



DISCUSSION PAPER SERIES

IZA DP No. 3973

A Distributional Analysis of Social Group Inequality in Rural India

Mehtabul Azam

January 2009

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

A Distributional Analysis of Social Group Inequality in Rural India

Mehtabul Azam

*Southern Methodist University
and IZA*

Discussion Paper No. 3973
January 2009

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

A Distributional Analysis of Social Group Inequality in Rural India^{*}

This paper examines the differences in welfare, as measured by per capita expenditure (PCE), between social groups in rural India across the entire welfare distribution. The paper establishes that the disadvantage suffered by two historically disadvantaged groups – Scheduled Castes (SCs) and Scheduled Tribes (STs) – is underestimated when the comparison group is Non-SCs/STs rather than general category (mostly higher castes). The ST households are the most disadvantaged followed by the SC and the Other Backward Caste households with respect to general category households, and the disadvantage exists across the entire distribution. Better covariates and better returns to those covariates contribute to the advantage of the general category households. The findings suggest that the policies to raise the human capital and strengthening the other productive assets of the SC and the ST households must remain a focus of attention besides promoting a more active labor market in rural India.

JEL Classification: C15, D63, 053

Keywords: India, social groups, inequality, quantile regression decomposition

Corresponding author:

Mehtabul Azam
Department of Economics
Southern Methodist University
Dallas, TX 75275
USA
E-mail: mazam@smu.edu

^{*} I thank Daniel L. Millimet for his comments. Any remaining errors remain mine.

1 Introduction

In this paper, we examine the differences in welfare, as measured by per capita household expenditure (PCE), between social groups in rural India across the entire distribution. This is important as we show in later sections that the differences vary across expenditure quantiles. In this case, it is important from welfare and policy perspectives to know that in which part of the distribution the differences are concentrated, and what drives those differences? This information will promote more specifically targeted policies. Specifically, we use a quantile regression decomposition technique proposed in Machado and Mata (2005, hereafter MM) to decompose the differences in distributions of PCE between households belonging to different social groups into a component that is due to differences in the distributions of the covariates and a component that is due to differences in the returns to those covariates.

Due to persistent underdevelopment, poverty and inequality have been important research areas in India. One of the forms through which inequality is manifested is through social groups. This inequality results from exclusion of certain groups from certain economic and social spheres, and the exclusion revolves around social processes and institutions that exclude, discriminate, isolate, and deprive some groups on the basis of groups' identities like caste and ethnicity. For example, recent work by Munsif and Rosenzweig (2006) shows how caste based labor markets have trapped individuals in narrow occupational categories for generations, and persist even today. In India, there are a number of such excluded groups, which constitute a sizable part of population. These include former untouchables or Scheduled Castes (SCs) and Scheduled Tribes (STs). These groups put together constituted more than 250 millions in 2001 (about 167 millions SCs, 86 millions STs), and have historically suffered in the past from exclusion in multiple spheres, which has led to their severe deprivations.

The Indian State recognized the backwardness and deprivation suffered by the SCs and the STs as back as in the 1950s. Legal safeguards were provided against discrimination in the Constitution. The government policy however, does not stop with legal safeguards against exclusion and discrimination, but goes beyond, and developed measures to provide equal opportunity and fair participation in the

economic and political process of the country. The most important of the affirmative policies has been the 'Reservation Policy' under which specific quotas are reserved in proportion to the population in government and other services, educational institutions, public housing, other public spheres, and in various democratic bodies including the Parliament, State assemblies, and panchayat institutions from district, taluk, and down to the village levels. These pro-active equal opportunity measures are thus, used to ensure proportional participation of the SCs and the STs in various public spheres, which otherwise may not have been possible due to the residual and continuing caste and untouchability based exclusion and discrimination against the SCs and STs, in some spheres if not all. (Thorat and Mallick, 2005).

Apart from these two disadvantaged groups, there exist a third group referred as the 'Other Backward Castes (OBCs)' which constitute the 'the backward classes' along with the SCs and the STs (Béteille, 2002). There were no direct provisions in the Constitution to benefit the OBCs, however, one of the directive principles of state policy said that some measures may also be adopted for the benefit of the OBCs without specifying what these measures should be and who the OBCs were.¹ So commissions were set up and since 1993, the OBCs are entitled for 27 percent reservation in public employment and higher education.²

So, broadly Indian society can be segmented into four social groups- SCs, STs, OBCs, and General. While the SCs, STs, and OBCs comprise together 'the backward classes', general category includes most of the higher castes in India. Unlike the backward classes, the members of general category do not benefit from any direct affirmative action by government.

The existence of affirmative policies for the SCs and the STs for so long has drawn a considerable amount of interest in social group inequality from the researchers as existence of any unfair differences among social groups indicate that the objectives of different existing policies to ameliorate the conditions

¹ In contrast, the actual complete listing of castes falling under the SCs and the STs is available under Article 341 and Article 342 of the Constitution of India since 1950.

² There is no political reservation for the OBCs, while recently (on April 10, 2008) Supreme Court of India upheld the government's move for initiating OBC quota in government funded educational institutions.

of these disadvantaged groups are not fulfilled yet. In addition any perceived unfair differences may have political implications, like social unrest and voting on the basis of social group identity.^{3,4}

The existing literature attempts to answer “is there exist any differences in living standards between Non-SC/ST and the SC/ST households” and if differences exist “what explains the differences”. Gang et al. (2002) decompose the poverty gap between the SC/ST and Non-SC/ST households using National Sample Survey (NSS) consumer expenditure data for year 1993-94. They find that for the SC households, differences in the characteristics explain the poverty gap more than the differences in the coefficients with 58 percent of the poverty gap attributable to the former. For the ST households, however, 59 percent of the poverty gap is attributable to the differences in the coefficients. Borooah (2005) explores the differences in mean level of income and average probability of being poor, between the SC/ST and Non-SC/ST Hindu households using data from NCAER HDI survey-1993. He uses the Oaxaca-Blinder decomposition and its extension to multinomial probability model, and finds that at least one-third of the average income/probability differences between Non-SC/ST Hindu and the SC/ST households were due to the unequal treatment of the SCs/STs. Kijima (2006b) decomposes the disparities in average per capita expenditure between the SC, ST and Non-SC/ST households using a regression based method. He finds that the component attributable to the differences in the returns to the characteristics contributes approximately half of the differentials. Also, the proportion of disparities explained by the different returns did not change much between 1983 and 1999.

However, the existing literature has two limitations. First, the existing literature focuses on averages only neglecting the remainder of the distribution. However, social groups are not homogeneous blocks- they are basically group of communities, which differ significantly within a group (Béteille, 2002). This suggests a high level of heterogeneity within a social group itself. In this case, average may hide much more than they reveal. Second, all the existing literature focuses on explaining the differences

³ Caste based politics and political parties are playing increasingly important role both at state and federal levels in India after 1990.

⁴ Some of the districts in India suffer from Naxalite movement, which is armed rebellion against official administration, mainly from members of lower sections of the society.

between the SC/ST and Non-SC/ST households. Non-SC/ST households include both households from general category and OBCs.⁵ According to NSS consumer expenditure survey, the STs and the SCs comprise 8.6 and 19.6 percent of Indian population in 2004-05, respectively, while the OBCs and general category comprise 41 and 30.8 percent of the Indian population, respectively.⁶ As the OBCs constitute a very large category and contain castes not very far from the SCs/STs in terms of social and economic backwardness, the disparities between the SC/ST and Non-SC/ST households may understate the gap between the bottom and top tiers of caste hierarchy. Comparison of the SCs and the STs with general category should be true measure of disadvantage suffered by these two disadvantaged groups.

We attempt to address both these issues using the most recent available NSS consumer expenditure data. We focus on the differences in the distributions of welfare between social groups in rural India, as almost 90 percent of the STs and 80 percent of the SCs and the OBCs live in rural India (Table 1).⁷ Since labor market in rural India is very thin and self-employment in agricultural or informal sectors is the primary source of livelihood for most households, the focus is on a broader measure of living standard. We measure welfare by PCE and conduct the analysis at the household level. Consumption expenditure is considered to be a better measure of well-being over a relatively longer time period than income since consumption tends to be smoothed against income fluctuations (Deaton, 1997). Expenditure is also often preferred over income as a measure of the current standard of living in agricultural economies, because of measurement errors in income earning data (Walle and Gunewardena, 2001).

