC X PartialTreat	0.25 ***	0.21 ***	0.22 ***	0.30 ***	0.25 ***	0.25 ***
	(0.05)	(0.05)	(0.08)	(0.05)	(0.04)	(0.08)
OBC X Treat	0.95 ***	0.79 ***	0.76 ***	0.83 ***	0.67 ***	0.64 ***
	(0.04)	(0.04)	(0.15)	(0.04)	(0.04)	(0.13)
Birthyear		-0.07 ***	-0.05 ***		-0.14 ***	-0.13 ***
		(0.00)	(0.01)		(0.01)	(0.01)
Constant	19.87 ***	17.29 ***	19.33 ***	21.50 ***	19.51 ***	20.50 ***
	(0.03)	(0.05)	(0.21)	(0.03)	(0.05)	(0.09)
Observations	178330	178330	178330	158212	158212	158212
R Squared	0.03	0.12	0.18	0.03	0.09	0.11
Control Variables	NO	YES	YES	NO	YES	YES
State Effects	NO	NO	YES	NO	NO	YES

Table 3B: Baseline Regression for Experiment B

	Age At Marriage			Age at First Birth		
	Base	Controls	State Effects	Base	Controls	State Effects
	(1)	(2)	(3)	(4)	(5)	(6)
OBC	-1.42 ***	-0.90 ***	-0.83 ***	-0.70 ***	-0.46 ***	-0.46 ***
	(0.06)	(0.06)	(0.20)	(0.05)	(0.05)	(0.10)
Treat	1.00 ***	2.83 ***	2.77	0.52 ***	3.32 ***	3.20 ***
	(0.05)	(0.09)	(0.23)	(0.04)	(0.08)	(0.14)
OBC X Treat	0.73 ***	0.61 ***	0.66 ***	0.53 ***	0.43 ***	0.47 ***
	(0.07)	(0.06)	(0.19)	(0.06)	(0.06)	(0.09)
Birthyear		-0.17 ***	-0.16 ***		-0.30 ***	-0.29 ***
		0.00	(0.03)		(0.01)	(0.02)
Constant	18.13 ***	14.89 ***	16.28 ***	19.44 ***	16.48 ***	16.99 ***
	(0.04)	(0.08)	(0.22)	(0.04)	(0.07)	(0.13)
Observations	64089	64089	63591	51158	51158	50777
R Squared	0.04	0.14	0.18	0.02	0.10	0.12
Control Variables	NO	YES	YES	NO	YES	YES
State Effects	NO	NO	YES	NO	NO	YES

Notes: For the outcome of age at marriage, the sample only consists of married women. For the outcome of age at first birth, the sample consists of women that have given birth. Columns 1 and 3 simply present the baseline specifications. Columns 2 and 4 add the appropriate control variables and Columns 3 and 6 add state specific fixed effects. Standard errors are clustered at the state level and are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

already 19-22 years old in 2008. Hence, the intensity of treatment for partially treated women is assumed to be lesser than the intensity of treatment for the fully treated women. Nevertheless, *OBCX PartialTreat* is another coefficient of interest. The column of interest in both sections of Tables 3A & 3B is Column 3, which includes the control variables and state-specific fixed effects.

In Table 3B Column 3, the coefficient on *OBC X Treat* is 0.66. This means that while age at marriage, on average, increased over time for all Indian women by 2.77 years (the coefficient on *Treat*), treated OBC women increased their age at marriage by an additional 0.66 years. With reference to Table 2 Column 3 Panel B, the difference between the mean age at marriage of untreated OBC women and untreated General Population women was 1.42 years. Therefore, this differential increase in the age at marriage for treated OBC women is responsible for closing the gap between the two castes by close to half (0.66 / 1.42). Similar interpretations can be made for the other coefficients of interest in Tables 3A & 3B.

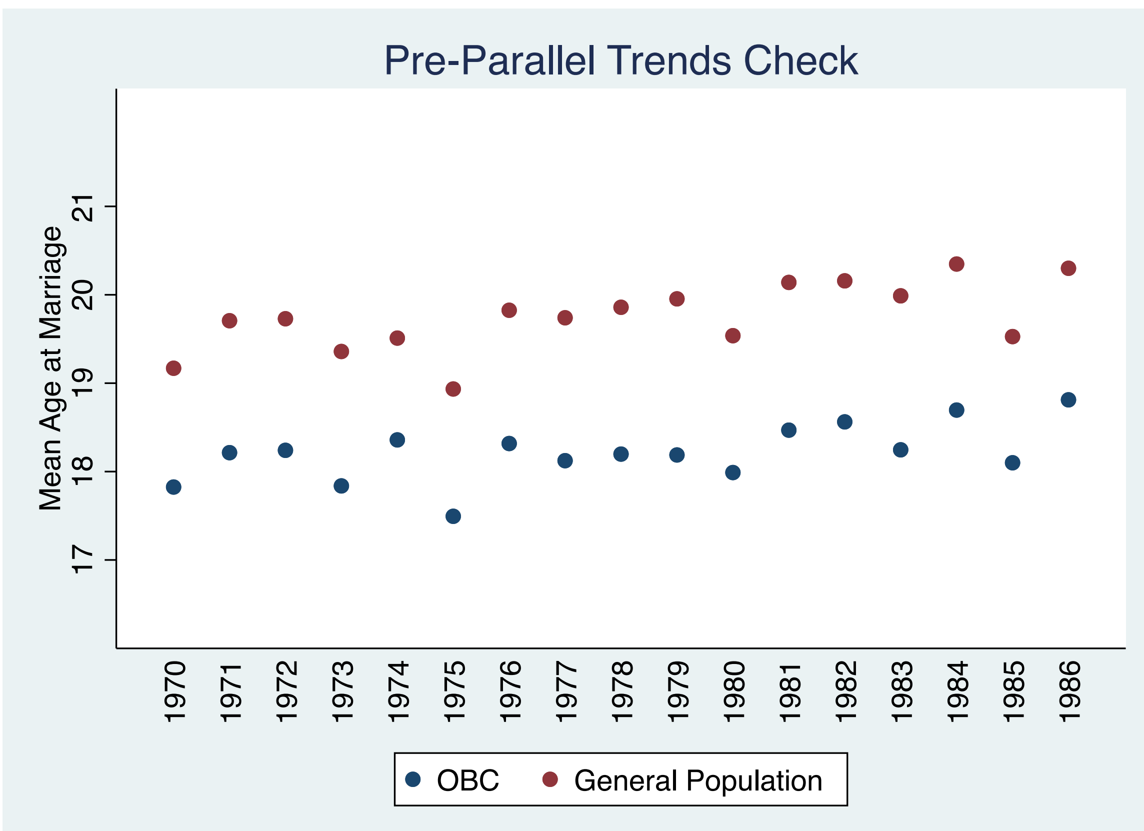
The magnitude of the differential impact is greater for the age at marriage compared to the age at first birth in both experiments. This could be because, in the Indian context, first birth tends to almost always occur after marriage. As the dataset may include some women that were married but who had not yet given birth, the downward sample selection bias for age at first birth is strengthened. Another potential explanation for this observation could be that after deferring their age at marriage, treated OBC women decreased the interval between marriage and first birth. However, since the results for age at marriage and age at first birth are similar, the rest of analysis predominantly focuses on age at marriage as the primary outcome of interest.

Another interesting observation from Tables 3A is that the effect of partial treatment is always positive yet smaller in magnitude. For example, the positive differential impact in the age at marriage for partially treated OBC women was a statistically significant 0.22 years while the coefficient on *OBC X Treat* was 0.76 years. This supports the hypothesis of the differing intensities of treatment between partially and fully treated women.

C. Parallel Trends

The causal interpretation of the above results relies on the validity of the parallel trends assumption: that there are no omitted time varying and caste or state-specific trends contemporaneous with the program. First, to observe pre-policy parallel trends, I plot the mean age at marriage for several birth years by caste for Experiment A[22]. In Figure 3, mean age at marriage trends for OBCs and the General Population seem to closely follow each other; the pre-policy parallel trend assumption between both castes seems to hold[23]. The slight positive slope

Figure 3: Pre-Policy Parallel Trend Check for Age at Marriage in Experiment A



Notes: This figures shows pre-policy parallel trends. The y-axis plots mean age at marriage and x-axis plots birth years. Blue data points represent OBCs and the red data points depict the General Population.

[22] Similar plots were plotted for Experiment B and can be found in the Appendix.
[23] Refer to the estimates of the event study, in Table 9A Column 1 for empirical evidence.

between mean age at marriage and birth year for both castes verifies the hypothesis that marriage patterns are gradually getting delayed in India.

Although the assumption of pre-policy parallel trends seems to be satisfied, the fact that India made substantial progress in reducing child marriage practices during the last few decades of the 20th century could lead to differential caste-specific trends over time. Furthermore, because India is a federal republic, age varying state-specific factors could further confound the estimates. Therefore, I must be cautious before concluding that the differential effect on treated OBC women in Tables 3A & 3B is the causal effect of the AA policy.

D. Baseline Regression with Controls for the Threats to Identification

Keeping these potential threats to identification in mind, I present the specifications in Tables 4 and 5. I include a *BirthYear X OBC* linear trend that controls for caste-specific differential trends over time, a *BirthYear X State* trend that accounts for state-specific time varying linear trends and a *BirthYear X OBC X Treat* interaction that tests whether treated OBCs are trending differentially compared to untreated OBCs.

