Information Asymmetries in Crop Insurance:
Theory and Experimental Evidence from the Philippines

Snaebjorn Gunnsteinsson*

September 7, 2016

Abstract

Asymmetric information can be costly in insurance markets and can even hinder market development, as is the case for most agricultural insurance markets. I study information asymmetries in crop insurance in the Philippines using a randomized field experiment. Using a combination of preference elicitation, a two-level randomized allocation of insurance and detailed data collection, I test for and find evidence of adverse selection, moral hazard and their interaction – that is, selection on anticipated moral hazard behavior. I conclude that information asymmetry problems are substantial in this context and that variations on this experimental design may be useful in future work for identifying interactions between choice and treatment effects.

JEL: O13; D82; G22; Q120
Keywords: insurance, adverse selection, moral hazard, selection on moral hazard, information asymmetries, selective trials, crop insurance, experiment, Philippines, agriculture

1 Introduction

The incomes of small-scale farmers in developing countries are often very volatile. The structure of agricultural production, combined with exposure to weather variation, pests and crop diseases, and fluctuations in input and output prices, results in incomes that are both periodic and highly uncertain. This risk has important short and long term negative welfare consequences for households (Maccini and Yang, 2009; Currie and Vogl, 2013; Rose, 1999). It also depresses

*University of Maryland, College Park; 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, 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. 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 for Insurance Economics and the Russell Sage Foundation.

investment in agriculture and thereby hinders the process of development in which improvements in agricultural productivity allow developing countries to shift their economy towards manufacturing and services.

Farming households and communities use a variety of means to manage this risk. To smooth consumption they borrow, save, and share risk with neighbors, friends, family members or traders through reciprocal gift networks or contingent credit arrangements. These strategies typically do not allow households to fully smooth their consumption[1] and they can be unreliable, offering little or no protection when communities experience large shocks such as widespread drought (Kazianga and Udry, 2006; Porter, 2012).

Given the incompleteness and unreliability of informal insurance mechanisms a great deal of effort has been put into developing formal insurance contracts to manage agricultural production risk. These fall into primarily two categories, traditional insurance that indemnifies based on farm-specific realizations and index-based insurance products that pay out based on an index such as local rainfall or average regional yield. Neither approach has developed into well functioning private markets for insuring major crops. In the traditional approach, payouts are highly correlated with farm-specific losses but verification of losses is costly and there is high potential for adverse selection and moral hazard. Large-scale programs, such as the Federal Crop Insurance Program in the United States, rely on large government subsidies. The index-based approach is free of adverse selection and moral hazard but basis risk – the risk that an insured farmer is not compensated for losses because they were not reflected in the index – is a major challenge. So far the experience with index insurance is that it typically has positive impact on agricultural production, allowing households to increase investment and shift production to riskier but higher return crops (Karlan et al., 2014; Cole, Giné and Vickery, 2013; Cai et al., 2015; Cai, 2016; Carter et al., 2014). Despite these positive effects demand has generally been low (Gine and Yang, 2009; Cole et al., 2013; Cole, Stein and Tobacman, 2014; Carter et al., 2014).[2] This low demand may be explained by basis risk driving risk averse consumers away (Clarke, 2016) or by a lack of trust or understanding (Cole et al., 2013). Index insurance with

[1] Townsend (1994); Attanasio and Davis (1996); Fafchamps and Lund (2003); Rosenzweig and Binswanger (1993); Rose (1999); Maccini and Yang (2009)

[2] The index insurance studied by Karlan et al. (2014) is an exemption where demand was high.

substantial basis risk may also simply fall into a marketing "dead zone." Because the premiums for small-scale farms is low, financially sustainable sales and marketing would have to rely on high purchase and repurchase rates, and positive word-of-mouth (Cole, Stein and Tobacman, 2014) but these channels are all hindered by high basis risk. The current situation is that index insurance for small-scale farmers has not been developed into a financially sustainable product with substantial market demand (Carter et al., 2014).

Given this, what are the ways forward in developing insurance for small-scale farmers? Future progress could be based on technological innovation, such as drones, satellite data or other new measurement strategies that reduce basis risk in index insurance or improve loss verification and pricing models in traditional insurance. Another avenue may be innovative contracts that meld index insurance with some degree of loss verification (Carter et al., 2014). To make progress on developing financially sustainable insurance for small-scale farmers it is critical (barring technological innovations that substantially solve the above problems) to understand the degree and type of asymmetric information in traditional crop insurance. This understanding can be leveraged to improve traditional crop insurance and develop new products that strike a new balance between basis risk and problems with asymmetric information.

In this paper I contribute to this understanding by studying information asymmetries in a traditional crop insurance contract in the Philippines using a series of randomized field experiments. 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 two stages. In the first stage, I elicited farmers' preference ranking for insurance on plots in their portfolio by asking them to rank the top three plots that they would prefer to be insured. The farmers were told their first-choice 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. 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 (about the first-choice plot) 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 part of 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.

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 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 floods.[3] I find evidence for moral hazard in the prevention of pests and crop diseases. Harvest loss due to these causes is about 22% 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 by 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 evidence that farmers use less fertilizer on insured plots though this impact is small (3-5%). 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 but that any such effect would be small. I do not find any evidence that insurance on one plot induces changes in investment on the farmers' other plots. The possible mechanisms for such an effect, such as scale economies (such as in fixed costs of obtaining inputs), wealth effects (from the reduced investment on insured plots) or important background risk effects (that is, incentives for greater investment on uninsured plots through reduced background risk from insured plots), appear either to cancel out or to be small in response

[3] 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).

to this insurance coverage.

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 30% 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 farmers appear to select on both of these dimensions.

This paper contributes primarily to the literature on agricultural insurance for small-scale farmers in developing countries by complementing the recent literature on index insurance.[4] In designing an insurance product an insurer must choose it's devil by trading off high basis risk against problems with asymmetric information. To design effective policies it is essential to understand the implications of this tradeoff. I contribute to this understanding in two ways. The key contribution is to identify and quantify the separate dimensions of asymmetric information in a crop insurance product in the Philippines. This evidence can be used to improve traditional crop insurance products and develop new products that minimize both basis risk and problems with asymmetric information. A second contribution is that I study the impact of this type of insurance on investment. In contrast to the index insurance literature I do not find increased investment (in fact the evidence supports a small decrease in fertilizer use). I interpret this as being due to the moral hazard inherent in the insurance for pests and crop diseases. Removing this coverage and focusing only on weather related risk may result in an insurance that provides incentives for investment.[5]

[4] See Gine and Yang (2009); Cole, Giné and Vickery (2013); Karlan et al. (2014); Mobarak and Rosenzweig (2013, 2012); Cole et al. (2013); Cole, Stein and Tobacman (2014); Cai et al. (2015); Cai, de Janvry and Sadoulet (2015); Dercon et al. (2014); Hill, Robles and Ceballos (2016); Cai (2016) and citations within Carter et al. (2014), who provide a recent review of this literature.

[5] Since this type of insurance is tied to a particular type of crop and a particular tract of land they would only provide incentives for intensifying production (such as through fertilizer use) as opposed to the type of investment response often observed from weather-index insurance, which is often based on extending production to a larger area and shifting to higher risk but higher return crops (Karlan et al., 2014; Cole, Giné and Vickery, 2013; Cai et al., 2015; Cai, 2016; Carter et al., 2014).

The paper also contributes in several ways to the more general literature on asymmetric information. First, because farmers in this study control multiple insurable units (plots) I am able to study their demand for insurance based on their understanding of the relative risk of loss across their plots. This allows a certain separation between risk on the one hand and the farmers' individual preferences and constraints on the other. 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.[6] Second, I study the issue of *selection on moral hazard* – where consumers demand insurance in part based on their anticipated moral hazard response – which has been done in only one existing paper (Einav et al., 2013). This effect can be identified directly from the experiment but with low statistical power and I rely on an alternative test that takes advantage of baseline data.

Finally, the experimental design and the discussion in Section 9 contribute to a recent literature on enhancing the information produced by randomized experiments. The design in this paper can be generalized as a two step procedure where incentivized choices are obtained in the first step and treatment is allocated according to preferences in the second step. This procedure is related to the one developed in Chassang, Padró i Miquel and Snowberg (2012) but focuses on the choice between two alternative treatments rather than on the willingness-to-pay for a single treatment or program.

The paper proceeds as follows. I will first describe the economic environment in Section 2. Next I discuss the literature on asymmetric information and describe the insurance contract and the experiment in Section 3. I will then present the model and derive empirical implications in Section 4. In Section 5 I discuss the implementation, describe the data and examine the integrity of the experiments. Next I present the three empirical sections. In Section 6 I separately estimate adverse selection and moral hazard, in Section 7 I investigate resource allocation over the farmers' portfolio of plots and mechanisms of moral hazard, and in Section 8 I disentangle selection on baseline risk from selection on moral hazard. In Section 9 I discuss how similar experimental

[6]There is a sizable body of 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, Finkelstein and McGarry (2008) and references within, e.g., Cawley and Philipson (1999); Finkelstein and Poterba (2004); Finkelstein and McGarry (2006); and Fang, Keane and Silverman (2008).

designs could be used in other contexts and I conclude in Section 10.

2 Economic Environment

Rice is the staple crop in the Philippines, and the major crop in the region where the study area is located. All of the farmers participating in this study are growing rice within the Tigman Hinagyaan Inarihan Regional Irrigation System north of Naga City in the Bicol region. The study area is located on low-lying planes and is characterized by a high density of contiguous, usually irrigated rice plots. The yield per hectare is typical for the Philippines. Production in this area is at risk due to floods, droughts, pests, crop diseases and, most importantly, typhoons (tropical cyclones) that hit the Philippines at a rate of about 15 per year.

