

Gang, Ira N.; Sen, Kunal; Yun, Myeong-Su

**Working Paper**

## Poverty in rural India: ethnicity and caste

Working Paper, No. 2006-34

**Provided in Cooperation with:**

Department of Economics, Rutgers University

*Suggested Citation:* Gang, Ira N.; Sen, Kunal; Yun, Myeong-Su (2006) : Poverty in rural India: ethnicity and caste, Working Paper, No. 2006-34, Rutgers University, Department of Economics, New Brunswick, NJ

This Version is available at:

<https://hdl.handle.net/10419/31274>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

## **Poverty in Rural India: Ethnicity and Caste\***

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 May 19, 2006

### **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 much higher than among non-scheduled households. By combining regression estimates for the ratio of per capita expenditure to the poverty line and an Oaxaca-type decomposition analysis, we study how these differences in the incidence of poverty arise. We find that for SC households, differences in characteristics explain the gaps in poverty incidence more than differences in transformed regression coefficients. In contrast, for ST households, the transformed regression coefficients play the more important role.

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

### **Correspondence:**

Ira N. Gang, Department of Economics, Rutgers University, 75 Hamilton St, New Brunswick NJ 08901-1248. Phone: +1 732 932-7405 Fax: +1 732 932-7416 Email: gang@economics.rutgers.edu

\*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, College of William and Mary, Graduate School of City University of New York, and the Midwest Economic Association. We thank Richard Palmer-Jones for helping us with the data.

## 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). Particularly troubling is the concentration of rural poverty in India in the ‘scheduled caste’ (SC) and ‘scheduled tribe’ (ST) populations.<sup>1</sup> The presence of such disparity in the incidence of poverty and wide-spread discrimination against scheduled groups have long histories in India. Affirmative action programs have been at the core of Indian social policy directed toward 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 43.3 percent of India’s rural poor are concentrated in these groups.<sup>2</sup> The incidence of poverty among scheduled caste and tribe households are much higher than for the rest of the population -- in 1993/1994 the proportion of rural SC and ST households below the poverty line were 49.2 and 50.3 percent respectively, as compared with a poverty rate of 33.1 percent for rural non-scheduled households. From Table 1 we

---

<sup>1</sup> 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.

<sup>2</sup> 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. These calculations used the Dubey-Gangopadhyay (DG) poverty lines. Using alternative Deaton-Tarozzi (DT) poverty lines, available for a subset of States and Union Territories, the composition of scheduled groups among the poor is 45.3 percent. We discuss the choice of poverty lines below.

see a gap in the proportion living in poverty (a poverty incidence gap) of 16.1 percent ( $= 49.2 - 33.1$ ) between SC and non-scheduled households, and a poverty incidence gap of 17.2 percent ( $= 50.3 - 33.1$ ) between ST and non-scheduled households.

One may attribute these disparities in poverty rates to discrimination, but this may be misleading since the disparities may arise from low income generating qualifications and credentials possessed by scheduled castes and tribes. Of course, it is possible that the level of these qualifications and credentials may be the result of discrimination. In this paper we study whether differences in the amounts of schooling, occupational choice and demographic characteristics hold the key to understanding the poverty incidence gap, and whether the poverty mitigating strength of household or individual characteristics (e.g., education and occupation) are different for each group. To answer these questions, we examine the determinants of poverty for scheduled households, SC and ST, and non-scheduled households and implement a methodology that allows us to examine causes of the disparity in poverty incidence. We address the causes of higher poverty amongst SC and ST households compared with non-scheduled households. This allows for the possibility that the determinants of poverty for these two social groups may not be the same, given the differing labor market attributes of these two groups.<sup>3</sup>

We use rural household survey data on 63,836 households from the 50th round of the National Sample Survey (NSS).<sup>4</sup> We estimate regression equations where the dependent variable

---

<sup>3</sup> The most important of these labor market attributes is that SC households interact in the same labor markets as non-scheduled households, as they often reside in the same villages, while ST households tend to be located in geographically distinct areas from non-scheduled households, and therefore, often operate in labor markets that isolated from the rest of the population.

<sup>4</sup> A more recent NSS round of the consumer expenditure survey is available – this is the 55<sup>th</sup> round conducted in 1999 and 2000. However, there have been questions asked about the accurate reporting of household spending on frequently purchased groups of food items in this round, since the same households were asked to report expenditures on these items for two alternative recall periods – 30 days and 7 days, with only expenditures for the 30 day recall period reported in the survey. There

is the natural logarithm of the ratio of (monthly) per capita expenditure to the poverty line, following an approach suggested in World Bank (2003). The likelihood of being in poverty can be calculated using the standard normal distribution function and transforming the regression coefficients by dividing them with the standard deviation of the error term. Based on this calculation of the likelihood of being in poverty for scheduled and non-scheduled groups, we can 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 differences in the incidence of poverty into the proportion explained by the differences in characteristics (characteristics effect) and the proportion explained by the differences in the "transformed regression" coefficients (coefficients effect).

The characteristics effect captures the amount of the poverty incidence gap caused by the differences in attributes. Though differences in characteristics are supposed to reflect differences in income generating qualifications and credentials possessed by scheduled and non-scheduled groups, it is possible 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 incidence gap caused by the differences in the effectiveness of characteristics in reducing poverty between the comparison groups.

Three recent studies – Bhaumik and Chakrabarty (2006), Kijima (2006) and Borooah (2005)

---

are concerns that the reporting may be biased if households were first canvassed on the 7 day reference period and this was subsequently extrapolated to the 30 day period by rough multiplicative adjustment, leading to an over-statement of consumer expenditures (Deaton 2003, Sundaram and Tendulkar 2003). If the multiplicative adjustment differed across households and by sampling unit, this would make consumption data across households and regions non-comparable. For this reason, we have decided to use the 50<sup>th</sup> round of the consumer expenditure survey where no such problem exists.

– investigate the living standards of the SC, ST and non-scheduled in India.<sup>5</sup> Bhaumik and Chakrabarty (2006) and Kijima (2006) examine differences in *earnings / mean consumption levels* among these social groups using Oaxaca's (1973) decomposition, but do not investigate the sharp differences in poverty incidence that exist.<sup>6</sup> Gang, Sen, and Yun (2002) and Borooah (2005) examine poverty incidence. While Gang, Sen, and Yun (2002) estimate the head count ratio using probit analysis, Borooah (2005) estimates a multinomial logit model of poverty, computing the average probability of being poor at different poverty lines for SC, ST and non SC/ST households. Borooah then decomposes poverty incidence into aggregate characteristics and coefficients effects using an extension of the Neilsen (1998) methodology applied to discrete choice models with multiple outcomes.

