Entrepreneurship and Welfare*

Jagannadha Pawan Tamvada†

Abstract

We examine returns to entrepreneurship using a standard measure of welfare-per-capita consumption expenditure. We find that entrepreneurs who employ others have the highest returns in terms of consumption, while those entrepreneurs who work for themselves, that is, self-employed individuals, have slightly lower returns than the salaried employees. Estimation of extended selection models (Bourguignon et al., 2007) suggests a process of endogenous non-random selection into occupation. In particular, it is found that the ablest individuals select into entrepreneurship (employers) followed by salaried employment, self-employment and casual labor, in that order.

JEL Codes: J24, J43, J44, L26.

Keywords: Entrepreneurship, Self-employment, Welfare, Developing Countries, Quantile Regressions, Selection Models.

*Discussions with David Audretsch, Stephan Klasen, and Christina Schneider have been very useful in the preparation of this manuscript. I am grateful to participants of a faculty seminar at the University of Göttingen for their helpful comments.

†Entrepreneurship, Growth and Public Policy Group, Max Planck Institute of Economics and Faculty of Economics, University of Göttingen. Address for Correspondence: Max Planck Institute of Economics, Kahlaische Strasse 10, 07745 Jena, Germany. Telephone: +49-3641-686730, Fax: +49-3641-686710, Email: tamvada@econ.mpg.de

1 Introduction

There is a rich literature providing insights into the determinants of entrepreneurship and its economic returns.[1] According to the expected utility theory, individuals choose self-employment when they expect higher returns from doing so relative to wage-employment (Rees and Shaw, 1986). In contrast, according to the non-pecuniary benefits theory, people select into entrepreneurship even if the expected returns are lower, in search of non-pecuniary benefits such as being their own boss (Hamilton, 2000). However, entrepreneurs are not a homogenous group of individuals and the type of entrepreneurship engaged in may have a significant effect on the returns.[2]

To date, there has been little research into the nature of entrepreneurship and its economic returns in developing countries. Therefore, the purpose of this paper is to examine the welfare effects of different types of entrepreneurship in a developing country context. Using a direct measure of welfare, per-capita consumption expenditure, this study links individuals' occupational choice to their welfare.[3] In addition, the underlying process of selection into occupations and subsequent returns in terms of welfare is examined, in an effort to determine whether certain people in a developing country are compelled to opt for low-productivity self-employment or whether there is voluntary self-selection based on unobserved abilities.

We find that entrepreneurs who employ others have the highest returns in terms of consumption, while those entrepreneurs who work for themselves, that is, self-employed individuals, have slightly lower returns than salaried employees. Estimation of extended

[1] See Parker (2004) for a synthesis of this literature.

[2] While some entrepreneurs employ others, the rest are solely self-employed individuals. Thus, the factors that influence the economic returns for entrepreneurs who are employers, may have different effects for the self-employed entrepreneurs.

[3] Most studies use income measures to examine the returns of occupations (Hamilton, 2000). In this paper, we use consumption measures. Income is usually highly correlated with consumption in a developing country. Furthermore, analyzing the consumption patterns itself has the advantage that variation is not so high as in income data. However, as people with higher incomes are likely to have greater savings, analyzing the consumption patterns for welfare comparisons may make their returns appear flattened to some extent.

selection models (Bourguignon et al., 2007) suggests a process of endogenous non-random selection into occupations. In particular, it is seen that the most able individuals select into entrepreneurship (employers) followed by salaried employment, self-employment and casual labor, in that order.

The structure of the paper is as follows. The next section provides an overview of the theoretical background on occupational choice and welfare and sets out the hypotheses. The third section discusses the methodology, in particular quantile regressions and selection models after multinomial logit estimation. Data and descriptive statistics are presented in the fourth section and the fifth section contains a discussion of the empirical analysis linking occupation and welfare. The paper concludes with a summary of the main findings linking occupation and welfare.

2 Theoretical Background

Murphy et al. (1991, p.505) suggest that individuals move into occupations that will give the greatest returns on their talents. They show that the economy grows fastest when the most able individuals become entrepreneurs. In their words, "which activities the most talented people choose can have significant effects on the allocation of resources. When talented people become entrepreneurs, they improve the technology in the line of business they pursue, and as a result, productivity and income grow." The seminal paper of Banerjee and Neuman (1993) suggests that the process of economic development is linked to the dynamics of occupational choice. They show that the initial wealth distribution of a population encourages or discourages people from starting new enterprises and thus has an impact on equilibrium returns to occupations and the long run distribution of

wealth.[4] According to these authors, occupational choice influences the capacity to save and bequest, and leads to persisting hysteresis, even though it is endogenously determined by the wealth distribution.[5]

A key observation of many studies, including Banerjee and Neuman (1993) and, more recently, Dabla-Norris et al. (2008), is the inherent hierarchy of occupational choice according to which the most productive individuals become entrepreneurs, the next best choose self-employment, and the rest become workers or subsistence workers. Dabla-Norris et al. (2008) propose that at equilibrium, the lowest productivity individuals are workers, individuals with intermediate productivity are informal entrepreneurs, and those who are most productive are formal sector entrepreneurs. These theoretical insights have yet to be empirically validated. The possibility of self-employment being worse off in the hierarchy relative to wage workers, as is traditionally assumed to be the case in less developed countries (Ranis and Fei, 1961; Harris and Todaro, 1970), or at least equal in returns, would make this theory inapplicable to less developed countries(LDCs). The literature on LDCs traditionally identifies self-employment as a distressed residual of people rationed out of jobs in the formal sector, though more recent literature on the nature of the labor market in developing countries is not monolithic on this point. Some scholars believe that the informal sector in LDCs consists of voluntarily self-selected competitive workers as well as disadvantaged individuals (Gindling, 1991; Magnac, 1991; Cunningham and Maloney, 2001; Maloney, 2004; Fields, 2005; Günther and Launov, 2006).[6]

Occupational choice is generally modeled as a utility maximizing decision of individ-

[4]They model four feasible occupational choices. Individuals can be self-employed or entrepreneurs, work for entrepreneurs or subsist. Self-employed are equivalent to workers but are not monitored by an external agent (the entrepreneur). They provide the labor for operating their firm and own the total output. Entrepreneurs, however, hire workers and monitor them though they themselves are not monitored. Employment contracts emerge only in the presence of a initial wealth distribution that has a certain degree of inequity, otherwise everyone chooses to be self-employed.

[5]More recently, Ghatak and Jiang (2002) argue that hysteresis, long run dependence on the initial conditions, depends on the size of the threshold level of wealth needed to start enterprises relative to the productivity of the modern and the subsistence technologies.

[6] Pratap and Quintin (2006) argue that there is no evidence of market segmentation in developing country labor markets.

uals (Lucas, 1978; Kihlstrom and Laffont, 1979).[7] While many models in the economics of entrepreneurship assume that individuals become self-employed as they expect higher returns relative to wage employment (Blau, 1987; Rees and Shaw, 1986; Parker, 1996), the labor and development literature suggests that in the LDC context, people are forced into self-employment in the absence of viable economic opportunities. However, Hamilton (2000) notes that entrepreneurs may trade lower earnings for the nonpecuniary benefits of business ownership. He finds no evidence of the earnings differential being a result of selection of low ability employees into self-employment. Further, he argues that self-employment offers significant nonpecuniary benefits, such as being one's own boss for most entrepreneurs. Evans and Leighton (1989) suggest that individuals who prefer greater autonomy are more likely to be entrepreneurs. Blanchflower and Oswald (1998) show that business owners have greater job satisfaction than paid-employees. According to Boháček (2006), as successful firms grow over time, individuals may enter self-employment even if the returns are lower. He claims that business households may have a higher saving rate in order to relax the wealth constraints in financing entrepreneurial projects and to operate their firms at an optimal size.

Thus three theories of returns to self-employment choice have emerged. First, the expected utility view claims that individuals choose self-employment when they expect higher returns in self-employment relative to wage-employment. Second, the non-pecuniary benefits view argues that individuals select into entrepreneurship even when the returns are lower, for non-pecuniary benefits such as being one's boss. Finally, the traditional low-productivity view suggests that individuals are compelled into self-employment in the absence of viable economic alternatives.

We hypothesize that, given the occupational structure of individuals in an econ-

[7]There are two main methods to model the returns of occupational choice. First is to estimate a mincer type wage equation for each occupation. Second is the structural probit method that estimates the reduced form probit and determines the wages corrected for selection. The sign of mill's ratio indicates the nature of selection. The predicted earnings differential are used to re-estimate the probit equation to predict self-employment choice as a function of expected utility (Rees and Shaw, 1986).

omy, the welfare returns to entrepreneurship are heterogeneous across the distribution. Entrepreneurs are either employers or solely self-employed. While employers are entrepreneurs who employ others as well, the self-employed only work for themselves. Employers are likely to have higher returns than salaried employees and self-employed people. However, the expected relative returns to self-employment compared to returns of salaried employment are unclear. This leads us to the second hypothesis. If self-employment is characterized by high skilled individuals voluntarily selecting into this occupation, the relative returns are likely to be higher than the returns to salaried employment. In the presence of segmented labor markets or if self-employment is a choice of low-skilled people, the returns to self-employment are likely to be lower. This hypothesis is tested in the empirical section using selection models. The issue of returns to occupation taking into consideration the selectivity issue has been examined in the literature by many studies. For instance, Hamilton (2000) tests for the selectivity issue considering self-employment as a binary variable. He finds that individuals of higher abilities select into entrepreneurship (also see Rees and Shaw, 1986).[8]

We hypothesize that there are locational as well as sectoral differences in returns to the entrepreneurship choice. We control for a number of other factors that have been found to influence the per-capita consumption of the households. Nelson (1988) shows the existence of economies of scale in all adult households. Such economies of scale are found to be more important in the consumption of shelter and less so in the consumption of clothing and transportation.[9] Furthermore, a vast literature is concerned with equivalence scales in the measurements of welfare for comparisons across households. Households with the

[8]Our model extends these studies to more than two occupations. In our analysis, individuals can select into one of the four occupations described earlier.

