Talent Allocation in the Indian Economy: Measurement and Policy Implications

Kopal Mathur*

March 2022

Abstract

When individuals are not engaged in occupations according to their ability or 'talent', we observe talent misallocation in an economy. This paper shows that talent allocation improved in India over 1983-2012. To explain this, I compute similarity indices and examine the patterns in educational attainment and wages for marginalised (relative to privileged) gender and caste groups. I then use the reduced form approach to relate these indices with talent allocation. The paper's findings suggest that the improvement in talent allocation is correlated with the decline in labor market discrimination and barriers to educational attainment towards marginalised groups. Furthermore, using the decomposition technique, I find that convergence in educational attainment among these two barriers accounts for most of the improvement in talent allocation.

Keywords: Talent Allocation, Labor Market Discrimination, Educational Barriers, Caste Discrimination, Gender Discrimination, Occupational Convergence, Decomposition

JEL Classification: I26, O15, J71, J31

---

*PhD student at The University of Queensland- IIT Delhi Academy of Research (UQIDAR).
Email: Kopal.Mathur@uqidar.iitd.ac.in

I am immensely grateful to my advisors Jorge Miranda-Pinto, Antonio Andrés Bellofatto, and Sourabh B. Paul for their indispensable support and guidance. I also thank participants of the 16^{th} Annual Conference on Economic Growth and Development at Indian Statistical Institute Delhi and Winter School 2021 at Delhi School of Economics and The Econometric Society. This paper is part of my doctoral thesis at UQIDAR. All errors are mine.

1 Introduction

There has been much focus on the prevalence of caste based occupation identity in India, thus, hinting towards the possibility of social identity playing a role in the occupational choices of an individual.[1] Historically, each caste has been associated with a traditional occupation, where individuals are 'born' into their family occupations. However, India has experienced significant changes in occupational sorting between 1983 to 2012 based on both caste and gender.[2] Why do we observe a change in the occupational distribution or talent allocation by social identity? The current literature conjectures that labor market discrimination and barriers to attaining higher education deter individuals from engaging in occupations that gives them the highest utility according to their ability or 'talent'. Thus, reflecting the presence of talent misallocation in the economy. This paper examines whether a decline in these barriers was related to the improvement in the allocation of talent.

If we specifically focus on the changes in the distribution of high skilled occupations, we observe that in 1983 there were fewer women from the general caste category and men and women from scheduled caste/tribe (SC/ST) relative to male workers from the general caste group (favorable group). In 1983, 92% of workers from the general caste category were working as scientists, architects, and engineers as opposed to only 8% belonging to SC/ST. In 2012, the share of workers from the general category engaged in these occupations reduced to 81%, as opposed to a rise in the SC/ST caste group's share to 19%.[3] Given that the educational qualification or innate ability to be in high skilled occupations does not change over time, why do we observe such changes in occupational distribution?

Along with social groups seen as a proxy to indicate productivity and skill levels by employers, gender is also another characteristic that exacerbates differences in the economic outcomes of individuals. The paper uses successive waves (1983-2012) of the Indian unemployment and employment survey rounds to identify the reasons behind this change in occupational distribution. I first try to quantify talent allocation at the rudimentary level by adopting Porzio (2017) approach. I do this by comparing the benchmark scenario where individuals sort occupations according to their years of

[1] See Banerjee and Knight (1985), Munshi and Rosenzweig (2006), Motiram and Singh (2012), Cassan et al. (2021), and Deshpande (2011), among others.

[2] India is broadly composed of four caste groups or social groups. In this paper I focus on workers belonging to three such groups- scheduled caste (SC), scheduled tribes (ST) and general category. Where SC/ST groups are considered to be marginalised social groups and the general caste group is considered as upper caste or privileged social group.

[3] Source: NSS-EUS 38th and 68th round, for employed individuals aged 25-60 years- based on usual principal activity status and occupation. Refer to Figure 4.

education with their actual occupational choice, which I observe in the Indian labor market surveys. Second, to highlight the presence of frictions for individuals from disenfranchised groups to attaining higher education, I estimate the transitional probability of each group to move up the educational ladder relative to men from the general caste (GM).[4] Further, to provide evidence of labor market discrimination for these groups, I estimate the wage differential model using a reduced form approach. Lastly, I build on Albelda (1986) to relate the aforementioned barriers with talent allocation and use decomposition techniques to examine the importance of each set of frictions in improving talent allocation.[5]

This paper has four set of findings. First, I observe that talent misallocation was higher in 1983 than in the year 2012 in India. Next, with respect to frictions in attaining higher education, it is observed that disenfranchised groups in 1983 as well as in 2012 had a lower probability of transitioning or enrolling in a graduation degree, however over time relative to the privileged group, this probability has increased. Thus, suggesting a decline in barriers in attaining higher education. In addition to this, the results of the wage differential model suggest that when engaged in high skilled occupations, individuals from disenfranchised caste groups earn significantly lower wages than men from the general group. Men from SC/ST group (SM) earned 23% less, and the difference in returns were much worse for females from the general (GF) and SC/ST category (SF), around 42% - 44%.

Paper's findings also indicate that, as expected, individuals with a graduation (or above) degree receive higher earnings when engaged in high skilled occupations. Over time, attaining higher education became even more attractive as returns in 2012 compared with 1983 rose by 16%.[6] If we compare the average returns/gains of attaining higher education, the returns for men from the general group moving towards higher education were more than individuals from other disadvantaged groups engaged in high skilled occupations between 1983 and 2012. Further, I investigate the patterns and convergence in occupation sorting by estimating an occupational similarity index. I find that overall, in terms of relative propensity, all groups experienced convergence in occupations relative to GM between 1983 and 2012.

Finally, I close the paper by relating the frictions with talent allocation. Findings indicate that decline in wage differences and educational barriers are significantly correlated with the convergence in occupational choice, thus suggesting that a decline in talent misallocation is related to the decline

[4] I assume that this group faces minimal or no frictions in the labor market and educational attainment

[5] I use occupational convergence as a proxy for a decline in talent misallocation.

[6] Returns of GM (the reference caste and gender category in the regression analysis) were 79% in 1983 and 95% in 2012

in these barriers. In addition, using the Blinder-Oaxaca non linear decomposition technique, I find that convergence in educational attainment among these two barriers accounted for most of the improvement in talent allocation. Overall, our results do not suggest that the aforementioned barriers were not present in 2012 but that these barriers have declined, which is correlated with the decline in the extent of talent misallocation in 2012 compared with 1983.

Contribution to the literature: This paper contributes to the literature on labor market discrimination initiated by Becker (1971) and Aigner and Cain (1977). These papers highlight how information asymmetry on individuals' productivity results in employers using social identity and visible characteristics (gender, race, or caste, among others), thus leading to "taste based discrimination". This is further explored in the Indian context in Banerjee and Knight (1985), Madheswaran and Attewell (2007), Das and Dutta (2007), Sengupta and Das (2014). Though gender wage gaps have been studied extensively, this paper complements these studies to explore caste based wage inequality, which has not been explored in great detail. This paper uses a mincerian type earnings model to show that wage gaps based on caste and gender existed in 2012. Although compared to 1983, the gaps have declined as we observe wage convergence among marginalised and privileged caste and gender based groups.

This paper also contributes to the literature on educational convergence discussed in Desai and Kulkarni (2008), Maitra and Sharma (2009), Sahoo and Klasen (2021) and Varughese and Bairagya (2020). This paper supports the above studies by providing evidence of barriers to educational attainment for the period 1983-2012. I also relate education attainment with wage gaps and highlight that individuals earn different wages based on their social identity even while having the same level of education.

Differential returns to education and hiring discrimination more often than not reinforce workers from backward groups to engage in informal jobs. Thus, resulting in a vicious loop of lower probability of being in high skilled jobs leading to lower investment in education and vice-versa. Finally, this paper also contributes to the literature on occupational choices studied in detail in Banerjee and Knight (1985), Deshpande (2011), Cassan et al. (2021) and Klasen and Pieters (2015). One key difference from the above mentioned studies is that in this paper I describe occupational convergence as a situation of improved talent allocation (Hsieh et al., 2019; Abdulla, 2019). Additionally, along with investigating the patterns of wage differential and education attainment separately, I relate them with talent allocation using similarity indices for both gender (Agrawal, 2020) and caste groups.

Layout of the paper: The remainder of the paper is organized as follows. Section 2 describes the data along with motivation. Section 3 quantifies talent misallocation at a rudimentary level. Section 4 and 5 focus on barriers to education and labor market discrimination, respectively. In section 6, I relate the barriers with talent allocation and further evaluate the share of each barrier using the decomposition technique in section 7, which is followed by the discussion on contribution and potential mechanism in section 8.

2 Data and Motivation

2.1 Data

Employment and Unemployment Surveys (EUS) conducted by the National Sample Survey Organization (NSSO) records detailed information on socioeconomic and background characteristics at individual and household level. I take advantage of nine quinquennial EUS rounds conducted between 1983-2012.[7] For the year 1983, NSS covered around 1,20,847 households and 6,23,446 individuals overall. Wage details of 69,655 employed individuals aged between 25-60 years were also collected. For 2012, 4,56,999 individuals' details were collected with wage data for 2,07,385 employed individuals in the age group 25-60.

The variables of interest at the individual level include completed years of education, age, gender, wage information, along with occupational data. National Classification of Occupations (NCO) 1968 and 2004 have been used to classify each currently working individual into various occupations. I could successfully map 20 broad and comparable occupations (both high and low skilled occupations) from both the classification lists.[8] At the household level, I include information on caste, religion, monthly household expenditure, household size, region in terms of rural/urban and state.

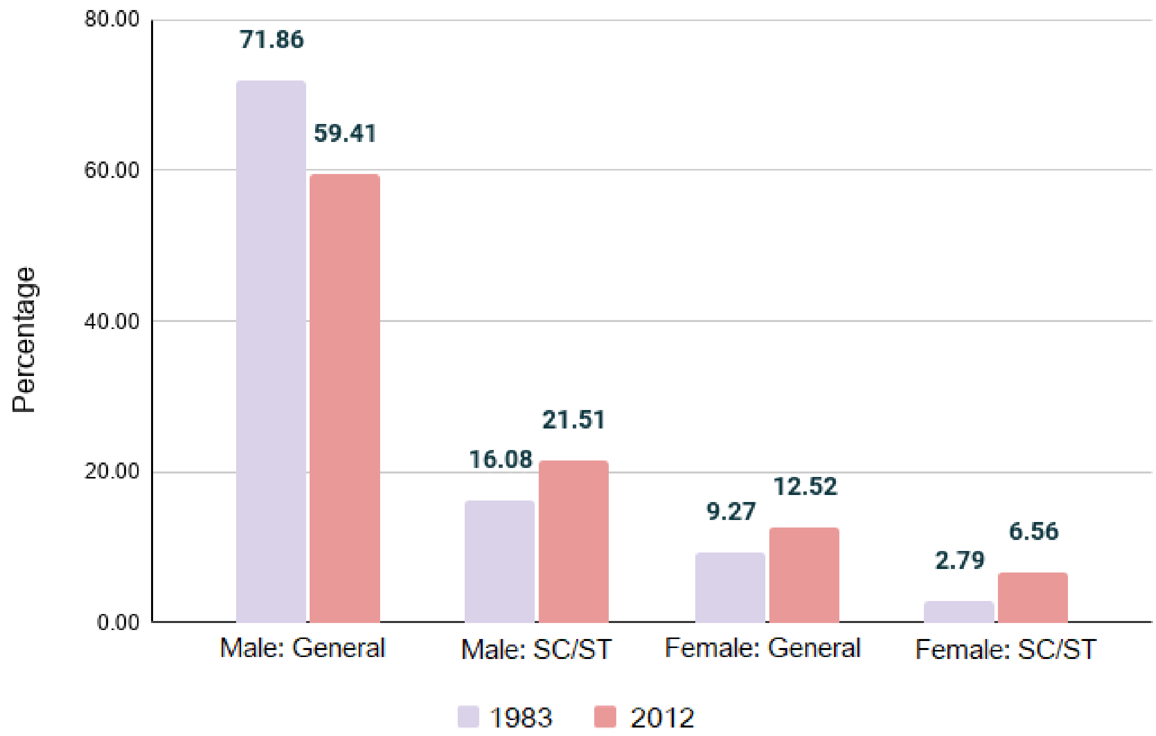
[7] NSS Rounds- 38^{th} (1983), 43^{rd} (1987 -'88), 50^{th} (1993 -'94), 55^{th} (1999 -2000), 61^{st} (2004 -'05), 62^{nd} (2005 -'06), 64^{th} (2007 -'08), 66^{th} (2009 -'10) and 68^{th} (2011 -'12)

[8] Refer to the mapping document here & details of the final occupations in A2. Occupations were mapped by referring to details provided for usual principal activity or occupation in which the individual was engaged in past 1 year. Refer to Appendix A1 & A3 for detailed steps undertaken for data preparation and summary statistics.

2.2 Motivating Evidence

To further motivate the research question, a preliminary descriptive analysis of changes in distribution for high skilled occupations by groups is displayed in Figure1.[9] I compare the data for 1983 with 2012 and observe an increase in the proportion of disadvantaged groups in high skilled occupations in 2012. However, the share of men from the general category in high skilled occupations has declined over time (by 12 percentage points). At first glance, this signals convergence or movement of workers from disadvantaged groups towards high skill occupations to some extent.

Figure 1: Group Share in High Skilled Occupations



As shown, the most substantial increase has been for men from the SC/ST category, whose group share increased by 5 percentage points. Depicted changes in occupational shares encourage us to think about the barriers in 1983 that restricted the disadvantaged groups' entry into high skilled occupations.

[9] This includes all high skilled professionals such as scientists, engineers, architects, medical professionals (including nurses), mathematicians, economists, jurists and accountants.

