CDE
February 2014

IS SELF-EMPLOYMENT THE ANSWER TO CASTE DISCRIMINATION? DECOMPOSING THE EARNINGS GAP IN INDIAN HOUSEHOLD NONFARM BUSINESSES

Ashwini Deshpande
Email: ashwini@econdse.org
Department of Economics
Delhi School of Economics

Smriti Sharma
Email: smriti@econdse.org
Department of Economics
Delhi School of Economics

Working Paper No. 236

Centre for Development Economics
Department of Economics, Delhi School of Economics

Is Self-Employment the Answer to Caste Discrimination? Decomposing the Earnings Gap in Indian Household Nonfarm Businesses

Ashwini Deshpande*
Smriti Sharma†
Department of Economics, Delhi School of Economics

February 11, 2014

*Email: ashwini@econdse.org. Address for correspondence: Department of Economics, Delhi School of Economics, University of Delhi, Delhi-110007. We are grateful to Shantanu Khanna for discussions on the quantile regression decomposition methodology and to Deepti Goel and seminar participants at the Delhi School of Economics for comments. We are responsible for all remaining errors and omissions.
†Email: smriti@econdse.org.

Abstract

Using the India Human Development Survey data for 2004-05, we employ two methodologies to estimate the earnings structure of household nonfarm businesses owned by Scheduled Castes and Tribes (SCSTs) and non-SCSTs: OLS estimation of mean earnings, and quantile regressions. Correspondingly, we use two decomposition methods: the conventional Blinder-Oaxaca decomposition and Melly's (2006) refinement of the Machado and Mata (2005) decomposition of quantile gaps. We find clear differences in characteristics between SCST-owned and non-SCST owned businesses. The Blinder-Oaxaca decomposition reveals that depending on the specification of explanatory variables, as much as 70 percent of the earnings gap could be attributed to the "unexplained" or the discriminatory component. Quantile regressions reveal that gaps are higher at lower deciles than the higher ones (both raw gaps, as well as after controlling for characteristics), and the decompositions show that the unexplained component is higher at the lower deciles than higher, suggesting that SCST-owned businesses at the lower end of the conditional distribution face greater discrimination, as compared to those at the higher end. Thus, we find strong evidence of a "sticky floor", a phenomenon observed for gender wage gaps in developing countries (in contrast to a "glass ceiling" in developed countries).

JEL classification codes: J31, J71, C21, O15, O17
Keywords: Caste, discrimination, household nonfarm business, earning gaps, quantile regressions, earnings decomposition

1 Introduction

Given the compelling evidence of labour market discrimination against marginalized groups the world over, there is a view that members of such groups should turn to self-employment, specifically towards entrepreneurship. There is evidence that the desire for entrepreneurship is also strong among marginalized groups (Fairlie, 2006). The argument has several different strands, which run from a focus on increasing individual wealth and providing role models for upward mobility to the rest of the community, to increasing group wealth. The former focuses on accumulation of individual wealth through one's own business, rather than be dependent on others for jobs that might either not be forthcoming at all, or be low paying, bottom-of-the-ladder, poor quality jobs; and even when members of marginalized groups get top-end jobs, they face serious discrimination at the workplace (Daniels, 2004). The latter set of arguments about group wealth encompasses several interrelated notions: one, business ownership is seen as a solution to the relative poverty of the group, rather than of individuals; two, by setting up their own businesses, they become job-givers, especially to members of their own community, and not job-seekers, and thus reduce their vulnerability to labour market discrimination; and three, marginalized communities enhance their representation among the economic elite by increasing their share of the pie, both through individual and collective businesses (e.g. co-operatives, partnerships).[1] Indeed, the concept of "Black Capitalism" in the USA encapsulated several of these views (e.g., Hill and Rabig 2012; Villemez and Beggs 1984). We see similar views being reiterated in India through an advocacy of "Dalit Capitalism", the first step of which has been the formation of the Dalit Indian Chamber of Commerce and Industry (DICCI), mirroring the old established big business association Federation of Indian Chambers of Commerce and Industry (FICCI).[2]

The implicit assumptions behind these discussions have been, (a) that a sufficiently large part of the self-employment activity would be concentrated

[1] Anderson (2001) discusses these aspects at length.

[2] Formerly untouchable castes in India (officially, Scheduled Castes or SCs) use the term Dalit (meaning oppressed) as a term of pride. Dalit Indian Chamber of Commerce and Industry (DICCI) was founded on April 14, 2005, the birth anniversary of Dr. B. R. Ambedkar, acknowledged by the DICCI as their "messiah and the intellectual father". More details can be found at their website: www.dicci.org

at the top-end, big industry level; and (b) that discriminatory tendencies that characterize the labour market would somehow be missing from other markets, such as land, credit or consumer markets, critical to the success of entrepreneurial activities. In a recent paper (Deshpande and Sharma, 2013), we examined unit-level data from two successive censuses of the Micro, Small and Medium Enterprises (MSME) sector for India to study the broad-based participation of marginalized castes in self-employment, and found reasons to question these assumptions. While there is a small and emerging section of Dalit industrialists at the top end, assisted ably by the DICCI, the bulk of Dalit businesses are engaged in small-scale, low productivity, survivalist activities according to our study, and the MSME sector exhibits very clear differences along business owners' caste (and gender), in virtually all business characteristics.

The evidence of systematic and persistent caste differences from our earlier work, however, does not prove discrimination along caste lines in the business sector; all the gaps could, in principle, be accounted for by the fact that businesses owned by Dalits and Adivasis[3] (SCST, henceforth) have characteristics that are systematically inferior to those of non-SCST-owned businesses. For instance, SCST owners could be less educated, poorer, lack strong business networks and so forth, and accounting for these characteristics might explain the entire gap in business earnings. Of course, the fact that SCSTs enter the field of self-employment with inferior characteristics indicates the presence of "pre-market" discrimination, of which there is plenty of independent evidence (e.g., Deshpande, 2011 and Thorat and Newman, 2010).

While the MSME census data used in Deshpande and Sharma (2013) has the advantage of being a complete enumeration of all registered MSMEs, one of its main shortcomings is that it does not collect information on owner characteristics except for caste, gender and religion. Thus, this precludes any meaningful examination of discrimination using that data. Our objective in this paper is to examine whether differences in earnings between SCST and non-SCST-owned household nonfarm businesses could be entirely explained by differences in characteristics, or whether there is a residual unexplained gap in earnings, which is indicative of discrimination.

[3]In addition to the SCs, tribal communities in India, also referred to as Adivasis (officially, Scheduled Tribes or STs), are officially recognized as marginalized communities.

Using the India Human Development Survey data for 2004-05, we employ two methodologies for understanding the earnings structure of household nonfarm businesses (referred to as simply "businesses" hereafter): OLS estimation of mean earnings for SCST and non-SCST businesses; and quantile regressions for a distributional analysis to look beyond the mean and to understand "what happens where" in the earnings distribution. Thereafter, we decompose the mean earnings gap through the conventional Blinder-Oaxaca decomposition. We then use the quantile regressions to decompose the gap at each percentile of the earnings distribution in order to examine whether and how the extent of discrimination changes along the earnings distribution.

Our main findings can be summarized as follows. There are clear differences in observable characteristics between SCST-owned and non-SCST-owned businesses. The latter are more urban, record larger number of total man-hours, owners are more educated, less poor, posses a greater number of assets, have better networks and are more likely to have a business in a fixed workplace. These disparities get reflected in both indicators of business performance in the data- gross receipts and net income- such that SCSTs, on average perform significantly poorly compared to non-SCSTs. The Blinder-Oaxaca decomposition reveals that depending on the specification of explanatory variables, as much as 70 percent of the net income gap could be attributed to the "unexplained" or the discriminatory component.

Quantile regressions reveal that gaps are higher at lower deciles than the higher ones (both raw gaps, as well as after controlling for characteristics), and the decompositions also reveal that the unexplained component is higher at the lower deciles than higher, suggesting that SCST-owned businesses at the lower end of the conditional earnings distribution face greater discrimination, as compared to those at the higher end. Thus, we find strong evidence of a "sticky floor", a phenomenon observed in the context of gender wage gaps in developing countries (defined in contrast to a "glass ceiling" in developed countries, where the gaps are larger at the higher end of the distribution). As is standard in decomposition exercises, the actual division of the wage gap into "explained" and "unexplained" (or "discriminatory") components is sensitive to the choice of the non-discriminatory earnings structure used to construct a counterfactual earnings distribution. In addition to earnings gaps being characterized by a sticky floor, both our

counterfactuals show that the unexplained or discriminatory part of the earnings gap is highest for the bottom two deciles and declines steadily thereafter. Thus, conditional on characteristics, low-earning SCST businesses are subject to greater discrimination than high earning ones, yet another manifestation of the sticky floor.

The rest of this paper is organized as follows: Section 2 contains a literature review; Section 3 outlines the two methodologies; Section 4 discusses the data and the key descriptive statistics; Section 5 presents the results, which are discussed further in Section 6. Section 7 offers concluding comments.

2 Related Literature

Estimates of caste differences in earnings mostly focus on wage earnings as the Employment-Unemployment Survey data from the National Sample Survey (NSS) has data on earnings of salaried employees and casual workers, but not on earnings of the self-employed.[4] For instance, Bhaumik and Chakrabarty (2006) decompose wage gaps between SCSTs and non-SCSTs using the 1986-87 and 1999-2000 NSS data. They find that the observed caste wage gap is on account of differences in characteristics and not due to discrimination. Further, they find that differences in educational endowments account for 28 percent and 44 percent of the differences in earnings in 1986-87 and 1999-2000 respectively. Deininger et al. (2013) use the 1999 Rural Economic and Demography Survey (REDS) and also find that caste differences in rural casual labour earnings can be explained by differences in endowments. They find no evidence of discrimination. However, Madheswaran and Attewell (2007) use the 1999-2000 NSS data and find that discrimination accounts for 15 percent of the average wage gap between SCSTs and non-SCSTs in urban salaried employment and this effect is stronger in the private sector than in the public sector. Das and Dutta (2007) use the 2004-05 NSS data and find that of the sizable average earnings gap between SCs and 'Others' in salaried employment, the unexplained

[4]In the 2004-05 round of the NSS Employment-Unemployment Survey, earnings from self-employment were indirectly assessed using the following questions: one, whether earning from self-employment was remunerative and two, what amount per month was considered remunerative.

discriminatory component is 59 percent.[5] On the other hand, differences in earnings from casual labour are small and can be largely attributed to differences in endowments.[6]

To the best of our knowledge, our paper is the first one to carry out a decomposition of caste differences in earnings from self-employment using Indian data. Also, as compared to other existing work, we go beyond the mean earnings gap to explore differences in earnings at various points of the earnings distribution. This has been enabled by the availability of the India Human Development Survey- a large-scale nationally representative survey- that canvasses data specifically on the performance of household non-farm businesses and also collects information on the background of the owner and the business which allows us to neatly decompose the income gap into components explained by differences in endowments and those due to discrimination.

There is ample evidence that convincingly shows that enterprises owned by SCSTs are relatively fewer and fare significantly worse than those owned by non-SCSTs. Iyer et al. (2013) and Thorat and Sadana (2009) document caste differences in non-agricultural enterprise ownership and performance. They find SCs and STs to be under-represented relative to their population shares. Enterprises owned by SCSTs are smaller in terms of number of workers, hire mostly family labour, rely less on external sources of finance and operate mostly in the unregistered unorganized sector as compared to enterprises owned by 'Others'. Thorat and Sadana (2009) also show that poverty rates are higher among SCST self-employed households. In a previous exercise, Deshpande and Sharma (2013), we use data on registered manufacturing MSMEs over 2001-02 and 2006-07 and confirm the findings of the above two studies. Further, through multivariate regressions, we also show that rates of growth of SCST owned firms are significantly smaller than those of non-SCST firms. Examining the caste mix of employees in firms, we find that proportions of SCST employees are highest in SCST-owned enterprises and significantly lower in enterprises owned by other caste groups. Singh (2011) uses the IHDS data to study caste differences

[5]The NSS data defines four broad social groups: Scheduled Castes (SCs), Scheduled Tribes (STs), Other Backward Classes (OBCs) and 'Others'. 'Others' is a reasonable approximation of the upper castes.

[6]The papers listed here use different population (rural versus urban, casual versus salaried) and data sets collected in different years. This is one apparent reason for divergent results.

in net farm income per acre of land cultivated. Decomposing the inequality in net farm income per acre into a between-group and within-group component, he finds that the between-group inequality accounts for most of the overall inequality. Jodhka (2010) through detailed interviews with Dalit entrepreneurs in Panipat, Haryana and Saharanpur, Uttar Pradesh finds that caste works as a direct and indirect barrier in the successful running of their businesses. Most of them report having difficulty on account of their Dalit identity in mobilizing finance and getting a space to start their enterprise. A majority of them felt that their caste identity was perceived as more important than their professional identity, which led to them being seen as "odd actors" in the local community.

There is also a large body of literature from the United States that studies racial differences in entrepreneurship using business creation rates, survival rates and performance measured through employment, sales, profits and net worth etc. as outcome variables. In general, the literature indicates low rates of entry into self-employment and high rates of exit from self-employment not just among blacks, but also among native Americans and Latinos, thereby leading to low rates of business ownership (Fairlie, 2006). Among endowments, racial disparities in asset ownership (with blacks lagging behind whites) are the most important factor leading to differences in business creation. Since blacks have also not traditionally been engaged in business, they are disadvantaged in terms of a family background in self-employment that has been found to increase the probability of moving into self-employment (Dunn and Holtz-Eakin 2000; Hout and Rosen 2000) and also improving business performance (Fairlie and Robb, 2007). While a third of the black-white gap in business creation has been attributed to differences in endowments, the remaining is the unexplained component that can be on account of lending discrimination (Blanchflower et al., 2003) or consumer discrimination (Borjas and Bronars, 1989) or some unobserved differences in behaviour.

3 Methodology

3.1 Blinder-Oaxaca Decomposition Framework

We first use the Blinder-Oaxaca method to decompose the mean difference in earnings from self-employment between SCSTs and non-SCSTs into

portions attributable to differences in the distribution of endowments (also known as the *composition effect* or *explained component*) and differences in returns to these endowments (also known as the *coefficients effect* or *unexplained component*) (Blinder, 1973 and Oaxaca, 1973). This methodology involves estimating Mincerian earnings equations separately for individuals i of the different groups g, SCSTs (group s) and non-SCSTs (group n):

$$w_{ig} = X_i^g \beta^g + u_i^g \tag{1}$$

where $g = (n, s)$ denotes the two groups. The dependent variable w is the natural log of earnings. X_i is the vector of covariates for individual i, which contains characteristics that would determine earnings, from personal characteristics such as age, education etc., to business characteristics such as sector of production, geographic location etc. β is the corresponding vector of coefficients and u is the random error term.

The gross difference in earnings between the two groups can be written as:

$$G = \bar{X}^n \hat{\beta}^n - \bar{X}^s \hat{\beta}^s \tag{2}$$

where $\bar{X}^n$ and $\bar{X}^s$ are average characteristics for non-SCSTs and SCSTs respectively and $\hat{\beta}^n$ and $\hat{\beta}^s$ are the coefficient estimates from separate OLS regressions for non-SCSTs and SCSTs respectively. In order to decompose this difference, one has to make some assumptions about the earnings structure that would apply to everyone in the absence of discrimination and construct counterfactual earnings functions. A possible counterfactual could be constructed by assuming that the non-discriminatory earnings structure is the non-SCST earning structure. In that case, the counterfactual earnings equation of the SCSTs would be written as:

$$w_{is}^c = X_i^s \beta^n + v_i^s \tag{3}$$

Adding and subtracting the counterfactual earnings to equation 2, we arrive at:

$$G = \bar{w}^n - \bar{w}^s = (\bar{X}^n - \bar{X}^s)\hat{\beta}^n + \bar{X}^s(\hat{\beta}^n - \hat{\beta}^s) \tag{4}$$

where the first term on the right hand side represents the part of the earnings differential due to differences in characteristics (the explained component) and the second term represents differences due to varying returns

to the same characteristics (different coefficients for non-SCSTs and SCSTs). The second term is the unexplained component and is considered to be a reflection of discrimination. Alternatively, another counterfactual could be constructed assuming that the non-discriminatory earnings structure applicable to all would be the SCST one. In that case, the earnings gap would be decomposed as follows:

$$G = (\bar{X}^n - \bar{X}^s)\hat{\beta}^s + \bar{X}^n(\hat{\beta}^n - \hat{\beta}^s) \tag{5}$$

The decomposition of the overall earnings gap into 'explained' and 'unexplained' components is sensitive to the choice of the non-discriminatory earnings structure, as the two counterfactuals would yield two different estimates of the two components. This is known as the 'index number problem' that has been extensively discussed in the literature. One could take the two estimates as the lower and upper bounds of the explained/unexplained components. One solution to the index number problem is to use the pooled estimates (estimating the earnings function for the whole population) as the single counterfactual. Another solution, suggested by Cotton (1988), is to construct the non-discriminatory earnings structure as a convex linear combination of the earnings structures of both groups.

3.2 Quantile Regression Decomposition Framework

The conventional Mincerian wage (or earnings) function focuses on average wage or earnings, as it uses the least squares method, which provides estimates of education or other attributes on the mean of conditional wage or earnings. However, moving away from the least squares method using quantile regressions introduced by Koenker and Bassett (1978), we can analyse the effect of a given attribute at different parts of the conditional earnings function, which need not be uniform. For instance, while OLS estimation shows that in the US education has a positive effect on average wages, quantile regressions have revealed that returns to education are greater at the higher quantiles of the conditional wage distribution (Buchinsky, 1994). This suggests that increased education levels in the population could lead to an increase in wage inequality, but least squares estimation would not have revealed this.

Generalising the traditional Blinder-Oaxaca decomposition that decom-

poses the earnings gap at the mean, Machado and Mata (2005) proposed a decomposition method that involves estimating quantile regressions separately for the two sub-groups and then constructing a counterfactual using covariates of one group and returns to those covariates for the other group.