[9]Economies of scale have a range of 0 to 1, with 1 indicating no economies of scale, and the measure of welfare considering the economies of scale is equal to per-capita income of the household in this case. We, however, use the standard measure of welfare, per-capita expenditure on consumption. One of the reasons for using the standard measure in the analysis is that although we use all nonagricultural households in the beginning, we restrict the rest of the analysis to those households where the sole economically activate member is the household head. Thus, it is plausible to assume economies of scale close to 1 in such households.

same income but different structures, in terms of the number of children and old people, are likely to have different consumption patterns. For instance, Lanjouw and Ravallion (1995, pp 1431-1432) suggest that the relationship between poverty and household size depends on the weight attached to child and adult welfare. [10] Hence we control for the household demographic structure in the analysis.In the Indian context, Dreze and Srinivasan (1997) find that the poverty head-count ratio is very robust to alternate equivalent scales. We also test the robustness of the results using adult equivalent scales.[11]

There are compelling reasons to hypothesize that female headed households are likely to be poorer. Dreze and Srinivasan (1997), using an earlier survey of India's National Sample Survey Organization(NSSO), also find that households that are female headed are more likely to be poor. Jenkins (2000) finds that changes in labor earnings from persons other than the household head, changes in non-labour income, changes in the earnings of the household head, and household composition are important determinants of the poverty dynamics. For these reasons, although we first analyze all nonagricultural households, we subsequently restrict the analysis to households that have only the household head economically active. Miles (1997) finds that uncertainty, education, and location matter. Using both durable and non-durable goods in the welfare measure, Glewwe (1991) finds high returns to education in urban areas compared to rural areas in Côte d'Ivory.[12] We also examine the returns to occupations in urban and rural areas separately.

[10]They find evidence against the conventional view that household size is negatively correlated to welfare when Rothbarth method based on non-food spending is used as a measure of welfare while a measure based on child stunting indicates that larger households tend to be poor. Browning (1992) notes though children may be endogenous to whatever we are interested in modeling, this can be circumvented by assuming that fertility is exogenous. See Browning and Crossley (2001) for recent developments in the life cycle model of consumption. More recent way of measuring poverty using perceptions of consumption adequacy are addressed in Pradhan and Ravallion (2000).

[11]The results are not reported in the paper but are available on request from the author.

[12]Benito (2006) finds that unemployment risks leads households to defer consumption using British Household Panel. The dataset we have, however, does not allow for such controls. We control for all these factors, other than uncertainty.

3 Methodology

We use two empirical methods to test the hypotheses of heterogenous returns of occupation across the welfare distribution and potential non-random endogenous selection into occupations.

3.1 Quantile Regressions

For testing the former, we employ quantile regressions (see Koenker and Hallock, 2001, and references therein). As Hamilton (2000) observes, superstar model of Rosen (1981) suggests that comparison of mean earnings of workers in self-employed sector and in wage sector would be highly influenced by few entrepreneurial superstars. Thus, mean earnings do not really characterize the returns of the majority of self-employed. The greatest advantage of using quantile regressions is their ability to show snapshots of relationships across different quantiles of the distribution and not only at the mean. This enables a comparison, for example, between the poorest selfemployed individual with the poorest salaried employee at the lowest quantile and the richest selfemployed individual with the richest salaried employee at the highest quantile.

3.2 Selection Models for Multiple Outcomes

In order to test for the selectivity issue, we employ the methods of selection bias correction based on the multinomial logit (see Bourguignon et al., 2007, for a survey). Lee (1983) and Dubin and McFadden (1984) suggested ways to extend the pioneering work of Heckman (1979) for the case of multinomial logit into a selection model. In what follows in this section, we summarize the method (refereed to as BFG) and describe the basic idea behind modeling the selection process, after multinomial logit estimation.[13]

Consider the following model:

[13]This section is based on Bourguignon et al. (2007).

$$y_1 = x\beta_1 + u_1, \tag{1}$$

$$y_j^* = z\gamma_j + \eta_j, \text{ for } j = 1, \dots M \tag{2}$$

where the disturbance u_i has mean 0 and variance σ^2, conditional on x and z. j is a categorical variable that describes the choice of an economic agent among M alternatives based on 'utilities' y_j^*. z determines the alternatives and x determines the outcome variable. The outcome variable y_1 is observed only if the first category is chosen and this happens if $y_1^* > max(y_j^*)$ for all $j \neq 1$. By defining $\varepsilon_1 = max(y_j^* - y_1^*) = max(z\gamma_j + \eta_j - z\gamma_1 - \eta_1)$, we have $\varepsilon_1 < 0$. Assuming that the IIA hypothesis holds, that is, η_js are independent and identically Gumbel distributed, $G(\eta) = exp(-e^{-n})$ is the cumulative function and $g(\eta) = exp(-\eta - e^{-n})$ is the density function. This gives the multinomial logit model with

$$P(\varepsilon_1 < 0|z) = \frac{exp(z\gamma_1)}{\sum_j exp(z\gamma_j)} \tag{3}$$

Maximum likelihood estimation then gives consistent estimates of γ. The selection problem involves the estimation of β_1, given that u_1 may depend on the the η_js. By defining $\Gamma = \{z\gamma_1, z\gamma_2, \dots, z\gamma_M\}$ and generalizing the Heckman (1979), bias correction can be based on the conditional mean of u_1,

$$E(u_1|\varepsilon_1 < 0, \Gamma) = \int_{-\infty}^{0} \frac{u_1 f(u_1, \varepsilon_1|\Gamma)}{P(\varepsilon_1 < 0|\Gamma)} d\varepsilon_1 du_1 = \lambda(\Gamma) \tag{4}$$

where $f(u_1|\varepsilon_1 < 0|\Gamma)$ is the conditional joint density.

If P_k be the probability of preferring an alternative k, then P_k is given by:

$$P_k = \frac{exp(z\gamma_k)}{\sum exp(z\gamma_j)} \tag{5}$$

As the M components of Γ and the M probabilities are invertible, a unique function exists such that $E(u_1|\varepsilon < 0, \Gamma) = \mu(P_1, P_2, \ldots, P_M) = \lambda(\Gamma)$. Then consistent estimation of β_1 can be based on

$$y_1 = x_1\beta_1 + \mu(P_1, P_2, \ldots, P_M) + w_1 \tag{6}$$

Bourguignon et al. (2007) survey the methods developed so far, including Lee (1983), Dubin and McFadden (1984) and Dahl (2002). They propose a variant of the Dubin and McFadden (1984) model by removing the restriction that all the correlation coefficients sum-up to zero. Using monte-carlo simulations they find that the Dubin and McFadden (1984) method and the variants of this method proposed by them have higher efficiency. Following them, we use the BFG variant of the Dubin and McFadden (1984) for the empirical analysis. Bourguignon et al. (2007) find that having the restriction in Dubin and McFadden (1984) is a source of bias when it is incorrectly specified. They recommend that as many correction parameters enter the outcome equation as the number of alternatives in the selection equation. Furthermore, they claim that system is still identified as the all the correction terms are non-linear in probabilities. Their study positively tests for the little efficiency loss when all the non-linear probabilities are introduced.

In the BFG method, the selection terms are given by,

$$E(\eta_1^*|y_1* > max(y_s^*), \Gamma) = m(P_1), \text{ for } s \neq 1 \tag{7}$$

$$E(\eta_j^*|y_j* > max(y_s^*), \Gamma) = m(P_1)\frac{P_j}{P_j - 1}, \text{ for all } j > 1 \tag{8}$$

and the outcome conditional on $j = 1$ is,

$$y_1 = x_1\beta_1 + \sigma\left[r_1^* m(P_1) + \sum_{j=2\ldots M} r_j^* m(P_j)\frac{P_j}{P_j - 1}\right] + w_1 \tag{9}$$

Thus, the selection terms entering the regression may be treated as corrections for the underlying process of alternatives being chosen based on latent utilities. A positive coefficient of a selection term $m(P_j)$ in the equation estimating β_1 suggests that an upward bias is caused by the alternative j being non-randomly chosen. In our case, this suggests that people of lesser abilities have selected alternative j and this is resulting in an upward bias in the estimation of β_1 and is corrected by introducing the selection term $m(P_j)$. A negative coefficient of the selection term $m(P_j)$ in the equation estimating β_1 similarly suggests that there is a downward bias caused by people with higher abilities not choosing alternative 1 but choosing the alternative j. If the coefficient of $m(P_1)$ is positive, this suggests that there is an upward bias in the estimation of β_1 caused by people with higher abilities choosing alternative 1.

4 Data

The data used for the analysis comes from the 60th round employment-unemployment survey of the National Sample Survey Organization (NSSO) of India. We only consider those households where the household heads have reported to be self-employed (includes own account workers and employers), salaried employees, casual laborers, and unemployed. We restrict the sample to those who are older than 15 years but younger than 70 years. We then consider only those households who work in nonagriculture. The final sample consists of 26,485 households. In these households, 13,782 households have only the household head economically active.

Figure 1 shows that kernel density plots of log per-capita consumption of households with heads working as self-employed, salaried, employers and laborers. While the distribution plots of salaried employees and employers are to the right of the self-employed, the density of the laborers is centered to their left. Furthermore, the plots show that the inequality observed in the employer group is substantially higher than others.

5 Empirical Analysis

As mentioned in the methodology section, the hypotheses are tested through two econometric frameworks. First, heterogeneity in returns to occupations across the distribution is examined using simultaneous quantile regressions. Second, the process of endogenous non-random selection of individuals into different occupations is tested using selection models after discrete choice models with multiple outcomes.

5.1 Entrepreneurship and Welfare

5.1.1 Household Level Analysis

As Browning and Lusardi (1996, p. 1801) note, "although consumption changes are uncorrelated with anticipated income changes, the actual path of consumption may follow quite closely the actual path of income if the latter displays some persistence." Hence, the consumption and income paths are assumed to be correlated. The empirical strategy is to estimate simultaneous quantile regressions, using the log of per-capita consumption of the household as dependent variable.[14]

The occupations of the members of the household enter the regression as independent variables. A series of controls that are found to influence the consumption of the household by earlier studies are introduced in the estimation. In particular, personal characteristics of the household head, demographics of the household including the proportion of children, adults and old persons, educational background of the members, urban location

[14]Wodon (2000) also uses per-capita consumption. Many alternate strategies to construct welfare measures that are comparable across households exist. For instance, Lazear and Michael (1980) develop a technique that converts families of different structures into single person equivalents. Also see Muellbauer (1974) and Deaton and Muellbauer (1980, 1986) for a theory of equivalence scales. The identification of correct equivalent scales is still an unresolved issue (Deaton and Paxson, 1995).

and land possessed are introduced as control variables.[15] State level dummies are also included to control for regional effects.

The results presented in Table 1 suggest that the entrepreneurship has a distinct relationship with welfare.[16] As mentioned earlier, economically active people have one of the five primary occupations. They are either employers, self-employed, salaried employees, casual laborers or unemployed. In this estimation, the left out category for the occupation variables is the proportion of economically active individuals in a household who are self-employed. As the positive coefficients suggest, households that have a higher proportion of employers and those that a higher proportion of salaried employees have higher per-capita consumption levels than self-employed households. However, households that have a higher proportion of casual laborers and unemployed people have lower welfare levels than self-employed households. This suggests the existence of a welfare hierarchy, that is determined by occupational choices of members of the household.

