Caste, Ethnicity and Poverty in Rural India

Ira N. Gang
Department of Economics, Rutgers University
gang@economics.rutgers.edu

Kunal Sen
School of Development Studies, University of East Anglia
k.sen@uae.ac.uk

Myeong-Su Yun
Department of Economics, Tulane University
msyun@tulane.edu

Revised February 2004

ABSTRACT

This paper analyzes the determinants of rural poverty in India, contrasting the situation of scheduled caste (SC) and scheduled tribe (ST) households with the non-scheduled population. The incidence of poverty in SC and ST households is significantly higher than among non-scheduled households. Using a probit decomposition analysis, we break down the difference in poverty incidence between the scheduled castes and tribes and the non-scheduled households into the proportion explained by the differences in characteristics and the proportion explained by the differences in probit coefficients. We find that for SC households, differences in characteristics explain the gap in poverty rates more than differences in coefficients; while for ST households, it is the converse. It is striking that differences in educational attainment explain approximately one quarter of the poverty rate gap for both social groups. Occupational structure is very influential in determining the poverty rate gap for both SC and ST, as are differences in returns to individual occupations.

Keywords: poverty, caste, ethnicity, probit decomposition
JEL Classifications: I32, O12, J15

This manuscript has benefitted from discussions with Shubhashis Gangopadhyay and Ajit Mishra, and from comments by seminar participants at University College – Dublin, Queens University – Belfast, Bar-Ilan University, Indian Statistical Institute-Calcutta Centre, Delhi School of Economics, University of East Anglia, and the Midwest Economic Association. We thank Richard Palmer-Jones for helping us with the data.

Correspondence: Myeong-Su Yun, Department of Economics, Tulane University, 206 Tilton Hall, New Orleans, LA 70118, USA; phone (+1 504) 862-8356; fax (+1 504) 865-5869; email: msyun@tulane.edu

1. Introduction

Since obtaining independence in 1947, Indian governments have been deeply concerned with widespread poverty and have implemented various anti-poverty schemes. However, rural poverty remains persistent, with the headcount ratio being 42.7 percent in 1993/94 (Dubey and Gangopadhyay, 1998). More troubling is the concentration of rural poverty in India in the 'scheduled caste' (SC) and 'scheduled tribe' (ST) population.[1] The presence of such disparity in the incidence of poverty and wide-spread discrimination against the scheduled groups have long histories in India. Affirmative action programs have been at the core of Indian social policy directed toward the scheduled groups.

According to the 1991 Census of India, scheduled castes and tribes comprise 16.5 percent and 8.1 percent, respectively, of India's population, yet 43.5 percent of India's rural poor are concentrated in these groups.[2] Poverty rates among scheduled caste and tribe households are significantly higher than the rest of the population – in 1993/1994, the proportion of rural SC and ST households below the poverty line were 49.0 and 49.5 percent respectively, as compared to a poverty rate of 32.8 percent for rural non-scheduled households. From Table 1 we see a gap in the

---

[1] The Indian Constitution specifies the list of castes and tribes included in these two categories, and accords the 'scheduled castes' and 'scheduled tribes' special treatment in terms of affirmative action quotas in state and central legislatures, the civil service and government-sponsored educational institutions (Revankar, 1971). The 'scheduled castes' correspond to the castes at the bottom of the hierarchical order of the Indian caste system and were subject to social exclusiveness in the form of 'untouchability' at Indian Independence (August 15, 1947), while the 'scheduled tribes' correspond to the indigenous tribal population mainly residing in the northern Indian states of Bihar, Gujarat, Maharashtra, Madhya Pradesh, Orissa, Rajasthan, West Bengal and in North-eastern India.

[2] These estimates are from the unit record data provided in the National Sample Survey's 50th round of the consumer expenditure survey. More details of the computations are provided in the next section.

proportion living in poverty (a poverty rate gap) of 16.2 percent (= 49.0 - 32.8) between SC and non-scheduled households, and a poverty rate gap of 16.7 percent (= 49.5 - 32.8) between ST and non-scheduled households. One major task in the fight to reduce rural poverty is to close the gap in poverty rates between scheduled castes and tribes and the non-scheduled group.

In spite of the existence of large gaps in poverty rates between scheduled and non-scheduled groups, very few studies have specifically and systematically investigated the causes of the disparity.[3] Just attributing the disparity in poverty rates to discrimination may be easy and natural, but it does not provide much insight into how to reduce poverty. In order to provide guidelines for reducing the differences in the incidence of poverty between scheduled and non-scheduled groups, we need a more detailed and systematic study which asks whether differences in the amounts of schooling, occupational choice or demographic characteristics hold the key to understanding the poverty rate gap, and whether the poverty mitigating strength of household or individual characteristics (e.g., education and occupation) are different. In order to answer these questions, this paper offers and implements a methodology that allows us to examine causes of the disparity in poverty incidence between scheduled and non-scheduled groups.

First, we explore the determinants of poverty for scheduled households, SC and ST, and non-scheduled households, using rural household survey data on 64,287 households from the 50th round of the National Sample Survey (NSS). This is done by estimating probit equations where the

---

[3] Three other studies examining the economic status of the scheduled castes and tribes are Deshpande (2000, 2001) and Meenakshi and Ray (2002). The first two studies compute measures of disparity between the scheduled castes and tribes and the non-scheduled households (in the first study, only for the state of Kerala), while the third paper studies poverty by constructing headcount ratios for scheduled caste and tribe households under alternative assumptions of economies of scale in consumption and equivalence scales. These studies do not examine the determinants of SC and ST poverty relative to non-scheduled households.

dependent variable has a value of one when a household is in poverty. Second, based on probit estimates for scheduled and non-scheduled groups, we are able to construct a decomposition equation that explains why poverty is much more prevalent among the scheduled casts and tribes than among the non-scheduled households. We decompose the differences in the poverty rates into the proportion explained by the differences in characteristics (characteristics effect) and the proportion explained by the differences in the probit coefficients (coefficients effect). The characteristics effect captures the amount of the poverty rate gap caused by the differences in attributes. Though the differences in characteristics are supposed to reflect differences in abilities or preferences between scheduled and non-scheduled groups, we suspect that the disparity in attributes might result from widespread discrimination against the scheduled groups in terms of educational opportunity and occupational choice. The coefficients effect captures the amount of the poverty rate gap caused by the differences in the effectiveness of the characteristics in reducing poverty between the comparison groups.