The conditional earnings distribution is estimated by quantile regression. The conditional quantile function $Q_\theta(w|X)$ can be expressed using a linear specification for each group as follows:

$$Q_\theta(w_g|X_g) = X_{i,g}^T \beta_{g,\theta} \; for \; each \; \theta \; \in (0,1) \tag{6}$$

where $g = (n, s)$ represents the groups. w is the dependent variable denoting the natural log of earnings, X_i represents the set of covariates for each individual i and β_θ are the coefficient vectors that need to be estimated for the different θ^{th} quantiles. For given $\theta \in (0,1)$, $\hat{\beta}_\theta$ can be estimated by minimizing in β

$$\frac{1}{n} \sum \rho_\theta (w_{ig} - X_{i,g}^T \beta_{g,\theta}) \tag{7}$$

with

$$\rho_\theta(u) = \begin{cases} \theta \; u, & \text{if } u \geq 0. \\ (\theta - 1)u, & \text{if } u < 0. \end{cases} \tag{8}$$

where $\rho_\theta(u)$ is the check function. In effect, minimizing the sum of absolute residuals yields the median. Estimating a conditional median function therefore means minimizing equation 7 at $\theta = 0.5$.

The quantile regression coefficients can be interpreted as the returns to various characteristics at different quantiles of the conditional earnings distribution. The assumption is that all quantiles of w, conditional on X, are linear in X. We can then estimate the conditional quantile of w by linear quantile regression for each specific percentile of $\theta \in (0,1)$.

The next step is to construct the counterfactual unconditional earnings distribution using estimates for the conditional quantile regressions. However, since the law of iterated expectations does not apply in the case of quantiles, the unconditional quantile is not the same as the integral of the conditional quantiles. Machado and Mata (2005) estimate the counterfactual unconditional earnings distribution using a simulation-based technique, which consists of the following steps:

1. Generate a random sample of size m from a uniform distribution

$U[0,1]$

2. For each group, separately estimate m different quantile regression coefficients, $\hat{\beta}_{s,\theta}$ and $\hat{\beta}_{n,\theta}$

3. Generate a random sample of size m with replacement from the empirical distribution of the covariates for each group, $X_{s,i}$ and $X_{n,i}$

4. Generate the counterfactual of interest by multiplying different combinations of quantile coefficients and distribution of observables between group s and group n after repeating this last step m times.

Standard errors for the estimated quantiles of the counterfactual distribution are computed using a bootstrapping technique proposed by Machado and Mata (2005).

In the Machado and Mata procedure, the larger the number of quantile regressions, the greater the precision of the estimated conditional distribution. However, their simulation-based estimator, which relies on the generation of a random sample to construct the counterfactual unconditional earnings distribution, comes at a cost of increased computational time. Melly (2006) proposed a procedure that is less computationally intensive. Instead of using a random sample with replacement, Melly (2006) integrates the conditional earnings distribution over the entire range of covariates to generate the marginal unconditional distribution of log earnings. Then, by inverting the unconditional distribution function, the unconditional quantiles of interest can be obtained. This procedure uses all the information contained in the covariates and makes the estimator more efficient than the one suggested by Machado and Mata (2005). This estimator has the advantage of not being simulation-based like the Machado and Mata estimator and is therefore computationally less demanding and faster. Melly (2006) shows that this procedure is numerically identical to the Machado and Mata decomposition method when the number of simulations used in the Machado and Mata procedure goes to infinity.

We construct a counterfactual for the SCST group using the characteristics of SCSTs and the earning structure for non-SCSTs here:

$$CF_{\theta}^{s} = X_{s,i}^{T}\beta_{n,\theta} \tag{9}$$

Using the abovementioned counterfactual, the decomposition of earnings gaps of the unconditional quantile function between groups s and n is as follows[7]:

$$\Delta_\theta = (Q_{n,\theta} - CF^s_\theta) + (CF^s_\theta - Q_{s,\theta}) \tag{10}$$

The first term on the right hand side represents the effect of characteristics (or the *quantile endowment effects*) and the second the effect of coefficients (or the *quantile treatment effects*).

4 Data and Descriptive Statistics

4.1 Data

The data used in this paper are from the India Human Development Survey (IHDS), which was conducted in 2004-05 by the University of Maryland in collaboration with the National Council of Applied Economic Research, New Delhi between November 2004 and October 2005.[8] The nationally representative data covers 1504 villages and 971 urban areas across 33 states and union territories of India.[9] The survey covering 41,554 households was carried out through face-to-face interviews by pairs of male and female enumerators in local languages. The respondents included a person who was knowledgeable about the household economic situation (usually the male head of the household) and an ever-married woman aged 15-49. The detailed modules of the survey canvass data on a wide range of questions relating to economic activity, income and consumption expenditure, asset ownership, social capital, education, health, marriage and fertility etc.

This paper is based on the household non-farm business module of the survey. The IHDS data do not allow us to identify the primary decision-maker regarding the household business. However, we can identify specific members in the household who worked in the business and the amount of time they spent, in terms of days per year and hours per day. Using that we try to arrive at one individual per household who we treat as being the decision-maker. The assumption we make is that the person who has spent

[7]The decomposition would be as follows if we construct a counterfactual distribution for the non-SCST group: $\Delta_\theta = (Q_{s,\theta} - CF^n_\theta) + (CF^n_\theta - Q_{n,\theta})$

[8]`http://ihds.umd.edu/index.html`

[9]Andaman and Nicobar and Lakshadweep were not included in the sample. These two union territories together account for less than 0.05 percent of India's population.

maximum number of hours in the business (i.e., days per year multiplied by hours per day) is the de-facto decision maker.[10]

For purposes of this analysis, we restrict the sample only to those states where there are at least 50 household businesses. With this restriction, we drop some of the north-eastern states and the union territories and are left with the following 22 states: Jammu and Kashmir, Himachal Pradesh, Punjab, Uttaranchal, Haryana, Delhi, Rajasthan, Uttar Pradesh, Bihar, Tripura, Assam, West Bengal, Jharkhand, Orissa, Chhattisgarh, Madhya Pradesh, Gujarat, Maharashtra, Andhra Pradesh, Karnataka, Kerala and Tamil Nadu.[11] Additionally, we consider only males in this analysis because factors affecting selection into self-employment vary along lines of gender, as well as the fact that to delineate the effect of caste we need to hold gender constant, so as not to confound the effect of overlapping identities or intersectionality. Note that we do not account, in our estimation, for selection bias, i.e. the possibility that the individuals in our sample are not self-employed randomly but on account of some individual and household characteristics.

The data canvasses information on two measures of financial performance of the business: net income and gross receipts. Our primary dependent variable is the (log) net income from business over the last 12 months.[12] Net income is computed as gross receipts less hired workers' wages less cost of materials, rent, interest on loans etc. While expenditure-based indicators have been found to be more reliable than income-based measures in developing countries- on account of recall errors, non-response and deliberate under- or over-reporting- for an analysis focusing on business performance, income from business is the most meaningful outcome variable to consider.

For explanatory variables, we have three broad categories. Firstly, we have the *individual specific variables* such as age, marital status and standard years of education completed. Secondly, we have the *household spe-*

[10]One can argue for using the characteristics of the head of the household. However, in a sufficient number of households, the head has not reported spending any time in the business.

[11]The following states have been dropped: Chandigarh, Sikkim, Arunachal Pradesh, Nagaland, Mizoram, Meghalaya, Daman and Diu, Dadra and Nagar Haveli, Pondicherry, Goa and Manipur.

[12]Results using (log) gross receipts from business over the last year as the dependent variable are available with the authors upon request.

cific variables such as wealth (in terms of ownership of assets), rural/urban status, whether someone close to or within the household is an official of the village panchayat/nagarpalika/ward committee and membership in the following: business or professional group; credit or savings group; caste association; development group or NGO and agricultural, milk or other co-operative. Finally, we have the *business specific variables* such as number of family members who worked in the business, total number of hours put into the business, work place type and industry.[13]

We construct an asset index using principal components analysis (hereafter, PCA) as a proxy for wealth of the household.[14] While other ways of constructing asset indices are possible, for instance, by assigning equal weights to all assets (an arbitrary method) or using asset prices to weight each asset (hard to implement in the absence of asset prices), we use PCA to determine weights to be assigned. PCA is a data reduction technique that extracts from a large number of variables those few orthogonal linear combinations of the variables that best capture the common information. The components are ordered so that the first principal component is the linear index of variables with the largest amount of information common to all of the variables. The second principal component has maximal variance among all unit length linear combinations that are uncorrelated to the first principal component etc. The last principal component has the smallest variance among all unit-length linear combinations of the variables. All principal components combined contain the same information as the original variables, but the important information is partitioned over the components in a particular way: the components are orthogonal, and earlier components contain more information than later components.

4.2 Descriptive Statistics

Table 1 lists the summary statistics for the full sample and for SCST and non-SCST businesses separately. It shows that of the 7363 business owners in our dataset, 7 percent are Brahmin, 42 percent are OBCs, 12.6 percent

[13]Definitions of variables are available in Appendix B.

[14]The IHDS data contain information on the ownership of the following 16 items (binary variables): cycle/bicycle, sewing machine, generator set, mixer/grinder, motorcycle/scooter, black and white television, colour television, air cooler, clock/watch, electric fan, chair or table, cot, telephone, cell phone, refrigerator and pressure cooker.

are SCs, 5 percent are STs and 33 percent are 'Others'.[15] We club OBCs, Brahmins and 'Others' together and refer to them as non-SCSTs in the analysis. Thus, of the total 7363 businesses, 1302 are owned by SCSTs (17.7 percent) and the remaining 6061 by non-SCSTs (82.3 percent). In most household and individual characteristics, there are clear differences between SCST and non-SCST self-employed business owners. Close to half of all household businesses are urban, with SCST (34 percent urban) and non-SCST businesses (53 percent urban) showing distinctly different patterns in the rural-urban distribution. Overall, 16.5 percent of businesses are owned by poor (i.e. below poverty line, or BPL) households, with corresponding proportions for SCST being close to 31 percent and for non-SCST being 13 percent. This disparity in material standard of living is also reflected in the average monthly per capita expenditure (MPCE), which for all households is Rs. 1068, for SCST Rs. 760 and for non-SCST is Rs. 1135.

Data on the total number of assets owned shows than on average, a household owns approximately 7 of the 16 assets in the data set (with a standard deviation of 3.5). Non-SCSTs own approximately 8 assets while SCSTs own around 5. T-tests show that this difference is statistically significant. For each individual asset, the proportion of SCST ownership is lower than that of non-SCSTs, with the exception of bicycles where 66 percent of SCSTs own one as compared to 64 percent of non-SCSTs.[16] Studies for the US indicate that blacks have substantially lower levels of asset ownership than whites, and that these differences contribute to racial differences in business ownership levels (e.g., Fairlie, 2006).

Table C.2 in Appendix C presents scoring coefficients (i.e. first principal component), means and standard deviations for each of the 16 assets used to construct the asset index. The last column in the table states values for scoring coefficient divided by the standard deviation. Since all asset variables take either the values of zero or one, the weights can be interpreted as the change in asset index as one moves from 0 to 1. Thus, negative values for bicycle and black and white television indicate that ceteris paribus, households owning these two goods have asset indices that are lower by 0.04 units and 0.16 units than those households that do not own these

[15]Readers familiar with NSS data should note that the 'Others' category in the IHDS dataset is different from that in the NSS data. Brahmins are not included in the 'Others' here, but are included in the 'Others' of NSS data.

[16]Table C.1 in Appendix C contains caste-wise ownership of each individual asset.

items. Similarly, owning a refrigerator raises the asset index by 0.78 units.

Following Filmer and Pritchett (2001), we use the first principal component as a measure of wealth, and divide the sample based on this into three groups: those lying in the bottom 40 percent, middle 40 percent and the top 20 percent, and call them poor, middle and rich respectively. However, it should be noted that these cutoffs are somewhat arbitrary and do not follow any standard definitions of poverty. By this definition, 65.2 percent of SCST households fall in the poor category while 34.6 percent of non-SCST households are poor. 27.4 percent and 42.7 percent of SCSTs and non-SCSTs respectively are in the middle category, and 7.4 percent of SCST households and 22.7 percent of non-SCSTs lie in the rich group.[17]

The summary of individual characteristics of the business owner (defined as the primary decision maker) reveals that he is on average 39 years old. 86 percent of business owners are married. These numbers are similar across SCST and non-SCST owners. However, average years of education for business owners, which for all is 7.8, differ significantly by caste, with 8.3 years for non-SCST and 5.7 years for SCST owners.

We examine the importance of business or professional networks for household businesses.[18] A small fraction of all owners are members of business or professional groups (8 percent), with the corresponding proportion for SCST owners being even smaller (5 percent). To the extent membership of such groups helps business positively, SCST households would be at a disadvantage. However, networks, depending their quality and strength, can have divergent effects on the decision to enter self-employment, as well as on the prospective success of the business. Allen (2000) suggests that in addition to individual characteristics and factors such as education, earnings, business cycles, liquidity constraints etc., the immediate social environment surrounding the entrepreneur, which includes the depth and breadth of family, friends and acquaintances, have an impact of self-employment choice. A social network could act as a source of venture capital, initial customers,

[17]Our calculations reveal that the first component has an eigenvalue of 1.01 and accounts for 35.7 percent of the variance. We retain only the first component since the second component has an eigenvalue of 0.33 (explaining 11.8 percent of the variance), well below the rule of thumb eigenvalue of 1.

[18]However, there is a possible endogeneity here that should be noted. Not only does a large network affect the probability of choosing self-employment, the self-employed also have incentives to develop a large network. It should be noted that we are not estimating the probability of starting a business or choosing self-employment over wage employment.

information and psychological support. We find that participation in credit or savings groups does not differ by caste, covering roughly 7 percent of owners. Membership of caste associations for all is at 14 percent, with the corresponding figure for SCST owners at 12.6 percent. Membership in development groups or NGOs is miniscule, while that in agricultural, milk or other co-operatives is slightly higher at 3.5 percent overall, and 2 percent for SCST.

The IHDS data set allows us to examine political networks as well. Overall, 11 percent of business owners have someone in, or close to, their households who has been an official of rural or urban local bodies. Interestingly, this proportion is higher for SCST (12.5 percent), than for non-SCST (10 percent). This could possibly reflect the operation of the mandatory 22.5 percent caste quotas in local bodies for SCSTs.

Coming to business characteristics, the average gross receipts for all businesses are Rs. 121262, but the receipts for non-SCST businesses (Rs. 132322) are 2.2 times higher those for SCST businesses (Rs. 60476). A similar pattern can be seen in the average net incomes with income for non-SCST businesses (Rs. 51879) being 1.94 times that for SCST businesses (Rs. 26635). Figure 1 plots the kernel density distribution of log income for SCST and non-SCST businesses. The distribution of incomes of non-SCST business lies distinctly to the right of the SCST businesses. Thus, by both indicators, non-SCST businesses are larger as compared to SCST businesses.

These gaps in incomes and receipts could be related to other characteristics, such as a) the number of family members who worked in the business: SCST businesses have greater than average number of family members working in the business (1.5), as compared to non-SCST businesses (1.39); and b) the total number of man hours put in by everyone working in the business: non-SCST businesses record 1.3 times more hours than their SCST counterparts.

About 25 percent of businesses are home-based, and this proportion does not differ by caste. 52 percent of businesses are located in a fixed workplace outside of the home, but with clear disparities by caste- non-SCST and SCST proportions being 55 and 39 respectively. To the extent a fixed workplace indicates permanency, it suggests that non-SCST businesses are more stable, less makeshift, and bigger. This dissimilarity is

mirrored in the proportions in 'moving' workplaces, where the proportion for SCST businesses is 34, and the corresponding proportion for non-SCST businesses is 19.

Distribution Across Industry Types and States

The most important industrial sector for household nonfarm businesses is wholesale, retail trade and restaurants and hotels, which include activities such as running of '*kirana*' (neighbourhood grocery) stores, other grocery stores, petty shops and general stores. Close to 54 percent of businesses are involved in this activity, with variation across caste groups, such that the proportion among non-SCST owned businesses is close to 56 and that for SCST owned businesses is 44.5. Close to 13 percent of all businesses are in manufacturing activities, and this proportion does not vary by caste. The major activities here are blacksmith (3-4 percent), carpenters (6-7 percent) and flour mill (7-8 percent). About 16 percent of businesses are in the 'community, social and personal services' sector. This includes activities such as barbers (7-8 percent), cycle repair shops (about 4 percent) and tailoring related activities (20-22 percent). These examples also corroborate our intuition that these businesses are engaged in low-end activities, and are more survivalist than entrepreneurial.

The next important activity is 'transport, storage and communication' with roughly 6.5 percent businesses in this sector, the proportion being the same across caste groups. Overall, only 4 percent of businesses are in the primary sector (agriculture, hunting, forestry and fishing), but 15 percent of SCST businesses are in this sector. Self-employment in construction is small, involving only about 2 percent of businesses. The overall proportion in 'financing, insurance, real estate and business services' is also 2 percent, but with clear caste differences, in that non-SCST proportion at 3 percent is double that of non-SCST. Businesses in mining and quarrying as well as electricity, gas and water sectors are practically non-existent, which is only to be expected, given that these highly capital intensive activities are not conducive to self-employment.

The distribution of businesses is marked by inter-state variation, with about 12 percent of businesses in Uttar Pradesh, followed by 8-9 percent in Maharashtra and Karnataka, 7-8 percent in West Bengal and Rajasthan;

4-7 percent in Bihar, Orissa, Madhya Pradesh and Gujarat.[19]

5 Results

5.1 Earnings Function Estimates

We use two alternative measures of business performance to be decomposed: net income (income, henceforth), and gross receipts. Table 1 indicates that the gap in average receipts of non-SCST-owned and SCST-owned businesses is Rs. 73845, and the gap in average net income is Rs. 25244 in current prices. This translates into a difference of 0.86 in log receipts of and of 0.77 in log income.

We present three sets of estimates. First, with only personal characteristics as explanatory variables (age, age squared, whether married or not, years of education; PC hereafter); second, to the personal characteristics equation, we add household and location characteristics as explanatory variables (whether urban or not, asset ownership, networks and state of residence; PHC hereafter); and third, in addition to personal and household characteristics, we use business characteristics (industry, whether the workplace is fixed or moving (reference category is home-based), total man-hours and number of people working in the business; PHBC hereafter).[20]

Table 2 shows the OLS estimates of the three specifications with log income as the dependent variable, for the pooled sample (all businesses), and separately for SCST and non-SCST businesses. As expected, earnings have a quadratic relationship with age such that earnings initially increase with age and start to decline thereafter (inverted-U relationship). Urban location and years of education are positively correlated with earnings. The SCST dummy is negative and significant in all specifications of the pooled model, indicating that controlling for characteristics, belonging to these marginalized groups is negatively correlated with log income. Education, asset ownership and urban location have a positive and significant association with log income. Business or professional group membership

[19]Tables on distribution of businesses across states and industries are in Appendix D.

