

Castes and Labor Mobility*

Viktoriya Hnatkovska[†], Amartya Lahiri[†], and Sourabh B. Paul[†]

August 2011

Abstract

We examine 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 Dufo, 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).

1 Introduction

In this paper we study the impact of the rapid transformation of the Indian economy on a particularly 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.

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 changes in India since the early 1980s affected this goal? 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 discrimination 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. Our analysis yields three main results. First, we find significant convergence in the education attainment levels and the occupation distributions of SC/STs and non-SC/STs between 1983 and 2004-05. Second, we find a statistically significant trend of convergence of consumption and wages of the two groups. The median wage premium of non-SC/STs relative to SC/STs has declined systematically from 36 percent in 1983 to 21 percent 2004-05. 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.

Third, we find that the overall consumption and wage convergence between the groups has been driven significantly by convergence in the education choices of the two groups. In fact for wages, attributes orthogonal to caste account for almost all of the relative wage movements of the two groups. In contrast, the consumption convergence was driven by both a convergence in the covariates as well as a decline in the caste specific penalty for consumption.

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 take advantage of new opportunities opening up in the economy. Using multiple decomposition methods for quantile wages and consumption, we find that there was convergence for the majority of the distribution. In fact, any widening of the gaps was mostly in the top-20 percentiles of households.

In summary, neither the lack of occupational mobility nor the lack of education have proved to be major impediments toward the SC/STs taking advantage of the rapid structural changes in India during this period in which they have rapidly narrowed their huge historical economic disparities with non-SC/STs.

There exists a large literature on labor market discrimination in India.¹ Our study differs from these in that we examine the data for all states and for both rural and urban areas. Moreover, 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.

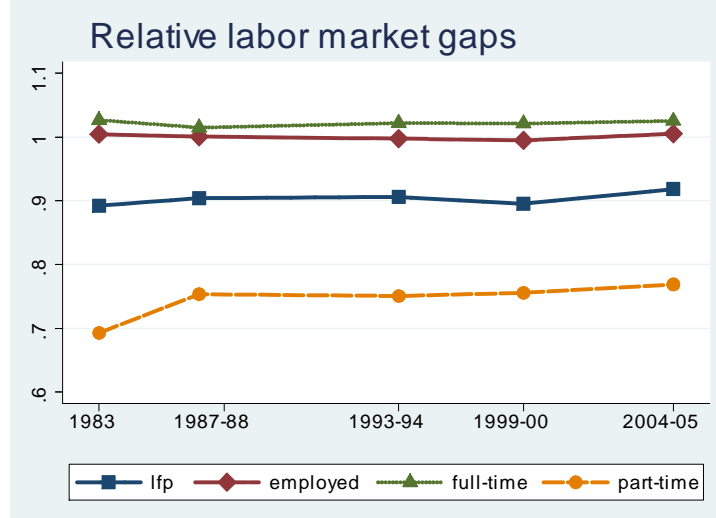
In the next section we describe the data and our constructed measures as well as some summary statistics. Section 3 contrasts SC/STs with their non-SC/ST cohorts in terms of the evolution of the distributions of education attainment rates, occupations, wages and consumption expenditures. There we also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

¹Amongst 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 discrimination in employment in the urban labor market. Ito (2009) has studied both wage and employment discrimination simultaneously by examining data from two Indian states – Bihar and Uttar Pradesh while Kijima (2006) has used NSS data to study consumption inequality of SC/ST households. 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.

2 The Data

Our data comes from various rounds of the National Sample Survey (NSS) of India. 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 per round. Our working sample consists of all individuals between the ages of 16 and 65 belonging to male-headed households.² The sample is restricted to individuals who provided their 3-digit occupation code and their education information. 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 sample of around 163,000-182,000 individuals, depending on the survey round. Importantly, it constitutes our working sample for all empirical work.³

Figure 1: Labor force participation and employment gaps



Note: "lfp" refers to the ratio of labor force participation rate of non-SC/STs to SC/STs. "employed" refers to the ratio of employment rates for the two groups; while "full-time" and "part-time" are, respectively, 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.1.5.

The focus on individuals who are employed full-time is primarily motivated by our interest in wage and consumption convergence patterns across the caste groups. To verify that this restriction does not drive the convergence results presented in the next section, Figure 1 plots the non-SC/ST

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

³Any variation in number of observations across the specifications in the paper are thus solely due to data availability rather than sample restrictions.

to SC/ST ratios in labor force participation rates, overall employment rates, as well as full-time and part-time employment rates.⁴ The noteworthy feature of the figure is that there was very little movement across the rounds in any of these indicators. Therefore, our focus on full-time employed individuals is unlikely to contaminate the results that follow.

Data on wages are more limited. The sub-sample with complete wage data consists of, on average across rounds, about 64,000 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 rural Maharashtra prices.

We also study consumption expenditures, data on which is available at the household level. To be included in the sample, the restrictions we imposed on our working sample must be satisfied (i.e. age of 16-65, employed full-time, availability of education and occupation information, and being from a male-headed household), but by the head of the household. Thus, our consumption expenditure sample consists of about 87,000 households per round. 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.⁵ 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 of our formal 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.⁶ To

⁴In our dataset, household members are classified into activity status based on activities pursued by them during the previous one week, relative to the survey week and each day of the reference week. We used this classification to assign each individual into the four different labor force and employment groups. The details as to how these groups were constructed are available in the Appendix A.1.5. The sample underlying 1 satisfies all the restrictions imposed for the working sample except for the full-time employment restriction.

⁵Even though consumption data is only available at the household level, it does not make the breakdown of households into SC/STs and non-SC/STs too problematic since inter-caste marriages in India are rare.

⁶We 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

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 these categories and their conversion into education years are in Appendix A.1.

Table 1 gives some summary statistics of the data on the individual and household characteristics of SC/STs and non-SC/STs and their changes over time. The table indicates that non-SC/STs are, on average, slightly older, have more years of education and are more likely to live in urban areas than SC/STs. While these have been stable over time, one characteristic in which there has been change is household size. The household size of non-SC/STs has remained larger than SC/ST households throughout but the difference narrowed significantly across the sample rounds.

