Caste Gender and Occupational Outcomes

Borooah, Vani

University of Ulster

July 2019

Chapter 4
Caste, Gender, and Occupational Outcomes

4.1 Introduction

An important concern of public policy in India is to ensure that all persons, regardless of gender, caste, or religion, are treated fairly in the jobs market. There are two aspects to this concern. The first is whether differences in remuneration between persons fully reflect their difference in productivities or whether these differences might, wholly or in part, be due to "earning discrimination". The second concerns inter-group differences in the likelihoods of, *ceteris paribus,* attaining different levels of occupational success. The question here is whether these differences in likelihoods are justified by differences in the distribution of *employee attributes* or whether they are, wholly or in part, due to *employer bias* or "occupational discrimination". Schmidt and Strauss (1975), Macpherson and Hirsch (1995), and Borooah (2001a) are examples of studies into this question. This chapter is concerned with the second issue — "occupational discrimination" — in the context of the Indian labour market.

In response to the burden of social stigma and economic backwardness borne by persons belonging to some of India's castes — the Scheduled Castes (SC) and the Other Backward Classes (OBC) — as well as the isolation from the modern world and from mainstream society of India's Scheduled Tribes (ST), the Constitution of India allows for special provisions for members of these groups.[1] One of these provisions is compensatory discrimination in favour of persons from the ST, SC, and the OBC. Compensatory discrimination has taken the form of guaranteeing seats in national and state legislatures and in village *panchayats*, places in educational institutions, and the reservation of a certain proportion of government jobs for the ST, the SC, and the OBC. Hereafter, the groups which are protected in terms of reserved places in educational institutions are referred to as "reserved categories" while those that are not offered such protection are referred to as "general categories". It is important to emphasise that the Constitution restricts SC status to *Hindu* groups in "unclean"

[1] For the history and evolution of caste-based preferential policies in India see Osborne (2001).

occupations: their non-Hindu equivalents are not accorded this status and, therefore, cannot benefit from reservation policies.[2]

In the mind of the Indian public, however, it is jobs reservation that is seen as the most important of the public concessions towards the reserved categories and it is the one which arouses the strongest of passions.[3] On the one hand, there is the demand to extend the boundaries of the reserved category to include persons who are not currently beneficiaries of reservation. [4] On the other hand, there is the demand to expand the scope of reservation by extending it to private sector jobs.[5] Given that issues relating to occupational discrimination and the alleged unfair treatment of people belonging to certain castes and religions dominate public debate and discourse in India, it is surprising how little academic research there is on this subject (see, however, Dhesi and Singh, 1989 Esteve-Bolart, 2004; Borooah *et al*., 2007; Thorat and Attewell, 2007; Ito, 2009). Are certain groups treated "unfairly" treated in the jobs market in India? And, if so, is it possible to quantify the extent of unfair treatment?

This chapter attempts to answer these questions using unit record data from the Indian Human Development Survey relating to the period 2011–12 (hereafter, IHDS-2011).[6] This is a nationally representative, multi-topic panel survey of 42,152 households in 384 districts, 1,420 villages and 1,042 urban neighbourhoods across India. Each household in the IHDS-2011 was the subject of two

---

[2] For example, converts to Islam from Hindu "unclean occupations": *halalkhors*, *helas*, *lalbegis*, *dhobis*, *hajjams*, *chiks*, *faqirs*. However, extensions were made to the reservations list for *Mazhabi Sikhs* (in 1956) and *neo-Buddhists* (in 1990).

[3] In arriving at a judgement about who should be eligible for reservation, the criterion has been a person's caste rather than his/her income or wealth. Consequently, groups belonging to what Article 115 of the Indian Constitution calls "socially and educationally backward classes" have benefited from reservation even though, in practice, many persons belonging to these classes could not be regarded as "socially and educationally backward"; at the same time, many persons belonging to non-backward classes could legitimately be regarded as "socially and educationally backward". Compounding this anomaly, many of the benefits of reservation have been captured by well-off groups from the depressed classes (for example, *chamars*) while poorer groups (for example, *bhangis*) have failed to benefit. Unfortunately, it is not possible to address this issue in this study since the data do not allow a breakdown of the SC by sub-caste.

[4] Article 340 of the Indian Constitution empowers the government to create such classes and in 1955, following the report of the "Kalelkar" Commission, 2,339 groups were designated as belonging to the OBC. The 1980 report of the "Mandal" Commission recommended that, in addition to the 23% of government jobs reserved for the SC and ST, a *further* 27% be reserved for the OBC. In 1990, V.P. Singh's government announced plans to implement this recommendation, triggering a wave of "anti-Mandal" rioting in India. In 1992, in *Sawhney v The Union of India*, India's Supreme Court upheld jobs reservation for the OBC but ruled that: (i) reservation was not to extend to more than 50% of the population and (ii) that groups within the OBC category who were manifestly not disadvantaged (the "creamy layer") were to be excluded from reservation.

[5] See Bhambri (2005); Thimmaiah (2005); Thorat *et. al.* (2005), Thorat *et al.* (2016).

[6] Desai *et al.* (2015).

hour-long interviews. These interviews covered *inter alia* issues of: health, education, employment, economic status, marriage, fertility, gender relations, and social capital. The IHDS-2011, like its predecessors for 2005 and 1994, was designed to complement existing Indian surveys by bringing together a wide range of topics in a single survey. This breadth permits analyses of associations across a range of social and economic conditions.

Of particular interest to this chapter is that the IHDS-2011 provides details about the occupations of approximately 62,500 persons by placing them in one or more of 99 occupations. These 99 occupational categories were aggregated in this chapter to form the following six broad occupational categories (with figures in parentheses representing the proportion of the total sample in each category):

1. Professional and Executive (6%)
2. Clerical (5%)
3. Sales and service (8%)
4. Agricultural labour and other farm (32%)[7]
5. Construction (31%)
6. Other non-farm (18%)

Using these data, this chapter focuses on men and women who were between the ages of 21 and 60 (hereafter, aged 21–60 years) and estimates, using the methods of multinomial logit, the probabilities of them being in these occupational categories, after controlling for their gender/caste/religion and their employment-related attributes. Of particular interest is the issue of differences between men and women, and differences between persons belonging to different social groups, in their likelihood of being in the different employment categories.

