Identifying Information Asymmetries in Insurance:

Experimental Evidence on Crop Insurance from the Philippines

– DRAFT: Preliminary and Incomplete –

Snaebjorn Gunnsteinsson*
Department of Economics, Yale University.

October 25, 2012

*This is a preliminary version prepared for NEUDC 2012. Please find current version at http://www.snaebjorn.com. I would to extend thanks and gratitude to my advisors, Christopher Udry and Dean Karlan, for invaluable advise and support throughout this research project. I am grateful for advice from David Atkin, Xavier Gine, Daniel Keniston, Mushfiq Mobarak, Mark Rosenzweig and Nancy Qian. I would also like to thank Peter Srouji and Zoe VanGelder, who did a terrific job on field management, as well as the whole survey team at IPA Naga. Thanks to Melanie Morten, Camilo Dominguez, Treb Allen and Yaniv Stopnitzky for comments and discussions. I also thank the Philippines Crop Insurance Corporation for a productive collaboration. This research received generous support from the Australian Agency for International Development, the National Science Foundation, the Sasakawa Fund at the Yale Economic Growth Center, the Russell Sage Foundation and the Geneva Foundation for Insurance Economics.

Abstract

Why have insurance markets for many important risks failed to develop? Economic theory suggests this is most likely due to asymmetric information problems, but their presence is hard to identify empirically. In this paper I provide empirical estimates of information asymmetries in the insurance for crops, a market that has failed to develop for the most important types of natural hazards to crops. The evidence is based on a randomized field experiment among rice farmers in the Philippines that also includes an incentivized choice experiment and is combined with administrative data and detailed survey data. The design provides several unique tests of contract theory. I find strong evidence for both adverse selection and moral hazard, and provide estimates of magnitudes.

Most of the worlds' poor live in rural areas and depend to a large degree on agriculture for their livelihood. This dependence on agricultural production leaves rural households exposed to enormous risks from floods or droughts, hurricanes, pests or crop diseases. This risk depresses investment and access to credit, inhibits the development of markets and has serious direct welfare consequences. Yet, despite the importance of this risk, formal insurance markets to manage this risk have failed to develop[1].

Previous research has identified adverse selection, moral hazard and spatial co-variability of risk as the most likely causes of the failure of crop insurance markets. But empirical evidence for these information asymmetries is scarce since they are hard to identify using observational data.

The paper makes contributions in two main areas:

I) As a contribution towards understanding the failure of crop insurance markets, this paper examines in detail information asymmetries that may be

[1]The main exceptions to the lack of insurance markets is the development of markets for insurance that is relatively free from the problems of adverse selection and moral hazard, such as the recent development of index insurance markets (which are more akin to hedging instruments than insurance) and specialized markets such as hail insurance in Australia and Europe. Farmers in many countries, such as the US and the Philippines, also have access to government subsidized insurance but these are very expensive to governments and reportedly mired in problems with adverse selection and moral hazard.

present in a market for crop insurance. Do farmers have private information about the susceptibility of their land to natural disasters? And do they take advantage of this information in insurance market transactions? Do farmers shirk on preventative action on insured plots? Are these information asymmetries extensive enough to explain the failure of such markets?

To answer these questions, I designed and implemented a series of randomized field experiments of insurance. To complement the interventions I collected detailed survey data at the farmer and plot level. The survey data is further complemented by administrative data from the insurance provider.

My design incorporates both a choice experiment and randomized provision of insurance (at no cost) to farms and to plots within farms (using a two level randomization). This allows me to disentangle how adverse selection and moral hazard contribute separately to losses observed by the insurance company, shedding light on what role they play in the unravelling of markets for crop insurance.

II) The study design allows several unique contributions to empirical contract theory. The choice experiment involves choosing a plot for insurance from among a portfolio of the farmers plots. This allows for a test of adverse selection in relative insurance demand that allows for a sharp test of adverse selection on the risk type of the land. The design also allows me to disentangle the portion of adverse selection that can be explained by baseline characteristics into two conceptually different types of selection; 1) selection based on baseline risk (that is the risk type of the plot independent of effort) and 2) selection based on the amount of effort saved (i.e., intended moral hazard).

The design also allows empirical estimates of moral hazard using random variation of insurance coverage across production units controlled by the same individual and across individuals. This is the first test of its kind and will allow an examination of resource allocation within the farm in response to changing incentives. This is also the first experimental estimate (to my knowledge) of moral hazard in an insurance market other than health insurance [2]. Finally, by randomizing within the farm we also obtain a test that is more robust than alternative cross-farm test (e.g. to non-random attrition of farmers).

I find that farmers have substantial private information about the susceptibility of their land to natural hazards and take advantage of this private information when selecting plots for insurance. These results provide strong evidence for adverse selection. I also find substantial evidence for moral hazard in the prevention of pest and crop disease damage. Damages from pests and crop diseases are estimated to be about 20% higher on insured plots. Overall, the results support theoretical models of information asymmetries in insurance and show that they pose a serious impediment to the development of crop insurance markets. In future revisions of this paper I will explore the mechanisms underlying these information asymmetries and discuss implications for possible progress on insuring crops for natural hazards.

1 Identification of information asymmetries

Under normal market conditions it is a formidable challenge to separately identify the different information asymmetries at play between an agent and a principal. A recent contribution by Karlan and Zinman (2008) made progress on identifying asymmetries in credit markets using a randomized field experiment. But most tests of information asymmetries in insurance markets rely on the positive correlation test(Chiappori and Salanie, 2000). The basic procedure is to test whether the choice of insurance coverage, conditional on all variables used in pricing the insurance contract, is positively correlated with risk occurrence. The test is based on estimating two equations, the first relating the amount of insurance coverage C to variables used in policy

[2] Only two known experiments have been done in health insurance, the famous RAND Health Insurance Experiments(Manning et al., 1987) and more recently the Oregon Health Insurance Experiment(Finkelstein et al., 2011)

pricing, X:

$$C = \alpha_1 + \alpha_2 X + \epsilon \tag{1}$$

and the second relates occurrence of accident (or loss amount) L to the same set of variables used in pricing:

$$L = \beta_1 + \beta_2 X + \eta \tag{2}$$

The positive correlation test is based on testing for a positive correlation between ϵ and η. This test has three main limitations: 1) it does not distinguish between adverse selection and moral hazard, 2) it is not able to distinguish selection based on risk from selection based on preferences, 3) it is unable to test between selection based on baseline risk (independent of effort) versus selection based on the costs of effort.