In the next section we discuss who are the poor among the scheduled castes, scheduled tribes and the non-scheduled group by studying the mean characteristics of each group. Section 3 investigates why they are poor by utilizing a probit analysis of the relative influence of various economic and non-economic variables on poverty. Section 4 employs a probit decomposition analysis to examine and explain the poverty rate gaps between scheduled and non-scheduled households. Finally, Section 5 provides a summary of our study and its main conclusions.

2. Data and Descriptive Statistics

For our analysis we use the 50th round of India's National Sample Survey (NSS) on consumer expenditure in rural areas collected in 25 states and 7 Union Territories. The survey period was from July 1993 to June 1994. The NSS data is a cross-section of a geographically distributed random sample of households across India. In addition to information on household consumer expenditure and demographic behavior, the NSS contains detailed questions on other household characteristics such as the educational level and occupation of the head of the household.

In this paper, we focus on rural poverty primarily for two reasons. Firstly, the majority of India's poor live in rural areas. Secondly, the NSS classifies a household as SC or ST if it is so indicated by the head of the household at the time of the survey. Such sorting criteria as indicators of a household's social status will be weaker in urban areas where intermingling or intermarriage between SC and ST individuals and non-scheduled individuals may occur. Since the NSS provides expenditure data by household, our estimates of poverty are at the level of the household, not at the level of the individual.[4]

We estimate the incidence of rural poverty across all three social groups, and relate these to the demographic, educational and occupational characteristics of these three groups. Poverty rates at the official state-specific poverty lines (taken from Dubey and Gangopadhyay, 1998) were obtained by enumerating the number of households with monthly per capita total expenditure below the poverty line.[5] We restrict our sample to households where the age of the head of the household is between 20 and 70 years. Since the information on age, education and occupation that we obtain from the NSS surveys are for the head of the household, the range 20-70 assumes an age structure

[4] This distinction may become important if there are significant differences in the intra-household consumption of food and other necessities across the SC, ST and non-scheduled households.

[5] These poverty lines are loosely based on a concept of minimum food (especially calorie) expenditure plus additional necessary expenditures. Households are classified as poor if they did not purchase at least 2400 calories per capita.

where the head of the household is expected to be economically active.

The poverty rates by social group and by age, household size, educational level and occupation are presented in Table 1.[6] Firstly, we observe that there is a non-linear relationship between age and poverty rates across all three social groups, with the poverty rate first increasing as we move from the age group 20-29 to 30-39, and then decreasing for ages 40 years and above. Secondly, poverty rates seem to increase with household size, with the highest poverty rates observed among households that have seven or more members. Thirdly, while literacy is negatively related to the incidence of poverty, the negative correlation between educational attainment and poverty incidence seems to be weaker for SC households as compared to ST and non-scheduled households. Approximately 23 percent of SC households with literacy levels of higher secondary and above are poor as compared to 14.9 percent of similarly educated ST households and 9.9 percent of non-scheduled households. Finally, there is a higher incidence of poverty among agricultural laborers across all three social groups as compared to other occupations, and in the case of ST households, for those households self-employed in agriculture.

Table 2 shows the mean characteristics of the sample households in our study. Considering the demographic characteristics of the three groups of households first, we find that SC and ST households have a lower mean age for the head of the household as compared to non-scheduled households. SC and ST households are also smaller in size than non-scheduled households -- the mean household size for SC and ST households are 4.7 and 4.8 respectively, as compared to a mean household size of 5.0 for non-scheduled households.

---

[6] Our poverty our estimates are weighted by the multiplier associated with each household. The NSS supplied multiplier for each household indicates the total number of households in the population represented by the sampled household.

A much higher proportion of SC and ST households are not literate (68 percent and 70 percent, respectively), as compared to non-scheduled households (48 percent of whom are not literate). With respect to occupation, 11 percent of SC households are self-employed in non-agriculture, 53 percent as agricultural laborers, 11 percent as non-agricultural laborers, 19 percent are self-employed in agriculture while 6 percent are classified in a residual category termed 'others'. For ST households, 6 percent are self-employed in non-agriculture, 39 percent are agricultural laborers, 10 percent are non-agricultural laborers, 39 percent are self-employed in agriculture while 6 percent are in other occupations. Finally, for non-scheduled households, 15 percent are self-employed in non-agriculture, 24 percent are agricultural laborers, 7 percent are non-agricultural laborers, 44 percent are self-employed in agriculture while 10 percent are in other occupations. Thus, a greater proportion of SC households are agricultural laborers as compared to ST and non-scheduled households. On the other hand, a greater proportion of ST households are employed as non-agricultural laborers, as compared to SC and non-scheduled households.

Although interesting, Table 2 is only suggestive as the observed bivariate connections may be caused by other variables. In order to keep other things constant, one must carry out a multivariate analysis of the factors determining poverty status. This we do in the section below.

3. Probit Analysis of the Incidence of Poverty

In this section, we employ probit analysis to help us understand why some households are in poverty.[7] The estimates from the probit analysis will be used for studying why households in some social groups, scheduled castes (SC) and scheduled tribes (ST), are more likely to be below the

[7] See Dreze and Srinivasan (1997) for a similar analysis of poverty incidence among widow-headed households in India.

poverty line than households in the non-scheduled population.

The focus of our analysis is on education and occupation. To capture the effect of education on the probability of a household being in poverty, we use dummy variables corresponding to the highest educational level completed by the head of the household. Thus, we include dummy variables corresponding to 'literate, below primary level', 'literate, below secondary level', 'literate up to secondary level', and 'literate, higher secondary and above' (the reference group in our case is households where the head of the household is not literate). With respect to occupation, we include dummy variables corresponding to four occupational groups -- self-employed in non-agriculture, self-employed in agriculture, agricultural labor and non-agricultural labor (with the reference group being the occupational category termed 'others' by the NSS).