The coefficients of controls variables are in accordance with what might be expected. Households with older household heads are more likely to have higher consumption rates and female headed households are poorer across quantiles. Female headed households are worse off most at the lowest quantile of the distribution. Households with a higher proportion of educated individuals have higher consumption rates and the returns are increasing along the quantiles as well as along higher levels of education. The quantile regression technique enables comparisons of the returns to characteristics at different quantiles of the distribution. In particular, the quantile plots in Figure 2 show that the estimates based on the quantile regression are non-linear, although for the occupational variables the estimates are mostly in 90% confidence intervals of the OLS estimates. As Figure 2 suggests, employers are increasingly better off at higher quantiles than self-

[15]Land variables proxy the wealth of the household. Wodon (2000) suggests that the land possessed by a household is also a determinant of the welfare. We also check for the robustness of the results with the land variables excluded from the analysis. Given that we have only nonagricultural households in the data set, the problem of endogeneity of the land variables is less severe.

[16]The estimates of the inter-quantile regressions are available from the author.

employed workers. Salaried employees who are in the middle of the distribution are most different than those at the extreme quantiles relative to the self-employed. At higher quantiles, casual laborers are increasingly worse off than the self-employed, and a similar phenomena is observed for the unemployed.[17] Nonlinearities with respect to high school and university education are distinct, so OLS estimates would not have given the right picture. The returns to education are comparatively much higher at higher quantiles. Figure 3 shows the estimates for the other control variables that represent the demographics and the characteristics of the household.

The proportion of children less than 15 years old in the household has a significant negative effect at the lowest two quantiles, but vanishes at higher quantiles. However, the proportion of old people in a household significantly increases the per-capita consumption expenditure. A 1% increase in the proportion of elderly people, increases the per-capita consumption by 18% at the lowest quantile and 38% at the highest quantile. The proportion of females has an insignificant effect in the lower two quantiles but has a significant positive effect at higher quantiles. Thus, at median, a 1% increase in the proportion of females, increases the per-capita consumption by 4.4% and at q(.9), by 9%. The plots of the household size variables show that the relationship between household size and welfare of the household is consistent with earlier studies that households of larger size have a lower per-capita consumption expenditure. However, the household size squared term is positive and increases across quantiles, indicating that households of larger size become worse off along the quantiles, but at decreasing rates. Thus, a convex relationship exists between household size and welfare, with households in the middle of the distribution showing the greatest negative effect of size on per-capita consumption. This could be the result of higher economies of scale at the tails of the income distribution.

[17] However, the unemployed variable slightly moves upward at the highest quantile but remains significantly negative.

5.1.2 Analysis Restricted to Household Heads

One of the main limitations of the analysis of the household level occupation data, is the simultaneous determination of occupation of the household members leading to potential endogeneity of the occupation variables. There are three possible endogeneity problems. First, occupation of members of household may not be independent of the occupation of head of the household, in the presence of intra-household dependence of occupation choice. Second, personal characteristics such as age and educational background of the household members may determine their occupational choice. Third, current income may determine future occupational choice. However, given the cross-sectional nature of our database, we are not in a position to test this third issue. Furthermore, data on wages are available for only a small fraction of individuals in the database. However, we control for wealth of the household and thus the results are conditional upon this factor. The first two issues concerning endogeneity are addressed using the following empirical strategies.

In order to reduce the potential endogenous determination of the occupational choice of the household based on the occupational choice of the household head, we re-estimate the simultaneous quantile regressions for a restricted sample of households that have only the household head as the economically active individual in Table 2. This is more likely to give the pure effect of occupation, and entrepreneurship in particular, on household welfare.[18] In order to address the second issue, we estimate models with corrections for selectivity in subsection 5.2.

We also drop the unemployed as there are only 90 heads of household who are unemployed. Furthermore as a check for robustness of the results in Table 1, we also control for the industry sector of the individuals in Table 2 as there may be sectoral differences in returns to self-employment.[19] The base category for the occupation variables is "salaried

[18] An alternate strategy would be to use instrumental variables techniques and instrument for the occupation of the household members using the occupation of the household head. However, as household heads themselves are in the sample and the occupation of their parents is not known, this is not viable.

[19] As the dataset had unemployed people earlier, industry effects could not be controlled.

employee". The estimation results are consistent with the estimations of the quantile regressions presented in Table 1. The results presented in Table 2 confirm the welfare hierarchy that the earlier regression suggested. Households headed by employers and salaried individuals have a higher per-captia consumption than households headed by self-employed individuals and casual laborers, after controlling for other factors that influence household welfare. The magnitude of the coefficient of "employer" suggests that households headed by entrepreneurs who employ others have the highest consumption levels. Although the coefficient of salaried employees is positive, it is small, and salaried employees are only slightly better off than those who are self-employed.[20] The casual laborers are last in the hierarchy.

Table 2 suggests that at lower quantiles, informal education has a significant positive effect on the per-capita consumption. The returns to primary school education increase along the quantiles. It is seen that at the lowest quantile, q(.1) primary schooling increases the per-capita consumption of the household by 14%. The coefficient however is higher at the highest quantile, q(.9), where it raises the per-capita consumption of household by 19%. A similar effect is observed for other education variables. If household head has high school education, per-capita consumption expenditure increases by 23% at the lowest quantile and 36% at the highest quantile. Similarly, if the household head has university education, the per-capita consumption of the household increases by 41% at the lowest quantile and 73% increase at the highest quantile. Thus, education has a positive effect on the per-capita consumption and increases as individuals move from the lower to higher quantiles. The returns to technical degree/diploma are also positive and increasing as individuals shift from the lower to the higher quantiles.[21] The estimates of the control variables are in accordance with the hypotheses and are consistent with the estimation in Table 1.

[20] Hamilton (2000) postulates that lower returns to self-employment may be attributed to individuals choice of freedom leading them to select self-employment.

[21] As there are very few individuals with technical degrees or diplomas, we merge these into one variable.

5.1.3 Entrepreneurship, Poverty and Inequality

Per-capita consumption of individuals is predicted after estimating the quantile regression at different quantiles.[22] The cumulative distribution plots of occupation wise predicted values are shown in Figure 4. As the plots suggest, per-capita consumption level is determined by occupation status. Entrepreneurs who are employers have the least probability of being under the poverty line.[23] Households headed by employers are followed by those headed by salaried employees, self-employed and the casual laborers, in that order, at all quantiles. The plot clarifies the status of the self-employed; they appear sandwiched between the salaried employees and the casual laborers. A direct implication of this observation is that, conditional on other characteristics, individuals in the informal sector, primarily comprising of the self-employed and the casual laborers, have lower returns to their occupations. Furthermore, the dataset is split into formal and informal sectors, with laborers and self-employed in the informal sector and salaried employees and employers in the formal sector, the plots suggest that in both sectors, entrepreneurship in the form of employers in the formal sector and self-employed in the informal sector entails higher relative consumption and an escape from poverty. The Lorenz curves in Figure 5(a) suggest that inequality is highest amongst the households with self-employed head. As the generalized Lorenz curves in Figure 5(b) suggest, the employers group has a distribution preferred by all equity respecting social welfare functions relative to the distributions of the other occupations. This is followed by the distribution of the salaried employees, self-employed people and the casual laborers.

Furthermore, we analyzed occupational choice as a determinant of poverty of households using a probit model. The poverty line was assumed to be given by half the median

[22]The log-inverse transformation of the predicted values gives the value of the normalized per-capita consumption expenditure. These transformed values are used in the poverty and inequality analysis.

[23]The plot does away with the necessity of having a poverty line to examine the poverty status of people based on their occupation and indicates the relative positions of the various occupation groups, in which we are primarily interested.

of per-capita consumption of the household.[24] The results suggest that households headed by employers, self-employed and salaried employees are less likely to be under the poverty line. Households headed by casual laborers are most likely to be under the poverty line, after controlling many characteristics that are likely to influence their poverty status.[25]

5.1.4 Rural and Urban Estimations

As we hypothesize that the returns to occupations in urban areas might be different than returns to occupations in rural areas, we estimate quantile regression welfare for rural and urban subsamples separately. It is seen in Table 3 that though the hierarchy is evident again in rural areas, the coefficients of the self-employment variable are insignificant in urban areas in Table 4. The difference between self-employed and salaried individuals disappears in urban areas at three of the five quantiles. The presence of several self-employed professionals in urban areas may be a reason for this. These results hold even after controlling for industrial sectors. Thus, it is possible that urban areas provide a more suitable environment for self-employment activities, while in rural areas, self-employment is primarily characterized by activities that inherently have lower returns. Thus, the results suggest that in the urban informal sector (UIS), there is no evidence that the returns to self-employment are lower than the returns to wage employment. It is also seen that in rural areas, the returns to education are lower than in urban areas. For instance, while university education increases per-capita consumption by 30% in rural households (at the lowest quantile), it increases the per-capita consumption by 46.5% in urban households. Furthermore, while university education increases the per-capita consumption by 49% in rural households (at the highest quantile), it increases the per-capita consumption by 81% in urban households. The returns to primary school education and high school education are also higher in urban areas. The returns to

[24]Using an alternate poverty line based on the number of adults has not significantly altered the main inferences.

[25]For brevity these results are not reported here but are available from the author.

technical education are also higher in urban areas.

In summary, these estimations suggest that self-employment in urban areas entails returns similar to salaried employment. Thus, these results support the hypothesis that self-employment in a less developed country is a blend of low-productivity and high-productivity activities.

5.2 Endogenous Non-random Occupational Selection

The main assumption underlying the quantile regression analysis is that occupations are exogenous. However occupation is rarely determined exogenously. Though we control for simultaneous determination of occupation within the household (in subsection 5.1) by selecting only those households where the household head alone is economically active, occupation itself might be endogenously determined by individual characteristics and their cultural contexts. Hence, for analyzing occupational choice, it is also appropriate to consider the selectivity issue in order to control for endogenous non-random selection into different occupations. This approach also provides insights into the selection of individuals into different occupations based on their unobserved abilities.

We estimate consumption functions separately for each of the occupational groups. A chow test rejects the null hypothesis that the coefficients of the consumption functions are same across the occupational groups. Hence, we use the method proposed by BFG to consider occupational choice as a selectivity problem in estimating the determinants of household welfare. Using this technique, the consumption equations are re-estimated for the four types of households.[26] Once again, we use the restricted sample of households

[26]Bourguignon et al. (2007) using Monte Carlo simulations show that the selection model after multinomial logit estimation provides good correction in the outcome equation even if the IIA hypothesis is violated.

in which only the household head is economically active.[27]

The multinomial logit selection equation is given in Table 9. The dependent variable is the primary occupation of the household head. The Sargan test rejects the poolability of the outcomes. The set of independent variables is same includes personal characteristics, educational background, household variables such as urban location and land possessed, religion and caste variables and regional dummies.[28] The selectivity corrected estimates for occupational groups are presented in Table 5, Table 6, Table 7 and Table 8. The following empirical approach is adopted to ensure proper identification of the models. For each occupational group, we estimate three different selection models after estimating the multinomial logit equation. In the first model, the outcome equation consists of demographics of the household and household size variables alone. Thus, all the other variables including the personal characteristics, land possessed, education, regional dummies and the religion and caste variables act as instruments for identifying the model. This is to ensure that we avoid the problem of multi-collinearity that arises as variables in the selection equation simultaneously enter the outcome equation. Hence, in the first model, there are no variables common to both the selection and outcome equations. In order to test for robustness, we estimate a second specification by introducing household characteristics such as urban location and land variables that proxy household wealth, and personal characteristics in the outcome equation. In the third specification, we introduce all the variables that form the selection equation, aside from religion and caste variables and regional dummies. Thus, religion, caste and region variables act as instruments in the third specification. We thus estimate three models for each occupational group, in order to check robustness of the selectivity coefficients.

