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

Ashwini Deshpande[*]

Smriti Sharma[†]

Department of Economics, Delhi School of Economics

August, 2014

Abstract

Using the India Human Development Survey data for 2004-05, we employ two methodologies to estimate the earnings structure of household 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-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 55 percent of the earnings gap could be attributed to the unexplained or the discriminatory component. Quantile regressions suggest that gaps are higher at lower deciles than the higher ones, and the quantile decompositions show that the unexplained component is greater at the lower and middle deciles than higher, suggesting that SCST-owned businesses at the lower and middle parts of the conditional earnings distribution face greater discrimination, as compared to those at the higher end.

1 Introduction

Given the compelling evidence of labor 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. In fact, there are several examples that show how ownership of small business has been an important factor in the economic success and mobility of immigrant groups such as the Chinese, Koreans, Jews and Italians in the United States (see Fairlie, 2006 and the references therein) and ethnic minorities in England and Wales (Clark and Drinkwater, 2000). 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.[1] 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, members of marginalized groups become job-givers, especially for their own community, and not job-seekers, and thus reduce their vulnerability to labor 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).[2] Indeed, the concept of "Black Capitalism" in the USA encapsulated several of these views (e.g., Bates 1973; 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 industry association Federation of Indian Chambers of Commerce and Industry (FICCI).[3]

---

[1]Even when members of marginalized groups get top-end jobs, they may face serious discrimination at the workplace (Daniels, 2004).

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

[3]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 www.dicci.org

The implicit assumptions underlying these claims have been that one, a sufficiently large part of the self-employment activity would be entrepreneurial, rather than survivalist; and two, discriminatory tendencies that characterize the labor market would somehow be missing from other markets, such as land, credit or consumer markets, critical to the success of entrepreneurial activities.

Our objective in this paper is to assess the presence of caste discrimination in self-employment. 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 businesses owned by Scheduled Castes and Tribes (SCs and STs[4]) 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. Correspondingly, we use two decomposition strategies to estimate discrimination. We decompose the mean earnings gap through the conventional Blinder-Oaxaca method. We also decompose the gaps estimated from quantile regressions at each percentile of the earnings distribution 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 and non-SCST businesses. The latter are more urban, record larger number of total man-hours, have better educated and less poor owners who possess a greater number of assets, 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 evidence of systematic differences, however, does not prove discrimination along caste lines; all the gaps could, in principle, be accounted for by these differences in characteristics of SCST and non-SCST businesses. Of course, the fact that SCSTs enter the field of self-employment with inferior characteristics indicates the presence of some "pre-market" discrimination, of which there is plenty of independent evidence (e.g., Deshpande, 2011 and

---

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

Thorat and Newman, 2010).

The Blinder-Oaxaca decomposition reveals that depending on the specification of explanatory variables, as much as 55 percent of the net income gap could be attributed to the unexplained or the discriminatory component. Raw gaps in earnings are higher at lower deciles than at the higher deciles, underscoring the importance of examining earnings gaps at different points of the distribution. Quantile regressions with one specification reveal that gaps are higher at lower deciles than the higher ones, after controlling for characteristics, and the quantile decompositions reveal that the unexplained component is greater at the lower and middle deciles than higher, suggesting that SCST-owned businesses at the lower and middle end of the conditional earnings distribution face greater discrimination, as compared to those at the higher end. Thus, we find some evidence similar to 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). Using the non-SCST earnings structure as the counterfactual, we find the unexplained component to be the highest for businesses in the middle deciles of the earnings distribution.[5]

This paper is a part of a larger enquiry. In a companion 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 nature of participation of marginalized groups in self-employment in order to evaluate the nature of Dalit Capitalism. While there is a small and emerging section of successful Dalit industrialists at the top end, ably assisted by the DICCI, our study reveals that the majority of Dalit (and Adivasi) businesses are engaged in small-scale, low productivity, survivalist activities, and the MSME sector exhibits very clear differences along business owners' caste (and gender), in virtually all business characteristics.

To the best of our knowledge, ours is the first paper to estimate discrimination through a decomposition of the caste gaps in earnings from self-employment using Indian data. Also, we go beyond the mean earnings gap to explore differences in earnings at various points of

---

[5]As is standard in decomposition exercises, the actual division of the earnings gap into explained and unexplained components is sensitive to the choice of the non-discriminatory earnings structure used to construct a counterfactual earnings distribution, as well as to the set of explanatory variables used in the earnings equation.

the earnings distribution. This has been enabled by the availability of the India Human Development Survey - a large-scale nationally representative survey - that among other things, also canvasses data on the performance of household non-farm businesses, as well as on owner and business characteristics, which allows us to neatly decompose the income gap into two components - one explained by differences in characteristics, and the residual unexplained gap, conventionally attributed to discrimination.

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   Review of 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.[6] 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 mainly on account of differences in characteristics and not so much 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 and also find that caste differences in rural casual labor 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

---

[6]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.

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 discriminatory component is 59 percent.[7] On the other hand, differences in earnings from casual labor are small and can be largely attributed to differences in endowments.[8]

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 labor, 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.

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

---

[7]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.

[8]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.

and Bronars, 1989) or some unobserved differences in behavior.

3 Methodology

3.1 Blinder-Oaxaca Decomposition Framework

We first use the Blinder-Oaxaca method to decompose the mean earnings gap from self-employment between SCSTs and non-SCSTs into portions attributable to differences in characteristics (also known as the explained component or composition effect) and differences in returns to these endowments (also known as the unexplained component or coefficients effect) (Blinder, 1973 and Oaxaca, 1973). It is conventional in decomposition exercises to attribute the unexplained component or the coefficients effect to discrimination. However, it is highly plausible that this residual includes the effects of either unmeasurable or unobservable characteristics. All decomposition exercises are subject to this caveat. However, it is equally true that pre-market discrimination affects the formation of characteristics, and thus, the explained component also embodies the effects of past discrimination. Therefore, estimates of the unexplained component from decomposition exercises should not be taken as precise measurements of "true" discrimination, but as rough estimates, providing orders of magnitude.

This method involves estimating 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. 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:

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

In order to decompose this gap, some assumptions have to be made about the earnings structure that would prevail 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 one applicable to non-SCSTs. 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 and the second term represents differences due to varying returns to the same characteristics. The second term is the unexplained component and is considered to be a reflection of discrimination.[9]

The decomposition is sensitive to the choice of the non-discriminatory earnings structure, as the two counterfactuals yield different estimates. To get around this 'index number problem', one solution is to use the pooled estimates (estimating the earnings function for the whole population) as the single counterfactual (Oaxaca and Ransom, 1994). 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 earnings function focuses on average wage or earnings. However, moving away from the least squares method using quantile regressions introduced by Koenker and Bassett (1978), we can analyze the effect of a given attribute at different parts of the conditional earnings function, which need not be uniform.

---

[9]One could also construct an alternative counterfactual by assuming that the non-discriminatory earnings structure is the one applicable to the SCSTs.

Generalizing the traditional Blinder-Oaxaca decomposition, 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 regressions. The conditional quantile function $Q_\theta(w|X)$ can be expressed using a linear specification for each group as follows:

$$Q_\theta(w_g|X_g) = X_{i,g}^T \beta_{g,\theta} \; for \; each \; \theta \; \in (0,1) \tag{5}$$

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. The quantile regression coefficients can be interpreted as the returns to various characteristics at different quantiles of the conditional earnings distribution.