We advance the understanding of social group inequality in India in following ways. First, we distinguish households belonging to the OBCs from households belonging to general category, and compare the SC and the ST households with general category households. In addition, we also examine

⁵ One possible reason for this kind of focus is limitation put by data prior to 1999-00, which do not distinguish between general category and OBCs. National Sample Surveys and National Family Health Surveys, two major datasets used by researchers, started distinguishing OBCs and general category from 1999-00.

⁶ Census data does not provide information about the OBCs. According to Census of India 2001, the STs and the SCs comprise 8.2 percent and 16.2 percent of India's total population.

⁷ According to Census of India 2001, 92 percent of the STs and 80 percent of the SCs live in rural India.

the differences in welfare between the OBC and general category households. Second, unlike previous literature, we examine the differences over the entire distribution. Our quantile regression framework allows for the covariates to have marginal effects (returns) that vary with household's position in the welfare distributions. Mean regression cannot reveal such variation. Third, we decompose the differences in the PCE distributions using the MM (2005) method into a part attributable to differences in the covariates (covariate effect) and a part attributable to differences in the returns to those covariates (coefficient effect). The advantage of the MM (2005) method is that it captures the heterogeneity in the coefficient and the covariate effects across the entire distribution.

The findings of the paper are following. First, the disadvantage suffered by the SC and ST households is underestimated if one compares these two disadvantaged groups with Non-SC/ST households because of presence of the OBC households. The OBC households are themselves disadvantaged compared with general category households across the entire distribution. Second, the ST households are the most disadvantaged, followed by the SC and the OBC households compared to general category households. For the SC and the OBC households, the disadvantage increases with the household position in the PCE distribution. But, for the ST households, the disadvantage is almost uniform over most of the lower half of the distribution before it increases with quantiles in upper half of the distribution. Third, for the differences in PCE between general category and OBC households, the contribution of the covariates (coefficient) is larger than the contribution of the coefficient (covariate) below the median (above the median). For the differences between general category and SC households, the covariate effect dominates over most of the distribution. For difference between general category and ST households, the coefficient (covariate) effect dominates in lower (upper) half of the distribution.

The findings bring out the startling disadvantage of the STs compared to general category. The findings suggest that the policies to raise the human capital and strengthening the other productive assets of households belonging to the backward classes (especially the STs and SCs) must remain a focus of attention. However, what is less certain is that whether this will be sufficient to eliminate the

disadvantage suffered by these two groups. Geographical isolation of the STs may be one major reason for different returns to the ST households; however, discrimination cannot be overruled. Moreover, for the SCs, the geographical isolation is not an important issue. It is important to promote a more active labor market in rural India as the market mechanism generally pushes towards similar returns to productive characteristics.

The remainder of the paper is organized as follows: Section 2 deals with the empirical strategy. Section 3 describes the data, Section 4 investigates the results, and Section 5 discusses the findings and concludes with some observations about future work.

2 Empirical Strategy

2.1 Quantile Regression

To begin, assume household welfare is a function of household level endowments and characteristics. We investigate how the relationship between log PCE and household characteristics differs between different social groups at various quantiles of the PCE distribution. We do this by estimating quantile regression in the form⁸

$$Q_{\theta}[y|X, SC, ST, OBC] = \alpha^0(\theta) + X.\beta(\theta) + SC.\alpha^{SC} + X.\beta^{SC}(\theta) + ST.\alpha^{ST} + X.\beta^{ST} + OBC.\alpha^{OBC} + X.\beta^{OBC}(\theta) + \varepsilon(\theta) \quad (1)$$

where y is log PCE, $Q_{\theta}[y|X, SC, ST, OBC]$ is the conditional θ^{th} quantile of y , $\alpha^0(\theta)$ is the regression intercept, X is the covariate matrix, and SC, ST and OBC are dummies for the SCs, STs and OBCs, respectively. $\alpha^{SC}(\theta), \alpha^{ST}(\theta), \alpha^{OBC}(\theta)$ and $\beta^{SC}(\theta), \beta^{ST}(\theta), \beta^{OBC}(\theta)$ represent quantile intercepts and slope differentials for SCs, STs and OBCs relative to general category.

⁸ See Koenker and Bassett (1978), Buchinsky (1998) for details about quantile regressions.

2.2 Decomposition of difference in PCE

Before discussing how to decompose the differences in PCE distributions of general category (GC) households and households belonging any of the three backward classes ($g = SC, ST, OBC$) at each percentile, it is useful to discuss the familiar case of mean where the standard Oaxaca-Blinder decomposition can easily be used. The (log) PCE for group τ depends on x :

$$y_\tau = x_\tau \beta_\tau + \varepsilon_\tau, \quad \tau = GC, g \quad (2)$$

where $x = (1, X)$.

Under the usual assumption that the error term ε_τ has a conditional mean of zero given the covariates x_τ , β_{GC} and β_g can be consistently estimated using OLS. The mean difference in PCE between general category and any of three disadvantaged group can be decomposed as:

$$\Delta = \overline{y_{GC}} - \overline{y_g} = \underbrace{\left\{ \frac{(\overline{x_{GC}} - \overline{x_g}) \beta_{GC}}{\text{Covariate effect}} \right\}}_{\text{Covariate effect}} + \underbrace{\left\{ \frac{\overline{x_g} (\beta_{GC} - \beta_g)}{\text{Coefficient effect}} \right\}}_{\text{Coefficient effect}} \quad (3)$$

where $\overline{y_{GC}}$ and $\overline{y_g}$ are the mean PCE by households belonging to general category and group g , respectively, and $\overline{x_{GC}}$ and $\overline{x_g}$ are the corresponding mean values of explanatory variables.

Next we move to decompose the difference in PCE at each percentile. For this, we use the MM (2005) method to decompose the differences in the log PCE distributions between general category households and SC, ST and OBC households, respectively. The MM method involves estimating a quantile regression for log PCE for each social group separately. We model the conditional quantile as:

$$Q_\theta(y|x) = x'_\tau \beta_\tau(\theta), \quad \tau = GC, SC, ST, OBC \quad (4)$$

where $\beta_\tau(\theta)$ is a vector of the quantile regression (QR) coefficients for social group τ .

Having fitted the conditional quantile function for each social group, the estimated parameters can

be used to simulate the conditional distribution of y given x via an application of the probability integral transformation theorem: If θ is a uniform random variable on $[0,1]$, then $F^{-1}(\theta)$ has the distribution F . Thus, if $\theta_1, \theta_2, \dots, \theta_m$ are drawn from a uniform $(0, 1)$ distribution, the corresponding m estimates of the conditional quantiles of log PCE at x_i , $\hat{y}_i = \{x'_i \beta(\theta)\}_{i=1}^m$, constitute a random sample from the (estimated) conditional distribution of log PCE given x . To 'integrate x out' and get a sample from the marginal distribution, instead of keeping x fixed at a given value, a random sample can be drawn from appropriate distribution. The algorithm below summarizes the procedure:

Let $y(\tau)$, $x(\tau)$, $\tau = GC, SC, ST, OBC$, denote log PCE and the m covariates of group τ . In addition, let $g(x; \tau)$ be the joint density of the covariates for group τ . We want to generate a random sample from the log PCE density that would prevail for group τ if the conditional fitted model (4) were true and the covariates were distributed as $g(x; \tau)$. To get this

1. Generate a random sample of size m from a uniform $(0, 1)$: $\theta_1, \theta_2, \dots, \theta_m$
2. For the data set for group τ (denoted by $x(\tau)$, a $n_\tau \times k$ matrix of data on the covariates), and each $\{\theta_i\}$ estimate

$$Q_{\theta_i}(y|x; \tau)$$

which yields m estimates of the QR coefficients $\hat{\beta}(\theta)$.

3. Generate a random sample of size m with replacement from the rows of the covariate matrix $x(\tau)$: $\{x_i^*(\tau), i = 1, 2, 3, \dots, m\}$.
4. Finally

$$\{y_i^*(\tau) = x_i^*(\tau)' \hat{\beta}(\theta_i)\}_{i=1}^m$$

is a random sample of size m from the desired distribution.

However, instead of drawing a random sample of size m from uniform distribution and estimating $\hat{\beta}(\theta)$ for each θ_i , we estimated $\hat{\beta}(\theta)$ on a grid of $\theta = [0.001, 0.002, \dots, 0.999]$. Then, we draw $m = 1000$ random draws from the distribution of covariates for each $\hat{\beta}$ and stack $x' \hat{\beta}$ to get the desired

distribution.^{9,10} In practice, this procedure yields the same estimate and removes the sampling error from the first step of MM procedure (Albrecht et al., 2007).