[27] As there are only 90 households in a sample of 14000 households that have an unemployed individual as a head, we drop this category in subsequent analysis. Keeping this category creates problems in the convergence of the multinomial logit selection equation, as many of the states have no such individuals. The number of households headed by employers is also very small, but they are about 250 such households and by including only those state dummies in the regressions that have such households, we are able to obtain sufficient convergence of the multinomial logit equation.

[28] The estimation results of the selection equation are are not discussed here, for brevity.

The results given in Table 5 provide selectivity corrected estimates of the consumption function for households headed by self-employed people. In the first column, demographic characteristics of the household and household size variables are introduced. The negative coefficient of the selectivity coefficient m(Employer) in the estimated consumption function of the self-employed group suggests a downward bias caused by people selecting into the employers group. If consumption is assumed to be correlated with unobserved abilities, this suggests that a process of non-random selection of individuals with higher unobservable abilities into the employer category is causing a downward bias in the consumption function of the self-employed group. Similarly, the positive coefficient on the selectivity variable m(Casual) suggests that there is an upward bias caused by non-random selection of people with lesser unobservable abilities into the casual labor category. The positive coefficient of m(Salaried) suggests that a positive bias is caused by non-random selection of people with lower unobserved abilities into the salaried category. However, in contrast to the selection coefficient of the casual labor category, the selection coefficient for the salaried employee category is much smaller.

In the remaining analysis, the estimation results of the first model for the employers, salaried employees, and the laborers are only discussed for brevity. The selection coefficients in the consumption function of the employers in Table 6 suggest an upward bias caused by the selection of individuals with lower unobservable abilities into the self-employed and laborer categories. There is no evidence of non-random selection associated with the salaried class in this case. Similarly, the consumption function of the salaried group in Table 7 suggests that a downward bias is caused by people with higher unobserved abilities selecting into the employer category. Furthermore, an upward bias is caused due to the selection of individuals with lower unobservable abilities into the self-employed group and people with the lowest abilities selecting into the labor class. The results for laborer category, however, show that there is no bias caused by selection into

the self-employed or the salaried class but there is a negative bias caused by non-random selection of people with higher abilities into the employer category.

The positive coefficient of a selection term on a particular category in the consumption of that category indicates that people with higher unobservable abilities have selected into that group and this is causing an upward shift in the consumption function. A negative coefficient of the selection term on a particular category in the consumption function of that specific group would suggest that people with lower abilities have selected into the group and this is causing a downward bias in the consumption function. Hence, the positive term, m(self-employed) in the self-employed equation in Table 5 suggests that people with higher unobservable abilities have moved into the self-employed group and this results in an upward shift of the consumption function. Thus, the table suggests non-random selection of people with lower-abilities into the casual labor and salaried employment as well as positive self-selection into self-employment. Similarly, the positive coefficient m(salaried) in the consumption equation of the salaried group in Table 7 suggests that people with higher unobservable abilities have selected into the salaried class and the estimation has an upward bias as low-ability people have selected into casual labor and self-employment. Thus, the selection models confirm the hypothesis that self-employment is a blend of the competitive and disadvantaged sectors (Fields, 2005; Günther and Launov, 2006).

The estimates of the second and third specifications in all the four tables consistently support the inferences drawn based on the first model though there are some deviations. For instance, in the Table 8, the third specification (given in the table as Model III) shows a significant positive selection term for the salaried variable, m(Salaried) in the estimated casual labor consumption function. This suggests that people with lower abilities are selected into the salaried class leading to an upward bias. However, this result is quite counter-intuitive and could be purely due to collinearity that is caused by the presence of

many variables in the outcome equation that are also present in the selection equation.

In summary, this analysis gives insights into the selection process of individuals differing in their abilities into different occupational categories and its subsequent impact on the consumption functions of each group. The analysis shows the presence of non-random selectivity of individuals, based on their unobserved abilities, into different occupations. In particular, the selection corrected consumption functions of the self-employed and the salaried employees suggest that there are biases caused by people selecting into the employers group, followed by the salaried category, the self-employed group and the casual labor group based on their unobserved abilities, in that order. The selection corrected consumption function of the employer group suggests the presence of bias caused by the selection of individuals with the lowest abilities becoming self-employed and casual laborers and the selection corrected consumption function of the casual laborer group suggests the presence of bias as people with the higher unobservable abilities move into the employer category.

6 Conclusion

This paper makes important contributions to the literature on the economics of entrepreneurship. We extensively examine the welfare consequences of entrepreneurship in a developing country, an area of study that has received little attention to date. We use recent empirical methodologies to examine returns to entrepreneurship and test for the process of endogenous non-random selection into occupations based on unobserved abilities.

Using simultaneous quantile regressions, we find that employers, those entrepreneurs who also hire others, have the highest returns in terms of consumption, while the self-employed, those entrepreneurs who work for themselves, have slightly lower returns than the salaried employees. This evidence suggests that self-employment is not a better

occupational option relative to salaried employment, a finding that clearly contradicts a key assumption of many theoretical studies including that of Banerjee and Neuman (1993). We do find evidence that the self-employed are more likely to escape poverty, as are salaried employees and entrepreneurs who are employers. The results are robust even after controlling for industrial sectors. The results suggest that the gap between salaried employees and the self-employed is higher in rural areas than it is in urban areas. Lower returns to self-employment, however, do not necessarily support the theory that people are compelled into self-employment, as even in developed countries, the self-employed have lower returns. As one explanation for this, Hamilton (2000) argues that self-employment is associated with freedom, and hence individuals might opt for it, in spite of lower returns.

Given the potential non-random selection of individuals into different occupational categories, we also use selection models after discrete choice models with multiple outcomes, to examine the selection process and its effect on the consumption patterns in the occupation subgroups. We find evidence of endogenous non-random selection into occupation and obtain selection corrected estimates to returns to occupations. Specifically, we find that the ablest of individuals select into entrepreneurship and become employers, with the next best choosing salaried employment, self-employment and casual labor, in that order. Though this is consistent with the quantile regressions performed in the first part of the analysis, it is observed that the these regressions overestimate returns to individual characteristics, if the selectivity issue is not considered. While the quantile regression considers all individuals together and examines returns to characteristics at different quantiles, the selection model estimates separate regression curves for each of the occupational groups, at the mean of the independent variables. Extending the selectivity correction issue into the quantile regression framework is an interesting avenue for future research.

References

Banerjee, Abhijit V. and Andrew F. Neuman, "Occupational Choice and the Process of Economic Development," *Journal of Political Economy*, 1993, *101*, 274–298.

Benito, Andrew, "Does Job Insecurity Affect Household Consumption?," *Oxford Economic Papers*, January 2006, *58* (1), 157–181.

Blanchflower, David G. and Andrew Oswald, "What Makes an Entrepreneur?," *Journal of Labor Economics*, 1998, *16* (1), 26–60.

Blau, David M., "A Time Series Analysis of Self-Employment in United States," *Journal of Political Economy*, 1987, *95* (3), 445–467.

Boháček, Radim, "Financial Constraints and Entrepreneurial Investment," *Journal of Monetary Economics*, 2006, *53* (8), 2195–2212.

Bourguignon, Francois, Martin Fournier, and Marc Gurgand, "Selection Bias Corrections Based on the Multinomial Logit Model: Monte Carlo Comparisons," *Journal of Economic Surveys*, 2007, *21* (1), 174–205.

Browning, Martin, "Children and Household Economic Behavior," *Journal of Economic Literature*, September 1992, *30* (3), 1434–75.

_ **and Annamaria Lusardi**, "Household Saving: Micro Theories and Micro Facts," *Journal of Economic Literature*, December 1996, *34* (4), 1797–1855.

_ **and Thomas F. Crossley**, "The Life-Cycle Model of Consumption and Saving," *Journal of Economic Perspectives*, Summer 2001, *15* (3), 3–22.

Cunningham, Wendy V. and William F. Maloney, "Heterogeneity in the Mexican Micro-enterprise Sector," *Economic Development and Cultural Change*, 2001, *50*, 131–156.

Dabla-Norris, Era, Mark Gradstein, and Gabriela Inchauste, "What Causes Firms to Hide Output? The Determinants of Informality," *Journal of Development Economics*, 2008, *85*, 1–27.

Dahl, Gordon B., "Mobility and the Return to Education: Testing a Roy Model with Multiple Markets," *Econometrica*, 2002, *70* (6), 2367–2420.

Deaton, Angus S. and Christina H. Paxson, "Measuring Poverty Among the Elderly," *NBER Working Paper*, 1995, *5296.*

_ **and John Muellbauer**, *Economics and Consumer Behavior*, Cambridge University Press, 1980.

_ **and** _ , "On Measuring Child Costs: With Applications to Poor Countries," *The Journal of Political Economy*, 1986, *94* (4), 720–744.

Dreze, Jean and P. V. Srinivasan, "Widowhood and Poverty in Rural India: Some Inferences from Household Survey Data," *Journal of Development Economics*, December 1997, *54* (2), 217–234.

Dubin, Jeffrey A. and Daniel L. McFadden, "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption," *Econometrica*, 1984, *52* (2), 345–362.

Evans, David S. and Linda S. Leighton, "Some Empirical Aspects of Entrepreneurship," *American Economic Review*, 1989, *79* (3), 519–535.

Fields, Gary S., "A Guide to Multisector Labor Market Models," Social Protection Discussion Paper Series No. 0505, World Bank 2005.

Ghatak, Maitreesh and Neville Nien-Huei Jiang, "A Simple Model of Inequality, Occupational Choice and Development," *Journal of Development Economics*, 2002, *69* (1), 205–226.

Gindling, T. H., "Labor Market Segmentation and the Determination of Wages in the Public, Private-Formal and Informal Sectors in San-Jose, Costa-Rica," *Economic Development and Cultural Change*, 1991, *39*, 585–605.

Glewwe, Paul, "Investigating the Determinants of Household Welfare in Côte d'Ivoire," *Journal of Development Economics*, April 1991, *35* (2), 307–337.

Günther, Isabel and Andrey Launov, "Competitive and Segmented Informal Labor Markets," IZA Discussion Papers 2349, Institute for the Study of Labor (IZA) September 2006.

