Occupational Choice in Rural Kenya: Using Subjective Expectations Data to Measure Credit and Insurance Constraints*

Anthony Keats†

February 2, 2013

DRAFT: Please do not cite or circulate without permission

Abstract

Millions of people living in rural areas of Sub-Saharan Africa supplement earnings from subsistence farming by operating microenterprises or by working as day laborers. However, little is known about how people choose which of these additional income-generating activities to do, and, in particular, to what extent market imperfections affect this choice and contribute to inefficiencies in production and welfare losses. This paper investigates the role of two market imperfections – lack of access to credit and lack of insurance – on these occupational choices among subsistence farmers in rural Kenya. Using a unique dataset of people's subjective beliefs about the distribution of returns and entry costs for the set of income-generating activities available to them, I estimate a random utility model of occupational choice and find evidence that risk, and the lack of insurance, prevents entry into high profit but high variance occupations. Calculations of the compensating variation associated with changes in the variance, or riskiness, of these occupations suggest that income losses from this market failure can be a considerable fraction of expected earnings. In contrast, the relative size of entry costs appears to have no effect on people's decisions, suggesting that access to credit does not constrain choice for the occupations available in the study area.

JEL Codes: D52, D84, D92, O16

*I am grateful to Pascaline Dupas, Robert Jensen, Adriana Lleras-Muney, and Jonathan Robinson for helpful comments on this draft. I thank Sarah Green, Kathy Nolan, and Kim Siegal for research assistance in the field. All errors are my own.

†Wesleyan University Department of Economics, email: akeats@wesleyan.edu

1 Introduction

Most people in Sub-Saharan Africa live in rural areas, where they typically own a small plot of land and do some subsistence farming. At the same time, many people also engage in other money-generating activities to supplement farm income, such as operating a small business or working as a day laborer. These additional occupations provide an important boost to rural livelihoods: they have been estimated to account for 30-50% of household income (Chuta and Liedholm, 1990), and they can be an *ex-ante* hedge against farm risk as well as an *ex-post* source of income diversification following a farming shock (Reardon, 1997). However, not all occupations are alike: some have larger entry costs, some are riskier than others, and profits can vary from job to job. How do people choose which additional occupation (if any) to enter? Do market failures – in particular, the lack of access to credit or to insurance – constrain this choice? This may be the case given the scarcity of formal financial services in rural areas and the limitations of more informal mechanisms. If so, how much potential income is lost because people cannot acquire investment capital needed to start a high profit business, or because without adequate insurance they are forced to choose less risky occupations that are also less profitable?

In this paper, I estimate the effects of both credit and insurance constraints on non-farm occupational choices for a sample of people living in rural Western Kenya. The analysis combines data on individuals' subjective expectations about the entry costs and the distribution of profits for a large set of occupations commonly available in the study area with data on their occupational choices three months later, when approximately 40% of the sample either started or stopped at least one occupation besides farming. There are several advantages to using subjective beliefs: 1) data on individual returns and costs for each occupation are difficult and costly to measure directly, 2) people are more likely to base their decisions on their own beliefs rather than aggregate data, and 3) the subjective measures help account for differences in ability and human capital that might otherwise be missing from data on the average distribution of returns for each occupation. The set of non-farm occupations includes three types of physical labor (both farm and non-farm) and more than a dozen types of micro-enterprises, including market vending, artisan trades, and small retail shops. Assuming an additive random utility model of occupational choice, I estimate the relationship between beliefs about entry costs, expected profits, and risk (the expected variance of profits) and the probability of starting or keeping one of these occupations.

There are two main findings. First, the evidence presented suggests that risk, and the lack of insurance, prevents entry into high profit but high variance occupations, such as owning a small retail shop. People in the study sample perceive large trade-offs between occupations with higher returns and risk. On average, the correlation coefficient between expected average returns and the expected variance of returns is 0.77. Given this, and likely because they lack adequate means to mitigate this risk, people in the sample (whether men or women) tend to avoid occupations with higher expected variances. Indeed, the estimation results of the random utility choice model reveal that while expected profits are positively and significantly associated with the probability of choosing an occupation, the expected variance of returns is also negatively and significantly associated with this choice. This suggests that people may forgo potential earnings in order to avoid risk. Calculations of the compensating variation associated with changes in the expected variance of the non-farm occupations show that people would be willing to give up, for example, 11% of their expected earnings for a 25% reduction in risk.

On the other hand, access to capital does not appear to be a first order determinant of occupational choice. In the study sample, expected returns and entry costs are also positively correlated, but the relationship is much weaker (the average correlation coefficient in the sample is 0.37). In line with this, while I find that the probability of choosing an occupation is negatively correlated with expected entry cost, this effect is quite small and is not distinguishable from zero. In comparison with the previous finding, the compensating variation associated with a 25% reduction in the entry costs of non-farm occupations is just 0.6% of expected earnings. One possible explanation for why credit constraints do not appear to impinge on entry into non-farm occupations is that these costs are relatively low to begin with. While there is considerable variation in entry costs in the study sample, the median value for these costs across all non-farm occupations is less than $40. In addition, although the set of occupations used in this study represents the majority of alternatives available to people, it does not include a small number of very high entry cost occupations, such as wholesalers, and a few low cost alternatives, such as grass mat weavers. Omitting these occupations could bias the results for entry costs toward zero.

These results are robust to the inclusion of controls, such as education and cognitive ability, sector specific work experience, and other demographic characteristics, which may be correlated both with people's subjective beliefs and their occupational choices. The covariates also provide insight on the productive efficiency of occupational selection. Under financial constraints, production may be inefficient if people are unable to enter the occupations for which they are particularly well suited. However, I find evidence of positive selection on skills. People with low levels of education tend to do more low-skilled physical labor, while people with high cognitive ability are more likely to be skilled entrepreneurs.

The main findings are also consistent with a growing literature on the role of credit and insurance constraints on productive choices in developing countries. The blossoming of microfinance institutions in recent years has been motivated in part by the idea that credit constraints prevent people from being able to acquire investment capital needed to start their own enterprises. Indeed, a key assumption of the Grameen Bank "is that if individual borrowers are given access to credit, they will be able to identify and engage in viable income-generating activities."[1] Supporting this, Banerjee and Newman (1993) present a dynamic model in which credit constraints and indivisible entry costs lead to a persistent poverty trap. However, the empirical evidence is more mixed. Paulson and Townsend (2004) find that individuals in Thailand with high levels of initial wealth are more likely to engage in entrepreneurial activity. Banerjee et al. (2010) evaluate a microfinance program in India in which access to credit was randomized and find that 7% of treated households started a new business, compared to 5.3% of control households (a 32% increase in the number of new microenterprises). In contrast, McKenzie and Woodruff (2008) show that in Mexico median start-up costs are sufficiently low relative to profits such that limited access to capital ought not impede small business formation (similar to this study, the authors find that median entry costs in Mexico were less than $50).

At the same time, a number of studies focus on the impact of risk and incomplete insurance markets on production decisions. There is a long literature examining the role of insurance failures on agriculture in developing countries (Bliss and Stern, 1982, ch. 8; Binswager and Rosenzweig, 1993; and Morduch, 1990). More recently, Karlan et al. (2011) find that households in Ghana randomly assigned to rainfall insurance expand farm production and have greater harvests than control households, while households assigned both rainfall insurance and cash grants increased the use of chemical fertilizer. In contrast, households randomly assigned access to cash grants only did not use the money to expand agricultural production, but instead invested in school fees and home repairs. In terms of non-farm occupational choices, Bianchi and Bobba (2010) show that Mexican households that receive transfers from the Progresa program are, in fact, more likely to start microenterprises. However, their analysis reveals that households are more sensitive in this choice to the stream of expected future transfers than to the value of the last transfer received. From this they conclude that it is the insurance value of the transfers rather than the easing of liquidity constraints that spurs households to start microenterprises.

Extending on this work, this paper is the first to my knowledge to examine the effects of both credit and insurance constraints on individuals' specific choice of non-farm occupation (whether it be a small enterprise or day labor, if any) in developing countries. The results presented here support findings that suggest a more modest role of credit constraints on productive choices. Instead, the evidence bolsters the notion that these decisions are driven

[1] Grameen Bank website.

more by “income-smoothing” behavior (Morduch, 1995), in which risk averse individuals with limited ability to insure against shocks shift production (or choose occupations) to less risky but also less profitable alternatives.

On the methodological side, this study builds on an emerging body of evidence on the efficacy of eliciting subjective expectations in developing countries where many respondents have low levels of education and lack numeracy skills thought to be important for answering such questions (see Delavande et al., 2011a, for a review). Respondents in the study sample were able to answer questions about subjective probabilities and their beliefs were quite accurate. The majority of responses about expected entry costs and the mean and variance of profits for more than a dozen occupations were clustered around the “true” values as measured through traditional survey questions asked of people in those occupations. In addition, consistent with a number of previous studies that use subjective beliefs in models of choice (e.g. McKenzie et al. 2007, and Giné et al. 2009), this paper finds that the expectations variables continue to provide predictive power even after controlling for correlates of expectations such as education and experience. While this paper does not provide causal evidence of the relationship between expectations and choice, a few experimental studies have demonstrated this link. Jensen (2010) shows that secondary school students in the Dominican Republic randomly assigned information about the returns to schooling derived from earnings data change their expectations and choose to stay in school longer. Nguyen (2008) finds similar results from another randomized experiment in Madagascar: students provided with actual returns update their beliefs and improve test scores, suggesting greater investment in schooling.

The remainder of this paper is organized as follows. Section 2 provides a conceptual framework for the occupational choice decision. Section 3 describes the sample, data, and subjective expectations elicitation procedure. Section 4 provides some background on the various occupations people can choose from, their experience with those occupations, as well as their beliefs about them. Section 5 provides the empirical strategy and results, and Section 6 concludes.

