

CDE
March 2015

DISADVANTAGE AND DISCRIMINATION IN SELF- EMPLOYMENT: CASTE GAPS IN EARNINGS IN INDIAN SMALL 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
(Revised Version)

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

Disadvantage and Discrimination in Self-Employment: Caste Gaps in Earnings in Indian Small Businesses*

Ashwini Deshpande[†]

Smriti Sharma[‡]

March, 2015

Abstract

Using the 2004-05 India Human Development Survey data, we estimate and decompose the earnings of household businesses owned by Scheduled Castes and Tribes (SCSTs) and non-SCSTs across the earnings distribution. We find clear differences in characteristics between the two types of businesses with the former faring significantly worse. The mean decomposition reveals that as much as 55 percent of the caste earnings gap could be attributed to the unexplained component. Quantile regressions suggest that gaps are higher at lower deciles, providing some evidence of a sticky floor. Finally, 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.

JEL classification codes: J31, J71, C21, O15, O17

Keywords: Caste, discrimination, household business, earning gaps, quantile decomposition, India

*We are grateful to Deepti Goel for detailed comments on an earlier draft; to Shantanu Khanna for discussions on the quantile regression decomposition methodology; and to seminar participants at Delhi School of Economics, Brown University, New Delhi office of the World Bank, UNU-WIDER, Nordic Conference on Development Economics 2014, IEA-World Bank Roundtable on Shared Prosperity, and the IHDS-NCAER 2014 conference on 'Human Development in India: Evidence from IHDS' where earlier versions of the paper were presented, for useful comments. We are responsible for all remaining errors and omissions.

[†]Department of Economics, Delhi School of Economics, University of Delhi, Delhi - 110007, India. Email: ashwini@econdse.org

[‡]UNU-WIDER, Katajanokanlaituri 6B, 00160 Helsinki, Finland. Email: smriti@wider.unu.edu

1 Introduction

Ethnic and racial discrimination in labour markets, as manifested in wage and occupational attainment gaps, has been widely examined (e.g., Altonji and Blank 1999; Antecol and Bedard 2004; Atal et al. 2009). In India too, labour market discrimination against marginalized caste groups (Scheduled Castes and Scheduled Tribes, SCs and STs hereafter) is well documented, with SCs and STs earning significantly lower wages and being allocated to less respectable jobs as compared to upper castes, even after controlling for their productive characteristics (Banerjee and Knight 1983; Madheswaran and Attewell 2007; Das and Dutta 2007). However, the disadvantage faced by these groups may not be limited just to wage employment, and could extend to the realm of self-employment as well. While there is a sizable literature from the United States that studies racial differences in entrepreneurship in terms of business creation rates, survival, employment, profits and net worth (e.g., Fairlie 2004; Fairlie 2006; Lofstrom and Bates 2013; Ahn 2011), examining such issues in the Indian context has not been feasible until recently, largely due to data constraints.¹

Our paper attempts to fill this gap for the Indian case, by assessing the presence of caste discrimination in household non-farm businesses ('businesses' hereafter), which has been possible due to the recent availability of good quality earnings data from such businesses. Given the nature and small scale of operations of these businesses, catering mostly to customers in the local community, it is highly plausible that businesses owned by low-caste owners, face discrimination at the hands of customers, suppliers and lenders, since their caste status is easily identifiable and salient, unlike in large businesses with complex ownership and management structures, where observing the caste of the owners might be less straightforward. However, discrimination can be directed towards larger low-caste businesses too: in personal interviews, rich SC entrepreneurs have discussed their individual battles with caste discrimination as they started their businesses.² There are other ethnographic accounts as well (Jodhka 2010; Prakash 2010) that indicate the presence of persistent disadvantage and discrimination in the self-employment arena, which forms the motivation for the present study.

To the best of our knowledge, ours is the first paper to examine caste gaps in earnings from businesses for India using decomposition methods. We use the India Human Develop-

¹The Employment-Unemployment Survey conducted by the National Sample Survey (NSS), one of the most widely used Indian micro-data sets, collects data on earnings of salaried employees and casual workers, but not of the self-employed.

²See, for instance, interviews in Outlook Business, May 2, 2009, p.25

ment Survey data for 2004-05, and employ two methodologies for understanding the earnings structure of businesses: OLS estimation of mean earnings for businesses owned by SCs and STs, 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 decomposition strategies to decompose the earnings gap between SCST and non-SCST businesses into explained and unexplained components (with the latter being indicative of discrimination), at the mean and at various quantiles of the earnings distribution.³

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 richer owners, 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 mean decomposition reveals that depending on the specification of 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 supporting a ‘sticky floor’, a phenomenon observed in the context of gender wage gaps in developing countries (e.g., Chi and Li 2008; Carrillo et al. 2014).

In addition to contributing to the broader literature on racial and ethnic disparities in small business ownership from a developing country perspective, this paper has significant policy implications, particularly in the context of the current discourse on ‘Dalit Capitalism’ in India - inspired by ‘Black Capitalism’ in the United States - by the Dalit Indian Chamber of Commerce

³Large Indian data sets such as the NSS and Economic Census define 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. Even though these large omnibus administrative categories mask intra-group heterogeneity, it is standard practice to use these for empirical estimation since data are available only for these categories.

and Industry (DICCI).⁴ DICCI believes that Dalits should enter business and industry sectors as entrepreneurs and use this route to become ‘job givers, and not job seekers’ especially for others in their own community, and enhance their wealth, instead of being dependent on the state for benefits. However, our results show that majority of Dalit businesses are small, owner-operated, survivalist household enterprises that do not have the potential to generate either employment or wealth. Further, our results also suggest that discriminatory tendencies that characterize labour markets are in fact not absent in other markets that are critical to the success of entrepreneurial activities, and businesses owned by SCs and STs fall behind.

The rest of this paper is organized as follows: Section 2 contains a literature review; Section 3 outlines the methodology; Section 4 discusses the data and descriptive statistics; Section 5 presents the results. Section 6 concludes.

2 Review of Related Literature

Recently, a few papers have relied on data sources such as the Economic Census of India to show 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) in descriptive analyses 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’. Deshpande and Sharma (2013) 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 and found that the MSME sector exhibits very clear differences along business owners’ caste and gender, in virtually all business characteristics.

This evidence of systematic differences, however, does not prove discrimination along caste lines; all the gaps in performance could, in principle, be accounted for by these differences in characteristics of SCST and non-SCST businesses.⁵ For example, in the United States, racial disparities in asset ownership and family background in self-employment (with blacks being

⁴SCs use the term Dalit (meaning oppressed) as a term of pride. More details about DICCI can be found at www.dicci.org.