Farmers in the area use a variety of income and consumption smoothing strategies to manage this production risk. As in other contexts, it is very common to till multiple parcels and to engage in other income generating activities, such as driving tricycles, operating shops or having family members work in the nearby town or city as income smoothing strategies(Rosenzweig and Binswanger, 1993; Dercon, 1996; Morduch, 1995). Fafchamps and Lund (2003) document, in a different region of the Philippines, a substantial role for gifts and informal loans as a way to smooth consumption. A very large literature describes how such income and consumption smoothing strategies are employed elsewhere.[7] As noted earlier, these informal strategies typically provide only partial consumption smoothing and can be unreliable in the event of large aggregate shocks.

To address this uninsured risk the Government of the Philippines established the Philippines Crop Insurance Corporation (PCIC). Among it's products is a multi-peril crop insurance to rice farmers. PCIC is fully owned by the government, which subsidizes the insurance product by about 55% of the premium. The insurance contract offered by PCIC covers rice production on a particular field and pays out in the event of damages to that specific field due to one of the covered causes. These include typhoons, floods, droughts, and various pests (rats and insects) and crop diseases (especially tungro, a crop disease spread by insects).[8] Any particular damage

[7] See Morduch (1995); Rosenzweig and Binswanger (1993); Dercon (1996); Kochar (1999, 1995); Deaton (1992); Udry (1994).

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

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 contracts have a maximum per-hectare payout (in this study, 20,000 pesos or about $430) and pay out based on the share of harvest lost and the timing of loss (farmers can often plant again if damages occur early in the season). In addition, if losses from pests and crop diseases are localized and not due to a wider outbreak affecting many farmers then the payouts are capped at 30% of the policy value. A copy of PCIC's informational flier for the rice crop insurance is included in the Online Appendix.

The fact that payouts are based on the percent of harvest lost rather than an evaluation of the absolute loss is important. It means that payouts are unrelated to underlying productivity or marginal investment (such as fertilizer). A total loss on a relatively unproductive plot that was minimally fertilized would bring the same payment as in a fully fertilized and productive plot provided they are of the same size and both had full standing crops of rice before the damage. This makes verification easier as the adjuster only has to assess the share of crops that are damaged rather than the value of counterfactual harvest. But, as I discuss later, it has potential adverse effects on investment and demand for insurance.

Even with the government premium subsidy demand for this insurance is limited (Reyes and Domingo, 2009). In the 2000's about 30,000 rice and corn farmers were covered each year. In early 1990's, when premium subsidies were even higher, these programs covered over 300,000 farmers. In the next section I describe the experiments that I designed and implemented to understand the degree to which asymmetric information increases the costs of providing this insurance, leading to lower demand and necessitating public subsidies.

3 Experimental Design and Implementation

A very extensive literature analyzes the reasons for the absence or underperformance of financial markets in developing countries (see e.g., Hoff and Stiglitz (1990); Besley (1994); Conning and Udry (2007)). In particular, the seminal contributions of Stiglitz and Weiss (1981) and

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 nontrivial but small, and occur primarily when plants are seedlings (before transplanting), so it is impossible to assign per-plot damage rates.

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 United States 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, Calvin and Quiggin, 1999; Makki and Somwaru, 2001).

Empirical identification of information asymmetries is hard using data normally available to insurance companies and researchers. It is particularly challenging to separately identify the role of each dimension of this asymmetric information, such as that based on heterogeneity in individual preferences, inherent risk or cost of effort. 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, Finkelstein and McGarry, 2008). Third, it is very hard to identify selection on private information that influences the degree of ex-post moral hazard (Einav et al. (2013) term this mechanism "selection on moral hazard"). This mechanism would be operating in our context if a farmer chooses to buy insurance on a plot that is for example close to residential areas (hence susceptible to rats), is next to a plot of a neighbor with lax pest management practices (hence susceptible to insects and other pests), or is far from her home (high fixed cost of monitoring), explicitly because, once the plot is insured, she can save a substantial amount of effort in preventing damages.

A positive correlation between choice of insurance coverage and accident occurrence conditional on data observable by an insurance provider has been shown to be a robust test of the presence of information asymmetries, but such tests can not distinguish between different dimensions of asymmetric information (Chiappori and Salanie, 2000; Chiappori et al., 2006).

Recent contributions have used dynamic data, difference-in-difference techniques, direct data on subjective believes, and structural estimation to make progress on identifying specific components of asymmetric information (Abbring et al., 2003; Finkelstein, McGarry and Sufi, 2005; Cardon and Hendel, 2001; Cohen and Einav, 2005; Einav et al., 2013; Finkelstein and McGarry, 2006).

In this paper, I build on previous experimental efforts, in particular the Rand Health Insurance Experiment[9] and more recently the work of Karlan and Zinman (2009) on consumer credit. In an effort to disentangle many of the relevant information asymmetries I introduce three key features into the experimental design: (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, (2) I introduce experimental variation across plot within the same farm and obtain incentivized choices at the plot level and (3) I introduce experimental variation in insurance coverage across farms.

The study design for each season was as follows:

Step 1: Each farmer was asked to rank, out of their portfolio of plots, the top 3 plots that they would prefer to have insured. They were told that their top choice plot would have a higher chance of receiving free insurance in the 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.

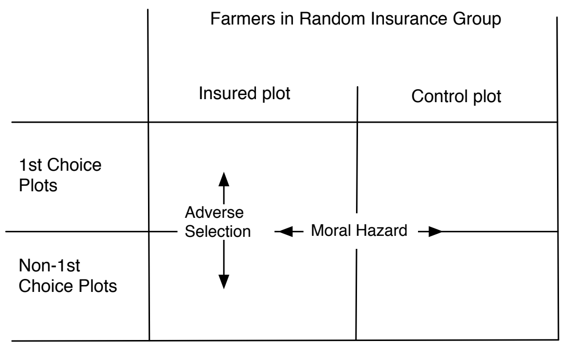
Group C (30%: Control): Received no insurance.

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

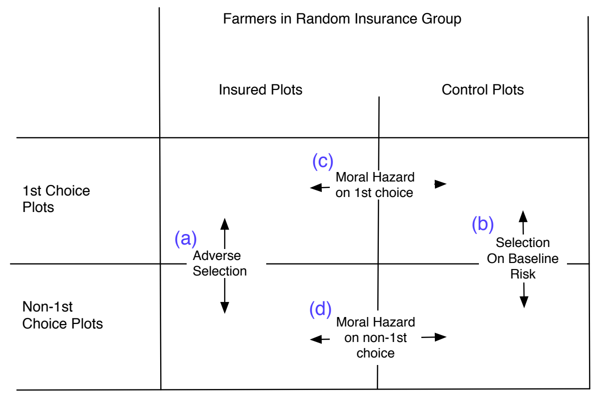
The farmers were not informed of the exact randomization probabilities but were told that their first-choice plot would have a higher chance of receiving insurance coverage. This (Group B)

[9] Key references include Manning et al. (1987) and Newhouse and Rand Corporation. Insurance Experiments group (eds.) (1993).

is a truth-telling mechanism. It ensures that it is incentive compatible for the farmer to reveal her true preference for their first choice plot. The farmer-level randomization was stratified by geographic location.[10] 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 the choice elicitation and the randomization. Two follow-up surveys were conducted in each season, one after planting and another after harvest.



(a) Identification of adverse selection and moral hazard



(b) Decomposition of adverse selection

Figure 1: This figure shows the basic identification strategies

Figure 1a depicts the basic identification strategy to separately identify adverse selection and moral hazard. To identify adverse selection I compare the first-choice plot of the farmer to her other plots, excluding first-choice plots of farmers in the Choice Group. Since insurance coverage is random on this sample of plots, this provides a test for adverse selection. I will test this both by comparing measures of predicted damages, actual damages and payouts. To identify moral hazard I compare randomly insured and uninsured plots within and across farmers. In principle, the design allows me to identify moral hazard separately for first-choice plots and for other plots and therefore identify whether the farmer selects a plot in part based on anticipated moral hazard behavior. Figure 1b depicts how this test would be carried out (here we also exclude first choice

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

plots of the Choice Group so insurance choice and insurance status are orthogonal). Comparing first-choice plots and other plots in the subsample that were not allocated insurance (comparison (b) in the Figure 1b) identifies what I term *selection on baseline risk*, that is, selection on risk characteristics of plots that do not interact with moral hazard. A similar comparison among insured plots (comparison (a) in Figure 1b) identifies the full degree of adverse selection, including the former effect and any interactions with moral hazard. This interaction exists if farmers choose plots in part based on their anticipated moral hazard behavior. In terms of the effects depicted in Figure 1b, we have effect (a) = effect (b) + (effect (c) - effect(d)). Practically, I do not have enough statistical power to carry out this test directly but I will discuss and carry out a modified test of the *selection on moral hazard* effect in Section 8.

4 A Model of Preventative Effort and Choice of Plot for Insurance

4.1 Introduction and Summary of the Model

In this section I develop a model of the decision problem faced by farmers in the experiments. The main building blocks of the model are the following. Each farmer has a portfolio of plots, $\Omega = ((A_1, \theta_1), ..., (A_N, \theta_N))$[11], where A_j is the size of plot j and θ_j indexes the risk characteristics of plot j. Aside from the farmers portfolio of plots, I model two other factors that are likely to be of first order importance in the effort and insurance choice decisions of farmers. First, each farmer has a degree of risk aversion that I model with the parameter ρ. Second, given the well documented role of informal risk sharing in a context such as this, I index the strength of each farmers risk sharing network with the parameter $\tau \in [0, 1]$. A farmer with $\tau = 1$ is fully insured informally and only cares about expected profits whereas the utility of a farmer with $\tau = 0$ is fully penalized (according to her risk aversion) for variability in farm profits.