In Table 4A, the coefficient on the *BirthYear X OBC* linear trend remains insignificantly different from zero, except in Column 4. For both age at marriage and first birth the coefficient on *OBC X Treat* is still positively signed but is attenuated after the inclusion of the control terms[24]. The fact that the coefficients of interest decrease after including these controls raises the fear that the coefficient on *OBC X Treat* in Column 3 is upwardly biased due to the pre-policy convergence of fertility patterns between castes. The estimates on *OBC X Treat* tend to remain robust to the inclusion of the *BirthYear X State* trend in Column 4.

[24] For example, in Table 4A, after including *BirthYear X OBC*, the coefficient of interest falls from 0.76 in Column 3 to 0.28 in Column 4. The coefficient slightly increases to 0.31 and 0.33 after including *BirthYear X State* and *BirthYear X OBC X Treat* in Columns 5 and 6 respectively.

Table 4A: Full Baseline Regression for Age at Marriage in Experiment A

	Base (1)	Controls (2)	State Effects (3)	Caste Trends (4)	State Trends (5)	Full (6)
OBC	-1.60 ***	-0.98 ***	-0.72 ***	-0.35 **	-0.46 ***	-0.62 ***
	(0.04)	(0.03)	(0.14)	(0.17)	(0.16)	(0.19)
Partial Treat	0.19 ***	0.49 ***	0.44 ***	0.61 ***	0.63 ***	0.15
	(0.04)	(0.05)	(0.09)	(0.10)	(0.10)	(0.10)
Treat	-0.90 ***	-0.02	0.01	0.33 **	0.32 **	-0.01
	(0.04)	(0.06)	(0.13)	(0.16)	(0.16)	(0.15)
OBC X Partial Treat	0.25 ***	0.21	0.22 ***	-0.03	-0.01	0.09
	(0.05)	(0.05)	(0.08)	(0.09)	(0.09)	(0.10)
OBC X Treat	0.95 ***	0.79 ***	0.76 ***	0.28 **	0.31 **	0.33 **
	(0.04)	(0.04)	(0.15)	(0.13)	(0.13)	(0.14)
Birthyear		-0.07 ***	-0.05 ***	-0.09 ***	-0.17 ***	-0.07 ***
		(0.01)	(0.01)	(0.02)	(0.01)	(0.02)
BY X OBC				0.05 ***	0.02	0.00
				(0.02)	(0.02)	(0.02)
BY X OBC X Treat						0.09 ***
						(0.03)
Constant	19.87 ***	17.30 ***	19.33 ***	19.08 ***	18.86 ***	19.56 ***
	(0.03)	(0.05)	(0.21)	(0.25)	(0.27)	(0.27)
Observations	178330	178330	178330	178330	178330	178330
R Squared	0.03	0.12	0.18	0.18	0.19	0.19
Control Variables	NO	YES	YES	YES	YES	YES
State Effects	NO	NO	YES	YES	YES	YES
State Linear Trends	NO	NO	NO	NO	YES	YES

Table 4B: Full Baseline Regression for Age at Marriage in Experiment B

	Base (1)	Controls (2)	State Effects (3)	Caste Trends (4)	State Trends (5)	Full (6)
OBC	-1.42 ***	-0.90 ***	-0.83 ***	0.09	0.04	0.05
	(0.06)	(0.06)	(0.20)	(0.21)	(0.20)	(0.32)
Treat	1.00 ***	2.83 ***	2.77	3.53 ***	3.60 ***	3.06 ***
	(0.05)	(0.09)	(0.23)	(0.26)	(0.26)	(0.27)
OBC X Treat	0.73 ***	0.61 ***	0.66 ***	-0.47 **	-0.42 *	-0.44
	(0.07)	(0.06)	(0.19)	(0.21)	(0.21)	(0.31)
Birthyear		-0.17 ***	-0.16 ***	-0.25 ***	-0.31 ***	-0.23 ***
		(0.00)	(0.03)	(0.03)	(0.03)	(0.03)
BY X OBC				0.13 ***	0.11 ***	0.11 **
				(0.02)	(0.02)	(0.05)
BY X OBC X Treat						0.01
						(0.05)
Constant	18.13 ***	14.89 ***	16.28 ***	15.66 ***	15.59 ***	16.17 ***
	(0.04)	(0.08)	(0.22)	(0.27)	(0.26)	(0.30)
Observations	64089	64089	63591	63591	63591	63591
R Squared	0.04	0.14	0.18	0.18	0.19	0.19
Control Variables	NO	YES	YES	YES	YES	YES
State Effects	NO	NO	YES	YES	YES	YES
State Linear Trends	NO	NO	NO	NO	YES	YES

Notes: The sample only consists of married women. Columns 1, 2 and 3 are exactly the same as columns 1, 2 and 3 from Table 3. Column 4 adds a caste specific linear trend while Column 5 adds state specific linear trends. Standard errors are clustered at the state level and are in parentheses.*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 5A: Full Baseline Regression for Age at First Birth in Experiment A

	Base	Controls	State Effects	Caste Trends	State Trends	Full
	(1)	(2)	(3)	(4)	(5)	(6)
OBC	-0.98 ***	-0.63 ***	-0.51 ***	-0.35 ***	-0.41 ***	-0.45 ***
	(0.03)	(0.03)	(0.11)	(0.12)	(0.12)	(0.14)
Partial Treat	-0.35 ***	0.37 ***	0.36 ***	0.43 ***	0.44 ***	0.01
	(0.04)	(0.04)	(0.07)	(0.07)	(0.07)	(0.07)
Treat	-1.70 ***	-0.14 **	-0.11 ***	0.03	0.01 ***	-0.21 *
	(0.03)	(0.06)	(0.12)	(0.12)	(0.12)	(0.11)
OBC X Partial Treat	0.30 ***	0.25 ***	0.25 ***	0.14 *	0.15 *	0.17 **
	(0.05)	(0.04)	(0.08)	(0.08)	(0.08)	(0.08)
OBC X Treat	0.83 ***	0.67 ***	0.64 ***	0.44 ***	0.45 ***	0.39 ***
	(0.04)	(0.04)	(0.13)	(0.13)	(0.12)	(0.12)
Birthyear		-0.14 ***	-0.13 ***	-0.15 ***	-0.16 ***	-0.07 ***
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
BY X OBC				0.02	0.01	0.00
				(0.02)	(0.01)	(0.02)
BY X OBC X Treat						0.06 **
						(0.02)
Constant	21.50 ***	19.51 ***	20.50 ***	20.39 ***	20.20 ***	20.83 ***
	(0.03)	(0.05)	(0.09)	(0.11)	(0.12)	(0.12)
Observations	158212	158212	158212	158212	158212	158212
R Squared	0.03	0.09	0.11	0.11	0.12	0.12
Control Variables	NO	YES	YES	YES	YES	YES
State Effects	NO	NO	YES	YES	YES	YES
State Linear	NO	NO	NO	NO	YES	YES

Table 5B: Full Baseline Regression for Age at First Birth in Experiment B

	Base	Controls	State Effects	Caste Trends	State Trends	Full
	(1)	(2)	(3)	(4)	(5)	(6)
OBC	-0.70 ***	-0.46 ***	-0.46 ***	0.09	0.04	0.05
	(0.05)	(0.05)	(0.10)	(0.21)	(0.20)	(0.32)
Treat	0.52 ***	3.32 ***	3.20 ***	3.53 ***	3.60 ***	3.06 ***
	(0.04)	(0.08)	(0.14)	(0.26)	(0.26)	(0.27)
OBC X Treat	0.53 ***	0.43 ***	0.47 ***	-0.47 **	-0.42 *	-0.44
	(0.06)	(0.06)	(0.09)	(0.21)	(0.21)	(0.31)
Birthyear		-0.30 ***	-0.29 ***	-0.25 ***	-0.31 ***	-0.23 ***
		(0.01)	(0.02)	(0.03)	(0.03)	(0.03)
BY X OBC				0.13 ***	0.11 ***	0.11 **
				(0.02)	(0.02)	(0.05)
BY X OBC X Treat						0.01
						(0.05)
Constant	19.44 ***	16.48 ***	16.99 ***	15.66 ***	15.59 ***	16.17 ***
	(0.04)	(0.07)	(0.13)	(0.27)	(0.26)	(0.30)
Observations	51158	51158	50777	63591	63591	63591
R Squared	0.02	0.10	0.12	0.18	0.19	0.19
Control Variables	NO	YES	YES	YES	YES	YES
State Effects	NO	NO	YES	YES	YES	YES
State Linear Trends	NO	NO	NO	NO	YES	YES

Notes: The sample only consists of married women. Columns 1, 2 and 3 are exactly the same as columns 4, 5 and 6 from Table 3 . Column 4 adds a caste specific linear trend while Column 5 adds state specific linear trends. Standard errors are clustered at the state level and are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Furthermore, the fact that the coefficient on *BirthYear X OBC X Treat* is positive and statistically significant is striking. The coefficient on *OBC X Treat* after the inclusion of *BirthYear X OBC X Treat* remains positive and significant. As 1990 is the base birth year, these findings suggest that there was an initial positive treatment effect in 1990 and the intensity of the treatment continued to increase within the treated cohort[25]. This crucial finding of the increasing treatment effect, even within the treated cohort, substantiates the hypothesis of younger OBC women being better equipped to alter their educational path to benefit from AA. It also gives credence to the theories that there were time lags between the passing of the bill and the ensuing implementation of the policy and that it took time for OBC women to become aware that they were eligible for AA in higher education.