Of the approximately 38,000 men and 16,500 women in the sample, aged 21–60 years, nearly one-third reported being in more than one occupation; consequently, in order to avoid ambiguity, this chapter focuses on the 25,941 men and the 11,252 women who reported being in a single occupation. Data on these men and women were used, in conjunction with the methodology detailed below, to

[7] Of the persons in this category, 95% were agricultural labourers with the remainder engaged in "other farm" activities.

decompose the observed difference between the groups, in their average proportions in the different occupations, into an "employer bias" and an "employee attributes" effect.

4.2 Analysing Disproportionality in Occupational Outcomes

The social groups which, in terms of their members' probabilities of being in various occupations, are the focus of this study were distinguished as follows: Scheduled Tribes (ST); Scheduled Castes (SC); Non-Muslim Other Backward Classes (OBC-NM); Muslims; and Forward Castes (FC).[8] The composition of the 25,941 men and 11,252 women between the ages of 21–60 years who reported a single occupation, are shown in terms of their social group in, respectively, Figures 4.1 and 4.2. Hereafter, single-occupation men and women, between the ages of 21–60 years, are referred to simply as "men" or "women".

<Figures 4.1 and 4.2>

Table 4.1 shows the proportionate presence of 25,763 men (upper panel) and 11,219 women (lower panel) in each of six occupations: (i) Professional and Executive (men: 8.7%; women: 10.2%); (ii) Clerical (men: 9%; women: 4.2%); (iii) Sales and service (men: 10.8%; women: 10.1%); (iv) Agricultural labour and other farm activity (men: 19.1%; women: 42.9%); (v) Construction (men: 24.3%; women: 21.2%); (vi) Other non-farm (men: 28.2%; women: 11.3%).

<Table 4.1>

Underlying these overall percentages, however, there is considerable inter-group disparity. For example, 18% of FC men — in contrast to 3% of ST, 5.6% of SC, 7.3% of OBC-NM, and 7.5% of Muslim men — were in professional and executive (P&E) jobs. At the other end of the scale, only 10.9% and 13.3% of FC men worked as, respectively, agricultural and construction workers in contrast to 23.1% and 29.4% of SC men. The numbers for women tell a similar story: 28.9% of FC women — in contrast to 3.4% of ST, 5.6% of SC, 8.3% of OBC-NM, and 11.2% of Muslim women — were in P&E jobs. At the other end of the scale, only 22.8% and 17.2% of FC women worked as, respectively, agricultural and construction workers, in contrast to the 48.1% and 23.6% of SC women.

[8] 94% of persons in the FC category were Hindu, 4% were Christian, and 2% were Sikh.

Table 4.1 shows that agriculture and construction collectively provided livelihoods to 67.7% of ST men and to 85.5% of ST women, to 52.5% SC men and 71.7% of SC women. By contrast, 49.6% of FC men, and 51.4% of FC women, worked in the "white collar" occupations of P&E, clerical, and sales and service.

<Table 4.2>

Table 4.2 provides an alternative perspective to the relationship between occupational outcomes and social group. This shows that the FC supplied 41.8% of men and 42.1% of women working in P&E occupations and 40.4% of men and 34.7% of women in clerical jobs; by contrast, the SC supplied 15.5% of men and 16.3% of women in P&E occupations and 19.7% of men and 22.4% of women in clerical occupations. At the other end of the scale, Table 4.2 shows that the SC supplied 29.3% of men and 33.3% of women in agriculture and 29.4% of men and 33.1% of women in construction. In contrast, the FC only supplied 11.5% of men and 7.9% of women in agriculture and 11.1% of men and 12.1% of women in construction.

Since, as Figure 4.1 shows, men from the FC and the SC comprised, respectively, 24.3% and 20.3% of the sample of men (aged 21–60) working in a single occupation and, as Figure 4.2 shows, women from the FC and the SC comprised, respectively, 29.7% and 14.9% of the sample of women (aged 21–60) working in a single occupation, there would appear to be fairly strong disproportionality between the presence of men and women from these groups in the overall sample and their presence in specific occupations.

Disproportionality between groups' representation in the population and in a particular outcome implies "access inequality". So, on this definition, there was, in the context of both men and women, access inequality to P&E jobs since men and women from the FC had disproportionately greater access to such jobs than did their counterparts from the other groups. When there are many groups, however, the relevant question is how to merge these group disproportionalities into a *single* measure of *access inequality*. As discussed in chapter 2, one way of measuring inequality in a variable is by *the natural logarithm of the ratio of the arithmetic mean of the variable to its geometric mean.* [9]

[9] See Theil (1967) and Bourguignon (1979).

This idea translates very naturally, from its usual application to income inequality, to measuring the degree of inequality in occupational outcomes by which people in different population groups meet with different degrees of success in securing a job in a "desirable" occupation — in this case, P&E. The variable of interest is the proportion of persons from each group that are in P&E jobs (the *access rate*) and it is inequality in the distribution of this rate between the groups that is sought to be measured.

Suppose that the sample is divided into M mutually exclusive and collectively exhaustive groups with N_m ($m=1…M$) persons in each group such that N_m and H_m are the numbers of persons from *each* group in, respectively the population and in P&E jobs. Then $N = \sum_{m=1}^{M} N_m$ and $H = \sum_{m=1}^{M} H_m$ are, respectively, the total numbers of persons in the population and in P&E jobs.

The success rate of group m (denoted e_m) is $e_m = H_m / N_m,\ 0 \le e_m \le 1$. Then the arithmetic and geometric means of e_m are, respectively:

$$\overline{e} = \sum_{m=1}^{M} e_m n_m \ \text{ and } \hat{e} = \prod_{m=1}^{M} (e_m)^{n_m} \ \textit{where}\ n_m = N_m / N,\ \sum_{m=1}^{M} n_m = 1 \qquad (4.1)$$

so that the measure of access inequality is:

$$J = \log(\overline{e} / \hat{e}) = \log(\overline{e}) - \sum_{m=1}^{M} n_m \log(e_m) \qquad (4.2)$$

Now from the definition of e_m:

$$e_m = H_m / N_m = \left(H_m / N_m\right)\left(N / H\right)\left(H / N\right) = \ (H_m / H)(N / N_m)(H / N) = \left(h_m / n_m\right)\overline{e} \qquad (4.3)$$

where: $h_m = H_m / H$ *and* $n_m = N_m / N$ are, respectively, group *m's* share of P&E jobs and of the population. Employing equation (4.3) in equation (4.2) yields:

$$J = \log(\overline{e} / \hat{e}) = \log(\overline{e}) - \sum_{m=1}^{M} n_m \log(e_m) = \log(\overline{e}) - \sum_{m=1}^{M} n_m \log\left[\frac{h_m}{n_m}\overline{e}\right] = -\sum_{m=1}^{M} n_m \log\left[\frac{h_m}{n_m}\right] \qquad (4.4)$$

