Identifying Information Asymmetries in Insurance:

Theory and Evidence on Crop Insurance in the Philippines

Snaebjorn Gunnsteinsson*

University of Maryland

March 31, 2014

Abstract

Crop insurance markets have the potential to improve the welfare of agricultural households but they have largely failed to develop. Information asymmetries may explain this market failure, but very little evidence is available on the relative importance or underlying mechanisms of the different dimensions of this asymmetric information. I separately identify and quantify information asymmetries in crop insurance in the Philippines. I designed and implemented a field experiment that first elicited farmers' choices of which plots they would prefer to insure, and then randomly allocated insurance to farmers and plots, generating across- and within-farm variation in which plots were insured. I model jointly the farmers' choice of a plot for insurance and their resource allocation across plots. In addition to adverse selection and moral hazard, the model implies the possible presence of an interaction between the two: that is, selection on plot-specific benefits from moral hazard behavior. I find strong evidence for adverse selection and moral hazard, and I find evidence that farmers select both on the inherent riskiness of plots and on the plot-specific benefits from moral hazard behavior. Further, possibly due to moral hazard, I find that farmers use less fertilizer on insured plots, suggesting that subsidies for this type of insurance may reduce aggregate investment.

*Department of Agricultural and Resource Economics; sgunnste@umd.edu. This paper is a revised version of Chapter 1 of my dissertation at the Department of Economics at Yale University. I would like to extend thanks and gratitude to my advisors, Christopher Udry and Dean Karlan, for invaluable advice and support throughout this research project. I am grateful for advice from David Atkin, Xavier Gine, Daniel Keniston, Mushfiq Mobarak, Michael Peters, 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 Treb Allen, Camilo Dominguez and Melanie Morten 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 Geneva Association and the Russell Sage Foundation.

1 Introduction

Agricultural incomes are highly volatile, in part due to natural hazards. In response, farming households shift to safer but lower-return production technologies and smooth consumption through borrowing, saving or other financial arrangements. Both strategies can be very costly and there is ample evidence of serious welfare consequences when households have limited ability to smooth consumption.[1] While crop insurance markets have the potential to protect households against some of the worst shocks, such markets have largely failed to develop.[2] Information asymmetries may explain this market failure, but empirical evidence is scarce.[3] More importantly, very little evidence is available on the relative importance or underlying mechanisms of the different dimensions of this asymmetric information. Although some progress has been made in insuring agricultural production risk that can be captured by aggregate variables (such as a rainfall index), this is not the most important part of production risk in many contexts. Therefore, a better understanding of the nature of this asymmetric information is fundamental for both policy making and better contract design.

In this paper, I combine data generated by a randomized field experiment with a model of farmers' insurance choice and resource allocation decisions to identify information asymmetries in crop insurance in the Philippines. The experiment and data collection strategy are explicitly designed to separate out the multiple dimensions of asymmetric information. I account for unobserved variation in preferences and constraints by collecting data on multiple plots for each farmer. By eliciting farmers' relative demand for insurance on plots in their portfolio and by inducing random variation in insurance coverage across a farmers' plots, I disentangle selection on risk characteristics of plots from moral hazard. Using this approach, I am able to 1) separately identify and quantify adverse selection and

[1] On farmers' production choices, see for example Rosenzweig and Binswanger (1993) and Dercon (1996). On limited consumption smoothing, see for example Fafchamps and Lund (2003); Rosenzweig and Binswanger (1993); Maccini and Yang (2009); and Kazianga and Udry (2006). For a larger review, see Morduch (1995).

[2] Private crop insurance markets for certain natural hazards have developed (e.g., for hail damage). Government subsidized programs also exist in many countries. In the US, annual subsidies for crop insurance are currently about 9 billion dollars (GAO, 2012).

[3] A very extensive literature analyzes the reasons for the absence or underperformance of financial markets in developing countries (see f.e., Hoff and Stiglitz (1990); Besley (1994); Conning and Udry (2007)). In particular, the seminal contributions of Stiglitz and Weiss (1981) and Rothschild and Stiglitz (1976) show how adverse selection can cause market failures in credit and insurance markets, respectively. In the case of crop insurance, previous research (mostly based on markets in the US and Canada) has identified adverse selection, moral hazard and spatial co-variability of risk as the main culprits for the failure of private markets and public schemes (L Hueth and Hartley Furtan, 1994; Miranda and Glauber, 1997; Just et al., 1999; Makki and Somwaru, 2001).

moral hazard, 2) determine that this adverse selection is caused by a deliberate selection on damage risk and not by a correlation between risk and preferences or constraints and 3) disentangle selection on the inherent riskiness of plots from selection on the plot-specific ability to engage in moral hazard.

Information asymmetries are very hard to identify empirically using data normally available to insurance companies and researchers. First, it is very hard to identify moral hazard without some exogenous shift in coverage. Second, since both preferences and risk type are (at least partly) unobserved, it is hard to identify to what degree selection is based on private information on risk type versus private information on preferences. This difference has crucial implications for the insurance provider and for market development. Selection on risk type leads to higher payouts and can cause the market to break down (Rothschild and Stiglitz, 1976), while selection on preferences is less likely to be a cause for higher payouts. In fact, in many markets (such as automobile insurance and life insurance), selection on risk preferences is likely to counteract selection on risk type (de Meza and Webb, 2001; Cutler et al., 2008). Third, it is very hard to identify any possible interactions between adverse selection and moral hazard. An example of this in our context would be when a farmer chooses to buy insurance on a plot that is far from her home, explicitly because, once the plot is insured, she can save a substantial amount of effort in preventing damages.

Although the experimental design is informed by the theory, for the purpose of exposition I will first discuss the context and the experimental design, and then illustrate the model. In the Philippines a government owned insurance company offers crop insurance for rice crops. This insurance covers crop losses due to specific natural hazards (such as typhoons, pests and crop diseases). Payments are based on an ex-post damage assessment by an agent of the insurance company. Since the insurance pays out based on the harvest losses on each particular plot, there is good reason to expect substantial asymmetric information. The experiment was based on three stages. In the first stage, I elicited farmers' choices of which plot, if they could choose only one, they would prefer to have insured. The farmers were told this plot would have a higher chance of receiving free insurance in a lottery. In the second stage, I randomly chose farmers to receive free insurance on a subset of their plots. In the third stage, I randomly selected which of their plots received insurance, but allowed their first choice plots to have a higher chance of receiving insurance coverage. This generated across- and within-farmer variation in which plots were insured and provided an incentive for truth-telling in the first stage. Finally, I combined the data generated through this process with geospatial data on the locations of plots and environmental characteristics, administrative data from the insurance company and comprehensive survey data.

The goal of this paper is to understand the behavior of farmers when faced with the incentives generated by a crop insurance contract of this type. The focus is on the degree and type of asymmetric information that leads to excess payouts by the insurance company. Since the insurance is provided for free, I do not study demand and therefore do not consider the partial or general equilibrium of the insurance market. Although studying insurance demand in this context would be worthwhile, there is no competitive equilibrium to study precisely because the market has failed to develop (except for the political economy equilibrium of government subsidized insurance). The fact that no equilibrium exists is not a limitation for this study but rather is the motivation.

I explicitly model behavior in the experiment and by using this model to understand the data I provide insights into the extent and type of private information in this context. Specifically, I model the joint determination of the plot choice decision and the farmers allocation of preventative effort across plots. I allow for heterogeneity in both the inherent riskiness of plots and in the plot-specific cost of effort. Farmers select plots taking into account their endogenous effort response to both plot characteristics and insurance. If the cost of effort is prohibitively large on all plots, then farmers select plots that are large and have high inherent riskiness. If the cost of effort is lower, allowing for a sizable effort response by the farmer, then farmers face a tradeoff between choosing plots that have high expected damages and those on which they can save a relatively large effort cost if insured. The model therefore implies that, in addition to classic moral hazard, two types of adverse selection may be present. First, selection on "baseline risk"; that is, selection on the expected damages on a plot, taking into account the endogenous effort response to plot characteristics but not the endogenous response to insurance. And second, "selection on moral hazard"; that is, selection on the plot-specific anticipated effort response to insurance. These effects are defined precisely in the model.[4]

In the first empirical section I use the experiment to separately estimate adverse selection in plot choice and classic moral hazard. I estimate moral hazard by comparing the damage experience on randomly insured and uninsured plots of the same farmer and estimate adverse selection by comparing damages on the farmers' first choice plot to damages on other plots of the same farmer. I find strong evidence for both. Farmers select plots that are prone to floods and crop diseases and this leads to about 20% higher

[4] Einav et al. (2013) also study this type of selection in the US health insurance market and coined the term "selection on moral hazard."

damages on first choice plots compared to the farmers' other plots. To investigate moral hazard, I separate the harvest losses into two components: loss due to typhoons and floods, and loss due to 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 flood.[5] I find evidence for moral hazard in the prevention of pests and crop diseases. Harvest loss due to these causes is about 24% higher on randomly insured plots compared to uninsured plots. In contrast, I find no evidence of moral hazard in the prevention of typhoon and flood damage, providing some confidence that the earlier estimate is not due to reporting bias.

In the second empirical section I investigate the impact of insurance on investment, as measured my fertilizer expenditures, and use the across-farm randomization to investigate whether insurance on one plot has implications for farming decisions on the farmers' other plots. I find that farmers use less fertilizer on insured plots. This is consistent with moral hazard, since under moral hazard insured plots are higher risk than uninsured plots, and provides further confidence that the observed moral hazard effect is indeed identifying moral hazard. This also implies that subsidies for this type of insurance may reduce aggregate investment. I do not find any evidence that insurance on one plot induces higher investment on the farmers' other plots. This evidence therefore does not support the presence of substantial wealth effects (from the reduced investment on insured plots) nor important background risk effects (that is, incentives for greater investment on uninsured plots through reduced background risk from insured plots).

In the third empirical section I develop an empirical strategy to disentangle selection on what I have termed "baseline risk" from "selection on moral hazard." The strategy uses plot characteristics collected at baseline, which predict about 37% of the observed adverse selection effect, to construct measures of predicted damages separately for randomly insured and uninsured plots. I then study whether selection is based on the predicted values for uninsured plots (i.e., baseline risk) or on the difference (i.e., selection on moral hazard). The difference is computed by subtracting predicted values on control plots from predicted values on insured plots and represents the predicted moral hazard based on baseline characteristics. I find that baseline risk is the key determinant of selection, but that farmers select on both dimensions.

This paper builds on previous experimental approaches to identify information asymmetries, including the RAND Health Insurance Experiment (Manning et al., 1987) and

[5] The insurance company makes the same distinction and offers an insurance package that only covers typhoons and floods as well as offering a comprehensive package the covers the full range of damages (all insurance coverage in this study was the comprehensive coverage).

Karlan and Zinman (2009)'s experiment on consumer credit in South Africa. As in both of these experiments I use random variation to identify moral hazard. However, by combining selection and randomization across production units within the same farm, and complimenting these with comprehensive data collection, I am able to overcome many of the limitations of the earlier work. First, I am able to isolate selection on plot risk characteristics from selection on farmer level characteristics (such as preferences or constraints). This is important as many papers that study adverse selection in insurance markets have found little evidence of adverse selection or have even found evidence of advantageous selection(de Meza and Webb, 2001).[6] Second, Karlan and Zinman (2009) note that selection may be based on unobserved ability to engage in moral hazard but they are unable to disentangle this from selection on risk type, as I do in this paper. This has been done for an insurance market in only one existing paper, a forthcoming paper byEinav et al. (2013). The present paper complements the work by Einav et al. by identifying this effect using experimental variation and data on multiple insurable units (i.e., plots) per decision maker in contrast to the difference-in-difference and structural estimation techniques employed by Einav et al. Third, I am able to identify a causal source of moral hazard heterogeneity through the random variation in distance between insured and uninsured plots induced by the experiment. Overall, I find that information asymmetries in crop insurance are pervasive and economically significant, operating through the channels of adverse selection and moral hazard, and through the interaction of the two. Finally, I experimentally identify the effect of insurance coverage on investment in the presence of moral hazard. I am not aware of a previous experimental identification of this effect and, as noted earlier, the findings imply that subsidies for this type of insurance may have unintended negative consequences for aggregate investment.[7]

The paper proceeds as follows. For expositional purposes, I will first describe the insurance contract and the experiment in Section 2. I will then present the model and derive empirical implications in Section 3. In Section 4 I discuss the implementation, describe the data and examine the integrity of the experiments. Next I present the three empirical sections. In Section 5 I separately estimate adverse selection and moral hazard, in Section 6 I investigate resource allocation over the farmers' portfolio of plots, and in

[6]There is a sizable literature confirming this possibility empirically with results largely diverging by insurance type. Health insurance and annuity markets tend to show adverse selection while the evidence points to advantageous selection in life and long-term care insurance (see Cutler et al. (2008) and references within, e.g., Cawley and Philipson (1999); Finkelstein and Poterba (2004); Finkelstein and McGarry (2006); and Fang et al. (2008)).

[7]This contrasts with recent work that finds that insurance leads to increased investment when not subject to moral hazard, e.g., Karlan et al. (2012).

Section 7 I disentangle selection on baseline risk from selection on moral hazard. Finally, I conclude in Section 8.

2 Insurance Contract and Experimental Design

2.1 Insurance Contract

Agricultural production in the Philippines is susceptible to not only the usual threats of floods, droughts, pests and crop diseases, but also the arrival of more than 15 typhoons or tropical cyclones per year. A government-owned company, the Philippines Crop Insurance Corporation, offers subsidized crop insurance to rice farmers throughout the Philippines. Take-up is low and no private market has developed.

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 and various insects) and crop diseases (e.g. tungro, a crop disease spread by insects).[8] Any particular damage event must cause at least 10% loss of harvest to be eligible for a claim. 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. 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%.

2.2 Experimental Design

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.

The study design for each season was the following:

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

[8]The insurance also covers rare events such as volcanic eruptions and earthquakes but excludes some minor pests such as birds and snails. 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.

would prefer to have insured. They were told that the plot they chose would have a higher chance of receiving free insurance in a lottery.

Step 2: Baseline survey (if not baselined in earlier seasons)

Step 3: 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 first-choice plot and a random half of remaining plots.

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

Step 4: Two follow up surveys, one after planting and another after harvest.