In Table 5B, the *BirthYear X OBC* linear trend is always positive and statistically significant. This means that, particularly in Experiment B, there is suggestive evidence for a differential upward trend for OBCs, in line with the hypothesis that fertility patterns had started to converge before the policy. More concerning, after the inclusion of caste linear trends, the coefficient on *OBC X Treat* flips and becomes negative. A plausible explanation for the flip is downward sample selection bias, as only women that likely never benefitted from the policy are observed. The next section accounts for sample selection bias and will investigate this further.

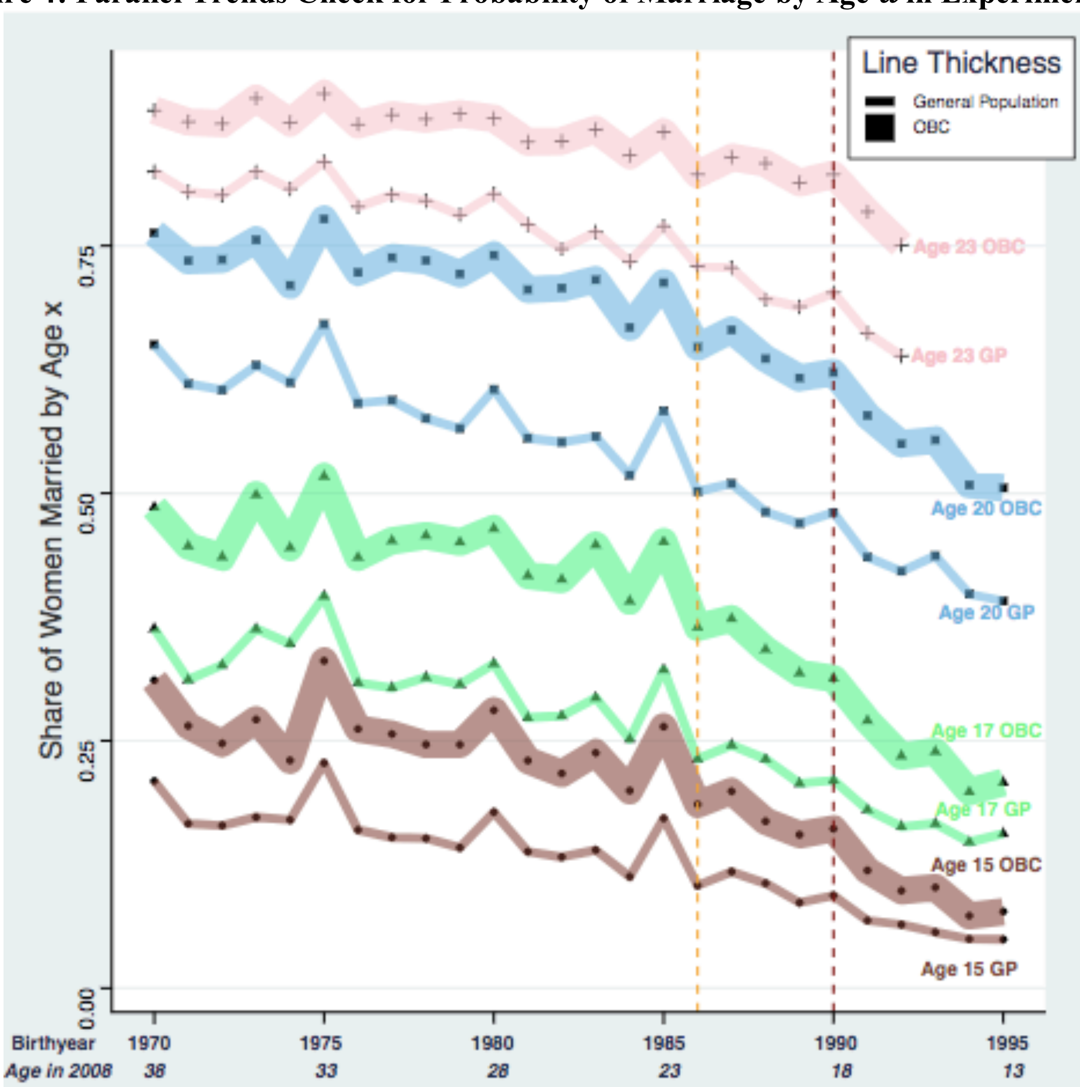
E. Accounting for Sample Selection Bias Using Logistic Regression

Now, I turn to Equation 2 to estimate the probability of a woman being married at or before a certain age a. This attempts to solve the sample selection bias issue as both married and

[25] With 1990 as the base birth year, the coefficient of 0.09 on *BirthYear X OBC X Treat* in Table 4A indicates that the differential effect for treated OBCs born in 1991 is 0.33 (the coefficient on *OBC X Treat*) $+ \ 0.09 \ = \ 0.42$ years. Likewise, the treatment effect for OBCs born in 1995 is $0.33 \ + \ 4 \ * \ (0.09) = \ 0.69$ years. Therefore, the treatment effect for OBC women born in 1990 is 0.33 years while the treatment effect of OBCs born in 1995 is 0.69 years, which represents $((0.69 - 0.33) \ / \ 0.33) \ * \ 100$ = 109.09% change in the treatment effect between 1990 and 1995.

unmarried women are needed to estimate the probability of being married by a certain age. First, it is imperative to check whether parallel trends assumption holds in the logistic regression framework. Figure 4 demonstrates how the share of women who are married at different ages trends over time. To better visualize the parallel trends, only four ages, 15, 17, 20 and 23, are chosen. Each age has its unique color. The thinner lines correspond to the trends for the General Population while the thicker lines correspond to the patterns of the OBCs. Figure 4 indicates that the pre-policy parallel trends assumption for the untreated cohort, observed to the left of the orange vertical line, holds for all ages. That is, the gap between the share of OBC and General Population

Figure 4: Parallel Trends Check for Probability of Marriage by Age a in Experiment



Notes: This figures shows pre-policy parallel trends for the probability of marriage at ages 15, 17, 20 & 23. The x-axis plots birth year / age in 2008 and the y axis plots the proportion of women married The thin lines correspond to the General Population and the thicker lines correspond to OBC woman.

women married at all ages remains fairly constant for women born between 1970 and 1985. For ages 15 and 17, this gap begins to close for the partially treated cohort. To the right of the red vertical line, OBCs in the treated cohort have almost completely closed the gap with the General Population. These dynamic trends for the probability of marriage at ages 15 and 17, are suggestive of the treatment's effect on delaying the age at marriage for OBC women. For age 20, the gap between OBCs and the General Population only slightly decreases for the partially and fully treated cohorts whereas for age 23, the gap is wider. The lack of a treatment effect for ages 20 and 23 can be explained by sample selection bias because it persists in the logistic regression framework for later ages since the marital status of several unmarried women in their 20s remains unobserved[26].

Turning to the regression results, Table 6 estimates the probability of being married at or before the age of 18 in Experiment B and is just one of the many logistic regressions, using Equation 2, run for each age[27]. Although there are threats to identification, including caste and state-specific differential trends over time, the strong pre-policy parallel trends between OBCs and the General Population in Figure 4 make Column 3, which includes the baseline regression

Table 6: Marginal Effects for Probability of Marriage by Age 18 in Experiment B

	Base (1)	Controls (2)	State Effects (3)	Caste Trends (4)	State Trends (5)	Full (6)
OBC	0.18***	0.10***	0.09***	0.08**	0.09***	0.15***
	(0.01)	(0.01)	(0.02)	(0.03)	(0.03)	(0.04)
Treat	-0.11***	0.03***	0.0200	0.01	0.00	0.05*
	(0.01)	(0.01)	(0.02)	(0.03)	(0.02)	(0.03)
OBC X Treat	-0.08***	-0.06***	-0.07***	-0.06**	-0.07***	-0.12***
	(0.01)	(0.01)	(0.02)	(0.03)	(0.03)	(0.03)
Observations	87,985	87,981	87,174	87,174	87,174	87,174
Control Variables	NO	YES	YES	YES	YES	YES
State Effects	NO	NO	YES	YES	YES	YES
Caste Linear Trends	NO	NO	NO	YES	YES	YES
State Linear Trends	NO	NO	NO	NO	YES	YES
Differential Treated Trend	NO	NO	NO	NO	NO	YES

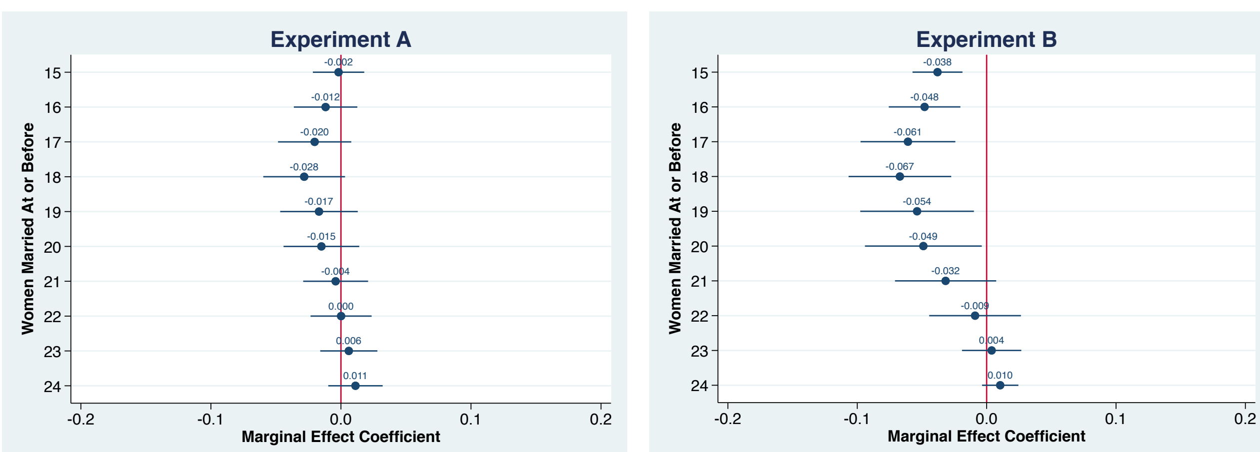
Notes: Marginal effects for the probability of marriage by age 18 in Experiment B, calculated at *OBC = 1, Treat = 1 and OBC X Treat = 1,* are displayed. Standard errors are clustered at the state level and are in paranthesis.
*** p<0.01, ** p<0.05, *p<0.1