[20]There could be some concern about endogeneity of the business characteristics variables in the third specification, but these are commonly used in decomposition exercises. In any event, the two other specifications do not share a similar concern, so readers uncomfortable with the use of business characteristics as controls could focus on the first two estimates.

is positively associated with log income for non-SCST businesses, but is insignificant for SCST businesses, suggesting that the kinds of business or professional groups that SCST households would be members of might not contribute substantially to increasing incomes, either due to their inexperience or lack of expert or specialized business knowledge. Somewhat perplexing is the fact that membership of credit or savings group is negatively and significantly associated with earnings of non-SCST businesses, but is insignificant for SCST businesses. One possible explanation for this might be that businesses that are not members of such organisations have deep pockets and are better-off in concomitant unobservable social capital. In other words, only those owners lacking in naturally favourable connections are joining such networks. Membership in caste associations has a positive and significant association for non-SCST businesses, but is negative, larger and significant for SCST businesses. This suggests that the social capital that caste networks embody seem to be associated with the two types of businesses in opposite directions. In the non-SCST case, the networks act as a source of advantage, whereas in the SCST case, these act as added sources of disadvantage.

5.2 Blinder-Oaxaca Decomposition

The aim of the decomposition exercise, as outlined in Section 3.1, is to decompose the average log receipts gap and log income gap into "explained" and "unexplained" parts, where the latter is taken as a measure of discrimination. The interpretation of the characteristics effect as "explained" and the coefficients effect as "unexplained" needs further explanation. The latter is used as an indicator for discrimination, since it is the residual gap that remains after all characteristics are accounted for. However, there are characteristics- for instance, ability or motivation- that cannot be measured but can affect the earnings gap.[21] In the presence of such unmeasurable characteristics, the unexplained component would overestimate discrimination. Similarly, it should be noted that the explained component, that is attributable to characteristics, might itself be a result of pre-market discrimination. Thus, pre-market discrimination would exhibit itself in lower educational qualifications, poorer quality of education, lower level of skills,

[21]It should be noted that *if* the average ability for the two groups is similar, then this would not be a factor in explaining the earnings gap.

and a whole range of adverse socio-economic characteristics that would result in worse characteristics for disadvantaged groups compared to the privileged groups. To the extent this is true, quantitative estimates of discrimination would underestimate the total magnitude of discrimination in the economy, especially *if* the two groups have similar average ability. Overall, therefore, the decomposition exercises provide a rough-and-ready measure of discrimination, since the "true" value of total discrimination is not amenable to neat quantitative estimation.

The results of the Blinder-Oaxaca decomposition with log income as the dependent variable are presented in Table 3.[22] Panel A of Table 3 displays the decomposition results of the 'pooled model', which uses coefficients from a pooled model over both groups as the reference coefficients. Panel B shows the results using the non-SCST coefficients, i.e., how SCST businesses would fare if they were treated like non-SCST businesses. Panel C shows the results based on SCST coefficients, i.e. how non-SCSTs would fare if they were treated like SCSTs.

If the non-discriminatory earnings structure were to be the one applicable to non-SCST businesses, (panel B), the unexplained part of the income gap would vary between roughly 71 percent (PC estimation), 39 percent (PHC estimation) and 26 percent (PHBC estimation). Using SCST coefficients (the other counterfactual non-discriminatory earnings structure), we see that the corresponding figures are 64, 22 and 11, and for the pooled model, the figures are 70, 35 and 21. Thus, the upper bound on the unexplained or discriminatory part of the income gap varies between 64-71 percent, depending on the set of explanatory variables and the assumption of the prevailing earning structure, and the lower bound on the discriminatory component varies between 26 and 11 percent. Following Banerjee and Knight (1985), we can take the geometric mean of the estimates from Panels B and C to yield a single estimate of the "unexplained" part for each of the specifications. These turn out to be 0.45, 0.20 and 0.11 respectively for each of the specifications, which correspond to unexplained or discrimination estimates of 66.5 percent, 28.3 percent and 16.5 percent.

Which of the explanatory variables contributes the most to the explained component? The lower panel of Table 3 shows the contribution of selected significant individual characteristics to the overall explained part

[22] All estimations are done using the STATA program "oaxaca" by Jann (2008).

of the income gap. In all three counterfactual scenarios, when only personal characteristics are used as predictors of earnings (PC specification), 'years of education' is the single most important explanatory factor, accounting for over 90 percent of the explained component. However, when household and location characteristics are added (PHC specification), the importance of education decreases drastically and the asset index becomes the most important component of the explained part. Urban residence and membership in business or professional group are also significant, but their absolute contribution to the explained component is small. Some state dummies are significant and contribute to a part of the explained component. When business characteristics are added (PHBC specification), the relative contribution of asset index drops and that of man-hours increases. Interestingly, the number of people working in the business is not significant under any specification.

5.3 Quantile Regressions

For quantile regressions, we use the same three specifications (PC, PHC and PHBC) of the earnings function with log income as the dependent variable that we used for the OLS regressions. We saw in the previous section that the average gap in log incomes of non-SCST-owned and SCST-owned business businesses is 0.77, which corresponds to a gap of 116 percent in raw unconditional net incomes of the two types of businesses.[23] This is instructive, but when we juxtapose this against the log income gap for the different quantiles, we see that the picture is more complex. Broadly speaking, as Figure 2 indicates, the raw uncontrolled caste gap in incomes is higher for low-income businesses as compared to high-income businesses, with the highest gap for those at the 10th percentile (300 percent). The gap reduces between the 10th percentile and the median to 81 percent (crossing the mean gap between the 37th and 38th percentile), to increase slightly and further dip to its lowest point at the 90th percentile (67 percent). From the 90th to the 99th percentile, it increases again to 100 percent, but

[23] In a quantile regression of log earnings on only the caste dummy, let $\hat{\beta}_\theta$ represent the coefficient of the caste dummy. For each quantile θ, $\hat{\beta}_\theta = lnw_{n\theta} - lnw_{s\theta} = ln(w_{n\theta}/w_{s\theta})$. Therefore, $[e^{\hat{\beta}_\theta} - 1] * 100 = [\frac{w_{n\theta} - w_{s\theta}}{w_{s\theta}}] * 100$. Since a log transformation is monotonic, $lnw_{n\theta} = [lnw_n]_\theta$ i.e. taking the log of the θth quantile of the raw earnings distribution is equivalent to the θth quantile of the log earnings distribution. Therefore, a caste earnings gap of 0.77 corresponds to a raw earnings gap of 116 percent.

remains below the mean income gap (116 percent).

Tables 4, 5 and 6 report quantile regression results for PC, PHC and PHBC specifications respectively for the pooled model at the 10th, 25th, 50th, 75th, 90th and 95th percentiles. The three sets of estimates show that controlling for various characteristics does not eliminate the significance of caste that we observed in the uncontrolled income gap in Figure 2. In all three specifications, we see that the caste dummy is negative and significant at all quantiles. For the PC and PHC specifications, it has the highest absolute magnitude for the 10th percentile, which declines till the 75th percentile and then increases to the 90th to remain roughly the same till the 95th. As successively more explanatory variables are added in the PHC and PHBC specifications, the caste dummy remains significant, but its magnitude becomes smaller, which is not surprising since there are caste differences in the significant explanatory variables, as noted in the descriptive statistics earlier. What the caste dummy indicates is that there is an independent association with net incomes, after controlling for various characteristics, which are in themselves differentiated by caste.[24]

Comparing the pooled quantile regression results with the pooled OLS results listed in Column 1 in each of the Tables 4, 5 and 6, we find that the significance of variables remains unchanged with some exceptions. In the PC specification, while the OLS coefficient for the married dummy is positive and significant, it is only significant at the 50th percentile in the quantile regression. In the PHC specification, the magnitude and significance of married dummy and memberships in caste associations, NGOs and co-operatives varies across quantiles. In the PHBC specification, total man-hours coefficient is significant according to OLS estimates and also across quantiles.

In the PC specification, 'years of education' is one of the highly significant explanatory variables, in line with the discussion of the OLS results. In the PHC specification, years of education continue to be significant, but the coefficient is smaller in magnitude, and asset index, urban residence and membership of professional and business groups become significant, similar to the OLS results. In the PHBC specification, in addition to these

[24]Tables A.1, A.2 and A.3 that present the quantile regression results separately for SCSTs and non-SCSTs show that the discussion for the pooled sample remains valid when the same specification is estimated for the two groups separately.

variables, 'total man-hours' is a significant variable in explaining log income.

5.4 Within-Group Inequality

Coefficients of the median regression reported in Tables 7, 8 and 9 indicate how earnings for the two groups depend on the different covariates. We can go further to analyze the effects of characteristics on the dispersion of earnings, or within-group inequality by examining the difference between the quantile regression coefficients at the 90th and the 10th percentiles (Melly, 2005). If the error term is independent of the characteristic, the coefficient on that covariate would not vary with the quantile, and thus, $\hat{\beta}(0.9) - \hat{\beta}(0.1)$ should not be significantly different from zero. If the difference between the 90th and 10th percentiles coefficient on a covariate is positive, a higher value of this variable increases within-group inequality.

Tables 7, 8 and 9 shows that some characteristics affecting within-group inequality are common to both groups, i.e. they are significant and have the same sign for both groups. Common factors that decrease within-group inequality (negative and significant in the inter-decile regressions for both groups) are urbanization and total man-hours. Asset index (in the PHBC specification) is the only factor that would increase within-group inequality for both groups. There are other factors that have different effects on the two groups. For instance, education would reduce within-group earnings inequality for SCST businesses but is not a significant factor for non-SCST businesses. Membership of caste associations would increase within-group inequality for SCST businesses but is not a significant factor for non-SCST businesses.

5.5 Melly Decomposition of Log Income

We conduct the decomposition separately using the three specifications.[25] Table 10 shows the summary results with the raw difference, characteristics (explained) and coefficients (unexplained or discriminatory) for each decile (10th to 90th percentile) for the case where the counterfactual distribution is constructed using the coefficients of non-SCST. As noted above,

[25]The decomposition of differences in earnings distributions is done using the STATA program "rqdeco" (Melly, 2007).

another counterfactual distribution could be constructed using the coefficients of SCST, with the 'true' value of the discriminatory component lying somewhere in-between the two estimates.[26] For all three specifications, we are now able to evaluate how the explained and unexplained proportions change across the entire income distribution. Thus, we can significantly enrich our understanding of discrimination beyond the single figure that the Blinder-Oaxaca decomposition yielded.

In all three specifications, we find that not only is the raw gap in incomes highest for the 10th percentile, the unexplained or discriminatory component is also the highest for the lower percentiles than higher ones. For instance, the difference in log incomes is 1.07 at the 10th percentile using the PC specification and continuously declines till the 80th percentile (0.56) and then slightly increases to 0.59 at the 90th percentile. A similar trend is observed using the PHC and PHBC specifications except that the log differences are somewhat smaller due to the successive inclusion of explanatory variables.

In each of the specifications, the proportion of the income gap due to differences in characteristics increases as one moves up to the higher percentiles of the distribution. In the PC specification, characteristics account for 31 percent of the gap (0.33 points out of 1.07 points) at the 10th percentile, 34 percent at the 30th percentile, 38 percent at the median, 42 percent at the 70th percentile and reducing to 38 percent at the 90th percentile. In the PHC specification, the gap attributable to characteristics increases from 66 percent at the 10th percentile to 86 percent at the 90th percentile. Finally, in the PHBC specification, the explained share of the difference increases from 75 percent at the 10th percentile to 104 percent at the 90th percentile.

An increase in the proportion of the explained component is mirrored in the decline in the observed log income gap that can be attributed to coefficients. In each of the 3 specifications, we find that the unexplained component is larger at the lower percentiles than at the higher percentiles. In the PC specification, the unexplained share falls from 69 percent at the 10th percentile to 61 percent at the 90th percentile. Using the PHC specification, the unexplained share declines from 34 percent at the 10th

[26]Melly decomposition results using the SCST coefficients are in Table A.4 in Appendix A.

percentile to 14 percent at the 90th percentile. Finally, in the PHBC specification, the unexplained component which is 25 percent at the 10th percentile drops to 5.7 percent at the 80th percentile and to -3.7 percent at the 90th percentile with the latter two values also being statistically insignificant. As is expected the proportion of the raw difference that is attributed to coefficients declines as we include more explanatory variables in moving from the PC to PHC and PHBC specifications.

Figure 3 plots the raw gap, the contribution of characteristics and that of coefficients at each percentile of the earnings distribution for the full specification (PHBC).[27] After the 86th percentile, we see that that the contribution of coefficients becomes negative (and correspondingly, the contribution of characteristics is greater than the gap in log incomes). This indicates that if the all businesses were evaluated at the same rate at which the market evaluates non-SCST characteristics, SCST businesses in the 86th percentile and above, on the basis of their characteristics, might face *negative* discrimination, and this would be true for the top 3 percentiles, as the coefficients term is significant for these businesses. Thus, the highest earning SCST businesses would be treated favourably under the non-SCST earnings structure. However, the alternative counterfactual (if all businesses were evaluated at the SCST rates of return) shows that the contribution of characteristics remains positive till the 96th percentile. Coefficients turn negative only for the 97th to 99th percentiles, but the contribution of coefficients is not significant.

Based on the counterfactual using the non-SCST rates of return, Figure 4 plots the proportion of the unexplained or discriminatory component estimated using the Blinder-Oaxaca decomposition and the Melly decomposition for the PHBC specification. This illustrates clearly that the question of "what happens where" in the earnings or income distribution is critical to the understanding of patterns of discrimination.

The Sticky Floor

Our analysis shows the presence of greater discrimination at the lower end of the conditional earnings distribution. The unexplained or discriminatory component peaks at the 10th percentile and declines steadily till the 90th percentile, again to increase between the 90th and 100th percentiles.

[27]Figures A.1 and A.2 in Appendix A correspond to the PC and PHC specifications respectively.

This suggests that conditional on observable characteristics, those starting at low-level activities might face greater constraints, but those constraints should ease for businesses with earnings up to the 90th percentile. Again, for businesses in the topmost decile, the unexplained component starts rising.

The phenomenon of higher caste gaps at lower levels of earnings is similar to the "sticky floor" phenomenon observed in the gender wage gap literature. Sticky floors are broadly defined as declining earning gaps as one moves from lower to higher quantiles of the earnings distribution (e.g., Arulampalam et al., 2007). Unlike gender wage gaps in most developed countries that are characterized by "glass ceilings" (i.e., increasing wage gaps as one moves from lower to higher quantiles), several developing countries reveal a sticky floor, for instance India (Khanna, 2013), and China (Chi and Li, 2008). This is revealed by both higher gaps at lower wage levels, as well as greater discrimination (higher unexplained component) at lower wage levels.

Sticky floors (or their opposite, glass ceilings) refer both to raw, unconditional wage gaps, as well as to gaps in the conditional wage distributions estimated by quantile regressions. Arulampalam et al. (2007) define a "sticky floor" if the 10th percentile wage gap is higher than the 25th percentile wage gap by at least two percentage points (p.170). In our results for all the three specifications, we find this to be the case. For labour markets, a possible explanation for the "sticky floor" (in contrast to the "glass ceiling" effect, which refers to increasing wage gaps as one moves from lower to higher quantiles) is statistical discrimination. For self-employment, reasons behind the sticky floor need to be explored further.

6 Discussion

Self-employment is often advocated as an alternative to wage employment, especially for marginalized groups, as they face discrimination in labour markets (see Fairlie, 2004 and references therein). This belief implicitly assumes that discrimination is somehow absent from land, credit and consumer markets that are critical for the success of entrepreneurial activities. There could be circumstances under which this might be plausible. For instance, Clark and Drinkwater (2000) investigate high rates of self-

employment among ethnic minorities in England and Wales. They find that differences in individuals' predicted earnings in paid and self-employment are strongly correlated with self-employment decisions, suggesting that the existence of discriminatory wages in the paid employment sector may 'push' minorities into entrepreneurship. Fairlie (2006) documents evidence that shows how ownership of small business has been an important factor in the economic success of earlier immigrant groups in the US (e.g. Chinese, Japanese, Jews, Italians and Greeks), and more recently Koreans have used business ownership for economic mobility.

Even for groups that face discrimination in both labour and credit markets, it is possible that eventually some of the groups end up achieving some measure of success in self-employment activities. Using a long-term panel data and comparing growth rate of earnings of young self-employed black and Hispanic men with their wage work counterparts, Fairlie (2004) finds that while black and Hispanic self-employed males initially experience slower rate of growth of earnings than wage workers, after several years they experience higher growth with rates for Hispanics being significant, but not so for blacks.

There is a larger merit presumably associated with self-employment or ownership of businesses. The low rates of business ownership are seen as the cause for the poverty of the group, as exemplified by the belief that blacks in the US are poor because they do not own businesses (Anderson, 2001). Low rates of business ownership also means that workers belonging to marginalized groups work for employers from dominant groups, instead of working in businesses owned by their own group members. This understanding underlies the DICCI's slogan "be job givers, not job seekers".

However, our results from both decomposition exercises strongly suggest that discrimination against SCSTs is very much characteristic of business activities or self-employment. Combining this with independent evidence of labour market discrimination discussed earlier, it appears that SCST individuals are subject to discrimination in both wage and self-employment.

Importance of Social Networks

In Section 5.1, we noted the differential association of caste and business networks with business earnings of the two caste groups. These differences constitute a part of the set of explanations about why some people choose

self-employment and others do not. Allen (2000) tests the hypothesis that a more effective social network increases the likelihood of self-employment, ceteris paribus and finds that differences in social networks might partially explain differences in self-employment between men and women in the US. Fairlie (2006) discusses how the lack of black traditions in business enterprises is a major cause of low levels of black business ownership. This is further suggested by the inter-generational link in business ownership through studies that show the probability of self-employment being substantially higher among children of self-employed individuals (Dunn and Holtz-Eakin, 2000).

As we noted earlier, networks can have divergent impacts on prospects of self-employment. Ethnic enclaves (i.e. "concentration of individuals from the same ethnic background within a specific geographical location", Clark and Drinkwater, 2000, *p.*606) have also been found to be an important determinant of entering self-employment, especially among immigrant populations. While on the one hand, a concentration of co-ethnics provides a captive market for producing ethnic goods that hold particular appeal for the community, on the other hand, enclaves may discourage self-employment on account of either tight competition from established immigrants or low wealth levels leading to low purchasing power among immigrants. Close-knit networks within enclaves also provide better opportunities for paid employment, thereby diverting people away from self-employment. For instance, Borjas (1986) and Toussaint-Comeau (2008) find that the size of ethnic enclaves and the amount of entrepreneurial capital within the ethnic community promotes self-employment. Conversely, Clark and Drinkwater (2000) find that individuals who live in enclaves have a lower probability of self-employment than individuals who live in less concentrated areas. Munshi (2003) finds that larger networks significantly improve paid employment outcomes for Mexican immigrants in the US.