From equation (4.4), inequality is minimised when *J=0*. This occurs when $n_m = h_m$, that is, when each group's share in the "population" *(n_m)* is equal to its share in P&E jobs (h_m). Otherwise, *J>0*. Inequality is at a maximum when one group (say group 1) has complete access to P&E jobs — all the P&E jobs are filled by persons from that group — with access denied to the other groups ($h_1 = 1,\ h_2 = h_3 ... = h_m = 0$). Then $J_{\max} = -n_1 \log(1/n_1) = n_1 \log(n_1)$ and, therefore, $0 \le J \le n_1 \log(n_1)$

Using the numbers shown in Table 4.2, under the last column (for the n_m of equation (4.4)) and under the column headed "Professional and Executive" (for the h_m of equation (4.4)), the computed value of *J* was 13.2 for men and 21.9 for women. To put these results in perspective, an earlier study (Borooah, 2001b) computed the value of *J* in equation (4.4) for employment inequality in Northern Ireland in the years when the Catholic share in employment fell well short of its share in the labour force. This shortfall, in turn, generated debate about labour market discrimination and spawned equal opportunities legislation that utterly transformed Northern Ireland's labour market. These results, with a highest *J* value of 6.89 in 1992, show that, even in those dark days, disproportionality in Northern Ireland's labour market was much less than that which existed in India in 2012 with respect to P&E jobs.

4.3 A Multinomial Logit Model of Occupational Outcomes

Suppose that there are *J* mutually exclusive possible occupational outcomes, indexed *j=1…J,* for each individual *i*, indexed *i=1…N*, in a sample of persons. Suppose that Y_i is a variable taking integer values such that outcome *j* occurs for individual *i* if and only if $Y_i = j$. If outcome *J* is taken as the base outcome then suppose for each individual (*i=1…N*) the logarithm of the odds ratio of outcome *j (j=1…J)* to the base outcome, *J*, can be written as a linear function of *K* determining variables (indexed, *k=1…K*) with X_{ik} representing the value of variable *k* for individual *i.* Then:

$$\log\left(\frac{p_{ij}}{p_{iJ}}\right) = \sum_{k=1}^{K} \beta_{jk} X_{ik} = Z_{ij},\ j = 1...J-1 \qquad (4.5)$$

where: $p_{ij} = \Pr(Y_i = j)$, $\sum_{j=1}^{N} p_{ij} = 1$ and β_{jk} are the coefficients associated with j^{th} outcome for the k^{th} determining variable, with by definition, $\beta_{Jk} = 0$ $(k = 1...K)$. The assumption is that these coefficients do not vary across the individuals in the sample.

The interpretation of the coefficients β_{jk} in equation (4.5) is that a positive/negative coefficient implies that the odds ratio $\left(\frac{p_{ij}}{p_{iJ}}\right)$ rises/falls with an increase in the value of the k^{th} variable. However, the signs of the coefficients β_{jk} are uninformative about the direction of travel of the underlying probabilities, p_{ij}, which are the real items of interest: $\beta_{jk} > 0$ implies that $\left(\frac{p_{ij}}{p_{iJ}}\right)$ rises with an increase in the value of the k^{th} variable but, if the odds ratio for some other outcome rises even faster, rise in the odds ratio implied by $\beta_{jk} > 0$ may be accompanied by a fall in the value of p_{ij}.

In order to obtain the underlying probabilities p_{ij} the estimated coefficients need to be employed in solving the equation (derived from equation (4.5)):

$$\Pr(Y_i = j) = p_{ij} = \frac{\exp(Z_{ij})}{[1 + \sum_{s=1}^{J} Z_{is}]} = \frac{\exp\left(\sum_{k=1}^{K} \beta_{jk} X_{ik}\right)}{1 + \exp\left(\sum_{s=1}^{J}\sum_{k=1}^{K} \beta_{sk} X_{ik}\right)} \tag{4.6}$$

In addition to the social group of persons, discussed in the previous section, it was hypothesised that their probabilities of being in particular occupations — the occupations as set out in Tables 4.1 and 4.2 with the non-farm occupation being the baseline occupation, *J* — would *inter alia* depend upon:

1. The person's gender.
2. The person's highest educational level. Five levels of education were distinguished: (i) none; (ii) up to primary; (iii) above primary and up to secondary; (iv) higher secondary; (v) graduate and above.
3. Fluency in English: (i) none; (ii) little fluency; (iii) fluent.

4. The location of the person in terms of: (i) metropolitan; (ii) non-metropolitan urban; (iii) developed village; (iv) less developed village.
5. The age of the person organised in four age bands: 21–30, 31–40, 41–50, and 51–60 years.
6. The state in which they resided: Andhra Pradesh, Assam, Bihar, Chhattisgarh, Delhi, Haryana, Himachal Pradesh, Gujarat, Jammu & Kashmir, Jharkhand, Karnataka, Kerala, Madhya Pradesh, Maharashtra, Orissa, Punjab (including Chandigarh), Rajasthan, Tamil Nadu, Uttarakhand, Uttar Pradesh, and West Bengal.