[26] Refer back to Section VI C, page 20-21.
[27] All tables have not been displayed in this thesis due space constraints but can be found in the Appendix.

framework, appropriate control variables and state-specific fixed effects, of most interest. It is important to note that because of the non-linear logistic function, the coefficients in Table 6 are marginal effects, calculated at *OBC = 1, Treat = 1* and *OBC X Treat = 1.*[28] Since the AA policy is hypothesized to delay fertility patterns for treated OBC women, I expect the differential marginal effect on *OBC X Treat* to be negative as it would be less likely for treated OBC women to be married by age a. In Table 6 Column 3, the marginal effect on *OBC X Treat* indicates that treatment led to a 7 percentage point differential decrease in the probability of marriage for OBC women at age 18. As the coefficient on *OBC* is 0.09, the differential decrease almost completely offsets the gap between the two castes. The rest of the results predominantly plot the marginal effects of *OBC X Treat* from Column 3 for all ages between 15 and 24.

In Figure 5, supporting the expectation of a differential decrease in the probability of marriage, the marginal effects on *OBC X Treat* are predominantly negative. In Experiment A, the marginal effects start out negative but very close to zero for age 15, become more negative in the

Figure 5: Marginal Effects on *OBC X Treat* for Probability of Marriage by Age a



Notes: Marginal effects on *OBC X Treat* in the specification with control variables and state fixed effects are plotted. Marginal effects are calculated at *OBC =1, Treat = 1 and OBC X Treat = 1.*Bars represent 90% confidence intervals.

[28] Marginal effects were presented and are of most interest because the non-linear logistic function makes the logit coefficients hard to interpret. Furthermore, an advantage of marginal effects is that they are not determined simply at the sample average but enable me to calculate the impact on the marginal woman of interest.

later teens, before starting to increase and eventually becoming positive for ages 23 and 24. However, none of the marginal effects are statistically different from zero. A comparable pattern is seen for Experiment B. For age 15, the marginal effect is -0.038, representing a 3.8 percentage point differential decrease. The marginal effects become more negative till age 19, before increasing and finally becoming positive again for age 23 and 24. The negative differential impact between the ages 15 and 19 is statistically significant. The fact that we see the negative differential impacts gradually becoming stronger from age 15 to age 20 indicates that reduction in child marriage is not the only driving force behind the observed treatment effect, as if this were the case, we would see a gradually decreasing differential effect as age increased. Furthermore, the fact that the most negative marginal effects are observed for the most common college attending years, between 17 and 20, reinforces the assumption that the observed treatment effect is caused by more OBC women attending higher education institutions. Therefore, Figure 5 accounts for sample selection bias and provides weak evidence of the differential impact on treated OBC women at the all India level.

F. *Difference in Difference Estimation by Region*

The analysis thus far has only focused on estimates at the all India level. However, due to regional variation in the amount of AA before the implementation of the CEIAB, I proceed with estimating the double difference for each region separately. As described in the Conceptual Framework, I expect the eastern region to have the largest impact from the policy, followed by the north-central region. The southern region should theoretically be unaffected as it had been implementing quotas in excess of the centrally mandated 27% for OBCs even before 2008.

Table 7 provides suggestive evidence to support this hypothesis by presenting the same differences in differences as Table 2, but for each region separately. OBCs in the southern region

Table 7: Difference in Difference Estimation by Region

	Age At First Marriage				Age At First Birth			
	South (1)	North Central (2)	East (3)	(East - South) (4)	South (5)	North Central (6)	East (7)	(East - South) (8)
Experiment A								
Difference in Difference	0.59***	1.00***	1.54***	0.94***	0.52***	0.77***	1.33***	0.80***
	(0.15)	(0.05)	(0.10)	(0.18)	(0.14)	(0.05)	(0.09)	(0.17)
Experiment B								
Difference in Difference	0.46**	0.87***	1.50***	1.05***	0.34**	0.50***	0.88***	0.54**
	(0.17)	(0.09)	(0.14)	(0.22)	(0.16)	(0.08)	(0.12)	(0.20)

Notes: This table only displays the difference in difference estimations for three regions. The single differences across cohorts and castes, as displayed in Table 3 were calculated but have not been shown. Columns 4 and 9 present the triple difference by subtracting the double difference of the eastern region by the same double difference from the southern region. Standard errors are clustered at the state level and are in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

had the least impact from the policy, with a 0.59 year differential increase in age at marriage over time compared to the southern General Population women. The largest differential increase of 1.54 years was seen for OBC women in the eastern region. In fact, the differential increase in the eastern region was twice as large compared to the southern region, and more pronounced in Experiment B. The differences in differences by region followed the same pattern for age at first birth, although the percentage change between the southern and eastern regions was smaller in magnitude.

G. Baseline Regression by Region

Next, I move to the familiar regression framework where I estimate Equation 1 separately for each region. Once again, the specification of most interest incorporates the appropriate control variables and state-specific fixed effects.

Table 8 displays the results of this specification[29]. The coefficient on *OBC X Treat* in Table 8A Column 2, for the southern region, is 0.19 and statistically insignificant, whereas the same coefficient in Column 3 for the north-central region is 0.77. The differential impact is greatest for

[29] Column 1 displays the results at the all India level, helping with comparisons,

Table 8A: Baseline Regression by Region for Experiment A

	All India (1)		South (2)		North - Central (3)		East (4)	
OBC	-0.72	***	-0.35	**	-0.80	***	-0.69	**
	(0.14)		(0.11)		(0.18)		(0.28)	
Partial Treat	0.44	***	0.41		0.59	***	-0.04	
	(0.09)		(0.32)		(0.07)		(0.14)	
Treat	0.01		-0.14		0.19		-0.44	
	(0.13)		(0.32)		(0.13)		(0.30)	
OBC X PartialTreat	0.22	***	0.11		0.13		0.56	***
	(0.08)		(0.32)		(0.09)		(0.13)	
OBC X Treat	0.76	***	0.19		0.77	***	1.15	**
	(0.15)		(0.24)		(0.15)		(0.39)	
Birthyear	-0.052	***	-0.042		-0.04	***	-0.095	**
	(0.01)		(0.03)		(0.01)		(0.04)	
Constant	19.33	***	16.44	***	19.14	***	17.72	***
	(0.21)		(0.17)		(0.22)		(0.37)	
Observations	178330		25815		117504		35011	
R Squared	0.18		0.16		0.19		0.16	
Control Variables	YES		YES		YES		YES	
State Effects	YES		YES		YES		YES	
State Linear Trends	NO		NO		NO		NO	

Table 8B: Baseline Regression by Region for Experiment B

	All India (1)		South (2)		North - Central (3)		East (4)	
OBC	-0.83	***	-0.63	*	-0.99	***	-1.07	**
	(0.20)		(0.27)		(0.26)		(0.35)	
Treat	2.77	***	2.93	***	2.68	***	2.87	***
	(0.23)		(0.71)		(0.26)		(0.44)	
OBC X Treat	0.66	***	0.38		0.73	***	1.36	***
	(0.19)		(0.37)		(0.21)		(0.34)	
Birthyear	-0.16	***	-0.24	**	-0.13	***	-0.20	***
	(0.03)		(0.08)		(0.03)		(0.05)	
Constant	16.28	***	14.05	***	16.21	***	14.59	***
	(0.22)		(0.58)		(0.29)		(0.35)	
Observations	63591		8892		41927		12772	
R Squared	0.18		0.19		0.18		0.19	
Control Variables	YES		YES		YES		YES	
State Effects	YES		YES		YES		YES	
State Linear Trends	NO		NO		NO		NO	

Notes: The sample only consists of married woman. Column 1 is at the All India a level and is exactly the same as Column 3 in Table 3A for Experiment A and as Column 3 in Table 3B for Experiment B. Columns 2, 3 and 4 estimate the baseline regression separately for each region. Standard errors are clustered at the state level and are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 9A: Pre-Policy Parallel Trends Check Using an Event Study for Experiment A

