Basis risk and Compound-risk Aversion: Evidence from a WTP Experiment in Mali

Ghada Elabed and Michael R. Carter

Preliminary and Incomplete

Abstract

In this paper, we present a novel way to understand the low uptake of index insurance using the interlinked concepts of ambiguity and compound lottery aversion. We begin our analysis by looking at index insurance from the farmer's perspective, noticing that index insurance is a compound lottery. Specifically, we use the smooth model of ambiguity aversion developed by Klibanoff, Marinacci, and Mukerji (2005) to derive an expression of the willingness to pay to reduce basis risk. Empirically, we implement the WTP measure using framed field experiments with cotton farmers in Southern Mali. In this sample, 57% of the surveyed farmers reveal themselves to be compound-risk averse to various degrees. Using the distributions of compound-risk aversion and risk aversion in this population, we then simulate the impact of basis risk on the demand for an index insurance contract whose structure mimics the structure of an actual index insurance contract distributed in Mali. Compound-risk aversion decreases the demand for index insurance relative to what it would be if individuals had the same degree of risk aversion but were compound-risk neutral. In addition, demand declines more steeply as basis risk increases under compound-risk aversion. Our results highlight the importance of designing contracts with minimal basis risk under compound-risk aversion. This would not only enhance the value and productivity impacts of index insurance, but would also assure that the contracts are popular and have the anticipated impact.

Keywords: Index Insurance, Risk and Uncertainty, Compound Risk, Ambiguity, Field Experiments

1 Introduction

Informal risk mitigation mechanisms in developing countries tend to only insure against idiosyncratic shocks, which affect a single individual, and are therefore costly in terms of forgone income (Alderman and Paxson 1992). Covariate shocks, which affect a group of people, remain widely uninsured in developing countries, making households particularly vulnerable to such risks (Jalan and Ravallion 2001). A growing body of research has produced compelling evidence that uninsured risk impedes economic growth; it leads to a persistence of inefficient traditional agricultural technologies (Morduch 1995) and may thereby contribute

to poverty traps (Dercon and Christiaensen 2011; Carter and Lybbert 2012). Therefore, formal insurance contracts could be a crucial instrument for improving welfare in developing countries.

Index insurance is an example of an innovative financial product designed to insure poor households against shocks. The index is chosen to be some variable that closely correlates with farmers' yields, and an individual farmer receives his/her indemnity if the index falls below a pre-determined strike point. Index insurance overcomes informational problems and reduces transaction costs, and is therefore cheaper than conventional indemnity insurance. This type of insurance can therefore offer coverage for poor, small-scale farmers who are typically excluded from existing formal insurance markets. However, uptake of the product remains unexpectedly low, despite a decade of efforts to promote index insurance as a tool for poverty reduction in developing countries (Gine and Yang 2007, Cole et al. 2010, Boucher and Mullally 2010, Meherette 2009).

This paper hypothesizes and tests a mechanism that can help explain the low uptake rates of index insurance. We begin our analysis by looking at index insurance from the farmer's perspective. Compared to conventional indemnity insurance, index insurance is itself a probabilistic investment: payouts are not perfectly correlated with the farmer's loss. For example, in the case of an area-yield insurance contract, the farmer's yield can be low when the average yield in the area is high, and vice versa. This imperfect correlation is known as *basis risk*. Because of basis risk, index insurance is a compound lottery. The first stage lottery determines the individual farmer's yield, and the second stage lottery determines whether or not the index triggers an indemnity payout. Under the Reduction of Compound Lotteries axiom of expected utility, individuals will reduce this compound lottery to a simple lottery. In this paper, we examine what happens when we relax the assumption that decision makers behave according to the predictions of expected utility theory.

There is a large body of literature on non-expected utility models of decision making under uncertainty, but we focus here on the interrelated concepts of ambiguity and compound risk aversion. Ambiguity aversion was first demonstrated by Ellsberg (1961), who showed that individuals react much more cautiously when choosing among ambiguous risks (with unknown probabilities) than when they choose among risks with known probabilities. While the individual probabilities under index insurance are known, for individuals who cannot computationally reduce a compound lottery to a single lottery, the final probabilities are unknown, as in the Ellsberg experiment. Halvey (2007) corroborates this intuition by experimentally establishing a relationship between ambiguity aversion and compound-risk aversion, showing that those who are ambiguity averse are also compound-risk averse.

Abdellaoui et al.(2011) notes that "(...), behavior towards compound risk is relatively understudied in the literature (...)", and our paper helps fill this gap by quantifying compound-risk attitudes in a previously unexplored context, and by using a framed field experiment to study the implications of ambiguity and compound-risk aversion on the demand for index insurance. To our knowledge, with the exception of the study by Bryan (2010), we are not aware of any other study that looks at the impact of ambiguity attitudes

on index insurance uptake.

Specifically, we use the smooth model of ambiguity aversion developed by Klibanoff, Marinacci, and Mukerji (2005) to derive an expression of the *willingness to pay* to reduce basis risk (WTP). We define this WTP as the maximum amount of money a farmer is willing to pay and still be indifferent between index insurance and the corresponding conventional indemnity insurance contract. We then show how this measure varies with compound-risk aversion.

Empirically, we implement the willingness to pay measure using framed field experiments with cotton farmers in Southern Mali. In this sample, 57% of the surveyed farmers reveal themselves to be compound-risk averse to various degrees. Using the distributions of compound-risk aversion and risk aversion in this population, we then simulate the impact of basis risk on the demand for an index insurance contract whose structure mimics the structure of an actual index insurance contract distributed in Mali.

Compound-risk aversion decreases the demand for index insurance relative to what it would be if individuals had the same degree of risk aversion but were compound-risk neutral. In addition, demand declines more steeply as basis risk increases under compound-risk aversion. Were basis risk as high as 50%, only 35% of the population would demand index insurance, as opposed to the 60% who would be willing to purchase the product if individuals were simply maximizing expected utility. Previous studies have investigated various other factors for the uptake of index insurance such as are lack of financial literacy and exposure to financial markets (Giné et al., 2008); lack of trust (Cole et al. 2010); liquidity constraints and ambiguity aversion. We provide evidence that an unstudied concept, compound-risk, could explain low uptakes of insurance. An implication of this result is the importance of designing contracts with minimal basis risk when farmers are compound-risk averse, to enhance the impacts of index insurance and to assure that farmers actually demand the contracts.

The remainder of the paper is structured as follows. In the next section, we review the relevant literature. We then present the theoretical framework and the derivation of the willingness to pay. In subsequent sections, we describe and present the results of the field experiment in Mali. We conclude with potential policy recommendations.

2 Related Studies

There is ample evidence that people do not behave according to expected utility theory when they face a risky prospect, and most departures from expected utility theory are likely to exacerbate the effects of basis risk. Because of the presence of basis risk, index insurance is a form of probabilistic insurance, a concept introduced by Kahneman and Tversky [1979]. Contrary to the predictions of the expected utility theory, studies of the uptake of probabilistic insurance have found that consumers dislike this type of insurance contract, instead preferring a regular insurance contract that pays with certainty when a loss occurs. Wakker et al. [1997] used survey data to show that the respondents demand about a 30% reduction in the premium to compensate for

a 1% probability of not getting a payment in the case of a loss. Expected utility theory cannot explain these findings. Under reasonable assumptions, an expected utility maximizer would be expected to demand only a 1% decrease in premium to compensate them for the 1% increase in the probability that the insurance contract fails. These observations lead us to wonder about the reasons behind this aversion to probabilistic insurance.

Two main theories explain the attitudes of consumers towards probabilistic insurance: prospect theory, namely the probability weighting function Kahneman and Tversky [1979], Wakker et al. [1997], and rank dependent utility functions Segal [1988]. In this paper, we focus on the inter-related concepts of ambiguity and compound risk aversion. The latter concept was first defined in Abdellaoui et al. [2011], by comparing certainty equivalents between a compound lottery and the equivalent reduced lottery. According to their definition, a decision maker is compound-risk averse (seeking) if the certainty equivalent for the compound lottery is equal (above) the certainty equivalent of the simple lottery. As mentioned in the introduction, even when the individual probabilities under index insurance are given objectively, for individuals who cannot computationally reduce a compound to a single lottery, the final probabilities are unknown.

The relationship between attitudes towards compound lotteries and ambiguity aversion was first established by the recursive non-expected utility model of Segal [1987] who had the novel idea of representing the Ellsberg problem as a compound lottery. In the first stage, the decision maker assigns the probability of getting the various lotteries in the second stage. Halevy [2007] confirmed experimentally the theoretical findings of Segal by demonstrating the existence of a strong link between ambiguity aversion and compound risk attitude. He finds that ambiguity neutral participants are more likely to reduce compound lotteries, behaving according to the expected utility theory. In the contrast, ambiguity averse participants fail to reduce compound objective lotteries.

In our modeling efforts, the smooth model of ambiguity aversion developed by Klibanoff et al. [2005] is most appropriate in our context. Unlike the other existing theories of decision making under ambiguity aversion, this modeling approach relaxes the reduction of compound lottery axiom and allows for a continuous objective function. Therefore, it allows the exploration of the implications of the relationship between ambiguity aversion and compound-risk aversion for the demand for index insurance, and especially the sensitivity of that demand to increases in basis risk.

While the existence of compound-risk aversion or ambiguity aversion is an important finding in and of itself, we further wish to understand the impact of this type of aversion on the demand for index insurance. The literature on this issue is still in its infancy, and tends to focus on the impact of ambiguity on technology adoption. The main assumptions of these recent studies are that new technologies are more uncertain in terms of risk and ambiguity than traditional ones. Therefore, ambiguity averse farmers are expected to be less likely to adopt Engle-Warnick et al. [2007], Ross et al. [2010], Alpizar et al. [2009], Barham et al. [2011]. In the special case of insurance decisions, Bryan [2010] shows that under some theoretical restrictions on the shape of preferences, households that are both ambiguity averse and risk averse will not value insurance

because they perceive it as increasing risk. Based on the results of two experiments in Malawi and Kenya, he finds that ambiguity aversion explains the low uptake of index insurance. The main assumption of his model is that the production function generating the household income is ambiguous (i.e. the relationship between the index and the income). Our work differs from Bryan [2010] in that the ambiguity in our case arises from the payoff structure of the index insurance contract and not the probability distribution of the production function.

3 Conceptual Framework

The goal of this section is to first study the impact of basis risk on the demand for index insurance. We first explore the demand for actuarially unfair index insurance under expected utility maximization. We then note that index insurance appears to the individual as a compound lottery. Compound lottery induces a behavior akin to ambiguity. Then we explore the impact of ambiguity or compound risk aversion on the demand for index insurance, using the model of ambiguity aversion developed by Kilbanof, Marinacci and Mukerji [2005] (referred to as the *KMM model*). This model captures risk preferences by the curvature of the utility of wealth function, and ambiguity preferences by a second-stage utility functional defined over the expected utility of wealth. Central to this analysis is the concept of *generalized uncertainty premium*, calculated by Maccheroni, Marinacci and Ruffino (2010), which we also briefly summarize in the following section.

In a second step, we will present a method that characterizes participants according to their compound-risk aversion. The crux of this method is to give the participant a choice between the index insurance and some equivalent conventional indemnity insurance. From this procedure, we will elicit the *willingness to pay to reduce basis risk*, that is, the maximum amount of money a farmer is willing to pay and still be indifferent between the index insurance and the conventional indemnity insurance contract.

3.1 Index insurance from the farmer's perspective

According to standard economic theory, an expected utility maximizer faced with an actuarially fair insurance contract will insure the entire amount at risk. If the risk can only be partially insured (as with an index insurance contract), the expected utility maximizing agent will still purchase whatever partial insurance is available if priced at an actuarially fair level. In a more realistic setting, however, insurance companies impose loadings to cover transaction costs. In that case, standard economic theory predicts that a utility maximizer will leave part of the risk uninsured. Index insurance contracts are an example of partial insurance, and typically have a loading of 20%. Therefore, a risk averse agent will purchase index insurance only if basis risk is small enough compared to the fraction of total risk he is exposed to. Moreover, the more risk averse he is, the higher the amount he insures. However, what happens when the individual is not an expected utility maximizer?

In order to understand the behavior of the farmer, let's first present the payoff structure under an index insurance contract. Consider a farmer who owns one hectare of land and has no sources of income other than what he produces on that land.

Figure 1 below reveals the payoff structure under an example of an index insurance contract. Under this structure, the individual farmer faces, for example, a probability p that yields are good, and with a probability $1-p$ that he incurs a yield loss L. If the individual yields are good, there is a probability q_1 strictly less than 1 that the index insurance triggers a payoff, resulting in an income of $(Y_0 - \tau_1 + \Pi)$ equal to the net income under good yields, less the insurance premium plus the value of the insurance indemnity payment. However, there is a probability $1-q_1$ that the index is not triggered. In that case, no insurance payments are made and the individual receives an income equal to the net income under good yields less the insurance premium $(Y_0 - \tau_1)$. If the individual experiences poor yields, there is a probability q_2 strictly less than 1 that the index insurance will trigger a payoff, resulting in an income of $(Y_0 - L - \tau_1 + \Pi)$ equal to the net income under bad yields, less the insurance premium plus the value of the insurance indemnity payment. However, there is a probability $1-q_2$ (basis risk) that conditional on poor yields, the insurance contract fails to payoff. In this case, the individual receives a net income of $Y_0 - L - \tau_1$(equal to the net income under bad yields minus the insurance premium).

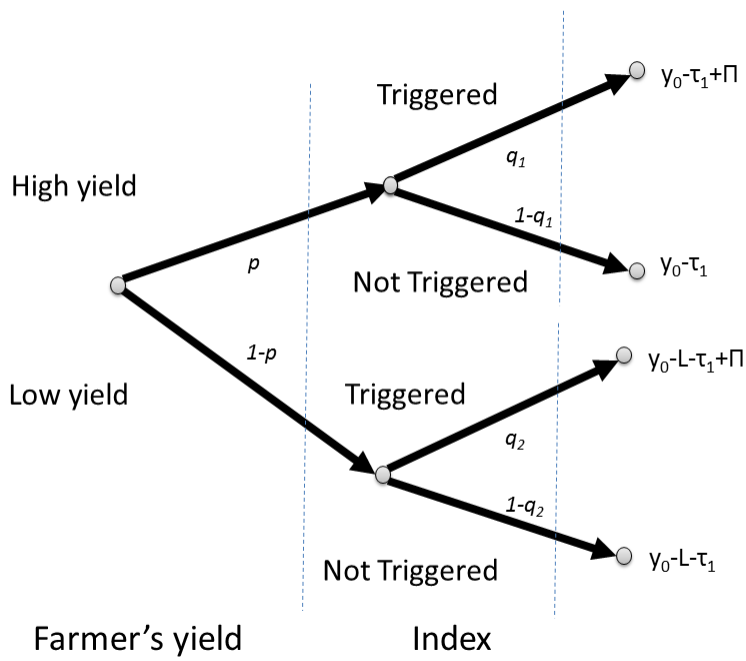


Figure 1: Index insurance from the farmer's point of view.

In the more general case of multiple states of the nature, let's define two marginal distributions: f_y and f_X are the respective pdfs of the farmer's yield and the index X. Let y_{IX} denote the final wealth of the

farmer after all payments are made under the index insurance contract.

While index insurance is effectively a compound lottery, under the Reduction of Compound Lotteries axiom of expected utility theory (ROCL), the individual will reduce the compound lottery to a simple lottery. Assuming that the individual's risk preferences are captured by the utility function u defined over final wealth, and assuming that the farmer is risk averse by imposing concavity of u (u is of course also increasing), the objective function of an expected utility maximizer is the following:

$$\mathbb{E}_{f_{yX}}\left[u\left(y_{IX}\right)\right] \tag{1}$$

where f_{yX} is the joint probability distribution function of the index and the yield.

However, as explained in the prior section, compound lotteries create something akin to ambiguity. While the individual probabilities under the index insurance are known, for poor farmers who cannot computationally reduce a compound lottery to a single lottery, the final probabilities are in fact unknown, as in the Ellsberg experiment.

Under the KMM model of smooth ambiguity aversion, for each realization of the index, the farmer's expected utility is evaluated by an increasing function v that captures compound risk preferences, and the farmer's objective function is the expected value of v given the probability distribution of the yield. Thus, the farmer's objective function is given by:

$$\mathbb{E}_{f_y}\left[v\left(\mathbb{E}_{f_{X/y}}u\left(y_{IX}\right)\right)\right] \tag{2}$$

where $\mathbb{E}_{f_y}$denotes the expectation with respect to f_y . The expectation $\mathbb{E}_{f_{X/y}}$ is taken with respect to $f_{y/x}$, the probability distribution function of the index conditional on the realization of the yield. As risk aversion is imposed by the concavity of u, under the KMM model, compound-risk aversion is obtained by imposing concavity of v: $v^{'} > 0$ and $v^{''} \leq 0$. In the compound-risk neutral case (i.e., v is linear), this expression reduces to the conventional Von Neumann-Morgenstern expected utility maximization represented by Equation 1.

3.2 The demand for index insurance

Having presented the objective functions with and without compound-risk aversion, we look now at the demand for index insurance under these two models of decision making. Central to the analysis in this section is the concept of *uncertainty premium* derived by Maccheroni, Marinacci and Ruffino [2010] (hereafter MMR), who extended the classic result of Pratt [1964] on the risk premium under uncertainty to the ambiguity framework assuming the KMM model.. In the case of index insurance, this uncertainty premium ρ_X is defined such that the farmer is indifferent between receiving the net revenue from the index insurance contract and the certain average revenue $\overline{y}_{IX}$. This premium solves the following equation:

$$\mathbb{E}_{f_y}\left[v\left(\mathbb{E}_{f_{X/y}}u\left(y_{IX}\right)\right)\right] = v\left(u\left(\bar{y}_{IX} - \rho_X\right)\right) \tag{3}$$

Proposition 1. *Call ρ_X^N the uncertainty premium of a compound-risk neutral farmer with a utility function u. The solution of Equation 3, if it exists, is lower bounded by ρ_X^N:*

$$\rho_X \geq \rho_X^N$$

The following proof follows Alary et al. [2012]; we provide it for the sake of completeness.

Proof. Since u is concave, using Jensen's inequality:

$$\begin{aligned} v\left(u\left(\bar{y}_{IX} - \rho_X\right)\right) &= E_f\left[v\left(E_{X/y}u(y_{IX})\right)\right] \\ &\leq v\left(\mathbb{E}_{f_y}E_{f_{y/X}}u(y_{IX})\right) \\ &= v\left(E_{f_{yX}}u(y_{IX})\right) \\ &= v\left(u\left(\bar{y}_{IX} - \rho_X^N\right)\right) \end{aligned}$$

□

Proposition 1 means that compound-risk aversion should decrease the demand for index insurance relative to what it would be if individuals had the same degree of risk aversion but were compound-risk neutral. Intuitively, compound-risk averse agents pay an extra premium to eliminate the ambiguity associated with the secondary lottery.

Now if we consider a compound-risk averse individual, what happens to his demand for index insurance as basis risk increases? The answer to this question is based on the findings of MMR who showed that the uncertainty premium ρ_X is the sum of an ambiguity premium and the classical risk premium:

$$\rho_X \simeq -\frac{1}{2}\sigma^2_{y_{IX}}\frac{u''\left(\bar{y}_{IX}\right)}{u'\left(\bar{y}_{IX}\right)} - \frac{1}{2}\sigma^2_{\bar{y}_{IX}}\frac{v''\left(u\left(\bar{y}_{IX}\right)\right)\left(u'(y_{IX}\right)}{v'\left(u\left(\bar{y}_{IX}\right)\right)} \tag{4}$$

where $\sigma^2_{y_{IX}}$ is the variance of the final net wealth when purchasing the index insurance contract :

$$\sigma^2_{y_{IX}} = \mathbb{E}_{f_y}\left[\mathbb{E}_{f_{X/y}}\left[y - \bar{y}_{IX}\right]^2\right]$$

$\sigma^2_{\bar{y}_{IX}}$ is the variance of the expected revenue under the probability distribution of the index:

$$\sigma^2_{y_{\bar{I}X}} = \mathbb{E}_{f_y}\left[\mathbb{E}_{f_{X/y}}\left[y - \bar{y}_{IX}\right]\right]^2$$

According to MMR, $\sigma^2_{\bar{y}_{IX}}$ reflects the "uncertainty on the expectation of the revenue, due to the model

uncertainty that the decision maker perceives". In the particular case of index insurance, this expected value varies because conditional on every yield realization, the distribution of the index yields a different expected wealth.

From Equation 4, it is straightforward to show that:

1. For a compound-risk neutral individual, the uncertainty premium reduces to the classical Pratt premium,
$$\rho_X^N \approx -\frac{1}{2}\sigma_{y_{IX}}^2 \frac{u''(\bar{y}_{IX})}{u'(\bar{y}_{IX})}$$

2. For a conventional indemnity insurance, the uncertainty premium also reduces to the classical Pratt premium, whether the farmer is compound-risk averse or not.

3. The results of proposition1 is verified: a compound-risk averse individual is willing to pay an extra premium to reduce basis risk compared to his compound-risk neutral counterpart, who has the same level of risk aversion. This extra premium is denoted ρ^c, and it is a function of the curvature of v, u and $\sigma_{y_{\bar{I}X}}^2$.

Proposition 2 states the impact of basis risk on the demand for index insurance.

Proposition 2. *As basis risk increases, the demand for the index insurance contract decreases for a compound-risk averse participant. This decrease in demand is higher than under expected utility theory.*

The next section describes a methodology to characterize the compound-risk attitudes of the participants.

3.3 Deriving a test to separate farmers by compound-risk attitudes

In order to classify the farmers by compound-risk attitudes, let us imagine the situation where a farmer has to choose between the index insurance contract and a conventional indemnity insurance contract. This latter contract yields a net wealth y_I and pays for sure when the farmer's yield is low.

From the farmer's point of view, the conventional indemnity insurance contract is a simple lottery. Therefore, when faced with this contract, his decision criterion reduces to the usual VMN objective function:

$$\mathbb{E}_{f_y}\left[u\left(y_I\right)\right] \tag{5}$$

Applying the increasing transformation v does not change the previous program, but has the advantage of making Equations 5 and 2 comparable. Therefore, the utility function of the farmer when facing the indemnity insurance is the following:

$$v\left(\mathbb{E}_{f_y}\left[u\left(y_I\right)\right]\right) \tag{6}$$

Suppose the farmer is facing the decision to purchase either the index insurance contract, or the individual insurance contract. What is the amount of money that makes the farmer indifferent between the two

contracts? By definition, this WTP w is the maximum amount of money the farmer is willing to give up in order to be indifferent between the index insurance contract, and the individual insurance contract. This WTP solves the following equation:

$$\mathbb{E}_{f_y}\left[v\left(\mathbb{E}_{f_{X/y}}u\left(y_{IX}\right)\right)\right] = v\left(\mathbb{E}_{f_y}u\left(y_I - w\right)\right)$$

Proposition 3. *If the farmer is compound-risk neural then his WTP w^N is the difference between the certainty equivalents of the two contracts :*

$$w^N = \bar{y_I} - \bar{y}_{IX} + \rho_X^N - \rho_I$$

Proof. If the farmer is compound-risk neural, then:

$$v\left(\mathbb{E}_{f_{X_y}}u\left(y_{IX}\right)\right) \quad = \quad v\left(\mathbb{E}_{f_y}u\left(y_I - w^N\right)\right)$$

By the definition of the certainty premia ρ_X and ρ_I, we have:

$$u\left(\bar{y}_{IX} - \rho_X^N\right) = u\left(\bar{y}_I - \rho_I - w^N\right)$$

where ρ_I is the regular Pratt uncertainty premium for the individual insurance contract:

$$\rho_I \approx -\frac{1}{2}\sigma_{y_I}^2 \frac{u''\left(\bar{y}_I\right)}{u'\left(\bar{y}_I\right)}$$

□

Proposition 4. *A compound-risk averse individual has a higher WTP compared to his compound-risk neutral counterpart, for the same level of risk aversion:*

$$w \geq w^N$$

Proof. If the farmer is compound-risk averse, then his WTP w satisfies the following equations, by Jensen inequality

$$\begin{aligned}
\mathbb{E}_{f_y}\left[v\left(\mathbb{E}_{f_{X/y}}u\left(y_{IX}\right)\right)\right] &= v\left(\mathbb{E}_{f_y}u\left(y_I - w\right)\right) \\
&\leq v\left(\mathbb{E}_{f_{yX}}u\left(y_{IX}\right)\right) \\
&= v\left(\mathbb{E}_{f_y}u\left(y_I - w^N\right)\right)
\end{aligned}$$

□

4 Experimental Design and Data

To test these hypothesis, 331 cotton farmers from 34 cotton cooperatives in Bougouni, Mali participated in a set of framed field experiments. A first game allowed the measurement of their risk aversion coefficients. A second game elicited their WTP as defined above, which allows the elicitation of compound-risk aversion attitudes. This last game closely resembles the theoretical framework described in Section 2 with one difference. If the individual yield is high, the index is no longer triggered. The reason is to mimic the structure of an area yield index insurance product that was designed as part of the ongoing project "Index insurance for Cotton farmers in Mali", and launched by the Index Insurance Innovation Initiative (I4). More details about this project and the structure of the distributed contract can be found in Elabed et al. 2013.

4.1 Experimental Procedure

The participants are 331 members of 34 cotton cooperatives selected at random from the list of cooperatives participating in the project mentioned above. In addition, a survey gathered detailed information on various socio-economic characteristics of the participating farmers such as demographic characteristics, wealth, assets owned, agricultural production and shocks. Data collection for the survey took place in December 2011 through January 2012, and the experiments took place in January and February 2012.

Three rural area animators translated the experimental protocol from French to Bambara, the local language, and ensured that it is accessible to a typical cotton farmer. Game trials were conducted with graduate students in Davis, CA, and with high school students and cotton farmers who were not part of the final experimental sample in Bougouni, Mali. Local leaders (secretaries of cotton cooperatives and/or village chiefs) assisted us in recruiting the eligible participants from a list of names that we provided.

The sessions took place in a classroom on weekends and in the village chief's office on weekdays. The sessions took place with members of the same cooperative, and they lasted around two and a half hours, not counting the time necessary to gather or to distribute of gains at the end of each session. We divided the sessions into two parts with a short break between each. Each participant played one pure luck game and four decision and luck games. Each decision and luck game started with a set of six "low stakes" rounds aimed at familiarizing them with the rules, which were followed by a set of six "high stakes" rounds. The only difference between these two types of rounds was the exchange rate used to compute the gains in cash: the gains from a high stake round were 5 times higher than the gains from a low stake round. At the end of the session, we paid the players for only one of the low stake rounds and one of the high stake rounds of every game by having a farmer roll a six-sided die. We used this random incentive device in order to encourage the players to choose carefully. The animator announced the selection procedure to the players at the beginning of every game. In order to incentivize the players to think more carefully about their decisions, we repeated

the following sentence "There is no right or wrong answer. You should do what you think is best for you and your family whether it is choice #1, choice #2, etc.".

At the end of the session, participants received their game winnings in cash, in addition to a show up fee of 100 CFA. Minimum and maximum earnings, excluding show up fee, were 85 CFA and 2720 CFA and mean earnings was 1905 CFA. The daily wage for a male farm labor in the areas we ran the experiments were between 500 CFA (0. 93 USD) and 2000 CFA (3.75 USD) and on average 1040 CFA (1.95 USD). Since literacy rates are very low in the area, we presented the games orally with the help of many visual aids. In addition to the main animator, two rural animators assisted the players with the various materials.

4.2 The Games

The players had to take decisions framed in terms most familiar to them: their decisions were centered on cotton-their main cash crop. Before playing the risk aversion game, the participants learned how to determine their yields and the resulting revenue. Then participants, endowed with one "hectare of land", had to choose among different insurance contracts.

4.2.1 Determining the yield:

Based on historical yield distributions and pooling all the available data across years and cooperatives, we discretized the density of cotton yields into six sections with the following probabilities (in percent): 5, 5, 5, 10, 25 and 50, respectively. The individual yield values corresponding to the mid-point of those sections are (in kg/ha): 250, 450, 645, 740, 880 and 1530, respectively. Table 1 shows the yield distribution and the corresponding revenue in d, the local currency.