⁵The fact that SCSTs possess inferior characteristics suggests some ‘pre-market’ discrimination (Deshpande 2011; Thorat and Newman 2010).

more disadvantaged than whites) are among the most important factors leading to differences in business creation and performance (Dunn and Holtz-Eakin 2000; Hout and Rosen 2000; Fairlie and Robb 2007). However, even after controlling for differences in endowments, a significant proportion of the performance gap remains unexplained, and that could be on account of discrimination or some unobserved differences in behaviour such as ability and risk aversion, or some factors not amenable to measurement.

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 from self-employment are reduced for ethnic minorities because they have to compensate white consumers by lowering prices charged. Coate and Tennyson (1992) study credit discrimination assuming that lenders are unable to observe entrepreneurial ability. Individuals from a group discriminated against in the labour market will receive less favorable terms in the credit market since lenders know that for such individuals, the opportunity cost of being self-employed 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, *ceteris paribus*. 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).

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. Kumar (2013) finds that public sector banks operating in upper-caste dominated areas 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 identity in mobilizing finance and leasing business space. Further, residential segregation, a by-product of historical discrimination, is still prevalent in India with Dalits living in their own segregated neighbourhoods. If the main customer base of SCST businesses is their own community - and

given that SCSTs are on average poorer with 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 that while a concentration of co-ethnics can be advantageous and provide a captive market for ethnic goods, minority businesses could also languish if the ethnic group is poor.

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 (the explained component or composition effect) and differences in returns to these endowments (the unexplained component or coefficients effect) (Blinder 1973; Oaxaca 1973). While the unexplained component can be attributed to discrimination, it is highly plausible that this residual also includes the effects of either unmeasurable or unobservable characteristics. All decomposition exercises are subject to this caveat. However, it is equally true that some 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):

$$w_{ig} = X_i^g \beta^g + u_i^g \quad (1)$$

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:

$$G = \bar{X}^n \hat{\beta}^n - \bar{X}^s \hat{\beta}^s \quad (2)$$

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:

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

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

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

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.⁶

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 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

Generalizing the traditional Blinder-Oaxaca decomposition to analyze earnings gaps at different parts of the earnings distribution, 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

⁶One could also construct an alternative counterfactual by assuming that the non-discriminatory earnings structure is the one applicable to the SCSTs.

quantile function $Q_\theta(w|X)$ can be expressed using a linear specification for each group as follows:

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

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, 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 with replacement to construct the counterfactual unconditional earnings distribution, and comes at the cost of increased computational time. Melly (2006) proposed a procedure that is less computationally intensive and faster by integrating 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). 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:

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

This yields the following decomposition:

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

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

We use the India Human Development Survey (IHDS) for 2004-05, which is a nationally representative data set covering 41,554 households. 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 household non-farm businesses 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.

We restrict the sample to those states where there are at least 50 household businesses, leaving us 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) 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.⁷

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 labour in these businesses, which would affect the calculation of 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 labour is not reported, it cannot be assumed that no labour was actually hired. Thus, data do not allow us to clearly distinguish between hired and unpaid family labour, resulting in the inability to estimate 'true' net income. We, thus, use the net income figures in the data as reported. 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 misreporting - for an analysis focusing on enterprise performance, income is the most appropriate outcome to consider.

As explanatory variables, we use individual specific variables such as age, marital status and standard years of education completed of the decision-maker; 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 and agricultural, milk or other co-operative; and 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.⁸

As our sample is limited to only those households that operate businesses, a potential limitation of our estimations is that coefficients of earnings regressions may be biased since individuals and households do not randomly select into self-employment. Unfortunately, our data set does not provide us with suitable instruments to correct for selection.

⁷We drop 1156 businesses that are run by female owners. This constitutes about 13 percent of the sample.

⁸Definitions of variables are available in Appendix A.

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. Of the total 7288 businesses, 1300 are owned by SCSTs (17.8 percent) and the remaining 5988 by non-SCSTs (82.2 percent).⁹

[Table 1 here]

In terms of performance, the average net income for non-SCST businesses (Rs. 45,218) is 1.76 times that for SCST businesses (Rs. 25,640). A similar pattern can be seen in the average gross receipts. 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.

[Figure 1 here]

This large difference in business performance could be on account of a variety of characteristics, in most of which there are clear differences between SCSTs and non-SCSTs. The primary decision maker is on average 39 years old and 86 percent of them are married. These numbers are similar across SCST and non-SCST decision makers. However, average years of education, 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 as reflected in asset ownership, in that out of the 16 assets in the questionnaire, non-SCSTs own approximately 8 while SCSTs own around 5.¹⁰ We create a wealth index using Principal Components Analysis, and divide the sample into three groups following Filmer and Pritchett (2001): those lying in the bottom 40 percent (poor), middle 40 percent (middle), and the top 20 percent (rich). By this somewhat arbitrary 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

⁹Since the decomposition methodology is applicable only to pairs of groups, we club together relatively similar groups, albeit with intragroup heterogeneity, into two broad dissimilar groups.

¹⁰The 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.

respectively are in the middle, and 7.4 percent of SCSTs and 22.7 percent of non-SCSTs are rich.

We also examine 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 and co-operatives is miniscule across the board. In terms of political networks, 12.5 percent of SCSTs have someone in, or close to, their households who has been an official in local bodies while for non-SCSTs, the corresponding figure is 10.6 percent.¹¹ 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 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.

In terms of business location, about 25 percent of businesses are home-based, and this proportion does not differ by caste. 34 percent of SCSTs and 20 percent of non-SCSTs have mobile workplaces, while the proportions of non-SCSTs and SCSTs with fixed workplaces are 55 and 39 respectively. To the extent a fixed workplace indicates permanency, it suggests that non-SCST businesses are more stable and less makeshift.

The most important sector for these businesses is ‘wholesale, retail trade and restaurants and hotels’, which include activities such as running of ‘kirana’ (neighbourhood grocery) stores, other grocery and general stores, and petty shops. 56 percent of non-SCST businesses and 44.5 percent of SCST businesses are involved in this sector. About 13 percent of businesses are in manufacturing activities, and this proportion does not vary by caste. The major activities here are blacksmiths, carpenters, and flour mills. About 16 percent of businesses are in the ‘community, social and personal services’ sector. This includes activities such as barbers, cycle

¹¹This could possibly reflect the operation of the mandatory 22.5 percent caste quotas in local bodies for SCSTs.

repair shops, and tailoring. These examples also corroborate our intuition that these businesses are engaged in low-end activities, and are more survivalist than entrepreneurial.

