Review of Income and Wealth
Series 54, Number 1, March 2008

POVERTY IN RURAL INDIA: CASTE AND TRIBE

BY IRA N. GANG

Department of Economics, Rutgers University

KUNAL SEN

IDPM, University of Manchester

AND

MYEONG-SU YUN*

Department of Economics, Tulane University

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, differences in the transformed regression coefficients play the more important role.

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 30.2 percent in 1999/2000 (Deaton, 2003a). Particularly troubling is the concentration of rural poverty in India in the "scheduled caste" (SC) and "scheduled tribe" (ST) populations.[1] The presence of such disparity in the incidence of poverty and widespread discrimination against scheduled groups have long histo-

Note: This manuscript has benefited from discussions with Shubhashis Gangopadhyay and Ajit Mishra, and from comments by the two referees and 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.

*Correspondence to: Myeong-Su Yun, Department of Economics, Tulane University, 206 Tilton Hall, New Orleans, LA 70118, USA (msyun@tulane.edu).

[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 exclusion 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, and West Bengal, and in North-Eastern India.

ries in India. Affirmative action programs have been at the core of Indian social policy directed toward scheduled groups.

Our focus is on rural poverty. Most of India's poor live in rural areas. Also, we can observe classification of SC, ST and non-scheduled households more clearly in rural areas. The data we use (discussed below) classify 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).

According to the 2001 Census of India, scheduled castes and tribes comprise 16.2 percent and 8.2 percent, respectively, of India's population, yet 47.3 percent of India's rural poor are concentrated in these groups.[2] The incidence of poverty among scheduled caste and tribe households is much higher than for the rest of the population—in 1999/2000 the proportion of rural SC and ST households below the poverty line were 30.1 and 39.4 percent respectively, as compared with a poverty rate of 17.7 percent for rural non-scheduled households. From Table 1 we see a gap in the proportion living in poverty (a poverty incidence gap) of 12.4 percent (= 30.1 – 17.7) between SC and non-scheduled households, and a poverty incidence gap of 21.7 percent (= 39.4 – 17.7) between ST and non-scheduled households.

We study the causes of higher poverty amongst SC and ST households compared with nonscheduled households. We ask 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 first examine the determinants of poverty for scheduled households, SC and ST, and non-scheduled households, and implement an Oaxaca-type decomposition methodology that allows us to examine causes of the disparity in poverty incidence.

We use rural household survey data on 67,942 households from the 55th round of the National Sample Survey (NSS). We estimate regression equations where the dependent variable is the natural logarithm of the ratio of (monthly) per capita expenditure to the poverty line, following an approach suggested in Coudouel *et al.* (2002), later referred to as the World Bank approach. 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 differences in the incidence of poverty in terms of

[2]These estimates are from the unit record data provided in the National Sample Survey's 55th round of the consumer expenditure survey. More details of the computations are provided in the next section. These calculations used the official poverty lines from the Indian Planning Commission. Using alternative Deaton–Tarozzi (DT) poverty lines, available for a subset of States and Union Territories, scheduled groups comprise 48.6 percent of India's rural poor. We discuss the choice of poverty lines below.

TABLE 1
POVERTY RATES

	Scheduled Castes	Scheduled Tribes	Non-Scheduled	All
Overall	30.1	39.4	17.7	22.7
Age				
20–29	26.7	36.6	17.9	22.5
30–39	36.9	45.6	22.7	28.7
40–49	30.5	39.5	17.1	22.4
50–59	23.2	34.0	14.3	18.0
60–70	26.4	33.5	14.0	17.9
Household size				
1	7.6	11.3	5.0	6.1
2	12.1	18.6	5.1	8.4
3	18.4	26.6	8.6	12.9
4	24.4	33.6	12.9	17.4
5	33.2	42.7	17.5	23.9
6	38.0	49.5	25.1	30.7
7 or more	47.6	56.0	29.0	35.4
Education				
Not literate	33.4	45.3	24.0	29.5
Literate, below primary	27.8	34.7	17.4	21.2
Literate, below secondary	24.1	27.3	13.4	16.2
Literate, secondary	23.0	14.7	7.9	10.1
Literate, higher secondary and above	13.6	15.9	5.3	6.8
Occupation				
Self-employed in non-agriculture	27.7	34.1	16.3	19.3
Self-employed in agriculture	23.1	32.4	13.2	16.5
Agricultural labor	34.9	48.7	28.6	33.5
Non-agricultural labor	27.1	38.0	17.8	22.5
Others	18.4	19.2	9.9	11.7

Notes: Observations are weighted by the multipliers assigned to each household in the unit record datafile.

Source: 55th round (1999/2000) of the consumer expenditure survey of the NSS.

differences in characteristics (characteristics effect) and differences in the "transformed regression" coefficients (coefficients effect).

Interpreting these two effects is always difficult and controversial as shown in studies decomposing wage differentials. The popular interpretation is that the characteristics effect is not due to discrimination while the coefficients effect represents an outcome of unequal treatment by society (discrimination). 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. On the other hand, it is not clear that discrimination is the only source for the existence of the coefficients effect. For example, educational quality may differ between scheduled and non-scheduled households for reasons not due to discrimination. Hence, the differences in the coefficients on education may also capture differences in the education quality between scheduled and nonscheduled in addition to capturing discrimination. Therefore, our interpretation is that 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. These caveats should be kept in mind in interpreting decomposition results.

