Castes and Labor Mobility*

Viktoria Hnatkovska†, Amartya Lahiri†, and Sourabh B. Paul†

June 2011

Abstract

Can large macroeconomic changes also alter the historical economic mobility patterns of various social groups? We examine this question by contrasting the relative fortunes of the historically disadvantaged scheduled castes and tribes (SC/ST) in India in terms of their education attainment, occupation choices, consumption and wages. We study the period 1983-2005 using household survey data from successive rounds of the National Sample Survey. We find that this period has been characterized by a significant convergence of education, occupation distribution, wages and consumption levels of SC/STs toward non-SC/ST levels. Using various decomposition approaches we find that the improvements in education account for a major part of the wage and consumption convergence.

JEL Classification: J6, R2

Keywords: Intergenerational mobility, wage gaps, castes

*We would like to thank two referees, the editor Esther Duflo, David Green and Thomas Lemieux for extremely helpful and detailed suggestions. Thanks also to Siwan Anderson, Nicole Fortin, Ashok Kotwal, Anand Swamy, and seminar participants at Delhi School of Economics, ISI-Delhi, Monash, Toronto, UBC and the UC Santa Barbara Growth and Development conference for comments. Hnatkovska and Lahiri thank SSHRC for research support.

†Department of Economics, University of British Columbia, 997 - 1873 East Mall, Vancouver, BC V6T 1Z1, Canada. E-mail addresses: hnatkovs@interchange.ubc.ca (Hnatkovska), alahiri@interchange.ubc.ca (Lahiri), sbikas@gmail.com (Paul).

"If there is no enthusiasm, life becomes drudgery - a mere burden to be dragged. Nothing can be achieved if there is no enthusiasm. The main reason for this lack of enthusiasm on the part of a man is that an individual loses the hope of getting an opportunity to elevate himself. Hopelessness leads to lack of enthusiasm. The mind in such cases becomes deceased...When is enthusiasm created? When one breathes an atmosphere where one is sure of getting the legitimate reward for one's labor, only then one feels enriched by enthusiasm and inspiration"

"This condition obtains even where there is no slavery in the legal sense. It is found where as in caste system, some persons are forced to carry on the prescribed callings which are not their choice..."

B. R. Ambedkar (Chief architect of the Indian Constitution.)

1 Introduction

Large macroeconomic and structural changes of the economy often go hand-in-hand with winners and losers at the level of individuals, sectors or social groups. Hence, managing the microeconomic distributional consequences of macroeconomic changes is a key challenge for policymakers. But do large-scale macroeconomic changes tend to accentuate or dampen historical inequities? Do these economic redistributions necessarily benefit the economically stronger sections of society or can they also coincide with a reduction in economic inequality? What are the key margins which account for these distributional changes?

The Indian economy provides a natural environment to investigate these questions due to a dramatic process of economic reforms, structural changes and rapid economic growth since the 1980s. Average annual GDP growth rates in India have climbed rapidly to above 8 percent during the last twenty years from the anaemic 3-3.5 percent that characterized the first 35 years since 1947. Accompanying this growth takeoff has been a hastening process of structural transformation of the economy. The agricultural sector, which historically had the largest employment and output share, has rapidly lost ground in both during this period. Such rapid structural changes often deeply affect the lives of people in these economies by redistributing income and economic opportunities from some groups to others. Hence, any assessment regarding the welfare effects of these reforms must necessarily involve an evaluation of their redistributive effects. As an example, a particularly emotive issue that has been debated energetically with regard to the Indian experience is the effect of the economic take-off on the fortunes of the poor.

In this paper we study the impact of the rapid transformation of the Indian economy on one such historically disadvantaged group: the Scheduled Castes and Scheduled Tribes (SC/STs). SC/STs were historically economically backward, mostly very poor, concentrated in low-skill (mostly agricultural) occupations and primarily rural. Moreover, they were also subject to centuries of systematic caste-based discrimination both economically and socially. Indeed, the historical tradition of social division through the caste system created a social stratification along education, occupation and income lines that has continued into modern India. In fact, this stratification was so endemic that the constitution of India aggregated these castes into a Schedule of the constitution and provided them with affirmative action cover in both education and public sector employment. This constitutional initiative was viewed as a key component of attaining the ultimate policy goal of raising the social and economic mobility of the SC/STs to the levels of the non-SC/STs. Amazingly, there has been almost no broad-based evaluation of the economic performance of SC/STs either for the reform phase since the mid-1980s or for the post-independence period since 1947. An ancillary goal of this paper is to partly fill this gap in the literature.

The existence of caste-based frictions in labor market allocations and social matching processes have been documented by a number of micro-level studies. Indeed, a key goal of the reservations policy was to make it easier for, say, the child of an illiterate SC or ST farm worker living below the poverty line to get educated and find productive employment in a better paying occupation. How have the tectonic changes in India since the early 1980s affected this goal? What has been the net effect on the fortunes of SC/STs of the interplay between these micro-level frictions and the massive aggregate macroeconomic changes in India over the past two decades? Has the rapid growth percolated down to the SC/STs in terms of tangible changes in their economic and social conditions? Is the primary reason for the economic deprivation of these underprivileged castes the types of occupations they tend to work in? Alternatively, is the key impediment the lack of education, i.e., do they get stuck in low wage jobs due to the lack of education? Or, is ongoing discrimination in occupations and wages the primary problem facing these groups? This paper attempts to answer these questions.

We use data from five successive rounds of the National Sample Survey (NSS) of India from 1983 to 2004-05 to analyze patterns of education attainment, occupation choices, wages and consumption expenditures of both SC/ST and non-SC/ST households. In particular, we contrast the time-series evolution of education, occupation choices, wages and consumption expenditures of SC/STs with their non-SC/ST counterparts both at an aggregated generation level of young (ages 16-40) and old

(ages 41-65) as well as at more disaggregated age groups.

Our analysis yields four main results. First, while SC/ST households are, on average, less educated than their non-SC/ST counterparts throughout the sample period and across age groups, the education attainment levels of SC/STs have been converging toward the levels of their non-SC/ST cohorts across all education categories and age groups. Moreover, the trend towards education convergence of the two groups emerges both in rural and urban sectors but is sharper in urban areas. We also find that *conditional* on their non-caste characteristics, the sharpest convergence between the education attainment levels of the two groups has occurred in the categories "primary school" and "middle school".

Second, there has been significant churning in the occupational distributions of SC/STs and non-SC/STs between 1983 and 2004-05. While both groups have been moving out of low skill agrarian occupations into relatively higher skill occupations, SC/STs have been doing so at a faster rate. As a result the relative proportions of the two groups in each of the occupations have been converging between 1983 and 2004-05. However, we also find that the difference in the probabilities of members of the two groups being in a given occupation have not changed uniformly across the occupation categories. While the unconditional probabilities of being in agrarian and blue-collar occupations have converged for the two groups , the corresponding probabilities of being in high-skill, white collar occupations have diverged with SC/STs becoming even less likelier to be employed in them.

Third, we find a clear and statistically significant trend of convergence of wages of the two groups. The median wage premium of non-SC/STs relative to SC/STs has declined systematically from 42 percent in 1983 to 22 percent 2004-05. The wage premium declined from 38 percent to 17 percent for the non-SC/ST young and from a whopping 70 percent to 48 percent for the old. While the gaps are typically larger for older age groups, the convergent trend is significant across both the aggregated generational age groups as well as more disaggregated age groups. We find that the overall wage convergence between the groups has been driven primarily by a convergence in the covariates of wages, in particular a convergence in the education and occupation choices of the two groups. In fact, attributes orthogonal to caste account for almost all of the relative wage movements of the two groups. To put these wage dynamics in perspective, the median white male to black male wage premium in the US has hovered stubbornly around 30 percent over the past 35 years, which makes the SC/ST relative wage behavior in India even more striking.

Fourth, the household consumption expenditures of the two groups have also been converging significantly during this period. In contrast to the wage convergence patterns however, we find that

the consumption convergence was driven by both a convergence in the covariates of consumption as well as a decline in the magnitude of the cast specific penalty for consumption. Put differently, both attributes and the consumption structure have played independently important roles in accounting for the relative changes in the consumption expenditures of the two groups.

An independent issue of interest is whether the past three decades have affected all sections of SC/STs the same or has the period tended to favor only some amongst them. This issue has been the subject of debate with concerns that only the relatively more affluent amongst the backward castes (often called the "creamy layer") take advantage of new opportunities opening up in the economy while the vast majority of the SC/STs continue to stagnate. This issue, combined with our findings that the convergence patterns in the education and occupation distributions of the two groups have been particularly strong in specific categories, led us to examine the convergence patterns in wages and consumption not just for the mean or the median but rather along the entire distribution. We do this by using multiple decomposition methods for quantile wages and consumption including the DiNardo, Fortin, and Lemieux (1996) aggregate decomposition technique as well as Recentered Influence Function quantile regressions (Firpo, Fortin, and Lemieux (2009)). Contrary to the conventional view that only the richest amongst the disadvantaged castes benefited, we find that there was convergence in wages and consumption for the majority of the distribution. In fact, if there is any widening of the gaps it was in the top-20 percentiles of households.

A common feature of the dynamics in the different indicators is that most of the overall convergence across the indicators is due to a convergence in attributes. However, for some sub-groups, attributes have at times converged more than the outcomes for some indicators indicating that significant caste specific obstacles continue to exist at least for some segments of SC/STs. This is often the case for those with relatively greater education, working in white collar jobs and earning higher incomes. We find this intriguing since the conventional view is that the relatively better-off amongst the SC/STs may have been better positioned to benefit from the undergoing changes in the economy.

In summary, these results suggest some uplifting answers to the questions we set out to answer. In the twenty-two year period since 1983, SC/STs have sharply narrowed their historical gaps in education attainment levels, occupation choices, wages and consumption levels relative to non-SC/STs. Moreover, the convergence in wages and consumption have been sharpest amongst the lower percentile of households. Overall, we conclude that neither the lack of occupational mobility nor the lack of education have been a major impediment toward the SC/STs taking advantage of the rapid structural changes in India during this period to rapidly narrow their huge historical economic disparities

with non-SC/STs.

To the best of our knowledge, our's is the first study to jointly analyze caste differences in education, occupation, wage and consumption outcomes in a single study, track the time series evolution of these outcomes, and do so using data that covers the entire country. It is worth reiterating that we do this using the NSS data which has the broadest coverage for India both spatially and over time.

There exists a large literature which has investigated the existence and extent of labor market discrimination in India. Amongst others, Banerjee and Knight (1985) and Madheswaran and Attewell (2007) have studied the extent of wage discrimination faced by SC/STs in the urban Indian labor market. Borooah (2005) has studied the extent of discrimination in employment in the urban labor market. Ito (2009) studies both wage and employment discrimination simultaneously by examining data from two Indian states – Bihar and Uttar Pradesh. Kijima (2006) uses NSS data to study consumption inequality of SC/ST households.[1] Our study differs from these in that we examine the data for all states and for both rural and urban areas. Moreover, as opposed to most of these studies, we analyze education, occupation, consumption and wage outcomes jointly. Lastly, by using data for five rounds of the National Sample Survey of households we are also able to provide a time series perspective on the evolution of SC/ST fortunes in India, a feature that other studies have typically not examined.

In the next section we describe the data and our constructed measures as well as some summary statistics. Section 3 contrasts SC/STs with their non-SC/ST cohorts in terms of the evolution of the distributions of education attainment rates, occupations, wages and consumption expenditures. There we also explore the gender differences in these dynamics. The last section concludes.

2 The Data

Our data comes from the National Sample Survey (NSS) of India and its various rounds. In particular, we use the NSS Rounds 38 (1983), 43 (1987-88), 50 (1993-94), 55 (1999-2000) and 61 (2004-05). The survey covers the whole country except for a few remote and inaccessible pockets. The rounds that we use include detailed information on over 120,000 households and 600,000 individuals. Our working sample consists of all individuals between the ages of 16 and 65 belonging to male-headed households. The sample is restricted to those individuals who provided their 3-digit occupation code information

[1] In related work Munshi and Rosenzweig (2009) document the lack of labor mobility in India. Also, Munshi and Rosenzweig (2006) show how caste-based network effects affect education choices by gender.

and their education information.[2] Our focus is on full-time working individuals who are defined as those that worked at least 2.5 days per week, and who are not currently enrolled in any education institution. This selection leaves us with a working sample of around 165,000-182,000 individuals, depending on the survey round.

Data on wages are more limited. The sub-sample with complete wage data consists of, on average across rounds, about 65,000 individuals which is considerably smaller than our working sample but large enough to facilitate formal analysis. Wages are obtained as the daily wage/salaried income received for the work done by respondents during the previous week (relative to the survey week). Wages can be paid in cash or kind, where the latter are evaluated at the current retail prices. We convert wages into real terms using state-level poverty lines that differ for rural and urban sectors. We express all wages in 1983 Maharashtra prices. Given the limited availability of wage data, we also study consumption expenditures. This variable, however, is only available at the household level. Nevertheless, it allows us to check the robustness of our findings for wages. Our consumption expenditure sample consists of about 105,000 observations per round. We convert consumption expenditure into per capita real terms to make it comparable with the wage data. Details regarding the dataset are contained in the Appendix A.1.

Our education variable contains many categories: not-literate; literate but below primary; primary education; middle education; secondary, higher secondary, diploma/certificate course, graduate and above, postgraduate and above. In the summary statistics table below as well as in some parts of our formal data analysis we convert these educational categories into years of education by using the following mapping: not-literate = 0 years; literate but below primary = 2 years; primary = 5 years; middle = 8 years; secondary and higher secondary = 10 years; graduate = 15 years; post-graduate = 17 years. Diplomas are treated similarly depending on the specifics of the attainment level.[3] To simplify the exposition when we present statistics on the categories themselves, we group them into 5 broader categories: not-literate; literate but below primary; primary; middle; and secondary and above education. These categories are coded as education categories 1, 2, 3, 4 and 5 respectively. Details on education categories and their conversion into education years is contained in the Appendix A.1.

Table 1 gives some summary statistics of the data. Panel (a) reports average age, education

[2] We also consider a narrower sample in which we restrict the sample to only males and find that our results remain robust.

[3] We are forced to combine secondary and higher secondary into a combined group of 10 years because the higher secondary classification is missing in the 38th and 43rd rounds. The only way to retain comparability across rounds then is to combine the two categories.

level in years, share of married and male individuals among the young workers; while panel (b) reports the corresponding statistics for the old. Panel (c) report the percentage of rural households and the average household size in the sample. Note that "All" refers to the full sample, while the "Non-SC/ST" and "SC/ST" panels refer to the corresponding sub-samples.

Amongst other features, the table indicates that about 78 percent of the surveyed households are rural. This number is slightly higher for SC/ST households, around 87 percent of whom live in rural areas on average. The working members of the surveyed households are mostly married males. The average age of the "old" (ages 41-65) is around 51 years, while the young are typically around 28 years old. The average education level of the young is consistently greater than that of the old for both SC/STs and non-SC/STs, and has increased over time. Non-SC/STs are also consistently more educated than SC/ST. The proportion of SC/ST households in the sample across the different rounds is around 24 percent.