Consequently, the equation that was estimated, in the context of the multinomial form of equation (4.5), was:

$$\log\left(\frac{\Pr(Y_i = j)}{\Pr(Y_i = J)}\right) = \log\left(\frac{p_{ij}}{p_{iJ}}\right) = \\ \boldsymbol{\alpha}_{jk} \times \mathbf{GEN}_i + \boldsymbol{\beta}_{jk} \times \mathbf{SOCGROUP}_i + \boldsymbol{\gamma}_{jk} \times \mathbf{EDUCATION}_i + \boldsymbol{\delta}_{jk} \times \mathbf{ENGLISH}_i \\ + \boldsymbol{\theta}_{jk} \times \mathbf{AGEBAND}_i + \boldsymbol{\psi}_{jk} \times \mathbf{LOCATION}_i + \boldsymbol{\zeta}_{jk} \times \mathbf{STATE}_i \qquad (4.7)$$

In equation (4.7), **GEN** is a variable such that GEN_i=1, for men and GEN_i=*2* for women with coefficient vector $\boldsymbol{\alpha}_{jk}$; **SOCGROUP** is a variable whose components are the five social groups (ST, SC, OBC-NM, Muslim, and FC) with associated coefficient vector $\boldsymbol{\beta}_{jk}$; **EDUCATION** is a variable whose components are the five education levels with associated coefficient vector $\boldsymbol{\gamma}_{jk}$; **ENGLISH** is a variable whose components are the three fluency levels with associated coefficient vector $\boldsymbol{\delta}_{jk}$; **AGEBAND** is a variable whose components are the age bands with associated coefficient vector $\boldsymbol{\theta}_{jk}$; **LOCATION** is a variable whose components are the four locations with associated coefficient vector $\boldsymbol{\psi}_{jk}$; and **STATE** is a variable whose components are the Indian states (listed above) with associated coefficient vector $\boldsymbol{\xi}_{jk}$.

Equation (4.7) was estimated on data for 36,943 persons, aged 21–60 years, all of whom were working in a single occupation, of whom 25,730 were men and 11,213 were women. The estimates from equation (4.7) were then used, via equation (4.6), to compute the probabilities of (single-occupation) persons in the sample of being in the different occupations; the results in this chapter are presented in terms of these probabilities rather than in terms of the underlying coefficients of equation

(4.5). This chapter focuses on the set of jobs which populate the P&E, clerical, and sales and service occupations since these may be legitimately regarded as the most desirable of jobs.

The coefficient vector $\boldsymbol{\beta}_{jk}$ in equation (4.7) may be interpreted as *employer bias*, for or against members of group *k*, with respect to occupation *j* because, even if the values of the other (non-social group) variables were the same between the individuals, differences in the component values of these vectors would imply that persons' probabilities of being in occupation *j* would be different simply by virtue of their belonging to different groups.

4.4 The Predicted and Synthetic Probabilities of Being in Professional and Executive Jobs

Differences in the observed proportions of persons in different groups belonging to a particular occupation could arise through employer bias so that, for example, inter-group differences in the proportions of persons in P&E jobs were, in part, the result of employers being biased against persons belonging to certain groups and in favour of persons belonging to other groups. They could, however, also be the result of the average level of attributes differing between persons from the various groups. For example, as Table 4.3 shows, compared to the nearly 26% of men and women from the FC who were graduates, only 7.8% of SC men and 3.2% of SC women, and only 7.4% of Muslim men and 5.5% of Muslim women were educated to this level. Inter-group differences in the proportions of persons in P&E jobs could, therefore, be due to differences in attributes between such persons.

<Tables 4.3 and 4.4>

So, the fact that, as Table 4.1 shows, 18% of FC men and 28.9% of FC women, compared to only 5.6% of SC men and women, were in P&E jobs may have been partly due to employer bias but it may have been partly the result of the educational qualifications of men and women from the FC being, on average, superior to those from the SC. Consequently, the observed outcome with respect to inter-group differences in the proportions in P&E jobs could be regarded as the outcome of the combined working of employer bias and employee attributes.

A major purpose of this chapter is to disentangle the effects of employer bias and employee attributes on the observed proportions of persons, belonging to different social groups, being in P&E

jobs. These observed proportions are referred to as the average predicted probabilities of being in P&E jobs because if the multinomial logit model was used to *predict* for each of the *N* persons in the sample the likelihood of being in a P&E job (denoted $\hat{p}_i \;\; i = 1...N$) then the average of these $\hat{p}_i$, computed over every subgroup, would equal the *observed* proportion of persons from that subgroup in a P&E job. This is because the multinomial logit model has the property of passing through the mean. So, $\hat{p}^{FC}, \hat{p}^{MU}, \hat{p}^{OBC}, \hat{p}^{SC}, \hat{p}^{ST}$, the *average predicted probabilities* from the multinomial logit model of, respectively, FC, Muslim, OBC-NM, SC, and ST persons being in P&E jobs would be the same as the *observed proportion* of persons from these groups being in such jobs. In contrast to the average predicted probabilities are average *synthetic* probabilities of being in P&E jobs — denoted $\tilde{p}^{FC}, \tilde{p}^{MU}, \tilde{p}^{OBC}, \tilde{p}^{SC}, \tilde{p}^{ST}$ for persons from the five groups, where these synthetic probabilities were computed on the basis of simulations based on the method of *recycled proportions* described in the previous chapter and summarised below.

In order to compute the synthetic probability of persons from the SC being in P&E jobs, assume that all the 36,943 persons in the sample were SC or, in other words, apply the "SC component" of the vector $\boldsymbol{\beta}_{jk}$ in equation (4.7) to every person in the sample. Then, holding the values of the other variables constant (either to their observed sample values, as in this chapter, or to their mean values over the estimation sample), compute the average probability of being in a P&E job and denote it $\tilde{p}^{SC}$. Similarly, in order to compute the synthetic probability of persons from the FC being in P&E jobs, assume that all the 36,943 persons in the sample were FC or, in other words, apply the "FC component" of the vector $\boldsymbol{\beta}_{jk}$ in equation (4.7) to every person in the sample. Then, holding the values of the other variables constant (either to their observed sample values, as in this chapter, or to their mean values over the estimation sample), compute the average probability of being in a P&E job and denote it $\tilde{p}^{FC}$.

Since the values of the non-social group variables (education, fluency in English, location, age, and state of residence) were unchanged between these two (all-SC and all-FC) hypothetical scenarios, the only difference between the two synthetic probabilities, $\tilde{p}^{SC}$ and $\tilde{p}^{FC}$ was that the first

probability was the result of applying "SC coefficients" to the entire sample while the second probability was the result of applying "FC coefficients" to the entire sample. Consequently, the difference between the two synthetic probabilities, $\tilde{p}^{SC}$ and $\tilde{p}^{FC}$, was *entirely* due to differences in caste because all other differences between persons from the SC and FC had been neutralised by assigning them the attributes of the entire sample.