In addition to the explanatory variables capturing occupation and educational levels, we include in our analysis a number of background and demographic variables. First, we include the generational impact reflected by the age of the person. We use two variables: age (number of years), and age-squared (number of years of age-squared divided by 100), to reflect the non-linear effects of age on poverty. Second, we incorporate the effect of household size on the probability of the household being in poverty, as previous studies have noted a negative relationship between per capita expenditures and the size of the household (Krishnaji, 1981, 1984). Given the possible presence of economies of scale in household consumption, we include household size squared as an additional control variable.[8]

---

[8] We do not include the child-adult ratio which is often used a control for household composition as inter-group poverty comparisons using NSS data are quite robust to different assumptions about equivalence scales (Dreze and Srinivasan, 1997, Meenakshi and Ray, 2002). When we include the child-adult ratio as an additional explanatory variable the results were broadly similar to the ones reported. Another variable of potential interest is land assets. While we have data on these, four of the five occupational groups – self-employed- non-agriculture, agricultural laborers,

Finally, we include controls for the location of the household. There are large differences in rural poverty rates within Indian states, with states in North-western India (Haryana, Punjab) along with the state of Kerala having lower poverty rates than the national average (Datt and Ravallion, 1998). In contrast, the poverty rates in Assam, Bihar and Orissa are significantly higher than the national average. The omission of state dummies to capture the location of the household may bias the results if the SC and ST households are mostly residing in the states where higher poverty is observed, and if this higher incidence of poverty is due to state-level factors exogenous to the household such as agro-climactic factors or the nature of state-level public policies toward poorer households. We present our results with and without the inclusion of state dummies.

The estimated probit coefficients are reported in Table 3, with columns (1), (3) and (5) containing the results for SC, ST and non-scheduled households without the inclusion of state dummies. Columns (2), (4) and (6) contain the results with the state dummies. The dependent variable is a binary variable with a value of one if the household is in poverty. The results are not sensitive to the inclusion of state dummies. The reported coefficients for each of the independent variables are broadly similar across all three social groups, though likelihood ratio tests (not reported) show that the coefficients of the probits for each group are significantly different from the other groups.

The estimated coefficients show that greater educational attainment is associated with a statistically significant reduction in the probability of being poor, with everything else held constant. This is true for all three groups of households. However, higher educational attainment from the

---

non-agricultural laborers, ‘others’ – own little or no land. The only occupational group with some land-holding are self-employed agriculturalists (farmers). Less than 9 % of poor households in the three groups are female-headed.

secondary level up seems to lead to a greater decline in the incidence of poverty among ST and non-scheduled households when compared to SC households.

We now turn our attention to occupation and its role in poverty status. The results reported in Table 3 are consistent and robust. Compared to the occupational category 'others', all other occupational categories lead to a higher incidence of poverty among all three social groups (except in the case of non-scheduled households who are self-employed in agriculture when state dummies are included). SC households who are agricultural laborers are more likely to be in poverty compared to agricultural laborer households in the other two social groups. In the case of ST households, those who are self-employed in agriculture are more likely to be in poverty than SC and non-scheduled households in the same occupational group. Overall, the results suggest that households who are laborers, whether involved in agricultural or non-agricultural work, are more likely to be in poverty when compared to households who are self-employed. With respect to demographic factors, higher aged the heads of households are associated with lower poverty. However, this relationship is non-linear, with further increases in age leading to less than proportionate decreases in poverty. A non-linear relationship is also found between poverty and household size, with poverty more evident in larger sized households.

To summarize, the probit results imply that households that are larger in size, where the head of the household is not literate, is an agricultural laborer, and is younger in age, are more likely to be in poverty. We also find that the effects of explanatory variables on the likelihood of a household being poor vary over social groups.

4. Accounting for Differences in Poverty Rates

In this section, we seek to explain why poverty is much more prevalent among the scheduled caste and tribe households, than among non-scheduled households. For the scheduled castes in comparison to the non-scheduled we are seeking to find the sources of a poverty rate gap of 16.2 (= 49.0 - 32.8); for scheduled tribes versus the non-scheduled the gap is 16.7 (= 49.5 - 32.8). Our analysis breaks down the poverty rate gap into its components.

There are two broad approaches to explaining the gap in poverty incidence: the characteristics effect and the coefficients effect. The characteristics effect relies on the possibility that the characteristics of individuals which give rise to poverty differ among groups. For example, one group may have less education than another group, or be in the "bad" jobs. The characteristics effect reflects how differences in the characteristics of individuals among groups affect the likelihood that someone is in poverty.

The coefficients effect relies on the possibility that the effectiveness of individual characteristics, reflected in probit estimates, may vary among the three groups. Therefore, the likelihood of being in poverty differs across groups. For example, education may be less effective in reducing the probability of being poor in scheduled households relative to non-scheduled households. The coefficients effect reflects how differences in the probit coefficients across groups affect the likelihood that someone is in poverty.

Studying characteristics and coefficients effects was formally introduced by Oaxaca (1973). Though the implementation and extensions of the Oaxaca decomposition have generally been in the context of wage differentials (in general, any continuous variable), recent extensions allow for discrete dependant variables (e.g., Even and Macpherson, 1993; Yun, 2004). Until recent innovations, decomposing when one had discrete dependant variables (e.g., employment status) was generally accomplished by so-called "simulation" (see Abowd and Killingsworth, 1984). In these analyses, logits or probits would be estimated for each group, and the coefficients for one group (e.g., scheduled caste) would be replaced with those of the other group (e.g., non-scheduled caste) in order to calculate a counter-factual predicted probability. Subtracting this counter-factual prediction from the observed probability for the former group (scheduled caste), one sees the effects of the differences in coefficients between the two groups, holding the characteristics constant. The coefficients can be switched one by one to see contribution of each variable, or completely to see the effect of overall change. However, this simulation method is not only tedious but also problematic since it may be sensitive to the order of switching (see Ham, Svejnar and Terrell, 1998, p. 1137 for a discussion of path-dependency). Also, the simulation method usually only partially explains the changes since it looks only at the effect of switching coefficients, that is, the coefficients effect explained above. The decomposition method employed in this paper can provide a complete picture of the changes.