Approximately 6.5 percent of businesses are in the ‘transport, storage and communication’ sector, with the proportion being the same across castes. 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. Proportions in ‘construction’ and ‘financing, insurance, real estate and business services’ are small, involving only about 2 percent of businesses each. Businesses engaged in ‘mining and quarrying’ and ‘electricity, gas and water’ sectors are practically non-existent, as expected, since 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, 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.¹²

[Table 2 here]

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

¹²As 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.

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 mobile are correlated with higher incomes than home-based businesses.

Pooled 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 also carry out caste-specific OLS regressions. 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 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 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 that 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 3.¹³ 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.

[Table 3 here]

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

¹³This is done using the STATA program 'oaxaca' (Jann, 2008).

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 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 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).¹⁴ 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), China (Chi and Li 2008), and Vietnam (Pham and Reilly 2007). In fact, Carrillo et al. (2014) find that gender wage gaps in poorer and more unequal countries exhibit sticky floors whereas glass ceilings characterize richer and less unequal ones, using a sample of 12 Latin American countries.

[Figure 2 here]

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 reduces but 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.

[Tables 4 and 5 here]

Results of caste-specific quantile regressions 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.

¹⁴Specifically, 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.

[Tables 6 and 7 here]

5.4 Quantile Decompositions of Log Income Gaps

We conduct the quantile decompositions separately using both specifications.¹⁵ Table 8 shows the summary results with the raw difference, characteristics effect and coefficients effect for each of the 9 deciles using the non-SCST coefficients. As noted above, another set of estimates could be obtained using the SCST coefficients (Table 9).

[Table 8 here]

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 explained proportion larger due to the inclusion of more explanatory variables.

[Table 9 here]

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.

¹⁵This is done using the STATA program ‘rqdeco’ (Melly, 2007).

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.

[Figures 3 and 4 here]

6 Conclusion

In this paper, our objective has been to assess the presence of caste based discrimination in small household businesses using the large-scale nationally representative India Human Development Survey of 2004-05. Our results show that businesses owned by SCSTs fare significantly worse in terms of owner's education, household economic status and business characteristics, as compared to their non-SCST counterparts. Depending upon the specification of variables used, 19-55 percent of the mean earnings gap between businesses owned by SCSTs and non-SCSTs cannot be explained by differences in characteristics. Further, we find that there is substantial heterogeneity in earnings gaps across the earnings distribution, thereby necessitating the use of quantile regression based decomposition methods. These indicate that the proportion of the earnings gap on account of differences in characteristics increases in the higher deciles of the conditional earnings distribution.

In addition to being the first to examine this question for India, this paper's findings confirm patterns that have been observed in the context of racial and ethnic differences in entrepreneurship in other countries such as the United States and United Kingdom. However, unlike the United States, for instance, where a number of migrant groups such as the Koreans and Japanese have used self-employment as a way to achieve upward economic and social mobility, that does not appear to be the case for India, as suggested by our findings and also those in Iyer et al. (2013) and Deshpande and Sharma (2013). This also suggests that the exuberance surrounding

Dalit Capitalism may be somewhat misplaced since the reality of most SC and ST businesses is in stark contrast to the success of a few established low-caste industrialists.

The simultaneous existence of discrimination against SCs and STs in self-employment and wage employment presents serious challenges for public policy. While caste-based job quotas in India target public sector salaried employment, that may not be the appropriate instrument to tackle discrimination faced by the self-employed. One such recent move is a 2012 public procurement policy for micro and small enterprises (MSEs) that mandates 4 percent of government procurement to be from MSEs owned by SCs and STs. Other multi-pronged measures need to be devised that would tackle discrimination in both spheres.

A 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) find that roughly two-thirds of the mean difference in wealth between blacks and whites in the US can be explained by differences in earnings from all sources, which suggests that substantial wealth gaps remain even after controlling for earning differences. Whether an increase in business ownership by SCs and STs translates into narrowing wealth gaps would have to be the subject matter of a future exercise.

References

- Ahn, T. (2011). Racial differences in self-employment exits. *Small Business Economics*, 36, 169-186.
- Allen, W.D. (2000). Social networks and self-employment. *Journal of Socio-Economics*, 29, 487-501.
- Altonji, J.G., & Blank, R.M. (1999). Race and gender in the labor market. In O. Ashenfelter & D. Card (Ed.) *Handbook of Labor Economics*, Edition 1, vol. 3, (pp. 3143-3259). Elsevier.
- Antecol. H., & Bedard, K. (2004). The Racial Wage Gap: The Importance of Labor Force Attachment Differences Across Black, Mexican, and White Men. *Journal of Human Resources*, 39 (2), 564-583.
- Arulampalam, W., Booth, A. L., & 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
- Atal, J.P., Nopo, H., & Winder, N. (2009). New Century, Old Disparities Gender and Ethnic Wage Gaps in Latin America. IDB Working Paper No. 109
- Banerjee, B., & Knight, J.B. (1985). Caste in Indian urban labour market. *Journal of Development Economics*, 17, 277-307
- Barsky, R., Bound, J., Charles, K.K., & Lupton, J.P. (2002). Accounting for the black-white wealth gap: A nonparametric approach. *Journal of the American Statistical Association*, 97, 663-673
- Blanchflower, D.G., Levine, P.B., & Zimmerman, D.J. (2003). Discrimination in the Small-Business Credit Market. *Review of Economics and Statistics*, 85 (4), 930-943.
- Blinder, A. S. (1973). Wage discrimination: Reduced form and structural estimates. *Journal of Human Resources*, 8, 436-455.
- Borjas, G., & Bronars, S. (1989). Consumer discrimination and self-employment. *Journal of Political Economy*, 97, 581-605.
- Carrillo, P., Gandelman, N., & Robano, V. (2014). Sticky floors and glass ceilings in Latin America. *Journal of Economic Inequality*, 12(3), 339-361

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

Coate, S., & 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.

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

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

Deshpande, A., & 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., & 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). Racial trends in ethnic and racial business ownership. *Small Business Economics*, 23, 203-218.