Table 1: Sample summary statistics

	(a) Individuals			(b) Households		
Non-SC/ST	age	edu years	married	male	rural	hh size
1983	35.34 (0.05)	3.62 (0.02)	0.80 (0.00)	0.82 (0.00)	0.72 (0.00)	6.67 (0.02)
1987-88	35.63 (0.04)	3.84 (0.02)	0.81 (0.00)	0.81 (0.00)	0.74 (0.00)	6.46 (0.02)
1993-94	36.00 (0.05)	4.58 (0.02)	0.80 (0.00)	0.81 (0.00)	0.72 (0.00)	6.07 (0.02)
1999-00	36.33 (0.05)	5.11 (0.02)	0.81 (0.00)	0.79 (0.00)	0.71 (0.00)	6.17 (0.02)
2004-05	37.04 (0.05)	5.52 (0.02)	0.80 (0.00)	0.80 (0.00)	0.72 (0.00)	5.98 (0.02)
SC/ST						
1983	34.69 (0.08)	1.41 (0.02)	0.83 (0.00)	0.73 (0.00)	0.85 (0.00)	6.16 (0.03)
1987-88	34.72 (0.07)	1.58 (0.02)	0.84 (0.00)	0.73 (0.00)	0.87 (0.00)	6.03 (0.02)
1993-94	35.26 (0.07)	2.08 (0.02)	0.82 (0.00)	0.74 (0.00)	0.86 (0.00)	5.74 (0.02)
1999-00	35.58 (0.08)	2.64 (0.03)	0.82 (0.00)	0.71 (0.00)	0.84 (0.00)	5.92 (0.03)
2004-05	36.28 (0.09)	3.18 (0.03)	0.81 (0.00)	0.74 (0.00)	0.84 (0.00)	5.88 (0.03)
Difference						
1983	0.64*** (0.09)	2.21*** (0.02)	-0.03*** (0.00)	0.09*** (0.00)	-0.13*** (0.00)	0.51*** (0.04)
1987-88	0.91*** (0.08)	2.27*** (0.02)	-0.03*** (0.00)	0.08*** (0.00)	-0.13*** (0.00)	0.42*** (0.03)
1993-94	0.75*** (0.09)	2.50*** (0.03)	-0.02*** (0.00)	0.07*** (0.00)	-0.14*** (0.00)	0.33*** (0.03)
1999-00	0.75*** (0.09)	2.48*** (0.03)	-0.02*** (0.00)	0.08*** (0.00)	-0.13*** (0.00)	0.26*** (0.03)
2004-05	0.76*** (0.10)	2.34*** (0.04)	-0.01*** (0.00)	0.06*** (0.00)	-0.12*** (0.01)	0.10*** (0.04)

Notes: This table reports summary statistics for our sample. Panel (a) gives the statistics at the individual level, while panel (b) gives the statistics at the level of a household. Panel labelled "Difference" reports the difference in characteristics between non-SC/STs and SC/STs. Standard errors are reported in parenthesis. * p-value \leq 0.10, ** p-value \leq 0.05, *** p-value \leq 0.01.

then is to combine the two categories.

3 How Have the Scheduled Castes Fared?

There are three issues of interest: (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. We examine them sequentially.

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.

Table 2: Education Gap: Years of Schooling

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

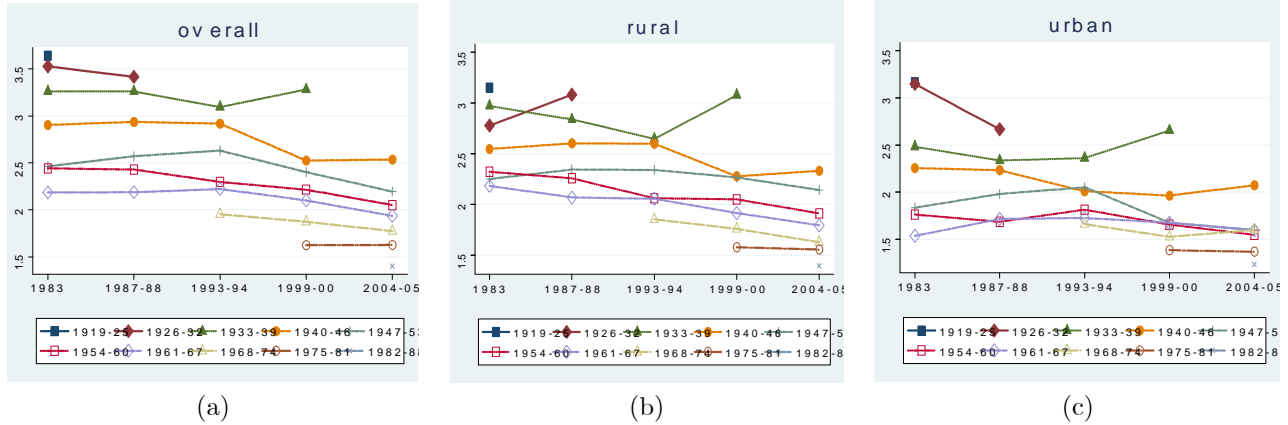
Notes: This table presents the average years of education for the overall sample and separately for non-SC/STs and SC/STs. The reported statistics are obtained for each NSS survey round which is shown in the first column. Standard errors are in parenthesis.

Do the overall trends in years of education mask key differences in the relative movements within more disaggregated cohorts? To investigate this we compute the average years of education of non-SC/STs relative to SC/STs within various birth cohorts across survey round. Figure 2 plots the result. Panel (a) reveals that for all rounds, education gaps decline as birth cohorts become younger. Importantly, the education gaps have declined even within the same birth cohort over time.⁷ Given the large concentration of households in rural areas, a related question is whether the trends in

⁷The education attainment gaps of a given birth cohort can change over successive rounds primarily due to individuals acquiring more education during their life-cycle. 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 older cohorts would change over time due to them.

education attainment are different between rural and urban households. Figure 2 shows that the years of education have been converging in both rural and urban areas.⁸

Figure 2: Education gaps by birth cohorts



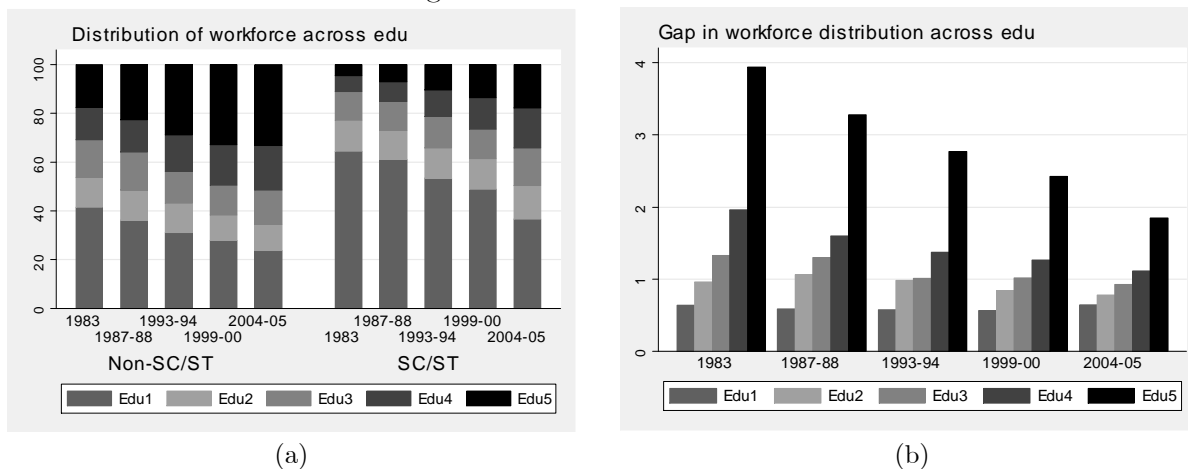
Notes: The figures show the evolution of the relative gap in years of education between non-SC/STs and SC/STs over time for different birth cohorts. Panel (a) presents the results for the overall sample, while panels (b) and (c) report the results for rural and urban households separately.

While the changes in the average years of education are interesting, they mask potentially interesting changes in the underlying composition of education attainment levels. 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 answer this question using Figure 3. Given the similar patterns across the birth cohorts and for ease of presentation, we report the statistics for the entire sample.

Panel (a) of Figure 3 shows the distribution of the workforce across five broad education categories as well as the evolution of the distribution over the successive rounds. Panel (a) shows the distributions of non-SC/STs (left set of bars) and SC/STs (right set of bars). Clearly, SC/STs are systematically less educated than non-SC/STs. 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 77 percent of SC/ST were either illiterate or had below primary level education while the corresponding number for non-SC/STs was 54 percent. These numbers declined to 50 percent for SC/ST and 34 percent for non-SC/ST by 2004-05.