2 Conceptual Framework

In order to provide the basic intuition for how credit and insurance market constraints affect occupational choice, I develop a simple model of an individual’s consumption and production decisions.[2] Agents maximize utility of a single, non-storable consumption good c with

[2]These decisions may instead be made at the household level. I address this concern in detail in Appendix A, where I provide justification and supporting evidence for using the individual model.

associated cost p. Labor does not enter the utility function and is supplied inelastically. However, agents choose how much of their labor endowment, E^L, to allocate into two productive activities. The first activity is farming, which is a concave function of labor and land, $f(L, A)$. Agents choose to allocate L_a^f units of their own labor and L_h^f units of hired labor to farming. The prevailing wage rate is w. I assume there is no land market, so agents farm only their endowment of land, E^A. The second activity is non-farm production, which depends on labor and capital, K. Agents have a capital endowment of E^K, which they can allocate to productive use as K_h or sell on the market as K_m at price r. They may also hire capital K_a at price r as well. Agents choose a non-farm activity j from a set of concave non-farm production functions, $g_j(L, K)$, $j = 1, ..., J$. This set includes wage work and one or more different types of microenterprise. For now, assume that, except for land, agents face a complete set of markets. Thus, the agent's maximization problem is

$$\max U(c) \tag{1}$$

subject to

$$pc + w(L_h^f + L_h^{nf}) + rK_h \leq f(L^f, E^A) + g_j(L^{nf}, K) + rK_m \tag{2}$$

$$L^f = L_h^f + L_a^f, L^{nf} = L_h^{nf} + L_a^{nf} \tag{3}$$

$$K = K_h + K_a \tag{4}$$

$$E^L = L_a^f + L_a^{nf}, E^K = K_a + K_m \tag{5}$$

$$c, L_a^f, L_a^{nf}, E^A, E^K, K_a, K_m \geq 0 \tag{6}$$

Substituting (3)-(5) into (2) yields the "full-income" constraint:

$$pc \leq \Pi^f + \Pi^{nf} + wE^L + rE^K \tag{7}$$

$$\Pi^f = f(L^f, E^A) - wL^f \tag{8}$$

$$\Pi^{nf} = g_j(L^{nf}, K) - wL^{nf} - rK \tag{9}$$

$$c, L^f, L^{nf}, E^A, K \geq 0 \tag{10}$$

This results in the standard "separation property" (Singh, Squire, and Strauss, 1986, and Benjamin, 1992): agents first choose to allocate labor and capital to maximize farm and non-farm profits given land holdings E^A, then they maximize utility subject to the full-income constraint. At the optimum, agents choose L^{f*} and L^{nf*} such that $f_{L^f}(\cdot) = g_{j,L^{nf}}(\cdot) = w$. In addition, agents choose K^* such that $g_{j,K}(\cdot) = r$. This ensures efficiency in production. Given these profit maximizing conditions, agents select the non-farm production alternative

j that is associated with the largest profits, $\Pi_j^{nf} \geq \Pi_k^{nf}$ $\forall k \in J$. Production decisions depend only on prices and the land endowment, and not on labor and capital endowments or risk preferences, while non-farm occupational choice depends solely on the relative size of profits associated with each activity.

2.1 Credit constraints

Credit constraints on entry require both that people have limited (or no) access to credit and that there is some non-convexity in production, such as an indivisible entry cost. Assume now that there is no market for capital such that non-farm production depends on labor and individual endowments of capital, E^K. In addition, assume that for each non-farm occupation j there is some minimum level of capital, $\underline{K}_j$, below which $g_j(L, K) = 0$ for all $K < \underline{K}_j$. Individuals now choose the non-farm occupational alternative that maximizes profits subject to the new constraint that $E^K \geq \underline{K}_j$ for each alternative j. Thus occupational choice under credit constraints will now depend on both expected profits and the relative size of entry costs and capital endowments (but not risk preferences).

2.2 Insurance constraints

Suppose both farming and non-farming production are risky. Assume that farm risk enters with a single multiplicative factor θ, and non-farm risk enters with multiplicative factor φ. Both are random variables with mean one and variance σ_θ^2 and σ_φ^2, respectively. Except for aggregate shocks, assume also that $cov(\sigma_\theta^2, \sigma_\varphi^2) = 0$. Agents now choose farming labor so that

$$EU'(c)\left[\theta f_{L^f}(\cdot) - w\right] = 0$$

Multiplying through, subtracting $f_{L^f}(\cdot)EU'(c)$ from both sides, and rearranging yields:

$$f_{L^f}(\cdot)\left(\frac{cov(U'(c), \theta)}{EU'(c)} + 1\right) = w$$

where the covariance term comes from the fact that $E(\theta) = 1$. Using the definition of correlation this becomes:

$$f_{L^f}(\cdot)\left(1 + \rho_{U'(c),\theta}\sigma_\theta \frac{\sigma_{U'(c)}}{EU'(c)}\right) = w$$

where $\rho_{U'(c),\theta}$ is the correlation coefficient between the marginal utility of consumption and the farming shock. This leads to a well-known result. In the case of full insurance, $\rho_{U'(c),\theta} = 0$

and we again have that agents allocate L^{f*} units of labor to equate the marginal product of farm labor to the wage. However, if insurance markets are incomplete, the marginal utility of consumption will be correlated with the farming shock. This correlation will be negative since a positive shock implies greater consumption and thus lower marginal utility. Since $\sigma_\theta > 0$ and $\frac{\sigma_{U'(c)}}{EU'(c)} > 0$, agents will choose $L^f < L^{f*}$ and farming will not be productively efficient and farm output will be lower. In particular, the efficiency cost will depend on the degree of the insurance failure, $\rho_{U'(c),\theta}$, the riskiness associated with farming, σ_θ, and the shape of the utility function, $\frac{\sigma_{U'(c)}}{EU'(c)}$.[3]

As with farming, in the case of insurance market failure the first-order conditions of the maximization problem with respect to non-farm labor and capital reduce to

$$g_{L^{nf}|j}(\cdot)\left(1+\rho_{U'(c),\varphi}\sigma_{\varphi|j}\frac{\sigma_{U'(c)}}{EU'(c)}\right)=w$$
$$g_{K|j}(\cdot)\left(1+\rho_{U'(c),\varphi}\sigma_{\varphi|j}\frac{\sigma_{U'(c)}}{EU'(c)}\right)=r$$

From the above equations it is clear that the optimal allocation of labor and capital for each non-farm occupational alternative will not be the productively efficient allocation. Moreover, non-farm occupational choice will now be a function of both expected profits and risk. That is, for any two non-farm production functions that are identical except for risk, agents will choose the occupation with lower risk since that would imply a smaller efficiency loss and thus correspond to greater earnings.

3 Sample and Data

3.1 Sample

The sample for this study was derived from a census that identified 1,898 households living within 4 km radii of three small market centers in rural Western Kenya. The census was part of a separate on-going research project on financial access (Dupas et al., 2011). As part of that project, households with prior access to formal financial institutions were excluded from further study. Polygamous households and households with no female head were also dropped[4]. In total, 989 households of the 1,898 households identified in the census met these eligibility criteria. Missing data from subsequent data collection resulted in the loss of 130 additional households and a final sample for this study of 859 households consisting of 1,156 people, which represents 45% of the census survey.

[3]For example, if agents are risk neutral $\frac{\sigma_{U'(c)}}{EU'(c)} = 0$ and the efficient allocation is again possible.

[4]These households are likely quite different from the rest of the sample, yet there are too few of them to allow for separate analysis.

Appendix Tables B1 and B2 present household and individual summary statistics comparing the census population to the study sample. In general, the study sample is slightly poorer and less educated. Households in the study sample population are less likely to have a cement floor or iron roof, or have a mobile phone in the house. They hold fewer assets and own less land. By construction, households in the study sample also had no access to formal financial services at baseline, whereas 20% of the census sample had at least one member with a bank account. At the individual level, both women and men in the study sample have slightly fewer years of schooling, and lower levels of literacy. In terms of labor market participation, people in the study sample are more likely to have identified farming as their primary occupation at baseline. They are less likely to be formal employees (such as government hired teachers) and less likely to have said owning a microenterprise was their main activity at baseline.

3.2 Data

The data come from three types of survey: a background survey carried out in late 2009-early 2010, four waves of a panel survey conducted approximately every three months after that, and a survey on individual subjective beliefs about the distribution of profits and the entry costs of various non-farm occupations in the study area, which was collected mid-way through 2010, after the first two waves of panel surveys (see Figure 1 for a timeline of the data collection).

3.2.1 Baseline survey

The baseline survey includes information on personal and household characteristics (such as educational attainment, family size, assets and land holdings), as well as measures of cognitive ability. Cognitive ability was measured using a digitspan recall test in which respondents are asked to repeat back increasingly long chains of random numbers. In addition, the baseline survey contains information about personal and parental work experience, including the type of experience (e.g. kind of microenterprise or wage work) and the length of the experience.

Summary statistics from the baseline survey are presented in Table 1. Seventy-one percent of the sample are women, and 71% of all individuals are married. The average respondent is 41 years old and has about 3.5 children living in the household. Educational attainment is low: women have 5 years of schooling and men have about 7 years. Approximately two-thirds of the sample can write a letter in Swahili. Just 12% of respondents report living in a home with a cement floor (instead of dirt), and 43% report having roof made of iron sheets (instead of thatch). Nearly everyone owns land, but land holdings are quite small and average slightly under 2 acres. In terms of use of informal credit and insurance mechanisms, around 40% of the sample participates in a ROSCA, and they receive nearly $50 on average in transfers from outside the household each year.

3.2.2 Panel surveys

The panel surveys contain detailed information about labor supply, income sources, and earnings over the prior month. To minimize the chance that respondents failed to report all of their income, enumerators asked respondents about their earnings from a comprehensive list of income sources (including sale of livestock and farm produce, wage work, self-employment, and pensions or other government transfers). Respondents reported their total revenues and hours worked for each source of income. For self-employment activities, respondents also reported their total profits over the same period.

3.2.3 Subjective beliefs survey

The survey on subjective beliefs elicited people's expectations about the characteristics of 16 non-farm occupations as well as farming. The set of non-own-farm occupations includes three types of physical labor (casual farm and non-farm work, and bicycle taxis), seven types of market vending businesses (including sale of dry or fresh fish, fruits and vegetables, grains, household goods, clothing, and charcoal), and six types of small enterprises (tailors, shopkeepers, metal and wood workers, food kiosks, and barbers), and represents more than 80% of the non-farm occupations people in the sample reported engaging in over the course of a year.

For each occupation, respondents were asked to report their beliefs about the entry costs and the distribution of weekly profits for the "average person" in the occupation and, to account for differences in human capital attributes or physical capital holdings, for themselves if they were to enter into the occupation. The full distribution of weekly profits was elicited using a visual aid and follows the procedure Delavande et al. (2011b) found to be the most accurate in a recent randomized control trial of subjective expectation survey techniques. The exact procedure used in this paper was as follows:

First, trained survey enumerators explained the visual aid. The visual aid consisted of 20 kernels of maize. Respondents were told that they could use these kernels of maize to represent how likely they thought something would happen. For example, enumerators explained that if they selected 0 kernels of maize that meant they thought the event in question was impossible, whereas if they selected all 20 kernels that meant they thought it was definitely going to happen, and finally, if they selected 10 kernels that meant they thought it had equal chances of happening or not.