The two randomized health insurance experiments (Manning et al., 1987; Finkelstein et al., 2011) conducted so far are an exception to the former. Finkelstein and McGarry (2006) is an exception to the latter, as they provide evidence of multiple dimensions of private information (preference for insurance versus risk type) in the long-term care insurance market based on data on individuals' subjective assessment of their risk type.

Many tests of adverse selection in insurance available in the literature fail to find evidence for adverse selection. It is likely that in many cases this is because of interdependence between preferences and risk which can cause zero adverse selection on average, or even advantageous selection, even if agents take advantage of their private information in risk(Meza and Webb, 2001). In this paper I will be able to separately identify moral hazard and adverse selection and test for adverse selection using a test that allows a shaper test of whether agents take advantage of private information on damage probabilities. In later drafts I will also separately identify selection based on baseline risk from selection based on the cost of effort.

2 Experimental Design and Implementation

2.1 Context

The data come from a series of field experiments among rice farming households in Bicol province in the Philippines that were designed and directed by the author and implemented by Innovations for Poverty Action[3].

Rice is cultivated twice a year in Bicol. The period from June through October (the ”wet season”) is characterized by heavy rains and – especially in September and October – by high incidence of typhoons. It is generally the less productive of the two. The period from November through May (the ”dry season”) is characterized by more moderate rains, low typhoon activity and is consequently usually more productive.

2.2 Insurance Contract

Agricultural production in the Philippines is susceptible to not only the usual threats of floods, droughts, pests and corp diseases, but also the arrival of more than (on average) 15 typhoons or tropical cyclones per year (Virola, 2008). A government agency, the Philippines Crop Insurance Corporation, offers crop insurance to farmers throughout the Philippines but take-up is very low and, as elsewhere, no private market has developed. PCIC's most important products are multi-peril insurance policies for rice and corn and we will focus on understanding the information asymmetries in PCIC's insurance scheme for rice, the most important staple crop in the Philippines.

The insurance offered by PCIC is a multi-peril crop insurance that covers all the major specific natural hazards in this context. This includes typhoons, floods, droughts, and various pests (e.g. rats, various insects) and crop diseases (e.g. tungro, a crop disease spread by insects). The insurance also

[3]Innovations for Poverty Action is a research NGO based in New Haven (CT, USA) that conducts field research on poverty and development policy; see www.poverty-action.org for more details.

covers rare events such as volcanic eruptions and earthquakes but excludes some minor pests such as birds and snails[4]. Any particular damage event must cause at least 10% loss of harvest to be considered by the insurance company. If a damage event causes more than 10% damage, an insured farmer files a Notice of Loss to the company, which sends an insurance adjuster to verify damages. The contract pays out per hectare of insured land in proportion to the share of harvest lost to specific causes. This structure of the contract has important implications for investment incentives, but we defer this discussion to later drafts. In this context, typhoons and floods are the major source of damages to crops, accounting for about 60% of indemnities, while various pests and crop diseases account for about 40%.

2.3 Experimental Design

In order to overcome the limitations of the standard positive correlation test I introduce two key features into the design that allow me to disentangle many of the relevant information asymmetries: 1) I take advantage of the fact that farmers in this context routinely till multiple plots of land and designed the experiment and data collection to consider the plot as the base unit of analysis and 2) I introduce experimental variation across plot within the same farm and obtain incentivized choices at the plot level. This will allow me to disentangle adverse selection and moral hazard, allow a test of adverse selection that is sign-independent of preferences and constraints (I will describe this below) and allows for moral hazard tests both across plots within the same farm and across farms.

The study design was the following:

Step 1: Each farmer was asked, if they could choose one plot to be

[4] We ignore damages from birds and snails in the analysis. The amount of damages from birds are trivial. Losses from snails are non-trivial but small, and occur primarily when plants are seedlings (before transplanting) so it is impossible to assign per-plot damage rates.

covered by free insurance, which plot they would choose. They were told that the plot they chose would have a higher chance of receiving free insurance in a lottery.

Step 2: Farmers were then entered into a lottery and randomly allocated to three groups:

Group A (66.5%; Full Randomization): Received insurance on a random half of plots.

Group B (3.5%; Choice): Received insurance on 1st choice plot and a random half of remaining plots.

Group C (30%: Control): Received no insurance.

Group B is a truth-telling mechanism; it ensures that it is incentive compatible for the farmer to reveal her true preference. The farmer level randomization was stratified by geographic location[5]. Insurance was allocated to plots in Group A using a block randomization within the farm such that half of the farmers plots received insurance. Farmers with an odd number of plots, n, were randomly selected to receive insurance on $\frac{n-1}{2}$ or $\frac{n+1}{2}$ plots. After insurance had been allocated to the 1st choice plots of farmers in Group B, their remaining plots were randomly allocated insurance using the same procedure as in Group A.

The study was explicitly designed to separately estimate adverse selection and moral hazard. Figure 1 depicts the basic identification strategy. To identify adverse selection I compare the 1st choice plot of the farmer to her other plots. We will test this both by comparing measures of predicted damages and actual damages and payouts. When comparing actual damages this comparison should, theoretically, be done between two insured plots. To identify moral hazard I compare an insured and uninsured plots of farmers in the fully random group (Group A).

[5] In the first season the experiments were conducted in a relatively small geographic location and we stratified by the number of plots instead.

In principle, the design allows me to identify moral hazard separately for 1st choice plots and for other plots, see Figure 2. I will discuss the identification of separate components of adverse selection (as depicted in the figure) in Section 6.