Hamilton, Barton H., "Does Entrepreneurship Pay? An Empirical Analysis of the Returns of Self-Employment," *Journal of Political Economy*, 2000, *108* (3), 604–631.

Harris, John R. and Michael P. Todaro, "Migration, Unemployment and Development: A Two Sector Analysis," *American Economic Review*, 1970, *60* (1), 126–142.

Heckman, James J., "Sample Selection Bias as a Specification Error," *Econometrica*, 1979, *47* (1), 153–162.

Jenkins, Stephen P., "Modelling Household Income Dynamics," *Journal of Population Economics*, 2000, *13* (4), 529–567.

Kihlstrom, Richard E. and Jean Jacques Laffont, "A General Equilibrium Entrepreneurial Theory of Firm Formation Based on Risk Aversion," *Journal of Political Economy*, 1979, *87*, 719–748.

Koenker, Roger and Kevin F. Hallock, "Quantile Regression: An Introduction," *Journal of Economic Perspectives*, 2001, *15* (4), 143–156.

Lanjouw, Peter and Martin Ravallion, "Poverty and Household Size," *Economic Journal*, November 1995, *105* (433), 1415–34.

Lazear, Edward P. and Robert T. Michael, "Family Size and the Distribution of Real Per Capita Income," *American Economic Review*, March 1980, *70* (1), 91–107.

Lee, Lung-Fei, "Generalized Econometric Models with Selectivity," *Econometrica*, 1983, *51* (2), 507–512.

Lucas, Robert, "On the Size Distribution of Business Firms," *Bell Journal of Economics*, 1978, *9*, 508–523.

Magnac, Th., "Segmented or Competitive Labor Markets," *Econometrica*, 1991, *59*, 165–187.

Maloney, William F., "Informality Revisited," *World Development*, 2004, *32* (7), 1159–1178.

Miles, David, "A Household Level Study of the Determinants of Incomes and Consumption," *Economic Journal*, January 1997, *107* (440), 1–25.

Muellbauer, John, "Household Composition, Engel Curves and Welfare Comparisons Between Households," *European Economic Review*, 1974, *5*, 103–122.

Murphy, Kevin M., Andrei Shleifer, and Robert W. Vishny, "The Allocation of Talent: Implications for Growth," *Quarterly Journal of Economics*, 1991, *106* (2), 503–530.

Nelson, Julie A., "Household Economies of Scale in Consumption: Theory and Evidence," *Econometrica*, November 1988, *56* (6), 1301–14.

Parker, Simon C., "A Time Series Model of Self-employment Under Uncertainty," *Economica*, 1996, *63*, 469–475.

_ , *The Economics of Self-Employment and Entrepreneurship*, Cambridge: Cambridge University Press, 2004.

Pradhan, Menno and Martin Ravallion, "Measuring Poverty Using Qualitative Perceptions Of Consumption Adequacy," *The Review of Economics and Statistics*, August 2000, *82* (3), 462–471.

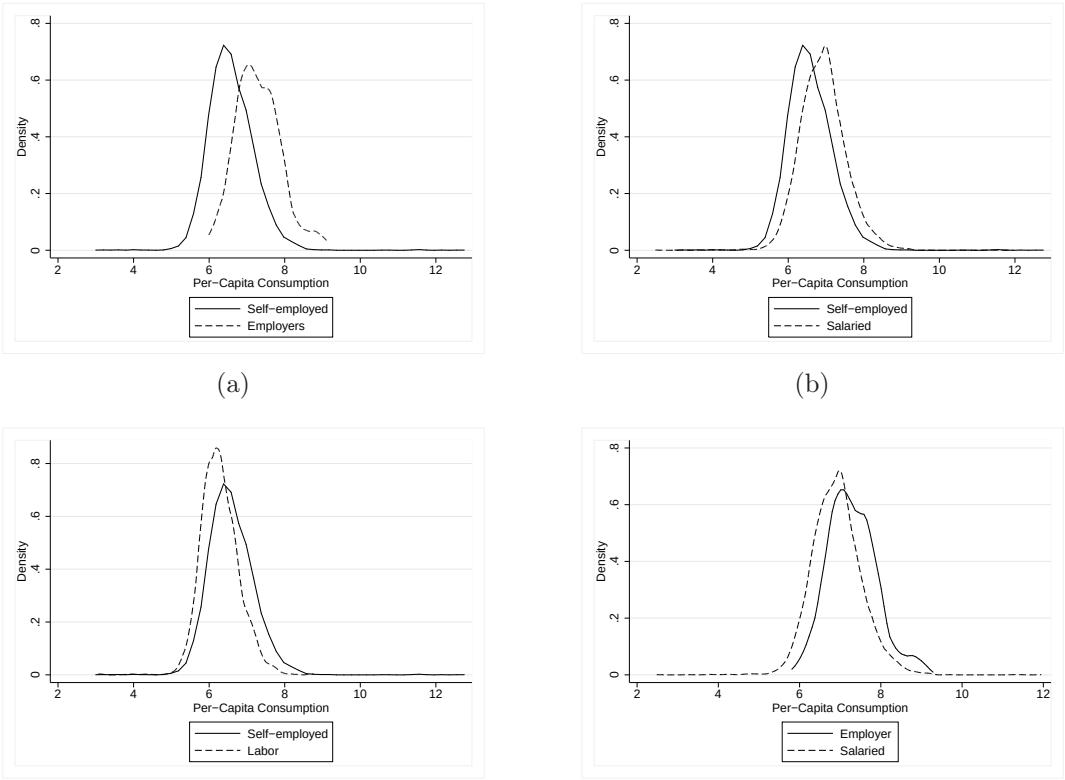
Pratap, Sangeeta and Erwaan Quintin, "Are Labor Markets Segmented in Developing Countries? A Semiparametric Approach," *European Economic Review*, 2006, *50*, 1817–1841.

Ranis, Gustav and John Fei, "A Theory of Economic Development," *American Economic Review*, 1961, *51*, 533–565.

Rees, H. and Anup Shaw, "An Empirical Analysis of Self-Employment in UK," *Journal of Applied Econometrics*, 1986, *1*, 95–108.

Rosen, Sherwin, "The Economics of Superstars," *The American Economic Review*, 1981, *71* (5), 845–858.

Wodon, Quentin T., "Microdeterminants of Consumption, Poverty, Growth, and Inequality in Bangladesh," *Applied Economics*, 2000, *32* (10), 1337–1352.



(a) (b)

(c) (d)

Figure 1: Consumption and Occupation(Un-normalised)

Table 1: Households, Occupation and Consumption

Estimates of Simultaneous Quantile Regression

Independent Var.	q10	q25	q50	q75	q90
Occupation					
Prop. Employers	0.336***	0.342***	0.405***	0.454***	0.461***
	(0.038)	(0.047)	(0.039)	(0.035)	(0.045)
Prop. Salaired	0.0816***	0.0945***	0.0996***	0.0841***	0.0778***
	(0.011)	(0.0081)	(0.0077)	(0.0069)	(0.013)
Prop. Laborers	-0.148***	-0.143***	-0.158***	-0.172***	-0.184***
	(0.012)	(0.011)	(0.010)	(0.012)	(0.016)
Prop. Unemployed	-0.192***	-0.187***	-0.208***	-0.242***	-0.182***
	(0.032)	(0.017)	(0.027)	(0.020)	(0.043)
Head's Characteristics					
Age	0.0164***	0.0162***	0.0184***	0.0204***	0.0163***
	(0.0038)	(0.0019)	(0.0016)	(0.0026)	(0.0050)
Age Square	-0.0163***	-0.0156***	-0.0174***	-0.0193***	-0.0146**
	(0.0042)	(0.0022)	(0.0018)	(0.0032)	(0.0057)
Female	-0.0912***	-0.0896***	-0.0738***	-0.0801***	-0.0573**
	(0.025)	(0.025)	(0.014)	(0.021)	(0.025)
Married	0.0516*	0.0459***	0.0495***	0.0261	0.00218
	(0.028)	(0.017)	(0.016)	(0.025)	(0.031)
Divorce/Widow	-0.0382	-0.0242	-0.0285	-0.0162	-0.0205
	(0.042)	(0.026)	(0.025)	(0.030)	(0.044)
Education					
Prop. Informal Education	0.196***	0.200***	0.220***	0.214***	0.238***
	(0.022)	(0.012)	(0.010)	(0.017)	(0.033)
Prop. Primary School	0.343***	0.344***	0.365***	0.381***	0.422***
	(0.021)	(0.014)	(0.013)	(0.017)	(0.024)
Prop. High School	0.565***	0.602***	0.661***	0.704***	0.758***
	(0.024)	(0.017)	(0.018)	(0.019)	(0.028)
Prop. University Education	0.958***	1.072***	1.187***	1.335***	1.519***
	(0.019)	(0.020)	(0.020)	(0.032)	(0.031)
Prop. Technical Degree	0.190***	0.235***	0.253***	0.281***	0.305***
	(0.020)	(0.017)	(0.033)	(0.038)	(0.035)
Demographics					
Prop. Children (less 5 years)	-0.133***	-0.0732***	-0.0156	0.00982	0.0198
	(0.025)	(0.023)	(0.032)	(0.027)	(0.053)
Prop. Children (6-10 years)	-0.125***	-0.0638**	0.0116	0.0301	0.0981*
	(0.036)	(0.025)	(0.028)	(0.037)	(0.052)
Prop. Children (11-15 years)	-0.140***	-0.0941***	-0.0601*	-0.0500*	-0.0402
	(0.035)	(0.022)	(0.032)	(0.027)	(0.048)
Prop. Females (15-60 years)	0.000581	0.0323	0.0442**	0.0604**	0.0900**
	(0.020)	(0.021)	(0.018)	(0.025)	(0.039)
Prop. Old (above 60 years)	0.188***	0.196***	0.212***	0.336***	0.383***
	(0.067)	(0.041)	(0.060)	(0.082)	(0.11)
Household Characteristics					
Urban	0.232***	0.233***	0.258***	0.277***	0.281***
	(0.0078)	(0.0044)	(0.0065)	(0.0066)	(0.0100)
0.2< Land <0.4 Hectares	0.0415***	0.0341***	0.0288***	0.0230**	0.0327***
	(0.0086)	(0.0059)	(0.0072)	(0.0091)	(0.013)

continued on next page...