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

[11] I omit the farmer subscript, i, here and later but with full indexing this expression would be $\Omega_i = ((A_{i,1}, \theta_{i,1}), ..., (A_{i,N_i}, \theta_{i,N_i}))$.

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. Since all insurance is free in the experiment, 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 was allowed to designate any of her eligible plots as her first-choice, regardless of size,[12] 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 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. This effect was termed *selection on moral hazard* by Einav et al. (2013), who identified it using a structural model and data on health insurance in the United States.

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 the insurance is randomly 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

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

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. Rather, they are typically positively correlated, particularly for aggregate shocks such as typhoons. If farmers take into account the likely positive correlation between shocks then they are likely to shift in some cases to choosing the largest plot rather than the plot with the highest expected damages to maximize their payment if, for instance, they experience total loss on all plots. This would lead to some downward bias in the adverse selection estimates reported later. In the empirical section, this issue is also addressed through the design of the experiment (the plot randomization) and through data collection (especially the collection of spatial coordinates of plots, allowing spatially corrected standard errors).

Even if shocks are independent 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'.

4.2 Setup and Maximization Problem

Consider a farmer with a portfolio of plots indexed by j. Each plot, j, is of size A_j hectares, has risk characteristics θ_j and is assumed to produce a maximum output of 1 per hectare (I relax this last assumption in Section 4.5). Some of this output may be lost to natural hazards. The share of harvest lost, S_j, is a random variable that I assume is uniformly distributed on $[0, \theta_j(1 - e_j)]$ 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 $\theta = (\theta_j)_{j=1}^N$ be the vector of plot risk characteristics and $\mathbf{e} = (e_j)_{j=1}^N$ the vector of effort levels across plots. I assume that, conditional on θ and $\mathbf{e}$ (which determine the support of the distribution of losses), the harvest losses are independent random draws across plots.[13]

A plot may be insured, in which case the farmer receives a payout of LS_j per hectare, where $L < 1$ is the per hectare insurance coverage.[14] 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).[15]

The total farm profits are stochastic and given by $\Pi(\alpha, \mathbf{e}) = \sum_j \{A_j((1-S_j)+\alpha_j LS_j)\} - C(\mathbf{e})$ where C is the cost-of-effort function. Given their stochastic nature, the resulting utility derived by the farmer is based on her risk aversion and the degree of other risk sharing arrangements that she has in place. I assume the farmers' preferences can be represented by a mean-variance utility over total future profits : $E\left[U(\Pi)\right] = E\left[\Pi\right] - \rho(1-\tau)Var(\Pi)$. In this setup, there is a utility penalty for variability in profits (according to her risk aversion) but this penalty is tempred by the degree of informal risk sharing that the farmer is engaged in. A farmer who's risk sharing network allows full informal risk sharing ($\tau = 1$) would only derive utility from the first term.

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, to maximize expected utility:

$$\max_{\alpha,\mathbf{e}} E\left[\Pi\right] - \rho(1-\tau)Var(\Pi) \tag{1}$$

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

[13]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. I make this assumption in the model for tractability. In the empirical section I use the spatial data to adjust standard errors to take account of the spatial correlation.

[14]I define $L < 1$ for simplicity but this can be thought of as the maximum payout divided by the typical harvest if no damages occur. The average harvest is valued at 47.3 thousand pesos, the value of the average damages are 15.5 thousand pesos and the maximum payout in the experiments is 20 thousand pesos. These numbers yield an $L = \frac{20}{15.5+14.7} = 0.32$.

[15]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.

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 after the insurance allocation in the experiment is known and as input into the first stage choice problem.

4.3 Optimal Effort

To derive the optimal effort, 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 plot-specific features that make it hard to prevent pests or insects, or they may represent how easy or hard it is to drain the plot after heavy rains. They may also incorporate other sources of the cost of effort, such as distance from home or scale economies (since they are per-hectare costs). In the case of distance from home, 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.[16] Total effort costs are assumed to be separable and additive: $C(\mathbf{e}) = \sum_{j=1}^{N} A_j \psi_j e_j$.[17] Given this setup, the 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 Appendix C that optimal effort is given by:

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

[16]Scale economies can be due to different plot sizes or due to the same farmer having two plots close to each other. About 35% of the plots in the sample are adjacent to at least one other plot of the same farmer. 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. 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.

[17]This assumption is of course less palatable in the case of plots that are adjacent or very close to each other but I maintain it here for the tractability of the model.

where $w_j = \frac{1}{2}(1 - \alpha_j L)\theta_j$. Figure 2 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 2. The model therefore implies moral hazard over this range.

In this section we have assumed that 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 the farmer is informed of the insurance allocation to her plots, her problem simplifies. Instead of the farmers problem in Section 4.2 she now maximizes only over $\mathbf{e}$ (effort). Then α (insurance) 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 Subsection 4.6. First, I use this characterization of optimal effort allocation to derive the optimal insurance choice.

4.4 Insurance Choice

To 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 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.[18] Based on the utility output of plot j (see Appendix C), the perceived utility gain from insurance on plot j for a farmer constrained by myopia of this type is:

$$\begin{aligned} \Delta u_j^{\text{myopic}} &= u_j^{\text{myopic}}(1, \theta_j, \psi_j, A_j, \rho, \tau) - u_j^{\text{myopic}}(0, \theta_j, \psi_j, A_j, \rho, \tau) \\ &= \frac{1}{2} A_j \theta_j L (1 - \hat{e}_j^0) - \frac{\rho(1-\tau)}{12} A_j^2 \theta_j^2 ((1-L)^2 - 1)(1 - \hat{e}_j^0)^2 \end{aligned} \tag{3}$$

[18]She does, on the other hand, anticipate how her effort level is influenced by plot characteristics. For example, 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.

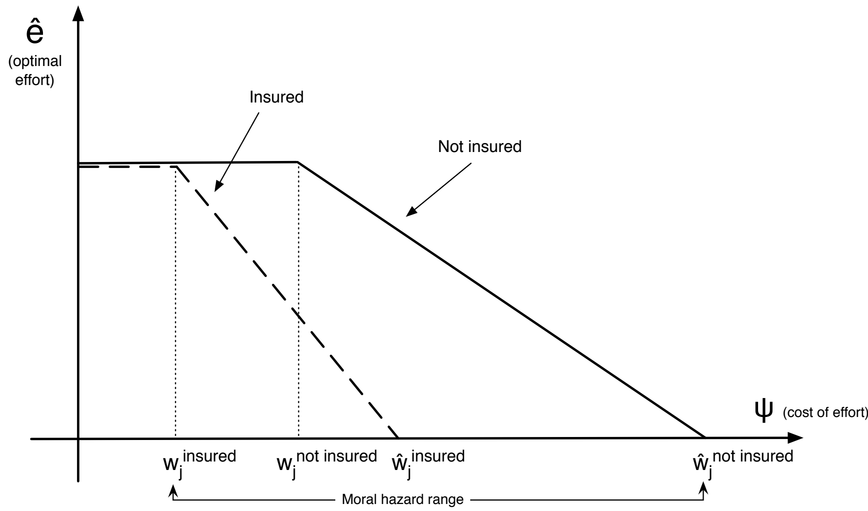


Figure 2: Optimal effort, $\hat{e}_j$, as a function of the plot-specific cost of effort for insured and uninsured plots. Here, w_j^{insured} and $w_j^{\text{not insured}}$ denote w_j for insured and uninsured plots, respectively. Therefore, $w_j^{\text{insured}} = \frac{1}{2}\theta_j(1-L)$ and $w_j^{\text{not insured}} = \frac{1}{2}\theta_j$. 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 first term is the expected payout on plot j if the farmer applies effort as she would without insurance. The second term is the expected gain in utility from the reduction in the variance of profits that the insurance provides (it contributes positively to utility since $(1-L)^2 - 1 < 0$). 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 (a proof can be found in the Online Appendix).[19]

Now contrast this with the insurance choice of a more sophisticated farmer who anticipates her effort response to insurance and takes an optimal decision with this in mind. The perceived

[19]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.

utility gain from insurance in this case is

$$\begin{aligned}\Delta u_j^{\text{sophisticated}} &= \frac{1}{2}A_j\theta_j\left[(1-\hat{e}_j^0)-(1-L)(1-\hat{e}_j^1)\right] \\ &+ \frac{\rho(1-\tau)}{12}A_j^2\theta_j^2\left[(1-\hat{e}_j^0)^2-(1-L)^2(1-\hat{e}_j^1)^2\right] \\ &+ A_j\psi_j(\hat{e}_j^0-\hat{e}_j^1)\end{aligned} \tag{4}$$

As before, the farmer derives utility gain from the increase in expected profits inclusive of insurance payouts (the first term above) and the decrease in the variance of profits (the second term) but in contrast to the myopic farmer she anticipates her moral hazard behavior when evaluating these terms. In contrast to the earlier case the farmer also takes into account the third term above that captures the utility gain from the effort that the farmer saves when the plot is insured. Therefore, in this case the farmer balances the gains from an insurance payout against the gains from saved effort.