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

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

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

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

- Jann, B. (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)
- Karunakaran, N. (2009). Bite the Caste Bullet. Outlook Business, May 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
- Kumar, S.M. (2013). Does Access to Formal Agricultural Credit Depend on Caste? *World Development*, 43, 315-328.
- Lofstrom, M. & Bates, T. (2013). African Americans' pursuit of self-employment. *Small Business Economics*, 40, 73-86.
- Machado, J., & J. Mata (2005). Counterfactual Decomposition of Changes in Wage Distributions using Quantile Regression. *Journal of Applied Econometrics*, 20, 445-465.
- Madheswaran, S., & 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. Unpublished manuscript, University of St. Gallen
- Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review*, 14, 693-709.
- Oaxaca, R., & Ransom, M. (1994). On discrimination and the decomposition of wage differentials. *Journal of Econometrics*, 61(1), 5-21.
- Pham, T., & Reilly, B. (2007). The gender pay gap in Vietnam: 1993-2002: A quantile regression approach. *Journal of Asian Economics*, 18(5), 775-808.
- 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*. London: Routledge
- Thorat, S., & Newman, K. (Eds.) (2010). *Blocked by Caste: Economic Discrimination in Modern India*. New Delhi: Oxford University Press.

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

Table 1: Summary Statistics

Variable	All enterprises	SCST enterprises	Non-SCST enterprises
Outcome Variables:			
Gross Receipts (in Rs.)	108015.8 (258019)	58804.02 (98524.46)	118708.7 (279809.3)
Net Income (in Rs.)	41726.15 (45158.62)	25640.14 (32726.04)	45218.44 (46704.93)
Explanatory Variables:			
Individual characteristics			
Age (in years)	39.13 (12.43)	38.6 (12.53)	39.25 (12.4)
Married	0.86 (0.34)	0.86 (0.34)	0.86 (0.34)
Years of Education	7.79 (4.64)	5.66 (4.57)	8.25 (4.53)
Household characteristics			
SCST	17.84 (0.38)		
Urban location	0.49 (0.5)	0.33 (0.47)	0.53 (0.5)
Monthly Per Capita Expenditure (in Rs.)	1054.4 (1084.9)	755.1 (887.1)	1119.3 (1112.8)
Business or professional group membership	0.08 (0.28)	0.06 (0.23)	0.09 (0.29)
Credit or savings group membership	0.07 (0.26)	0.07 (0.26)	0.07 (0.26)
Caste association membership	0.14 (0.35)	0.13 (0.33)	0.15 (0.35)
Development group/NGO membership	0.02 (0.14)	0.01 (0.1)	0.02 (0.15)
Co-operative membership	0.03 (0.18)	0.02 (0.15)	0.04 (0.19)
Village Panchayat or Ward Committee	0.11 (0.31)	0.13 (0.33)	0.11 (0.31)
Business Characteristics			
Number of family workers	1.39 (0.7)	1.48 (0.8)	1.37 (0.67)
Number of hours	2585.73 (1614.59)	2065.16 (1480.38)	2698.74 (1620.45)
Workplace: home-based	0.25 (0.43)	0.26 (0.44)	0.25 (0.43)
Workplace: other fixed	0.52 (0.5)	0.4 (0.49)	0.55 (0.5)
Workplace: moving	0.23 (0.42)	0.35 (0.48)	0.2 (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*** (0.06)			-0.10*** (0.04)		
Age	0.02** (0.01)	0.04** (0.02)	0.02* (0.01)	0.02*** (0.01)	0.04** (0.02)	0.02** (0.01)
Age squared/100	-0.02* (0.01)	-0.04** (0.02)	-0.02* (0.01)	-0.02*** (0.01)	-0.04** (0.02)	-0.02** (0.01)
Married	0.14** (0.06)	-0.08 (0.11)	0.19*** (0.06)	0.06 (0.05)	-0.05 (0.13)	0.09* (0.05)
Years of education	0.05*** (0.00)	0.06*** (0.01)	0.05*** (0.00)	0.01*** (0.00)	0.01 (0.01)	0.01*** (0.00)
Urban location	0.72*** (0.05)	0.77*** (0.09)	0.71*** (0.05)	0.25*** (0.04)	0.26*** (0.08)	0.25*** (0.04)
Asset ownership				0.15*** (0.01)	0.14*** (0.03)	0.15*** (0.01)
Business or professional group membership				0.16** (0.06)	0.19 (0.13)	0.16** (0.07)
Credit or savings group membership				-0.12** (0.05)	-0.12 (0.10)	-0.13** (0.06)
Caste association membership				-0.05 (0.06)	-0.07 (0.09)	-0.05 (0.07)
Development group/NGO membership				0.07 (0.08)	0.53* (0.29)	0.06 (0.08)
Co-operative membership				-0.10 (0.10)	-0.24 (0.24)	-0.09 (0.10)
Village panchayat or ward committee				-0.05 (0.06)	-0.04 (0.07)	-0.05 (0.08)
Log(number of hours)				0.55*** (0.03)	0.59*** (0.05)	0.53*** (0.03)
Number of workers				-0.04* (0.03)	-0.08 (0.05)	-0.03 (0.03)
Workplace-other fixed				0.27*** (0.05)	0.26*** (0.08)	0.27*** (0.06)
Workplace-moving				0.16*** (0.05)	0.09 (0.09)	0.18*** (0.06)
Constant	9.50*** (0.27)	9.97*** (0.34)	9.41*** (0.30)	5.28*** (0.28)	5.12*** (0.50)	5.51*** (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*** (0.08)	0.64*** (0.08)	0.67*** (0.08)	0.64*** (0.08)	0.67*** (0.08)	0.64*** (0.08)