Table 1: Sample summary statistics

	(a) Young (16-40)				**(b) Old (41-65)**				**(c) Households**	
ALL	age	edu	% married	% male	age	edu	% married	% male	% rural	hh size
1983	28.14	3.35	0.78	0.77	50.84	2.28	0.82	0.89	0.78	6.07
	(0.03)	(0.02)	(0.00)	(0.00)	(0.04)	(0.02)	(0.00)	(0.00)	(0.00)	(0.02)
1987-88	28.37	3.51	0.77	0.78	50.89	2.56	0.82	0.90	0.79	5.90
	(0.03)	(0.02)	(0.00)	(0.00)	(0.03)	(0.02)	(0.00)	(0.00)	(0.00)	(0.01)
1993-94	28.62	4.19	0.78	0.76	50.86	3.19	0.81	0.91	0.78	5.54
	(0.03)	(0.02)	(0.00)	(0.00)	(0.04)	(0.02)	(0.00)	(0.00)	(0.00)	(0.01)
1999-00	29.07	4.69	0.75	0.76	50.58	3.67	0.79	0.92	0.77	5.65
	(0.03)	(0.02)	(0.00)	(0.00)	(0.04)	(0.03)	(0.00)	(0.00)	(0.00)	(0.02)
2004-05	29.2	5.27	0.77	0.75	50.51	4.09	0.80	0.93	0.76	5.45
	(0.03)	(0.02)	(0.00)	(0.00)	(0.04)	(0.03)	(0.00)	(0.00)	(0.00)	(0.02)
Non-SC/ST										
1983	28.19	4.01	0.80	0.75	50.91	2.78	0.84	0.89	0.75	6.20
	(0.03)	(0.02)	(0.00)	(0.00)	(0.04)	(0.03)	(0.00)	(0.00)	(0.00)	(0.02)
1987-88	28.46	4.19	0.80	0.76	50.91	3.10	0.84	0.90	0.76	6.01
	(0.03)	(0.02)	(0.00)	(0.00)	(0.04)	(0.02)	(0.00)	(0.00)	(0.00)	(0.02)
1993-94	28.68	4.93	0.8	0.75	50.87	3.85	0.83	0.92	0.74	5.64
	(0.03)	(0.02)	(0.00)	(0.00)	(0.04)	(0.03)	(0.00)	(0.00)	(0.00)	(0.02)
1999-00	29.18	5.47	0.78	0.75	50.63	4.40	0.81	0.93	0.73	5.72
	(0.04)	(0.03)	(0.00)	(0.00)	(0.04)	(0.03)	(0.00)	(0.00)	(0.00)	(0.02)
2004-05	29.3	5.96	0.79	0.74	50.57	4.78	0.81	0.93	0.73	5.47
	(0.04)	(0.03)	(0.00)	(0.00)	(0.05)	(0.04)	(0.00)	(0.00)	(0.00)	(0.02)
SC/ST										
1983	28.02	1.64	0.71	0.80	50.63	0.86	0.76	0.88	0.87	5.72
	(0.05)	(0.02)	(0.00)	(0.00)	(0.07)	(0.02)	(0.00)	(0.00)	(0.01)	(0.04)
1987-88	28.14	1.81	0.72	0.82	50.82	1.00	0.76	0.89	0.89	5.57
	(0.05)	(0.02)	(0.00)	(0.00)	(0.07)	(0.02)	(0.00)	(0.00)	(0.00)	(0.03)
1993-94	28.46	2.39	0.73	0.79	50.84	1.37	0.77	0.90	0.88	5.29
	(0.05)	(0.03)	(0.00)	(0.00)	(0.07)	(0.03)	(0.00)	(0.00)	(0.00)	(0.03)
1999-00	28.84	2.98	0.69	0.79	50.48	1.89	0.73	0.91	0.86	5.47
	(0.05)	(0.03)	(0.00)	(0.00)	(0.08)	(0.04)	(0.01)	(0.00)	(0.00)	(0.03)
2004-05	28.97	3.64	0.73	0.77	50.34	2.26	0.76	0.92	0.85	5.39
	(0.06)	(0.03)	(0.00)	(0.00)	(0.08)	(0.04)	(0.01)	(0.00)	(0.00)	(0.03)

Notes: This table reports summary statistics for our sample. Panel (a) gives the statistics for the young (ages 16-40), while panel (b) gives the statistics for the old (ages 41-65). Standard errors are reported in parenthesis.

3 How Have the Scheduled Castes Fared?

We start our analysis by comparing SC/STs with non-SC/STs both overall and across age groups. We construct age groups in each round by splitting the sample into ages 16-40 and 41-65. We call these generational groups "Young" and "Old", respectively. For each group we compute the occupation distribution, the average education attainment level, the average daily wage, and daily per capita household consumption expenditure earned for the entire sample as well as for SC/STs and non-SC/STs separately. Issues of particular interest to us are: (a) whether the education attainment levels of the SC/ST's are converging to the levels of their non-SC/ST cohorts? (b) whether their occupation choices are converging over time; (c) whether wages and consumption levels of SC/STs are converging to non-SC/ST levels.

3.1 Education Attainment

We start with the record on education attainment rates. Table 2 shows the average years of education of the overall population as well as those for non-SC/ST and SC/STs separately. In 1983, the average years of education of non-SC/STs was 3.62 relative to 1.41 years for SC/STs – a 157 percent relative discrepancy. However, over the sample period, there was a clear trend towards convergence in education levels of SC/STs toward their non-SC/ST counterparts as the gap declined to just 74 percent by 2004-05.[4] The table also shows that most of the relative pick-up in the SC/ST education attainment has occurred since the 1990's. Hence, the rate of convergence during the sample period has not followed a monotonic time path.

Table 2: Education Gap: Years of Schooling

	Average years of education		
	Overall	**Non-SC/ST**	**SC/ST**
1983	3.02	3.62	1.41
	(0.01)	(0.02)	(0.02)
1987-88	3.21	3.84	1.58
	(0.01)	(0.02)	(0.02)
1993-94	3.86	4.58	2.08
	(0.01)	(0.02)	(0.02)
1999-2000	4.36	5.12	2.64
	(0.02)	(0.02)	(0.02)
2004-05	4.87	5.55	3.19
	(0.02)	(0.02)	(0.03)

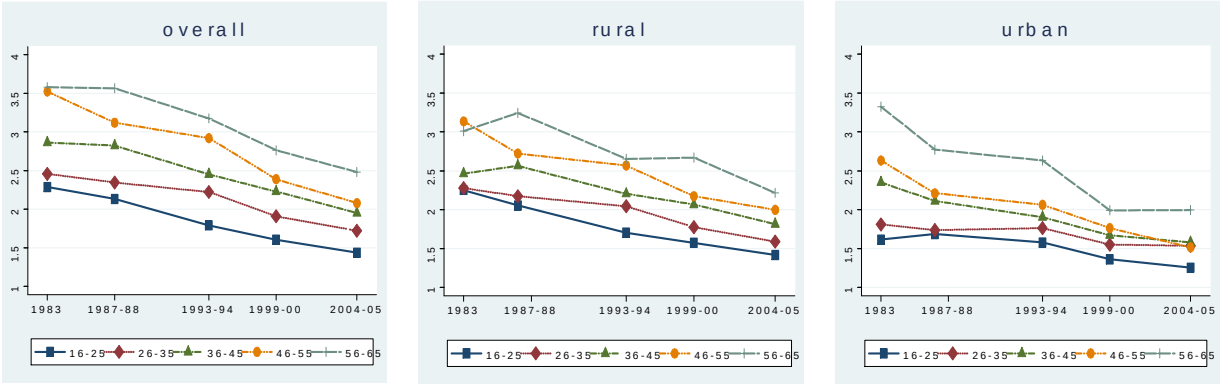
Notes: This table presents the average years if education for the overall sample and separately for non-SC/STs and SC/STs. The reported statistics are obtained for each NSS survey round which is shown in the first column. Standard errors are in parenthesis.

[4] We should note this trend was strong for both the young and the old. For the young the difference declined from 144 percent to just 64 percent while for the old it went from 225 percent to 112 percent during this period. Importantly, both groups increased their education attainment levels over the period with the SC/STs levels rising faster.

Do the overall trends in years of education mask key differences in the relative movements within more disaggregated age cohorts? To investigate this we compute the average years of education of non-SC/STs relative to SC/STs within five age-cohorts for each survey round. Figure 1 plots the result. Panel (a) reveals a clear pattern of education convergence across the different age-cohorts over time. Importantly, for all the rounds, the older the cohort the larger the education gap.

Given the large concentration of households in rural areas, a related question is whether the trends in education attainment are different between rural and urban households. To address this, we split the age-cohorts into rural and urban households and then plot the gap in the average years of education for the two sectors separately in panels (b) and (c) of Figure 1. Three features of the figure are noteworthy. First, the variation in years of education across cohorts is somewhat smaller in rural areas than in urban areas. Second, the years of education have been converging in both rural and urban areas. Third, the convergence rates are, on average, faster in urban areas especially for the older cohorts.[5]

Figure 1: Education gaps by age cohorts



(a) (b) (c)

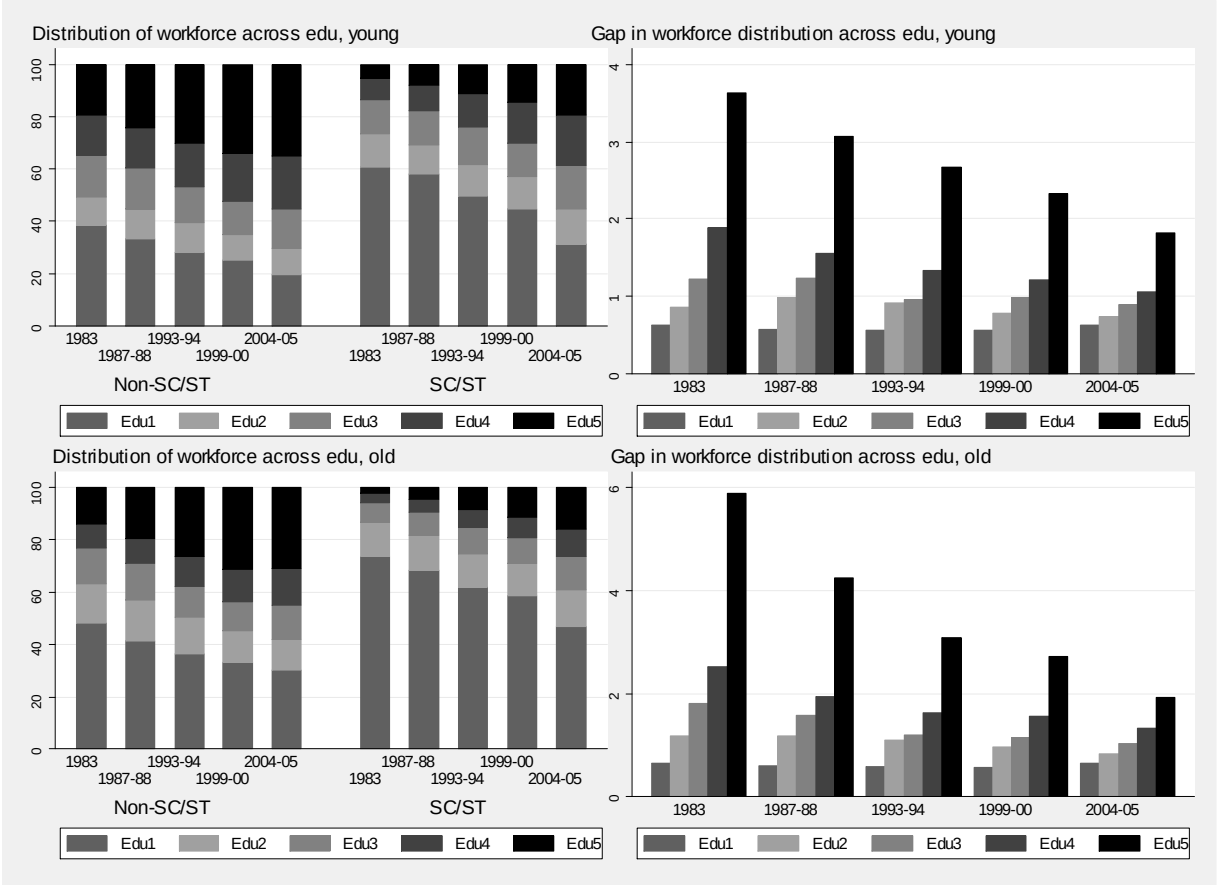
Notes: The figures show the evolution of the relative gap in years of education between non-SC/STs and SC/STs over time for different age groups. Panel (a) presents the results for the overall sample, while panels (b) and (c) report the results for rural and urban households separately.

While the changes in the years of education are interesting, they mask potentially interesting changes in the composition of education attainment levels underlying the aggregate change in years of education. Thus, is the change in the average years of education due to more illiterates going to primary school or is it primarily due to more people going on to middle school or higher? We

[5]The education attainment gaps of the "old" cohort can change over successive rounds for two reasons. First, as some of the "young" become "old" in subsequent rounds, the education composition of the "old" will clearly change. Second, since 1951 India has introduced a series of literacy initiatives (such the National Literacy Mission) with a special focus on adult literacy. In as much as these programs had a positive effect on adult literacy, the education composition of the old would change over time due to them as well.

answer this question using Figure 2. Given the similar patterns across the age groups and for ease of presentation we report the statistics on the more aggregated age groups of young (ages 16-41) and old (ages 41-65).

Figure 2: Education distribution of children and parents



(a) (b)

Notes: Panel (a) of this figure presents the distribution of the workforce of young and old cohorts across five education categories across different NSS rounds. The left set of bars on each figure refers to non-SC/STs, while the right set is for SC/STs. Panel (b) presents gaps in the distribution of non-SC/STs relative to SC/STs across five education categories. The gaps are also reported for young and old. See the text for the description of how education categories are defined (category 1 is the lowest education level - illiterate).

Panel (a) of Figure 2 shows the distribution of the workforce across 5 broad education categories and the evolution of the distribution over the successive rounds. The top graph of Panel (a) shows the distribution of non-SC/ST young (left set of bars) and the corresponding distribution of SC/ST young (right set of bars). The bottom graph of Panel (a) reports the same distributions for the old. It is clear that SC/ST young are systematically less educated than their non-SC/ST counterparts. The difference is most glaring in the lowest and highest categories. In category 1 (the illiterate groups),

SC/STs are hugely over-represented while in category 5 (secondary education or above) they are strongly under-represented. The scale of the lack of education in India, both in general and amongst SC/STs, is probably best summarized by the fact that as recently as in 1983, about 72 percent of SC/ST young were either illiterate or had below primary level education while the corresponding number for non-SC/STs was 49 percent. These numbers declined to 42 percent for SC/ST young and 30 percent for non-SC/ST young by 2004-05.

The figure also makes clear that there has been a sustained decrease over time in the share of illiterates amongst both SC/STs and non-SC/STs. Between 1983 and 2004-05, the proportion of young illiterate SC/ST (category 1) fell from 60 percent to 30 percent while for non-SC/STs the proportion of illiterates fell from just under 40 percent to 20 percent. These were by far the sharpest changes amongst all education categories for either group. The largest increases for SC/STs occurred in education categories 4 (middle school) and 5 (secondary or above) while for non-SC/STs they occurred in the secondary or above education category 5.

Panel (b) of Figure 2 reports the relative gaps in education distribution for the two social groups. The top graph of Panel (b) shows the numbers for the young, while the bottom graph does the same for the old. The bar for any education category j is obtained by dividing the non-SC/STs in category j as a share of all non-SC/STs by the SC/STs in category j as a share of all SC/STs. The deviation of the height of the bar for any category j from one then indicates the degree to which there is a disproportionate presence of one group in that category. Thus, the first bar from the left in the top graph of Panel (b) shows that the percentage of all SC/ST young who belonged to education category 1 in 1983 exceeded the corresponding percentage of non-SC/ST young in category 1 in that year by over 20 percentage points, i.e., SC/STs were over-represented in education category 1 (relative to their overall numbers). Over time, a movement of the heights of the bars toward one would indicate a trend towards convergence in the education distribution across the two groups.

The top graph of panel (b) of Figure 2 shows that, without exception, the differences in the proportion of young in the different education categories either stayed constant or tended towards convergence for the two groups. Perhaps the sharpest trend toward convergence has been in category 5 (secondary or higher). In 1983 about 19 percent of non-SC/ST young had secondary school or higher levels of education while the number was just around 5 percent for SC/STs. By 2004-05, 35 percent of non-SC/ST young had secondary or higher levels of education while the number for SC/STs had risen to 19 percent. Clearly, for both groups there has been an increase in the share of young with secondary or higher education. However, the absolute difference between the two groups

still remains very high.

The bottom panels of Figure 2 show the same information for the old. There are a few key differences between the education distribution patterns of the old and the young. First, the share of illiterates and those with less than primary education (education categories 1 and 2) is higher for both SC/ST and non-SC/ST old throughout and declined at a slower rate than that of the young. Thus, amongst the old in 1983, the combined share of categories 1 and 2 was 62 percent for non-SC/STs and 85 percent for SC/STs. These numbers fell over time but still remained at a very high 41 percent and 60 percent, respectively, in 2004-05. Second, at the high end of the education distribution the changes have been much more tepid for the old of both groups relative to that of the young. The share of those with secondary education or higher amongst the non-SC/ST old rose from 13 percent in 1983 to 30 percent in 2004-05. Correspondingly, the share for the SC/ST old rose from 1 percent to 17 percent. One should note though that the convergence patterns are there for the old as well, as the bars in the bottom graph of panel (b) have all moved towards one over time.

Figure 2 has revealed a clear trend towards convergence of the educational composition of SC/STs and non-SC/STs. Is this convergence statistically significant? In order to examine this issue, we estimated ordered probit regressions of the broad education categories 1 to 5 on an SC/ST dummy for each sample round. It is important to note that these ordered probit regressions estimate the absolute gaps in the probabilities of the different education categories between SC/STs and non-SC/STs. The bars that were shown in Panel (b) of Figure 2, on the other hand, depicted the relative gaps in education distribution for the two social groups. We believe both the absolute and the relative gaps are informative and hence have chosen to examine both.