4.5 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 farmer jointly determines the level of effort and investment across her portfolio of plots. Her profit function is now defined as $\Pi(e, f) = \sum_j^N \{G(f_j)A_j(1-S_j) + \alpha_j L S_j A_j\} - C(\mathbf{e}) - F(\mathbf{f})$. Using the properties of the exponential utility as before the farmers maximization problem becomes:

$$\max_{\mathbf{e},\mathbf{f}} \sum_{j}^{N} A_j \left[G(f_j) - \frac{1}{2}(G(f_j) - \alpha_j L)\theta_j(1-e_j) \right]$$
$$- \rho(1-\tau)\frac{1}{12}A_j^2(G(f_j) - \alpha_j L)^2\theta_j^2(1-e_j)^2 - C(\mathbf{e}) - F(\mathbf{f}) \tag{5}$$

Because of 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.[20] However, insurance coverage can impact investment through the joint determination of effort and investment. The insurance coverage incentivizes less effort to prevent damages which in turn makes additional productive investments (such as fertlizer) less cost effective and more risky. Under plausible assumptions, this gives the prediction that insurance coverage reduces productive investment.[21]

4.6 Empirical Implications

The data allows me to test various features of the model. Some of these are not specific to this model (almost any model would for example predict adverse selection and moral hazard in this data) but they are listed here for completeness.

Adverse Selection The model predicts that farmers will prefer insurance on plots that are large and risky. Unless there is a strong negative correlation between plot size and risk of damage,

[20] 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.

[21] To illustrate this, first note that the first order condition with respect to investment is $p_f = G'(f_j)\{A_j\left[1 - \frac{1}{2}\theta(1-e)\right] - \frac{1}{6}\rho(1-\tau)A_j^2(G(f_j) - \alpha_j L)\theta_j^2(1-e_j)^2\}$. Now, taking the derivative of this equation with respect to effort, we have:

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

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 $\frac{\partial f}{\partial e} > 0$, that is, that reduced preventative effort reduces investment.

this translates into a prediction of adverse selection. Empirically this correlation seems to be small and positive (plot size and total damages have a correlation of about 0.05). Nevertheless, I condition on plot size in the specification later on to prevent a false positive test of adverse selection. However, the fact that farmers choose in part on plot size (due to the structure of the experiment) could still lead to false negative tests of adverse selection and will bias estimates downward. Given the results I report later (where I show strong evidence of adverse selection) the former is not a major concern. I will discuss the latter when I interpret the adverse selection estimates in Section 6.

A key feature of the model is the possibility that farmers choose not only on the risk profile of their plots but may also select on plot-specific heterogeneity in cost of effort, inducing a "selection on moral hazard" effect. In Section 8 I will empirically investigate whether the data fits better with a model where farmers select only on the risk profiles of plots (and their area) or whether they are more sophisticated, anticipating their effort response to insurance, and choosing in part on this basis.

Moral Hazard The model predicts that we will observe moral hazard behavior for hazards that can be prevented at a cost that falls within a specific range (See Figure 2). Actions that prevent damages and have negligible costs will be taken by most farmers regardless of insurance status and therefore do not lead to moral hazard. Likewise, there is no room for moral hazard in actions that are so costly that they are never performed. Many actions that help to prevent pests and crop diseases (such as using pesticides and insecticides, or removing infected plants) would fall between these two poles, leading to potential moral hazard behavior for these types of damages. In contrast, preventative measures against typhoons and floods are likely to be so costly that they are not undertaken regardless of insurance status. In this context, preparing for floods by erecting barriers or digging ditches is usually not feasible because the plots are part of a large plain of contiguous plots and each farmer has little ability to control the environment around her plot.

Investment Section 4.5 shows that farmers have an incentive to reduce the use of non-preventative productive investment (such as fertilizer) on insured plots. This highlights a negative implication of the insurance contract design, which doesn't insure marginal productive investment since payouts are based on the percent of harvest lost (rather than absolute loss).[22] It also provides another test for moral hazard.

The mean-variance utility assumed in characterizing the optimal effort and the insurance choice decision implies that getting insurance on one plot does not influence the farmers decisions on her other plots. This does not hold for general utility functions and may not fit the data well, for example 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. By removing some risk in income from an insured plot, the insurance coverage may allow a farmer to take more risk on an uninsured plot. This concept of *background risk* and the related concept of *risk vulnerability* have been studied extensively in the theoretical literature (Gollier and Pratt, 1996; Christian, 2006; Heaton and Lucas, 2000; Eeckhoudt, Gollier and Schlesinger, 1996) but the empirical evidence is more limited.[23] Cardak and Wilkins (2009) find that background risk due to labor income and health status risk are important for the financial portfolio choice of Australian households. This concept has also been studied in lab experiments by Harrison, List and Towe (2007); Lee (2008) and Herberich and List (2012). In Section 7 I test which of these different predictions fit the data better.

5 Sample, Experimental Integrity and a Description of the Data

Under the direction of the author, Innovations for Poverty Action (IPA)[24] implemented the experiments and data collection from the spring of 2010 through mid-2012. IPA staff invited farmers in the study area that fulfilled certain eligibility criteria (described below) to participate.

[22]This design feature is likely not there by mistake but rather due to the difficulty that insurance adjusters would face in evaluating expected yields, particularly for damage that occurs early in the cropping season.

[23]In the typical use of the term "background risk" it refers to risk in a different domain than the decision under study, such as considering risks to labor income as 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.

[24]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.

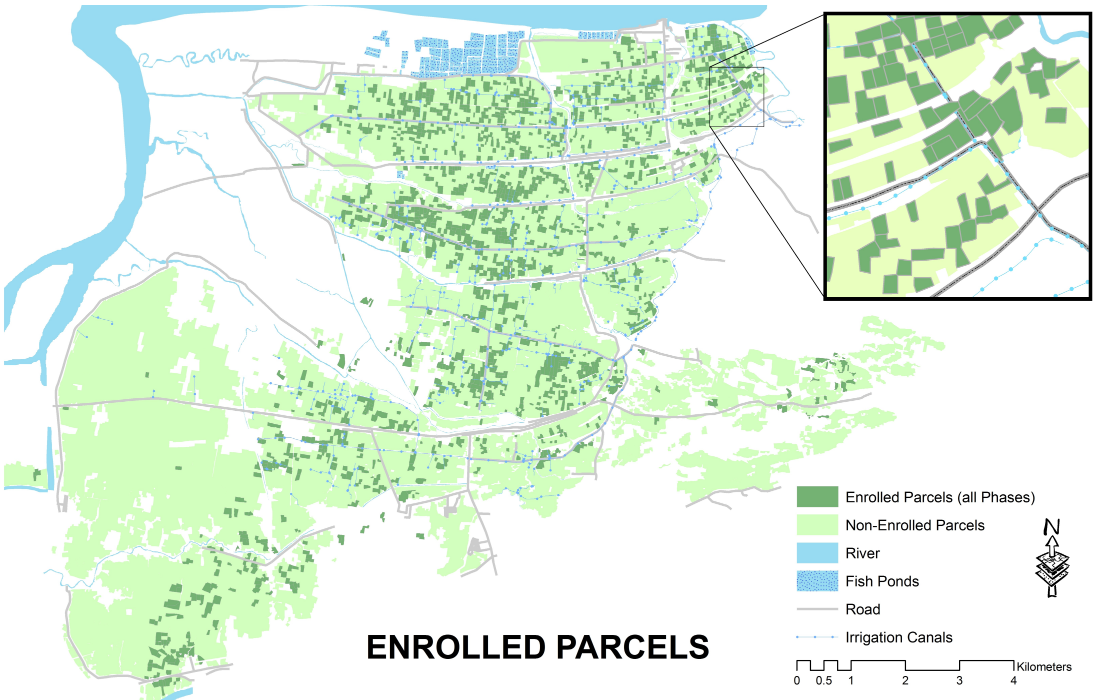


Figure 3: This figure shows a map of the study area. Dark green plots are those that were a part of the study in at least one season while the light green plots are other rice plots.

The implementation started in the 2010 wet season (June - 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. Figure 3 shows parcels that were part of the study in at least one of the seasons.

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.[25] We recruited farmers principally through door-to-door canvassing and, to a lesser extent, at regular farmer meetings. Although we do not have a full census of farmers in the area, based on reports from field staff, we enrolled a large majority of farmers in the target areas that fulfilled the eligibility criteria.

5.1 Sample, Attrition and Integrity of the Experiments

A total of 839 farmers were enrolled in any of the three experimental seasons (counting repeat enrollees multiple times).[26] Of those, 10 farmers fell out of the experiment before farmers were informed of their insurance status due to sickness, death, mistake in enrollment or (in 4 cases) because they were already insured by the company (see Table 11 in the Appendix for the breakdown by season). These farmers are left out of the intent-to-treat sample. Of the remaining 829 farmers, 698 are in the final analysis sample. Of the 131 farmers outside of the sample, 87 dropped out and 44 participated through the end (and were surveyed) but were unable to give information about output or damages. These 87 farmers dropped out because they refused surveys (44), due to sickness, death or migration (14), because they did not planting that season (16), or for unknown reasons (13). Table 11 gives a breakdown of the reasons for attrition for each season. Attrition of farmers overall is greater in the control group (21%) than the treatment group (14%). The attrition improved over time and in the largest and final season was 15% in the control group and 8% in the treatment group. This attrition is not trivial and may affect estimates based on the farmer-level randomization. Most of the analyses in this paper are based on within-farmer comparisons taking advantage of the plot-level randomization and the farmers' preferences for insurance over her portfolio. For farmers in the analysis sample, I

[25]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.

[26]I use the term 'farmer' for farmer-season observations. That is, a farmer in multiple seasons is treated here as a separate observation in each case.

have damage and output data for 90% of plots and this is balanced across treatment (90.5%) and control (89.5%) plots (Table 12 shows the breakdown of plot attrition by season and treatment status). Table 4 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.