	All India (1)	South (2)	North - Central (3)	East (4)
1977 X OBC	0.00	-0.48	-0.07	0.44
	(0.16)	(0.60)	(0.21)	(0.39)
1978 X OBC	-0.03	0.36	0.01	-0.12
	(0.13)	(0.53)	(0.16)	(0.24)
1979 X OBC	-0.09	-0.52	-0.08	0.16
	(0.17)	(0.73)	(0.22)	(0.34)
1980 X OBC	0.06	-0.14	0.06	0.44
	(0.13)	(0.29)	(0.16)	(0.35)
1981 X OBC	-0.15	-0.09	-0.33 **	0.53
	(0.13)	(0.35)	(0.15)	(0.39)
1982 X OBC	0.07	0.41	-0.12	0.61
	(0.17)	(0.80)	(0.15)	(0.44)
1983 X OBC	-0.11	0.28	-0.28	0.49
	(0.14)	(0.49)	(0.16)	(0.29)
1984 X OBC	-0.07	-0.52	-0.18	0.58
	(0.14)	(0.49)	(0.15)	(0.37)
1985 X OBC	0.12	-0.12	0.03	0.66
	(0.16)	(0.64)	(0.18)	(0.41)
Observations	107742	17,450	68,771	21,521
R-squared	0.19	0.16	0.19	0.17
Control Variables	YES	YES	YES	YES
State Effects	YES	YES	YES	YES
State Linear Trends	NO	NO	NO	NO

Table 9B: Pre-Policy Parallel Trends Check Using an Event Study for Experiment B

	All India (1)	South (2)	North - Central (3)	East (4)
1968 X OBC	0.20	0.38	0.15	-0.04
	(0.19)	(0.41)	(0.22)	(0.49)
1969 X OBC	0.15	0.56	0.24	-0.84
	(0.33)	(0.77)	(0.41)	(0.70)
1970 X OBC	0.10	-0.43	0.00	0.50
	(0.28)	(0.78)	(0.33)	(0.59)
1971 X OBC	0.18	-0.57	0.44	0.12
	(0.26)	(0.71)	(0.34)	(0.59)
1972 X OBC	0.35	-0.58	0.53 **	0.70
	(0.28)	(0.98)	(0.24)	(0.51)
1973 X OBC	0.39	0.72	0.31	0.28
	(0.28)	(0.67)	(0.32)	(0.65)
1974 X OBC	0.27	0.16	0.43	0.13
	(0.25)	(0.44)	(0.32)	(0.54)
1975 X OBC	0.29	0.10	0.14	0.72
	(0.24)	(0.29)	(0.30)	(0.58)
1976 X OBC	0.24	0.22	0.21	0.18
	(0.27)	(0.75)	(0.30)	(0.67)
Observations	22,036	5,374	11,267	5,395
R-squared	0.30	0.29	0.29	0.30
Control Variables	YES	YES	YES	YES
State Effects	YES	YES	YES	YES
State Linear Trends	NO	NO	NO	NO

Notes: Interaction terms between birth year dummies and OBC are displayed. Column 1 is at All India level and the subsequent Columns are at the regional level. Outcome variable is age at marriage. Standard errors are clustered at the state level and are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

eastern OBCs, suggesting that treated eastern OBC women increased their age at marriage by an additional 1.15 years compared to treated eastern General Population women during the same time period. It is also notable that the eastern region is the only region that had a statistically significant coefficient on *OBC X Partial Treat*, which is indicative of the strong impact the policy had in the east. The results from Tables 8A & 8B perfectly line up with the hypothesis of the differing impacts the CEIAB had by region, even without accounting for downward bias from sample selection.

To check the pre-policy parallel trends at the regional level, I conduct an event study. I estimate the following equation for each region separately:

$$FP_{icst} = \beta_1 OBC_{ic} + \sum_{\tau=1977}^{1985} OBC_{ic}\, X\, D_t^{\tau} + \alpha_s + \delta_t + C_i + \epsilon_{icst} \qquad (4)$$

where all the variables are the same as above. The new addition D_t^{τ} is a binary variable equal to one if the birth year t equals the specific year, τ and zero otherwise. Birth year fixed effects δ_t are also included. A selection of women who were between 29-39 years old when observed, born between 1976 and 1985, for Experiment A and born between 1967 and 1976 for Experiment B[30] were chosen. The interaction terms for 1976 and 1967 are respectively omitted to identify the models. Tables 9A & 9B show the interaction coefficients between OBC and birth year dummies. All the interactions at both the regional and the all-India level are statistically insignificant (except two years in the north-central region). However, upon closer look, estimates for the eastern region are systematically positive, large in magnitude and only insignificant due to large standard errors. The interaction terms for the southern and north-central region are smaller in magnitude and tend to oscillate between positive and negative. Therefore, although the event study confirms that the

[30] Earlier birth years were chosen for the pre-trends check in Experiment B since several women born between 1976 and 1985 were not married when observed in 2006.

parallel trends assumption holds for the southern and north-central regions[31], the threat of a pre-policy differential upward trend in the fertility patterns of eastern OBC women persists.

H. Accounting for Sample Selection Bias Using Logistic Regression by Region

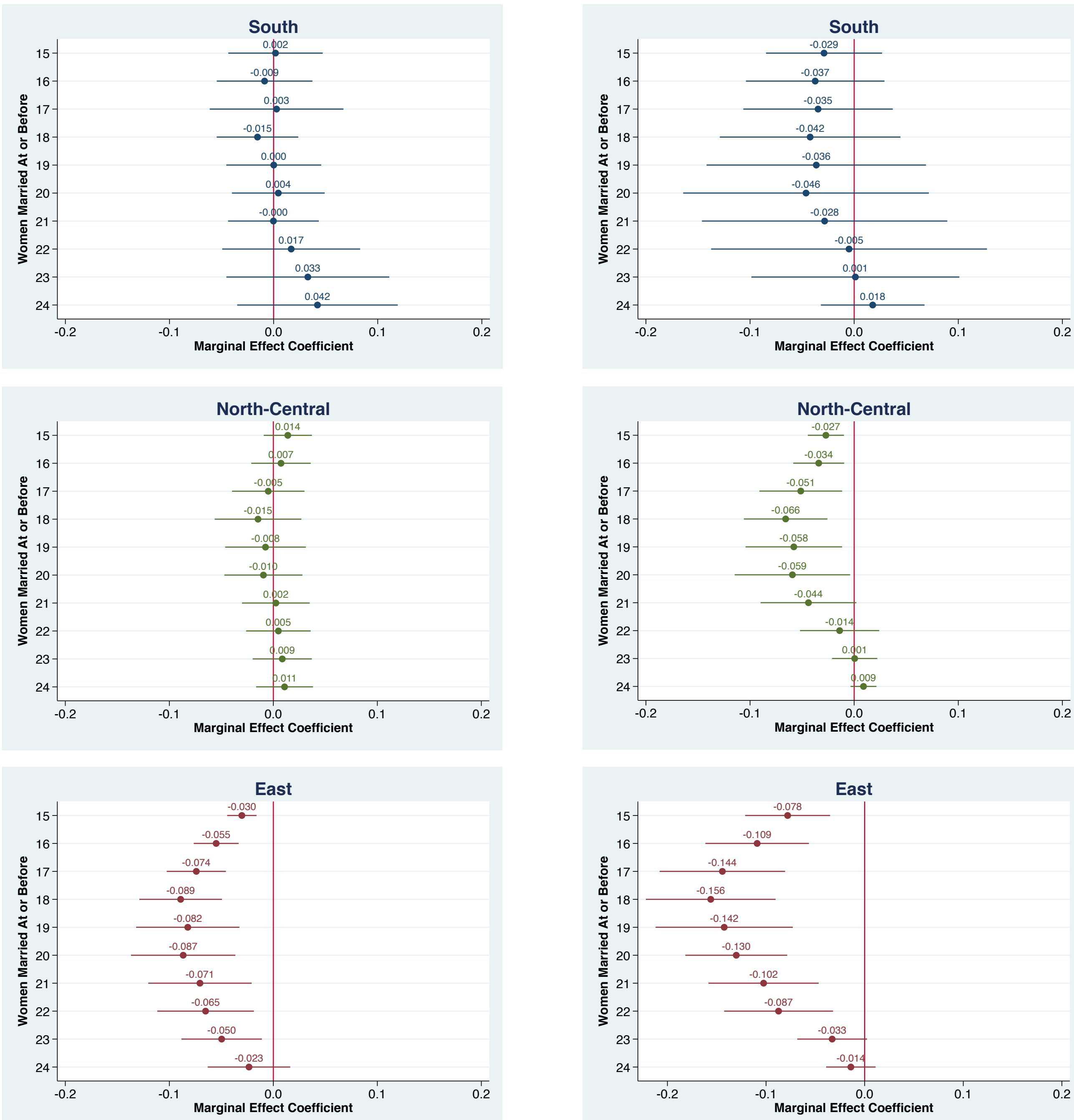
Before dealing with differential upward trends in the eastern region, I turn to the familiar logistic regression framework at the regional level to solve the issue of sample selection bias.

Figure 6 is a substantial confirmation of the primary hypothesis of interest. Treated OBC women showed substantial declines in the likelihood of getting married in the eastern region. For example, in Experiment B, eastern OBC women had a 14.2 percentage point decrease in the likelihood of getting married by age 19. This differential decrease entirely offsets the difference in the probability of marriage between OBCs and the General Population. Furthermore, as hypothesized, there appears to be no observed treatment effect in the southern region. The marginal effects on *OBC X Treat* in the southern region are all estimated to be very close to zero and are all insignificant. The results are less clear for the north-central region. In Experiment B, there is a statistically significant differential impact between the ages 15 and 19, but in Experiment A the marginal effects are all insignificant and very close to zero. Additionally, it is also worth noting that for the eastern region, the magnitude of the differential decrease becomes more negative till age 19 before beginning to increase again. As explained in the description of Figure 5, this pattern is unlikely to occur if the observed treatment effect was solely driven by the effects of reducing child marriage. Therefore, taken together, Figure 6 provides strong evidence of the effect of AA in higher education on the fertility patterns of treated women.