Next, the survey enumerators guided respondents through a series of practice questions to ensure that they understood the visual aid and to give them experience using it. The practice questions included a zero probability event, an event with relatively equal probability of happening, and several questions with multiple possible outcomes. For example, for the relatively equal probability practice question, respondents were asked: If a baby is born, how likely do you think that it will be a girl? For each question, respondents were presented with a laminated sheet of paper with different squares printed on it that represented the outcomes described in the practice questions. Thus, for the equal probability practice question, respondents were presented with two squares, one labeled "girl" and the other "boy", into which they could place the kernels of maize. The survey enumerators were instructed to give assistance to respondents who at first struggled with the visual aid and to make sure that the concepts were understood before continuing.

Once the respondent was well practiced with the visual aid, they moved on to the main part of the survey. For each occupation, respondents were asked to state the minimum and maximum weekly profits they would expect the average person in this occupation to earn in a week. Respondents were then given the laminated sheet of paper with different squares printed on it, each of which now represented a profit interval. The size of these profit intervals was predetermined and guided by extensive pre-testing in the field. Respondents were asked to allocate the 20 kernels of maize across these profit intervals to represent the likelihood of earning the amount in the interval during one week. Respondents had to use all 20 kernels of maize and allocate them only into the profit intervals that fell between the minimum and maximum weekly profits they expected the average person in that occupation would earn in a week.

To give an example, suppose a respondent said that the minimum weekly profit for the average person in occupation A was 150 Ksh and the maximum profit was 800 Ksh. The respondent would then be directed to the squares on the laminated sheet representing 101-200 Ksh, 201-300 Ksh, and so on up to 700-800 Ksh. Next, the respondent would allocate the 20 kernels into these 7 squares to represent the likelihood the average person in occupation A would earn this amount in a week.

Finally, after respondents allocated the maize kernels for a given occupation for the "average person", they were then asked to use a second set of 20 kernels of maize (painted a different color) to represent the weekly profits they believed they would earn if they were to work in this occupation themselves. During this latter exercise, the first set of maize kernels was left on the printed grid so that respondents could compare their responses for the "average person" and for themselves.

Using the results of this exercise, I calculate each individual's expected average return and the variance of that return for each occupation for both the "average person" in the occupation and for the respondent him/herself.

4 Labor participation and beliefs about occupations

Putting these three strands of data together, a picture emerges of the pattern of farm and non-farm labor participation in the sample, and how people view the non-farm occupations available to them. Nearly everyone in the sample does some farming. Non-farm occupations are also common, but there is a lot of entry into and exit from these occupations. Finally, perhaps as a result of the high turnover, people report having experience in a number of different occupations, and they have fairly accurate beliefs about the characteristics of these jobs.

4.1 Labor participation

Table 2 presents evidence from the baseline survey on individuals' participation in both farming and non-farming. In the baseline year, 98% percent of the sample reported farming their own plots. Maize and beans (both staples) are the most common crops grown, and much of this farm production is used for own consumption. At the same time, it is quite common for respondents to engage in additional income generating activities. At baseline, close to half of the sample population reported earning money in a non-own-farm activity. In addition, people report having experience in several different types of non-farm labor. The average respondent has approximately 2 years of experience market vending, just under 2 years experience owning their own small enterprise (other than market vending), and nearly 6 years experience working as a casual laborer.

The fact that people commonly report having experience with a wider set of occupations than those they are currently engaged in suggests that occupation turnover is relatively important. The panel data largely confirms this, as shown in Table 3. Between each survey wave, around 40% of the sample either starts or stops an occupation other than farming. This is not limited to people switching from farming only to farming plus some additional activity, or vice versa – people also switch between non-farm activities at the same frequency. This amount of turnover is consistent with Liedholm and Mead (1998), who find a lower-bound of 20% turnover for MSEs in Kenya. Maybe surprisingly, in my sample these changes do not appear to follow any seasonal patterns and the overall distribution of occupations remains relatively stable over the course of the year.[5]

Table 4 provides a snap-shot of the occupational distribution in detail at the time of the third wave of the panel survey.[6] While 59% of the sample population during this time was engaged in farming only, approximately 11% also were market vendors, 5% operated their own enterprise (other than market vending), and 26% hired themselves out for physical labor (both farm and non-farm). There are some gender differences across the occupational distribution as well. Women are twice as likely as men to do farming only, and they comprise most of the market vendors. Men, on the other hand, are much more likely to be hired to perform non-farm physical labor (which typically includes tasks such as hauling, construction and repair, and other handyman services), while women are more likely to perform day labor on someone else's farm.

4.2 Beliefs about occupations

4.2.1 Accuracy of beliefs

How close are the elicited expectations to the truth? In order to investigate this, I pool the self-reported profits of individuals working in each occupation across the four panel surveys. This data is then used to compute the mean and variance of profits for each occupation. The same procedure applies for entry costs, but, because these were asked only during the first panel survey, there are many fewer observations with data to compute the average entry cost for each occupation and so these values are rather imprecise. In total, there were 12 non-farm occupations for which there was sufficient data to produce estimates of these "true" values. I then compare these values to the respondents' beliefs.

Figures 2, 3, and 4 plot the distribution of people's beliefs about mean profits, the variance of profits, and mean entry costs, respectively, for the "average person" in the occupation against the "true" values computed from the panel survey data. In each plot, the "true" value is represented by the vertical line. For the sake of visual clarity, I have not plotted beliefs of people who gave responses outside the predetermined profit interval support in the case of mean and variance of profits, and I have trimmed entry costs at 40,000 Ksh. However, in each plot I include information on the number of observations that were dropped in this manner, as well as the number of observations used to compute the "true" values. For almost every occupation, the mode of the distribution of people's beliefs is centered on the "true"

[5] There are two agricultural seasons in this part of Kenya, a "long rains" season (with planting in February/March and harvesting in July/August), and a "short rains" season (with planting in September/October, and harvesting in December/January).

[6] The occupational choices in this panel are used in the analysis of this paper. However, as discussed above, the distribution of occupations remains relatively unchanged across the panel rounds.

values. Moreover, almost 60% of beliefs about mean profits and entry costs are within 2 standard deviations of the "true" values computed from the data. This suggests that, in general, people have quite realistic beliefs about the characteristics of each occupation.

It is not altogether surprising that people are knowledgeable about these occupations. As noted earlier, many people have quite a bit of experience working in each sector and have had a chance to learn about them first hand. In addition, people also interact with market vendors and skilled tradesmen frequently, either as friends or family members or as paying customers, which provides another opportunity to find out about how lucrative, risky, or costly different occupations are. For example, as shown at the bottom of Table 2, 35% of respondents in our sample have a parent with experience in market vending, and 28% have a parent with entrepreneurship experience.

4.2.2 Do people perceive trade-offs between profits and entry costs or risk?

Respondents associate higher profits both with larger entry costs and greater risk. Figure 5 plots median beliefs about own profits for each of the 16 occupations against median beliefs about entry costs and the variance of profits. While in both cases there is a positive relationship, the relatively greater degree of dispersion in the plot relating profits and entry costs suggests that the trade-off that is most salient to people is the one between profits and risk. Indeed, except for the four most profitable occupations, there appears to be almost no relationship between expected profits and entry costs.

4.2.3 Do people enter occupations they believe are the most profitable?

As argued in Section 2, if credit and insurance markets are perfect people will enter the occupation which they associate with the highest expected earnings. On the other hand, if there are financial constraints, we should observe too little entry into high profit occupations that also carry high entry costs or large risks, and too much entry into low return occupations that have low costs or low risk. Column 1 of Table 5 presents the fraction of people in the sample employed in each occupation as of panel wave 3. Almost 60% of the sample is engaged in farming only, which carries no additional entry costs or risk, but also earns no additional income. Another quarter of the population is engaged in farm and non-farm casual labor, which are both associated with the lowest entry costs and moderate levels of risk. Another 16% of the sample operates some type of microenterprise, with a larger concentration in those enterprises that carry some of the lowest risk but not necessarily the lowest entry costs. In contrast, very few people engage in the more profitable own enterprise occupations such as metal and wood working shops and small retail shops. These latter occupations tend also to be associated with both the highest risk and the largest entry costs.

Column 2 of Table 5 presents the fraction of people who believe a given occupation would yield them the highest average returns as compared to all other occupations. Significantly, not all people believe the same occupation is the most profitable. That is, if people faced no credit or insurance constraints, and were free to choose the occupation with the highest return, the would not all become metal shop owners. These differences in beliefs may be associated with differences in individual human capital, and underscore the importance of using own beliefs rather than some aggregate measure of returns in the estimation procedure. At the same time, these data also suggest that there is potentially too little entry into high-profit occupations such as retail and metal shops, which carry both very high costs and high risks, but also certain higher-return market vending activities such as clothing and household good sellers that are associated with more reasonable entry costs but still carry a lot of risk.

5 Empirical Strategy and Results

5.1 Empirical strategy

In order to measure the relative effect of credit and insurance market constraints on people's non-farm labor market decisions, I assume and estimate an additive random utility model (ARUM) of occupational choice. Individuals choose from among 17 alternatives, including the 16 different non-farm occupations and non-participation. Given that nearly everyone in the sample is a farmer, non-participation in this context refers to choosing to do farming only, while participation means that the individual both farms and does some additional income-generating activity.

Individuals choose the occupational alternative that provides them with the greatest utility. As is standard in ARUM models, the utility associated with occupation k in sector j is given by

$$U_{jk} = V_{jk} + \varepsilon_{jk}$$

where V_{jk} is the deterministic component of utility, and ε_{jk} corresponds to an unobserved random component of utility (the individual subscript i has been omitted for clarity). Assuming that the ε's come from a generalized extreme value distribution, and allowing for these errors to be correlated within the three non-farm occupational sectors, this is the nested logit model of McFadden (1978) and the probability of choosing a given occupation has a well-known closed form solution.