Next, Machado and Mata (2005) construct the counterfactual unconditional earnings distribution using estimates for the conditional quantile regressions 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 are computed using a bootstrapping technique.

This simulation-based estimator relies on the generation of a random sample to construct the counterfactual unconditional earnings distribution, and comes at the cost of increased com-

putational 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. 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). Since the estimator is not simulation-based, it is computationally faster. The Melly (2006) and Machado and Mata (2005) decompositions are numerically identical when the number of simulations in the latter goes to infinity.

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

$$CF_\theta^s = X_{s,i}^T \beta_{n,\theta} \tag{6}$$

This yields the following decomposition:

$$\Delta_\theta = (Q_{n,\theta} - CF_\theta^s) + (CF_\theta^s - Q_{s,\theta}) \tag{7}$$

The first term on the right hand side represents the effect of characteristics (explained component) and the second the effect of returns to characteristics (coefficients effect or unexplained component).

4 Data and Descriptive Statistics

4.1 Data

The data come 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. The nationally representative data covers 1504 villages and 971 urban areas across 33 states and union territo-

ries of India.[10] The survey covering 41,554 households was carried out through face-to-face interviews. The respondents usually included the male head of the household and an ever-married woman aged 15-49. The modules of the survey collect 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.

The survey module on the household non-farm business does not identify the primary decision-maker in the 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 information we assume that the person who has spent maximum number of hours in the business is the de-facto decision maker.[11]

We restrict the sample only to those states where there are at least 50 household businesses. With this restriction, we 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. We consider only male businesses (i.e. where men are the primary decision makers according to our assumption above) in this analysis because factors affecting selection into self-employment vary along lines of gender; additionally, in order to delineate the effect of caste we need to hold gender constant, so as not to confound the effect of overlapping identities.

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 of net income from the business over the last 12 months. Net income is computed as gross receipts less hired workers' wages less cost of materials, rent, interest on loans etc. One issue on which the data are patchy is the use of unpaid family labor in these businesses, which would affect the calculation of net income. It is highly likely that the opportunity cost of SCST labor would be different from that of non-SCST labor, and if that could be accounted for through imputing shadow prices, the

---

[10]Andaman and Nicobar and Lakshadweep were not included in the sample. These Union Territories account for less than 0.05 percent of India's population.

[11]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.

"true" net income might differ from the reported net income. While some businesses in the data report the individual components as well as a net income, others report only the net income. However, our queries with the IHDS team revealed that when hired labor is not reported, it can not be assumed that no labor was actually hired. Thus, data do not allow us to clearly distinguish between hired and unpaid family labor, resulting in the inability to estimate "true" net income. We, thus, use the net income figures in the data as reported.[12] 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 enterprise performance, income is the most appropriate outcome to consider.

As explanatory variables, firstly, we have the individual specific variables such as age, marital status and standard years of education completed of the decision-maker. Secondly, we have the household specific variables such as wealth (proxied by asset ownership), 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 type.[13]

We are unable to account for selection bias in our estimations due to the lack of an appropriate instrument for selection into self-employment. Borjars and Bronars (1989) found positive self-selection into self-employment among whites and neutral self-selection among blacks, using Heckman's sample correction method, and suggest that consumer discrimination might be the underlying mechanism at work. Kawaguchi (2005) using panel data found selection by both low-ability and high-ability whites into self-employment (with middle-ability whites being in salaried jobs), but negative selection among blacks, suggesting credit market and consumer discrimination as possible channels. If indeed there is differential selection by SCSTs and non-

---

[12]We are grateful to Rinku Murgai for raising this issue.

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

SCSTs into self-employment, then our estimates of discrimination would need to be adjusted, especially if part of the difference in selection is accounted for by discrimination.

4.2 Descriptive Statistics

Table 1 lists the summary statistics for the whole sample and for the sample of SCST and non-SCST businesses separately. 7 percent of business owners are Brahmin, 42 percent are Other Backward Classes (OBCs), 12.7 percent are SCs, 5.1 percent are STs and 33 percent are 'Others'.[14] We club OBCs, Brahmins and 'Others' together and refer to them as non-SCSTs in the analysis. Thus, of the total 7288 businesses, 1300 are owned by SCSTs (17.8 percent) and the remaining 5988 by non-SCSTs (82.2 percent).

In terms of performance, the average gross receipts for all businesses is Rs. 108,015, but the receipts for non-SCST businesses (Rs. 118,708) are more than double those for SCST businesses (Rs. 58,804). A similar pattern can be seen in the average net incomes with income for non-SCST businesses (Rs. 45,218) being 1.76 times that for SCST businesses (Rs. 25,640). Figure 1 plots the kernel density distribution of log income for SCST and non-SCST businesses. The distribution of incomes of non-SCST businesses lies distinctly to the right of the SCST businesses.

This large difference in business performance can be on account of a variety of individual, household and business characteristics, in most of which there are clear differences between SCSTs and non-SCSTs. The summary of individual characteristics of the primary decision maker reveals that he is on average 39 years old. 86 percent of them are married. These numbers are similar across SCST and non-SCST decision makers. However, average years of education, which for all is 7.8, differ significantly by caste, with 8.3 years for non-SCSTs and 5.7 years for SCSTs. There is a distinctly different pattern in the rural-urban distribution across castes with 33 percent of SCST households and 53 percent of non-SCST households being located in urban areas.

There is also disparity in material standard of living, reflected in the average monthly per

---

[14]The 'Others' category in the IHDS data 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.

capita expenditure, which for SCSTs is Rs. 755 and for non-SCST is Rs. 1119. This is further reflected in asset ownership, in that out of the 16 assets listed in the questionnaire, non-SCSTs own approximately 8 assets while SCSTs own around 5.[15] We create a wealth index using Principal Components Analysis, and divide the sample based on this into three groups following Filmer and Pritchett (2001): 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.

We examine the importance of business, professional and political networks since these can affect the decision to become self-employed, as well as the prospective success of the business (Allen, 2000). In general, participation in such networks is low. 8 percent of all businesses are members of business or professional groups with membership of SCST businesses being below average (5 percent). Participation in credit or savings groups does not differ by caste, covering roughly 7 percent of owners. Membership in caste associations is 14 percent and 12 percent for non-SCST and SCST businesses respectively. Membership in development groups or NGOs is miniscule across the board, while that in co-operatives is slightly higher at 3.5 percent overall, and 2 percent for SCSTs. In terms of political networks, 12.5 percent of SCSTs have someone in, or close to, their households who has been an official of rural or urban local bodies while for non-SCSTs, the corresponding figure is 10.6 percent. This could possibly reflect the operation of the mandatory 22.5 percent caste quotas in local bodies for SCSTs. Overall, there is no discernible pattern in network participation of the two groups in our data.

These gaps in performance could also be related to other characteristics, such as a) the number of family members who worked in the business: SCST businesses have greater than

---

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

average number of family members working in the business (1.47), as compared to non-SCST businesses (1.37); and b) the total number of 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 and less makeshift. This dissimilarity is mirrored in the proportions in moving or mobile workplaces, where the proportion for SCST businesses is 34, and the corresponding proportion for non-SCST businesses is 20.

Finally, there are some differences by caste in the industries that these businesses operate in. The most important industrial sector for 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 businesses is close to 56 and that for SCST 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), tailoring related (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 SCST. Businesses in "mining and

quarrying" as well as "electricity, gas and water" sectors is practically non-existent, which is only to be expected, given that these highly capital intensive activities are not conducive to self-employment.