4.1. Probit Decomposition Methodology

Algebraically, the differences in the average probability of being poor between groups A and B, $(\overline{P}_A - \overline{P}_B)$, where A = scheduled castes or tribes and B = non-scheduled, may be decomposed into two components which represent the characteristics effect and coefficients effect. Asymptotically, this is,[9]

$$\overline{P}_A - \overline{P}_B = \left(\overline{\Phi(\sum_{i=1}^{i=k} X_A^i \beta_B^i)} - \overline{\Phi(\sum_{i=1}^{i=k} X_B^i \beta_B^i)} \right) + \left(\overline{\Phi(\sum_{i=1}^{i=k} X_A^i \beta_A^i)} - \overline{\Phi(\sum_{i=1}^{i=k} X_A^i \beta_B^i)} \right),$$

[9] A decomposition equation with a different parameterization is also possible; our results with it are not substantially different from those presented here. The results of the other version of the decomposition equation are available from the authors upon request.

where Φ is standard normal cumulative distribution function, β_A, and β_B are sets of estimated coefficients for each group, and X_A and X_B are the various explanatory variables used in the probit equations.

The above decomposition provides us with the overall coefficients and characteristics effects. In order to find the relative contribution of each variable to the poverty rate gap, in terms of characteristics and coefficients effects, we employ a decomposition equation for the probit model proposed by Yun (2004);[10]

$$\begin{aligned}\overline{P}_A - \overline{P}_B &= \sum_{i=1}^{i=k} W_{\Delta X}^{i} \left(\overline{\Phi(\sum_{i=1}^{i=k} X_A^i \beta_B^i)} - \overline{\Phi(\sum_{i=1}^{i=k} X_B^i \beta_B^i)} \right) \\ &+ \sum_{i=1}^{i=k} W_{\Delta\beta}^{i} \left(\overline{\Phi(\sum_{i=1}^{i=k} X_A^i \beta_A^i)} - \overline{\Phi(\sum_{i=1}^{i=k} X_A^i \beta_B^i)} \right),\end{aligned}$$

where $W_{\Delta X}^{i} = \dfrac{(\overline{X}_A^i - \overline{X}_B^i)\ \beta_B^i}{\sum_{i=1}^{i=k} (\overline{X}_A^i - \overline{X}_B^i)\ \beta_B^i}$ and $W_{\Delta\beta}^{i} = \dfrac{\overline{X}_A^i\ (\beta_A^i - \beta_B^i)}{\sum_{i=1}^{i=k} \overline{X}_A^i\ (\beta_A^i - \beta_B^i)}$,

where $\overline{X_A^i}$ and $\overline{X_B^i}$ are average values of explanatory variables i for groups A and B, respectively.

---

[10] Because of the non-linearity of the standard normal cumulative distribution function, we find the weights ($W_{\Delta X}^{i}$, $W_{\Delta\beta}^{i}$) using the following approximations: (1) an approximation of the value of the average of the standard normal distribution function, $\overline{\Phi(\sum_{i=1}^{i=k} X^i \beta^i)}$, with that of the standard normal distribution evaluated at the average value of exogenous variables, $\Phi(\sum_{i=1}^{i=k} \overline{X^i} \beta^i)$; (2) a first order Taylor expansion to linearize the characteristics and coefficients effects around $\overline{X}_B \beta_B$ and $\overline{X}_A \beta_A$, respectively (for details, see Yun, 2004).

4.2. Explaining Differences in Poverty Rates

In this section, we discuss empirical findings from the decomposition analysis. We focus on the percentage share which tells us what percentage of the total poverty rate gap is accounted for by that particular element or group of elements. We discuss the overall effects first, and then breakdown the overall effects into successively smaller subgroups. We discuss the poverty rate gap of scheduled castes relative to the non-scheduled in Tables 4 and 6, and that of scheduled tribes relative to the non-scheduled in Tables 5 and 7. In Tables 4 and 5 we find the results of the aggregate breakdown, and of key groups of variables, both when we do not include state dummies and when we do. In Tables 6 and 7 we decompose further into individual variables effects.

We proceed by first discussing the aggregate effects and sub-aggregate effects without state dummies for SC households respectively (Table 4). The *Aggregate Effects* row in the tables show the overall effects of characteristics versus coefficients in explaining differences in poverty. The majority (58 percent) of the difference in poverty incidence between the SC and non-scheduled castes is explained by the differences in the levels of characteristics possessed by the two groups, while 42 percent by the differences in probit coefficients. One could refer to the latter as the difference in effectiveness since it is a summary measure of the differences in the strength of the various individual characteristics influencing poverty.[11] If in both groups the various variables

---

[11] The coefficients effect in the well-known Oaxaca decomposition of the wage gap is conventionally considered discrimination (Oaxaca, 1973). The coefficients effect in our decomposition analysis of the binary choice model may be interpreted as discrimination or behavioral differences, among other interpretations. We interpret the coefficients effect as the differences in the effectiveness of characteristics to reduce poverty. Since our probits estimate the probability of being in poverty, a positive value for the coefficients effect implies that the influence of the poverty reducing power of a characteristic in the scheduled groups is weaker than that of the non-scheduled group. It may be argued that the differences in poverty mitigating power (coefficients effect) results from discrimination. Furthermore, one can argue that a positive value for the characteristics effect may reflect discrimination in a broad sense, if the opportunities for obtaining

influencing poverty status had the same strength (their coefficients in the probit equation had been equal), then 42 percent of the increased probability of being in poverty for SC households would disappear. On the other hand, if both groups had the same characteristics, 58 percent of the poverty rate gap would disappear. When we include state dummies, the aggregate effects do not change much, with the coefficients effect being 43.7 percent and the characteristics effect being 56.3 percent. Thus, the higher poverty rate observed among SC households is not because of their location - that is, it is not because they are disproportionately located in the poorer Indian states.