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.[9] Insurance was allocated to plots in Group A using 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 first-choice plots of farmers in Group B, their remaining plots were randomly allocated insurance using the same procedure as in Group A. A baseline was conducted between Steps 1 and 2. Two follow-up surveys were conducted in each season, one after planting and another after harvest.

The study was explicitly designed to separately estimate adverse selection and moral hazard. Figure 2 depicts the basic identification strategy. To identify adverse selection I compare the first-choice plot of the farmer to her other plots. I will test this both by comparing measures of predicted damages, actual damages and payouts. To identify moral hazard I compare insured and uninsured plots of farmers in the fully random group (Group A). In principle, the design allows me to identify moral hazard separately for first-choice plots and for other plots. I will discuss the identification of separate components of adverse selection (as depicted in the Figure 3) in the Sections 3.7 and 7.

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

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

3.1 Introduction and Summary of the Model

In this section, I develop a model describing optimal decisions of farmers in the experiment. In the model, farmers are faced with the possibility that they may lose part of each plot's harvest to a natural hazard. Farmers make two decisions. First they choose one plot to designate as their "first choice". Next they allocate preventative effort (to reduce crop loss from natural hazards) to each of their plots. I assume that plot characteristics and effort levels are unobserved by the insurance provider. This is consistent with the context: per-hectare prices only depend on the season and the geographic area; furthermore, no monitoring of farm practices (such as pesticide or insecticide use) takes place. The study area is fully contained in one pricing area, so all farmers face the same per-hectare prices. Of course, all insurance is free in the experiment, and the tradeoff that the farmer faces in selecting a plot for insurance is the opportunity cost of not insuring one of her other plots.

I consider two versions of the model for insurance choice. In the first, farmers are partially myopic such that they do not take into account their possible moral hazard response when choosing a plot for insurance. In the second, farmers are more sophisticated and fully take into account their anticipated endogenous effort response to insurance when making their insurance choice. In the first scenario, the insurance decision of the farmer is straightforward: she chooses the plot that maximizes the expected payout from the insurance company. Since the farmer could choose one plot, regardless of size,[10] she maximizes the payout by choosing the plot that maximizes the product of plot size and the expected share of harvest lost.

In the second version, the farmers' insurance choice takes into account the her anticipated effort response to the insurance coverage. In this case, farmers derive two types of benefits from insurance coverage on a specific plot: the payout in case of harvest loss and the ability to save some cost of effort. This implies that farmers may select not only on the inherent riskiness of plots but also on the ability to engage in moral hazard.

The key feature of the experiment's design is that the insurance choice is only probabilistic. The plot chosen may or may not get insurance and insurance is randomly

[10] Within the limit that only plots between .25 and 2.5 hectares were eligible to be included in the experiment.

allocated to plots (though the first-choice plots have a higher chance of being insured). Given this feature of the data, I start by modeling insurance choice and effort as a joint decision for the purpose of studying insurance choice. I then consider the insurance to be exogenously determined to study moral hazard and extend the model to consider together the farmers' effort and variable investment decisions. In the model I assume that, conditional on plot characteristics and effort, shocks are uncorrelated between plots of the same farmer and that the farmer maximizes a mean-variance utility. This implies that effort and investment decisions on plot j of farmer i are independent of whether plot j' of the same farmer is insured. These assumptions provide tractability, but of course shocks are not uncorrelated across plots. In the empirical section, this issue is addressed both through the design of the experiment (i.e., plot randomization) and through data collection (especially the collection of spatial coordinates of plots, allowing spatially corrected standard errors). However, even with independent shocks across plots the farmers input decisions on plot j are not independent of whether plot j' is insured for general utility functions. The design of the experiment, in particular the two-stage randomization procedure, allows us to test these implications of the model – that is, whether reducing production risk on plot j has implications for production decisions on plot j'. This can be thought of as a test of whether background risk influences investment decisions. [11]

3.2 Setup and Maximization Problem

Consider farmers, indexed by i, each farming a portfolio of N plots, indexed by j. I omit the farmer subscript when possible. Plot j produces a maximum output of 1 per hectare (I relax this assumption in Section 3.5) and is A_j hectares in size. Some of this output may be lost to natural hazards. The share of harvest lost, D_j, is a random variable and follows $U\left[0, \theta_j(1-e_j)\right]$ where $\theta_j \in (0,1]$ indexes the risk characteristics of the plot and $e_j \in [0,1]$ is the effort put forth to reduce damages. Let $\theta = (\theta_j)_{j=1}^N$ and $\mathbf{e} = (e_j)_{j=1}^N$. I assume that, conditional on θ and $\mathbf{e}$ (which determine the support of the distribution of

[11] Normally background risk is thought of as in a different domain that then risk being studied, such as considering risks to labor income as the background risk for investment decisions in the stock market (e.g., Heaton and Lucas (2000)). In this case I consider the investment risk on one production unit as the background risk for investment decisions on another production unit.

losses), the harvest losses are independent random draws across plots.[12] A plot may be insured, in which case the farmer receives a payout of LD_j per hectare, where $L < 1$ is the per hectare insurance coverage. I denote the indicator for insurance coverage with $\alpha_j \in \{0, 1\}$ and define $\alpha = (\alpha_j)_{j=1}^N$. This is now a choice variable, with the restriction that $\sum_{j=1}^N \alpha_j = 1$, representing the choice that the farmer faces in choosing one plot as their first choice (later on I replace α with $\alpha^{assigned}$ to represent the exogenously assigned insurance allocation).[13] Therefore, profits of farmers are stochastic and given by

$$\Pi(\alpha, \mathbf{e}) = \sum_j \{A_j((1 - D_j) + \alpha_j L D_j)\} - C(\mathbf{e}) \tag{1}$$

where C is the cost-of-effort function. I assume the farmers' preferences can be represented by a mean-variance utility over total future profits:

$$E[U(\Pi)] = E[\Pi] - \rho Var(\Pi) \tag{2}$$

It is convenient to note that, given the uniform distribution for D, we have:

$$\begin{aligned} E[D_j|e_j] &= \tfrac{1}{2}\theta_j(1 - e_j) \\ Var[D_j|e_j] &= \tfrac{1}{12}\theta_j^2(1 - e_j)^2 \end{aligned}$$

This implies that

$$\begin{aligned} E[\Pi|(\alpha, \mathbf{e})] &= \textstyle\sum_{j=1}^N A_j \left(1 - \frac{1}{2}(1 - \alpha_j L)\theta_j(1 - e_j)\right) - C(\mathbf{e}) \\ Var[\Pi|(\alpha, \mathbf{e})] &= \tfrac{1}{12} \textstyle\sum_{j=1}^N A_j^2(1 - \alpha_j L)^2\theta_j^2(1 - e_j)^2 \end{aligned}$$

[12] For some farmers, with two or more plots close to each other, this assumption is clearly unrealistic. For others, with more spread out plots, it is more reasonable. For tractability of the model I use this assumption for now. In the empirical section I take advantage of geospatial data on plots to investigate how the extent of information asymmetry is affected by how spread out farms are; however, even with this geospatial data is is hard to separately identify spatial correlation in inherent riskiness from spatial correlation in shocks.

[13] The farmers choice is only probabilistic but I assume that the farmer chooses a plot in the same way as she would do if insurance were to be assigned with probability 1.

The farmers maximization problem is to choose one plot as her preferred plot for insurance and then choose effort level on each plot conditional on its insurance coverage:

$$\begin{aligned}\max_{\alpha,\mathbf{e}} \sum_{j=1}^{N} \Bigg[A_j \left(1 - \frac{1}{2}(1-\alpha_j L)\theta_j(1-e_j)\right) \\ - \rho\frac{1}{12}A_j^2(1-\alpha_j L)^2\theta_j^2(1-e_j)^2 \Bigg] \\ - C(\mathbf{e}) \end{aligned} \tag{3}$$

subject to $\sum_{j=1}^{N} \alpha_j = 1$, $\alpha_j \in \{0,1\}$ and $e_j \in [0,1]$. The core of the research design is that the experiment allows us to break this maximization problem into two parts, identifying the two choice variables separately – that is, identifying insurance choice based on inherent plot characteristics and anticipated effort allocation, and then separately (from selection) identifying effort and investment responses to insurance. In the next section I first analyze the optimal effort allocation as a function of insurance coverage. This both serves as an analysis of optimal behavior once the insurance allocation in the experiment is know and serves as input into the first stage choice problem.

3.3 Optimal Effort

Optimal effort given insurance coverage is:

$$\begin{aligned}\widehat{\mathbf{e}}(\alpha) = \arg\max_{\mathbf{e}} \sum_{j=1}^{N} [\; A_j \left(1 - \frac{1}{2}(1-\alpha_j L)\theta_j(1-e_j)\right) \\ - \rho\frac{1}{12}\sum_{j=1}^{N} A_j^2(1-\alpha_j L)^2\theta_j^2(1-e_j)^2 \;] - C(\mathbf{e})\end{aligned}$$

The first order condition for effort is:

$$\begin{aligned}\frac{\partial C}{\partial e_j} &= A_j\left[\frac{(1-\alpha_j L)\theta_j}{2} - \rho A_j(1-\alpha_j L)^2\theta_j^2\frac{-2(1-e_j)}{12}\right] \\ &= A_j(1-\alpha_j L)\theta_j\left[\frac{1}{2} + \rho A_j(1-\alpha_j L)\theta_j\frac{1-e_j}{6}\right] \\ &= W_j\left[1 + \rho W_j\frac{2(1-e_j)}{3}\right]\end{aligned}$$

where $W_j = A_j w_j$, and I define $w_j = \frac{1}{2}(1-\alpha_j L)\theta_j$ as the per-hectare harvest at risk on plot j (i.e., the expected monetary loss if no effort is applied). This amount will be prominent in the calculations below and I use it as a convenient shorthand.[14]

I assume that the per-hectare cost-of-effort function is separable and of the form $c(e_j) = \psi_j e_j$ where ψ_j represents the plot-specific cost of effort. Here the ψ's may, for example, represent distances from home. In this case, the ψ's are not characteristics of the plot, per se, but from the perspective of the farmer they can be treated as plot characteristics.[15,16] Total effort costs are:

$$C(\mathbf{e}) = \sum_{j=1}^{N} A_j \psi_j e_j. \tag{4}$$

Given this setup, effort of farmer i on plot j is a function of the farmers' risk aversion (ρ, omitting the farmer subscript i) and plot-level attributes: the insurance coverage (α_j), the inherent riskiness of the plot (θ_j), the parameter of the cost function (ψ_j) and area (A_j). I show in Section A.1 that optimal effort is given by:

$$\hat{e}_j(\alpha_j, \theta_j, \psi_j, A_j, \rho) = \begin{cases} 0 & \text{if } \psi_j \geq w_j + \frac{2}{3}\rho A_j w_j^2 \\ 1 - \frac{3}{2}\frac{\psi_j - w_j}{\rho A_j w_j^2} & \text{if } w_j < \psi_j < w_j + \frac{2}{3}\rho A_j w_j^2 \\ 1 & \text{if } \psi_j \leq w_j \end{cases} \tag{5}$$

where, as defined earlier, $w_j = \frac{1}{2}(1 - \alpha_j L)\theta_j$. Figure 1 illustrates optimal effort as a function of the plot-specific cost of effort (ψ) for insured and uninsured plots. Effort is lower on insured plots in the range where (1) cost of effort is large enough so that effort is less than 1 if the plot is insured but (2) small enough so that effort is positive if the plot is uninsured – that is, if $\psi \in (w^1, \hat{w}^0))$ in Figure 1. The model therefore implies moral hazard over this range.

In this section we have assumed the α_j's are given. These findings therefore describe both (1) the maximization problem the farmer faces after she learns of the insurance allocation in the experiment and (2) the problem that the farmer expects to face during the cropping season as she is taking her insurance choice decision. In the experiment, after

[14]The second order condition for a local maximum is $-\rho\frac{W_j^2}{6} < 0$, which is always satisfied.

[15]As the ψ's are per-hectare costs, they may also represent scale economies when the farmer tills two or more adjacent (or close by) plots. About 35% of the plots in the sample are adjacent to at least one other plot of the same farmer.

[16]Although the model considers for now only one type of damage, in reality farmers face multiple natural hazards, each associated with a different plot-specific cost of preventative effort. I discuss this further in Section 5.6. The primary distinction in the paper will be between cost of effort in preventing typhoons and floods versus pests and crop diseases. A priori, one might expect ψ to be very high for all plots in the case of typhoons and floods, but lower (and possibly variable across plots) for pests and crop diseases.

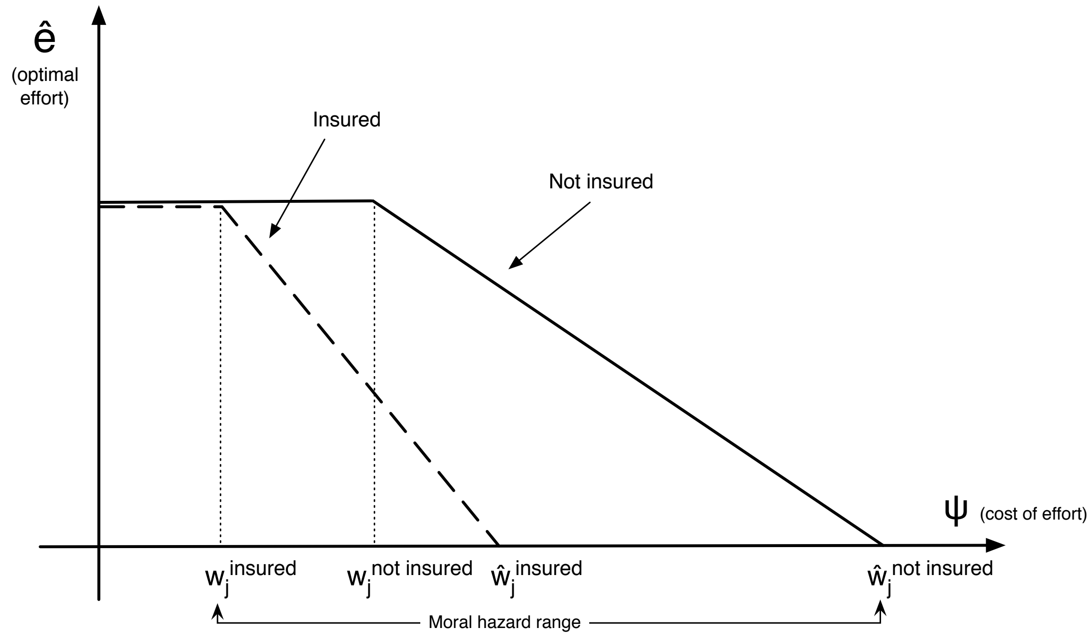

Figure 1: Optimal effort, $\hat{e}_j$, as a function of the plot-specific cost of effort for insured and uninsured plots. Here, w_j^{insured} and $w_j^{\text{not insured}}$ denote w_j for insured and uninsured plots, respectively. Therefore, $w_j^{\text{insured}} = \frac{1}{2}\theta_j(1-L)$ and $w_j^{\text{not insured}} = \frac{1}{2}\theta_j$. The upper boundaries are defined by $\hat{w}_j^{\text{insured}} = w_j^{\text{insured}} + \frac{2}{3}\rho A_j(w_j^{\text{insured}})^2$ and $\hat{w}_j^{\text{not insured}} = w_j^{\text{not insured}} + \frac{2}{3}\rho A_j(w_j^{\text{not insured}})^2$. The policy functions imply that, for plot j, effort is lower when the plot is insured if $w_j^{\text{insured}} < \psi < \hat{w}_j^{\text{not insured}}$, and otherwise equal.

the farmer is informed of the insurance allocation to her plots, her problem simplifies. Instead of the farmers problem in 3 she now maximizes only over $\mathbf{e}$ and α is no longer a choice variable but is replaced by $\alpha^{assigned}$, which is exogenous and is not limited to adding up to one over her plots. I discuss the empirical implications for analyzing moral hazard in Section 3.7. First, I use this characterization of optimal effort allocation to derive the optimal insurance choice.