Yield range (kg/ha)	Mid point	Probability	Revenue (in d)
<300	250	5%	2400
300-600	450	5%	10400
600-690	645	5%	18200
690-790	745	10%	22000
790-780	880	25%	27600
>980	1530	50%	53600

Table 1: Yield distribution and corresponding revenues

Understanding the notion of probability associated with the yield determination is a challenge that we addressed by using the randomization procedure used by Galarza and Carter [2011] in Peru to simulate the realizations of the individual yields. Every participating farmer drew his yield realizations from a bag containing 20 blocks (1 black, 1 yellow, 1 red, 2 orange, 5 green and 10 blue) which reproduce the probability distribution mentioned earlier, going from the lowest to the highest yield. Figure 2 shows the visual aid provided to farmers to help them understand the game better. Equation 7 computes the individual farmer's per hectare profits in d without any insurance contract:

$$profit_i = p * y_i - Inputs \tag{7}$$

where the price (p) of a kg of cotton is set at d40, the cost of the inputs is set at d7600 in order to guarantee that the players never incur a real loss in the games with the different contracts.

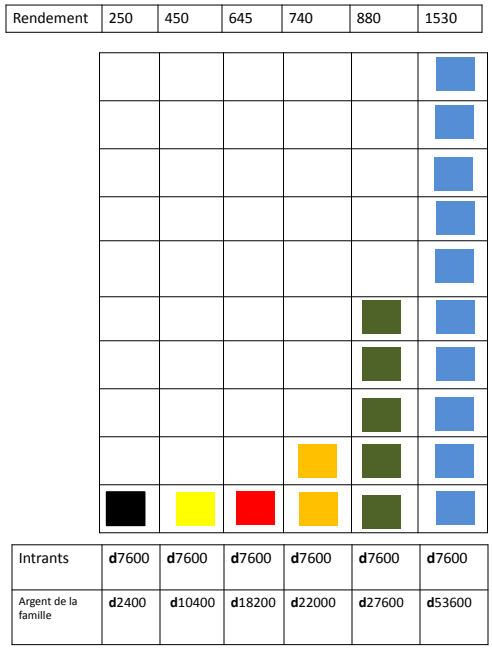


Figure 2: Visual aid for yield distribution

4.2.2 Conventional indemnity insurance contract

After having practiced determining their yields and the corresponding revenue, the player, indexed by i had to decide whether to purchase an insurance contract. The contract is linear and the payment occurs if the yield falls below the strike point T. In case the farmer is eligible for an insurance payment, the insurance reimburses the difference between the individual yield and the strike point such that the farmer is guaranteed to have an income corresponding to yield T. The premium is set to include a loading cost of 20%, such that the amount paid is 120% the amount received on average. Thus, the payment schedule is the following:

$$payment(y_i) = \begin{cases} p * (T - y_i), & y_i \leq T \\ 0 & y_i > T \end{cases} \tag{8}$$

4.2.3 The index insurance contract

The index insurance contract is characterized by a strike point T at the individual level, and by another strike point T_z at the ZPA (aggregate agricultural area) level. Every participant farmer was explicitly told

that he represents a separate agricultural production area in order to emphasize the fact that the index is independent from the realizations of the other farmers in the group. Thus, compared to the regular indemnity insurance, in order to be eligible for a payment, the farmer has to satisfy an extra condition. The payment schedule is the following:

$$payment(y_i) = \begin{cases} p * (T - y_i) : & y_i \leq T \text{ and } y_z \leq T_z \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

Thus, from the player's point of view, once he suffers a loss (i.e. his yield is below the individual strike point), he risks not getting a payment with positive probability. Based on historical data from the area, this probability is set at 20%. Further, the individual-level trigger is set at 70% of the median historical yield, and the contract was priced with a loading cost of 20%. If a farmer decides to purchase an index insurance contract, then he faces a two-stage game. First, he determines his own yield by drawing a block from the yield sack. Then, if the yield is below the individual strike point, he draws another block from a second sack which contains 4 brown blocks (i.e. the index triggered) and one green block (i.e. the index is not triggered) .

Note that, contrary to the more general model presented in Section 2, the farmer does not get a payment in case his yield is high. This choice is made in order to mimic the area based yield contract that was distributed in the area to cotton cooperatives.

4.2.4 Game 1: Eliciting risk preferences

The risk aversion game was framed in terms of an insurance decision to elicit risk preferences. While alternative unframed methodologies exist in the literature, this framed design is chosen for pedagogical reasons. Each subject had six different possibilities: don't purchase an insurance contract, or choose among five different insurance contracts that differ in their strike points, which were 100%, 80%, 70%, 60%, and 50% of the median historical yield (980 kg/ha). In terms of actual yields, this corresponds to 980 kg/ha, 790 kg/ha, 690 kg/ha, 600 kg/ha, and 300 kg/ha, respectively.

The net revenue of farmer i if he purchases contract j is given by the following formula:

$$profit_{ij} = p * y_i + Indemnity_j - premium_j \tag{10}$$

where *indemnity* is an indicator function for the insurance payment, and *premium* is the premium of the insurance contract. Table 2 shows the different revenues associated with each choice and the corresponding risk aversion ranges.

In this game, each player had to determine whether he wanted to purchase an insurance contract, and if so which one. Then, an assistant asked him to draw a block in order to determine his revenue.

Table 2 shows the characteristics of every contract: trigger defined as a percentage of the median yield

in the region, the premium and the net revenue.

Contract #	Trigger (% ybar)	Premium (d)	Net Profit (d) (d)						CRRA range
Yield (kg/ha)			250	450	645	740	880	1530	
Proba.			5%	5%	5%	10%	25%	50%	
0	0	0	2400	10400	18200	22000	27600	53600	$(\infty; 0.08)$
1	50	600	4280	10280	18080	21880	27480	53480	$(0.08; 0.16)$
2	60	1200	15200	15200	17000	20800	26400	52400	$(0.16; 0.27)$
3	70	1740	18260	18260	18260	20260	25860	52860	$(0.27; 0.36)$
4	80	2700	21300	21300	21300	21300	24900	50900	$(0.36; 0.55)$
5	100	6180	25420	25420	25420	25420	25420	47420	$(0.55; \infty)$

Table 2: Individual insurance contracts and risk aversion coefficient

The last column of Table 2 exhibits the CRRA ranges corresponding to every contract choice, assuming a CRRA utility function. Let's assume that the player chose the third contract. Assuming monotonic preferences, this implies that he preferred this contract to contracts 2 and contract 4. The upper (lower) bounds of the CRRA range is found by equalizing the expected utility that the farmer derives from contract 2 and 3 (3 and 4). In this case, as Table 2 shows, the CRRA range of the player is (0.27; 0.36). Note that as the level of coverage (measured by the trigger as percentage of the median yield increases, the CRRA increases.

4.2.5 Game 2: Eliciting the WTP to reduce basis risk

After having practiced determining his revenue under the index insurance contract, every participant played a game that aimed at eliciting the WTP measure defined above (the amount of money the farmer is willing to pay above the price of the indemnity insurance contract). Specifically, we wanted to see whether the player, whom we call Mr. Toure, preferred the indemnity contract to the index contract as we increase the price of the individual contract from its base price (d1340) to d3540, by increments of 200d.

The elicitation procedure was the following: The animator presented players with the following scenario: Mr. Toure's friend, Mr. Cisse, is going to Bamako (the capital of Mali, 90 miles away). Mr. Toure asks Mr. Cisse to buy an insurance contract for Mr. Toure. Mr. Toure knows that the price of the individual contract can vary depending on the day, but the price of an index contract is always the same. After highlighting the fact that at the price of d1340, it is always more profitable to buy the individual insurance contract, Mr. Toure was asked to tell Mr. Cisse at which price Mr. Toure should switch to favoring the index insurance contract over the individual insurance contract. Thus, by the end of the game, we have the switching price for every player from which we deduce his willingness to pay to reduce basis risk.

The game reduces to ten choices between 10 paired insurance contracts whose net revenues are listed in table 3. Notice that the price of the index insurance contract does not vary, whereas the price of the individual insurance contract increases by d200 as we move down the table.

Index Insurance contract	Indemnity insurance contract	Implied WTP	Implied CRRA under EUT
d1400	d1740	0	$(0; 0.49)$
d1400	d1940	d200	$(0.49; 0.71)$
d1400	d2140	d400	$(0.71; 0.87)$
d1400	d2340	d600	$(0.87; 0.99)$
d1400	d2540	d800	$(0.99; 1.09)$
d1400	d2740	d1000	$(1.09; 1.18)$
d1400	d2940	d1200	$(1.18; 1.25)$
d1400	d3140	d1400	$(1.25; 1.32)$
d1400	d3340	d1600	$(1.32; 1.37)$
d1400	d3540	d1800	$(1.37; +\infty)$

Table 3: Game 2: Eliciting WTP measure.

The last column of Table 3 presents the CRRA ranges implied by the measured WTP if the player behaves according to the predictions of EUT, i.e if he reduces the index insurance compound-lottery to a simple lottery. However, if a participant is compound-risk averse, then the elicited CRRAs are not true.

In order to deduce the compound-risk aversion of a player, we impose a functional form on the function v we defined earlier. For computational convenience, we impose constant relative compound risk aversion. Thus, the function v defined in Section 2 is given by:

$$v(y) = \begin{cases} \frac{y^{1-g}}{1-g} & \text{if } g \in [0,1) \\ \log(y) & \text{if } g = 1 \end{cases} \tag{11}$$

where g is the coefficient of constant relative compound-risk aversion.

Table 4 below lists the predicted coefficients of compound-risk aversion based on the player's choices in Games 1 and 2. To simplify the calculations, these measures are made after taking the midpoint of every risk aversion range. For example, if the player chose contract 4 in Game 1, then the corresponding CRRA is 0.45. The corresponding g is obtained by equalizing equations 2 and 6.

	Contract choice in Game 1:					
WTP (d)	0	1	2	3	4	5
0	0.01	0.00	0.00	0.00	0.00	0.00
200	0.08	0.07	0.06	0.05	0.01	0.00
400	0.14	0.14	0.14	0.13	0.10	0.00
600	0.21	0.21	0.21	0.21	0.20	0.00
800	0.27	0.28	0.29	0.29	0.29	0.00
1000	0.34	0.35	0.36	0.38	0.39	0.00
1200	0.40	0.42	0.44	0.46	0.48	0.13
1400	0.47	0.49	0.51	0.54	0.58	0.29
1600	0.53	0.56	0.59	0.62	0.67	0.46
1800	0.59	0.62	0.66	0.70	0.76	0.63

Table 4: Predictions of the Coefficients of Compound-Risk Aversion.

5 Descriptive analysis of the experimental results

5.1 Participants characteristics

Table 5 provides the descriptive statistics for the experiment participants. All the participants are male, which is not surprising given the division of labor in the area of study: cotton is a male crop. The average participant is approximately 47 years old, has limited formal education (three years of schooling), and belongs to a household with almost 19 members. 71% of the participants are the head of their households, and almost all of them have heard of the cotton insurance contract distributed in the field. The average household head has been a member in the cooperative for almost 8.6 years. The average household economic status is represented by a total livestock value of 1.8 million CFA, a house worth 400,000 CFA and a total land area of 9.62 ha.

	Variable	Definition	mean	sd/percent
Participant Characteristics	head	1 if the participant is head of household	0.7	
	age	Participant's agent	47.07	13.21
	gender	1 if participant is male	1	
	education	Participant's years of schooling	4.55	6.57
	knowledge_ins	1 if participant heard about cotton insurance before	0.92	
Head Characteristics	age_hh	Head of the household's age	55.55	15.22
	gender_hh	1 if head of the household is male	1	
	coop_years	Number of years of household's head membership in the cotton cooperative	8.62	6.28
Household Characteristics	hh_size	Size of the household	18.82	11.88
	livestock_2012	Value of livestock in CFA	1,822,602	5,634,664
	ag_value	Value of agricultural equipment in CFA	171,299	247,236
	assets_value	Value of household's assets in CFA	204,200	164,468
	house_value	Value of the house in CFA	396,952	1,042,061
	land_owned	Total area of land owned in ha	9.62	7.81

Table 5: Descriptive Statistics of the Participants

5.2 Description of the results of Game 1

The last column of Table 6 below shows the distribution of the levels of CRRA of the participants, based on the results of Game 1. The majority of the farmers (78%) chose an insurance contract, and 30% of them chose the highest level of coverage which corresponds to a CRRA level of more than 0.55.

Contract #	CRRA range	%
0	$(\infty; 0.08)$	22.56
1	$(0.08; 0.16)$	7.32
2	$(0.16; 0.27)$	9.76
3	$(0.27; 0.36)$	10.67
4	$(0.36; 0.55)$	17.99
5	$(0.55; \infty)$	31.71

Table 6: Distribution of the CRRAs in the sample

6 Empirical Analysis

6.1 Testing the hypothesis of compound-risk neutrality

As we saw in Proposition 3 of Section 3, a compound-risk averse farmer is willing to pay more money to switch from the index insurance contract to the individual insurance contract, than his compound-risk neutral counterpart who has the same level of risk aversion. Therefore, in order to empirically test the hypothesis that farmers are compound-risk neutral (i.e. expected utility maximizers), one should compare the distribution of the CRRA coefficients elicited from Game 1 (last column of Table 2) to those elicited from Game 2 (last column of Table 3). Games 1 and 2 do not elicit the actual CRRA coefficients; rather they provide CRRA classes that are not directly comparable. Therefore, before performing the hypothesis test, we begin by fitting a continuous probability distribution to the CRRAs elicited from both games.

Instead of conducting an exhaustive search of every possible probability distribution, it is more practical to fit a general class distribution to the data. Ideally, this distribution will be flexible enough to reasonably represent the underlying parameters. This section uses the Beta of the first kind (B1), a three-parameter distribution, as the continuous model that represents the data. The Beta distribution of the first kind is one member of a class of distributions called Generalized Beta distributions (GB), a family of five-parameter distributions that encompasses a number of commonly used distributions (Gamma, Pareto, etc.). The GB is a flexible unimodal distribution and is widely used when modeling bounded continuous outcomes, such as income distributions.

Since the B1 distribution is defined for bounded variables, one should make assumptions about the range of the CRRAs. The participants are assumed to be risk-averse since they are poor cotton farmers from a developing country. We allow the upper bound of the elicited CRRA to be 1.7.

Let $B1(b, p_1, q_1)$ and $B1(b, p_1, q_1)$ be the probability distribution functions of the CRRAs elicited from Game 1 and Game 2 respectively. The parameter b is the upper bound of the CRRAs . The Appendix explains the methodology used to estimate these parameters

Table 7 presents the results of the estimation method:

Game	Game 1	Game 2
First parameter	0.68	2.07
Second parameter	1.99	4.36

Table 7: Estimated parameters of the distribution

We estimate the confidence intervals for the different parameters using the bootstrap method. Table 8 shows the confidence intervals of parameters p_1, q_1, p_2 and q_2 at the 5% significance level, obtained after 10000 simulations. The bootstrap parameters appear to be consistent estimates for the actual parameters.

		mean	[95% conf	Interval]
Game 1	p_1 parameter	0.67	0.63	0.84
	q_1 parameter	1.98	1.80	2.58
Game 2	p_1 parameter	2.07	1.92	2.57
	q_1 parameter	4.37	4.16	5.09

Table 8: Bootstrap confidence intervals for the parameters.

From Figure 3, it is clear that the estimated parameters follow a normal distribution whose mean is close to the observed values. Therefore, the estimation strategy provides a good fit for the data.

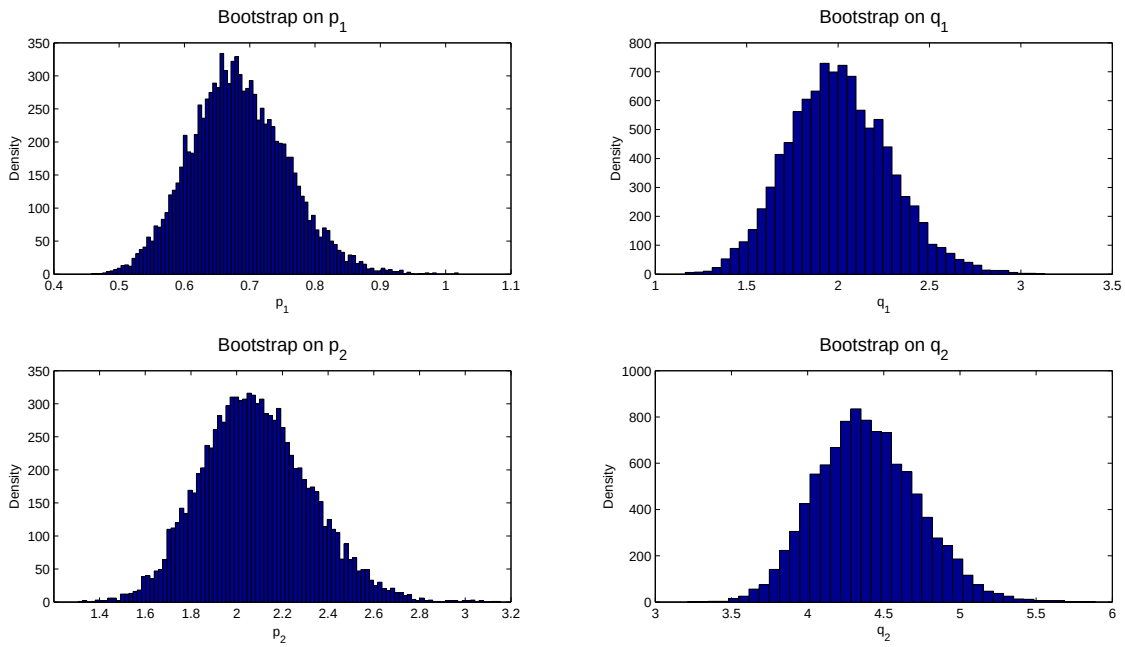


Figure 3: Histogram of bootstrap for parameter p and q.

The test of equality of the distributions of the two CRRAs elicited from the games is performed using 10 000 bootstrapped simulations of the data. We reject the hypothesis that the parameters of the two distributions are the same at the 5% level. Therefore, we reject the hypothesis that the sample of farmers compound-risk neutral.

6.2 Participants have different compound-risk attitudes

Overall, only 40.18% of the participants were indifferent between the index insurance contract and the equivalent individual insurance contract. This supports the hypothesis that basis risk reduces the demand for index insurance. The remaining 60.82% participants have an average WTP of 395d, which represents 22% of the price of the individual insurance contract.

We presented the coefficient of compound-risk aversion for each demonstrated category of WTP in Table 4. Using the Table 4 coefficients of compound-risk aversion, we derive the number of participants who are

compound-risk averse and disaggregate this number by risk aversion range. As shown in Table 9, 57% of the players are compound-risk averse. Furthermore, most of the compound-risk averse farmers are also the least risk averse (22.39%). While the existence of compound-risk aversion is important in and of itself, we will study its impact on the demand for index insurance in the next section.

	CRRA Range							
	0	0.08	0.16	0.27	0.36	0.55	1.7	
Compound-risk averse participants	73	24	32	35	59	103		186
% of CRRA range	*100*	*37.5*	*75.0*	*74.2*	*66.1*	*14.6*		
% of total participants	*22.39*	*2.76*	*7.36*	*7.98*	*11.96*	*4.60*		***57.07***

Table 9: Distribution of Compound- risk Attitudes by CRRA levels.

6.3 Simulating the Impact of Basis Risk

Drawing on the findings of the experiments described above, this section simulates the impact of basis risk on the demand for index insurance under expected utility maximization (equivalently, compound-risk neutrality), and compound-risk aversion. In the following discussion, we assume that the distributions of risk aversion and of compound-risk aversion among the participating farmers reflect the distributions in the overall population.

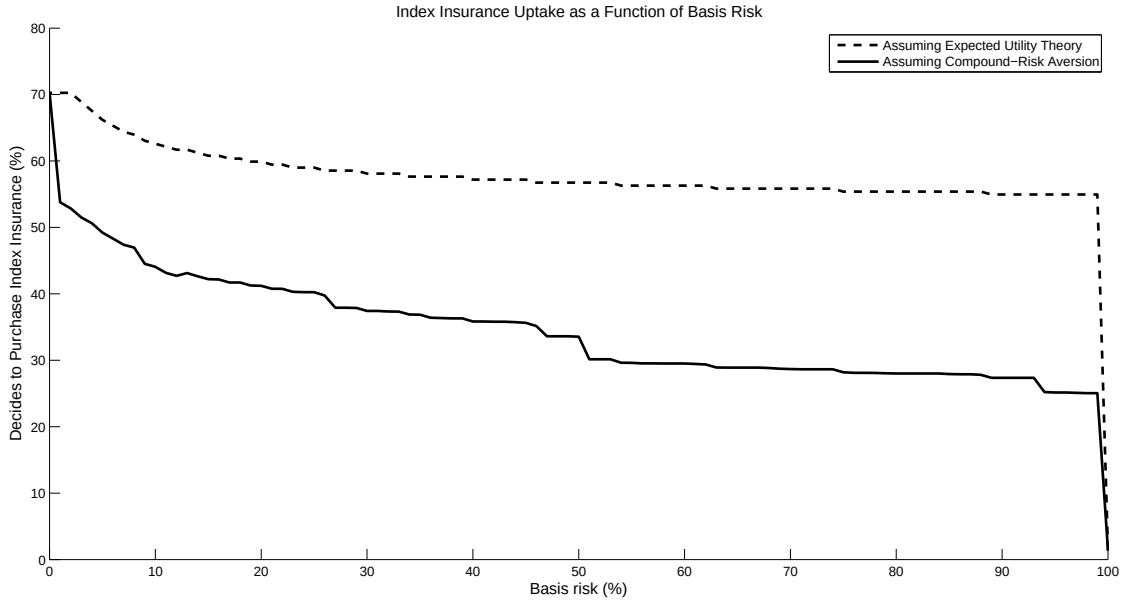


Figure 4: Basis Risk and the Demand for Index Insurance.

The dotted curve of Figure 4 illustrates the impact of basis risk on the demand for index insurance assuming that:

1. Individuals are expected utility maximizers,

2. The price of index insurance is 20% above the actuarially fair price, and

3. The distribution of risk aversion in the population of farmers matches the distribution revealed by the experimental games played in Mali.

Here, basis risk is the probability of not getting a payment conditional on the farmer experiencing a loss. As the basis risk increases under this contract structure, the probability of a payout decreases, and the price of the insurance contract in turn declines. However, because the contract is not actuarially fair, a number of agents drop out of the market as basis risk increases. As can be seen in Figure 4 , increasing basis risk in an index insurance contract will discourage demand because it fails to sufficiently reduce the risk of collateral loss. For a contract with zero basis risk, i.e. one that pays off for sure in case of a loss, moderately and highly risk averse farmers (70% of the population in the Mali experiment) ask for index insurance. As basis risk increases, the farmers with the highest risk aversion coefficient are the first to stop demanding the contract. This drop in demand reaches as high as 15% for extremely high levels of basis risk (90%). Despite this decrease in demand, the demand for the partial insurance provided by this index insurance contract remains relatively robust even as basis risk increases (assuming that individuals maximize expected utility).

Basis risk matters even more when people are compound-risk averse. Using the distribution of compound-risk aversion in the population of the farmers, the solid line in Figure 4shows the impact of basis risk on demand for index insurance. As expected, compound-risk aversion decreases the demand for index insurance relative to what it would be if individuals had the same degree of risk aversion but were compound-risk neutral. In addition, as can be seen in the figure, demand declines more steeply as basis risk increases under compound-risk aversion. Were basis risk as high as 50% (a not unreasonably high number given the kind of rainfall index insurance contracts that have been utilized in a number of pilots), demand would be expected to be only 35% of the population as opposed to the 60% predictions implied by EUT. In short, under compound-risk aversion, designing contracts with minimal basis risk is important, not only to enhance the value and productivity impacts of index insurance, but also to assure that the contracts are demanded.

7 Conclusion

In the absence of traditional insurance markets, poor households in developing countries rely on costly risk-managing mechanisms. Although index insurance provides a good alternative to these households in theory, demand has been surprisingly low. In this paper, we presented a novel way to understand these low uptake rates, using the interlinked concepts of *ambiguity* and *compound lottery aversion*.

In a framed field experiments conducted with cotton farmers in Bougouni, Mali we elicited the coefficients of risk-aversion and the WTP measure, and we derived the compound-risk aversion coefficients of the farmers. Individuals generally did not behave in accordance with expected utility theory. Instead we observed 57% of game participants revealed themselves to be compound-risk averse to varying degrees. In fact, the willingness

to pay of those individuals who demand index insurance is on average considerably higher than the predictions of expected utility theory.

Using the distribution of compound risk aversion and risk aversion in this population, we simulated the impact of basis risk on the demand for index insurance. As we expected we found that compound risk aversion decreases the demand for index insurance relative to what it would be if individuals had the same degree of risk aversion but were compound-risk neutral. In addition demand declines more steeply as basis risk increases under compound-risk aversion.

Our results highlight the importance of designing contracts with minimal basis risk under compound-risk aversion. This would not only enhance the value and productivity impacts of index insurance, but would also assure that the contracts are popular and have the anticipated impact.

References

M. Abdellaoui, P. Klibanoff, and L. Placido. Ambiguity and compound risk attitudes: an experiment. 2011.

D. Alary, C. Gollier, and N. Treich. The effect of ambiguity aversion on insurance and self-protection. 2012.

H. Alderman and C.H. Paxson. *Do the poor insure?: a synthesis of the literature on risk and consumption in developing countries*, volume 1008. World Bank Publications, 1992.

F. Alpizar, F. Carlsson, and M. Naranjo. *The Effect of Risk, Ambiguity, and Coordination on Farmers' Adaptation to Climate Change: A Framed Field Experiment*. Institutionen för nationalekonomi med statistik, Handelshögskolan vid Göteborgs universitet, 2009.

B. Barham, J.P. Chavas, D. Fitz, V. Rios Salas, and L. Schechter. The roles of risk and ambiguity in technology adoption. 2011.

S. Boucher and C. Mullally. Evaluating the impact of index insurance on cotton farmers in peru. *Technical Notes IDB TN 198, Designing Impact Evaluations for Agricultural Projects, InterAmerican Development Bank (IDB), Washington, DC*, 2010.

G. Bryan. Ambiguity and insurance. *Unpublished Manuscript*, 2010.

M.R. Carter and T.J. Lybbert. Consumption versus asset smoothing: testing the implications of poverty trap theory in burkina faso. *Journal of Development Economics*, 2012. ISSN 0304-3878.

S. Cole, X. Gine, J. Tobacman, P. Topalova, R.M. Townsend, J.I. Vickery, and Harvard Business School. *Barriers to household risk management: evidence from India*. World Bank, 2010.

S. Dercon and L. Christiaensen. Consumption risk, technology adoption and poverty traps: evidence from ethiopia. *Journal of Development Economics*, 96:159–173, 2011. ISSN 0304-3878. 2.

Ghada Elabed, Marc Bellemare, Michael Carter, and Catherine Guirkinger. Managing basis risk with multi-scale index insurance contracts for cotton producers in mali. 2013.

D. Ellsberg. Risk, ambiguity, and the savage axioms. *The Quarterly Journal of Economics*, page 643–669, 1961. ISSN 0033-5533.

J. Engle-Warnick, S.C. Laszlo, J.A. Escobal, and CIRANO. *Ambiguity aversion as a predictor of technology choice: Experimental evidence from Peru*. CIRANO, 2007.

Francisco B. Galarza and Michael R. Carter. Risk preferences and demand for insurance in peru: A field experiment. 2011. URL `http://www.aae.wisc.edu/mwiedc/papers/2011/Galarza_Francisco.pdf`.

X. Gine and D. Yang. Insurance, credit, and technology adoption: Field experimental evidence from malawi. *Policy Research Working Paper*, 4425, 2007.

X. Giné, R. Townsend, and J. Vickery. Patterns of rainfall insurance participation in rural india. *The World Bank Economic Review*, 22:539–566, 2008. ISSN 0258-6770. 3.

Y. Halevy. Ellsberg revisited: An experimental study. *Econometrica*, 75:503–536, 2007. ISSN 1468-0262. 2.

J. Jalan and M. Ravallion. Behavioral responses to risk in rural china. *Journal of Development Economics*, 66:23–49, 2001. ISSN 0304-3878. 1.