According to the framework described in Section 2, the deterministic component of utility V_{jk} will depend on some, or all, of the observable characteristics of each occupational alternative. It can be written as

$$V_{jk} = \alpha_{jk} + \beta \ln(avgprofit_{jk}) + \gamma \ln(varprofit_{jk}) + \delta(entrycost_{jk})$$

where α_{jk} is an occupational alternative-specific constant that reflects the benefit of occupation k in sector j not captured by the other regressors. The expected variance of profits also enters in the natural log because of the concavity of the non-farm production functions. As the variance increases, adjustments in the marginal product of labor and capital require smaller and smaller changes in allocations of labor and capital, and thus have smaller impacts on production and earnings. Entry costs, on the other hand, enter linearly since what matters for choice is the level of these costs. The results presented below look very similar under alternative specifications.

When markets are complete, occupational choice will be determined solely by the relative size of average profits. In this case, we would expect $\beta > 0$ and $\gamma = \delta = 0$. If, however, people are credit constrained (i.e., they lack access to credit and face indivisible entry costs) then occupational choice will be negatively related with entry costs and $\delta < 0$. Similarly, if insurance markets are imperfect and individuals are risk averse, utility will be decreasing in the variance of profits and we would expect $\gamma < 0$.

The characteristics of each occupational alternative (average profits, variance of profits, and entry costs) used in the estimation are given by people's elicited subjective beliefs. Since this data was collected just after round 2 in the panel, people's occupational choices are defined as those occupations they were observed in during round 3 of the panel three months later. In this sense, and given the high rate of turnover in non-farm occupations, the data reflect people's *ex ante* subjective beliefs before making their occupational choices. For each of these variables the elicited values correspond to those for "respondent him/herself" rather than for "the average person." Using the respondent's own returns I am thus able to control for person-specific differences in human capital that may influence occupational choices.

5.2 Results: The role of Missing Insurance and Credit Markets

Table 6 presents the main results from estimating the nested logit model. The coefficient on the expected average profit is positive and significant, indicating that people do indeed select non-farm occupations they believe will be more profitable. At the same time, there is also evidence that people choose occupations that they associate with less risk. The coefficient on the expected variance of profits is negative and significant. This is consistent with the framework developed in Section 2 in which the lack of complete insurance markets forces a trade-off between potential earnings and risk.

In contrast, it does not appear that imperfect credit markets constrain non-farm occupational choice. The coefficient on the expected entry costs is both small and statistically indistinguishable from zero. This does not preclude a role of credit constraints on profit sizes or growth, but it does suggest that access to credit does not impede entry into the different occupations generally available in the study area.

To check the robustness of these results, I include several sets of additional regressors that may be correlated both with occupational choice and people's beliefs, or that may be predictors of occupational choice in their own right. These regressors are added sequentially and include demographic characteristics (gender, marital status, age, education, cognitive ability, and land holdings), sector-specific work experience, and parental sector-specific work experience. The results for the main parameters of interest after including these controls are presented in columns 2-4 of Table 6. In each case, adding covariates only increases the point estimates and makes them more statistically significant. In addition, I run the same regressions again, this time trimming outliers. Outliers are defined as beliefs about returns that fell outside the predetermined profit interval, or beliefs about entry costs that were above 40,000 Ksh. These results are similar to main findings and are presented in Table 6a.

5.3 Compensating variation and WTP to relax constraints

To investigate the potential welfare costs of the insurance and credit constraints, I compute the compensating variation associated with changes in the risk or entry costs that individuals face. This is, of course, a partial equilibrium effect, but it is still useful in terms of measuring the perceived size of the trade-offs that people face.

The compensating variation is the amount of money that a person would need to receive (or pay) to equate maximum utility in the initial state with maximum utility in some new state. The utility derived from each alternative j in the choice set can be rewritten as

$$U_j = V(y - p_j, \mathbf{X}_j, \mathbf{C}) + \varepsilon_j$$

where y is income, p_j is the price of alternative j, $\mathbf{X}_j$ is a vector of alternative-specific characteristics, and $\mathbf{C}$ is a vector of individual-specific characteristics. Assuming budget exhaustion, the sum $y - p_j$ represents income spent on the numeraire good. In this application, the sum $y - p_j$ represents the expected earnings from choosing occupation j and the numeraire good is total consumption. As before, $\mathbf{X}_j$ consists of the expected variance and entry cost terms as well as an alternative-specific constant. In general, the compensating variation associated with a change from the initial conditions $\{y^0, p^0, \mathbf{X^0}\}$ to some new condition $\{y^0, p^0, \mathbf{X}^1\}$ is the amount of money cv such that

$$U^0 = max_j \left\{ \left[V(y^0 - p_j^0, \mathbf{X}_j^0, \mathbf{C}) + \varepsilon_j \right] \right\} = max_j \left\{ \left[V(y^0 - cv - p_j^0, \mathbf{X}_j^1, \mathbf{C}) + \varepsilon_j \right] \right\}$$

Thus, the compensating variation measures how much people would be willing to pay (or would need to be paid in the case of a worsening) to be restored to their original utility after a change in one of the determinants of choice. For instance, in this application the compensating variation associated with a reduction in the variance of the returns (or a reduction in the entry costs) can be thought of as the willingness-to-pay for an insurance product (or for a loan). This can also be thought of as equivalent to the amount of returns that people forgo because of the existence of risk and/or credit constraints.

In order to make these calculations, I follow the procedure presented in Karlstrom and Morey (2004) who provide an exact formula for the expected compensating variation in a model with income effects.[7] The general formula is given as

$$E[cv] = y - E[m]$$

where

$$E[m] = -\sum_j \int_{\underline{\mu}}^{\mu_{jj}} y dP_j(y)$$

Here j indexes the alternatives, μ_{jj} is the amount of money needed in the new state to restore utility to the level it was in the old state if the individual chooses alternative j both before and after the change, $\underline{\mu} = min_k \mu_{kk}$, and $P_j(y)$ is the probability of choosing alternative j given the vector of the deterministic components of utility. Calculations for changes in the expected variance and expected entry costs were calculated using this formula and carried out using Ox code (Doornik, 1998).

Figure 5 plots the computed compensating variation as a percentage of expected earnings in the original state against percentage changes to the expected variance and entry costs of the non-farm occupations. This relationship is nonlinear and has a concave shape, reflecting the nonlinear relationship between the indirect utilities and returns as well as the variance of returns. As can be seen, the $E[cv]$ for changes in the variance of non-farm occupations dominates that for changes in the entry costs. Indeed, a 25% change in the variance is associated with an $E[cv]$ that is 11% of expected income. That is, people would be willing to give up 11% of their expected income in order to avoid an occupation that carried 25%

[7] Income effects are implied by the model since $y - p_j$, or expected earnings, enter the indirect utility function non-linearly in the natural log.

more risk. On the other hand, the average willingness to pay for changes in entry costs is much more modest. The same percentage change in entry costs is associated with an $E[cv]$ that is just 0.6% of expected income.

5.4 Other Determinants of Occupational Choice

This section discusses the role of other determinants of occupational choice, besides credit and insurance constraints.[8] Indeed, there may be non-pecuniary costs and benefits that also affect people's occupational choices and these may vary by individual demographic characteristics. For example, people who have large land holdings, or women who have greater home production responsibilities (such as caring for young children), may be systematically less likely to do any non-farm work if these traits increase the opportunity cost of their time. In terms of people's sector-specific work experience, the effect of a person's work history on the probability of choosing an occupation could go in either direction. Individuals who have longer experience in a given sector, may have a better appreciation for its non-pecuniary costs and benefits. For instance, one might learn that while profits are good for, say, physical labor, the work is simply too physically taxing. On the other hand, people who have worked for a long time in one occupation may be reluctant to switch to another occupation, even if they recognize larger profits could be had, because the effort of reestablishing contacts and networks or developing new skills appears too onerous. Finally, occupational choice may be dynastic. Individuals may choose the same occupations as their parents or close family members. This could be a direct effect because parents or family members encourage the choice. Or, it could be indirect if family members influence individual beliefs and then people choose occupations based on those beliefs.

The point estimates for these covariates are presented in Table 7. There are separate estimates for each occupational sector, and these can be interpreted as in a binary logit model where the omitted category is farming only. Compared to people who farm only, physical laborers are more likely to have completed less years of schooling, while people who have their own enterprise have higher cognitive ability scores. This suggests that despite the presence of financial constraints, there is some positive selection on skills. Market vendors, on the other hand, are more often female and unmarried, and they come from households with larger land holdings relative to farmers (perhaps indicating that they come from richer households as well). Across all three sectors, own-sector experience is a strong predictor of current occupational choice. Finally, only market vending appears to have a dynastic

[8] Since the beliefs data used in the estimation are for the returns and costs for the respondent him/herself, and since the estimation is within individual, differences across individuals in human capital that may affect choice are already accounted for.

element. People whose parents or close family friends were market vendors are also more likely to be market vendors themselves. Table 7a presents the same results using the trimmed beliefs data.

6 Conclusion

This paper has shown that market imperfections can have significant effects on occupational choice and productive efficiency in rural areas of developing countries. In particular, insurance market failures prevent people from entering occupations that are associated with high profits but also carry large risks. Instead, most people are forced to make do in safer but much less remunerative pursuits, such as short-term physical labor contracts. On average individuals give up a substantial portion of their potential income because of these risks and the lack of insurance.

In contrast, the results of this paper suggest that the effect of credit constraints is more muted. Although most non-own-farm employment is concentrated in occupations that have the lowest entry costs, these costs do not appear to be a barrier for people's occupational choices. However, this paper does not rule out the possibility that a lack of access to credit may prevent small enterprises from growing larger.

These results highlight some of the challenges faced by policymakers who wish to encourage the growth of microenterprises in developing countries as a means to alleviate poverty. On the one hand, providing access to credit (or enabling savings) appears to be an attractive approach to mitigating both credit and insurance market constraints. If people have cash on hand, or are able to get a microloan during a bad shock, they may be enabled to enter high profit occupations that also carry high risk. However, recent evidence from experimental studies of microfinance programs suggest that these efforts have a relatively small impact on new small business formation.

On the other hand, given the results of this paper and other recent literature, providing insurance may have a more direct and larger impact on these outcomes. But, in designing insurance products for non-farm occupations, the problems of moral hazard and adverse selection appear more difficult than in the case of rainfall insurance for farmers. Further research is needed on the effect of providing people in developing countries with access to insurance for non-farm occupations, or to safety net programs that help people mitigate business risks (such as NREGA in India, which guarantees wage work in the case of business failure). In these cases, design features that address moral hazard and adverse selection issues should also be investigated.