To construct the counterfactual density, i.e., the density function of log PCE for general category corresponding to the any of the disadvantaged group's distribution of covariates, we follow the algorithm above for general category, but in the third step, instead of drawing the sample from general category covariate matrix, $x(\tau = GC)$, we sample it from the rows of the disadvantaged group covariate matrix, $x(\tau = g)$, $g = SC, ST, OBC$. After obtaining the desired counterfactual densities we can decompose the overall difference in log PCE between two groups into a part attributable to the coefficients, another to the covariates, and a residual.

Let $f(y(GC))$ denote the estimate of the marginal density of log PCE for general category based on observed sample, i.e., the empirical density, and $f^*(y(GC))$ denote an estimate of the density of log PCE for general category based on the generated sample, i.e., the marginal implied by the model. Extending this notation to the counterfactual distributions, we may have $f^*(y(GC); x(g))$ - density that would have prevailed if all covariates would have been distributed as in group $g(= SC, ST, OBC)$ and the prices paid are as in general category.

Let α be a usual summary statistic (for instance, quantile or scale measure), we may decompose differences in log PCE as:

$$\begin{aligned}
 & \alpha\{f(y(GC))\} - \alpha\{f(y(g))\} \\
 &= \alpha\{f^*(y(GC))\} - \alpha\{f^*(y(g))\} + residual \\
 &= \alpha\{f^*(y(GC))\} + \alpha\{f^*(y(GC); x(g))\} - \alpha\{f^*(y(GC); x(g))\} - \alpha\{f^*(y(g))\} + residual \\
 &= \underbrace{\alpha\{f^*(y(GC))\} - \alpha\{f^*(y(GC); x(g))\}}_{Covariate\ Effect} + \underbrace{\alpha\{f^*(y(GC); x(g))\} - \alpha\{f^*(y(g))\}}_{Coefficient\ Effect} + residual^{11}
 \end{aligned}$$

⁹ To take into account the household survey weights, we implemented unequal probability sampling with replacement.

¹⁰ We end up with 990,000 observations on y_i^* .

¹¹ There is another possible counterfactual which can be used in the decomposition, i.e., $f^*(y(g); x(GC))$ - density that would have prevailed if all covariates were distributed as in general category and prices being paid as in group g . It is well known that the decomposition results may not be invariant with respect to the choice of the involved counterfactual (Oaxaca and Ransom, 1994). Therefore, the choice of a counterfactual should be guided by the

This decomposition will then give us the contribution of the coefficients, the covariates and an unexplained part (residual).

3 Data

The analysis is based on household-level data from the Consumer Expenditure Schedule conducted by the National Sample Survey Organization (NSSO), Government of India. Data from 61st Round (2004-05) is used. The advantage of the data is that it is the most recent available quinquennial round survey and contains information whether a household belongs to OBCs.¹² The data contain information on household size and composition, social group, religion, monthly consumption, landholdings, demographic variables (age, gender, and marital status), and educational attainment of household members. The data covers 79,298 households residing in rural India. The sample of households is drawn based on a stratified random sampling procedure and all the analysis is done using survey weights.

In the data, households are divided into four mutually exclusive groups - SCs, STs, OBCs and others. Others are general category who do not benefit from any affirmative action from the government, and include most of the higher castes in India.

Our dependent variable is log of per capita expenditure, while the independent variables (the covariate matrix X) include household characteristics, household human capital measures, land cultivated by household, household main occupation and dummies for states. The household characteristics include household size, proportion of members below age 14 and proportion of adult (age 21-60) male and adult female in household.¹³ Household human capital is measured as a series of dummy variables for the education level achieved by the household head. Thus, we include dummy

question of economic interest. We believe that the prices paid to general category are the non-discriminatory prices, hence we chose the first counterfactual, i.e., $f^*(y(GC); x(g))$ for our analysis.

¹² NSSO conducts thick round surveys at five-year intervals (called 'Quinquennial Rounds'). The previous quinquennial rounds were conducted in 1972-73, 1977-78, 1983, 1987-88, 1993-94, 1999-00. The NSSO data started distinguishing the OBCs in the quinquennial round conducted in 1999-00.

¹³ It is illegal to employ a child below the age 14, although child labor is not very uncommon in India.

variables corresponding to primary, middle, secondary, higher secondary, graduate and postgraduate levels of education (reference group is households whose head is illiterate/below primary). For occupation, we include dummies for self employed in non agriculture, self employed in agriculture, other labor and others (with reference group being agricultural laborer). 'Others' covers households whose major source of income arises mostly from contractual employment with regular wages and salaries.

Table 3 presents the descriptive statistics of variables used. The unconditional means from our data establish that the 'backward classes' do indeed have lower standards of living on average than general category. The general category households also have better educational attainment than the households belonging to the 'backward classes'. Incidence of poverty is highest among the STs, followed by the SCs and the OBCs. The general category has the lowest poverty incidence (Table 2). In addition, the poor belonging to the 'backward classes' suffer from a much higher poverty gap compared to the poor belonging to general category (Figure 1). A poor belonging to the STs is much worse than a poor belonging to general category/OBCs/SCs. The greater poverty depth suffered by the STs is probably the reason for the little decline in poverty among STs during the 1990s (the poverty ratio for the STs decreased from 52 to 47.6 percent between 1993-94 and 2004-05 (table 2)).

Table 4 presents the difference in log PCE for the SC and the ST households when compared to Non-SC/ST and general category households, respectively. For both the SCs and the STs, the gap increased by an average 10 log points when general category is the comparison group instead of Non-SCs/STs. Also, the increase in gap is across the entire distribution. This confirms that using Non-SCs/STs as comparison group rather than general category underestimates the disadvantage suffered by the SCs and the STs.

4 Results

4.1 Quantile Regressions

We begin by running restricted version of (1) that includes only dummies for the SCs, STs, and OBCs as explanatory variables (other X's are omitted). The results are given in Panel A of Table 4. The constant term is basically an estimate of log PCE for the general category households. The coefficients on the social group dummies give the unconditional difference in log PCE between a household belonging to that particular social group and a household belonging to general category at the θ^{th} quantile of log PCE distribution. If the coefficients are small, they can be multiplied by 100 and interpreted as approximately the percentage by which particular social group households' PCE exceed those of general category households' PCE.¹⁴ All social group dummies have negative and statistically significant coefficients at the select quantiles. This implies that households belonging to the 'backward classes' are disadvantaged compared to same ranked households belonging to general category, and the disadvantage is larger at higher quantiles. The disadvantage suffered by the ST households compared to general category households is greatest, followed by the SC households and the OBC households. More importantly, households belonging to the OBCs also experience a statistically significant disadvantage compared to households belonging to general category.

Panel B of Table 5 presents the results when the full set of control variables is added to the model, but the slope coefficients are restricted to be identical for all the social groups. The included social group dummies in these regressions are interpreted as the difference in the log PCE of a household belonging to that particular social group relative to a household belonging to general category after we control for differences in the characteristics. Interestingly, when we control for the covariates, the coefficients on the social group dummies reduce in absolute value relative to the raw gaps at all the select quantiles. The

¹⁴ The approximation to percent difference in PCE is poor for larger coefficient values; exact percentage difference is $100 * [\exp(\hat{\beta}) - 1]$.

OLS coefficients also decrease. However, all three social group dummies remain statistically significant. This implies that although the general category households have ‘better’ covariates compared to households belonging to the ‘backward classes’, this fails to explain fully the advantages enjoyed by households belonging to general category.

Next we estimate the full model including interactions of the social group dummies with all the covariates as given in equation (1). This allows for different returns across covariates by social group. The results are given for select quantiles in appendix Table A2, while description of the variables is given in appendix Table A1. The results show that there exist statistically significant differences in returns of many covariates by social group. Hence, there exist differences in the distribution of the covariates and differences in the returns across social groups. To aggregate the differences in the covariates and the returns to those covariates, we turn to the quantile regression decomposition method proposed in MM (2005).

4.2 The Decomposition

We assume that general category is being paid the non-discriminatory prices and compare the general category with all three disadvantaged groups separately to summarize what explains the higher PCE of households belonging to general category compared to households belonging to the ‘backward groups’.

4.2.1 Differences between General and OBCs

Figure 2 plots the decomposition results for the differences in PCE between general category and the OBC households, and panel I of Table 6 report the results for select quantiles. The Oaxaca-Blinder decomposition reveals that the average differential in PCE between general category and the OBC households is 16 log points: the differences in the covariates explain 8.6 log points, while differences in the coefficients explain 7.6 log points.