D. Kahneman and A. Tversky. Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, page 263–291, 1979. ISSN 0012-9682.

P. Klibanoff, M. Marinacci, and S. Mukerji. A smooth model of decision making under ambiguity. *Econometrica*, 73:1849–1892, 2005. ISSN 1468-0262. 6.

F. Maccheroni, M. Marinacci, and D. Ruffino. Alpha as ambiguity: Robust mean-variance portfolio analysis. *Working Papers*, 2010.

J.B. McDonald and Y.J. Xu. A generalization of the beta distribution with applications. *Journal of Econometrics*, 66:133–152, 1995. ISSN 0304-4076. 1.

E. Meherette. *Providing weather index and indemnity insurance in Ethiopia*. International Food Policy Research Institute (IFPRI), 2009.

J. Morduch. Income smoothing and consumption smoothing. *The journal of economic perspectives*, 9: 103–114, 1995. ISSN 0895-3309. 3.

J.W. Pratt. Risk aversion in the small and in the large. *Econometrica: Journal of the Econometric Society*, page 122–136, 1964. ISSN 0012-9682.

N. Ross, P. Santos, and T. Capon. Risk, ambiguity and the adoption of new technologies: Experimental evidence from a developing economy. *Unpublished Manuscript*, 2010.

U. Segal. The ellsberg paradox and risk aversion: An anticipated utility approach. *International Economic Review*, 28:175–202, 1987. ISSN 0020-6598. 1.

U. Segal. Probabilistic insurance and anticipated utility. *Journal of Risk and Insurance*, page 287–297, 1988. ISSN 0022-4367.

P. Wakker, R. Thaler, and A. Tversky. Probabilistic insurance. *Journal of Risk and Uncertainty*, 15:7–28, 1997. ISSN 0895-5646. 1.

A Appendix: Fitting a B1 distribution to the CRRA

In this section, we estimate the probability density function f of the coefficient of constant relative risk aversion r we elicited from an experiment.

We use Maximum Likelihood estimation assuming that r follows a Generalized Beta distribution of first kind (GB1). The GB1 distribution is defined by the following pdf:

$$f\left(r;b,p,q\right)=\frac{\left(r^{p-1}\left(1-\frac{r}{b}\right)^{q-1}\right)}{b^{p}B\left(p,q\right)}$$

for $0<r<b$ where b, p and q are positive. The scaling factor $B\left(p,q\right)$ is the Beta function:$B\left(p,q\right)=\frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$ where $\Gamma\left(p\right)=\left(p-1\right)!$.

By construction, our data is paritioned in 6 intervals. Therefore, we do not observe the continuous variable r. Following McDonald and Xu [1995], we obtain the parameters of interest (p and q) using a Maximum Likelihood estimator based on a multinomial with an underlying density $f(r;b,p,q)$ and cumulative function $F\left(r;b,p,q\right)$.

We now derive the log-likelihood function. Let j denote the risk aversion interval $[r_{j,},\, r_{j1}]$. Player i's true risk aversion coefficient r has a probability $p_i=F\left(r_{j+1};a,b,p,q\right)-F\left(r_j;a,b,p,q\right)$ of being in interval j. Denoting m_j the number of observations in interval j, the likelihood function L_N is the joint probability function:

$$\mathbf{L}_N=\prod_{i=1}^{N}p_i$$

Maximizing $\mathrm{L_N}$is equivalent to maximizing the log-likelihood function:

$$\begin{aligned}\mathcal{L}_N(b,p,q) &= \log \mathbf{L}_N(b,p,q) \\ &= \sum_{j=1}^{6} m_j \log(p_j)\end{aligned}$$

Where m_j is the number of observations in the interval $[r_{j,}, r_{j1}]$. The probability p_j of being in that interval is

$$p_j = F(r_{j+1}; a, b, p, q) - F(r_j; a, b, p, q)$$

Since r is a Beta distribution of the first kind, its cumulative F is:

$$\begin{aligned}F(r; b, p, q) &= \int_0^{\frac{r}{b}} \frac{t^{p-1}(1-t)^{q-1}}{B(p,q)} \mathrm{d}t \\ &= I_{\left(\frac{r}{b}\right)(p,q)}\end{aligned}$$

where $I_{\left(\frac{r}{b}\right)(p,q)}$ the regular beta function is the cumulative distribution function of the Beta variable with parameters $p$and q evaluated at $\frac{r}{b}$.

Proof. By definition:

$$F(r; a, b, pq) = \int_0^r \frac{t^{p-1}\left(1-\frac{r}{b}\right)^{q-1}}{b^p B(p,q)} dt$$

using the change of variable $x = \frac{t}{b}$, we obtain the result. □

Moral Hazard, Risks and Index Insurance in the Rural Credit Market: A Framed Field Experiment in China

Lan Cheng

01/19/2013

Abstract

This paper explores how the availability of index insurance affects borrowers' moral hazard behavior in the rural credit market using a framed field experiment with 450 Chinese farmers. The experiment focuses on one specific type of moral hazard: credit diversion, which occurs when borrowers violate loan contracts and use production loans for consumption purpose. This paper builds a theoretical model showing that, first, credit diverters would be self-selected out of index insurance market; second, index insurance can mitigate credit diversion by reducing the fluctuation and increasing the expected level of consumption under full investment. Coinciding with theoretical predictions, experimental results show that index insurance reduces the number of credit diverters from 15.6% to 3.8% of the overall sample. The treatment effect is heterogeneous across farmers depending on their risk preferences and ethical costs associated with violating loan contracts. In addition, index insurance reduces the number of risk-rationed farmers from 21.3% to 10.9%, and the heterogeneity of treatment effect on risk rationing only depends on farmers' risk preferences. The theoretical and empirical results have important policy implications, which suggest that index insurance can be substituted for collateral requirements and thus lessen both quantity and risk rationing.

1. Introduction

Asymmetric information together with risks has hampered the development of the rural credit market in developing countries (Udry and Conning, 2007; Ghosh et. al, 2000). One of the most prominent problems due to asymmetry information is ex-ante moral hazard caused by hidden information about borrowers' action before the realization of investment profits. The literature has

discussed three types of ex-ante moral hazard. The first type refers to the case when borrowers prefer risky projects to safe projects as discussed by Stigliz and Weiss (1981). The second type is called non-contractible effort, meaning borrowers do not put as much effort as possible to make sure the success of their investment. The third one is credit diversion, which occurs when borrowers violate loan contracts and divert production loans for consumption purpose. This paper focuses on the third type, which is one of the most prominent and pervasive issues in the rural credit market. Credit diversion reduces borrowers' ability to repay loans and results in higher loan default rates. Lenders respond by asking for collateral to control credit diversion. However, collateral requirement causes quantity rationing and risk rationing, as poor households who do not have assets or fear of risking their assets to put as collateral are excluded from the credit market (Boucher et al., 2008). Group liability, a key feature in microfinance, is an alternative to overcome credit diversion through peer monitoring and social collateral. But recent empirical studies show that the popularity of group liability has been diminished in the rural credit market due to problems of free-riding and collision (Giné et al. 2010). It seems neither of the above strategies can effectively mitigate credit diversion without restricting farmers' access to credit. As a result, formal financial institutions limit credit supply to rural areas and smallholder farmers are severely credit constrained. The poor have to forgo income-enhancing and costly investment opportunities due to the lack of capital.

This paper proposes to use index insurance, an innovative risk management instrument, to solve moral hazard problems in the rural credit market. Specifically, we investigate the impact of the availability of index insurance on borrowers' credit diversion behavior using a framed field experiment with Chinese farmers. Index insurance contract is based on a variety of external indices that are correlated with but not influenced by the insured individual's outcome. It overcomes adverse selection and moral hazard problems in the insurance market, which are usually associated with conventional agricultural insurance. Given unique advantages of index insurance and costly consequences of credit rationing, the recent literature has started to pay increasing attention to the interaction between index insurance and the rural credit market. All the discussions focus on how index insurance can reduce the risks in the credit market, either risk rationing on the demand side or covariant risks on the supply side. Giné and Young (2009) empirically examine whether index insurance can improve loan take-ups by reducing farmers' risk exposure using a randomized field experiment in Malawi. Miranda and Gonzales-Vega (2011) theoretically compare the benefit between individual farmers' purchases of index insurance and agricultural banks' purchases of index insurance, where the former affects farmers' risk exposure and the later affects lenders' covariant risks of loan portfolio. Carter et al. (2012) argue that interlinking index insurance with loan contracts can simultaneously reduce the risks on both demand and supply side. However, these studies have neglected the potential of index insurance as an instrument to overcome information asym-

metry between farmers and lenders. This paper is the first one that investigates the impact of index insurance on borrowers' moral hazard problems in the rural credit market.

To better interpret experiment results, this paper builds a simple two-period model where farmers decide whether to borrow a production loan and, if so, whether to divert loans for consumption. Carter (1988) formally models credit diversion where the choice of credit diversion is a continuous variable. His model suggests that farmers would divert less credit for consumption when production risks are reduced, which provides a basic intuition of how index insurance influences credit diversion. We model credit diversion as a discrete choice of either investing the full amount or diverting half of the loan for consumption, to closely resembles the choices in the experiment[1]. The model shows that credit non-diverters benefit from purchasing index insurance, while credit diverters would be worse-off if they purchased index insurance. For credit non-diverters, index insurance can reduce their risk exposure by raising consumption levels in bad years and increase expected consumption by improving access to future loans. For credit diverters, however, index insurance is likely to increase their consumption risks or lower expected consumption. The fundamental reason for the difference of the impact between credit diverters and non-diverters is that index insurance pays indemnities based on an external index rather than farmers' realized income level. Therefore, the availability of index insurance encourages farmers to choose full investment of loans instead of credit diversion. Furthermore, the model predicts that the effect of index insurance on credit diversion is heterogeneous depending on farmers' risk preferences and ethical costs associated with violating loan contracts. In addition to credit diversion, we also analyze the impact of index insurance on the risk rationed, who do not take a loan due to risks. The model shows index insurance can reduce risk rationing and this effect only depends on farmers' risk preference.

To test these theoretical predictions empirically, we conducted a framed field experiment with 450 rural households in the north region of China. In the experiment, subjects were invited to participate in a set of games. Among the games, control game and treatment game are the primary ones used to identify the treatment effect of index insurance. Index insurance was not available in the control game and became available in the treatment game. In both games, subjects decided whether to borrow a production loan and, if so, whether to divert loans for consumption. Each subject played both control and treatment game. After the experiment, subjects took a detailed household survey, which recorded demographics and social-economics information as well as household activities of agricultural production, formal and informal borrowing and insurance purchase.

Based on theoretical predictions, the empirical analysis estimates the average and the heterogeneous treatment effect of index insurance on credit diversion as well as risk rationing. The main

[1]This also reflects the reality in the study area, where partial credit diversion is much more common than diverting the whole amount of loans.

results are that, first, index insurance reduces the number of credit diverters from 15.6% to 3.8% of the overall sample. Second, the treatment effect on credit diversion is heterogeneous depending on farmers' risk preferences and their ethical costs associated with violating loan contracts. For those with high level of ethical costs, the treatment effect is increasing in risk aversion. For those with middle level of ethical costs, the treatment effect is quadratic in risk aversion. For those with low level of ethical costs, the treatment effect is decreasing in risk aversion. Third, index insurance reduces the risk rationed from 21.3% to 10.9% of the overall sample, and the treatment effect on risk rationing is increasing in risk aversion.

The experiment context, rural China, has been plagued by asymmetric information problems characterized by its unique political environment. Although observing a fast economic growth during the past three decades, most Chinese smallholder farmers in the rural area still find it extremely difficult to borrow from formal financial institutions. Since Chinese farmers do not have property rights in land and cannot use land as collateral, there is no effective instrument to inhibit credit diversion[2]. Based on an empirical study in China, Feder et al. (1990) estimate that about one third of formal production credit is used for consumption purpose[3]. Credit diversion along with other asymmetric information problems makes formal lenders very reluctant to lend to the rural area. According to a national representative survey in 2003 by Chinese statistical bureaus, 28.3% of rural households obtained formal or informal loans. Among the borrowers, only 34.6% get formal loans. Among the formal and informal loans, only 27.2% are from formal institutions (Li & Zhu, 2010). The central government bank, People's Bank of China, reports that less than 10% of formal loans were used in agriculture sector in 2005 (Gale, 2006). The limited supply of formal credit results in severe credit constraints. The credit constrained in the formal credit market reported in the literature ranges from at least 30% to as high as 70% (Feder et al. 1990, Dong et al. 2012, Li and Zhu 2010, Zhang et al. 2008).

Index insurance is very new in China. Since 2007, only a couple of pilot index insurance programs have been launched at small scales. The attempt to develop index insurance is partly motivated by problems associated with the existing conventional agricultural insurance. Although more than half of rural households have enrolled in conventional agricultural insurance programs, the existing programs are far from efficient and effective. First, due to asymmetric information problems, government has to take a large share of insurance payout responsibility when the loss ra-

[2]Although without property rights in land, Chinese farmers have renting rights in their land for a long period, currently 30 years

[3]Feder et al. (1990a) show that although the household survey reports only 3.4% of formal credit is used for consumption purpose (including construction, consumption, social activity and others), the estimation results have a much higher diversion ratio, which is about one third. This indicates that the self-reported credit diversion activity is underestimated and can be used as a lower bound of real credit diversion.

tio exceeds a certain amount. Otherwise insurance companies are not willing to provide agricultural insurance. Second, because of high transaction costs and lack of transparency, the huge government expenditure on insurance program does not lead to significant welfare improvement of rural households as expected. Village officers, who are designated as agents to sell contracts, are often reported to embezzle insurance premium. Farmers do not understand the insurance contract and thus lack of informed demand and behavior response to insurance. Index insurance is a promising candidate to avoid the above problems. Thus the results from this study suggest important policy implications that can help improve the development of both credit and insurance market in China and other developing countries.

The rest of this paper is organized as follows. Section 2 presents a simple two-period model of household investment choice. Section 3 describes experiment design and survey data. Section 4 presents empirical model and Section 5 summarizes regression results. Section 6 concludes and discusses policy implications.

2. Theoretical Model

In this section, Section 2.1 lays out assumptions about the environment farmers are facing, including technology options, loan contracts and index insurance contracts. Based on these assumptions, Section 2.2 constructs a two-period model and analyzes how farmers choose the optimal investment plan when index insurance is and is not available. Finally, given the calibration of the experiment, Section 2.3 predicts farmers' decision-making in experimental games.

2.1 Environment

Technology

Assume farmers face two types of technology, a low-profit safe technology and a high-profit risky technology. Safe technology, the fall-back choice, generates a constant annual profit of y_m. Assuming its input costs are neglectable, the profit of safe technology ρ_m satisfies $\rho_m = y_m$. The risky technology requires an input of b to generate an income $y_s(b)$. Given an input level, the value of y_s depends on a stochastic shock θ, which equals to 1 in good years and 0 in bad years. y_s reaches a high level of income at y_s^h when $\theta = 1$ and a low level at y_s^l when $\theta = 0$. When b equals to a

constant B, the income from risky technology can be written as

$$y_s(b=B) = \begin{cases} y_s^h(B) & \text{, if } \theta\text{=1} \\ y_s^l(B) & \text{, if } \theta = 0 \end{cases}$$

, where y_s^h is the high income function when $\theta = 1$ and y_s^l is the low income function when $\theta = 0$. Both y_s^h and y_s^l are functions of b. Assume $\frac{\partial y_s^h}{\partial b} > 0$, $\frac{\partial y_s^l}{\partial b} > 0$, and $y_s^h(b) > y_s^l(b)$. The probability of $\theta = 1$ is p .

Loan Contract

Assume farmers are liquidity constrained and they have to borrow loans to invest in the risky technology. Suppose lenders offer a fixed loan contract denoted as $\ell < B, r, \chi = 0 >$ with loan size B, interest rates r and zero collateral level ($\chi = 0$). All the loan contract terms are fixed, because the goal is not to solve for the optimal contract under information asymmetries as Boucher et al. (2008), but rather to derive the farmer's optimal behavior with and without index insurance. Interest rates r satisfy $y_s^h(B) > (1+r)B > y_s^l(B)$, so that the income from risky technology is sufficient to repay the loan in good years but insufficient in bad years. Assume that the high-yielding technology is more profitable than the safe technology after taking capital costs into account, such that $py_s^h(B) + (1-p)y_s^l - (1+r)B > y_m$. If investing the whole amount of the loan B in the risky technology, farmers' net income after repaying loans, $\rho_s(B)$, can be written as

$$\rho_s(B) = \begin{cases} y_s^h(B) - (1+r)B & \text{, if } \theta\text{=1} \\ 0 & \text{, if } \theta = 0 \end{cases}$$

. Finally, the loan contract has dynamic incentives. If they default in the first year, borrowers will not get access to future loans and have to choose the safe technology in the second year. Following Stigliz and Weiss (1981), default happens when $y_s < (1+r)B$ given zero collateral, which is equivalent to $\theta = 0$.

Credit Diversion

Instead of investing all the credit, borrowers may choose to use part of the loan for consumption right after obtaining the loan and only invest the rest in risky technology. This is called credit diversion, when the amount invested in production is less than loan size B. To closely resemble the

experiment and simplify the analysis, our model specifies the way farmers divert credit by making the following two assumptions:

- Credit diverters divert half of the loan, $\frac{1}{2}B$, for consumption.[4]
- Credit diversion only affects the magnitude of default, rather than the probability of default. In other words, credit diverters and non-diverters have the same probability of default p. But the amount that diverters default in bad years is larger than non-diverters, because the former has less income to repay than the latter. [5]

Therefore, credit diverters' income from risky technology becomes

$$y_s(b = \frac{1}{2}B) = \begin{cases} y_s^h(\frac{1}{2}B) = (1+r)B & \text{, if } \theta\text{=1} \\ y_s^l(\frac{1}{2}B) = 0 & \text{, if } \theta = 0 \end{cases}$$

, where the income from partial investment is just enough to repay the loan when $\theta = 1$ and equals to 0 when $\theta = 0$. Due to information asymmetry between lenders and borrowers, lenders cannot differentiate credit diverters and non-diverters. Thus credit diverters face the same loan contract as non-diverters. Diverters' net income from risky technology becomes

$$\rho_s(\frac{1}{2}B) = 0$$

. The consumption level under credit diversion is equal to the sum of the net income from risky technology and the diverted amount, which is equal to $\rho_s(\frac{1}{B}) + \frac{1}{2}B = \frac{1}{2}B$.

Note that lenders' expected profit is smaller if borrowers choose credit diversion rather than full investment. Lenders' profit π can be written as

$$\pi = \begin{cases} (1+r)B & \text{, if } \theta = 1 \\ y_s^l(b) & \text{, if } \theta = 0 \end{cases}$$

. Since y_s^l is increasing in input b, credit diversion reduces lenders' expected profit $E(\pi)$.

[4] A focus group study in the sample area reveals that, borrowers rarely divert all the credit for consumption but usually divert part of the credit.

[5] Under partial credit diversion, it is reasonable to assume diverters can still repay the loan in good years given the amount of their investment.

Index Insurance Contract

Index insurance contract is denoted as $A < \theta, I, q >$, where I represents indemnities and q insurance premium. The indemnification depends on θ. Farmers receive indemnities only when $\theta = 0$. The contract is actuarially fair, so that $q = (1-p)I$. If the farmer invests all the loan in risky technology and purchases index insurance, his income y_s^I can be written as

$$y_s^I(B) = \begin{cases} y_s^h(B) - q & \text{, if } \theta\text{=1} \\ y_s^l(B) - q + I & \text{, if } \theta = 0 \end{cases}$$

. To effectively reduce production risks, index insurance contract satisfies $\frac{(1+r)B - y_s^l(B)}{p} < I < y_s^h(B) - y_s^l$, which means insurance indemnity is large enough to pay the loan but not too big to create new income risks. Assuming borrowers first repay the loan before any consumption occurs, the insurance can help borrowers under full investment to avoid default. Under index insurance, non-diverters' net income ρ_s^I becomes

$$\rho_s^I(B) = \begin{cases} y_s^h(B) - q - (1+r)B & \text{, if } \theta\text{=1} \\ y_s^l(B) + I - q - (1+r)B & \text{, if } \theta = 0 \end{cases}$$

. In this case, lenders' expected profit $E(\pi)$ is equal to $(1+r)B$, higher than that if borrowers choose credit diversion or full investment but without insurance.

2.2 A Two-period Model

Based on the above setup, we construct a two-period model with year 1 and year 2. The farmer chooses the optimal investment and production option j from the choice set Φ to maximize his expected utility over the two years:

$$\underset{j \in \Phi}{Max}\, EU(c_j^1 + c_j^2 - EC_j)$$

, where c_j^1 and c_j^2 are consumptions of year 1 and 2 for option j, and EC_j is the consumption equivalent of ethical costs associated with option j. Ethical costs measure the disutility due to ethical reasons, which the borrower suffers if he chooses to violate loan contract terms after agreeing to the

contract [6]. In this model, the violation refers to diverting production loans for consumption purpose. Farmer i's ethical costs depend on his choice of credit diversion and his ethical sensitivity. If the farmer chooses the fall-back technology or full investment, $EC_j = 0$. If the farmer i chooses credit diversion, $EC_j = e_i$, where $e_i \geq 0$. Assume the utility function is a constant-relative-risk-aversion utility function and r_i is the CRRA coefficient for farmer i.

Without Index Insurance

In absence of index insurance, the choice set Φ has three options. Option 1 is the fall-back choice. Option 2 is to take a loan and invest the full amount of the loan. Option 3 is to take a loan but divert half of it for consumption. Table 1 lists each year's consumption level and ethical costs for each option. Note that, due to dynamic incentives of the loan contract, if default occurs with $\theta = 0$ in year 1, borrowers will not have access to loans and thus have to take the safe technology in year 2.

The Table 1 shows that Option 2 is riskier than the other two options. Without considering ethical costs, highly risk averse farmers are likely to choose Option 1 (risk rationing) or Option 3 (credit diversion). Risk neutral or risk loving farmers are likely to choose Option 2, since full investment has higher expected total consumption than the fall-back choice [7]. If considering ethical costs, as e_i is increasing, farmer i is more likely to choose Option 1 or Option 2.

With Index Insurance

In this section, we analyze the impact of index insurance on farmers' choices in two steps. First, assuming that farmers are not allowed to change production plans, we examine who will purchase index insurance contract. Based on the results, we construct the choice set Φ when index insurance becomes available. Then we relax the assumption and analyze how index insurance affects farmers' investment choices.

Purchase of Index Insurance

The first step answers the question that who benefit from and purchase index insurance if farmers' production and investment plans are fixed. In the following we discuss each of the three options.

For farmers under the fall-back choice, the purchase of an actuarially fair insurance will increase the fluctuation of their net income, because the fall back choice already offers a constant net income.

[6] During a focus group study, some villagers stated a good person should not divert credit and should exactly follow loan contract, while some villagers did not show such believes.

[7] The expected total consumption over two years for the three options are $(1+p)y_m + (1-p)y_m$, $(1+p)[py_s^h(B) - p(1+r)B] + (1-p)y_m$, and $(1+p)\frac{1}{2}B + (1-p)y_m$ respectively.

Risk averse farmers will always prefer not to purchase insurance if choosing Option 1.

Credit diverters will not purchase index insurance for two reasons. First, when the net insurance indemnity satisfies $I - q = pI < (1+r)B$, index insurance cannot help credit diverters to avoid default because the net indemnity is smaller than the debt. Index insurance will not reduce consumption risks but lower expected consumption of credit diverters, since they still need to pay insurance premium. The fundamental reason for credit diverters' losses caused by index insurance is that the magnitude of indemnities is based on the loss under full investment rather than farmers' realized income. Second, even when the net indemnity is large enough to cover their loss in bad years, credit diverters will not purchase because the timing of indemnities is based on θ rather than the occurrence of low consumption. For example, if $pI = (1+r)B$, insurance premium has to be as high as $q = \frac{(1-p)}{p}(1+r)B$. The maximum insurance premium that credit diverters can afford is $\frac{1}{2}B$. Therefore, credit diverters either cannot afford the insurance premium or end up with a very low consumption. Table 2 compares credit diverters' consumption with and without index insurance, when $pI = (1+r)B$ and $\frac{1}{2}B = q$. Index insurance raises the riskiness of credit diversion to the extent that it even becomes riskier than full investment. Besides, index insurance may also lower the expected consumption of credit diversion[8]. The above discussion suggests credit diverters are likely to be risk averse. Thus, credit diverters are not likely to purchase index insurance.

If the farmer under full investment purchases index insurance, his consumption over the two years is shown in the second row of Table 3 . Now we compare the full investment between with and without index insurance. The expected total consumption with and without index insurance is denoted as $E_2^I(c_1+c_2)$ and $E_2(c_1+c_2)$ respectively, which can be written as

$$E_2(c_1+c_2) = (1+p)[py_s^h(B) - p(1+r)B] + (1-p)y_m$$

and,

$$\begin{aligned} E_2^I(c_1+c_2) &= 2[py_s^h(B) + (1-p)y_s^l - (1+r)B] \\ &= (1+p)[py_s^h(B) - p(1+r)B - (1-p)((1+r)B - y_s^l)] + (1-p)[py_s^h(B) + (1-p)y_s^l - (1+r)B] \end{aligned}$$

. Since $(1+r)B > y_s^l$, the first term of $E_2(c_1+c_2)$ is bigger than that of $E_2^I(c_1+c_2)$. It means that, borrowers can have a higher mean consumption if not purchasing insurance, due to the limited liability of loan contract. Since $py_s^h(B) + (1-p)y_s^l - (1+r)B > y_m$, the second term of $E_2(c_1+c_2)$ is smaller than that of $E_2^I(c_1+c_2)$. This means that insurance can increase expected consumption

[8]The expected consumption without index insurance is equal to $(1+p)\frac{1}{2}B + (1-p)y_m$, and that with index insurance is equal to $(1-p)B$.

by providing access to future loans and income-enhancing technology. In other words, the first term of $E_2^I(c_1+c_2)$ represents the loss due to limited liability of loans, and the second term represents the gain due to dynamic incentives. Putting the two factors together, the difference of expected consumption between with and without index insurance is ambiguous. If considering multiple periods more than two years, it is more likely that index insurance raises the expected consumption level for farmers under full investment. The comparison of riskiness between with and without insurance can be seen from the minimum total consumption. Full investment without insurance has a minimum consumption of y_m, and with insurance has a minimum consumption of $2[pI+y_s^l-(1+r)B]$. This means when insurance indemnity satisfies $pI > (1+r)B - y_s^I + \frac{y_m}{2}$, index insurance not only reduces the probability of default and increase expected consumption but also reduces income risks. In this case, farmers under full investment are very likely to purchase index insurance.

Overall, although insurance company does not know farmers' production and investment activity, only farmers choosing full investment (Option 2) are likely to purchase index insurance, while those choosing the fall-back option (Option 1) or credit diversion (Option 3) will not purchase. This self-selection into index insurance market is a major advantage of index insurance and plays a key role in how index insurance influences credit diversion. Due to the self-selection, we can eliminate the two cases: Option 1 plus index insurance and Option 3 plus index insurance. Thus when index insurance is available, farmers essentially face four choices as shown in Table 4, which compose the choice set Φ under index insurance.

Impact of Index Insurance

When allowing farmers to switch their production plan, those choosing the fall-back option or credit diversion when index insurance is not available are likely to switch to full investment when index insurance becomes available, since index insurance can reduce consumption risks and raises expected consumption level of full investment as discussed above. Therefore, index insurance can reduce credit diversion by inducing the switch from Option 3 to Option 4, and risk rationing by inducing the switch from Option 1 to Option 4.

The model also suggests that the impact of index insurance is not universal. The difference of expected utility between Option 3 and 4 depends on farmers' risk preference r_i and ethical costs e_i, while the difference of expected utility between Option 1 and 4 only depends on r_i. The exact pattern of the heterogeneity depends on the calibration of the environment.

2.3 Predictions

This section predicts the impact of index insurance and the heterogeneity of the impact in the experiment. Given the calibration of the experiment, Figure 1 depicts the CDFs of total consumption $(c_1 + c_2)$ for Option 1, 2 and 3 when index insurance is not available. Since Option 2 has a higher mean but more down-side risks than Option 1, the choice between the two options depends on farmers' risk preference, r_i. When $e_i = 0$, Option 3 first-order stochastically dominates Option 1. As e_i rises, farmers are more likely to choose Option 1 or 2. In general, a farmer's choice depends on his risk preference, r_i and ethical costs associated with credit diversion, e_i .