Earnings and Wealth

Ceteris paribus, increase in SCST business ownership should lead to an increase in group wealth, but the present data-set allows us to estimate and comment only on earnings differences between SCST and non-SCST businesses. The larger question is the relationship between earnings and wealth, and whether an increase in earnings (from businesses and elsewhere) is suf-

ficient to close the wealth gap between communities. Barsky et al. (2002) analyze the racial wealth gap in the US using a non-parametric method that simulates white wealth over the black earnings distribution and find that roughly two-thirds of the mean difference in wealth between blacks and whites can be explained by differences in earnings from all sources. Among the middle-aged, 90 percent of black households have less wealth than the median white household even after controlling for the earnings differential. The relationship between earnings and wealth for different caste groups would have to be the subject matter of a future exercise.

7 Concluding Comments

The present paper focuses on one part of the IHDS data set, viz., data related to household nonfarm business, where we see clear evidence of caste-based disparities in earnings and other business characteristics, as well as the existence of discrimination. Desai and Dubey's (2011) analysis of "caste in 21st century India" based on the entire IHDS data set suggests that our findings fit into the larger pattern of persistence of caste inequalities, which results in inequalities in opportunities as well as inequalities in outcomes that their paper documents. They find an increase in civic and political participation by marginalized groups, but also document how economic and educational disparities continue to flourish. As result, they find that Brahmins are ahead of everyone else, even of other forward castes, in terms of total income and wage income.

The importance of evidence pointing to persistent disparities cannot be overemphasized. As Teltumbde (2011) suggests, in order to assess the impact of Dalit capitalism, we need to establish the improvement or deterioration in Dalit conditions in relation to the non-Dalit population. He argues "the celebration of Dalit capitalists and their Chamber of Commerce on the basis of some 100-odd individuals (out of more than 170 million) in businesses, the cumulative value of which may not even be a droplet in the corporate ocean.." ($p.10$). He points out that the presence of a few rich individuals among the Dalit communities is, historically speaking, not a new phenomenon. Thus, the celebration of Dalit capitalists should be placed in context. As Guru (2012) argues, the reason Dalit capitalists appear spectacular is because their success is juxtaposed against the mass

of poor Dalits; thus, the poverty of the group is necessary to highlight the success of the entrepreneurs. Moving to the larger point of whether the success of Dalit capitalists represents the triumph of markets (Prasad and Kamble 2013), Guru (2012) suggests that it is state and political patronage, rather than the free and competitive context of the market that provided the initial conditions for the mobility of the Dalit millionaire in India ($p.43$).

This paper establishes the presence of discrimination against Dalits in self-employment. The simultaneous existence of discrimination in wage employment, according to evidence presented earlier in the paper, presents serious challenges for public policy, further complicated by the existence of "pre-market" discrimination for Dalits which results in lower and poorer quality of educational and skill attainment. For this reason, as well as based on our finding that education would reduce within-group inequality for Dalit businesses, focusing on improvement in educational outcomes ought to be a key component of the strategy to enhance both employability as well as earnings of Dalits.

While job quotas target one part of the salaried employment market, that is not the appropriate instrument to tackle discrimination in other segments of the labour market, as well as discrimination faced by businesses. More research is needed before we can pronounce a balanced and appropriate policy package. However, international evidence presents some pointers. For instance, looking at the entrepreneurial success of migrant groups in countries such as the US and UK indicates that Fairlie's (2006) suggestion of stimulating business creation in sectors with high growth potential (e.g. construction, wholesale trade and business service) might be one effective element of public policy for promoting job creation and increasing earnings, especially in areas where marginalized groups are concentrated. There could be other such measures. What is clear is that given the various spheres marked by discrimination, an anti-discriminatory public policy, in order to be successful, needs to be multi-pronged.

References

Allen, W.D. (2000). Social networks and self-employment. *Journal of Socio-Economics*, 29, 487-501.

Anderson, C. (2001) *PowerNomics: The national plan to empower Black America*. PowerNomics Corporation of America

Arulampalam, W., Booth, A. L., and Bryan, M. L. (2007). Is there a glass ceiling over Europe? Exploring the gender pay gap across the wage distribution. *Industrial and Labor Relations Review*, 60(2), 163-186.

Banerjee, B. and Knight, J.B. (1985). Caste in Indian urban labour market. *Journal of Development Economics*, 17, 277-307

Barsky, R., Bound, J., Charles, K.K., and Lupton, J.P. (2002). Accounting for the black-white wealth gap: A nonparametric approach. *Journal of the American Statistical Association*, 97, 663-673

Bhaumik, S., and Chakrabarty, M. (2009). Earnings Inequality in India: Has the Rise of Caste and Religion Based Politics in India had an Impact? In A. Shariff and R. Basant (Eds.) *Handbook of Muslims in India*, Oxford University Press, New Delhi

Blanchflower, D.G., Levine, P.B., and Zimmerman, D.J. (2003). Discrimination in the Small-Business Credit Market. *Review of Economics and Statistics*, 85 (4), 930-43.

Blinder, A. S. (1973). Wage discrimination: Reduced form and structural estimates. *Journal of Human Resources*, 8, 436-455.

Borjas, G. (1986). The self-employment experience of immigrants. *Journal of Human Resources*, 21(4), 485-506.

Borjas, G., and Bronars, S. (1989). Consumer discrimination and self-employment. *Journal of Political Economy*, 97, 581-605.

Buchinsky, M. (1994). Changes in the U.S. wage structure 1963-1987: application of quantile regression. *Econometrica*, 62(2), 405-458

Chi, W., and Li, B. (2008). Glass ceiling or sticky floor? Examining the gender earnings differential across the earnings distribution in urban China, 1987-2004. *Journal of Comparative Economics*, 36(2), 243-263

Clark, K., and Drinkwater, S. (2000). Pushed out or pulled in? Self-employment among ethnic minorities in England and Wales. *Labour Economics*, 7, 603-628

Cotton, J. (1988). On the decomposition of wage differentials. *Review of Economics and Statistics*, 70, 236-243.

Daniels, C. (2004) *Black Power Inc: The new voice of success.* John Wiley and sons, NJ

Das, M.B., and P. Dutta (2007). Does caste matter for wages in the Indian labour market? Washington, DC, USA: The World Bank

Deininger, K., Jin, S., and Nagarajan, H. (2013). Wage Discrimination in India's Informal Labor Markets: Exploring the Impact of Caste and Gender. *Review of Development Economics*, 17(1), 130-147

Desai, S., and Dubey, A. (2011). Caste in 21st Century India: Competing Narratives. *Economic and Political Weekly*, XLVI(11), 40-49.

Deshpande, A. (2011). *The Grammar of Caste: Economic Discrimination in Contemporary India.* New Delhi: Oxford University Press

Deshpande, A., and Sharma, S. (2013). Entrepreneurship or Survival? Caste and Gender of Small Business in India. *Economic and Political Weekly*, XLVIII (28), 38-49

Dunn, T.A., and Holtz-Eakin, D.J. (2000). Financial capital, human capital, and the transition to self-employment: evidence from intergenerational links. *Journal of Labor Economics*, 18(2), 282-305.

Fairlie, R. (2004). Does Business Ownership Provide a Source of Upward Mobility for Blacks and Hispanics? in Douglas Holtz-Eakin and Harvey S. Rosen (eds.) *Public Policy and the Economics of Entrepreneurship*, MIT Press.

Fairlie, R. (2006). Entrepreneurship among Disadvantaged Groups: An Analysis of the Dynamics of Self-Employment by Gender, Race and Education in Simon C. Parker, Zoltan J. Acs, and David R. Audretsch (eds.) *International Handbook Series on Entrepreneurship*, Vol. 2. New York: Springer

Fairlie, R., and Robb, A.M. (2007). Why are black-owned businesses less successful than white-owned business? The role of families, inheritances, and business human capital. *Journal of Labor Economics*, 25, 289-323.

Filmer, D., and Pritchett, L. (2001). Estimating wealth effect without expenditure data-or tears: An application to educational enrollments in states of India. *Demography*, 38, 115-132.

Guru, G. (2012). Rise of the Dalit Millionaire: A low intensity spectacle. *Economic and Political Weekly*, XLVII(50), 41-49.

Hill, L. W. and Rabig, J. (Eds.) (2012). *The Business of Black Power: Community Development, Capitalism, and Corporate Responsibility in Postwar America*. University of Rochester Press.

Hout, M., and Rosen, H. (2000). Self-employment, family background, and race.*Journal of Human Resources*, 35(4), 670-692.

Iyer, L., Khanna, T. and Varshney, A. (2013). Caste and Entrepreneurship in India. *Economic and Political Weekly*, XLVIII(6), 52-60

Jann, Ben (2008). The Blinder-Oaxaca decomposition for linear regression models. *STATA Journal*, 8(4), 453-479

Jodhka, S. (2010). Dalits in Business: Self-Employed Scheduled Castes in Northwest India. Indian Institute of Dalit Studies Working Paper, 4(2)

Khanna, S. (2013). Gender wage discrimination in India: glass ceiling or sticky floor. Centre for Development Economics Working Paper 214, Delhi School of Economics

Koenker, R. and G. Bassett (1978). Regression Quantiles. *Econometrica*, 46, 33-50.

Machado, J. and J. Mata (2005). Counterfactual Decomposition of Changes in Wage Distributions using Quantile Regression. *Journal of Applied Econometrics*, 20, 445-465.

Madheswaran, S., and Attewell, P. (2007). Caste Discrimination in the Indian Urban Labour Market: Evidence from the National Sample Survey. *Economic and Political Weekly*, 4146-4153.

Melly, B. (2005). Decomposition of differences in distribution using quantile regression. *Labour Economics*, 12(4), 577-590.

Melly, B. (2006). Estimation of counterfactual distributions using quantile regression. *Review of Labor Economics*, 68, 543-572.

Melly, B. (2007). Rqdeco: A Stata Module to Decompose Differences in Distribution. Mimeo, University of St. Gallen

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review*, 14, 693-709.

Prasad, C.B., and Kamble, M. (2013). Manifesto to End Caste: Push Capitalism and Industrialization to Eradicate this Pernicious System. Times of India, January 23.

Singh, A. (2011). Farm income inequality and the role of caste: new evidence from India. *Economics Bulletin*, 31(4), 2847-2862

Teltumbde, A. (2011). Dalit Capitalism and Pseudo Dalitism. *Economic and Political Weekly*, XLVI(10), 10-11.

Thorat, S., and Newman, K. (Eds.) (2010). *Blocked by Caste: Economic Discrimination in Modern India*. New Delhi: Oxford University Press.

Thorat, S. and Sadana, N. (2009). Caste and Ownership of Private Enterprises. *Economic and Political Weekly*, XLIV (23), 13-16

Villemez, W.J., and Beggs, J.J. (1984). Black Capitalism and Black Inequality: Some Sociological Considerations. *Social Forces*, 63(1), 117-144

Figure 1: Kernel Density of Log Income

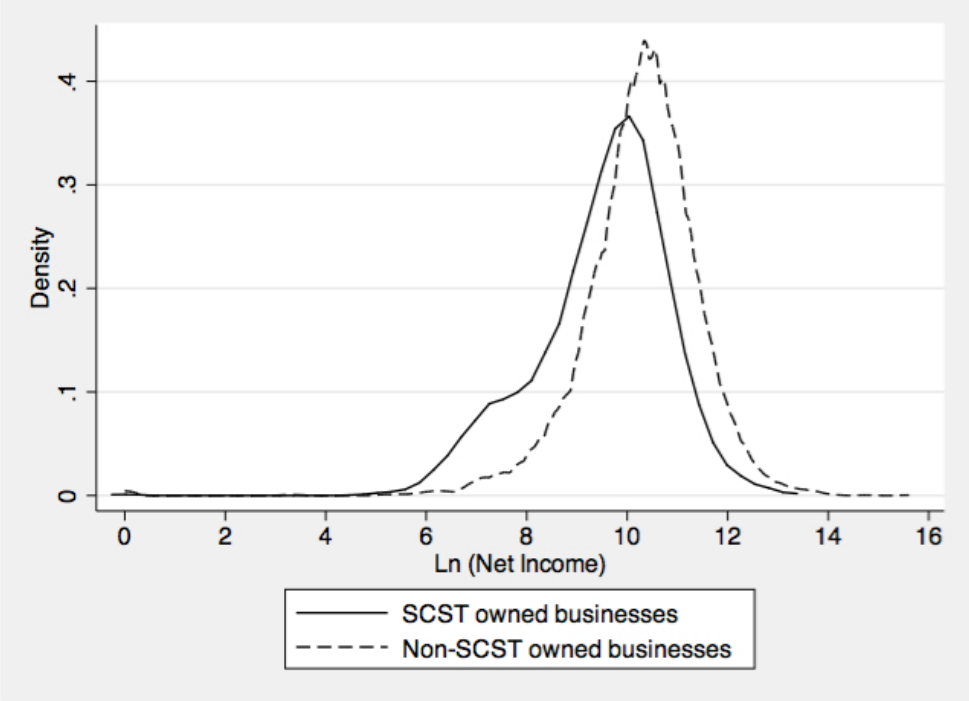

Figure 2: Caste Log Income Gap Across Percentiles & Average Gap

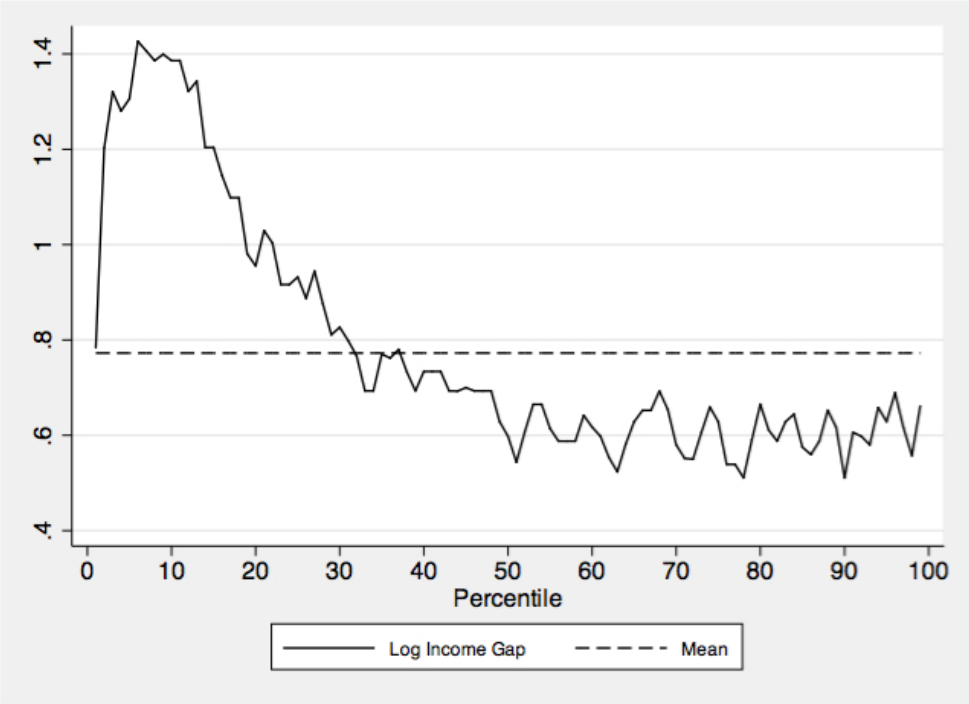

Figure 3: Melly Decomposition of Log Income Gap: PHBC specification (non-SCST coefficients)

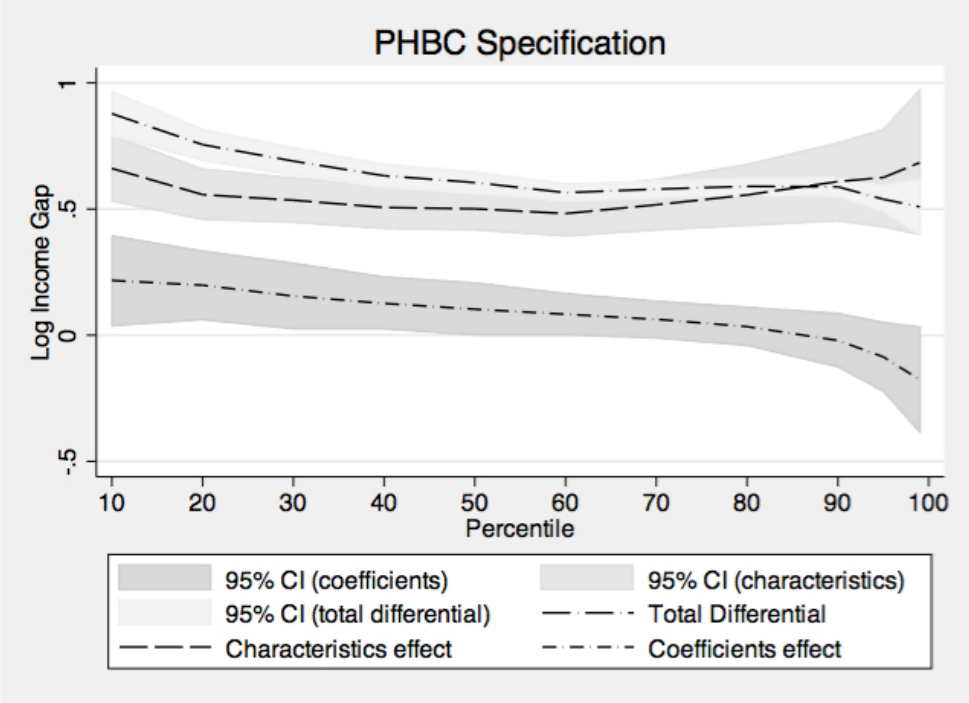

Figure 4: Blinder-Oaxaca & Melly proportion of discrimination estimates: PHBC specification (non-SCST coefficients)