Table 1: (*continued*)

Independent Var.	q10	q25	q50	q75	q90
0.4< Land < 2 Hectares	0.0763***	0.0594***	0.0430***	0.0439***	0.0518**
	(0.015)	(0.011)	(0.013)	(0.017)	(0.021)
Land > 2 Hectares	0.127***	0.126***	0.148***	0.147***	0.173***
	(0.018)	(0.022)	(0.027)	(0.016)	(0.030)
Household Size	-0.118***	-0.140***	-0.162***	-0.184***	-0.206***
	(0.0045)	(0.0049)	(0.0048)	(0.0080)	(0.0086)
Householdsize Square	0.00447***	0.00578***	0.00686***	0.00838***	0.00985***
	(0.00029)	(0.00029)	(0.00032)	(0.00062)	(0.00064)
Region Controls					
North & East States					
Punjab	0.162***	0.109***	0.0714***	0.0571***	0.0433
	(0.013)	(0.021)	(0.015)	(0.022)	(0.037)
Delhi	0.184***	0.180***	0.135***	0.0970***	0.0604**
	(0.016)	(0.024)	(0.021)	(0.021)	(0.030)
Rajasthan	0.0802***	0.0535***	-0.00930	-0.0596***	-0.102***
	(0.019)	(0.012)	(0.015)	(0.012)	(0.036)
Uttar Pradesh	-0.0687***	-0.0729***	-0.103***	-0.130***	-0.149***
	(0.011)	(0.0096)	(0.0073)	(0.014)	(0.018)
Bihar	-0.171***	-0.197***	-0.257***	-0.281***	-0.330***
	(0.018)	(0.016)	(0.016)	(0.019)	(0.019)
Manipur	0.0381	-0.0538***	-0.126***	-0.195***	-0.265***
	(0.032)	(0.018)	(0.013)	(0.019)	(0.034)
Assam	-0.0702***	-0.0766***	-0.111***	-0.159***	-0.221***
	(0.025)	(0.019)	(0.014)	(0.012)	(0.021)
West Bengal	-0.0712***	-0.0617***	-0.106***	-0.132***	-0.160***
	(0.012)	(0.013)	(0.0079)	(0.0080)	(0.020)
Orissa	-0.310***	-0.328***	-0.324***	-0.343***	-0.352***
	(0.020)	(0.013)	(0.015)	(0.020)	(0.018)
Central & West & South States					
Chhattisgar	-0.163***	-0.202***	-0.254***	-0.231***	-0.243***
	(0.028)	(0.015)	(0.019)	(0.028)	(0.051)
Madhya Pradesh	-0.218***	-0.209***	-0.227***	-0.262***	-0.292***
	(0.023)	(0.019)	(0.012)	(0.018)	(0.028)
Gujrat	0.118***	0.124***	0.0822***	0.0212*	-0.0526***
	(0.022)	(0.017)	(0.011)	(0.013)	(0.014)
Maharastra	-0.0118	-0.0174	-0.0281**	-0.0335*	-0.0493**
	(0.015)	(0.013)	(0.012)	(0.020)	(0.022)
Karnataka	-0.0671***	-0.0749***	-0.117***	-0.130***	-0.150***
	(0.018)	(0.015)	(0.012)	(0.014)	(0.026)
Kerala	0.0381	0.0830***	0.0664***	0.0711***	0.0981***
	(0.026)	(0.019)	(0.016)	(0.018)	(0.032)
Tamil Nadu	-0.143***	-0.126***	-0.154***	-0.148***	-0.146***
	(0.014)	(0.017)	(0.012)	(0.011)	(0.020)
Constant	5.726***	5.963***	6.181***	6.443***	6.807***
	(0.069)	(0.030)	(0.038)	(0.041)	(0.094)
Observations	26485	26485	26485	26485	26485

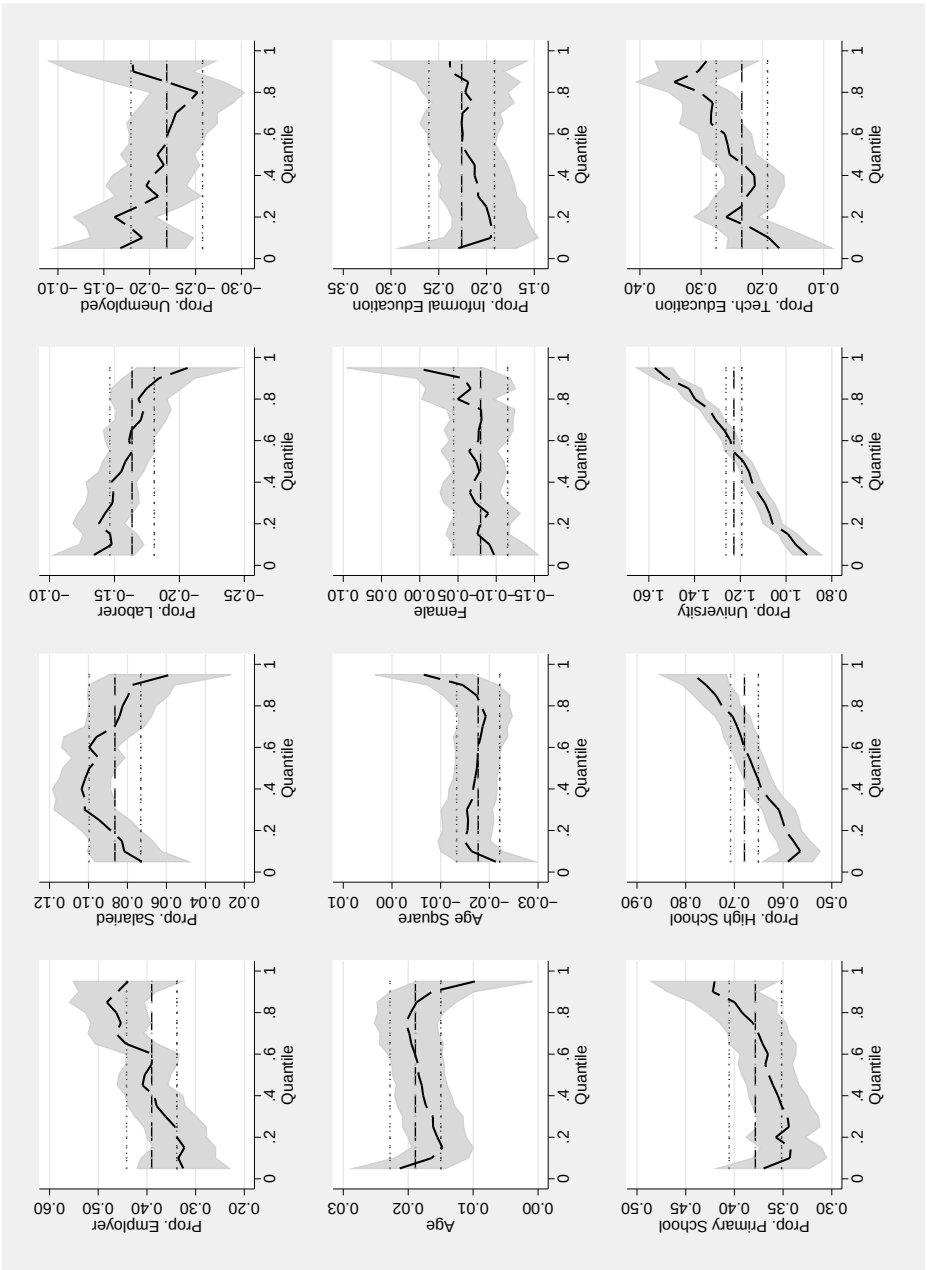


Figure 2: Quantile Plots-Households

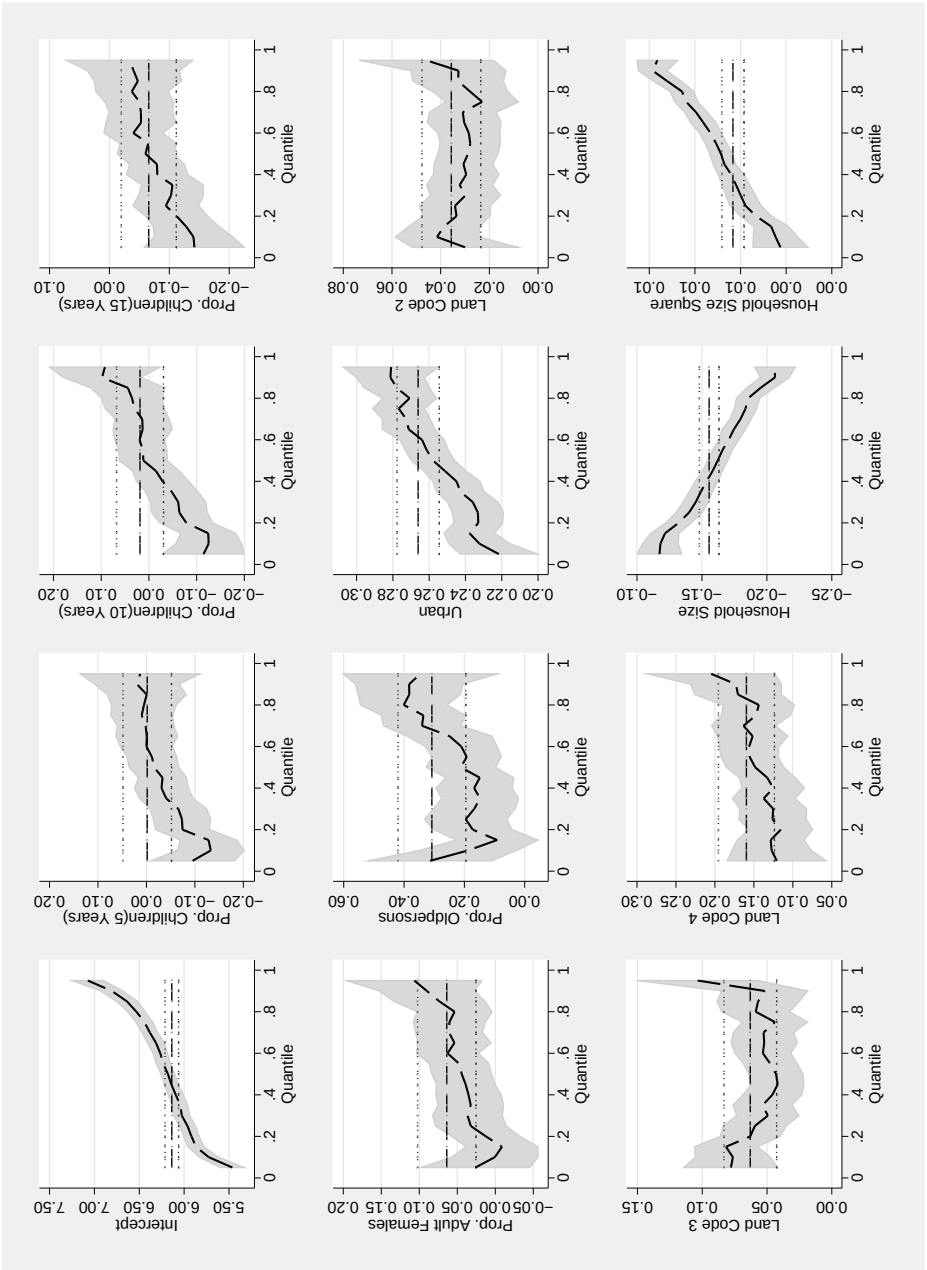


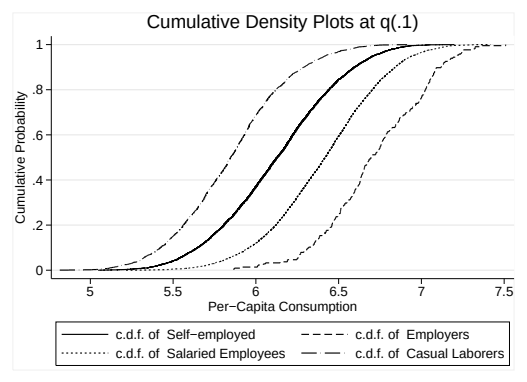
Figure 3: Quantile Plots-Households (continued)