5.2 Damage Measures and a Summary of the 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) geospatial data collected by research 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. The most direct way to construct measures that correspond to the model is to compute the percent of harvest lost to the insured events. I will call this the "damage ratio" and define it as damages (total or due to specific causes) divided by harvest plus total damages.[27] In this way, harvest plus total damages are thought of as representing the counterfactual harvest (if no damages had occurred). I separate all-cause 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 floods – and by the fact that this is a categorization the company uses already.[28]

Collecting a panel of plot-specific information can bring certain practical challenges, such as misunderstandings between the farmer and the surveyor about which plot is which and

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

[28] 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.

whether specific information refers to a plot or the whole farm. The survey team employed various measures to minimize this risk. This included collecting information for each plot on the farmers tilling neighboring plots and then reminding the farmer in subsequent survey rounds of the neighbors to the plot being discussed. Nevertheless it seems clear from the data that some errors were made. To limit the impact of these errors on the estimates, in the main sample I defined as missing damage ratio observations that were either more than 10 SD above the mean damage or where the damage reported was more than three times the mean output in the sample. In both of these cases it is likely that the damage reported refers to the larger farm but was mistakenly assigned to a specific plot. One observation of pest and disease damage fits both criteria. In addition, 5 observations of pest and disease damage and 6 observations of typhoon and flood damage fit the second criteria.[29] Section 6.5 discusses robustness of the main findings with respect to including these outliers.

The way the damage data was collected, where farmers were asked about total damages (over the cropping season) due to a specific cause rather than specifying damages for each 'damage event', also presents challenges in comparing damages with payouts. Among insured plots the correlation of total damages and payouts per hectare is 0.51, for typhoon and flood damages versus payouts for typhoons and floods it is 0.49 and for pest and crop diseases 0.37. One issue is that a series of small scale crop losses could add up to a substantial total loss over the cropping season but this would not be covered by the insurance contract and could partly explain these low correlations. We also do not know the specific timing of each loss event. Both issues prevent us from creating damage measures that should correlate stronger with the actual payouts.[30]

Table 3 presents summary statistics of the key outcome variables (Table 4 shows summary statistics of baseline variables). Harvest losses due to the various natural hazards are large in this context. During the three seasons of experiments all-cause harvest losses were 24% on average, with 16% due to typhoons and floods and 8% due to pests and crop diseases. The typical farmer has 2-3 plots of size 0.6 hectares and her harvest value was 47 thousand pesos

[29] These exclusions affect the parameter estimates primarily through the one observation that fits both criteria. This observation is for an unusually small plot (0.15 hectares) that doesn't satisfy the normal eligibility criteria (requiring plots to be more than 0.25 hectares) but was included by exception early on in the study.

[30] There are also some errors or irregularities in the damage and payout data. In particular, we have 7 plots (out of 756 randomly insured plots) that have positive payout even though recorded total damages are less than 10%.

per hectare. Per-hectare insurance payouts were about 5000 pesos ($108) on average or about 650 pesos ($14) per hectare for all insured plots. Conditional on a payout, the average payout amount was 10.3% of the average harvest value.

The baseline survey contains a series of questions on plot characteristics that are used in Section 8 to construct measures of predicted damages by plot. These questions 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, we asked questions on whether the plot is easy, medium, or hard to drain after heavy rains, compared to the farmer's other plots, and whether the plot is low-, medium-, or high-lying, compared to the farmer's other plots.[31] I combine the questions pertaining to floods (flood risk, low-lying and hard to drain) into an index (hereafter "the flooding index") by taking the first principal component from a principal components analysis of three binary variables that signify that the plot is high risk for floods, is low-lying, and is hard to drain after floods. Table 3 shows summary statistics of these variables where I have recoded the rat and tungro variables into indicators for medium and high risk and I've omitted the variable on risk of strong wind as it showed very little variation.

6 Results on Adverse Selection and Moral Hazard

In this section I will first describe how farmers in the study made their plot choice decisions (based on baseline data). I will then empirically estimate the magnitude of moral hazard using data on harvest losses (self reported) and test for adverse selection using data on both harvest losses (again, self reported) and payouts (from administrative data).

6.1 Preference Ranking of Plots for Insurance

The empirical tests for adverse selection will use the fact that by design the farmers were incentivized to reveal their top choice for insurance. In the data we also have their ranking of their top three choices for insurance. Even though the 2nd and 3rd choice is not incentivized the

[31]These characteristics were only collected for seasons 2 and 3.

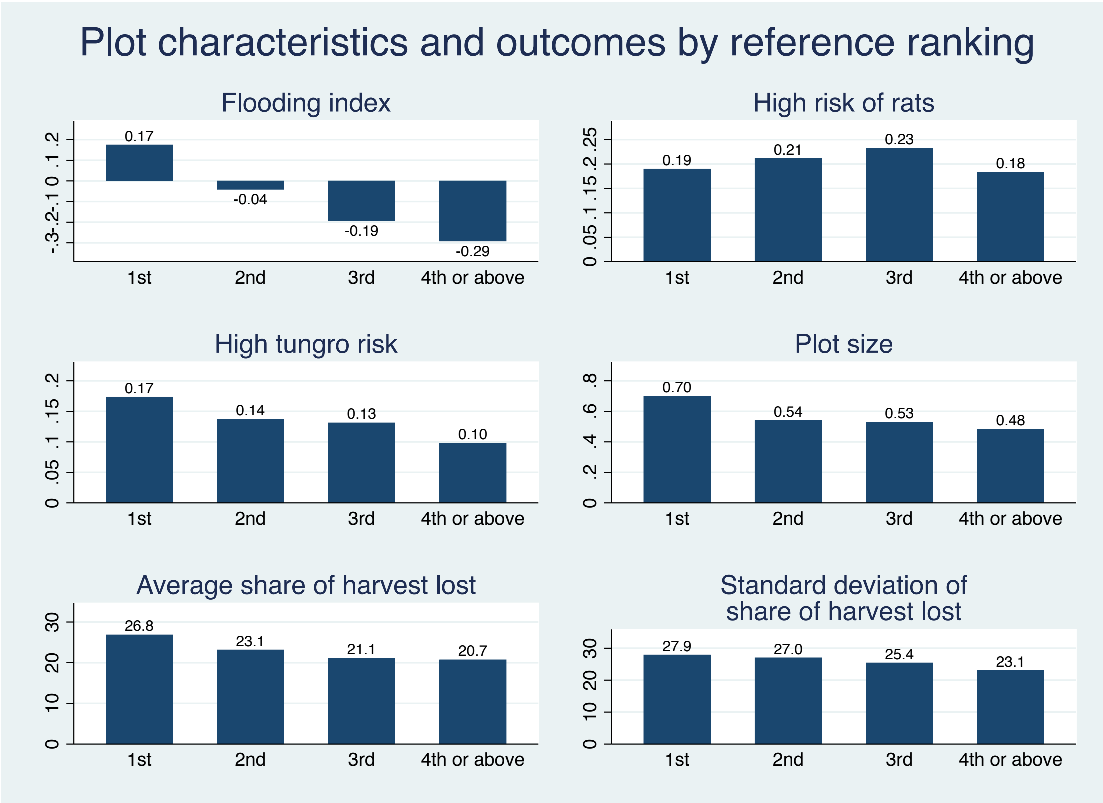


Figure 4: Average share of harvest lost by preference ranking for insurance. The left panel shows the mean and the right panel the standard deviation of the share of harvest lost.

farmers have little reason to not reveal their preference so it is instructive to look at how they made these choices. In Figure 4 I investigate these choices (In this figure I exclude first choice plots of farmers in the choice group so that choice and insurance status are independent). In the first four panels I use baseline survey data while the last two are based on damage data reported after harvest. In the top right panel I report average values of the flooding index by preference rank ordering. It is clear that farmers strongly prefer insurance on plots that they deem at high risk of floods as seen by the monotonic relationship from +.17 for first-choice plots down to -.29 for plots not ranked in the top three. In the top right panel I see no pattern of plot ranking depending on the risk of rat infestations (even though this was reported as very important in qualitative interviews). In the middle panel on the left I plot the average values of an indicator of the plot being of high risk of harvest losses due to tungro, a crop disease spread by insects that often brings devastating harvest losses in this area. This is clearly an important characteristics that farmers use in their choice: 17% of first-choice plots are deemed (by the farmer) to be of high risk of tungro infestation compared 14% and 13% for the second and third choices, and only 10% of the plots that are not in the top 3. The middle panel on the right shows that, due to the features of the experiment, plot size is an important attribute that farmers use in their choice. This will have an impact on how we understand the estimates of adverse selection later in this section and I will discuss those implications in Section 6.3. In the last two panels I use follow-up data on self-reported damages to plot the average and standard deviation of damages by plot ranking. These panels suggest that indeed farmers know which plots are likely to be damaged and can successfully rank them in order. Both panels show a monotonic relationship from first choice plots with an average harvest loss of 27% (and standard deviation of 28) to second and third choice plots with harvest losses of 24.4% (SD 27.2) and 21.2% (SD 25.5), and finally to harvest losses of 20.7% (SD 23.1) for plots that are not ranked as one of the top three.

6.2 Baseline Predictors of Insurance Choice

To get a fuller picture of how farmers in the study chose their top choice plot I estimate a conditional logit model of insurance choice. The outcome variable is a binary indicator of whether the plot was chosen as the farmers' first choice and the choice is conditional on the portfolio of plots the farmer is tilling in that season. I include the size of the plot (standardized to zero mean and unit standard deviation) and the following risk characteristics of the plot: the flooding index (a standardized variable) and indicators denoting that the farmer evaluated the plot as at high risk of rats or tungro. Aside from the risk characteristics of the plot, whether the farmer owns the plot, or if not the type of contractual arrangements the farmer has with its owner, could be important determinants of plot choice. I capture this dimension by including indicators for land ownership and land contractual arrangements. The categories included are sharecropping, mortgaged in, lent for free and owned.[32] The remaining plots, those under fixed rent contracts, are the reference category in the estimation. In the first season the baseline survey did not include questions on risk characteristics of plots so data from that season is excluded from the analysis and the estimation is performed on 486 farmer-season portfolios with

[32] When a plot is 'mortgaged in' it is tilled by this farmer as interest payment for an outstanding loan to another farmer. Plots are lent for free primarily within families, such as children tilling their parents plots.

a total of 1263 plots.