Our study adds to this literature by utilizing an alternative estimation strategy, by formulating the characteristics and coefficients effects at a highly disaggregated level that is consistent with calculations at the aggregate level, and by focusing attention on rural poverty. Above, we briefly described our estimation method and our decomposition approach, and will do so in greater detail below. A strength of the decomposition methodology we employ is that it allows us not only to calculate the aggregate characteristics and coefficients effects, but also these effects for groups of variables and even specific variables. Thus we will be able to say, for example, how much differences in schooling contribute to the gap in poverty between the groups, and how much

---

<sup>5</sup> See also the studies by Deshpande (2000, 2001) and Meenakshi and Ray (2002), which examine the economic status of scheduled castes and tribes. These studies do not examine the determinants of living standard disparities between scheduled and non-scheduled households.

<sup>6</sup> Bhaumik and Chakrabarty use individual level data on earnings from the employment and unemployment surveys of the NSS in their decomposition exercise. Kijima uses household level data drawn from the consumption surveys of the NSS to decompose differences in mean consumption levels between the SC/ST and the non-SC/ST into the components explained by differences in economic characteristics on one hand and differences in returns to characteristics on the other.

of the gap is related to the effectiveness of the education attainment differing between the scheduled and the non-scheduled groups.

Our focus is on rural poverty. Most of India's poor live in rural areas and the data we use (discussed below) 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 are weaker in urban areas where intermingling or intermarriage between SC, ST, and non-scheduled individuals occurs with greater frequency. Moreover, in urban areas there is less certainty about caste affiliation and thus room for false claims of belonging to lower castes to take advantage of the jobs reserved for them (Gatade, 2005).

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, examining the relative influence of various economic and non-economic variables on poverty. Section 4 employs decomposition analysis using the transformed regression coefficients to examine and explain the poverty incidence 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 extended from July 1993 to June 1994. The NSS data is a cross-section of a geographically distributed random sample of households. Besides 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. 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.<sup>7</sup>

We estimate the incidence of rural poverty across all three social groups, and relate these to their demographic, educational and occupational characteristics. We restrict our sample to households where the age of the head of the household is between 20 and 70 years.

An important issue that we need to address in determining the poverty status of households is the choice of the poverty line. Traditionally, poverty rates in India have been estimated using official rural and urban state-level poverty lines, which incorporate state to state differences in price levels, provided by the Indian Planning Commission (Government of India 1993).<sup>8</sup> A significant limitation of the official poverty lines is that the price indices used to update them are based on fixed commodity ‘weights’ that have become outdated over time. Deaton and Tarozzi (2000) have proposed an alternate set of poverty lines based on unit values and quantities consumed, obtained from the NSS expenditure surveys themselves. These price indexes have the advantage of allowing for substitution among goods as households adapt to relative price changes over time. The Deaton-Tarozzi poverty lines are not available for all States and Union Territories in India -- in particular, they are not available for North-East India where 37 per cent of ST households in our sample are found. In this paper, we use poverty lines provided by Dubey and Gangopadhyay (1998) that are available for all States and Union Territories in India. Unlike Deaton-Tarozzi, these are based on actual price data for individual items obtained from the Consumer Price surveys undertaken by the

---

<sup>7</sup> This distinction becomes important when there are significant differences in the intra-household consumption of food and other necessities across the SC, ST and non-scheduled households.

<sup>8</sup> 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.



Central Statistical Organisation, India, along with the use of current weighting diagrams from the NSS expenditure surveys.

The poverty rates by social group and by age, household size, educational level and occupation are presented in Table 1.<sup>9</sup> We observe that there is a non-linear relationship between age and incidence of poverty across all three social groups, with the poverty rate first increasing as we move from age group 20-29 to 30-39, and then decreasing for ages 40 years and above. Poverty increases with household size, with the highest poverty rates observed among households that have seven or more members. While literacy is negatively related to the incidence of poverty, the negative correlation between educational attainment and poverty incidence seems weaker for SC households as compared with ST and non-scheduled households. Approximately 23 percent of SC households with literacy levels of higher secondary and above are poor as compared to 15.7 percent of similarly educated ST households and 10.2 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 for 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 compared to non-scheduled households. SC and ST households are also smaller than non-scheduled households – the mean household size for SC and ST households are 4.7 and 4.8 respectively, compared with a mean household size of 5.0 for non-scheduled households.

A much higher proportion of SC and ST households are not literate (66 percent and 68

---

<sup>9</sup> Our poverty 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.

percent, respectively), compared with non-scheduled households (47 percent). 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, 20 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, 40 percent are agricultural laborers, 10 percent are non-agricultural laborers, 39 percent are self-employed in agriculture while 5 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 11 percent are in other occupations. Thus, a greater proportion of SC households are agricultural laborers than are ST and non-scheduled households.

Although interesting, Table 2 is only suggestive as the observed bivariate connections have not controlled other variables. We carry out a multivariate analysis of the factors determining poverty status below.

### 3. Determinants of Poverty Incidence

We employ an approach proposed in World Bank (2003) to understand why households are in poverty.<sup>10</sup> According to World Bank (2003), poverty incidence can be computed using the following two step method. First, construct the ratio of per capita expenditure ( $Y$ ) to the poverty line ( $Z$ ), i.e.,  $R = Y/Z$ . The regression equations is  $\log R = X\beta + e$ , where  $R$ ,  $X$ , and  $\beta$  are, respectively, an  $N \times 1$  vector, an  $N \times K$  matrix of independent variables, and a  $K \times 1$  vector of coefficients. Second, the probability of being in poverty is obtained by computing  $Pr(\log R < 0)$ ;

---

<sup>10</sup> See Dreze and Srinivasan (1997) for a similar analysis of poverty incidence among widow-headed households in India. Gang, Sen and Yun (2002) employed probit analysis to examine poverty incidence.

usually this probability is computed using the standard normal distribution function,  $\Phi(\cdot)$ , i.e.,  $Pr(e < -X\beta) = \Phi(X\tilde{\beta})$ , where  $\tilde{\beta} = -\beta/\sigma$  and  $\sigma$  is the standard deviation of the error term ( $e$ ). Obviously, if  $X\beta$  is larger or  $X\tilde{\beta}$  is smaller, then it is likely that the ratio of per capita expenditure to the poverty line increases and the likelihood of being in poverty decreases. In this section, we discuss the specification our regression equation, which we estimate using maximum likelihood for households in the non-scheduled group, scheduled castes (SC) and scheduled tribes (ST), separately.<sup>11</sup> We also discuss the implications of the estimated coefficients on the likelihood of being in poverty.