3.4 Insurance Choice

In this section I characterize the optimal insurance choice of farmers in the experiment. I consider and contrast two different levels of sophistication on part of the farmer. First, in Section 3.4.1, I consider the insurance choice of a farmer that is partially myopic in that she does not take into account her anticipated effort response to insurance and instead chooses insurance assuming she will farm the plot in the same manner as she would normally do without insurance.[17] Next, in Section 3.4.2, I contrast this insurance choice with that of a more sophisticated farmer who anticipates her effort response to insurance and takes an optimal decision with this in mind.

Before considering these two cases I first derive the value function for insurance choice. Given the optimal effort $\hat{e}_j(\alpha_j, \theta_j, \psi_j, A_j, \rho)$, the utility output on plot j is:

$$\begin{aligned} u_j(\alpha_j, \theta_j, \psi_j, A_j, \rho) = A_j &- \frac{1}{2} A_j w_j (1 - \hat{e}(\alpha_j, \psi_j, \rho, A_j, w_j)) \\ &- \frac{\rho}{12} A_j^2 w_j^2 (1 - \hat{e}(\alpha_j, \psi_j, \rho, A_j, w_j))^2 \\ &- A_j \psi_j \hat{e}(\alpha_j, \psi_j, \rho, A_j, w_j) \end{aligned} \tag{6}$$

Therefore, the correct value function for insurance choice (used by the fully sophisticated farmer) is

$$V(\alpha) = \sum_{j=1}^{N} u_j(\alpha_j, \theta_j, \psi_j, A_j, \rho). \tag{7}$$

Therefore, the sophisticated farmers' maximization problem when choosing the plot to designate as first choice is $\max_\alpha V(\alpha)$ subject to $\alpha_j \in \{0, 1\}$ and $\sum_{j=1}^{N} \alpha_j = 1$. In contrast, the less sophisticated (partially myopic) farmer bases her insurance choice decision on plot specific utility that does not take into account the effect of insurance on effort – that is, she assumes an effort function $\hat{e}^{myopic}(\theta_j, \psi_j, A_j, \rho) = \hat{e}(0, \theta_j, \psi_j, A_j, \rho)$ and an associated utility (u_j^{myopic}) and value function (V^{myopic}), obtained by substituting $\hat{e}^{myopic}$ for $\hat{e}$ in equation 6 and substituting u_j^{myopic} for u_j in equation 7. Since the second case is simpler I reverse the discussion and consider first the optimal insurance choice of a farmer constrained by myopia of this type.

3.4.1 Insurance choice of a (partially) myopic farmer

I now characterize the optimal choice of a farmer that is constrained by myopia in the sense that she does not take into account her anticipated effort response to insurance. In the below I will simply the notation by using $\hat{e}_j^0$ and $\hat{e}_j^1$ in place of $\hat{e}(0, \theta_j, \psi_j, A_j, \rho)$ and $\hat{e}(1, \theta_j, \psi_j, A_j, \rho)$. The perceived utility gain from insurance on plot j for a farmer

[17] She does, on the other hand, anticipate how her effort level is influenced by plot characteristics. E.g., if a plot is of high risk of floods but this is easily prevented by low-cost effort she anticipates this and may prefer insurance on a plot that has a medium risk of damage but for which no low-cost preventative solution is available.

constrained by myopia of this type is:

$$\begin{aligned}\Delta u_j^{\text{myopic}} &= u_j^{\text{myopic}}(1,\theta_j,\psi_j,A_j,\rho) - u_j^{\text{myopic}}(0,\theta_j,\psi_j,A_j,\rho) \\ &= \frac{1}{2}A_j\theta_j L(1-\hat{e}_j^0) - \frac{\rho}{12}A_j^2\theta_j^2((1-L)^2-1)\end{aligned} \tag{8}$$

The first term is the expected payout on plot j if the farmer applies effort as she would without insurance. The second term (which is positive) is the expected gain in utility from the reduction in the variance of profits that the insurance provides in this case.

In this case, the only utility gain from insurance is the payout received and this is maximized by choosing the plot that has the highest expected damages – that is, the highest product of area and expected damages per hectare.[18]

3.4.2 Insurance choice of a fully sophisticated farmer

Consider now the full model where the farmer anticipates her effort response to insurance. In this case, the farmer balances the gains from the expected payout against the gains from saved effort. The payout, P, is given by

$$P = LA_jE\left[D_j|\theta_j,\psi_j,\alpha_j=1\right] = \frac{1}{2}LA_j\theta_j(1-\hat{e}(\theta_j,\psi_j,\alpha_j=1)) \tag{9}$$

and the saved effort, ΔC, is given by

$$\Delta C = \psi_jA_j(\hat{e}(\theta_j,\psi_j,\alpha_j=0)-\hat{e}(\theta_j,\psi_j,\alpha_j=1)) \tag{10}$$

It may be intuitive to suppose that this additional degree of sophistication would lead greater adverse selection. However, since the full cost of saved effort is the product of the unit cost (ψ) and effort it is in general ambiguous whether this leads to higher or lower effort, and it is therefore generally ambiguous whether this type of selection leads to higher or lower payouts for the insurance provider. The key conclusion is that this tradeoff between payouts and saved effort exists, leading to the possibility of selection on the anticipated moral hazard response. I will empirically investigate whether this is the

[18] Note that the expectations of damages are conditional on expected efforts that in turn are based on all aspects of the model other than insurance status. In particular, the farmer anticipates any effect that plot characteristics may have on her effort. In other words, the farmer is assumed to anticipate farming the plot as she would normally do (without insurance). For example, if a particular plot is highly susceptible to a particular pest but she knows that this can easily be prevented through low-cost actions on her part, then she may prefer insurance on other plots, knowing that this will not be an issue for this particular plot.

case in Section 7.

3.5 Extending the Model with Productive Investment

Farmers expend effort and resources not only to prevent damages but also to increase yield through other means. I now extend the model to allow for the use of a productive investment input, such as fertilizer. In this section α (insurance) is not a choice variable. This is because the goal of this subsection is to understand how effort and investment interact in response to exogenous insurance provision and to empirically test these implications using the randomized experiment. Output on a plot when no damages occur are now assumed to be $G(f_j)$ instead of 1, where G is increasing and concave and f_j is the amount of investment input applied to plot j. I assume the price of the investment input is p_f so that the cost function for investment is $F(\mathbf{f}) = p_f \sum_j^N f_j$. The farmers' profit function is now defined as

$$\Pi(e, f) = \sum_j^N \{G(f_j)A_j(1 - D_j) + \alpha_j L D_j A_j\} - C(\mathbf{e}) - F(\mathbf{f}) \tag{11}$$

Using the properties of the exponential utility as before the farmers maximization problem becomes:

$$\begin{aligned}\max_{\mathbf{e},\mathbf{f}} = &\sum_j^N A_j \left[G(f_j) - \frac{1}{2}(G(f_j) - \alpha_j L)\theta_j(1 - e_j) - \rho \frac{1}{12} A_j^2 (G(f_j) - \alpha_j L)^2 \theta_j^2 (1 - e_j)^2 \right] \\ &- C(\mathbf{e}) - F(\mathbf{f}) \end{aligned} \tag{12}$$

That is, the farmer jointly determines the level of effort and investment across her portfolio of plots. Consider the FOC with respect to investment:

$$p_f = G'(f_j)\{A_j \left[1 - \frac{1}{2}\theta(1 - e)\right] - \frac{1}{6}\rho A_j^2 (G(f_j) - \alpha_j L)\theta_j^2 (1 - e_j)^2\} \tag{13}$$

The way the insurance contract is structured, insurance coverage doesn't impact the marginal expected return to the investment input except through changes in effort provision. Insurance coverage does reduce the variance of returns and can therefore impact investment directly (i.e., not through incentives for less effort provision). Given farmers risk aversion, this direct effect provides incentives for more investment. Insurance coverage also impact investment through the joint determination of effort and investment. This channel will operate in the opposite direction. Under the most likely scenarios (see more precisely below), insurance provides incentives for using less effort, which leads to higher chance of damage that in turn provides incentives for less investment. To illustrate this, taking the derivative of 13 with respect to effort, we have:

$$\frac{\partial^2 G}{\partial f^2}\frac{\partial f}{\partial e} = \frac{-p_f\{\overbrace{\frac{1}{2}\theta_j A_j}^{>0} - \frac{1}{12}\rho\theta_j^2 A_j^2\left[\overbrace{2\frac{\partial G}{\partial f}\frac{\partial f}{\partial e}(1-e_j)^2}^{\text{assumed small}} - \overbrace{2(G(f)-\alpha_j L)2(1-e)}^{>0}\right]\}}{(A_j\left[1-\frac{1}{2}\theta(1-e)\right] - \frac{1}{6}\rho A_j^2(G(f_j)-\alpha_j L)\theta_j^2(1-e_j)^2)^2} < 0 \tag{14}$$

To obtain the final inequality I assume the first term in the bracket is small relative to the second term. This seems reasonable since the first term is the product of two marginal effects (on G and f) whereas the second term includes the level of $G(f) - \alpha_j L$ and insurance coverage is far from complete. Given that G is assumed concave, we have $\frac{\partial f}{\partial e} > 0$.

Overall the preceding discussing shows that in this model insurance coverage provides both incentives for increased and decreased investment, and the outcome depends on which term dominates. Empirically this implies that applying this model to an observation of reduced investment on insured versus uninsured plots of each farmer would imply that the second term dominates and that moral hazard in effort is causing reduced investment.

3.6 Portfolio Investment

So far I have assumed a mean-variance utility function for the farmer. It provides tractability but is restrictive. In particular, it implies that getting insurance on one plot does not influence the farmers decisions on her other plots. It therefore predicts that effort and investment are identical on uninsured plots of insured farmers compared to plots of uninsured (control) farmers. This may be unrealistic in important ways, e.g. if the farmer puts more weight on preventing outcomes below a certain threshold. This could be the case if the farmer is close to subsistence level or if, as is common in the study area, she takes out an informal production loan that has high penalties for late payment. In these cases insurance on one plot would provide incentives for increasing investment inputs (e.g. fertilizer) on her uninsured plots. Since those plots are not insured, there is no counteracting moral hazard incentive to reduce investment. In contrast to the mean-variance case, these preferences would predict higher effort and fertilizer use on uninsured plots of insured farmers compared to plots of uninsured (control) farmers. We test for these differing predictions in Section 6.

3.7 Empirical Implications

The model has implications for the empirical specifications designed to separate adverse selection from moral hazard and those designed to understand whether selection is based on the ability to engage in moral hazard. The first implication is that that farmers may select on the plot-specific heterogeneity in cost of effort, possibly inducing a "selection on moral hazard" effect. A second set of implications is caused by the fact that farmers have an incentive to select larger plots. This implies that when studying selection based on ex-post damages I must control for the area to prevent me from misattributing selection on area as selection on risk characteristics (This problem arises if area and the plot specific inherent risk or cost of effort are correlated). This also suggests that if I study adverse selection using predicted damages (as I do in Section 7), then the correct specification includes a term for area by itself and a term that interacts area and predicted damages. The correct test for adverse selection in this case is to test for a positive interaction term.

4 Implementation, Experimental Integrity and Description of the Data

4.1 Implementation

Under the direction of the author, Innovations for Poverty Action (IPA)[19] implemented the experiments and data collection from the spring of 2010 through mid-2012. IPA staff invited farmers fulfilling certain eligibility criteria (described below) that were located within the Tigman Hinagyanan Inarihan Regional Irrigation System and surrounding communal irrigation systems in the Camarines Sur region of Bicol province in the Philippines to participate in a research project on crop insurance. The implementation started in the 2010 wet season (July - September) with a small pilot experiment with 52 farmers, followed by full scale experiments and data collection over the following three cropping seasons. The sample was gradually expanded, from 106 farmers with 291 plots in the dry season (December - April) of 2010-11, to 285 farmers with 806 plots in the wet season of 2011 and 447 farmers with 1302 plots in the dry season of 2011-12. After each round, farmers were invited to participate in subsequent rounds. Therefore, for a subsample of farmers and plots, we have up to 3 rounds of panel data.

[19]Innovations for Poverty Action (IPA) is a US-based non-profit organization that specializes in conducting impact evaluations that aim to inform programs and policies to reduce poverty and improve well-being, primarily in developing countries. See more at www.poverty-action.org.