3 Measuring Talent Allocation in 1983 and 2012

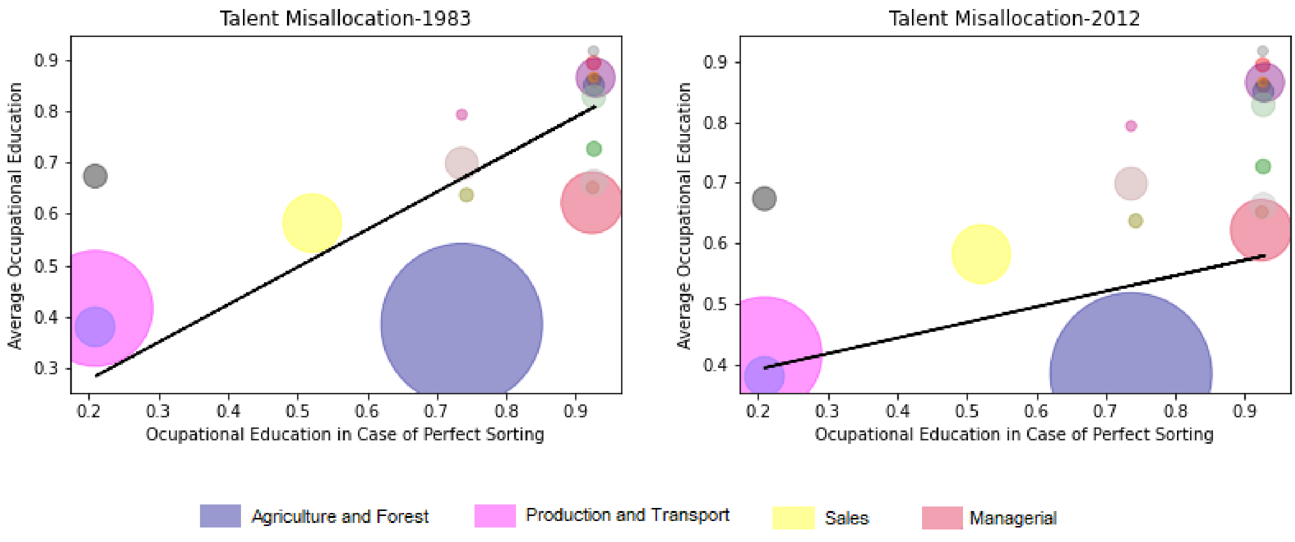
Our analysis will add to the limited literature that hints at the presence of talent or skill mismatch for the Indian economy (Mukherjee and Paul, 2012; Sengupta, 2017). Since investments in education enhance individuals' talent, and due to the lack of direct data measures on talent, I use years of schooling to measure talent allocation.

This section quantifies talent misallocation and estimates the benchmark case where individuals sort occupations according to the completed years of education and compare it with what is observed in the survey data. I do this for both the years 1983 and 2012 and discuss the trend of talent allocation over time. For this purpose I adopt empirical estimation of talent concentration in Porzio (2017). First, I normalise the measure of skills, given by $x_i = F(s_i)$, where s_i is the schooling years of each individual i and F is the year specific cumulative density function. Second, I compute the average skill in each occupation j: $\bar{x_j}$ = E $[x_i|I_{ij} = 1]$, where I_{ij} is an indicator function equal to 1 if individual i works in occupation j. Occupations with higher average skills are classified as high skilled occupations. In step three, a perfectly sorting counterfactual is estimated where talent is segregated according to occupations. Keeping the size of each occupation constant, all the most talented individuals are assigned to the occupations with the highest average skills, and sequentially, keep assigning the occupations as we go lower in the distribution of talented individuals. Step four includes estimating each occupation's average education or talent post perfectly sorting counterfactual assignment is performed, $\bar{p_j}$ = E $[x_i|I^c_{ij} = 1]$. To get a measure of the expected talent gap ' $\hat{\beta}_1$' across occupations relative to the benchmark case of perfect sorting I regress the average talent in each occupation in the benchmark case on average talent in each occupation as observed i.e., $\bar{x_j} = \beta_0 + \beta_1 \bar{p_j} + \epsilon$

The slope coefficient $\hat{\beta}_1$ can be written as $\hat{\beta}_1 = \dfrac{E[\bar{x_j}|\bar{p_j}] - E[\bar{x_{j'}}|\bar{p_{j'}}]}{\bar{p_j} - \bar{p_{j'}}}$. The slope coefficient here can also be interpreted as the expected ability gap across occupations relative to the benchmark case in which workers sort perfectly across occupations. Alternatively, $\hat{\beta}_1$ captures how close the observed allocation of talent is to the case in which individuals perfectly sort across occupations by their talent. A value closer to 1 of the slope coefficient will indicate higher talent misallocation whereas a value closer to 0 will indicate improved talent allocation.

When I fit the regression line for 1983 and 2012, the estimated slope coefficients are 0.73 and 0.25, respectively. Thus suggesting that the expected ability gap across occupations relative to the case where occupations are perfectly sorted was higher in 1983 (closer to one or perfect talent mis-

Figure 2: Talent Misallocation 1983-2012



allocation) compared with 2012. Another interpretation of this could be that talent misallocation was higher in 1983 when compared with 2012. Figure 2 plots the average skill in each occupation, where each bubble represents an occupation. Any increase or decrease in the size of the bubble is determined by the proportion of individuals employed in the occupation.[10]

Through the measurement of talent misallocation presented in this section, I try and measure the misallocation when talented individuals do not choose the occupation that maximises their utility. Hence, over time convergence in occupational distribution between the disenfranchised groups and GM would indicate a decline in talent misallocation.

The paper aims to show that convergence in educational attainment and reduction in wage differences are correlated with the decline in talent misallocation. To do this, in the next section, I first show the existence and the decline in educational barriers and labor market discrimination for disenfranchised groups.

4 Educational Mobility: Barriers in Attaining Higher Education

[10]The sample includes only employed individuals within the age bracket 25-60 years with information on completed years of education. In both the years agriculture and production and transport laborers remains the highest employed occupations. The most concentration or increase in employed proportion has been in the ‘Managerial’ and ‘Sales’ occupation.

One of our research questions pertains to gender and caste differences in educational attainment. Given the inequality in educational opportunities (Desai and Kulkarni, 2008; Maitra and Sharma, 2009), I hypothesize that the transitional probability of education (or academic progression) from school (*higher secondary*) to college (*graduation and above*) differs and is lower for marginalised communities than men from the general caste category. I estimate this for 6-29 years old in the 1983 and 2012 NSS rounds.

I adopt the sequential probit model approach by Lillard and Willis (1994) and Maitra and Sharma (2009) to model this educational stratification. This methodology is specifically useful because, in our analysis, transitioning to the next class or grade in school is conditional on the completion of previous classes. This censoring at each transitioning node is captured in the sequential probit model, making it a superior methodology choice than any other discrete choice model. Since our focus is mainly on the transition to higher educational categories, I concentrate on two levels of academic progression- 'secondary/sr. secondary' and 'graduation and above'. Where the following holds:

$$\mathrm{d} = \begin{cases} 0, & \text{if in middle grade i.e years of schooling is 8 yrs} \\ 1, & \text{if in secondary/sr. secondary grade i.e years of schooling is 10yrs} \\ 2, & \text{if graduate and above i.e years of schooling is 15yrs} \end{cases}$$

The first decision is to be in middle school (for only enrolled students who have completed primary schooling).[11] To establish the existence and decline/increase of educational barriers based on caste and gender, we are interested in the transition of the above stated stages. Our primary interest is from stage one to stage two. For each individual belonging to the defined group ' i ' a probit index function at each decision point is defined as below:

$$\mathrm{I}_{id} = \beta_d X_{id} + \delta_i + \epsilon_{id} \;\; ; d = 0, 1, 2 \tag{1}$$

Similar to Maitra and Sharma (2009), the heterogeneity in the propensity to continue their education or keep pursuing their education is considered in the δ, which is constant in all schooling decisions, and X_{id} term, which varies across decisions. In our analysis δ includes variables like religion, gender, social caste, and gender and years of education of the household head. X_{id} includes

[11] A "no- detention" policy is implemented via The Right to Education Act (2009) for elementary grades (1- 8) and enrollment in grade 1 at age 6. Here primary schooling means classes I to V, and middle schooling means classes VI to VIII; secondary and higher secondary classes IX to XII.

covariates like age, household monthly expenditure, and region. Individuals from the group 'i' will move to the next level of d, i.e., from d to $d+1$ if $I_{id} > 0$ and drop out from the education system otherwise. Hence the transition nodes are specified as:

$$\text{P(d)} = \begin{cases} P[I_{i1} \leq 0], & \text{if d=0} \\ P[I_{i1} > 0, I_{i2} \leq 0], & \text{if d=1} \\ P[I_{i1} > 0, I_{i2} > 0], & \text{if d=2} \end{cases}$$

The three transitions would imply:

1. Whether the respondents have completed middle school;
2. Those who have enrolled in secondary/higher secondary schooling conditioned on them attaining middle school education;
3. Those respondents who have enrolled in higher education, i.e., graduation (and above) conditioned on having fulfilled their requirements of attaining secondary/sr. secondary education.

Separate regressions are estimated for each academic category for both 1983 and 2012. In addition to identifying caste and gender groups, I also control for age, monthly household per capita expenditure, religion, education of the household head, the number of female and male members in the household, and controlling for state fixed effects. Table 1 displays the estimates of the model.

The results suggest that all others have a significantly lower transition probability of moving up the educational ladder with reference to men from the general category. However, over time the difference between the probability of transition between marginalised groups and GM has declined. This decline was substantial for the female category from both SC/ST and general category in the middle grade. In the senior secondary category, the substantial decline in difference was for women from the general category.

Specific to our discussion for attaining higher education, I discuss in detail the transition to college (or graduation and above educational category) in columns (3) and (6). In 1983, there was a significant difference in transition probabilities between GM and non GM categories. The lowest transition probability of enrolling for a graduation degree is for SC/ST females, with around 5% lower transition probability. For men from SC/ST category, the difference was 3% (significant at 1% level of significance), and for females from the general category was 0.8% significant at a 10% level

Table 1: Sequential Probit Estimates for Academic Progression

Variables	Middle	Secondary/ Sr. Secondary	Graduation and above	Middle	Secondary/ Sr. Secondary	Graduation and above
	(1)	(2)	(3)	(4)	(5)	(6)
	1983			2012		
Group [Base=General: Male]						
SC/ST:Male	-0.060***	-0.068***	-0.030***	-0.032***	-0.051***	-0.026***
	(0.003)	(0.006)	(0.010)	(0.003)	(0.005)	(0.005)
General:Female	-0.064***	-0.035***	-0.008*	-0.016***	-0.016***	0.0025
	(0.002)	(0.004)	(0.005)	(0.003)	(0.005)	(0.004)
SC/ST:Female	-0.134***	-0.137***	-0.057***	-0.064***	-0.078***	-0.032***
	(0.004)	(0.009)	(0.014)	(0.004)	(0.006)	(0.006)
Constant	YES	YES	YES	YES	YES	YES
Other Control Variables	YES	YES	YES	YES	YES	YES
Observations	166,229	61,591	26,759	106,429	58,322	35,978

Note: The table records average marginal effects; robust standard errors are in parenthesis.
*** p<0.01, ** p<0.05, * p<0.1

of significance. In 2012, the difference declined, which was significant for men and women from SC/ST categories. However, this is insignificant for women from the general category.

Overall, estimates of the sequential probit model suggest that men from the general category have a higher probability of transition to higher education, i.e., graduation and above degree, compared with marginalised categories. In other words, the estimates show that individuals of all caste groups are less likely to transition to college than GM.[12] This is consistent with the findings of Desai and Kulkarni (2008). They did a similar analysis using logit specification for data until 2000. Our analysis suggests that these results hold even after a decade (2012). In addition to this, the findings of our paper indicate that individuals from the historically backward groups are soon catching up with GM as the difference between the probability overtime has declined.

In our analysis, the improvement of a decline in these barriers lagged for SM, whereas SF fared better in the upward educational mobility from sr. secondary transition to college. I also find that even though GM and non GM differences existed, over time, these barriers declined. This decline can be attributed to the affirmative action policy, which was implemented to equalise educational opportunities and to overcome caste based differences. Along with affirmative action, many other policies encouraging the construction of schools/colleges in areas dominated by underprivileged

[12]The likelihood of GM enrolling in a graduate degree was 20% in 1983, this significantly increased to 22% in 2012.

communities, provision of mid day meals, free distribution of sanitary pads, stationery, and financial benefits were also carried out. Higher demand for skilled workers might have also led to a reduction of indicated educational barriers.[13]

In conclusion, even though GM's likelihood of completing schooling and enrolling (and attaining) for higher education is greater even in 2012, overall, individuals from marginalised groups' likelihood of moving up the education ladder have improved. Thus, indicating that educational barriers for all groups have declined over time compared with 1983, though they still existed in 2012.

5 Wage Differentials: Labor Market Discrimination

The prior section discussed the presence of educational barriers in the Indian labor market. In this section, I establish the presence of labor market discrimination. Since direct measures of labor market discrimination are unattainable for national samples of the population, in our analysis I use wage differentials (conditional on observed characteristics of the workers) as a proxy for labor market discrimination.[14]

I adopt a reduced form approach and estimate a wage differential model while restricting our sample to individuals employed in high skilled occupations. Through the results of this model, the goal is to infer whether there exist wage differences among workers employed in high skilled occupations according to their social identity and whether these differentials have reduced over time. I estimate the following three specifications of the wage differential model, and the results are detailed in table 2:

$$\log(\text{wage})_{it} = \beta_0 + \beta_1 \text{Group}_i + \beta_2 \text{Education attained}_i + \beta_3 \text{Time}_t + \beta_4 X_i + \epsilon_{it} \tag{2}$$

$$\begin{aligned}\log(\text{wage})_{it} = &\beta_0 + \beta_1 \text{Group}_i + \beta_2 \text{Education attained}_i + \beta_3 \text{Time}_t + \beta_4 (\text{Group*Education})_{it} + \\ &\beta_5 X_i + \epsilon_{it}\end{aligned} \tag{2.1}$$

[13] Jong-Wha and Wie (2017) do find evidence of an increase in the demand for skilled workers since the early 1980s in India. This has increased the wage premia of high skilled workers and might also be responsible as a motivation to individuals to move up the education ladder.