Four recent studies—Bhaumik and Chakrabarty (2006), Kijima (2006), Borooah (2005) and Gang *et al.* (2002)—investigate the living standards of the SC, ST and non-scheduled in India.[3] 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.[4] Gang *et al.* (2002) and Borooah (2005) examine poverty incidence. While Gang *et al.* (2002) estimate probit model of head count ratio using an older round (50th) of NSS, 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 (2005) then decomposes poverty incidence into aggregate characteristics and coefficients effects using an extension of the Nielsen (1998) methodology applied to discrete choice models with multiple outcomes.

Our study adds to this literature by utilizing an alternative estimation strategy, formulating the characteristics and coefficients effects at a highly disaggregated level in a way that is consistent with calculations at the aggregate level, and by focusing attention on rural poverty. 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 of the gap is related to the effectiveness of the education attainment differing between the scheduled and the non-scheduled groups. We describe our estimation method and our decomposition approach in detail below.

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 socio-economic variables on poverty. Section 4 employs decomposition analysis using transformed regression coefficients to examine and explain the poverty incidence gaps between scheduled and nonscheduled 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 55th 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 1999 to June 2000. The NSS data is a cross-section of a geographically distributed random sample of house-

[3]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.

[4]Bhaumik and Chakrabarty (2006) use individual level data on earnings from the employment and unemployment surveys of the NSS in their decomposition exercise. Kijima (2006) 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.

holds. 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.[5]

We estimate the incidence of rural poverty across all three social groups, and relate this 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. In this paper, we use the official poverty lines provided by the Indian Planning Commission. These are available for all States and Union Territories in India, based on actual price data for individual items obtained from the Consumer Price Surveys undertaken by the Central Statistical Organization, India, and are estimated at the state level for rural and urban households separately along with the use of weights from the NSS expenditure surveys (Government of India, 1993).[6] A 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 (2005) have proposed an alternate set of poverty lines based on unit values and quantities consumed, obtained from the NSS expenditure surveys themselves (the poverty lines for 1999–2000 are available in Deaton, 2003b). These price indexes have the advantage of allowing for substitution among goods as households adapt to relative price changes over time. However, 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 percent of ST households in our sample are found. We will use the Deaton–Tarozzi poverty lines as a robustness test of the validity of our results obtained from using official poverty lines.[7]

[5]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.

[6]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 2,400 calories per capita.

[7]A further issue in the calculation of poverty rates is that unadjusted per capita expenditure as we have used in the computation of the poverty rates in Table 1 may be problematic if economies of scale are present in consumption. The use of unadjusted per capita expenditure is equivalent to assuming that there are no economies of scale in consumption. Drèze and Srinivasan (1997) provide a simple way to test whether the relaxation of the assumption of zero economies of scale can lead to significant changes in poverty incidence gaps between different social groups. Following their procedure, we define scale-adjusted per capita expenditure (say, y^*) for a household of size n, as $y^* \equiv Y/n^{\theta}$, where θ is a parameter between zero and one and captures the extent of scale economies in consumption, with lower values indicating greater economies of scale. We then compute poverty rates for the SC, ST and non-scheduled households for different values of θ where a household was considered poor if y^* fell below a pre-specified threshold $z(\theta) \equiv z(1)m^{1-\theta}$, where m is the mean household size, and $z(1)$ is the official poverty line. We find that when we relax the assumption that there are no economies of scale in consumption, the poverty incidence gaps between SC and ST households on one hand, and non-scheduled households on the other are a little larger than in the baseline case of zero economies of scale in consumption ($\theta = 1$). The largest poverty incidence gap of 16.2 percent between SC and non-scheduled households is when θ is 0.1. The largest poverty incidence gap of 24.4 percent between ST and non-scheduled households is when θ is 0.5. We find no rank reversals in poverty rates between social groups, and the large poverty incidence gaps between SC and ST households on one hand and non-scheduled households on the other, remain even under the assumption of some economies of scale in consumption.

TABLE 2
Sample Means

	Scheduled Castes		Scheduled Tribes		Non-Scheduled	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Poverty rates	0.301	0.459	0.394	0.489	0.177	0.381
Monthly per capita expenditure	449.965	345.939	412.587	221.443	562.328	375.220
Ratio of monthly per capita expenditure to the poverty line	1.381	1.013	1.271	0.664	1.724	1.136
Demographic and other control variables						
Age	42.657	12.214	41.998	11.786	44.293	12.443
Household size	4.884	2.270	4.966	2.283	5.136	2.590
Land owned (hectares)	0.313	0.806	0.817	1.296	0.826	1.833
Education variables						
Not literate	0.617	0.486	0.653	0.476	0.440	0.496
Literate, below primary	0.144	0.351	0.137	0.344	0.154	0.361
Literate, below secondary	0.171	0.376	0.152	0.359	0.252	0.434
Literate, secondary	0.040	0.197	0.028	0.164	0.084	0.277
Literate, higher secondary and above	0.028	0.165	0.029	0.169	0.071	0.256
Occupation variables						
Self-employed in non-agriculture	0.121	0.327	0.053	0.225	0.153	0.360
Self-employed in agriculture	0.164	0.371	0.367	0.482	0.380	0.485
Agricultural labor	0.540	0.498	0.441	0.497	0.266	0.442
Non-agricultural labor	0.096	0.295	0.081	0.272	0.072	0.258
Others	0.078	0.269	0.057	0.232	0.130	0.336
Number of observations	12,325		9,713		45,904	

Notes: Sample means are calculated using the individual household multiplier.
Source: 55th round (1999/2000) of the consumer expenditure survey of the NSS.