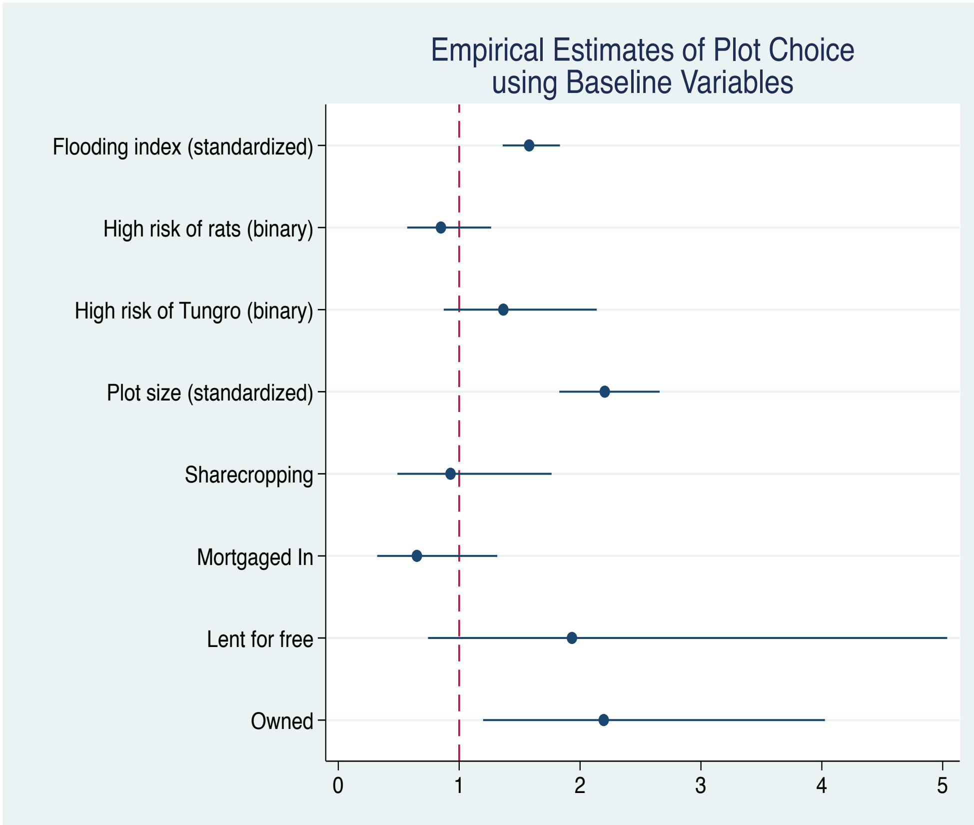


Figure 5: This figure summarizes results of a conditional logit model of plot choice based on baseline predictors.

Figure 5 summarizes the results of the estimation. The figure shows parameter estimates as odds ratios and gives 90% confidence intervals. Significance at the 10% level can be visually assessed from the figure by whether the confidence interval includes 1. Three characteristics of the plot are statistically significant in predicting the plot choice: (1) the flooding index, (2) the size of the plot and (3) whether the farmer owns the plot. I estimate that a plot with a one standard deviation higher flooding index value than another plot is 1.57 times more likely to be chosen but other risk characteristics are not related to plot choice in a statistically significant way. Likewise, a plot that is one standard deviation (0.4 hectares) larger than another plot is more likely to be chosen by a factor of 2.2. Plots that are tilled under fixed rent, sharecropping or are mortgaged in are chosen at similar rates but those owned by the farmer are clearly favored (the estimated odds ratio is 2.2). Plots lent in for free also seem to be favored but this effect is statistically insignificant.

The strong association between the flooding index and insurance choice is additional evidence that the farmers are engaged in a substantial amount of adverse selection. In the next section I confirm this finding using data on damages and payouts, and test for and estimate the degree of moral hazard.

6.3 Main Empirical Specification and Results

The main empirical specification is a within-farm specification (that is, including farm-season fixed effects) with indicators for insurance status and first-choice plot:

$$D_{ij} = \beta_0 + \beta_1 \alpha_{ij} + \beta_2 C_{ij} + \beta_4 A_{ij} + \lambda_i + \epsilon_{ij} \tag{6}$$

The outcome variables (damages, payouts or plot characteristics) are described in the next section. Here α is an indicator for insurance coverage, and C is an indicator for the plot chosen as the farmers' first-choice plot. A is the area of the plot in hectares (centered at the sample mean) and λ_i is a farm-season fixed effect. 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.

I estimate this equation excluding first-choice plots of farmers that were in the choice group. This means, given the randomization of insurance, that plot choice (and the indicator C_{ij}) is independent of the insurance allocation (α_{ij}).[33] This provides unique variation in the data that can be used to separately identify adverse selection and moral hazard. Given the randomized allocation of insurance in this sample the β_1 coefficient captures the effect of insurance on a plot relative to uninsured plots in the farmers portfolio (since farmer-season fixed effects are included). When the outcome variable is damages, a positive β_1 coefficient suggests moral

[33] As a reminder, for those not in the choice group the insurance is allocated at random using block randomization within the farm. Those in the choice group get insurance on the first choice but then the rest of the plots are allocated insurance in the same way as if their farm consistent only of the plots excluding the first-choice plot. Therefore, excluding the first-choice plot of the choice group provides a sample of plot where insurance status is allocated at random.

hazard behavior in preventing damages on insured plots. The β_2 coefficient likewise captures the degree to which damages (or payouts) are higher on first-choice plots relative to other plots in the farmers portfolio (or, in the case of plot characteristics, more adverse). I can therefore test for adverse selection by evaluating whether this coefficient is positive.

The specification above implicitly assumes a constant treatment effect of the insurance coverage across first-choice and other plots. I relax this assumption in Section 8 and allow for heterogeneity in treatment effects and for the possibility that farmers choose plots based on their knowledge of this heterogeneity.

Figure 6 summarizes the results of empirically estimating Equation 6. The figure reports parameter estimates and 90% confidence intervals for β_1 (left panel) and β_2 (right panel) for four outcome variables (listed to the left of the figure). In each case the confidence interval is constructed using standard errors that are corrected for spatial correlation using the spatial GMM method in Conley (1999). These results are also reported in Columns 1, 3, 5 and 7 in Table 5.

6.3.1 Moral Hazard Results

I first describe the findings in the left side panel of Figure 6, which reports the coefficient on insurance allocation for a plot in the above regressions. Since the insurance is allocated at random within the farm it is orthogonal to plot characteristics and to the plot choice decision of the farmer (the latter because we exclude the first choice plots of choice farmers). The coefficient on insurance therefore identifies the causal effect of insurance on the difference in farming practices between insured and uninsured plots. I find evidence for moral hazard in preventing pest and crop disease damage but not in preventing typhoon and flood damage. Damages from pests and crop diseases are on average 1.66 percentage points (21.5%) larger on insured plots compared to uninsured plots of the same farmer. This finding is robust to using alternative outcome variables (such as using the value of harvest lost due to these specific causes, or using the log of this value) and to inclusions of the full set of plot controls. These findings suggest that moral hazard accounts for $\frac{21.5}{100+21.5} = 18\%$ of insurance payouts under the pest and crop disease coverage and 7% of all payouts under the comprehensive insurance

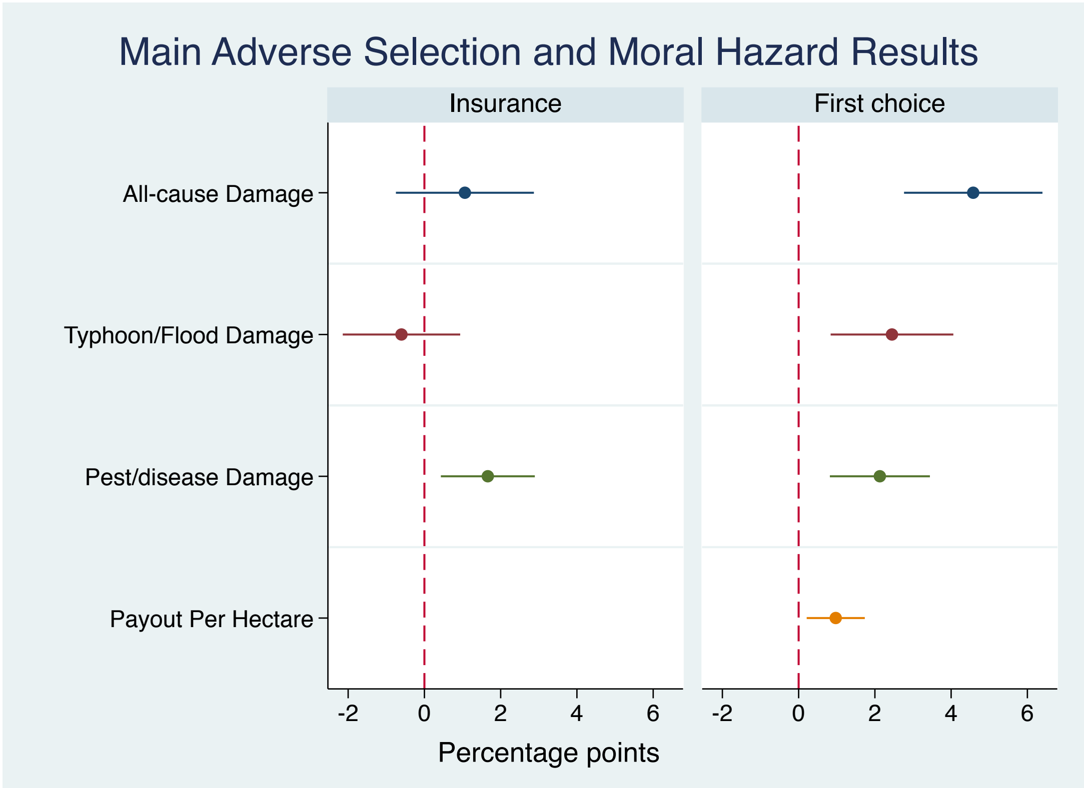


Figure 6: This figure summarizes results of regressions based on Equation 6. Each row represents a separate regression, differing only in the outcome variable. Estimated coefficients and 95% confidence intervals for β_1 (insurance) are shown in the left panel and for β_2 (first choice) in the right panel. The estimations with damages as an outcome variable (the first three rows) are based on 1739 observations with 695 farm-season fixed effects in which I use the full sample except that I exclude first-choice plots of farmers in the choice group. The last estimation (payout per hectare) is based on 691 insured plots and, although all 492 farmers with some insurance are included in the regression, the identification is based on only 92 farmers who had both their first choice plot and at least one other plot insured. Standard errors are corrected for spatial dependence using the method developed by Conley (1999).