Eligibility Criteria and Recruiting Rice is grown in this area by owner-operators or through a variety of informal contractual arrangements between tillers and owners. This necessitated a clear definition of "farmer". We defined a person to be the farmer of an agricultural plot only if they were both (1) the principal decision maker for farming decisions, and (2) the bearer of a majority of the production risk. Because of the design of the experiment (involving within-farm plot randomization) we focused only on farmers with two or more agricultural plots. We attempted to recruit as many farmers as possible in the sample area that satisfied the eligibly criteria of farming two eligible plots within the geographic area of the study. Plots in the study area were eligible if they were irrigated, traditionally rice growing plots, and were of size between 0.25 and 2.5 hectares.[20] We recruited farmers principally through door-to-door canvassing and, to a lesser extent, at regular farmer meetings. Based on administrative data from the local office of the National Irrigation Administration, we were able to recruit about X% of farmers in this area that fulfill the eligibility requirements.[21] Overall, the study covered about Y% of agricultural land in the key target areas.

4.2 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.

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:[22]

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

[20] 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. We chose to have an upper bound both because we did not want a large amount of our funds for insurance premiums to be used for a small set of plots and because this seemed more acceptable to the community based on conversations during the pilot phase.

[21] This data is somewhat outdated but allows us to have a rough sense of the sample relative to all farmers in the area.

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

where losses and harvest are measured by value (in pesos) and ”Type of damage” is one of 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).

4.3 Integrity of the Experiments

In Table 12, 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 which leads us to have some damage data on 90% and 85% of treatment and control groups in the last season. 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 11 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 first-choice plot. Overall, the evidence suggests that the integrity of the plot randomization was maintained.

5 Separate Estimation of Adverse Selection and Moral Hazard

In this section I empirically estimate both adverse selection and moral hazard using data on harvest losses (self reported) and payouts (from administrative data). As can be seen in Table 10, harvest losses due to the various natural hazards are large in this context. All-cause harvest loss is estimated at 31% in the wet season and 21% in the dry. I separate all-cause harvest losses into two components: loss due to typhoons and floods, and due loss to 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.[23] In the wet seasons the losses from typhoons and floods dominate, accounting for about 80%, while in the dry season losses are more equally distributed between typhoons and floods (about 60%) and pests and crop disease (the remaining 40%).

5.1 Empirical Specification

I estimate two separate equations. All regressions will be based on the sample of farmers that got insurance by random. In the first equation, I include indicators for plot choice, insurance status and their interaction. In the second equation I constrain the interaction term to be zero. The first equation is the following:

$$D_{ij} = \beta_0 + \beta_1\alpha_{ij} + \beta_2 C_{ij} + \beta_3 C_{ij} \cdot \alpha_{ij} + \beta_4 A_{ij} + \lambda_i + \epsilon_{ij} \tag{15}$$

Here α is an indicator for insurance coverage and C is an indicator that is one if the plot was chosen as the farmers' first-choice plot. The reason for the additional area control is a possible correlation between area and plot risk characteristics. If A is positively correlated with θ_{ij} (inherent riskiness) or ψ_{ij} (cost of effort), the additional control for area guards me against the mistake of attributing selection on area to selection on other characteristics. The second specification (see below) captures better the moral hazard effect so I defer the discussion on moral hazard.[24] The model illustrates the two conceptually different types of selection, selection on "baseline risk" and "selection on moral hazard." The former is based on the farmers' desire to choose a plot that has the maximum expected payout ignoring any effort response. In the framework of the model, farmers select plots that maximize the product of area (A_j), inherent riskiness (θ) and a linearly decreasing function in the effort that is applied when the plot is not covered by insurance: $A_j\theta_j(1-\hat{e}(\theta_j, \psi_j, \alpha_j = 0))$.

[23] The insurance company offers 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.

[24] In this specification, β_1 identifies the moral hazard effect for non-first-choice plots and $\beta_1 + \beta_3$ identifies the effect for first-choice plots. The second equation allows for a direct estimation of the average moral hazard effect.

This selection effect is identified by β_2. To illustrate this in the framework of the model, consider two plots of a farmer, the first-choice plot with characteristics (A_1, θ_1, ψ_1) and a non-first-choice plot with characteristics $(A_{-1}, \theta_{-1}, \psi_{-1})$ (for farmers with more than two plots, think of this latter plot as a random draw of her non-first-choice plots). Then β_2 identifies the selection on "baseline risk", that is selection on the difference in damages between first-choice and non-first-choice plots assuming no effort response to insurance (but effort responds to plot characteristics). That is, it identifies selection on[25]

$$
\begin{aligned}
&E\left[D_1|(A_1,\theta_1,\psi_1),\alpha_1=0\right]-E\left[D_{-1}|(A_{-1},\theta_{-1},\psi_{-1}),\alpha_{-1}=0\right]\\
&=\underbrace{\frac{1}{2}\theta_1(1-\hat{e}(\theta_1,\psi_1,\alpha_1=0))}_{\text{Baseline risk on first-choice}}-\underbrace{\frac{1}{2}\theta_{-1}(1-\hat{e}(\theta_{-1},\psi_{-1},\alpha_{-1}=0))}_{\text{Baseline risk on non-first-choice}}
\end{aligned}
\tag{16}
$$

The second selection effect is based on the farmers' desire to save effort. The β_3 coefficient captures this "selection on moral hazard" effect. This effect is precisely captured by the difference between the moral hazard effect on her first-choice plot and the moral hazard effect on her non-first-choice plot. This effect is the difference (c - d) in effects (c) and (d) in Figure 3. In the framework of the model, β_3 identifies selection on:

$$
\begin{aligned}
&\left(E\left[D_1|(A_1,\theta_1,\psi_1),\alpha_1=1\right]-E\left[D_1|(A_1,\theta_1,\psi_1),\alpha_1=0\right]\right)\\
&-\left(E\left[D_{-1}|(A_{-1},\theta_{-1},\psi_{-1}),\alpha_{-1}=1\right]-E\left[D_{-1}|(A_{-1},\theta_{-1},\psi_{-1}),\alpha_{-1}=0\right]\right)\\
=&\overbrace{\frac{1}{2}\theta_1\left[(1-\hat{e}(\theta_1,\psi_1,\alpha_1=1))-(1-\hat{e}(\theta_1,\psi_1,\alpha_1=0))\right]}^{\text{Moral hazard effect on first-choice plots}}\\
&-\underbrace{\frac{1}{2}\theta_{-1}\left[(1-\hat{e}(\theta_{-1},\psi_{-1},\alpha_{-1}=1))-(1-\hat{e}(\theta_{-1},\psi_{-1},\alpha_{-1}=0))\right]}_{\text{Moral hazard effect on the non-first-choice plot}}\\
=&\frac{1}{2}\theta_1\overbrace{\left[\hat{e}(\theta_1,\psi_1,\alpha_1=0)-\hat{e}(\theta_1,\psi_1,\alpha_1=1)\right]}^{\text{Reduction in effort on first-choice plots when insured}}\\
&-\frac{1}{2}\theta_{-1}\underbrace{\left[\hat{e}(\theta_{-1},\psi_{-1},\alpha_{-1}=0)-\hat{e}(\theta_{-1},\psi_{-1},\alpha_{-1}=1)\right]}_{\text{Reduction in effort on non-first-choice plots when insured}}
\end{aligned}
\tag{17}
$$

This effect may be non-zero if the farmer chooses a plot in part based on the plot-specific cost of effort (ψ) due to her desire to save on the costs associated with effort. Her benefit of saved effort is captured in the model for a plot j by the term $\psi_j\,[\hat{e}(\theta_j,\psi_j,\alpha_j=1)$ -

[25]Remember, in the model, damages (D) are distributed as $U\left[0,\theta(1-e)\right]$ where θ is the inherent riskiness of the plot and e is effort. This implies that $E\left[D|\theta,e\right]=\frac{1}{2}\theta(1-e)$.

$\hat{e}(\theta_j, \psi_j, \alpha_j = 0)]$, which is positive. However, since this term is the product of the cost and the difference in effort, it is in general ambiguous whether selection on this term is associated with higher or lower damages.

The second equation, where I add the constraint that β_3 (the interaction term) is zero, is:

$$D_{ij} = \beta_0^{'} + \beta_1^{'}\alpha_{ij} + \beta_2^{'}C_{ij} + \beta_3^{'}A_{ij} + \lambda_i + \epsilon_{ij}. \tag{18}$$

This specification illustrates more clearly the moral hazard effect and the separation between adverse selection and moral hazard. In this specification $\beta_1^{'}$ identifies the average moral hazard effect. Here $\beta_2^{'}$ captures the adverse selection effect, except that it does not fully capture any possible selection on moral hazard. More precisely, since insurance was allocated at random to precisely half of the farmers' plots (on average), $\beta_2^{'} = \beta_2 + \frac{1}{2}\beta_3$. That is, $\beta_2^{'}$ captures the full selection on "baseline risk" but only half of the "selection on moral hazard" effect.

5.2 Results

In Table 1, I estimate Equations 15 and 18. I first discuss Columns 1, 3, and 5, which are based on Equation 18 (no interaction term). I find strong evidence for adverse selection in overall damages as well as separately for both typhoons and floods, and also for pests and crop diseases. Damages are estimated to be 4.4 percentage points (95% CI: 1.8 - 7.0) 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 7. 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 their first-choice plot and at least one other plot insured (a total of 144 farmers). I find evidence for moral hazard in preventing pests and diseases but perhaps unsurprisingly no evidence for moral hazard in preventing typhoons and floods. Damages from pests and crop diseases are estimated to be 1.5 percentage points higher on insured plots compared to randomly uninsured plots of the same farmer of a baseline of 8.4%. This translates into about 24% increase in damages.

In Columns 2, 4 and 6, I estimate Equation 15 (allowing for the interaction term). In these regressions I lose some precision when estimating individual coefficients but the main effects are still statistically significant.[26] In addition, although the interaction coefficient is identified, the statistical power to test for this effect is low. If we take the estimates at face value, they imply that damages on first-choice plots are 4.4 percentage points higher on first-choice plots due to selection on "baseline risk," and a further 0.5 percentage points higher due to "selection on moral hazard." In Section 7 I construct a test, based on predicted damages, that is able to distinguish these two effects.

5.3 Robustness of Moral Hazard Estimates

In Table 2, I investigate the robustness of the moral hazard estimation. For comparison, in Column 1 of Table 2, I repeat the estimation in Column 5 in Table 1 (excluding the indicator for choice). In Column 2, I add controls for baseline risk characteristics (see summary statistics in Table 11). This limits the sample to plots of farmers in Seasons 2 and 3 (as these were not collected in Season 1) for which I have data on each of these plot characteristics. In Columns 3-4 I repeat the analysis in Columns 1-2, but restrict the sample to plots that lie between the 2.5th and 97.5th percentile on the full harvest (harvest plus damages) per hectare distribution. The reason for this restriction is to exclude plots for which rice cultivation was not seriously attempted (below 2.5th percentile) and plots that very likely have erroneous data (above 97.5th percentile). Those above the 97.5th percentile (particularly those well above) are likely the result of misunderstanding between surveyor and farmer in which the farmer mistakenly reports the combined harvest or damages on multiple plots when asked about the outcomes for a particular plot. The moral hazard estimate is consistent across these specifications and remains statistically significant in all cases.

5.4 Heterogeneity by Season

The wet and dry seasons are considerably different. Damages are substantially higher in the wet season due to the high incidence of typhoons. Columns 1 and 2 in Table 4 show that the moral hazard effect found in Table 1 is entirely based on the dry seasons. This is in part because the high damages due to typhoons hide damages due to pests and crop diseases (I observe a moral hazard effect in the wet season if I treat typhoon damage as exogenous and adjust for this) but it nonetheless motivates analyzing separately the

[26] An F-test that tests whether both the indicator for first-choice and the interaction are significant yields a one-sided p-value of 0.0015 (suggesting adverse selection). An F-test on whether both the treatment effect and the interaction effect are zero for Column 6 yields a one-sided p-value of 0.043 (suggesting moral hazard).

mechanisms of moral hazard and investment under moral hazard by type of season. I do this in Sections 5.5 and 6.1.

5.5 Mechanisms of Moral Hazard

In this subsection I investigate the possible underlying mechanisms for the observed moral hazard effect in preventing pests and crop diseases. There are several options, including: 1) Lower expenditure on pesticides and insecticides; 2) earlier planting on insured plots; 3) lower use of pest, insect or diseases resistant seeds on insured plots; or 4) less preventative effort (through the re-allocation of labor).

Lower expenditure on pesticides and insecticides is perhaps the most direct possible mechanism. Evidence on this mechanism is presented in columns 3-6 in Table 4. I do not find a reduction in expenditure on pesticides and insecticides. I similarly find no difference between the number of insured and uninsured plots on which any such expenses are reported (this includes expenses for the chemicals themselves and for labor for their application). This may be because these expenses are quite low in the sample (insecticide and pesticide expenses are less than 1% of total expenses for control plots).

In Columns 7-8 I investigate whether farmers plant seeds that are less resistant to pests, insects and crop diseases on insured plots. This data is based on self-reports that are available only in the last season. The farmer is coded as using pest, insect or disease resistant seeds if they reported selecting the seeds because they are resistant to pests, insects or the various crop diseases (e.g., tungro and bacterial leaf blight).[27] I estimate that insured plots are about 10-15% less likely to be planted with resistant seeds but I can not reject the hypothesis that insured and uninsured plots are planted with the same seeds.

Planting earlier, relative to neighbors, can lead to higher incidence of pests and insects but harvesting early can also allow the farmer to fetch better prices for her output (and an earlier payout). This suggests a possible moral hazard mechanism based on planting time, where insurance induces earlier planting and higher damages by influencing the trade-off between risk and expected profits. I find in Columns 9 and 10 that this mechanisms seems to be operating in this environment. I estimate that farmers plant 1.2 days earlier on insured plots. If we consider only plots that are islands (not adjacent to another plot of the same farmer) then I estimate that farmers plant 2.4 days earlier on insured plots.

[27] The farmer was able to report many reasons for using these particular seeds. This included, in addition to resistance to pests, insects and crop diseases: fast growing, drought resistant, less prone to floods, high yielding, high quality (high market price), saline resistant, can withstand strong winds.

This evidence is only suggestive because: 1) it is not clear to what degree each day of earlier planting leads to higher damage risk; and 2) the proper analysis would take into account planting times of neighboring plots. This spatial economic analysis is possible with the available data and may be a fruitful avenue for future work.