The poverty rates by social group and by age, household size, educational level and occupation are presented in Table 1.[8] We observe that there is a non-linear relationship between age and poverty incidence across all three social groups, with the poverty rate 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 to ST and non-scheduled households. Approximately 19 percent of SC households with literacy levels of secondary and above are poor as compared to 15 percent of similarly educated ST households and 7 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.

Table 2 shows the mean characteristics of the sample households in our study. Considering the demographic characteristics of the three groups of households, we

[8]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.

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 is 4.88 and 4.97 respectively, compared with a mean household size of 5.14 for non-scheduled households.

A much higher proportion of SC and ST households are not literate (61.7 percent and 65.3 percent, respectively), compared with non-scheduled households (44 percent). With respect to occupation, 12 percent of SC households are self-employed in non-agriculture, 54 percent are agricultural laborers, 10 percent are non-agricultural laborers, 16 percent are self-employed in agriculture, and 8 percent are classified in a residual category termed "others." For ST households, 5 percent are self-employed in non-agriculture, 44 percent are agricultural laborers, 8 percent are non-agricultural laborers, 37 percent are self-employed in agriculture, and 6 percent are in other occupations. Finally, for non-scheduled households, 15 percent are self-employed in non-agriculture, 27 percent are agricultural laborers, 7 percent are non-agricultural laborers, 38 percent are selfemployed in agriculture, and 13 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 the World Bank approach proposed in Coudouel *et al.* (2002) to understand why households are in poverty.[9] According to the World Bank approach, 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 equation is $\log R = X\beta + e$, where R, X, and β 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 $Pr\ (\log R < 0)$; 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 σ 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. We now discuss the specification of our regression equation, which we estimate using maximum likelihood for households in the non-scheduled group, SC and ST, separately.[10] We also discuss the implications of the estimated coefficients on the likelihood of being in poverty.

[9]See Drèze and Srinivasan (1997) for a similar analysis of poverty incidence among widow-headed households in India. Gang *et al.* (2002) employed probit analysis to examine poverty incidence.

[10]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 (β, σ) which is used to compute the covariance matrix for $\tilde{\beta}$, allowing us to perform significance tests for the decomposition equation.

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).[11]

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.[12] 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 significant differences in rural poverty rates across 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.[13] Furthermore, there is non-negligible variation in rural poverty rates *within* Indian states across NSS regions, and these variations are crudely associated with differences in agro-ecological conditions which may be vastly different within a state, parts of which may be more similar to those prevailing in geographically contiguous states (Palmer-Jones and Sen, 2003).[14] The omission of state and NSS region dummy variables to capture the location of the household may bias the results if the SC and ST households are mostly residing in Indian states and NSS regions

[11]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 percent of household income. The category "others" includes those where no one income source exceeds 50 percent or more of total income. Thus, the households in this category have very diversified income sources or more than one earning member.

[12]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 alternative assumptions about equivalence scales (Drèze 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.

[13]There are 32 States and Union Territories in the 55th round of the NSS consumer expenditure survey.

[14]NSS regions are groupings of contiguous districts within states. There are 82 regions in the 55th round of the NSS consumer expenditure survey.

where higher poverty is observed, and if this higher incidence of poverty is due to state-level and sub-state NSS regionlevel factors exogenous to the household such as the nature of state-level public policies toward poorer households or agro-climactic factors. We present our results with the inclusion of both state and NSS region fixed effects where these fixed effects are included separately.[15]

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 with the inclusion of state dummy variables.[16] Columns (2), (4) and (6) contain the results with the inclusion of NSS region dummy variables.[17] 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.[18]

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 non-scheduled households when compared with SC and ST 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 "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 households are more likely to be poor among all occupational groups, controlling for other determinants. ST households who are self-employed in agriculture have a much higher incidence of poverty than SC and non-scheduled households in the same occupational category. 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

[15]Van de Walle and Gunewardena (2001) also raise the possibility of geographical fixed effects operating at the level of the survey cluster in which the household is found. In the NSS dataset the survey clusters for rural households are 5,997 villages. Though not reported, we also estimate the model in Table 3 with village fixed effects and obtain results similar to those we present. However, the main purpose of this paper is explaining sources of differences in poverty incidence between SC/ST and non-scheduled households. This kind of explanation typically relies on Oaxaca-type decomposition which requires having identical explanatory variables for each comparison group. Because of the NSS survey design, the average village consists of a small number of households, hence 86 percent of the villages in our sample do not include at least one of the three social groups. Thus, the inclusion of village fixed effects will render the decomposition exercise to a large extent meaningless as it requires a large reduction in our sample so as to include only those villages where all three groups are surveyed. For these reasons, we confine our discussion of the results to the cases where state and NSS region dummy variables are included. We do, however, allow for the possible presence of unobserved village-level effects in the error term in the estimated equations originating from the stratified nature of the sampling design by using a cluster-correlated robust estimate of the variance–covariance matrix within each village (Wooldridge, 2002, pp. 409–10).

[16]We omit Lakshadweep as all three social groups are not present in this union territory.

[17]Along with the union territory of Lakshadweep, we omit region 5 in Bihar, region 3 in Punjab, and region 5 in West Bengal as all three social groups are not present in these regions.

[18]This likelihood ratio test supports our approach of studying SC and ST separately.