⁸The trends towards convergence in education attainments of SC/STs and non-SC/STs remain robust to using age groups instead of birth cohorts. In particular, when relative education gaps for various age groups are computed over time, we find that older age groups have larger gaps, but there is a clear pattern of education convergence across the different age-cohorts over time.

Figure 3: Education distribution



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

The figure also shows 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 illiterate SC/ST (category 1) fell from 65 percent to 37 percent while for non-SC/STs the proportion of illiterates fell from just under 42 percent to 24 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 3 reports the relative gaps in the education distribution of the two groups. The bar for any education category j is obtained by dividing the share of non-SC/STs in category j (among all non-SC/STs) by the share of SC/STs in category j (among all SC/STs). The deviation of the height of the bar from one for any category j indicates the degree to which there is a disproportionate presence of one group in that category. Panel (b) of Figure 3 clearly shows that the heights of the bars either stayed constant or tended towards one over time indicating a convergent trend for the two groups. Perhaps the sharpest convergence has been in category 5 (secondary or higher). However, the absolute difference between the two groups still remains very high.

Is this convergence statistically significant? We examine this by estimating ordered probit regressions of the broad education categories 1 to 5 on a constant and an SC/ST dummy for each sample round (where the dummy is equal to 1 if the individual belongs to SC/STs, and zero otherwise). Panel (a) of Table 3 reports the marginal effect of the SC/ST dummy on the probability of an indi-

vidual belonging to each education category; while panel (b) reports changes in the marginal effects for the consecutive decades and over the entire sample period of 1983 to 2005. Clearly, SC/STs are more likely to be illiterate (education category 1), and less likely to be in the higher education categories relative to non-SC/STs. This remained true across all the rounds. However, there is a significant trend towards convergence in education attainments of the two groups over time in every category except secondary education. In particular, the marginal effects of the 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.⁹

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

	Panel (a): Marginal effects, unconditional					Panel (b): Changes		
	1983	1987-88	1993-94	1999-2000	2004-05	83 to 93	93 to 05	83 to 05
Edu 1	0.2650 (0.0031)	0.2688 (0.0028)	0.2622 (0.0030)	0.2410 (0.0034)	0.2112 (0.0035)	-0.0029 (0.0043)	-0.0509*** (0.0047)	-0.0538*** (0.0047)
Edu 2	-0.0308 (0.0006)	-0.0281 (0.0006)	-0.0158 (0.0005)	-0.0040 (0.0003)	0.0072 (0.0003)	0.0150*** (0.0008)	0.0230*** (0.0005)	0.0380*** (0.0007)
Edu 3	-0.0619 (0.0010)	-0.0587 (0.0008)	-0.0405 (0.0007)	-0.0251 (0.0006)	-0.0167 (0.0005)	0.0214*** (0.0012)	0.0238*** (0.0009)	0.0453*** (0.0011)
Edu 4	-0.0687 (0.0010)	-0.0646 (0.0009)	-0.0645 (0.0009)	-0.0587 (0.0010)	-0.0529 (0.0011)	0.0042*** (0.0014)	0.0115*** (0.0014)	0.0158*** (0.0014)
Edu 5	-0.1036 (0.0012)	-0.1174 (0.0012)	-0.1414 (0.0015)	-0.1531 (0.0020)	-0.1488 (0.0023)	-0.0378*** (0.0020)	-0.0074*** (0.0027)	-0.0452*** (0.0026)
N	165034	182384	163132	171153	167857			

Notes: Panel (a) reports 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 for each survey round. Panel (b) of the table reports the change in the marginal effects over successive decades and over the entire sample period. N refers to the number of observations. Standard errors are in parenthesis. * p-value \leq 0.10, ** p-value \leq 0.05, *** p-value \leq 0.01.

Despite the overall significant trend toward convergence in education levels across castes, it is still important to note that for SC/STs, the probability of secondary school education has declined over the sample period. This suggests that not all sections of SC/STs may have benefited equally during this period.

We should note that the marginal effects of the SC/ST dummy reported in Table 3 estimate the absolute gaps between the two groups in the probabilities of belonging to a given education category. The bars in the panel (b) 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 while the absolute gap has widened in the top education category.

⁹The detailed estimation results from this regression and the ones that will follow are available in an online supplement from <http://faculty.arts.ubc.ca/vhmatkovska/Research/castes-supplement.pdf>

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. 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 have 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 4 shows the occupation distribution, by caste, for the working population in our sample. Panel (a) shows that, there has been a systematic decline in Occ 3 (farming/pastoral activities) between 1983 and 2004-05 for both groups. Correspondingly, the largest expansion in the employment share of both groups has been in Occ 2 which comprises mostly low skill blue collar and service sector jobs. These trends reflect the undergoing structural transformation in India characterized by a gradual decline in the agricultural sector and an expansion in services.

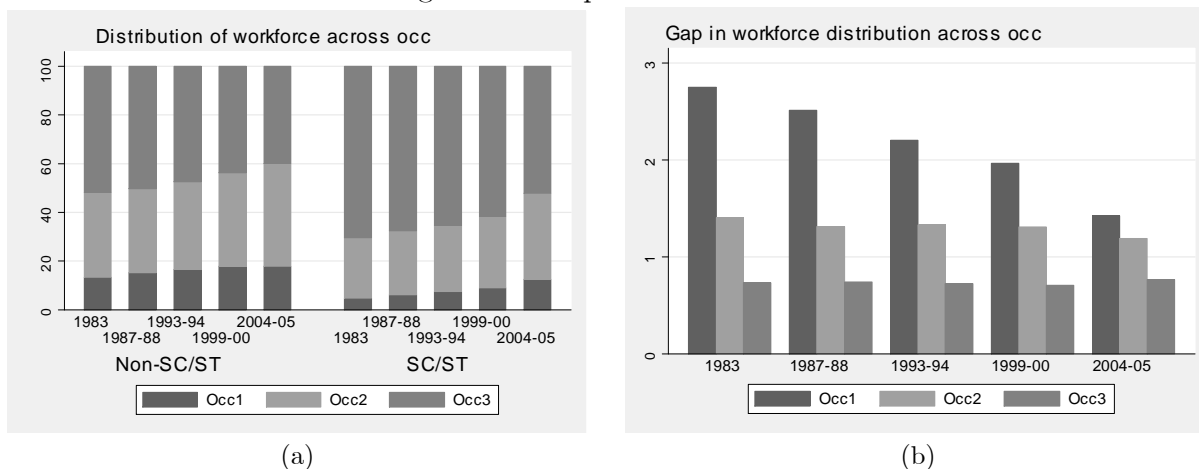
Since SC/STs were over-represented in Occ 3 and under-represented in Occ 1 and 2 in 1983, the trends in occupation shares of the two groups imply that the overall occupation distributions have been converging. This can be seen in Panel (b) of Figure 4 where the bar for any occupation category j is obtained by dividing the share of non-SC/STs in category j by the share of SC/STs in that category. Clearly, most of the bars are converging toward one over time.