5.6 Heterogeneity in Asymmetric Information By Geographic Distance

It is natural to assume that the inherent riskiness of plots is positively spatially correlated. That is, that θ_j and $\theta_{j'}$ are positively correlated for two plots that are adjacent or close. The cost of effort (in the model ψ_j) may also be spatially correlated. Consider two farmers in the experiment, one with two plots adjacent (or very close) to each other and the other with two plots some distance apart. Our expectation for the magnitude of adverse selection and moral hazard might be smaller in both cases for the first farmer. The adverse selection estimate may be low simply because the two plots have similar inherent risk (i.e., similar θ) and at the same time we may detect little moral hazard because of scale economies (promoting the farmer to treat both plots similarly) or due to the farmers' concern about damage on her uninsured plot through physical externalities.

In Table 3 I investigate whether this may be the case. Column 1 reproduces the main adverse selection effect from Column 1 in Table 1 for comparison. In Column 2 I interact the indicator for first-choice plot with the log of the average distance between plots in the farmers' portfolio. All log-variables in the table are standardized to zero mean and unit variance. I estimate that the interaction term is positive, consistent with positive spatial correlation in inherent riskiness. The first-choice plots of farmers with a portfolio one standard deviation above the mean on log(Average Distance Between Plots) is estimated to have 7.3 percentage points higher damages than the farmers' other plots, compared to the 4.8 percentage point difference between first-choice plots and other plots for farmers at the mean of the log(Average Distance Between Plots) distribution. The difference, 2.6 (95% CI: 0.1 - 5.0), is statistically significant. This measure is very potentially correlated with other farmer characteristics and therefore should be interpreted with caution. In Column 3, I add interaction terms between first-choice and the respondents' age and quartiles of the respondents' years of education. The estimated interaction is consistent with Column 2 (now estimated to be 2.1 percentage points). Although further controls could beneficially be employed, this is suggestive evidence that farmers with portfolios of plots spread out over a larger geographic area are able to take greater advantage of their private information about the inherent riskiness of plots in their portfolio when selecting a plot in the experiment.

In Columns 4 - 8 I investigate the heterogeneity in the moral hazard estimate across farmers with more or less geographically spread out farms and across the two rice cropping seasons (the wet and the dry season). Compared to the baseline estimate in Column 4, as noted earlier the estimates in Column 5 show that the moral hazard effect is almost entirely driven by the dry seasons (I have data from two dry seasons and one wet season). In Columns 6-8 I estimate regression equations of the form:

$$
\begin{aligned}
&\text{Loss (\%) Due to Pests and Diseases} = \alpha_0 + \alpha_1 \text{Insurance}_{ij} \\
&+ \alpha_2 \text{Insurance}_{ij} \text{ X } \log(\text{Average Distance Between Plots})_i \\
&+ \alpha_3 \text{Insurance}_{ij} \text{ X } \log(\text{Avg. Dist. Btw. Insured and Control Plots} \\
&/ \text{ Avg. Dist. Btw. Plots})_i \\
&+ \underbrace{\lambda_i}_{\text{Farmer Fixed Effect}} + \epsilon_{ij}
\end{aligned}
\tag{19}
$$

As in the case for selection, log(Average Distance Between Plots) is likely correlated with various unobserved farmer characteristics. However, the random allocation of insurance to plots generates random variation in the relative average distance between insured and control plots compared to the average distance for all plots. The α_3 coefficient is therefore identified and by the argument above (e.g. scale economies, physical externalities or correlated plot characteristics) I expect $\alpha_3 > 0$. For the full sample I find that this is true, but the effect is not statistically significant. However, in the dry season, which drives the overall moral hazard result, I find that this effect is both positive and statistically significant. This suggests, therefore, that farmers are able to engage in a greater degree of moral hazard if insured and control plots are some distance apart compared to when they are adjacent (or close), suggesting a role for scale economies, physical externalities or other related mechanisms.

6 Moral Hazard and Investment: Testing Alternative Models

6.1 Investment under Moral Hazard

The insurance contract does not provide incentives for investment because the payout is based on the share of harvest lost rather than the absolute loss. In fact, because of the inherent incentives for moral hazard, insured plots are more likely to be damaged than other plots and given that the insurance is partial this implies that farmers have an incentive to use less (non-preventative) variable investment (e.g., fertilizer) on insured plots. I investigate whether this is true in Table 5. I find that farmers spend less on fertilizer on insured plots, observing a reduction of 250 pesos of a baseline of 5300 pesos, or about 5% (this effect is significant at the 0.05 level using a one-sided test). I find that this effect is entirely driven by the data from the dry seasons where farmers spend 440 pesos less for fertilizer on insured plots, compared to 5200 pesos used on uninsured plots, amounting to about 9% reduction in fertilizer application (this test is significant with p = 0.002). The observed reduced investment in the dry season coincides with the moral hazard effect that is also driven by the dry seasons. This is further evidence to suggest that the moral hazard effect observed is a true effect and not the result of reporting bias and suggests that subsidies for this type of insurance may result in lower aggregate investment.

6.2 Testing for independence in inputs across plots

As discussed in Subsection 3.6 the assumptions of our baseline model (namely the mean-variance utility) imply that optimal effort and investment on a farmers' plot is unaffected by receiving insurance on another plot in the farmers' portfolio. In Table 6 I test whether this prediction can be rejected. In the top panel of the table I estimate

$$Y_{ij} = \beta_0 + \beta_1 T_{ij} + \beta_2 TG_j + \epsilon_{ij} \tag{20}$$

where Y is either damages or fertilizer investment, T_{ij} is insurance on plot j of farmer i and TG_j indicates whether the farmer is in the main treatment group (where insurance was fully randomly allocated to plots). The farmers that got insurance on their first choice plot by design were excluded from the estimation and I cluster the standard errors within the strata used for farm-level randomization. The coefficient estimate for β_2 is an estimate for the difference between uninsured plots of treatment farmers and plots of control farmers.

Damages I estimate that total damages are 1.07 percentage point higher on uninsured plots of farmers in the treatment group compared to plots of control farmers but, with a standard error of 2.3, this effect is indistinguishable from zero. When I separate the damages by typhoons/floods and pests/diseases the effects are also indistinguishable from zero, though they are larger and operate in opposite directions. I find a negative coefficient estimate for β_2 for pests and crop diseases, suggesting a possible shift in preventative effort from insured to uninsured plots. However, because of the high variability in damages across farmers I do not have enough statistical power to determine whether this is a true effect.

Investment I find a negative β_2 coefficient on the amount spent on fertilizer per hectare, equal to a drop of 150 pesos (with standard error of 310 pesos). This estimate is indistinguishable from zero but I can reject a 10% increase in investment at a 5% significance level. The data does therefore not support a model in which lower background risk due to insurance on the farmers' other plots leads to higher investment and suggests that the assumptions underlying the model in Section 3 that lead to optimal farming decisions being independent across plots may be reasonable approximation of actual behavior.

7 Selection on Moral Hazard

In this section, I test for overall adverse selection and separately for selection on "baseline risk" and for "selection on moral hazard." The key idea for the decomposition is presented in Figure 3. 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 first-choice plots to insured other plots of the same farmer (that is, effect (a) in Figure 3). This I can disentangle into two effects: 1) selection on baseline risk by comparing predicted damages on uninsured first-choice plots to predicted damages on uninsured other plots (effect (b) in Figure 3) 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 (that is, effect (c) minus effect (d) on Figure 3).

In Section 7.1 I show that plot characteristics observed by me through a baseline survey predict 37% of harvest losses. These characteristics are not observed by the insurance company and I assume in this section that they are a good proxy for the full information set that the farmer has about each plot. Some of these characteristics (or their proxies) might in principle be observable by the insurance company. However, at the moment the company does not condition prices on any characteristics, likely because they are too expensive to collect. Although this is not the result of a competitive market this is suggestive evidence that these characteristics are at least expensive to collect compared to the premiums that could be sustained in this market. In Sections 7.2 and 7.3 I map the model develop in Section 3 into estimable equations based on the above approach. I have in mind an estimation based on the conditional logit but these equations may be estimated with a linear model, and I will show estimates using both approaches. In Section 7.5 I test for adverse selection using this approach and test for the presence of the two separate components.

7.1 Baseline Characteristics Determine 37% of Selection

In Table 7 I estimate equations of the form

$$D_{ij} = \beta_0 + \beta_1 C_{ij} + \beta_2 X_{ij} + \beta_3 X_{ij} 1(\text{dry season}) + \lambda_i + \eta_{ij}. \tag{21}$$

I only use seasons 2 and 3 for this estimation since the relevant baseline characteristics were not collected in the first 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, strong wind, 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 variables that signify that the plot is high risk for 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.[28]

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 use these variables to

[28] The question on high wind showed little variability and is not used.

construct a measure of predicted damages that I can use to decompose the selection into the two conceptually distinct components.

7.2 Adverse Selection Based on Baseline Characterisics

In Section 3.4, I found that if cost of effort is very high and the farmer therefore exerts no effort in any scenario, then the farmer chooses the plot that maximizes expected payouts. That is, given that expected damages according to the model when no effort is applied are $\frac{1}{2}A_j\theta_j$, she chooses the plot that has the highest $A_j\theta_j$. In this section I test for adverse selection by comparing this model to the null hypothesis of no adverse selection, where instead farmers simply choose the largest plot. Based on the per-plot utility output derived in the model (Equation 6), the utility of insurance on plot j is:

$$v_j^* = u_j(\alpha_j = 1) - u_j(\alpha_j = 0) = cA_j\theta_j \tag{22}$$

where $c = \frac{1}{2}L$ is constant. Let $\bar{\theta} = \frac{1}{N}\sum_{j=1}^{N}\theta_j$. Then we can decompose this utility into:

$$v_j^* = cA_j\bar{\theta} + cA_j(\theta_j - \bar{\theta}) \tag{23}$$

Now since $E\left[D_j|\theta_j\right] = \frac{1}{2}A_j\theta_j$ (effort is zero), I can empirically compute the utility (up to a scaling factor) by:

$$\hat{u}_j = A_j\hat{E}\left[D_j|\theta_j^{obs}\right] \tag{24}$$

where θ_j^{obs} is the portion of risk that is observable to me based on the baseline characteristics.

The empirical analog of equation 23, where I use a conditional logit with the choice conditioned to the portfolio of each farmer, is of the form:

$$\Lambda(C_{ij}) = \alpha_0 + \alpha_1 A_{ij}\hat{E}\left[D|X, I = 1\right] + \alpha_2 A_{ij}\hat{E}\left[D|I = 1\right] + \epsilon_{ij} \tag{25}$$

where $C_{ij} = 1$ if farmer i chose plot j as her first choice. To test the model I include a term for area (multiplied by the overall damage rate for ease of interpretation) since, if no adverse selection is present, farmers are predicted to choose their largest plot. Here $\alpha_1 > 0$ provides a test for adverse selection. I also estimate an analogous linear model for comparison and this shows qualitatively the same results.

7.3 Decomposition of Selection on Baseline Characteristics

I assume now that 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. Now, again based on Equation 6, the utility of insurance coverage on plot j can in this case be written as:

$$\begin{aligned} v_j^{**} = u_j(\alpha_j = 1) - u_j(\alpha_j = 0) = &\overbrace{\frac{1}{2}A_j\theta_j(1-\hat{e}_j^0)L - \frac{\rho}{12}\left[(1-L)^2 - 1\right] A_j^2\theta_j^2(1-\hat{e}_j^0)^2}^{\text{Utility from coverage of inherent risk}} \\ + \underbrace{\frac{1}{2}A_j\theta_j(\hat{e}_j^0 - \hat{e}_j^1)L}_{\substack{\text{Utility of coverage}\\ \text{for moral hazard}}} + &\underbrace{\frac{\rho}{12}A_j^2\theta_j^2(1-L)^2\left[(1-\hat{e}_j^0)^2 - (1-\hat{e}_j^1)^2\right]}_{\substack{\text{Utility loss due to higher}\\ \text{variance through lower effort}}} + \underbrace{A_j\psi_j(\hat{e}_j^0 - \hat{e}_j^1)}_{\substack{\text{Utility gain from}\\ \text{saved effort}}} \end{aligned} \tag{26}$$

I can estimate the first and third terms in this utility from data and use this to test for the presence of selection on the ability to engage in moral hazard. That is, if I define $v_b = \frac{1}{2}A_j\theta_j(1-\hat{e}_j^0)L$ and $v_m = \frac{1}{2}A_j\theta_j(\hat{e}_j^0 - \hat{e}_j^1)L$ then the empirical analog of these expressions are:

$$\hat{v}_b = A_{ij}\hat{E}\left[D|X, I = 0\right] \tag{27}$$

and:

$$\hat{v}_m = A_{ij}(\hat{E}\left[D|X, I = 1\right] - \hat{E}\left[D|X, I = 0\right]). \tag{28}$$

To flexibly test the model, in addition to a term for area, I also add terms for v_b and v_m and estimate a conditional logit (or an analogous linear model) of the form:

$$\begin{aligned} \Lambda(C_{ij}) &= \alpha_0 + \alpha_1\hat{v}_b + \alpha_2\hat{v}_m + \alpha_3 A_{ij} + \alpha_4\hat{v}_b + \alpha_5\hat{v}_m + \epsilon_{ij} \\ &= \alpha_0 + \alpha_1 A_{ij}\hat{E}\left[D|X, I = 0\right] \\ &\quad + \alpha_2 A_{ij}(\hat{E}\left[D|X, I = 1\right] - \hat{E}\left[D|X, I = 0\right]) \\ &\quad + \alpha_3 A_{ij}\hat{E}\left[D|I = 1\right] + \alpha_4\hat{E}\left[D|X, I = 0\right] \\ &\quad + \alpha_5(\hat{E}\left[D|X, I = 1\right] - \hat{E}\left[D|X, I = 0\right]) + \epsilon_{ij}. \end{aligned} \tag{29}$$

Now, since I include the term for area, $\alpha_1 > 0$ provides a test for selection based on what could be called baseline risk, that is, on $\frac{1}{2}\theta_j(1-\hat{e}_j^0) = E\left[D|I = 0\right]$. The last term in the Equation 26 is positive if and only if $\alpha_2 > 0$. Therefore, given that $\psi > 0$, $\alpha_2 > 0$ provides a test for selection based on the plot-specific utility of saved effort. I estimate 25 and 29 in Section 7.5 (along with linear models for comparison).