5   Results

5.1   Earnings Function Estimates

Table 2 reports the OLS estimates with log income as the dependent variable, for the pooled sample, and separately by caste. We present estimates using two specifications. The first specification uses only exogenous explanatory variables. This includes age, age squared, whether married or not, years of education, whether urban or not, and state of residence. The second specification is more exhaustive and also includes potentially endogenous variables. In addition to variables in the first specification, we include the asset ownership/wealth index, memberships in: business or professional groups, credit or savings groups, caste associations, development groups or NGO, co-operatives, political networks, number of hours spent by everyone working in the business, number of family members working in the businesses, whether workplace is fixed or moving (reference category is home-based) and industry type.[16]

The SCST dummy is negative and significant in both specifications, indicating that ceteris paribus, belonging to these marginalized groups is negatively correlated with income. As expected, earnings have a quadratic relationship with age such that earnings initially increase with age and start to decline thereafter. Urban location, asset ownership and years of education are positively correlated with earnings. The number of hours spent working is positively correlated with income, as expected. Businesses based in other fixed locations (outside of the home) and that are moving/mobile are correlated with higher incomes than home-based businesses.

Caste-specific OLS estimates indicate that some variables correlate in different ways with performance of SCST and non-SCST businesses. For instance, business or professional group

---

[16]As a robustness check, we also estimated three specifications: one with purely personal characteristics; second with personal and household characteristics and third one being the same as the full specification with all variables. The results, robust to alternative specifications, are available from the authors upon request.

membership is positively associated with income for non-SCST businesses, but is insignificant for SCST businesses, suggesting that the kinds of business or professional groups that SCST businesses are members of might not contribute substantially to increasing incomes, either due to their inexperience or lack of expert or specialized business knowledge. Development group/NGO membership is positively correlated with earnings for SCST businesses but not for non-SCST businesses. Somewhat perplexing is the fact that membership of credit or savings group is negatively associated with earnings of non-SCST businesses, but is insignificant for SCST businesses. One possible explanation for this might be that businesses that are selecting into such groups are the ones who are lacking in some unobservable social capital.

5.2   Decomposition of the Mean Earnings Gap

The results of the Blinder-Oaxaca decomposition with log income as the dependent variable are presented in Table 3.[17] Panel A of Table 3 displays the decomposition results using 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.

In the presence of non-SCST coefficients, with the first specification, more than half of the mean log income gap remains unexplained (55.4 percent), while with the second specification, the unexplained component reduces to 19 percent. This is expected since in the latter specification, with more explanatory variables, a greater proportion of the average income gap is being accounted for. Using SCST coefficients, we see that the unexplained proportions for the two specifications are 41.7 percent and 10.5 percent (the latter not significant), and for the pooled model, the corresponding values are 52.3 percent and 16 percent. Thus, depending on the specification and the counterfactual earning structure, the unexplained component varies between 55.4 and 10.5 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 component for each specification. These are 0.32 and 0.08 respectively, which correspond to unexplained

---

[17]This is done using the STATA program "oaxaca" (Jann, 2008).

estimates of 48.1 percent and 13.8 percent for the first and second specification respectively.

Which of the variables contributes the most to the explained component? The lower panel of Table 3 shows the contribution of selected significant characteristics to the overall explained part of the income gap. Using the first specification, years of education contributes 39-42 percent of the explained component, depending on the counterfactual earnings structure. Urban location also accounts for 37-44 percent. However, in the second specification, the importance of years of education and urban location declines significantly to around 5-8 percent and number of hours and asset index are the dominant variables, each accounting for approximately 40 percent of the explained component.

5.3   Quantile Regressions

For quantile regressions, we use the same two specifications of the earnings function with log income as the dependent variable that we used for the OLS regressions. The average gap in log incomes of non-SCST-owned and SCST-owned businesses is 0.75, which corresponds to a gap of 112 percent in raw net incomes of the two types of businesses. This is instructive, but when we juxtapose this against the log income gap for the different quantiles, we see that restricting the analysis to only mean gaps misses a large part of the bigger picture. Broadly speaking, as Figure 2 indicates, while the uncontrolled log income gap is positive throughout the distribution, the gap is higher for low-income businesses as compared to high-income businesses, with the gap for those at the 10th percentile (300 percent) and 25th percentile (154 percent) being substantially higher than the gap at the 75th and 90th percentiles (87 percent and 66 percent respectively). This phenomenon of higher gaps at lower levels of the earnings distribution 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).[18] 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

---

[18]Specifically, Arulampalam et al. (2007) define a sticky floor as the 10th percentile wage gap being higher than the 25th percentile wage gap by at least two percentage points.

floor, for instance India (Khanna, 2013) and China (Chi and Li, 2008).

Tables 4 and 5 report quantile regression results for the two specifications respectively for the pooled model at the 10th, 25th, 50th, 75th and 90th percentiles. The estimates show that controlling for various characteristics does not eliminate the caste gap observed in Figure 2. In both specifications, we see that the caste dummy is negative at all quantiles. For the first specification, the SCST dummy reflects an income gap of 53 percent at the 10th percentile, which continues to decline to 27 percent at the 90th percentile. Therefore, the sticky floor still persists even after controlling for variables such as age, marital status, years of education, urban location and state of residence. As more variables are added in the second specification, the caste dummy remains significant, but its magnitude becomes smaller at each of the percentiles. The sticky floor no longer prevails, as we do not observe a declining income gap as we move up the earnings distribution. The caste income gap increases from 10 percent at the 10th percentile to 16 percent at the median, declines to 10 percent at the 75th percentile and increases again up to 14 percent at the 90th percentile.

Pooled quantile regressions impose the restriction that the returns to included characteristics are the same for the two caste groups. Since, this assumption is not realistic, particularly in the Indian context, we carry out caste-specific quantile regressions, results of which are reported in Tables 6 and 7. Columns 1-5 in both tables report results for the SCST sample while Columns 6-10 contain results for non-SCSTs. While being married is mostly associated positively with income for non-SCSTs, it is either negative or insignificant for SCSTs. Being located in urban areas and number of hours spent in the business seem to confer greater benefits at the lower end of the earnings distribution than at the higher end, for both groups. On the other hand, gains from asset ownership are increasing across the distribution for both groups. Returns to other fixed or moving workplaces are higher at all percentiles for SCSTs as compared to non-SCSTs.

5.4   Decomposition of the Log Income Gap at Different Quantiles

We conduct the Melly decomposition separately using both specifications.[19] Table 8 shows the summary results with the raw difference, characteristics effect and coefficients effect for each of

---

[19]This is done using the STATA program "rqdeco" (Melly, 2007).

the 9 deciles using the non-SCST coefficients. As noted above, another set of estimates could be obtained using the SCST coefficients (Table 9). For both specifications, we are now able to evaluate how the explained and unexplained proportions change across the entire income distribution.

Based on non-SCST coefficients, we find that the raw log income gap shows a generally declining trend, decreasing from 0.99 at the 10th percentile to 0.6 at the median, 0.54 at the 80th percentile and then increasing slightly to 0.55 at the 90th percentile. The proportion of the income gap due to differences in characteristics increases as one moves up to the higher percentiles of the distribution although the increase is not steady. For instance, the characteristics effect declines from 54 percent at the 1st decile to hover at around 51 percent at the 2nd, 3rd, 4th and 5th deciles and then increases to about 58 percent at the highest deciles. A similar trend is observed using the second specification except that the log difference is somewhat smaller and the proportion of the characteristics effect larger due to the inclusion of more explanatory variables.