Our focus 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).<sup>12</sup>

---

<sup>11</sup> OLS can be used and the OLS estimates are virtually identical to ML estimates. The merit of ML is that it provides the covariance matrix of  $(\beta, \sigma)$  which is used to compute the covariance matrix for  $\tilde{\beta}$ , allowing us to perform significance tests for the decomposition equation.

<sup>12</sup> The NSS classifies rural households in occupational categories according to the main source of income reported for each surveyed household. This is called the “principal occupation code” of the household. The principal occupation is defined to be that which contributes at least 50% of household income. The category ‘others’ includes those where no one income source exceeds 50% or more of total income. Thus, the households in this category have very diversified income sources or more than one earning member.

Besides the explanatory variables capturing occupation and educational levels, we include in our analysis a number of background and demographic variables. 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. 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.<sup>13</sup> We also include total cultivated land owned by the household as a measure of the household's wealth status.

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 much higher than the national average. The omission of state dummy variables 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 dummy variables.

A further potential econometric problem that we may encounter is the presence of

---

<sup>13</sup> We do not include the child-adult ratio that is often used to 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 are broadly similar to the ones reported.

geographical fixed effects operating at the level of the village where the household is found. These may be latent factors correlated with the included variables, for example, inter-village differences in land quality, local infrastructural development, geo-environmental attributes and other village level factors (van de Walle and Gunewardena 2001). We address the possible presence of such unobserved village-level effects in the error term in the estimated equations by using a cluster-correlated robust estimate of the variance-covariance matrix (Wooldridge 2002, pp. 409-410).<sup>14</sup>

The ML estimates of the regression equation 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 dummy variables. Columns (2), (4) and (6) contain the results with the state dummy variables. Though the reported coefficients for each of the independent variables are broadly similar across all three social groups, likelihood ratio tests (not reported) show that the coefficients for each group are significantly different from the other groups.<sup>15</sup>

The estimated coefficients show that greater educational attainment is associated with a statistically significant increase in the ratio of per capita expenditure to the poverty line, implying a reduction in the probability of being poor, with everything else held constant. This is true for all three household groups. However, higher educational attainment from the secondary level up seems to lead to a greater decline in the incidence of poverty among ST and non-scheduled households when compared with SC households.

We now turn our attention to occupation and its impact on the ratio of per capita expenditure to the poverty line, and its implications for poverty status. Compared with the occupational category

---

<sup>14</sup> We use the robust cluster option in Stata 8. The modified sandwich estimator associated with this option does not assume any particular kind of within-cluster correlation or a particular form of heteroskedasticity.

<sup>15</sup> This likelihood ratio test supports our approach of studying SC and ST separately.

‘others’, all other occupational categories lead to a lower ratio of per capita expenditure to the poverty line, i.e., those households are more likely to have a higher poverty incidence for all three social groups. Agricultural laborer household are more likely to be poor among all occupational groups, controlling for other determinants. Overall, the results suggest that households that contain laborers, whether involved in agricultural or non-agricultural work, are more likely to be in poverty when compared with households where there are self-employed, since the coefficients on laborers are more negative. With respect to demographic factors, higher aged heads of households are associated with a higher ratio of per capita expenditure to the poverty line (i.e., lower poverty). However, this relationship is non-linear, with further increases in age leading to less than proportionate increases in the ratio of per capita expenditure to the poverty line. A non-linear relationship is also found between poverty and household size; the ratio of per capita expenditure to the poverty line decreases within reasonable household size (about ten) then increases; poverty is more evident in larger sized households within the “reasonable” range. The possession of cultivable land seems to have a positive effect of similar magnitude on the ratio of per capita expenditure to the poverty line across all three social groups, once the location of the household is controlled for.

To summarize, the results imply that households that are larger, where the head of the household is not literate, is an agricultural laborer, and is younger in age, and possess a smaller amount of land are more likely to be in poverty. We also find that the effects of explanatory variables on the ratio of per capita expenditures to the poverty line vary over social groups.

#### 4. Accounting for Differences in Poverty Incidence

In this section, we seek to explain why poverty is so much more prevalent among the scheduled caste and tribe households, than among non-scheduled households. For the scheduled groups in comparison to the non-scheduled we are seeking to find the sources of the poverty incidence gap. The observed gap is 16.1 percent ( $= 49.2 - 33.1$ ) for the scheduled castes versus the non-scheduled; for scheduled tribes versus the non-scheduled the observed gap is 17.2 percent ( $= 50.3 - 33.1$ ). In practice, for the various explanatory elements, the share explained is calculated as a proportion of the “predicted” poverty incidence gap, i.e., it is based on our estimates. The predicted gap is 14.9 percent ( $= 47.3 - 32.3$ ) for the scheduled castes versus the non-scheduled; for scheduled tribes versus the non-scheduled the predicted gap is 16.2 percent ( $= 48.5 - 32.3$ ). Our analysis breaks down the predicted poverty incidence gap into its components, at different levels of aggregation.<sup>16</sup>

In examining the gap in poverty incidence we focus on the characteristics effect and the coefficients effect. The characteristics effect relies on the possibility that the characteristics or attributes of households that cause poverty differ among groups. For example, one group may have less education than another group, or be in “bad” jobs. The characteristics effect reflects how differences in the attributes of households among groups affect the likelihood that someone is in poverty.