In essence, therefore, in evaluating the effect of two characteristics *X* and *Y* on the likelihood of a particular outcome, the method of "recycled proportions" compares two probabilities: first, under an "all have the characteristic *X*" scenario, and then under an "all have the characteristic *Y*" scenario, *with the values of the other variables unchanged between the scenarios.* The difference between the two synthetic probabilities is then entirely due to the effect of the different attributes represented by *X* and *Y* (in this case, differences in caste between the SC and FC).[10]

4.5 Results from the Multinomial Logit

The average synthetic probabilities (hereafter, SP) of being in P&E jobs , in clerical jobs, and in sales/service jobs were computed using the estimates from equation (4.7) and the method of recycled proportions, described above, and are shown in, respectively, Table 4.5, 4.6, and 4.7, with respect to every determining variable in equation (4.7).[11] This follows from the advice by Long and Freese (2014) that it is more meaningful to present the results from the estimated equation in the form of the synthetic probabilities from the estimated multinomial logit coefficients rather than in terms of the estimates themselves. As discussed earlier, the multinomial logit estimates of equation (4.5) themselves do not have a natural interpretation — they exist mainly as a basis for computing more meaningful statistics and, in this case, these are the synthetic probabilities (SP) computed using equation (4.5) in conjunction with equation (4.6).

<Tables 4.5, 4.6, and 4.7>

The columns headed "Synthetic Probability" in Tables 4.5–4.7 show the SP associated with the various categories of variables. So, in the social group category, Table 4.5 shows that the synthetic

[10] STATA's margin command performs these calculations.
[11] The equations were estimated using the *svy* command in STATA or, in other words, by grossing up the sample observations using weights in IHDS-2011 contained in its FWT variable.

probability of persons aged 21–60 being in a P&E job was 7.4% for the ST, 8.5% for the SC, and 10.5% for the FC. Similarly, Table 4.6 shows that the synthetic probability of persons aged 21–60 being in clerical jobs was 6.2% for the ST, 7.7% for the SC, and 9% for the FC. Lastly, Table 4.7 shows that the synthetic probability of persons aged 21–60 being in sales/service jobs was 10.5% for the ST, 11.2% for the SC, and 12.5% for the FC.

The columns in Tables 4.5, 4.6, and 4.7 headed "Marginal Probability" represent, for the social group category, the *difference* between the synthetic probability of the individuals in the first four social groups and those in the reference group, FC, denoted by [R]. Table 4.5 shows that, for P&E jobs, the marginal probability for persons from the SC was -2 points (=8.5-10.5) points. Similarly, Table 4.6 shows that, for clerical jobs, the marginal probability for Muslims was -3.8 points (=5.2-9) points. Lastly, Table 4.7 shows that, for sales/service, the marginal probability for the OBC-NM was -1.9 points (=10.6-12.5) points. Dividing these marginal probabilities by their standard errors yielded the t-values. These showed whether these marginal probabilities were significantly different from zero in the sense that the likelihood of observing these values under the null hypothesis of no difference was less than 5% (superscript ** in Tables 4.5–4.7) or 10% (superscript * in Tables 4.5–4.7).

4.5.1 *Gender*

In terms of gender, the important finding was that for P&E jobs, the SP for women (Table 4.5: 16.6%) exceeded that for men (Table 4.5: 7.6%) by 9 points and this difference was significantly different from zero. This result needs to be understood in the context of the cultural norms and attitudes in India towards working women. As the *Economist* newspaper reported:

> Patriarchal social mores [in India] supersede economic opportunity in a way more associated with Middle Eastern countries. Outside a small urban elite, the default position is for a woman not to work unless there is no other way for a family to make ends meet. This reflects an enduring stigma of women being seen as "having" to work. A family's social standing

partly derives from women being able to stay at home. Such restrictions become more rigid higher up the caste hierarchy.[12]

In consequence, female employment rates in India, at 26%, are very low compared to the 50% of women who work worldwide.[13] This is reflected in the fact that only 11,213 (or 30%) of the estimation sample of 36,943 persons aged 21–60 years and reporting a single occupation, were women. It is reasonable to infer that many of the 1,261 of these 11,213 women who held P&E jobs would not have worked had they not been in employment which they deemed suitable. It is likely that they were highly motivated, had the support of their families, and that employers were *a priori* well-disposed towards employing them.

This favourable employer bias towards employing women was also evident with regard to sales/service where the SP for women (Table 4.7: 12.9%) was also greater than that for men (Table 4.7: 9.8%) by 3 points, with this difference being significantly different from zero. There was, however, no gender-based employer bias with regard to clerical jobs for which the SP for men and women (Table 4.6: respectively, 7.6% and 7.4%) were virtually indistinguishable.

4.5.2 *Social Groups*

In terms of the social groups, an important finding of Table 4.5 was that the synthetic probability of being in a P&E job was highest for Muslims and persons from the FC (Table 4.5: respectively, 10.8% and 10.5%) — though these were not significantly different — and lowest for persons from the ST (Table 4.5: 7.4%), the SC (Table 4.5: 8.5%), and the OBC-NM (Table 4.5: 8.5%) and these were all significantly lower than the SP for Muslims and FC persons.

For clerical jobs, the synthetic probability was lowest for Muslims (Table 4.6: 5.2%) and highest for persons from the FC (Table 4.6: 9%). Sandwiched in between were, as Table 4.6 shows, the synthetic probabilities of clerical jobs of 6.2% for the ST, 7.7% for the SC, and 7.3% for the OBC-

[12] *The Economist*, "A Job of Her Own: Culture and the Labour Market Keep India's Women at Home", *Briefing: Indian Women*, 5 July 2018, https://www.economist.com/briefing/2018/07/05/culture-and-the-labour-market-keep-indias-women-at-home
[13] *The Economist*, *op. cit*.

NM; all these synthetic probabilities were significantly lower than that for persons from the FC being in a clerical job.

For sales/service jobs, too, the SP was lowest for Muslims (Table 4.7: 8.7%) and highest for persons from the FC (Table 4.7: 12.5%) with the difference between the two groups being significantly different from zero. Sandwiched in between were the synthetic probabilities of sales/service of 10.5% for the ST, 11.2% for the SC, and 10.6% for the OBC-NM, but it was only for persons from the OBC-NM that the synthetic probability of being in sales/service jobs was *significantly* lower than for persons from the FC.

Summarising these findings, there was for all three types of jobs — P&E, clerical, and sales/service — a favourable employer bias towards persons from the FC. For P&E jobs, employers were most biased against persons from the ST but did not display any antipathy towards Muslims. For clerical and sales/service jobs, however, employer bias was strongest against Muslims.