When index insurance becomes available, farmers face a forth choice of Option 4. The experiment is calibrated so that index insurance raises the expected level and lowers the riskiness of consumption under full investment. The CDFs of the four choices are depicted in Figure 2. In fact Option 4 second-order stochastically dominates Option 2. In presence of index insurance, risk-averse farmers will be more likely to take loans and invest the full amount in risk technology.

Assume a population distributed on r_i and e_i. Figure 3 graphs the distribution of the choices of the population. The vertical axis is the consumption equivalent of ethical costs e_i as a percentage of the average total consumption of Option 1. The horizontal axis is the CRRA coefficient in the range of $[0,4]$. The model predicts that less risk-averse farmers on the left hand side of Figure 3 will choose to take the loan and invest the full amount. For risk-averse farmers on the right hand side, their choices depend on their ethical costs . If ethical costs are high, risk-averse farmers will choose the safe fall-back option and be risk rationed, who are on the upper-right corner of Figure 3. If ethical costs are relatively low, they will choose to divert credit for consumption and become credit diverters, who are on the lower right corner of Figure 3.

The distribution of farmers' choices under index insurance is depicted in Figure 4. The slashed area denotes the population who change production and investment plans, either from the fall-back safe technology to the risky technology or from credit diversion to full investment. All the risk-rationed switch from Option 1 to Option 4, indicating the treatment effect of reducing risk rationing is increasing in r_i in the range of $[0,4]$. Most credit diversion farmers switch to Option 4, except those with very high risk aversion coefficient and very low ethical costs.

The heterogeneity of the treatment effect on credit diversion is more complicated. Given a certain level of ethical costs, there are three possible patterns of who are likely to switch from credit diversion to full investment. When e_i is high, the larger the risk aversion, the more likely the switching occurs, indicating the magnitude of the treatment effect is increasing in r_i. When e_i decreases, farmers with middle level risk aversion are more likely to switch, indicating a quadratic form of treatment effect in r_i. In this case, the risk neutral or highly risk-averse farmers do not

respond to index insurance, either because they already choose full investment or they are too afraid of risks. When e_i is low, the magnitude of the treatment effect is decreasing in r_i. Section 4 and 5 empirically test these predictions using experiment and survey data.

3. Experiment Design and Survey Data

This section describes experiment design and procedure as well as statistics summaries.

3.1 Game Design

The experiment consists of 5 games: Practice Game I, Control Game, Practice Game II, Treatment Game, and Risk Preference Elicitation Game. Each subject played all the five games in the same order[9]. The choices subjects are facing in the games are summarized in Table 5, which corresponds to Table 1 in the model.

Practice Game I

Subjects first took Practice Game I, where they chose between two choice, Option 1 and Option 2. Option 1 is to grow maize without either loan or insurance in both the first and the second year. The costs of growing maize is neglected. The annual net income from growing maize is 12,000 in good years and 7,500 in bad years. The year status was realized by farmers drawing a pin-pong ball from a bag of 10 balls. The bag has 6 orange and 4 white balls. Orange balls represent good years and white ones represent bad years.

Option 2 is to take a loan of 30,000 and invest all of them in raising sheep. Interest rates are 20% without collateral requirements. The probability of good years for raising sheep is the same as that for growing maize. If the borrower draws an orange ball in the first year, the income from raising sheep is enough to repay the loan and he/she has a net income of 29,000 after loan repayment. In the second year, the farmer borrows another loan and invests in raising sheep with full amount again. If the borrower draws a white ball in the first year, the income from raising sheep is not sufficient to repay the loan. The borrower ends up with zero net income in the first year and has to grow maize in the second year. The purpose of this game is to introduce the two types of technology and help subjects to get familiar with the experiment setting.

[9]Ideally, we should randomize the order of games. However, the order is designed to assist farmers to gradually learn and get familiar with the experiment. So we keep the order the same for everyone.

Control Game

In the control game, subjects faced a third choice of Option 3, in addition to Option 1 and 2. Subjects chose one out of the three options. Option 3 represents credit diversion, where subjects take the loan, use half of it for consumption and leave the other half for investment in raising sheep. If the first year is a good year, the income from partial investment in sheep is just enough to repay the loan and the borrower has 15,000 for consumption, which equals to the amount he/she diverts. In the second year, he/she repeats to divert half of credit for consumption. If the first year is a bad year, the income from partial investment is not enough to repay the loan but the borrower still has 15,000 for consumption. In the second year the borrower has to grow maize due to default in the first year. If choosing Option 3, regardless of the year status, the borrower always has 15,000 for consumption in the first year. The control game is used to elicit farmers' investment choices in absence of index insurance.

Practice Game II

In the second practice game, index insurance for raising sheep was introduced. Given full investment of loans in raising sheep, subjects chose whether to purchase index insurance. The insurance pays indemnity based on the year status instead of income level. The insurance premium is 6,500. The indemnity is 16,250 and only paid in bad years. Subjects chose between Option 2 and Option 4. Option 4 has the same investment and production activity as Option 2, plus purchasing index insurance. If choosing Option 4, the borrower is always able to repay the loan and has access to loans in the second year. The net income is 22,500 in good years and 6,750 in bad years. The purpose of this game is to introduce index insurance and assist subjects to understand the difference between with and without insurance.

Treatment Game

In the treatment game, subjects faced all the four options (Option 1, 2, 3 and 4) and chose one out of the four. The treatment game elicits farmers' production and investment choices when index insurance is available.

Risk Preference Game

The last game estimates subjects' risk preference, following the conventional routine of Holt and Laury (2002). Subjects were presented with two lotteries. Lottery A rewards subjects 60,000 game

money if they draw an orange ball and 5,000 game money if they draw a white ball. Lottery B rewards 20,000 if subjects draw an orange ball and 15,000 if they draw a white ball. Subjects chose one between the two lotteries for 10 times. Each time, we changed the number of orange and white balls in the bag. The purpose of this game is to calibrate subjects' CRRA coefficient.

3.2 Experimental Procedures

We conducted the games with agricultural population in the north region of China. The games were organized in sessions. In each session, we recruited 20-40 household representatives from local villages. Each session consisted of the five games discussed above and a household survey after the games. At the beginning of each game, instructions were presented orally in Chinese to all the subjects simultaneously with the aid of posters. The posters listed the possible income, net income and consumption of each option, highlighting the key points specific to each option and the rules of the game. In each game, before making the final decision, subjects were asked to practice twice. In each practice, subjects wrote down their decisions on a worksheet. Each subject realized his/her own year status by drawing a pin-pang ball from a bag twice, the first drawing for the first year and the second drawing for the second year. Based on his decision and the color of the balls he drew, subjects determined his total consumption according to posters and wrote it down on the worksheet. After two practices, subjects then made their final choices. The ball-drawing for the final decision was conducted at the end of the session. Subjects were not allowed to switch between options after the realization of year status of the first year.

After subjects finished two practices and the final decision for each game, their worksheets were collected. A representative subject was invited to randomly pick one game from the five. The picked game is the payoff game, according to which subjects got paid. If the payoff game is one of the first four games, each subject draws balls from a bag twice to determine his year status of the first and second year. If the payoff game is the Risk Preference Elicitation game, another representative is invited to draw a ball to determine the number of orange balls in the bag. Then each subject draws a ball from the bag. Assistants recorded the payoff game and the result of drawing for each subject. Subjects did not know his game earnings before finishing the final decision for all the five games.

Subjects were paid based on a fixed ratio of the game money they earned in the payoff game. The maximum of cash earning is ¥120 RMB (about $20) and the average is around ¥60RMB (about $10). Each session including household survey takes an average of 3 hours. The average cash earning per hour per subject is about twice of the hourly minimum wage in nearby cities.

While assistants were calculating and preparing cash award, subjects were taking a household survey that asks demographic and social-economic information, agricultural activities and financial market activities. After finishing the survey, subjects came to assistants and claimed cash earning.

We played 17 sessions with an average of 27 subjects per session. The total number of experiment subjects is 450.

3.3 Study Area and Subject Statistics

The experiment was conducted in three districts in the Province of Inner Mongolia of China: Chifeng District, Tongliao District and Linxi District, as shown in Figure 5 . The three districts have diversified agricultural structure and demographics. In the Chifeng district, most villagers rely on cropping and off-farm working. In Tongliao District, most rural population are engaged in half-cropping-half-herding agriculture. In the Linxi District, villagers mainly rely on herding.

The statistics summaries of the sample pool are provided in Table 6, including demographics, maize production, sheep production, wealth and income. Among the sample, the minority (who are not "Han") takes 38%. The education level on average is 6.78 years, equivalent to just finishing elementary school. Most of the sample are either household head or household head spouse. The average off-farming working time is only around 1 month per year, indicating that the study area relies heavily on agriculture. The size of farmland is much larger in the study area compared to other regions of China. The average size of cultivated land is 42 mu, of which 22 mu is used to grow maize. But note that the median size of total farmland and maize crop is just 22 and 12 mu, which suggests a skewed distribution of crop production. The distribution of sheep production is even more skewed than crop production. The average value of all livestock is ¥20245 RMB but the median becomes only ¥50 RMB. The average number of sheep stock is 13 but the median is 0. This skewness is also reflected in the wealth and income statistics, such as the value of house and durables. The median of annual gross household income per capita is ¥6833 RMB (about $1093). If production costs is 50% of gross income, the net income per capital per year is $546.5, less than $1.5 per day. The median of cash held by household is ¥750 RMB (about $120).

Overall, the data suggests that the sample are very poor and rely on agriculture. Comparing mean and median, there are many outliers in terms of agricultural production and wealth indicators. Since every subject takes both control and treatment games, nonrandom selection into sample pool does not affect the internal validity of our results. However, selection into our experiment may affect the external validity of this exercise.

Table 7-9 describes the local credit market based on household survey. Table 7 shows that more

than half of the sample do not obtain any loans. Only 12.7% of the sample obtain loans from formal financial institutions. Table 8 shows that 44.1% of the sample are credit constrained and most credit constrained are quantity rationed. Table 9 summarizes loan terms and loan purpose of formal and informal loans in the study area. Note that there is only 20.3% of formal loans asking for collateral due to the lack of collateralizable assets. 7.8% of formal loans were defaulted and 15.6% of them were used for consumption purpose. For formal loans, we find a significant correlation between loan default and credit diversion. We also ask the quantity rationed of their loan purpose if they were granted a formal production loan. 18% of them stated they would divert all or part of the loan for consumption.

4. Empirical Model

This section presents the empirical model used to estimate the average and heterogeneous treatment effect of index insurance on credit diversion and risk rationing.

4.1 Average effect

Credit diversion

The average treatment effect of index insurance on credit diversion can be estimated by the following regression equation:

$$Y^c_{ijg} = \beta_0 + \beta_1 T_g + \beta_2 R_{ij} + \beta_3 X_{ij} + \beta_4 A_j + \varepsilon^c_{ijg} \tag{1}$$

where Y^c_{ijg} is a binary random variable that equals 1 if the subject i of game session j chooses credit diversion (Option 3) in game g and equals 0 otherwise. Game g equals 1 if it is the treatment game and equals 0 if it is the control game. T_g represents the treatment of the availability of index insurance in game g, which equals 1 if index insurance is available and equals 0 if not available. R_{ij} is the estimated subject i's constant relative risk aversion coefficient based on the risk preference elicitation game. X_{ij} are individual-level control variables. A_j are regional fixed effects. The error term ε^c_{ijg} can be written as

$$\varepsilon^c_{ijg} = v^c_{jg} + e^c_{ijg}$$

where $v^c_{jg}$is a random effect of session j in game g, and e^c_{ijg} is a mean-zero error term.

Risk rationing

The average treatment effect of index insurance on risk rationing can be estimated by the following regression equation:

$$Y^r_{ijg} = \alpha_0 + \alpha_1 T_g + \alpha_2 R_{ij} + \alpha_3 X_{ij} + \alpha_4 A_j + \varepsilon^r_{ijg} \tag{2}$$

, where Y^r_{ijg} is a binary random variable that equals 1 if the subject i of game session j chooses the safe fall-back choice (Option 1) in game g and equals 0 otherwise. Similarly, the error term ε^r_{ijg} can be written as

$$\varepsilon^r_{ijg} = v^r_{jg} + e^r_{ijg}$$

.

The parameters of interest are β_1 and α_1, which are the estimates of the average treatment effect on credit diversion and risk rationing respectively. The theoretical model suggests that $\beta_1 < 0$ and $\alpha_1 < 0$. We estimate equation (1) and (2) with OLS, clustering standard errors at the game and session level jg. Since each subject participated in both control and treatment game, the OLS estimators of the parameters β_1 and α_1 are unbiased.

4.2 Heterogeneous effect

Credit diversion

The theoretical model predicts that the treatment effect of index insurance on credit diversion depends on both risk preference and individual ethical costs. The difficulty of estimating the heterogeneous effect is that ethical costs are not observable. To overcome this difficulty, we estimate the heterogeneous treatment effect on credit diversion in two steps. In the first step, we construct an index of predicted ethical costs, $\hat{e}_i$, based on borrowers' past experience of credit diversion recorded in household survey. In the second step, the whole sample is divided into five groups based on the quintiles of $\hat{e}_i$. For each quintile group, we run the following regression equation:

$$Y^c_{ijg} = \beta_0 + (\delta_0 + \delta_1 R_{ij} + \delta_2 R^2_{ij}) * T_g + \beta_2 R_{ij} + \beta_3 X_{ij} + \beta_4 A_j + \varepsilon^c_{ijg} \tag{3}$$

, where β_1 is replaced with a quadratic function of R_{ij} to see how the treatment effect changes when risk aversion varies.

The parameters of interest are δ_0, δ_1 and δ_2. δ_0 measures the treatment effect through the channel of expected consumption. δ_1 and δ_2 measure the treatment effect through the channel

of consumption risk. The model suggests that $\delta_0 < 0$ for all quintile groups. δ_1 is changing from being negative for groups with high ethical costs to being positive for groups with low ethical costs. δ_3 is positive only for groups with middle level of ethical costs. We estimate regression equation (3) with OLS, clustering standard errors at the game and session level. Again, since each subject participated in both control and treatment game, the OLS estimators of the parameters δ_0, δ_1 and δ_2 are unbiased.

The index $\hat{e}_i$ is estimated based on borrowers' experience of diverting formal loans recorded in household survey. The theoretical model suggests credit diversion behavior is also influenced by risk preference and ethical costs. Denote the real experience of credit diversion as a binary variable CD_i, which equals to 1 if borrower i diverted any amount of credit for consumption and equals to 0 if not. Assume ethical costs e_i can be written as a linear function of individual i's demographic characteristics Z_i, as $e_i = e_i(Z_i) = -\eta_2 Z_i - v_e$. We run the following regression equation:

$$CD_i = \eta_0 + \eta_1 R_i - e_i + \eta_3 X_i + v_i = \eta_0 + \eta_1 R_i + \eta_2 Z_i + \eta_3 X_i + v_i + v_E \tag{4}$$

,where X_i are agricultural production history and credit market activities that influence borrowers' decisions of credit diversion. Since not every subject of the sample has experience of borrowing formal loans, we run the OLS regression on equation (4) only for formal loan borrowers plus the quantity rationed who expressed their planned loan purpose if they were granted a formal loan in the survey. Once obtaining the estimators of η_2, we construct the estimated $\hat{e}_i$ for the whole sample as an index of ethical costs by

$$\hat{e}_i = -\hat{\eta}_2 Z_i \tag{5}$$

.

Risk rationing

The model indicates that the treatment effect of index insurance on risk rationing depends only on farmers' risk preference. To test this heterogeneous effect, we estimate the following regression equation:

$$Y^r_{ijg} = \alpha_0 + (\gamma_0 + \gamma_1 R_{ij} + \gamma_2 R^2_{ij}) * T_g + \alpha_2 R_{ij} + \alpha_3 X_{ij} + \alpha_4 A_j + \varepsilon^r_{ijg} \tag{6}$$

, where α_1 is replaced with a quadratic function of R_{ij}. The coefficient γ_0 represents the effect of index insurance through changing farmers' expected consumption. Coefficient γ_1 and γ_2 measure the treatment effect through changing farmers' risk exposure. The hypotheses are $\gamma_0 < 0$, $\gamma_1 < 0$

and $\gamma_2 = 0$. Equation (6) is estimated by OLS, clustering at the game and session level.

5. Experiment and Estimation Result

5.1 Experiment Result

The number and percentage of each option chosen by subjects in control and treatment game are listed in Table 10. In control game, 63.1% of subjects chose to raise sheep with full investment, 21.3% are risk rationed and chose to grow maize, and 15.6% chose credit diversion. When Option 4 with index insurance becomes available in treatment game, the risk rationed are reduced to 10.9%. Almost half of the risk rationed switched to raising sheep. Credit diverters are reduced to 3.8%, which means 75.8% of credit diverters were induced to invest the full amount of loans. In treatment game, the percentage of subjects choosing to raise sheep with full investment reached 85.3% including both with and without index insurance, which is 20% higher than that in the control game.

The distribution of estimated CRRA coefficient based on the Risk Preference Elicitation game is graphed in Figure 6. The estimated CRRA coefficients concentrate around 1. 95% of the sample exhibit risk aversion. The mean of the CRRA coefficient is 0.92 and standard deviation is 0.45.

5.2 Average Effect of Index Insurance

The regression results of equation 1 and 2 are reported in Table 11. The estimated average treatment effects are -11.8% on credit diversion and -10.4% on risk rationing, both statistically significant at the 5% level.

Table 11 also reports the control variables that have significant correlations with credit diversion and risk rationing. Household size, years of cropping and size of owned grassland are negatively correlated with credit diversion. This suggests that the demand for capital input into production increases as other input goes up such as labor, experience and land. The amount of informal loans is positively correlated with credit diversion. As shown in statistics summaries, most of informal loans are used for consumption purpose. A higher amount of informal loans suggests a higher demand of liquidity for consumption, thus resulting in higher probability of credit diversion.

For risk rationing, higher age is associated with higher probability of risk rationing, since the elderly are likely to be more risk aversion although CRRA itself is not significant. Education level, the membership of Chinese Communist Party (CCP) and years of cropping and raising sheep are

all negatively correlated with risk rationing, suggesting human capital may help improve technology adoption. It is not straightforward why the number of informal loans and private lending are significantly correlated with risk rationing.

5.3 Heterogeneous effect

Before going into the detail of heterogeneous effect results, it is important to examine the quality of estimated CRRA coefficients in the Risk Preference Elicitation Game, since the estimated CRRA coefficients play a key part in the estimation of heterogeneous treatment effect. In addition to the Risk Preference Elicitation game, subjects also revealed their risk preference in Practice Game I and in household survey. In Practice Game I, subjects chose between the safe technology of growing maize and the risky technology of raising sheep. In household survey, subjects were asked to choose between a safe payment and a lottery. Highly risk-averse farmers are likely to choose growing maize in Practice Game I and the safe payment in the survey question, while risk-loving farmers are likely to choose the other one [10]. However, we find that some subjects who exhibited extreme risk-loving in the Risk Preference Game chose the safe choice in both Practice Game I and the survey question[11]. Although there are multiple possible reasons for this inconsistency such as lack of understanding of the risk preference game, irrational behavior and some unknown factors, we call those subjects "irrational" and find 9 such irrational subjects in the sample. In the following analysis, we run regressions with two samples. One is the full sample with 450 subjects, and the other is the partial sample with 441 subjects excluding the 9 irrational subjects.

The OLS regression result of equation 4 is shown in Table 12, which is used to construct ethical costs index. Variable Z_i are the highlighted bold regressors in Table 12, including gender, age, minority and religion status. It shows farmers who are old, belong to ethnic minorities and believe in certain religion are more likely to divert credit in the real life, because they have lower ethical costs of violating loan contracts. The estimated CRRA coefficient has a significant positive relation with real credit diversion history, which provides a real-life evidence beyond the experiment that risks are a key issue of causing credit diversion . These results do not change much between full and partial sample.

The heterogeneous treatment effect on credit diversion is shown in Table 13 for full sample and Table 14 for partial sample, which are estimated based on equation 3. Under the full sample,

[10] Based on the constant-relative-risk-aversion utility model, the CRRA thresholds between the safe and risky choices is 1.52 in the Practice Game I and 0.24 in the survey question.

[11] We do not find the other extreme case, where farmers who exhibit extreme risk averse choose the risk choice in both the practice game and the survey question.

without dividing the sample into quintile groups, only parameter δ_0 is significant. The two interaction terms with treatment indicator are not significant, indicating there is no heterogeneous effect. However, when we divide the sample into five quintile groups based on $\hat{e}_i$, the estimators of δ_1 and δ_2 become significant and change the sign when quintile group changes. For groups with high ethical costs, δ_1 is significantly negative and δ_2 is not significantly different from zero, meaning that the treatment effect on reducing credit diversion is increasing in risk aversion. For groups with middle level of ethical costs, δ_2 becomes significantly positive, indicating that the magnitude of treatment effect is first increasing and then decreasing in risk aversion. For groups with low ethical costs, δ_1 becomes positive although not significant and δ_2 is not significantly different from zero, suggesting that the magnitude of treatment effect is decreasing in risk aversion. These results are consistent with out theoretical predictions. When restricted to the partial sample, the results show the same pattern of heterogeneity as under full sample. But estimates are more significant and the magnitude of estimates is larger.

The regression result of equation 6 for heterogeneous treatment effect on risk rationing is shown in Table 15. Under full sample, the estimator of γ_0 and γ_1 are negative, which is consistent with model prediction but not significant. Under partial sample, $\hat{\gamma}_1$ becomes significantly negative at the 5% level. This means the treatment effect of index insurance is increasing as farmers' risk aversion rises. Also, the coefficient of CRRA is significantly positive under partial sample, which verifies that risk-averse farmers are more likely to be risk rationed.

6. Conclusion

This paper explores how the availability of index insurance affects borrowers' moral hazard behavior in the rural credit market using a framed field experiment with 450 Chinese farmers. We build a simple two-period model to show that credit diverters and the risk rationed would be self-selected out of the index insurance market. Only farmers who choose the risky technology and invest the full amount of loans are willing to purchase index insurance. The experiment results show that the availability of index insurance reduces the number of credit diverters from 15.6% to 3.8% over the whole sample. The treatment effect on credit diversion is heterogeneous across farmers depending on their risk preferences and ethical costs associated with violating loan contracts. The magnitude of the treatment effect is decreasing at low level of ethical costs, quadratic at middle level of ethical costs and increasing at high level of ethical costs in risk aversion. In addition, index insurance reduces the number of risk-rationed farmers from 21.3% to 10.9%. The treatment effect on risk

rationing is increasing in risk aversion.

The theoretical and empirical results have important policy implications for stimulating credit supply to agriculture and reducing credit rationing. They suggest that lenders can substitute index insurance for collateral requirement, since index insurance can effectively control moral hazard problem. If so, poor farmers who were previously quantity rationed or risk rationed can obtain credit simply by purchasing index insurance. In other words, lenders can use index insurance as a signal to identify "good" clients and overcome information asymmetry. Compared to other signaling mechanisms such as collateral, the purchase of index insurance is much less "expensive" and unlikely to exclude the poor from the credit market.

References

[1] Boucher, Stephen R., Michael R. Carter and Catherine Guirkinger. 2008. "Risk Rationing and Wealth Effects in Credit Markets: Theory and Implications for Agricultural Development." *American Journal of Agricultural Economics*, 90(2): 409-423.

[2] Carter, Michael R. 1988. "Equilibrium credit rationing of small farm agriculture." *Journal of Development Economics*, 28(1): 83-103.

[3] Carter, Michael R., Lan Cheng and Alexander Sarris. 2012. "The Impact of Interlinked Index Insurance and Credit Contracts on Financial Market Deepening and Small Farm Productivity." Working Paper.

[4] Conning, J. and Udry, C., 2007. "Rural financial markets in developing countries." In: Evenson R., Pingali P., Schultz T. (Eds.), *The Handbook of Agricultural Economics*, 3: 2857–2908.

[5] Dong, Fengxia, Jing Lu, and Allen M. Featherstone. 2012. "Effects of credit constraints on household productivity in rural China". *Agricultural Finance Review*. 72(3): 402 - 415.

[6] Feder, Gershon, Lawrence Lau, Justin Lin, and Xiaopeng Luo. 1990. "The Relation between Credit and Productivity in Chinese Agriculture: A Model of Disequilibrium." *American Journal of Agricultural Economics,* 72, no. 5:1151–57.

[7] Gale, F., and Collender, R. 2006. "New directions in China's agricultural lending." the Economic Research Service Electronic Outlook Report

WRS-06-01. United States Department of Agriculture (USDA). Available at: www.ers.usda.gov/publications/WRS0601/WRS0601.pdf , last accessed 6 June 2007.

[8] Ghosh, P., D. Mookherjee, and D. Ray. 2000. " Credit Rationing in Developing Countries: An Overview of the Theory." In D. Mookherjee and D. Ray, eds. *Readings in the Theory of Economic Development*, 283–301. Malden, MA: Blackwe.

[9] Giné, Xavier and Dean Young. 2009. "Insurance, credit, and technology adoption: Field experimental evidencefrom Malawi", *Journal of Development Economics*, 89: 1–11 .

[10] Giné, Xavier, Pamela Jakiela, Dean Karlan, and Jonathan Morduch. 2010. "Microfinance Games." *American Economic Journal: Applied Economics*, 2(3): 60-95.

[11] Holt, Charles A. and Susan K. Laury. 2002. "Risk Aversion and Incentive Effects." *American Economic Review*, 92(5): 1644-1655.

[12] Li, Rui and Xi Zhu. 2010. "Econometric Analysis of Credit Constraints of Chinese Rural Households and Welfare Loss." *Applied Economics.* 42:(13):1615-1625.

[13] Miranda, Mario J. and Claudio Gonzalez-Vega. 2011. "Systemic Risk, Index Insurance, and Optimal Management of Agricultural Loan Portfolios in Developing Countries". *American Journal of Agricultural Economics*. 93 (2): 399-406.

[14] OECD report (2004) "Rural Finance and Credit Infrastructure in China."

[15] Stiglitz, Joseph E. and Andrew M. Weiss. 1981. "Credit Rationing in Markets with Imperfect Information." *American Economic Review*, 71(3): 393-410.

[16] Udry, C., and J. Conning. 2007. " Rural Financial Markets in Developing Countries." In R. Even- son and P. Pingali, eds. *Agricultural Development: Farmers, Farm Production and Farm Markets*, pp. 2857–2908. Amsterdam: North Holland.

[17] Zhang, Bing, Guo-yu Xu, Ping-gui Zuo, Yang Cao, and Shuai Yuan. 2008. "Empirical study on the financial repression of rural households' debit and credit and the effect on their welfare in less developed regions—Take Suqian city of Jiangsu province as an example." *Journal of Modern Accounting and Auditing*. 4 (12): 27-34.