The results of the ordered probit regressions are shown in 3. Panel 1a of the Table reports the marginal effect of the SC/ST dummy on the probability of an individual belonging to each education category. We call these the unconditional effects of the SC/ST dummy since there are no other conditioning variables included in the regressions. Panel 1b reports the changes in those marginal effects over the relevant sample period. There are two features of the results to note. First, SC/STs are more likely to be illiterate (education category 1), and are less likely to be in the higher education categories relative to non-SC/STs. This remained true across all the rounds. Second, there is a significant trend towards convergence in education attainments of the two groups over time in all, but category 5. In particular, the marginal effects of SC/ST dummy have declined in absolute value in categories 1 to 4, indicating lower education gaps between the two social groups in those categories. This decline is especially pronounced in categories 2 and 3 (below primary and primary).

Only in the secondary education and above category (category 5) has the gap between SC/STs and non-SC/STs widened over the sample period. [6]

A second question of interest to us is whether the observed convergence in the education distribution due to a convergence in the covariates of education or has there been convergence in education levels between the castes even after controlling for the covariates? To examine this we ran the ordered probit regressions of education categories on an expanded list of controls including age, age squared, an SC/ST dummy, region dummies, a rural dummy, a muslim dummy, and a state-specific SC/ST quota variable. We then tested whether the estimated coefficient on the SC/ST dummy changed significantly across the sample rounds.[7]

By including state-level SC/ST reservation quotas, we aim to control for differences in reservation policies across states. The introduction of reservations for SC/STs in public sector employment and in higher education institutions was a key policy initiative in India. The reservations were provided in proportion to the population shares of SCs and STs. Due to their explicit purpose of redressing the historical inequities against SCs and STs, reservations have been a focus of attention for a number of researchers. Thus, Pande (2003) examines the effects of reservations on actual policies while Prakash (2009) studies the effects of reservations on the labor market outcomes of SC/STs. Both authors find evidence of positive effects of reservations on the targeted communities. Hence, it is important to control for state level reservation quotas in the regressions.[8]

We also include a Muslim dummy in our regression specification. This is intended as a control for the fact that Muslims, on average, have done poorly in modern India (post independence in 1947). If we do not control for a Muslim fixed factor explicitly, then part of the measured catch-up of SC/STs that we find in the data may be attributed to the poor performance of Muslims who would be assigned into non-SC/ST group.

We control for regional differences by grouping states into six regions – North, South, East, West, Central and North-East – and include region dummies in the regression specification.[9] In combination with the state-level reservation policy, this allows us to decompose state-level differences into those

[6] The detailed estimation results from this regression and the ones that will follow are available in an online supplement from http://faculty.arts.ubc.ca/vhnatkovska/Research/castes-supplement.pdf

[7] We also ran ordered logit regressions and obtained qualitatively similar results to the ordered probit results reported below.

[8] State-level reservations can change over time due to changes in SC/ST population shares. In 1991 the Indian government extended the reservation policy to include other backward castes (OBCs). In our analysis we focus only on the group of SC/STs while OBCs are included in the non-SC/ST reference group. If reservations benefited OBCs then our results potentially understate the true degree of convergence between SC/STs and non-SC/STs (excluding OBCs), especially since the extension of reservations to OBCs in 1991.

[9] This grouping reflects similarities across states along their geographic characteristics, and characteristics that are shared based on proximity.

attributable to reservations policy, and those due to other time-invariant factors that are common to all states within a given region. The identifying assumption behind this strategy is that the states within a region are broadly similar but differ in terms of the reservation quota they implement.

Table 3 Panels 2a and 2b shows the marginal effects on the SC/ST dummy from those regressions for each survey round, as well as the changes in the SC/ST marginal effects over the successive decades and over the entire sample period. The results for the conditional probit regressions are qualitatively the same as the unconditional regressions (Panels 1a and 1b). In particular, the results suggest that for all except the highest education categories, there has been significant convergence in the education distribution of the two groups over the sample period.

Table 3: Marginal Effect of SC/ST dummy in ordered probit regression for education categories

	Panel 1a: Marginal effects, unconditional					Panel 1b: Changes		
	1983	1987-88	1993-94	1999-2000	2004-05	83 to 93	93 to 05	83 to 05
Edu 1	0.2650	0.2688	0.2622	0.2410	0.2112	-0.0029	-0.0509***	-0.0538***
	(0.0031)	(0.0028)	(0.0030)	(0.0034)	(0.0035)	(0.0043)	(0.0047)	(0.0047)
Edu 2	-0.0308	-0.0281	-0.0158	-0.0040	0.0072	0.0150***	0.0230***	0.0380***
	(0.0006)	(0.0006)	(0.0005)	(0.0003)	(0.0003)	(0.0008)	(0.0005)	(0.0007)
Edu 3	-0.0619	-0.0587	-0.0405	-0.0251	-0.0167	0.0214***	0.0238***	0.0453***
	(0.0010)	(0.0008)	(0.0007)	(0.0006)	(0.0005)	(0.0012)	(0.0009)	(0.0011)
Edu 4	-0.0687	-0.0646	-0.0645	-0.0587	-0.0529	0.0042***	0.0115***	0.0158***
	(0.0010)	(0.0009)	(0.0009)	(0.0010)	(0.0011)	(0.0014)	(0.0014)	(0.0014)
Edu 5	-0.1036	-0.1174	-0.1414	-0.1531	-0.1488	-0.0378***	-0.0074***	-0.0452***
	(0.0012)	(0.0012)	(0.0015)	(0.0020)	(0.0023)	(0.0020)	(0.0027)	(0.0026)
	Panel 2a: Marginal effects, conditional					Panel 2b: Changes		
Edu 1	0.2556	0.2626	0.2540	0.2331	0.2139	-0.0016	-0.0401***	-0.0417***
	(0.0034)	(0.0031)	(0.0033)	(0.0036)	(0.0038)	(0.0048)	(0.0050)	(0.0051)
Edu 2	-0.0327	-0.0296	-0.0155	-0.0028	0.0098	0.0171***	0.0254***	0.0425***
	(0.0007)	(0.0006)	(0.0005)	(0.0004)	(0.0004)	(0.0009)	(0.0006)	(0.0008)
Edu 3	-0.0681	-0.0645	-0.0445	-0.0277	-0.0195	0.0235***	0.0251***	0.0486***
	(0.0012)	(0.0010)	(0.0008)	(0.0007)	(0.0007)	(0.0014)	(0.0010)	(0.0013)
Edu 4	-0.0720	-0.0690	-0.0699	-0.0647	-0.0619	0.0020	0.0080***	0.0101***
	(0.0011)	(0.0010)	(0.0011)	(0.0012)	(0.0013)	(0.0015)	(0.0017)	(0.0017)
Edu 5	-0.0828	-0.0996	-0.1240	-0.1378	-0.1424	-0.0412***	-0.0184***	-0.0596***
	(0.0011)	(0.0012)	(0.0015)	(0.0019)	(0.0022)	(0.0019)	(0.0027)	(0.0025)

Notes: Panels 1a and 2a of this table report the marginal effects of the SC/ST dummy in an ordered probit regression of education categories 1 to 5 on a constant and an SC/ST dummy (panel 1a); a constant, an SC/ST dummy and a set of individual, household and aggregate-level characteristics (panel 2a) for each survey round. Panels 1b and 2b of the table report the change in the marginal effects over successive decades and over the entire sample period. Standard errors are in parenthesis. * p-value≤0.10, ** p-value≤0.05, *** p-value≤0.01.

Our results indicate that there has been a significant trend toward convergence in education levels of SC/STs toward the levels of non-SC/STs even after controlling for covariates. It is important though to note the differential convergence patterns of the two groups across the different education categories, especially the fact that the marginal negative effect of SC/ST status on the probability of secondary school education has worsened over the sample period (after conditioning on the non-caste covariates of education). This potentially indicates that not all sections of SC/STs may have benefited equally during this period.

3.2 Occupation Choices

We now turn to the occupation choices of the two groups. In order to facilitate ease of presentation, we aggregate the 3-digit occupation codes that individuals report into a one-digit code. This leaves us with ten categories which are then grouped further into three broad occupation categories.[10] Our groupings, while subjective, are based on combining occupations with similar skill requirements. Thus, Occ 1 comprises white collar administrators, executives, managers, professionals, technical and clerical workers; Occ 2 collects blue collar workers such as sales workers, service workers and production workers; while Occ 3 collects farmers, fishermen, loggers, hunters etc..

The difference in skills across the occupation groups is reflected in the education attainment levels of those working in them. For both the young and old, those working in occupation 1 are the most educated (about 10 years of education) while occupations 2 and 3 employ people with progressively lesser education, on average (4 years and 2 years of education, respectively).[11] Moreover, the average years of education in all occupations has risen throughout the period with the sharpest increases being in blue collar jobs (Occ 2) and farming/agricultural jobs (Occ 3). Lastly, the differences in education are also reflected in the wage distribution by occupation: Occ 1 is characterized by the highest mean wage in our sample, followed by Occ 2, and Occ 3.

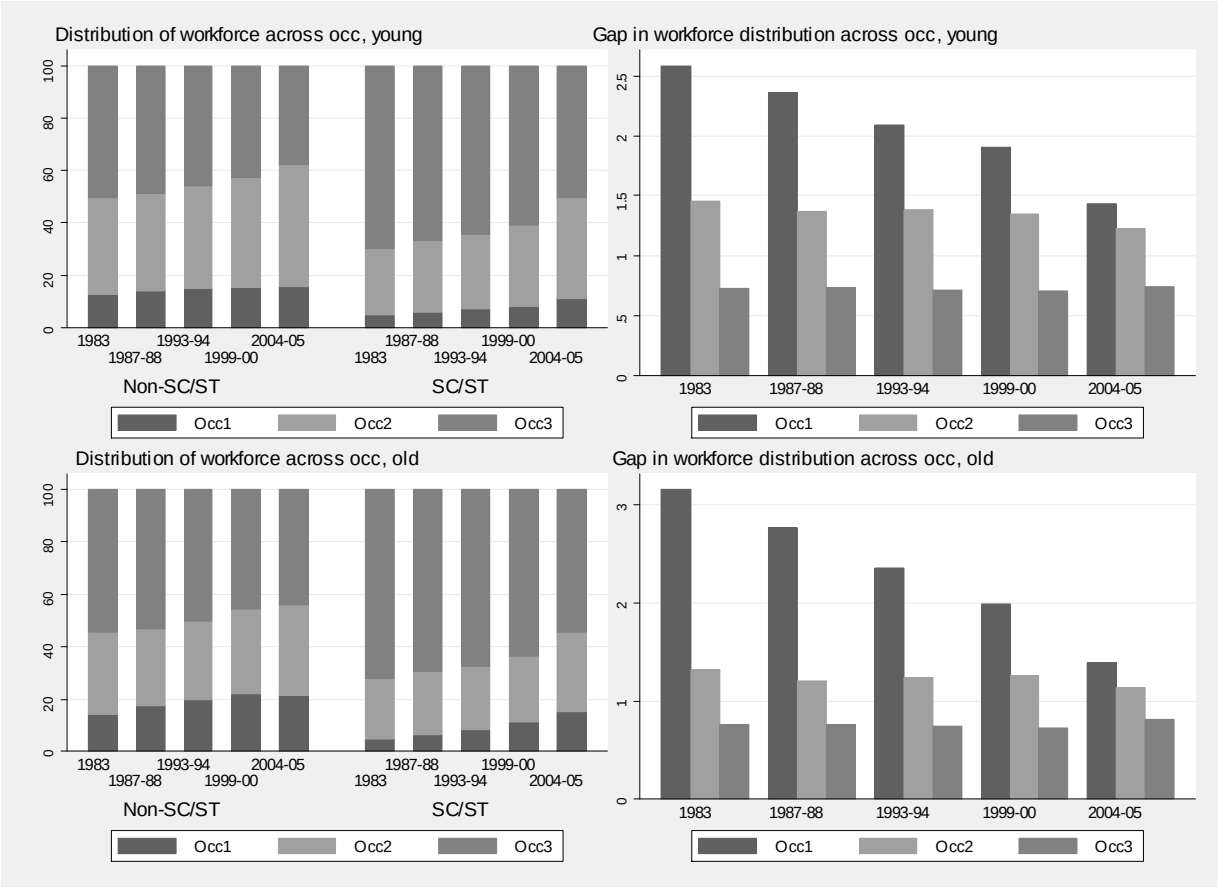
Figure 3 shows the occupation distribution for the working population in our sample, and the differences between non-SC/STs and SC/STs in this distribution. The top panel of Figure 3 refers to young, while the bottom panel refers to old. There are two features to note from Panel (a) of the Figure. First, there has been a systematic decline in Occ 3 (farming/pastoral activities) between 1983 and 2004-05 across all groups. This decline has been marginally sharper for SC/STs – both young and old. This reflects the structural transformation at the aggregate level for India wherein there has been a gradual decline in the output and employment share of the agricultural sector. Second, the largest expansion in the employment share has been in Occ 2 which comprises mostly low skill blue collar and service sector jobs. This phenomenon too has been common to both groups.

Since SC/STs were over-represented in Occ 3 and under-represented in Occ 1 and 2 in 1983, the trends in occupation shares of the two groups imply that the overall occupation distribution has become more similar over the sample period for both old and young, i.e., the distributions have been converging. This can be seen in Panel (b) of Figure 3 where the bar for any occupation category j is obtained by dividing the non-SC/STs in category j as a share of all non-SC/STs by the SC/STs in

[10] See Appendix A.1 for more details on the definitions of occupation categories.

[11] As before, we impute years of education from the reported education attainment levels by using the conversion formula outlined in section 2 and in the Appendix A.1.

Figure 3: Occupation distribution



(a) (b)

Notes: Panel (a) of this figure presents the distribution of workforce of young and old cohorts across three occupation categories for different NSS rounds. The left set of bars on each figure refers to non-SC/STs, while the right set is for SC/STs. Panel (b) presents gaps in the distribution of non-SC/STs relative to SC/STs across three occupation categories. The gaps are also reported for young and old. Occ 1 collects white collar workers, Occ 2 collects blue collar workers, while Occ 3 refers to farmers and other agricultural workers.

category j as a share of all SC/STs. Clearly, most of the bars are converging toward one over time.

To evaluate whether the convergent trends in the occupation distributions are significant, we estimate a simple multinomial probit model of occupation on a constant and an SC/ST dummy for every survey round. Panels 1a and 1b of Table 4 report the estimated marginal effect of the SC/ST dummy on the probability of being employed in each occupation (panel (1a)), as well as the changes in the marginal effects over successive decades and the entire sample period (panel (1b)). SC/STs are less likely to be employed in Occ 1 (white collar occupations) and 2 (blue collar occupations) and are more likely to work in Occ 3 (agrarian occupations) as compared to non-SC/STs throughout our sample period. However, these probabilities have been changing over time. In particular, the

negative effects of being an SC/ST on the probability of being employed in Occ 2 has declined, indicating that the share of SC/STs in that occupation has been growing over time, and converging to the share of non-SC/STs. Occ 3, on the other hand, has been releasing SC/ST workers as the marginal positive effect has fallen. Given the initial over-representation of SC/STs in Occ 3, this has led to converging employment shares of SC/STs and non-SC/STs. Lastly, there has been a very small but statistically significant increase in the negative effect of the SC/ST dummy on the probability of being employed in white collar jobs (Occ 1). While small, this effect represents an increase in disparities in their employment shares in this occupation. Hence, there again appears to have been a non-uniform convergence in the occupation distribution of the two groups across the occupation categories.