In Table 4 we also see the breakdown of characteristics and coefficients effects into important variable groupings. First and foremost we see the importance of both the characteristics and the coefficients effects for occupation in determining the poverty rate gap. The former contributes 42.7 percent to the poverty rate gap, while the latter contributes 46.4 percent. One of the salient features of the caste system is the generally undesirable and low-paying jobs scheduled castes are allowed to undertake. This would explain the characteristics effect, as SC households generally are in less-remunerative occupations. If anthropological evidence about the lack of job choice for individuals belonging to scheduled castes is accurate, this may represent pre-market discrimination (Srinivas, 1962; Beteille, 1965).[12] At the same time the coefficients effect tells us that even if the distribution of jobs was the same between SC and non-scheduled households, SC households are being rewarded less than non-scheduled households for the same occupation (controlling for education and demographic characteristics). Or, there is a difference in the strength of the poverty reducing effect

---

human capital (e.g., education) and other characteristics are themselves limited due to discrimination. Attributing differences in poverty rates to discrimination is always a difficult and controversial issue.

[12] Such discrimination may generate an ‘equilibrium trap’ where those who break caste customs suffer economically (Akerlof, 1976).

of occupation for SC and non-scheduled households that leaves more SC households into poverty. Though the sizes of the effects are somewhat less, we also find large characteristics and coefficients effects for the occupational distribution when state dummies are included. Education is remarkable in that both with and without state dummies the coefficients effect is negligible, while the characteristics effect is 21 – 22 percent. SC households simply attain lower levels of schooling, and that puts them at greater risk of being poor.[13]

Age and household size are included as control variables, but the results are interesting in and of themselves. The coefficients effect of age structure (age and age-squared taken together) is negative, while the characteristics effect is positive. For household size we find the characteristics effect is negative, and the coefficients effect is positive. Household size differences reduce the poverty rate gap, but differences in coefficients increase the poverty rate gap.[14] The constant term also contributes to reducing the poverty rate gap. The constant term may reflect underlying differences between the two groups which are not captured by the other explanatory variables.

Up to now we have discussed what accounts for differences in poverty incidences between SC and non-scheduled households. We now turn to a discussion of what explains differences in poverty rates between ST and non-scheduled households, shown in Table 5. Approximately 41 percent of the poverty rate gap is explained by differences in households' characteristics between the two groups. This means that if ST and non-scheduled households had the same characteristics, then

---

[13] This is supported by the finding of Dreze and Kingdon (2001) that SC and ST children are less likely to go to school, even after controlling for household wealth, parental education and motivation, school quality, and related variables.

[14] As seen in Table 2, SC and ST households are smaller in size than non-scheduled households, and the probit analysis suggests that the likelihood of being poor is positively related to household size.

the poverty rate gap would have been 41percent less. Differences in educational attainment accounts for 22.5% of the poverty rate gap. The occupational distribution explains 19.7% of the higher poverty among the ST households as compared to the non-scheduled. These results are basically the same when state dummies are included.

Fifty-nine percent of the poverty rate gap of 16.7 percent between ST and non-scheduled is explained by the differences in probit coefficients between ST and non-scheduled households. If in both groups the various variables influencing poverty status had the same strength (if their coefficients in the probit equation had been equal), then about 59 percent of the increased probability of being in poverty for ST vis-a-vis non-scheduled households would disappear.

Once we break down the aggregate coefficients effect into contributions of individual variables, we find that the coefficients effect of educational attainment is negative. This means that the strength of the ameliorating impact of education on poverty is greater for ST than non-scheduled households.[15] The coefficients effect of the occupational distribution group of variables is large between the ST and non-scheduled, accounting for 53 percent of the difference in the probability of being in poverty. This suggests that for ST households, more than occupational structure, what has contributed to the greater incidence of poverty among such households has been the significantly lower returns they have received for the jobs they hold as compared to non-scheduled households.

With respect to demographic control variables, as in the case of SC households, for ST households the coefficients effect of age structure (age and age-squared taken together) is negative. Thus, the age structure of ST households is better for reducing poverty than that of non-scheduled

[15] This finding should be interpreted with some caution as it only accounts for -1.7 percent of the difference in poverty incidence between ST and non-scheduled, and with state dummies included it becomes slightly positive.

households. This helps to reduce poverty incidence, but clearly not enough to compensate for the rest of the coefficients effects. Household size, both excluding and including state dummies, has a high positive coefficients effect and about -5 percent characteristics effect. Differences in constant terms also contribute to reducing the poverty rate gap.

The inclusion of state dummies provides some interesting results. In the case of SC households, we have seen that with the inclusion of state dummies both characteristics and coefficients effects are more or less the same. However, in the case of ST households, 26% of the difference in poverty rates is explained by differences in characteristics with the inclusion of state dummies, as compared to a 41% characteristics effect when state dummies are not included. These results may imply that ST households are more likely to be located in states where the average poverty rate is low (such as the North-eastern states of India). Thus, for households in this social group, location is working to their advantage in reducing the poverty rate gap.

Tables 6 and 7 present further decompositions, now with more disaggregated effects. In Table 6 we have SC versus non-scheduled, while in Table 7 we have ST versus non-scheduled. For both of these tables we restrict ourselves to the situation where we have excluded the state dummies. We find that the coefficients effect of various degrees of literacy is quite small, while the characteristics effect is substantial. This underlines the importance of obtaining higher level of education for scheduled castes and tribes as it is the gap in educational attainment between the scheduled and non-scheduled groups, and not the differential effectiveness of education in poverty reduction, that is one of the major causes of the poverty rate gap. Since access to education itself might be blocked for the scheduled groups, it is not surprising that once they obtain education our results imply they will do well (assuming that the distribution of ability is the same across groups).

We also find that in the case of SC households, being an agricultural laborer matters both for characteristics and coefficients, and explains much of the effect of occupational structure we observed in Table 4. For ST households, while being an agricultural laborer contributes to the characteristics effect of occupational structure, the coefficients effect of the latter can be largely explained by 'self-employed in agriculture'. This is an interesting finding, as we have already observed in Table 1 that the self-employed in agriculture (i.e., cultivators) have the lowest poverty rate among all occupational types, after the category 'others'. Thus, the coefficients effect of 'self-employed in agriculture' in increasing the poverty rate gap may indicate severe institutional impediments that ST households face in cultivating their land, possibly linked to underlying weaknesses in technology adoption.