Figures and Tables

Figure 1: CDFs without Index Insurance

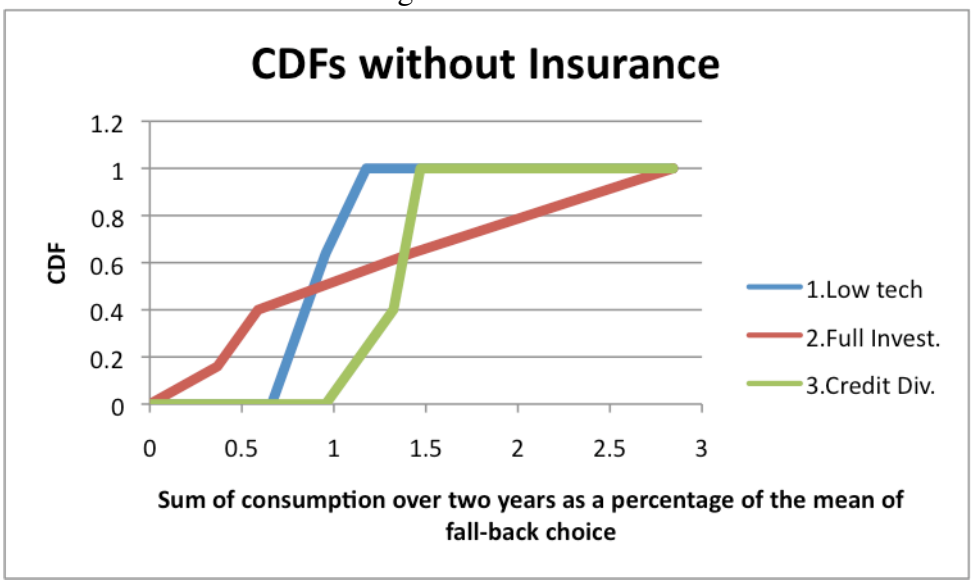


Figure 2: CDFs with Index Insurance

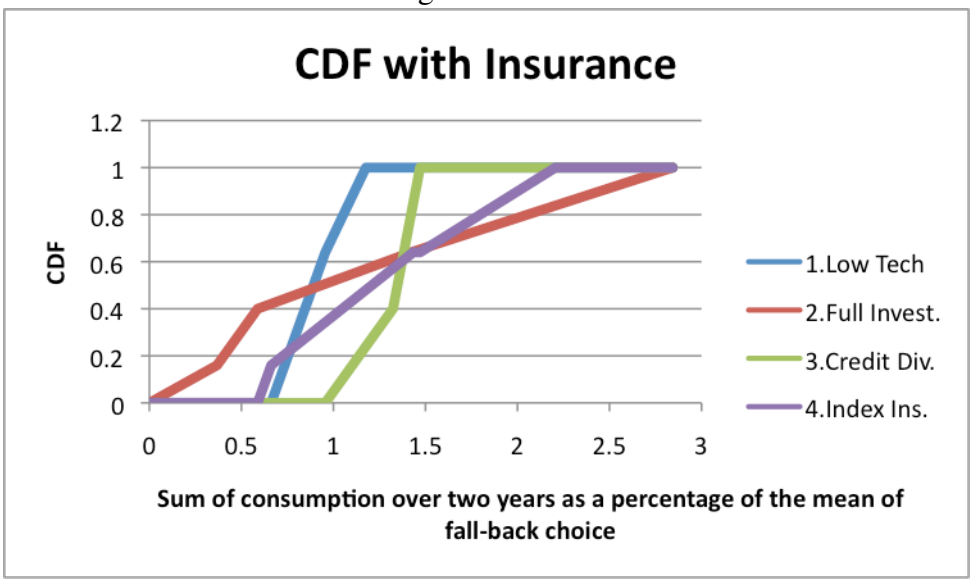
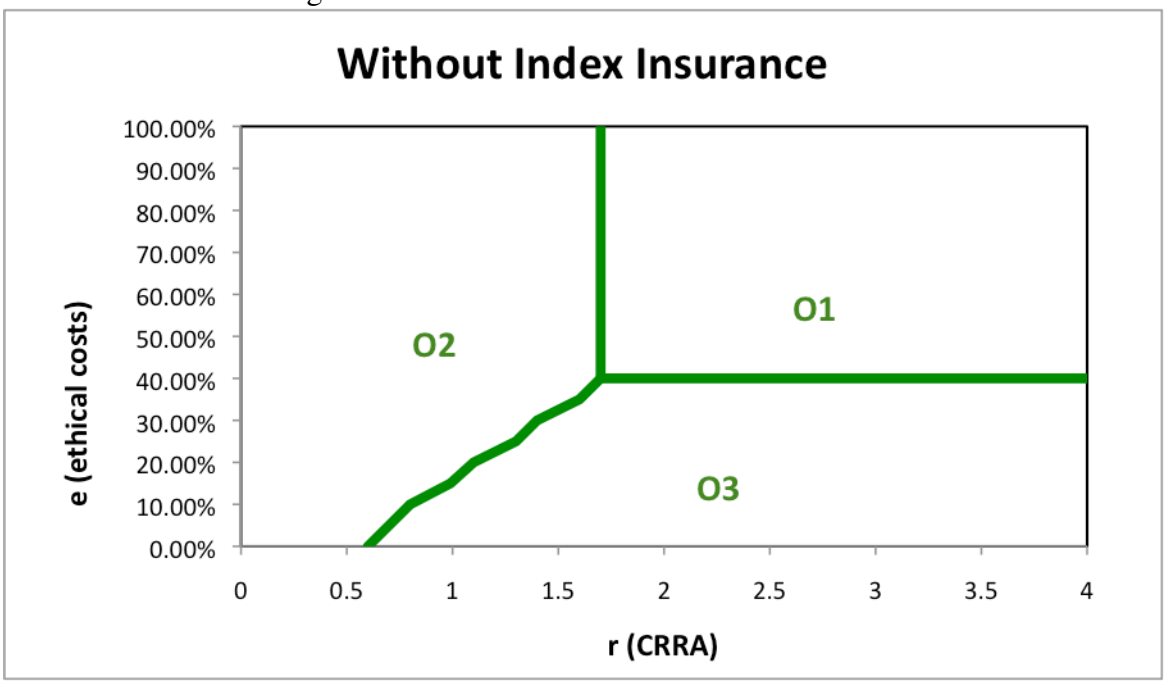


Figure 3: Choice distribution under no index insurance



Note: "O1", "O2" and "O3" represent Option 1, 2 and 3.

Figure 4: Choice distribution under index insurance

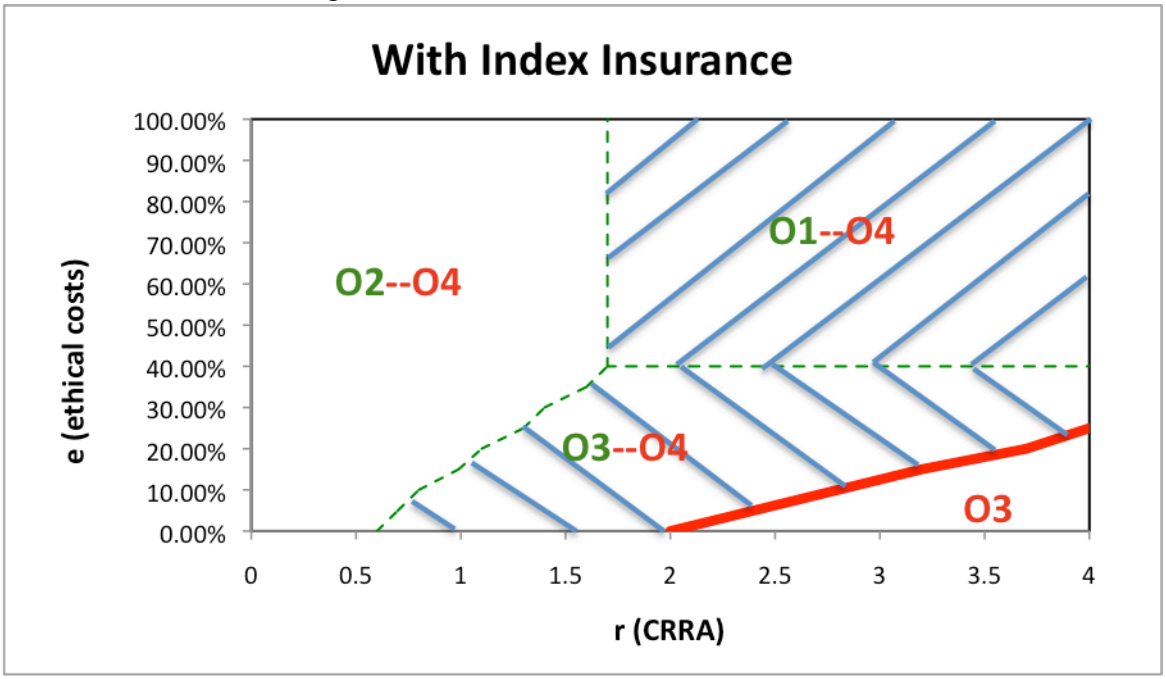


Figure 5: Study area

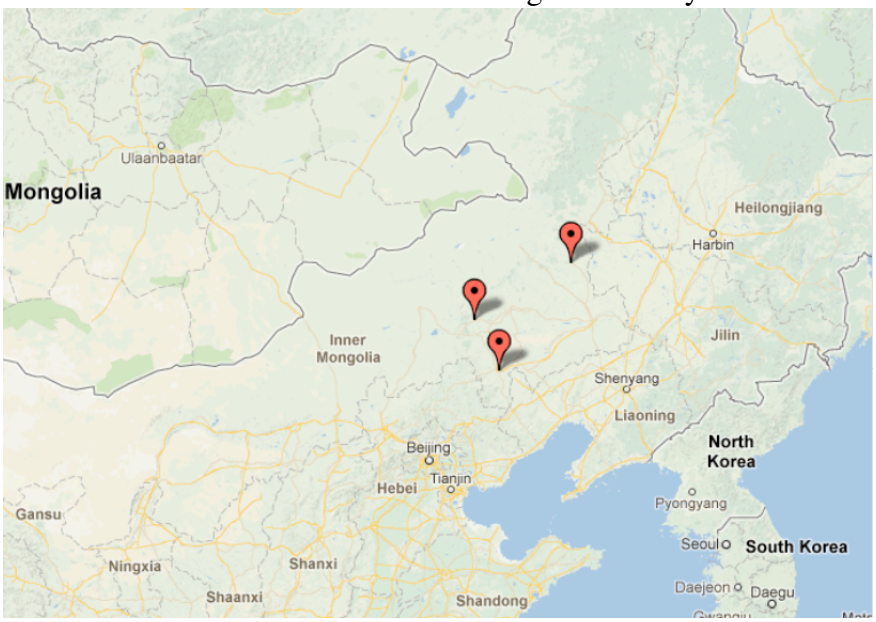


Figure 6: Distribution of estimated CRRA

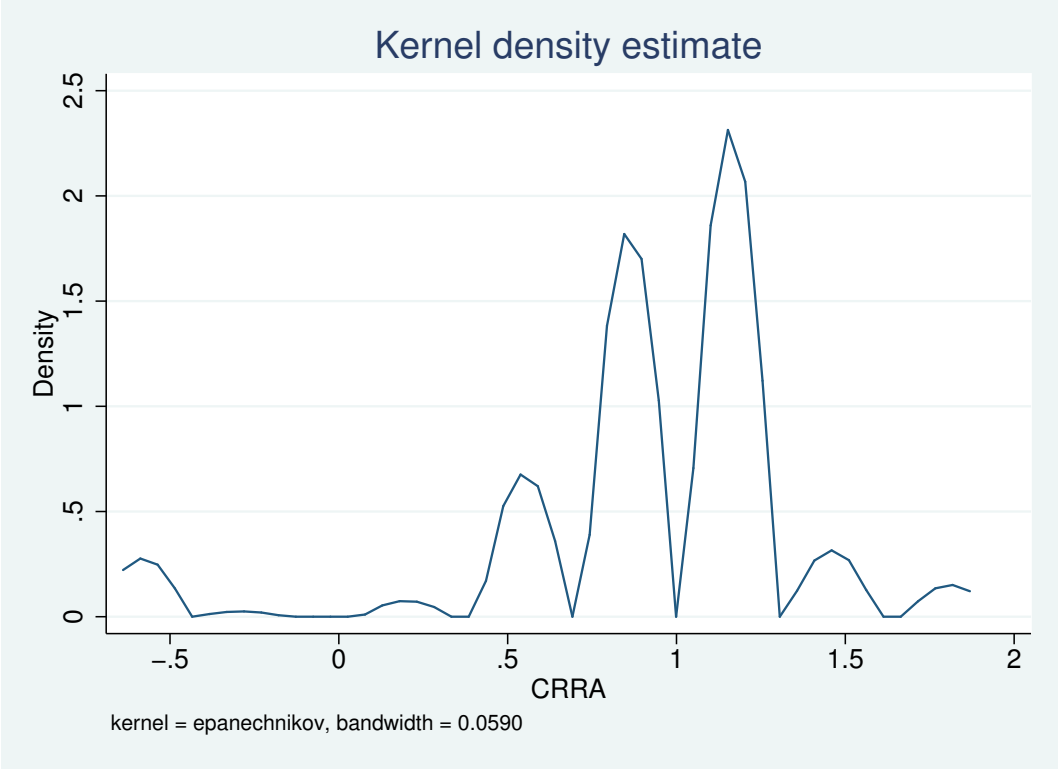


Table 1: Farmers' options when index insurance is not available

Options	θ	c_1	θ	c_2	E
1. Fall Back	1	y_m	1	y_m	0
			0		
	0		1		
			0		
2. Full Investment	1	$y_s^h(B)-(1+r)B$	1	$y_s^h(B)-(1+r)B$	0
			0	0	
	0	0	1	y_m	
			0		
3. Credit Diversion	1	$\frac{1}{2}B$	1	$\frac{1}{2}B$	e_i
			0		
	0	$\frac{1}{2}B$	1	y_m	
			0		

Table 2: Credit diversion with and without index insurance

	θ	c_1	θ	c_2
Credit Diversion	1	$\frac{1}{2}B$	1	$\frac{1}{2}B$
			0	
	0		1	y_m
			0	
Credit Diversion + Index Insurance	1	0	1	0
			0	$\frac{1}{2}B$
	0	$\frac{1}{2}B$	1	0
			0	$\frac{1}{2}B$

Table 3: Farmers' net income under different options

Options	θ	c_1	θ	c_2	E
Full Investment	1	$y_s^h(B)-(1+r)B$	1	$y_s^h(B)-(1+r)B$	0
			0	0	
	0	0	1	y_m	
			0		
Full Inv.+ Index Insur.	1	$y_s^h(B)-q-(1+r)B$	1	$y_s^h(B)-q-(1+r)B$	0
			0	$y_s^l+I-q-(1+r)B$	
	0	$y_s^l+I-q-(1+r)B$	1	$y_s^h(B)-q-(1+r)B$	
			0	$y_s^l+I-q-(1+r)B$	

Table 4: Farmers' options when index insurance is available

Options	θ	c_1	θ	c_2	E
1. Fall Back	1	y_m	1	y_m	0
			0		
	0		1		
			0		
2. Full Investment	1	$y_s^h(B)-(1+r)B$	1	$y_s^h(B)-(1+r)B$	0
			0	0	
	0	0	1	y_m	
			0		
3. Credit Diversion	1	$\frac{1}{2}B$	1	$\frac{1}{2}B$	e_i
			0		
	0	$\frac{1}{2}B$	1	y_m	
			0		
4. Full Inv.+ Index Insur.	1	$y_s^h(B)-q-(1+r)B$	1	$y_s^h(B)-q-(1+r)B$	0
			0	$y_s^l+I-q-(1+r)B$	
	0	$y_s^l+I-q-(1+r)B$	1	$y_s^h(B)-q-(1+r)B$	
			0	$y_s^l+I-q-(1+r)B$	

Table 5: Options in the experiment

Option	Project	Loan	Full Investment	Insurance
1	Maize	NO	-	NO
2	Sheep	YES	YES	NO
3	Sheep	YES	NO	NO
4	Sheep	YES	YES	YES

Table 6: Summary Statistics

Variable Name	Obs	Mean	Std. Dev.	Quantiles				
				Min	25%	Median	75%	Max
Demongraphic Characteristics								
Male	450	0.42	0.49	0	0	0	1	1
Age	450	45.32	11.32	17	39	45	53	80
Minority	450	0.38	0.49	0	0	0	1	1
Years of education	450	6.78	3.13	0	5	7	9	13
Household head	450	0.49	0.5	0	0	0	1	1
Household head spouse	450	0.44	0.5	0	0	0	1	1
Annual off-farming work (months)	450	1.09	2.44	0	0	0	1	12
Village leader	450	0.01	0.08	0	0	0	0	1
Member of CCP	450	0.04	0.2	0	0	0	0	1
Religion	450	0.04	0.2	0	0	0	0	1
Household size	450	3.91	1.29	1	3	4	5	9
Number of household labor	450	2.21	0.98	0	2	2	3	7
Children under middle school	450	0.41	0.64	0	0	0	1	3
Children above high school	450	0.32	0.61	0	0	0	1	4
Risk aversion	450	2.77	0.62	1	3	3	3	3
Maize								
Years of growing crops	450	22.26	12.07	0	15	20	30	65
Size of all crops in 2012	450	42.3	45.51	0	10	22	62	270
Years of growing maize during 2010-2012	450	2.64	0.94	0	3	3	3	3
Size of miaze in 2012	450	21.96	24.43	0	7	12	30	200
Average maize yield (500g/mu)	386	1051.11	496.54	200	500	1000	1500	2000
Maize yield risks during 2010-2011	450	0.75	0.85	0	0	0	2	2
Sheep								
Years of raising livestock	450	6.76	9.68	0	0	1	10	50
Value of all livestock in 2012	450	20245.8	62610.33	0	0	50	6200	5.20E+05
Years of raising sheep during 2010-2012	450	0.56	1.09	0	0	0	0	3
Average sheep stock	450	13.27	44.18	0	0	0	0	360
Value of sheep in 2012	89	80983.2	1.00E+05	1200	12000	30000	1.20E+05	3.60E+05
Risks of raising sheep during 2010-2011	450	0.09	0.38	0.00	0.00	0.00	0.00	2.00
Wealth and Income								
House ownership	450	0.92	0.28	0	1	1	1	1
House value	450	82625.6	89120.25	1000	30000	60000	1.00E+05	1.00E+06
Size of owned farmland	450	26.97	19.33	2	11	22.5	39.00	140
Size of owned forest	450	1.97	13.99	0	0	0	0.00	280
Size of owned grassland	450	15.55	39.55	0	0	0	27.00	700
Value of ag equip	450	5529.11	21663.27	0	0	2000	5000	3.10E+05
Value of durables	450	7597.61	32505.11	0	950	2600	5500	6.30E+05
Annual household income	450	36779.7	45029.85	1300	16300	26600	41500	6.40E+05
The share of agricultural income	450	0.71	0.29	0	0.49	0.86	0.95	1
Annual income per capita	450	10275.2	13182.58	260	4250	6833.3	11140	1.60E+05
Current cash amount	450	2725.56	5333.34	250	250	750	1500	2.00E+04
Current saving amount	450	6733.33	19888.22	0	0	0	2500	3.00E+05
Current lending amount	450	1858.89	9172.84	0	0	0	0	1.00E+05

Table 7: Formal and informal borrowers

	Formal loan borrower	Informal loan borrower	Non-borrower
Number of borrowers	57	178	240
% of the sample	12.7%	40.0%	53.3% [12]
Average number of the corresponding loans per borrower	1.12 (0.33)	1.71 (0.96)	–
Average amount of the corresponding loans per borrower (RMB)	21192.98 (13787.21)	23153.37 (31971.16)	–

Note: Standard deviation reported in parentheses

Table 8: The share of credit rationing

Credit Constrained			Credit Unconstrained	
Quantity	Risk	Transaction costs	Borrower	Non-borrower
35.9%	2.0%	6.2%	3.3%	52.5%
44.1%			55.9%	

Table 9: Formal and informal loan terms and purpose

	Formal loans	Informal loans
Ave. annual interest rates per loan	12.2% (0.04)	4.0% (0.08)
Ave. loan size (RMB)	18875 (10950.83)	13432.67 (24123.21)
Ave. loan duration (months)	14.64 (8.55)	–
% of installment loans	45.3%	–
% of collateralized loans	20.3%	1.3%
% of loans with guarantors	32.8%	4.6%
% of loans with group liability	42.2%	0%
% of default loans	7.8%	4.3%
% of loans for consumption purpose	15.6%[13]	65.3%
% of loans for production purpose	91%	40.3%[14]

Note: Standard deviation reported in parentheses

Table 10: Experiment results

Option	Control Game	Treatment Game	Treatment effect
1	*96* *(21.3%)*	*49* *(10.9%)*	-47 (-10.4%)
2	*284* *(63.1%)*	*55* *(12.2%)*	
3	*70* *(15.6%)*	*17* *(3.8%)*	-53 (-11.8%)
4	–	*329* *(73.1%)*	

Table 11: Average treatment effect of index insurance on credit diversion and risk rationing

	Dependent Variable	
	Credit diversion	Risk rationing
treatment	-0.118***	-0.104**
	(0.024)	(0.039)
CRRA	-0.004	0.009
	(0.030)	(0.043)
Age	0.002	0.007***
	(0.002)	(0.002)
Education	0.001	-0.007**
	(0.003)	(0.003)
CCP	0.044	-0.135***
	(0.059)	(0.038)
HH size	-0.034**	-0.007
	(0.016)	(0.016)
Years of cropping	-0.003*	-0.004**
	(0.001)	(0.002)
Size of cropping	0.0009*	-0.0008
	(0.0005)	(0.0006)
Years of raising sheep	0.012	-0.062***
	(0.021)	(0.015)
Amount of formal loans (1000 yuan)	-0.002	0.005*
	(0.002)	(0.002)
Number of informal loans	-0.022	-0.046**
	(0.021)	(0.019)
Amount of informal loans (1000 yuan)	0.002**	-0.0005
	(0.0007)	(0.0005)
Lending (1000 yuan)	0.0005	0.005**
	(0.001)	(0.002)
Size of owned grassland	-0.0004**	0.0002
	(0.0002)	(0.0004)
Value of durables (1000 yuan)	0.0005	-0.0003*
	(0.0004)	(0.0001)
Saving (1000 yuan)	-0.0007	-0.001*
	(0.0005)	(0.0006)
Observations	900	900
R-squared	0.109	0.137

Table 12: Relation between credit diversion history and demographic characteristics

VARIABLES	Full sample	Partial sample
CRRA	0.191***	0.169**
	(0.0613)	(0.0782)
HH head	-0.216*	-0.207
	(0.128)	(0.132)
HH spouse	-0.120	-0.119
	(0.124)	(0.126)
Gender	0.0462	0.0476
	(0.0977)	(0.0992)
Age	0.0172**	0.0173**
	(0.00712)	(0.00720)
Minority	0.167**	0.174**
	(0.0719)	(0.0753)
Education	-0.0111	-0.0105
	(0.0143)	(0.0146)
Off-farm work (months)	0.00184	0.00239
	(0.0111)	(0.0112)
Village leader	0.301	0.292
	(0.248)	(0.251)
CCP	-0.123	-0.117
	(0.129)	(0.131)
Religion	0.384**	0.380**
	(0.155)	(0.157)
HH size	0.0285	0.0275
	(0.0504)	(0.0506)
HH labor	-0.0271	-0.0228
	(0.0413)	(0.0432)
Chidlren under high school	0.0771	0.0759
	(0.182)	(0.182)
Children above high school	0.0438	0.0563
	(0.0652)	(0.0766)
Dependent ratio	-0.432	-0.411
	(0.702)	(0.705)
Area1	-0.0749	-0.0874
	(0.258)	(0.258)
Area2	0.124	0.107
	(0.230)	(0.229)
Observations	172	169
R-squared	0.423	0.424

Table 13: Heterogeneous treatment effect of index insurance on credit diversion under full sample

		Quintiles of Estimated Ethical Costs				
VARIABLES	All	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
treatment	-0.144**	-0.133	-0.133	-0.174	-0.186**	-0.122*
	(0.061)	(0.088)	(0.084)	(0.125)	(0.082)	(0.067)
treatment*CRI	0.0255	0.0472	0.0121	0.0609	-0.0804	-0.138*
	(0.054)	(0.082)	(0.100)	(0.145)	(0.078)	(0.080)
treatment*CRI	0.00249	-0.00652	0.0314	-0.0446	0.135**	0.129
	(0.032)	(0.063)	(0.057)	(0.109)	(0.057)	(0.102)
CRRA	-0.0178	0.0774	-0.0385	-0.0578	-0.0393	-0.0617
	(0.053)	(0.060)	(0.063)	(0.138)	(0.059)	(0.066)
Observations	900	180	180	180	180	180
R-squared	0.11	0.359	0.348	0.28	0.432	0.298

Table 14: Heterogeneous treatment effect on credit diversion under partial sample

		Quintiles of Estimated Ethical Costs				
VARIABLES	All	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
treatment	-0.177**	-0.256**	-0.346*	-0.243*	-0.184**	-0.0935
	(0.0672)	(0.117)	(0.199)	(0.140)	(0.0824)	(0.0634)
treatment*CRI	0.0633	0.204	0.453	0.102	-0.129	-0.161**
	(0.0656)	(0.189)	(0.329)	(0.160)	(0.0780)	(0.0779)
treatment*CRI	-0.00287	-0.0334	-0.176	-0.0340	0.178*	0.132
	(0.0345)	(0.104)	(0.148)	(0.0984)	(0.0888)	(0.0985)
CRRA	-0.0450	-0.0539	-0.0227	-0.0499	0.0291	-0.00611
	(0.0611)	(0.0908)	(0.129)	(0.130)	(0.0750)	(0.0977)
Observations	882	178	176	176	176	176
R-squared	0.113	0.384	0.360	0.326	0.433	0.309

Table 15: Heterogeneous treatment effect of index insurance on risk rationing under full and partial sample

VARIABLES	Full Sample	Partial Sample
treatment	-0.0725	-0.0184
	(0.093)	(0.057)
treatment*CRRA	-0.156	-0.184**
	(0.125)	(0.087)
treatment*CRRA^2	0.107	0.0863
	(0.065)	(0.057)
CRRA	0.0284	0.103*
	(0.070)	(0.054)
Observations	900	882
R-squared	0.143	0.141

Risk and Economic Under-Specialization: Why the Pin-Maker Grows Cassava on the Side

Ajay Shenoy*
University of Michigan, Ann Arbor

December 28, 2012
First Version: October 08, 2012

Abstract

Why do the poor have so many economic activities? One prominent theory states the poor inefficiently diversify to self-insure against the risk of relying on one income source. I test the theory by measuring the response of Thai rice farmers expecting a harvest to conditional volatility in the international rice price. Among uninsured households a 10% decrease in volatility increases specialization as much as a 4% increase in returns. I use the resulting exogenous variation to confirm extra activities are costly. I find no evidence for an alternative theory: that under-specialization is caused by the need for lumpy investments.

[PRELIMINARY AND INCOMPLETE]

*Email at shenoy@umich.edu. File formatting based on a stylesheet by Charles Jones. Special thanks to Raj Arunachalam, David Weil, Jeff Smith, David Lam, Stefanie Stantcheva, and many seminar participants at Michigan and Brown for helpful comments and advice.

1. Introduction

> *To take...the trade of the pin-maker...in the way in which this business is now carried on...it is divided into a number of branches...One man draws out the wire, another straights it, a third cuts it...I have seen a small manufactory of this kind where ten men only were employed... Those ten persons, therefore, could make among them upwards of forty-eight thousand pins in a day...But if they had all wrought separately and independently...they certainly could not each of them have made twenty, perhaps not one pin in a day...*
>
> -Adam Smith, *Wealth of Nations*

What made the Industrial Revolution so revolutionary? Though the steam engine and power loom were remarkable innovations, the most revolutionary breakthrough was one of management rather than machinery: economic specialization. The specialization revolution illustrated in Smith's parable spread beyond the assembly line and the factory. In a modern economy, most individuals channel their talents towards a single profession that makes a single portion of a single product. Like Smith's pin-maker we specialize to raise our collective output.

But households in the developing world do not specialize. Few poor households in the developing world rely on just one economic activity for their income (**?**). But if Smith and his pin-maker are right, the overabundance of activities must cost poor households foregone income. Why might the poor, who by definition have little income to spare, leave any on the table through inefficient under-specialization?

One theory is that having many economic activities, like diversifying a stock portfolio, will hedge against risk. When the returns to any activity are uncertain, perfect specialization is as dangerous as pouring a lifetime's savings into one stock. Lacking the privilege of steady jobs and unemployment insurance, the poor choose quite sensibly to spread risk. Economists have speculated that risk might explain under-specialization, but have also proposed other theories. One alternative theory states that expanding any activity to a profitable scale requires lumpy investments; for example, becoming a large-scale tailor requires

a sewing machine. Poor households lack the credit to make these investments. Unable to draw on one large income source, their only option is to pool income from many small sources. These two theories suggest very different constraints on poor households and very different policy recommendations to improve social welfare. But until now, neither theory has been formalized and rigorously tested.

I formalize the intuition of risk theory in a simple model that makes testable predictions. A household must choose how many side activities to pursue and how to allocated its labor endowment between them and its primary activity. It makes its decisions before the returns to any activity are realized, turning its revenue stream into a gamble whose variance falls with the number of activities. Weighed against the safety of more activities is the fixed cost it must pay for each extra activity. The model predicts a household responds to greater riskiness in the returns to its primary activity by increasing its number of side activities; it makes several more predictions about the nature of this response and the effect on its revenue.