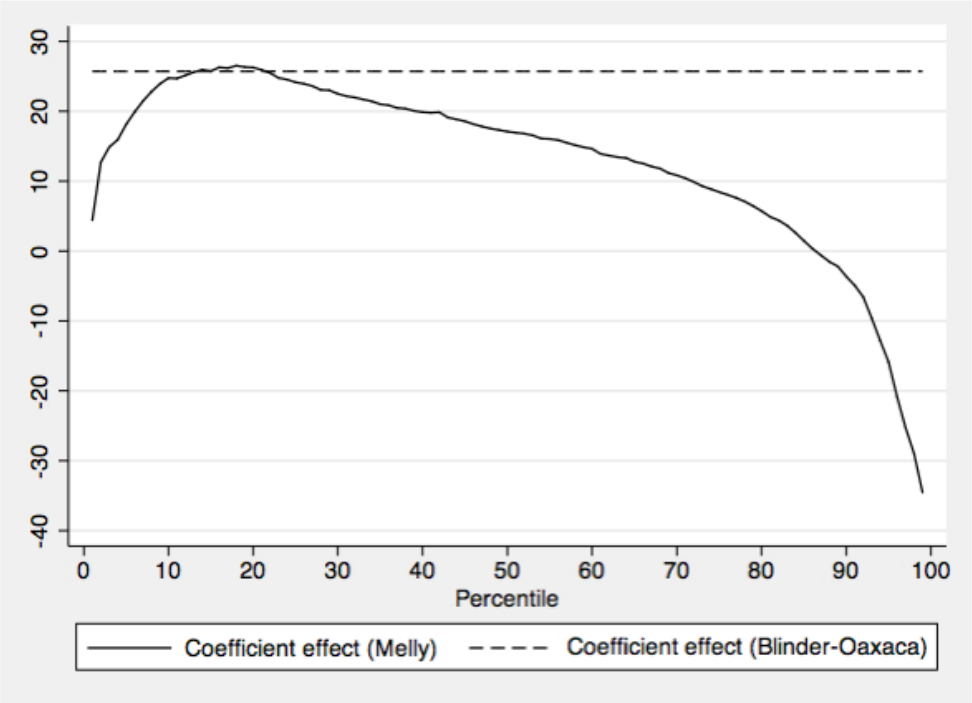

Table 1: Summary Statistics

Variable	All businesses	SCST owned businesses	Non-SCST owned businesses
Outcome Variables:			
Gross Receipts (in Rs.)	121261.9	60476.2	134321.6
	(344826.7)	(102119.8)	(375836.3)
Net Income (in Rs.)	47415.3	26635.2	51879.2
	(102084.9)	(38946.05)	(110553.3)
Explanatory Variables:			
Individual characteristics			
Age (in years)	39.15	38.6	39.27
	(12.41)	(12.52)	(12.39)
Married	0.86	0.86	0.86
	(0.35)	(0.34)	(0.35)
Years of Education	7.83	5.67	8.29
	(4.65)	(4.58)	(4.53)
Household characteristics			
SCST	17.68		
	(0.38)		
Urban location	0.49	0.34	0.53
	(0.5)	(0.47)	(0.5)
Monthly Per Capita Expenditure (in Rs.)	1068.8	760.2	1135.1
	(1105.2)	(899.3)	(1133.8)
Poor (below poverty line)	0.17	0.31	0.14
	(0.37)	(0.46)	(0.34)
Business or professional group membership	0.08	0.06	0.09
	(0.28)	(0.23)	(0.29)
Credit or savings group membership	0.07	0.07	0.07
	(0.26)	(0.26)	(0.26)
Caste association membership	0.14	0.13	0.15
	(0.35)	(0.33)	(0.35)
Development group/NGO membership	0.02	0.01	0.02
	(0.13)	(0.09)	(0.15)
Co-operative membership	0.04	0.02	0.04
	(0.18)	(0.15)	(0.19)
Village Panchayat or Ward Committee	0.11	0.12	0.11
	(0.31)	(0.33)	(0.31)
Business Characteristics			
Number of workers	1.41	1.52	1.39
	(0.74)	(0.89)	(0.71)
Number of hours	2609.74	2079.31	2723.68
	(1640.19)	(1498.21)	(1647.08)
Workplace: home-based	0.25	0.25	0.24
	(0.43)	(0.44)	(0.43)
Workplace: other fixed	0.52	0.4	0.55
	(0.5)	(0.49)	(0.5)
Workplace: moving	0.22	0.35	0.2
	(0.42)	(0.48)	(0.4)

Note: Standard errors are reported in parentheses. Net income is defined as gross receipts less hired workers' wages less all other expenses such as costs of materials, rent, interest on loans etc.

Table 2: OLS (pooled sample & separately for caste groups), for different specifications

Dependent variable: Ln(income)	**PC specification**			**PHC specification**			**PHBC specification**		
	Col.1 Pooled Sample	Col.2 SCST	Col.3 Non-SCST	Col.4 Pooled Sample	Col.5 SCST	Col.6 Non-SCST	Col.7 Pooled Sample	Col.8 SCST	Col.9 Non-SCST
SCST	-0.54***			-0.27***			-0.15***		
	(0.04)			(0.03)			(0.03)		
Age	0.04***	0.05***	0.04***	0.03***	0.04***	0.03***	0.03***	0.04***	0.03***
	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Age square/100	-0.04***	-0.04**	-0.04***	-0.04***	-0.04***	-0.04***	-0.03***	-0.05***	-0.03***
	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.01)
Married	0.08*	-0.18	0.14***	0.13***	-0.12	0.17***	0.10***	-0.09	0.14***
	(0.05)	(0.12)	(0.05)	(0.04)	(0.11)	(0.05)	(0.04)	(0.10)	(0.04)
Years of Education	0.08***	0.10***	0.08***	0.02***	0.02***	0.02***	0.02***	0.02**	0.02***
	(0.00)	(0.01)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.01)	(0.00)
Index of asset ownership				0.44***	0.48***	0.43***	0.35***	0.34***	0.35***
				(0.02)	(0.04)	(0.02)	(0.02)	(0.04)	(0.02)
Urban location				0.44***	0.49***	0.42***	0.29***	0.24***	0.30***
				(0.03)	(0.06)	(0.03)	(0.02)	(0.05)	(0.03)
Business or professional group membership				0.16***	-0.01	0.18***	0.12***	0.08	0.12**
				(0.05)	(0.18)	(0.05)	(0.04)	(0.11)	(0.05)
Credit or savings group membership				-0.13***	-0.10	-0.14***	-0.15***	-0.13	-0.15***
				(0.05)	(0.13)	(0.05)	(0.04)	(0.09)	(0.05)
Caste association membership				0.05	-0.24**	0.10**	0.03	-0.18**	0.07*
				(0.04)	(0.10)	(0.04)	(0.04)	(0.08)	(0.04)
Development group/NGO membership				0.03	0.27	0.01	0.11	0.45	0.08
				(0.08)	(0.29)	(0.08)	(0.08)	(0.31)	(0.08)
Co-operative membership				-0.13	0.05	-0.16*	-0.06	0.02	-0.10
				(0.08)	(0.21)	(0.09)	(0.08)	(0.20)	(0.08)
Village panchayat or ward committee				-0.03	-0.09	-0.01	-0.02	-0.06	-0.01
				(0.04)	(0.08)	(0.05)	(0.04)	(0.07)	(0.05)
Total number of hours							0.00***	0.00***	0.00***
							(0.00)	(0.00)	(0.00)
Number of workers							-0.12***	-0.14***	-0.10***
							(0.02)	(0.04)	(0.02)
Workplace-other fixed							0.26***	0.27***	0.26***
							(0.03)	(0.07)	(0.03)
Workplace-moving							0.12***	0.23***	0.11***
							(0.04)	(0.07)	(0.04)
Constant	8.52***	7.98***	8.52***	9.03***	9.33***	8.90***	8.13***	7.99***	8.29***
	(0.13)	(0.31)	(0.14)	(0.14)	(0.33)	(0.16)	(0.16)	(0.36)	(0.19)
Observations	7342	1299	6043	7223	1286	5937	7101	1253	5848
R^2	0.156	0.120	0.105	0.378	0.485	0.315	0.465	0.633	0.392
State dummies	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Industry dummies	No	No	No	No	No	No	Yes	Yes	Yes

Note: Dependent variable is log of income. Robust standard errors are reported in parentheses. PC specification includes only personal characteristics; PHC specification includes personal, household and location characteristics and PHBC specification includes personal, household, location and business characteristics. State dummy variables included in PHC and PHBC specifications. Industry dummy variables included in PHBC specification. *** significant at 1%,** significant at 5%,* significant at 10%.

Table 3: Blinder-Oaxaca Decomposition of Log Income

Ln (income)	Panel A Pooled			Panel B Non-SCST Coefficients			Panel C SCST Coefficients		
Variable	Col.1 PC	Col.2 PHC	Col.3 PHBC	Col.4 PC	Col.5 PHC	Col.6 PHBC	Col.7 PC	Col.8 PHC	Col.9 PHBC
Difference	0.68***	0.67***	0.65***	0.68***	0.67***	0.65***	0.68***	0.67***	0.65***
	(0.08)	(0.07)	(0.07)	(0.08)	(0.07)	(0.07)	(0.08)	(0.07)	(0.07)
Explained	0.21***	0.43***	0.51***	0.20***	0.41***	0.48***	0.24***	0.52***	0.57***
	(0.02)	(0.05)	(0.06)	(0.03)	(0.05)	(0.06)	(0.04)	(0.07)	(0.07)
Unexplained	0.47***	0.24***	0.14***	0.48***	0.26***	0.17***	0.43***	0.15***	0.07
	(0.08)	(0.05)	(0.04)	(0.08)	(0.06)	(0.05)	(0.07)	(0.05)	(0.05)
% Unexplained	69.61	35.17	20.91	70.74	38.98	25.72	63.82	22.11	11.54
Contribution to the explained component:									
Years of education	0.20***	0.04***	0.03***	0.19***	0.04***	0.04**	0.23***	0.08***	0.04**
	(0.02)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.04)	(0.03)	(0.02)
Urban		0.08***	0.05***		0.08***	0.05***		0.10***	0.06**
		(0.01)	(0.01)		(0.01)	(0.01)		(0.02)	(0.02)
Index of asset ownership		0.27***	0.20***		0.27***	0.21***		0.25***	0.17***
		(0.03)	(0.02)		(0.03)	(0.02)		(0.04)	(0.04)
Business & professional group membership		0.01**	0.01*		0.01**	0.01*		0.00	0.00
		(0.00)	(0.00)		(0.00)	(0.00)		(0.00)	(0.01)
Total number of hours			0.12***			0.11***			0.18***
			(0.02)			(0.02)			(0.03)
Number of workers			0.00			0.00			0.01
			(0.00)			(0.00)			(0.01)

Note: Robust standard errors in parentheses. PC specification includes only personal characteristics; PHC specification includes personal, household and location characteristics and PHBC specification includes personal, household, location and business characteristics. *** significant at 1%,** significant at 5%,* significant at 10%.

Table 4: Quantile Regression: PC Specification (Pooled Sample)

Dependent variable: Ln(income)	Col.1 OLS	Col.2 Q10	Col.3 Q25	Col.4 Q50	Col.5 Q75	Col.6 Q90	Col.7 Q95
SCST	-0.54***	-1.08***	-0.69***	-0.46***	-0.33***	-0.39***	-0.38***
	(0.04)	(0.10)	(0.06)	(0.05)	(0.03)	(0.04)	(0.07)
Age	0.04***	0.06***	0.06***	0.04***	0.04***	0.04***	0.04***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Age squared/100	-0.04***	-0.06***	-0.06***	-0.03***	-0.04***	-0.03***	-0.04***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Married	0.08*	0.05	0.07	0.09*	0.03	0.05	0.06
	(0.05)	(0.09)	(0.07)	(0.05)	(0.05)	(0.05)	(0.07)
Years of Education	0.08***	0.09***	0.08***	0.08***	0.08***	0.08***	0.09***
	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
Constant	8.52***	6.94***	7.83***	8.83***	9.33***	9.78***	9.91***
	(0.13)	(0.24)	(0.19)	(0.13)	(0.11)	(0.14)	(0.16)
Observations	7342	7342	7342	7342	7342	7342	7342

Note: Robust standard errors are reported in parentheses for OLS. Quantile regression standard errors in parentheses are bootstrapped using 100 replications. PC specification includes only personal characteristics. *** significant at 1%,** significant at 5%,* significant at 10%.

Table 5: Quantile Regression: PHC Specification (Pooled Sample)

Dependent variable: Ln(income)	Col.1 OLS	Col.2 Q10	Col.3 Q25	Col.4 Q50	Col.5 Q75	Col.6 Q90	Col.7 Q95
SCST	-0.27***	-0.44***	-0.37***	-0.28***	-0.14***	-0.17***	-0.20***
	(0.03)	(0.07)	(0.05)	(0.04)	(0.02)	(0.04)	(0.05)
Age	0.03***	0.05***	0.05***	0.03***	0.02***	0.02***	0.02**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Age squared/100	-0.04***	-0.05***	-0.06***	-0.03***	-0.02***	-0.02***	-0.03**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Married	0.13***	0.12	0.13***	0.13***	0.05	0.07	0.05
	(0.04)	(0.08)	(0.05)	(0.05)	(0.04)	(0.05)	(0.07)
Years of Education	0.02***	0.02***	0.01***	0.02***	0.02***	0.02***	0.02***
	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
Index of asset ownership	0.44***	0.39***	0.40***	0.40***	0.45***	0.47***	0.53***
	(0.02)	(0.03)	(0.02)	(0.01)	(0.02)	(0.02)	(0.03)
Urban location	0.44***	0.70***	0.50***	0.38***	0.29***	0.26***	0.17***
	(0.03)	(0.05)	(0.03)	(0.03)	(0.02)	(0.03)	(0.05)
Business or professional group membership	0.16***	0.21***	0.20***	0.16***	0.12***	0.14**	0.12
	(0.05)	(0.08)	(0.05)	(0.04)	(0.04)	(0.06)	(0.08)
Credit or savings group membership	-0.13***	-0.04	-0.11*	-0.14***	-0.13***	-0.14**	-0.10
	(0.05)	(0.08)	(0.06)	(0.05)	(0.05)	(0.06)	(0.08)
Caste association membership	0.05	-0.06	-0.00	0.06	0.08**	0.07	0.15**
	(0.04)	(0.08)	(0.05)	(0.04)	(0.04)	(0.05)	(0.07)
Development group/NGO membership	0.03	0.22**	-0.08	0.02	-0.08	-0.20**	-0.15
	(0.08)	(0.11)	(0.11)	(0.08)	(0.07)	(0.10)	(0.17)
Co-operative membership	-0.13	-0.31**	-0.17	-0.07	-0.03	0.05	0.06
	(0.08)	(0.16)	(0.11)	(0.06)	(0.05)	(0.09)	(0.12)
Village panchayat or ward committee	-0.03	-0.06	-0.05	-0.01	0.02	0.03	-0.07
	(0.04)	(0.08)	(0.06)	(0.04)	(0.04)	(0.05)	(0.08)
Constant	9.03***	7.59***	8.24***	9.25***	9.92***	10.29***	10.34***
	(0.14)	(0.23)	(0.20)	(0.17)	(0.14)	(0.19)	(0.25)
Observations	7223	7223	7223	7223	7223	7223	7223

Note: Robust standard errors are reported in parentheses for OLS. Quantile regression standard errors in parentheses are bootstrapped using 100 replications. PHC specification includes personal, household and location characteristics. *** significant at 1%,** significant at 5%,* significant at 10%.

Table 6: Quantile Regression: PHBC Specification (Pooled Sample)

Dependent variable: Ln(income)	Col.1 OLS	Col.2 Q10	Col.3 Q25	Col.4 Q50	Col.5 Q75	Col.6 Q90	Col.7 Q95
SCST	-0.15***	-0.10*	-0.18***	-0.20***	-0.09***	-0.18***	-0.14***
	(0.03)	(0.06)	(0.04)	(0.04)	(0.02)	(0.05)	(0.05)
Age	0.03***	0.06***	0.04***	0.02***	0.02***	0.01**	0.01*
	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)
Age squared/100	-0.03***	-0.06***	-0.04***	-0.03***	-0.02***	-0.01**	-0.02*
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Married	0.10***	0.10	0.12**	0.13***	0.09**	0.06	0.04
	(0.04)	(0.07)	(0.05)	(0.04)	(0.04)	(0.05)	(0.07)
Years of Education	0.02***	0.02***	0.01***	0.01***	0.02***	0.02***	0.02***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
Index of asset ownership	0.35***	0.29***	0.30***	0.33***	0.38***	0.38***	0.43***
	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)
Urban location	0.29***	0.42***	0.34***	0.26***	0.22***	0.20***	0.14***
	(0.02)	(0.05)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)
Business or professional group membership	0.12***	0.14**	0.09	0.15***	0.08**	0.17**	0.15**
	(0.04)	(0.06)	(0.06)	(0.05)	(0.04)	(0.07)	(0.07)
Credit or savings group membership	-0.15***	-0.02	-0.07	-0.13***	-0.15***	-0.18***	-0.21**
	(0.04)	(0.08)	(0.06)	(0.05)	(0.04)	(0.07)	(0.10)
Caste association membership	0.03	-0.04	-0.03	0.04	0.08**	0.10*	0.06
	(0.04)	(0.07)	(0.05)	(0.04)	(0.04)	(0.05)	(0.06)
Development group/NGO membership	0.11	0.16	0.09	0.03	0.11	0.02	0.11
	(0.08)	(0.14)	(0.07)	(0.08)	(0.09)	(0.13)	(0.20)
Co-operative membership	-0.06	-0.14	-0.17	-0.00	0.01	0.10	0.20
	(0.08)	(0.15)	(0.11)	(0.06)	(0.07)	(0.11)	(0.13)
Village panchayat or ward committee	-0.02	-0.04	-0.02	0.01	0.00	-0.03	-0.10
	(0.04)	(0.07)	(0.04)	(0.04)	(0.04)	(0.05)	(0.06)
Total number of hours	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Number of workers	-0.12***	-0.22***	-0.15***	-0.11***	-0.06***	-0.04**	-0.05*
	(0.02)	(0.05)	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)
Workplace-other fixed	0.26***	0.39***	0.34***	0.25***	0.15***	0.15***	0.15***
	(0.03)	(0.05)	(0.05)	(0.03)	(0.03)	(0.04)	(0.06)
Workplace-moving	0.12***	0.22***	0.16***	0.12***	0.01	-0.01	-0.04
	(0.04)	(0.07)	(0.05)	(0.04)	(0.03)	(0.05)	(0.06)
Constant	8.13***	6.18***	7.28***	8.35***	9.26***	9.86***	9.93***
	(0.16)	(0.33)	(0.27)	(0.16)	(0.16)	(0.18)	(0.27)
Observations	7101	7101	7101	7101	7101	7101	7101

Note: Robust standard errors are reported in parentheses for OLS. Quantile regression standard errors in parentheses are bootstrapped using 100 replications. PHBC specification includes personal, household, location and business characteristics. *** significant at 1%,** significant at 5%,* significant at 10%.