Mirroring these trends, we find in both specifications, that the unexplained component is larger at the lower end of the conditional earnings distribution than at the higher end. In the first specification, the share of the unexplained component falls from 46 percent at the 10th percentile to 42 percent at the 90th percentile. Using the second specification, the unexplained share declines from 12 percent at the 10th percentile to approximately 8 percent at the 90th percentile. However, it should be noted that while for the first specification, the coefficients effect remains statistically significant throughout the distribution, in the second specification, the coefficients effect is significant between the 40th and 80th percentiles i.e., for the businesses in the middle range of the earnings distribution, with the highest proportion of the unexplained component at the 6th decile.

Results using SCST coefficients reveal more clear trends in the share of the characteristics and coefficients effects. For instance, using the first specification, we find that the explained share increases gradually from 25 percent at the 10th percentile to about 54-55 percent at the 80th and 90th percentiles. This translates into a steady decline in the unexplained share from 75 percent at the lowest decile to 45 percent at the highest decile. Similarly, for the second

specification, we note a steady increase (decrease) in the explained (unexplained) share as we move up the earnings distribution. Figures 3 and 4 plot the raw gap, the contribution of characteristics and that of coefficients at each percentile of the earnings distribution using the second specification for the non-SCST and SCST coefficients respectively.

6 Discussion

Our results from both decomposition exercises strongly suggest that discrimination against SC-STs is very much characteristic of business activities or self-employment. Combining this with independent evidence of labor market discrimination discussed earlier, it appears that SCST individuals are subject to discrimination in both wage and self-employment, and therefore, self-employment as it stands today, may not be an answer or a solution to ending caste discrimination.

Discrimination in Self-Employment

As stated in Section 1, the unexplained component could contain the effects of characteristics that are not amenable to measurement. For instance, our data do not contain information on family background in running a business, prior experience of owner, type of customer base, risk aversion of the owner and so forth, all of which could affect business performance. Another unobserved factor that could constrain business performance is geographical segregation. Residential segregation, a by-product of historical discrimination, is still prevalent in India with Dalits living in their own segregated neighborhoods. If the main customer base of SCST businesses is their own community - and given that SCSTs are on average poorer and have lower purchasing power - they may have to keep their prices low in order to cater to members of their own group. For example, Clark and Drinkwater (2000) discuss how ethnic enclaves could be a source of advantage or disadvantage. 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, if the ethnic group is poor, then businesses set up by members of these groups may actually languish.

Discrimination manifests itself in self-employment primarily in the form of consumer and

credit market discrimination. Borjas and Bronars (1989) study consumer discrimination, and find that relative gains of entering self-employment are reduced for ethnic minorities because they have to compensate white consumers by lowering prices charged for goods and services. Coate and Tennyson (1992) study credit discrimination assuming that lenders are unable to observe entrepreneurial ability. Individuals from a group discriminated against in the labor market will receive less favorable terms in the credit market since lenders know that for such individuals, the opportunity cost of entering self-employment is lower, and, thus, they are willing to take more risks. Such groups will be charged higher interest rates, thus reducing the expected returns from self-employment. Empirical analyses using data from the United States show that the probability of loan denials and rates of interest charged on approved loans are higher for black-owned businesses than whites (Blanchflower et al., 2003) and probability of loan renewals is lesser for black and Hispanic-owned businesses (Asiedu et al., 2012).

While we cannot empirically test the specific channels through which discrimination operates in our data, there are other studies that qualitatively document the presence of active discrimination against SCST businesses. Prakash (2010) in his 2006-07 survey of 90 Dalit businesses in 13 districts spread across 6 states in India reports difficulty faced by them in obtaining initial formal credit in order to set up an enterprise, resulting in informal loans being taken at high interest rates. The ones who did successfully obtain institutional credit were those in partnerships with upper castes, or had local political contacts that could help in facilitating loan approvals. Kumar (2013) uses data from a 2002-03 All-India Debt and Investment survey and finds that public sector banks operating in areas with more upper castes tend to discriminate more against low caste loan applicants. Prakash (2010) also cites Dalit entrepreneurs who reported often charging less for their products than their upper-caste peers so that customers "forget" their castes. Jodhka (2010) through detailed interviews with Dalit entrepreneurs in two towns in northwest India 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.

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 sufficient to close the wealth gap between communities. Barsky et al. (2002) analyze the racial wealth gap in the US use 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 (2011)'s 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 upper 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 Dalit individuals 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. Moving to the larger point of whether the success of Dalit capitalists represents the triumph of markets as suggested by 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.

The simultaneous existence of discrimination in self-employment and wage employment, presents serious challenges for public policy, further complicated by the existence of pre-market discrimination for Dalits and Adivasis which results in lower and poorer quality of educational and skill attainment. For this reason, focusing on improvement in educational outcomes ought to be a key component of the strategy to enhance both employability as well as earnings of these marginalized groups.

While job quotas target salaried employment in the public sector, that may not be the appropriate instrument to tackle discrimination in others segments of the labor market, as well as discrimination faced by the self-employed. More research is needed before we can suggest a balanced and multi-faceted policy package. However, international evidence offers 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*. Power-Nomics 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.

Asiedu, E., Freeman, J.A., Nti-Addae, A. (2012). Access to credit by small businesses: How relevant and race, ethnicity and gender.*American Economic Review: Papers and Proceedings*, 102(3), 532-537

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

Bates, T.M. (1973). *Black Capitalism: A quantitative Analysis*. New York: Prager

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

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

Coate, S. and Tennyson, S. (1992). Labor Market Discrimination, Imperfect Information and Self-Employment. *Oxford Economic Papers*, 44, 272-288.

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. (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.

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.

Kumar, S.M. (2013). Does Access to Formal Agricultural Credit Depend on Caste? *World Development*, 43, 315-328.

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. (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.

Oaxaca, R., and Ransom, M. (1994). On discrimination and the decomposition of wage differentials. *Journal of Econometrics*, 61(1), 5-21.

Prakash, A. (2010). Dalit entrepreneurs in middle India. In Barbara Harriss-White and Judith Heyer (eds.) *The Comparative Political Economy of Development Africa and South Asia*, Routledge

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.