To test the model I observe the response of Thai rice farmers to changes in the mean and volatility of the international rice price. I model volatility in rice prices as a first-order ARCH process and use its predicted values as a monthly measure of volatility. I identify households in a monthly survey who expect a rice harvest in the near future. Higher volatility in the price of rice directly raises the riskiness of these households income, and their differential response to volatility (controlling for the mean price) identifies the causal effect of riskier income on economic specialization. I compare the response of harvest-expecting households to that of households who do not farm rice and households that do but expect no immediate harvest. The estimated triple-difference isolates the response to a riskier income from the response to a more variable cost of consumer goods and unobserved differences between rice farmers and non-rice farmers. I further test the model by seeing if the response to risk is strongest among households with poor insurance, where insurance is the tendency to receive money from other households when one's own revenue is low. I find suggestive evidence to support this prediction, and estimate that the response to 10% less volatility equals the response to 4% higher returns among poorly insured households.

Since a harvest-expecting household does not yet have its crop in hand, the mean and variance of the rice price cannot directly affect revenue. Then the variation they induce in the number of economic activities is exogenous and can form a valid first-stage in an IV regression of total revenue on the number of economic activities. Whereas OLS regressions indicate that more economic activities predict higher revenue, IV estimates reverse the sign of the effect; a failure to specialize costs households income exactly as the model predicts. The upward bias of OLS confirms another of the model's predictions.

Finally, I test the alternative theory of under-specialization the theory of lumpy investments and credit constraints using a quasi-experimental credit injection. I exploit the Million Baht Program, used most prominently by **?**, to test whether a relaxation of credit constraints causes a decrease in the number of businesses. I find no evidence to support the lumpy investment theory.

2. Literature

Economists have long studied the effects of risk on household decision-making, so I restrict my discussion to the most relevant studies from developing world contexts. **?** use variability in the onset of monsoon rains to test the effect of risk on profitable agricultural investment. They find farmers in riskier areas choose safer but less profitable bundles of investment unless they are wealthy enough to cope with income shocks. **?** examine the effects of risk on human capital investment. They use a theoretical model and expected versus unexpected shocks to separate the effects of imperfect credit from imperfect insurance on children's schooling. They find neither credit nor insurance markets function perfectly.

The study of income diversification in the developing world is much more limited. Though many scholars—most visibly **?**—have reported the lack of specialization among the poor, few studies have attempted to identify its causes. **?** use rainfall variability to test if risk affects specialization. They use occupational focus—whether members of a household choose the same profession—as their measure of specialization. They find lower focus in places with higher rainfall variability except in areas where flooding is common, arguing that flooding re-

duces a farmer's dependence on rainfall. They also find some evidence that forms of insurance like a government safety net mitigate the effect of rainfall risk. They give evidence that occupation focus may increase consumption. **?** also uses rainfall variability to study specialization and finds similar effects in Nepal. Many other scholars have examined the effects of climate variability on occupational choice.

But neither of these studies can deal with the possibility that different kinds of people live in places with highly variable rainfall, or that rainfall variability affects aspects of life beyond income risk. My triple-difference estimator isolates the effect of short-run income risk. I also focus on number of activities rather than occupational focus because it more directly answers my original question: why the poor have so many economic activities.

3. A Model of Risk and Specialization

In this section I formalize the intuition of costly under-specialization with a simple model. The model makes testable predictions I confirm in the empirical section of the paper.

3.1. Setup

Each household has a single primary economic activity and the option to take up secondary (or side) economic activities. The household has a single unit of labor; it must choose both the number of side activities (if any) and how much labor it allocates to each activity. Labor allocated to an activity produces income at constant returns; the marginal return for any activity is unknown when the household chooses its side activities and labor allocations. Finally, the household must pay a fixed-cost for each new side activity it undertakes.

The household's objective is

$$\max_{M, L^p, \{L^s_m\}} \mathbb{E}[-e^{-\alpha C}]$$

subject to the constraints

$$C = Y = w^p L^p + \sum_{m \in M} w_m^s L_m^p - MF$$
$$L^p + \sum_m L_m^s = 1$$

The choice variables are M, the number of side activities (if any); L^p, the fraction of the labor endowment devoted to the primary economic activity; and $\{L_m^s\}_{m \in M}$, the fraction of labor devoted to each side activity. Households have constant absolute risk-aversion preferences with parameter α. The household simply consumes its income, which is the sum of income from primary (p) and side (s) activities minus the total fixed cost it pays. The returns to the primary and secondary activities are w^p and $\{w_m^s\}_{m \in M}$. These are independent normal random variables with $w^p \sim N(\bar{w}^p, \sigma_p^2)$ and $w_m^s \sim N(\bar{w}^s, \sigma_s^2)$ for all m. Assume $w^p \geq w^s$, so greater specialization will not lower expected revenue.

Without any algebra it is clear that if returns to the primary activity were certain—from a salaried (preferably government) job—the household would perfectly specialize ($M = 0$). Specialization spares the household fixed-costs and may earn the household higher returns to labor. In this simple universe, the only reason for side economic activities is to reduce the variance of income.

3.2. Example: Choosing Between Specialization and One Side Activity

Focus for now on the simplified case where $\bar{w}^p = \bar{w}^s = \bar{w}$ and the household can only choose between no side activities and a single side activity. The household effectively must choose between two "bundles" of mean of consumption (call it $\bar{C}$) and variance of consumption (call it V):

	$\boldsymbol{M = 0}$	$\boldsymbol{M = 1}$
$\boldsymbol{\bar{C}}$	$\bar{w}$	$\bar{w}$ - F
$\boldsymbol{V}$	σ_p^2	$(L^p)^2\sigma_p^2 + (1 - L^p)^2\sigma_s^2$

Since $L^p < 1$, the variance of consumption with a side activity is lower than the variance under perfect specialization. With a positive fixed cost the house-

hold must trade a higher expected consumption for a less variable level of consumption.

Suppose the household has chosen to pursue a side activity and must now choose its optimal labor allocation conditional on that choice. Since consumption is a normal random variable, expected utility is (the negative of) a log normal random variable. The household's intermediate problem is

$$\max_{L^p} -e^{-\alpha\bar{C}+\frac{\alpha^2}{2}V}.$$

The first-order condition is

$$\begin{aligned} & 0 = -\frac{\alpha^2}{2}e^{-\alpha\bar{C}+\frac{\alpha^2}{2}V} \cdot \frac{\partial V}{\partial L^p} \\ \Rightarrow \quad & 0 = L^p\sigma_p^2 + (1-L^p)\sigma_s^2 \\ \Rightarrow \quad & L^p = \frac{\sigma_s^2}{\sigma_p^2+\sigma_s^2} \end{aligned}$$

The next step is to find a condition stating when the household wants to enter a side activity. To avoid a degenerate outcome (everyone either specializes or diversifies), consider heterogeneity in the fixed cost of the side activity. Suppose F is uniformly distributed on the interval $[0, \mathcal{F}]$ for some upper-bound $\mathcal{F}$. Then the optimal decision rule is a threshold condition; all households with a fixed cost below the threshold (call it $\bar{F}_0$) will start a side activity.

Write the mean and variance of consumption as functions of the number of side activities $(\bar{C}(M), V(M))$. The household whose fixed cost is exactly $\bar{F}_0$ is indifferent between zero and one side activities:

$$\begin{aligned} & -e^{-\alpha\bar{C}(0)+\frac{\alpha^2}{2}V(0)} = -e^{-\alpha\bar{C}(1)+\frac{\alpha^2}{2}V(1)} \\ \Rightarrow \quad & -\alpha\bar{C}(0) + \frac{\alpha^2}{2}V(0) = -\alpha\bar{C}(1) + \frac{\alpha^2}{2}V(1) \\ \Rightarrow \quad & \frac{\alpha}{2}[V(0)-V(1)] = \bar{C}(0) - \bar{C}(1) \end{aligned}$$

Substitute the expressions from the table above and rearrange to find the threshold fixed cost:

$$\bar{F}_0 = \frac{\alpha}{2} \frac{\sigma_p^4}{\sigma_p^2 + \sigma_s^2}$$

The threshold rises with the variance of the primary activity σ_p^2, and Figure **??**.a shows the effect on the number of households with a side activity. When their primary activity becomes riskier, households are willing to pay more to hedge their bets; hence the threshold rises. The mass of households with fixed costs between the old and new thresholds then enter a side activity. The change in the average number of activities is

$$\frac{\partial \mathbb{E}[M]}{\partial \sigma_r^2} = \frac{\alpha}{2} \cdot \frac{\sigma_r^2 \sigma_r^2 + 2\sigma_s^2}{2\left(\sigma_r^2 + \sigma_s^2\right)^2} \cdot \frac{1}{\mathcal{F}} > 0$$

This first prediction of the model is just a formalization of the intuition of the introduction:

Test 1 (Risk) *Households increase their number of economic activities in response to riskier returns to their primary activity.*

3.3. Adding Insurance

A household trapped in the model of Section **??** has no choice but to insure against risk with additional activities. But real households often have access to some other means of insurance. Workers in the rich world enjoy unemployment compensation, and households throughout the world can often count on friends and family to send money when needed.

At its core, insurance is simply a means of reducing the dependence of consumption on (transitory) income. I accordingly model it by adjusting household consumption to be

$$C = pY + (1 - p)\, y^p$$

Consumption is now partly drawn from current revenue and partly from the household's (constant) permanent income, where a smaller portion due to current revenue (smaller ρ) corresponds to a higher degree of insurance. For simplicity I do not model the source of the insurance (in the empirical section I use transfers from outside the household). I also assume permanent income

has the same expectation as current income (as must be true if income is stationary). Then the threshold for the cost of entering a new activity—that is, the largest cost a household is willing to bear for switching from perfect specialization to having a single side activity—becomes

$$\bar{F}_0 = \rho^2 \frac{\alpha}{2} \frac{\sigma_r^4}{\sigma_r^2 + \sigma_s^2}$$

Assuming as before the fixed cost is distributed uniformly over $[0, \mathcal{F}]$, a rise in the variance of the primary activity causes a smaller average increase in number of activities whenever $\rho < 1$:

$$\frac{\partial \mathbb{E}[M]}{\partial \sigma_r^2} = \rho^2 \cdot \frac{\alpha}{2} \cdot \frac{\sigma_r^2 \sigma_r^2 + 2\sigma_s^2}{2\left(\sigma_r^2 + \sigma_s^2\right)^2} \cdot \frac{1}{\mathcal{F}}$$

The expression provides a new testable prediction:

Test 2 (Insurance) *Uninsured households respond more to a rise in the variance of the primary activity than insured households.*

The prediction is hardly counter-intuitive: more outside insurance means less need for insurance through costly under-specialization. In the limit when insurance is perfect ($\rho = 0$), a change in the variance has no effect and a household will behave as though risk-neutral. Given its simplicity, this prediction is also a reality check—it must hold if the effects I find in the empirical portion of the paper are actually responses to risk.

3.4. The Effect of Higher Expected Returns

The empirical approach in Section **??** exploits variation in both the volatility of returns and the expected returns. To derive a prediction about changes in expected returns, relax the assumption that expected returns to the primary and secondary activities are equal and call the expected premium $w^+ = \bar{w}^p - \bar{w}^s$. To make the model interesting, assume $\alpha\sigma_p^2 - w^+ > 0$. This means the average premium—the extra marginal gains to specialization—are not too large relative to the utility cost of the riskiness of the primary activity. If the assumption does not hold, the additional gains from the primary activity are so attractive the household specializes in a single activity despite the risk.

The labor allocation to the primary activity conditional on having a side activity is now

$$L^p = \frac{\alpha\sigma_s^2 + w^+}{\alpha\left(\sigma_p^2 + \sigma_s^2\right)}$$

which is strictly less than one under the assumption on the average premium. The threshold cost for having a side activity is now

$$\bar{F}_0 = \frac{1}{2\alpha} \cdot \frac{\left(\alpha\sigma_p^2 - w_+\right)^2}{\sigma_p^2 + \sigma_s^2}$$

Then the change in the average number of activities is

$$\frac{\partial \mathbb{E}[M]}{\partial w^+} = \frac{w_+ - \alpha\sigma_p^2}{\alpha\sigma_p^2 + \alpha\sigma_p^2} < 0$$

In words the result states that a higher marginal return to the primary activity causes households to abandon their side activity and focus on the primary activity. Intuitively a higher premium raises the opportunity cost of self-insurance through an extra activity, and fewer households are willing to pay this cost.

Test 3 (Returns) *Households decrease their number of economic activities in response to higher expected returns to their primary activity.*

3.5. Linking Economic Activities to Revenue

By construction, taking on additional side activities lowers the household's total revenue. But the empirical strategy of Section **??** examines rice farmers who expect but have not yet collected a harvest; it cannot test predictions about total revenue by construction. Instead I formalize several statements about the effect of additional activities on revenue from side activities.

Consider again the simple case where returns are equal but suppose households can choose any number of activities. The general expression for each threshold is

$$\bar{F}_M = \frac{\alpha\sigma_p^4\sigma_s^2}{2\left(M\sigma_p^2 + \sigma_s^2\right)\left((1+M)\sigma_p^2 + \sigma_s^2\right)}$$

Let $y^s = \sum_{m \in M} w_m L_m^p M F$ denote the total revenue from side activities. For simplicity treat the number of activities M as continuous; holding a household's cost of additional activities fixed, an infinitesimal change in the number of activities changes side revenue on average by

$$\mathbb{E}\left[\frac{\partial y^s}{\partial M}\right] = -\mathbb{E}[F] + \mathbb{E}\left[\frac{w\sigma_p^2\sigma_s^2}{\left(M\sigma_p^2 + \sigma_s^2\right)^2}\right]$$

This is not equal to the fixed cost because a change in the number of activities also induces a change in labor allocated to side activities. This means the average derivative, which corresponds to the instrumental variables coefficient I estimate in Section **??**, might not be negative: a decrease in side revenue from additional activities implies a cost to specialization, yet the converse need not be true. But for the current endeavor equivalency is not necessary—a negative response will validate the idea of costly under-specialization.

Test 4 (Cost) *The average effect on revenue of more activities is negative only if under-specialization is costly.*

This last test draws insight from an instrumental variables regression, but the model also makes a prediction about the ordinary least squares coefficient. The OLS coefficient estimates the average effect of increasing the number of activities without holding their cost fixed; that is, it estimates the average total derivative:

$$\begin{aligned}
\mathbb{E}\left[\frac{dy^s(M,F)}{dM}\right] &= \mathbb{E}\left[\frac{\partial y^s}{\partial M} + \frac{\partial y^s}{\partial F}\cdot\frac{\partial F}{\partial M}\right] \\
&= \mathbb{E}\left[\frac{\partial y^s}{\partial M}\right] + \mathbb{E}\left[\frac{\partial y^s}{\partial F}\cdot\mathbb{E}\left[\frac{\partial F}{\partial M} \mid M\right]\right] \\
&= \mathbb{E}\left[\frac{\partial y^s}{\partial M}\right] + \mathbb{E}\left[\frac{\partial y^s}{\partial F}\cdot\frac{\partial}{\partial M}\mathbb{E}\left[F \mid M\right]\right]
\end{aligned}$$

The term $\frac{\partial y^s}{\partial F}$ is clearly negative. The term $\frac{\partial}{\partial M}\mathbb{E}\left[F \mid M\right]$, which is the change in the average fixed cost of households with more and more activities, is also negative. Figure **??** illustrates why: one household will have more activities than

another only if each additional activity is less costly. The number of activities is informative about their cost—indeed, $\mathbb{E}[F \mid M]$ is nothing more than the demand curve for insurance through economic under-specialization. Like any demand curve its slope is negative. This means

$$\begin{aligned}\beta_{OLS} &= \mathbb{E}\left[\frac{\partial y^s}{\partial M}\right] + \mathbb{E}\left[\frac{\partial y^s}{\partial F} \cdot \frac{\partial}{\partial M}\mathbb{E}[F \mid M]\right] \\ &> \mathbb{E}\left[\frac{\partial y^s}{\partial M}\right] \\ &= \beta_{IV}\end{aligned}$$

which is the final theoretical test of the model:

Test 5 (OLS Bias) *The OLS estimate of the effect of additional activities on side revenue is biased positively compared to the IV estimate.*

For the right set of parameters the bias can be strong enough to make the OLS coefficient positive, which is exactly what I find in Section **??**.

4. Data

My main source of household panel data is the Townsend Thai Monthly Survey (**?**), and I supplement the monthly data with the simultaneously collected Annual Survey (**?**). A baseline household survey administered in May 1997 reported on over two thousand rural households throughout four provinces in rural Thailand, and the annual resurvey series tracked households from one-third of the districts originally surveyed in each province through 2010. Of the districts not included in the annual resurvey, one from each province was chosen for the monthly survey. The households in these districts were surveyed every month to document changes in their incomes, crop conditions, and many other characteristics. I use the monthly price of rice from the IMF's commodity price dataset, and a monthly consumer price index from the Bank of Thailand to deflate all currency variables.

I use the monthly data for the risk response tests. I labeled a household as a rice farmer if it harvested rice at any point in the sample. I labeled a household

as expecting a harvest if it harvested rice in the next 1-3 months. I labeled a household as having had a recent harvest if it harvested rice in the previous 0-3 months (this includes a harvest in the current month). I defined the number of economic activities as the sum of the number of "large" businesses, crop-plots cultivated, types of livestock raised, number of jobs held by all members, number of miscellaneous or small businesses, and an indicator for whether the household engages in acquaculture (raising fish or shrimp). I define the household's total revenue as the sum of revenue from all the components of the number of economic activities. Finally, in merging aggregate time series data with the household survey and defining month dummies I adopt this convention: a household-month observed through a survey administered in the first half of the month is tagged as occurring in the previous month. Since the rice price is a monthly average, this convention better reflects the actual time span of the response period.

I use the annual data for the lumpy investment tests. I constructed the number of activities to be as similar as possible to the measure in the monthly sample: the sum of the number of large businesses, crop-plots, jobs, herds, an indicator for aquaculture, and a subset of the miscellaneous income sources. [1]

5. Suggestive Evidence of Risk and Under-Specialization

The model suggests risk will cause households to under-specialize under three conditions:

1. Risky revenue
2. Imperfect insurance
3. A tendency to reduce consumption risk with under-specialization

[1] Miscellaneous income sources in the annual survey often include remittances and other sources that do not meet my definition of economic activities (namely, revenue generating activities that require labor). I filter these unwanted sources using regular expressions on the textual descriptions of sources. The 1999 survey unfortunately does not contain textual descriptions, but the year dummies in the annual regressions should account for any 1999-specific measurement error.

If any of these conditions fails, Section **??** would find no causal effect of risk on economic activities. But the quasi-experimental variation of Section **??** uses short-run variation—changes in monthly price volatility—to measure effects on short-run outcomes—number of activities worked in a month. The short-term and long-term phenomena of under-specialization have similar mechanisms,. The model of Section **??** specifies no time horizon, so significant results in the short-run are informative about the long-run.

But knowing similar mechanisms determine the long- and short-run is not the same as seeing their effects on daily life. In this section I present basic in-equilibrium evidence to clarify the role of risk and under-specialization in the lives of my sample. Economic under-specialization is a defining characteristic of the poor; if the theory is true, then the three conditions above must be visible in a poor household's every-day life.

5.1. Risky Revenue

Do households suffer risk in their revenue? Figure **??** shows the average coefficient of variation of revenue for four dichotomies of households: rice farmers versus non-rice farmers; households with steady jobs versus those without; households with cash savings versus those without; and households in the top quartile of cash savings versus everyone else. More precisely, I compute the coefficient of variation of total revenue for each household across all months where it meets or does not meet the indicated criterion (e.g. holds a steady job). I then average the coefficients across all households.

No one—not even households with steady jobs—has completely stable revenue. Even households with steady jobs see fluctuations in their revenue nearly equal to their average revenue. Those without steady jobs can expect their revenue to vary twice as much from month-to-month. Rice farmers unsurprisingly suffer a much more variable stream of revenue than non-rice farmers. Finally, individuals with high cash savings enjoy a much less variable revenue stream. The most likely explanation is that individuals with more stable revenue are able to accumulate enough cash reserves to reach the top quartile of savers.

5.2. Imperfect Insurance

Will the fluctuations in revenue from Section **??** affect household consumption? Figure **??** graphs the spontaneous responses of households to the question "What did your household do in the worst year [for income] of the last five to get by?"[2] By far the most popular response was to take on an extra occupation, followed by working harder than usual. These responses do not prove households respond to risk through under-specialization, only that they respond to shocks through under-specialization. But if households must smooth their revenue by varying labor supply, they must not have access to any better smoothing mechanism. Borrowing to counter the shock is only the third most popular response and using savings only the fifth. The fourth most popular response is to consume less, meaning many households lack even a second-rate mechanism to smooth consumption.

Figure **??** provides more direct evidence: the correlation between revenue and consumption. For each household I compute the correlation coefficient between monthly revenue and consumption expenditure. If a risk-averse household has perfect insurance, its consumption should be independent of its current revenue; in fact, it should be constant. A household without perfect insurance must consume less when its revenue falls, driving the revenue-consumption correlation above zero. A higher correlation is evidence of worse insurance. The figure plots estimated kernel densities of household risk-consumption correlations among rice farmers and non-rice farmers. Since zero is modal it appears many households do have near-perfect insurance, but many more do not. The distribution is heavily skewed to the right (less insurance), and rice farmers in particular have worse insurance overall.

5.3. Risk and Under-Specialization

Finally and most importantly, do households respond to consumption risk through under-specialization? A simple test is to compare the economic activities of rice

[2]This question is from the 1997 baseline household survey. Responses are "spontaneous" in that households responded without being read a list of options. I graph only the seven most popular responses; as expected for an unprompted question, there were nearly one hundred unique responses.

farmers to non-rice farmers. Recall that rice farmers have more variable revenue and less effective insurance than everyone else (see Sections **??** and **??**). The theory claims they should have more economic activities. The histogram in Figure **??** agrees: rice farmers are more likely to have a large number of economic activities than everyone else.

The final test is the simplest of all: to check the correlation between revenue variability and number of economic activities. Figure **??** plots each household's average number of economic activities against the log of the standard deviation of its revenue throughout the sample. The positive correlation suggests households with more variable income are more likely to have many economic activities. The evidence of this section suggests the risk and under-specialization theory may have merit. Fluctuating revenue and imperfect insurance plague the lives of the poor, and the households that suffer the most appear to specialize the least. By themselves, the simple correlations and patterns of this section mean little; any number of stories and models could reproduce them. But combined with the quasi-experimental evidence of Section **??**, they argue forcefully that risk holds poor households back from the benefits of economic specialization.

6. Causal Evidence of Risk and Under-Specialization

The ideal test for the model of risk and under-specialization is to randomize people's lives: make the products of one group's labor perfectly stable and another's very risky. The theory predicts people cursed with riskier lives will become jacks of all trades while the lucky control group can specialize like Smith's pin-maker.

Lacking the capacity and amorality to randomize risk into people's lives, I instead observe their responses to imminent changes in the riskiness of their income. Changes in the monthly volatility of the international price of rice will directly alter the variance of revenue for a household expecting a rice harvest. By observing their responses I can isolate the effect of imminent income risk on specialization.

6.1. Estimating Risk Response

To estimate the causal effect of risk on number of economic activities, I employ a triple-differences estimator. A standard double-difference (or difference-in-differences) estimator measures the response of one (control) group to the response of another (treatment) group. Researchers typically leverage the response to a program or event, but the logic applies equally to this paper's lever: the response to conditional volatility in the international price of rice. A triple-difference estimator extends the approach by using three groups. Each incremental comparison of responses between groups removes an additional source of bias.

The estimator first compares the response of rice farmers to the responses of everyone else. Since rice is the consumption staple in Thailand, every household cares about its price and might respond to price volatility. The first comparison removes the portion of a households response due to consumption.

The estimator then compares the response of rice farmers expecting a harvest to the response of rice farmers in general. Although rice farming is the default occupation in rural Thai villages, households can leave the default for wage work or start a non-farm business. Those that remain rice farmers may differ in risk preferences or other unobservables that cause them to respond differently to rice price volatility even leaving aside income risk. Comparing the response of rice farmers in general to that of rice farmers expecting a harvest eliminates the portion of a households response due to occupational choice. Figure **??** summarizes the concept graphically; using the non-rice farmers and rice farmers not expecting a harvest as controls is like holding the mean and volatility of the rice price constant outside of the pre-harvest window of each farmer. The estimator then compares a farmer to himself when he faces volatile versus stable rice prices before his harvest.

Estimating the triple-difference in a regression framework is easy:

$$
\begin{aligned}
[Activities]_{it} = {} & [FE]_i + [\boldsymbol{Mean}] + [\boldsymbol{Volatility}] \\
& + [Expecting\ Harvest]_{it} + [Had\ Harvest]_{it} \\
& + [Rice\ Farmer]_i \times [Mean]_t + [Rice\ Farmer]_i \times [Volatility]_t \\
& + [Expecting\ Harvest]_{it} \times [Mean]_t + [Expecting\ Harvest]_{it} \times [Volatility]_t \\
& + [Had\ Harvest]_{it} \times [Mean]_t + [Had\ Harvest]_{it} \times [Volatility]_t + \varepsilon_{it}
\end{aligned}
$$

Ignore the indicator for $Had\ Harvest$ momentarily. Then the terms in the second line of the regression do the first comparison; controlling for the response of all households to rice price volatility (absorbed by the month dummies), the coefficients on the interaction of expected mean and expected volatility measure the differential response of rice farmers. Likewise, the third line estimates the parameters of interest: the response to volatility and expected returns of rice farmers expecting a harvest after controlling for the responses of households in general and rice farmers in general. The $Had\ Harvest$ indicator accounts for the fact that some farmers have fairly short gaps between their harvests; these farmers might simultaneously be expecting a new harvest and be taking the output of a recent harvest to market. The main effect and interactions prevent the post-harvest response (which may include arbitrage, etc.) from contaminating the pre-harvest response. The main effect of being a rice farmer is absorbed into the household fixed-effects, so the regression is indeed a triple-difference estimator. In most of my specifications I replace the bolded main effects of the mean and volatility with time dummies $\sum_t [\boldsymbol{Month\ Dummy}]_t$ for a more conservative specification.

In the theoretical framework of Section **??** the coefficient on $[Expecting\ Harvest]_{it} \times [Volatility]_t$ implements the Risk Test. The model predicts it should be positive and significant. Likewise, the coefficient on $[Expecting\ Harvest]_{it} \times [Mean]_t$ implements the Returns Test and should be negative and significant.

To implement the Insurance Test—that households with insurance respond less strongly to volatility—I leverage between-household transfers. In both the developing and developed worlds households rely on cash from friends and

family in difficult times, making it a form of insurance.[3] I define net incoming transfers as the cash value of gifts the household receives minus gifts it gives. Figure [fig] shows that households harvesting rice can expect more incoming transfers when the international rice price is low than high, confirming that households use transfers to shield themselves from income shocks.

To act as insurance, transfers must come when a household's own revenue is low.[4] I define a household as insured if the correlation between the household's revenue and incoming transfers is negative. I run the triple-difference regression separately for households with and without insurance; Test **??** states the coefficient on $[Expecting\ Harvest]_{it} \times [Volatility]_t$ should be smaller for households with insurance.

6.2. A Monthly Measure of Conditional Rice Price Volatility

The risk and under-specialization theory predicts that an all-else-equal increase in the variance of a households income should cause it to enter more side activities—that is, decrease its degree of specialization. Since I use rice farmers expecting a harvest as my treatment group, I need a monthly measure of the volatility of their income.

The measure I use is the conditional volatility of the international rice price. Although each farmer's exact return to a kilo of rice need not be the international price—transport costs for example can cause the gate price to differ from the international price—the actual return will move with the international price. Local prices will moreover be contaminated by local shocks that affect other aspects of the village economy; the international market for rice is large enough that no shock at the level of a single village in my sample can create endogeneity.[5]

No standard dataset of volatility for rice prices exists, so I must construct a measure from the time series of prices. I model the monthly price with the Autoregressive Conditional Heteroskedasticity (ARCH) model of **?** with one modi-

[3] **?**, for example, find that exogenous shocks to household income in the rural Phillipines cause changes of opposite sign in international remittance flows.