Explained	0.32*** (0.05)	0.54*** (0.07)	0.3*** (0.04)	0.52*** (0.07)	0.39*** (0.06)	0.58*** (0.08)
Unexplained	0.35*** (0.05)	0.10*** (0.04)	0.37*** (0.06)	0.12*** (0.04)	0.28*** (0.05)	0.06 (0.04)
<u>Contribution to the explained component:</u>						
Years of education	0.13*** (0.02)	0.03*** (0.01)	0.12*** (0.02)	0.03*** (0.01)	0.15*** (0.03)	0.03 (0.02)
Urban location	0.13*** (0.02)	0.05*** (0.01)	0.13*** (0.02)	0.05*** (0.01)	0.14*** (0.03)	0.05*** (0.02)
Asset ownership		0.21*** (0.02)		0.21*** (0.02)		0.19*** (0.04)
Total number of hours		0.22*** (0.04)		0.21*** (0.04)		0.23*** (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*** (0.06)	-0.53*** (0.07)	-0.47*** (0.05)	-0.37*** (0.13)	-0.28*** (0.04)	-0.27*** (0.04)
Age	0.02** (0.01)	0.04*** (0.01)	0.04*** (0.01)	0.03*** (0.01)	0.02*** (0.01)	0.02*** (0.01)
Age squared/100	-0.02* (0.01)	-0.05*** (0.01)	-0.04*** (0.01)	-0.04*** (0.01)	-0.02*** (0.01)	-0.02** (0.01)
Married	0.14** (0.06)	0.13 (0.09)	0.15*** (0.05)	0.11 (0.16)	0.08 (0.05)	0.05 (0.05)
Years of education	0.05*** (0.00)	0.05*** (0.01)	0.05*** (0.00)	0.05*** (0.01)	0.06*** (0.00)	0.06*** (0.00)
Urban location	0.72*** (0.05)	0.94*** (0.06)	0.74*** (0.04)	0.60*** (0.05)	0.53*** (0.03)	0.49*** (0.03)
Constant	9.50*** (0.27)	7.82*** (0.22)	8.68*** (0.27)	9.50*** (0.18)	10.24*** (0.18)	10.93*** (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*** (0.04)	-0.10** (0.05)	-0.12*** (0.04)	-0.16*** (0.03)	-0.10*** (0.03)	-0.14*** (0.03)
Age	0.02*** (0.01)	0.05*** (0.01)	0.04*** (0.01)	0.02*** (0.01)	0.02*** (0.00)	0.01 (0.01)
Age squared/100	-0.02*** (0.01)	-0.06*** (0.01)	-0.05*** (0.01)	-0.02*** (0.01)	-0.02*** (0.01)	-0.01 (0.01)
Married	0.06 (0.05)	0.17*** (0.06)	0.09* (0.05)	0.10** (0.04)	0.08* (0.04)	0.09* (0.05)
Years of education	0.01*** (0.00)	0.02*** (0.00)	0.01*** (0.00)	0.01*** (0.00)	0.02*** (0.00)	0.02*** (0.00)
Asset ownership	0.15*** (0.01)	0.12*** (0.01)	0.14*** (0.01)	0.15*** (0.01)	0.17*** (0.01)	0.17*** (0.01)
Urban location	0.25*** (0.04)	0.32*** (0.04)	0.27*** (0.03)	0.23*** (0.02)	0.20*** (0.02)	0.19*** (0.03)
Business or professional group membership	0.16** (0.06)	0.12** (0.06)	0.06 (0.05)	0.13*** (0.04)	0.10** (0.04)	0.17*** (0.05)
Credit or savings group membership	-0.12** (0.05)	-0.06 (0.06)	-0.10** (0.05)	-0.16*** (0.04)	-0.17*** (0.04)	-0.19*** (0.07)
Caste association membership	-0.05 (0.06)	-0.02 (0.06)	0.00 (0.05)	0.04 (0.04)	0.06* (0.03)	0.05 (0.05)
Development group/NGO membership	0.07 (0.08)	0.27** (0.11)	0.12 (0.09)	0.01 (0.08)	0.04 (0.08)	-0.01 (0.09)
Co-operative membership	-0.10 (0.10)	-0.17 (0.11)	-0.12 (0.09)	-0.02 (0.07)	-0.04 (0.07)	0.12* (0.08)
Village panchayat or ward committee	-0.05 (0.06)	-0.06 (0.07)	0.01 (0.04)	-0.01 (0.03)	0.00 (0.04)	-0.03 (0.05)
Log(number of hours)	0.55*** (0.03)	0.69*** (0.03)	0.63*** (0.02)	0.54*** (0.02)	0.42*** (0.02)	0.32*** (0.03)
Number of workers	-0.04* (0.03)	-0.15*** (0.04)	-0.10*** (0.02)	-0.06*** (0.02)	-0.03 (0.02)	0.00 (0.02)
Workplace-other fixed	0.27*** (0.05)	0.32*** (0.05)	0.27*** (0.03)	0.24*** (0.03)	0.13*** (0.03)	0.12*** (0.04)
Workplace-moving	0.16*** (0.05)	0.28*** (0.06)	0.19*** (0.04)	0.16*** (0.04)	0.00 (0.03)	-0.01 (0.04)
Constant	5.28*** (0.28)	2.59*** (0.33)	3.84*** (0.26)	5.52*** (0.23)	7.25*** (0.20)	8.38*** (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	Non-SCSTs									
	SCSTs					Non-SCSTs				
	Q10	Q25	Q50	Q75	Q90	Q10	Q25	Q50	Q75	Q90
Age	0.09*** (0.03)	0.05** (0.03)	0.03* (0.01)	0.05*** (0.02)	0.05** (0.02)	0.04*** (0.01)	0.04*** (0.01)	0.03*** (0.01)	0.02*** (0.01)	0.02*** (0.01)
Age squared/100	-0.10*** (0.03)	-0.06* (0.03)	-0.03* (0.02)	-0.04** (0.02)	-0.05** (0.02)	-0.04*** (0.01)	-0.04*** (0.01)	-0.04*** (0.01)	-0.02** (0.01)	-0.02** (0.01)
Married	-0.14 (0.17)	-0.28* (0.16)	-0.14 (0.10)	-0.12 (0.13)	-0.27* (0.16)	0.19** (0.09)	0.19*** (0.05)	0.14*** (0.04)	0.11* (0.06)	0.08 (0.06)
Years of education	0.05*** (0.01)	0.05*** (0.01)	0.05*** (0.01)	0.06*** (0.01)	0.06*** (0.01)	0.05*** (0.01)	0.05*** (0.00)	0.05*** (0.00)	0.06*** (0.00)	0.06*** (0.00)
Urban location	0.97*** (0.14)	0.81*** (0.09)	0.62*** (0.07)	0.52*** (0.06)	0.38*** (0.09)	0.92*** (0.05)	0.72*** (0.03)	0.58*** (0.03)	0.53*** (0.03)	0.49*** (0.04)
Constant	8.22*** (0.60)	9.45*** (0.57)	10.66*** (0.37)	10.41*** (0.30)	11.01*** (0.46)	7.87*** (0.27)	8.48*** (0.30)	9.40*** (0.14)	10.10*** (0.19)	10.53*** (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** (0.02)	0.05*** (0.02)	0.02 (0.02)	0.03*** (0.01)	0.02 (0.02)	0.05*** (0.01)	0.04*** (0.01)	0.02*** (0.01)	0.02*** (0.01)	0.01 (0.01)