[31] For the north-central region, the parallel trends assumption seems to hold for Experiment A, whereas the estimates in Experiment B are also systematically positive.

Figure 6: Marginal Effects on *OBC X Treat* for Probability of Marriage by Age *a* by Region

Experiment A **Experiment B**



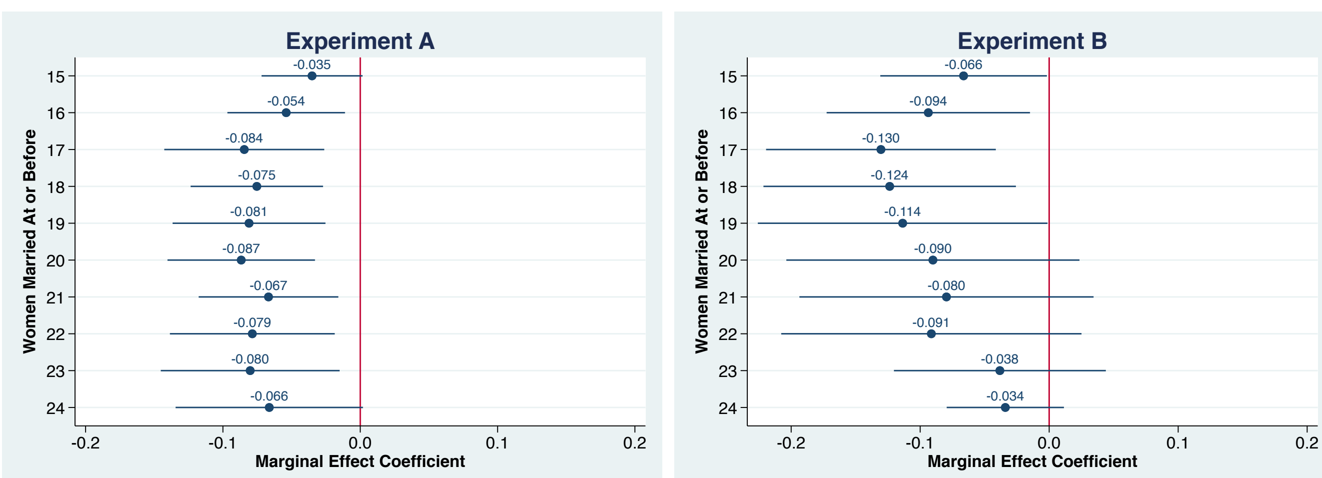
Notes: Marginal effects on *OBC X Treat* in the specification with control variables and state fixed effects are plotted. Marginal effects are calculated at *OBC =1, Treat = 1 and OBC X Treat = 1.* Age at marriage is the outcome variable. Bars represent 90% confidence intervals.

I. *Triple Difference Approach*

Before interpreting the results of Figure 6 as the causal effect of the CEIAB, I need to account for possible differential upward trends in OBC fertility patterns in the eastern region. Therefore, I turn to a triple difference approach using Equation 3. The double difference (as computed above) for the eastern region is compared to the double difference for the southern region[32]. The southern region serves as an especially useful counterfactual since it was exposed to all the other national level changes and policies during the period of interest but was not affected by the AA policy, as seen in Figure 6[33]. Figure 7 displays the marginal effects on *OBC X East X Treat* for the logistic regression framework in Experiment A and B, respectively.

The coefficients on *OBC X East X Treat* for all ages in both experiments were negative. This means that relative to women from the southern region, the decrease in the probability that a woman was married by a certain age between the treated and untreated cohorts was greater for

Figure 7: Marginal Effects on *OBC X East X Treat* for Probability of Marriage by Age *a*



Notes: Marginal effects on *OBC X East X Treat* in the specification with control variables and state fixed effects are plotted. The eastern region is compared to the southern region in the triple difference estimates. Bars represent 90% confidence intervals.

[32]The north-central states are left out of the triple difference estimate because of the evidence of partial treatment in the region.

[33] The parallel trends assumption for the triple difference specification was tested using an event study. The estimates can be found in the Appendix. The parallel trend assumption seems to hold.

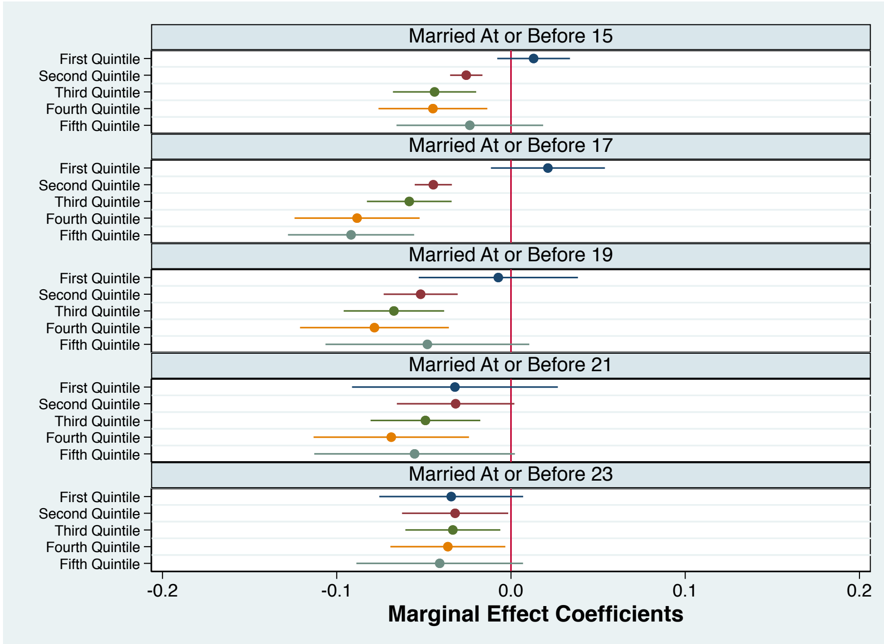
OBCs compared to the General Population in the eastern region. Noticeably, the magnitudes of the negative differential effect for the triple difference are smaller than the magnitudes of the marginal effects in the double difference for the eastern region. For example, in Experiment B, the differential impact at age 19 decreased from 14.2 percentage points[34] to 11.2 percentage points. Furthermore, in Experiment B, the triple difference estimates for ages 20 to 24 are no longer statistically significant. These findings suggest that the double difference estimates for the eastern region were upwardly biased due to differential upward trends in OBC fertility patterns or omitted national level changes or policies. As the triple difference approach, to a certain extent, accounts for the upward bias, it is the preferred estimation strategy.

J. The Creamy Layer

As mentioned in the Introduction, there is a prevalent notion in India that AA only helps the most affluent minorities and that it does not benefit the poorest members of the lower castes. To empirically investigate this, I interact the triple interaction term *OBC X East X Treat* with a Wealth Quintile categorical variable. The first quintile is the poorest and the fifth quintile is the richest. Figure 8 shows that there is no evidence of AA only benefitting the 'creamy layer', a small section of the OBCs that are financially well-off. Although OBCs in the fourth and fifth quintile seem to benefit the most from AA, Figure 8 shows that there is a considerable treatment effect for all OBCs in or above the second quintile of the wealth distribution. However, the lack of a treatment effect for OBCs in the first and poorest quintile is concerning as it suggests that the poorest OBCs did not benefit from the CEIAB. Although Figure 8 busts the myth that AA only benefits the 'creamy layer', it gives credence to the criticism that AA is ineffective in improving the outcomes of the poorest members of the targeted group.

[34] As seen in Figure 6.

Figure 8: Interactions between Wealth Quintile & *OBC X East X Treat* in Experiment A



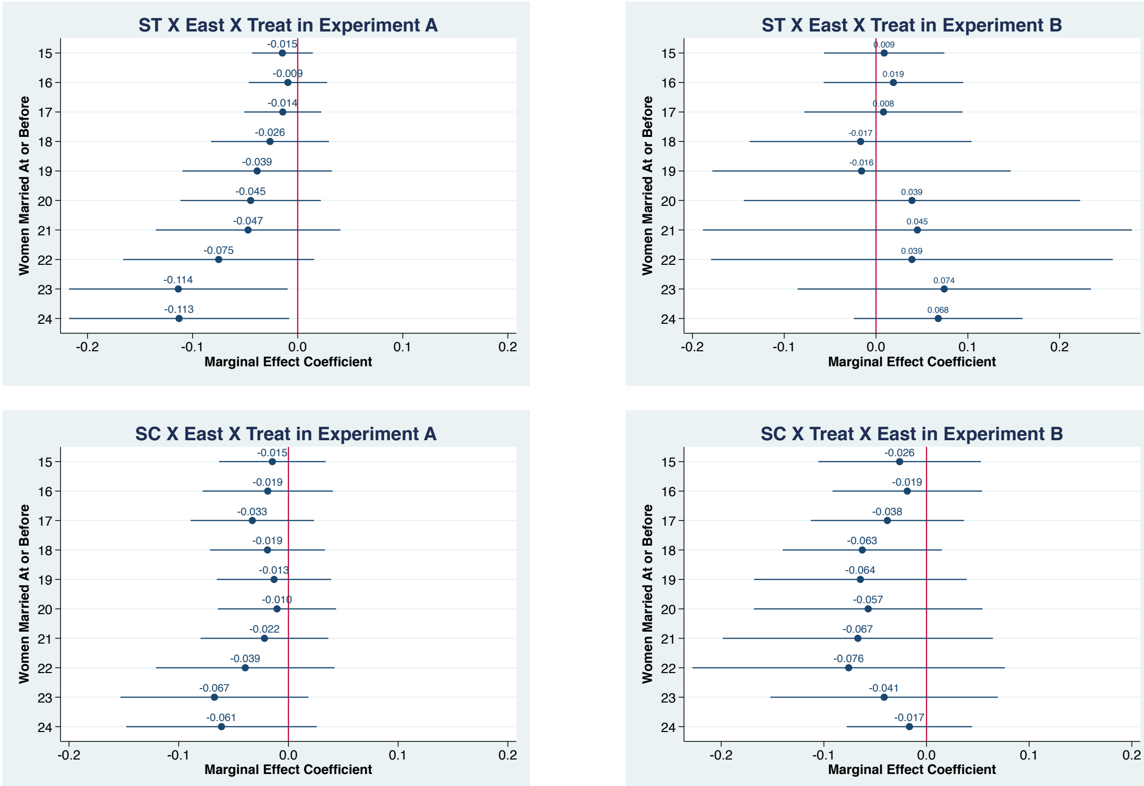
Notes: Marginal effects on *OBC X East X Treat X Wealth Quintile* in the specification with control variables and state fixed effects. Probability of marriage 15, 17, 19 & 23 selected. The first quintile is the poorest while the fifth quintile is the wealthiest. Bars represent 90% confidence intervals.