5. Summary and Conclusions

This paper has examined the relative significance of some of the key forces that shape the poverty profiles of the scheduled castes (SC), scheduled tribes (ST) and non-scheduled households in India. These profiles vary significantly among these three groups. While there is little difference in the incidence of poverty between SC and ST households, poverty rates of SC and ST households are 16.2 percentage points and 16.7 percentage points higher than non-scheduled households. Our analysis decomposes the poverty rate gap between SC (or ST) and non-scheduled households, into a part explained by differences in attributes of households (characteristics effects) and a part explained by differences in effectiveness of the attributes of households (coefficients effect), using household survey data from the 50th round of the National Sample Survey conducted in 1993-1994.

The decomposition analysis indicates that for SC households differences in characteristics explain the poverty rate gap more than differences in coefficients, with 58 percent of the poverty rate gap attributable to the former. For ST households, however, it is the reverse, with 59 percent of the poverty rate gap attributable to the differences in coefficients. Thus, while there is little difference in the poverty rates between SC and ST households, the causes of the incidence of poverty in these two social groups are not the same. The key difference in the causes of poverty between these two social groups lies in the characteristics effect of occupational structure. Its impact on the poverty rate gap between SC households and non-scheduled households is twice its corresponding impact on the poverty rate gap between ST and non-scheduled households. Further disaggregated analysis suggests that the higher incidence of poverty among SC households is due to the higher proportion of such households who are agricultural laborers, and the lower returns they obtain from agricultural labor as compared to non-scheduled households. In the case of ST households, a large proportion of the poverty rate gap can be explained by the lower returns they receive as cultivators as compared to non-scheduled households, with the occupational structure of such households being relatively less important in explaining the poverty rate gap.

We also see that differences in educational attainments explain about one quarter of the poverty rate gap for both social groups, but that returns to education do not matter in explaining the poverty rate gap. The disaggregated analysis shows that being literate matters for reducing the poverty rate gap; the disparity in educational attainments between the scheduled and the non-scheduled groups is responsible for a substantial gap in the poverty rates. Of course, we cannot rule out that poverty might be responsible for lower education. However, the caste system drives opportunities, and it is more likely that education is determined generally by caste affiliation, with the higher incidence of poverty a product of the lower average educational attainment.

Among demographic factors, we find that household size is working in the favor of SC and ST households, with differences in age structure having little effect on the poverty rate gap. Using state dummies to capture the effect of household location on the poverty rate gap, we find that the large poverty rate gap between SC and non-scheduled households cannot be attributed to where these households are located. On the other hand, for ST households, they seem to be located in the states with lower poverty on average so that differences in characteristics have a smaller effect on the poverty rate gap when state dummies are introduced.

The analysis of this paper has a few important implications for policies that will help combat the significant economic deprivation that is observed amongst SC and ST households as compared to the rest of the population. Firstly, though subsidies to higher education may contribute to India's current surge in high-tech industries, focusing on higher education does not help in mitigating poverty, at least in reducing the poverty rate gaps between scheduled and non-scheduled groups. Therefore, policies should aim to achieve higher educational attainments among SC and ST households; such policies should generally aim to increase the enrollment of SC and ST children, especially focusing on the primary and secondary level rather than on higher education, and preventing the high drop out rate that is evident in the school-age population in India (PROBE, 1999). It is clear from our analysis that greater amounts of literacy even below the primary school level will have a significant impact in reducing the poverty rate gap for both SC and ST households.

Secondly, occupations in rural India are largely socially and culturally determined. Our analysis suggests that in the case of SC households, policies that allow such households greater occupational choice, and in particular, diversification away from agricultural labor may pay significant poverty reduction dividends among these households (policies that encourage self-employment among SC households may be successful in this regard). Further, there is also a need to redress the lower returns that SC households seem to obtain from agricultural labor as compared to the non-scheduled households (possibly via stricter observance of minimum wage legislation).

Thirdly, in the case of ST households, policies need to address the lower returns these households receive from cultivation as compared to the non-scheduled cultivators. This could be in the form of targeted agricultural extension services, input subsidy programs, and directed rural credit programs. Finally, the analysis of the paper suggests that while poverty rates are not very different between SC and ST households, the underlying factors for the higher incidence of poverty in these social groups are to an appreciable extent different, and policy-makers need to be aware of these differences in the causes of poverty incidence when devising policies for poverty alleviation.

REFERENCES

Abowd, J.M. and M.R. Killingsworth (1984), "Do Minority/White Unemployment Differences Really Exist?," *Journal of Business and Economic Statistics*, 2, 1984, 64-72.

Akerlof, G. (1976), "The Economics of Caste and the Rat Race and Other Woeful Tales," *Quarterly Journal of Economics*, 90, 600-617.

Beteille, A. (1965), *Caste, Class and Power: Changing Patterns of Social Stratification in a Tanjore Village*, Berkeley: University of California Press.

Datt, G. and M. Ravallion (1998), "Why Have Some Indian States Done Better Than Others at Reducing Rural Poverty?," *Economica*, 65, 17-38.

Deshpande, A. (2001), "Caste at Birth? Redefining Disparity in India," *Review of Development Economics*, 5, 130-144.

Deshpande, A. (2000), "Does Caste Still Define Disparity? A Look at Inequality in Kerala, India," *American Economic Review*, 90, 322-325.

Dreze, J. and G. Kingdon (2001), "School Participation in Rural India," *Review of Development Economics*, 5, 1-24.

Dreze, J. and P.V. Srinivasan (1997), "Widowhood and poverty in rural India: some inferences from household survey data", *Journal of Development Economics*, 54, 217-234.

Dubey, A. and S. Gangopadhyay (1998), *Counting the Poor: Where are the Poor in India?*, Department of Statistics, Government of India.

Even, W.E. and D.A. Macpherson (1993), "The Decline of Private-Sector Unionism and the Gender Wage Gap," *Journal of Human Resources*, 28, 279-296.

Ham, J.C., J. Svejnar and K. Terrell (1998), "Unemployment and the Social Safety Net During Transitions to a Market Economy: Evidence from the Czech and Slovak Republics," *American Economic Review*, 88, 1117-1142.

Krishnaji, N. (1984), "Family-size, Levels of Living and Differential Mortality in Rural India - Some Paradoxes," *Economic and Political Weekly*, 19, 248-258.