[4] Recall from Section **??** that my definition of revenue excludes remittances and aggregates returns from economic activities that require effort.

[5] Talk about national shocks like floods in 2011

fication: I assume the level of the price follows a random walk.

More formally, suppose P_t is the price in month t and z_t is an orthogonal standard normal innovation. The model is

$$\begin{aligned} P_t &= P_{t-1} + \varepsilon_t \\ \varepsilon_t &= z_t \sqrt{h_t}, \qquad z_t \sim N(0,1) \\ h_t &= \tau_0 + \tau_1 \varepsilon_{t-1}^2. \end{aligned}$$

I estimate this model using conditional maximum likelihood.[6] Assuming a farmer must choose the number of economic activities for month t at the beginning (before the price for that month is announced), he will base his decision on $\mathbb{E}[\sigma_t^2 \mid \varepsilon_{t-1}^2] = h_t$. The predicted values of h of the estimated model are consistent estimates of the true conditional volatilities. In the empirical specifications of Section **??** I use the square root of this measure to make its units comparable to the mean.[7]

Figure **??** plots the actual price of rice, the predicted mean, and the predicted standard deviation. Simple though it is, the random walk assumption makes very accurate predictions about the mean. The red lines demarcate the start and end of the time period covered in the monthly panel data. The sample spans a time when prices are relatively stable, ending well before the massive food price spike of 2008.

6.3. The Costs of Under-Specialization

Section **??**'s empirical approach will determine whether the poor enter additional economic activities to mitigate risk. It will not determine whether doing so costs them foregone income. But Smith's pin-maker parable has been canonized precisely because it illustrates why specialization creates economic benefit. To verify the full story I must demonstrate that households could increase

[6]The true distribution of z_t need not be normal; the (quasi) maximum likelihood estimator based on a normal distribution is still consistent.

[7]The reader may question the assumption that rural farmers base their decisions on the predictions of an ARCH process. A less complicated measure like the absolute change in prices between months controlling for the price level has similar effects on economic activities. These results are in Table **??**.

their average revenue if only they did specialize. In other words, I must show that risk not only impacts specialization but causes under-specialization.

Identifying the causal impact of more economic activities on revenue requires instruments. Luckily, Section **??** has already produced a set: the responses to the expected mean and volatility of the rice price among farmers expecting a harvest. Since these farmers do not yet have their harvests in hand, movements of the price cannot affect their current revenue directly. Since the time span I examine is very short—the observations are monthly—changes in investment behavior due to heightened risk will not have any immediate effect. This leaves the change in the number of economic activities as the only channel of effect.

Though plausibly exogenous, the responses of rice farmers expecting a harvest by construction cannot test any predictions about the effect of more activities on total revenue. The farmers' primary source of revenue has not yet been harvested, so all I observe is their revenue from secondary activities. But I can nevertheless verify if under-specialization is costly because the Cost Test is about the effect of more activities on revenue from side activities. If the instrumental variables estimate of the effect of activities on revenue is negative, it means under-specialization must be costly.

The first-stage is exactly the regression presented in Section **??**. The second stage is

$$
\begin{aligned}
[Revenue]_{it} = [FE]_i + [\widehat{Activities}]_{it} + \sum_t [Month\ Dummy]_t \\
+ [Expecting\ Harvest]_{it} + [Had\ Harvest]_{it} \\
+ [Rice\ Farmer]_i \times [Volatility]_t + [Rice\ Farmer]_i \times [Mean]_t \\
+ [Had\ Harvest]_{it} \times [Volatility]_t + [Had\ Harvest]_{it} \times [Mean]_t + \varepsilon_{it}.
\end{aligned}
$$

The model passes the Cost Test if the coefficient on $[\widehat{Activities}]_{it}$ is negative.

To implement the OLS Bias Test, I must estimate the simple OLS regression[8]

[8] I use OLS because it maps most closely to the theory, but fixed-effects estimates are similar.

$$[Revenue]_{it} = \sum_t [Month\ Dummy]_t + [Activities]_{it} + \varepsilon_{it}$$

and compare the OLS coefficient on [Activities] to its analog from the IV regression. The test predicts the OLS coefficient should be more positive than the IV coefficient.

6.4. Is the Empirical Approach Valid?

Does a Thai Rice Farmer Really Care About the International Rice Price?
A rice farmer's income depends on the price she receives for her harvest, which is not the same as the international price. Is it plausible that farmers would actually respond to movements in the international price? Anecdotal evidence from the villages of the data sample suggests farmers actually follow the international price carefully in newspapers and on television. Aggregate information is more easily accessible than the local price of un-milled rice, which a farmer only learns when her acquaintances report back from visiting the market. More importantly, I find that a rise in the international price does predict a rise in sample-wide average of prices received at harvest (see Column 1 of Table **??** in Appendix **??**).

Do Farmers Manipulate their Harvest or Selling Date?
If farmers delay their harvest to coincide with high prices, the variation I exploit may not be exogenous. This is unlikely to be a concern given the biology of rice in Thailand. Rice is a semi-aquatic plant that while growing (largely underwater) is relatively safe from pests and weather damage. But once the plant is mature it must be harvested and dried quickly to prevent spoilage and damage from pests. I also find strong evidence against strategic selling. The correlation between rice sold and rice harvested (conditional on a harvest) is over .85, implying farmers immediately sell most of their harvest (see Column 2 of Table **??** in Appendix **??**).

Is Rice Near Harvest-Time Really an Activity?
The model assumes a household must allocate labor between its primary activity and side activities; does a household really work on its rice paddy in the lead-up to a harvest? The version of the data released as of this writing does not

contain crop-by-crop labor allocation data. However, I can observe the amount of labor a household puts towards its fields as a whole in the months before and after harvest. Rice farming households work no less in their fields until the month after harvest (see Figure **??** in Appendix **??**).

Other Concerns

In Appendix **??** I explore whether several other potential issues may be driving the results. I find no evidence that non-independent returns, general equilibrium effects, or changes in household composition or labor supply bias my results. The reader may also worry that Thai rice farmers cannot or do not compute conditional expectations and volatilities. I demonstrate that the results stand when I use simple heuristics in place of the ARCH predicted values: current price and absolute price change from the previous month.

7. Results: Risk and Under-Specialization

Tables **??** and **??** report the results of the regressions described in Sections **??** and **??**. All regressions except OLS in Column (1) of Table **??** use a two-stage bootstrap described in Appendix **??** to produce 95% confidence intervals and significance levels. The bootstrap accounts for the use of generated regressors (the predicted mean and volatility) estimated in Section **??** and allows for arbitrary within-household correlation in the errors. The coefficients relevant to tests from Section **??** are color-coded by test; see the table notes for the key.

Table **??** reports the coefficients relevant to Tests 1, 2, and 3. Exactly as the theory predicts in Test 1, the coefficient on $[Expecting\ Harvest]_{it} \times [Volatility]_t$ is positive; households increase the number of economic activities in response to a rise in risk. The coefficient implies a 10% rise in the standard deviation causes households to enter roughly .04 additional economic activities. Given the average number of economic activities is 4.6, the local elasticity is roughly .085. The theory also predicts in Test 3 the coefficient on $[Expecting\ Harvest]_{it} \times [Mean]_t$ should be negative, which it is. The magnitude implies a 10% rise in the expected price causes a decrease of .2 activities, implying a local elasticity of -.44. In other words, a decrease in the standard deviation of 10% increases specialization as much as a 2% increase in the returns.

Specifications (3) and (4) split the sample into the insured and uninsured as described in Section **??** to implement the Insurance Test with separate regressions. As predicted, the response of insured households is smaller and less statistically significant. Specification (5) implements the test in a single regression (the second column records the coefficients on each variable interacted with an indicator for insurance). As predicted, the risk response coefficient on the interaction is negative, but I am unable to reject the null of no difference. Given that my sample contains so few households, the likely cause is a lack of power. Specification (5) shows that the response of the uninsured is double that of the population as a whole: a 10% decrease in the standard deviation has the same effect as a 4% increase in returns.

Columns (1) and (2) of Table **??** report the coefficients relevant to Tests 4 and 5. According to the theory a negative coefficient on activities in the IV regression is possible only if under-specialization is costly; Column (2) shows the theory passes Test 4. It is difficult to interpret the magnitude of the coefficient. In Section **??** I show it is not the fixed cost of new activities because it also captures the household's allocation of labor between activities. But even if it were, the variation in fixed costs between households and the centrality of each household's cost in its specialization decision together mean the estimated coefficient is not an average cost but a local average cost.

Column (2) reports the OLS regression of total revenue on number of economic activities. For the reasons I explain in Section **??**, OLS is upward biased. The upward bias produces a positive coefficient even though Smith's theory of specialization predicts less specialization should decrease revenue. The IV estimates in Column (3) reverse the sign: more activities lowers revenue, exactly as the theory predicts. The size of the coefficient is large; if this were the average cost of an additional economic activity, the coefficient implies every new activity costs a household 63% of the average monthly revenue. But as the model in Section **??** demonstrates, it is not the average cost but rather a continuous analog to the "local average treatment effect" of the heterogeneous treatment effect literature.

8. The Alternative Theory: Lumpy Investments

The lumpy investment theory of under-specialization asserts that "the poor cannot raise the capital they would need to run a business that would occupy them fully" (**?**). Consider the following example: an individual can work as a tailor, a carpenter, or as a baker. Anyone can learn the basics skills of any trade, but expanding a business—sewing more than a few shirts or baking more than a few loaves every day—requires buying a large and lumpy piece of capital. Would-be tailors need sewing machines and would-be bakers need ovens. But binding credit constraints prevent anyone from buying the necessary capital and expanding to the necessary scale. Since its tailoring business must remain too small to cover its consumption, the household must also start a bakery. Then although each household would have preferred to specialize and trade, the lack of capital forces everyone to do everything to the detriment of all. The theory that poor people can only profitably work at any given task for so long and take up other crafts to occupy their free time is actually a consequence of the lumpy investment theory.[9]

To test the lumpy investment theory I exploit a government program that produced quasi-experimental variation in credit availability. If the theory is true, households would specialize if only they had the credit to make the necessary lumpy investment. Then we would expect a relaxation of credit constraints would cause a decrease in each household's number of economic activities. The program I exploit is the Million Baht Program. The program was rapidly implemented in all the villages of the Townsend Thai annual survey (among others) in the latter part of 2001 and provided one million Thai baht to each village's community lending facility. Since the size of the transfer was the same for each village regardless of size, a smaller village received a larger per-household transfer. **??**, who are the first to exploit the program, argue that villages in Thailand

[9]Assuming labor and capital (or whatever factor requires the lumpy investment) are complements. To see this most simply, take an extreme example: perfect complementarity. Suppose an activity m produces revenue with production function $y^m = A^m \min[L, K]$, with $m = T, B$ for tailoring or baking. Suppose WLOG $A^T > A^B$ for some household. If the household's labor endowment is $\bar{L}$, it will specialize in tailoring with $K^* = \bar{L}$. But suppose increasing capital beyond $\tilde{K} < K^*$ requires a lumpy investment the household cannot afford. If the household specialized, it would be left with $\bar{L} - \tilde{K}$ units of unused labor. In other words, it would be idle. The alternative is to spend its remaining time baking, so its total revenue is $A^T\tilde{K} + A^B(\bar{L} - \tilde{K}) < A^T\bar{L}$.

were delineated decades prior to the program by bureaucratic fiat for administrative convenience. Since the sizes of villages are effectively random, the per-household increase in credit availability is also random. The authors find the program had little effect on average investment largely because decreased investment by some households offset the increased investment by others. More recently, **?** used the program to test measures of within-sector factor allocative efficiency in the rice sector. I find the program had large but transient effects on overall efficiency acting largely through improved financial market efficiency.[10]

Since I do not know the exact month when the program reached any given village, I use the annual data and treat 2001 as the year of program implementation. The program effect is captured by the interaction of the year of (and year after) implementation interacted with some measure of village size. In one specification I use an indicator for whether the household is in the bottom quartile in number of households; in the other I use the actual per-household injection (1 million/number of households). The lumpy investment theory predicts the signs of the coefficients should be negative and significant.

Table **??** reports that the coefficients have the wrong sign. The positively signed effect of the program in the year immediately after the program is significant in Specification (1), but the effect does not appear in Specification (3), which restricts estimation to a balanced panel. The other specifications find no statistically significant effects of credit availability on the number of economic activities. Lumpy investments do not appear to explain economic under-specialization in Thailand.

9. Conclusion

If the benefits of specialization are so remarkable—and the Industrial Revolution proved they are—why would the poor fail to exploit them? The answer is fear: the fear of specializing in an activity with risky returns.

I find that Thai rice farmers expecting a harvest increase their number of

[10]Is the program still a valid source of exogenous variation if it affected within-sector efficiency? Improved within-sector efficiency would likely increase the returns to any particular activity and so should increase the incentives to specialize. If anything, my estimates should be biased towards finding an effect, making the lack of evidence that much more pronounced.

economic activities when confronted with more volatile prices, which is exactly what a simple model of risk and specialization predicts. I find suggestive evidence that risk affects uninsured households the most and confirm the model's prediction that higher expected returns lower the number of activities. The size of an uninsured household's response to a 10% fall in volatility is equivalent to a 4% rise in expected returns. I use the induced change as exogenous variation in the number of activities to verify under-specialization does cost households foregone revenue. Finally, I test an alternative theory of under-specialization—that the poor must run many small businesses because they lack the credit to make lumpy investments—and find no supporting evidence.

Muhammad Yunus, founder of the Grameen Bank, claims the poor are "natural entrepreneurs" because of their many economic activities. My results suggest instead that the poor are driven into entrepreneurship by the fear of a risky income stream. Like any choice made under duress, the choice to diversify economic activities is inefficient. The poor become jacks of all trades as a second-best option to what they really want: a single profession with the safety of a paycheck.

References

Bandyopadhyay, S. and Skoufias, E. (2012). Rainfall variability, occupational choice, and welfare in rural bangladesh. *World Bank Policy Research Working Paper*, (6134).

Banerjee, A. and Duflo, E. (2007). The Economic Lives of the Poor. *The Journal of Economic Perspectives: A Journal of the American Economic Association*, 21(1):141.

Engle, R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society*, pages 987–1007.

Jacoby, H. and Skoufias, E. (1997). Risk, financial markets, and human capital in a developing country. *The Review of Economic Studies*, 64(3):311–335.

Kaboski, J. and Townsend, R. (2009). The impacts of credit on village economies. *MIT Department of Economics Working Paper No. 09-13*.

Kaboski, J. and Townsend, R. (2011). A structural evaluation of a large-scale quasi-experimental microfinance initiative. *Econometrica*, 79(5):1357–1406.

Menon, N. (2009). Rainfall uncertainty and occupational choice in agricultural households of rural nepal. *The Journal of Development Studies*, 45(6):864–888.

Murphy, K. and Topel, R. (2002). Estimation and inference in two-step econometric models. *Journal of Business & Economic Statistics*, 20(1):88–97.

Rosenzweig, M. and Binswanger, H. (1993). Wealth, weather risk and the composition and profitability of agricultural investments. *The Economic Journal*, 103(416):56–78.

Shenoy, A. (2012). Market Failures and Misallocation: Decomposing Factor Misallocation by Source.

Townsend, R., Paulson, A., Sakuntasathien, S., Lee, T. J., and Binford, M. (1997). Questionnaire Design and Data Collection for NICHD Grant *Risk, Insurance and the Family* and NSF grants.

Townsend, R. M. (2012). Townsend Thai Project Monthly Survey (1-60) Initial Release. Murray Research Archive [Distributor] V1 [Version].

A Detailed Data Appendix

A1. Time Series Variables

- **Consumer Prices:** From Bank of Thailand monthly index, acquired from Global Financial Data database. Data were used with permission of Global Financial Data.
- **International Rice Price:** Acquired from IMF monthly commodity price data. Deflated using monthly consumer price index.

A2. Panel Variables

- **Rice Harvest:** From module 7 (Crop Harvest) section of the monthly survey. Keep only un-milled rice (both sticky and non-sticky). Define rice harvest soon as a reported positive harvest of unmilled rice in the subsequent three months. Define rice harvest past as having had positive harvest of unmilled rice in the current or previous three months. Define rice farmer (or rice harvest ever) as having had a positive rice harvest at any point in the survey span.

- **Crop-Plots:** From module 5 (Crop Activities) section of the monthly survey. Make the monthly aggregate of "value transacted" for each households sale of each crop. This is the revenue from crops. For number of crop plots, I use the "projected harvest" table, which asks farmers to predict revenue for each productive crop. Every entry corresponds to a different perceived revenue stream for the farmer, so I take number of crop-plots as simply the count of these for each household in each month.

- **Aquaculture:** From module 10 (Fish-Shrimp) of the monthly survey. For each household, make monthly aggregates of the value of fish and shrimp output; this is the revenue from aquaculture. I compute whether a household does aquaculture as whether it reports raising fish/shrimp or having shrimp ponds in a given month.

- **Large Businesses:** From module 12 (Household Business) of the monthly survey. For each household, make monthly aggregates of the cash and in-kind revenue plus the value of products/services consumed by the household; this is the revenue from large businesses. Compute the number of businesses for each household as the number of entries in the household report of revenues.

- **Small/Miscellaneous Businesses:** From module 24 (Income) of the monthly survey. For each household, make monthly aggregates of the cash and in-kind revenue for each "other" income source; this is the revenue from miscellaneous businesses. Compute the number of miscellaneous activities for each household as the number of entries in the household report of revenues.

- **Number of Jobs:** From module 11 (Activities-Occupation). For each person and each job number in any month, mark if it was worked the previous two and the following two months (note that jobs are not assigned job numbers in their first months, so technically I only check the previous one month as it must have been worked the month before to have an ID). If so, it is a steady job. I count each households total number of jobs and steady jobs each month, then compute the number of unsteady jobs as the difference. For each job and each month, sum the cash and in-kind payments and aggregate by household-month. This is the monthly job revenue.

- **Number of Activities:** I define number of activities as simply the sum of the number of crop plots, the number of livestock activities, the indicator for practice of aquaculture, the number of large businesses, the number of jobs, and the number of miscellaneous activities.

- **Total Revenue, Consumption, and Transfers:** Total revenue is the sum of revenue from crop activities, livestock activities, aquaculture, large businesses, jobs, and miscellaneous activities. Total consumption is the sum of all domestic expenditures by both cash and credit plus consumption of home-produced goods. Expenditures reported at a weekly rather than monthly frequency (in module 23W, Weekly Expenditures Update) are aggregated by month for each household and added to those reported at a monthly frequency (in module 23M, Monthly Expenditures Update). Transfers are defined as the household's net incoming transfers. More precisely, I aggregate by household-month the transfers from people inside and outside the village and subtract similarly aggregated transfers to people inside and outside the village (all found in module 13 on Remittances). I use only transfers not earmarked for a specific event because these unplanned transfers are more like insurance.

B Verifying the Validity of Assumptions

C Inference: The Two-Stage Bootstrap

The predicted mean and volatility are both generated regressors, so I must adjust my inference to account for their presence. It is easy to see that under my assumptions the full estimators match the conditions for **?**. Directly applying their analytic expressions is inconvenient and also problematic because small sample bias in the time series estimates might produce an abnormal small sample distribution for the estimated parameters. But the asymptotic normality their propositions guarantee also ensures the validity of bootstrapped confidence intervals and hypothesis tests.

I implement the procedure as outlined in Figures **??**-**??**. First, I prepare the time series of rice prices for resampling. I form "blocks" consisting of the contemporaneous price and however many lags I need to estimate the time series model. I then group every observation into one or more "blocks of blocks," contiguous interlocking sets of observations and their associated lags.

Next, I run the bootstrap replications. Each replication follows five intermediate steps. First, I sample with replacement the blocks of blocks of rice prices to construct a bootstrapped time series of equal length to the original time series. I estimate the parameters of the time series model on the bootstrapped data. I then resample with replacement households (together with all their monthly observations) from the panel to construct a bootstrapped panel with as many households as the original panel. Then I use the estimated time series model to predict the conditional mean and variance of the international rice price for each household-month observation. Finally, I estimate the panel specification and record the resulting coefficients. I run 1000 replications for the risk specification and 2000 replications for the IV specifications.

The final step is to compute confidence intervals and p-values. To construct confidence intervals, I use the dataset of estimated parameters from bootstrap replications to find the 2.5th and 97.5th percentiles. These are the boundaries of the 95% confidence interval. To construct p-values, I compute the absolute t-statistic centered around the original parameter estimate for each replication. The fraction of these absolute t-statistics that is greater than the original t-statistic is the p-value.

D Other Tests of Robustness

Table 1
Rice Prices and Sales

	Avg. Transaction Price b/se	Rice Sold b/se
Int. Rice Price	0.333**	
	(0.14)	
Rice Harvested		0.856***
		(0.01)
Constant	1.500	-2043.744***
	(1.53)	(70.44)
N	62	2126

Note: **Column 1 —** The unit of observation is a survey round. The dependent variable is the average price of a kilogram of rice based on actual transactions; the independent variable is the international price of rice in baht per kilogram. Not all survey rounds include any sales of rice—hence the number of observations is smaller than the number of survey rounds. Standard errors are heteroskedasticity-robust. **Column 2 —** The unit of observation in this regression is the household-month conditional on positive rice harvest.

Table 2

Risk Causes Under-Specialization, Especially Among the Uninsured

	(1) Activities (All)	(2) Activities (All)	(3) Activities (No Insurance)	(4) Activities (Insurance)	(5) Activities (Main Effect)	(5) Activities (× Insurance)
Mean	-0.00*					
	[-0.01,0.00]					
Volatility	-0.08***					
	[-0.26,-0.05]					
Rice Farmer						
- × Mean	0.01***	0.00	0.01**	-0.00	0.01*	-0.01
	[0.00,0.01]	[-0.00,0.01]	[0.00,0.02]	[-0.01,0.01]	[-0.00,0.01]	[-0.01,0.00]
- × Volatility	-0.20***	-0.07**	-0.09*	-0.10**	-0.10**	0.01
	[-0.68,-0.12]	[-0.21,-0.03]	[-0.31,-0.01]	[-0.31,-0.04]	[-0.29,-0.03]	[-0.09,0.11]
Expecting Harvest						
- Main	1.82***	1.95***	1.65***	1.99***	1.63***	0.40
	[-1.14,2.59]	[0.98,2.52]	[-0.09,2.46]	[0.95,2.78]	[0.14,2.49]	[-0.67,1.85]
- × Mean	-0.02***	-0.02***	-0.02***	-0.02***	-0.02***	0.00
	[-0.03,-0.02]	[-0.02,-0.01]	[-0.02,-0.01]	[-0.02,-0.01]	[-0.02,-0.01]	[-0.01,0.01]
- × Volatility	0.18***	0.05*	0.09*	0.04	0.10**	-0.06
	[0.10,0.55]	[0.01,0.18]	[0.01,0.32]	[-0.03,0.16]	[0.02,0.29]	[-0.25,0.02]
Recent Harvest						
- Main	-0.76	-0.60	-0.77	-0.39	-0.65	0.18
	[-8.01,0.74]	[-2.90,0.16]	[-3.19,0.27]	[-3.54,0.71]	[-3.07,0.32]	[-1.23,1.30]
- × Mean	-0.03***	-0.01**	-0.01	-0.01***	-0.01	-0.01
	[-0.04,-0.02]	[-0.01,-0.00]	[-0.02,0.00]	[-0.02,-0.01]	[-0.02,0.00]	[-0.02,0.01]
- × Volatility	0.41***	0.15***	0.14**	0.19***	0.14**	0.05
	[0.25,1.32]	[0.08,0.42]	[0.05,0.46]	[0.11,0.57]	[0.05,0.45]	[-0.07,0.23]
Household Fixed-Effects	Yes	Yes	Yes	Yes	Yes	
Month Fixed-Effects	No	Yes	Yes	Yes	Yes	
Households	743	743	270	473	743	
Observations	48329	48329	16933	31396	48329	

Note: These results estimate the triple-difference specification, where all regressions contain household fixed-effects and month dummies. The main effects of the conditional mean and volatility are colinear with the month dummies, and the main effect of being a rice farmer is colinear with the household fixed effects, so neither is estimated. The bracketed ranges are 95% confidence intervals. I compute the confidence intervals and the significance levels represented by the asterisks using a two-stage bootstrap that corrects for generated regressors and clusters at the household level (see Appendix **??**). The estimated coefficients are color-coded to reflect the tests they implement from the theory (see Section **??**): Test 1 (Risk),Test 2 (Insurance), Test 3 (Returns).

Table 3
Under-Specialization is Costly

	(1) Revenue	(2) Revenue
Activities	1851.257***	-13883.30**
	[949.50,2753.01]	[-30990.86,-3612.91]
Rice Farmer		
- × Mean		-128.97
		[-329.87,20.00]
- × Volatility		-272.90
		[-2165.46,948.74]
Expecting Harvest		
- Main		4993.52
		[-3632.31,17951.96]
- × Mean		(Excluded Instrument)
- × Volatility		(Excluded Instrument)
Recent Harvest		
- Main		-34753.04*
		[-78172.39,-4899.71]
- × Mean		300.00
		[-70.79,767.66]
- × Volatility		234.48
		[-2107.58,4204.75]
Household Fixed-Effects	No	Yes
Month Fixed-Effects	Yes	Yes
Households	743	743
Observations	48329	48329
F-stat Exc. Inst.		13.604
Hansen's J Stat		0.125

Note: The first column implements a naive OLS estimate of monthly revenue on number of economic activities. The second column estimates the triple-difference IV specification with (unreported) household fixed-effects and month dummies. The main effects of the conditional mean and volatility are colinear with the month dummies, and the main effect of being a rice farmer is colinear with the household fixed effects, so neither is estimated. Column 1 of Table **??** is the first-stage, where the indicated coefficients are the excluded instruments. The bracketed ranges are 95% confidence intervals. I compute the confidence intervals and the significance levels represented by the asterisks using a two-stage bootstrap that corrects for generated regressors and clusters at the household level (see Appendix **??**). The value of the F-statistic on the excluded instruments from the first stage meets common standards for strength. The value of the J-statistic for overidentification is much too small to reject the null of exogenous instruments. The estimated coefficients are color-coded to reflect the tests they implement from the theory (see Section **??**): Test 4 (Cost),Test 5 (OLS Bias).

Table 4
Testing the Theory of Lumpy Investments

	(1) Activities b/se	(2) Activities b/se	(3) Activities b/se	(4) Activities b/se
Small Village	-0.010		0.102	
	(0.11)		(0.11)	
2001 X Small	0.132		0.175	
	(0.16)		(0.16)	
2002 X Small	0.213		0.144	
	(0.14)		(0.15)	
Credit/HH		3.010		9.977
		(13.47)		(13.28)
2001 X Credit/HH		11.540		11.569
		(8.63)		(8.20)
2002 X Credit/HH		21.857*		16.619
		(11.55)		(11.86)
Household Fixed-Effects	Yes	Yes	Yes	Yes
Month Fixed-Effects	Yes	Yes	Yes	Yes
Villages	80	80	64	64
Households	1502	1502	706	706
Observations	15340	15340	9884	9884

All standard errors clustered by village

Note: The regressions use the larger impact of the Million Baht Program on small villages. The coefficient on the interaction of size with the implmentation year (2001) captures the effect of relaxed credit on number of activities. I compute number of activities as closely as possible to the methodology in the risk regressions and mark a village as "small" if it is in the bottom quartile of number of households. The alternative specification uses the average per-household credit injection (one million divided by number of households). The first two columns use the largest possible sample of households while the last two use a balanced panel. The lumpy investment theory predicts the program's impact should be negative and significant, which it is not.