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	0.643***	0.658***	0.636***	0.518***	0.611***	0.696***
	(0.062)	(0.062)	(0.085)	(0.094)	(0.041)	(0.042)
Demographic control variables						
Age	0.013***	0.014***	0.012***	0.012***	0.015***	0.014***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)
Age square	–0.011***	–0.012***	–0.009**	–0.009***	–0.011***	–0.011***
	(0.003)	(0.003)	(0.004)	(0.003)	(0.002)	(0.002)
Household size	–0.146***	–0.147***	–0.138***	–0.138***	–0.129***	–0.128***
	(0.006)	(0.006)	(0.011)	(0.009)	(0.004)	(0.004)
Household size squared	0.006***	0.006***	0.005***	0.005***	0.005***	0.005***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)
Land owned (hectares)	0.034***	0.036***	0.034***	0.034***	0.030***	0.032***
	(0.006)	(0.006)	(0.006)	(0.005)	(0.002)	(0.002)
Education variables—reference group: "not literate"						
Literate, below primary	0.080***	0.069***	0.095***	0.090***	0.084***	0.076***
	(0.012)	(0.011)	(0.015)	(0.014)	(0.007)	(0.007)
Literate, below secondary	0.124***	0.124***	0.160***	0.146***	0.163***	0.160***
	(0.012)	(0.011)	(0.016)	(0.015)	(0.007)	(0.006)
Literate, secondary	0.208***	0.207***	0.275***	0.238***	0.291***	0.283***
	(0.022)	(0.021)	(0.034)	(0.032)	(0.010)	(0.010)
Literate, higher secondary and above	0.338***	0.340***	0.348***	0.324***	0.422***	0.420***
	(0.033)	(0.032)	(0.035)	(0.035)	(0.014)	(0.014)
Occupation variables—reference group: "others"						
Self-employed in non-agriculture	–0.081***	–0.080***	–0.110***	–0.112***	–0.056***	–0.053***
	(0.022)	(0.022)	(0.038)	(0.035)	(0.011)	(0.010)
Self-employed in agriculture	–0.054**	–0.048**	–0.136***	–0.130***	–0.042***	–0.041***
	(0.022)	(0.021)	(0.033)	(0.030)	(0.011)	(0.010)
Agricultural labor	–0.196***	–0.192***	–0.222***	–0.227***	–0.225***	–0.214***
	(0.020)	(0.020)	(0.032)	(0.030)	(0.012)	(0.011)
Non-agricultural labor	–0.113***	–0.120***	–0.162***	–0.146***	–0.087***	–0.096***
	(0.027)	(0.025)	(0.038)	(0.037)	(0.014)	(0.013)
Standard deviation of error	0.346***	0.333***	0.331***	0.314***	0.366***	0.356***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.003)	(0.003)
State dummy variables?	Yes	No	Yes	No	Yes	No
NSS region dummy variables?	No	Yes	No	Yes	No	Yes

Notes: Observations are weighted by the individual household multiplier.

Dependent variable is the natural logarithm of the ratio of monthly per capita expenditure to the poverty line.

Standard errors in parentheses are robust to heteroskedasticity and clustered residuals within villages.

*** and **denote significance at the 1 and 5 percent levels, respectively.

Joint hypothesis test for coefficients that all state dummy variables and region dummy variables are zero is rejected by a log-likelihood test.

Source: 55th round (1999/2000) of the consumer expenditure survey of the NSS; our calculations.

more negative. With respect to demographic factors, older 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.

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 12.4 percent (= 30.1 – 17.7) for the scheduled castes versus the non-scheduled; for scheduled tribes versus the non-scheduled the observed gap is 21.7 percent (= 39.4 – 17.7). 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 11.3 percent (= 30.1 – 18.8) for the scheduled castes versus the non-scheduled; for scheduled tribes versus the non-scheduled the predicted gap is 18.9 percent (= 37.7 – 18.8). Our analysis breaks down the predicted poverty incidence gap into its components, at different levels of aggregation.[19]

In examining the sources of 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 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

[19]As noted in Coudouel *et al.* (2002), 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, the gain using estimates from the continuous variable regression may not outweigh the loss from the reduction in fit.

probability of being poor in scheduled households compared with nonscheduled 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; Fairlie, 2005). 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. nonscheduled 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 *et al.* 1998, p. 1137 for a discussion of path-dependency). 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 using the World Bank approach, discussed in 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 non-linear 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 (β) and the standard deviation of the error term (σ) for each group. By transforming the estimates to $\tilde{\beta} = -\beta/\sigma$, we can compute the probability of being in poverty as $\Phi(X\tilde{\beta})$, where Φ 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,

$$(1) \qquad \bar{P}_A - \bar{P}_B = \left[\overline{\Phi(X_A \tilde{\beta}_B)} - \overline{\Phi(X_B \tilde{\beta}_B)}\right] + \left[\overline{\Phi(X_A \tilde{\beta}_A)} - \overline{\Phi(X_A \tilde{\beta}_B)}\right],$$

where Φ is the standard normal cumulative distribution function; $\tilde{\beta}_A = -\beta_A/\sigma_A$ and $\tilde{\beta}_B = -\beta_B/\sigma_B$, β_A and β_B are sets of estimated coefficients for each group, and σ_A and σ_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);[20]

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

where

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

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

To complete the decomposition analysis, we calculate the standard errors of the components of the decomposition equation 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 (β) and the standard deviation of the error term (σ) which is used for deriving the asymptotic covariance matrix for (β/σ) using the delta method. The covariance of β/σ is, in turn, used for hypothesis testing (Yun, 2005a). In our discussion, we take account of both the size and significance of the components.