References

[1] Banerjee, A. V., E. Duflo, R. Glennerster, and C. Kinnan (2010, June). The miracle of microfinance? Evidence from a randomized evaluation. Mimeo, MIT.

[2] Banerjee, A.V. and A. Newman (1993). Occupational Choice and the Process of Development. *The Journal of Political Economy 101*(2), 274-298.

[3] Benjamin, D. (1992). Household Consumption, Labor Markets, and Labor Demand: Testing for Separation in Agricultural Household Models. *Econometrica 60*(2), 287-322.

[4] Bianchi, M. and M. Bobba (2010). Liquidity, Risk, and Occupational Choices. HAL-PSE Working Papers.

[5] Binswanger, H. and M. Rosenzweig (1993). Wealth, Weather Risk and the Composition and Profitability of Agricultural Investments. *Economic Journal 103*, 56-78.

[6] Bliss, C. and N. Stern (1982). *Palanpur: The Economy of an Indian Village*. Oxford: Oxford University Press.

[7] Chuta, E. and C. Liedholm (1990). Rural small-scale industry: Empirical evidence and policy issues. In C. Eicher and J. Staatz (Eds.), *Agriculture Development in the Third World*. Johns Hopkins University Press, Baltimore.

[8] Delavande, A., X. Giné, and D. McKenzie (2011a). Measuring subjective expectations in developing countries: A critical review and new evidence. *Journal of Development Economics 94*, 151-163.

[9] Delavande, A., X. Giné, and D. McKenzie (2011b). Eliciting Probabilistic Expectations with Visual Aids in Developing Countries: How Sensitive are Answers to Variations in Elicitation Design? *Journal of Applied Econometrics 26*, 479-497.

[10] Dupas, P., S. Green, A. Keats, and J. Robinson (2011). Challenges in Banking the Rural Poor: Evidence from Kenya's Western Province. Forthcoming in *NBER Africa Project Conference Volume*.

[11] Giné, X., R. Townsend, and J. Vickery (2009). Forecasting when it matters: evidence from semi-arid India. Mimeo, World Bank.

[12] Jensen, R. (2010). The perceived returns to education and the demand for schooling. *Quarterly Journal of Economics 125*(2), 515–548.

[13] Karlan, D., I. Osei-Akoto, R. Osei, and C. Udry (2011). Examining underinvestment in agriculture: Measuring returns to capital and insurance. Mimeo, Yale University.

[14] Karlstrom, A. and E. Morey (2004). Calculating the Exact Compensating Variation in Logit and Nested-Logit Models with Income Effects: Theory, Intuition, Implementation, and Application. Presented at the meeting of the American Economic Association, New Orleans, January 2001. Available SSRN: http://ssrn.com/abstract=562802.

[15] Liedholm, C. and D. Mead (1998). The dynamics of micro and small enterprises in developing countries. *World Development 26*(1), 61-74.

[16] McFadden, D. (1978). Modelling the Choice of Residential Location. In A. Karlquist, L. Lundquist, F. Snickars, and J. W. Weibull et al. (Eds.), *Spatial Interaction Theory and Planning Models*, 75–96, Amsterdam, New York, North-Holland.

[17] McKenzie, D., J. Gibson, and S. Stillman (2007, May). A land of milk and honey with streets paved with gold: do emigrants have over-optimistic expectations about incomes abroad? World Bank Policy Research Working Paper No. 4141; IZA Discussion Paper No. 2788.

[18] McKenzie, D. and C. Woodruff (2008). Experimental evidence on returns to capital and access to finance in Mexico. *World Bank Economic Review 22* (3), 457–482.

[19] Morduch, J. (1990). Risk, Production, and Saving: Theory and Evidence from Indian Households. Mimeo, Harvard University.

[20] Morduch, J. (1995). Income Smoothing and Consumption Smoothing. *Journal of Economic Perspectives 9*(3), 103-114.

[21] Nguyen, T. (2008). Information, role models and perceived returns to education: experimental evidence from Madagascar. Mimeo MIT.

[22] Paulson, A. and R. Townsend (2004). Entrepreneurship and Financial Constraints in Thailand. *Journal of Corporate Finance 10*(2), 229-262.

[23] Reardon, T. (1997). Using Evidence of Household Income Diversification to Inform Study of the Rural Nonfarm Labor Market in Africa. *World Development 25*(5), 735-747.

[24] Singh, Inderjit, L. Squire, and J. Strauss, eds. Agricultural Household Models: Extensions, Applications, and Policy. Baltimore: Johns Hopkins Press (for World Bank), 1986.

Appendix A

Household or Individual Model

Occupational choices may be made jointly within the household. In particular, individuals may coordinate to choose occupations, or agree to share returns, in such a way as to insure their spouse's risk. This would suggest that occupational choice should be modeled instead at the household level. However, under certain conditions, the individual model will perform equally well.

First, consider the case where the individual model will be incorrect. The individual model will provide incorrect estimates if spouses jointly choose occupations whose risks are negatively correlated. This will be a problem if, for instance, we observe a woman choosing an occupation with relatively large risk compared to her other alternatives. Is this because she is risk-loving, as the individual model would suggest? Or is it because the risks of that occupation are negatively correlated with those of the occupation chosen by her spouse and they are insuring one another? In this case, occupational choices depend not only on the relative sizes of the returns and risks an individual faces, but also on the sizes of returns and risks the spouse faces, and the degree to which risks are positively or negatively correlated across occupations.

But if households were behaving in this way, it should be evident in the data. It should be true that husbands and wives rarely work in the same occupation, or at least not in the same occupational sector. Comparing the occupational choices of spouses within all four waves of the panel survey, I find this is not the case. The correlation between spouses' occupational choices is typically quite small and insignificant from zero at standard confidence levels. When it is significant from zero, the correlation tends to be positive. Spouses' sector choices follow the same pattern. These sector-level correlations for panel wave 3 are presented in Appendix Table A1 (other panel waves provide similar results). On the diagonal we see that, if anything, spouse sector choice is positively correlated. This is especially true for households where one individual has their own enterprise. In any case, there is no evidence that spouses are choosing occupations that have risks negatively correlated with one another.

If this is true, the individual model will provide the correct measure of the average effect of risk, and the lack of insurance, on occupational choices. In the individual's problem described in Section 2, it was shown that the extent to which risk affects occupational choice depends also on the level of insurance failure that the person faces. As an individual weighs occupations of varying returns and risks, she implicitly takes into consideration the ability of her spouse to mitigate those risks through his own occupational choice. To give a simple if extreme example, if the household coordinates so that a woman's husband chooses a low-risk occupation in order to fully insure her against the risks she faces, then there will be no correlation between her non-farm production shock and consumption, and there will be no effect of risk on her choice. Going the other way, if a woman knows that her spouse will choose the occupation with the highest return, even if it carries lots of risk, then her implicit level of insurance is lower, and the riskiness of the occupations she faces will play a larger role in her choice. In any such scenario, the parameters in the individual model estimate the trade offs people make between the returns and risks of occupations they face, given the level of insurance provided by their spouse (if any). As long as spouses do not systematically choose occupations with negatively correlated risks, and as long as both husband and wife are included in the regressions, the individual model will provide the correct estimate of the average effect of risk, and the lack of insurance, in the sample population.

Table 1. Sample descriptive statistics

	Mean	Median	Std. Dev.	# obs
Female	0.71	1.00	0.45	1156
Married	0.71	1.00	0.46	1156
Age	40.85	37.00	16.53	1076
Number of children	3.50	3.00	2.31	1137
Years education	5.63	7.00	3.40	1106
Can write a letter in swahili	0.64	1.00	0.48	1021
Land holdings (acres)	1.82	1.00	1.92	1123
Owns land	0.96	1.00	0.18	1156
HH has iron roof	0.43	0.00	0.50	1156
HH has cement floor	0.12	0.00	0.32	1156
Value of physical assets (in Ksh)	8993	7150	7530	1105
Value of animals (in Ksh)	4407	150	9394	1112
ROSCA participant	0.38	0.00	0.49	1102
ROSCA contributions in past year (for those who participate)	3823	2400	4902	424
Transfers into HH in past year (in Ksh)	3731	1950	5340	1143
Transfers out of HH in past year (in Ksh)	863	420	1223	1135

Data: Baseline survey. The exchange rate at the time of the study was around 80 Kenyan Shillings (Ksh) to US$1 on average.

Table 2. Experience with farming and non-farming activities

	All	Women	Men
Participation in Farming and Non-Farming Activities at Baseline			
Farmed during baseline year	0.98	0.98	0.99
Had non-own-farm income at baseline	0.46	0.41	0.59
Non-own-farm experience			
Ever market vendor	0.48	0.57	0.27
Ever entrepreneur (other than market vendor)	0.29	0.26	0.35
Ever physical labor	0.72	0.69	0.79
If ever, years market vendor	4.31	4.52	3.23
	(5.94)	(6.09)	(0.44)
If ever, years entrepreneur (other than market vendor)	6.32	5.76	7.33
	(8.55)	(8.47)	(0.48)
If ever, years physical labor	7.76	7.33	8.68
	(8.52)	(8.45)	(8.60)
Parental non-own-farm experience			
Parent market vendor	0.35	0.35	0.37
	(0.48)	(0.48)	(0.49)
Parent own enterprise	0.28	0.27	0.31
	(0.45)	(0.44)	(0.46)
Number of Individuals	1156	822	334

Data: Baseline survey. Standard deviations in parentheses

Table 3. Sample occupational distribution at each panel survey wave

	Wave 1 (Feb-Mar 2010)	Wave 2 (Jun-Jul 2010)	Wave 3 (Oct-Nov 2010)	Wave 4 (Feb-Mar 2011)
Farming Only	0.61	0.59	0.59	0.60
Farming + Market Vending	0.08	0.11	0.11	0.11
Farming + Own Enterprise (other than Market Vending)	0.03	0.04	0.05	0.06
Farming + Physical Labor	0.28	0.26	0.26	0.24
Turnover between consecutive waves				
% who switched from "farming only" to "farming + non-farm activities"		0.17	0.13	0.14
% who switched from "farming + non-farm activities" to "farming only"		0.16	0.16	0.13
% who switched from one non-farm activity to another		0.11	0.09	0.10
Number of Individuals	1024	1063	1156	1065

Data: Four waves of panel surveys.