[14] When I use wage differentials as a proxy, I assume that the wage structure of all the groups is the same, which is a simplifying assumption of what is observed in reality. However, keeping in mind our broader goal of the paper, I stick to this assumption. There exists many studies which decompose wage differentials into 'explained' (covariates included in the model) and 'unexplained' which indicates discrimination (Deshpande et al., 2018; Deininger et al., 2013; Hnatkovska et al., 2012; Madheswaran and Attewell, 2007, among others.)

$$\begin{aligned}\log(\text{wage})_{it} =& \beta_0 + \beta_1\text{Group}_i + \beta_2\text{Education attained}_i + \beta_3\text{Time}_t + \beta_4(\text{Group*Time})_{it} + \\ & \beta_5(\text{Group*Education})_i + \beta_6(\text{Education*Time})_{it} + \\ & \beta_7(\text{Group*Education*Time})_{it} + \beta_8 X_i + \epsilon_{it}\end{aligned} \tag{2.2}$$

In all three specifications, the log of daily wages is the dependant variable. Where the independent variable 'group' takes a value 0 if a worker belongs to the category 'General men (GM)'; takes a value 1 if a worker belongs to the category 'SC/ST men (SM)'; value 2 if a worker belongs to the 'general women (GF)' category and a value of 3 if a worker belongs to 'SC/ST women (SF)' category. Variable 'education' includes details of years of schooling. NSS collects data on educational categories such as below primary, primary, middle, secondary and graduation (and above). For our analysis, I convert each of these categories into years of education.[15] Along with these, I control for the time period by including a binary variable 'Time', which takes a value 0 if the year is 1983 and takes a value 1 if the year is 2012. In addition, I also include information on religion, location, gender of the household head, years of education of the head of the household, no. of female and male members in the household, household per capita monthly expenditure (log) and also control for state fixed effects.

Results of model specification 2 (table 2; col:1) suggest that the returns of individuals from disadvantaged groups (GF/SF) are significantly lower, around 30%, relative to GM when employed in high skilled occupations. This was also true for men from SC/ST group, though the difference was only around 4% for them. Over time, for all groups, the remuneration of being employed in high skill occupations has increased, making such occupations more attractive. Attaining higher education also leads to higher returns.

Next, I estimate a difference-in-difference (DD) type specification of earning model.[16] Results of specification 2.1 (table 2; col:3) suggest that the β_4 coefficient of the interaction between '*group*education*' is insignificant for all other group categories suggesting that on average there is no difference between earnings for individuals from the disadvantaged groups when compared with men from the general category, as they move towards attaining higher education. In specification 2.2 (table 2; col:5) which is a difference-in-difference-in-difference(DDD) type specification model, the find-

[15] Illiterate=0 years, literate but below primary=2 years, primary=5 years, middle= 8 years, secondary and higher secondary=10 years, graduate and above=15 years. Note: I exclude diploma and literate w/o formal schooling.

[16] I write difference-in-difference 'type' because I rely on the interaction terms to inform our analysis. I am not studying the treatment or control effects of any policy in the proposed analysis.

Table 2: Wage Differential Model Estimates[OLS]

Variables	Log(wage)					
	Model 1		Model 2 (DD)		Model 3 (DDD)	
	OLS	Selection Corrected OLS	OLS	Selection Corrected OLS	OLS	Selection Corrected OLS
	(1)	(2)	(3)	(4)	(5)	(6)
Group [Base=General: Male]						
SC/ST:Male	-0.0490*	-0.0104	0.0303	-0.150***	-0.265***	-0.346***
	(0.0274)	(0.0116)	(0.134)	(0.0525)	(0.0761)	(0.0807)
General:Female	-0.357***	-0.383***	-0.399**	-0.360***	-0.566***	-0.428***
	(0.0234)	(0.0116)	(0.175)	(0.0984)	(0.155)	(0.137)
SC/ST:Female	-0.368***	-0.394***	-0.114	-0.246***	-0.585***	-0.503***
	(0.0498)	(0.0196)	(0.183)	(0.0935)	(0.145)	(0.152)
Years of Education [Base=Illiterate]						
Graduation and above	0.808***	0.715***	0.861***	0.695***	0.582***	0.707***
	(0.0834)	(0.0323)	(0.126)	(0.0389)	(0.0685)	(0.0523)
Time [Base: 1983]						
2012	1.557***	1.961***	1.561***	1.97***	1.047***	3.081***
	(0.0483)	(0.0249)	(0.0484)	(0.0261)	(0.110)	(0.158)
Interaction: Education * Time						
Graduation and above*2012					0.669***	-0.688***
					(0.102)	(0.119)
Interaction: Group * Education						
SC/ST:Male* Graduation			-0.132	0.0616	0.244***	0.302***
			(0.141)	(0.055)	(0.0817)	(0.093)
General:Female* Graduation			0.0308	-0.044	0.257*	0.0897
			(0.177)	(0.099)	(0.156)	(0.141)
SC/ST:Female* Graduation			-0.138	-0.105	0.438***	0.334*
			(0.193)	(0.099)	(0.162)	(0.193)
Interaction: Group * Time						
SC/ST:Male*2012					0.713***	0.328***
					(0.165)	(0.126)
General:Female*2012					0.365	0.768***
					(0.317)	(0.235)
SC/ST:Female*2012					1.239***	0.712***
					(0.266)	(0.202)
Interaction: Group * Education * Time						
SC/ST:Male* Graduation*2012					-0.851***	-0.599***
					(0.176)	(0.135)
General:Female* Graduation*2012					-0.470	-1.046***
					(0.321)	(0.243)
SC/ST:Female* Graduation*2012					-1.401***	-1.173***
					(0.284)	(0.243)
Lambda (inverse mills ratio)		-0.163***		-0.172***		-0.769***
		(0.0217)		(0.0238)		(0.0798)
R squared	0.859		0.859		0.861	
Constant	YES	YES	YES	YES	YES	YES
Other Control Variables	YES	YES	YES	YES	YES	YES
Observations	27,482	37,726	27,482	37,726	27,482	37,726

ings suggest that the β_7 (which is the coefficient of interest) coefficient of the interaction between *'group*education*time'* is significant for both women and men from SC/ST group at higher educational categories. This suggests that over time the average increase in earnings for attaining higher education for SC/ST men and women is significantly lower, by around 57% (SM) to 75% (SF), than general men in high skilled occupations.

In conclusion, for 1983-2012, conditional wage differences exist among workers employed in high skilled occupations according to their caste and gender. Individuals from disadvantaged groups earn significantly lower wages than general men when employed in high skilled occupations. On average returns/gains of GM moving towards higher education (graduation and above) were more than individuals from other disadvantaged groups engaged in high skilled occupations over time.[17]

5.1 Robustness checks: Correction of Selection Bias

There is a high possibility of selection bias in the models discussed briefly in the previous section. In this section, I discuss in detail and outline cases where I think there is selectivity bias, the method used to correct it, and discuss the selection corrected results.

In the proposed gender and caste wage gap model, the information of the dependant variable (log of wage) will be present only for the employed individuals at the time of the survey. Hence, we are using a subsample for individuals for whom we can observe the data point. For the individuals who are not present in the workforce, we cannot observe the wage data. This is a case of incidental truncation. Individuals self select themselves into being employed, which makes the sample non random, this makes the β estimates inconsistent.

In addition to only considering employed individuals, we restrict our sample to individuals in a specific type of job categorized as 'High skilled jobs'. Motivation or ability can play a critical role for individuals to be engaged in a type of job. These attributes are un-observable and can be correlated with the education levels. To correct the selection bias, I apply the heckman correction technique (heckit two step estimation). In this two-step model, first, a probit model is estimated with all observations. The selection model is then estimated with all independent variables in step 1 and at least one additional variable (to avoid multicollinearity) which directly affects labor force participation but does not determine wages. To fulfill this exclusion restriction, variables such as marital status and ownership or possession of land are included.

[17]Similar analysis was performed to understand the variation in occupational choices of the different social and gender groups. Refer to appendix A4 for the results.

The Heckman selection model uses all observations (total sample ,i.e., unemployed individuals and those not in the labor force are also included) to compute the mills ratio by estimating the probit selection model in the first stage. Then using the selected sample (individuals employed in high skilled occupations), the OLS regression model is estimated, inclusive of the coefficient of the inverse mills ratio.

If we note the coefficient of λ in Table 2 in col(2, 4 and 6) which is the coefficient of the inverse mills ratio, it is significant, suggesting that the models suffer from the issue of selection bias. After correcting for selection bias, the estimates of the DDD model indicate similar results as discussed in the previous section, in terms of sign and significance, even though the magnitude differs. If we note the coefficients of *'group*education*time'* we now obtain more substantial results for females from the general caste category, as now the difference for this group is significant, which was not the case previously. Post correcting for selection bias, we observe that the difference of wages of being employed in high skilled occupations would persist when individuals move to higher education categories- ranging from 45%-70%.

6 Linking Barriers with Talent Allocation

I have empirically shown that overtime talent misallocation has declined and separately provided evidence of labor market discrimination and educational barriers present in the Indian labor market. In this section, I first observe the trend in the barriers for each group. Secondly, I establish a link between these barriers with occupational convergence to argue that the trend in these aforementioned barriers is significantly correlated with the decline we observe in talent misallocation.

To formally show the degree of convergence in occupation sorting between GM and other groups, I adopt the following definition of the occupational similarity index from Hsieh et al. (2019), which is a variation of the proposed index by Duncan and Duncan (1955):

$$\text{Occupational Similarity Index}_g = 1 - \frac{1}{2}\sum_{i=1}^{N} |P_{i,gm} - P_{i,g}| \tag{3}$$

Here $P_{i,g}$ is the propensity of people in a group g and occupation i. The propensity is defined, for example- for females from the general category (GF) as the fraction of the total number of GF in an occupation i / total number of GF in all occupations. The index measures the difference between the propensity of men from the general category with a given group to be in occupation. The value of the index lies between 0 and 1. A value closer to one indicating a similar propensity of non GM groups and GM; the opposite will be true for a value closer to 0.[18] I also divide the individuals into two groups according to their educational attainment, those with a graduation degree in one group and the remaining in another.

Table 3 suggests that overall, in terms of propensity, all groups experienced convergence in occupations relative to GM between 1983 and 2012 (col:4). Consistent with the findings of Abdulla (2019) for the period 1993-2004, overall occupational choices of SC/ST men were similar to GM. Though, I also observe substantial convergence in occupational distribution for the period 1983-2005 for SM and SF.

When I examine the convergence for groups based on educational attainment, it was higher for low educated individuals than those with a degree of graduation. Relative to GM, this was more substantial for SC/ST women. In 1983, low educated SC/ST women primarily worked as agricultural and forest workers. At the same time, low educated GM were more concentrated in production and transport along with clerical and agricultural related occupations. In 2012, the movement of SF towards production and transport laborers was observed, whereas GM moved towards being managers or sales workers.

In 1983, GM were more concentrated in teaching, managerial, and sales occupations for higher educated workers. I observe a substantial divergence between SC/ST women and GM (graduation and above) in the period 1983 and 2005. Higher educated SF in 1983 were primarily concentrated in teaching and service occupation along with low skilled jobs, whereas GM were scattered among teaching and clerical jobs. This changed as GM worked primarily in sales and managerial jobs in 2005. In contrast SF mostly remained in teaching jobs and were concentrated in low skilled jobs even though they had a graduation degree. In conclusion, even though occasional divergence is observed, overall for 1983-2012, the index values suggest substantial convergence in occupational distribution for groups relative to GM.

To link educational and wage convergence with talent allocation, I compute the educational similarity index and wage similarity index for each group relative to GM. I first measure the educational

[18]However, since the index is weighted according to the share or population structure of a group in an occupation, the change in the index can also be due to the change in occupational structure rather than differences due to gender and caste group composition (Sloane et al., 2019). This is one of the limitations of using the index.

Table 3: Occupational Similarity Index Relative to Men from the General Category

Group	1983	2005	2012	2012- 1983 Difference	2005- 1983 Difference	2012 - 2005 Difference
	(1)	(2)	(3)	(4)	(5)	(6)
General Women (GF): All	0.57	0.63	0.63	0.06	0.06	0.00
General Women (GF): Graduation and above	0.60	0.62	0.66	0.06	0.02	0.04
General Women (GF): Below Graduation	0.57	0.64	0.62	0.05	0.07	-0.02
SC/ST Men (SM): All	0.68	0.79	0.79	0.11	0.11	0.00
SC/ST Men (SM): Graduation and above	0.86	0.73	0.73	0.13	0.13	0.00
SC/ST Men (SM): Below Graduation	0.70	0.83	0.84	0.14	0.13	0.01
SC/ST Women (SF): All	0.42	0.57	0.62	0.20	0.15	0.05
SC/ST Women (SF): Graduation and above	0.78	0.60	0.57	-0.21	-0.18	-0.03
SC/ST Women (SF): Below Graduation	0.45	0.60	0.63	0.19	0.16	0.03

Note: Authors' calculation. To estimate the above index, sample includes only those individuals who were employed, aged between 25-60 years and had information about their completed highest years of education.

similarity index by defining it as similar to the occupational similarity index.

$$\text{Educational Similarity Index}_g = 1 - \frac{1}{2}\sum_{i=1}^{N} |P_{i,gm} - P_{i,g}| \tag{4}$$

Here, $P_{i,g}$ is the propensity of people in a group g and educational category i. The propensity is defined, for example- for females from the general category (GF) as the fraction of the total number of GF in an educational category i / total number of GF in all educational categories. The index measures the difference between the propensity of men from the general category with a given group to be in an education category.

Table 4 presents the values for the defined indices over time for each group. As discussed previously, I find evidence of convergence in the occupational similarity index indicating that over time, the occupational choices of marginalised groups are altering and moving closer to the choices of GM. When I examine the educational similarity index, we again observe similar patterns. The convergence in the similarity index with GM for the period 1983-2012 is highest, around 64% for women from the SC/ST social group, followed by women from the general category (21%) and then men from the SC/ST group (15%).