Table 7: Median Regression Coefficients & Inter-decile Ranges: PC Specification

Dependent variable: Ln(income)	Col.1 Q50	Col.2 Q50	Col.3 Q90-Q10	Col.4 Q90-Q10
	SCST	Non-SCST	SCST	Non-SCST
Age	0.03	0.04***	-0.01	-0.03*
	(0.02)	(0.01)	(0.05)	(0.02)
Age squared/100	-0.02	-0.04***	0.01	0.04**
	(0.02)	(0.01)	(0.05)	(0.02)
Married	-0.09	0.15***	0.04	0.04
	(0.08)	(0.05)	(0.41)	(0.09)
Years of Education	0.09***	0.08***	-0.05**	0.01
	(0.01)	(0.00)	(0.02)	(0.01)
Constant	8.57***	8.74***	3.39***	2.82***
	(0.34)	(0.13)	(0.82)	(0.34)
Observations	1299	6043	1299	6043

Note: Standard errors in parentheses are bootstrapped using 100 replications. PC specification includes only personal characteristics. *** significant at 1%,** significant at 5%,* significant at 10%.

Table 8: Median Regression Coefficients & Inter-decile Ranges: PHC Specification

Dependent variable: Ln(income)	Col.1 Q50	Col.2 Q50	Col.3 Q90-Q10	Col.4 Q90-Q10
	SCST	Non-SCST	SCST	Non-SCST
Age	0.02	0.03***	-0.04	-0.03**
	(0.02)	(0.01)	(0.03)	(0.01)
Age squared/100	-0.03	-0.04***	0.05	0.03**
	(0.02)	(0.01)	(0.03)	(0.02)
Married	0.01	0.14***	0.06	-0.04
	(0.11)	(0.05)	(0.20)	(0.09)
Years of Education	0.02***	0.01***	-0.02	0.00
	(0.01)	(0.00)	(0.02)	(0.01)
Index of asset ownership	0.44***	0.40***	0.07	0.09**
	(0.06)	(0.01)	(0.11)	(0.04)
Urban location	0.42***	0.37***	-0.49***	-0.43***
	(0.08)	(0.03)	(0.15)	(0.06)
Business or professional group membership	0.13	0.17***	0.15	-0.07
	(0.14)	(0.05)	(0.28)	(0.12)
Credit or savings group membership	0.01	-0.17***	-0.09	-0.07
	(0.14)	(0.06)	(0.23)	(0.10)
Caste association membership	-0.29**	0.10**	0.53***	0.05
	(0.13)	(0.05)	(0.20)	(0.10)
Development group/NGO membership	-0.09	0.02	-0.76	-0.41***
	(0.46)	(0.10)	(0.59)	(0.15)
Co-operative membership	0.04	-0.08	0.17	0.44**
	(0.30)	(0.07)	(0.41)	(0.20)
Village panchayat or ward committee	-0.04	0.02	-0.11	0.20*
	(0.09)	(0.05)	(0.23)	(0.12)
Constant	9.82***	9.11***	2.21***	2.82***
	(0.41)	(0.17)	(0.75)	(0.30)
Observations	1286	5937	1286	5937

Note: Standard errors in parentheses are bootstrapped using 100 replications. PHC specification includes personal, household and location characteristics. *** significant at 1%,** significant at 5%,* significant at 10%.

Table 9: Median Regression Coefficients & Inter-decile Ranges: PHBC Specification

Dependent variable: Ln(income)	Col.1 Q50	Col.2 Q50	Col.3 Q90-Q10	Col.4 Q90-Q10
	SCST	Non-SCST	SCST	Non-SCST
Age	0.03***	0.02***	-0.04	-0.04***
	(0.01)	(0.01)	(0.03)	(0.01)
Age squared/100	-0.04***	-0.03***	0.05	0.05***
	(0.01)	(0.01)	(0.04)	(0.02)
Married	-0.01	0.15***	-0.32*	0.01
	(0.09)	(0.04)	(0.17)	(0.09)
Years of Education	0.01	0.01***	-0.03**	0.01
	(0.01)	(0.00)	(0.01)	(0.01)
Index of asset ownership	0.29***	0.34***	0.19**	0.10***
	(0.05)	(0.02)	(0.09)	(0.04)
Urban location	0.23***	0.28***	-0.25*	-0.26***
	(0.07)	(0.03)	(0.14)	(0.06)
Business or professional group membership	0.21	0.15***	0.09	0.06
	(0.17)	(0.05)	(0.24)	(0.09)
Credit or savings group membership	-0.16	-0.16***	-0.13	-0.08
	(0.10)	(0.05)	(0.22)	(0.12)
Caste association membership	-0.19	0.06	0.39**	0.10
	(0.12)	(0.04)	(0.19)	(0.09)
Development group/NGO membership	0.23	-0.01	0.10	-0.23
	(0.48)	(0.08)	(0.74)	(0.17)
Co-operative membership	0.08	-0.01	-0.20	0.33
	(0.23)	(0.07)	(0.67)	(0.20)
Village panchayat or ward committee	-0.05	0.03	-0.23	0.12
	(0.08)	(0.04)	(0.20)	(0.10)
Total number of hours	0.00***	0.00***	-0.00*	-0.00***
	(0.00)	(0.00)	(0.00)	(0.00)
Number of workers	-0.14***	-0.07***	0.16	0.17***
	(0.05)	(0.02)	(0.11)	(0.05)
Workplace-other fixed	0.43***	0.24***	-0.12	-0.24***
	(0.09)	(0.03)	(0.16)	(0.08)
Workplace-moving	0.27***	0.13***	-0.27	-0.20**
	(0.09)	(0.04)	(0.16)	(0.09)
Constant	8.24***	8.69***	2.94***	3.73***
	(0.59)	(0.20)	(0.87)	(0.49)
Observations	1253	5848	1253	5848

Note: Standard errors in parentheses are bootstrapped using 100 replications. PHBC specification includes personal, household, location and business characteristics. *** significant at 1%,** significant at 5%,* significant at 10%.

Table 10: Melly Decomposition of Log Income (Using the Non-SCST Coefficients)

	Panel A PC			Panel B PHC			Panel C PHBC		
Decile	Col.1 Difference	Col.2 Characteristics	Col.3 Coefficients	Col.4 Difference	Col.5 Characteristics	Col.6 Coefficients	Col.7 Difference	Col.8 Characteristics	Col.9 Coefficients
10	1.07***	0.33***	0.74***	0.95***	0.63***	0.32***	0.88***	0.66***	0.22**
	(0.05)	(0.09)	(0.1)	(0.04)	(0.07)	(0.08)	(0.04)	(0.07)	(0.09)
20	0.79***	0.27***	0.52***	0.79***	0.51***	0.28***	0.76***	0.56***	0.20***
	(0.03)	(0.06)	(0.07)	(0.03)	(0.06)	(0.06)	(0.03)	(0.05)	(0.07)
30	0.67***	0.23***	0.44***	0.69***	0.46***	0.23***	0.69***	0.53***	0.16**
	(0.02)	(0.05)	(0.06)	(0.03)	(0.05)	(0.06)	(0.03)	(0.04)	(0.06)
40	0.62***	0.23***	0.39***	0.63***	0.45***	0.19***	0.63***	0.51***	0.13**
	(0.02)	(0.05)	(0.05)	(0.02)	(0.05)	(0.05)	(0.02)	(0.04)	(0.05)
50	0.59***	0.22***	0.37***	0.60***	0.45***	0.15***	0.60***	0.50***	0.10**
	(0.02)	(0.05)	(0.05)	(0.02)	(0.05)	(0.05)	(0.02)	(0.04)	(0.05)
60	0.57***	0.21***	0.35***	0.58***	0.45***	0.12***	0.57***	0.48***	0.08**
	(0.02)	(0.05)	(0.04)	(0.02)	(0.05)	(0.04)	(0.02)	(0.05)	(0.04)
70	0.57***	0.23***	0.34***	0.58***	0.47***	0.11**	0.58***	0.52***	0.06*
	(0.02)	(0.06)	(0.04)	(0.02)	(0.05)	(0.04)	(0.02)	(0.05)	(0.04)
80	0.56***	0.24***	0.33***	0.58***	0.50***	0.08*	0.59***	0.56***	0.03
	(0.02)	(0.07)	(0.05)	(0.02)	(0.05)	(0.04)	(0.02)	(0.06)	(0.04)
90	0.59***	0.23***	0.36***	0.59***	0.51***	0.08*	0.59***	0.61***	-0.02
	(0.02)	(0.07)	(0.06)	(0.02)	(0.06)	(0.05)	(0.02)	(0.08)	(0.05)

Note: Standard errors in parentheses are based on bootstrapping with 100 replications. PC specification includes only personal characteristics; PHC specification includes personal, household and location characteristics and PHBC specification includes personal, household, location and business characteristics. *** significant at 1%,** significant at 5%,* significant at 10%.

A Supplementary Figures & Tables

Figure A.1: Decomposition of Log Income Gap: PC specification (non-SCST coefficients)

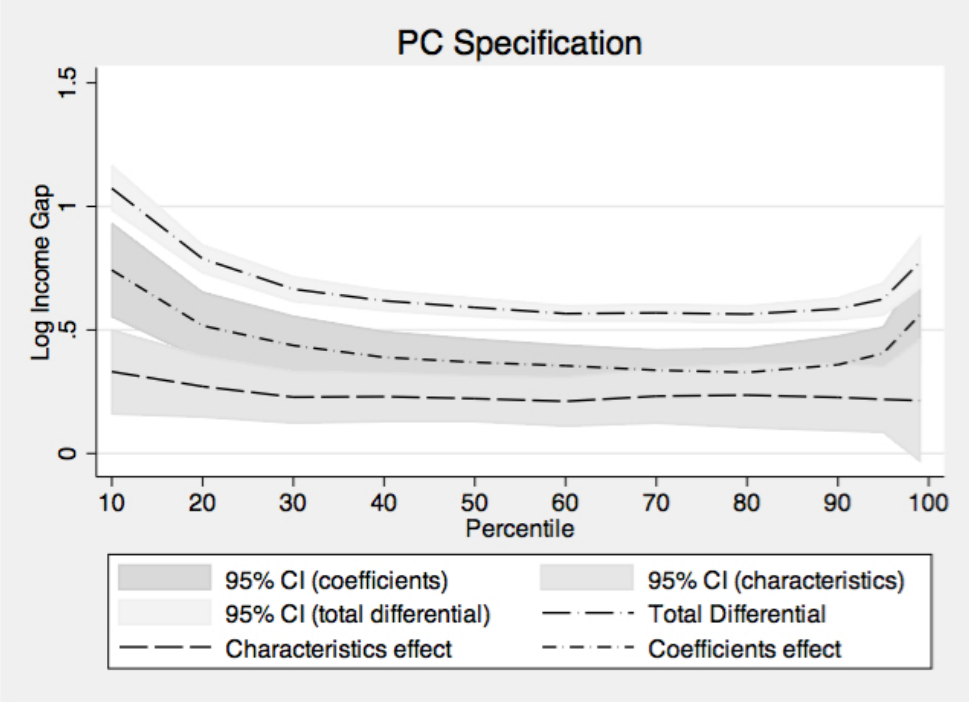

Figure A.2: Decomposition of Log Income Gap: PHC specification (non-SCST coefficients)

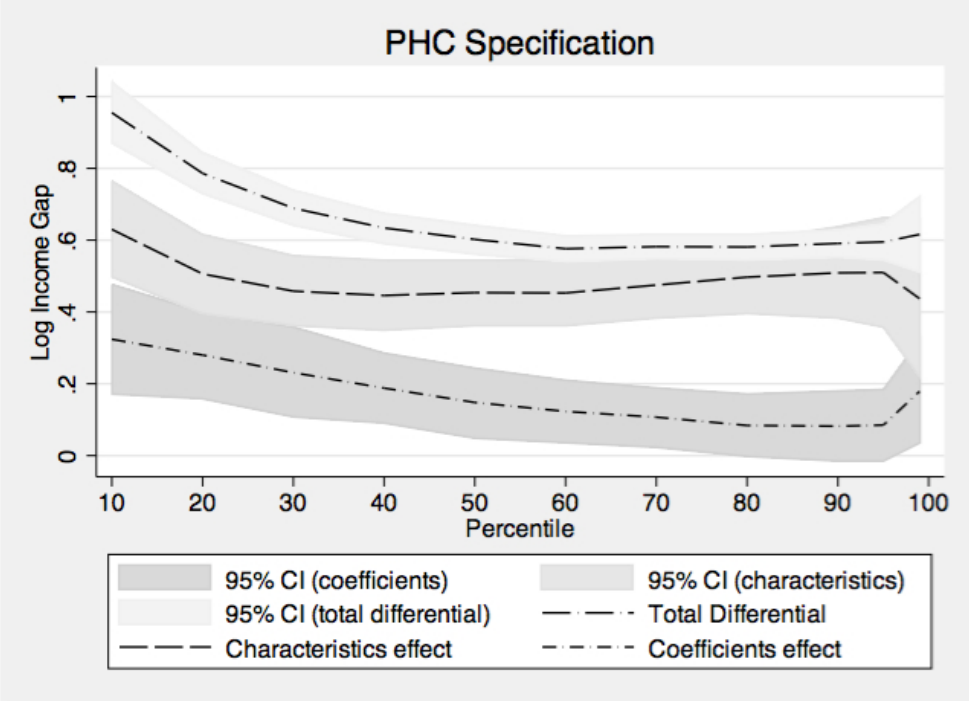

Figure A.3: Decomposition of Log Income Gap: PC specification (SCST coefficients)

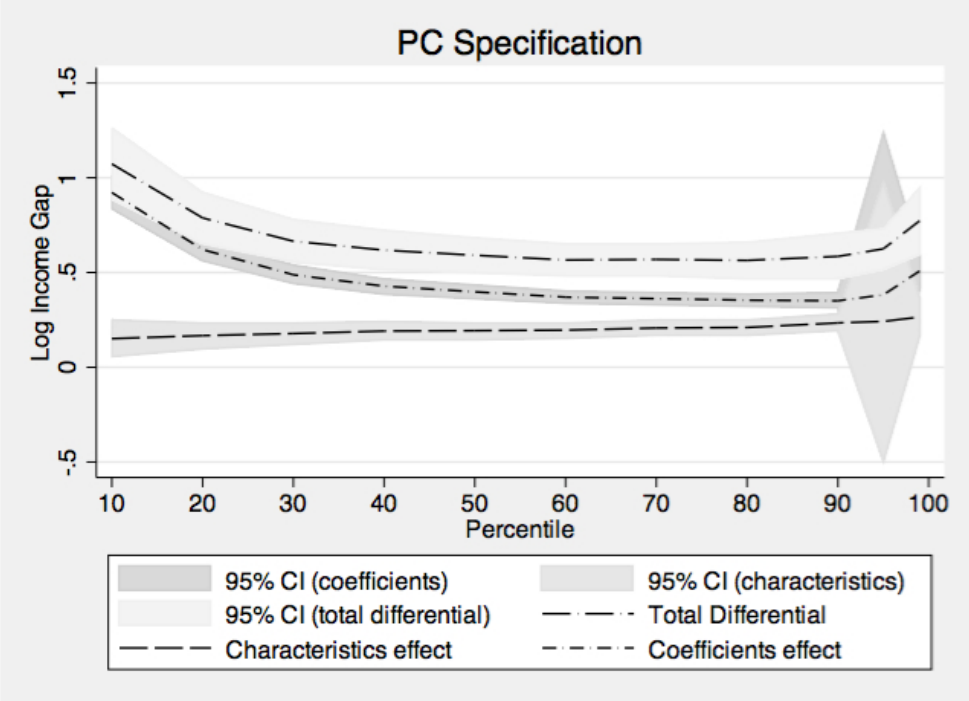

Figure A.4: Decomposition of Log Income Gap: PHC specification (SCST coefficients)

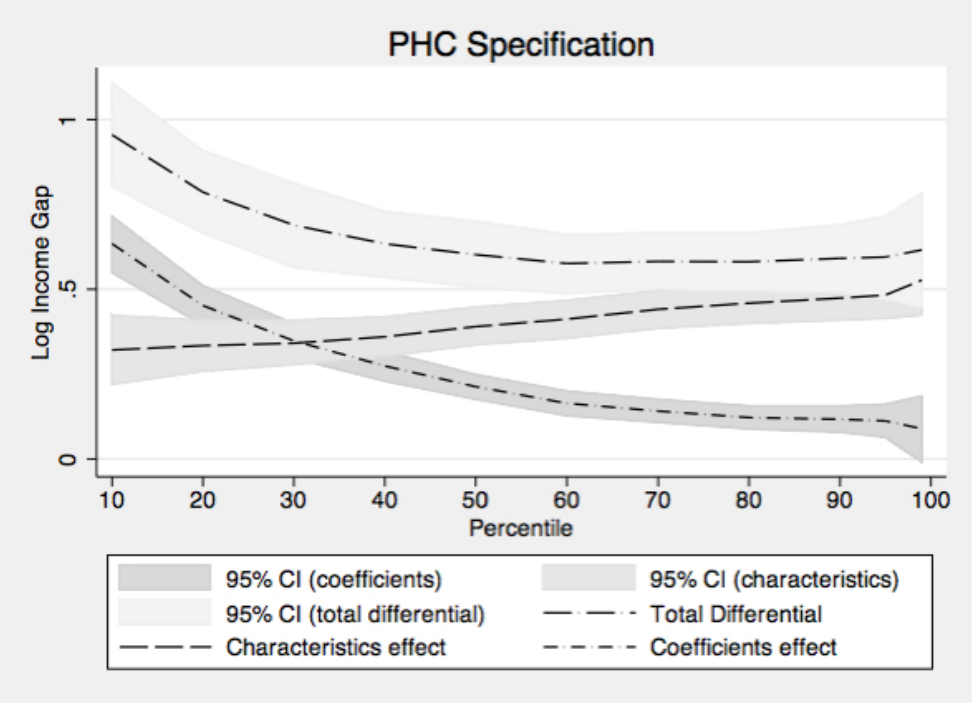

Figure A.5: Decomposition of Log Income Gap: PHBC specification (SCST coefficients)