Furthermore, we deal with robustness issues, known as the index or parameterization problem and the identification problem in detailed decompositions. A decomposition equation with a different parameterization, that is, $\left[\overline{\Phi(X_A\tilde{\beta}_A)} - \overline{\Phi(X_B\tilde{\beta}_A)}\right] + \left[\overline{\Phi(X_B\tilde{\beta}_A)} - \overline{\Phi(X_B\tilde{\beta}_B)}\right]$, is possible; our results with it are not substantially different from those presented here and 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). We follow a solution suggested by Yun (2005b) that, if alternative reference groups yield different estimates of the coefficients effects for each individual variable, it is natural to obtain estimates of the coefficients effects for every possible specification of the reference groups and take the average of the estimates of the coefficients effects with various reference groups as the "true" contributions of individual variables to differentials. While appearing cumbersome, this can be accomplished with a single estimation. We can transform

[20]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.

our regression estimates into a normalized equation and use the normalized equation for our decomposition. See the Appendix for an overview of this solution.

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 break down the overall effects into smaller subgroups. We discuss the poverty incidence gap of scheduled castes relative to non-scheduled castes in Table 4, and that of scheduled tribes compared with non-scheduled tribes 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 include state and NSS region dummy variables.

We proceed by first discussing the aggregate effects and sub-aggregate effects with state dummy variables for SC households respectively (Table 4). The *Aggregate Effects* row shows the overall effects of characteristics versus coefficients in explaining differences in poverty. The top panel shows that when state fixed effects are controlled, 56.6 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 43.4 percent is explained by the differences in the transformed regression coefficients. Both aggregate characteristics and coefficients effects are significant at the 1 percent level of significance. If in both groups the various variables influencing poverty status had the same strength (their transformed coefficients had been equal), then 43.4 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, 56.6 percent of the poverty incidence gap would disappear. When we include NSS region dummy variables (bottom panel), the aggregate coefficients effect is 40.4 percent and the aggregate characteristics effect is 59.6 percent.

In Table 4 we also see the breakdown of characteristics and coefficients effects into important variable groupings. We confine our discussion of the results in the case where NSS region dummies are included; there are relatively minor differences in the results when state dummies are included.[21] We see the importance of the characteristics effect for occupation in determining the poverty incidence gap, contributing 24.7 percent. One of the salient features of the caste system is the generally undesirable and low-paying jobs scheduled castes are allowed or forced to perform. SC households generally are in less-remunerative occupations. This may confirm anthropological evidence about the lack of job choice for individuals

[21]While we do not report the disaggregated results on the individual state dummy variables, it is interesting to note that among the major Indian states included in the analysis, the combined characteristics and coefficient effects of the dummy variables for Bihar, Uttar Pradesh and West Bengal are the largest, contributing 10.0, 18.7 and 10.2 percent respectively to the poverty incidence gap between SC and non-scheduled households. This means that for these three states in particular, the poverty incidence gap between SC and non-scheduled households also depends on unobserved factors at the state level, not controlled by included variables in the decomposition analysis.

TABLE 4

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

	Characteristics Effect		Coefficients Effect	
	Estimate	Share (%)	Estimate	Share (%)
With state dummy variables				
Aggregate effects	*0.064**** *(0.002)*	*56.6*	*0.049**** *(0.006)*	*43.4*
Intercept	–	–	–0.107** (0.054)	–95.1
Land owned (hectares)	0.008*** (0.001)	7.4	–0.001 (0.002)	–1.2
Age	0.004*** (0.000)	3.9	0.048 (0.048)	42.5
Household size	–0.007*** (0.000)	–6.6	0.051*** (0.019)	44.6
Education	0.025*** (0.001)	21.8	–0.017*** (0.006)	–14.7
Occupation	0.029*** (0.001)	25.8	–0.009** (0.004)	–7.8
State dummy variables	0.005*** (0.001)	4.3	0.085*** (0.014)	75.1
With NSS region dummy variables				
Aggregate effects	*0.067**** *(0.002)*	*59.6*	*0.046**** *(0.006)*	*40.4*
Intercept	–	–	–0.049 (0.054)	–43.1
Land owned (hectares)	0.009*** (0.001)	8.1	–0.001 (0.002)	–1.3
Age	0.004*** (0.000)	3.8	0.021 (0.049)	18.3
Household size	–0.007*** (0.000)	–6.5	0.061*** (0.020)	53.8
Education	0.024*** (0.001)	21.5	–0.016** (0.006)	–13.7
Occupation	0.028*** (0.001)	24.7	–0.007* (0.004)	–6.3
NSS region dummy variables	0.009*** (0.001)	8.1	0.037*** (0.007)	32.7

Notes: Share is calculated as a proportion to the predicted poverty incidence gap of 11.3% (= 30.1% – 18.8%). Observed poverty incidence gap is 12.4% (= 30.1% – 17.7%).

Standard errors in parentheses.