Table 4: Convergence in Occupation, Education and Wage Patterns by Caste & Gender

Year	Occupational Similarity Index (Panel A)			Educational Similarity Index (Panel B)			Wage Index (Panel C)		
	SC/ST Male	General Female	SC/ST Female	SC/ST Male	General Female	SC/ST Female	SC/ST Male	General Female	SC/ST Female
1983	0.685	0.573	0.425	0.659	0.638	0.351	-0.140	-0.104	-0.209
2005	0.794	0.639	0.572	0.753	0.715	0.507	-0.046	-0.071	-0.097
2012	0.797	0.631	0.621	0.763	0.775	0.576	-0.033	-0.039	-0.062

To measure similarity in wages, I adopt the index definition from Sloane et al. (2019). One benefit of this index is that its units are in the wage space, giving ease in interpreting the economic magnitude of differences in occupational choices by different gender and caste groups. $\bar{Y}_{i,gm}$ is defined as the median labor market wage within an occupational category (irrespective of their education). Since I assume that men from the general caste category face minimum or no labor market discrimination, their wage is considered the median wage in an occupational category 'i'. $P_{i,g}$ is the propensity of individuals in a group 'g' for occupational category 'i'.

$$\text{Wage Index}_g = \frac{\sum_{i=1}^{N} P_{i,g}\bar{Y}_{i,gm}}{\sum_{i=1}^{N} P_{i,gm}\bar{Y}_{i,gm}} - 1 \tag{5}$$

The potential wage index measures the differential log wage of group 'g'. A value of 0 of this index will reflect that the occupational choice of groups yields the same log wage as men from the general group. A negative value indicates that individuals from group 'g' choose occupational categories associated with lower relative wages.

Panel C of table 4 displays the estimates of the wage index. Relative to GM, individuals from all other groups choose occupations in which they earn lower daily wages (presence of labor market discrimination). However, there is evidence of convergence in the wage index over time. In 1983, men from the SC/ST group chose occupations where they earned 14% lower than GM, and the potential wage gap was highest for females from SC/ST group. In 2012, the potential wage gap fell to 3% lower for SM and 6% lower for SF. For females from the general group, the potential wage gap reduced to 3% in 2012 from 10% in 1983.

After examining the patterns of the indices through descriptive statistics, I empirically link the decline in educational barriers and labor market discrimination with occupational convergence. I argue that the decline in the barriers is correlated with the decline I observe in talent misallocation. To do this, I build upon the methodology proposed by Albelda (1986) and use data from all NSS-EUS rounds from 1983-2012. I first measure the similarity of education attainment (ESI) and regress it on the occupational similarity index (OSI). Where ‘ i ’ varies for each group (SM/SF/GF), and ‘ t ’ varies for each time period.

$$\mathrm{OSI}_{it} = \beta_0 + \beta_1 \mathrm{ESI}_{it} + \beta_2 \mathrm{Time}_t + \beta_3 X_i + \epsilon_{it} \tag{6}$$

$$\mathrm{OSI}_{it} = \beta_0 + \beta_1 \mathrm{WI}_{it} + \beta_2 \mathrm{Time}_t + \beta_3 X_i + \epsilon_{it} \tag{6.1}$$

$$\mathrm{OSI}_{it} = \beta_0 + \beta_1 \mathrm{ESI}_{it} + \beta_2 \mathrm{WI}_{it} + \beta_3 \mathrm{Time}_t + \beta_4 X_i + \epsilon_{it} \tag{6.2}$$

I run separate OLS regressions (equation 6) for the three groups in comparison with GM. X_i in eq(6) controls for household and socioeconomic characteristics (discussed in the previous sections). Table 5 - Panel A, where col (1) displays the results for SC/ST male, col(2) and col(3) display the results for women from the general and SC/ST category. Our primary coefficient of interest is of the variable educational similarity index (ESI) in each column. Essentially a positive sign of the ESI variable would suggest that similarity in educational choices/attainment (value closer to 1) of groups with respect to GM has a positive relation to the similarity in occupational choices. Once we control for time and other factors, I observe significant positive relation between the similarity of educational attainment for both men and women from the SC/ST category relative to GM. Suggesting that convergence in educational similarity is correlated with the observed occupational convergence over time. For women from the general category, I observe the opposite sign of the ESI variable. The negative sign of the significant ESI coefficient suggests that even though we observe the similarity in educational attainment between them and GM, this has impeded occupational convergence. Even though the patterns of educational attainment are becoming similar to GM, on average, this impedes occupational convergence among the two groups.

Our second variable of interest, 'Time' captures the effect of changes in occupational convergence over time. It is a categorical variable that takes a value 0 for the 38^{th} survey round (1983), 1 for the 43^{rd} round, and goes up to a value 8 for the 68^{th} round (2012). Given the policies implemented by the government to reduce the gap between the social and gender based groups, such as affirmative action policy, we expect the coefficient of 'Time' to have a positive sign. From the results, I observe that once we control for education and other household and socio-economic factors, the occupational choices for men from the SC/ST category and women from the general category relative to GM have significantly converged over time. However, the opposite is true for women from the SC/ST category.

Overall the results imply that convergence in patterns of educational attainment is statistically significant and positively associated with occupational convergence for SC/ST group. Suggesting that barriers to education have declined between them and GM, which is positive and significantly correlated with a decline in talent misallocation between these groups. The opposite is true for women from the general category, as the decline in educational barriers negatively correlates with occupational convergence.

I now attempt to link the decline in wage gaps with occupational convergence. To do this, I estimate a simple OLS regression model (equation 6.1), where the dependent variable is again the occupational similarity index. Our coefficient of interest is of the independent variable wage index (WI). Where 'i' varies for each group (SM/SF/GF), and 't' varies for each time period.

The results of equation 6.1 are displayed in Panel B of table 5. Controlling for time and other socioeconomic factors, a positive coefficient of the variable 'Wage Index' indicates that convergence in relative potential wages for SC/ST men and women have a positive bearing on occupational convergence. Potential wage convergence, thus, has been complementary to the reduction in talent misallocation for the SC/ST group. However, an opposite relationship exists for women from the general group. Coefficients of the 'time' variables demonstrate that various policies targeting to reduce the gap between various social groups and gender have been significantly negatively related to occupational convergence for SC/ST groups. Again, the opposite is true for women from the general group.

Even when I estimate equation 6.2 and include both educational convergence and potential wage indices, the above results hold. We observe that both significantly explain the decline in talent misallocation between SC/ST groups and GM. This is true, even when changes in laws and policies targeting the inclusion of disenfranchised groups over time have impeded occupational convergence. However, again, an opposite relationship exists for females from the general group. When the similarity of their educational distributions with men from the general category increases, their convergence in occupational distributions decreases. This is similar to convergence in relative potential wages. These results align with the stagnant and low female labor force participation rate that we observe in the Indian labor market (Afridi et al., 2018a; Datta et al., 2020).

Collectively, regardless of the direction of relation across different social caste and gender groups, the above analysis confirms that differences in education and wage gaps are related to similarity in occupational distributions.

7 Decomposing Changes in Talent Allocation

The previous section of the paper showed how the barriers in terms of educational attainment and labor market discrimination are correlated with the improvement in talent allocation over time. This section decomposes the convergence in occupational choices into proportional shares of convergence in educational attainment and narrowing wage differentials or labor market discrimination. To formally quantify and evaluate the share of these covariates in improvement of talent allocation, I use Blinder–Oaxaca decomposition technique (Blinder, 1973; Oaxaca, 1973).

This technique has been heavily cited to study labor market outcomes, such as to quantify average differences in log wages, consumption, labor force participation among other patterns by groups (Hnatkovska et al., 2012; Deininger et al., 2013; Afridi et al., 2018b; Deshpande et al., 2018). The method essentially quantifies the share of covariates or observables and the share of unobservables. To determine the factors contributing to the convergence in occupational choices between GM and non GM groups I estimate the following logit regression model for both 1983 and 2012:

$$\widehat{Y}_i = F(\widehat{\beta} X_i) \tag{7}$$

Where the dependent variable is a binary variable determining the occupational choice ,i.e., Y=1 if individuals are engaged in high skilled occupations and is 0 otherwise.[19] X includes explanatory variables discussed in the previous sections.

[19]Where individuals engaged in occupations as scientists, engineers, architects, nurses and other medical professionals like surgeons etc., economists, accountants, jurists, teachers, managers etc., are considered to be engaged in high skilled occupations.

Table 5: Determinants of Occupational Convergence between 1983-2012 [OLS]

	Dependent Variable: Occupational Similarity Index		
	SC/ST Male (1)	General Female (2)	SC/ST Female (3)
Independent Variables			
		Panel A	
ESI	0.744***	-1.245***	1.110***
	(0.00254)	(0.00743)	(0.00297)
Time	0.003***	0.029***	-0.005***
	(3.87e-05)	(0.000214)	(6.22e-05)
Other Control Variables	YES	YES	YES
Constant	YES	YES	YES
R Squared	0.640	0.326	0.874
Observations	254,768	118,656	133,540
		Panel B	
Wage Index	1.669***	-1.559***	1.775***
	(0.00294)	(0.0266)	(0.00826)
Time	-0.008***	0.016***	-0.003***
	(3.38e-05)	(0.000263)	(9.73e-05)
Other Control Variables	YES	YES	YES
Constant	YES	YES	YES
R Squared	0.677	0.158	0.709
Observations	247,721	112,545	127,037
		Panel C	
ESI	0.410***	-1.140***	1.006***
	(0.00237)	(0.00669)	(0.00472)
Wage Index	1.143***	-0.687***	0.409***
	(0.00298)	(0.0233)	(0.00762)
Time	-0.005***	0.033***	-0.007***
	(4.14e-05)	(0.000286)	(7.07e-05)
Other Control Variables	YES	YES	YES
Constant	YES	YES	YES
R Squared	0.737	0.345	0.884
Observations	247,721	112,545	127,037

Note: Authors' calculation. The table records OLS coefficients; with robust standard errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1. We also include an intercept term and control for household and socio- economic characteristics such as age, monthly household income, religion, gender and educational level of the household head, no. of male and female members of the households, and state fixed effects. The sample includes only those individuals who were employed, & in the age bracket 25-60 years.

The non-linear Blinder-Oaxaca (B-O) decomposition uses $\widehat{\beta}$ of group GM to decompose the difference into the explained and unexplained components:[20]

$$\bar{Y}_{i,gm} - \bar{Y}_{i,g} = \underbrace{\left\{ \sum_{i,gm=1}^{N} \frac{F(\widehat{\beta}^{gm} X_{i,gm})}{N^{gm}} - \sum_{i,g=1}^{N} \frac{F(\widehat{\beta}^{gm} X_{i,g})}{N^{g}} \right\}}_{\texttt{explained}} + \underbrace{\left\{ \sum_{i,g=1}^{N} \frac{F(\widehat{\beta}^{gm} X_{i,g})}{N^{g}} - \sum_{i,g=1}^{N} \frac{F(\widehat{\beta}^{g} X_{i,g})}{N^{g}} \right\}}_{\texttt{unexplained}} \quad (8)$$

Where $\bar{Y}$ is the mean predicted probability of being engaged in high skilled occupations and N is the population size. The explained term represents the difference in occupational choice between GM and non GM group attributed to changes in observables or X_i constant over the groups holding $\widehat{\beta_{gm}}$ constant. The unexplained term represents the difference in occupational choice attributed to changes in $\widehat{\beta}$ while keeping X_i constant. Equation 2 can alternatively, be specified as the following where the $\widehat{\beta}$ is of the reference group 'g' or non GM group:

$$\bar{Y}_{i,gm} - \bar{Y}_{i,g} = \underbrace{\left\{ \sum_{i,gm=1}^{N} \frac{F(\widehat{\beta}^{g} X_{i,gm})}{N^{gm}} - \sum_{i,g=1}^{N} \frac{F(\widehat{\beta}^{g} X_{i,g})}{N^{g}} \right\}}_{\texttt{explained}} + \underbrace{\left\{ \sum_{i,gm=1}^{N} \frac{F(\widehat{\beta}^{gm} X_{i,gm})}{N^{gm}} - \sum_{i,gm=1}^{N} \frac{F(\widehat{\beta}^{g} X_{i,gm})}{N^{gm}} \right\}}_{\texttt{unexplained}} \quad (9)$$

Since there is some debate as to which groups $\beta's$ should be chosen as the reference category, a pooled regression with a group membership indicator as the reference coefficients is used for the decomposition analysis (Jann, 2008; Fortin, 2006). The results of B-0 decomposition for non linear models (logit) for men from General group (GM) and men from SC/ST group (SM) can be referred to in Table 6. We first estimate the mean difference in the predicted probability of engaging in high skilled occupations in 1983 (col:1) and then in 2012 (col:2). The change in occupational choices between groups and over time is included in col (3) and col(4) displays the proportion of each explained component of the total predicted change in occupation probability of engaging in high skill occupations.

The results in Table 6 suggest that the mean prediction of engaging in high skilled occupations in 1983 was higher for GM (0.28) than SM (0.10), yielding the mean predicted difference of 0.172. In 2012, this difference declined to 0.12. For both 1983 and 2012 the share of difference in occupational choices explained by observables is more than 100% (pooled coefficients). This implies that education, wages, along with other socio-economic characteristics fully explain the difference

[20]Along with two-fold decomposition (as described in the text) this decomposition can be written as three fold. Refer to Appendix A6 for details.