However, actual differences in PCE between general category and the OBC households differ across quantiles: the difference is larger at higher quantiles. The differential is 11 log points at 10th

percentile, 15 log points at the median, and 25 log points at the 90th percentile. The positive difference in PCE across the entire distribution suggests that households belonging to general category are better off corresponding to the same ranked households belonging to the OBCs. Also, the rich belonging to general category are better off than their OBC counterparts to a greater extent than the general category poor are better off than their OBC counterparts. The MM (2005) decomposition reveals that the covariate effect dominates the coefficient effect until about median, and the coefficient effect dominates in most of the remainder part of the distribution. Importantly, both the covariate and the coefficient effects are larger at higher quantiles. The residuals are around zero across the entire distribution, which implies that our fitted model does reasonably well in predicting the difference in log PCE.

4.2.2 Differences between General and SCs

Figure 3 plots the decomposition results for the differences in PCE between general category and the SC households, and panel II of Table 6 reports those for select quantiles. The Oaxaca-Blinder decomposition reveals that average differential in PCE between general category and the SC households is 31 log points: differences in the returns (coefficient effect) explain 14 log points, while differences in the covariates (covariate effect) explain 17 log points.

However, the differential in PCE between general category and the SC households is not only positive throughout but differs across quantiles: the differential is larger at higher quantiles. The differential is 22 log points at 10th quantile, 29 log points at the median, and 46 log points at 90th quantile. The MM decomposition reveals that the covariate effect dominates the coefficient effect over the entire distribution. Also both the coefficient and the covariate effects are larger at higher quantiles.

4.2.3 Differences between General and STs

Figure 4 plots the decomposition results for the differences in PCE between general category and the ST households, and panel III of Table 6 reports those for select quantiles. The average differential between

PCE of general category and the ST households is 42 log points: both differences in the covariates and differences in the coefficients contribute equally to total differential observed in average PCE.

Moving beyond the average, we find that the differential in PCE of general and the ST households is positive across the entire distribution: the differential is uniform until about 40th quantile before it starts to increase with the quantile. The differential is 34 log points at the 10th quantile, 37 log points at the median and 54 log points at 90th quantile. The MM decomposition reveals that the coefficient effect dominates the covariate effects until median, while the covariate effect is larger than the coefficient effect after the median. While the coefficient effect is uniform over the distribution, the covariate effect is larger at higher quantiles.

5 Discussion

The results suggest that the ST households are the most disadvantaged, followed by the SC and the OBC households compared to the households belonging to general category. Moreover, the disadvantage exists across the entire distribution, and it is larger at higher quantiles. The disadvantage results not only due to disadvantage in terms of covariates but also due to less return to those characteristics. The differences in returns cannot be taken as discrimination, although we cannot rule out some contribution of discrimination in the existing differences. As pointed out by Walle and Gunewardena, 2001, the interpretation of differences in returns as discrimination is questionable when one thinks of the likely dynamics of the income generation process. Differences in returns may exist in the absence of current discrimination, due, for instance, to a history of past group disadvantage, or simply differential cultural development—possibly perpetuated by policies such as schooling—with a continuing legacy for the returns to economic characteristics. Longstanding differences in group behavior will be embodied in the model parameters for current levels of living.

Nevertheless, apart from issues of discrimination, understanding how much disparities are due to returns versus different characteristics remains the key to explaining the causes of inequality and

designing appropriate policy. The study finds that the differences in the characteristics contribute majorly to the gap experienced by the 'backward classes'. Hence, the policies to raise the human capital and strengthening the other productive assets of households belonging to the 'backward classes' must clearly remain a focus of attention. Findings from prior studies suggest that the existing reservation policy helps the 'backward classes'. For example, Borooah et al. (2005) finds that job reservation improves the probability of the SC/ST getting regular salaried job. Pande (2003) finds an increase in transfers targeted towards the disadvantaged groups due to political reservation mandated towards them. Bertrand et al. (2008) studies the effects of affirmative action in admissions to engineering colleges in India, and finds that the affirmative policy based on caste successfully targets the financially disadvantaged. Given that the differences in living standards still exist and differences in characteristics contribute majorly to the existing differences, there is a valid case of continuance of reservation policies for more time.

However, it is also important to investigate why even after more than fifty years of affirmative policies, the disadvantage suffered by the STs and the SCs compared to general category remains high? Do the differences in the human capital variables result from different demand by different social groups or different supply of facilities? The different supply of facilities is more important in the case of the STs as they live in isolated areas. Census data indicate that the districts with a higher proportion of the STs are associated with poorer public goods such as schools, tapped water, paved roads, electricity, and health facilities (Banerjee and Somanathan, 2001). It is essential that the supply side bottlenecks are removed fast and integration of STs with the mainstream is expedited. Less demand by the disadvantaged groups may result from economic, cultural, or outlook about future prospects. For example, Dréze and Kingdon (2001) find that the overall educational disadvantage of children belonging to underprivileged social groups is partly mediated by lower parental motivation. There may also be information problems. There are a number of programs sponsored by the government which provide economic incentives to weaker sections of the society. What is needed is more effective transmission of information to people by local government and NGOs. The less demand due to cultural reasons or outlook about future prospects

needs sustained efforts by NGOs to educate people and correct their perceptions about acquiring education.

However, what is less certain is that whether this will be sufficient to eliminate the disadvantage suffered by these two groups. Geographical isolation of STs may be partly responsible for different returns to the ST households; however, discrimination cannot be overruled. Kijima (2006b) finds that the differences between the ST and Non-SC/ST households are partly due to geographical differences. Nevertheless, for the SCs, the geographical isolation is not an important issue as they are spread over the entire country. To reduce the differences in returns, it is essential to promote a more active labor market in rural India as the market mechanism generally pushes towards similar returns to productive characteristics.

The findings that the backward classes especially the STs and SCs suffer a large disadvantage compared to general category across the entire distribution even today, suggest the importance of research on the “how welfare of different social groups have evolved over time”. Kijima (2006b) investigates how the average differences between social groups have evolved between 1983 and 1999. However, the average differences may increase or decrease with all groups gaining or all groups losing. It is important to ensure that high growth rate experienced by India is shared by the rural population as well as all social groups. It is also important to look at the entire welfare distribution rather than concentrating on averages as averages may not reveal the complete picture. For example, Azam (2008) finds that in urban India, while male workers at lower and higher quantiles experienced increase in real wages between 1993 and 2004, the workers at the middle quantiles didn't experience any increase, and the female workers at middle quantiles did experience a decrease in real wage between the same time period.

Apart from the issues of improvement in welfare, the priority should be on ensuring that the poor get out of the poverty, and the non-poor do not get into poverty again. Very little improvement in the poverty ratio of STs is a concern area for policy makers. Moreover, one should keep in mind that people