Table 2: Household Heads, Occupation and Consumption

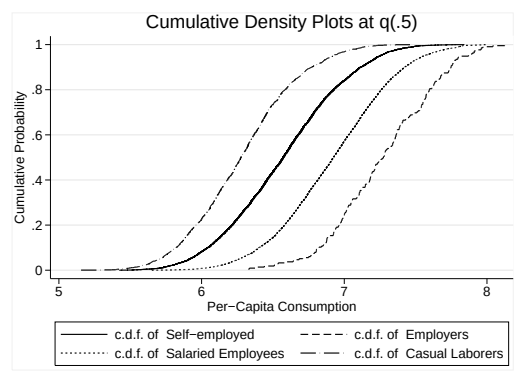
Estimates of Simultaneous Quantile Regression

Independent Var.	q10	q25	q50	q75	q90
Occupation					
Self-employed	-0.0491***	-0.0579***	-0.0631***	-0.0564***	-0.0225
	(0.013)	(0.012)	(0.012)	(0.012)	(0.019)
Employer	0.224***	0.226***	0.258***	0.252***	0.306***
	(0.058)	(0.044)	(0.037)	(0.077)	(0.069)
Laborer	-0.228***	-0.229***	-0.246***	-0.225***	-0.203***
	(0.016)	(0.017)	(0.012)	(0.019)	(0.018)
Personal Characteristics					
Age	0.0340***	0.0324***	0.0395***	0.0405***	0.0282***
	(0.0047)	(0.0039)	(0.0039)	(0.0043)	(0.0066)
Age Square	-0.0371***	-0.0329***	-0.0409***	-0.0399***	-0.0240***
	(0.0061)	(0.0050)	(0.0048)	(0.0051)	(0.0083)
Female	-0.0144	-0.0296	-0.0653	0.0125	0.0811
	(0.035)	(0.031)	(0.043)	(0.041)	(0.060)
Married	-0.0301	-0.0312	-0.0321	-0.0658***	-0.0435
	(0.037)	(0.021)	(0.029)	(0.022)	(0.053)
Divorce/Widow	-0.212***	-0.233***	-0.176***	-0.220***	-0.184**
	(0.037)	(0.034)	(0.042)	(0.034)	(0.075)
General Education					
Informal Education	0.0479*	0.0390**	0.0219	0.0339*	0.0233
	(0.027)	(0.019)	(0.025)	(0.018)	(0.024)
Primary School	0.142***	0.146***	0.137***	0.172***	0.191***
	(0.018)	(0.013)	(0.018)	(0.018)	(0.016)
High School	0.235***	0.268***	0.292***	0.341***	0.361***
	(0.017)	(0.014)	(0.016)	(0.015)	(0.017)
University Education	0.413***	0.483***	0.559***	0.640***	0.732***
	(0.025)	(0.015)	(0.019)	(0.023)	(0.022)
Technical Degree or Diploma	0.170***	0.180***	0.169***	0.191***	0.235***
	(0.021)	(0.015)	(0.016)	(0.017)	(0.024)
Demographics	YES				
Household Characteristics	YES				
Region Controls	YES				
Sector Controls	YES				
Constant	5.773***	6.081***	6.237***	6.478***	6.923***
	(0.085)	(0.071)	(0.072)	(0.068)	(0.12)
Observations	13692	13692	13692	13692	13692

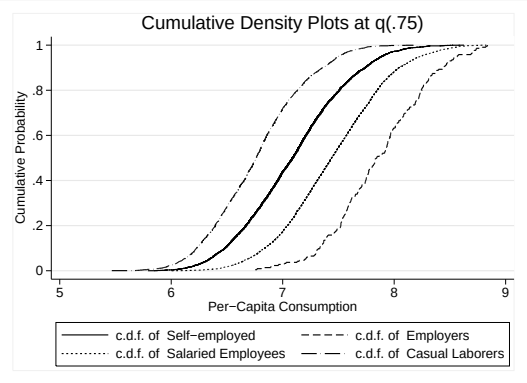
Notes: *Signifies p< 0.05; ** Signifies p<0.01;*** Signifies p<0.001. Standard errors are reported in parentheses. Dependent variable is log per-capita consumption expenditure. Base categories for occupation is salaried employee, for marital status is unmarried, for general/technical education is no general/technical education. Full set of state level regional dummies are also included in the regression with the excluded state being Andhra Pradesh.



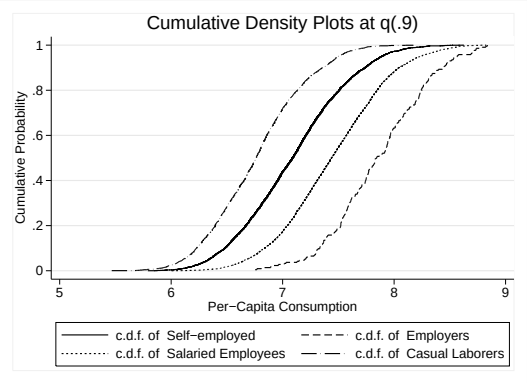
(a) Normalised Poverty Plots



(b) Normalised Poverty Plots



(c) Normalised Poverty Plots



(d) Normalised Poverty Plots

Figure 4: Occupation and Poverty Plots

(a)

(b)

Figure 5: Occupation and Inequality Plots at Median

Table 3: Occupation and Consumption in Rural Areas

Estimates of Simultaneous-Quantile Regression

	q10	q25	q50	q75	q90
Occupation					
Self-employed	-0.0737***	-0.0885***	-0.122***	-0.125***	-0.0771**
	(0.022)	(0.022)	(0.019)	(0.026)	(0.031)
Employer	0.266***	0.152*	0.165*	0.185*	0.263
	(0.088)	(0.078)	(0.087)	(0.10)	(0.23)
Laborer	-0.207***	-0.232***	-0.242***	-0.250***	-0.224***
	(0.028)	(0.024)	(0.024)	(0.027)	(0.025)
Personal Characteristics					
Age	0.0219***	0.0286***	0.0367***	0.0397***	0.0415***
	(0.0059)	(0.0034)	(0.0057)	(0.0045)	(0.0092)
Age Square	-0.0258***	-0.0315***	-0.0399***	-0.0430***	-0.0442***
	(0.0069)	(0.0035)	(0.0067)	(0.0056)	(0.011)
Female	0.0635	0.0552	0.00795	-0.00698	0.0296
	(0.044)	(0.039)	(0.040)	(0.058)	(0.075)
Married	0.0426	-0.0211	-0.0747**	-0.0877*	-0.101
	(0.064)	(0.053)	(0.036)	(0.048)	(0.092)
Divorce/Widow	-0.164**	-0.244***	-0.242***	-0.220***	-0.254**
	(0.082)	(0.070)	(0.040)	(0.042)	(0.11)
General Education					
Informal Education	0.0503	0.0319	0.0242	0.0316	0.00941
	(0.032)	(0.029)	(0.020)	(0.027)	(0.031)
Primary School	0.153***	0.159***	0.136***	0.164***	0.158***
	(0.029)	(0.016)	(0.015)	(0.020)	(0.022)
High School	0.178***	0.199***	0.208***	0.255***	0.275***
	(0.027)	(0.023)	(0.020)	(0.030)	(0.034)
University Education	0.303***	0.313***	0.384***	0.449***	0.491***
	(0.028)	(0.015)	(0.027)	(0.029)	(0.064)
Technical Degree or Diploma	0.169***	0.204***	0.179***	0.180***	0.191***
	(0.045)	(0.028)	(0.024)	(0.041)	(0.062)
Demographics	YES				
Household Characteristics	YES				
Region Controls	YES				
Constant	6.135***	6.289***	6.461***	6.671***	6.929***
	(0.11)	(0.10)	(0.093)	(0.097)	(0.15)
Observations	5202	5202	5202	5202	5202

Notes: *Signifies p< 0.05; ** Signifies p<0.01;*** Signifies p<0.001. Standard errors are reported in parentheses. Dependent variable is log per-capita consumption expenditure. Base categories for occupation is salaried employee, for marital status is unmarried, for education is no education. Full set of state level regional dummies are also included in the regression with the excluded state being Andhra Pradesh.

Table 4: Occupation and Consumption in Urban Areas

Estimates of Simultaneous-Quantile Regression

	q10	q25	q50	q75	q90
Occupation					
Self-employed	-0.0173	-0.0367**	-0.0370*	-0.0141	0.0244
	(0.025)	(0.016)	(0.020)	(0.015)	(0.023)
Employer	0.244***	0.223***	0.249***	0.269***	0.312***
	(0.055)	(0.059)	(0.045)	(0.078)	(0.060)
Laborer	-0.215***	-0.240***	-0.278***	-0.244***	-0.228***
	(0.034)	(0.025)	(0.015)	(0.024)	(0.028)
Personal Characteristics					
Age	0.0338***	0.0307***	0.0392***	0.0309***	0.0172**
	(0.0058)	(0.0054)	(0.0039)	(0.0058)	(0.0080)
Age Square	-0.0359***	-0.0297***	-0.0389***	-0.0274***	-0.0100
	(0.0076)	(0.0070)	(0.0043)	(0.0072)	(0.0098)
Female	-0.0805	-0.0726	-0.0653	0.0120	0.135
	(0.073)	(0.049)	(0.076)	(0.049)	(0.083)
Married	-0.0362	-0.0389	-0.0159	-0.0142	0.0168
	(0.042)	(0.032)	(0.033)	(0.032)	(0.047)
Divorce/Widow	-0.196***	-0.208***	-0.160**	-0.153**	-0.118
	(0.074)	(0.056)	(0.066)	(0.062)	(0.088)
General Education					
Informal Education	0.0529	0.0463**	0.0268	0.0532	0.0305
	(0.034)	(0.023)	(0.027)	(0.037)	(0.039)
Primary School	0.160***	0.159***	0.148***	0.187***	0.211***
	(0.029)	(0.025)	(0.022)	(0.022)	(0.028)
High School	0.286***	0.314***	0.349***	0.399***	0.405***
	(0.029)	(0.025)	(0.021)	(0.024)	(0.026)
University Education	0.465***	0.562***	0.633***	0.717***	0.812***
	(0.040)	(0.025)	(0.024)	(0.031)	(0.031)
Technical Degree or Diploma	0.175***	0.160***	0.152***	0.201***	0.215***
	(0.026)	(0.017)	(0.025)	(0.030)	(0.032)
Demographics	YES				
Household Characteristics	YES				
Region Controls	YES				
Constant	5.974***	6.337***	6.446***	6.826***	7.293***
	(0.13)	(0.092)	(0.073)	(0.11)	(0.15)
Observations	8490	8490	8490	8490	8490

Notes: *Signifies p< 0.05; ** Signifies p<0.01;*** Signifies p<0.001. Standard errors are reported in parentheses. Dependent variable is log per-capita consumption expenditure. Base categories for occupation is salaried employee, for marital status is unmarried, for education is no education. Full set of state level regional dummies are also included in the regression with the excluded state being Andhra Pradesh.