To evaluate the statistical significance of the convergent trends in the occupation distributions, we estimate a multinomial probit model of occupation on a constant and an SC/ST dummy for every survey round. Table 4 reports the marginal effect of the SC/ST dummy on the probability of being

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

¹¹As 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 4: Occupation distribution



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

employed in each occupation, as well as the changes in the marginal effects over time. Panel (a) shows that SC/STs were less likely to be employed in Occ 1 (white collar occupations) and 2 (blue collar occupations) and more likely to work in Occ 3 (agrarian occupations) as compared to non-SC/STs throughout our sample period. However, Panel (b) shows that the negative SC/ST effect on the probability of being employed in Occ 2 and the positive SC/ST effect on the probability of Occ 3 employment have both declined significantly over time. Hence, the convergence in the employment shares has been significant.

It is important to note though that this period has also been marked by a 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). Hence, there again appears to have been a non-uniform convergence in the occupation distribution of the two groups across the categories.¹²

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

¹²We should emphasize 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 4, on the other hand, depict the relative gaps in the probabilities.

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

	Panel (a): Marginal effects, unconditional					Panel (b): Changes		
	1983	1987-88	1993-94	1999-2000	2004-05	83 to 93	93 to 05	83 to 05
Occ 1	-0.0701*** (0.0016)	-0.0734*** (0.0014)	-0.0778*** (0.0016)	-0.0760*** (0.0021)	-0.0780*** (0.0021)	-0.0077*** (0.0022)	-0.0002 (0.0026)	-0.0079*** (0.0026)
Occ 2	-0.0735*** (0.0033)	-0.0663*** (0.0030)	-0.0833*** (0.0031)	-0.0769*** (0.0034)	-0.0472*** (0.0039)	-0.0098** (0.0045)	0.0362*** (0.0050)	0.0264*** (0.0051)
Occ 3	0.1436*** (0.0034)	0.1397*** (0.0031)	0.1611*** (0.0032)	0.1529*** (0.0036)	0.1251*** (0.0040)	0.0175*** (0.0047)	-0.0360*** (0.0052)	-0.0185*** (0.0053)
N	165034	182384	163132	171153	167857			

Note: Panel (a) of the 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 for each survey round. Panel (b) reports the change in the marginal effects of SC/ST dummy over successive decades and over the entire sample period. N refers to the number of observations. 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.

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 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.

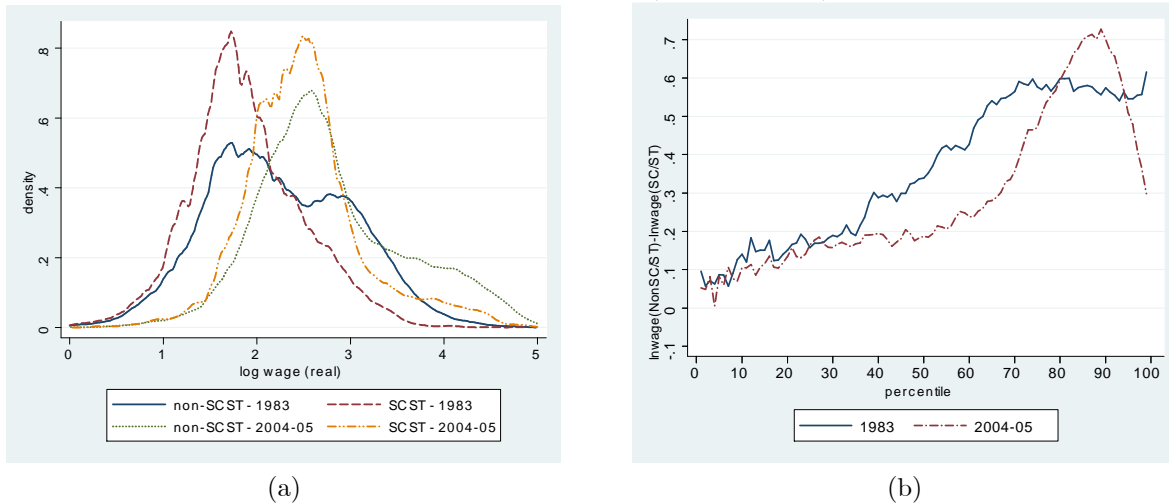
3.3.1 Wages

We start 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 while Panel (b) examines the changes in wage inequality by plotting the differences in log wages between non-SC/STs and SC/STs for different percentiles of their wage distributions for the two survey rounds. The figure shows that the wage distribution of both groups 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. Moreover, panel (b) shows that the density functions for the two groups have come much closer together over time.¹³

Panel (b) of Figure 5 reveals several additional features of the wage distributions. Both lines

¹³We 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.

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



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.

slope up and to the right, indicating that the wage distribution of non-SC/STs is more unequal than 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 as the line for 2004-05 round is well below the line for 1983 round for all except the 80th to 95th percentiles.¹⁴

We now examine the caste wage gap more closely by contrasting SC/STs with non-SC/STs at different points of the wage distribution for all the survey rounds.¹⁵ We estimate simple regressions of log wages on a constant and an SC/ST dummy variable. Since wages are likely changing over the course of one’s lifecycle, in our estimation we control flexibly for age by including individual age and

¹⁴For 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).

¹⁵It is important to note an oddity in the wage data for the 43th round (1987-88). 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 decline in rural wage observations for this round. We are not sure as to the reasons for this sudden decline. In order to avoid spurious comparisons, we drop the 43rd round from all our statistical analysis of the wage data.

age squared terms in the regression. We assess the effect of caste on the mean and unconditional quantiles of wages such as the median, 10th and 90th quantiles. For the latter we use the Recentered Influence Function (RIF) method of Firpo, Fortin, and Lemieux (2009). We use RIF regressions as we are interested in determining the effect of caste on the unconditional quantile of wages. Since the law of iterated expectations does not apply to quantiles, we cannot use a conditional regression approach to determine the effect of covariates on the unconditional quantiles. A RIF regression gets around this issue. Firpo, Fortin, and Lemieux (2009) show that for any quantile, it amounts to running a linear probability regression of the probability of wages exceeding that quantile on a vector of regressors. Details regarding the RIF method are provided in the Appendix A.3.

Table 5 reports our results. Panel (a) presents the estimated coefficient on the SC/ST dummy for each survey round. 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. Those results are presented in panel (b) of Table 5 over the two consecutive decades and over the entire sample period.

Table 5: Wage gaps and changes

	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	-0.1486*** (0.0127)	-0.0651*** (0.0101)	-0.0936*** (0.0087)	-0.1022*** (0.0133)	0.0835*** (0.0162)	-0.0371** (0.0167)	0.0464*** (0.0184)
50th quantile	-0.3563*** (0.0099)	-0.2884*** (0.0087)	-0.2473*** (0.0089)	-0.2102*** (0.0090)	0.0679*** (0.0132)	0.0782*** (0.0125)	0.1461*** (0.0134)
90th quantile	-0.4417*** (0.0103)	-0.4645*** (0.0126)	-0.5048*** (0.0203)	-0.5778*** (0.0214)	-0.0228 (0.0163)	-0.1133*** (0.0248)	-0.1361*** (0.0237)
mean	-0.3524*** (0.0075)	-0.3222*** (0.0086)	-0.3080*** (0.0088)	-0.2927*** (0.0087)	0.0302*** (0.0114)	0.0295*** (0.0122)	0.0597*** (0.0115)
N	64009	63366	66426	61004			

Note: Panel (a) of this table reports the estimates of coefficients on SC/ST dummy from RIF regressions of log wages on an SC/ST dummy, age, age squared, and a constant. 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. N refers to the number of observations. Standard errors are in parenthesis. * p-value \leq 0.10, ** p-value \leq 0.05, *** p-value \leq 0.01.