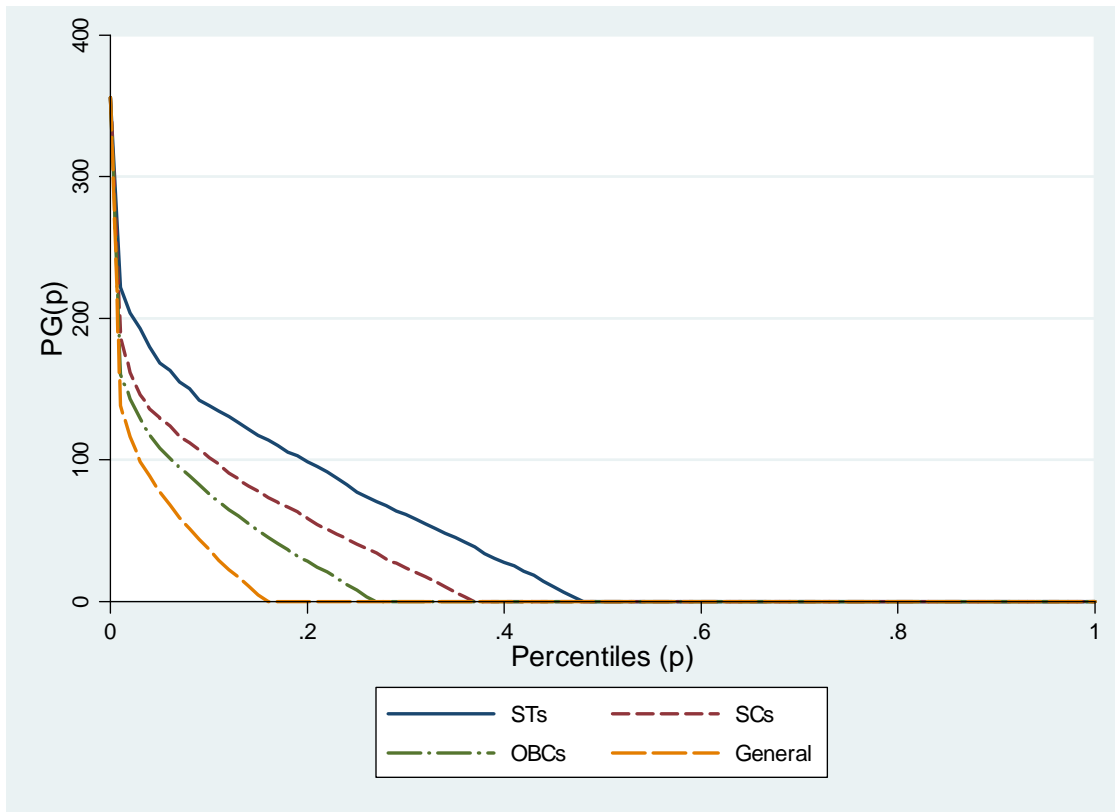
are often hurting behind the averages (Ravallion, 2001). In cases in which the survey data have tracked the same families over time ("panel data"), it is quite common to find considerable churning under the surface; Baulch and Hoddinott (2000) provide evidence of this for a number of countries. One can find that many people have escaped poverty while others have fallen into poverty, even though the overall poverty rate may move rather little. For example, comparing household incomes immediately after the 1998 financial crisis in Russia with incomes of the same households two years earlier, one finds a seemingly small two percentage point increase in the poverty rate. However, this was associated with a large proportion of the population (18%) falling into poverty, while a slightly smaller proportion (16%) escaped poverty over the same period (Lokshin and Ravallion, 2000a). A further analysis of income mobility in rural population and in different social groups residing in rural India will prove insightful.

References

- Albrecht, J., Vuuren, A. and Vroman, S. (2007), "Counterfactual Distributions with Sample Selection Adjustments: Econometric Theory and an Application to the Netherlands," *Georgetown University Working Paper*.
- Autor, D.H., Katz, L.F. and Kearney, M.S. (2005), "Rising Wage Inequality: The Role of Composition and Prices," *NBER Working Paper*, 11628.
- Azam, M. (2008), "Changes in Wage Structure in Urban India during 1983-2004: A Quantile Regression Decomposition," *Southern Methodist University Working Paper*.
- Banerjee, A. and Somanathan, R. (2001), "Caste, Community, and Collective Action: The Political Economy of Public Goods Provision in India," *Unpublished manuscript*, Department of Economics, MIT.
- Béteille, A. (2002), "Caste, Inequality, and Affirmative Action," *Institute of Labor Studies*, Geneva.
- Bertrand, M., Hanna, R. and Mullainathan, S. (2008), "Affirmative Action in Education: Evidence from Engineering College Admissions in India," *NBER Working paper*, 13926.
- Bhaumik, S. and Chakrabarty, M. (2006), "Earning Inequality in India: Has the Rise of Caste and Religion Based Politics in India Had an Impact," *IZA Discussion paper*, 2008.
- Borooah, V.K. (2005), "Caste, Inequality, and Poverty in India," *Review of Development Economics*, 9(3), 399-414.
- Borooah, V., Dubey, A. and Iyer, S. (2007), "The Effectiveness of Jobs Reservation: Caste, Religion and Economic Status in India," *Development and change*, 38(3), 423-445.
- Chamraborty, R. (2006), "Economic Liberalization and Wage Inequality in India," *World Development*, 34(12), 1997-2015.
- Deaton, A. (1997), "The Analysis of Household Surveys - A Microeconomic Approach to Development Policy," *Johns Hopkins University Press*, Baltimore.
- Despande, A. (2000), "Recasting Economic Inequality," *Review of Social Economy*, LVIII(3), 381-399.
- Gang, I.N., Sen, K. and Yun, M. (2002), "Caste, Ethnicity and Poverty in Rural India," *IZA Discussion Paper*, 629.
- Gang, I.N., Sen, K. and Yun, M. (2007), "Poverty in Rural India: Caste and Tribe," *Review of Income and Wealth*, forthcoming.
- Kijima, Y. (2006a), "Why did Wage Inequality Increase? Evidence from Urban India 1983-99," *Journal of Development Economics*, 81(1), 97-117.
- Kijima, Y. (2006b), "Caste and Tribe Inequality: Evidence from India, 1983--1999," *Economic Development and Cultural Change*, 54(2), 369-404.

- Koenker, R. and Bassett, G. (1978), "Regression Quantiles," *Econometrica*, 46, 33-50.
- Machado, J. and Mata. J. (2005), "Counterfactual Decomposition of Changes in Wage distributions using Quantile Regression," *Journal of Applied Econometrics*, 20, 445-65.
- Munshi, K. and Rosenzweig, M. (2006), "Traditional Institutions Meet the Modern World: Caste, Gender, and Schooling Choice in a Globalizing Economy," *American Economic Review*, 96, 1225-1252.
- Mutatkar, R. (2005), "Social Group Disparities and Poverty in India," *Indira Gandhi Institute of Development Research Working Paper*, 2005-004.
- Nguyen, B.T., Albrecht, J.W., Vroman, S.B. and Westbrook, M.D. (2007), "A Quantile Regression Decomposition of Urban--Rural Inequality in Vietnam," *Journal of Development Economics*, 83, 466-490.
- Oaxaca, R. (1973), "Male-Female Differentials in Urban Labor Markets," *International Economic Review*, 14, 693-709.
- Oaxaca, R. and Ransom, M. (1994), "On Discrimination and the Decomposition of Wage Differentials," *Journal of Econometrics*, 61, 5-21.
- Pande, R. (2003), "Can Mandated Political Reservation Provide Disadvantaged Minorities Policy Influence? Theory and Evidence from India," *American Economic Review*, 93, 1132-1151.
- Ravallion, M. (2001), "Growth, Inequality, and Poverty: Looking Beyond Averages," *World Bank Policy Research Working Paper*, 2558.
- Sundaram, K. and Tendulkar, S.D. (2003), "Poverty among Social and Economic Groups in India in the Nineteen Nineties," *Centre for Development Economics Working Paper*, 118.
- Thorat, S. and Mahamallick, M. (2005), "Persistent Poverty-Why do Scheduled Castes and Scheduled Tribes Stay Chronically Poor," *Paper presented at Indian Institute of Public administration*.
- Walle, D.V.D. and Gunewardena, D. (2001), "Sources of Ethnic Inequality in Viet Nam," *Journal of Development Economics*, 65, 177-207.

Figure 1: Poverty Gap for Different Social Groups, 2004-05



Notes: The poverty gap is in Indian Rupees at current prices. Official rural poverty line (356.30 Rupees) is used to measure poverty gap.