VIII. Falsification Tests

A. Falsification Test Using Scheduled Castes and Scheduled Tribes

As members of the lower castes were more likely to be living in poverty, eastern region-specific poverty reduction policies, remain an unaddressed threat to identification. If the impacts of these omitted policies were driving the treatment effects observed in the regional double difference and triple difference figures above, a similar impact should be observed on the SCs and STs. In fact, since SCs and STs have lower socio-economic outcomes than the OBC caste category, the observed treatment effects would be greater for SCs and STs. Therefore, as a falsification test, I run the triple difference specification with SCs and STs separately instead of the OBCs. Figure 9 indicates that most coefficients are negative but insignificantly different than zero. All coefficients in Figure 9 also seem to be greater (less negative) than the coefficients in Figure 7. The statistically significant and negative coefficients for STs at ages 23 and 24 in Experiment A

Figure 9: Contemporaneous Effects on SCs and STs



Notes: Marginal effects on *SC/ST X East X Treat* in the specification with control variables and state fixed effects are plotted. The eastern region is compared to the southern region in the triple difference estimates. Bars represent 90% confidence intervals.

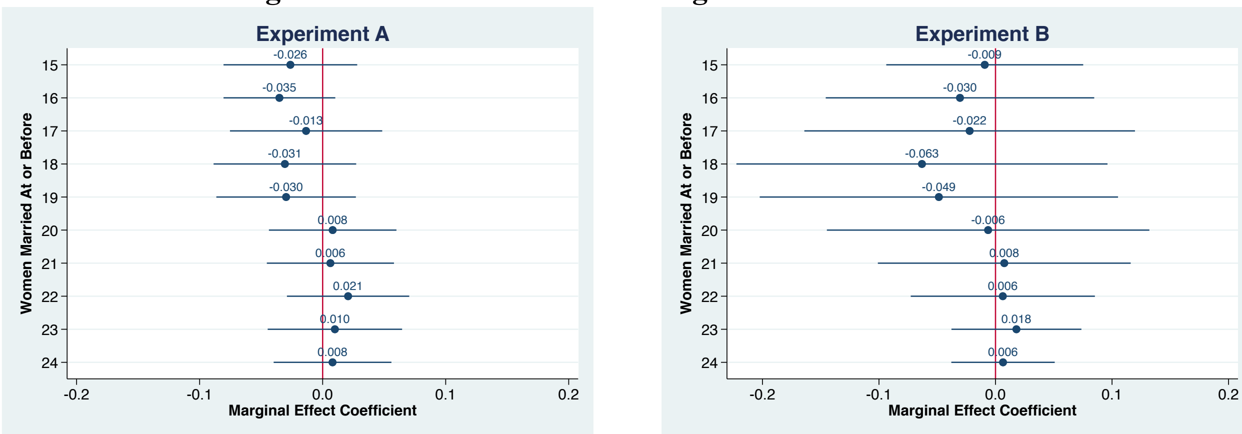
are concerning. However, taken together, the treatment effect seems to be insignificantly different than zero for the SCs and STs.

B. Falsification Test Using Two Untreated Cohorts

As another falsification test, I assume that the CEIAB is passed in 1998. I conduct the same analysis on two untreated cohorts. For this falsification test, I assume that women born between 1970 and 1975 are untreated and that women born between 1980 and 1985 are treated in both Experiments A and B[35]. I return to comparing OBCs and the General Population. The intuition behind running this falsification test was that there should be no observed treatment effect on two untreated cohorts. Figure 10 exactly corroborates this hypothesis. All estimates are close to zero,

[35] I still used the NFHS-3 for the untreated women in Experiment B for this falsification test

Figure 10: Falsification Test Using Two Untreated Cohorts



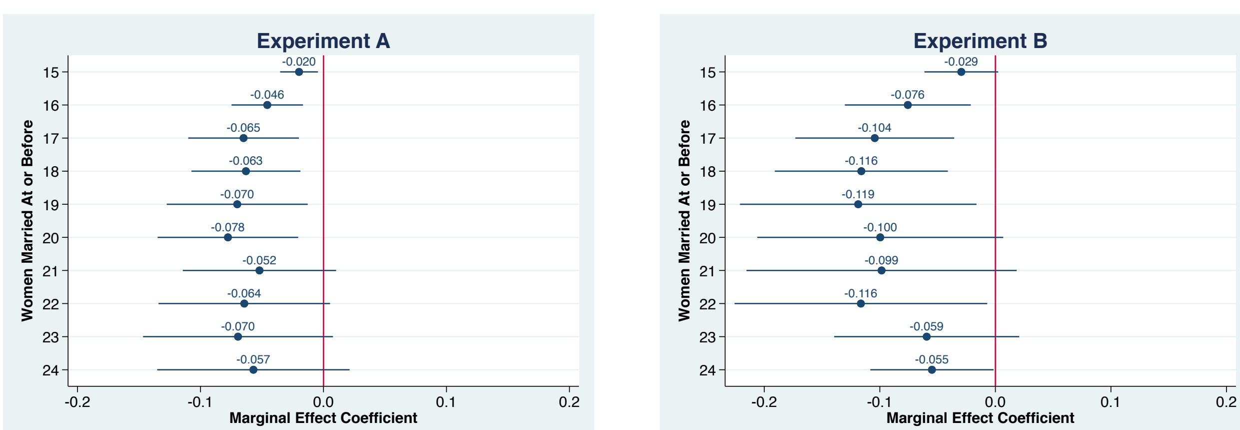
Notes: Marginal effects on *OBC X East X Treat* in the specification with control variables and state fixed effects are plotted. The eastern region is compared to the southern region in the triple difference estimates. Bars represent 90% confidence intervals.

marginally positive for later ages and statistically insignificant. These results make the treatment effect observed in Figure 7 more convincing.

C. Restricting The Sample to High School Educated Women

There remains a concern that the observed treatment effects seen in Figures 7 are due to the reduction of child marriage practices. As mentioned earlier, social reforms are a threat to identification, especially because the practice of child marriage was more prevalent amongst OBCs compared to the General Population. Hence, I restrict the sample to women that have studied at least till high school (7^{th} grade and above) to investigate if the results in Figure 7 are robust to a sample of women who are educated and who likely have low rates of child marriage. This excludes

Figure 11: Treatment Effect on High School Educated Women



Notes: Marginal effects on *OBC X East X Treat* in the specification with control variables and state fixed effects are plotted. The eastern region is compared to the southern region in the triple difference estimates. Bars represent 90% confidence intervals.

about two fifths of my sample. Looking at Figure 11, the estimates on *OBC X East X Treat* are all negative. The coefficients on the younger, college attending years, are statistically significant, while the early 20s are marginally insignificant. It is interesting to note that the magnitude of the estimates is slightly less compared to Figures 7, suggesting that the effects of discouraging child marriage could be mixed with the effects of AA. However, it is worth mentioning that any spillover effects of the policy had on uneducated women cannot be captured in this test. Overall, the estimates are robust to only including high-school educated women.

IX. Discussion & Conclusion

The results suggest that AA in higher education has benefitted the OBC caste category, the effect of which can be seen in the differential decrease in the probability of marriage at most ages between 15 and 24 for treated OBC women. The triple difference estimates in Experiment A indicate that the probability of marriage at age 19 for treated eastern OBCs decreased by 8.1 percentage points, which completely offset the pre-policy gap in the probability of marriage between OBCs and the General Population. Similar differential decreases at other ages between 15 and 24 were also observed. Furthermore, the finding that OBCs in the southern states, who were already benefitting from AA before the CEIAB, experienced no differential impact in their age at marriage makes the effect of treatment more convincing.

Firstly, the results provide evidence for fertility patterns starting earlier for lower castes compared to higher castes. This implies that lower castes tend to have higher fertility rates and lower socio-economic outcomes to begin with, justifying the government's rationale for AA. Secondly, this paper shows that AA positively impacts the OBCs. The previous literature has already shown that AA in higher education increased the participation rates and wages of targeted groups. These results provide suggestive evidence for the fact that the benefits of AA also spilled over to broader socio-economic outcomes. Therefore, it seems as if AA is effective in reducing the caste divide in India.