7.4 Predicted Damages Based on Baseline Characteristics

To empirically estimate 25 and 29 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). In Table 8 I estimate models of the form:

$$D_{ij} = \beta_0 + \beta_1 X_{ij} + \beta_2 X_{ij} 1(\text{dry season}) + \eta_{ij} \tag{30}$$

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

7.5 Estimated Selection Effects

Using these predicted damages I can now empirically estimate equations 25 and 29. Table 9 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 first stage (farmer randomization) treatment group or differential reporting across (farmer randomization) treatment and control groups from affecting estimates. In Column 1, I estimate equation 25 and find strong evidence for adverse selection. A one percentage point increase in predicted damages increases the odds of a plot being chosen by 12%.

In Column 2, I estimate equation 29. 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 13%. 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 10% greater odds of choosing a particular plot. As noted earlier, this implies that farmers select not only on the baseline risk of plots (i.e. $\theta(1-\hat{e}_0)$ in the model) but also on the cost of effort since a positive α_3 coefficient in Equation 29 implies that the farmer chooses a plot with high potential of saved effort, i.e., high $A_j \psi_j (\hat{e}_j^0 - \hat{e}_j^1)$.

8 Conclusions

I designed and implemented a randomized controlled trial of crop insurance in the Philippines to identify and quantify the different dimensions of asymmetric information in this type of insurance. Using a model that jointly considers the farmers choice of a plot for insurance and their resource allocation decisions across plots, I illustrate three conceptually different information asymmetries that are relevant in this context: classic adverse selection (on the risk profile of the plot), classic moral hazard (on preventative effort), and the interaction of the two – that is, selection on anticipated moral hazard.

I find that all three are present in this context: 1) The plot that is selected by the farmer as her first-choice for insurance coverage has 20% higher damages than other plots in her portfolio, suggesting the presence of adverse selection; 2) I find evidence of moral hazard in the prevention of pests and crop diseases – damages from these causes are 24% higher on randomly insured than uninsured plots – but unsurprisingly, I find no evidence of moral hazard in damages from typhoons and floods; and 3) I find that the majority of selection is based on the inherent riskiness of plots but that farmers also select on plot-specific anticipated moral hazard.

I find that, due to moral hazard, farmers use less fertilizer on insured plots. This highlights a problem with the contract structure when the coverage includes damages subject to moral hazard (i.e., pests and crop diseases) and suggests that possible disincentives for investment should be taken into account in the cost-benefit analysis of subsidies for insurance for pests and crop diseases. I don't find any evidence to support that investment is shifted to uninsured plots, suggesting that the reduction in background risk associated with insurance on the farmers' other plots does not induce greater investment on uninsured plots.

By considering these information asymmetries together I am able to illustrate the different channels through which asymmetric information contributes to excess damages and to understand their relative importance. In addition, by identifying empirically the presence of selection on anticipated moral hazard, I highlight the possibility of this issue, which may have important implications in other markets (e.g., automobile and property insurance, and credit markets).

References

Besley, T, "How do market failures justify interventions in rural credit markets?," *The World Bank Research Observer*, January 1994, *9* (1), 27–47.

Cawley, John and Tomas Philipson, "An Empirical Examination of Information Barriers to Trade in Insurance," *The American Economic Review*, September 1999, *89* (4), 827–846.

Conning, Jonathan and Christopher Udry, *Handbook of Agricultural Economics*, Vol. 3 of *Handbook of Agricultural Economics*, Elsevier, 2007.

Cutler, D, Amy Finkelstein, and K McGarry, "Preference heterogeneity and insurance markets: Explaining a puzzle of insurance," *American Economic Review*, January 2008.

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

Dercon, S, "Risk, crop choice, and savings: Evidence from Tanzania," *Economic Development and Cultural Change*, January 1996, *44* (3), 485–513.

Einav, Liran, Amy Finkelstein, Stephen P Ryan, Paul Schrimpf, and Mark R Cullen, "Selection on Moral Hazard in Health Insurance," *The American Economic Review*, 2013, *103* (1), 178–219.

Fafchamps, Marcel and Susan Lund, "Risk-sharing networks in rural Philippines," *Journal of Development Economics*, 2003, *71* (2), 261–287.

Fang, Hanming, Michael P Keane, and Dan Silverman, "Sources of Advantageous Selection: Evidence from the Medigap Insurance Market," *Journal of Political Economy*, April 2008, *116* (2), 303–350.

Finkelstein, Amy and James Poterba, "Adverse selection in insurance markets: Policyholder evidence from the UK annuity market," *Journal of Political Economy*, 2004, *112* (1), 183–208.

_ and Kathleen McGarry, "Multiple Dimensions of Private Information: Evidence from the Long-Term Care Insurance Market," *The American Economic Review*, September 2006, *96* (4), 938–958.

GAO, *Crop Insurance: Savings Would Result from Program Changes and Greater Use of Data Mining*, United States Government Accountability Office, March 2012.

Heaton, John and Deborah Lucas, "Portfolio choice in the presence of background risk," *The Economic Journal*, 2000, *110* (460), 1–26.

Hoff, Karla and Joseph Stiglitz, "Introduction: Imperfect Information and Rural Credit Markets: Puzzles and Policy Perspectives," *The World Bank Economic Review*, September 1990, *4* (3), 235–250.

Hueth, Darrell L and W Hartley Furtan, *Economics of Agricultural Crop Insurance: Theory and Evidence*, Springer, January 1994.

Just, Richard E, Linda Calvin, and John Quiggin, "Adverse selection in crop insurance: Actuarial and asymmetric information incentives," *American Journal of Agricultural Economics*, 1999, *81* (4), 834–849.

Karlan, D S, R Osei, I Osei-Akoto, and C Udry, "Agricultural Decisions after Relaxing Credit and Risk Constraints," *Center for Economic Policy Research, Discussion Paper 9173*, 2012.

Karlan, Dean and Jonathan Zinman, "Observing unobservables: Identifying information asymmetries with a consumer credit field experiment," *Econometrica*, 2009, *77* (6), 1993–2008.

Kazianga, Harounan and Christopher Udry, "Consumption smoothing? Livestock, insurance and drought in rural Burkina Faso," *Journal of Development Economics*, 2006, *79* (2), 413–446.

Maccini, Sharon and Dean Yang, "Under the Weather: Health, Schooling, and Economic Consequences of Early-Life Rainfall," *The American Economic Review*, June 2009, *99* (3), 1006–1026.

Makki, Shiva S and Agapi Somwaru, "Evidence of adverse selection in crop insurance markets," *Journal of Risk and Insurance*, 2001, pp. 685–708.

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.

Miranda, Mario and Joseph Glauber, “Systemic Risk, Reinsurance, and the Failure of Crop Insurance Markets,” *American Journal of Agricultural Economics*, February 1997, *79* (1), 206–215.

Morduch, Jonathan, “Income Smoothing and Consumption Smoothing,” *The Journal of Economic Perspectives*, July 1995, *9* (3), 103–114.

Rosenzweig, Mark and Hans Binswanger, “Wealth, Weather Risk and the Composition and Profitability of Agricultural Investments,” *The Economic Journal*, January 1993, *103* (416), 56–78.

Rothschild, Michael and Joseph Stiglitz, “Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information,” *The Quarterly Journal of Economics*, November 1976, *90* (4), 629–649.

Stiglitz, Joseph and Andrew Weiss, “Credit Rationing in Markets with Imperfect Information,” *The American Economic Review*, June 1981, *71* (3), 393–410.

Tables

Table 1: Empirical Estimates of Adverse Selection and Moral Hazard

	Harvest loss (%) due to:						
	All causes		Typhoons and floods		Pests and diseases		Payout ($) per hectare
First-choice	4.4 * **	4.0*	2.6 * *	2.4	1.7*	1.6	9.6*
	(1.3)	(2.4)	(1.1)	(2.0)	(0.9)	(1.7)	(5.6)
Insurance	1.3	1.0	−0.2	−0.4	1.5*	1.4	
	(1.2)	(1.9)	(1.0)	(1.6)	(0.8)	(1.3)	
Area (hectares)	2.5	2.5	0.6	0.7	1.8	1.8	−4.5
	(2.2)	(2.2)	(1.8)	(1.8)	(1.5)	(1.5)	(7.5)
txXfirstchoice		0.8		0.5		0.3	
		(4.0)		(3.3)		(2.8)	
Farmer-season FE	Yes		Yes		Yes		Yes
Sample:	Plots of farmers in the fully random group						Insured plots
Mean of non-first choice plots	23.6	23.6	15.7	15.7	8.0	8.0	14.3
Mean of non-insured plots	24.6	24.6	17.0	17.0	7.6	7.6	
Num FE's	474	474	474	474	474	474	510
Observations	1213	1213	1213	1213	1213	1213	720

This table shows separate estimation adverse selection and moral hazard using estimation equations 15 and 18. 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. OLS standard errors are reported in parentheses. The identification of the selection coefficient in Column 4 is based on the 144 farmers who have their first choice plot insured as well as at least one other plot.
Significance: * < 0.1; ** < 0.05; *** < 0.01.

Table 2: Robustness of Moral Hazard Estimates

Outcome Variable:	**Harvest losses (%) due to Pests and Crop Diseases**			
	Full Sample		Restricted Sample	
Insurance	2.0 **	1.8 **	1.9 **	1.6*
	(0.9)	(0.9)	(0.9)	(0.9)
Flooding index		1.0		0.8
		(0.6)		(0.6)
High risk from wind		0.6		0.4
		(2.8)		(2.8)
High risk of rats		1.0		1.8
		(1.6)		(1.6)
High risk of tungro		1.7		0.7
		(1.9)		(1.9)
Area (hectares)	2.0	1.9	1.5	1.4
	(1.5)	(1.6)	(1.5)	(1.6)
Sample:	Plots of farmers in fully random group		Plots of farmers in fully random group excluding top and bottom 2.5% of full harvest per hectare distribution	
Mean of control plots of insured farmers	8.1	8.1	7.6	7.6
Num FE's	462	455	454	446
Observations	1163	1093	1106	1040

The table explores the robustness of the moral hazard finding. 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. Full harvest per hectare is calculated as $\frac{\text{Value of loss 'All causes'+Value of harvest}}{\text{Area}}$, see Figure 4.

Table 3: Heterogeneity of Information Asymmetries by Geographic Distance

Loss (%) due to:	All Causes			Pests and Diseases				
Sample Restriction:							Wet Season	Dry Season
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Base Selection	4.7 * **	4.8 * **	5.7 * *					
	(1.4)	(1.4)	(2.2)					
First-choice X log(Average Distance Between Plots)		2.5 * *	2.2*					
		(1.3)	(1.3)					
Insurance				1.9 * *	0.2	2.0 * *	0.5	2.8 * *
				(0.9)	(1.6)	(0.9)	(1.0)	(1.2)
Insurance X Dry Season					2.4			
					(1.9)			
Insurance X log(Average Distance Between Plots)						1.1	−1.2	2.0
						(0.9)	(1.0)	(1.2)
Insurance X log(Av. Dist. Btw. Insured and Control Plots / Av. Dist. Btw. Plots)						1.4	−1.0	2.3*
						(0.9)	(1.2)	(1.2)
Area controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
First-choice X Education and age controls	No	No	Yes	No	No	No	No	No
Mean of non-first-choice plots	23.7	23.7	23.7					
Mean of control plots				7.8	7.8	7.8	4.6	9.2
Num FE's	475	413	401	475	475	412	125	287
R-squared	0.68	0.70	0.70	0.65	0.65	0.67	0.67	0.67
Observations	1217	1061	1019	1222	1222	1064	323	741

The average distance between a farmers' plots is computed using the centroids of the plots. In column (3) I add interactions between the first-choice plot indicator and quartiles of the respondents years of education. All log variables are standardized to zero mean and unit standard deviation.

Table 4: Mechanisms of Moral Hazard

	Harvest loss (%) due to PCD		Expenses for chemicals		Any expenses for chemicals		Used pest or disease resistant seeds		Transplanting day	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: All Seasons										
Insurance	1.9**	2.1**	9.3	17.7	1.8*	1.8	-0.02*	-0.02*	-0.8	-2.4***
	(0.9)	(1.2)	(17.4)	(24.3)	(1.2)	(1.6)	(0.010)	(0.01)	(0.8)	(1.0)
Mean for uninsured plots	7.84	7.86	324.6	332.1	48.4	52.6	0.11	0.099	184.1	184.4
Num FE's	474	384	304	242	468	372	300	244	462	367
Observations	1222	817	815	527	1290	825	830	540	1235	801
	(1)	(2)	(3)	(4)	(5)	(6)			(7)	(8)
Panel B: Wet Season										
Insurance	0.2	-0.2	3.4	5.4	0.7	0.5			-0.6	-2.0**
	(1.1)	(1.5)	(33.6)	(55.1)	(1.9)	(3.0)			(0.7)	(1.1)
Mean for uninsured plots	4.58	5.02	357.5	397.5	26.6	29.3			175.2	175.9
Num FE's	141	114	101	75	116	88			164	127
Observations	359	242	270	167	308	192			427	275
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel C: Dry Seasons										
Insurance	2.6**	3.2**	12.1	23.8	2.2*	2.3	-0.02*	-0.02*	-1.0	-2.6**
	(1.1)	(1.5)	(20.1)	(24.6)	(1.5)	(2.0)	(0.010)	(0.01)	(1.1)	(1.4)
Mean for uninsured plots	9.23	9.12	307.6	298.0	55.3	60.3	0.11	0.099	188.7	189.1
Num FE's	332	269	202	166	351	283	300	244	297	239
Observations	863	575	545	360	982	633	830	540	808	526

All regressions are estimated on the sample of farmers that got insurance fully by random and include both farmer fixed effects and a control for the plot area. PCD in the first two columns refers to pests and crop diseases. Expenses for chemicals are given in pesos per hectare (peso = 0.02 USD) and Any expenses (columns 5-6) are set to 1 if expenses for chemicals are positive. Using pest or disease resistant seeds is based on self reports about reasons for using the particular seed (this data is only available in the last season). Transplanting day is the date of planting seedlings in the field. Even numbered columns exclude plots that are adjacent to another plot of the same farmer.
Significance (one-sided tests): * < 0.1; ** < 0.05; *** < 0.01.