Table 4: Marginal effect of SC/ST dummy in multinomial probit regressions for occupations

	Panel 1a: Marginal effects, unconditional					Panel 1b: Changes		
	1983	1987-88	1993-94	1999-2000	2004-05	83 to 93	93 to 05	83 to 05
Occ 1	-0.0701***	-0.0734***	-0.0778***	-0.0760***	-0.0780***	-0.0077***	-0.0002	-0.0079***
	(0.0016)	(0.0014)	(0.0016)	(0.0021)	(0.0021)	(0.0022)	(0.0026)	(0.0026)
Occ 2	-0.0735***	-0.0663***	-0.0833***	-0.0769***	-0.0472***	-0.0098**	0.0362***	0.0264***
	(0.0033)	(0.0030)	(0.0031)	(0.0034)	(0.0039)	(0.0045)	(0.0050)	(0.0051)
Occ 3	0.1436***	0.1397***	0.1611***	0.1529***	0.1251***	0.0175***	-0.0360***	-0.0185***
	(0.0034)	(0.0031)	(0.0032)	(0.0036)	(0.0040)	(0.0047)	(0.0052)	(0.0053)
	Panel 2a: Marginal effects, conditional					Panel 2b: Changes		
Occ 1	0.0005	-0.0032**	-0.0047***	-0.0060***	-0.0095***	-0.0052**	-0.0048	-0.0100***
	(0.0017)	(0.0015)	(0.0018)	(0.0023)	(0.0025)	(0.0024)	(0.0030)	(0.0030)
Occ 2	0.0152***	0.0158***	0.0057	0.0149***	0.0353***	-0.0095	0.0296***	0.0201***
	(0.0043)	(0.0040)	(0.0042)	(0.0045)	(0.0050)	(0.0060)	(0.0065)	(0.0066)
Occ 3	-0.0157***	-0.0126***	-0.0010	-0.0089*	-0.0258***	0.0147***	-0.0248***	-0.0101
	(0.0044)	(0.0042)	(0.0162)	(0.0048)	(0.0054)	(0.0044)	(0.0054)	(0.0070)

Note: Panels 1a and 2a of this table present the marginal effects of SC/ST dummy from a multinomial probit regression of occupation choices on a constant and an SC/ST dummy (panel 1a); on a constant, SC/ST dummy and a set of other individual, household and aggregate characteristics (panel 2a) for each survey round. Panels 1b and 2b report the change in the marginal effects of SC/ST dummy over successive decades and over the entire sample period. Occupation 1 (Occ 1) has white collar workers, while Occupation 2 (Occ 2) collects blue collar workers. Occupation 3 (Occ 3) includes all agrarian jobs. Occ 3 is the reference group in the regression. Standard errors are in parenthesis. * p-value$\leq$0.10, ** p-value$\leq$0.05, *** p-value$\leq$0.01.

As in the ordered probit regressions for education categories, it is again important to note that the marginal effects of the SC/ST dummy reported in Table 4 estimate the absolute gaps between the two groups in the probabilities of being employed in the different occupation categories. The bars in the right panel of Figure 3, on the other hand, depict the relative gaps in the probabilities. This difference accounts for the fact that the relative probability gap has declined in Occ 1 while the absolute gap in Occ 1 has widened. We choose to compute both measures since the results can be used a check for robustness of the results.

Given that most of the workforce is employed in Occ 2 and Occ 3, the results above suggest that there has been a process of occupation convergence between SC/STs and non-SC/STs. The key question again is whether this is due to a convergence in the covariates of occupation choices or

is there convergence even after conditioning on the covariates? To answer this question we extend the multinomial probit regressions of the occupation choice to include individual, household and aggregate-level characteristics. The extended list of regressors includes individual's age, age squared, an SC/ST dummy, education category dummies, region dummies, a rural dummy, a muslim dummy, and a state-specific SC/ST quota variable. As before, we report only the marginal effect of belonging to the SC/ST social group on the probability of opting into the particular occupation. Those results are in Panels 2a and 2b of Table 4.

In line with our earlier findings, the effect of SC/ST status on the probability of being in Occ 1 is mostly negative and significant, implying that SC/STs tend to be under-employed in Occ 1, even after conditioning upon their characteristics. At the same time, the effect of SC/ST status on the probability of being in occupation 2 is mostly positive and significant, while its effect on being in occupation 3 is negative for all rounds. This reverses our findings for these two occupation categories in unconditional evaluations. It implies that after controlling for their attributes, SC/STs are more likely to be employed in blue collar jobs, and are less likely to work in agriculture, relative to non-SC/STs. Moreover, the magnitude of these conditional SC/ST effects have been rising over time for all occupations.

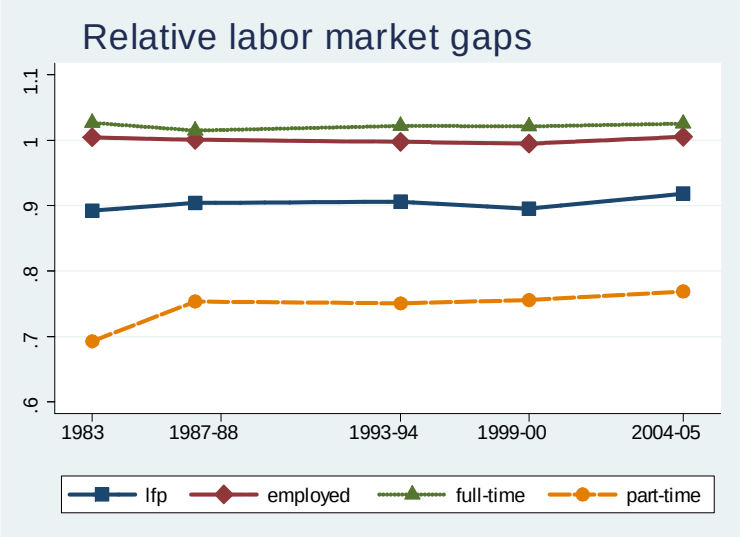
These results suggest to us that the unconditional convergence in the occupation distribution shown in the right panel of Figure 3 and in Table 4 is mostly due to convergence in the common covariates of occupation choices across the castes rather than a convergence in explicit caste based occupation allocations. If anything, the caste-based allocation factors may have been exacerbating the pre-existing disparities in the occupation distributions, as the marginal effect of SC/ST status on occupations, after conditioning on the attributes, appears to be significant and has been departing further from zero over our sample period. These changes over the entire sample period (1983 to 2004-05) are significant only for Occ 1 and Occ 2.

Before closing this section, it is worthwhile briefly describing developments along two other dimensions of the labor market, namely the labor force participation rates and employment rates of the caste groups. This is important since it could well be the case that while conditional on being employed, occupation choices are converging across the castes, gaps in both the labor force participation rates and employment rates of SC/STs and non-SC/STs may have remained unchanged or even widened during this period.

Figure 4 plots the non-SC/ST to SC/ST ratios in labor force participation rates, overall em-

ployment rates, as well as full-time and part-time employment rates.[12] There are three noteworthy features of the figure. First, the labor force participation rates of SC/STs are consistently greater than non-SC/STs. Second, the employment rates of the two groups are very similar. However, SC/STs are slightly less likely to work full-time relative to non-SC/STs, and are more likely to be employed part-time. Third, there was very little movement in either the relative participation rate or the relative employment rates over the sample period. If anything, there was a very slight trend toward convergence in both rates.

Figure 4: Labor force participation and employment gaps



Note: This figure shows the evolution of relative labor market gaps. "lfp" refers to the ratio of labor force participation rate of non-SC/STs to the corresponding rate for SC/STs. "employed" refers to the ratio of employment rates for the two groups; while "full-time" and "part-time" are, respectively, for the ratios of full-time employment rates and part-time employment rates of the two social groups. Details on how these rates were computed are available in the Appendix A.1.

3.3 Wages and Consumption

The previous two categories – education and occupation choices – are related to choices being made by households and individuals regarding their training and vocational pursuits. But what have been the economic rewards of these choices? The two standard ways of examining this is to study either wages or consumption expenditures of the concerned groups. Both measures have some problems. The wage data in the NSS is only available for the non-self-employed individuals thereby ruling out a large segment of the rural land-owning farmers. Moreover, there are often concerns about the accuracy of

[12] The details as to how these rates were computed are available in the Appendix A.1.

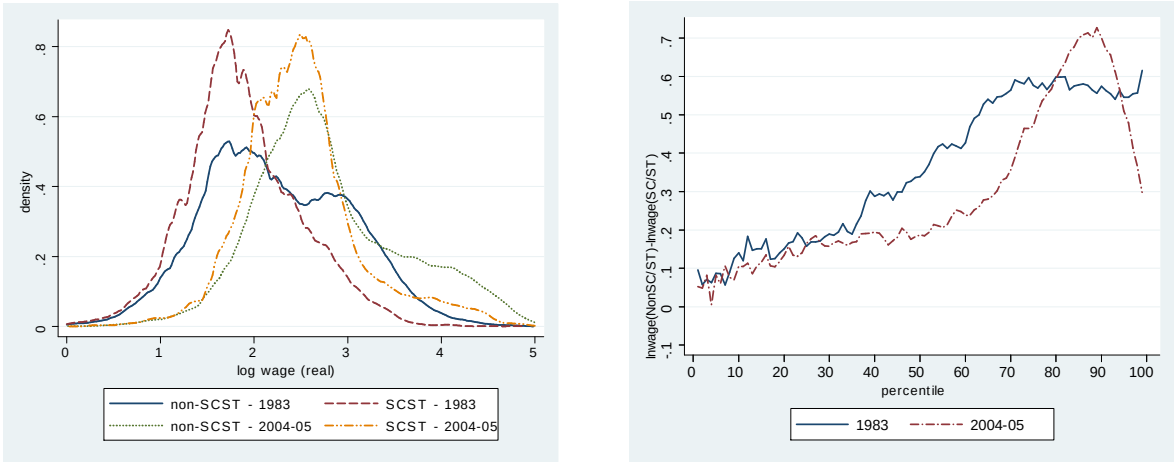
the reported wages in sample surveys (see Banerjee and Piketty (2005)). Consumption expenditure data, on the other hand, includes the self-employed and hence has a wider coverage. However, consumption data is available only at the household level which makes it difficult to accurately map consumption expenditures to individual characteristics. If household characteristics are different then two individuals with identical incomes will consume different amounts. Since neither approach is individually problem free, we choose to focus on both measures. Examining both also provides an automatic check of robustness of the results.

3.3.1 Wages

We first examine the trend in wages. We are particularly interested in determining whether the rising educational attainment rates and changing occupation distribution of SC/STs towards relatively higher skilled occupations have also resulted in a change in the wage gap relative to non-SC/STs.

At the outset it is important to note one important oddity in the wage data for the 43th round (1987-88). In particular, the number of observations for wages in this round falls precipitously to about half the level of the other rounds. This occurs due to a very large and disproportionate decline in the rural wage observations for this round. We are not sure as to the reasons for this sudden increase in the number of missing observations in the 43th round. In order to avoid spurious comparisons, we drop the 43rd round from all our statistical analysis of the wage data.

Figure 5: The log wage distributions of non-SC/STs and SC/STs for 1983 and 2004-05



(a) (b)

Notes: Panel (a) shows the estimated kernel densities of log real wages for non-SC/STs and SC/STs, while panel (b) shows the difference in percentiles of log-wages between non-SC/STs and SC/STs plotted against the percentile. The plots are for 1983 and 2004-05 NSS rounds. The line that slopes upward and to the right indicates more unequal distribution for non-SC/STs compared to SC/STs. The lines that are above the horizontal axis indicate stochastic dominance in non-SC/STs wage distribution.

It is instructive to start our analysis of the wage data by presenting the distribution of wages for the first and last rounds of our sample, i.e., for 1983 and for 2004-05. Panel (a) of Figure 5 plots the kernel densities of the wage distribution for SC/STs and non-SC/STs for these two rounds. Panel (b) examines the changes in wage inequality more closely by looking at the differences in log wages between non-SC/STs and SC/STs for different percentiles of their wage distributions for the two survey rounds.

Two features of panel (a) of Figure 5 deserve comment. First, for both groups the wage distribution has shifted sharply to the right. This is to be expected as the period 1983-2005 coincides with the rapid takeoff of the Indian economy. Second, the density functions for the two groups have come much closer together in 2004-05 relative to 1983.[13]

Panel (b) of Figure 5 reveals several additional features of the wage distributions. First, it shows the first-order stochastic dominance of the non-SC/ST wage distribution relative to the SC/ST wage distribution since wages are almost uniformly higher for non-SC/STs than for SC/STs for every percentile. However, the degree of the stochastic dominance has declined over time as the line for 2004-05 is much closer to zero for almost all percentiles up to the 80th percentile.[14] Second, both lines slope up and to the right, indicating that the wage distribution of non-SC/STs is more unequal that the wage distribution of SC/STs. An upward sloping line indicates that the difference in wages of the two groups is smaller for lower percentiles than for higher percentiles. But this implies that higher percentile non-SC/STs must earn not only more than higher percentile SC/STs, but that their wage mark-up relative to lower percentile non-SC/STs must also be greater than the wage mark-up of higher percentile SC/STs relative to their lower percentile counterparts. Hence, an upward sloping line indicates a more unequal wage distribution for non-SC/STs than SC/STs. The flattening out of the lines over time indicates a decrease in the wage inequality of the two distributions even though the sharp positive slope towards the right tail indicates continued wage inequality at the top-end of the income distribution. Overall, the plot suggests convergence in the two distributions over time as the line for 2004-05 round is well below the line for 1983 round for all except the 80th to 95th percentiles.

We now examine the wage gap between SC/STs and non-SC/STs more closely by contrasting the evolution of the median wage of SC/STs with non-SC/STs over finer sub-groups of age and

[13] We should note though that a formal Kolmogorov-Smirnov test of the equality of SC/ST and non-SC/ST wage distributions rejects the null hypothesis of equality both for 1983 and 2004-05. Moreover, the test also rejects the null hypothesis of the SC/ST distribution in 1983 being the same as the SC/ST distribution in 2004-05. This conclusion carries over to a comparison of the non-SC/ST wage distributions in these two rounds as well.

[14] For the top quintile there appears to have been some non-monotonicity in the distribution. This may be partly reflecting sampling issues at the top end of the income distribution as pointed out by Banerjee and Piketty (2005).

generation age groups as well as for all the survey rounds under study. Table 5 reports the median wage evolution of the generational age groups of old and young across the survey rounds. The columns labelled "median" show the ratios of median daily wages of non-SC/STs to the median daily wages of SC/STs for the overall sample, as well as separately for the working young and old age groups during 1983 to 2004-05. The adjoining columns labelled "mean" report the ratio of mean daily wages of the two groups. The lower panel called "Changes" computes the change in the relative gap over the 1990s decade and over the entire sample period. Standard errors for the estimates are reported in parenthesis under the estimates.

There are three broad takeaways from Table 5. First, the wage premium of non-SC/STs is much greater for the old than for the young across the rounds. Second, there is a clear trend towards convergence of wages of SC/STs toward non-SC/ST levels for both generational groups. The median wage premium of the non-SC/ST young has secularly declined from 38 percent in 1983 to 17 percent by 2004-05. The non-SC/ST median wage premium for the old has fallen from 70 percent in 1983 to 48 percent in 2004-05. Overall, the median wage premium of non-SC/STs has declined from 42 percent to 22 percent during this period. Third, the standard errors reported in the "change" column clearly show that except for the 1993-2000 period, all the changes in the median wage gaps are statistically significant indicating that the wage premium of non-SC/STs have declined significantly during this period. Fourth, the gaps in mean wages are smaller than the median wage gaps. Moreover, the changes in the mean wage gaps are also smaller than the changes in the median gaps and are sometimes insignificant. This last result suggests that it may be important to examine the wage gaps and the changes therein along the entire wage distribution rather than just the mean or the median since the answers may be different for different percentiles.