Table 4. Snapshot of occupational distribution (Panel Wave 3)

	All	Women	Men
Farming Only	**0.59**	**0.66**	**0.42**
Farming + Market Vending	**0.11**	**0.14**	**0.03**
Fresh fish	0.01	0.01	0.00
Dried fish	0.03	0.04	0.00
Clothing/shoes	0.00	0.00	0.00
Grains	0.02	0.02	0.01
Fruits/vegetables	0.03	0.04	0.01
Household goods	0.01	0.02	0.01
Charcoal	0.01	0.01	0.01
Farming + Own Enterprise (Other than Market Vending)	**0.05**	**0.05**	**0.05**
Tailor	0.01	0.01	0.00
Restaurant/Prepared food kiosk	0.02	0.02	0.01
Metal shop	0.00	0.00	0.01
Wood shop	0.00	0.00	0.01
Barber	0.01	0.01	0.01
Retail shop	0.01	0.01	0.01
Farming + Physical Labor	**0.26**	**0.16**	**0.49**
Bicycle taxi	0.01	0.00	0.05
Casual non-farm	0.11	0.02	0.32
Casual farm	0.13	0.14	0.12
Number of Individuals	1156	822	334

Data: Wave 3 of panel survey, conducted in Oct/Dec 2010.

Table 5. Comparing occupational choices to beliefs about highest profit

	Fraction of sample in this occupation in Panel Wave 3	Fraction of sample who believes this occupation has highest expected profit
Nothing	59.0%	0.0%
Casual farm	13.3%	5.7%
Casual non-farm	10.8%	5.9%
Fruits/vegetables	3.1%	1.2%
Dried fish	2.5%	7.4%
Restaurant/Prepared food kiosk	1.9%	3.2%
Grains	1.6%	3.6%
Bicycle taxi	1.4%	1.0%
Household goods	1.4%	7.9%
Barber	1.1%	3.7%
Fresh fish	1.0%	9.9%
Charcoal	0.7%	3.5%
Tailor	0.8%	6.9%
Retail shop	0.8%	6.8%
Clothing/shoes	0.3%	15.7%
Wood shop	0.3%	9.3%
Metal shop	0.2%	9.8%

Data: Panel Wave 3 and subjective beliefs survey responses for the respondent him/herself.

Table 6. Nested-logit estimation of expectations and occupational choice

	(1)	(2)	(3)	(4)
ln expected average profit	0.148	0.151	0.184	0.187
	(0.093)	(0.089)*	(0.100)*	(0.100)*
ln expected variance of profit	-0.107	-0.109	-0.133	-0.134
	(0.046)**	(0.061)*	(0.059)**	(0.060)**
expected entry cost (100 Ksh)	0.000	0.000	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)
demographic controls		Yes	Yes	Yes
personal experience controls			Yes	Yes
parental experience controls				Yes
Observations	18918	18918	18918	18918
N	1156	1156	1156	1156
Wald	5.65	249.00	798.65	824.26
p-wald	0.13	0.00	0.00	0.00
Pseudo-loglikelihood	-1754.10	-1608.55	-1550.18	-1544.98

*Data: study sample. Demographic controls include age, gender, marital status, education, digitspan recall score, and land holdings. Personal experience controls include sector-specific years of experience. Parental experience controls include dummy =1 if parent experience in that sector, =0 otherwise. Standard errors clustered at the household level. * significant at 10%; ** significant at 5%; *** significant at 1%.*

Table 6a. Nested-logit estimation of expectations and occupational choice using trimmed data

	(1)	(2)	(3)	(4)
ln expected average profit	0.052	0.063	0.117	0.122
	(0.121)	(0.123)	(0.123)	(0.122)
ln expected variance of profit	-0.088	-0.104	-0.133	-0.135
	(0.055)	(0.054)*	(0.054)**	(0.054)**
expected entry cost (100 Ksh)	0.000	0.000	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
demographic controls		Yes	Yes	Yes
personal experience controls			Yes	Yes
parental experience controls				Yes
Observations	16171	16171	16171	16171
N	1122	1122	1122	1122
Wald	7.53	1735.38	3403.22	3190.05
p-wald	0.06	0.00	0.00	0.00
Pseudo-loglikelihood	-1598.33	-1458.74	-1399.19	-1393.26

*Data: study sample with outliers trimmed. Outliers are defined as beliefs about returns that were greater than the predetermined profit support or entry costs that were greater than 40,000 Ksh (or approx. US$500). Demographic controls include age, gender, marital status, education, digitspan recall score, and land holdings. Personal experience controls include sector-specific years of experience. Parental experience controls include dummy =1 if parent experience in that sector, =0 otherwise. Standard errors clustered at the household level. * significant at 10%; ** significant at 5%; *** significant at 1%.*

Table 7. Nested-logit continued - Parameter estimates for covariates from full model

	Physical Labor	Market Vending	Own Enterprise
Land (acres)	0.021	0.168	0.006
	(0.045)	(0.059)***	(0.077)
Female	-1.751	0.981	-0.131
	(0.231)***	(0.388)**	(0.450)
Female x Married[1]	0.065	-0.379	0.101
	(0.226)	(0.257)	(0.396)
Age	0.131	0.120	0.161
	(0.036)***	(0.049)**	(0.079)**
Age squared	-0.002	-0.002	-0.002
	(0.000)***	(0.001)***	(0.001)**
Years of education	-0.102	0.030	0.052
	(0.030)***	(0.045)	(0.060)
Digitspan recall score	0.005	0.015	0.190
	(0.026)	(0.040)	(0.051)***
Physical labor experience (years)	0.052	-0.009	0.021
	(0.010)***	(0.018)	(0.026)
Market vending experience (years)	-0.043	0.139	-0.051
	(0.030)	(0.023)***	(0.052)
Own enterprise experience (years)	-0.012	-0.016	0.069
	(0.017)	(0.025)	(0.019)***
Parent was/is market vendor	0.040	0.073	0.050
	(0.028)	(0.030)**	(0.036)
Parent owns(ed) own enterprise	0.160	0.172	0.068
	(0.085)*	(0.118)	(0.152)
Observations	18918	18918	18918
N	1156	1156	1156

*Coefficients and standard errors presented in this table are estimates from the full nested-logit specification and correspond to the model in Table 6 column 5. Standard errors clustered at the household level. * significant at 10%; ** significant at 5%; *** significant at 1%.*

[1] *All men in the sample are married.*

Table 7a. Nested-logit continued - parameter estimates for covariates from full model using trimmed data

	Physical Labor	Market Vending	Own Enterprise
Land (acres)	0.023	0.170	-0.108
	(0.048)	(0.062)***	(0.101)
Female	-1.747	0.878	0.119
	(0.234)***	(0.408)**	(0.512)
Female x Married[1]	0.071	-0.324	0.191
	(0.229)	(0.264)	(0.450)
Age	0.128	0.116	0.136
	(0.036)***	(0.051)**	(0.089)
Age squared	-0.002	-0.002	-0.002
	(0.000)***	(0.001)***	(0.001)*
Years of education	-0.106	0.024	0.119
	(0.031)***	(0.047)	(0.066)*
Digitspan recall score	0.000	0.019	0.212
	(0.027)	(0.042)	(0.058)***
Physical labor experience (years)	0.051	-0.010	0.017
	(0.011)***	(0.019)	(0.030)
Market vending experience (years)	-0.039	0.153	-0.013
	(0.031)	(0.024)***	(0.046)
Own enterprise experience (years)	-0.012	-0.012	0.076
	(0.016)	(0.027)	(0.023)***
Parent was/is market vendor	0.039	0.066	0.065
	(0.028)	(0.031)**	(0.035)*
Parent owns(ed) own enterprise	0.184	0.183	-0.124
	(0.086)**	(0.123)	(0.183)
Observations	16171	16171	16171
N	1122	1122	1122

*Data: study sample with outliers trimmed. Outliers are defined as beliefs about returns that were greater than the predetermined profit support or entry costs that were greater than 40,000 Ksh (or approx. US$500). Coefficients and standard errors presented in this table are estimates from the full nested-logit specification and correspond to the model in Table 6 column 5. Standard errors clustered at the household level. * significant at 10%; ** significant at 5%; *** significant at 1%.*