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
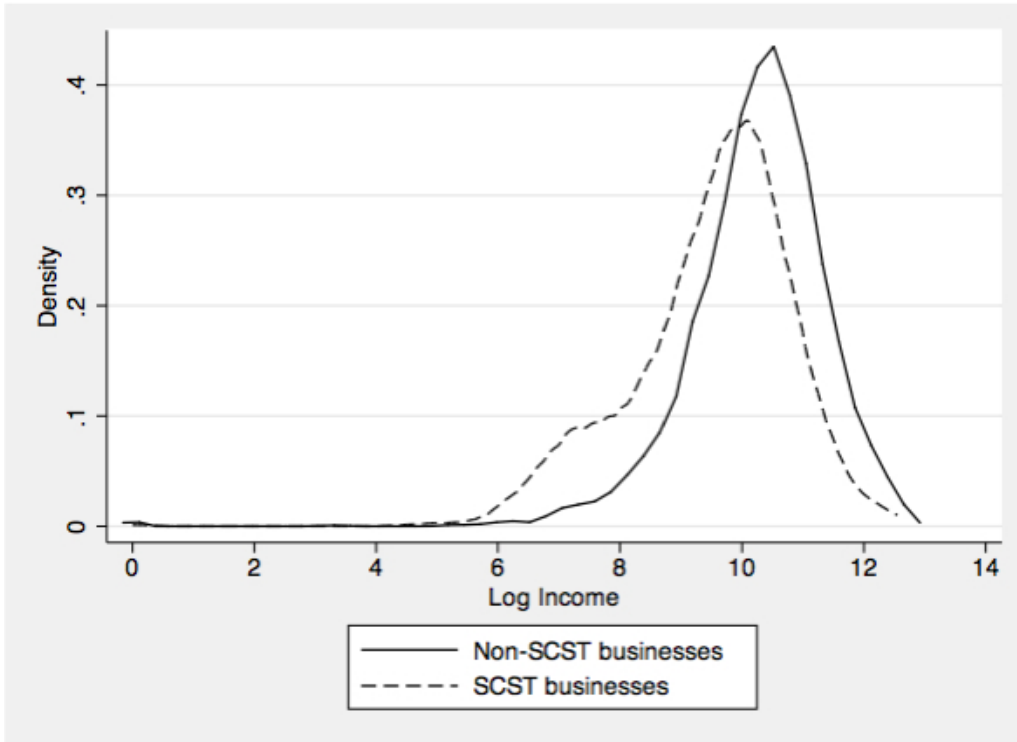
Figure 2: Caste Log Income Gap Across Percentiles & Average Gap
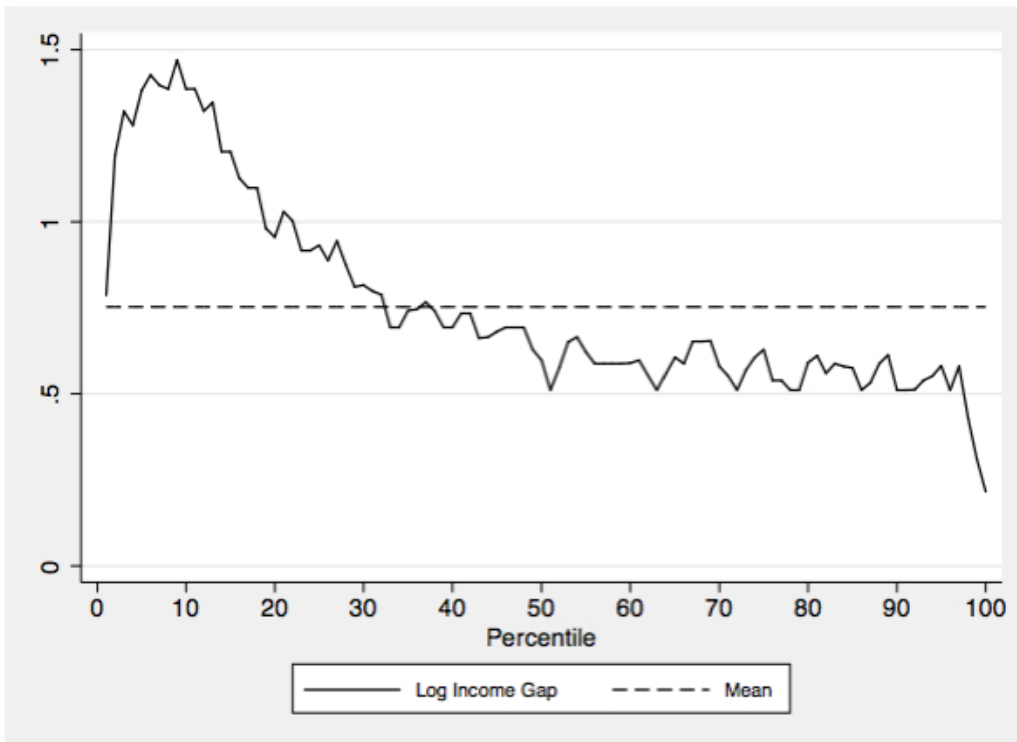
Figure 3: Melly Decomposition of Log Income Gap: Non-SCST coefficients
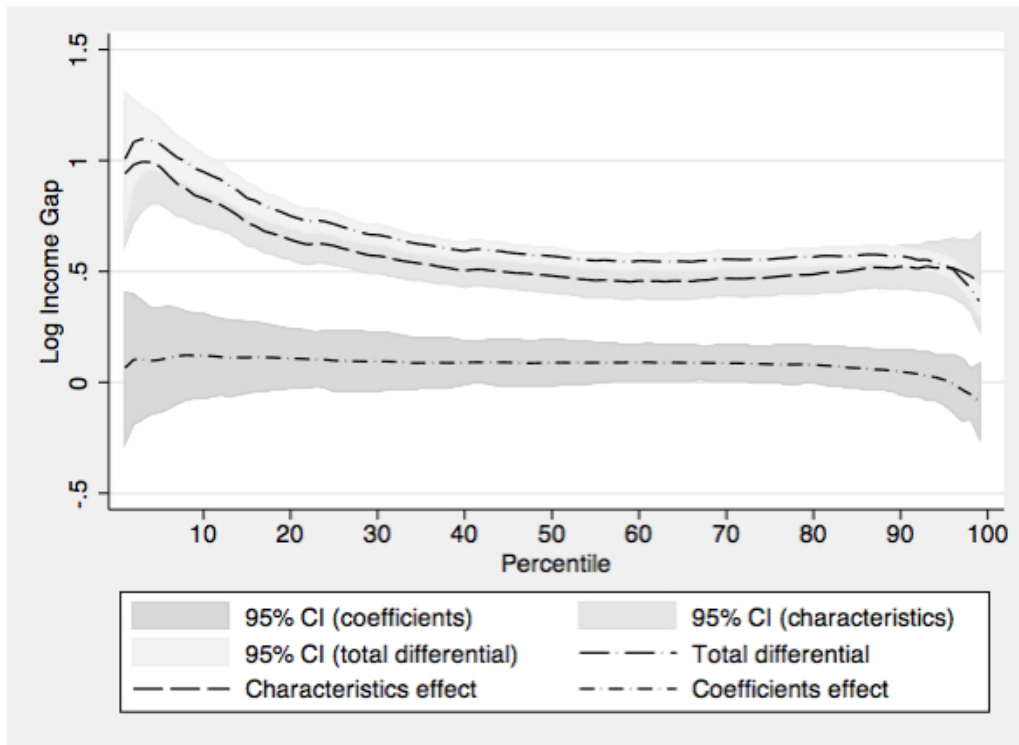
Figure 4: Melly Decomposition of Log Income Gap: SCST coefficients
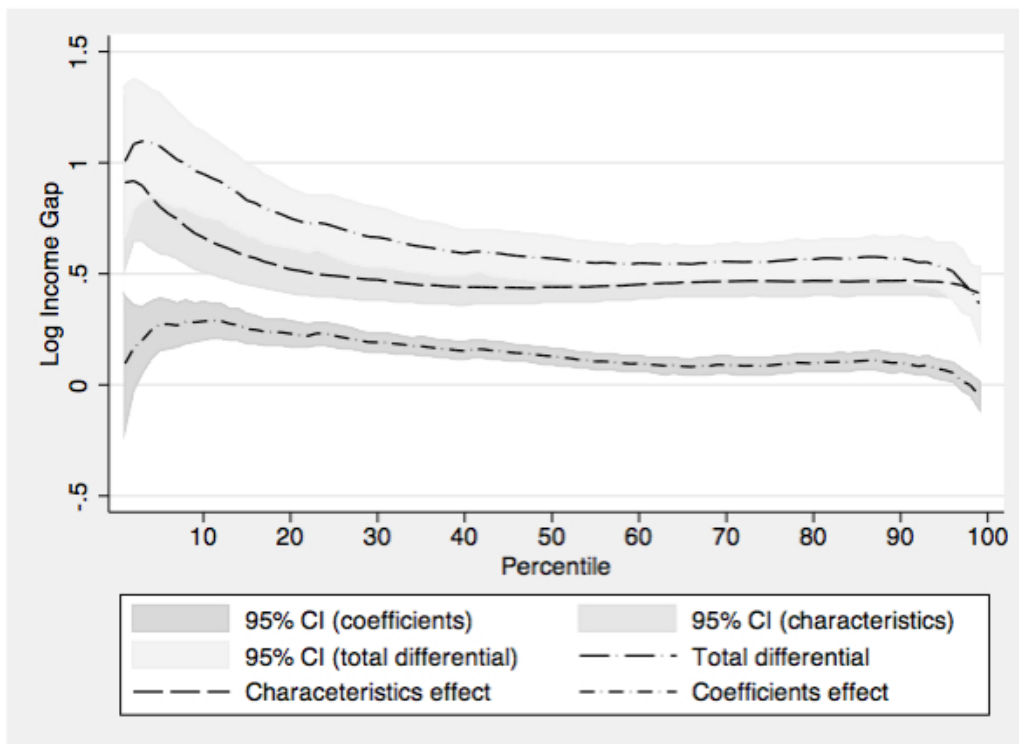
Table 1: Summary Statistics