Table 5: Investment under Moral Hazard

	Harvest loss (%) due to PCD		Expenses for fertilizer		Log expenses for fertilizer	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: All Seasons						
Insurance	1.93 ∗ ∗	2.15 ∗ ∗	−0.22 ∗ ∗	−0.15	−0.43 ∗ ∗	−0.30
	(0.86)	(1.18)	(0.13)	(0.17)	(0.25)	(0.33)
Mean for uninsured plots	7.84	7.86	5.10	5.24	10.4	10.6
Num FE's	474	384	371	293	371	293
Observations	1222	817	944	634	944	634
	(1)	(2)	(3)	(4)	(5)	(6)
Panel B: Wet Season						
Insurance	0.23	−0.21	0.10	0.19	0.20	0.37
	(1.08)	(1.54)	(0.22)	(0.33)	(0.44)	(0.65)
Mean for uninsured plots	4.58	5.02	5.12	5.39	10.5	11.0
Num FE's	141	114	146	113	146	113
Observations	359	242	383	248	383	248
	(1)	(2)	(3)	(4)	(5)	(6)
Panel C: Dry Seasons						
Insurance	2.64 ∗ ∗	3.16 ∗ ∗	−0.46 ∗ ∗∗	−0.39 ∗ ∗	−0.90 ∗ ∗∗	−0.77 ∗ ∗
	(1.14)	(1.55)	(0.15)	(0.17)	(0.30)	(0.34)
Mean for uninsured plots	9.23	9.12	5.09	5.14	10.3	10.4
Num FE's	332	269	224	179	224	179
Observations	863	575	561	386	561	386

All regressions are estimated on the sample of farmers that got insurance fully by random and include both farmer fixed effects and a control for the plot area. PCD in the first two columns refers to pests and crop diseases. Expenses for fertilizer are given in 1000's of pesos per hectare (peso = 0.02 USD). Log fertilizer is defined by $log(fertilizer + \sqrt{fertilizer^2 + 1})$, which can be interpreted like the log function. Even numbered columns exclude plots that are adjacent to another plot of the same farmer.
Significance (one-sided tests): * < 0.1; ** < 0.05; *** < 0.01.

Table 6: Empirical Estimates of Resource Allocation Across Plots

	Damages			
	Total	Typhoon/floods	Pests/diseases	Fertilizer
Insurance	1.38	−0.68	2.05 * *	−0.095
	(1.23)	(0.99)	(0.87)	(0.097)
Farmer in TX Group	1.07	3.25	−2.18	−0.15
	(2.30)	(1.96)	(1.36)	(0.31)
Constant	23.4 * **	14.0 * **	9.45 * **	5.35 * **
	(1.64)	(1.37)	(0.99)	(0.23)
Num FE's	85	85	85	85
Observations	1654	1654	1654	1647
Insurance	1.72	−0.18	1.89 * *	−0.13
	(1.41)	(1.03)	(0.93)	(0.16)
Share of area insured	0.44	4.19	−3.75	−0.13
	(3.71)	(3.25)	(2.26)	(0.53)
Constant	24.0 * **	14.9 * **	9.10 * **	5.30 * **
	(1.12)	(1.00)	(0.72)	(0.15)
Num FE's	85	85	85	85
Observations	1654	1654	1654	1647

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. Standard errors are clustered at the (farm-level) randomization strata level.
Significance: * < 0.1; ** < 0.05; *** < 0.01.

Table 7: Estimated Share of Adverse Selection Explained by Baseline Characteristics

	Loss (%) Due to: All causes	
	Eq. (1)	Eq. (2)
First-choice	5.2 ***	3.4 **
	(1.4)	(1.5)
Area (hectares)	2.5	3.8
	(2.2)	(2.4)
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	23.6	23.6
Num FE's	434	425
Observations	1128	1051

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 8: Estimation of Predicted Damages by Treatment Group

	Loss (%) Due to All Causes	
	Insured Plots	Control Plots
	(1)	(2)
Flooding index	3.9*	6.9 * **
	(2.1)	(2.1)
Medium risk of rats	5.2	4.2
	(4.2)	(4.1)
High risk of rats	5.4	0.5
	(5.2)	(4.8)
Medium risk of tungro	7.9*	−1.0
	(4.5)	(4.2)
High risk of tungro	13.1 * *	3.4
	(5.6)	(6.0)
Season 3 X Flooding index	0.1	−4.9*
	(2.7)	(2.5)
Season 3 X Medium risk of rats	−6.0	−7.1
	(4.6)	(4.6)
Season 3 X High risk of rats	−9.3	−9.0
	(6.0)	(5.7)
Season 3 X Medium risk of tungro	−0.3	3.0
	(5.3)	(4.9)
Season 3 X High risk of tungro	−7.0	−2.4
	(7.0)	(7.2)
Constant	21.6 * **	26.4 * **
	(2.9)	(2.6)
Observations	525	526

Estimates of Equation 30 for insured (Column 1) and control (Column 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.

Table 9: Empirical Estimates of a Selection Model Based on Predicted Damages

	Outcome: Indicator for first choice plot			
	Conditional Logit		Linear FE	
	Odds-ratio (SE)	Odds-ratio (SE)	β(SE)	β(SE)
	(1)	(2)	(3)	(4)
Area X $\hat{E}(D\|X, I = 1)$	$1.12***$		$0.025***$	
	(0.044)		(0.0078)	
Area X $\hat{E}(D\|X, I = 0)$		$1.13***$		$0.026***$
		(0.052)		(0.0090)
Area X $(\hat{E}(D\|X, I = 1) - \hat{E}(D\|X, I = 0))$		$1.10*$		$0.023*$
		(0.061)		(0.012)
$\hat{E}(D\|X, I = 1)$	1.01		0.0045	
	(0.024)		(0.0052)	
$\hat{E}(D\|X, I = 0)$		1.01		0.0066
		(0.028)		(0.0062)
$(\hat{E}(D\|X, I = 1) - \hat{E}(D\|X, I = 0))$		0.99		-0.0018
		(0.032)		(0.0080)
Area (hectares) X $\hat{E}(D\|I = 1)$	0.96	0.95	-0.0073	-0.0084
	(0.035)	(0.040)	(0.0081)	(0.0089)
Number of farmers	462	462	493	493
Observations	1254	1254	1311	1311

Columns 1 and 2 present empirical estimates of Equations (25) and (29) and columns 3 and 4 are their linear analogs. Coefficients in the first two columns are given as odds-ratios. Standard errors, calculated using a bootstrap procedure with 500 repetitions, 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.

Table 10: Summary Statistics of Damages and Payouts

	Wet season			Dry season		
	Mean	(SD)	N	Mean	(SD)	N
Harvest Loss (%)						
All causes	31.7	(26.3)	518	21.2	(26.8)	1264
Typhoons and floods	25.6	(25.4)	518	12.0	(21.5)	1264
Pests and crop diseases	6.0	(13.5)	518	9.4	(19.9)	1267
Payouts (in USD) per hectare (insured plots)	22.9	(52.9)	208	11.8	(36.9)	512

This table shows average harvest losses based on self reports in a follow-up survey and average payout amounts per hectare based on data from the insurance provider. Losses as a percentage of potential harvest due to all causes are calculated as $\frac{\text{Value of loss (all causes)}}{\text{Value of loss (all causes)}+\text{Value of harvest}}$. To calculate typhoon and flood losses I replace the value of loss in the numerator with the value of loss due to this specific cause, and likewise for pests and crop diseases. The wet season data is based on the 2011 wet season and the dry season data is a combination of the 2010-11 and 2011-12 dry seasons. The latter dry season contributes about 85% of the data for the two dry seasons.

Table 11: Summary Statistics and Treatment Balance

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	

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)	0.33	0.33	0.00	0.34	0.34	0.01
			(0.84)			(0.73)
Area (hectares)	0.59	0.60	−0.02	0.59	0.60	−0.01
			(0.39)			(0.62)
Owns plot (1 = yes)	0.23	0.25	−0.02	0.20	0.22	−0.02
			(0.36)			(0.51)
Flooding index (unit SD)	0.02	−0.02	0.04	−0.01	−0.04	0.03
			(0.40)			(0.59)
High Rat Risk (1 = yes)	0.19	0.21	−0.02	0.20	0.22	−0.02
			(0.23)			(0.42)
High Tungro Risk (1 = yes)	0.15	0.14	0.01	0.16	0.15	0.01
			(0.63)			(0.60)
High Wind Risk (1 = yes)	0.04	0.04	−0.00	0.05	0.05	−0.00
			(1.00)			(0.87)
Observations	852	852		683	677	

This 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.

Table 12: Attrition in the experiments

Attrition

Farmer randomization group: (Insurance includes Groups A and B)	Season 1				Season 2				Season 3				All Seasons			
	Insurance		Control		Insurance		Control		Insurance		Control		Insurance		Control	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
A. Farmer attrition																
At randomization	71		36		199		86		336		111		466		211	
Dropouts:																
Did not farm this season		1%		3%		5%		4%		2%		1%		3%		2%
Refused survey		1%		8%		5%		7%		3%		4%		4%		7%
Not found, died, fell ill or other		6%		14%		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				Season 2				Season 3				All Seasons			
	Insurance		Control		Insurance		Control		Insurance		Control		Insurance		Control	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	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%

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.

Table 13: Empirical Estimates of Resource Allocation Across Plots

	Damages			
	Total	Typhoon/floods	Pests/diseases	Fertilizer
Phase 2				
Insurance	4.53	4.35*	0.18	−0.24
	(4.42)	(2.29)	(3.15)	(0.68)
Farmer in TX Group	−13.1*	−11.0*	−2.11	0.41
	(7.01)	(5.32)	(3.49)	(0.73)
Constant	18.1***	13.2***	4.87**	5.33***
	(3.28)	(2.72)	(1.56)	(0.40)
Num FE's	9	9	9	9
Observations	141	141	141	139
Phase 3				
Insurance	1.54	0.73	0.81	0.018
	(2.32)	(2.04)	(0.99)	(0.20)
Farmer in TX Group	−0.64	4.10	−4.73**	−0.66
	(3.76)	(4.30)	(2.01)	(0.55)
Constant	31.5***	22.3***	9.21***	5.85***
	(2.59)	(2.93)	(1.32)	(0.38)
Num FE's	23	23	23	23
Observations	502	502	502	491
Phase 4				
Insurance	1.17	−1.64	2.81**	−0.14
	(1.56)	(1.20)	(1.27)	(0.099)
Farmer in TX Group	4.02	4.67**	−0.65	0.047
	(2.94)	(2.10)	(1.93)	(0.41)
Constant	19.6***	9.64***	9.95***	5.11***
	(2.24)	(1.55)	(1.49)	(0.32)
Num FE's	51	51	51	51
Observations	1011	1011	1011	1017

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. OLS standard errors are reported in parentheses.
Significance: * < 0.1; ** < 0.05; *** < 0.01.