Panel (a) of Table 5 shows that the coefficient on the SC/ST dummy variable is negative and significant throughout. Hence, the mean or quantile wage of SC/STs was significantly lower than similarly aged non-SC/ST. Interestingly, the wage gaps are the lowest at the bottom end of the distribution, and the highest at the top end. Panel (b) of Table 5 shows that relative wage gaps between the two groups have declined over time along the entire distribution, except the very top end. The sharpest decline occurred at the median of the distribution, where the wage gap has fallen from 36 percent to 21 percent. Furthermore, the majority of these changes are significant. Hence, there appears to have been a significant convergence in wages across castes during 1983-2005 period.

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.¹⁶ In contrast, our computations above imply that the median wage of SC/STs relative to non-SC/STs has increased secularly from 64 percent in 1983 to 79 percent in 2004-05.¹⁷ 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 been converging due to convergence in the determinants of wages such as education, 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 education?

To address these questions, we proceed in two steps. First, we apply 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 within a sample round into two components – the part that is explained by differences in 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 regressors X 's while the rest of the gap is unexplained and arises due to differences in the coefficients β 's or the wage structure of the two groups. The advantage of the DFL method is that it can be used to study the entire distribution rather than a subset of points in the distribution. Second, we use a Oaxaca-Blinder decomposition of *changes* in the wage gap *over time* for different quantiles of the distribution in order to assess the relative importance of different attributes in accounting for the observed wage convergence. We use the RIF methodology of Firpo, Fortin, and Lemieux (2009) due to its appropriateness for quantile regressions.

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

¹⁶These numbers are from US Current Population Survey.

¹⁷For 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.

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. Details regarding the procedure are provided in the Appendix A.2.

To evaluate the contribution of different attributes to the wage gap between SC/STs and non-SC/STs, we introduce attributes sequentially. First, we evaluate the role of demographic characteristics such as individual’s age, age squared, rural sector of residence, and region of residence. 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 DFL decomposition. Next, we add education to the set of characteristics and obtain the incremental contribution of education to the observed wage convergence.¹⁸ Finally, we introduce state-level SC/ST reservation quotas to obtain the share of convergence that is accounted for by reservation policies. Reservations for SC/STs (in proportion to their population shares) in public sector employment and in higher education institutions was a key policy initiative in India to correct historical inequities.¹⁹ Hence, it is important to assess their contribution to the observed wage convergence across castes.²⁰

Figure 6 shows the actual wage gaps between non-SC/STs and SC/STs for each percentile and the gaps explained by differences in attributes of the two groups, where we introduced the attributes sequentially. The explained gap in each case is computed as the difference between the actual and re-weighted non-SC/ST wage percentiles, where we re-weight the wage density of non-SC/STs by giving them the attributes of SC/STs. Panel (a) shows the results for 1983, while panel (b) reports them for 2004-05.

Strikingly, Figure 6 shows that attributes can account for almost all of the wage gap between the groups. Differences in demographic characteristics do explain a substantial portion of the SC/ST wage gap, especially in 2004-05; however, the majority of the observed difference in wages between the two groups is attributed to difference in education attainments.²¹ Reservation quotas, however,

¹⁸The region groupings reflect similarities across states along their geographic characteristics, and characteristics that are shared based on proximity. Education is introduced through a set of education dummies reflecting education categories 1-5.

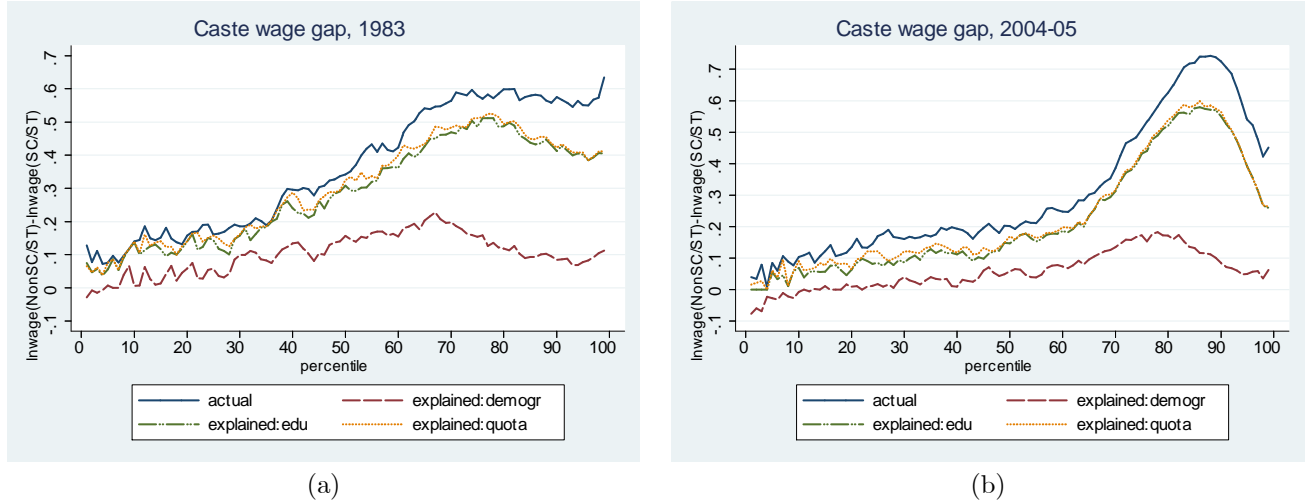
¹⁹Reservations have been a focus of attention by a number of researchers. Thus, Pande (2003) examines the effects of reservations on actual policies while Prakash (2009) studies the effects of reservations on the labor market outcomes of SC/STs. Both authors find evidence of positive effects of reservations on the targeted communities.

²⁰Since reservations for SC/STs were provided in proportion to their population shares, state-level reservations can change over time due to changes in SC/ST population shares in the states.

²¹We should note that results from the sequential introduction of regressors can be sensitive to the order in which regressors are introduced, as has been shown by Gelbach (2009). However, we found that the Oaxaca-Blinder decomposition of the wage gap using OLS and RIF regressions for the mean and quantile wages, respectively, yielded similar results with the majority of the wage gap being accounted for by differences in the attributes of the two caste groups, especially their education attainments.

play a negligible role.^{22,23}

Figure 6: The percentile wage gaps: Decompositions



Notes: Each panel shows the actual log wage gap between non-SC/STs and SC/STs for each percentile, and the counterfactual percentile log wage gaps when non-SC/STs are sequentially given SC/ST attributes. Three sets of attributes are considered: demographic (denoted by "demogr"), demographics plus education ("edu"), and all of the above plus SC/ST reservation quotas ("quota"). The left panel is for 1983 while the right panel shows the decomposition for 2004-05.

What factors account for the changes in the wage gap between 1983 and 2004-05? We use the Oaxaca-Blinder method to decompose the observed changes in the mean and quantile wage gaps into explained and unexplained components as well as to quantify the contribution of the key individual covariates. We rely on OLS regressions for the decomposition at the mean, and on the RIF regressions for decompositions at the 10th, 50th, and 90th quantiles. Our DFL results for the entire wage distribution suggest that focusing on the two tails and the middle of the distribution, while parsimonious, should provide a representative view of relevant distributional developments.²⁴

The set of covariates that we use is the same as the one we used in DFL decompositions. In addition, we include a Muslim dummy in our evaluation to control for the fact that Muslims, on average, have done poorly in modern India (post independence in 1947). If we do not control for

²²In 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.