Figure 2: Differences in log PCE for General and OBCs

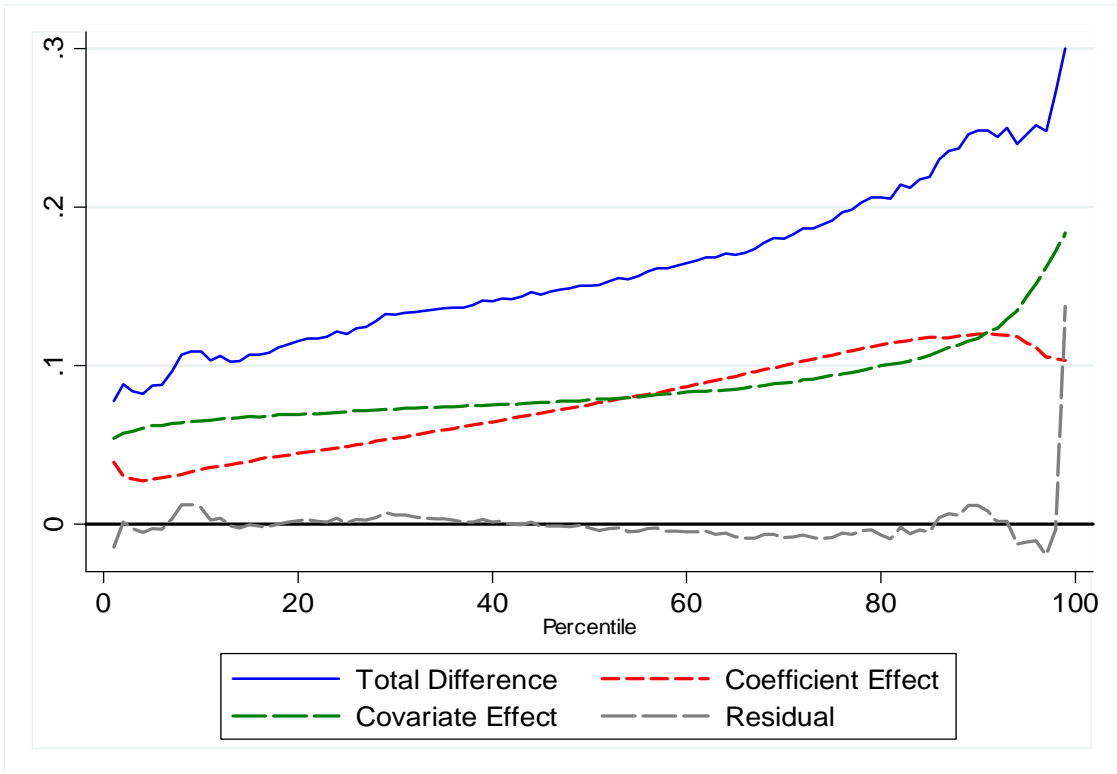
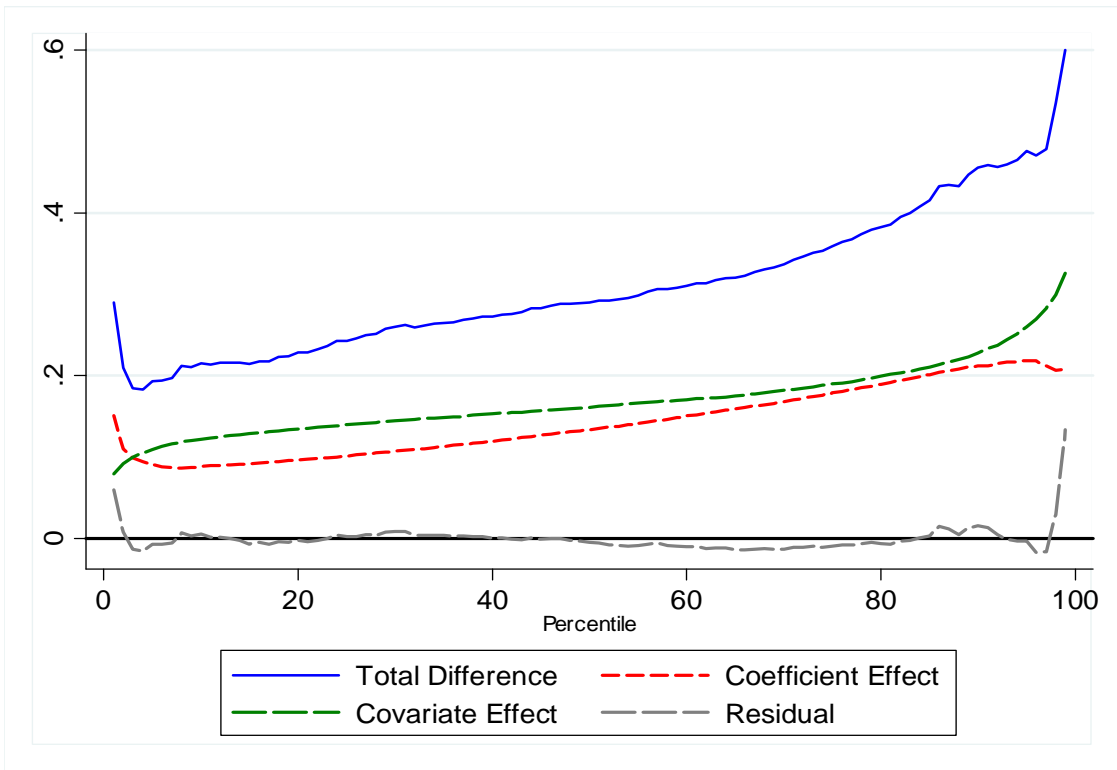
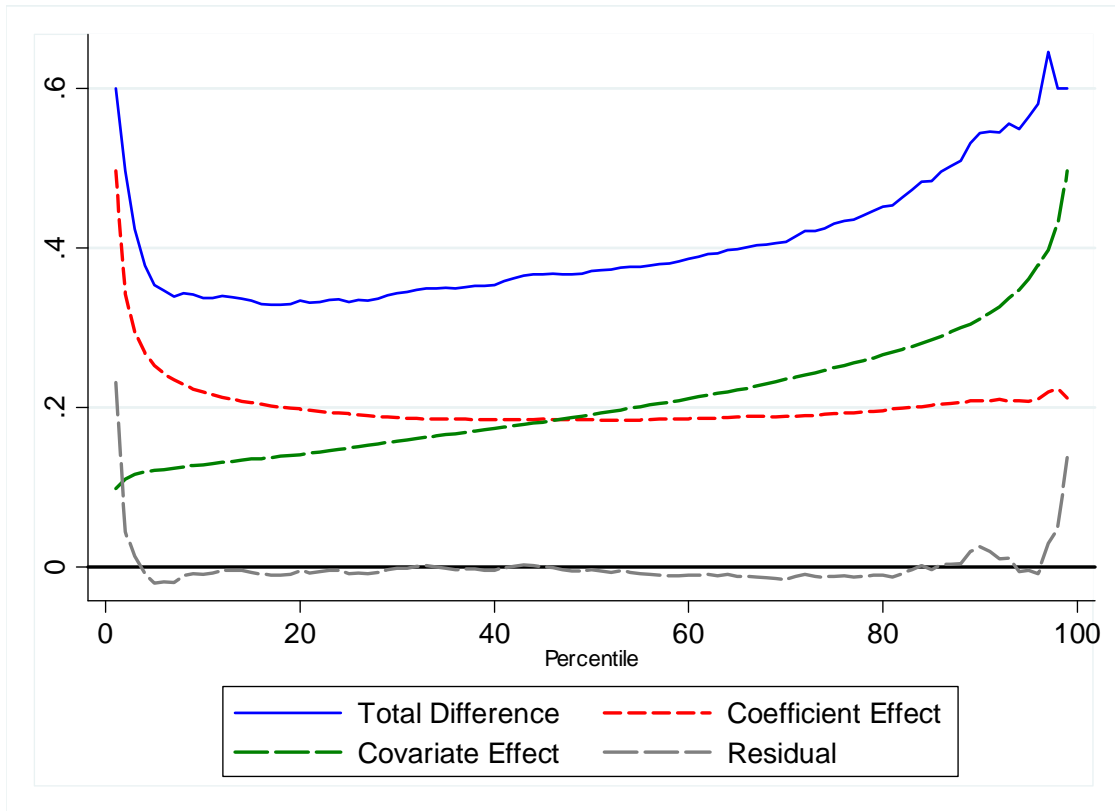


Figure 3: Differences in log PCE for General and SCs



Note : See notes after Figure 3.

Figure 4: Differences in log PCE for General and STs



- Notes: (1) Total difference is the difference of actual empirical densities of general category and group g at each percentile, i.e., $\alpha(f(y(\text{Gen}))) - \alpha(f(y(g)))$; where g is OBCs, SCs and STs in Figure 1, Figure 2 and Figure 3, respectively.
- (2) The coefficient effect is difference between the counterfactual density-the density that would have prevailed if general category would have the characteristics of group g -and the fitted density for general category, i.e., $\alpha(f^*(y(\text{Gen});x(g))) - \alpha(f^*(y(\text{Gen})))$.
- (3) Covariate effect is difference of fitted density of general category and counterfactual density-the density that would have prevailed if general category would have the characteristics of group g , i.e., $\alpha(f^*(y(\text{Gen}))) - \alpha(f^*(y(\text{Gen});x(g)))$.
- (4) Residual is the difference of difference in actual empirical densities and difference in fitted densities at each percentile, i.e., $\{\alpha(f(y(\text{Gen}))) - \alpha(f(y(g)))\} - \{\alpha(f^*(y(\text{Gen}))) - \alpha(f^*(y(g)))\}$.

See text for further details.

Table 1: Distribution of Social Group population in Rural and Urban

Group	Rural	Urban
STs	91.43	8.57
SCs	79.78	20.22
OBCS	77.98	22.02
General	62.34	37.66
All	74.68	25.32

Notes: 1) Author's calculation from NSS 61st round consumption expenditure survey.
2) Individual level data is used.