Variable	All enterprises	SCST enterprises	Non-SCST enterprises
Outcome Variables:			
Gross Receipts (in Rs.)	108015.8	58804.02	118708.7
	(258019)	(98524.46)	(279809.3)
Net Income (in Rs.)	41726.15	25640.14	45218.44
	(45158.62)	(32726.04)	(46704.93)
Explanatory Variables:			
Individual characteristics			
Age (in years)	39.13	38.6	39.25
	(12.43)	(12.53)	(12.4)
Married	0.86	0.86	0.86
	(0.34)	(0.34)	(0.34)
Years of Education	7.79	5.66	8.25
	(4.64)	(4.57)	(4.53)
Household characteristics			
SCST	17.84		
	(0.38)		
Urban location	0.49	0.33	0.53
	(0.5)	(0.47)	(0.5)
Monthly Per Capita Expenditure (in Rs.)	1054.4	755.1	1119.3
	(1084.9)	(887.1)	(1112.8)
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
	(0.14)	(0.1)	(0.15)
Co-operative membership	0.03	0.02	0.04
	(0.18)	(0.15)	(0.19)
Village Panchayat or Ward Committee	0.11	0.13	0.11
	(0.31)	(0.33)	(0.31)
Business Characteristics			
Number of family workers	1.39	1.48	1.37
	(0.7)	(0.8)	(0.67)
Number of hours	2585.73	2065.16	2698.74
	(1614.59)	(1480.38)	(1620.45)
Workplace: home-based	0.25	0.26	0.25
	(0.43)	(0.44)	(0.43)
Workplace: other fixed	0.52	0.4	0.55
	(0.5)	(0.49)	(0.5)
Workplace: moving	0.23	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 Estimation: Pooled Sample & Caste-wise

Dependent variable: Log Income	Spec. 1			Spec. 2		
	Pooled Sample	SCST	Non-SCST	Pooled Sample	SCST	Non-SCST
SCST	-0.35***			-0.10***		
	(0.06)			(0.04)		
Age	0.02**	0.04**	0.02*	0.02***	0.04**	0.02**
	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.01)
Age squared/100	-0.02*	-0.04**	-0.02*	-0.02***	-0.04**	-0.02**
	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.01)
Married	0.14**	-0.08	0.19***	0.06	-0.05	0.09*
	(0.06)	(0.11)	(0.06)	(0.05)	(0.13)	(0.05)
Years of education	0.05***	0.06***	0.05***	0.01***	0.01	0.01***
	(0.00)	(0.01)	(0.00)	(0.00)	(0.01)	(0.00)
Urban location	0.72***	0.77***	0.71***	0.25***	0.26***	0.25***
	(0.05)	(0.09)	(0.05)	(0.04)	(0.08)	(0.04)
Asset ownership				0.15***	0.14***	0.15***
				(0.01)	(0.03)	(0.01)
Business or professional group membership				0.16**	0.19	0.16**
				(0.06)	(0.13)	(0.07)
Credit or savings group membership				-0.12**	-0.12	-0.13**
				(0.05)	(0.10)	(0.06)
Caste association membership				-0.05	-0.07	-0.05
				(0.06)	(0.09)	(0.07)
Development group/NGO membership				0.07	0.53*	0.06
				(0.08)	(0.29)	(0.08)
Co-operative membership				-0.10	-0.24	-0.09
				(0.10)	(0.24)	(0.10)
Village panchayat or ward committee				-0.05	-0.04	-0.05
				(0.06)	(0.07)	(0.08)
Log(number of hours)				0.55***	0.59***	0.53***
				(0.03)	(0.05)	(0.03)
Number of workers				-0.04*	-0.08	-0.03
				(0.03)	(0.05)	(0.03)
Workplace-other fixed				0.27***	0.26***	0.27***
				(0.05)	(0.08)	(0.06)
Workplace-moving				0.16***	0.09	0.18***
				(0.05)	(0.09)	(0.06)
Constant	9.50***	9.97***	9.41***	5.28***	5.12***	5.51***
	(0.27)	(0.34)	(0.30)	(0.28)	(0.50)	(0.32)
Observations	7271	1298	5973	7035	1252	5783
R^2	0.304	0.378	0.252	0.514	0.653	0.454

Note: Robust standard errors clustered at the district level are reported in parentheses. *** significant at 1%,** significant at 5%,* significant at 10%. State of residence dummy variables included.

Table 3: Blinder-Oaxaca Decomposition of Log Income

Log Income	Panel A Pooled		Panel B Non-SCST Coefficients		Panel C SCST Coefficients	
Variable	Spec.1	Spec.2	Spec.1	Spec.2	Spec.1	Spec.2
Difference	0.67***	0.64***	0.67***	0.64***	0.67 ***	0.64***
	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
Explained	0.32***	0.54***	0.3***	0.52***	0.39***	0.58***
	(0.05)	(0.07)	(0.04)	(0.07)	(0.06)	(0.08)
Unexplained	0.35***	0.10***	0.37***	0.12***	0.28***	0.06
	(0.05)	(0.04)	(0.06)	(0.04)	(0.05)	(0.04)
Contribution to the explained component:						
Years of education	0.13***	0.03***	0.12***	0.03***	0.15***	0.03
	(0.02)	(0.01)	(0.02)	(0.01)	(0.03)	(0.02)
Urban location	0.13***	0.05***	0.13***	0.05***	0.14***	0.05***
	(0.02)	(0.01)	(0.02)	(0.01)	(0.03)	(0.02)
Asset ownership		0.21***		0.21***		0.19***
		(0.02)		(0.02)		(0.04)
Total number of hours		0.22***		0.21***		0.23***
		(0.04)		(0.04)		(0.04)

Note: Robust standard errors clustered at the district level are reported in parentheses. *** significant at 1%,** significant at 5%,* significant at 10%.