***, **, and *denote significance at the 1, 5 and 10 percent levels, respectively.

Source: 55th round (1999/2000) of the consumer expenditure survey of the NSS.

belonging to scheduled castes (Srinivas, 1962; Beteille, 1965).[22] However, the coefficients effect contributes a negative 6.3 percent of the gap (not significant), which tells us that there is no significant difference in the manner SC households are being rewarded as compared to non-scheduled households for the same occupation (controlling for education and demographic characteristics). In other words, the strength of the poverty reducing effect of occupation for SC and non-scheduled households cannot explain why the incidence of poverty is higher for SC house-

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

TABLE 5

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

	Characteristics Effect		Coefficients Effect	
	Estimate	Share (%)	Estimate	Share (%)
With state dummy variables				
Aggregate effects	*0.068****	*36.2*	*0.120****	*63.8*
	(0.003)		*(0.009)*	
Intercept	–	–	–0.092	–48.7
			(0.074)	
Land owned (hectares)	0.0002***	0.1	–0.005	–2.8
	(0.000)		(0.005)	
Age	0.006***	3.1	0.037	19.5
	(0.000)		(0.070)	
Household size	–0.004***	–2.0	0.075**	39.7
	(0.000)		(0.032)	
Education	0.029***	15.2	–0.002	–0.8
	(0.001)		(0.010)	
Occupation	0.019***	10.1	0.002	1.0
	(0.001)		(0.008)	
State dummy variables	0.018***	9.7	0.106***	55.9
	(0.002)		(0.020)	
With NSS region dummy variables				
Aggregate effects	*0.092****	*48.8*	*0.096****	*51.2*
	(0.005)		*(0.009)*	
Intercept	–	–	–0.086	–45.5
			(0.075)	
Land owned (hectares)	0.0002***	0.1	–0.004	–2.4
	(0.000)		(0.005)	
Age	0.006***	3.2	0.022	11.5
	(0.000)		(0.070)	
Household size	–0.004***	–2.1	0.100***	53.4
	(0.000)		(0.032)	
Education	0.030***	15.9	–0.009	–4.5
	(0.001)		(0.010)	
Occupation	0.019***	10.3	0.008	4.4
	(0.001)		(0.008)	
NSS region dummy variables	0.041***	21.6	0.065***	34.3
	(0.004)		(0.013)	

Notes: Share is calculated as a proportion to the predicted poverty incidence gap of 18.9% (= 37.7% – 18.8%). Observed poverty incidence gap is 21.7% (= 39.4% – 17.7%).

Standard errors in parentheses.

***, **, and *denote significance at the 1, 5 and 10 percent levels, respectively.

Source: 55th round (1999/2000) of the consumer expenditure survey of the NSS.

holds. Education is remarkable in that the characteristics effect is a positive 21.5 percent while the coefficient effect is a negative 13.7 percent. SC households attain lower levels of schooling, and that puts them at greater risk of being poor.[23] However, the poverty-reducing effectiveness of education is higher for SC households than for non-scheduled households.

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

[23]This is supported by the finding of Drèze 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.

owned contributes 8 percent to the poverty incidence gap. However, there is almost no coefficients 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 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 and large. Household size differences reduce the poverty incidence gap, but differences in coefficients increase the poverty incidence gap.[24]

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 poverty rates between ST and non-scheduled households, shown in Table 5. Unlike the case of SC households, the aggregate coefficients effect is larger in magnitude than the aggregate characteristics effect in explaining the differences in poverty between ST and non-scheduled households. Taking the case where state dummies are included, approximately 36 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 36 percent less. It is interesting to note that the aggregate coefficients effect is 63.8 percent with the inclusion of state dummies but falls to 51.2 percent with the inclusion of NSS region dummies. This suggests that the location of ST households in regions with adverse agro-climactic factors (as captured by the NSS region dummies) matters to a large extent in explaining the higher incidence of poverty among ST households.[25]

Again, we confine our discussion of the detailed results to the case with the inclusion of NSS region dummy variables. Differences in educational attainment account for 15.9 percent of the poverty incidence gap. The occupational distribution explains 10.3 percent of the higher poverty among the ST households as compared to the non-scheduled households. We also find that the coefficients effects of educational attainment and occupational structure are negligible.

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, including both state and regional dummy variables, has a high positive coefficients effect and about a negative 2 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.

[24]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.

[25]As in the case of SC households, the combined characteristics and coefficient effects of a few individual states are particularly important in explaining the poverty incidence gap between ST and non-scheduled households. These states are Madhya Pradesh and Orissa, where the individual state dummy variables contribute 27.7 and 24.2 percent, respectively, to the poverty incidence gap between ST and non-scheduled households.

In order to test whether our specification is robust, we estimate the same without occupation variables, 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 the size of the aggregate coefficients (characteristics) effect increases (decreases). When we look at changes in detailed decomposition, 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.

To examine how sensitive our results are to the choice of poverty lines, we also undertook the decomposition analysis (both with state and regional dummy variables) using Deaton–Tarozzi (DT) and official poverty lines only for those states where DT poverty lines were available. With the inclusion of state dummies, for SC households the aggregate coefficients effect was 41.3 percent when we used the DT poverty lines and 40.6 percent when we used the official poverty lines, and for ST households, the aggregate coefficients effect was 50.8 percent when we used the DT poverty lines and 51.7 percent when we used the official poverty lines We obtained virtually identical results with the inclusion of regional dummies. Thus, our results are robust to the choice of the poverty line.[26]

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. Observed poverty rates of SC and ST households are 12.4 percentage points and 21.7 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 55th round of the National Sample Survey conducted in 1999–2000.

The decomposition analysis shows that for SC vs. non-scheduled households differences in characteristics explain the poverty incidence gap more than differences in coefficients, with 60 percent of the poverty incidence gap attributable to the former. When NSS region fixed effects are controlled, for ST vs. non-scheduled households, however, it is the reverse, with 51 percent of the poverty incidence gap attributable to the differences in coefficients. Thus, the causes of higher incidence of poverty in these two social groups relative to non-scheduled households are not identical.