The patterns in the aggregated generational age groups we found above could also be masking important underlying changes in more disaggregated groups. To examine this, we break the sample along two margins – five age groups and three education category. We then study the wage convergence patterns for the five age groups within each education category. Figure 6 depicts the behavior of median wages of non-SC/STs relative to SC/STs by age groups across education categories. Panel (a) of the Figure shows the results for illiterates, panel (b) collects below primary, primary and middle school educated individuals, while panel (c) refers to those with secondary education and above. The wage gaps of illiterates were, on average, the smallest but also showed the least movement during this period. Age groups with primary and middle education on the other hand experienced a sustained convergence for all. The secondary and above cohorts reveal a slightly mixed picture with

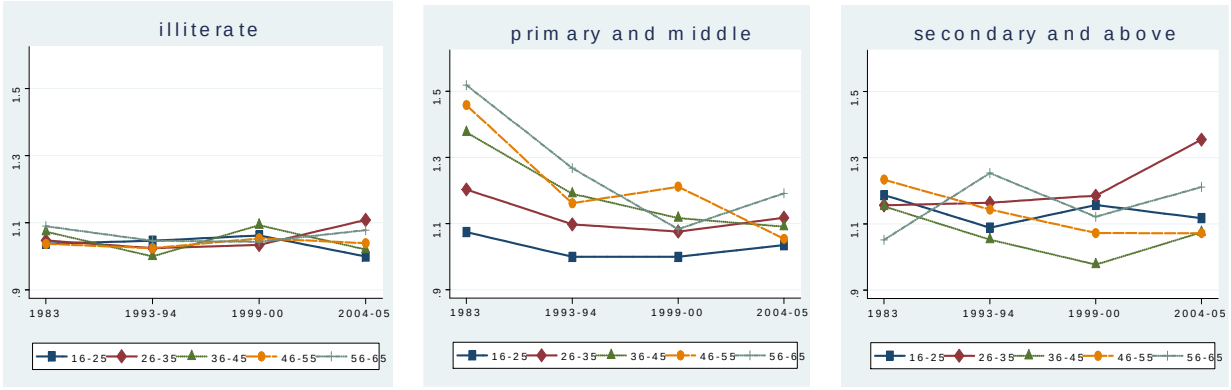
Table 5: Median wage gaps and changes

	Overall		Young (16-40)		Old (41-65)	
	median	mean	median	mean	median	mean
1983	1.42	1.61	1.38	1.53	1.70	1.81
	(0.01)	(0.03)	(0.01)	(0.04)	(0.03)	(0.06)
1993-94	1.27	1.54	1.23	1.39	1.59	1.83
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.03)
1999-00	1.26	1.55	1.19	1.41	1.59	1.77
	(0.01)	(0.03)	(0.01)	(0.03)	(0.02)	(0.06)
2004-05	1.22	1.56	1.17	1.41	1.48	1.77
	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.04)
Changes:						
1983 to 1993-94	-0.15***	-0.07***	-0.14***	-0.14***	-0.11***	0.02
	(0.01)	(0.03)	(0.02)	(0.04)	(0.04)	(0.07)
1993-94 to 2004-05	-0.05***	0.01	-0.06***	0.02	-0.11***	-0.06
	(0.01)	(0.03)	(0.01)	(0.03)	(0.03)	(0.05)
1983 to 2004-05	-0.20***	-0.05	-0.21***	-0.12***	-0.22***	-0.04
	(0.01)	(0.04)	(0.01)	(0.04)	(0.04)	(0.07)

Notes: Columns labeled "median" and "mean" in this table present, respectively, the relative median and mean wage gaps for our overall benchmark sample (column "overall") and separately for the two generational groups – young and old. The median (mean) wage gaps are computed as the ratios of median (mean) real wages of the two groups. Panel "Changes" computes the decade changes in the relative wage gap and its change over the entire sample period (1983 to 2004-05). The reported statistics are obtained for each NSS survey round which is shown in the first column. Standard errors are in parenthesis. * p-value≤0.10, ** p-value≤0.05, *** p-value≤0.01.

the gaps rising in two age groups and falling in the other three. Given the relative population shares of these three education categories, the figure suggests an overall narrowing of the median wage gap between SC/STs and non-SC/STs across most age groups.

Figure 6: Median wage gaps by age cohorts and education categories



(a) (b) (c)

Notes: The figures show the evolution of the relative median wage gap between non-SC/STs and SC/STs over time for different age and education categories. Panel (a) contains illiterates, panel (b) collects below primary, primary and middle school educated individuals, while panel (c) refers to those with secondary education and above.

The evolution of the wage gaps between SC/STs and non-SC/STs provides an interesting counterpoint to the racial wage gaps that are typically reported in the USA. Between 1980 and 2006, the median wage of black males relative to white male workers has remained relatively stable around 75

percent. During the same period, the median wage of Hispanic men relative to white men declined from 71 percent to under 60 percent.[15] In contrast, our computations above imply that the median wage of SC/STs relative to non-SC/STs has increased secularly from 70 percent in 1983 to 82 percent in 2004-05.[16] Amongst the younger groups the wage catch-up has been even faster with the relative median wages of SC/ST young having risen from 72 percent to 85 percent during this period. Clearly, the rate of wage convergence for SC/STs since 1983 has been quite striking both at an absolute level as well as in comparison to historically disadvantaged minority groups in more developed countries like the USA.

While the trends documented above are instructive, they leave unexplained the factors behind the converging trends. Have wages converged across the entire wage distribution or is the convergence restricted to specific quantiles? Have wages been converging due to convergence in the determinants of wages such as education, occupation choices etc. or have they been converging due to changes in unmeasured factors such as discrimination? How much of the wage convergence can be accounted for by convergence in education? We approach these issues in two different ways.

In our first approach, we follow the methods developed by DiNardo, Fortin, and Lemieux (1996) (DFL from hereon) to decompose the overall difference in the observed wage distributions of non-SC/STs and SC/STs into two components – the part that is explained by differences in the composition of attributes and the part that is explained by differences in the wage structure of the two groups. Borrowing loosely from the two-fold Oaxaca-Blinder decomposition terminology, the gap accounted for by attributes is the explained part arising due to differences in the $X's$ while the rest of the gap is unexplained and arises due to differences in the $\beta's$ or the wage structure of the two groups. Details regarding the procedure are provided in the Appendix A.2.

Intuitively, the DFL method works by first estimating the wage densities of the two groups separately using standard kernel density methods and then constructing a counterfactual wage density by "giving" one group the attributes of the other group. We construct the counterfactual density by reweighting the distribution of non-SC/STs to give them the SC/ST distribution of attributes. While one can also do the reverse reweighting by giving SC/STs the non-SC/ST attribute distribution, we choose not to because of a common support problem, i.e., there may not be enough SC/STs at the top end of the distribution to be able to mimic the non-SC/ST distribution.

Figure 7 shows the actual wage gaps between non-SC/STs and SC/STs for each percentile, the

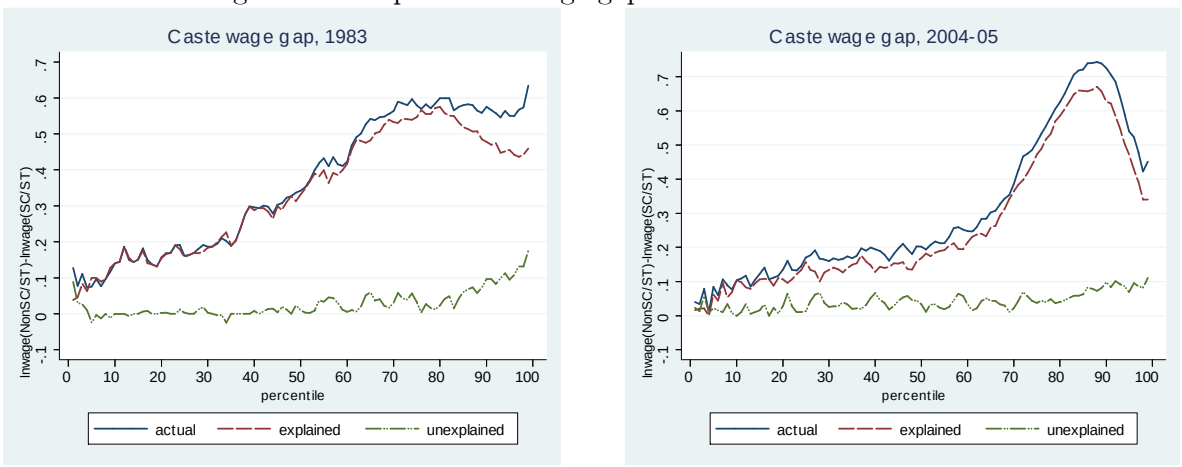
[15] These numbers are from US Current Population Survey.

[16] For ease of comparison with the typical wage gap numbers reported for the USA, the SC/ST wage gaps reported here are the inverses of the non-SC/ST to SC/ST relative wage gaps we reported above.

gap explained by differences in attributes of the two groups, and the unexplained gap. The explained gap is computed as the difference between the actual and re-weighted non-SC/ST wage percentiles, where we re-weighted the wage density of non-SC/STs by giving them the attributes of SC/STs. The unexplained wage gap is obtained as the difference between the re-weighted non-SC/ST wage percentiles and the actual wage percentiles of SC/STs. Panel (a) shows the plots for 1983, while panel (b) reports them for 2004-05.

The most striking feature about Figure 7 is that it shows that attributes can account for almost all of the wage gap between the groups. This is an interesting result in its own right as it indicates that the convergence in attributes of the two groups were the primary reason for the convergence in wages between the groups. There are two additional points to note. First, as we noted before using Figure 5 (b), the wage gaps in 2004-05 are universally lower than the gaps in 1983 until the 80th percentile. Between the 80th and 95th percentiles however, the wage gaps appear to be larger in the last round compared to 1983. Second, for 1983, the unexplained component tends to rise at the highest percentiles.

Figure 7: The percentile wage gaps: actual and counterfactual



(a) (b)

Notes: Each panel shows the actual log wage gap between non-SC/STs and SC/STs for each percentile, the counterfactual percentile log wage gap when non-SC/STs are given SC/ST attributes, and the residual wage gap for each percentile. The left panel is for 1983 while the right panel shows the decomposition for 2004-05.

While the DFL decomposition is useful for decomposing the wage into the attributes and wage structure effects, we are also interested in a more detailed decomposition wherein we can assess the effect of specific attributes such as education or caste on unconditional wage quantiles such as the median. As is well known, standard methods like the Oaxaca-Blinder decomposition for mean

wages do not work for quantiles. This is because the law of iterated expectations does not apply to quantiles: if w^m is the median wage and $E(w^m|X)$ is the expectation conditional on attributes X, then $E_X[E(w^m|X)] \neq E[w^m]$. Hence, we cannot use a conditional regression of the median such that $E(w^m|X) = \hat{\beta}X$ to determine the effect of X on the unconditional median. This only works for the means.

To get around this problem, we use the Recentered Influence Function (RIF) regression method developed by Firpo, Fortin, and Lemieux (2009) (FFL henceforth) for computing unconditional quantiles. The influence function of any τ^{th} quantile of wages, q_τ, is defined as $IF(w, q_\tau) = \frac{\tau - I[w \leq q_\tau]}{f(q_\tau)}$ where I is the indicator function which takes the value 1 when $w \leq q_\tau$ and f is the density function of wages.[17] The recentered influence function is $RIF(w, q_\tau) = q_\tau + IF(w, q_\tau)$. Since $E[\tau - I(w \leq q_\tau)] = 0$, it follows that $E[RIF(w, q_\tau)] = q_\tau$. FFL demonstrate that this property of the RIF also goes through for its expectation conditional on X. In particular, they show that

$$q_\tau = E_X[E\{RIF(w, q_\tau)|X\}] = E[X]\gamma^\tau$$

where γ^τ are the estimated coefficients from a regression of $RIF(w, q_\tau)$ on X. Hence, the τth quantile RIF regressions aggregates to the unconditional quantile.

While the details regarding this procedure can be found in FFL, the RIF methodology essentially amounts to first recentering the influence function of the quantile wage q_τ by adding the quantile wage and then using the recentered influence function as the dependent variable in a standard linear regression. In effect, the RIF regression for any quantile then is equivalent to running a linear probability regression of the probability of wages exceeding that quantile on a vector of regressors X.

The RIF regressions can be used to assess both the effect of individual covariates on the unconditional quantiles as well as their individual contributions to the observed quantile wage gaps. We proceed by first running quantile wage RIF regressions for each sample round on a set of controls that consists of individual's age and age squared, dummies for education category, SC/ST dummy, Muslim dummy, rural dummy, region and occupation specific dummies and a state-level reservation quota. We then test whether the coefficient on the caste dummy has changed significantly over the sample period and in which direction. This provides a test of conditional convergence of quantile wages. The results for the 10th, 50th and 90th quantiles are reported in Table 6. We also include

[17] The influence function of a statistic provides a measure of the influence of extreme observations on the statistic. For the mean μ of a random variable w, the influence function is $w - \mu$.

the results from the same regression for the mean of log wages.

Table 6: Effect of SC/ST dummy in the mean and RIF quantile regressions of log wages

	Panel (a): SC/ST dummy coefficient				**Panel (b): Changes**		
	1983	1993-94	1999-2000	2004-05	1983 to 1993-94	1993 to 2004-05	1983 to 2004-05
10th quantile	-0.0022	0.0081	-0.0169*	-0.0350**	0.0103	-0.0431***	-0.0328*
	(0.0141)	(0.0112)	(0.0097)	(0.0142)	(0.0180)	(0.0181)	(0.0200)
50th quantile	-0.0378***	-0.0314***	-0.0020	-0.0439***	0.0064	-0.0125	-0.0061
	(0.0091)	(0.0085)	(0.0083)	(0.0087)	(0.0125)	(0.0122)	(0.0126)
90th quantile	-0.0786***	-0.0401***	-0.0478***	-0.0646***	0.0385***	-0.0245	0.0140
	(0.0094)	(0.0115)	(0.0168)	(0.0182)	(0.0149)	(0.0215)	(0.0205)
mean	-0.0468***	-0.0424***	-0.0253***	-0.0521***	0.0044	-0.0097	-0.0053
	(0.0066)	(0.0080)	(0.0072)	(0.0073)	(0.0104)	(0.0108)	(0.0098)

Note: Panel (a) of this table reports the estimates of coefficients on SC/ST dummy from RIF regressions of log wages on a set of individual, household and aggregate characteristics. Results are reported for the 10th, 50th and 90th quantiles. Row labelled "mean" reports the SC/ST coefficient from the conditional mean regression. Panel (b) reports the changes in the estimated coefficients over successive decades and the entire sample period. Standard errors are in parenthesis. * p-value≤0.10, ** p-value≤0.05, *** p-value≤0.01.

Panel (a) of Table 6 shows that the coefficient on the SC/ST dummy variable is mostly negative and significant throughout. The negative estimates for the SC/ST dummy indicate that the corresponding quantile wage of SC/STs was significantly lower than similarly endowed non-Muslim non-SC/STs. Interestingly, the wage gaps are the lowest at the bottom end of the distribution, and increase for the upper quantiles during 1983-1994 period. Panel (b) of Table 6 however shows that the change in the estimated coefficient on the caste dummy over the entire period 1983-2005 was mostly insignificant except for the 10th quantile where the gap has increased. Hence, there does not appear to have been a convergence in wages across castes after conditioning for characteristics. Overall, these results corroborate those obtained using the DFL decomposition: the unconditional convergence in wages is mostly due to a convergence in attributes across the castes.

Next, we conduct a detailed decomposition of wage gaps between non-SC/STs and SC/STs for various quantiles and for the mean of the wage distribution. We use the RIF (OLS) regressions to decompose the observed quantile (mean) gaps into explained and unexplained components as well as to quantify the contribution of the key individual covariates to the explained component. This decomposition is in the spirit of the Oaxaca-Blinder method and is formalized in FFL for quantiles.[18] For quantiles the procedure consists of two main steps: (1) Run RIF regressions for the wage quantiles separately for SC/STs and non-SC/STs. (2) Use the results to perform the usual Oaxaca-Blinder decomposition for the wage gap at that quantile. We used a similar approach for the mean. All decompositions are performed using a pooled model across SC/STs and non-SC/STs

[18] Details of the decompositions can be found in Appendix A.3.

as the reference model.[19] Table 7 presents our results.[20]

Table 7: Decomposing SC/ST wage gaps

				explained	
(a). 1983	(i) measured gap	(ii) explained	(iii) unexplained	(iv) education	(v) occupation
10th quantile	0.1228***	0.1550***	-0.0322**	0.1188***	0.0227***
	(0.0122)	(0.0107)	(0.0162)	(0.0064)	(0.0087)
50th quantile	0.3476***	0.3310***	0.0166	0.1497***	0.1669***
	(0.0209)	(0.0145)	(0.0160)	(0.0060)	(0.0085)
90th quantile	0.5771***	0.3793***	0.1978***	0.2370***	0.1681***
	(0.0160)	(0.0103)	(0.0166)	(0.0082)	(0.0080)
mean	0.3661***	0.3193***	0.0468***	0.1789***	0.1350***
	(0.0118)	(0.0079)	(0.0089)	(0.0041)	(0.0053)
(b). 2004-05					
10th quantile	0.0977***	0.0722***	0.0255*	0.0862***	0.0132**
	(0.0134)	(0.0103)	(0.0164)	(0.0060)	(0.0066)
50th quantile	0.1989***	0.1729***	0.0260**	0.0938***	0.1047***
	(0.0149)	(0.0081)	(0.0124)	(0.0051)	(0.0059)
90th quantile	0.7249***	0.5304***	0.1945***	0.2419***	0.3682***
	(0.0248)	(0.0275)	(0.0262)	(0.0139)	(0.0216)
mean	0.3023***	0.2503***	0.0521***	0.1334***	0.1480***
	(0.0120)	(0.0092)	(0.0090)	(0.0049)	(0.0064)
(c). Change (1983 to 2004-05)					
10th quantile	-0.0251	-0.0828***	0.0577***	-0.0326***	-0.0095
	(0.0169)	(0.0140)	(0.0206)	(0.0085)	(0.0104)
50th quantile	-0.1487***	-0.1581***	0.0094	-0.0559***	-0.0622***
	(0.0249)	(0.0163)	(0.0203)	(0.0076)	(0.0103)
90th quantile	0.1479***	0.1511***	-0.0033	0.0049	0.2001***
	(0.0284)	(0.0293)	(0.0307)	(0.0169)	(0.0231)
mean	-0.0638***	-0.0691***	0.0053	-0.0455***	0.0130*
	(0.0173)	(0.0124)	(0.0126)	(0.0063)	(0.0084)
(d). Change in explained component					
10th quantile	-0.0828***	-0.0319***	-0.0509***	-0.0247***	-0.0127***
	(0.0140)	(0.0068)	(0.0114)	(0.0035)	(0.0035)
50th quantile	-0.1581***	-0.0645***	-0.0937***	-0.0230***	-0.0469***
	(0.0163)	(0.0134)	(0.0122)	(0.0040)	(0.0077)
90th quantile	0.1511***	0.0021	0.1490***	0.0149***	-0.0148**
	(0.0293)	(0.0122)	(0.0248)	(0.0077)	(0.0062)
mean	-0.0691***	-0.0406***	-0.0285***	-0.0143***	-0.0300***
	(0.0124)	(0.0107)	(0.0065)	(0.0047)	(0.0055)

Note: Panels (a) and (b) of this table present decomposition results of log wage gap between non-SC/STs and SC/STs in 1983 and 2004-05, respectively. Panel (c) presents the change in the SC/ST wage gap between 1983 and 2004-05. Panel (d) reports the decomposition of the time-series change in the explained component of the change in the SC/ST wage gap over 1983-2004-05 period. All gaps are decomposed into explained and unexplained components using RIF regression approach of FFL for the 10th, 50th and 90th quantiles; and using a standard OLS decomposition for the mean. All panels also report the contribution of education and occupation to the explained gaps. Bootstrapped standard errors are in parenthesis. * p-value≤0.10, ** p-value≤0.05, *** p-value≤0.01.