Age squared/100	-0.06*** (0.02)	-0.06*** (0.02)	-0.03 (0.02)	-0.04** (0.01)	-0.02 (0.02)	-0.06*** (0.01)	-0.05*** (0.01)	-0.03*** (0.01)	-0.02*** (0.01)	-0.01 (0.01)
Married	0.11 (0.14)	0.00 (0.11)	0.08 (0.08)	-0.01 (0.12)	-0.24 (0.15)	0.18** (0.07)	0.10* (0.06)	0.13*** (0.04)	0.11*** (0.04)	0.12** (0.06)
Years of education	0.02 (0.01)	0.00 (0.01)	0.00 (0.01)	0.01* (0.01)	0.00 (0.01)	0.02*** (0.01)	0.01*** (0.00)	0.01*** (0.00)	0.02*** (0.00)	0.02*** (0.00)
Asset ownership	0.13*** (0.04)	0.11*** (0.03)	0.14*** (0.02)	0.16*** (0.02)	0.20*** (0.03)	0.12*** (0.01)	0.14*** (0.01)	0.15*** (0.01)	0.17*** (0.01)	0.17*** (0.01)
Urban location	0.21** (0.10)	0.27*** (0.07)	0.16*** (0.06)	0.14** (0.06)	0.03 (0.06)	0.35*** (0.05)	0.28*** (0.03)	0.23*** (0.03)	0.21*** (0.03)	0.19*** (0.04)
Business or professional group membership	-0.13 (0.22)	0.00 (0.16)	0.26* (0.15)	0.10 (0.10)	0.04 (0.16)	0.14** (0.06)	0.05 (0.05)	0.14*** (0.05)	0.13*** (0.04)	0.18*** (0.07)
Credit or savings group membership	0.15 (0.18)	-0.01 (0.08)	-0.15 (0.09)	-0.22** (0.11)	-0.27*** (0.09)	-0.11 (0.07)	-0.12** (0.05)	-0.16*** (0.05)	-0.16*** (0.05)	-0.19*** (0.07)
Caste association membership	-0.15 (0.17)	-0.12 (0.09)	-0.20** (0.09)	-0.12 (0.10)	-0.02 (0.12)	-0.02 (0.06)	0.04 (0.06)	0.07* (0.04)	0.06 (0.04)	0.07 (0.05)
Development group/NGO membership	0.70 (0.56)	0.58* (0.32)	0.48 (0.49)	0.52 (0.54)	0.52 (0.59)	0.24* (0.13)	0.10 (0.11)	0.01 (0.07)	-0.02 (0.09)	-0.04 (0.09)
Co-operative membership	-0.30 (0.31)	-0.21 (0.14)	-0.36 (0.27)	-0.04 (0.22)	0.05 (0.45)	-0.15 (0.12)	-0.10 (0.08)	-0.03 (0.08)	-0.02 (0.06)	0.09 (0.07)
Village panchayat or ward committee	-0.02 (0.16)	0.00 (0.10)	-0.05 (0.06)	-0.14 (0.08)	-0.12 (0.13)	-0.11 (0.08)	0.01 (0.05)	0.02 (0.04)	0.05 (0.04)	-0.01 (0.05)
Log(number of hours)	0.66*** (0.06)	0.64*** (0.04)	0.61*** (0.04)	0.51*** (0.05)	0.38*** (0.05)	0.70*** (0.03)	0.60*** (0.03)	0.50*** (0.02)	0.37*** (0.02)	0.30*** (0.03)
Number of workers	-0.19* (0.10)	-0.03 (0.06)	-0.09** (0.04)	-0.01 (0.05)	-0.02 (0.04)	-0.16*** (0.04)	-0.08*** (0.02)	-0.04* (0.02)	0.00 (0.02)	0.01 (0.02)
Workplace-other fixed	0.36*** (0.14)	0.29*** (0.07)	0.36*** (0.10)	0.23*** (0.08)	0.18** (0.09)	0.31*** (0.05)	0.28*** (0.03)	0.23*** (0.03)	0.12*** (0.04)	0.14*** (0.04)
Workplace-moving	0.33** (0.13)	0.25** (0.10)	0.27*** (0.08)	0.18* (0.10)	0.08 (0.11)	0.26*** (0.06)	0.21*** (0.05)	0.16*** (0.04)	-0.00 (0.04)	-0.00 (0.06)
Constant	3.58*** (0.79)	3.91*** (0.52)	5.10*** (0.48)	6.06*** (0.51)	7.71*** (0.50)	2.56*** (0.35)	4.32*** (0.33)	5.88*** (0.23)	7.68*** (0.23)	8.73*** (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: Quantile Decompositions of Log Income (Non-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*** (0.05)	0.54*** (0.09)	0.45*** (0.1)	0.95*** (0.04)	0.83*** (0.06)	0.12 (0.1)
20	0.82*** (0.03)	0.42*** (0.07)	0.40*** (0.07)	0.75*** (0.03)	0.65*** (0.05)	0.1 (0.07)
30	0.72*** (0.02)	0.37*** (0.06)	0.35*** (0.06)	0.66*** (0.02)	0.57*** (0.04)	0.09 (0.07)
40	0.65*** (0.02)	0.33*** (0.05)	0.32*** (0.05)	0.59*** (0.02)	0.5*** (0.04)	0.09* (0.05)
50	0.6*** (0.02)	0.31*** (0.05)	0.29*** (0.04)	0.57*** (0.02)	0.48*** (0.04)	0.09* (0.05)
60	0.58*** (0.02)	0.32*** (0.05)	0.26*** (0.04)	0.55*** (0.02)	0.46*** (0.04)	0.09** (0.04)
70	0.55*** (0.02)	0.32*** (0.04)	0.23*** (0.04)	0.55*** (0.02)	0.47*** (0.04)	0.08** (0.04)
80	0.54*** (0.02)	0.31*** (0.05)	0.23*** (0.04)	0.57*** (0.02)	0.49*** (0.04)	0.08* (0.04)
90	0.55*** (0.02)	0.32*** (0.05)	0.23*** (0.05)	0.57*** (0.02)	0.52*** (0.05)	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: Quantile Decompositions 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*** (0.1)	0.25*** (0.05)	0.74*** (0.05)	0.95*** (0.1)	0.66*** (0.08)	0.29*** (0.04)
20	0.82*** (0.07)	0.27*** (0.04)	0.55*** (0.03)	0.75*** (0.07)	0.52*** (0.06)	0.23*** (0.03)
30	0.72*** (0.06)	0.29*** (0.03)	0.43*** (0.02)	0.66*** (0.07)	0.47*** (0.05)	0.19*** (0.02)
40	0.65*** (0.05)	0.3*** (0.03)	0.35*** (0.02)	0.59*** (0.05)	0.44*** (0.04)	0.15*** (0.02)
50	0.6*** (0.04)	0.3*** (0.03)	0.3*** (0.02)	0.57*** (0.05)	0.44*** (0.04)	0.13*** (0.02)
60	0.58*** (0.04)	0.3*** (0.03)	0.28*** (0.02)	0.55*** (0.04)	0.45*** (0.03)	0.1*** (0.02)
70	0.55*** (0.04)	0.31*** (0.03)	0.24*** (0.02)	0.55*** (0.04)	0.46*** (0.03)	0.09*** (0.02)
80	0.54*** (0.04)	0.3*** (0.03)	0.24*** (0.02)	0.57*** (0.04)	0.47*** (0.03)	0.1*** (0.02)
90	0.65*** (0.05)	0.3*** (0.03)	0.25*** (0.02)	0.57*** (0.05)	0.47*** (0.03)	0.1*** (0.02)