The coefficients effect relies on the possibility that the effectiveness of household

---

<sup>16</sup> As noted in World Bank (2003), the binary choice models (e.g., probit) typically have better predictive power in classifying households as poor or non-poor than fitting poverty incidence using regression estimates (ML or OLS). It should be noted that though theoretically the continuous variable contains more information than using just binary information, it may not be clear whether the gain from using estimates from the continuous variable regression outweighs the reduction in fit.

characteristics, reflected in transformed regression 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 compared with non-scheduled households. The coefficients effect reflects how differences in the transformed regression coefficients across groups affect the likelihood that someone is in poverty.

The study of characteristics and coefficients effects was formally introduced by Oaxaca (1973). Though the implementation and extensions of the Oaxaca decomposition have generally been in wage differentials (in general, any continuous variable), the methodology has been extended to allow for discrete dependant variables (e.g., Even and Macpherson 1993; Yun 2004). Decomposing differences in the mean value of a binary dependent variable (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 characteristics constant. 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 explains only the coefficients effect since it looks only at the effect of switching coefficients. The decomposition method proposed by Yun (2004) provides a systematic treatment for differences in binary outcomes.

We can easily incorporate the computation of the probability of poverty incidence proposed in World Bank (2003), discussed in the section 3, into the decomposition methodology developed



by Yun (2004) when comparing poverty incidence across groups. This is because the decomposition equation of Yun (2004) is an extension of the Oaxaca decomposition to a nonlinear model, e.g., probit, and both the probit model and the computation of the probability of poverty incidence described above use the standard normal distribution function.

#### 4.1. Decomposing the Differences in Poverty Incidence using Regression Estimates

As discussed in Section 3, we first estimate the regression coefficients ( $\beta$ ) and the standard deviation of the error term ( $\sigma$ ) for each group. By transforming the estimates to  $\tilde{\beta} = -\beta/\sigma$ , we can compute the probability of being in poverty as  $\Phi(\mathbf{X}\tilde{\beta})$ , where  $\Phi$  is the standard normal cumulative distribution function. Algebraically, the differences in the average probability of being poor between groups  $A$  and  $B$ ,  $(\bar{P}_A - \bar{P}_B)$ , where  $A$  = scheduled castes or tribes and  $B$  = non-scheduled, may be decomposed into two components that represent the characteristics effect and coefficients effect. Asymptotically, this is,

$$\bar{P}_A - \bar{P}_B = \left[ \overline{\Phi(\mathbf{X}_A \tilde{\beta}_B)} - \overline{\Phi(\mathbf{X}_B \tilde{\beta}_B)} \right] + \left[ \overline{\Phi(\mathbf{X}_A \tilde{\beta}_A)} - \overline{\Phi(\mathbf{X}_A \tilde{\beta}_B)} \right], \quad (1)$$

where  $\Phi$  is the standard normal cumulative distribution function;  $\tilde{\beta}_A = -\beta_A/\sigma_A$  and  $\tilde{\beta}_B = -\beta_B/\sigma_B$ ,  $\beta_A$  and  $\beta_B$  are sets of estimated coefficients for each group, and  $\sigma_A$  and  $\sigma_B$  are the standard deviation of error term ( $e_A$  and  $e_B$ );  $X_A$  and  $X_B$  are the various explanatory variables used in the regression equations; “over bar” represents the value of the sample’s average.

The above decomposition gives us the overall coefficients and characteristics effects. To find the relative contribution of each variable to the predicted poverty incidence gap, in terms of

characteristics and coefficients effects, we employ a decomposition equation proposed by Yun (2004);<sup>17</sup>

$$\bar{P}_A - \bar{P}_B = \sum_{k=1}^K W_{\Delta X}^k \left[ \overline{\Phi(X_A \tilde{\beta}_B)} - \overline{\Phi(X_B \tilde{\beta}_B)} \right] + \sum_{k=1}^K W_{\Delta \tilde{\beta}}^k \left[ \overline{\Phi(X_A \tilde{\beta}_A)} - \overline{\Phi(X_A \tilde{\beta}_B)} \right], \quad (2)$$

where

$$W_{\Delta X}^k = \frac{(\bar{X}_A^k - \bar{X}_B^k) \tilde{\beta}_B^k}{(\bar{X}_A^k - \bar{X}_B^k) \tilde{\beta}_B^k}, \quad W_{\Delta \tilde{\beta}}^k = \frac{\bar{X}_A^k (\tilde{\beta}_A^k - \tilde{\beta}_B^k)}{\bar{X}_A^k (\tilde{\beta}_A^k - \tilde{\beta}_B^k)}, \quad \text{and} \quad \sum_{k=1}^K W_{\Delta X}^k = \sum_{k=1}^K W_{\Delta \tilde{\beta}}^k = 1, \quad (3)$$

where  $\bar{X}_A^k$  and  $\bar{X}_B^k$  are average values of explanatory variables  $k$  for groups  $A$  and  $B$ , respectively.<sup>18</sup>

Decomposition analysis does not usually calculate the standard errors of the components of the decomposition equation. We compute the standard errors and implement hypothesis testing. For doing this, we estimate the regression equation using maximum likelihood (ML) instead of OLS. The ML provides the covariance matrix of estimates ( $\beta$ ) and the standard deviation of the error term ( $\sigma$ ) which is used for deriving the asymptotic covariance matrix for ( $\beta/\sigma$ ) using the delta method.

---

<sup>17</sup> In order to obtain a proper weight, the following approximations are used; first, an approximation of the value of the average of the function,  $\overline{\Phi(X\tilde{\beta})}$ , with that of the function evaluated at the average value of exogenous variables  $\Phi(\bar{X}\tilde{\beta})$ ; second, a first order Taylor expansion to linearize the characteristics and coefficients effects around  $\bar{X}_B \tilde{\beta}_B$  and  $\bar{X}_A \tilde{\beta}_A$ , respectively. See Yun (2004) for details.

<sup>18</sup> A decomposition equation with a different parameterization, that is,  $\bar{P}_A - \bar{P}_B = \left[ \overline{\Phi(X_A \tilde{\beta}_A)} - \overline{\Phi(X_B \tilde{\beta}_A)} \right] + \left[ \overline{\Phi(X_B \tilde{\beta}_A)} - \overline{\Phi(X_B \tilde{\beta}_B)} \right]$ , 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. Another issue when interpreting the decomposition results is that the coefficients effect in the detailed decomposition is not invariant to the choice of omitted groups when dummy variables are used (see Oaxaca and Ransom 1999, for details of this issue). Therefore, the readers should keep in mind the invariance issue when reading the results of the decomposition analysis.