Panels (a) and (b) report the detailed decomposition of the wage gaps for the first (1983) and the last (2004-05) survey rounds for the 10th, 50th and 90th quantiles, and for the mean. We report the difference in log wages between non-SC/STs and SC/STs (column (i)), how much of that differential

[19] Following Fortin (2006) we allow for a group membership indicator in the pooled regressions.

[20] The standard errors for the unexplained part of the decomposition are generally hard to derive if a pooled model is used. In addition, in our decompositions we also interested in intertemporal comparisons, which complicates the computation of the standard errors further. Therefore, in our decompositions for wages and consumption expenditures, the standard errors are computed using bootstrap procedure. In the computations we accounted for the complex survey design of the NSS data. We also use adjusted sampling weights that account for the pooled sampling (over rounds) in our decompositions. The variance is estimated using the resulting replicated point estimates (see Rao and Wu (1988) and Rao, Wu, and Yue (1992)). See also Kolenikov (2010)) for STATA implementation.

is explained by characteristics (column (ii)), and how much is unexplained (column (iii)), and how much of the explained difference is due to education (column (iv)) and occupation (column (v)) gaps between the two social groups.

The main feature to note is that education and occupation jointly account for most of the wage gap between the castes for all rounds. For the median gap, the joint contribution of education and occupation to the overall gap has in fact steadily increased from about 90 percent in 1983 to 99 percent in 2004-05. For the mean gap, the corresponding numbers are 86 percent in 1983 and 93 percent in 2004-05. The table also reports the decomposition for the 10th and 90th percentiles. Relative to the median gaps, we find that education consistently accounted for a much larger share of the observed gaps for the 10th percentile while the contribution of the occupation is larger for the 90th percentile.

What is behind the changes in the wage gap between 1983 and 2004-05? Panel (c) of Table 7 reports the changes in the wage gap over time as well as the changes in its explained and unexplained components, including education and occupation.[21] Clearly, the mean wage gap has declined, and so has the explained component. The unexplained component, on the other hand, has shown a slight increase over time. The part of the explained component attributable to education has declined, while that due to occupation has increased. These variations, however, mask changes in both the individual endowments and returns to these endowments over our sample period. Thus, we decompose the changes in the explained component of the wage gap during the 1983-2005 period further into a part attributable to changes in gaps in the observables (explained component), and a part that is due to changes in the returns to these observables over time (unexplained component).[22] We find that changes in education and occupation have accounted for about 45 percent of the change in the explained gap at the 10th and 50th quantiles. That proportion is larger at 2/3 for the mean. In contrast, for the 90th quantile, the change in the explained gap is mostly due to changes in the group-specific structure of wages.

Overall, the results obtained under the various decomposition methods suggest that the caste wage gap is mostly explained by differences in attributes of the two social groups, and that the convergence in wages between the castes during 1983-2005 period occurred due to a convergence in the education attainments and occupation choices of the two groups.

[21] In these decompositions we used 1983 round as the benchmark sample.

[22] This inter-temporal decomposition of outcome differentials is in the spirit of Smith and Welch (1989) who used such decomposition techniques in their analysis of the change in the black-white wage differential over time. The details of our decomposition are presented in the Appendix.

3.3.2 Consumption

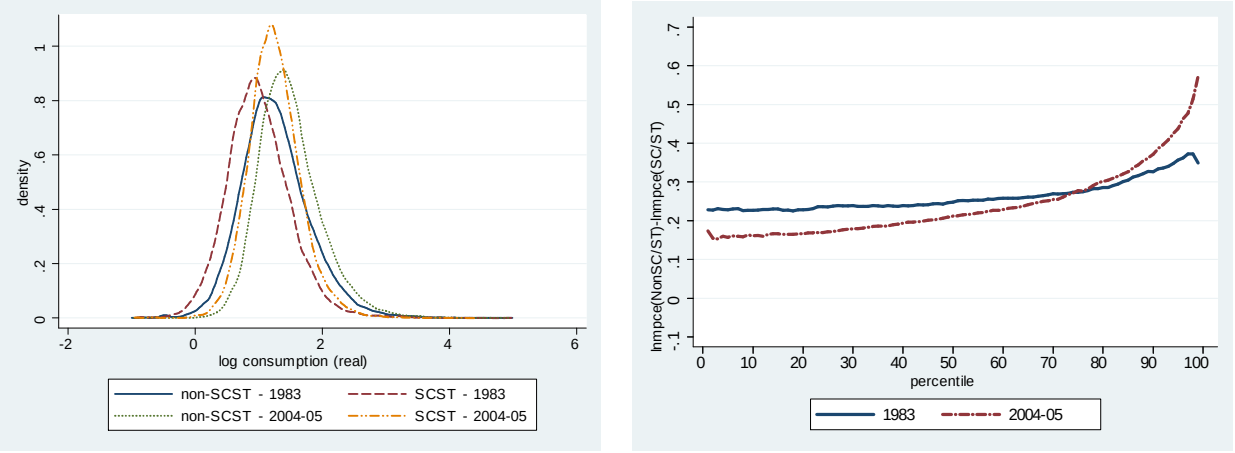
A key drawback of the wage data we used in the previous subsection is that it is available for a limited number of individuals. The sample size for the wage data is, on average, a third of the sample size for the education and occupation data. The missing wage observations are primarily due to the large segment of the rural population who identify themselves as being self-employed and correspondingly do not report any wages. The missing wage data raises a natural concern about sample selection. In particular, if non-SC/ST rural households are more likely to be land-owning and hence self-employed, then the wage data (particularly for rural households) would be skewed towards landless SC/ST households. The problem would be compounded by the fact that the wage earning non-SC/ST households may also be the worst off amongst the non-SC/STs. In this event we would be biasing our results toward finding low wage gaps between the two groups. In order to examine the robustness of our findings for the wage analysis above, we now compare per capita real consumption expenditures of SC/STs with non-SC/STs and its evolutions over time.

Consumption expenditure data is available at the household level, giving us a sample size of around 105,000 respondents on average across rounds. We convert consumption expenditures into real terms using official state-poverty lines, with rural Maharashtra in 1983 as the base – same as we did for wages. Further, to make real consumption results comparable with the wage data, we convert consumption into per-capita daily value terms.

We start by plotting the distribution of consumption for the two groups for 1983 and 2004-05 in Panel (a) of Figure 8 and the consumption gaps by percentiles in Panel (b) of the Figure, where we compute the differences in percentiles of consumption distributions between non-SC/STs and SC/STs in the same way as we did for wages.

Panel (a) of the figure reveals a shift rightward in the consumption distributions of both groups between 1983 and 2004-05. Panel (b) of Figure 8 reveals clear stochastic dominance of the non-SC/ST consumption distribution relative to the SC/ST consumption distribution in both periods, as both plots are above zero. However, it also shows that the consumption distributions of the two social groups have come closer together. In particular, the consumption gap between non-SC/STs and SC/STs declined over time for all percentiles below the 80th percentile. The slope of the consumption gap line has also increased in 2004-05 relative to 1983, indicating a more unequal distribution of non-SC/ST consumption in comparison with SC/ST consumption in the later years, especially at the right tail of the distribution. Overall, these patterns broadly mirror the patterns we found in the evolution of the wage distributions of the two groups during this period.

Figure 8: The log consumption distributions of non-SC/STs and SC/STs for 1983 and 2004-05



(a) (b)

Notes: Panel (a) shows the estimated kernel densities of log real consumption expenditure for non-SC/STs and SC/STs, while panel (b) shows the difference in percentiles of log-consumption between non-SC/STs and SC/STs plotted against the percentile. The plots are for 1983 and 2004-05 NSS rounds.

Are the changes in the consumption expenditure gaps significant? Table 8 reports the ratios of median per capita consumption of non-SC/STs relative to that of SC/STs for each round in our sample. As before we report the numbers for the overall sample as well as for the subsamples broken down by the generational age groups "young" and "old". The table shows that the relative median consumption gaps have fallen secularly over the rounds for both the overall sample as well as the generational age groups. Moreover, as the bottom panels show, the declines in the relative consumption gaps between 1983 and 2004-05 are all significant. Clearly, the significant convergence in wages has also been mirrored in convergence in consumption.

Next, we examine the determinants of the trends in the consumption patterns of the two groups. Specifically, we are interested in examining whether SC/ST households have lower consumption levels after one controls for household characteristics. As with wages, we proceed in two ways. First, we use the DFL method to decompose the consumption expenditure gap into the composition and wage structure components. Second, we conduct a detailed decomposition of the consumption gaps using unconditional quantile RIF regression methods.

Figure 9 presents our results for the DFL decomposition. This figure reports the actual log consumption gap for non-SC/STs relative to SC/STs computed for every percentile; the explained gap, computed as the difference between the actual and reweighted non-SC/ST consumption percentiles, where we reweighted the consumption distribution of non-SC/STs by assigning them with SC/ST attributes. Finally, the figure also reports unexplained consumption gap which is the difference be-

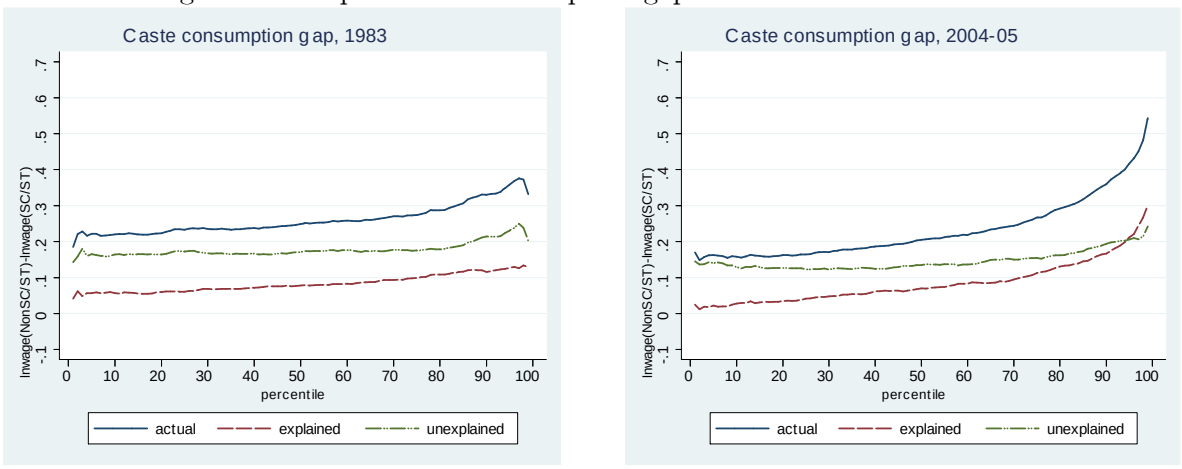
Table 8: Consumption gaps

	Overall		Young (16-40)		Old (41-65)	
	median	mean	median	mean	median	mean
1983	1.28	1.30	1.27	1.17	1.29	1.45
	(0.01)	(0.07)	(0.01)	(0.10)	(0.02)	(0.08)
1987-88	1.25	1.27	1.22	1.21	1.26	1.43
	(0.01)	(0.34)	(0.01)	(0.48)	(0.01)	(0.38)
1993-94	1.24	1.31	1.23	1.29	1.24	1.32
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
1999-00	1.21	1.27	1.22	1.26	1.2	1.27
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
2004-05	1.23	1.36	1.21	1.40	1.22	1.33
	(0.01)	(0.04)	(0.01)	(0.09)	(0.01)	(0.01)
Changes:						
1983 to 1993-94	-0.04***	0.01	-0.03	0.11	-0.05***	-0.13*
	(0.01)	(0.07)	(0.02)	(0.10)	(0.02)	(0.08)
1993-94 to 2004-05	-0.01	0.05	-0.02**	0.11	-0.02	0.01
	(0.01)	(0.04)	(0.01)	(0.09)	(0.02)	(0.02)
1983 to 2004-05	-0.06***	0.06	-0.05***	0.22*	-0.07***	-0.12
	(0.01)	(0.08)	(0.02)	(0.13)	(0.02)	(0.08)

Notes: Columns labeled "median" and "mean" in this table present, respectively, the relative median and mean consumption gaps for our overall benchmark sample (column "overall") and separately for the two generational groups – young and old. The median (mean) consumption gaps are computed as the ratios of median (mean) real consumption expenditure of the two groups. Panel "Changes"computes changes in the relative gap over the relevant period. The reported statistics are obtained for each NSS survey round which is shown in the first column. Standard errors are in parenthesis. * p-value≤0.10, ** p-value≤0.05, *** p-value≤0.01.

tween the re-weighted non-SC/ST consumption percentiles and the actual consumption percentiles of SC/STs. Panel (a) shows the plots for 1983, while panel (b) reports them for 2004-05.

Figure 9: The percentile consumption gaps: Actual and counterfactual



(a) (b)

Notes: Each panel shows the actual log consumption gap between non-SC/STs and SC/STs for each percentile, the counterfactual percentile log consumption gap when non-SC/STs are given SC/ST attributes, and the residual consumption gap for each percentile.

Several features of the data stand out from the plots. First, as we pointed out before, the actual consumption gap in 2004-05 is smaller than in 1983 for all except the top 20 percentiles. Second,

the composition effect ("explained" components) is clearly important in both periods. It accounts for 25% of the actual consumption gap at the bottom end of the consumption distribution and for about 35% at the top end in 1983. In 2004-05 the contribution of the composition effects was only 15% at the bottom end of the consumption distribution, but increased to over 50% at the top end of the consumption distribution. Third, in contrast to the wage decompositions, the unexplained component of the consumption gap is larger than the explained component, both in 1983 and in 2004-05, although its relative importance has declined over time. Fourth, relative to the wage gaps between the two groups, a much larger share of the consumption gap between them is explained by differences in the consumption structure (as opposed to attributes). Hence, convergence in the consumption structures have clearly played a larger role in the consumption convergence between SC/STs and non-SC/STs. Overall, our results suggest a growing importance of attributes in accounting for the convergence in the consumption distribution for the two social groups over time.

Next, we examine which variables can explain the trends we observed in the consumption data. As we did for wages, we use the RIF regression approach to assess both the individual contribution of different covariates for unconditional quantiles of consumption distribution; and to perform a Oaxaca-Blinder-type decomposition at these quantiles. Panel (a) of Table 9 reports the coefficients on the SC/ST dummy in various quantile RIF regressions of log consumption expenditure on a set of household-level and aggregate attributes. These include age and age squared of the male head of household, his education and occupation dummies, the maximum education obtained by members of the household (dummies), the size of the household and the number of earning household's members, rural dummy, muslim dummy, a set of zone dummies and the SC/ST quota. Panel (b) of Table 9 reports the changes in the SC/ST dummy coefficient over various time periods. We report the results for the 10th, 50th and 90th quantiles.