²³To further understand the effect of reservations, we also computed SC/ST wage gaps separately for public and non-public sector employees. The nature of employer, unfortunately, is only available for 1987-88, 1993-94 and 2004-05 survey rounds and only for salaried workers. This significantly reduces the number of observations in the wage sample. We find that wages of public sector workers are systematically higher for both SC/STs and non-SC/STs, but the wage gaps in public and non-public sectors are comparable in magnitude.

²⁴All decompositions are performed using a pooled model across SC/STs and non-SC/STs as the reference model. Following Fortin (2006) we allow for a group membership indicator in the pooled regressions. We also used 1983 round as the benchmark sample. Details of the decomposition method can be found in Appendix A.4.

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. Table 6 presents our results.

Table 6: Decomposing changes in SC/ST wage gaps over time

(a). Change (1983 to 2004-05)		explained			
	(i) measured gap	(ii) explained	(iii) unexplained	(iv) education	(v) quota
10th quantile	-0.0227 (0.0198)	-0.0827*** (0.0126)	0.0599*** (0.0233)	-0.0377*** (0.0086)	-0.0161** (0.0075)
50th quantile	-0.1408*** (0.0266)	-0.1431*** (0.0140)	0.0023 (0.0227)	-0.0842*** (0.0103)	-0.0311*** (0.0043)
90th quantile	0.1527*** (0.0377)	0.0997*** (0.0265)	0.0530 (0.0353)	0.1370*** (0.0264)	-0.0066 (0.0053)
mean	-0.0639*** (0.0197)	-0.0749*** (0.0116)	0.0110 (0.0165)	-0.0340*** (0.0097)	-0.0141*** (0.0028)
(b). Change in explained component					
10th quantile	-0.0827*** (0.0126)	-0.0232*** (0.0064)	-0.0594*** (0.0110)	-0.0278*** (0.0048)	-0.0028 (0.0029)
50th quantile	-0.1431*** (0.0140)	-0.0342*** (0.0134)	-0.1089*** (0.0104)	-0.0349*** (0.0080)	-0.0034 (0.0037)
90th quantile	0.0997*** (0.0265)	0.0135 (0.0163)	0.0863*** (0.0204)	0.0131 (0.0150)	-0.0005 (0.0007)
mean	-0.0749*** (0.0116)	-0.0205* (0.0124)	-0.0545*** (0.0062)	-0.0213** (0.0091)	-0.0025 (0.0025)

Note: Panel (a) presents the change in the SC/ST wage gap between 1983 and 2004-05. Panel (b) 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. Both panels also report the contribution of education and SC/ST reservation quota to the explained gaps. Bootstrapped standard errors are in parenthesis. * p-value \leq 0.10, ** p-value \leq 0.05, *** p-value \leq 0.01.

Panel (a) of Table 6 reports the changes in the SC/ST wage gap between 1983 and 2004-05 for the 10th, 50th, 90th quantiles, and the mean (column (i)). Column (ii) reports how much of that differential is explained by characteristics, while column (iii) shows how much is unexplained; columns (iv) and (v) report, respectively, how much of the explained difference is due to education and how much is due to reservation quotas. Bootstrapped standard errors are in parenthesis.²⁵ Clearly, the mean wage gap has declined, except at the very top end of wage distribution, and almost all of it is accounted for by changes in measured characteristics, in particular education. These variations, however, mask changes in both the individual endowments and returns to these endowments over our sample period.

Panel (b) shows the decomposition of the changes in the explained component of the wage gap during 1983-2005 period into a part attributable to changes in gaps in the observables (explained

²⁵The standard errors for the unexplained part of the decomposition are generally hard to derive if a pooled model is used. Furthermore, in our decompositions we are 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.

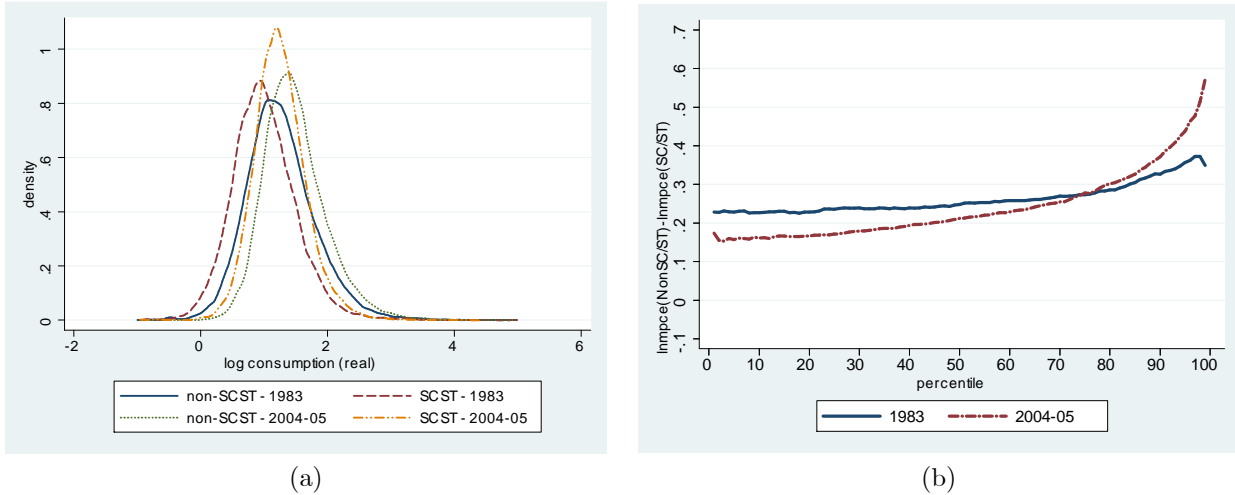
component), and a part that is due to changes in the returns to these observables over time (unexplained component).²⁶ We find that changes in education and reservation quota have accounted for about 40 percent of the change in the explained gap at the 10th quantiles and for about 30 percent at the mean and 50th quantile. 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 characteristics of the two social groups with education having played a large role in driving the convergence in wages during the 1983-2005 period.

3.3.2 Consumption

The last indicator of interest to us is household consumption. In particular, we compare per capita real consumption expenditures of SC/STs with non-SC/STs and their evolutions over time. We start with the distribution of (log) consumption for the two groups for 1983 and 2004-05 in Panel (a) of Figure 7 and the (log) 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.

Figure 7: The log consumption distributions of non-SC/STs and SC/STs for 1983 and 2004-05



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.

Panel (b) of Figure 7 reveals that the consumption gap between non-SC/STs and SC/STs has 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

²⁶This 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.

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 wage data.²⁷

Are the changes in the consumption expenditure gaps significant? Table 7 reports the results from the mean and quantile regressions of log consumption expenditure on a constant and an SC/ST dummy (panel (a)), as well as the changes in that coefficient over consecutive decades and over the entire sample period (panel (b)). The results mirror the convergence results for wages. The SC/ST dummy is significant and negative across all the rounds. The coefficient is quantitatively similar for all the quantiles we considered and for the mean, indicating that the SC/ST consumption gap is of similar magnitude along the entire distribution. The estimated value of the SC/ST dummy coefficient has declined over time for the entire distribution, except for the very top, where it has widened somewhat. Over the entire period, the consumption gap has narrowed by almost 8 percent for the bottom quantiles and about 4 percent at the median.