Table 6: B-O Decomposition of Occupational Convergence between GM and SM over time

Decomposition	1983	2012	Change: 2012-1983	%
	(1)	(2)	(3)	(4)
General: Men (GM)	0.281*** (0.0021)	0.3615*** (0.0044)		
SC/ST: Men (SM)	0.109*** (0.0027)	0.2403*** (0.0037)		
Total Difference	0.172*** (0.0033)	0.1212*** (0.0055)	-0.0508 (Pr=0.000)	100
Explained Component	0.173*** (0.0021)	0.1421*** (0.0038)	-0.0309 (Pr=0.000)	60.8
Covariates in explained component				
Years of education	0.122*** (0.0030)	0.0904*** (0.0028)	-0.0316 (Pr=0.000)	102.2
Log of wage	0.0378*** (0.0018)	0.0322*** (0.0016)	-0.0056 (Pr=0.0209)	18.12
Other control variables	0.0136*** (0.0025)	0.0195*** (0.0024)	0.0059 (Pr=0.0877)	-19.09

Note: Authors' calculation. The table records results of non linear decomposition (logit); with bootstrapped standard errors in parenthesis. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Coefficients from a pooled regression with a group membership indicator as the reference coefficients is used for decomposition. Col (3) records the change in 2012-1983 for each component and includes the results of the chi test in parenthesis (p value) to signify whether the change is significant. Other control variables includes household and socio- economic characteristics such as age, monthly household income, religion, gender and educational level of the household head, no. of male and female members of the households, and state fixed effects. The sample includes only those individuals who were employed, & in the age bracket 25-60 years.

in occupational probabilities.[21] In 1983 among the explained components, differences in education attained contributed to 70% of the difference in probability to be engaged in high skilled occupations between SM and GM. Whereas differences in wages (log) attributed around 22%. The share of explained component in 2012 was 117%. Among the observables, 63% attributed to differences in education and 23% by differences in wages. If we observe in detail, most of the overall convergence in occupational choice can be attributed to the decline in educational barriers between SM and GM.

[21]Proportions greater than 100 imply that if only the attributes taken into consideration were responsible for changes in the dependent variable, then the decline in talent misallocation should have been larger than what we observe (Afridi et al., 2018b). The negative contribution of an explanatory, such as the difference in wage(log), will imply that the difference in wages increased hence it would explain a negative amount of convergence in occupational choices

Table 7: B-O Decomposition of Occupational Convergence between GM and SF over time

Decomposition	1983	2012	Change: 2012-1983	%
	(1)	(2)	(3)	(4)
General: Men (GM)	0.281***	0.3615***		
	(0.0021)	(0.0044)		
SC/ST: Females (SF)	0.0484***	0.1989***		
	(0.0028)	(0.0049)		
Total Difference	0.233***	0.1626***	-0.070	100
	(0.0032)	(0.0070)	(Pr=0.000)	
Explained Component	0.241***	0.2298***	-0.0112	16
	(0.0026)	(0.0046)	(Pr=0.0423)	
Covariates in explained component				
Years of education	0.1537***	0.1578***	0.0041	-36.6
	(0.0042)	(0.0049)	(Pr=0.5817)	
Log of wage	0.0695***	0.0577***	-0.0118	105.3
	(0.0030)	(0.0031)	(Pr=0.0045)	
Other control variables	0.0177***	0.0142***	-0.0035	31.2
	(0.0031)	(0.0029)	(Pr=0.435)	

Note: Authors' calculation. The table records results of non linear decomposition (logit); with bootstrapped standard errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1. Coefficients from a pooled regression with a group membership indicator as the reference coefficients is used for decomposition. Col (3) records the change in 2012-1983 for each component and includes the results of the chi test in parenthesis (p value) to signify whether the change is significant. Other control variables includes household and socio- economic characteristics such as age, monthly household income, religion, gender and educational level of the household head, no. of male and female members of the households, and state fixed effects. The sample includes only those individuals who were employed, & in the age bracket 25-60 years.

Similarly, in Table 7, we observe that over time the difference in the mean predicted probability of engaging in high skilled occupation between GM and females from the SC/ST group (SF) has significantly decreased over time. Thus, suggesting a decline in talent misallocation for the group between 1983-2012. In 1983, the explained component share was around 103%. Among the explained component individually in both 1983 and 2012, the differences in education attainment attributed the most to the difference in mean occupational probabilities (63% and 68%). Though over time, explained component contributes to only 16% of the decline in talent misallocation.[22] Out of the

[22]Thus, suggesting that over time unobservable factors such as social norms and cultural differences explain a substantial share (84%) of occupational convergence between SC/ST females and men from general group.

explained component only changes in wage differentials significantly attributed to the decline in the difference in occupational probability for females from the SC/ST group.

Table 8: B-O Decomposition of Occupational Convergence between GM and GF over time

Decomposition	1983	2012	Change: 2012-1983	%
	(1)	(2)	(3)	(4)
General: Men (GM)	0.2814*** (0.0021)	0.3615*** (0.0044)		
General: Females (GF)	0.2442*** (0.0042)	0.4900*** (0.0087)		
Total Difference	0.0372*** (0.0043)	-0.1284*** (0.0105)	-0.1656 (Pr=0.000)	100
Explained Component	0.1141*** (0.0026)	0.0067 (0.0068)	-0.107 (Pr=0.000)	64.6
Covariates in explained component				
Years of education	0.065*** (0.0019)	0.0019 (0.0002)	-0.0631 (Pr=0.000)	58.9
Log of wage	0.0444*** (0.0019)	0.0042 (0.0041)	-0.0402 (Pr=0.000)	37.5
Other control variables	0.0047*** (0.0013)	0.0005 (0.0008)	-0.0042 (Pr=0.0193)	3.9

Note: Authors' calculation. The table records results of non linear decomposition (logit); with bootstrapped standard errors in parenthesis. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Coefficients from a pooled regression with a group membership indicator as the reference coefficients is used for decomposition. Col (3) records the change in 2012-1983 for each component and includes the results of the chi test in parenthesis (p value) to signify whether the change is significant. Other control variables includes household and socio- economic characteristics such as age, monthly household income, religion, gender and educational level of the household head, no. of male and female members of the households, and state fixed effects. The sample includes only those individuals who were employed, & in the age bracket 25-60 years.

For females from the general group (GF), the results in Table 8 suggests, the majority of the total difference can be attributed to the observables or explained components in 1983. Out of which, differences in education have a larger share, around 57%. If we focus on col(4), we observe that the decline in talent misallocation between GF and GM can be significantly contributed to the observables. Among decline in frictions, again decline in educational barriers had a larger share contributing to the changes in talent allocation over time for this group.

Overall for both men from SC/ST group and females from general group, the decline in educational barriers contributed the most to the decline in talent misallocation over time. The decomposition analysis suggests that this was not true for females from SC/ST group as only the decline in wage differentials contributed significantly in the decline in talent allocation over time.

8 Contribution and Potential Mechanism

This paper quantifies talent misallocation in India based on social identity and gender. I first provide evidence of barriers in attaining higher education and wage discrimination towards marginalised groups. Further, I relate these barriers with talent misallocation for the Indian economy, which is the paper's main contribution.

Highlighting important findings from the paper. First, I show that even in 2012, men from the general group (GM) are more likely to transition to a higher education degree even though the barriers for other groups have declined over time. However, SC/ST men still fare the worst in attaining higher education among the SC/ST group. Second, wage differences indicate labor market discrimination towards non GM groups in high skilled occupations, and over time they persist even though there was some level of convergence, still GM earns significantly more than individuals from other groups. Specifically, results indicate that attaining higher education plays a significant role in determining wages and occupational choice. As expected, the results show that workers with higher education get higher returns when employed in high skilled occupations. However, at higher education level, women and SC/ST workers earn lesser when compared with men from the general category. This disparity exists in both 1983 and 2012. In parallel, we also observe an increase in the average predicted probability of workers from these groups to be employed in high skilled occupations.

On the one hand, disadvantaged groups are more likely to be employed in high skilled occupations in 2012 than in 1983 in comparison to general men. However, on the other hand, the return of employment (wages) would still be lower than men from the general category. Results of the reduced form approach and the occupational similarity index imply convergence of workers from other groups engaging in higher skilled occupations even in the presence of wage gaps.[23]

[23] Klasen and Pieters (2015) recorded a change in occupational choices by educated women in their study. Where the authors observe a movement towards business and financial sectors of the group as their participation in the public sector declined in 2011 compared with 1987.

The results indicate that even in the presence of barriers, there exist returns to higher education. So why was it the case that overtime workers from minority groups could access institutions imparting higher education? Were there any exogenous factors or policies which encouraged this movement? Did the structural change in India push forward this movement? One such policy implemented was the reservation policy introduced to equalize opportunities between the various caste groups.[24] There is some evidence of higher enrolment and ease in access of SC/ST individuals to higher education institutions due to the reservation policy (Deshpande, 2013; Cassan, 2019; Weisskopf, 2004; Lee, 2019).

Restrictive traditional networks and community based networking can also be a potential reason why previously, disadvantaged groups were not moving towards higher education even when there were clear premiums of attaining higher education. The literature suggests that individuals from disadvantaged groups are locked in traditional caste based networks and occupations, it can be the case that disadvantaged groups choose education based on their rigid career choices. Overtime decaying of these traditional community based networking can be a reason for workers to prefer higher education (Munshi and Rosenzweig, 2006; Munshi, 2011; Cassan et al., 2021).

If we focus on gender inequality and talent allocation, there is evidence that women are underrepresented in top jobs (or at the managerial level). After taking into account the ability gap, Ashraf et al. (2021), find that overall the gender wage gap seems to be underestimated in their study across nations. In a scenario where men and women have the same skills or ability, traditional norms reinforce the equilibrium where women should pursue staying and working at home rather than encouraging them to participate in the labor market. There is also evidence showing that the ability gap is correlated with the prevalence of gender norms. For example, men make better executives than women, a pre- school child suffers with working mother or men have more rights to a job than women. This correlation can be even more severe or prevalent when we intertwine gender with caste. Hence, this encourages us to think about both short term and long term effects of policies mitigating inequalities of access to opportunities, be it due to gender or caste discrimination. A policy pushing from one low level equilibrium to a high level equilibrium point can help improve talent

[24] Quota in educational institutions and employment seats for SC/ST was first introduced in the early 1950s and was extended to Other Backward Classes (OBCs) in the early 1990s. Gender based quota in the political sphere was also introduced in 1994, where 33% of seats in local government were reserved for women.

allocation, which might be more relevant for developing nations.

The third finding of the paper suggests that educational convergence and convergence in wages are significantly correlated with occupational convergence. Thus, suggesting that a decline in barriers can be correlated with better resource allocation and further can have an effect at the aggregate level. Furthermore, I find that convergence in educational attainment among these two barriers accounts for most of the improvement in talent allocation for men from SC/ST group and females from general group. For females from SC/ST group convergence in wage differentials accounted for significant contribution in better talent allocation.

As a final thought, the stylized facts from this paper motivate us to extend the analysis and provide a theoretical framework and a structural model to understand the reasons behind them. Future extensions can include a more comprehensive framework that focuses on other factors, such as specific policies which had a role in reducing these barriers. One such policy is an affirmative action policy. Examining the linkages of this policy in India's education and labor market can help us identify the mechanisms that played a role in improving talent allocation in the economy.[25]

Similarly, there can be other potential reasons for improved talent allocation, which I have not discussed in the paper. I do not discuss the role of mobility or migration patterns of disenfranchised groups to locations where there is lower labor market discrimination that enables them to participate in previously unavailable jobs. Measurement and inclusion of traditional social norms towards women and marginalised caste groups in Indian society can provide a holistic understanding of the decline in talent misallocation. It might also be relevant to extend the analysis by focusing on soft skills v/s hard skills or abstract tasks v/s contact based jobs (Hurst et al., 2021). This might be especially relevant in explaining the growth in the Indian economy, which is attributed to the growth in the service sector.

[25] The next chapter of my doctoral thesis focuses on this particular extension.

References

Abdulla, K. (2019). Productivity gains from reallocation of talent in brazil and india. *Journal of Macroeconomics*, 62:103160.

Afridi, F., Dinkelman, T., and Mahajan, K. (2018a). Why are fewer married women joining the work force in rural india? a decomposition analysis over two decades. *Journal of Population Economics*, 31(3):783–818.

Afridi, F., Dinkelman, T., and Mahajan, K. (2018b). Why are fewer married women joining the work force in rural india? a decomposition analysis over two decades. *Journal of Population Economics*, 31(3):783–818.

Agrawal, T. (2020). Gender segregation and wage differentials in india: the role of educational attainment and occupational choices. *International Journal of Manpower*.

Aigner, D. J. and Cain, G. G. (1977). Statistical theories of discrimination in labor markets. *Ilr Review*, 30(2):175–187.

Albelda, R. P. (1986). Occupational segregation by race and gender, 1958–1981. *ILR Review*, 39(3):404–411.

Ashraf, N., Bandiera, O., Minni, V., and Quintas-Martínez, V. (2021). The misallocation of women's talent across countries: Evidence from personnel data.

Banerjee, B. and Knight, J. B. (1985). Caste discrimination in the indian urban labour market. *Journal of development Economics*, 17(3):277–307.

Becker, G. S. (1971). *The Economics of Discrimination*. The University of Chicago Press Books.

Blinder, A. S. (1973). Wage discrimination: reduced form and structural estimates. *Journal of Human resources*, pages 436–455.

Cassan, G. (2019). Affirmative action, education and gender: Evidence from India. *Journal of Development Economics*, 136(C):51–70.

Cassan, G., Keniston, D., and Kleineberg, T. (2021). A division of laborers: Identity and efficiency in india. Working Paper 28462, National Bureau of Economic Research.

Das, M. B. and Dutta, P. V. (2007). Does caste matter for wages in the indian labor market. *Washington, DC, USA: The World Bank*.

Datta, A., Endow, T., and Mehta, B. S. (2020). Education, caste and women's work in india. *The Indian Journal of Labour Economics*, 63(2):387–406.

Deininger, K., Jin, S., and Nagarajan, H. (2013). Wage discrimination in india's informal labor markets: Exploring the impact of caste and gender. *Review of Development Economics*, 17(1):130–147.

Desai, S. and Kulkarni, V. (2008). Changing educational inequalities in india in the context of affirmative action. *Demography*, 45(2):245–270.

Deshpande, A. (2011). *The grammar of caste: Economic discrimination in contemporary India*. Oxford University Press.

Deshpande, A. (2013). Social justice through affirmative action in india: An assessment. In *Capitalism on Trial*. Edward Elgar Publishing.

Deshpande, A., Goel, D., and Khanna, S. (2018). Bad karma or discrimination? male–female wage gaps among salaried workers in india. *World Development*, 102:331–344.