Krishnaji, N. (1981), "On Measuring Incidence of Undernutrition - What Is a Consumer Unit," *Economic and Political Weekly*," 16, 1509-1511.

Meenakshi, J.V. and R. Ray (2002), "Impact of household size and family composition on poverty in rural India", *Journal of Policy Modeling*, 24, 539-559.

Oaxaca, R. (1973), "Male-Female Differentials in Urban Labour Markets", *International Economic Review*, 14, 693-709.

PROBE Team (1999), *The Public Report on Basic Education in India*, Delhi: Oxford University Press.

Revankar, R. (1971), *The Indian Constitution - A Case Study of Backward Classes*, Rutherford: Fairleigh Dickinson University Press.

Srinivas, M.N. (1962), *Caste in Modern India*, Bombay: Asia Publishing House.

Yun, M.-S. (2004), "Decomposition Differences in the First Moment", *Economics Letters*, 82, 273-278.

Table 1
Poverty Rates

	Scheduled Castes	Scheduled Tribes	Non-Scheduled	All
Overall	49	49.5	32.8	38.3
Age				
20-29	45.3	48.6	30.2	36.5
30-39	55.9	56.6	38.2	44.7
40-49	48.9	48.2	32.1	37.6
50-59	43.9	43	29.8	34.2
60-70	44.3	43.4	30	34.1
Household Size				
1	21.5	11.3	14.7	16
2	28.9	24.8	16.7	20.9
3	36	39.1	23.5	28.4
4	49.3	49.5	27.5	35.2
5	54.1	54.5	36.2	42.3
6	59.3	61.6	42.2	48.2
7 or more	65.2	64.3	44	50
Education				
Not Literate	53.6	54	40.6	46.1
Literate, below primary	44.5	43.8	32.5	36.9
Literate, below secondary	38.4	40.6	27.2	30
Literate, secondary	32.8	25.3	16.3	18.7
Literate, higher secondary & above	23.3	14.9	9.9	11.6
Occupation				
Self-employed in non-agriculture	41.9	40.8	30.1	32.9
Self-employed in agriculture	37.4	44.9	26.5	29.9
Agricultural labor	58.6	58.3	50.8	54.7
Non agricultural labor	45	52.6	37.3	41.7
Others	23.3	22.6	19.4	20.2

Notes: a) Observations are weighted by the multipliers assigned to each household in the unit record datafile containing the consumer expenditure survey of the NSS in the 50th round.
Source: 50th round (1993/94) of the consumer expenditure survey of the NSS; our calculations.

Table 2
Sample Means (weighted)

	Scheduled Castes		Scheduled Tribes		Non-Scheduled	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Poverty Rates	0.49	0.5	0.495	0.5	0.328	0.469
Demographic Control Variables						
Age	42.23	12.32	41.36	11.89	43.69	12.44
Household Size	4.691	2.175	4.774	2.145	5.013	2.435
Education Variables						
Not Literate	0.679	0.467	0.695	0.46	0.482	0.5
Literate, below primary	0.117	0.322	0.121	0.326	0.148	0.356
Literate, below secondary	0.156	0.363	0.142	0.35	0.25	0.433
Literate, secondary	0.03	0.165	0.03	0.155	0.07	0.246
Literate, higher secondary & above	0.02	0.141	0.02	0.131	0.05	0.227
Occupation Variables						
Self-employed in non-agriculture	0.106	0.307	0.06	0.238	0.146	0.353
Agricultural labor	0.529	0.499	0.395	0.489	0.238	0.426
Non-agricultural labor	0.106	0.308	0.1	0.3	0.07	0.255
Self-employed in agriculture	0.195	0.396	0.39	0.488	0.438	0.496
Others	0.06	0.245	0.06	0.227	0.107	0.309
Number of Observations	12213		9861		42213	

Table 3
The Determinants of Poverty: Probit Analysis

	Scheduled Castes		Scheduled Tribes		Non-Scheduled	
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	-0.897 (0.0034)	-0.707 (0.0036)	-1.040 (0.0049)	-1.703 (0.0062)	-0.843 (0.0021)	-0.757 (0.0022)
DEMOGRAPHIC CONTROL VARIABLES						
Age	-0.027 (0.0002)	-0.034 (0.0002)	-0.024 (0.0002)	-0.027 (0.0002)	-0.023 (0.0001)	-0.028 (0.0001)
Age Square	0.020 (0.0002)	0.027 (0.0002)	0.014 (0.0003)	0.016 (0.0003)	0.015 (0.0001)	0.021 (0.0001)
Household size	0.329 (0.0004)	0.364 (0.0004)	0.424 (0.0006)	0.461 (0.0006)	0.304 (0.0002)	0.322 (0.0002)
Household size squared	-0.013 (0.0000)	-0.014 (0.0000)	-0.021 (0.0000)	-0.022 (0.0000)	-0.013 (0.0000)	-0.013 (0.0000)
EDUCATION VARIABLES - REFERENCE GROUP: 'NOT LITERATE'						
Literate, below primary	-0.260 (0.0008)	-0.290 (0.0009)	-0.309 (0.0012)	-0.345 (0.0012)	-0.247 (0.0005)	-0.301 (0.0005)
Literate, below secondary	-0.439 (0.0008)	-0.491 (0.0008)	-0.390 (0.0011)	-0.435 (0.0012)	-0.382 (0.0004)	-0.458 (0.0004)
Literate, secondary	-0.517 (0.0017)	-0.549 (0.0018)	-0.715 (0.0027)	-0.716 (0.0028)	-0.726 (0.0008)	-0.805 (0.0008)
Literate, higher secondary & above	-0.729 (0.0021)	-0.791 (0.0022)	-1.018 (0.0036)	-0.969 (0.0037)	-1.020 (0.0010)	-1.113 (0.0010)
OCCUPATION VARIABLES - REFERENCE GROUP: 'OTHERS'						
Self-employed in non-agriculture	0.317 (0.0014)	0.265 (0.0015)	0.264 (0.0024)	0.241 (0.0025)	0.037 (0.0007)	0.049 (0.0007)
Self-employed in agriculture	0.130 (0.0013)	0.075 (0.0014)	0.278 (0.0020)	0.255 (0.0020)	-0.137 (0.0006)	0.107 (0.0006)
Agricultural labor	0.787 (0.0012)	0.740 (0.0013)	0.684 (0.0020)	0.689 (0.0021)	0.574 (0.0007)	0.583 (0.0007)
Non-agricultural labor	0.418 (0.0014)	0.435 (0.0015)	0.495 (0.0022)	0.514 (0.0023)	0.247 (0.0008)	0.297 (0.0009)
With State Dummies?	No	Yes	No	Yes	No	Yes
Log Likelihood	-15222311	-14334263	-7747445	-7085045	-41014760	-38294722

Notes: a) Observations have been weighted by individual household multiplier. b) Dependent variable equals 1 if the household is below the poverty line. c) Standard Errors in parentheses. d) all coefficient estimates are significant at .01 level.