A few features of the results are important to note. First, the SC/ST dummy is significant and negative across all the rounds. Hence, consumption of SC/ST households is significantly lower than that of non-SC/ST households. The coefficient is also quantitatively similar for all the quantiles we considered, indicating that the SC/ST consumption gap is of similar magnitude along the entire distribution. Second, the estimated value of the SC/ST dummy coefficient has declined over time for all quantiles, and this decline is especially pronounced and significant between 1983 and 1993-94. Over the entire period under study, the consumption gap, conditional on characteristics, has narrowed by almost 5% for the bottom quantiles and about 3% at the median. The conditional gap has remained roughly unchanged at the top end of the consumption distribution.

Table 9: Effect of SC/ST dummy in the mean and RIF quantile regressions for consumption expenditures

	Panel (a): SC/ST dummy coefficient				
	1983	1987-88	1993-94	1999-2000	2004-05
10th quantile	-0.1894***	-0.1632***	-0.1324***	-0.1460***	-0.1421***
	(0.0102)	(0.0091)	(0.0087)	(0.0088)	(0.0103)
50th quantile	-0.1810***	-0.1574***	-0.1607***	-0.1366***	-0.1519***
	(0.0072)	(0.0059)	(0.0059)	(0.0060)	(0.0068)
90th quantile	-0.1843***	-0.1654***	-0.1609***	-0.1293***	-0.1709***
	(0.0097)	(0.0090)	(0.0095)	(0.0104)	(0.0093)
mean	-0.1831***	-0.1614***	-0.1544***	-0.1361***	-0.1544***
	(0.0060)	(0.0052)	(0.0047)	(0.0045)	(0.0051)
	Panel (b): Changes				
		1983 to 1993-94	1993 to 2004-05	1983 to 2004-05	
10th quantile		0.0570***	-0.0097	0.0473***	
		(0.0134)	(0.0135)	(0.0145)	
50th quantile		0.0203***	0.0088	0.0291***	
		(0.0093)	(0.0090)	(0.0099)	
90th quantile		0.0234*	-0.0100	0.0134	
		(0.0136)	(0.0133)	(0.0134)	
mean		0.0287***	0.0000	0.0287***	
		(0.0076)	(0.0069)	(0.0079)	

Note: Panel (a) reports the estimates of the coefficient on SC/ST dummy from RIF regressions of log consumption expenditures on a set of household and aggregate characteristics. Panel (b) reports the changes in the estimated coefficients over successive decades and the entire sample period. Standard errors are in parenthesis. * p-value≤0.10, ** p-value≤0.05, *** p-value≤0.01.

Finally, we apply the RIF approach in combination with Oaxaca-Blinder decomposition to evaluate the contribution of various attributes to the consumption gap, at different quantiles and for the mean. We follow the same steps as we did for wages, but using a consumption-relevant set of controls in the regression step as specified above. All decompositions are performed by assigning the coefficients from a pooled regression (with a group membership indicator) as the reference. Panels (a) and (b) in Table 10 report the results for the 1983 and 2004-05 survey rounds for the 10th, 50th, 90th quantiles, and the mean. We report the difference in log consumption expenditures between non-SC/STs and SC/STs (column (i)), how much of that differential is explained by characteristics (column(ii)), and how much is unexplained (column (iii)), and how much of the explained difference is due to education and occupation gaps between the two social groups (columns (iv) and (v), respectively). Bootstrapped standard errors are in parenthesis.

In 1983, the median log consumption gap between non-SC/STs and SC/STs was 25 percent. 30 percent of it was due to difference in characteristics, especially education, which accounted for the majority of the explained gap. Differences in occupation choices of household heads accounted for about 6 percent of the explained gap. Similar decomposition results are obtained for the 10th and 90th quantiles. Note, however, that the gap is small at the lower end of the distribution, but significantly larger at the upper end. Moving on to 2004-05, we find that the overall gaps decreased over time, except at the very top of the distribution. The explained component is between 20 percent

Table 10: Decomposing SC/ST consumption expenditure gaps

				explained	
(a). 1983	(i) measured gap	(ii) explained	(iii) unexplained	(iv) education	(v) occupation
10th quantile	0.2202***	0.0607***	0.1595***	0.0778***	0.0084***
	(0.0122)	(0.0053)	(0.0111)	(0.0031)	(0.0015)
50th quantile	0.2492***	0.0740***	0.1753***	0.1001***	0.0043***
	(0.0077)	(0.0054)	(0.0076)	(0.0035)	(0.0016)
90th quantile	0.3308***	0.0953***	0.2356***	0.1271***	0.0126***
	(0.0139)	(0.0073)	(0.0147)	(0.0055)	(0.0033)
mean	0.2585***	0.0754***	0.1832***	0.0997***	0.0078***
	(0.0085)	(0.0047)	(0.0084)	(0.0030)	(0.0016)
(b). 2004-05					
10th quantile	0.1569***	0.0312***	0.1257***	0.0493***	0.0133***
	(0.0104)	(0.0052)	(0.0107)	(0.0028)	(0.0016)
50th quantile	0.2044***	0.0652***	0.1392***	0.0698***	0.0185***
	(0.0063)	(0.0046)	(0.0062)	(0.0029)	(0.0015)
90th quantile	0.3608***	0.1531***	0.2077***	0.1042***	0.0569***
	(0.0116)	(0.0084)	(0.0098)	(0.0046)	(0.0041)
mean	0.2324***	0.0793***	0.1530***	0.0732***	0.0269***
	(0.0061)	(0.0045)	(0.0054)	(0.0024)	(0.0015)
(c). Change (1983 to 2004-05)					
10th quantile	-0.0634***	-0.0296***	-0.0338**	-0.0285***	0.0049**
	(0.0159)	(0.0069)	(0.0156)	(0.0043)	(0.0023)
50th quantile	-0.0448***	-0.0087	-0.0361***	-0.0303***	0.0143***
	(0.0092)	(0.0069)	(0.0095)	(0.0047)	(0.0021)
90th quantile	0.0299*	0.0578***	-0.0279	-0.0229***	0.0443***
	(0.0184)	(0.0119)	(0.0179)	(0.0070)	(0.0054)
mean	-0.0262***	0.0039	-0.0301***	-0.0265***	0.0191***
	(0.0103)	(0.0065)	(0.0100)	(0.0039)	(0.0021)
(d). Change in explained component					
10th quantile	-0.0296***	-0.0075*	-0.0221***	-0.0158***	-0.0003
	(0.0069)	(0.0044)	(0.0060)	(0.0030)	(0.0010)
50th quantile	-0.0087	0.0026	-0.0114***	-0.0084***	0.0003
	(0.0069)	(0.0048)	(0.0046)	(0.0034)	(0.0008)
90th quantile	0.0578***	0.0342***	0.0236***	0.0052	0.0061***
	(0.0119)	(0.0073)	(0.0093)	(0.0048)	(0.0020)
mean	0.0039	0.0091*	-0.0052	-0.0063*	0.0016*
	(0.0065)	(0.0049)	(0.0039)	(0.0034)	(0.0009)

Note: Panels (a) and (b) of this table present decomposition results of log consumption expenditure gap between non-SC/STs and SC/STs in 1983 and 2004-05, respectively. Panel (c) presents the change in the SC/ST consumption gap between 1983 and 2004-05. Panel (d) reports the decomposition of the time-series change in the explained component of the change in the SC/ST consumption gap over 1983-2004-05 period. All gaps are decomposed into explained and unexplained components using RIF regression approach of FFL for the 10th, 50th and 90th quantiles; and using a standard OLS decomposition for the mean. Bootstrapped standard errors are in parenthesis. * p-value≤0.10, ** p-value≤0.05, *** p-value≤0.01.

for the 10th quantile and 42 percent for the 90th quantile with most of it being accounted for by education.

Is the change in the consumption gap between 1983 and 2004-05 mostly due to changes in characteristics or changes in the returns differentials for the two social groups? Panel (c) of Table 10 shows that the median consumption gap between the two groups declined by about 0.045 during our sample period, while the mean gap fell by 0.03. The decline was most pronounced at the bottom of the distribution, where it fell by 0.06, while the gap increased by about 0.03 at the very top of the distribution. Interestingly, our decomposition suggests that the unexplained component of the consumption gaps fell over the sample period for the entire consumption distribution. The

explained component, on the other hand, has fallen for the lower end of the distribution, while it has increased for the upper end of the distribution, and for the mean. The contribution of education to the explained gap has uniformly declined, while the contribution of occupation choices has uniformly increased for the entire distribution. Finally, from panel (d) of Table 10 we see that at the mean and the 90th quantile, most of the change in the explained gap between 1983 and 2004-05 is due to changes in the differential of individual attributes. For the lower end of the distribution however, inter-temporal changes in the return differentials contribute the most to the observed change in the explained gap.

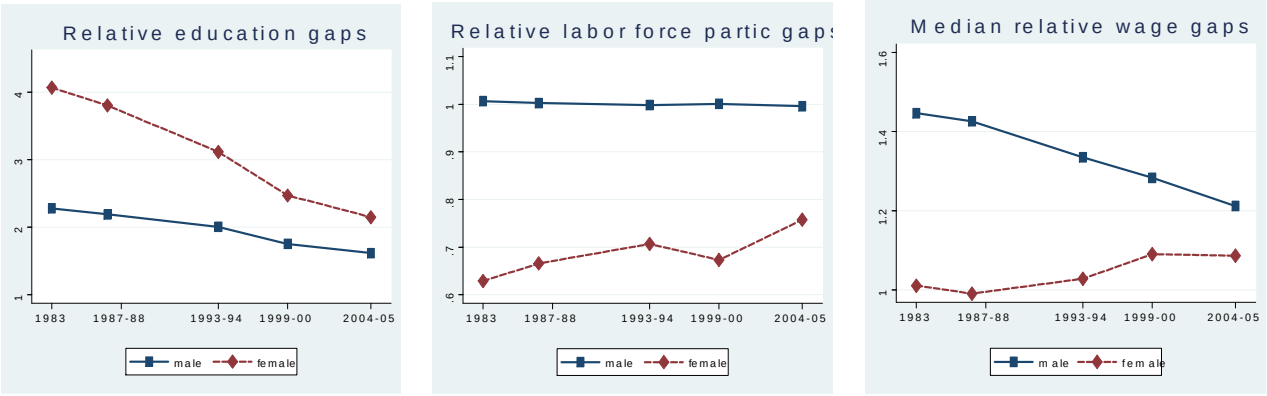
Our main conclusion from this section is that both consumption and wages of SC/STs have been converging toward the levels of non-SC/STs. The wage convergence process has been sharper reflecting both a larger initial gap as well as the fact that consumption levels adjust more slowly over time than wages. However, we also find that conditional on household and individual attributes, SC/STs are likely to have significantly lower consumption and wages relative to similarly endowed non-SC/STs. The degree of this conditional caste gap has remained relatively unchanged during this period for wages (except at the bottom of the distribution), but has declined for consumption, especially at the lower end of the distribution. Most of the convergence in wages is accounted for by changes in the measured attributes of SC/STs relative to non-SC/STs, while for consumption it is a combination of convergence in attributes and unobservable caste-based factors.

3.4 Gender Differences

The last issue of interest to us is whether the overall patterns of caste-based disparities we have described above mask important underlying differences along gender lines. More specifically, recent work along the lines of Munshi and Rosenzweig (2006) have argued that the behavior of women is key to understanding the education and labor market outcomes of different castes. In order to shed light on this issue, we compared the education, labor force participation rates and wages of SC/STs with their non-SC/ST counterparts for men and women separately. Figure 10 plots the results. All the plotted series show non-SC/STs relative to SC/STs.

In terms of years of education, the trends are similar for men and women. Specifically, the education gaps were larger for women throughout but became smaller over time. In sharp contrast to the education patterns, the labor force participation rates and the median wage gaps however reveal a fascinating contrast between the genders. Panel (b) shows that the labor force participation rates of SC/ST women were consistently larger than the participation rates of non-SC/ST women.

Figure 10: Gaps by Gender



(a) (b) (c)

Notes: The figures show the evolution of the relative gap of non-SC/STs to SC/STs over time for males and females separately. Panel (a) is the relative education gap (in years), panel (b) is the gap in the labor force participation rates, while panel (c) shows the relative median wage gaps.

However, this gap became smaller over time. The participation rates of men of the two groups were however quite similar and didn't change much over the rounds either.

Perhaps the most interesting contrast between the genders is in the relative median wage gaps shown in panel (c) of Figure 10. The caste wage gap was smaller throughout for women relative to the gap for men. Curiously though, the wage gap for women actually rose marginally from an initial position of parity. The contrasting behavior of some of the labor market gaps across gender lines in India during this period is a fascinating area of potential research. In particular, the widening wage gap for women despite a decrease in the relative labor supply of SC/ST women is a puzzle and deserves to be explored much more closely. We intend to address this issue in future work.

4 Conclusion

In this paper we have studied the evolution of education attainment rates, occupation choices, and wages in India between 1983 and 2004-05 with a special focus on the fortunes of scheduled castes and scheduled tribes (SC/STs). We have found that this has been a period of dramatic changes for these historically disadvantaged groups. SC/STs have systematically and significantly reduced the gap with non-SC/STs in their average education attainment levels and in their relative representations in different occupations. Moreover, the median wage and consumption gaps between SC/STs and non-SC/STs have narrowed sharply during this period. While the convergence in occupation and wages is mostly due to a convergence in attributes, especially education, the convergence in education and consumption expenditures is also accounted for by a significant reduction in the effect of caste

after conditioning for other attributes.

As a sub-text, we have also found that not all groups have been affected the same in the process of this convergence. For both wages and consumption, we have found that the convergence is much sharper for the lower percentile groups. In fact, our evidence suggests that there may have been a slight widening of the gaps for the highest percentiles (after conditioning for the attributes of the groups). This heterogeneity finds an echo in the education and occupation changes as well. In education, there has been convergence in the representations of the two groups in the relatively lower education categories but a slight increase in the gaps for the highest education categories. For occupation choices, we have found that there has been convergence in the probabilities of the two groups being in blue collar occupations but a divergence in the corresponding probabilities of being in agrarian and white collar occupations. These results suggest that, in contrast to the conventional view, the relatively better-off amongst the SC/STs may have faced more obstacles in benefiting from the undergoing changes in the economy than the relatively poorer SC/STs.

We have found that the gaps were usually smallest for the younger age groups. This last fact when combined with the convergence in education attainment levels of the two groups suggest that intergenerational mobility rates of SC/STs may have risen faster than that of non-SC/STs between 1983 and 2005. In work in Hnatkovska, Lahiri, and Paul (2011), we confirm this impression by using formal evaluation of changes in the intergenerational mobility rates in education, occupation and wages of the two groups. Given that the remaining gaps are typically the smallest amongst the younger age groups, and the changes have often been the sharpest amongst urban households, we find the evidence particularly reassuring in terms of the future prospects of SC/STs in India.

What explains these significant changes in the Indian social landscape? The rapid structural changes in the Indian economy over the past 25 years may be a key reason. The liberalization of the previously restricted economy has opened up new opportunities for the private sector. The more rapid response of SC/STs to these opportunities probably reflects a confluence of factors. One factor may be the competitive pressures that were unleashed on markets by the economic liberalization. As argued by Becker (1957), increasing competition raises the losses to businesses from pursuing discriminatory labor market practises including wage discrimination. The resultant decline in the wage gap could then also induce these disadvantaged groups to increase their education attainment rates since the returns to education rise. A second factor may be that the rapidly changing socioeconomic environment in India has presented SC/STs with a historic opportunity to break out of a centuries-old cycle of illiteracy and poverty, and they have been acting proactively to take

advantage. A strengthening of community based networks of SC/STs along the lines suggested in Munshi (2010) may have also been at play in accelerating this process. The third possibility is that the reservations policy in place since 1950 for public sector jobs and higher education seats may have played a key role. The first two possibilities imply that caste may be becoming a less important factor in economic allocations in India while the third factor would put caste-based policies at the center of the explanation. We intend to examine these potential explanations in future work.

References

Banerjee, A., and T. Piketty (2005): "Top Indian Incomes, 1922-2000," *World Bank Economic Review*, 19(1), 1–20.

Banerjee, B., and J. Knight (1985): "Caste Discrimination in the Indian Urban Labour Market," *Journal of Development Economics*, 17(3), 277–307.

Becker, G. (1957): *The Economics of Discrimination*. University of Chicago Press, Chicago.

Borooah, V. K. (2005): "Caste, Inequality, and Poverty in India," *Review of Development Economics*, 9(3), 399–414.

DiNardo, J., N. M. Fortin, and T. Lemieux (1996): "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach," *Econometrica*, 64(5), 1001–44.

Firpo, S., N. M. Fortin, and T. Lemieux (2009): "Unconditional Quantile Regressions," *Econometrica*, 77(3), 953–973.

Fortin, N. M. (2006): "Greed, Altruism, and the Gender Wage Gap," Working papers, University of British Columbia.