2.4 Implementation

The experiments were implemented over three cropping seasons from fall of 2010 through mid-2012. IPA staff invited farmers who were tilling (as primary tillers) two or more rice plots within the Tigman Hinagyanan Inarihan Regional Irrigation System and surrounding communal irrigation systems to participate in research project on crop insurance. A farmer was considered a primary tiller of a plot if he was the main decision maker for production decisions and was the person taking on the majority of the production risk. Eligible plots were restricted to those between 0.25 and 2.5 hectares[6].

A total of 106, 285 and 447 farmers were enrolled in the three seasons, with a total of 291, 806 and 1302 parcels, or about three parcels per farmer on average. Season 1 is the dry season of 2010-11, Season 2 the wet season of 2011 and Season 3 the dry season of 2011-12. In some cases estimations are based on Seasons 2 and 3 only since data collection was more limited in Season 1.

3 A Model of Preventative Effort and Insurance Choice on a Portfolio of Plots

Consider an agricultural economy where each farmer, indexed by i, farms a portfolio of plots, indexed by j. I omit the farmer subscript when possible. Plot j produces a maximum output of G_j per hectare and is A_j hectares in

[6]The vast majority of plots fall into this range. The lower bound is an eligibility requirement of the insurance company. Some exceptions from this lower bound were given in the first season.

size. Some of this output can be lost if the farm is hit by a natural hazard. The damages from natural hazards are $D_j \in [0,1]$ which follow a distribution $F_{(\theta_j,e_j)}$. The shape of the distribution is a function of θ_j, which indexes the baseline risk of damage on plot j and e_j, the effort that the farmer puts forth to prevent damages on the plot. We denote the mean and variance of this distribution by $(\mu(\theta_j,e_j),\sigma^2(\theta_j,e_j))$ and assume that the mean is increasing in risk ($\frac{\partial\mu}{\partial\theta} > 0$) and that μ and σ^2 are decreasing and convex in effort[7]. A plot may be insured ($\alpha_j = 1$), in which case the farmer receives a LD_j payout. Therefore, profits of farmers are stochastic and given by

$$\Pi = \sum \Pi_j = \sum_j A_j(G_j(1-D_j) + \alpha_j LD_j - e_j) \tag{3}$$

The farmer has a utility function over output, u_ρ, indexed by a risk aversion parameter (ρ). Her maximization problem is to choose one plot to be insured and then choose effort on each of her plots:

$$\max_{(\alpha_1,..,\alpha_n,e_1,..,e_n)} E\left[u(\sum_j \Pi_j)\right] \tag{4}$$

subject to $\alpha_j \in \{0,1\}$, $\sum_j \alpha_j = 1$ and $e_j \geq 0$. For tractability I assume that utility is exponential and that the distribution F is such that $E(u(c)) \approx u(\mu_c - \frac{1}{2}\rho\sigma_c^2)$[8]

3.1 Implications

An appendix with detailed comparative statics will be provided in later drafts. The key implications are that effort is decreasing in insurance coverage (moral hazard) and that, in first approximation, the farmer chooses the plot that maximizes her expected payout. Given the experimental setup,

[7]That is, $\frac{\partial\mu}{\partial e} > 0$, $\frac{\partial^2\mu}{\partial e^2} > 0$, $\frac{\partial\sigma^2}{\partial e} > 0$ and $\frac{\partial^2\sigma^2}{\partial e^2} > 0$.

[8]For exponential utility, the relation is exact if the distribution is normal and is a good approximation for many other distributions(Kirkwood, 1997).

I make a distinction between insurance choice and actual insurance coverage. Let $\alpha_j^C = 1$ if the farmer chose plot j and let $\alpha_j^A = 1$ if the farmer got insurance in the experiment. Then she chooses the plot that maximizes $A_j\mu(\theta_j, \hat{e}(\theta_j, \alpha_j^A = 1))$ where $\hat{e}(\theta_j, \alpha_j^A = 1)$ is the optimal effort on plot j if it is insured.

3.2 Empirical Strategy

In the empirical analysis in Section 5 I will use ex-post damages to test for information asymmetries. To obtain an estimation equation that allows for estimation of adverse selection and moral hazard within the farmers portfolio I derive approximations of effort and damages around the average plot of the farmer. First, approximate effort around the average plot is

$$\hat{e}(\theta_j, \alpha_j^A, \rho) = \frac{\partial \hat{e}_j}{\partial \alpha}(\alpha_j^A - \bar{\alpha}^A) + \frac{\partial \hat{e}_j}{\partial \theta_j}(\theta_j - \bar{\theta}) + \hat{e}(\bar{\theta}, \bar{\alpha}^A, \rho). \tag{5}$$

A first approximation of expected damages around the average plot is

$$\mu(\theta_j, \hat{e}_j(\theta_j, \alpha_j^A)) \approx \frac{\partial \mu}{\partial \theta}(\theta_j - \bar{\theta}) + \frac{\partial \mu}{\partial e}(\hat{e}_j(\theta_j, \alpha_j^A) - \bar{e}) + \mu(\bar{\theta}, \bar{e}). \tag{6}$$

Using the approximation for effort this gives:

$$\begin{aligned} \mu(\theta_j, \hat{e}_j(\theta_j, \alpha_j^A)) \approx & \left[\frac{\partial \mu}{\partial \theta} + \frac{\partial \mu}{\partial e}\frac{\partial e}{\partial \theta}\right]\theta_j + \frac{\partial \mu}{\partial e}\frac{\partial e}{\partial \alpha_j^A}\alpha_j^A \\ & + \underbrace{\mu(\bar{\theta}, \bar{e}) + \frac{\partial \mu}{\partial e}\frac{\partial e}{\partial \bar{\alpha}^A}\bar{\alpha}^A + \frac{\partial \mu}{\partial e}\frac{\partial e}{\partial \theta}\bar{\theta}}_{\lambda_i} \end{aligned}$$

This motivates the following estimation equation:

$$D_{ij} = \beta_0 + \overbrace{\beta_1}^{\frac{\partial \mu}{\partial e}\frac{\partial e}{\partial \alpha_j^A}} \alpha_{ij}^A + \beta_2 \alpha_{ij}^C + \beta_3 Area_{ij} + \lambda_i + \overbrace{\epsilon_{ij}}^{\left[\frac{\partial \mu}{\partial \theta} + \frac{\partial \mu}{\partial e}\frac{\partial e}{\partial \theta}\right]\theta_j + \overbrace{\nu_{ij}}^{\text{idiosyncratic}}} \tag{7}$$

Consider estimating this equation on the sample of farmers who got their plots insured by random. Then α^C and α^A are independent and β_1 identifies the causal effect of insurance coverage on damages. Now since $\frac{\partial \mu(\theta, e)}{\partial e} \leq 0$ by assumption, $\hat{\beta}_1 > 0$ is a test for moral hazard since it implies that the average change in effort that follows insurance coverage is negative, i.e., $\int \frac{\partial e(\theta, \alpha_A)}{\partial \alpha^A} d\theta < 0$.