[1] *All men in the sample are married.*

Figure 1. Timeline of data collection

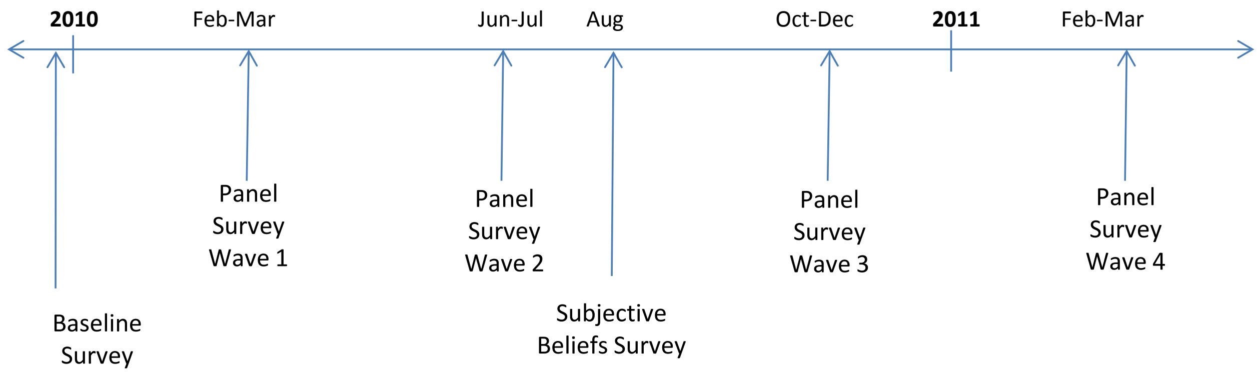
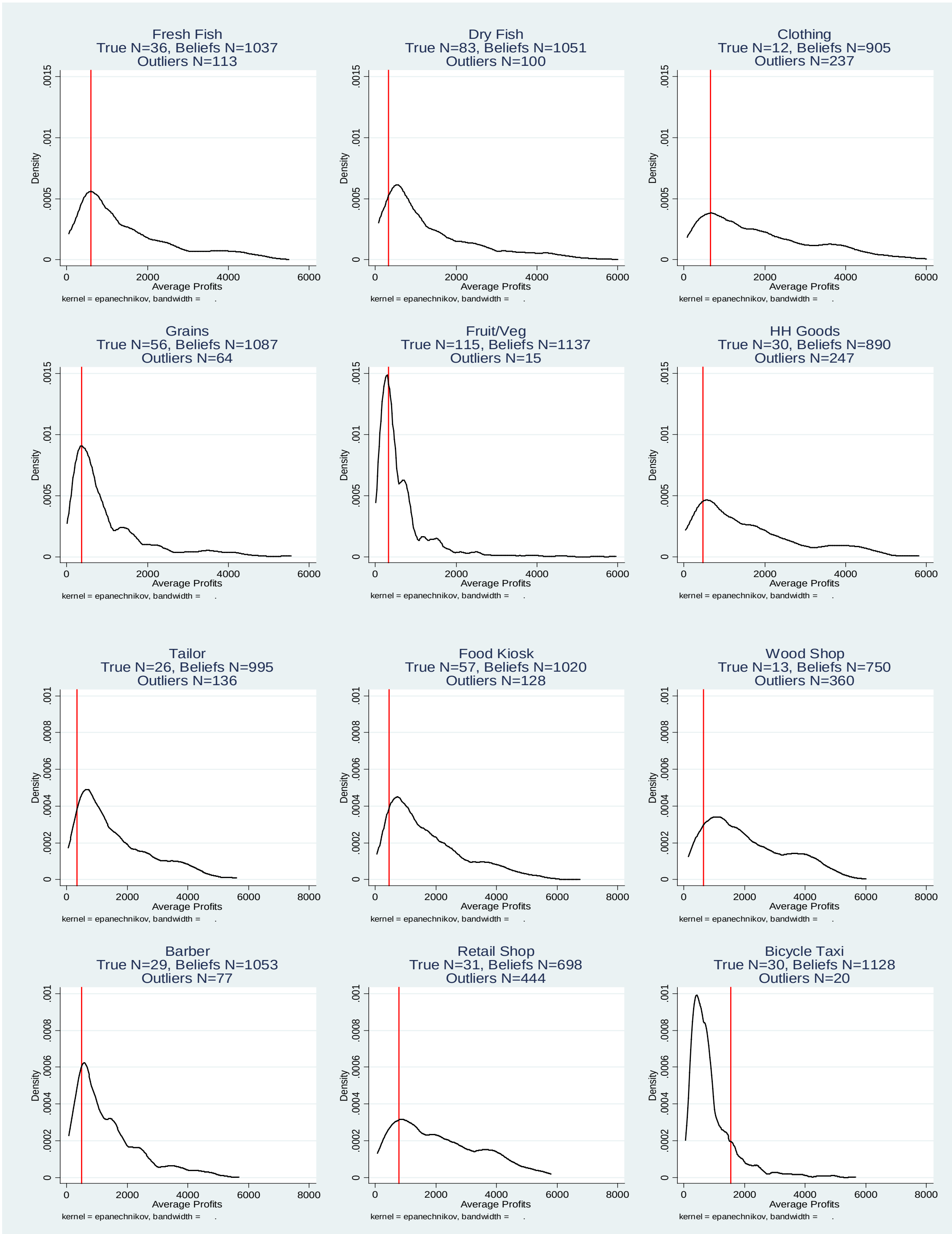
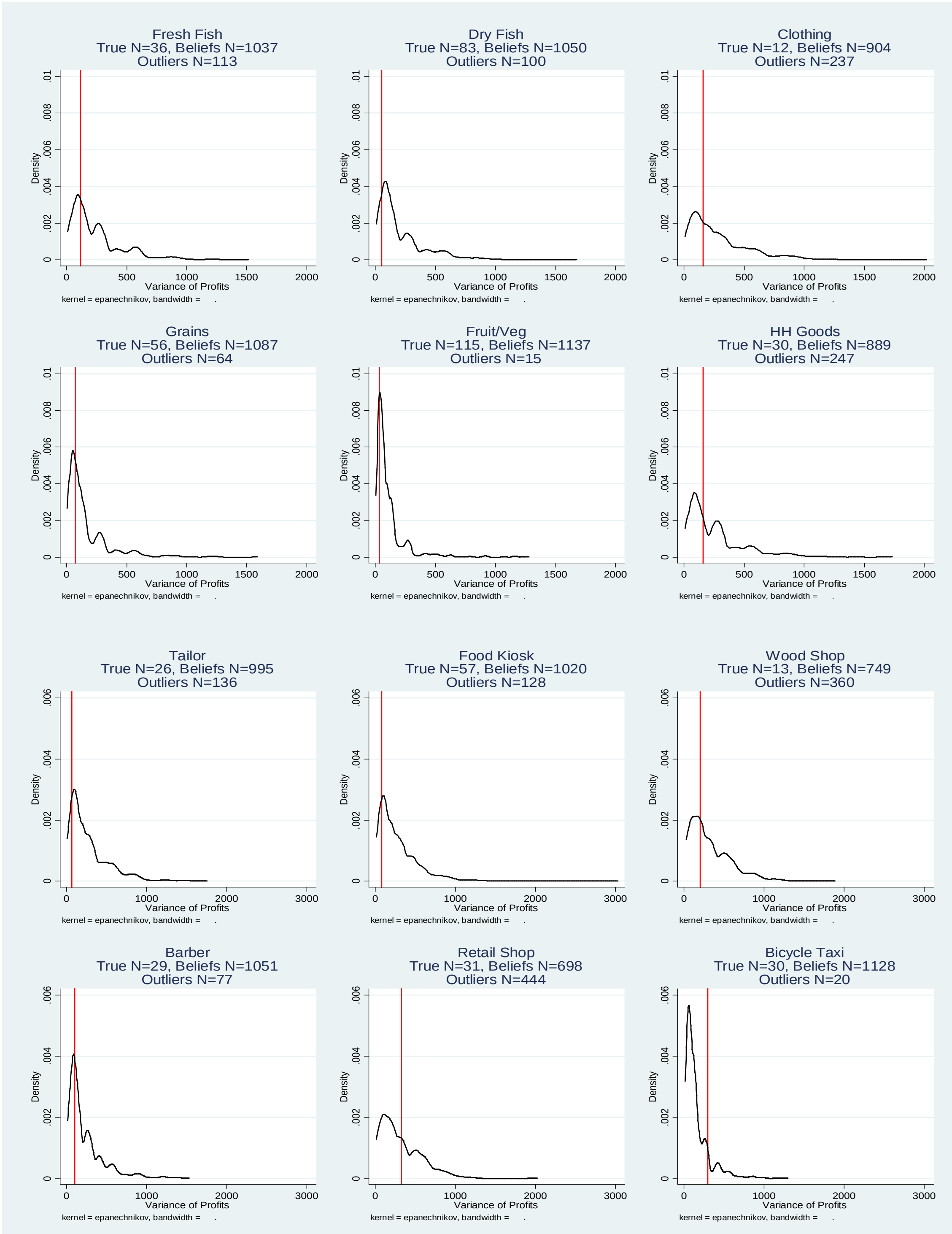


Figure 2. Distribution of beliefs about expected weekly profits and "true" weekly profits



The distributions plot the subjective beliefs data for the "average" person in each occupation.
"True" values are computed using reported profits from the full panel. In each plot, these "true" values are indicated by the verticle line.
Outliers are defined as beliefs that were greater than the predetermined profit support. They are not included in these plots.

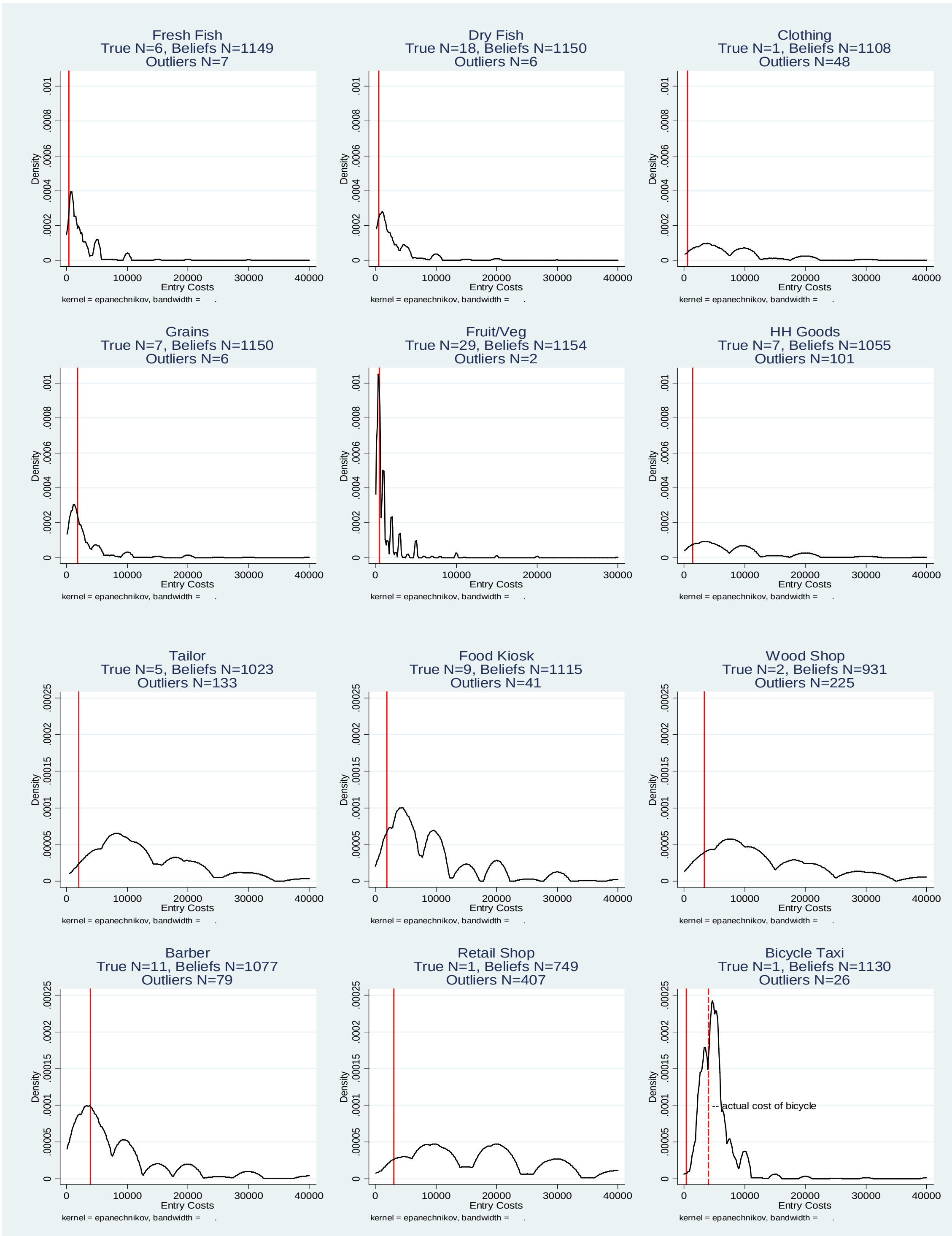
Figure 3. Distribution of beliefs about expected variance of weekly profits and "true" variance of weekly profits



The distributions plot the subjective beliefs data for the "average" person in each occupation. "True" values are computed using reported profits from the full panel. In each plot, these "true" values are indicated by the verticle line.