The results also give an idea of the rate at which AA closes the gap between the age at marriage of OBCs and the General Population. This can be visualized, even at the all-India level, in Figure 4 where in about a decade after treatment the gap between the proportion of OBC and General Population women married at age 15 and 17 practically closed. However, it must be noted that other socio-economic outcomes may not be as responsive as age at marriage to AA in higher education. Nonetheless, these results can help policy makers start to understand the time range for which AA for OBCs[36] is necessary, which could make AA less controversial in India.

It is important to remember that the results of this paper are Intent-to-Treat (ITT) effects. An ITT analysis captures the effect of being assigned to treatment but ignores the fact that some individuals assigned to treatment never got treated (Angrist and Pischke 2014). This is pertinent to the context because the ambit of OBC affirmative action remains narrow due to limited college seats and existing quotas for SCs and STs. In fact, only about two fifths of OBC women born between 1990 and 1995 (the treated cohort) from this sample had passed the 10th grade. As a result, the observed treatment effect is conservative suggesting that the impact of AA on those women that directly benefitted from quotas during the application process was even greater.

Due to the limited scope of AA, critics in India often argue that quotas are filled up by financially well off lower castes and that AA does little to help the poorest sections of society. My results suggest that AA deferred the fertility patterns of women belonging to the second through

[36] I draw a distinction between OBC affirmative action and SC and ST affirmative action because SCs and STs face a particular stigmatization on account of their untouchability that goes beyond the economic and social marginalization faced by the OBCs. This can be seen by the fact that AA for SCs and STs has been implemented since the 1950s but these groups still have the worse socio-economic outcomes as seen in the Descriptive Statistics table.

fifth quintiles of the wealth distribution, with the strongest differential effect for women in the fourth quintile. However, the findings evince that OBCs in the poorest quintile were mostly unaffected. This has implications for future policy decisions because it is indicative of AA not being enough to break the vicious cycle of disadvantage caused via discrimination. AA can break it at the upper end by making it easier for OBCs to access coveted college seats, but anti-poverty and social programs need to complement AA to break the cycle at the lower end (Deshpande 2013).

The demonstrated effectiveness of AA coupled with its limited scope calls for the extension of quotas to the private sector. As AA has historically only been applicable to the public sector and has been characterized by poor implementation and a lack of political will, these results suggest that the time is right for the burgeoning Indian private sector to take on additional social responsibility and lead the way in fostering caste equality.

A major limitation of this study is that only a short period of time had passed between the implementation of the CEIAB in 2008 and the collection of the NFHS-4 in 2015. This results in a very narrow treatment window and leads to the long-term outcomes, including eventual fertility, of treated women remaining unobserved. The narrow treatment window makes age at marriage and first birth the primary outcomes of interest, which is another shortcoming since these fertility patterns could have been deferred only due to enrolment, leaving the eventual fertility of treated women unchanged. The emphasis on fertility rates also inadvertently excludes men from the analysis. Therefore, this study is not a final evaluation of the efficacy of the policy. Rather, this study can be viewed as the mid-term effects of the CEIAB and a similar methodology, using eventual fertility and other broader socio-economic outcomes, can be used in the future to understand the policy’s long-term impact on its intended beneficiaries.

Productivity of lower castes, the mismatch hypothesis and the cost-benefits of AA remain beyond the scope of this paper and have been addressed by the previous literature. However, the empirical evidence that demonstrates the effectiveness of AA raises the question if these arguments are relevant in the Indian context, given that the primary aim of redressing the effects of past discrimination is being achieved by AA.

In conclusion, the results of this paper show the effectiveness of state-sponsored AA in higher education in closing the gap between the age at marriage and age at first birth between OBCs and the General Population. Delayed fertility patterns for OBC women provide suggestive evidence of the efficacy of the CEIAB in reducing the socio-economic divide between lower and upper castes. As OBCs that belong to the lowest quintile of the wealth distribution remain largely unaffected by the CEIAB, policy makers concerned with improving the socio-economic outcomes of lower castes may want to expand the scope of AA, beyond the public sector, to the private sector. Given the long history of the discriminatory caste system, this paper provides empirical evidence of the importance of AA policies like the CEIAB in helping India move towards the goal of a casteless society.

X. References

Abarcar, Paolo, and Caroline Theoharides. "Medical Worker Migration and Origin-Country Human Capital: Evidence from U.S. Visa." *Working Paper,* (March, 2021).

Abeberese, Ama Baafra, Todd J. Kumler, and Leigh L. Linden. "Improving Reading Skills by Encouraging Children to Read in School: A Randomized Evaluation of the Sa Aklat Sisikat Reading Program in the Philippines." *Journal of Human Resources* 49, no. 3 (2014): 611–33.

Amonker, Ravindra G, and Gary Brinker. "Reducing Fertility in India." *International Journal Of Sociology Of The Family* 33, no. 2 (2007): 327 – 348.

Angrist, Joshua David, and Jörn-Steffen Pischke. *Mastering 'metrics: The Path from Cause to Effect*. Princeton; Oxford: Princeton University Press, 2015.

Bagde, Surendrakumar, Dennis Epple, and Lowell Taylor. "Does Affirmative Action Work? Caste, Gender, College Quality, and Academic Success in India." *American Economic Review* 106, no. 6 (June 1, 2016): 1495–1521. https://doi.org/10.1257/aer.20140783.

Bains, Karan. "Delayed Gratification and Parental Decision-Making." *Submitted to the Department of Economics at Amherst College* (April 16, 2014).

Basant, Rakesh, and Gitanjali Sen. "Quota-Based Affirmative Action in Higher Education: Impact on Other Backward Classes in India." *The Journal of Development Studies* 56, no. 2 (February 1, 2020): 336–60. https://doi.org/10.1080/00220388.2019.1573987.

Bayly, Susan. "State policy and 'reservations': The politicisation of caste-based social welfare schemes." In *Caste, society and politics in India from the eighteenth century to the modern age*, by Susan Bayly, 266-305. Cambridge: Cambridge University Press , 1999.

Bertrand, Marianne, Rema Hanna, and Sendhil Mullainathan. "Affirmative Action in Education: Evidence from Engineering College Admissions in India." *Journal of Public Economics* 94, no. 1 (February 1, 2010): 16–29. https://doi.org/10.1016/j.jpubeco.2009.11.003.

Bloom, David E., and P. H. Reddy. "Age Patterns of Women at Marriage, Cohabitation, and First Birth in India." *Demography* 23, no. 4 (November 1, 1986): 509–23.

Cassan, Guilhem. "Affirmative Action, Education and Gender: Evidence from India." *Journal of Development Economics* 136, no. C (2019): 51–70.

Chauhan, Chandra Pal Singh. "Education and Caste in India." *Asia Pacific Journal of Education* 28, no. 3 (September 2008): 217–34. https://doi.org/10.1080/02188790802267332.

Chhetri, Durga P. "Politics of Social Inclusion and Affirmative Action: Case of India," *The Indian Journal of Political Science* 73, no. 4 (December, 2012): 587 – 600.

Desai, Sonalde, and Veena Kulkarni. "Changing Educational Inequalities in India in the Context of Affirmative Action." *Demography* 45, no. 2 (May 1, 2008): 245–70.

Deshpande, Ashwini. *Affirmative Action In India.* New Delhi: Oxford University Press, 2013.

Deshpande, Ashwini, and Thomas E. Weisskopf. "Does Affirmative Action Reduce Productivity? A Case Study of the Indian Railways." *World Development* 64 (December 2014): 169–80.

Duflo, Esther. "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment." *The American Economic Review* 91, no. 4, (2001): 795–813.

Frisancho, Veronica, and Kala Krishna. "Affirmative Action in Higher Education in India: Targeting, Catch up, and Mismatch." *Higher Education* 71, no. 5 (May 2016): 611–49.

Gandotra, M M, Robert D Retherford, Arvind Pandey, Norman Y Luther, and Vinod K Mishra. "Fertility in India," *National Family Health Survey Subject Reports,* no. 9 (May 1998): 1–70.

International Institute for Population Sciences – IIPS/India and ICF/ 2017. National Family Health Survey NFHS-4, 2015-16: India. Mumbai:IIPS.

International Institute for Population Sciences – IIPS/India and Macro International. 2007. National Family Health Survey NFHS-3, 2005-06: India: Volume1. Mumbai, India: IIPS.

Jayal, Niraja Gopal. "Affirmative Action in India: Before and after the Neo-Liberal Turn." *Cultural Dynamics* 27, no. 1 (March 2015): 117–33.

Jr, Roland G Fryer, and Glenn C Loury. "Affirmative Action and Its Mythology," *Journal of Economic Perspectives* 19, no. 3 (2005): 147 – 162.

Khanna, Gaurav. "Does Affirmative Action Incentivize Schooling? Evidence from India." *The Review of Economics and Statistics* 102, no. 2 (May 2020): 219–33.

Muralidharan, Karthik, and Nishith Prakash. "Cycling to School: Increasing Secondary School Enrollment for Girls in India," *American Economic Journal* 9, no. 3 (July, 2017): 321 – 350.

Shrestha, Slesh A. "No Man Left Behind: Effects of Emigration Prospects on Educational and Labour Outcomes of Non-Migrants." *The Economic Journal* 127, no. 600 (March 2017): 495–521. https://doi.org/10.1111/ecoj.12306.

Young, Emily. "The Effects of School Finance Reform on Input Spending and Student Achievement." *Submitted to the Department of Economics at Amherst College* (April 2020).