Table 7: Consumption gaps and changes

Panel (a): SC/ST dummy coefficient					
	1983	1987-88	1993-94	1999-2000	2004-05
10th quantile	-0.2497*** (0.0101)	-0.2081*** (0.0089)	-0.1747*** (0.0085)	-0.1867*** (0.0086)	-0.1721*** (0.0101)
50th quantile	-0.2550*** (0.0067)	-0.2252*** (0.0057)	-0.2295*** (0.0057)	-0.2014*** (0.0061)	-0.2157*** (0.0065)
90th quantile	-0.2805*** (0.0092)	-0.2883*** (0.0083)	-0.2845*** (0.0091)	-0.2611*** (0.0108)	-0.3229*** (0.0091)
mean	-0.2585*** (0.0059)	-0.2402*** (0.0052)	-0.2314*** (0.0048)	-0.2128*** (0.0051)	-0.2324*** (0.0054)
N	87364	93702	87098	87725	87102
Panel (b): Changes					
	1983 to 1993-94	1993 to 2004-05	1983 to 2004-05		
10th quantile	0.0750*** (0.0132)	0.0026 (0.0132)	0.0776*** (0.0143)		
50th quantile	0.0255*** (0.0088)	0.0138* (0.0086)	0.0393*** (0.0093)		
90th quantile	-0.0040 (0.0129)	-0.0384*** (0.0129)	-0.0424*** (0.0129)		
mean	0.0271*** (0.0076)	-0.0010 (0.0072)	0.0261*** (0.0080)		

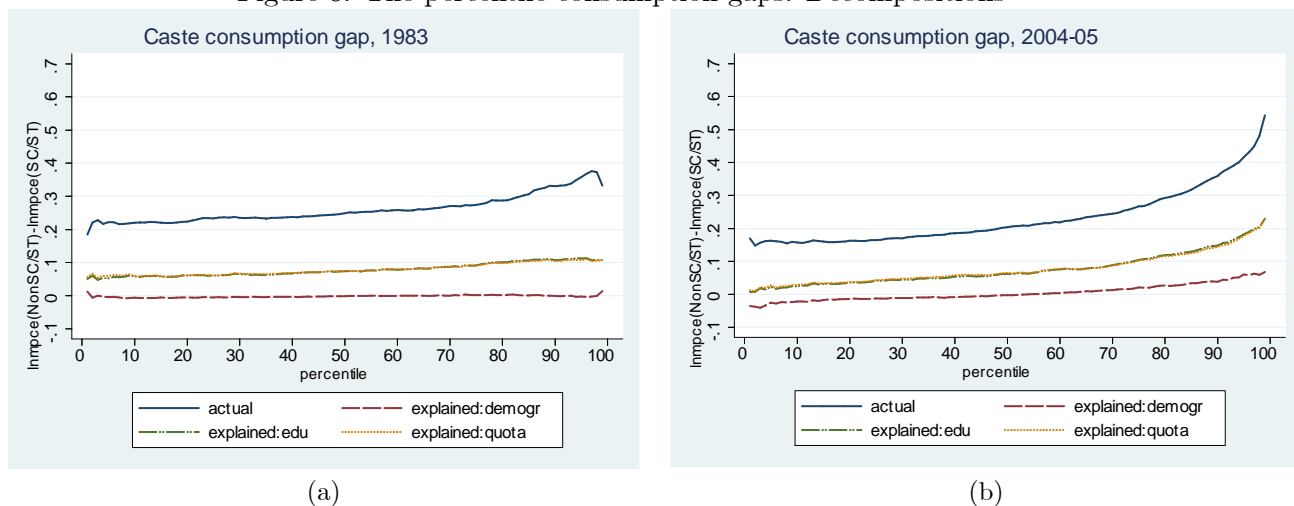
Note: Panel (a) reports the estimates of the coefficient on SC/ST dummy from RIF regressions of log consumption expenditures on an SC/ST dummy and a constant. Panel (b) reports the changes in the estimated coefficients over successive decades and the entire sample period. N refers to the number of observations. Standard errors are in parenthesis. * p-value \leq 0.10, ** p-value \leq 0.05, *** p-value \leq 0.01.

Do SC/ST households have lower consumption levels after one controls for household characteristics? We use the DFL decomposition to answer this. Figure 8 reports the actual log consumption gap

²⁷Note that the convergence in per capita consumption expenditures between SC/STs and non-SC/STs is not due to uneven changes in the household size of the two groups over time. While both types of households saw their size fall over time, the decline was faster for non-SC/STs. If anything, this would have contributed to a divergence in consumption expenditures when measured in per capita terms. Thus, the convergence in aggregate household-level consumption must be even more pronounced.

for non-SC/STs relative to SC/STs computed for every percentile; and the explained gaps, 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 SC/ST attributes. As with wages, we examine the influence of individual covariates by introducing them sequentially. We first consider demographic characteristics, which include household size, the number of earning members of the household, rural dummy and regional dummies. Then we add education attainments, which consist of the education attainments of the household head and the highest level of education attained in the household. Finally, we include state-level reservation quotas for SC/STs to account for policy. Panel (a) shows the results for 1983, while panel (b) reports them for 2004-05.

Figure 8: The percentile consumption gaps: Decompositions



Notes: Each panel shows the actual log consumption gap between non-SC/STs and SC/STs for each percentile, and the counterfactual percentile log consumption gaps when non-SC/STs are sequentially given SC/ST attributes. Three sets of attributes are considered: demographic (denoted by "demogr"), demographics plus education ("edu"), and all of the above plus SC/ST reservation quotas ("quota"). The left panel is for 1983 while the right panel shows the decomposition for 2004-05.

Demographic characteristics do not contribute much to the observed consumption gaps between SC/STs and non-SC/STs in either period. However, when education is added to the set of characteristics, the explained component becomes much more important, accounting for 27 percent of the actual consumption gap at the bottom end of the consumption distribution and for about 33 percent at the top end in 1983. In 2004-05 the contribution of the demographic characteristics and education was only 15 percent at the bottom end of the consumption distribution, but increased to about 43 percent at the top end of the consumption distribution. As with wages, reservation quotas do not contribute much to the consumption gaps. Overall, these results are similar to the corresponding ones for wages, except that the consumption gaps have a larger unexplained component.

As with wages, we now formally evaluate the contribution of the various attributes to changes in the consumption gap over time by applying the RIF and OLS approaches in combination with the Oaxaca-Blinder decomposition to the 10th, 50th and 90th consumption quantiles as well as the mean. We use the complete set of characteristics from the DFL approach as control variables. As for wages, we also include a Muslim dummy.

Panel (a) of Table 8 shows that the mean, *median* and 10th quantile of consumption gap between non-SC/STs and SC/STs have all declined significantly, while the gap has widened somewhat at the top of the consumption distribution. Changes in education attainments have accounted for a large fraction of the decline in the measured gaps; but the unexplained component has also been substantial. Panel (b) of Table 8 shows that at the 90th quantile, most of the change in the explained gap between 1983 and 2004-05 is due to changes in individual attributes. For the lower end of the distribution however, intertemporal changes in the return differentials contributed the most to the change in the explained gap.