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

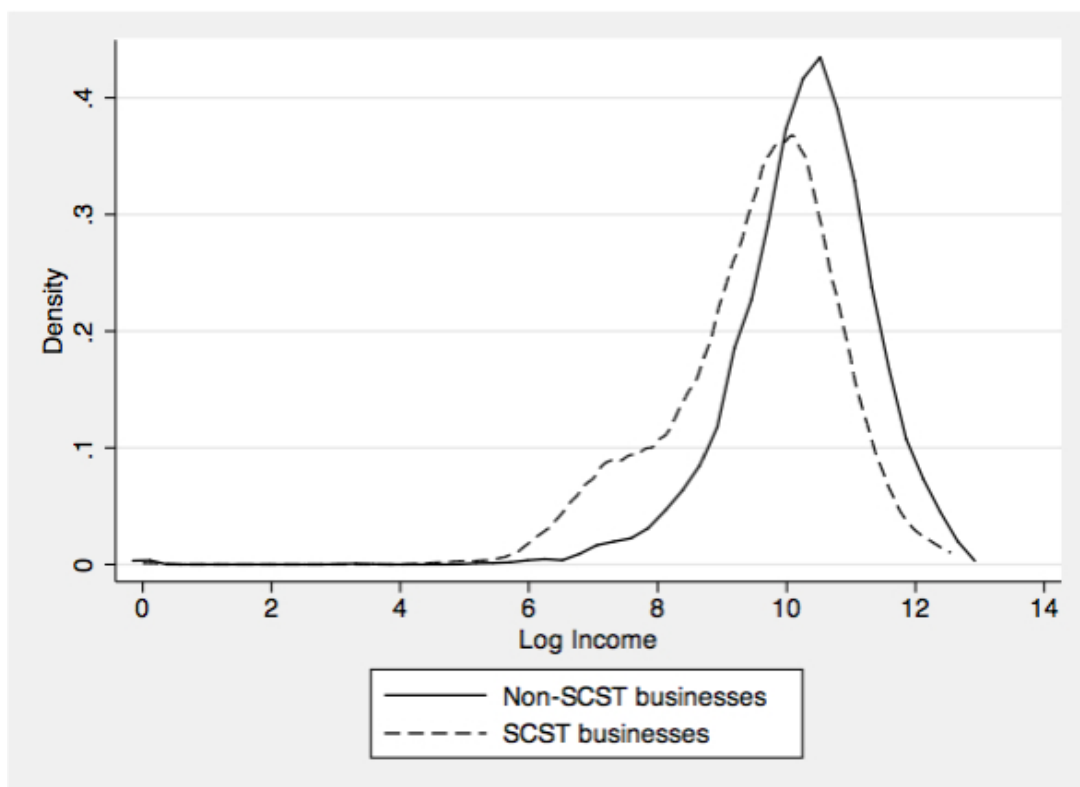


Figure 1: Kernel Density of Log Income

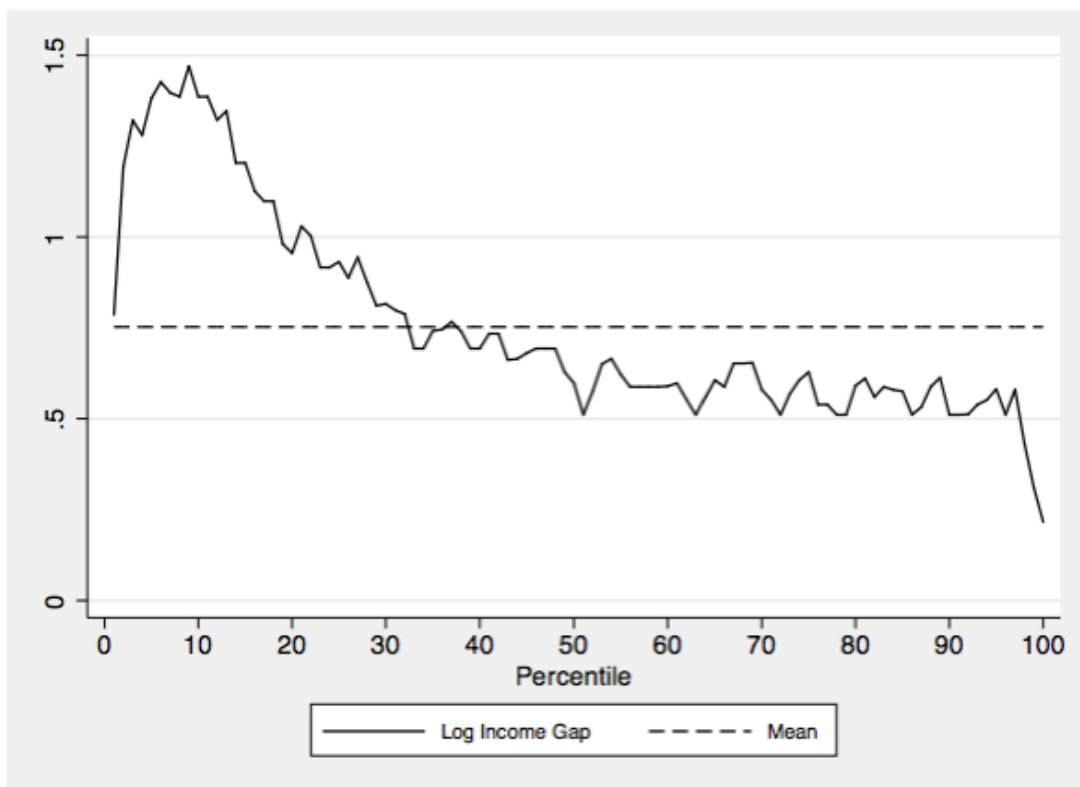


Figure 2: Caste Log Income Gap Across Percentiles & Average Gap