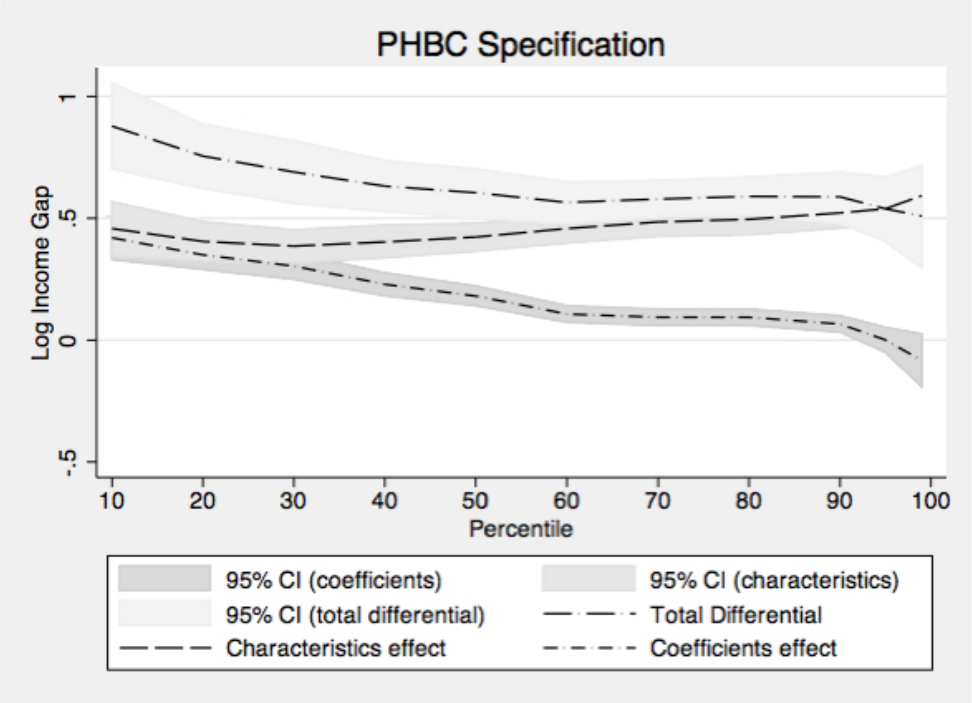

Figure A.6: Blinder-Oaxaca & Melly proportion of discrimination estimates: PC specification (non-SCST coefficients)

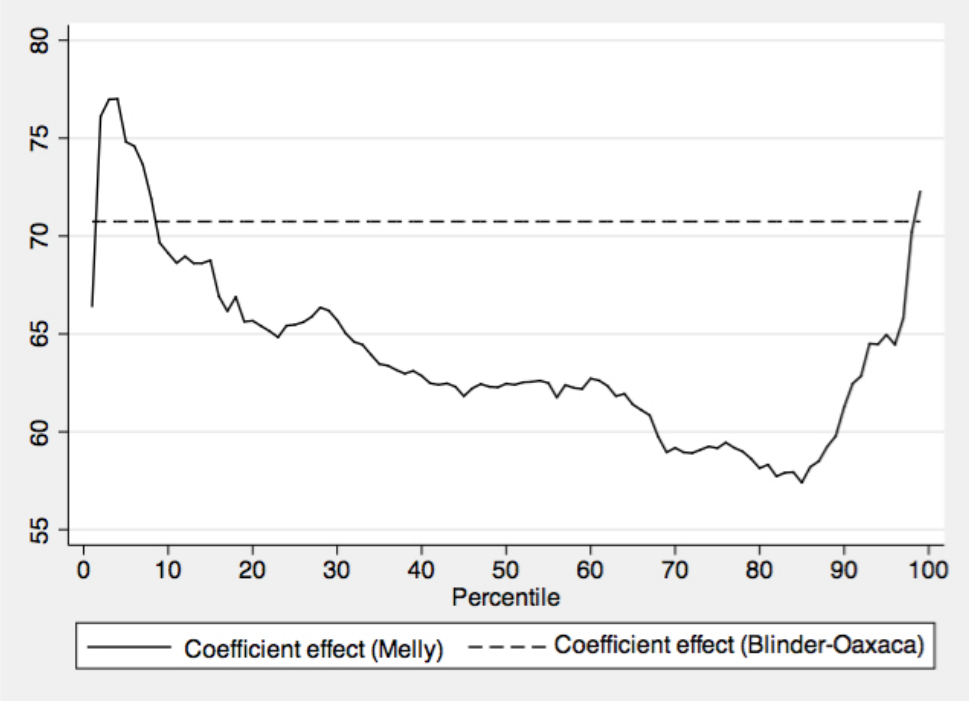

Figure A.7: Blinder-Oaxaca & Melly proportion of discrimination estimates: PHC specification (non-SCST coefficients)

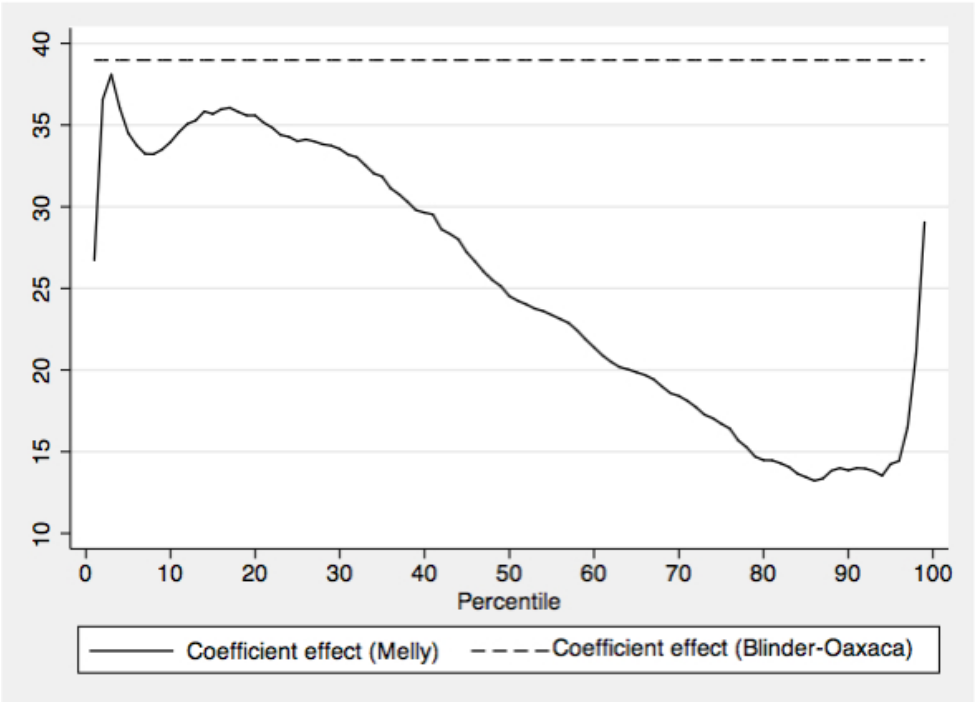

Figure A.8: Blinder-Oaxaca & Melly proportion of discrimination estimates: PC specification (SCST coefficients)

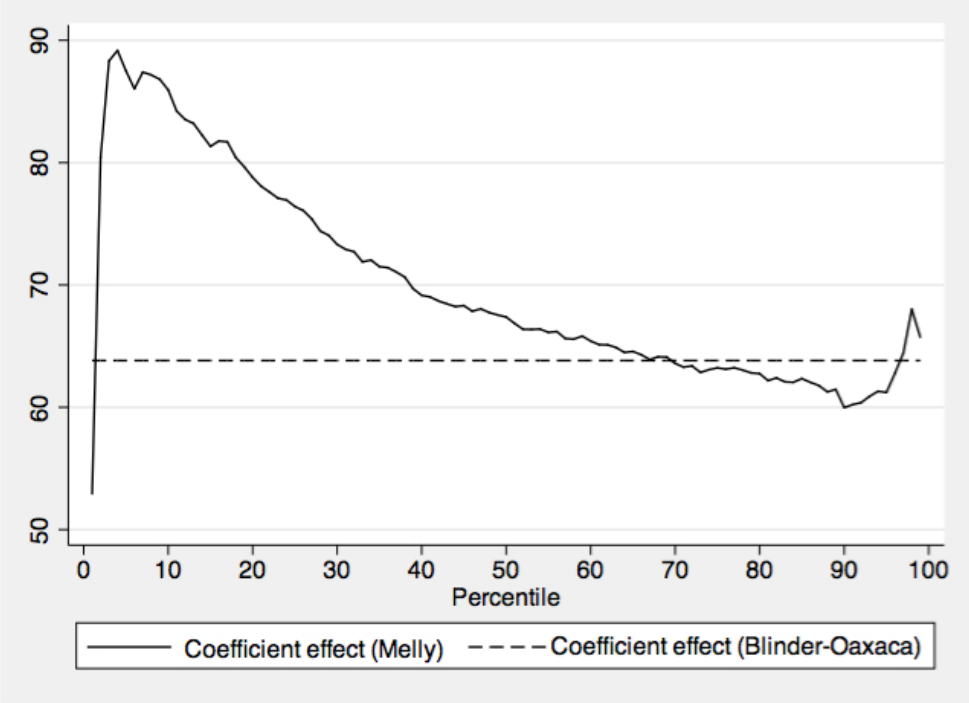

Figure A.9: Blinder-Oaxaca & Melly proportion of discrimination estimates: PHC specification (SCST coefficients)

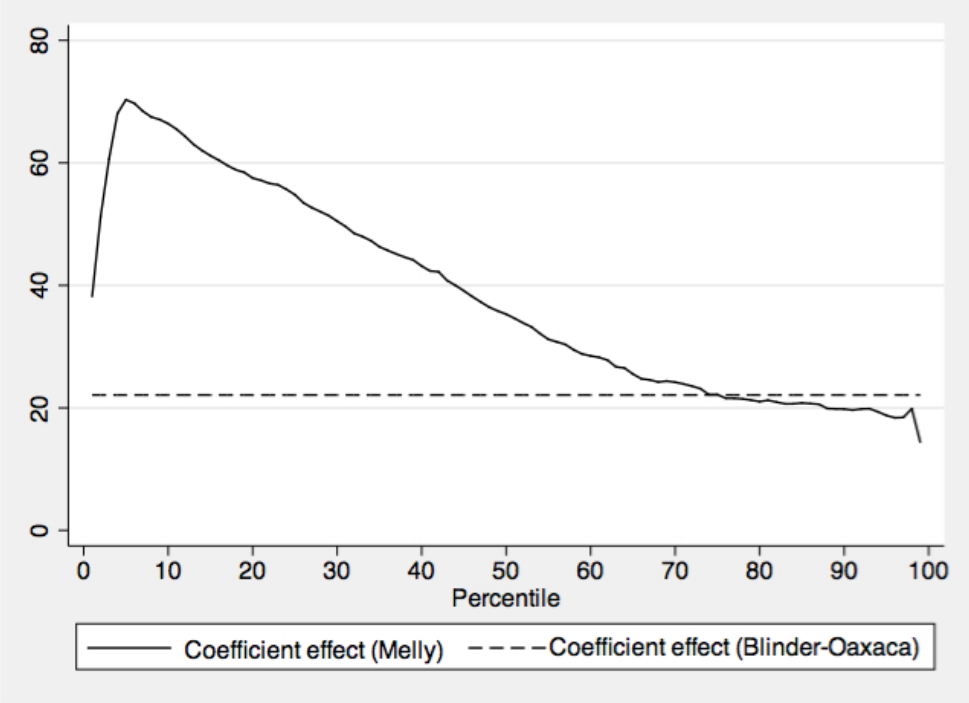

Figure A.10: Blinder-Oaxaca & Melly proportion of discrimination estimates: PHBC specification (SCST coefficients)

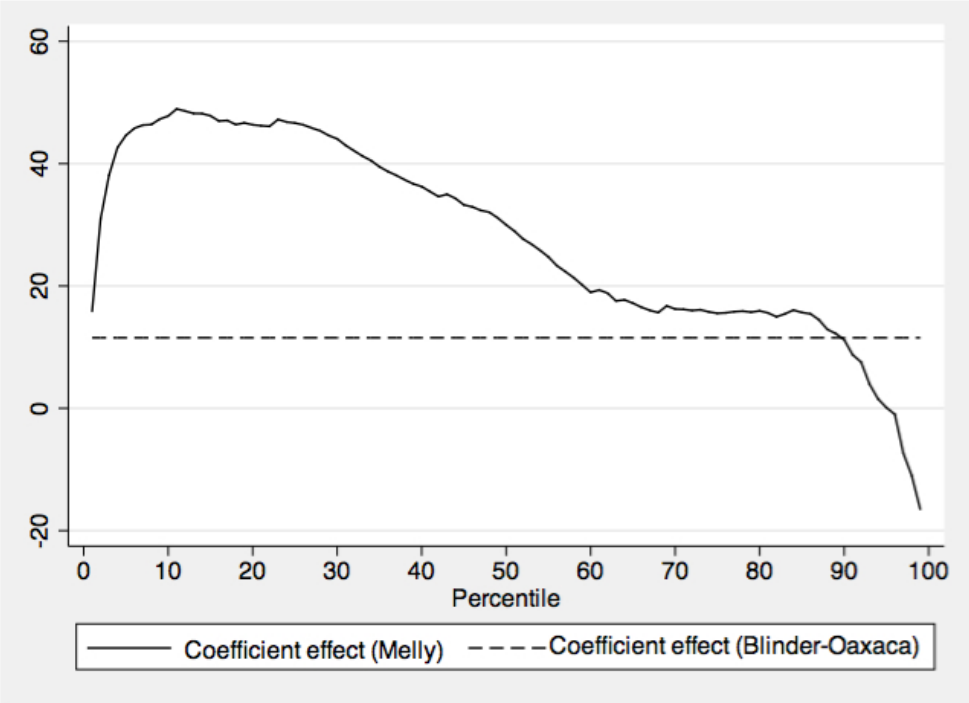

Table A.1: Caste-Wise Quantile Regressions: PC specification

Dep variable: Ln(income)	**SCST**						**Non-SCST**					
	Col.1 Q10	Col.2 Q25	Col.3 Q50	Col.4 Q75	Col.5 Q90	Col.6 Q95	Col.7 Q10	Col.8 Q25	Col.9 Q50	Col.10 Q75	Col.11 Q90	Col.12 Q95
Age	0.05	0.08**	0.03	0.05***	0.05**	0.10***	0.06***	0.06***	0.04***	0.04***	0.03***	0.04***
	(0.04)	(0.04)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Age squared/100	-0.05	-0.08*	-0.02	-0.05***	-0.04*	-0.10***	-0.07***	-0.06***	-0.04***	-0.03***	-0.03**	-0.03***
	(0.04)	(0.04)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Married	-0.30	-0.51***	-0.09	-0.03	-0.26	-0.40**	0.05	0.15**	0.15***	0.03	0.10	0.16*
	(0.36)	(0.18)	(0.08)	(0.12)	(0.17)	(0.16)	(0.08)	(0.07)	(0.06)	(0.05)	(0.06)	(0.09)
Years of Education	0.13***	0.11***	0.09***	0.08***	0.09***	0.08***	0.08***	0.07***	0.08***	0.08***	0.08***	0.09***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Constant	6.05***	6.95***	8.57***	8.83***	9.44***	8.86***	7.00***	7.79***	8.74***	9.36***	9.82***	10.00***
	(0.71)	(0.67)	(0.32)	(0.23)	(0.31)	(0.41)	(0.30)	(0.17)	(0.12)	(0.14)	(0.18)	(0.22)
Observations	1299	1299	1299	1299	1299	1299	6043	6043	6043	6043	6043	6043

Note: Standard errors in parentheses are bootstrapped using 100 replications. PC specification includes only personal characteristics. *** significant at 1%,** significant at 5%,* significant at 10%.

Table A.2: Caste-Wise Quantile Regressions: PHC specification

Dep variable: Ln(income)	**SCST**						**Non-SCST**					
	Col.1 Q10	Col.2 Q25	Col.3 Q50	Col.4 Q75	Col.5 Q90	Col.6 Q95	Col.7 Q10	Col.8 Q25	Col.9 Q50	Col.10 Q75	Col.11 Q90	Col.12 Q95
Age	0.08***	0.05**	0.02	0.03**	0.04**	0.04*	0.05***	0.05***	0.03***	0.02***	0.02***	0.02**
	(0.03)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Age squared/100	-0.09***	-0.06**	-0.03	-0.03**	-0.04*	-0.05*	-0.06***	-0.06***	-0.04***	-0.03***	-0.02***	-0.03**
	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Married	-0.18	-0.11	0.01	-0.10	-0.12	-0.22	0.12	0.15***	0.14***	0.05	0.08	0.07
	(0.18)	(0.16)	(0.13)	(0.10)	(0.16)	(0.16)	(0.08)	(0.06)	(0.05)	(0.05)	(0.05)	(0.07)
Years of Education	0.03**	0.02*	0.02***	0.03***	0.01	0.01	0.02***	0.01***	0.01***	0.02***	0.02***	0.02***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
Index of asset ownership	0.45***	0.47***	0.44***	0.45***	0.51***	0.53***	0.38***	0.40***	0.40***	0.44***	0.46***	0.51***
	(0.10)	(0.08)	(0.05)	(0.05)	(0.07)	(0.10)	(0.04)	(0.02)	(0.02)	(0.02)	(0.03)	(0.04)
Urban location	0.66***	0.52***	0.42***	0.24***	0.17**	0.16*	0.68***	0.46***	0.37***	0.29***	0.26***	0.21***
	(0.13)	(0.09)	(0.08)	(0.07)	(0.07)	(0.09)	(0.06)	(0.03)	(0.03)	(0.03)	(0.04)	(0.07)
Business or professional group membership	-0.05	0.08	0.13	0.10	0.10	0.14	0.23***	0.22***	0.17***	0.13**	0.15**	0.14
	(0.23)	(0.24)	(0.16)	(0.17)	(0.17)	(0.18)	(0.08)	(0.05)	(0.03)	(0.05)	(0.07)	(0.11)
Credit or savings group membership	-0.06	0.01	0.01	-0.19	-0.15	-0.25	-0.06	-0.14***	-0.17***	-0.10**	-0.14**	-0.10
	(0.20)	(0.17)	(0.12)	(0.12)	(0.15)	(0.24)	(0.10)	(0.05)	(0.05)	(0.05)	(0.07)	(0.09)
Caste association membership	-0.45**	-0.51***	-0.29**	0.04	0.08	0.09	0.04	0.06	0.10**	0.10**	0.09*	0.13*
	(0.18)	(0.14)	(0.14)	(0.14)	(0.11)	(0.12)	(0.09)	(0.05)	(0.05)	(0.04)	(0.05)	(0.08)
Development group/NGO membership	0.38	0.31	-0.09	-0.09	-0.38	0.08	0.21*	-0.07	0.02	-0.09	-0.20**	-0.18
	(0.46)	(0.37)	(0.43)	(0.42)	(0.51)	(0.47)	(0.12)	(0.10)	(0.09)	(0.07)	(0.10)	(0.20)
Co-operative membership	0.07	-0.26	0.04	0.01	0.24	0.14	-0.38***	-0.19	-0.08	-0.03	0.05	0.08
	(0.25)	(0.29)	(0.29)	(0.29)	(0.35)	(0.37)	(0.15)	(0.14)	(0.06)	(0.07)	(0.09)	(0.13)
Village panchayat or ward committee	0.05	-0.01	-0.04	-0.27***	-0.07	-0.13	-0.13	-0.07	0.02	0.08*	0.06	0.04
	(0.15)	(0.12)	(0.10)	(0.09)	(0.14)	(0.12)	(0.10)	(0.07)	(0.04)	(0.04)	(0.07)	(0.08)
Constant	7.57***	8.58***	9.82***	10.00***	9.78***	9.77***	7.49***	8.18***	9.11***	9.74***	10.31***	10.41***
	(0.67)	(0.59)	(0.42)	(0.31)	(0.39)	(0.55)	(0.29)	(0.26)	(0.17)	(0.21)	(0.20)	(0.26)
Observations	1286	1286	1286	1286	1286	1286	5937	5937	5937	5937	5937	5937

Note: Standard errors in parentheses are bootstrapped using 100 replications. PHC specification includes personal, household and location characteristics.*** significant at 1%,** significant at 5%,* significant at 10%.