It also follows that $\hat{\beta}_2 = \left[\frac{\partial \mu}{\partial \theta} + \frac{\partial \mu}{\partial e}\frac{\partial e}{\partial \theta}\right](\theta_j - \bar{\theta}_{-j})$. I assume, naturally, that $\frac{\partial \mu}{\partial \theta} + \frac{\partial \mu}{\partial e}\frac{\partial e}{\partial \theta} > 0$ so that $\hat{\beta}_2 > 0$ is a test of adverse selection.

4 Data and Experimental Integrity

4.1 Data

The data come from the following sources: 1) Plot choices obtained at enrollment in the study (if a farmer participated in multiple seasons, a new choice was obtained before each season); 2) Plot characteristics from a baseline survey; 3) Input data from mid-season and follow-up surveys; 4) Output and damage data from a follow-up survey; 5) Administrative data from the insurance provider; and 6) Geo-spatial data collected by study staff.

The Data Appendix A provides a description of the construction key measures.

4.2 Integrity of experiments

In Table 6 I report the sample sizes and attrition for each stage of the randomization. Panel A shows attrition by treatment in the first stage of the randomization that allocated farmers to treatment or control groups. The sample size grew from 107 farmers in the first season to 447 farmers in the last season. Attrition among farmers is not trivial, particularly in the control group in the first two seasons where only about 75-79% of farmers complete a follow-up survey, compared to 88-92% of the treatment groups. In the last season attrition is considerably improved and 89% and 92% of control and treatment groups complete the follow-up survey, respectively. Some follow-up surveys contain no data on damages (in some cases the farmer doesn't know) which leads us to have some damage data on 90% and 85% of treatment and control groups in the last season, respectively. Overall we have some damage data from 87% and 79% of farmers in the treatment and control groups.

Panel B shows attrition of plots conditional on us having some damage data from the farmer and conditional on the plots being in the randomization group (that is, not in Group C and not a first choice plot of Group B). This is therefore relevant for examining the internal validity of estimates based on the second stage randomization. Overall we observe data on about 87-88% of these plots and this is balanced across the two treatment groups.

Table 5 shows balance checks across the two stages of randomization. In both cases the randomization is well balanced on baseline observables both at randomization and for the sample of farmers and plots for which we have harvest data. The plot randomization is also clearly orthogonal to the choice of a 1st-choice-plot.

In this version of the paper I show only treatment estimates using the plot randomization along with farmer fixed effects. Overall, the evidence suggests that the integrity of the plot randomization was maintained.

5 Empirical Estimates of Adverse Selection and Moral Hazard

In Table 1 I estimate Equation 7 for total damages and separately for typhoons and floods, and for pests and crop diseases. This distinction is motivated by expectations at the start of the project that pests and crop diseases would be more preventable than typhoons and floods, and by the fact that this is a categorization the company uses already[9].

I find strong evidence for adverse selection in overall damages and separately for both typhoons & floods and pests & crop diseases. Damages are estimated to be 4.7 percentage points (95% CI: 2.0 - 7.3) greater on first choice plots compared to other plots of the same farmer from a base of 24%. The first choice plots therefore have about 20% higher damages. I estimate damages to be 17% higher due to typhoons and floods and 21% higher due to pests and crop diseases. This evidence is consistent with evidence based on administrative data from the insurance provider, shown in column 4. I estimate that payouts per hectare are about 50% higher on insured first choice plots compared to other insured plots of the same farmer. This estimate is quite imprecise, (95% CI: -0.8 - 14.4) and only estimated from a small set of farmers who have two insured plots (one of which is the first choice plot).

I find evidence for moral hazard in preventing pests and diseases but no evidence for moral hazard in preventing typhoons and floods. Damages from pests and crop diseases are estimated to be 1.92 percentage points (95% CI: 0.2 - 3.6) higher on insured plots compared to randomly un-insured plots of the same farmer of a baseline of 8.4%. This translates into about 24% increase in damages. At the moment standard errors are calculated using regular OLS. In future revisions these will be calculated using spatial standard errors following the procedures developed in (Conley, 1999). Earlier

[9] The insurance company offer two types of coverage, a basic coverage that covers only typhoons and floods, and a comprehensive coverage that also includes coverage for pests and crop diseases. The insurance studied in this paper is the comprehensive coverage.

estimation suggests this will reduce standard errors on the treatment effect by about 10-15%.

6 The Nature of Selection

In this section I test for adverse selection and separately estimate adverse selection based on risk characteristics of the land from adverse selection based on the cost of preventative effort. The key idea for the decomposition is presented in Figure 2. I first compute predicted damages based on baseline information for insured and uninsured plots separately (for farmers in the fully random group). Then I can identify overall adverse selection by comparing the predicted damages for insured 1st choice plots to insured other plots (all comparisons here are across plots of the same farmer). This I can disentangle into two effects: 1) selection on baseline risk (independent of effort) by comparing predicted damages on uninsured 1st choice plots to predicted damages on uninsured other plots and 2) selection on intended moral hazard by taking the difference in predicted change in damages, when moving from being uninsured to being insured (moral hazard), between first choice plots and other plots.

I section 6.2 I develop a model of selection in the experiment and show how this model maps into the framework for conditional logit analysis. I estimate the conditional logit model and test for overall adverse selection and test for the presence of the two separate components in Section 6.3.