Table 2: Poverty Ratio in Rural India by Social groups

Group	1993-94	2004-05
STs	52.0	47.64
SCs	48.2	36.8
OBCS	-	26.7
General	-	16.0
General + OBCs	31.4	22.7
All	37.2	28.3

Notes: 1) Author's calculation from 50th and 61st round of NSS consumption expenditure survey.
2) Individual level data and official rural poverty lines are used.

Table 3: Descriptive Statistics of Sample used

	Scheduled Tribes		Scheduled Castes		Other Backward Castes		General		ALL	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
log PCE	5.80	0.51	5.90	0.46	6.05	0.50	6.21	0.55	6.03	0.52
Household size	4.73	2.28	4.77	2.24	4.97	2.52	4.90	2.49	4.88	2.43
% of child in hh (pchild)	0.34	0.26	0.32	0.24	0.31	0.24	0.28	0.23	0.31	0.24
% of adult male (padmale)	0.31	0.20	0.31	0.20	0.30	0.20	0.32	0.21	0.31	0.20
% of adult female (padfemale)	0.30	0.18	0.30	0.18	0.31	0.18	0.32	0.18	0.31	0.18
Land cultivated (in hectares)	0.80	1.22	0.33	1.15	0.74	1.68	0.94	1.84	0.71	1.60
Female head*	0.10	0.30	0.11	0.31	0.12	0.32	0.11	0.31	0.11	0.31
Age of head	41.96	13.74	43.64	13.48	45.20	13.66	46.49	13.93	44.85	13.77
Education variables										
Below Primary*	0.71	0.45	0.66	0.47	0.56	0.50	0.42	0.49	0.56	0.50
Primary*	0.13	0.33	0.13	0.34	0.15	0.36	0.16	0.37	0.15	0.36
Middle*	0.10	0.30	0.12	0.33	0.15	0.35	0.19	0.39	0.15	0.35
Secondary*	0.03	0.18	0.05	0.21	0.07	0.26	0.11	0.32	0.07	0.26
Higher Secondary*	0.02	0.15	0.03	0.16	0.04	0.19	0.06	0.24	0.04	0.20
Graduate*	0.01	0.11	0.01	0.11	0.02	0.13	0.04	0.20	0.02	0.15
Post Graduate*	0.00	0.04	0.00	0.06	0.01	0.08	0.01	0.12	0.01	0.08
Occupation variables										
Others*	0.09	0.29	0.09	0.29	0.11	0.31	0.15	0.36	0.11	0.32
Self employed in non-agriculture (selfnonagr)*	0.07	0.25	0.14	0.35	0.18	0.38	0.17	0.38	0.16	0.36
Agricultural Labor*	0.35	0.48	0.43	0.49	0.23	0.42	0.16	0.36	0.27	0.44
Other labor (othlab)*	0.11	0.31	0.15	0.36	0.10	0.30	0.08	0.27	0.11	0.31
Self-employed in agriculture (selfagr)*	0.38	0.49	0.19	0.39	0.38	0.49	0.44	0.50	0.36	0.48
Number of Observations	12,817		13,711		29,992		22,678		79,198	

Note: * denotes dummy variable.

Table 4: Difference in log PCE between different social groups

Quantile	Difference between STs and Non-SCs/STs	Difference between STs and general	Under-estimation of Gap for STs	Difference between SCs and Non-SCs/STs	Difference between SCs and general	Under-estimation of Gap for SCs
10 th	-0.27	-0.34	-0.07	-0.14	-0.22	-0.07
25 th	-0.25	-0.33	-0.08	-0.16	-0.24	-0.08
50 th	-0.28	-0.37	-0.10	-0.19	-0.29	-0.10
75 th	-0.31	-0.43	-0.12	-0.24	-0.36	-0.12
90 th	-0.40	-0.54	-0.14	-0.31	-0.46	-0.14
95 th	-0.43	-0.56	-0.14	-0.34	-0.48	-0.14
Mean	-0.32	-0.42	-0.10	-0.21	-0.31	-0.10

Note: 1. Difference in log PCE refers to difference in empirical CDFs of log PCE for two groups at different quantiles.

2. Non-SCs/STs refer to General category and Other Backward Castes.

Table 5: Difference in log PCE for different Social Groups at different Quantiles

	Quantiles					
	OLS	10 th	25 th	50 th	75 th	90 th
Panel A						
Constant	6.410 (0.000)	5.797 (0.000)	6.041 (0.000)	6.348 (0.000)	6.707 (0.000)	7.110 (0.000)
SC	-0.313 (0.000)	-0.216 (0.000)	-0.243 (0.000)	-0.290 (0.000)	-0.359 (0.000)	-0.456 (0.000)
ST	-0.418 (0.000)	-0.337 (0.000)	-0.332 (0.000)	-0.371 (0.000)	-0.430 (0.000)	-0.544 (0.000)
OBC	-0.162 (0.000)	-0.109 (0.000)	-0.120 (0.000)	-0.151 (0.000)	-0.192 (0.000)	-0.248 (0.000)
Panel B						
Constant	6.095 (0.000)	5.591 (0.000)	5.827 (0.000)	6.116 (0.000)	6.423 (0.000)	6.681 (0.000)
SC	-0.145 (0.000)	-0.113 (0.000)	-0.112 (0.000)	-0.131 (0.000)	-0.129 (0.000)	-0.149 (0.000)
ST	-0.223 (0.000)	-0.209 (0.000)	-0.180 (0.000)	-0.190 (0.000)	-0.189 (0.000)	-0.189 (0.000)
OBC	-0.060 (0.000)	-0.044 (0.000)	-0.048 (0.000)	-0.057 (0.000)	-0.051 (0.000)	-0.066 (0.000)

- Note: 1. OLS = ordinary least squares.
2. P-values in the parentheses.
3. In panel A, the gap is estimated by quantile regression where log of monthly per capita expenditure is dependent variable and social group dummies (excluded is general category) are independent variables.
4. In panel B, the gap is estimated by quantile regression where log of monthly per capita expenditure is dependent variable and social group dummies (excluded is general category), state dummies and all explanatory variables as independent variables.

Table 6: The Decomposition of the Differences in log per capita expenditure

<i>I. Difference between General and OBCs</i>						
Quantile	Marginal- General	Marginal- OBCs	Difference	Coefficient effect	Covariate effect	Residual
Mean*	6.41	6.25	0.16	0.76 (0.47)	0.86 (0.53)	0.00 (0.00)
10 th	5.60	5.49	0.11	0.03 (0.31)	0.06 (0.59)	0.01 (0.09)
25 th	5.85	5.73	0.12	0.05 (0.41)	0.07 (0.59)	0.00 (0.00)
50 th	6.15	6.00	0.15	0.08 (0.52)	0.08 (0.50)	0.00 (-0.02)
75 th	6.51	6.32	0.19	0.11 (0.56)	0.09 (0.49)	-0.01 (-0.05)
90 th	6.91	6.67	0.25	0.11 (0.48)	0.11 (0.47)	0.02 (0.05)
<i>II. Difference between General and SCs</i>						
Quantile	Marginal- General	Marginal- SCs	Difference	Coefficient effect	Covariate effect	Residual
Mean*	6.41	6.10	0.31	0.14 (0.46)	0.17 (0.54)	0.00 (0.00)
10 th	5.60	5.39	0.22	0.09 (0.41)	0.12 (0.56)	0.01 (0.03)
25 th	5.85	5.60	0.24	0.10 (0.42)	0.14 (0.57)	0.00 (0.01)
50 th	6.15	5.86	0.29	0.13 (0.46)	0.16 (0.56)	-0.01 (-0.02)
75 th	6.51	6.15	0.36	0.18 (0.50)	0.19 (0.53)	-0.01 (-0.03)
90 th	6.91	6.46	0.46	0.21 (0.47)	0.23 (0.50)	0.02 (0.03)

Continued.....

Table 6..... (Cont.)

III. Difference between General and STs						
Quantile	Marginal- General	Marginal- STs	Difference	Coefficient effect	Covariate effect	Residual
Mean*	6.41	5.99	0.42	0.21 (0.49)	0.21 (0.51)	0.00 (0.00)
10 th	5.60	5.26	0.34	0.22 (0.65)	0.13 (0.38)	-0.01 (-0.03)
25 th	5.85	5.51	0.33	0.19 (0.58)	0.15 (0.45)	-0.01 (-0.03)
50 th	6.15	5.78	0.37	0.18 (0.50)	0.19 (0.51)	0.00 (-0.01)
75 th	6.51	6.08	0.43	0.19 (0.45)	0.25 (0.58)	-0.01 (-0.03)
90 th	6.91	6.37	0.54	0.21 (0.38)	0.31 (0.57)	0.03 (0.05)

Note: Marginal refers to marginal distribution of log PCE. Proportion of difference Explained is in the parenthesis. See text for further details.