Table A.3: Caste-Wise Quantile Regressions: PHBC specification

Dep variable: Ln(income)	**SCST**						**Non-SCST**					
	Col.1 Q10	Col.2 Q25	Col.3 Q50	Col.4 Q75	Col.5 Q90	Col.6 Q95	Col.7 Q10	Col.8 Q25	Col.9 Q50	Col.10 Q75	Col.11 Q90	Col.12 Q95
Age	0.07***	0.05***	0.03**	0.03**	0.03**	0.00	0.06***	0.04***	0.02***	0.02**	0.01	0.02
	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.03)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Age squared/100	-0.08***	-0.06***	-0.04**	-0.04**	-0.03*	0.00	-0.06***	-0.05***	-0.03***	-0.02**	-0.01	-0.02
	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Married	0.01	0.02	-0.01	-0.09	-0.31**	-0.17	0.12	0.11*	0.15***	0.13***	0.13**	0.10
	(0.15)	(0.14)	(0.09)	(0.12)	(0.14)	(0.23)	(0.07)	(0.06)	(0.04)	(0.04)	(0.05)	(0.07)
Years of Education	0.04***	0.01	0.01	0.02***	0.01	-0.00	0.02***	0.01***	0.01***	0.02***	0.02***	0.03***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
Index of asset ownership	0.28***	0.26***	0.29***	0.37***	0.47***	0.50***	0.28***	0.30***	0.34***	0.38***	0.38***	0.42***
	(0.08)	(0.06)	(0.05)	(0.04)	(0.06)	(0.08)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)
Urban location	0.33***	0.34***	0.23***	0.17***	0.08	0.01	0.46***	0.35***	0.28***	0.25***	0.21***	0.15**
	(0.12)	(0.08)	(0.07)	(0.06)	(0.06)	(0.08)	(0.06)	(0.03)	(0.03)	(0.03)	(0.05)	(0.06)
Business or professional group membership	-0.08	0.19	0.21	0.00	0.01	0.21	0.14**	0.07	0.15***	0.09**	0.20***	0.15
	(0.19)	(0.24)	(0.16)	(0.13)	(0.16)	(0.14)	(0.07)	(0.05)	(0.05)	(0.05)	(0.07)	(0.12)
Credit or savings group membership	-0.09	0.00	-0.16	-0.14	-0.22*	-0.42***	-0.08	-0.11*	-0.16***	-0.15***	-0.16**	-0.17
	(0.22)	(0.14)	(0.10)	(0.12)	(0.12)	(0.15)	(0.10)	(0.06)	(0.06)	(0.04)	(0.07)	(0.10)
Caste association membership	-0.34*	-0.27***	-0.19*	-0.10	0.05	0.06	0.03	0.04	0.06	0.10***	0.13**	0.10
	(0.18)	(0.10)	(0.11)	(0.12)	(0.13)	(0.13)	(0.08)	(0.06)	(0.04)	(0.04)	(0.06)	(0.07)
Development group/NGO membership	0.51	0.34	0.23	0.86	0.61	0.41	0.18	0.10	-0.01	-0.04	-0.05	-0.02
	(0.64)	(0.44)	(0.48)	(0.53)	(0.61)	(0.68)	(0.16)	(0.08)	(0.07)	(0.09)	(0.15)	(0.20)
Co-operative membership	0.15	-0.07	0.08	-0.05	-0.05	0.44	-0.27	-0.17	-0.01	0.06	0.07	0.19
	(0.48)	(0.23)	(0.24)	(0.19)	(0.48)	(0.48)	(0.18)	(0.14)	(0.07)	(0.07)	(0.09)	(0.14)
Village panchayat or ward committee	0.08	-0.05	-0.05	-0.13*	-0.15	-0.08	-0.10	-0.05	0.03	0.06	0.02	-0.05
	(0.15)	(0.12)	(0.08)	(0.08)	(0.11)	(0.12)	(0.08)	(0.05)	(0.04)	(0.04)	(0.06)	(0.07)
Total number of hours	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Number of workers	-0.25***	-0.14**	-0.14***	-0.07	-0.09**	-0.10*	-0.18***	-0.16***	-0.07***	-0.04*	-0.02	-0.03
	(0.09)	(0.07)	(0.05)	(0.04)	(0.04)	(0.05)	(0.05)	(0.04)	(0.02)	(0.02)	(0.03)	(0.03)
Workplace-other fixed	0.23*	0.33***	0.43***	0.27***	0.10	0.09	0.44***	0.36***	0.24***	0.14***	0.20***	0.20***
	(0.12)	(0.08)	(0.10)	(0.07)	(0.09)	(0.10)	(0.07)	(0.04)	(0.03)	(0.04)	(0.04)	(0.06)
Workplace-moving	0.33***	0.14	0.27***	0.16**	0.07	-0.01	0.20**	0.18***	0.13***	-0.00	0.00	0.01
	(0.12)	(0.12)	(0.10)	(0.08)	(0.11)	(0.11)	(0.08)	(0.06)	(0.04)	(0.04)	(0.06)	(0.07)
Constant	6.36***	6.92***	8.24***	8.76***	9.30***	9.72***	6.19***	7.48***	8.69***	9.47***	9.91***	10.02***
	(0.71)	(0.66)	(0.55)	(0.37)	(0.49)	(0.59)	(0.39)	(0.33)	(0.20)	(0.18)	(0.23)	(0.30)
Observations	1253	1253	1253	1253	1253	1253	5848	5848	5848	5848	5848	5848

Note: Standard errors in parentheses are bootstrapped using 100 replications. PHBC specification includes personal, household, location and business characteristics. *** significant at 1%,** significant at 5%,* significant at 10%.

Table A.4: Melly Decomposition of Log Income (Using the SCST Coefficients)

	Panel A PC			Panel B PHC			Panel C PHBC		
Decile	Col.1 Difference	Col.2 Characteristics	Col.3 Coefficients	Col.4 Difference	Col.5 Characteristics	Col.6 Coefficients	Col.7 Difference	Col.8 Characteristics	Col.9 Coefficients
10	1.07***	0.15***	0.92***	0.95***	0.32***	0.63***	0.88***	0.46***	0.42***
	(0.1)	(0.05)	(0.5)	(0.08)	(0.05)	(0.04)	(0.09)	(0.06)	(0.04)
20	0.79***	0.17***	0.62***	0.78***	0.33***	0.45***	0.76***	0.41***	0.35***
	(0.07)	(0.03)	(0.03)	(0.06)	(0.04)	(0.03)	(0.07)	(0.04)	(0.03)
30	0.67***	0.18***	0.49***	0.69***	0.34***	0.35***	0.69***	0.39***	0.30***
	(0.06)	(0.03)	(0.04)	(0.06)	(0.03)	(0.03)	(0.06)	(0.04)	(0.03)
40	0.62***	0.19***	0.43***	0.63***	0.36***	0.27***	0.63***	0.40***	0.23***
	(0.05)	(0.03)	(0.02)	(0.05)	(0.03)	(0.02)	(0.05)	(0.03)	(0.02)
50	0.59***	0.19***	0.40***	0.60***	0.39***	0.21***	0.60***	0.42***	0.18***
	(0.05)	(0.02)	(0.02)	(0.05)	(0.03)	(0.02)	(0.05)	(0.03)	(0.02)
60	0.57***	0.20***	0.37***	0.57***	0.41***	0.16***	0.57***	0.46***	0.11***
	(0.04)	(0.02)	(0.02)	(0.04)	(0.03)	(0.02)	(0.04)	(0.03)	(0.02)
70	0.57***	0.21***	0.36***	0.58***	0.44***	0.14***	0.58***	0.49***	0.09***
	(0.04)	(0.02)	(0.02)	(0.04)	(0.03)	(0.02)	(0.04)	(0.03)	(0.02)
80	0.56***	0.21***	0.35***	0.58***	0.46***	0.12***	0.59***	0.5***	0.09***
	(0.05)	(0.02)	(0.02)	(0.04)	(0.03)	(0.02)	(0.04)	(0.03)	(0.02)
90	0.58***	0.23***	0.35***	0.59***	0.47***	0.12***	0.59***	0.52***	0.07***
	(0.06)	(0.02)	(0.02)	(0.05)	(0.03)	(0.02)	(0.05)	(0.03)	(0.02)

Note: Standard errors in parentheses are based on bootstrapping with 100 replications. PC specification includes only personal characteristics; PHC specification includes personal, household and location characteristics and PHBC specification includes personal, household, location and business characteristics. *** significant at 1%,** significant at 5%,* significant at 10%.

B Definitions of Variables

In order to canvass data on non-farm businesses, the survey explicitly asks "Does anybody in this household run their own business, however big or small? Does anybody make something for sale, such as cloth or some food like pickles? Or does anybody sell something in a market or to customers of any sort? Or does anybody provide a service to others for a price, either a skilled service like a doctor or an unskilled service like a barber?"

1. **Net income**: Gross receipts less hired workers' wages less all other expenses such as costs of materials, rent, interest on loans etc.
2. **Age**: in years (of the de-facto decision-maker)
3. **Marital status**: equals 1 if married, 0 otherwise (of the de-facto decision-maker)
4. **Years of education**: standard number of years of education completed (of the de-facto decision-maker)
5. **Urban**: equals 1 if household is in an urban area, 0 otherwise
6. **Business or professional group membership**: equals 1 if the household is a member, 0 otherwise
7. **Credit or savings group membership**: equals 1 if the household is a member, 0 otherwise
8. **Caste association membership**: equals 1 if the household is a member, 0 otherwise
9. **Development group/NGO membership**: equals 1 if the household is a member, 0 otherwise
10. **Co-operative membership**: equals 1 if the household is a member, 0 otherwise
11. **Village panchayat or ward committee**: equals 1 if someone in, or close to the household is a member, 0 otherwise
12. **Total number of hours**: total number of hours spent in the business by all workers involved in the business

13. **Number of workers**: number of household members that worked in the business

14. **Workplace type**: dummy variables for each of the 3 categories - home; other fixed place; other moving place

15. **Industry type (NIC-1987)**: dummy variables for each of the 9 categories - agriculture, hunting, forestry, fishing; mining and quarrying; manufacturing; electricity, gas and water; construction; wholesale trade, retail trade, restaurants and hotels; transport, storage and communication; finance, insurance, real estate and business services; community, social and personal services.

16. **State**: dummy variables for each of the 22 states - Jammu and Kashmir, Himachal Pradesh, Punjab, Uttaranchal, Haryana, Delhi, Rajasthan, Uttar Pradesh, Bihar, Tripura, Assam, West Bengal, Jharkhand, Orissa, Chhattisgarh, Madhya Pradesh, Gujarat, Maharashtra, Andhra Pradesh, Karnataka, Kerala and Tamil Nadu.

C Constructing the Asset Index using Principal Components Analysis

Table C.1: Asset ownership, by caste

Variable	All	SCST	Non-SCST
Cycle/bicycle	64.6	66.1	64.3
Sewing machine	38.4	25.4	41.1
Generator set	3	1.4	3.4
Mixer/grinder	36.7	16.2	41.1
Motorcycle	32	16.7	35.3
Black & white TV	31.6	29	32.2
Colour TV	40.7	23.7	44.4
Air cooler	20.7	11	22.8
Clock/watch	92.3	82	94.5
Electric fan	76.8	53.8	81.7
Chair/table	80.2	63.8	83.7
Cot	90.9	84.3	92.4
Telephone	28.6	12.7	32
Cell Phone	16.1	6.2	18.3
Refrigerator	28.7	13	32
Pressure cooker	60.1	38.2	64.8

Note: Values represent proportions of households owning each of the listed items.

Using all of the assets listed in the table above, we create an asset index using Principal Components Analysis (PCA) as follows. Suppose we have a set of n correlated variables, X_1^* to X_n^*, representing the ownership of n assets by each household. PCA creates uncorrelated indices or components, where each component is a linear weighted combination of the initial variables. PCA starts by specifying each variable normalized by its mean and standard deviation: for example, $X_1=(X_1^* - \bar{X_1^*})/s_1^*$ where $\bar{X_1^*}$ is the mean of X_1^* across households and s_1^* is its standard deviation. Therefore, for a set of variables X_1 to X_n:

$$PC_1 = a_{11}X_1 + a_{12}X_2 + ... + a_{1n}X_n$$

$$PC_m = a_{m1}X_1 + a_{m2}X_2 + ... + a_{mn}X_n$$

where a_{mn} represents the weight for the m^{th} principal component and the n^{th} variable.

The weights for each principal component are given by the eigenvectors

of the correlation matrix or the covariance matrix, as the case may be. Eigenvalues of the corresponding eigenvectors measure the amount of the variation explained by each PC and will be largest for the first PC and smaller for the subsequent PCs. The components are ordered so that the first component (PC_1) explains the largest possible amount of variation in the original data. The second component (PC_2) is completely uncorrelated with the first component, and explains additional but less variation than the first component. Each component captures an additional dimension in the data, while explaining smaller and smaller proportions of the variation of the original variables. A common method used to select PCs is where the associated eigenvalue exceeds one.

We use the first principal component as a measure of wealth. Our calculations reveal that the first component accounts for 35.7 percent of the variance and has an eigenvalue of 1.01. We retain only the first component since the second component has an eigenvalue of 0.33 (explaining 11.8 percent of the variance), well below the rule of thumb eigenvalue of 1.

Table C.2: Scoring coefficients & summary statistics for variables to compute PC_1

Variable	Scoring coefficients	Mean	SD	Scoring coefficient/SD
Cycle/Bicycle	-0.0184	0.646	0.478	-0.038
Sewing machine	0.2447	0.384	0.486	0.503
Generator set	0.0421	0.030	0.172	0.245
Mixer / Grinder	0.3536	0.367	0.482	0.733
Motor Cycle/Scooter	0.3187	0.320	0.466	0.683
Black & white TV	-0.0735	0.316	0.465	-0.158
Colour TV	0.3829	0.407	0.491	0.779
Air Cooler	0.2338	0.207	0.405	0.577
Clock or watch	0.0908	0.923	0.267	0.340
Electric Fan	0.2457	0.768	0.422	0.582
Chair or table	0.2127	0.802	0.398	0.534
Cot	0.0762	0.909	0.287	0.265
Telephone	0.3173	0.286	0.452	0.702
Cell Phone	0.2034	0.161	0.368	0.553
Fridge/Refrigerator	0.3504	0.287	0.452	0.775
Pressure Cooker	0.3467	0.601	0.490	0.708

D Supplementary Descriptive Statistics

Table D.1: Distribution of businesses across industry types, by caste

Variable	All	SCST	Non-SCST
Agriculture, hunting, forestry & fishing	4.93	15.3	2.7
Mining & quarrying	0.2	0.07	0.2
Manufacturing	13	13.2	12.9
Electricity, gas & water	0.3	0.2	0.3
Construction	2.2	2.2	2.2
Wholesale, retail trade & restaurants & hotels	53.7	44.5	55.7
Transport, storage & communication	6.5	6.4	6.5
Financing, insurance, real estate & business services	2.7	1.4	3
Community, social & personal services	16.4	16.6	16.3

Note: Values represent percentage of businesses engaged in each of the industries.

Table D.2: State-wise distribution, by caste

Variable	All	SCST	Non-SCST
Jammu & Kashmir	1.41	1	1.5
Himachal Pradesh	3.35	4.07	3.2
Punjab	3.69	4.15	3.6
Uttaranchal	1.09	1.15	1.07
Haryana	2.80	1.77	3.02
Delhi	2.13	1.61	2.24
Rajasthan	7.04	6.3	7.19
Uttar Pradesh	12.40	9.22	13.08
Bihar	6.03	4.15	6.43
Tripura	1.14	2.15	0.92
Assam	1.91	1.92	1.91
West Bengal	7.55	11.83	6.63
Jharkhand	2.85	3.99	2.61
Orissa	5.51	7.76	5.03
Chhattisgarh	3.22	7.53	2.29
Madhya Pradesh	6.65	11.37	5.64
Gujarat	5.20	3.3	5.61
Maharashtra	8.28	7.14	8.53
Andhra Pradesh	3.54	1.54	3.98
Karnataka	8.84	5.68	9.52
Kerala	2.74	0.38	3.25
Tamil Nadu	2.59	2	2.72

Note: Values represent percentage of businesses in each of the listed states.

Centre for Development Economics
Department of Economics, Delhi School of Economics
Delhi 110 007 INDIA
Tel.: 27667005/146, 27666703-705, 27666533-535
Fax: 91-11-27667159
Email: office@econdse.org