Duncan, O. D. and Duncan, B. (1955). A methodological analysis of segregation indexes. *American sociological review*, 20(2):210–217.

Fortin, N. M. (2006). Greed, altruism, and the gender wage gap. *Unpublished manuscript, Department of Economics, University of British Columbia*.

Gang, I. N., Sen, K., and Yun, M.-S. (2017). Is caste destiny? occupational diversification among dalits in rural india. *The European Journal of Development Research*, 29(2):476–492.

Hnatkovska, V., Lahiri, A., and Paul, S. (2012). Castes and labor mobility. *American Economic Journal: Applied Economics*, 4(2):274–307.

Hsieh, C., Hurst, E., Jones, C. I., and Klenow, P. J. (2019). The allocation of talent and us economic growth. *Econometrica*, 87(5):1439–1474.

Hurst, E., Rubinstein, Y., and Shimizu, K. (2021). Task-based discrimination.

Jann, B. (2008). The blinder–oaxaca decomposition for linear regression models. *The Stata Journal*, 8(4):453–479.

Jong-Wha, L. and Wie, D. (2017). Wage structure and gender earnings differentials in china and india. *World Development*, 97:313–329.

Klasen, S. and Pieters, J. (2015). What explains the stagnation of female labor force participation in urban india? *The World Bank Economic Review*, 29(3):449–478.

Lee, A. (2019). Does affirmative action work? evaluating india's quota system. *Comparative Political Studies*, page 0010414021989755.

Lillard, L. A. and Willis, R. J. (1994). Intergenerational educational mobility: Effects of family and state in malaysia. *The Journal of Human Resources*, 29(4):1126–1166.

Madheswaran, S. and Attewell, P. (2007). Caste discrimination in the indian urban labour market: Evidence from the national sample survey. *Economic and political Weekly*, pages 4146–4153.

Maitra, P. and Sharma, A. (2009). Parents and children: Education across generations in india. In *5th annual conference on economic growth and development, ISI Delhi, Delhi*. Citeseer.

Motiram, S. and Singh, A. (2012). How close does the apple fall to the tree? some evidence from india on intergenerational occupational mobility. *Economic and Political Weekly*, pages 56–65.

Mukherjee, A. and Paul, S. (2012). Community identity and skill mismatch: A study on indian labour market. In *Annual Conference on Economic Growth and Development (New Delhi, 2012)*.

Munshi, K. (2011). Strength in numbers: Networks as a solution to occupational traps. *The Review of Economic Studies*, 78(3):1069–1101.

Munshi, K. and Rosenzweig, M. (2006). Traditional institutions meet the modern world: Caste, gender, and schooling choice in a globalizing economy. *American Economic Review*, 96(4):1225–1252.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International economic review*, pages 693–709.

Porzio, T. (2017). Cross-country differences in the optimal allocation of talent and technology. *Working Paper*.

Sahoo, S. and Klasen, S. (2021). Gender segregation in education: Evidence from higher secondary stream choice in india. *Demography*, 58(3):987–1010.

Sengupta, A. (2017). Mismatch between skills and jobs in indian labour market during the post-reform era: estimates with unit level data. In *IARIW ICRIER Conference,(New Delhi)*.

Sengupta, A. and Das, P. (2014). Gender wage discrimination across social and religious groups in india: Estimates with unit level data. *Economic and Political Weekly*, pages 71–76.

Sloane, C., Hurst, E., and Black, D. (2019). A cross-cohort analysis of human capital specialization and the college gender wage gap.

Varughese, A. R. and Bairagya, I. (2020). Group-based educational inequalities in india: Have major education policy interventions been effective? *International Journal of Educational Development*, 73:102159.

Weisskopf, T. E. (2004). Impact of reservation on admissions to higher education in india. *Economic and Political Weekly*, pages 4339–4349.

Appendix

A1. Data Preparation

In this section I discuss the steps taken for data preparation. For every round, data collected for different blocks/sections is stored in separate Stata files (for example household details, member details, occupational details, employment status and monthly expenditure details are stored in separate .dta files). For our analysis, I use both individual and household level files hence I merge each block to make a master file for that particular round. However merging blocks becomes a bit tricky for some rounds as the unique identifiers can be mixed up or not defined in a similar manner for each block (issue more prominent for the 38th round).

Next step in the data preparation process is to clean the data. One issue in the NSS data was that over the rounds the definition of several variables changed or many additional categories in a variable were introduced (social group, educational categories, monthly expenditure and the occupations). So mapping these changes was an essential part of the process.

To prepare the wage data, it is important to note how NSS specifies an occupation code. NSS collects two weekly activity details of each member and records the intensity of their work each day and then estimates the total number of days work in a week (summing the intensity for each day) for that particular activity. An activity which is pursued for more than 1 hour but less than 4 hours is considered to have been pursued with 'half' intensity, if pursued for 4 hours or more, the activity is considered to have been pursued with 'full' intensity. In a case where a member of the household is engaged in more than two activities, details of only those activities are recorded in which they spend their majority time. An occupational code will be based on the economic activity in which the highest number of days have been spent. To map the occupations over NCO 1968 and 2004 the occupational codes corresponding to these activities (current weekly occupation status) have been used. I then keep the wages of the highest intensity activity corresponding to the weekly occupational code. So the data has a unique entry (row) for each member with occupational code details, wages and other household and individual characteristics. A master data file is then prepared by appending data files of all the rounds.

A2. Occupational Codes

I referred to ICSSR conversion of three digit occupation codes between 1968 and 2004 NCO lists.[26] Following are the details of the 20 occupations mapped:

Broad Occupation 0: "Scientists"

(Includes: Physicist, Chemists, Geologists, Astronomer (physical scientists), Biologists, Zoologists, Botanists,Bacteriologists, Pharmacologists, Silviculturists,Agronomists and Agricultural Scientists and Related Scientists)

Broad Occupation 1: "Aircraft/Ship officers and Science and Engineering Technicians"

(Includes: Laboratory Assistant, Slide Examiner, Draughtsmen; Civil Engineering,Electrical and Electronic Engineering, Metallurgical Overseers and Technicians, Aircraft Pilots, Flight Engineers/Navigators, Ship and Deck Officers and Pilots)

Broad Occupation 2: "Architects and Engineers"

(Includes: Architects and Town Planners, Civil Engineers, Electrical and Electronic Engineers, Mechanical Engineers,Chemical Engineers,Metallurgists, Mining and Industrial Engineers, Surveyor)

Broad Occupation 3: "Nurses and Other medical staff"

(Includes: Vaccinators, Inoculators and Medical Assistants,Dental,Veterinary,Pharmaceutical Assistants,Nurses,Optometrists and Opticians, Dietitians and Nutritionists)

Broad Occupation 4: "Physicians & Surgeons"

Broad Occupation 5: "Mathematicians/Economists and other social scientists"

(Includes: Economists, sociologists, Historians, Archaeologists Political Scientists Related Workers, Labor, Social Welfare Political Workers etc)

Broad Occupation 6: "Accountants and auditors"

(Includes: Personnel manager, manpower manager, Accountants and Auditors,Cost and Works Accountants)

Broad Occupation 7: "Jurists"

(Includes: Lawyers, Judges and Magistrates)

[26] Access the document here: http://www.icssrdataservice.in/datarepository/index.php/catalog/1/download/12

Broad Occupation 8: "Teachers"

(Includes: University and college professors,School teachers-Higher, middle, secondary school, primary)

Broad Occupation 9: "Artists"

(Includes: Poets, Authors and Critics,Editors and Journalists,Sculptors, Painters, Photographers, Music composers and performing artists, Sportsperson)

Broad Occupation 10: "Priests and other religious workers"

(Includes: Astrologers, Palmists and Related Workers, Ordained Religious Workers,Non-ordained Religious Workers)

Broad Occupation 11: "Elected and legislative officials/Administrative Executives, Govt. and local bodies"

(Includes: Union, state, local bodies, Administrative Executive Officials-Union, state, local bodies)

Broad Occupation 12: "Managers"

(Includes: Working proprietors, Directors and Managers of wholesale and retail trade, financial institutions, mining, construction, manufacturing, transport, storage, farm and dairy)

Broad Occupation 13: "Operators"

(Includes: Machine operators, Book keepers, Cashiers, Stenographers, Typists, Ticket Sellers, Duplicating/embossing/addressing Machine operator,bill collectors, money lender and pawn brokers)

Broad Occupation 14: "Clerical"

(Includes: Clerical and related workers, Guards and Breaks Men, Railway,Postmen and Telephone operators)

Broad Occupation 15: "Sales workers"

(Includes: Merchants and Shopkeepers, Wholesale and retail traders, Manufacturer's agents, Salesman, Shop assistants, Insurance, Real estate, securities and business service salesman and auctioneers)

Broad Occupation 16: "Service workers"

(Includes: Hotel and Restaurant keepers, Matrons and Stewards, Cooks, Waiters, Bartenders, Maids, Launderers, Dry cleaners, Building caretakers, Cleaners, Barbers, Beauticians)

Broad Occupation 17: "Protective service workers"

(Includes: Fire Fighters, Policemen and Detectives,Customs Examiners, Patrollers Related Workers,Watchmen, Gate Keepers,Protection Force, Home Guard and Security Workers)

Broad Occupation 18: "Agriculture and Forest Workers"

(Includes: Cultivators, Other farmers, Agricultural laborers, Plantation laborers, Forestry and fisherman, Hunters)

Broad Occupation 19: "Production and Transport equipment laborers"

(Includes: Miners, Quarry men, Well drillers, Metal processors, Wood preparation workers,Spinners, Weavers, Knitting, Tanners,Tailors,Shoemakers,Carpenters, Plumbers, Welders etc)

I classify these occupations under two major categories- high and low skilled occupations, details of each of these classifications are as following:

- **High Skilled Occupations**: Scientists, Architects and Engineers, Aircraft and ship officers, Physicians and Surgeons, Mathematicians, Economists and other social scientists, Accountant and auditors, Jurists, Teachers, Artists and Managers, Engineering technicians, Nurses and other medical health officers, Dietitians, Elected legislative officials and govt. local bodies, Operators and Protective workers.

- **Low Skilled Occupations**: Priests and other religious workers, Clerical, Sales workers, Service workers, Agriculture and Forest Workers and Production and Transport equipment laborers.

A3. Summary Statistics

To give a sense of how the Indian data looks like, this section presents some basic summary statistics. Essentially I divide the social groups into two, first General and second one combining both SC and ST group and understand how the distribution changes for both men and women.

In the Indian dataset, on average, General men are slightly older than SC/ST men. This difference is significant at the 10 percent level of significance.[27] Since one of our objective is to investigate whether educational barriers based on caste and gender is related to talent misallocation in the Indian labor market, I first focus on the trends in mean years of education over time. NSS collects data on educational categories such as below primary, primary, middle, secondary and graduation and above.[28] For the purpose of our analysis instead of educational categories, I convert each of these categories into years of education.[29]

Figure 3: Mean years of education

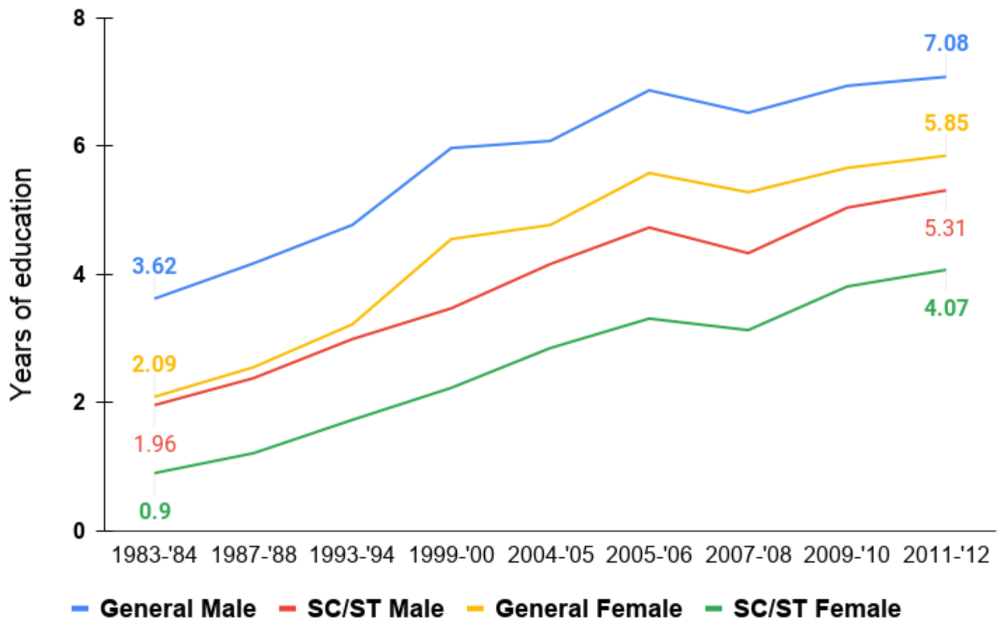


Figure3 shows how educational attainment has changed over the years. There has been a clear upward trend for all four groups. Overall, men and women from the general category have more years of education than SC/ST men and women on average. This difference is significant at 5% level

[27]Refer Table A1 in appendix

[28]However, additional categories such as higher secondary, diploma, post graduation and above have been introduced in recent rounds but unavailable in the early rounds.

[29]Illiterate=0 years, literate but below primary=2 years, primary=5 years, middle= 8 years, secondary and higher secondary=10 years, graduate and above=15 years. Note: I exclude diploma and literate w/o formal schooling.

of significance. If we focus on general and SC/ST men, the gap was at 84% in 1983, which reduced to 33% in 2012. Similar convergence in the attainment of education but with a higher magnitude can be seen for women. The gap between general and SC/ST women was 132% in 1983 which reduced to 43% in 2012. If we look in detail only for employed individuals, it seems that the Secondary and Higher secondary educational category drives this result. Individuals with secondary and higher secondary education increased by around 13 percentage points for general and 9.7 percentage points for SC/ST men and for educational category- Graduation and above by 9 percentage points (192% increase) and 2.53 percentage points (356% increase) respectively.