4.5.3 *Education*

In a manner analogous to the synthetic probabilities for the social groups, one can also construct synthetic probabilities for the categories of education by considering, in succession, scenarios in which *all* the persons in the estimation sample were assigned to a particular educational category. The synthetic probability of having a P&E job rose with the level of education, with graduates having the highest synthetic probability of being in a P&E job (Table 4.5: 39.8%) followed by those with higher secondary qualifications (Table 4.5: 18.9%), with those without any education having virtually no chance of being in a P&E occupation (Table 4.5: 0.7%). Moreover, the synthetic probability of having a P&E job as a graduate was significantly greater than the synthetic probabilities associated with lower educational levels.

Table 4.6 shows that having a graduate degree (average SP: 18.7%) had less impact on being in clerical jobs than it did on P&E jobs. Nonetheless, as for P&E jobs, the SP of having a clerical job rose with the level of education, with graduates having the highest chance of success (Table 4.6: 18.7%), followed by those with higher secondary qualifications (Table 4.6: 13.5%). Moreover, as with

P&E jobs, the SP of having a clerical job as a graduate was significantly greater than the SP associated with lower educational levels.

For sales and service occupations, as Table 4.7 shows, it was persons with higher secondary qualifications (Table 4.7: 13.1%) that had the highest SP of holding such jobs, followed by those with secondary level qualifications (Table 4.7: 12.7%), followed by those with either no education or primary education (Table 4.7: respectively, 11% and 11.8%), with graduates bringing up the rear (Table 4.7: 10.9%). There was, however, no significant difference between these probabilities and it is therefore safe to conclude that differences in education levels between persons did not exert significant influence on differences between them on their SP of being in sales/service occupations.

4.5.4 *Fluency in English*

It is reasonable to suppose that fluency in English (no fluency; some fluency; fluent) would play an important part in obtaining jobs P&E and clerical jobs but, perhaps, be less important in sales/service occupations. Analogous to the simulations carried out for social groups, one could compute the synthetic probabilities of being in P&E, in clerical, and in sales and service occupations under the successive assumptions that the entire sample was: (i) fluent in English; (ii) had some fluency in English; (iii) had no fluency in English. Since the only difference between the simulations was fluency in English, differences between the three probabilities would be *entirely* due to differences in English fluency.

Table 4.5 shows that the SP of being in a P&E occupation for persons who were fluent in English was 17.7% and this was significantly higher than the 9.8% SP for those who had some fluency which, in turn, was significantly higher than the 6% SP for those who had no fluency. The results in Table 4.6 show also that the SP of being in clerical occupations for persons fluent in English was, at 11.6%, significantly higher than the 9.9% SP for those who had some fluency. In turn, the SP of being in clerical occupations for persons with some English fluency was significantly higher than the 6.1% SP for those who had no fluency. Table 4.7 shows that English language fluency did not have any significant effect on the SP of being in sales/service occupations.

4.5.5 *Location and Age Effects*

Table 4.7 shows that the average SP of being in a clerical job or in a sales/service occupation was significantly higher in a metropolitan area than in a non-metropolitan urban location which, in turn, was significantly higher than that in villages (both more developed and less developed). Lastly, the average synthetic probability of being in a clerical job or in a sales/service occupation was significantly higher in a more developed village than in a less developed village.

Tables 4.5–4.7 show a strong association between the SP of being in P&E, and clerical occupations. For the first two occupations, the association was positive, with the SP of being in a P&E and in clerical occupations being significantly higher in the 51–60 age band than in the 21–30 age band (Table 4.5: respectively, 12% and 7.7% for P&E, and Table 4.6: respectively, 11.4% and 6.1% for clerical). For the sales/service occupations, there was no significant association between changes in age and changes in the SP of being in these occupations.

In terms of age, the SP of being in P&E and in clerical jobs rose steadily with age: for P&E occupations from 7.7% for persons aged 21-30 to 12% for those aged 51-60 (Table 4.5) and, for clerical occupations, from 6.1% for persons aged 21-30 to 11.4% for those aged 51-60 (Table 4.6). For sales and service jobs, however, the SP fell steadily with age: from 11.1% for persons aged 21-30 to 9.9% for those aged 51-60 (Table 4.7).

4.6 Decomposing Employer Bias and Employee Attributes

Notwithstanding employer bias in favour of women for P&E jobs (Table 4.5: 9.8% and 12.9% for, respectively, men and women), the proportion of women in the estimation sample who held such jobs was, at 10.2%, only 1.5 points higher than the corresponding proportion of 8.7% for men. So, the observed difference between women and men in their proportions in P&E jobs ($\hat{p}_W - \hat{p}_M = 1.5$) was smaller than the difference between their synthetic probabilities ($\tilde{p}_W - \tilde{p}_M = 3.1$). To understand this result requires referring to the decomposition set out in the previous chapter and reproduced below in the context of gender:

$$\overbrace{\hat{p}_W - \hat{p}_M}^{Z} = \overbrace{(\tilde{p}_W - \tilde{p}_M)}^{A} + \left[\overbrace{(\hat{p}_W - \tilde{p}_W)}^{B} - \overbrace{(\hat{p}_M - \tilde{p}_M)}^{C} \right] \quad (4.8)$$

The terms *Z* and *A* in equation (4.8) represent, respectively, the difference between women and men in their *predicted* (observed proportions) probabilities (*Z)* and in their *synthetic* probabilities (*A*) of being in P&E jobs where, as discussed earlier, the term *A* represents the difference which is due *solely* to differences in gender. Consequently, it would be legitimate to regard the term *A* as resulting from employer bias.[14]

The situation with respect to gender and P&E jobs is that *Z=1.5>0 A=3.1*>0 but *Z<A*. This is because *B-C=-1.6.*[15] The fact that *B-C<0* implies that, on average, women's employment attributes, with respect to P&E jobs, were not as good as those of men and this diluted employer bias in their favour. For example, as Table 4.3 shows, 12.2% of men, compared to 7.7% of women, were graduates.