Hnatkovska, V., A. Lahiri, and S. Paul (2011): "Breaking the Caste Barrier: Intergenerational Mobility in India," Working papers, University of British Columbia.

Ito, T. (2009): "Caste Discrimination and Transaction Costs in the Labor Market: Evidence From Rural North India," *Journal of Development Economics*, 88(2), 292–300.

Kijima, Y. (2006): "Caste and Tribe Inequality: Evidence from India, 1983-1999," *Economic Development and Cultural Change*, 54(2), 369–404.

Kolenikov, S. (2010): "Resampling variance estimation for complex survey data," *Stata Journal*, 10(2), 165–199(35).

Madheswaran, S., and P. Attewell (2007): "Caste Discrimination in the Indian Urban Labour Market: Evidence from the National Sample Survey," *Economic and Political Weekly*, 42(41), 4146.

Munshi, K. (2010): "Strength in Numbers: Networks as a Solution to Occupational Traps," *Review of Economic Studies, forthcoming*.

Munshi, K., and M. Rosenzweig (2006): "Traditional Institutions Meet the Modern World: Caste, Gender, and Schooling Choice in a Globalizing Economy," *American Economic Review*, 96(4), 1225–1252.

——— (2009): "Why is Mobility in India so Low? Social Insurance, Inequality, and Growth," Working papers, Brown University, Department of Economics.

Pande, R. (2003): "Can Mandated Political Representation Increase Policy Influence for Disadvantaged Minorities? Theory and Evidence from India," *American Economic Review*, 93(4), 1132–1151.

Prakash, N. (2009): "The Impact of Employment Quotas on the Economic Lives of Disadvantaged Minorities in India," Working papers, Dartmouth College, Department of Economics.

Rao, J. N. K., and C. F. J. Wu (1988): "Resampling inference with complex survey data," *Journal of the American Statistical Association*, 83, 231–241.

Rao, J. N. K., C. F. J. Wu, and K. Yue (1992): "Some recent work on resampling methods for complex surveys," *Survey Methodology*, 18, 209Ű217.

Smith, J. P., and F. R. Welch (1989): "Black Economic Progress after Myrdal," *Journal of Economic Literature*, 27(2), 519–64.

A Appendix

A.1 Data Appendix

A.1.1 National Sample Survey (NSS)

The National Sample Survey Organization (NSSO), set up by the Government of India, conducts rounds of sample surveys to collect socioeconomic data. Each round is earmarked for particular subject coverage. We use the latest five large quinquennial rounds – 38(Jan-Dec 1983), 43(July 1987-June 1988), 50(July 1993-June 1994), 55(July 1999-June 2000) and 61(July 2004-June 2005) on Employment and Unemployment (Schedule 10). The survey covers the whole country except for a few remote and inaccessible pockets. The NSS follows multi-stage stratified sampling with villages or urban blocks as first stage units (FSU) and households as ultimate stage units. The field work in each round is conducted in several sub-rounds throughout the year so that seasonality is minimized. The sampling frame for the first stage unit is the list of villages (rural sector) or the NSS Urban Frame Survey blocks (urban sector) from the latest available census. We describe the broad outline of sample design – stratification, allocation and selection of sample units - with a caveat that the details have changed from round to round.

The whole country is divided politically into states and union territories, and each state is further divided into districts for administrative purpose. The NSSO also constructs regions by grouping contiguous districts within a state which are similar in population density and crop pattern for the sampling purpose. Two different stratification methods are used for rural and urban sector in each state. In the rural sector, each district is generally counted as a separate stratum (populous districts are split into two or more strata) whereas in the urban sector, strata are formed within the NSS region based on population size of cities. For example, all towns with population less than 50,000 in a region will form stratum 1 and so on. In the 61st round, the stratification method was changed substantially. For this round, each district is divided into two basic strata – rural and urban. Then the rural and urban strata are further divided into sub-strata.

The total sample size of first stage unit (villages/urban blocks) is allocated to the states and union territories in proportion to population. The subsequent allocations to rural and urban sector and at stratum level within a state are based on population size as well. In rural sectors, sample FSUs are selected with probability proportional to population from each stratum (sub-stratum for

61st round). In urban sectors, they are selected by simple random sampling without replacement in 38th and 61st round and circular systematic sampling with equal probability in the 43rd, 50th and 55th round. Within each stratum (sub-stratum for 61st round), samples are drawn in the form of two independent sub-samples for both rural and urban sectors. Once the FSUs are randomly drawn, the large FSUs are subdivided into certain number of parts (hamlet-group/sub-block) with approximately equal population and one of them selected randomly for listing of households. Complex second stage stratification based on "means of livelihood class" is implemented to select households randomly from the sample frame of households in each FSU (or hamlet-group/sub-block).

As the sample design changes over the rounds, estimation without considering the complex design may be misleading. The NSSO supplies household level multipliers with the unit record data for each round to help minimize estimation errors on the part of researchers. The questionnaire collects demographic details like age, sex, marital status, education, etc. and information about occupation, industry, activity, time disposition in reference week, wage, etc. of household members. It also collects monthly total household expenditure along with other household level characteristics.

The data are given in fixed format text files with a list of variable names and byte positions. We have checked the validity of our data extraction process by comparing the statistics on a number of the variables with numbers reported in published works by other authors. However, there is some miscoding which is typical for any survey data and we tried our best to clean it. Other notable changes over the rounds are formation of new states, deletion of the social group called "Neo-Buddhist" and formation of new social group called "Other Backward Class" or "OBC" (see below), and changes in coding for education, enrolment in educational institution, activity status and industry. We recoded all these changes to make it uniform and consistent over the time.

A.1.2 Sample Selection

We drop all households for which we have no information on social group or whose social group is miscoded (3/ 120706 households in 38th round, 43/ 129060 households in 43rd round, none for 50th and 55th rounds (115409 and 120386 households, respectively), and 86/124680 households for 61st round are dropped). The classification of Scheduled Castes (SC) and Scheduled Tribes (ST) groups remain unchanged over the rounds. However, there is a new classification of "Other Backward Classes" (OBC) from the 55th round while the "Neo-Buddhist" classification was discontinued from the 50th round. We club these groups with non-SC/ST so that the SC/ST group remains uniform throughout.

In our data work, we only consider individuals that report their 3-digit occupation code and education attainment level. Occupation codes are drawn from the National Classification of Occupation (NCO) – 1968. We use the "usual" occupation code reported by an individual for the usual principal activity over the previous year (relative to the survey year). The dataset does not contain information on the years of schooling for the individuals. Instead it includes information on general education categories given as (i) not literate -01, literate without formal schooling: **EGS/ NFEC/ AEC -02, TLC** -03, others -04; (ii) literate: below primary -05, primary -06, middle -07, secondary -08, higher secondary -10, diploma/certificate course -11, graduate -12, postgraduate and above -13. We aggregate those into five similarly sized groups as discussed in the main text. We also convert these categories into years of education. The mapping we used is discussed in the main text and section A.1.3 of this Appendix.

The NSS only reports wages from activities undertaken by an individual over the previous week (relative to the survey week). Household members can undertake more than one activity in the reference week. For each activity we know the "weekly" occupation code, number of days spent working in that activity, and wage received from it. We identify the main activity for the individual as the one in which he spent maximum number of days in a week. If there are more than one activities with equal days worked, we consider the one with paid employment (wage is not zero or missing). Workers sometimes change the occupation due to seasonality or for other reasons. To minimize the effect of transitory occupations, we only consider wages for which the weekly occupation code coincides with usual occupation (one year reference). We calculate the daily wage by dividing total wage paid in that activity over the past week by days spent in that activity.

Lastly, we identify full time workers in our dataset. We assume that an individual is a full time worker if he is employed (based on daily status code) for at least two and half days combined in all activities during the reference week. We drop observations if total number of days worked in the reference week is more than seven.

To summarize, our working sample imposes the following restrictions on the data:
1) The working sample includes all household members in the 16-65 age group who are not enrolled in an educational institution, have education and occupation information and are working full-time.
2) The wage sub-sample includes only those household members from the overall sample for whom wage data are non-missing and non-zero, and for whom the occupation code reported for the last year (relative to the survey year) coincided with the occupation code for which wages over the last week (relative to the survey week) were collected.

A.1.3 Education Categories

Table A1 reports the mapping we used to convert the education categories reported in the NSS data into broader categories that we used in the paper.

Table A1: Education categories

Reported category description	Education category
Not-literate	1
Literate but below primary	2
Primary	3
Middle	4
Secondary	5
Higher secondary	5
Diploma/certificate course	5
Graduate	5
Postgraduate and above	5

A.1.4 Occupation Categories

Table A2 summarizes the one-digit occupation categories in our dataset and presents our grouping of these categories into the Occ 1 - "white collar", Occ 2 - "blue collar" and Occ 3 - "agriculture" groups that we used in the text.

Table A2: Occupation categories

Occupation code	Occupation description	Group
0-1	Professional, technical and related workers	Occ 1
2	Administrative, executive and managerial workers	Occ 1
3	Clerical and related workers	Occ 1
4	Sales workers	Occ 2
5	Service workers	Occ 2
6	Farmers, fishermen, hunters, loggers and related workers	Occ 3
7-8-9	Production and related workers, transport equipment operators and labourers	Occ 2

A.1.5 Employment Status

Here we describe how we classified individuals in our data into those in the labor force, employed, and those who are employed part-time or full-time. This classification was used in Figure 4 in the text. In our dataset, household members are classified into various activity status based on activities pursued by them during previous one week, relative to the survey week (current weekly status) and each day of the reference week (current daily status). We use the "current weekly status" code to classify each individual in working age population into three broad categories of activity status – employed, unemployed and not in labour force. A member is assumed to be "employed" if his/her weekly status is any one from the following: 1) working with an employer under obligation/ not specifically compensated by prevalent wage or salary (38th round only), 2) worked in household

enterprise, 3) worked as helper in household enterprise, 4) worked as regular salaried/wage employee, 5) worked as casual wage labour in public works, 6) did not work though there was work in household enterprise, 7) did not work but had regular salaried/wage employment. A member is classified as "unemployed" if his/her weekly status is either 1) sought work or 2) did not seek but was available for work. The status "not in labour force" is based on the following weekly status code: 1) attended educational institution, 2) attended domestic duty only, 3) too young to work/to attend school/to seek employment, 4) old and disabled, 5) renters, pensioners, remittance recipients etc., 6) beggars, prostitutes, etc., 7) others and 8) did not work due to sickness.

Therefore, "labour force" consists of 1) employed and 2) unemployed. It is implied from the above classification that "self-employed" and "casual workers" are included in "employed" group. Individuals with weekly status "worked in household enterprise" and "worked as helper in household enterprise" are considered as "self-employed". We define a worker as being employed "full-time" if his/her total days of work in the reference week is more than 2.5.

A.2 The DFL method

Our first decomposition of the wage gap between SC/STs and non-SC/STs follows the DFL methodology due to DiNardo, Fortin, and Lemieux (1996). The DFL method in effect involves constructing counterfactual wage densities for the two groups. One such counterfactual density function could be: what would be the wage density function of non-SC/STs if they had SC/ST attributes/characteristics? This counterfactual density function would allow us to assess the contribution of differences in attributes as well as the differences in the underlying wage structure of the two groups in accounting for the wage difference along the entire density function.

In our case of two castes, for each member i of caste $c = S, N$ (where S denotes SC/STs and N denotes non-SC/STs), we observe wages W_{ic} and a vector of attributes X_i. Denote the joint conditional distribution of W_c and X as $F_{W_c,X|c}(.,.)$. The marginal wage distribution of caste c is then obtained as

$$F_{Y_c}(y) = \int F_{Y_c|X}(y|X)\, dF_{X|c}(X) \ , \qquad c = S, N$$

We can also construct a counterfactual wage distribution of SC/STs if they had non-SC/ST attributes but had been paid according to the SC/ST wage structure:

$$F_{Y_S^N}(y) = \int F_{Y_S|X}(y|X)\, dF_{X|N}(X)$$

Analogously, we can also construct the wage distribution of non-SC/STs if they had SC/ST characteristics but been paid according to the non-SC/ST wage structure:

$$F_{Y_N^S}(y) = \int F_{Y_N|X}(y|X)\, dF_{X|S}(X).$$

The key requirement for these to be valid counterfactual distribution functions is that the conditional distribution function $F_{Y_c|X}(y|X)$ remains unchanged when the marginal distribution function $F_{X|c}(X)$ is changed from one group to another.

The main innovation underlying the DFL methodology is the demonstration by DiNardo, Fortin, and Lemieux (1996) that these counterfactual densities can be constructed using a simple reweighting of the actual densities. In particular, it is easy to see that

$$F_{Y_N^S}(y) = \int F_{Y_N|X}(y|X)\, \psi(X)\, dF_{X|N}(X)$$

where $\psi(X) \equiv \frac{dF_{X|S}(X)}{dF_{X|N}(X)}$ is the re-weighting function. Hence, the re-weighting function can be used to write the counterfactual density in terms of the actual density function.

To construct these counterfactual distributions we need to estimate the re-weighting function. To do this, note that z is a dummy variable with $z = 1$ if $c = S$ and $z = 0$ if $c = N$. Bayes' rule implies that

$$\frac{dF_{X|S}(X)}{dF_{X|N}(X)} = \frac{\Pr(X|z=1)}{\Pr(X|z=0)} = \frac{\Pr(z=1|X)/\Pr(z=1)}{\Pr(z=0|X)/\Pr(z=0)}$$

We estimate $\Pr(z = i|X)$ for $i = 0, 1$ by running a linear regression of the probability of belonging to the group on a vector of individual attributes. $\Pr(z = 0)$ and $\Pr(z = 1)$ are estimated using the sample proportions.

A.3 Decompositions of the caste gaps in wages and consumption

We are interested in performing a time-series decomposition of non-SC/STs to SC/STs wage and consumption expenditure gaps between 1983 and 2004-05. We employ a two-fold Oaxaca-Blinder procedure where we use coefficients from a pooled regression with a group membership indicator (as in Fortin, 2006) as the reference coefficients. We use 1983 as the base year for the inter-temporal decomposition, so 1983 is the benchmark sample in our analysis.

Our econometric model for caste group c and round t is given by

$$y_{ct} = X'_{ct}\beta_{ct} + e_{ct}, \qquad c = 1, 2; \text{ and } t = 1, 2,$$

where y_{ct} is a vector of outcomes (log wage or log consumption expenditure) while X_{ct} is the matrix of regressors for caste group c in round t. Here β_{ct} is a coefficient vector, and e_{ct} is the vector of residuals. The differential in expected outcomes between the castes in round t is then given by:

$$\Delta y^e_t = \Delta X'_t\tilde{\beta}_t + X'_{1t}(\beta_{1t} - \tilde{\beta}_t) + X'_{2t}(\tilde{\beta}_t - \beta_{2t}),$$

where $\tilde{\beta}_t$ is the vector of coefficients from the model with both groups pooled The first term above is the explained part while the last two terms give the unexplained parts of the decomposition. Denote E_t to be the explained component of the decomposition, and U_t to be the unexplained part, then

$$\begin{aligned} E_t &= \Delta X'_t\tilde{\beta}_t, \qquad t = 1, 2, \\ U_t &= X'_{1t}(\beta_{1t} - \tilde{\beta}_t) + X'_{2t}(\tilde{\beta}_t - \beta_{2t}), \qquad t = 1, 2. \end{aligned}$$

These decompositions for the first (1983) and last (2004-05) rounds are presented, respectively, in Panels (a) and (b) in Table 7 for wages and Table 10 for consumption.

Next, the inter-temporal change in the outcome differentials can be written as the sum of changes in the explained, E and unexplained, U components:

$$\Delta y^e_2 - \Delta y^e_1 = (E_2 - E_1) + (U_2 - U_1) = \Delta E + \Delta U$$

These differentials are reported in Panel (c) of Tables 7 and 10. Note, however, that inter-temporal changes in the explained and unexplained components may be due to changes in either the attribute gaps or in the returns to those attributes. Since the unexplained part is typically small in our decompositions, we focus on the inter-temporal decomposition of the explained part, ΔE, in the main text. $\Delta E = \Delta X'_2\tilde{\beta}_2 - \Delta X'_1\tilde{\beta}_1$ can be broken down as

$$\Delta E = \Delta X'_2\left(\tilde{\beta}_2 - \tilde{\beta}_1\right) + \left(\Delta X'_2 - \Delta X'_1\right)\tilde{\beta}_1,$$

where the first term is the unexplained part of ΔE, while the second term is the explained part of ΔE. This decomposition is presented in Panel (d) of Tables 7 for wages and 10 for consumption.