Table 5: Occupational Selection and Consumption (*Self-employed*)

Selection after Multinomial Logit

Independent Var.	Model I	Model II	Model III
Personal Characteristics			
Age		0.0333***	0.0394***
		(0.0063)	(0.0060)
Age Square		-0.0420***	-0.0430***
		(0.0077)	(0.0069)
Female		-0.0645	0.0890
		(0.078)	(0.077)
Married		0.151***	0.0142
		(0.052)	(0.054)
Divorce/Widow		0.0854	-0.102
		(0.078)	(0.089)
Education			
Informal Education			0.110***
			(0.036)
Primary School			0.273***
			(0.049)
High School			0.419***
			(0.083)
University Education			0.633***
			(0.11)
Technical Degree or Diploma			0.0949**
			(0.044)
Demographics	YES	YES	YES
Household Characteristics			
Urban		0.214***	0.327***
		(0.018)	(0.029)
Land Variables		YES	YES
Household Size	-0.0956***	-0.140***	-0.141***
	(0.023)	(0.020)	(0.019)
Household Size Square	0.000948	0.00449***	0.00457***
	(0.0020)	(0.0017)	(0.0015)
Selection Coefficients			
m(Self-employed)	0.655***	0.703***	0.368***
	(0.087)	(0.079)	(0.12)
m(Employer)	-2.338***	-1.714***	-1.696***
	(0.40)	(0.33)	(0.30)
m(Salaried Employee)	0.697***	0.829***	1.017***
	(0.22)	(0.20)	(0.23)
m(Labor)	1.715***	1.559***	0.586**
	(0.20)	(0.18)	(0.27)
Constant	7.286***	6.492***	6.181***
	(0.12)	(0.15)	(0.17)
Observations	5047	5047	5047

Notes: *Signifies p< 0.05; ** Signifies p<0.01;*** Signifies p<0.001. Dependent variable is log per-capita consumption expenditure. State dummies are included only in the selection equation.

Table 6: Occupational Selection and Consumption (*Employers*)

Selection after Multinomial Logit

Independent Var.	Model I	Model II	Model III
Personal Characteristics			
Age		0.0152	0.0142
		(0.036)	(0.037)
Age Square		-0.0254	-0.0148
		(0.039)	(0.040)
Female		-0.628**	-0.419
		(0.27)	(0.38)
Married		-0.0853	-0.437
		(0.35)	(0.43)
Divorce/Widow		-0.0214	-0.384
		(0.40)	(0.51)
Education			
Informal Education			-0.0861
			(0.23)
Primary School			0.157
			(0.23)
High School			0.552*
			(0.32)
University Education			0.883**
			(0.40)
Technical Degree or Diploma			0.333**
			(0.16)
Demographics	YES	YES	YES
Household Characteristics			
Urban		0.176	0.408***
		(0.13)	(0.15)
Land Variables		YES	YES
Household Size	-0.0477	-0.0865	-0.0504
	(0.12)	(0.13)	(0.13)
Household Size Square	-0.00906	-0.00448	-0.00814
	(0.011)	(0.013)	(0.014)
Selection Coefficients			
m(Self-employed)	1.917***	1.630**	-1.225
	(0.70)	(0.74)	(1.27)
m(Employer)	-0.168	-0.214*	-0.317**
	(0.11)	(0.11)	(0.13)
m(Salaried Employee)	1.448	1.203	0.332
	(0.98)	(1.12)	(1.31)
m(Labor)	1.623**	1.303	-1.078
	(0.80)	(0.86)	(1.26)
Constant	9.855***	9.380***	7.102***
	(0.82)	(1.12)	(1.37)
Observations	215	215	215

Notes: *Signifies p< 0.05; ** Signifies p<0.01;*** Signifies p<0.001. Dependent variable is log per-capita consumption expenditure. State dummies are included only in the selection equation.

Table 7: Occupational Selection and Consumption (*Salaried*)

Selection after Multinomial Logit

Independent Var.	Model I	Model II	Model III
Personal Characteristics			
Age		0.0323***	0.0382***
		(0.0066)	(0.0064)
Age Square		-0.0365***	-0.0388***
		(0.0081)	(0.0080)
Female		-0.193**	-0.0298
		(0.087)	(0.21)
Married		0.113***	-0.0332
		(0.040)	(0.040)
Divorce/Widow		0.0544	-0.146**
		(0.071)	(0.066)
Education			
Informal Education			0.0795*
			(0.045)
Primary School			0.246***
			(0.049)
High School			0.375***
			(0.070)
University Education			0.595***
			(0.086)
Technical Degree or Diploma			0.184***
			(0.029)
Demographics	YES	YES	YES
Household Characteristics			
Urban		0.174***	0.273***
		(0.018)	(0.023)
Land Variabels		YES	YES
Household Size	-0.0841***	-0.165***	-0.158***
	(0.022)	(0.018)	(0.017)
Household Size Square	-0.00282	0.00397**	0.00348**
	(0.0020)	(0.0017)	(0.0015)
Selection Coefficients			
m(Self-employed)	1.844***	1.960***	0.614***
	(0.17)	(0.15)	(0.24)
m(Employer)	-0.930***	-0.311	-0.750***
	(0.31)	(0.26)	(0.23)
m(Salaried Employee)	0.748***	0.649***	0.414***
	(0.099)	(0.090)	(0.11)
m(Labor)	2.609***	2.165***	0.763***
	(0.18)	(0.16)	(0.24)
Constant	8.115***	7.378***	6.227***
	(0.058)	(0.13)	(0.19)
Observations	6391	6391	6391

Notes: *Signifies p< 0.05; ** Signifies p<0.01;*** Signifies p<0.001. Dependent variable is log per-capita consumption expenditure. State dummies are included only in the selection equation.

Table 8: Occupational Selection and Consumption (*Laborers*)

Selection after Multinomial Logit

Independent Var.	Model I	Model II	Model III
Personal Characteristics			
Age		0.0157*	0.0210**
		(0.0090)	(0.0084)
Age Square		-0.0226*	-0.0225**
		(0.012)	(0.011)
Female		0.0532	0.167
		(0.094)	(0.11)
Married		0.0401	-0.0531
		(0.057)	(0.062)
Divorce/Widow		-0.106	-0.254**
		(0.086)	(0.11)
Education			
Informal Education			0.0928**
			(0.041)
Primary School			0.257***
			(0.053)
High School			0.329***
			(0.100)
University Education			0.451***
			(0.16)
Technical Degree or Diploma			0.127
			(0.080)
Demographics	YES	YES	YES
Household Characteristics			
Urban		0.113***	0.217***
		(0.022)	(0.034)
Land Variables		YES	YES
Household Size	-0.147***	-0.170***	-0.178***
	(0.027)	(0.032)	(0.031)
Household Size Square	0.00590**	0.00760***	0.00831***
	(0.0026)	(0.0028)	(0.0028)
Selection Coefficients			
m(Self-employed)	0.162	0.190	0.249
	(0.20)	(0.20)	(0.22)
m(Employer)	-3.378***	-3.075***	-2.065***
	(0.50)	(0.49)	(0.50)
m(Salaried Employee)	-0.116	0.153	1.021***
	(0.21)	(0.22)	(0.30)
m(Labor)	0.0897	0.133*	0.104
	(0.074)	(0.072)	(0.090)
Constant	6.962***	6.738***	6.860***
	(0.11)	(0.18)	(0.18)
Observations	2036	2036	2036

Notes: *Signifies p< 0.05; ** Signifies p<0.01;*** Signifies p<0.001. Dependent variable is log per-capita consumption expenditure. State dummies are included only in the selection equation.

Table 9: Occupational Selection

Base Multinomial Selection Equation

Independent Var.	Self-employed	Employer	Laborer
Personal Characteristics			
Age	-0.0294**	0.0536	-0.0497**
	(0.015)	(0.057)	(0.021)
Age Square	0.0225	-0.0300	-0.00269
	(0.018)	(0.064)	(0.026)
Female	-0.685***	-1.460**	-0.970***
	(0.14)	(0.68)	(0.19)
Married	0.807***	0.484	0.619***
	(0.10)	(0.43)	(0.13)
Divorce/Widow	1.042***	0.411	1.178***
	(0.17)	(0.73)	(0.22)
Education			
Informal Education	-0.345***	-0.488	-0.626***
	(0.10)	(0.58)	(0.11)
Primary School	-0.744***	-0.175	-1.541***
	(0.078)	(0.39)	(0.086)
High School	-1.301***	-0.238	-2.919***
	(0.079)	(0.39)	(0.10)
University Education	-1.772***	-0.351	-4.496***
	(0.086)	(0.40)	(0.20)
Technical Degree or Diploma	-0.535***	-0.339	-0.903***
	(0.083)	(0.24)	(0.22)
Household Characteristics			
Urban	-0.376***	0.486**	-0.874***
	(0.046)	(0.20)	(0.063)
0.2< Land <0.4 Hectares	0.342***	0.619***	0.180***
	(0.046)	(0.18)	(0.067)
0.4< Land < 2 Hectares	0.0457	0.984***	-0.0912
	(0.087)	(0.31)	(0.13)
Land > 2 Hectares	-0.127	0.813	-0.986**
	(0.17)	(0.55)	(0.38)
Hindu	-0.255***	-0.368**	0.0340
	(0.049)	(0.16)	(0.073)
Backward	-0.181***	-0.740***	0.340***
	(0.044)	(0.17)	(0.069)
Region Variables	YES	YES	YES
Constant	1.147***	-6.354***	2.104***
	(0.28)	(1.18)	(0.38)
Observations	13700	13700	13700
LR $\chi^2(81)$	4193.61		
Log likelihood	-12593.207		
Pseudo R-squared	0.1427		

Notes: Marginal effects after multinomial logit estimation. Dependent variable is primary occupation of the household head. Set of state level regional dummies are also included in the regression.