Differences in educational attainments explain about 21 (16) percent of the poverty incidence gap for SC (ST) vs. non-scheduled households, which suggests that allocating more resources toward SC and ST children in education will decrease the discrepancy in poverty incidence between the scheduled groups and

[26]The results of the regression and the decomposition when occupation variables are omitted, and when the DT poverty lines are used instead of the official poverty lines are available from the authors upon request.

non-scheduled households. Stronger emphasis on primary education is necessary since the greatest difference between scheduled and non-scheduled households can be found in the proportion of the non-literate. 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 nonscheduled 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. The impact of the characteristics effect on the poverty incidence gap between SC households and non-scheduled households is substantially higher than 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 less likely 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 explain why, for such households, locational disadvantage rather than the type of occupation that they are in, explains much of the poverty incidence gap between these households and non-scheduled households. The analysis of the paper suggests that the underlying factors for the higher incidence of poverty in the SC and ST social groups are to an appreciable extent different; policymakers need to be aware of these differences in the causes of poverty while devising policies for poverty alleviation.

Appendix: Solving the Decomposition's Invariance Issue using Normalized Regression

The coefficients effect in the detailed decomposition is not invariant to the choice of omitted groups when dummy variables are used (Oaxaca and Ransom, 1999). Yun (2005b) suggests obtaining estimates of the coefficients effects for every possible specification of the reference groups and taking their average as the "true" contributions of individual variables to wage differentials.

In practice, rather than estimate all possible specifications, a normalized equation which can identify all coefficients, including categories omitted in standard specifications, is constructed by transforming a single set of regression estimates from a standard specification. We can construct the Oaxaca (1973) decomposition equation using the normalized equation. This resolves the identification problem for the detailed decomposition.

To illustrate this approach and the derivation of the normalized equation, we suppose there are two sets of dummy variables in addition to continuous variables in the regression equation (incorporating more sets of dummy variables is trivial). The usual regression equation—suppressing the individual subscripts is, $y = \alpha + \left[\Sigma_{g=2}^{G} D_g \kappa_g + \Sigma_{t=2}^{T} Q_t \zeta_t\right] + \Sigma_{l=1}^{L} V_l \delta_l + e$, where there are two sets of categorical variables (D's and Q's) and L continuous variables (V's); the first and second sets of dummy variables (D's and Q's) have G and T categories and $G - 1$ and $T - 1$ dummy variables in the equation, respectively; without loss of generality, the reference group is the first category for each set of dummy variables, i.e. $\kappa_1 = \zeta_1 = 0$. The identification problem in the detailed Oaxaca type decomposition is that the sum of the coefficients effects for dummy variables of the D's and/or the Q's is not invariant when the reference group is changed (Oaxaca and Ransom, 1999).

As long as we obtain consistent estimates, we can manipulate these to obtain a normalized regression equation (Suits, 1984; Yun, 2005b), allowing separate identification of the intercept and coefficients of all dummy variables including the reference category (Gardeazabal and Ugidos, 2004; Yun, 2005b). The normalized regression equation is,

$$y = \left(\alpha + \bar{\kappa} + \bar{\zeta}\right) + \left[\sum\nolimits_{g=1}^{G} D_g \left(\kappa_g - \bar{\kappa}\right) + \sum\nolimits_{t=1}^{T} Q_t \left(\zeta_t - \bar{\zeta}\right)\right] + \sum\nolimits_{l=1}^{L} V_l \delta_l + e,$$

where $\bar{\kappa} = \Sigma_{g=1}^{G} \kappa_g / G$, $\bar{\zeta} = \Sigma_{t=1}^{T} \zeta_t / T$, and $\kappa_1 = \zeta_1 = 0$. The covariance of the estimates in the normalized equation is obtained using the covariance matrix of the consistent estimates for the usual regression equation. The sums of coefficients for sets of dummy variables (e.g. D's or Q's including the omitted category variable) in the normalized equation are restricted to be zero. X and β in the paper consist of (1, D's, Q's and V's), and $\left(\alpha + \bar{\kappa} + \bar{\zeta}\right.$, $\left(\kappa_g - \bar{\kappa}\right)$'s, $\left(\zeta_t - \bar{\zeta}\right)$'s, and δ's), respectively.

We test hypotheses regarding characteristics and coefficients effects the same way as we do with usual estimates and their covariance matrix. The use of the normalized equation does not change our inference for the overall characteristics and coefficients effects (Yun, 2005c).

References

Abowd, John M. and Mark R. Killingsworth, "Do Minority/White Unemployment Differences Really Exist?" *Journal of Business and Economic Statistics*, 2, 64–72, 1984.

Akerlof, George, "The Economics of Caste and the Rat Race and Other Woeful Tales," *Quarterly Journal of Economics*, 90, 600–17, 1976.

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

Bhaumik, Sumon K. and Manisha Chakrabarty, "Earnings Inequality in India: Has the Rise of Caste and Religion based Politics in India had an Impact?" IZA Discussion Paper 2008, 2006.

Borooah, Vani K., "Caste, Inequality and Poverty in India," *Review of Development Economics*, 9, 399–414, 2005.