The covariance of  $\beta/\sigma$  is, in turn, used for hypothesis testing (Yun, 2005). In our discussion, we take account of both the size and significance of the components.

#### **4.2. Explaining Differences in Poverty Incidence**

We now discuss our empirical findings from the decomposition analysis. We focus on the percentage share that tells us what percentage of the (predicted) total poverty incidence 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 smaller subgroups. We discuss the poverty incidence gap of scheduled castes relative to the non-scheduled in Table 4, and that of scheduled tribes compared with the non-scheduled in Table 5. 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 dummy variables and when we do.

We proceed by first discussing the aggregate effects and sub-aggregate effects without state dummy variables for SC households respectively (Table 4). The *Aggregate Effects* row in the top panel of the table shows the overall effects of characteristics versus coefficients in explaining differences in poverty. The majority (62.5 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 37.5 percent by the differences in the transformed regression coefficients. Both aggregate characteristics and coefficients effects are significant at the 1 percent level of significance. One could call the latter the difference in effectiveness since it is a summary measure of the differences in the strength of the various individual characteristics influencing

poverty.<sup>19</sup> If in both groups the various variables influencing poverty status had the same strength (their transformed coefficients had been equal), then 37.5 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, 62.5 percent of the poverty incidence gap would disappear. When we include state dummy variables (bottom panel), the aggregate coefficients effect is 41.7 percent and the aggregate characteristics effect is 58.3 percent.

In Table 4 we also see the breakdown of characteristics and coefficients effects into important variable groupings. We see the importance of the characteristics effect for occupation in determining the poverty incidence gap, contributing 35.1 percent. One of the salient features of the caste system is the generally undesirable and low-paying jobs scheduled castes are allowed to perform. SC households generally are in less-remunerative occupations. This may confirm anthropological evidence about the lack of job choice for individuals belonging to scheduled castes (Srinivas 1962; Beteille 1965).<sup>20</sup> While the coefficients effect, contributing 19.1 percent of the gap (but not significant), 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

---

<sup>19</sup> The coefficients effect in the well-known Oaxaca decomposition of the wage gap is conventionally considered discrimination (Oaxaca 1973). We interpret the coefficients effect as the differences in the effectiveness of characteristics to reduce poverty. Since we compute 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 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.

<sup>20</sup> Such discrimination may generate an ‘equilibrium trap’ where those who break caste customs suffer economically (Akerlof 1976).

difference in the strength of the poverty reducing effect of occupation for SC and non-scheduled households that leaves more SC households in poverty. Though the size of the effect is somewhat less, we also find a large characteristics effect for the occupational distribution when state dummy variables are included; the coefficients effect is smaller, but still positive. Education is remarkable in that both with and without state dummy variables the characteristics effect is 23 - 25 percent. SC households attain lower levels of schooling, and that puts them at greater risk of being poor.<sup>21</sup> The coefficients effect is a small 4 - 5 percent, indicating that the effectiveness of education level achieved differs between SC and non-scheduled households, but more modestly.

Ownership of land, age and household size are included as control variables, yet the results are interesting in and of themselves. The characteristics effect of land owned contributes 8 - 12 percent to the poverty incidence gap. However, there is almost no coefficient effect for land owned, suggesting that differences in land owned, rather than differences in the quality of owned land may be a contributing factor to the differences in poverty rates between SC and non-scheduled households. The coefficients effect of age structure (age and age-squared taken together) is negative and not significant while the characteristics effect is positive and significant, though small. For household size we find the characteristics effect is negative, and the coefficients effect is positive. Household size differences reduce the poverty incidence gap, but differences in coefficients increase the poverty incidence gap.<sup>22</sup>

Up to now we have discussed what accounts for differences in poverty incidence between SC and non-scheduled households. We now turn to a discussion of what explains differences in

---

<sup>21</sup> 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.

<sup>22</sup> As seen in Table 2, SC and ST households are smaller in size than non-scheduled households, and our analysis suggests that the likelihood of being poor is positively related to household size.

poverty rates between ST and non-scheduled households, shown in Table 5. Approximately 39 percent of the poverty incidence gap is explained by differences in households' characteristics between the two groups, and this difference is statistically significant. Thus, if ST and non-scheduled households had the same characteristics, then the poverty incidence gap would have been 39 percent less (33 percent less when state dummy variables are included). Differences in educational attainment account for 23.5 percent of the poverty incidence gap. The occupational distribution explains 18.1 percent of the higher poverty among the ST households as compared to the non-scheduled. These results are basically the same when state dummy variables are included.

Sixty-one percent of the predicted poverty incidence gap of 16.2 percent between ST and non-scheduled is explained by the differences in the transformed regression coefficients between ST and non-scheduled households (67 percent when state dummy variables are included). Again, the differences in coefficients between the two groups are statistically significant. If in both groups the various variables influencing poverty status had the same strength (if their transformed coefficients had been equal), then about 61 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 the contributions of subgroups to the poverty incidence gap, we find that the coefficients effect of educational attainment is negligible. The coefficients effect of the occupational distribution group of variables is large between the ST and non-scheduled, accounting for 29 percent of the difference in the probability of being in poverty, though not significant when state dummy variables are omitted. With state dummy variables the effect is significant at the 10 percent level and explains 35.7 percent of the poverty incidence gap. This suggests that for ST households, more than occupational structure, what have contributed to the greater incidence of poverty among such households have been the much

lower returns they have received for the jobs they hold as compared with non-scheduled households.<sup>23</sup>

With respect to demographic control variables, both the characteristics and the coefficients effects of age structure (age and age-squared taken together) are positive. Thus, the age structure of ST households is worse for reducing poverty than that of non-scheduled households. Household size, both excluding and including state dummy variables, have a high positive coefficients effect and about a negative six percent characteristics effect. Land owned has a minor role to play in explaining the poverty incidence gap between ST and non-scheduled households -- both the characteristics and coefficients effects are quite small.<sup>24</sup>

---

<sup>23</sup> One may wonder what happens if occupation variables are omitted, as occupations are often considered endogenous. Once occupation variables are excluded from the specification of the regression and the computation of the decomposition equation, then, roughly speaking, the characteristics and coefficients effects previously attributed to occupations are shifted to differences in intercepts while the two effects of the other variables are not changed substantially. As the constant term is included in the coefficients effect, for decomposing differences in poverty incidence between the SC or ST and non-scheduled, this increases the size of the aggregate coefficients effect and decreases the size of the aggregate characteristics effect. Note that the characteristics and coefficients effects attributed to occupations are positive when occupations are included, so that excluding them increases the differences in the intercepts and increasing the size of the coefficients effect. This may be a natural consequence of not controlling for occupations in the regression. The results of the regression and the decomposition without occupation variables are available from authors upon request.