Table 4
Decomposition of the 16.2% Gap in Poverty Rates
Between Scheduled Castes vs. Non-Scheduled Aggregate and Sub-Aggregate Effects

Without State Dummies

	Characteristics Effect		Coefficients Effect	
	Estimate	Share(%)	Estimate	Share(%)
Aggregate Effects	*0.094*	*58*	*0.068*	*42*
Intercept			-0.019	-11.9
Demographic Control Variables				
Age	0.004	2.7	-0.031	-19.6
Household Size	-0.015	-9	0.044	26.8
Education	0.036	21.6	0	0.3
Occupation	0.069	42.7	0.074	46.4

With State Dummies

Aggregate Effects	*0.091*	*56.3*	*0.07*	*43.7*
Intercept			-0.016	-9.8
Demographic Control Variables				
Age	0.004	2.6	-0.037	-23
Household Size	-0.014	-8.7	0.058	35.7
Education	0.037	22.8	0.003	1.9
Occupation	0.062	38.6	0.05	30.8
Net Effect of State Dummies	0.002	-1	-0.019	-11.6

Note: Share is percentage share of difference in the probability of being poor.

Table 5
Decomposition of the 16.7% Gap in Poverty Rates
Between Scheduled Tribes vs. Non-Scheduled Aggregate and Sub-Aggregate Effects

Without State Dummies

	Characteristics Effect		Coefficients Effect	
	Estimate	Share(%)	Estimate	Share(%)
Aggregate Effects	*0.068*	*40.8*	*0.099*	*59.2*
Intercept	----	----	-0.071	-42.8
Demographic Control Variables				
Age	0.006	3.8	-0.039	-23
Household Size	-0.009	-5.2	0.123	73.9
Education	0.036	22.5	-0.003	-1.7
Occupation	0.032	19.7	0.088	52.8

With State Dummies

	Characteristics Effect		Coefficients Effect	
	Estimate	Share(%)	Estimate	Share(%)
Aggregate Effects	*0.043*	*25.7*	*0.124*	*74.3*
Intercept	—	—	-0.295	-176.2
Demographic Control Variables				
Age	0.007	4.2	-0.011	-7.1
Household Size	-0.009	-5.7	0.136	81.7
Education	0.044	26.6	0.001	0.5
Occupation	0.034	20.5	0.067	40.3
Net effect of state dummies	-0.033	-19.9	0.36	135.1

Note: Share is percentage share of difference in the probability of being poor.

Table 6
Decomposition of the 16.2% Gap in Poverty Rates
Between Scheduled Castes vs. Non-Scheduled Individual Effects
(without state dummies)

	Characteristics Effect		Coefficients Effect	
	Estimate	Share(%)	Estimate	Share(%)
Total	*0.094*	*58*	*0.068*	*42*
Intercept			-0.019	-11.9
DEMOGRAPHIC CONTROL VARIABLES				
Age	0.011	6.8	-0.061	-38
Age-squared	-0.007	-4.1	0.03	18.4
Household size	-0.033	-20.1	0.043	26.3
Household size squared	0.018	11.1	0.001	0.5
EDUCATION VARIABLES - REFERENCE GROUP: ‘NOT LITERATE’				
Literate, below primary	0.003	1.6	-0.001	-0.3
Literate, below secondary	0.012	7.4	-0.003	-1.9
Literate, secondary	0.009	5.5	0.002	1.3
Literate, higher secondary & above	0.012	7.2	0.002	1.3
OCCUPATION VARIABLE - REFERENCE GROUP: ‘OTHERS’				
Self-employed in non-agriculture	0	-0.3	0.01	6.5
Self-employed in agriculture	0.011	6.9	0.018	11.4
Agricultural labor	0.055	34.3	0.04	24.5
Non-agricultural labor	0.003	1.8	0.006	4

Note: Share is percentage share of difference in the probability of being poor.

Table 7
Decomposition of the 16.7% Gap in Poverty Rates
Between Scheduled Tribes vs. Non-Scheduled Individual Effects
(without state dummies)

	Characteristics Effect		Coefficients Effect	
	Estimate	Share(%)	Estimate	Share(%)
Aggregate Effect	*0.068*	*40.8*	*0.099*	*59.2*
Intercept			-0.071	-42.8
DEMOGRAPHIC CONTROL VARIABLES				
Age	0.017	10.2	-0.026	-15.4
Age-squared	-0.011	-6.4	-0.013	-7.6
Household size	-0.024	-14.2	0.206	123.6
Household size squared	0.015	9	-0.083	-49.7
EDUCATION VARIABLES - REFERENCE GROUP: 'NOT LITERATE'				
Literate, below primary	0.002	1.3	-0.003	-1.6
Literate, below secondary	0.013	8.1	0	-0.2
Literate, secondary	0.009	5.7	0	0.1
Literate, higher secondary & above	0.012	7.4	0	0
OCCUPATION VARIABLE - REFERENCE GROUP: 'OTHERS'				
Self-employed in non-agriculture	-0.001	-0.6	0.005	3
Self-employed in agriculture	0.002	1.3	0.058	35
Agricultural labor	0.029	17.6	0.016	9.4
Non-agricultural labor	0.002	1.4	0.009	5.4

Note: Share is percentage share of difference in the probability of being poor.