Appendix

Table A1: Definition of variables

<i>Acronym</i>	<i>Definition</i>
SC	Dummy=1 if household belong to Scheduled Castes
ST	Dummy=1 if household belong to Scheduled Tribes
OBC	Dummy=1 if household belong to Other Backward Castes
PCE	Per capita household consumption expenditure
hhsiz	Household Size
pchild	Proportion of child below age 14 in the household
padmale	Proportion of adult male in household
padfemale	Proportion of adult female in household
femalehead	Dummy=1 if head is female
headage	Age of household head
primary	Dummy=1 if household head's education level is Primary
middle	Dummy=1 if household head's education level is Middle
secondary	Dummy=1 if household head's education level is Lower Secondary
hrsecondary	Dummy=1 if household head's education level is Higher Secondary
graduate	Dummy=1 if household head's education level is Graduate
postgrad	Dummy=1 if household head's education level is Post Graduate
cultland	Land Cultivated in hectares
selfnonagr	Dummy=1 if self employed in non-agriculture
others	Dummy=1 if employed in others
othlab	Dummy=1 if work in other labor activities
selfagr	Dummy=1 if self employed in agriculture

Table A2: OLS and Quantile Regression, Full Model.

	<i>Quantiles</i>					
	OLS	10 th	25 th	50 th	75 th	90 th
Constant	6.087***	5.486***	5.734***	6.073***	6.407***	6.689***
SC	-0.134***	0.036	0.075	-0.101	-0.082	-0.193*
ST	-0.245***	0.085	-0.109	-0.188*	-0.244***	-0.428***
OBC	-0.058	0.079	0.051	-0.002	-0.043	-0.061
hhsz	-0.045***	-0.042***	-0.043***	-0.046***	-0.047***	-0.046***
SC X hhsz	0.008***	0.009**	0.003	0.003	-0.001	-0.003
ST X hhsz	0.015***	0.014**	0.002	0.002	0.005	-0.005
OBC X hhsz	0.006**	0.006*	0.001	0.003	0.006*	0.004
pchild	-0.313***	-0.127***	-0.225***	-0.287***	-0.394***	-0.452***
SC X pchild	0.022	-0.094	-0.032	0.055	0.076	0.178*
ST X pchild	-0.059	-0.388***	0.011	0.066	0.139*	0.335***
OBC X pchild	0.059	-0.039	0.039	0.034	0.088	0.077
padmale	0.215***	0.359***	0.270***	0.165***	0.149***	0.135***
SC X padmale	0.026	-0.083	-0.052	0.085	0.011	0.162*
ST X padmale	0.071	-0.250***	0.022	0.126*	0.086	0.227**
OBC X padmale	0.084**	-0.062	0.025	0.071	0.082*	0.118*
padfemale	0.069**	0.182***	0.123***	0.063*	-0.082**	-0.046
SC X padfemale	0.120***	0.040	-0.014	0.063	0.119*	0.142*
ST X padfemale	0.015	-0.219***	-0.103	0.02	0.099	0.137
OBC X padfemale	0.042	-0.006	-0.001	0.004	0.089*	0.064
femalehead	0.049***	0.031*	0.043***	0.031*	0.065***	0.079***
SC X femalehead	-0.051**	-0.068**	-0.072***	-0.056*	-0.051*	0.021
ST X femalehead	0.016	-0.018	-0.037	0.000	0.020	0.025
OBC X femalehead	-0.022	-0.049**	-0.053**	-0.028	-0.015	-0.008
headage	0.003***	0.003***	0.003***	0.003***	0.002***	0.002***
SC X headage	0.000	-0.001*	-0.002**	-0.001	-0.001	-0.001
ST X headage	0.000	0.000	0.000	-0.001	-0.001	0.002
OBC X headage	-0.001	-0.001*	-0.001**	-0.001*	-0.001*	-0.001
Primary	0.095***	0.084***	0.089***	0.077***	0.097***	0.100***
SC X primary	-0.009	-0.027	-0.018	0.005	-0.007	-0.008
ST X primary	-0.026	0.028	0.052*	0.038	0.057*	0.021
OBC X primary	-0.015	-0.009	-0.002	0.011	-0.031*	-0.034
Middle	0.171***	0.179***	0.170***	0.150***	0.154***	0.164***
SC X middle	-0.051***	-0.093***	-0.077***	-0.037	-0.004	-0.019
ST X middle	0.03	-0.013	0.015	0.026	0.059*	0.062
OBC X middle	-0.022*	-0.036*	-0.032*	-0.01	-0.014	0.003

Continued.....

Table A2..... (cont.)

	<i>Quantiles</i>					
	OLS	10th	25th	50th	75th	90th
secondary	0.268***	0.234***	0.270***	0.248***	0.267***	0.279***
SC X secondary	-0.061**	-0.099***	-0.114***	-0.089***	-0.075**	-0.061
ST X secondary	0.071*	0.095*	-0.001	0.060	0.004	-0.035
OBC X secondary	-0.028*	-0.048*	-0.085***	-0.028	-0.035	-0.019
hrsecondary	0.384***	0.335***	0.385***	0.359***	0.394***	0.435***
SC X hrsecondary	-0.018	-0.105**	-0.156***	-0.072*	-0.02	0.036
ST X hrsecondary	0.110***	-0.015	-0.062	0.002	0.026	0.052
OBC X hrsecondary	-0.070***	-0.076**	-0.123***	-0.084**	-0.087***	-0.056
graduate	0.474***	0.407***	0.423***	0.434***	0.487***	0.526***
SC X graduate	0.002	-0.098	-0.123**	-0.015	0.015	-0.009
ST X graduate	0.223***	0.047	0.180***	0.045	0.098*	0.166**
OBC X graduate	-0.017	-0.051	-0.026	-0.006	-0.031	-0.081*
postgrad	0.774***	0.552***	0.611***	0.700***	0.948***	1.137***
SC X postgrad	-0.287***	-0.472***	-0.430***	-0.264**	-0.260*	-0.543***
ST X postgrad	0.164	0.517***	0.222**	-0.044	-0.319**	-0.077
OBC X postgrad	-0.175***	-0.062	-0.065	-0.085	-0.365***	-0.478***
cultland	0.037***	0.027***	0.030***	0.039***	0.046***	0.054***
SC X cultland	-0.021***	-0.022***	-0.019**	-0.008	0.021***	0.024***
ST X cultland	-0.005	-0.015**	-0.005	-0.011	-0.003	0.008
OBC X cultland	-0.006*	-0.002	0.003	0.007**	0.006*	0.012***
selfnonagr	0.252***	0.166***	0.208***	0.227***	0.271***	0.331***
SC X selfnonagr	-0.112***	-0.039*	-0.087***	-0.091***	-0.119***	-0.149***
ST X selfnonagr	-0.042	-0.04	-0.058*	-0.055*	-0.008	-0.053
OBC X selfnonagr	-0.052***	-0.025	-0.048**	-0.046*	-0.055**	-0.063**
others	0.334***	0.216***	0.267***	0.336***	0.418***	0.470***
SC X others	-0.180***	-0.188***	-0.129***	-0.113***	-0.154***	-0.113**
ST X others	-0.295***	-1.082***	-0.278***	-0.090**	-0.065*	-0.06
OBC X others	-0.091***	-0.080***	-0.066***	-0.090***	-0.103***	-0.090**
othlab	0.143***	0.094***	0.145***	0.124***	0.164***	0.192***
SC X othlab	-0.074***	-0.046*	-0.097***	-0.066**	-0.095***	-0.078*
ST X othlab	-0.026	0.024	-0.038	-0.025	-0.059*	-0.090*
OBC X othlab	-0.041**	-0.031	-0.068***	-0.028	-0.031	-0.057
selfagr	0.196***	0.167***	0.180***	0.186***	0.216***	0.219***
SC X selfagr	-0.035**	-0.039*	-0.029	-0.022	-0.081***	-0.063*
ST X selfagr	-0.075***	-0.075***	-0.071***	-0.056**	-0.057**	-0.046
OBC X selfagr	-0.016	-0.036*	-0.023	-0.027	-0.032	-0.044*

Observations 79,224.

*p<0.05, **p<0.01, ***p<0.001

Note: The model also include state dummies not reported in the table.