6.1 Baseline Characteristics Determine 37% of Selection

In Table 3 I estimate equations of the form

$$D_{ij} = \beta_0 + \beta_1 C_{ij} + \beta_2 X_{ij} + \beta_3 X_{ij} 1(\text{dry season}) + \lambda_i + \eta_{ij}. \tag{8}$$

I only use seasons 2 and 3 for this estimation since the relevant baseline characteristics were not collected in the first (pilot) season. The observables used to predict damages are taken from the baseline and are based on a series of questions that asked "Compared to your other plots, does this plot have low, medium or high risk of ______?" where I ask separately for floods, rats and tungro (a crop disease). In addition I have questions that asked whether the plot is easy, medium or hard to drain after heavy rains, compared to the farmers other plots, and whether the plot is low, medium or high-lying, compared to the farmers other plots. I combine the questions pertaining to floods (flood risk, low-lying and hard to drain) into one index by taking the first principal component from a PCA of three binary variable that signify that the plot is high risk of floods, is low-lying, and is hard to drain after floods. The questions for rats and tungro are added by using binary indicators for the medium and high categories.

The estimated selection effect in Column 2 is 37% lower than in Column 1, where β_2 and β_3 are constrained to zero. In what follows of this section I will decompose this portion of selection into two conceptually distinct effects.

6.2 Decomposition of Selection on Baseline Characteristics

The structure of the selection problem is analogous to the one analyzed by McFadden (1974). Instead of the case of consumers who select between transportation options based on their characteristics (such as, e.g., price and time-to-destination), I apply the model to the farmers choice problem of selecting one plot for insurance from her portfolio of plots based on plot characteristics. The key assumptions that underlie McFadden's model – i.e., that choice probabilities are positive, independent of irrelevant alternatives and what McFadden terms "irrelevance of alternative set" – seem reasonable for this context but we will keep these in mind when we discuss robustness later on.

A key distinction in this problem in contrast to the problem analyzed by McFadden is the role of effort. I assume farmers are sophisticated and take into account their endogenous provision of effort on insured plots. Let $\hat{e}_j^I$ be the farmers optimal choice of effort on plot j if the plot is insured and likewise $\hat{e}_j^0$ for an uninsured plot. Given the assumption on independence across plots, the solution to the maximization problem in 4 is that the farmer chooses plot j if

$$\alpha * = \arg \max_{\alpha_j, j \in \{1,..,n\}} u(\Pi_j^{'}(1, e_{j,\alpha_j=1})) - u(\Pi_j^{'}(0, e_{j,\alpha_j=0})) \tag{9}$$

subject to $\alpha_j \in \{0, 1\}$ and $\sum_j \alpha_j = 1$. Applying the mean-value theorem this implies that

$$\alpha * = \arg \max_{\alpha_j, j \in \{1,..,n\}} u'(c) \left[\Pi_j^{'}(1, e_{j,\alpha_j=1}) - \Pi_j^{'}(0, e_{j,\alpha_j=0}) \right] \tag{10}$$

for $c \in \left[\Pi_j^{'}(1, e_{j,\alpha_j=1}), \Pi_j^{'}(0, e_{j,\alpha_j=0} \right]$. The second term combines the two key adverse selection effects – selection based on baseline risk versus selection based on opportunities for moral hazard. Rewriting I obtain an expression that illustrates the three key effects that impact the selection decision:

$$\alpha * = \arg \max_{\alpha_j, j \in \{1,..,n\}} u'(c) \left[\overbrace{\Pi_j^{'}(1, e_{j,\alpha_j=0}) - \Pi_j^{'}(0, e_{j,\alpha_j=0})}^{\text{Utility from coverage of baseline risk}} \right. \tag{11}$$

$$\left. + \underbrace{\Pi_j^{'}(1, e_{j,\alpha_j=1}) - \Pi_j^{'}(1, e_{j,\alpha_j=0})}_{\text{Utility from coverage of moral hazard}} \right] \tag{12}$$

Based on equation 3 and the approximation above, the mean and variance of $\Pi_j^{'}$ are:

$$\mu_{\Pi_j} = A_j G_j - A_j G_j \mu_j + \alpha_j A_j L \mu_j - c_j e_j$$

$$\sigma_{\Pi_j}^2 = A_j^2 (\alpha_j L - G_j)^2 \sigma_j^2$$

where μ_j and σ_j are the mean and variance of $F(\theta_j, e_j)$. Using these equation 12 can be rewritten as:

$$\alpha* = \arg \max_{\alpha_j, j \in \{1,..,n\}} u'(c) \Bigg[\overbrace{\mu(\theta_j, \hat{e}_j^0) A_j L - \frac{1}{2} \rho A_j^2 (L - G_j)^2 \sigma(\theta, \hat{e}_j^0)}^{\text{Utility from coverage of baseline risk}} \tag{13}$$

$$+ \underbrace{A_j(L - G_j) \left[\mu(\theta_j, \hat{e}_j^I) - \mu(\theta_j, \hat{e}_j^0)\right] - \frac{1}{2} \rho A_j^2 (L - G_j)^2 \left[\sigma^2(\theta_j, \hat{e}_j^I) - \sigma^2(\theta_j, \hat{e}_j^I)\right] - c_j(\hat{e}_j^I - \hat{e}_j^0)}_{\text{Utility from coverage of moral hazard}} \Bigg]$$

The expression illustrates the three key tradeoffs that the farmer faces in choosing a plot for insurance: 1) The benefits of insurance coverage for baseline risk (independent of effort), 2) The benefits of saved effort and 3) The benefits of insurance coverage on a plot based on other plot characteristics, such as plot area. The farmers preferences play a role in assigning relative weights to each of these benefits. The first two involve the farmer taking advantage of private information that will lead to higher damage probabilities and ultimately to increased exposure for the insurance provider. In the next three subsections I map this model into an empirical conditional logit specification to estimate the importance of each of these effects.