<sup>24</sup> To examine how sensitive our results are to the choice of poverty lines, we also undertook the decomposition analysis (both with and without state dummy variables) using Deaton-Tarozzi (DT) and Dubey-Gangopadhyay (DG) poverty lines only for those states where DT poverty lines were available. For SC households, we obtained identical results with respect to decomposition of the poverty incidence gap between aggregate characteristics and coefficients effects. For ST households, when we used the DT poverty line, the aggregate coefficients effect was 57.7 per cent while it was 60.3 per cent when we used the DG poverty line. Thus, our results are robust to the choice of the poverty line.

## 5. Summary and Conclusions

This paper has examined the relative significance of some key forces that shape the poverty profiles of the scheduled castes (SC), scheduled tribes (ST) and non-scheduled households in India by combining regression estimation with decomposition analysis. While there is little difference in the incidence of poverty between SC and ST households, observed poverty rates of SC and ST households are 16.1 percentage points and 17.2 percentage points higher than non-scheduled households. Our analysis decomposes the predicted poverty incidence 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 shows that for SC households differences in characteristics explain the poverty incidence gap more than differences in coefficients, with 63 percent of the poverty incidence gap attributable to the former. For ST households, however, it is the reverse, with 61 percent of the poverty incidence 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 different.

Differences in educational attainments explain about one quarter of the poverty incidence gap for both social groups, though returns to education do not matter in explaining the poverty incidence gap. Our results suggest that allocating more resources towards scheduled group children and shifting the educational focus from higher education to primary and secondary schools will decrease the discrepancy in poverty incidence between the scheduled groups and non-scheduled households. Though subsidies to higher education may contribute to India's current surge in high-



tech industries, this policy favors children from more affluent households, as the children of the poor reach higher education with relatively less frequency (PROBE Team 1999).

The difference in the social and economic attributes of SC and ST households may explain why the causes of the difference of poverty incidence between these social groups and the non-scheduled households are different. A major source of the difference in the causes of poverty between these two social groups lies in the characteristics effect of occupational structure, compared with its coefficient effect. The impact of the characteristics effect on the poverty incidence gap between SC households and non-scheduled households is twice its corresponding impact on the poverty incidence gap between ST and non-scheduled households. SC households operate in similar labor markets as non-scheduled households, and are unlikely to obtain significantly different income flows than non-scheduled households for the same occupation and level of education. Thus, it is more likely that for SC households, it is social constraints to occupational diversification (because of the caste system), rather than returns to occupational structure, that explains much of the poverty incidence gap. For ST households who often operate in geographically distinct labor markets from non-scheduled households, and who do not face similar constraints to occupational diversification, their location in unfavorably endowed areas in terms of agricultural potential and their relative lack of access to superior technology may help explain why for such households, low returns to occupational structure, rather than the type of occupation that they are in, explains much of the poverty incidence gap between these households and non-scheduled households.

## REFERENCES

- Abowd, John M. and Mark R. Killingsworth. 1984. "Do Minority/White Unemployment Differences Really Exist?" *Journal of Business and Economic Statistics* 2: 64-72.
- Akerlof, George. 1976. "The Economics of Caste and the Rat Race and Other Woeful Tales." *Quarterly Journal of Economics* 90: 600-617.
- Beteille, André. 1965. *Caste, Class and Power: Changing Patterns of Social Stratification in a Tanjore Village*. Berkeley: University of California Press.
- Bhaumik, Sumon K. and Manisha Chakrabarty (2006), "Earnings Inequality in India: Has the Rise of Caste and Religion based Politics in India had an impact?", Bonn: Institute for the Study of Labour (IZA) Discussion Paper No. 2008, pp. 1-30.
- Borooh, Vani K. (2005), "Caste, Inequality and Poverty in India", *Review of Development Economics*, Vol. 9(3), pp. 399-414.
- Datt, Gaurav and Martin Ravallion. 1998. "Why Have Some Indian States Done Better Than Others at Reducing Rural Poverty?" *Economica* 65: 17-38.
- Deaton, Angus (2003), "Adjusted Indian Poverty Estimates for 1999-2000", *Economic and Political Weekly*, January 25, pp. 322-326.
- Deaton, Angus and Alessandro Tarozzi. 2000. "Prices and poverty in India." Working Paper, Research Program in Development Studies, Princeton University.
- Deshpande, Ashwini. 2001. "Caste at Birth? Redefining Disparity in India." *Review of Development Economics* 5: 130-144.
- Deshpande, Ashwini. 2000. "Does Caste Still Define Disparity? A Look at Inequality in Kerala, India." *American Economic Review* 90: 322-325.
- Dreze, Jean and Geeta Kingdon. 2001. "School Participation in Rural India." *Review of Development Economics* 5: 1-24.
- Dreze, Jean. 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, Amaresh. and Shubhashis Gangopadhyay. 1998. *Counting the Poor: Where are the Poor in India?* Department of Statistics, Government of India.
- Even, William E. and David A. Macpherson. 1993. "The Decline of Private-Sector Unionism and the Gender Wage Gap." *Journal of Human Resources* 28: 279-296.

Gang, Ira N., Kunal Sen, Myeong-Su Yun. (2002). Caste, Ethnicity and Poverty in Rural India. IZA Discussion paper 629.

Gatade, Subhash. 2005. "Phenomenon of False Caste Certificates." *Economic and Political Weekly*, October 22, 2005.

Government of India. 1993. *Report of the Expert Group on Estimation of Proportion and Number of Poor*. Planning Commission: New Delhi.

Ham, John C., Jan Svejnar and Katherine 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.

Kijima, Yoko (2006), "Caste and Tribe Inequality: Evidence from India, 1983-1999", *Economic Development and Cultural Change*, Vol. 54, No. 2, pp. 369-404.