<Table 4.8>

Table 4.8 attaches numbers to the components of equation (4.8) by showing, with respect to P&E and clerical occupations, the differences between the observed proportions and the synthetic probabilities for persons from the five social groups: ST, SC, OBC-NM, Muslims, and FC. For P&E occupations, the gaps in the estimation sample proportions of persons from the FC and persons from other groups were 17.4, 15.5, 13, and 12.3 points for, respectively, the ST, SC, OBC-NM, and Muslims. The gaps in the synthetic probabilities of being in P&E jobs between persons from the FC and persons from other groups were much smaller: 3.1, 2, 2, and -0.3 points for, respectively, the ST, SC, OBC-NM, and Muslims. Since the gaps in synthetic probabilities are interpreted as discrimination, the largest amount of discrimination in P&E jobs was faced by persons from the ST: 18% (=3.1/17.4) of the gap between ST and FC persons in their sample proportion in P&E occupations was due to discrimination; on the other hand, Muslims did not face any discrimination in P&E jobs.

[14] See the previous chapter for a more detailed discussion.
[15] *B=10.2-12.9=-2.7* and *C=8.7-9.8=-1.1* so that *B-C=-2.7+1.1=-1.6.*

Table 4.8 shows that for clerical jobs, the gaps in the estimation sample proportions of persons from the FC and persons from other groups were smaller than for P&E jobs: 13, 10, 9.3, and 11.8 points for, respectively, the ST, SC, OBC-NM, and Muslims. The gaps in the synthetic probabilities of being in clerical jobs between persons from the FC and persons from other groups were also much smaller: 2.8, 1.3, and 1.7 points for, respectively, persons from the ST, SC, and the OBC-NM. For Muslims, however, the gap in the synthetic probabilities of being in clerical jobs between FC persons and them was 3.8 points. Consequently, Muslims faced the largest amount of discrimination in clerical jobs jobs: 32% (=3.8/11.8) of the gap between Muslims and FC persons in their sample proportion in clerical occupations could be blamed on discrimination. On the other hand, persons from the SC, perhaps aided by jobs reservation, faced much less discrimination: 13% (=1.3/10) of the gap between SC and FC persons in their sample proportion in clerical occupations could be attributed to discrimination.

4.7 Inequality Decomposition

The econometric analysis of P&E occupations, encapsulated in Table 4.5, highlighted several factors which affected the likelihood of persons being in such jobs. Of these, three in particular stood out: the social group, fluency in English, and level of education. This section examines, using the tools of inequality decomposition, the relative contribution of these three factors to inter-personal inequality in the predicted likelihood of being in a P&E occupation.

The estimated multinomial logit equations (equation (4.8)) could be used to predict for each of the 36,943 persons in the estimation sample, the probability of their being in a P&E occupation. Armed with a knowledge of these individual probabilities, it is possible to estimate how much of the *overall inequality* in these 36,943 probabilities, or some subset thereof, could be explained by a particular factor.

This section provides an answer to this question, using the methodology of *inequality decomposition*. This method decomposes overall inequality into "between-group" and "within-group"' inequality. When the decomposition is *additive*, overall inequality can be written as the *sum* of within-group and between-group inequality:

$$\underbrace{I}_{\text{overall ineqality}} = \underbrace{A}_{\text{within group inequality}} + \underbrace{B}_{\text{between group inequality}} \tag{4.9}$$

When inequality is additively decomposed, one can say that the basis on which the individuals were subdivided (say, household wealth) contributed [(B/I)×100] % to overall inequality, the remaining inequality, [(A/I)×100] %, being due to inequality *within* the groups. So, inequality decomposition provides a way of analysing the extent to which inter-personal inequality (in this case, in the probabilities of being in P&E jobs) is "explained" by a factor or a set of factors. If, indeed, inequality can be "additively decomposed", then, as Cowell and Jenkins (1995) have shown, the proportionate contribution of the between-group component (**B**) to overall inequality is the income inequality literature's analogue of the R^2 statistic used in regression analysis: the size of this contribution is a measure of the amount of inequality that can be "explained" by the factor (or factors) used to subdivide the sample.

One of the most popular ways of measuring inequality is by the *Gini coefficient* which is computed as follows. If N is the number of persons, p_i is the (predicted) probability of person i being in a P&E occupation and $\overline{p}$ is the mean probability, computed over the N persons, the Gini coefficient is defined as:

$$G = \frac{1}{2N^2\overline{p}} \sum_{i=1}^{N} \sum_{j=1}^{N} | p_i - p_j | \tag{4.10}$$

In other words, the Gini coefficient is computed as half the mean of the difference in probabilities between pairs of respondents, divided by the average probability ($\overline{p}$). The Gini coefficient associated with the distribution of the probabilities of being in a P&E job of the 4,016 persons who were graduates was 0.201 ($\overline{p}$ =0.526) while the Gini coefficient associated with the distribution of the probabilities of being in a P&E job of the 2,016 persons who were both graduates *and* were fluent in English was 0.133 ($\overline{p}$ =0.625).

In other words, the distribution of probabilities of being in a P&E occupation was compressed as the qualifications of persons were tightened from being graduates to being graduates who were fluent in English. In terms of equation (4.10), if two graduates were chosen at random, the mean

difference in their probabilities of being in a P&E occupation would be 0.402×0.526=0.21 points. On the other hand, if two graduates , who were also fluent in English, were chosen at random, the mean difference in their probabilities of being in a P&E occupation would be 0.266×0.625=0.17 points.

The Gini coefficient, however, is not additively decomposable in terms of equation (4.9). In order to decompose inequality additively it has to be measured in a very specific way. Only inequality indices which belong to the family of *Generalised Entropy Indices* are additively decomposable (Shorrocks, 1980) and one of these indices is Theil's (1967) Mean Logarithmic Deviation (MLD) Index which is used for the analysis in this section. The MLD index is defined over N persons as $\left(\sum_{i=1}^{N}\log(p_i / \overline{p})\right) / N$ where p_i is the probability of person i $(i=1…N)$ being in a P&E occupation and $\overline{p} = \sum p_i \big/ N$ is the mean probability.

If one considered the 4,016 persons in the estimation sample *who were graduates* and separated them into their social groups (ST, SC, OBC-NM, Muslims, and FC) then the values of *I, A*, and *B* of equation (4.9) would be, respectively: 0.072, 0.067, and 0.005. In other words, the division of the sample by social group "explains" only 7% of overall inequality in the probabilities of graduates being in P&E occupations. Similarly, if one considered the 2,016 persons in the estimation sample *who were both graduates and fluent in English* and separated them into their social groups (ST, SC, OBC-NM, Muslims, and FC) then the values of *I, A*, and *B* of equation (4.9) would be, respectively: 0.028, 0.027, and 0.001. In other words, the division of the sample by social group "explains" only 3.6% of overall inequality in the probabilities of graduates with English fluency being in P&E occupations. In summary, therefore, in terms of "explaining" the unequal chances that graduates faced in securing jobs in P&E occupations, their membership of one or the other of the five delineated social groups was not a major factor.