6.3 Adverse Selection Based on Predicted Damages

In this and the next subsection I assume risk preferences have an ignorable effect on the relative preference ranking of plots. In this case 13 simplifies to

$$\alpha* = \arg \max_{\alpha_j, j \in \{1,..,n\}} \overbrace{\mu(\theta_j, \hat{e}_j^1) A_j L}^{\text{Overall Utility}} \tag{14}$$

$$= \arg \max_{j \in \{1,..,n\}} \Bigg[\overbrace{\mu(\theta_j, \hat{e}_j^0) A_j L}^{\text{Utility from baseline risk coverage}} \tag{15}$$

$$+ \underbrace{A_j(L - G_j) \left[\mu(\theta_j, \hat{e}_j^I) - \mu(\theta_j, \hat{e}_j^0)\right] - c_j(\hat{e}_j^I - \hat{e}_j^0)}_{\text{Utility from moral hazard coverage}} \Bigg]$$

The empirical analog of the expression in 14 is (L is constant and is ignored)

$$\hat{v}(X) = A_{ij} \hat{E}\left[D | X, I = 1\right]$$

Empirically I estimate a conditional logit model, where the choice is conditioned to the portfolio of each farmer, of the form:

$$\Lambda(C_{ij}) = \alpha_0 + \alpha_1 A_{ij} \hat{E}\left[D | X, I = 1\right] + \alpha_2 A_{ij} + \epsilon_{ij} \tag{16}$$

where $C_{ij} = 1$ if farmer i chose plot j as her first choice. Here $\alpha_1 > 0$ provides a test for adverse selection. The empirical analog of 15 is

$$\begin{aligned} \Lambda(C_{ij}) = \alpha_0 &+ \alpha_1 A_{ij} \hat{E}\left[D | X, I = 0\right] \\ &+ \alpha_2 A_{ij} (\hat{E}\left[D | X, I = 1\right] - \hat{E}\left[D | X, I = 0\right]) \\ &+ \alpha_3 A_{ij} + \epsilon_{ij} \end{aligned} \tag{17}$$

Now $\alpha_1 > 0$ provides a test for selection based on baseline risk and $\alpha_2 > 0$ provides a tests for selection based on opportunities for moral hazard (e.g., selection on hidden cost of effort). I will also estimate analogous linear models for comparison. To empirically estimate 16 and 17 I must first obtain empirical estimates of predicted damages $\hat{E}\left[D|X, I=0\right]$ (for uninsured plots) and $\hat{E}\left[D|X, I=1\right]$ (for insured plots).

6.3.1 Predicted damages based on baseline characteristics

In Table 3 I estimate models of the form:

$$D_{ij} = \beta_0 + \beta_1 X_{ij} + \beta_2 X_{ij} 1(\text{dry season}) + \lambda_i + \eta_{ij} \tag{18}$$

separately for insured and uninsured plots of farmers in the pure randomization group (Group A). Here X indicates the baseline characteristics used for prediction and λ_i are farmer fixed effects. The key notable difference is that medium and high risk of tungro (a crop disease) predicts damages strongly for the insured group but not for the control group.

6.4 Estimated selection effects

Using these predicted damages I can now empirically estimate equations 16 and 17. Table 4 presents these results. The sample for these regressions includes only farmers who were in the full randomization group (Group A: Received insurance on half of plots at random). The reason for this is that since the predicted values are calculated using this group this prevents differential attrition by treatment group or differential reporting across treatment and control groups to affect estimates. In Column 1 I estimate equation 16 and find strong evidence for adverse selection. A one percentage point increase in predicted damages increases the odds of a plot being chosen by 8%.

In column 2 I estimate equation 17. I find very strong evidence for adverse selection on baseline risk. I now estimate that a one percentage point increase in baseline risk increases the odds of a plot being chosen by 9%. I also find evidence of selection on opportunity for moral hazard and estimate that a one percentage point higher moral hazard effect ex-post leads to a 5% greater odds of choosing a particular plot.

7 Conclusions

I find strong evidence of adverse selection. Farmers have substantial private information about the susceptibility of their land to natural hazards and take advantage of this information when selecting plots for insurance. Decomposing the part of selection that is explained by baseline characteristics I find strong evidence that farmers select on baseline risk (risk independent of effort) and some evidence that farmers also select plots that are allow for greater moral hazard (f.e., plots that are require high effort to keep free of pests and insects).

I find substantial evidence of moral hazard in the prevention of pest and crop disease damage. Damages from pests and crop diseases are estimated to be about 20% higher on insured plots. An analysis of the mechanisms is in progress but preliminary analysis suggests the primary mechanism of moral hazard may be through shifting planting dates forward.

Overall, the results show that farmers in this context generally know their plots well and are sophisticated in their risk management decisions. The results align well with models in contract theory and suggest the presence of substantial information asymmetries in this context. In future revisions of this paper I will investigate the mechanisms that underlie these information asymmetries and discuss implications for making progress on insuring crops for natural hazards.

References

Chiappori, PA and B Salanie, "Testing for asymmetric information in insurance markets," *Journal of Political Economy*, January 2000.

Conley, T.G., "GMM estimation with cross sectional dependence," *Journal of Econometrics*, 1999, *92* (1), 1–45.

Finkelstein, Amy and Kathleen McGarry, "Multiple Dimensions of Private Information: Evidence from the Long-Term Care Insurance Market," *The American Economic Review*, September 2006, *96* (4).

_ , Sarah Taubman, Bill Wright, Mira Bernstein, Jonathan Gruber, Joseph P Newhouse, Heidi Allen, Katherine Baicker, and The Oregon Health Study Group, "The Oregon Health Insurance Experiment," NBER WORKING PAPER 2011.

Karlan, D and J Zinman, "Observing Unobservables: Identifying Information Asymmetries with a Consumer Credit Field Experiment," *Econometrica*, 2008.

Kirkwood, C W, "Notes on attitude toward risk taking and the exponential utility function," 1997.

Manning, W G, J P Newhouse, N Duan, E B Keeler, and A Leibowitz, "Health insurance and the demand for medical care: evidence from a randomized experiment," *American Economic Review*, 1987, pp. 251–277.

McFadden, D, "Conditional Logit Analysis of Qualitative Choice Behavior," in P Zarembka, ed., *Frontiers in Econometrics*, Academic Press: New York, 1974.