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, Jonnalagadda Venkata and Ranjan Ray. 2002. "Impact of household size and family composition on poverty in rural India." *Journal of Policy Modeling* 24: 539-559.

Nielsen, Helena S. (1998), "Discrimination and Detailed Decomposition in a Logit Model", *Economic Letters*, Vol. 61, pp. 115-120.

Oaxaca, Ronald. 1973. "Male-Female Differentials in Urban Labor Markets." *International Economic Review* 14: 693-709.

Oaxaca, Ronald and Michael R. Ransom. 1999. "Identification in Detailed Wage Decompositions." *Review of Economics and Statistics* 81: 154-157.

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

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

Srinivas, Mysore Narasimhachar. 1962. *Caste in Modern India*. Bombay: Asia Publishing House.

Sundaram, K. and Suresh D. Tendulkar (2003), "Poverty Has Declined in the 1990s: A Resolution of Comparability Problems in NSS Consumer Expenditure Data", *Economic and Political Weekly*, January 25, pp. 327-337.

Van de Walle, Dominique and Dileni Gunewardena. 2001. "Sources of Ethnic Inequality in Vietnam." *Journal of Development Economics* 65: 177-207.

Wooldridge, Jeffrey 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA.: The MIT Press.

World Bank. 2003. "Poverty Reduction Strategy Sourcebook." <http://www.worldbank.org/poverty/>.

Yun, Myeong-Su. 2004. "Decomposition Differences in the First Moment." *Economics Letters* 82: 273-278.

Yun, Myeong-Su. 2005. "Hypothesis Tests when Decomposing Differences in the First Moment." *Journal of Economic and Social Measurement* 30: 295-304.

**Table 1: Poverty Rates**

	Scheduled Castes	Scheduled Tribes	Non- Scheduled	All
<b>Overall</b>	49.2	50.3	33.1	38.7
<b>Age</b>				
20-29	45.6	49	30.6	36.9
30-39	56.1	57.3	38.5	45
40-49	49.1	49.3	32.4	37.9
50-59	44.3	44	30.1	34.6
60-70	44.5	44.8	30.6	34.6
<b>Household Size</b>				
1	21.4	11.9	15.8	16.7
2	28.9	25.7	17	21.2
3	36.2	40.4	23.5	28.7
4	49.5	49.9	27.7	35.4
5	54.5	55.3	36.6	42.8
6	59.7	62.7	42.8	48.8
7 or more	65.4	65.2	44.2	50.3
<b>Education</b>				
Not Literate	54.1	54.9	41.1	46.7
Literate, below primary	44.9	45.4	33.2	36.7
Literate, below secondary	38.6	41.4	27.5	30.3
Literate, secondary	32.8	25.9	16.5	18.9
Literate, higher secondary & above	23.3	15.7	10.2	11.8
<b>Occupation</b>				
Self-employed in non-agriculture	42.1	41.5	30.6	33.3
Self-employed in agriculture	37.6	45.7	26.7	30.2
Agricultural labor	58.9	59	51.1	55.1
Non agricultural labor	45.1	53.2	37.7	42
Others	23.5	24.4	19.9	20.8

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**

	Scheduled Castes		Scheduled Tribes		Non-Scheduled	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Poverty Rates	0.492	0.500	0.503	0.500	0.331	0.471
Ratio of Monthly Per Capita Expenditure to the Poverty Line	1.160	0.722	1.159	1.362	1.473	1.427
<b>Demographic and Other Control Variables</b>						
Age	42.23	12.32	41.33	11.89	43.69	12.44
Household Size	4.692	2.174	4.773	2.146	5.015	2.435
Land Owned (hectares)	0.362	0.912	0.99	2.146	0.993	2.069
<b>Education Variables</b>						
Not Literate	0.662	0.473	0.683	0.465	0.466	0.499
Literate, below primary	0.134	0.341	0.135	0.342	0.164	0.371
Literate, below secondary	0.156	0.363	0.141	0.348	0.25	0.433
Literate, secondary	0.030	0.165	0.02	0.154	0.07	0.246
Literate, higher secondary & above	0.020	0.141	0.02	0.128	0.05	0.227
<b>Occupation Variables</b>						
Self-employed in non-agriculture	0.106	0.308	0.06	0.238	0.146	0.353
Self-employed in agriculture	0.195	0.396	0.389	0.488	0.438	0.496
Agricultural labor	0.529	0.499	0.397	0.489	0.238	0.426
Non-agricultural labor	0.106	0.308	0.101	0.301	0.07	0.255
Others	0.060	0.245	0.05	0.224	0.107	0.309
<b>Number of Observations</b>	12228		9437		42171	

Notes: Sample means are calculated using the individual household multiplier.

Source: 50th round (1993/94) of the consumer expenditure survey of the NSS; our calculations.

**Table 3: The Determinants of (log) Ratio of Monthly Per Capita Expenditure to the Poverty Line, Maximum Likelihood Estimates**