When we observe the composition by location, we observe that SC/ST as a group is more likely to live in the rural sector when compared with individuals belonging to the general category. This difference is statistically significant (Table A1). Over the years, we can see a trend of people moving out from rural to urban areas. Overall summary statistics (Table A1: Part A) and descriptive graphs suggest that general men, when compared with SC/ST men, have higher years of education, less likely to live in rural areas and are older on average. Similar is true for general and SC/ST women. (Table A1: Part B).

NSS has used National Classification of Occupations (NCO) 1968 and 2004 to classify each individual's occupational details over various rounds. I could successfully map 20 broad comparable occupations (at 3 digits) from both the classification lists.[30]

To get a broad understanding of the proportion of workers in these occupations I have clubbed them into two categories, high skilled occupation and low skilled occupations.From the data we observe that over the years, general men in the labor force, in the 25-60 years of age bracket are moving towards being employed in high skilled occupations (change of 17 percentage points). [31] They are approximately three times more likely to get engaged in high skilled occupations when compared with 1983. Similar is true for SC/ST men.

[30] Used- http://www.icssrdataservice.in/datarepository/index.php/catalog/2/download/47. Details in Appendix A3

[31] Refer figure A3 and A4 in appendix.

Table A1: Sample Summary Statistics (Full Sample)

Part A: Difference between General Male and SC/ST Male

	Years of Education	Age	Married	Rural
Round:38	-1.62***	-1.055***	0.011***	0.19***
(1983'-84)	(0.01)	(0.07)	(0.00)	(0.00)
Round:43	-1.72***	-1.55***	0.00	0.18***
(1987 -'88)	(0.01)	(0.07)	(0.00)	(0.00)
Round:50	-1.78***	-1.62***	-0.005***	0.18***
(1993 -'94)	(0.01)	(0.07)	(0.00)	(0.00)
Round:55	-2.51**	-2.51*	-0.02***	0.23***
(1999 -2000)	(0.02)	(0.08)	(0.00)	(0.00)
Round:61	-1.97***	-2.78***	-0.03***	0.199***
(2004 -'05)	(0.02)	(0.08)	(0.00)	(0.00)
Round:62	-2.17***	-2.90*	-0.03***	0.28***
(2005 -'06)	(0.02)	(0.10)	(0.00)	(0.00)
Round:64	-2.10***	-3.11***	-0.042***	0.27***
(2007 -'08)	(0.02)	(0.09)	(0.00)	(0.00)
Round:66	-1.94***	-3.166***	-0.03***	0.22***
(2009 -'10)	(0.02)	(0.98)	(0.00)	(0.00)
Round:68	-1.80**	-3.081***	-.040***	0.206***
(2011 -'12)	(0.02)	(0.10)	(0.00)	(0.00)

Note: Authors' own calculation
This table reports summary statistics for all samples for all EUS NSS rounds.
Panel A reports the differences between General male and SC/ST male characteristics.
Parentheses reports the standard errors of t-tests.
*** Significant at 1% level; ** Significant at 5% level; * Significant at 10% level.

Table A1: Sample Summary Statistics (Full Sample) Contd.

Part B: Difference between General Female and SC/ST Female

	Years of Education	Age	Married	Rural
Round:38	-1.17***	-1.05***	0.009***	0.17***
(1983'-84)	(0.01)	(0.08)	(0.00)	(0.00)
Round:43	-1.31***	-1.51***	-0.002	0.16***
(1987 -'88)	(0.01)	(0.07)	(0.00)	(0.00)
Round:50	-1.48***	-1.78***	-0.008***	0.17***
(1993 -'94)	(0.01)	(0.08)	(0.00)	(0.00)
Round:55	-2.31***	-2.83*	-0.03***	0.22***
(1999 -2000)	(0.01)	(0.08)	(0.00)	(0.00)
Round:61	-1.91***	-2.95***	-0.04***	0.12***
(2004 -'05)	(0.02)	(0.09)	(0.00)	(0.01)
Round:62	-2.25***	-2.81***	-0.038***	0.20***
(2005 -'06)	(0.02)	(0.1)	(0.00)	(0.01)
Round:64	-2.18**	-3.26*	-0.047***	0.21***
(2007 -'08)	(0.02)	(0.09)	(0.00)	(0.00)
Round:66	-1.88***	-3.5***	-0.050***	0.20***
(2009 -'10)	(0.02)	(0.10)	(0.00)	(0.01)
Round:68	-1.798***	-3.630***	-0.053***	0.189***
(2011 -'12)	(0.02)	(0.10)	(0.00)	(0.00)

Notes: Authors' own calculation
This table reports summary statistics for all samples for all EUS NSS rounds.
Panel B reports the differences between General female and SC/ST female characteristics.
Parentheses reports the standard errors of t-tests.
* Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level

Figure 4: Share in High Skilled Occupations, by Caste Group

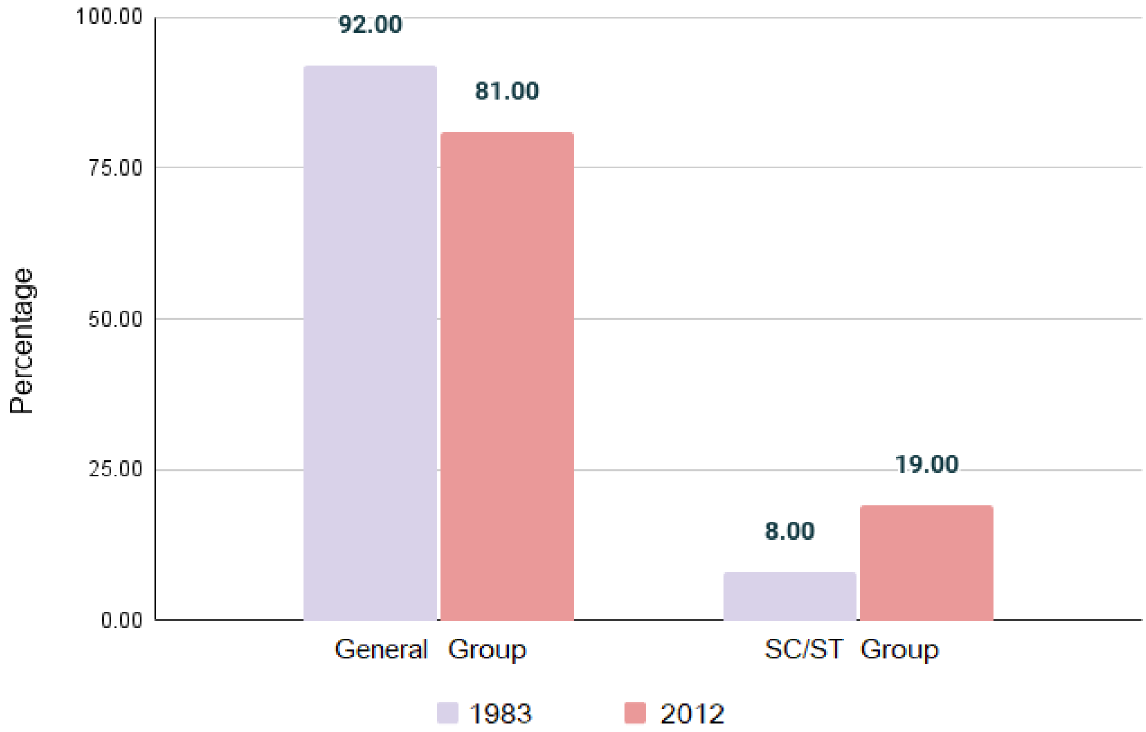


Figure A1: Education Decomposition for General Male (in proportion)

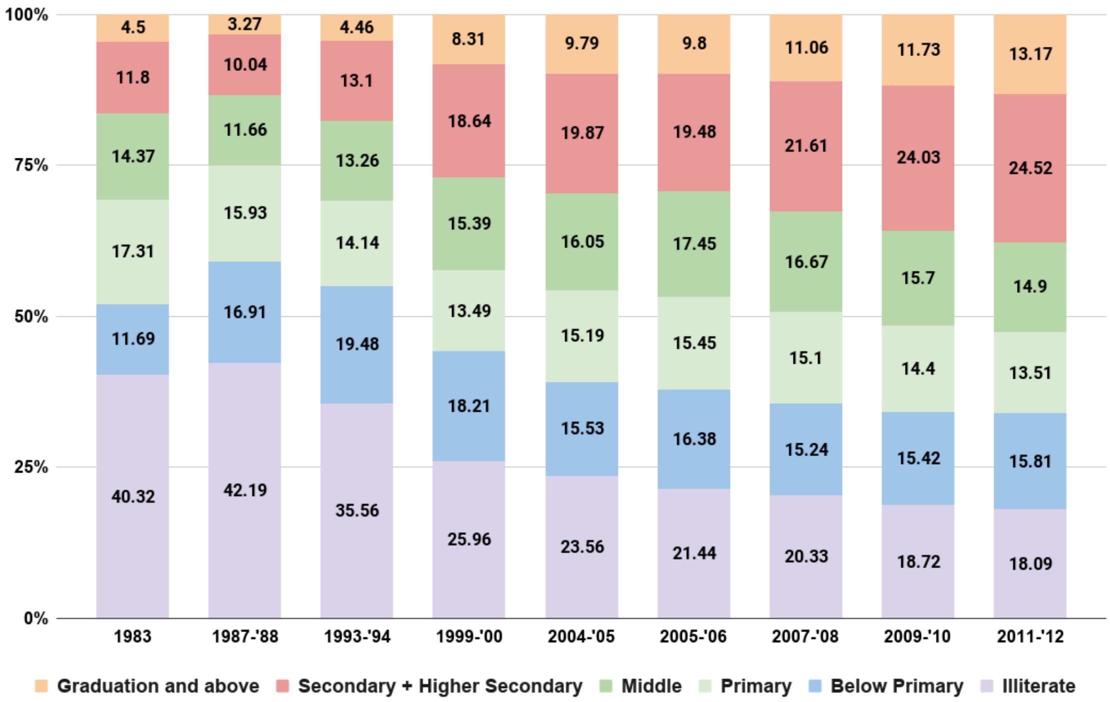


Figure A2: Education Decomposition for SC/ST Male (in proportion)