Meza, David and David Webb, "Advantageous Selection in Insurance Markets," *The RAND Journal of Economics*, July 2001, *32* (2), 249–262.

Virola, Romulo, "Statistically Speaking... Some things you better know about typhoons in the Philippines!," August 2008.
World Development Report 2008: Agriculture for Development

World Development Report 2008: Agriculture for Development*, Technical Report 2007.*

Tables

Table 1: Empirical Estimates of Adverse Selection and Moral Hazard

	Loss (%) Due to:			
	All Causes	Typhoons and floods	Pests and diseases	Payout ($) per hectare
1st choice	$4.65***$	$2.65**$	$1.77*$	$6.81*$
	(1.36)	(1.13)	(0.99)	(3.89)
Insurance	1.26	-0.33	$1.92**$	
	(1.19)	(0.99)	(0.86)	
Area (hectares)	2.28	0.58	1.62	-2.68
	(2.19)	(1.82)	(1.60)	(5.60)
Farmer-season FE	Yes	Yes	Yes	Yes
Sample:	Plots of farmers in the fully random group			Insured plots
Mean of non-first choice plots	23.7	15.8	8.4	12.2
Mean of non-insured plots	24.6	17.0	7.8	
Num FE's	474	474	474	606
Observations	1217	1217	1222	899

The table shows separate estimation adverse selection and moral hazard using estimation equation 7. Percentage loss is calculated as $\frac{\text{Value of loss due to 'Type of cause'}}{\text{Value of loss 'All causes'+Value of harvest}}$ and are based on self-reports from a follow-up survey. The payouts are based on administrative data from the insurance provider.

Table 2: Estimated Share of Adverse Selection Explained by Baseline Characteristics

	Loss (%) Due to: All Causes	
	Eq. (1)	Eq. (2)
1st choice	5.50 * **	3.76 * *
	(1.40)	(1.47)
Area (hectares)	2.32	3.57
	(2.22)	(2.39)
Plot Risk Characteristics	No	Yes
Farmer-season FE	Yes	Yes
Sample:	Plots of farmers in the fully random group	
F-test All Risk Characteristics		p = 0.00
Mean of dependent variable for non-first choice plots	23.7	23.7
Num FE's	434	425
Observations	1132	1055

The table only includes data from Season 2 and 3 since these baseline characteristics were not collected in Season 1. Significance: * $<$ .1; ** $<$ 0.05; *** $<$ 0.01.

Table 3: Estimation of Predicted Damages by Treatment Group

	Loss (%) Due to All Causes	
	Insured Plots	Control Plots
Flooding index	3.95*	6.97***
	(2.13)	(2.06)
Medium risk of rats	5.21	4.83
	(4.19)	(4.16)
High risk of rats	5.47	1.04
	(5.25)	(4.85)
Medium risk of tungro	7.99*	−1.28
	(4.54)	(4.16)
High risk of tungro	13.2**	2.96
	(5.68)	(5.98)
Season 3 X Flooding index	0.17	−4.56*
	(2.73)	(2.54)
Season 3 X Medium risk of rats	−6.16	−7.40
	(4.62)	(4.60)
Season 3 X High risk of rats	−8.59	−9.44
	(6.05)	(5.75)
Season 3 X Medium risk of tungro	0.086	3.02
	(5.37)	(4.92)
Season 3 X High risk of tungro	−7.04	−2.65
	(7.05)	(7.18)
Area (hectares)	−1.11	5.17*
	(2.95)	(2.81)
Constant	22.1***	23.2***
	(3.34)	(3.13)
Observations	525	526

Estimates of Equation 18 for insured (Column 1) and control (2) plots of farmers in the fully random group. The table only includes data from Season 2 and 3 since these baseline characteristics were not collected in Season 1. Significance: * < .1; ** < 0.05; *** < 0.01.

	Outcome: Indicator for first choice plot			
	Conditional Logit		Linear FE	
	Odds-ratio (SE)	Odds-ratio (SE)	β(SE)	β(SE)
Area X $\hat{E}(D\|X, I = 1)$	$1.08***$		$0.018***$	
	(0.016)		(0.0045)	
Area X $\hat{E}(D\|X, I = 0)$		$1.09***$		$0.022***$
		(0.019)		(0.0055)
Area X $(\hat{E}(D\|X, I = 1) - \hat{E}(D\|X, I = 0))$		$1.05*$		0.011
		(0.028)		(0.0087)
Area (hectares)	0.99	0.99	-0.000014	-0.0030
	(0.016)	(0.017)	(0.0056)	(0.0061)
Number of farmers	462	462	493	493
Observations	1254	1254	1311	1311

Table 4: Columns 1 and 2 present empirical estimates of Equations (16) and (17) and columns 3 and 4 are their linear analogs. Coefficients in the first two columns are given as odds-ratios. Standard errors are reported in parenthesis. The sample for these regressions includes only farmers who were in the full randomization group (Group A: Received insurance on half of plots at random). Significance: * < .1; ** < 0.05; *** < 0.01.

Summary Statistics by Treatment Group

A. Randomization of farmers

	At Randomization			Analysis sample		
	Mean		Difference	Mean		Difference
	In Insurance Group (A+B)	In Control Group (C)	(p-value)	In Insurance Group (A+B)	In Control Group (C)	(p-value)
Total enrolled area (hectares)	1.72	1.57	0.15*	1.71	1.57	0.14
			(0.09)			(0.14)
Number of enrolled plots	2.88	2.84	0.04	2.88	2.82	0.06
			(0.69)			(0.61)
Education (years)	10.20	10.45	−0.26	10.11	10.40	−0.29
			(0.41)			(0.39)
Age (years)	53.82	52.93	0.90	53.76	53.17	0.59
			(0.37)			(0.59)
Gender (1 = female)	0.17	0.16	0.00	0.16	0.15	0.01
			(0.91)			(0.68)
Observations*	607	233		532	184	
					p(F) =	0.71