Coudouel, Aline, Jesko S. Hentschel, and Quentin T. Wodon, "Using Linear Regressions for Analyzing the Determinants of Poverty," in Jeni Klugman (ed.), *A Sourcebook for Poverty Reduction Strategies*, Annex A.8, World Bank, 2002.

Datt, Gaurav and Martin Ravallion, "Why Have Some Indian States Done Better Than Others at Reducing Rural Poverty?" *Economica*, 65, 17–38, 1998.

Deaton, Angus, "Adjusted Indian Poverty Estimate for 1999–2000," *Economic and Political Weekly*, 38, 322–6, 2003a.

———, "Prices and Poverty in India, 1987–2000," *Economic and Political Weekly*, 38, 362–8, 2003b.

Deaton, Angus andAlessandro Tarozzi, "Prices and Poverty in India," in Angus Deaton and Valerie

Kozel (eds), *Data and Dogma: The Great Indian Poverty Debate*, Chapter 16, Macmillan India, New Delhi, 2005.
Deshpande, Ashwini, "Does Caste Still Define Disparity? A Look at Inequality in Kerala, India," *American Economic Review*, 90, 322–5, 2000.
———, "Caste at Birth? Redefining Disparity in India," *Review of Development Economics*, 5, 130–44, 2001.
Drèze, Jean and Geeta Kingdon, "School Participation in Rural India," *Review of Development Economics*, 5, 1–24. 2001.
Drèze, Jean and P. V. Srinivasan, "Widowhood and Poverty in Rural India: Some Inferences from Household Survey Data," *Journal of Development Economics*, 54, 217–34, 1997.
Even, William E. and David A. Macpherson, "The Decline of Private-Sector Unionism and the Gender Wage Gap," *Journal of Human Resources*, 28, 279–96, 1993.
Fairlie, Robert W., "An Extension of the Blinder-Oaxaca Decomposition Technique to Logit and Probit Models," *Journal of Economic and Social Measurement*, 30, 305–16, 2005.
Gang, Ira N., Kunal Sen, and Myeong-Su Yun, "Caste, Ethnicity and Poverty in Rural India," IZA Discussion Paper 629, 2002.
Gardeazabal, Javier and Arantza Ugidos, "More on Identification in Detailed Wage Decompositions," *Review of Economics and Statistics*, 86, 1034–6, 2004.
Gatade, Subhash, "Phenomenon of False Caste Certificates," *Economic and Political Weekly 40*, October 22–28, Commentary, 2005.
Government of India, *Report of the Expert Group on Estimation of Proportion and Number of Poor*, Planning Commission, New Delhi, 1993.
Ham, John C., Jan Svejnar and Katherine Terrell, "Unemployment and the Social Safety Net During Transitions to a Market Economy: Evidence from the Czech and Slovak Republics," *American Economic Review*, 88, 1117–42, 1998.
Kijima, Yoko, "Caste and Tribe Inequality: Evidence from India, 1983–1999," *Economic Development and Cultural Change*, 54, 369–404, 2006.
Krishnaji, N., "On Measuring Incidence of Undernutrition—What Is a Consumer Unit," *Economic and Political Weekly*, 16, 1509–11, 1981.
———, "Family-size, Levels of Living and Differential Mortality in Rural India—Some Paradoxes," *Economic and Political Weekly*, 19, 248–58, 1984.
Meenakshi, Jonnalagadda Venkata and Ranjan Ray, "Impact of House Size and Family Composition on Poverty in Rural India," *Journal of Policy Modeling*, 24, 539–59, 2002.
Nielsen, Helena S, "Discrimination and Detailed Decomposition in a Logit Model," *Economic Letters*, 61, 115–20, 1998.
Oaxaca, Ronald, "Male-Female Differentials in Urban Labor Markets," *International Economic Review*, 14, 693–709, 1973.
Oaxaca, Ronald and Michael R. Ransom, "Identification in Detailed Wage Decompositions," *Review of Economics and Statistics*, 81, 154–7, 1999.
Palmer-Jones, Richard. and Kunal Sen, "What has Luck Got To Do with it? A Regional Analysis of Poverty and Agricultural Growth in Rural India," *Journal of Development Studies*, 40, 1–33, 2003.
PROBE Team, *The Public Report on Basic Education in India*, Oxford University Press, Delhi, 1999.
Revankar, Ratna G., *The Indian Constitution—A Case Study of Backward Classes*, Fairleigh Dickinson University Press, Rutherford, 1971.
Srinivas, Mysore Narasimhachar, *Caste in Modern India*, Asia Publishing House, Bombay, 1962.
Suits, Daniel B., "Dummy Variables: Mechanics v. Interpretation," *Review of Economics and Statistics*, 66, 177–80, 1984.
Van de Walle, Dominique and Dileni Gunewardena, "Sources of Ethnic Inequality in Vietnam," *Journal of Development Economics*, 65, 177–207, 2001.
Wooldridge, Jeffrey, *Econometric Analysis of Cross Section and Panel Data*, The MIT Press, Cambridge, MA, 2002.
Yun, Myeong-Su, "Decomposition Differences in the First Moment," *Economics Letters*, 82, 273–8, 2004.
———, "Hypothesis Tests when Decomposing Differences in the First Moment," *Journal of Economic and Social Measurement*, 30, 295–304, 2005a.
———, "A Simple Solution to the Identification Problem in Detailed Wage Decomposition," *Economic Inquiry*, 43, 766–72, 2005b.
———, "Normalized Equation and Decomposition Analysis: Computation and Inference," IZA Discussion Paper 1822, 2005c.