Table 4: Quantile Regression: Specification 1 (Pooled Sample)

Log Income	Mean	Q10	Q25	Q50	Q75	Q90
SCST	-0.35***	-0.53***	-0.47***	-0.37***	-0.28***	-0.27***
	(0.06)	(0.07)	(0.05)	(0.13)	(0.04)	(0.04)
Age	0.02**	0.04***	0.04***	0.03***	0.02***	0.02***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Age squared/100	-0.02*	-0.05***	-0.04***	-0.04***	-0.02***	-0.02**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Married	0.14**	0.13	0.15***	0.11	0.08	0.05
	(0.06)	(0.09)	(0.05)	(0.16)	(0.05)	(0.05)
Years of education	0.05***	0.05***	0.05***	0.05***	0.06***	0.06***
	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.00)
Urban location	0.72***	0.94***	0.74***	0.60***	0.53***	0.49***
	(0.05)	(0.06)	(0.04)	(0.05)	(0.03)	(0.03)
Constant	9.50***	7.82***	8.68***	9.50***	10.24***	10.93***
	(0.27)	(0.22)	(0.27)	(0.18)	(0.18)	(0.21)
Observations	7271	7271	7271	7271	7271	7271
R^2	0.304					

Note: Robust standard errors clustered at the district level are reported in parentheses for OLS. Quantile regression standard errors in parentheses are bootstrapped using 100 replications. *** significant at 1%,** significant at 5%,* significant at 10%. State of residence dummy variables included.

Table 5: Quantile Regression: Specification 2 (Pooled Sample)

Log Income	Mean	Q10	Q25	Q50	Q75	Q90
SCST	-0.10***	-0.10**	-0.12***	-0.16***	-0.10***	-0.14***
	(0.04)	(0.05)	(0.04)	(0.03)	(0.03)	(0.03)
Age	0.02***	0.05***	0.04***	0.02***	0.02***	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)
Age squared/100	-0.02***	-0.06***	-0.05***	-0.02***	-0.02***	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Married	0.06	0.17***	0.09*	0.10**	0.08*	0.09*
	(0.05)	(0.06)	(0.05)	(0.04)	(0.04)	(0.05)
Years of education	0.01***	0.02***	0.01***	0.01***	0.02***	0.02***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Asset ownership	0.15***	0.12***	0.14***	0.15***	0.17***	0.17***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Urban location	0.25***	0.32***	0.27***	0.23***	0.20***	0.19***
	(0.04)	(0.04)	(0.03)	(0.02)	(0.02)	(0.03)
Business or professional group membership	0.16**	0.12**	0.06	0.13***	0.10**	0.17***
	(0.06)	(0.06)	(0.05)	(0.04)	(0.04)	(0.05)
Credit or savings group membership	-0.12**	-0.06	-0.10**	-0.16***	-0.17***	-0.19***
	(0.05)	(0.06)	(0.05)	(0.04)	(0.04)	(0.07)
Caste association membership	-0.05	-0.02	0.00	0.04	0.06*	0.05
	(0.06)	(0.06)	(0.05)	(0.04)	(0.03)	(0.05)
Development group/NGO membership	0.07	0.27**	0.12	0.01	0.04	-0.01
	(0.08)	(0.11)	(0.09)	(0.08)	(0.08)	(0.09)
Co-operative membership	-0.10	-0.17	-0.12	-0.02	-0.04	0.12*
	(0.10)	(0.11)	(0.09)	(0.07)	(0.07)	(0.08)
Village panchayat or ward committee	-0.05	-0.06	0.01	-0.01	0.00	-0.03
	(0.06)	(0.07)	(0.04)	(0.03)	(0.04)	(0.05)
Log(number of hours)	0.55***	0.69***	0.63***	0.54***	0.42***	0.32***
	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)
Number of workers	-0.04*	-0.15***	-0.10***	-0.06***	-0.03	0.00
	(0.03)	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)
Workplace-other fixed	0.27***	0.32***	0.27***	0.24***	0.13***	0.12***
	(0.05)	(0.05)	(0.03)	(0.03)	(0.03)	(0.04)
Workplace-moving	0.16***	0.28***	0.19***	0.16***	0.00	-0.01
	(0.05)	(0.06)	(0.04)	(0.04)	(0.03)	(0.04)
Constant	5.28***	2.59***	3.84***	5.52***	7.25***	8.38***
	(0.28)	(0.33)	(0.26)	(0.23)	(0.20)	(0.25)
Observations	7035	7035	7035	7035	7035	7035
R^2	0.514					

Note: Robust standard errors clustered at the district level are reported in parentheses for OLS. Quantile regression standard errors in parentheses are bootstrapped using 100 replications. *** significant at 1%,** significant at 5%,* significant at 10%. State of residence and industry dummy variables included.

Table 6: Caste-wise Quantile Regressions: Specification 1

Log Income	SCSTs					Non-SCSTs				
	Q10	Q25	Q50	Q75	Q90	Q10	Q25	Q50	Q75	Q90
Age	0.09***	0.05**	0.03*	0.05***	0.05***	0.04***	0.04***	0.03***	0.02***	0.02***
	(0.03)	(0.03)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Age squared/100	-0.10***	-0.06*	-0.03*	-0.04**	-0.05**	-0.04***	-0.04***	-0.04***	-0.02**	-0.02**
	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Married	-0.14	-0.28*	-0.14	-0.12	-0.27*	0.19**	0.19***	0.14***	0.11*	0.08
	(0.17)	(0.16)	(0.10)	(0.13)	(0.16)	(0.09)	(0.05)	(0.04)	(0.06)	(0.06)
Years of education	0.05***	0.05***	0.05***	0.06***	0.06***	0.05***	0.05***	0.05***	0.06***	0.06***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
Urban location	0.97***	0.81***	0.62***	0.52***	0.38***	0.92***	0.72***	0.58***	0.53***	0.49***
	(0.14)	(0.09)	(0.07)	(0.06)	(0.09)	(0.05)	(0.03)	(0.03)	(0.03)	(0.04)
Constant	8.22***	9.45***	10.66***	10.41***	11.01***	7.87***	8.48***	9.40***	10.10***	10.53***
	(0.60)	(0.57)	(0.37)	(0.30)	(0.46)	(0.27)	(0.30)	(0.14)	(0.19)	(0.26)
Observations	1298	1298	1298	1298	1298	5973	5973	5973	5973	5973

Note: Standard errors in parentheses are bootstrapped using 100 replications. *** significant at 1%, ** significant at 5%, * significant at 10%. State of residence dummy variables included.