	Scheduled Castes		Scheduled Tribes		Non-Scheduled	
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.350*** (0.064)	0.259*** (0.059)	0.389*** (0.087)	0.605*** (0.102)	0.380*** (0.038)	0.335*** (0.036)
<b>Demographic Control Variables</b>						
Age	0.011*** (0.003)	0.014*** (0.002)	0.008** (0.004)	0.008*** (0.003)	0.012*** (0.002)	0.014*** (0.001)
Age Square	-0.009*** (0.003)	-0.012*** (0.003)	-0.004 (0.004)	-0.005 (0.004)	-0.010*** (0.002)	-0.012*** (0.002)
Household size	-0.135*** (0.007)	-0.139*** (0.006)	-0.136*** (0.016)	-0.136*** (0.015)	-0.134*** (0.004)	-0.132*** (0.004)
Household size squared	0.006*** (0.001)	0.006*** (0.001)	0.006*** (0.001)	0.005*** (0.001)	0.005*** (0.000)	0.005*** (0.000)
Land Owned (hectares)	0.041*** (0.006)	0.023*** (0.005)	0.020*** (0.006)	0.021*** (0.004)	0.038*** (0.002)	0.027*** (0.002)
<b>Education Variables – Reference Group: ‘Not Literate’</b>						
Literate, below primary	0.085*** (0.013)	0.100*** (0.012)	0.105*** (0.018)	0.110*** (0.018)	0.096*** (0.008)	0.118*** (0.008)
Literate, below secondary	0.140*** (0.013)	0.156*** (0.012)	0.142*** (0.020)	0.153*** (0.018)	0.173*** (0.008)	0.194*** (0.007)
Literate, secondary	0.247*** (0.028)	0.248*** (0.028)	0.334*** (0.045)	0.324*** (0.045)	0.330*** (0.013)	0.338*** (0.012)
Literate, higher secondary & Above	0.350*** (0.032)	0.365*** (0.031)	0.517*** (0.062)	0.493*** (0.063)	0.471*** (0.018)	0.488*** (0.017)
<b>Occupation Variables - Reference Group: ‘Others’</b>						
Self-employed in non-agriculture	-0.103*** (0.026)	-0.077*** (0.023)	-0.098* (0.056)	-0.105** (0.050)	-0.062*** (0.013)	-0.064*** (0.012)
Self-employed in agriculture	-0.091*** (0.025)	-0.048** (0.022)	-0.161*** (0.036)	-0.160*** (0.037)	-0.047*** (0.013)	-0.039*** (0.012)
Agricultural labor	-0.288*** (0.024)	-0.248*** (0.021)	-0.265*** (0.038)	-0.258*** (0.039)	-0.280*** (0.014)	-0.258*** (0.013)
Non-agricultural labor	-0.144*** (0.026)	-0.135*** (0.024)	-0.169*** (0.041)	-0.185*** (0.043)	-0.136*** (0.018)	-0.142*** (0.016)
Standard Deviation of Error	0.405*** (0.005)	0.386*** (0.005)	0.409*** (0.012)	0.383*** (0.010)	0.441*** (0.004)	0.419*** (0.004)
<b>State Dummy Variables?</b>	No	Yes	No	Yes	No	Yes

Notes: a) Observations have been weighted by the individual household multiplier. b) Dependent variable is the natural logarithm of ratio of monthly per capita expenditure to poverty line. c) Heteroskedasticity-robust standard errors in parentheses. d) \*\*\*, \*\*, and \* denotes significance at the 1, 5 and 10 percent respectively. e) Joint hypothesis test for coefficients that all state dummy variables are zero is rejected when log-likelihood test is performed.

Source: 50th round (1993/94) of the consumer expenditure survey of the NSS; our calculations.

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

*Without State Dummy Variables*

	Characteristics Effect		Coefficients Effect	
	Estimate	Share(%)	Estimate	Share(%)
<i>Aggregate Effects</i>	0.093*** (0.002)	62.5	0.056*** (0.007)	37.5
Intercept	-	-	-0.001 (0.064)	-0.5
Land Owned (hectares)	0.017*** (0.001)	11.6	-0.002 (0.002)	-1.4
Age	0.004*** (0.000)	2.5	-0.003 (0.056)	-1.7
Household Size	-0.015*** (0.000)	-9.9	0.027 (0.019)	18
Education	0.035*** (0.001)	23.2	0.006 (0.004)	3.8
Occupation	0.052*** (0.002)	35.1	0.029 (0.021)	19.1

*With State Dummy Variables*

<i>Aggregate Effects</i>	0.087*** (0.002)	58.3	0.062*** (0.006)	41.7
Intercept	-	-	0.044 (0.059)	29.2
Land Owned (hectares)	0.012*** (0.001)	8.1	0.000 (0.002)	0.3
Age	0.004*** (0.000)	2.6	-0.008 (0.053)	-5.3
Household Size	-0.014*** (0.000)	-9.5	0.046** (0.019)	30.4
Education	0.037*** (0.001)	24.7	0.007** (0.003)	4.8
Occupation	0.049*** (0.002)	32.9	0.009 (0.019)	5.8
Overall Effect of State Dummy Variables	-0.001 (0.001)	-0.6	-0.035*** (0.013)	-23.5

Note: a) Share is percentage share of difference in the probability of being poor. b) standard errors in parentheses. c) \*\*\*, \*\*, and \* denotes significance at the 1, 5 and 10 percent respectively. d) Observed poverty incidence gap is 16.1% (=49.2% - 33.1%). However, the share is calculated as a proportion to predicted poverty incidence gap of 14.9% (=47.3% - 32.3%).

Source: 50th round (1993/94) of the consumer expenditure survey of the NSS; our calculations.



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

*Without State Dummy Variables*

	Characteristics Effect		Coefficients Effect	
	Estimate	Share(%)	Estimate	Share(%)
<i>Aggregate Effects</i>	<i>0.064***</i> (0.002)	39.3	<i>0.098***</i> (0.012)	60.7
Intercept	-	-	-0.032 (0.083)	-19.9
Land Owned (hectares)	0.000*** (0.000)	0	0.013** (0.005)	7.7
Age	0.005*** (0.000)	3.4	0.043 (0.068)	26.3
Household Size	-0.009*** (0.000)	-5.7	0.030 (0.038)	18.5
Education	0.038*** (0.001)	23.5	-0.001 (0.004)	-0.8
Occupation	0.029*** (0.001)	18.1	0.047 (0.030)	29

*With State Dummy Variables*

<i>Aggregate Effects</i>	<i>0.054***</i> (0.004)	33	<i>0.109***</i> (0.010)	67
Intercept	-	-	-0.265*** (0.097)	-163
Land Owned (hectares)	0.000*** (0.000)	0	0.003 (0.004)	1.8
Age	0.006*** (0.000)	3.4	0.067 (0.064)	41.3
Household Size	-0.009*** (0.000)	-5.3	0.050 (0.036)	30.7
Education	0.039*** (0.001)	24.2	0.002 (0.005)	1.1
Occupation	0.026*** (0.001)	16.2	0.058* (0.032)	35.7
Overall Effect of State Dummy Variables	-0.009*** (0.003)	-5.5	0.194*** (0.061)	119.3

Note: a) Share is percentage share of difference in the probability of being poor. b) standard errors in parentheses. c) \*\*\*, \*\*, and \* denotes significance at the 1, 5 and 10 percent respectively. d) Observed poverty incidence gap is 17.2% (=50.3% - 33.1%). However, the share is calculated as a proportion to predicted poverty incidence gap of 16.2% (=48.5% - 32.3%).

Source: 50th round (1993/94) of the consumer expenditure survey of the NSS; our calculations.