Table 5
Robustness: Main Results Excluding Pre-Harvest Rice Sales

	(1) Activities	(2) Revenue
	(1) Activities	(2) Revenue
Activities		-14195.18**
		[-30152.21,-4348.14]
Rice Farmer		
- × Mean	0.00	-86.06
	[-0.00,0.01]	[-270.98,71.63]
- × Volatility	-0.09**	-368.73
	[-0.29,-0.04]	[-2437.34,878.75]
Expecting Harvest		
- Main	1.37***	-238.27
	[0.39,1.92]	[-6361.03,8661.24]
- × Mean	-0.01***	(Excluded Instrument)
	[-0.02,-0.01]	
- × Volatility	0.05*	(Excluded Instrument)
	[-0.00,0.18]	
Recent Harvest		
- Main	-0.63	-27329.39*
	[-3.42,0.25]	[-68381.34,706.23]
- × Mean	-0.01***	147.78
	[-0.02,-0.00]	[-161.76,521.56]
- × Volatility	0.17***	1095.82
	[0.10,0.51]	[-1053.76,5783.96]
Household Fixed-Effects	No	Yes
Month Fixed-Effects	Yes	Yes
Households	743	743
Observations	47395	47395
F-stat Exc. Inst.		10.054
Hansen's J Stat		0.015

Note:

Table 6
Robustness: Non-Generated Volatility Measure

	(1) Activities	(2) Revenue
Activities		-9305.776*
		[-19145.72,534.17]
Rice Farmer		
- × Mean	0.009***	-21.253
	[0.00,0.01]	[-181.62,139.12]
- × Volatility	-0.055***	-397.123
	[-0.07,-0.04]	[-898.06,103.82]
Expecting Harvest		
- Main	2.578***	1537.605
	[1.96,3.20]	[-7017.51,10092.71]
- × Mean	-0.019***	
	[-0.02,-0.01]	
- × Volatility	0.024***	
	[0.01,0.04]	
Recent Harvest		
- Main	1.300***	148.018
	[0.59,2.01]	[-28041.52,28337.55]
- × Mean	-0.018***	-44.617
	[-0.02,-0.01]	[-328.34,239.10]
- × Volatility	0.053***	965.185**
	[0.04,0.07]	[211.07,1719.30]
Constant	4.270***	
	[3.84,4.70]	
Households	743	743
Observations	48329	48329
F-stat Exc. Inst.		20.660
Hansen's J Stat		0.369

Note: These regressions use heuristic measures of mean and variance of prices in place of the predicted values from the ARCH(1) model. I measure the mean as simply the current price and the variance as the absolute change in the price from the previous month. The p-values and confidence intervals are computed using asymptotic standard errors that cluster by households.

Table 7

Robustness: Risk Results Controlling for Labor and Household Size

	(1) Activities	(2) Activities
Rice Farmer		
- × Mean	0.00	0.00
	[-0.00,0.01]	[-0.00,0.01]
- × Volatility	-0.10***	-0.10**
	[-0.26,-0.05]	[-0.32,-0.05]
Expecting Harvest		
- Main	1.78***	1.86***
	[1.00,2.37]	[0.82,2.46]
- × Mean	-0.01***	-0.01***
	[-0.02,-0.01]	[-0.02,-0.01]
- × Volatility	0.05*	0.05*
	[-0.00,0.15]	[0.01,0.18]
Recent Harvest		
- Main	-0.54	-0.61
	[-2.47,0.28]	[-3.53,0.25]
- × Mean	-0.01***	-0.01***
	[-0.02,-0.00]	[-0.02,-0.00]
- × Volatility	0.16***	0.17***
	[0.09,0.40]	[0.10,0.52]
Total Household Labor	0.01***	
	[0.00,0.01]	
Household Size		0.13***
		[0.07,0.19]
Household Fixed-Effects	Yes	Yes
Month Fixed-Effects	Yes	Yes
Households	743	743
Observations	48160	48164

Note:

Table 8

Robustness: Revenue Results Controlling for Labor and Household Size

	(1) Revenue	(2) Revenue	(3) Revenue
Activities	-17060.61***	-11232.10**	-14121.34**
	[-34233.53,-6230.30]	[-26498.70,-2679.13]	[-30065.12,-4755.21]
Rice Farmer			
- × Mean	-91.88	-112.59	-78.17
	[-281.52,66.28]	[-317.53,41.12]	[-244.91,69.41]
- × Volatility	-791.17	-493.25	-956.84
	[-3198.94,403.45]	[-2728.30,562.07]	[-3603.86,141.36]
Expecting Harvest			
- Main	4897.58	7295.65	7192.17
	[-3305.05,17278.34]	[-339.65,20379.13]	[-497.92,18084.96]
Recent Harvest			
- Main	-34442.87**	-28550.62*	-28256.34**
	[-75470.80,-5690.34]	[-72140.06,-2296.48]	[-68749.30,-1940.35]
- × Mean	288.46	167.55	157.39
	[-45.32,722.62]	[-176.42,578.93]	[-151.10,517.92]
- × Volatility	161.63	1018.61	946.90
	[-2312.48,3298.51]	[-1337.55,4839.11]	[-1686.56,5013.26]
Total Household Labor	275.42***		253.15***
	[202.43,385.70]		[183.24,352.32]
Household Size	950.03		692.91
	[-637.48,3533.41]		[-835.94,2614.27]
Household Fixed-Effects	Yes	Yes	Yes
Month Fixed-Effects	Yes	Yes	Yes
Households	743	743	743
Observations	48160	48329	48160
F-stat Exc. Inst.	13.154	13.604	13.154
Hansen's J Stat	0.273	0.002	0.002

Note:

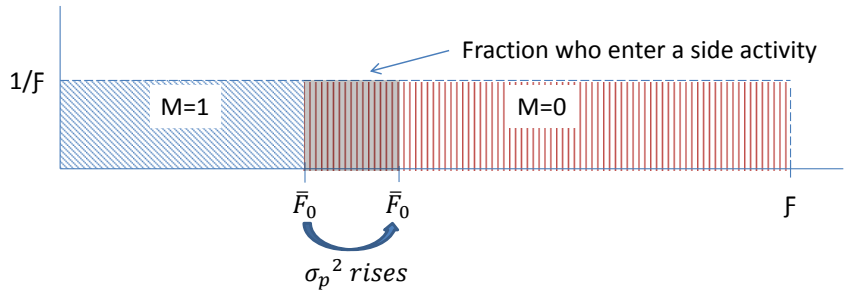
Table 9

Robustness: Regressor of Interest Does not Affect Wages

	(1)
	Activities
Mean	0.01
	[-0.02,0.05]
Volatility	0.66
	[-0.02,2.32]
Rice Farmer	
- Main	-43.31
	[-167.84,12.38]
- × Mean	0.01
	[-0.16,0.23]
- × Volatility	0.52
	[-3.07,5.92]
Expecting Harvest	
- Main	24.62
	[-23.44,116.85]
- × Mean	-0.06
	[-0.23,0.08]
- × Volatility	-2.90
	[-14.31,4.24]
Recent Harvest	
- Main	20.65
	[-0.94,72.20]
- × Mean	-0.07
	[-0.25,0.13]
- × Volatility	-1.80
	[-9.82,1.66]
Village Fixed-Effects	Yes
Month Fixed-Effects	No
Villages	16
Observations	1152

Note:

Figure 1
Intuition of the Simplified Case



Note: M is the number of side activities; $\bar{F}_0$ the threshold fixed cost for moving from zero to one side activity; σ_p^2 is the variance of the primary economic activity. A rise in the variance of the return to the primary economic activity causes the threshold fixed cost to rise. Since households are now willing to pay more for insurance, the highlighted mass of individuals switches from specialization to having a side activity.

Figure 2
Why is OLS Upward-Biased?

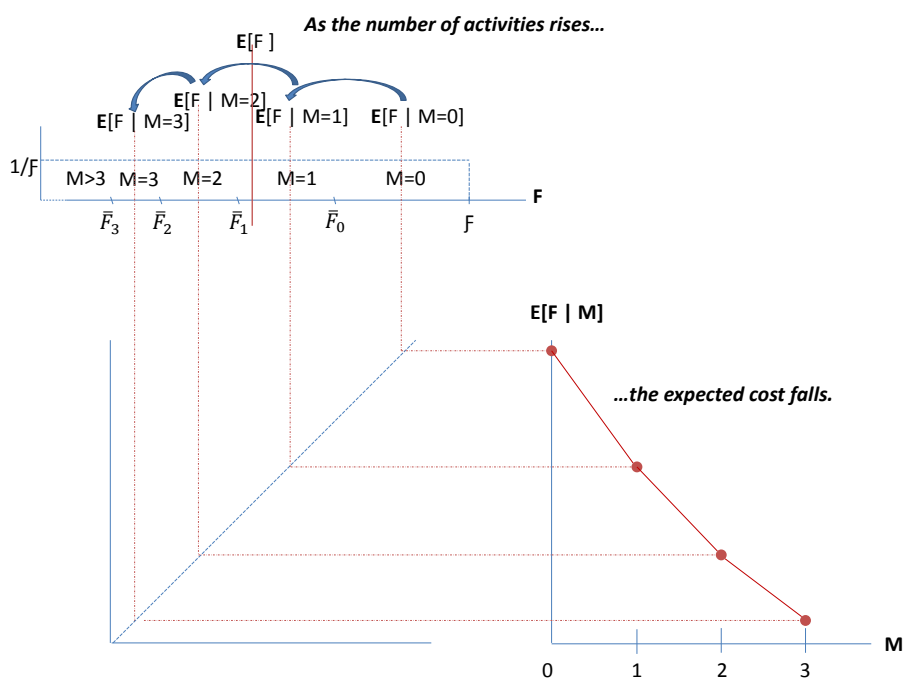
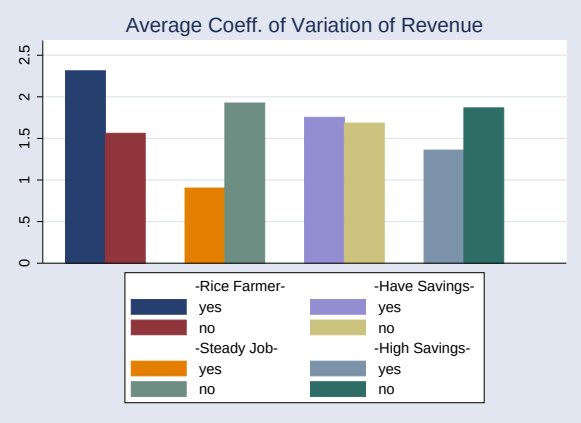
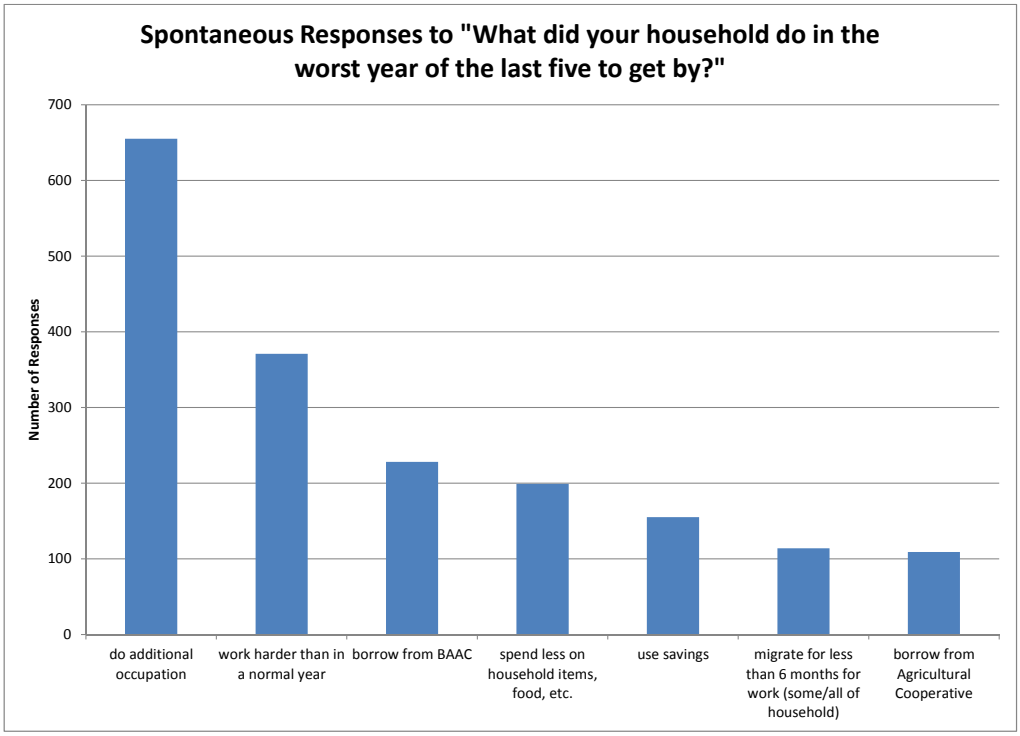


Figure 3
Variation in Household Revenue Across Subgroups



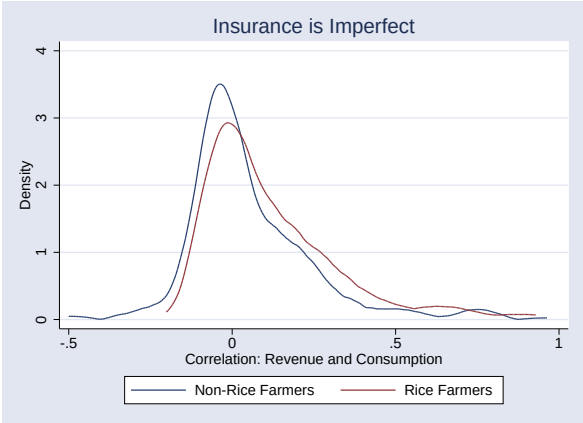
Note:

Figure 4
Household Response to Negative Income Shocks



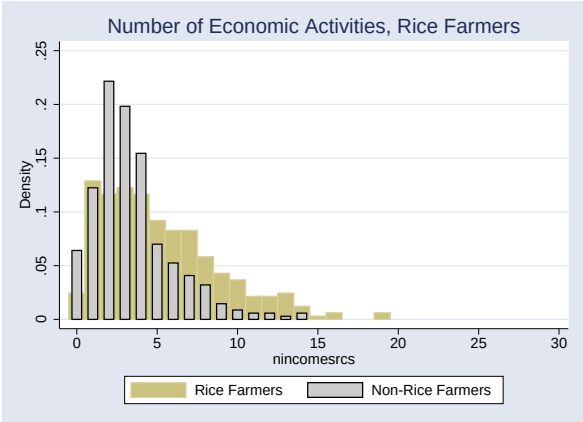
Note:

Figure 5
Correlation Between Monthly Revenue and Consumption



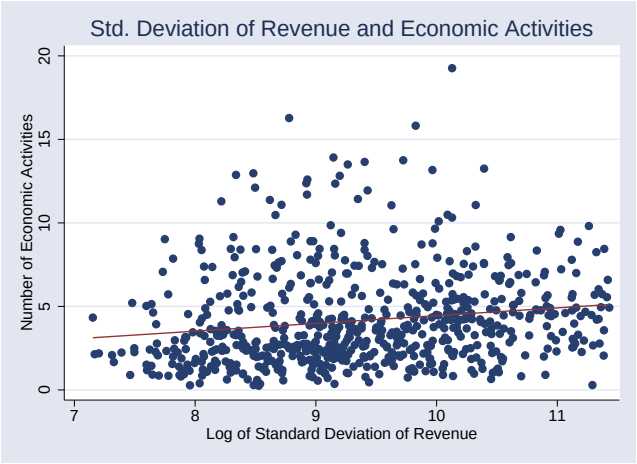
Note:

Figure 6
Number of Economic Activities, Rice Farmers and Non-Rice Farmers



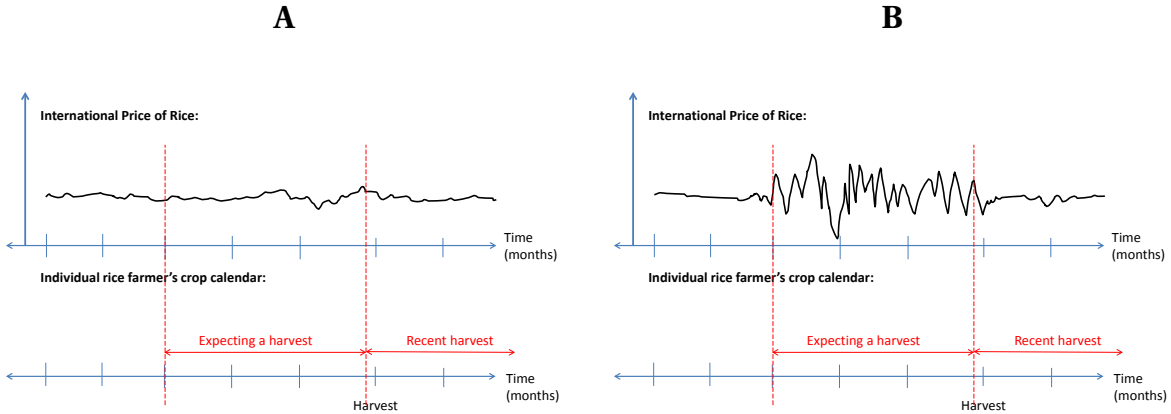
Note:

Figure 7
Riskiness of Revenue and Number of Economic Activities



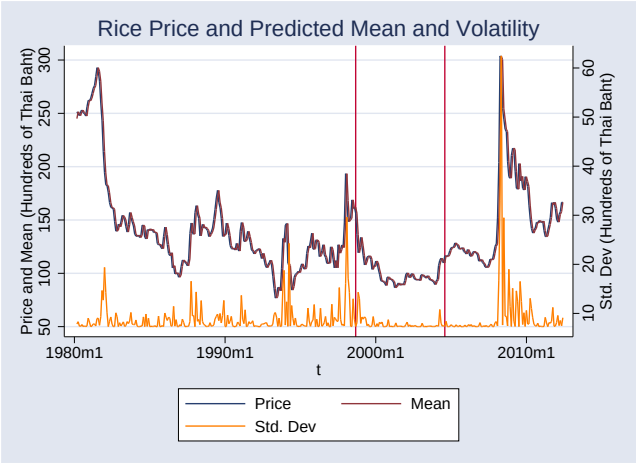
Note:

Figure 8
Response to Conditional Volatility



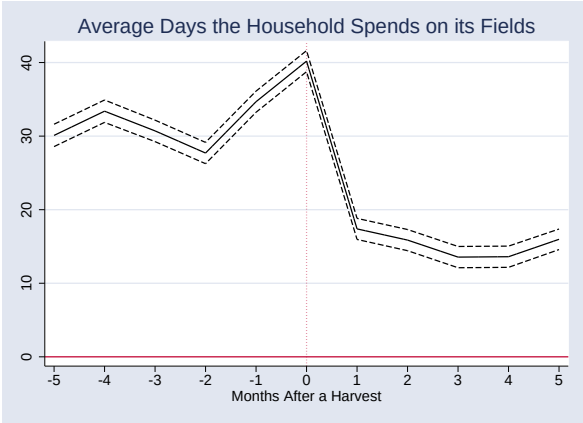
Note: The triple-difference estimator compares the response of a rice farmer expecting a harvest when rice prices are stable (**A**) to when they are volatile (**B**).

Figure 9
Rice Price and Predicted Mean and Conditional Standard Deviation



Note: Red lines denote start and end of the time span covered by the panel data.

Figure 10
An Impending Rice Harvest Requires Labor

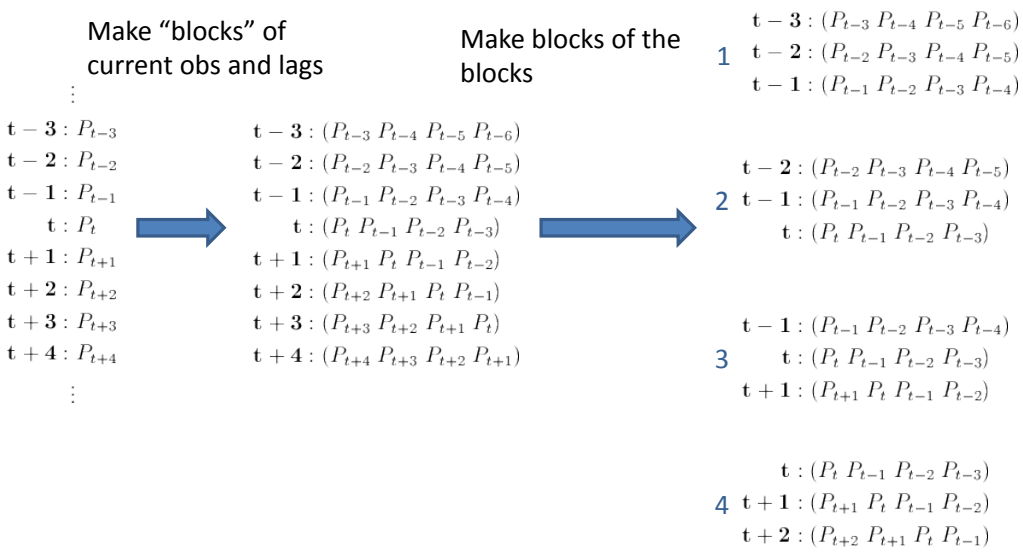


Note: The figure plots coefficients on dummies for periods before and after the harvest in a regression. Dashed lines are 95% confidence intervals.

Figure 11
Households Receive More Transfers when Prices are Low at Harvest

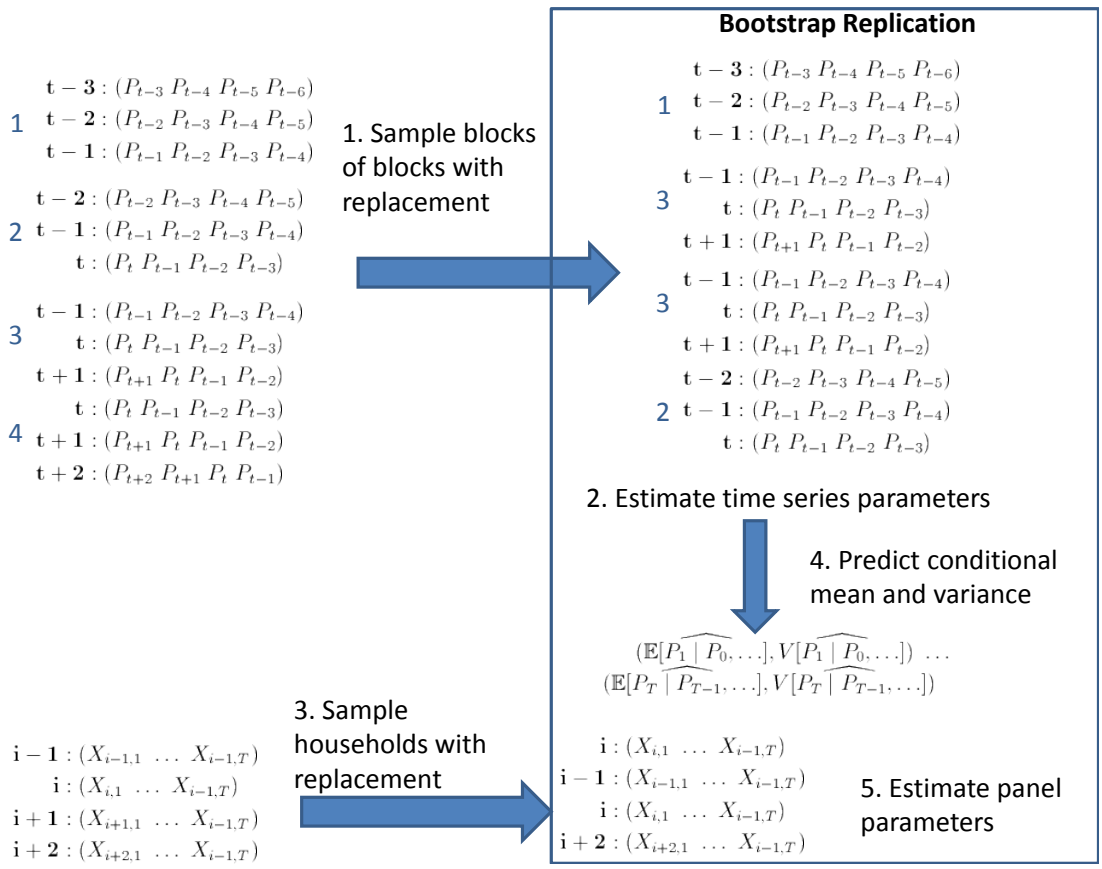
Note: The first bar depicts average incoming transfers for households harvesting rice when the international rice price is in the upper three quartiles of the periods of the sample. The second bar depicts the bottom quartile.

Figure 12
Bootstrap, Step 1: Forming Blocks of Blocks



Note: First, I prepare the time series of rice prices for resampling. I form "blocks" consisting of the contemporaneous price and however many lags I need to estimate the time series model. I then group every observation into one or more "blocks of blocks," contiguous interlocking sets of observations and their associated lags.

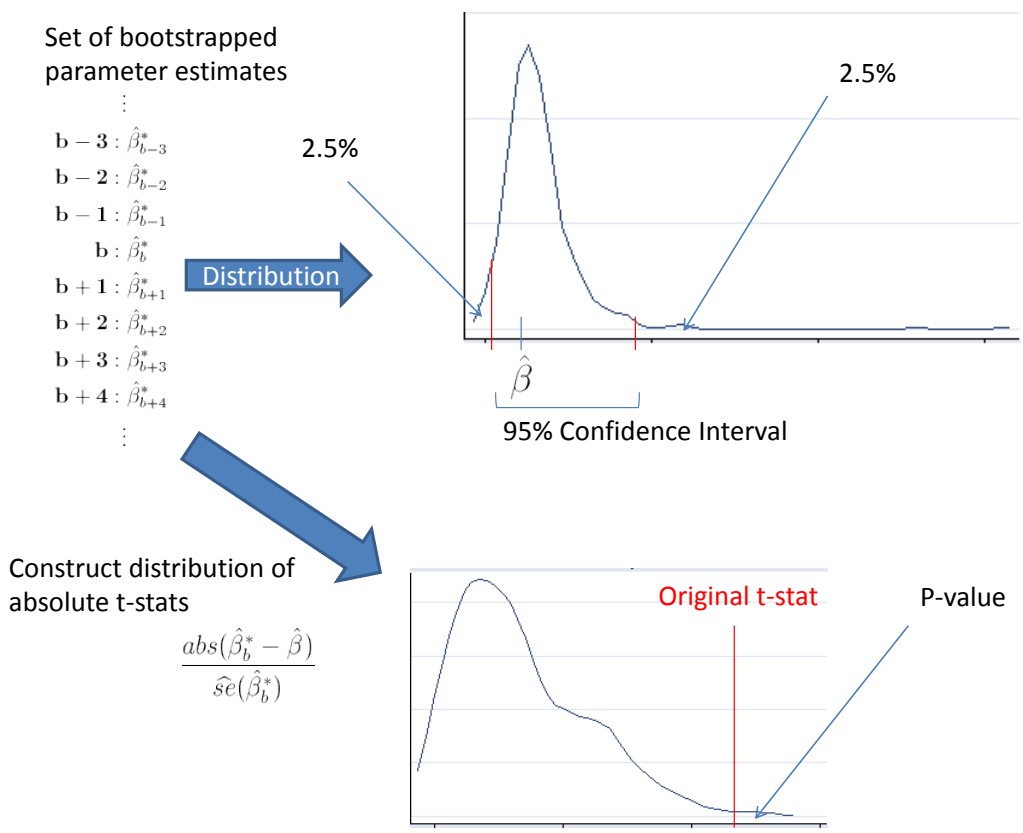
Figure 13
Bootstrap, Step 2: Bootstrap Replications



Note: Next, I run the bootstrap replications. Each replication follows five intermediate steps. First, I sample with replacement the blocks of blocks of rice prices to construct a bootstrapped time series of equal length to the original time series. I estimate the parameters of the time series model on the bootstrapped data. I then resample with replacement households (together with all their monthly observations) from the panel to construct a bootstrapped panel with as many households as the original panel. Then I use the estimated time series model to predict the conditional mean and variance of the international rice price for each household-month observation. Finally, I estimate the panel specification and record the resulting coefficients. I run 1000 replications for the risk specification and 2000 replications for the IV specifications.

Figure 14

Bootstrap, Step 3: Constructing Confidence Intervals and P-Values



Note: The final step is to compute confidence intervals and p-values. To construct confidence intervals, I use the dataset of estimated parameters from bootstrap replications to find the 2.5th and 97.5th percentiles. These are the boundaries of the 95% confidence interval. To construct p-values, I compute the absolute t-statistic centered around the original parameter estimate for each replication. The fraction of these absolute t-statistics that is greater than the original t-statistic is the p-value.