4.8 Conclusions

The issue of under-attainment by certain social groups has concerned Indian society and a measure of this concern is the special provisions in the Indian Constitution for members of the Scheduled Tribes,

the Scheduled Castes, and the Other Backward Classes. These provisions are mainly in the form of reserved seats in the national parliament, state legislatures, municipality boards, and village councils (*panchayats*); reserved places in public higher educational institutions; and job reservations in the public sector.

Notwithstanding these concerns, questions about the relation between the caste and religion of persons and their occupational attainment in India have not, as yet, elicited a considered answer. This chapter represented a first attempt at answering these questions within a systematic theoretical framework. This framework comprised four elements: the construction of an overall measure of disproportionality between representation in the population and in "desirable" occupations; a multinomial logit model for determining the likelihood of occupational outcomes in terms of synthetic probabilities; a methodology for decomposing the overall disparity between groups in their representation in "desirable" occupations into one part due to "employer bias" and another part due to "employee attributes"; and, lastly, a methodology for decomposing overall inequality in the individual probabilities of attaining a particular occupational status into a within-group and a between-group effect.

Unit record data from the IHDS-2011, encompassing nearly 37,000 persons aged 21–60 years, helped put empirical flesh on this framework. This marriage of data and methodology meant that an area of India's economic life which, hitherto, had been in shadow could now be illuminated. The results showed that employer bias (or discrimination) did not explain more than 18% of the gap in observed proportions in P&E jobs between persons from the FC and persons from other groups. On the other hand, the substantial part of this observed difference could be explained by differences in employee attributes between persons from the FC and persons from other groups. The overall conclusion of this study is that the occupational attainment of persons in India is largely a matter of *what* they are, though, undeniably, there is a part which is determined by *who* they are. Unfortunately, the sole thrust of social policy in India in the area of employment has been to focus on the identity — and specifically, the caste identity — of persons rather than on their attainments and, alarmingly, to seek an extension of identity-based employment to private sector jobs.

On the basis of the analysis contained in this and the previous chapter, this is barking up the wrong tree with, arguably, damaging consequences for India's labour market. There are three reasons for disquiet about jobs reservation policy. First, a pernicious effect of discrimination against members of a group is that under a *negative stereotype*, a prior, negative belief about a group's average value in terms of of some relevant characteristic is used to assess the ability of all individuals from this group. Consequently, discrimination based on prior beliefs about the average performance of a group penalises talented individuals from a group by ascribing to them the average quality of their group.

This can then change the behaviour of a group. If the employer is going to associate a negative stereotype with a group then none of it members will see any value to making human capital investments (education, study, diligence, work habits, attitudes to work) since such investments will be crushed on the anvil of discrimination. Jeffery and Jeffery (1997) in their study of Muslims in Bijnor, argued that many Muslims regarded their relative economic weakness as stemming from their being excluded from jobs due to discriminatory practices in hiring. The belief that their sons would not get jobs then led Muslim parents to devalue the importance of education as an instrument of upward economic mobility. It was with such considerations in mind that Myrdal (1944) spoke of the "vicious circles of cumulative causation": the failure of discriminated groups to make progress justifies the prejudicial attitudes of dominant groups. As Elmslie and Sedo (1996, p. 474) observe in their development of this argument, "One initial bout of unemployment that is not productivity based can lay the foundation for continued future unemployment and persistently lower job status even if no future discrimination occurs".

This argument, made in the context of being discriminated against, could also apply to *positive* discrimination (affirmative action) which, in India, is implemented by reserving a certain proportion of jobs for persons from "reserved categories". The awareness by members of a group that, provided they meet the most minimal of qualifications, a certain proportion of jobs are reserved for them and that their employment is guaranteed could act as a disincentive to invest in themselves with a view to improving their employability. There is no point, under *any* kind of discrimination, in improving one's capacity as an employee: when one is discriminated against, employment is more likely than not to be denied; under positive discrimination, with jobs reservation, employment is all

but assured. Under both kinds of discrimination, investment in oneself is a waste of resources and effort.

The second pernicious effect of positive discrimination is the shadow of stigma which falls on anyone who benefits from such discrimination. Heilman *et al.* (1992) showed that affirmative action stigmatises its beneficiaries who are held to be incompetent. A subsequent study (Heilman *et al.,* 1997) provided further support for the idea that there is a stigma of incompetence associated with affirmative action. They also indicate that people tend to discount qualifications as the basis for the hiring of those associated with affirmative action and suggest that this discounting process influences competency evaluations. Leslie *et al.* (2014) also concluded that affirmative actions, while designed to facilitate workplace success for members of the groups they target (e.g., women, ethnic minorities), may have the ironic effect of stigmatising their beneficiaries and, in turn, decreasing their performance outcomes. As Riley (2012) writes:

> Liberal supporters of affirmative action like to pretend that there is no shame in being hired to meet a racial or ethnic quota and not for your job skills alone, or in being admitted to a college with SAT scores well below those of your white and Asian peers. But the reality is that nobody who has any pride wants to be that "diversity" hire in the office or that token minority on campus, especially if it allows others to dismiss your success as having resulted from a tilted playing field.

The third pernicious effect of jobs reservation relates to inefficiency — vacancies are divided into two groups: those which are to be filled by persons from the general category — that is, anyone can fill them — and those which can only be filled by persons from the reserved categories. Indeed, the Government of India explicitly demands that "if the required number of SC/ST/OBC candidates is not available, the vacancies which could not be filled up shall remain unfilled until the next recruitment year. These vacancies will be treated as 'backlog vacancies'".[16] Consequently, the possession of a caste certificate is more valuable than a degree certificate: vacancies could remain unfilled because of a dearth of suitably qualified *reserved category* candidates in spite of there being

[16] See p. 46 of: http://persmin.gov.in/DOPT/Brochure_Reservation_SCSTBackward/Ch-06_2014.pdf (accessed 20 October 2018).

an abundance of *general category* candidates who could fill these positions. For all these reasons, Indian policy makers — goaded by a group frenzy which sees reservation as the gateway to prosperity — have an infinite capacity to inflict harm on Indian society through a well-meaning but misguided policy of offering quotas to its "backward classes".