coverage. This moral hazard effect is observed even though the insurance coverage is far from complete, suggesting that moral hazard is a significant constraint to offering insurance contracts with higher coverage levels.

6.3.2 Adverse Selection Results

In the results reported in the right panel, I estimate that total damages are 4.6 percentage points (20 percent) greater on first choice plots compared to other plots, which have a damage rate of 22.2 percent. This is in equal parts due to higher damages from typhoons and floods (2.45 percentage points) and pests and crop diseases (2.1 percentage points). Both of these estimates are statistically significant and suggest that first choice plots have 16.3 percent higher damages from typhoons and floods and 29.2 percent higher damages from pests and crop diseases. Finally, in the last row of Figure 6 (and Column 7 of Table 5) I report estimates of Equation 6 on the sample of insured plots using payouts as an outcome variable. The outcome variable here is payouts per hectare as a share of the average harvest value per hectare (for all plots). I find a statistically significant difference with first choice plots having one percentage point (as a share of the average harvest) greater payout than other plots, which have an average payout of 1.4 percentage points.[34]

6.4 Adverse Selection Discussion

Taken together, the evidence on damages and payouts, along with the evidence from the last section on the associations between baseline variables and insurance choice, shows that the farmers have substantial private information about the risk profiles of their land and are able to take advantage of this information in their relationship with the insurance provider. But given that this selection occurs within an experimental setup where farmers choose among plots (for free insurance) rather than purchasing insurance in a marketplace, what can this tell us about adverse selection in the market?

First, because the insurance is not purchased it is possible that the evidence here overstates the possible adverse selection in the market if individuals with high demand for insurance due to their individual characteristics or personal circumstances tend to have land that's lower risk than average (or tend to take better care in preventing damages). The primary worry here might be that individuals with high degree of risk aversion would choose to farm low risk land and work harder to prevent damages but also have high demand for insurance, possibly resulting in what de Meza and Webb (2001) term *advantageous selection*. A second possible issue with extrapolating the findings is that the evidence presented provides an idea of the heterogeneity in risk within the portfolio of plots held by the farmers rather than the heterogeneity across plots

[34] As mentioned earlier, conditional on any payout, the average payout is 10% of the average harvest value (see details in Table 3).

within the pricing region.[35] Finally, an important issue derives from the fact that the insurance was given for free for a specific plot rather than covering a specific acreage (The latter would have either necessitated a revised insurance contract and altered procedures by the company, adding substantial complexity, or it would have required matching payments from the farmers, requiring a much larger experiment). The farmers therefore have an incentive to choose a large plot as their first choice and the tradeoff between size and risk of damage implies that the adverse selection estimates reported in Figure 6 are biased downward.

To investigate the first issue, I use a question from the baseline that asked the farmer how willing she is, on a scale from 1 to 7, to take risks on her farm. I categorized 27% of farmers as having relatively high risk aversion based on this measure. I use this measure in two ways. First, I estimated a simple regression of damages on risk aversion, reported in Table 1. There is little correlation between the risk aversion measure and typhoon and flood damages, but pest and crop disease damage is lower on the farms of more risk averse farmers. This relationship holds, and is almost identical, after controlling for age, education and the number of plots. This is likely due to risk averse farmers taking better care to prevent damages from pests and crop diseases. Although this simple analysis is likely subject to omitted variables bias, this pattern is consistent with the potential for risk preferences to reduce or eliminate the adverse selection in the insurance for pests and crop diseases, or even to induce advantageous selection (this could be the case if, for instance, the true effect of risk aversion is much stronger but I estimate a smaller effect due to measurement error in risk aversion). At the same time, there appears to be little evidence of a risk preference effect in the insurance for typhoons and floods.

Second, to investigate this issue further as well as the second issue identified above I estimated models of the form

$$D_{ij} = \beta_0 + \beta_1 C_{ij} + \beta_2 C_{ij} \times Z_i + \beta_3 A_{ij} + \lambda_i + \eta_{ij} \tag{7}$$

where D_{ij}, C_{ij}, A_{ij} and λ_i are as before but I now include an interaction term between first choice and a farmer or farm level variable Z_i. I estimate six models, varying the outcome

[35] The insurance company sets regional per-hectare prices. As mentioned earlier, all farmers in this study are located within one pricing region.

Table 1: Average Damages by Risk Aversion

	Harvest Loss (percent) Due to:	
	Typhoons and floods	Pests and diseases
Risk averse	−0.36	−2.80 ∗ ∗
	(1.99)	(1.17)
Constant	15.8 ∗ ∗∗	8.85 ∗ ∗∗
	(0.91)	(0.77)
Observations	1654	1654

This table reports estimates of a simple regression of damages, for typhoons and floods in Column 1 and pests and crop diseases in Column 2, on the risk aversion measure. Standard errors are clustered at the farm-season level.

between overall damages, typhoon and flood damage or pest and disease damage, and varying the Z_i variable among ρ_i, the risk aversion measure above and d_i, which is an indicator if the distance between the two plots furthest apart on the farm is larger than the median for the sample (a measure of how spread out the farm is). Here η_{ij} is a disturbance term and I cluster standard errors at the level of the farm-season (rather than using the spatial clustering since the variation of the interaction variables is at the farmer level make within-farm correlations a greater concern).

The findings are reported in Figure 7. In the three figures on top we see that there is no interaction effect with risk aversion for typhoon and flood damage but a substantial negative (and statistically significant) interaction effect with pest and disease damage. That is, I find less adverse selection in pest and disease damage among the more risk averse farmers. In fact, this negative interaction effect is large enough to cancel out the previous adverse selection findings so that I find no adverse selection in pest and disease damage for risk averse farmers. This pattern again supports the view that risk preferences may ameliorate the adverse selection problem for the insurance of pests and crop diseases but not for insuring typhoon and flood damage.

Now turning to the second issue identified above on how to understand what the baseline adverse selection estimate reported in Figure 6 tells us about potential adverse selection across farmers within the same pricing area. In the three figures at the bottom I show that there is

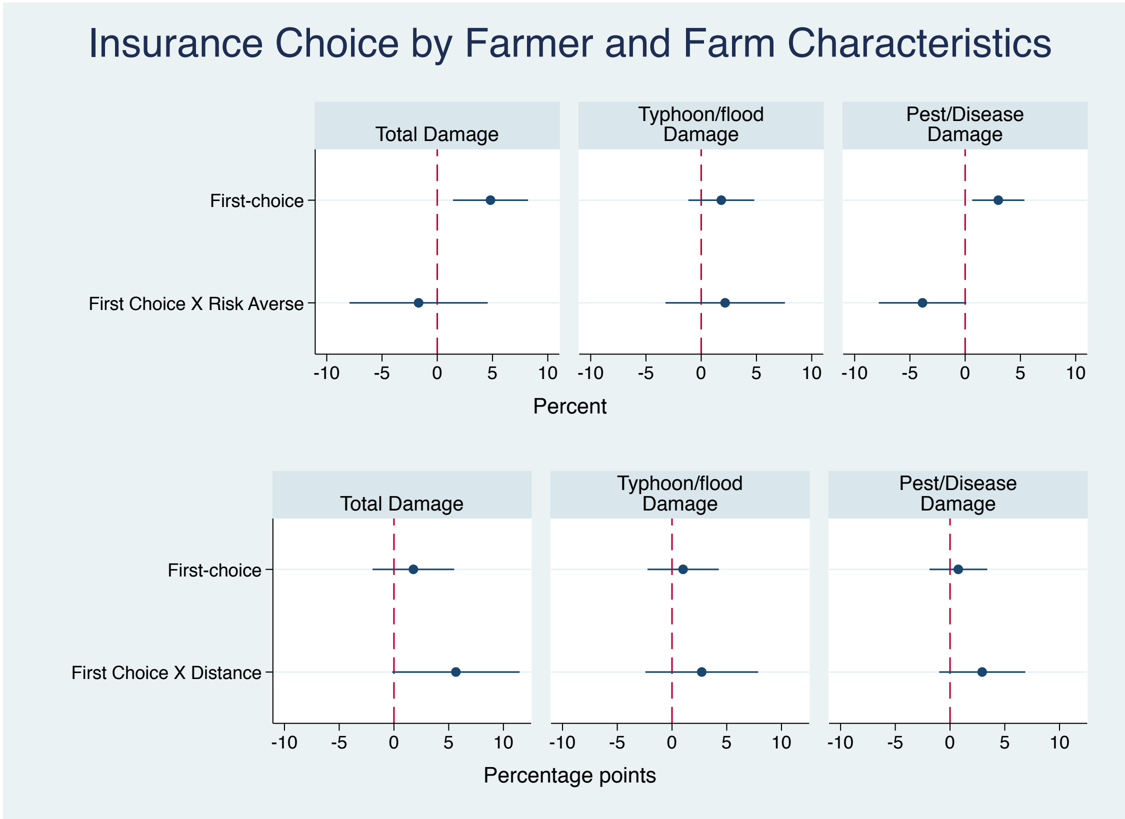


Figure 7: This figure summarizes results of regressions based on Equation 7. The outcome variable are the usual share of harvest losses due to all causes (left), typhoons and floods (middle) and pests and crop diseases (right). The risk aversion measure is an indicator based on a baseline question of how willing the farmer is in taking risks on her farm. The "Distance" variable is an indicator that is one if largest distance between two plots on the farm falls above the median for such distances across all farms. The top three panels are based on 670 farmers with 1654 plots and the bottom three on 612 farmers with 1497 plots (due to missing observations in the risk aversion and distance measures).