Graphs

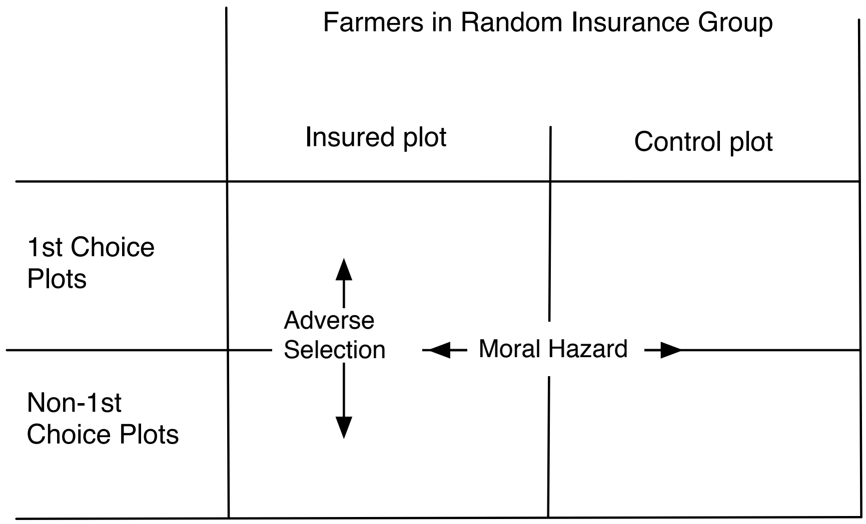

Figure 2: Identification of adverse selection and moral hazard

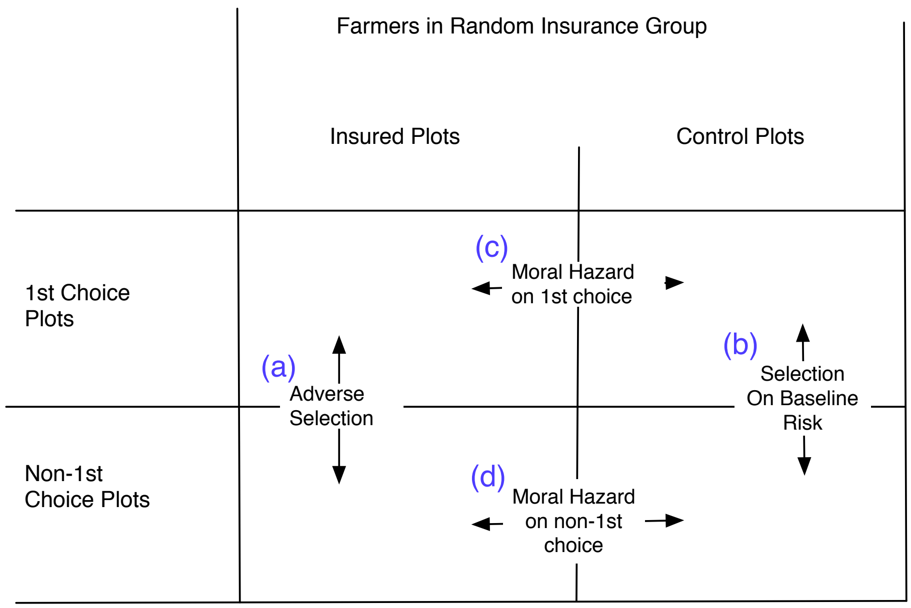

Figure 3: Decomposition of adverse selection

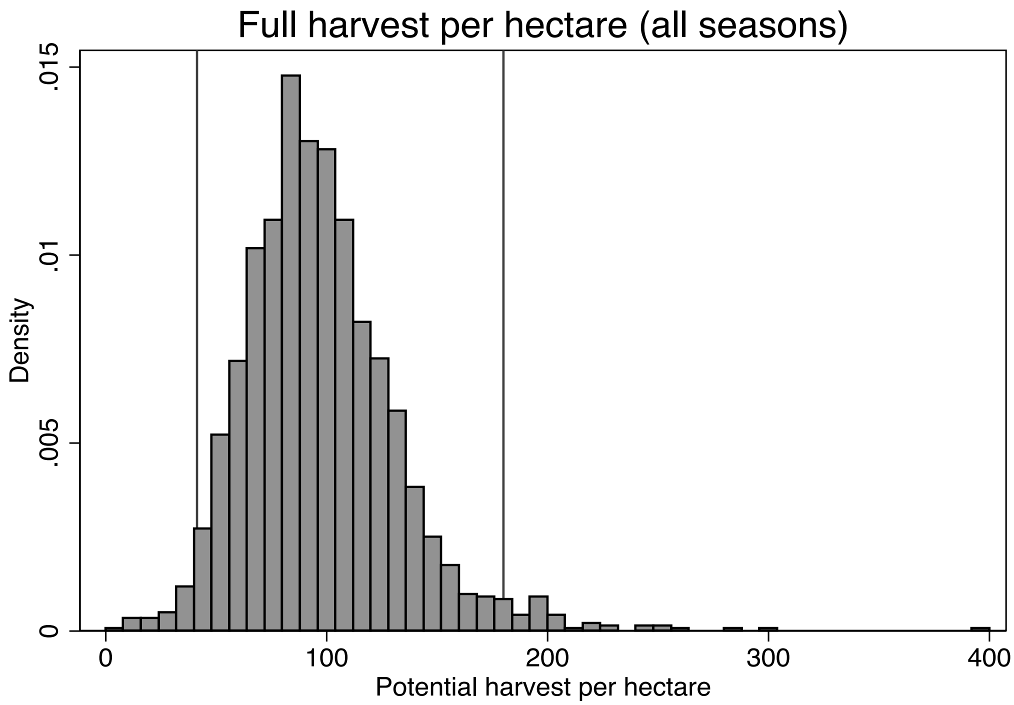

Figure 4: Histogram of full harvest (Harvest + Damages) per hectare for all seasons. The two vertical lines mark the 2.5th and 97.5th percentile. For some robustness regression specifications I exclude plots that fall below the 2.5th or above the 97.5th percentile. The reason is to exclude plots for which rice cultivation was not seriously attempted (below 2.5th percentile) and plots that very likely have erroneous data (above 97.5th percentile). Those above the 97.5th percentile (particularly those well above) are likely the result of misunderstanding between surveyor and farmer in which the farmer mistakenly reports the combined harvest or damages on multiple plots as that from one particular plot.

A Model Derivations

A.1 Interior First Order Condition for Effort

The first order condition for effort is:

$$\begin{aligned}\frac{\partial C}{\partial e_j} &= A_j\left[\frac{(1-\alpha_j L)\theta_j}{2} - \rho A_j(1-\alpha_j L)^2\theta_j^2\frac{-2(1-e_j)}{12}\right]\\ &= A_j(1-\alpha_j L)\theta_j\left[\frac{1}{2} + \rho A_j(1-\alpha_j L)\theta_j\frac{1-e_j}{6}\right]\\ &= W_j\left[1+\rho W_j\frac{2(1-e_j)}{3}\right]\end{aligned}$$

where $W_j = \frac{1}{2}A_j(1-\alpha_j L)\theta_j$. The first order condition for effort implies that:

$$\begin{aligned} A_j\psi_j &= W_j\left[1+\rho W_j\frac{2(1-e_j)}{3}\right]\\ \Leftrightarrow \quad \rho W_j\frac{2(1-e_j)}{3} &= \frac{A_j\psi_j}{W_j} - 1\\ \Leftrightarrow \quad 1-e_j &= \frac{3A_j\psi_j}{2\rho W_j^2} - \frac{3}{2\rho W_j}\\ \Leftrightarrow \quad e_j &= 1 - \frac{3A_j\psi_j}{2\rho W_j^2} + \frac{3}{2\rho W_j}\\ \Leftrightarrow \quad e_j &= 1 - \frac{3}{2}\frac{\psi_j - w_j}{\rho A_j w_j^2} \end{aligned} \tag{31}$$

For an interior solution we must have $e_j \in (0,1)$. This implies for an interior solution we must have

$$e_j > 0 \Leftrightarrow \psi_j < w_j + \frac{2}{3}\rho A_j^2 w_j^2 \tag{32}$$

and

$$e_j < 1 \Leftrightarrow w_j < \psi_j. \tag{33}$$

A.2 Comparative statics

If the solution is interior, the comparative statics are the following:

$$\begin{aligned}
\frac{\partial e}{\partial \theta_j} &= \frac{\partial e}{\partial W_j}\frac{\partial W_j}{\partial \theta_j} \\
&= \left[-(-2)\frac{3A_j\psi_j}{2\rho W_j^2} - \frac{3}{2\rho W_j^2}\right](\frac{1}{2}A_j(1-\alpha_j L)) \\
&= \frac{6A_j\psi_j - 3W_j}{2\rho W_j^3}(\frac{1}{2}A_j(1-\alpha_j L)) > 0 \quad \text{(by 33)} \frac{\partial e}{\partial \psi_j} &= -\frac{3A_j}{2\rho W_j^2} < 0 \qquad (34) \\
\frac{\partial \hat{e}}{\partial A_j} &> 0 \qquad (35) \\
\frac{\partial e}{\partial \rho} &= \frac{3(A_j\psi_j - W_j)}{4\rho^2 W_j^2} > 0 \qquad (36) \\
& \qquad (37)
\end{aligned}$$

That is:

1. $\frac{\partial \hat{e}}{\partial \theta_j} > 0$ (effort is increasing in the inherent riskiness)
2. $\frac{\partial \hat{e}}{\partial \psi_j} < 0$ (effort is decreasing in cost of effort)
3. $\frac{\partial \hat{e}}{\partial A_j} > 0$ (effort is increasing in area)
4. $\frac{\partial \hat{e}}{\partial \rho} > 0$ (effort is increasing in risk aversion)

These comparative statics also apply at the lower corner – that is, for the probability that effort is positive. More formally, if ψ follows a distribution F, we define $\bar{\psi}_j = \frac{3W_j + 2\rho W_j^2}{3A_j}$ and $p_j = Prob(\hat{e}_j > 0) = Prob(\psi_j < \bar{\psi}_j) = F(\bar{\psi}_j)$, then $\frac{\partial p_j}{\partial \theta_j} > 0$, $\frac{\partial p_j}{\partial \rho} > 0$ and $\frac{\partial p_j}{\partial A_j} > 0$. Then

$$\frac{\partial p}{\partial w} = \frac{\partial F}{\partial w} = F'(\bar{\psi})\frac{3 + 4\rho W}{3A} > 0 \qquad (38)$$

$$\frac{\partial p}{\partial \theta} = F'(\bar{\psi})\frac{3 + 4\rho W}{3A}\frac{1}{2}A(1-\alpha L) > 0 \qquad (39)$$

$$\frac{\partial p}{\partial \rho} = F'(\bar{\psi})\frac{2W^2}{3A} > 0 \qquad (40)$$

$$\frac{\partial p}{\partial A} = F'(\bar{\psi})\frac{1}{6}\rho\theta^2(1-\alpha L)^2 > 0 \qquad (41)$$

At the upper corner, the probability that $\hat{e}_j = 1$ is increasing in θ_j and w_j but unaffected by ρ or A_j. Then

$$\frac{\partial q}{\partial w} = F'(\hat{\psi}) > 0 \tag{42}$$
$$\frac{\partial q}{\partial \theta} = F'(\hat{\psi})\frac{1}{2}A(1-\alpha L) > 0 \tag{43}$$
$$\frac{\partial q}{\partial \rho} = 0 \tag{44}$$
$$\frac{\partial q}{\partial A} = 0 \tag{45}$$

A.3 Derivation of Optimal Insurance Choice for a Partially Myopic Farmer

In this section I show that the optimal insurance choice of a farmer that does not anticipate her endogenous effort response to insurance is to choose the plot that has the highest expected payout. I first define a loss function, Λ, that represents the total harvest losses net of insurance payouts and net of effort costs used to prevent damages. Define

$$\begin{aligned}\Lambda(\alpha, \theta, \psi, \mathbf{A}, \rho) = \sum_{j=1}^{N} & A_j w_j (1 - \hat{e}(\alpha_j, \theta_j, \psi_j, A_j, \rho)) \\ & + \frac{\rho}{3} A_j^2 w_j^2 (1 - \hat{e}(\alpha_j, \theta_j, \psi_j, A_j, \rho))^2 \\ & + A_j \psi_j (\hat{e}(\alpha_j, \theta_j, \psi_j, A_j, \rho))\end{aligned} \tag{46}$$

With this definition, total profits are equal to potential harvest less total losses: $\Pi(\alpha, \theta, \psi, \mathbf{A}, \rho) = \sum_{j=1}^{N} A_j - \Lambda(\alpha, \theta, \psi, \mathbf{A}, \rho)$. Since the first term is not impacted by the farmers' actions, she chooses insurance to minimize costs (effort and damages) from natural hazards:

$$\hat{\alpha} = \arg\min_{\alpha} \Lambda(\alpha, \psi_j, ,\rho, \mathbf{A}, \mathbf{w}) \tag{47}$$

Since the farmer does not take into account her anticipated moral hazard response to insurance then she chooses a plot for insurance assuming she will apply effort equal to $\hat{e}(0, \psi_j, \rho, A_j, w_j)$ on plot j (i.e., effort as if the plot will not be insured). Below I will use $\hat{e}_0^j$ as a shorthand for $\hat{e}(0, \psi_j, \rho, A_j, w_j)$. I define the function λ by $\lambda(x, y) = \frac{1}{2}xy + \frac{\rho}{12}x^2y^2$.

Then

$$\Lambda(\alpha, \theta, \psi, \mathbf{A}, \rho) = \sum_{j=1}^{N} A_j \left\{ A_j w_j (1 - \hat{e}_0^j) + \frac{\rho}{3} A_j^2 w_j^2 (1 - \hat{e}_0^j)^2 + A_j \psi_j \hat{e}_0^j \right\}$$
$$\equiv \sum_{j=1}^{N} \lambda(A_j \theta_j (1 - \hat{e}_0^j), (1 - \alpha_j L)) - \sum_{j=1}^{N} A_j \psi_j \hat{e}_0^j$$

Since $\sum_{j=1}^{N} A_j \psi_j \hat{e}_0^j$ is independent of the insurance choice the λ function will determine the plot chosen. Now consider two plots, h and l with $A_h \theta_h (1 - \hat{e}_0^h) > A_l \theta_l (1 - \hat{e}_0^l)$. I will show that plot h is chosen as first choice plot if this inequality holds for all other plots l in the portfolio. Let $\Lambda((\alpha_h = 1, \alpha_{-h} = 0))$ represent the total loss if plot h is insured but all other plots are not insured. Now the difference in total losses between choosing plot h and plot l for insurance is

$$\begin{aligned}
&\Lambda((\alpha_h = 1, \alpha_{-h} = 0)) - \Lambda((\alpha_l = 1, \alpha_{-l} = 0)) \\
= \quad & \lambda(A_h \theta_h (1 - \hat{e}_0^h), (1 - L)) - \lambda(A_l \theta_l (1 - \hat{e}_0^l), (1 - L)) \\
& + \lambda(A_l \theta_l (1 - \hat{e}_0^l), 1) - \lambda(A_h \theta_h (1 - \hat{e}_0^h), 1) \\
\equiv \quad & M(A_h \theta_h (1 - \hat{e}_0^h))
\end{aligned} \tag{48}$$

where I define the function M relative to a given plot l. Now to show that plot h will be chosen I must show that $M(A_h \theta_h (1 - \hat{e}_0^h) < 0$. Now we have

$$\begin{aligned}
\frac{\partial \lambda(A\theta(1 - \hat{e}_0), (1 - \alpha L))}{\partial A\theta(1 - \hat{e}_0)} &= \frac{1}{4}(1 - \alpha L) + \frac{\rho}{6} A\theta(1 - \hat{e}_0)(1 - \alpha L)^2 > 0 \\
\frac{\partial \lambda(A\theta, (1 - \alpha L))}{\partial (1 - \alpha L)} &= \frac{1}{4} A\theta(1 - \hat{e}_0) + \frac{\rho}{6} (A\theta(1 - \hat{e}_0))^2 (1 - \alpha L) > 0 \\
\frac{\partial^2 \lambda(A\theta, (1 - \alpha L))}{\partial A\theta \partial (1 - \alpha L)} &= \frac{1}{4} + \frac{\rho}{3} (A\theta(1 - \hat{e}_0))(1 - \alpha L) > 0
\end{aligned} \tag{49}$$

Given that $M(A_l\theta_l(1-\hat{e}_0)) = 0$ we have

$$\begin{aligned}
M(A_h\theta_h(1-\hat{e}_0^h)) &= M(A_h\theta_h(1-\hat{e}_0^h)) - M(A_l\theta_l(1-\hat{e}_0^l)) \\
&= \int_{A_l\theta_l(1-\hat{e}_0^l)}^{A_h\theta_h(1-\hat{e}_0^h)} \frac{\partial M(s)}{\partial A\theta(1-\hat{e}_0)} ds \\
&= \int_{A_l\theta_l(1-\hat{e}_0^l)}^{A_h\theta_h(1-\hat{e}_0^h)} \left(\frac{\partial \lambda(s, 1-L)}{\partial A\theta(1-\hat{e}_0)} - \frac{\partial \lambda(s, 1)}{\partial A\theta(1-\hat{e}_0)} \right) ds \\
&= -\int_{A_l\theta_l(1-\hat{e}_0^l)}^{A_h\theta_h(1-\hat{e}_0^h)} \left(\frac{\partial \lambda(s, 1)}{\partial A\theta(1-\hat{e}_0)} - \frac{\partial \lambda(s, 1-L)}{\partial A\theta(1-\hat{e}_0)} \right) ds \\
&= -\int_{A_l\theta_l(1-\hat{e}_0^l)}^{A_h\theta_h(1-\hat{e}_0^h)} \int_{1-L}^{1} \underbrace{\frac{\lambda(s, m)}{\partial s \partial m}}_{>0 \text{ (by 49)}} dmds \\
&< 0
\end{aligned}$$

Therefore farmers prefer the plot with the largest $A\theta(1-\hat{e}_0)$. That is, since expected damages per hectare on the plot when not insured are equal to $\frac{1}{2}\theta(1-\hat{e}_0)$ this implies that the farmer chooses the plot that has the highest expected payout (area times expected damages per hectare).