Table 7: Caste-wise Quantile Regressions: Specification 2

Log Income	SCSTs					Non-SCSTs				
	Q10	Q25	Q50	Q75	Q90	Q10	Q25	Q50	Q75	Q90
Age	0.05**	0.05***	0.02	0.03**	0.02	0.05***	0.04***	0.02***	0.02***	0.01
	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Age squared/100	-0.06***	-0.06***	-0.03	-0.04**	-0.02	-0.06***	-0.05***	-0.03***	-0.02***	-0.01
	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Married	0.11	0.00	0.08	-0.01	-0.24	0.18**	0.10*	0.13***	0.11***	0.12**
	(0.14)	(0.11)	(0.08)	(0.12)	(0.15)	(0.07)	(0.06)	(0.04)	(0.04)	(0.06)
Years of education	0.02	0.00	0.00	0.01*	0.00	0.02***	0.01***	0.01***	0.02***	0.02***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
Asset ownership	0.13***	0.11***	0.14***	0.16***	0.20***	0.12***	0.14***	0.15***	0.17***	0.17***
	(0.04)	(0.03)	(0.02)	(0.02)	(0.03)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Urban location	0.21**	0.27***	0.16***	0.14**	0.03	0.35***	0.28***	0.23***	0.21***	0.19***
	(0.10)	(0.07)	(0.06)	(0.06)	(0.06)	(0.05)	(0.03)	(0.03)	(0.03)	(0.04)
Business or professional group membership	-0.13	0.00	0.26*	0.10	0.04	0.14**	0.05	0.14***	0.13***	0.18***
	(0.22)	(0.16)	(0.15)	(0.10)	(0.16)	(0.06)	(0.05)	(0.05)	(0.04)	(0.07)
Credit or savings group membership	0.15	-0.01	-0.15	-0.22**	-0.27***	-0.11	-0.12**	-0.16***	-0.16***	-0.19***
	(0.18)	(0.08)	(0.09)	(0.11)	(0.09)	(0.07)	(0.05)	(0.05)	(0.05)	(0.07)
Caste association membership	-0.15	-0.12	-0.20**	-0.12	-0.02	-0.02	0.04	0.07*	0.06	0.07
	(0.17)	(0.09)	(0.09)	(0.10)	(0.12)	(0.06)	(0.06)	(0.04)	(0.04)	(0.05)
Development group/NGO membership	0.70	0.58*	0.48	0.52	0.52	0.24*	0.10	0.01	-0.02	-0.04
	(0.56)	(0.32)	(0.49)	(0.54)	(0.59)	(0.13)	(0.11)	(0.07)	(0.09)	(0.09)
Co-operative membership	-0.30	-0.21	-0.36	-0.04	0.05	-0.15	-0.10	-0.03	-0.02	0.09
	(0.31)	(0.14)	(0.27)	(0.22)	(0.45)	(0.12)	(0.08)	(0.08)	(0.06)	(0.07)
Village panchayat or ward committee	-0.02	0.00	-0.05	-0.14	-0.12	-0.11	0.01	0.02	0.05	-0.01
	(0.16)	(0.10)	(0.06)	(0.08)	(0.13)	(0.08)	(0.05)	(0.04)	(0.04)	(0.05)
Log(number of hours)	0.66***	0.64***	0.61***	0.51***	0.38***	0.70***	0.60***	0.50***	0.37***	0.30***
	(0.06)	(0.04)	(0.04)	(0.05)	(0.05)	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)
Number of workers	-0.19*	-0.03	-0.09**	-0.01	-0.02	-0.16***	-0.08***	-0.04*	0.00	0.01
	(0.10)	(0.06)	(0.04)	(0.05)	(0.04)	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)
Workplace-other fixed	0.36***	0.29***	0.36***	0.23***	0.18**	0.31***	0.28***	0.23***	0.12***	0.14***
	(0.14)	(0.07)	(0.10)	(0.08)	(0.09)	(0.05)	(0.03)	(0.03)	(0.04)	(0.04)
Workplace-moving	0.33**	0.25**	0.27***	0.18*	0.08	0.26***	0.21***	0.16***	-0.00	-0.00
	(0.13)	(0.10)	(0.08)	(0.10)	(0.11)	(0.06)	(0.05)	(0.04)	(0.04)	(0.06)
Constant	3.58***	3.91***	5.10***	6.06***	7.71***	2.56***	4.32***	5.88***	7.68***	8.73***
	(0.79)	(0.52)	(0.48)	(0.51)	(0.50)	(0.35)	(0.33)	(0.23)	(0.23)	(0.26)
Observations	1252	1252	1252	1252	1252	5783	5783	5783	5783	5783

Note: Standard errors in parentheses are bootstrapped using 100 replications. *** significant at 1%, ** significant at 5%, * significant at 10%. State of residence and industry dummy variables included.

Table 8: Melly Decomposition of Log Income (Non-SCST Coefficients)

| Decile | Panel A Spec.1 | | | Panel B Spec.2 | | |
	Col.1 Difference	Col.2 Characteristics	Col.3 Coefficients	Col.4 Difference	Col.5 Characteristics	Col.6 Coefficients
10	0.99***	0.54***	0.45***	0.95***	0.83***	0.12
	(0.05)	(0.09)	(0.1)	(0.04)	(0.06)	(0.1)
20	0.82***	0.42***	0.40***	0.75***	0.65***	0.1
	(0.03)	(0.07)	(0.07)	(0.03)	(0.05)	(0.07)
30	0.72***	0.37***	0.35***	0.66***	0.57***	0.09
	(0.02)	(0.06)	(0.06)	(0.02)	(0.04)	(0.07)
40	0.65***	0.33***	0.32***	0.59***	0.5***	0.09*
	(0.02)	(0.05)	(0.05)	(0.02)	(0.04)	(0.05)
50	0.6***	0.31***	0.29***	0.57***	0.48***	0.09*
	(0.02)	(0.05)	(0.04)	(0.02)	(0.04)	(0.05)
60	0.58***	0.32***	0.26***	0.55***	0.46***	0.09**
	(0.02)	(0.05)	(0.04)	(0.02)	(0.04)	(0.04)
70	0.55***	0.32***	0.23***	0.55***	0.47***	0.08**
	(0.02)	(0.04)	(0.04)	(0.02)	(0.04)	(0.04)
80	0.54***	0.31***	0.23***	0.57***	0.49***	0.08*
	(0.02)	(0.05)	(0.04)	(0.02)	(0.04)	(0.04)
90	0.55***	0.32***	0.23***	0.57***	0.52***	0.05
	(0.02)	(0.05)	(0.05)	(0.02)	(0.05)	(0.05)

Note: Standard errors in parentheses are based on bootstrapping with 100 replications. ***
significant at 1%,** significant at 5%,* significant at 10%.

Table 9: Melly Decomposition of Log Income (SCST Coefficients)

	Panel A Spec.1			Panel B Spec.2		
Decile	Col.1 Difference	Col.2 Characteristics	Col.3 Coefficients	Col.4 Difference	Col.5 Characteristics	Col.6 Coefficients
10	0.99***	0.25***	0.74***	0.95***	0.66***	0.29***
	(0.1)	(0.05)	(0.05)	(0.1)	(0.08)	(0.04)
20	0.82***	0.27***	0.55***	0.75***	0.52***	0.23***
	(0.07)	(0.04)	(0.03)	(0.07)	(0.06)	(0.03)
30	0.72***	0.29***	0.43***	0.66***	0.47***	0.19***
	(0.06)	(0.03)	(0.02)	(0.07)	(0.05)	(0.02)
40	0.65***	0.3***	0.35***	0.59***	0.44***	0.15***
	(0.05)	(0.03)	(0.02)	(0.05)	(0.04)	(0.02)
50	0.6***	0.3***	0.3***	0.57***	0.44***	0.13***
	(0.04)	(0.03)	(0.02)	(0.05)	(0.04)	(0.02)
60	0.58***	0.3***	0.28***	0.55***	0.45***	0.1***
	(0.04)	(0.03)	(0.02)	(0.04)	(0.03)	(0.02)
70	0.55***	0.31***	0.24***	0.55***	0.46***	0.09***
	(0.04)	(0.03)	(0.02)	(0.04)	(0.03)	(0.02)
80	0.54***	0.3***	0.24***	0.57***	0.47***	0.1***
	(0.04)	(0.03)	(0.02)	(0.04)	(0.03)	(0.02)
90	0.65***	0.3***	0.25***	0.57***	0.47***	0.1***
	(0.05)	(0.03)	(0.02)	(0.05)	(0.03)	(0.02)

Note: Standard errors in parentheses are based on bootstrapping with 100 replications. ***
significant at 1%,** significant at 5%,* significant at 10%.

A 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 household members 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.