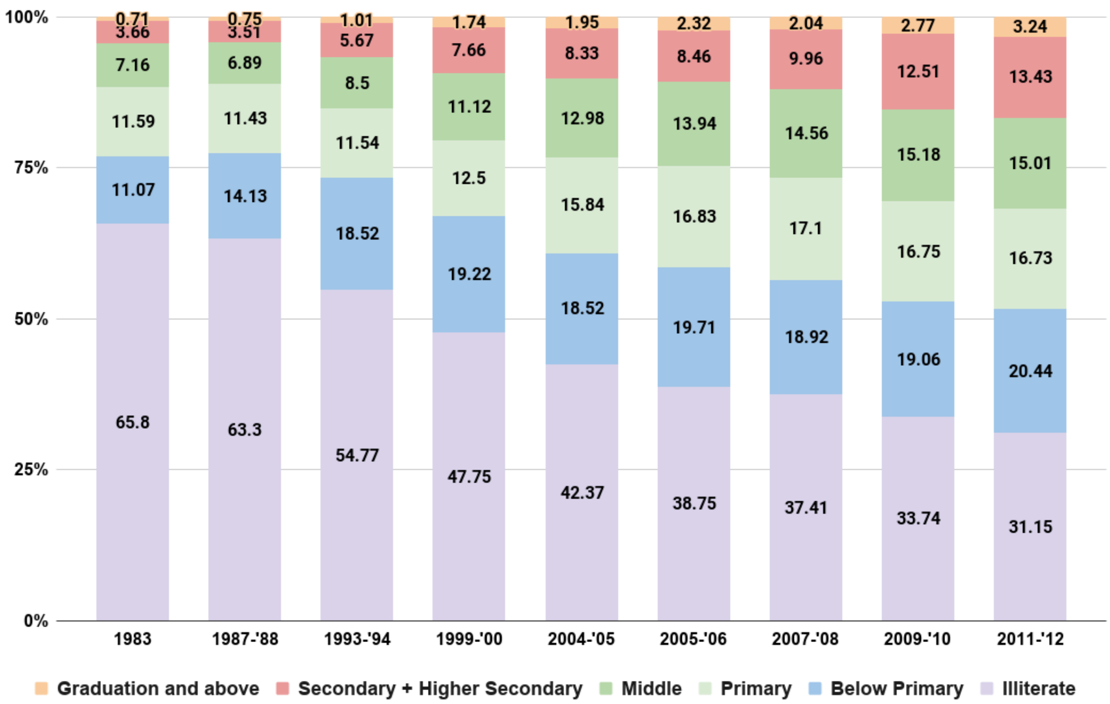


Figure A3: Proportion of General Male Workers by Types of Occupations

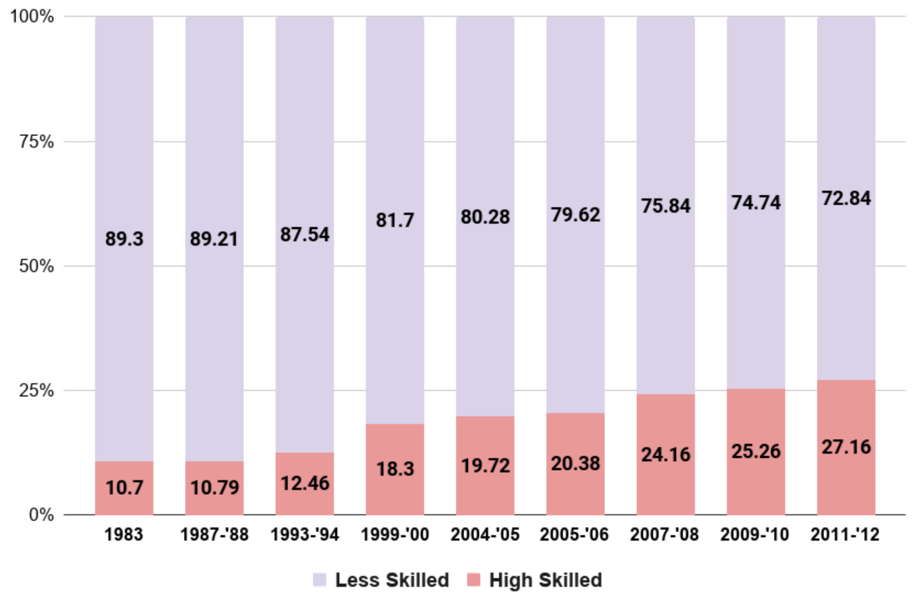
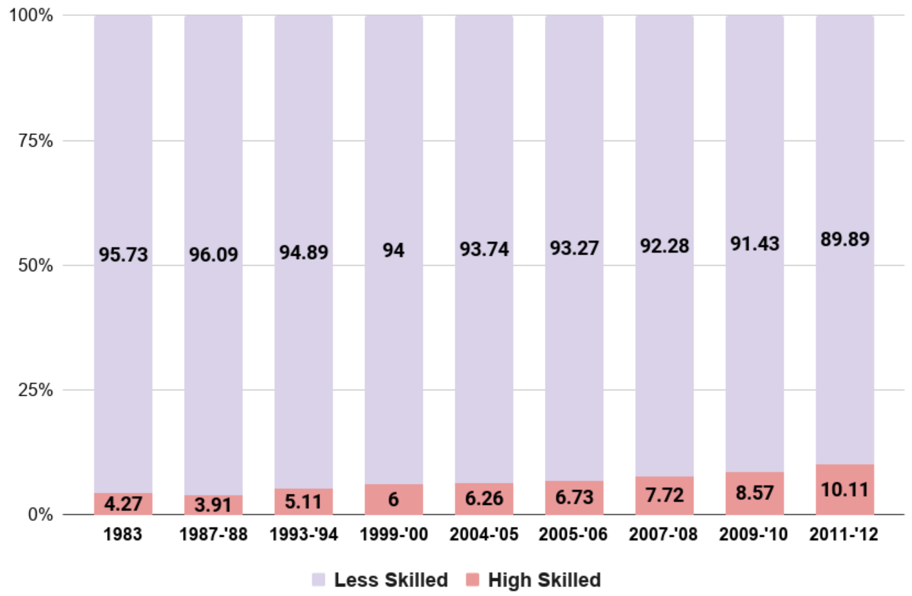


Figure A4: Proportion of SC/ST Male Workers by Types of Occupations



A4. Occupational Choice Model

Similar to the regressions discussed in the previous section, I estimate the following three specifications of the occupational choice model. The results are shown in table 9.

$$\text{High Skilled Occupation}_{it} = \beta_0 + \beta_1 \text{Group}_i + \beta_2 \text{Education attained}_i + \beta_3 \text{Time}_t + \beta_5 X_i + \epsilon_{it} \quad \text{(A1)}$$

$$\begin{aligned} \text{High Skilled Occupation}_{it} =& \beta_0 + \beta_1 \text{Group}_i + \beta_2 \text{Education attained}_i + \beta_3 \text{Time}_t + \beta_4 X_i + \\ & \beta_5 (\text{Group * Education attained})_i + \epsilon_{it} \end{aligned} \quad \text{(A2)}$$

$$\begin{aligned} \text{High Skilled Occupation}_{it} =& \beta_0 + \beta_1 \text{Group}_i + \beta_2 \text{Education attained}_i + \beta_3 \text{Time}_t + \beta_4 (\text{Group*Time})_{it} + \\ & \beta_5 (\text{Group*Education})_i + \beta_6 (\text{Education*Time})_{it} + \\ & \beta_7 (\text{Group*Education*Time})_{it} + \beta_8 X_i + \epsilon_{it} \end{aligned} \quad \text{(A3)}$$

Results of model specification A1 (table 9; col:1) suggest that the probability of individuals from disenfranchised groups (GF/SF/SM) to be employed in high skilled occupations are 0.5% to 1.4% higher relative to GM. Over time, for all groups, this probability has declined. Attaining higher education increases the probability of being engaged in such occupations.

In specification A2 (table 9; col:3) the coefficient of the interaction between *'group*education'* is significant and positive for all groups at higher educational categories relative to general men category. This indicates that individuals from disadvantaged groups are significantly more likely to be employed in high skilled occupations than men from the general category when they move up in the education ladder and attain graduation degree.

Specification A3 (table 9; col:5) suggests that the β_7 coefficient is significant for women and SC/ST group individuals at higher educational categories (graduation and above). Thus indicating that over time women and SC/ST group moving towards getting a graduate degree have a higher predicted probability, around 0.01% to 0.5%, than GM to be in high skilled occupations.

Estimation of the occupational choice model indicates that, on average, over time, the predicted probability of being employed in high skilled occupations is higher for disadvantaged groups when they attain higher education levels relative to GM.

Results of this section indicates that on average, over time, the predicted probability of being employed in high skilled occupations is higher for backward caste groups (relative to GM) as they move to higher education categories. This suggests some level of convergence in occupational choice of workers from marginalised groups and supports the results of Hnatkovska et al. (2012) which shows how occupational convergence occurs when segregated on the basis of caste. Our analysis extends their results by including the gender dimension. This also reveals that caste ties with occupations are slowly decaying over time and that attainment of higher education is enabling them to move out of archaic caste based networks. The average marginal effects of probit occupational choice model suggest that, on average, over time, the change in probability of SC/ST men being engaged in high skilled occupation is 0.18 percentage points higher than GM if they attain a degree of graduation (or above). The change in probability is greater for women (from both the caste groups), between 0.4 to 0.5 percentage points.

The result hints that the rewards of getting higher education (graduation or above) in terms of likelihood of engaging in skilled jobs have increased over time- this is more for disadvantaged groups relative to GM. Gang et al. (2017) particularly looks at the interaction term between SC households and year time dummies for agricultural laborers and suggests that over time SC workers are moving out of this profession and into those which are at the higher end of the occupational spectrum. This result is consistent with the findings of this paper.

However, there is one peculiar result of our study which suggests that even in 1983 individuals from disadvantaged groups (SM/SF) were more likely to be in high skilled occupations relative to GM. The literature does not support this. As even though the results indicate that they are more likely to be in high skilled occupations when I look at the proportion of each group- I still observe general men to have a higher representation in high skilled occupations.

This brings us to some necessary robustness checks which are needed to be certain about our results. We can't ignore the issue of endogeneity in the model specified. When I look at models estimating the returns to wage or occupational choice in a regression analysis, incidental truncation is something that has been talked about repeatedly in the literature. To check whether the findings are robust, I correct for selection bias through the proposed two-step procedure of heckman selection model (for the wage differential model)and bivariate probit model (for the occupational choice model). The results, after correcting for selection, are similar to what I found in our initial analysis in terms of the sign of the coefficient and their significance, even though the magnitude differs.

A4.1. Robustness Checks: Correction of Selection Bias

As discussed previously, I restrict our analysis to occupational choice for only high skilled occupations. Identifying whether an individual is engaged in high skilled occupations will only be observed if they are employed in the labor force. Again, I look at truncated data, and getting engaged in high-skilled jobs can be correlated with unobservable attributes.

When the dependant variable is a binary variable I apply "heckprobit" specification. In the occupational choice model, since our dependant variable takes a value 1 if individuals are employed in high skilled occupations and 0 otherwise, I use the bivariate probit model with selection. Again, the intuition is the same as discussed in the previous section, if the correlation between the structural and the selection model errors is not equal to 0 then there exists selection bias, essentially this will be given by ρ.

The coefficient of artrho and wald test in Table 9 signals whether we can reject presence of sample selection. Significant statistics for model 1 and 2 suggests correlation between the error term of the structural and the selection model. Overall the estimated coefficients for women and men from the SC/ST again have a similar signs as the probit model with selection bias except the magnitude differs. However, presence of selection bias is rejected for model 3.

Concluding the findings of both wage differentiation and occupational choice models, despite the fact that the increase in wages for attaining higher education overtime is more for the group GM, the change in probability of engaging in high skilled occupations for disadvantaged groups is greater. Even if the returns are not as attractive for non GM group it seems that over time there is convergence in occupational choices compared with GM. These changes can be potentially be related to the decline in talent misallocation.

Table 9: Occupational Choice Model Estimates [Probit]

Variables	Pr(High Skilled Occupations)=1					
	Model 1		Model 2 (DD)		Model 3 (DDD)	
	Probit	Selection Corrected Probit	Probit	Selection Corrected Probit	Probit	Selection Corrected Probit
	(1)	(2)	(3)	(4)	(5)	(6)
Group [Base=General: Male]						
SC/ST:Male	0.014***	0.0139***	0.010***	0.0103***	.008***	.0065**
	(0.0023)	(0.0024)	(0.0023)	(0.0024)	(.0025)	(0.0025)
General:Female	0.011***	0.034***	0.015***	0.024***	.0195***	-0.005
	(0.0028)	(0.009)	(0.0028)	(0.0046)	(.0029)	(0.0089)
SC/ST:Female	0.005	0.026***	0.014***	0.023***	.039***	0.02**
	(0.0035)	(0.0085)	(0.0039)	(0.0052)	(.0056)	(0.0089)
Years of Education [Base=Illiterate]						
Graduation and above	0.533***	0.545***	0.533***	0.559***	.532***	0.536***
	(0.0064)	(0.0074)	(0.0069)	(0.0073)	(.0074)	(0.0107)
Time [Base: 1983]						
2012	-0.186***	-0.189***	-0.180***	-0.181***	-0.184***	-0.208***
	(0.0046)	(0.0053)	(0.0046)	(0.0049)	(.0043)	(0.006)
Interaction: Education * Time						
Graduation and above*1983					.0314***	.0344***
					(.0042)	(.0045)
Graduation and above*2012					.00423***	.00393***
					(.00089)	(.0009)
Interaction: Group * Education						
SC/ST:Male* Graduation			.0085***	.0086***	.00098	.00091
			(.0017)	(.0017)	(.0041)	(.0041)
General:Female* Graduation			.0231***	.0276***	.0423***	.0402***
			(.0034)	(.0041)	(.0076)	(.0076)
SC/ST:Female* Graduation			.023***	.027***	7.57e-06	-0.0007
			(.0053)	(.0059)	(.0111)	(.011)
Interaction: Group * Time						
SC/ST:Male*1983					-.000014**	-.000013**
					(5.80e-06)	(5.57e-06)
SC/ST:Male*2012					-3.31e-06	-3.18e-06
					(2.44e-06)	(2.29e-06)
General:Female*1983					-.000033***	-.000032***
					(7.59e-06)	(7.34e-06)
General:Female*2012					-.000011***	-.0000104***
					(3.79e-06)	(3.34e-06)
SC/ST:Female*1983					-.000032***	-.0000312***
					(7.50e-06)	(7.24e-06)
SC/ST:Female*2012					-.000011***	-0.000011***
					(3.73e-06)	(3.68e-06)
Interaction: Group * Education * Time						
SC/ST:Male* Graduation*2012					.00188***	.00169***
					(.00047)	(.00045)
General:Female* Graduation*2012					.00402***	.0022*
					(.00099)	(.0012)
SC/ST:Female* Graduation*2012					.00551***	.0037**
					(.0016)	(.0016)
athrho		-0.150***		-0.072***		0.108
		(0.053)		(0.024)		(0.075)
Wald test of indep eqn(rho=0):chi2(1)		7.5***		8.65***		2.06
		(Pr=0.006)		(Pr=0.003)		(Pr=0.015)
Constant	YES	YES	YES	YES	YES	YES
Other Control Variables	YES	YES	YES	YES	YES	YES
Observations	166,059	217,569	166,059	217,569	166,059	217,569

Note: The table records average marginal effects; where robust standard errors are in parenthesis. *** p<0.01, ** p<0.05, * p<0.1

A5. Three Fold Blinder-Oaxaca Decomposition

The three fold non linear B-0 decomposition for our model can be written as:

$$\bar{Y}_{i,gm} - \bar{Y}_{i,g} = \Big\{ \sum_{i,gm=1}^{N} \frac{F(\widehat{\beta}^{gm} X_{i,gm})}{N^{gm}} - \sum_{i,g=1}^{N} \frac{F(\widehat{\beta}^{g} X_{i,g})}{N^{g}} \Big\} \quad , or \tag{A4}$$

$$\begin{aligned}
\bar{Y}_{i,gm} - \bar{Y}_{i,g} = & \Big\{ \sum_{i,gm=1}^{N} \frac{F(\widehat{\beta}^{g} X_{i,gm})}{N^{gm}} - \sum_{i,g=1}^{N} \frac{F(\widehat{\beta}^{g} X_{i,g})}{N^{g}} \Big\} \\
& + \Big\{ \sum_{i,g=1}^{N} \frac{F(\widehat{\beta}^{gm} X_{i,g})}{N^{g}} - \sum_{i,g=1}^{N} \frac{F(\widehat{\beta}^{g} X_{i,g})}{N^{g}} \Big\} \\
& + \Big\{ \Big[\sum_{i,gm=1}^{N} \frac{F(\widehat{\beta}^{gm} X_{i,gm})}{N^{gm}} - \sum_{i,g=1}^{N} \frac{F(\widehat{\beta}^{gm} X_{i,g})}{N^{g}} \Big] - \Big[\sum_{i,gm=1}^{N} \frac{F(\widehat{\beta}^{g} X_{i,gm})}{N^{gm}} - \sum_{i,g=1}^{N} \frac{F(\widehat{\beta}^{g} X_{i,g})}{N^{g}} \Big] \Big\}
\end{aligned} \tag{A5}$$

The first term in equation A5 is attributed to the explained or variables which are observable. The second component is the difference attributed by variables not taken into account. The last term in the curly brackets is the interaction effect which is the difference between explained component referenced at coefficients of GM with the explained component referenced at non GM coefficients. The aim to use B-0 decomposition methodology is to identify the share of each barrier in the total decline in talent misallocation which we observe over time. For this purposes in the main text of the paper I focus on the results of the explained component of the decomposition from the two-fold decomposition.