B. Randomization of plots
(excludes plots not randomized (Group C and 1st choice plots of Group B))

	At Randomization			Analysis sample		
	Mean		Difference	Mean		Difference
	Insured	Control	(p-value)	Insured	Control	(p-value)
Is first choice plot (1 = yes)	0.35	0.35	0.00	0.37	0.36	0.01
			(0.99)			(0.82)
Area (hectares)	0.59	0.61	−0.02	0.60	0.61	−0.01
			(0.38)			(0.61)
Owns plot (1 = yes)	0.24	0.26	−0.02	0.21	0.22	−0.02
			(0.41)			(0.49)
Flooding index (unit SD)	0.05	0.01	0.03	0.03	0.00	0.02
			(0.52)			(0.67)
High Rat Risk (1 = yes)	0.19	0.21	−0.02	0.21	0.22	−0.01
			(0.28)			(0.57)
High Tungro Risk (1 = yes)	0.15	0.14	0.01	0.16	0.15	0.01
			(0.60)			(0.63)
High Wind Risk (1 = yes)	0.04	0.05	−0.00	0.05	0.05	−0.00
			(0.86)			(0.84)
Observations*	802	791		642	631	
					p(F) =	0.97

Table 5: The table shows summary statistics by treatment condition and tests for treatment balance. Observations are given for the full sample. Some rows are based on a smaller sample due to missing values.

Attrition

Farmer randomization group: (Insurance includes Groups A and B)	Season 1 Insurance N	%	Control N	%	Season 2 Insurance N	%	Control N	%	Season 3 Insurance N	%	Control N	%	All Seasons Insurance N	%	Control N	%
A. Farmer attrition																
At randomization	71		36		199		86		336		111		466		211	
Dropouts:																
Did not farm this season		-		-		5%		4%		2%		1%		3%		2%
Refused survey		-		-		5%		7%		3%		4%		4%		7%
Not found, died or fell ill		-		-		5%		11%		4%		6%		5%		9%
Total endline surveys	65	92%	27	75%	175	88%	68	79%	309	92%	99	89%	420	90%	176	83%
Endlined but no harvest data		0%		0%		6%		6%		1%		5%		4%		4%
Endlined with harvest data but no damage data		3%		0%		7%		2%		0%		0%		3%		1%
Farmers with any damage data (requires harvest data)	63	89%	27	75%	149	75%	61	71%	304	90%	94	85%	404	87%	166	79%
Plot randomization group:	Season 1 Insurance N	%	Control N	%	Season 2 Insurance N	%	Control N	%	Season 3 Insurance N	%	Control N	%	All Seasons Insurance N	%	Control N	%
B. Plot attrition. Excludes control farmers (Group C) and 1st choice plot of choice farmers (Group B)																
Plots of farmers with any damage data	84		75		225		232		436		439		745		746	
Missing harvest data		11%		12%		9%		11%		7%		8%		8%		9%
Missing damage data		5%		3%		9%		9%		1%		1%		4%		4%
Ineligible plots (respondent found to be worker, not farmer)		0%		0%		0%		0%		0%		0%		0%		0%
Plots with damage data	71	85%	64	85%	183	81%	186	80%	401	92%	399	91%	655	88%	649	87%

Table 6: The table shows attrition in the experiments. Panel A shows attrition of farmers. Panel B shows, conditional on the farmer completing a follow-up survey with any data on damages (i.e. data on damages for at least one plot), the plot attrition. Only farmers in the two treatment groups (A and B) are included in panel B, and plots that got insurance for sure (1st choice plots in group B) are also excluded. Panel A is therefore relevant for comparisons using the cross-farmer randomization and panel B for comparisons using the randomization of insurance to plots within the same farm.

Graphs

Figure 1: Identification of Adverse Selection and Moral Hazard

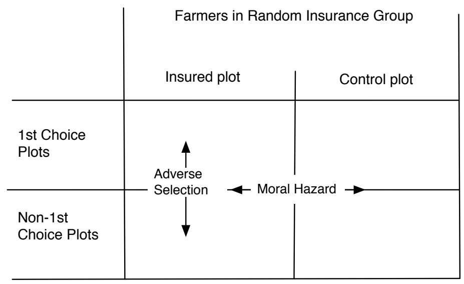
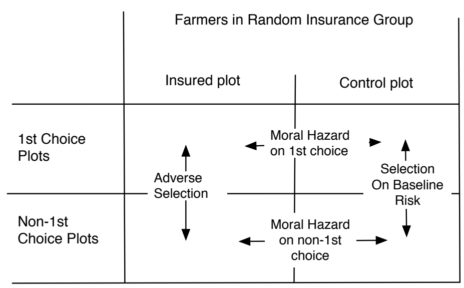


Figure 2: Decomposition of Adverse Selection



A Data Appendix

A.1 Measures of Risk Aversion

I constructed a measure of risk aversion based on two likert-scale responses (collected at baseline) of how willing a farmers is to take risk on the farm and in life. The questions asked, on a scale of 1 (avoid risk) to 7 (fully prepared to take risk) how willing the farmer is to take risks on the farm and in life in general. I label farmers who choose 6 or 7 on both questions as less risk averse (coded as 0) and other farmers as risk averse (coded as 1). A detailed module of questions on preferences was administered at baseline which will be used to incorporate a richer set of preferences in later drafts.

A.1.1 Damage Measures

To obtain a survey measure of the share of harvest lost to the various causes we asked each farmer how much they lost on each plot to each cause. Because most farmers do not have a good grasp of percentages we asked about damages in terms of number of sacks of palay (unmilled rice) lost. I calculate the percentage loss as[10]:

$$\text{Loss (\%) 'Type of damage'} = \frac{\text{Losses due to 'Type of damage'}}{\text{Harvest + Harvest loss from 'All Causes'}}$$

where losses and harvest are measured by value (in pesos) and 'Type of damage' is one of the three aggregate measures. The three aggregate measures are 'All Cause'(a combination of all categories), 'Typhoons and floods' and 'Pests and crop diseases' (A combination of damages from rats, insects, tungro and other crop diseases).

[10] These measures follow naturally from the model. Given that AG is the harvest and AGD is the loss, a natural measure for D is $D = \frac{AGD}{AG} = \frac{AGD}{AGD+AG(1-D)} = \frac{\text{Total loss}}{\text{Total loss+Harvest}}$.