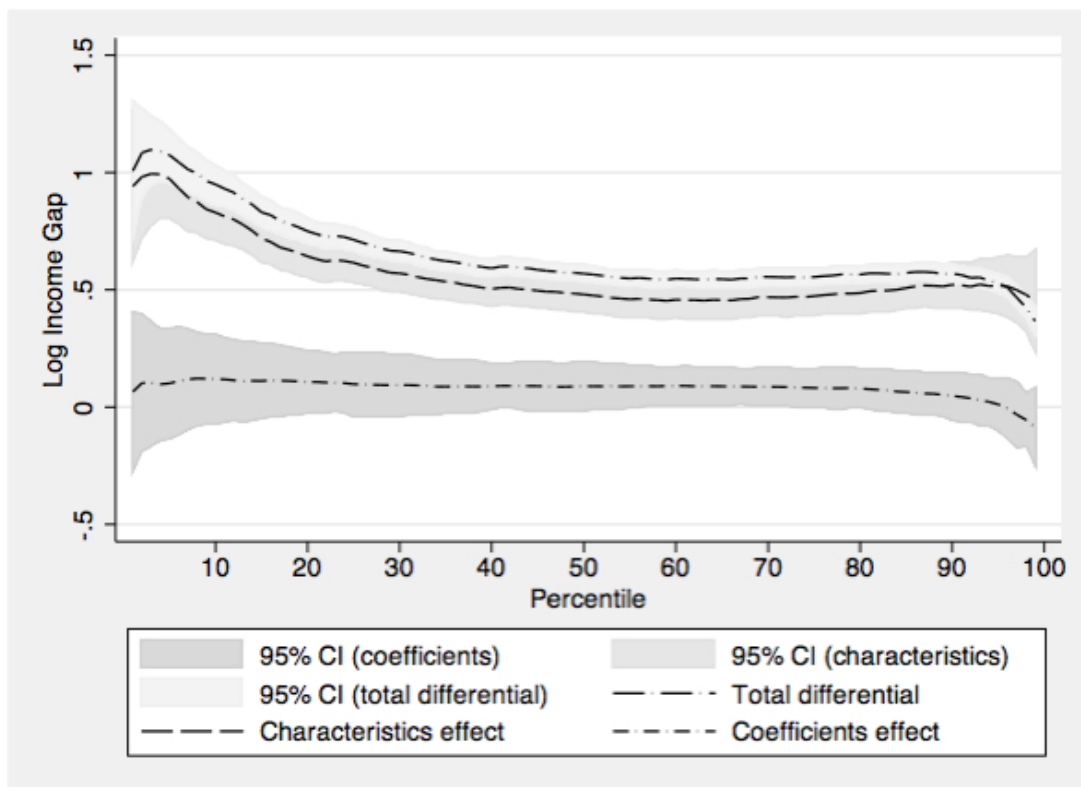


Figure 3: Quantile Decomposition of Log Income Gap: Non-SCST coefficients

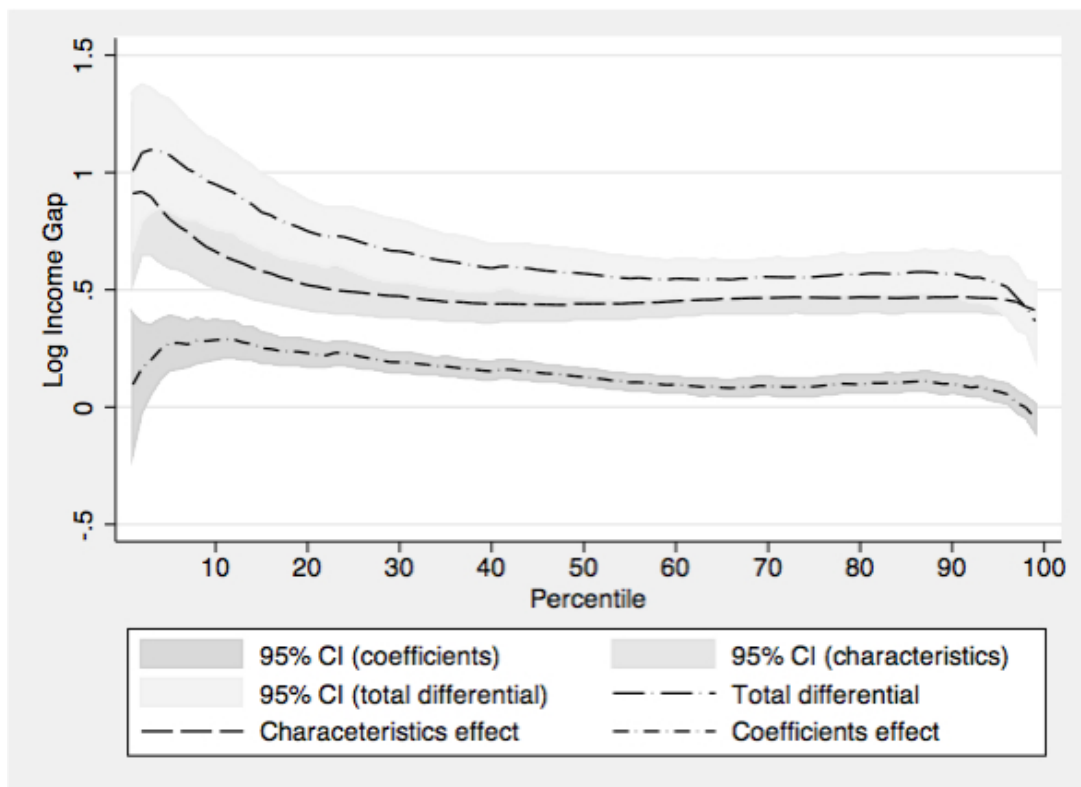


Figure 4: Quantile Decomposition of Log Income Gap: SCST coefficients

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.