Table 8: Decomposing changes in SC/ST consumption expenditure gaps over time

(a). Change (1983 to 2004-05)					
	(i) measured gap	(ii) explained	(iii) unexplained	(iv) education	explained (v) quota
10th quantile	-0.0653*** (0.0169)	-0.0342*** (0.0070)	-0.0311** (0.0155)	-0.0258*** (0.0046)	-0.0044 (0.0033)
50th quantile	-0.0449*** (0.0107)	-0.0183*** (0.0067)	-0.0267*** (0.0104)	-0.0231*** (0.0046)	-0.0025 (0.0020)
90th quantile	0.0301* (0.0185)	0.0368*** (0.0101)	-0.0067 (0.0177)	-0.0029 (0.0082)	-0.0071*** (0.0029)
mean	-0.0268** (0.0110)	-0.0079 (0.0061)	-0.0189* (0.0101)	-0.0175*** (0.0044)	-0.0039** (0.0020)
(b). Change in explained component					
10th quantile	-0.0342*** (0.0070)	-0.0102** (0.0042)	-0.0240*** (0.0058)	-0.0180*** (0.0031)	0.0023*** (0.0008)
50th quantile	-0.0183*** (0.0067)	0.0000 (0.0049)	-0.0183*** (0.0052)	-0.0097*** (0.0036)	0.0005* (0.0003)
90th quantile	0.0368*** (0.0101)	0.0267*** (0.0076)	0.0101 (0.0077)	0.0055 (0.0054)	-0.0003 (0.0004)
mean	-0.0079 (0.0061)	0.0053 (0.0050)	-0.0132*** (0.0039)	-0.0073* (0.0038)	0.0008** (0.0004)

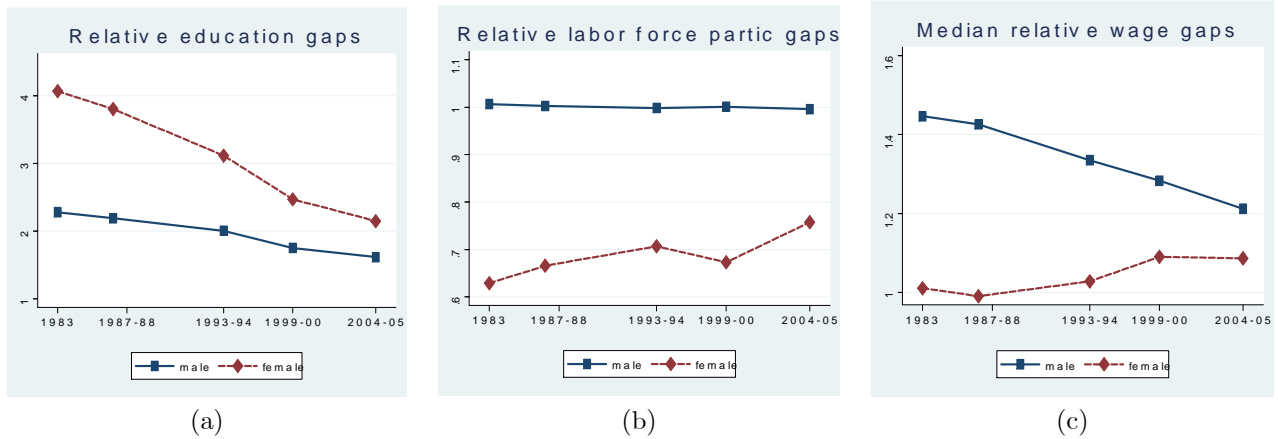
Note: Panel (a) presents the change in the SC/ST consumption gap between 1983 and 2004-05. Panel (b) reports the decomposition of the time-series change in the explained component of the change in the SC/ST consumption gap over 1983–2004-05 period. All gaps are decomposed into explained and unexplained components using RIF regression approach of FFL for the 10th, 50th and 90th quantiles; and using a standard OLS decomposition for the mean. Both panels also report the contribution of education and SC/ST reservation quota to the explained gaps. Bootstrapped standard errors are in parenthesis. * p-value \leq 0.10, ** p-value \leq 0.05, *** p-value \leq 0.01.

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 at every point of their respective distributions, except at the very top. 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 mask important 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. We compared the education, labor force participation rates and wages of SC/STs with their non-SC/ST counterparts for men and women separately.²⁸ Figure 9 plots the results. All the plotted series show non-SC/STs relative to SC/STs.

Figure 9: Gaps by Gender



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. Details on construction of the labor force participation rates are contained in A.1.5.

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. 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. 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 9. 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 widening caste wage gap for women despite a decrease in the relative labor

²⁸Education and wage comparisons are conducted using our working sample. Labor force participation gaps are computed from the same sample but without the full-time employment restriction.

supply of SC/ST women is a puzzle. 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. The convergence in occupation and wages is mostly due to a convergence in attributes, especially education.

We have also found that the convergence patterns have not been uniform. For both wages and consumption, the convergence is much sharper for the lower percentile groups. This heterogeneity finds an echo in the education and occupation changes as well. In education, there has been caste convergence in the relatively lower education categories but a slight increase in the gaps for the highest education categories. For occupation choices, there has been convergence in representations of the two groups in blue collar occupations but a divergence in their representations in white collar occupations. Hence, in contrast to the conventional view, the relatively better-off SC/STs may not have benefited more from the undergoing changes in the economy than the poorer SC/STs.

These results suggest that intergenerational mobility rates of SC/STs may have risen faster than non-SC/STs between 1983 and 2005. In work in Hnatkovska, Lahiri, and Paul (2011), we confirm this impression by examining changes in the intergenerational mobility rates in education, occupation and wages of the two groups.

What factors explain these changes? One 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 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 a strengthening of community based networks of SC/STs along the lines suggested in Munshi (2010). 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, though our results here raised doubts about this channel. 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.
- DI NARDO, 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.
- GELBACH, J. B. (2009): “When do covariates matter? And which ones, and how much?,” Ssrn working papers, Yale University.
- 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, 209–217.
- 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 Figures 1 and 9 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 $F_{W_c, X|c}(\cdot, \cdot)$. The marginal wage distribution of caste c is then obtained as

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

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 The RIF method

In order to assess the effect of the caste on unconditional wage quantiles we apply the method of Firpo, Fortin, and Lemieux (2009) (FFL henceforth). We resort to using this method as standard regression techniques 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) 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.²⁹ 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 τ th 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 .

A.4 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

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

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

²⁹The 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 $w - \mu$.

residuals. The differential in expected outcomes between the castes in round t is then given by:

$$\Delta y_t^e = \Delta X_t' \tilde{\beta}_t + X_{1t}'(\beta_{1t} - \tilde{\beta}_t) + X_{2t}'(\tilde{\beta}_t - \beta_{2t}),$$

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

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

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

$$\Delta y_2^e - \Delta y_1^e = (E_2 - E_1) + (U_2 - U_1) = \Delta E + \Delta U$$

These differentials are reported in Panel (a) of Tables 6 and 8. 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

$$\Delta E = \Delta X_2' (\tilde{\beta}_2 - \tilde{\beta}_1) + (\Delta X_2' - \Delta X_1') \tilde{\beta}_1,$$

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 (c) of Tables 6 for wages and 8 for consumption.