Outliers are defined as beliefs that were greater than the predetermined profit support. They are not included in these plots.

Figure 4. Distribution of beliefs about expected entry costs and "true" entry costs



The distributions plot the subjective beliefs data for the "average" person in each occupation.
"True" values are computed using reported entry costs from the first wave of the panel suvey. In each plot, these "true" values are indicated by the verticle line.
Outliers are defined as beliefs about entry costs that were greater than 40,000 Ksh (or approx. US$500). They are not included in these plots.

Figure 5. Trade-offs between expected profits and entry costs and risk

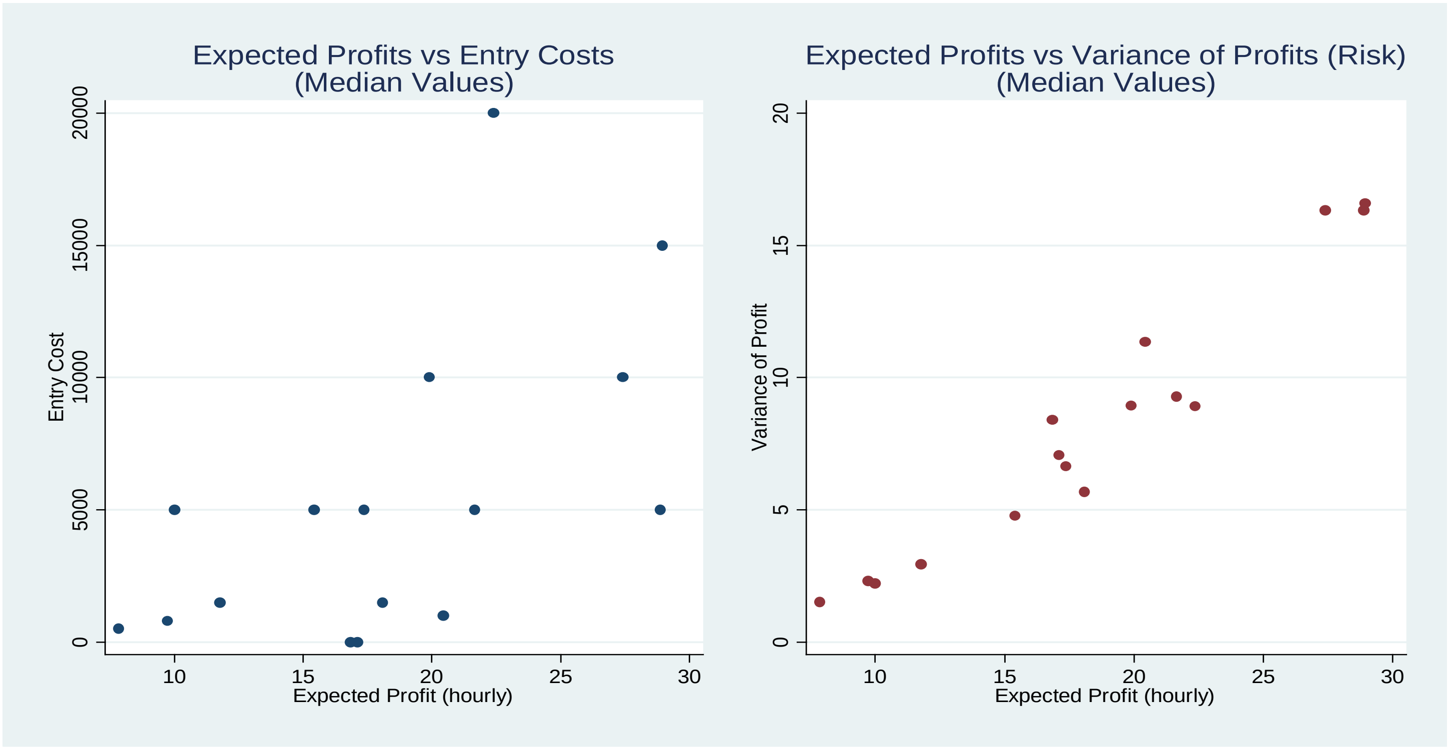
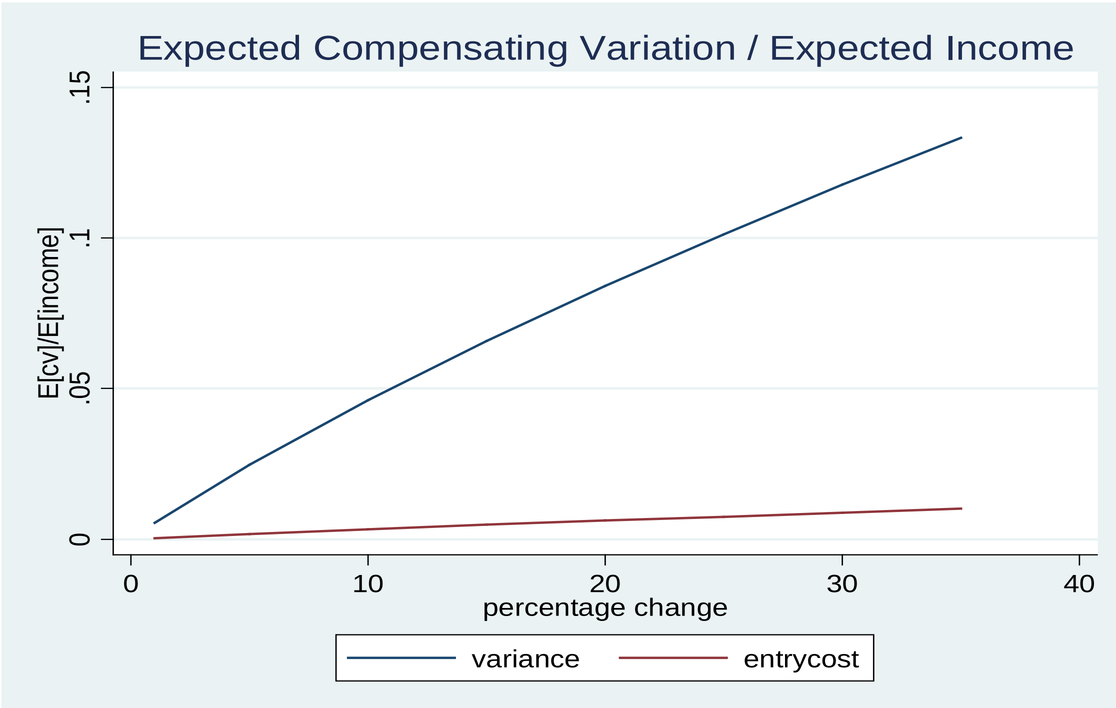


Figure 6. Expected compensating variation as a fraction of expected income



Notes: This graph presents the results of the compensating variation computation for changes to the variance (blue) and changes to the entry costs (red) of the non-own-farm occupations. The x-axis corresponds to percentage changes to the original values for the variance and entry costs. The y-axis corresponds to the ratio of the computed average expected variation in the new state to average expected earnings in the original state.
Data: full sample.

Table A1. Correlation between spouse occupational sector choices

	Husb farm only	Husb farm + physical labor	Husb farm + market vending	Husb farm + own enterprise
Wife farm only	0.073	0.003	-0.073	-0.124
	(0.16)	(0.96)	(0.17)	(0.02)
Wife farm + physical labor	-0.046	0.063	-0.045	-0.003
	(0.39)	(0.23)	(0.39)	(0.96)
Wife farm + market vending	-0.061	-0.026	0.140	0.093
	(0.25)	(0.62)	(0.01)	(0.08)
Wife farm + own enterprise	0.011	-0.086	0.040	0.150
	(0.84)	(0.10)	(0.44)	(0.00)
Number of households	859	859	859	859

P-values in parentheses.

Data: Wave 3 of panel survey, conducted in Oct/Dec 2010.

Table B1. Census and study sample household characteristics

	(1) Census Sample	(2) Study Sample
Total Household Size	5.83	5.72
	(3.05)	(2.94)
No Male Head	0.31	0.37
No Female Head	0.11	0.03
Polygamous household	0.08	0.00
Number of children	3.38	3.40
	(2.34)	(2.26)
Iron roof at home	0.48	0.45
Cement floor at home	0.17	0.13
HH has cell phone	0.47	0.41
Value of physical assets (in Ksh)	10338	9064
	(9547)	(8068)
Value of animals (in Ksh)	4037	4402
	(9009)	(9369)
Land holdings (acres)	1.90	1.79
	(2.86)	(1.87)
At least one member of household has a bank account	0.20	0.00
Number of households	1898	859

Standard deviations in parentheses. The exchange rate at the time of the study was around 80Ksh to US$1 on average.

Table B2. Census and study sample individual characteristics

	(1) Census Sample	(2) Study Sample
Panel A. Women		
Age	39.46	40.53
	(16.12)	(16.93)
Years of education	6.05	5.28
	(3.89)	(3.58)
Can write in Swahili[1]	0.64	0.57
Primary occupation at baseline:		
Farming	0.72	0.79
Own enterprise	0.19	0.15
Physical labor	0.02	0.02
Employee	0.03	0.00
None	0.05	0.05
ROSCA participant	0.45	0.41
Has a bank account	0.09	0.00
Number of Women	1726	822
Panel B. Men		
Age	41.81	41.49
	(15.32)	(15.40)
Years of education	8.08	7.14
	(3.59)	(3.22)
Can write in Swahili	0.90	0.88
Primary Occupation:		
Farming	0.39	0.54
Own enterprise	0.36	0.27
Physical labor	0.10	0.12
Employee	0.09	0.02
None	0.06	0.04
ROSCA participant	0.36	0.32
Has a bank account	0.20	0.00
Number of Men	1334	334

Standard deviations in parentheses.

[1] I use writing in Swahili as a proxy for literacy because data on being able to read in Swahili is not as complete. However, these results are not substantively different under alternate definitions.