greater adverse selection on farms that are more spread out. I estimate that for the less spread out farms the adverse selection effect in total damages is 1.8 but that it is 7.4 on the more spread out farms, and the difference between the two is statistically significant (as can be seen on the bottom left panel as the confidence interval for the interaction term does not contain zero). The average damage on plots not chosen as first choice in the more spread out farms is 20.2 so the first choice plots on these farms have 37% higher damages. This suggests that the risk heterogeneity that the farmers can take advantage of is larger across different parts of the pricing area than it is across the plots of the typical farm. That is, due to this issue, we could expect greater adverse selection in the market than we observe in the experiment. A natural question here is whether the insurance company can collect better information and price these contracts at a finer level (or even at the plot level). Data on how much it would cost to obtain such information is not available but the type of information that can be collected profitably is substantially constrained by the small size of the premium payments for a typical plot.[36]

Finally, to get a sense of how the tradeoff between risk and plot size affects the adverse selection estimates I performed a simple simulation exercise. In the simulation we focus on the simpler case where there is no capacity for selection on moral hazard. In this case the farmer chooses plot j if it has the largest value of $A_j\theta_j$ (area times risk) among the plots in the portfolio. The simulation exercise is designed to answer the question: what difference in damages would we observe between the first choice plot and other plots if farmers instead choose the most risky plot (the one with the largest θ_j)? I simulated a portfolio of plots, ranging from 2 to 5 plots, where each plot consists of a pair (A_j, θ_j) that is drawn from a bivariate uniform distribution with correlation ρ (between A_j and θ_j). I then simulated per-hectare damages, D_j by drawing from a distribution that is uniform from zero to θ_j. For each portfolio, I identify the plot with the largest $A_j\theta_j$ and compute the difference in damages between this plot and (the average of) the other plots, denoting this amount ΔD_j^1. Likewise, I identify the plot with the largest θ_j and compute the difference in damages in the same way to obtain ΔD_j^2. The results for 2 and 3 plots are reported in Table 2.

The first column gives the correlation that is assumed between plot size and risk. The second and fifth column give the average difference in damages (in the simulations) between first choice plots and other plots if farmers choose the plot with the largest $A_j\theta_j$ (as we expect to be the case in the data) whereas columns three and six give the average difference in damages between first choice plots and other plots if farmer choose always the plot with the largest θ_j. Columns four and seven then give the ratio between the earlier two columns. This ratio represents the counterfactual damages (in the simulation) that we would have observed if farmers selected

[36] For instance, the premium (including the government subsidy) for a typical 0.5 hectare plot is in the range of \$20-30 (depending on the season)). For comparison, if an assessor were hired to visit the farm and estimate risk of damage she would likely be paid about \$25 per day (given current salaries in the Philippines). Given travel time and transportation costs the assessment would likely require a very large part of the premium. Going forward it is possible that new technology, such as satellite images or photography from drones, could reduce the cost of collecting this information.

Table 2: Simulation Results

	Two plots			Three plots		
ρ	$\Delta\bar{D}^1$	$\Delta\bar{D}^2$	$\frac{\Delta\bar{D}^2}{\Delta\bar{D}^1}$	$\Delta\bar{D}^1$	$\Delta\bar{D}^2$	$\frac{\Delta\bar{D}^2}{\Delta\bar{D}^1}$
-0.2	0.093	0.166	1.776	0.113	0.189	1.776
-0.1	0.100	0.167	1.662	0.123	0.190	1.551
0.0	0.106	0.166	1.562	0.131	0.191	1.459
0.1	0.113	0.165	1.458	0.139	0.191	1.377
0.2	0.120	0.165	1.378	0.145	0.191	1.319

This table reports the simulation results. The first row gives the correlation that is assumed between plot size (A_j) and risk (θ_j). I performed 10,000 simulations in each case and $\Delta\bar{D}^k$ is the average of Δ^k over those simulations (for $k = 1, 2$).

without regard for the plot size. In the data farmers have on average 2.6 plots and the empirical correlation between the share of harvest lost to all causes and plot size is 0.05. Therefore, rows three and four may be the most relevant. I find that $\frac{\Delta\bar{D}^2}{\Delta\bar{D}^1}$ goes from 1.377 for $\rho = 0.1$ and three plots to 1.562 for $\rho = 0$ and two plots. These results suggest that, if the insurance had been given in a way that did not reward choosing a larger plot (e.g., if all farmers got insurance on a specific acreage) then the adverse selection estimates reported here would be on the order of 38-56% larger.

The model and the simulations above assume no correlation in shocks between plots. In reality these shocks are positively correlated (particularly for typhoon damage) and this might shift some farmers away from choosing the plot with the largest expected damages and towards simply the largest plot to maximize the payout after large shocks, such as total harvest loss on all plots. This would produce a further downward bias in the main adverse selection estimates. However, to evaluate this bias I am limited by the fact that I do not have good measures of this correlation over time (since I only have three cropping seasons).

Based on the above, what does this data tell us about the adverse selection problem that the insurance company faces? First, using the median of the 38-56% interval from the simulations, it suggests that for the more spread out farms the damages on first choice plots would have been twice that of the other plots (obtained by multiplying 1.37 (the adverse selection estimate above) by 1.49 (the median ratio)). This is a reasonable lower bound for the heterogeneity in risk across different farms within a single pricing region that can be identified by the farmers. Selection based on risk preferences (advantageous selection) would counteract this effect (through it's effect on adverse selection in insuring pests and crop diseases), possibly reducing this effect by as much as half (since pests and crop diseases are almost half of the overall adverse selection and I find no adverse selection in pest and crop disease damage among the more risk averse farmers). The degree to which farmers who purchase insurance have higher damages than other farmers naturally depends on the premium that is charged in the market. But at a price in which the minority of farmers would buy insurance, the evidence above suggests that those farmers are likely to have at least 50% larger damages than the uninsured farmers. This would rise to 100% larger if we only consider the typhoon and flood insurance (in the market, farmers can choose to buy only the typhoon and flood coverage, or to pay additional for the comprehensive coverage that also includes pests and crop diseases).

One possible adjustment in the contract would be to focus on catastrophic losses. For example, to only pay out if losses are above $\frac{2}{3}$ of the expected harvest. This would save on verification costs and possibly improve demand through lower price and by focusing on the events with the highest utility cost. Unfortunately this is likely to lead to even more adverse selection. In fact, using specification in Equation 6 I estimate that 6.7% of first choice plots versus 4.2% of other plots have catastrophic typhoon or flood losses (above 66%). The first choice plots therefore have 60% higher chance of such catastrophic loss. This is compared to only a 7% higher chance of typhoon or flood loss above 10%. Taken together it is clear that the company faces a very difficult adverse selection problem.

6.5 Robustness of Adverse Selection and Moral Hazard Results

In Tables 6 and 7, I investigate the robustness of the above evidence on adverse selection and moral hazard. I estimate an equation of the same form as 6 but (in even columns) add controls for plot characteristics. The characteristics included are an index of flooding risk, and indicators for the plot being of high risk of rats, tungro (a crop disease) or strong winds.[37] I perform the

[37]These baseline characteristics were not collected in the small first season experiments and are missing for some plots in the later two seasons. In those cases I replace the indicator values with zero and the flooding index value with the sample mean.

estimation for three different outcome variables in two separate samples. The outcome variables are (1) the damage ratio (as before), (2) the damage ratio winsorized at the 97.5th percentile, and (3) the log of the damage ratio. The sample used in the top panels is the full sample where, in contrast to the sample used for the main results, I do not exclude the outliers discussed in Section 5.2. In the bottom panels of the two tables I restrict the earlier sample to those plots that fall in the middle 95% of a per hectare counterfactual harvest distribution, defined as the value of harvest plus damages divided by the plot size. This is one way to focus in on the sample of plots that are more likely to give accurate results since it excludes very marginal plots or plots that the farmer did not seriously attempt to farm (the bottom 2.5%) and plots where the harvest and damage data together suggest that either one may be inaccurate, for example when a farmer responds to a question about a particular plot with figures that refer to the whole farm (the top 2.5%).

Columns 1, 3 and 5 in the top panel of Table 6 show that there is very strong evidence for adverse selection even if these outliers are included. The same is true for pests and crop diseases (see Columns 1, 3 and 5 of Table 7) but including these outliers slightly reduces the adverse selection estimate. Interestingly, the even columns of these two tables show that the observable characteristics can account for almost all of the adverse selection based on typhoon and flood damage but essentially none for pest and crop disease damage. These observables are self reported by the farmer and so generally not available to the insurance company. It could collect some related data (such as whether the plot lies low relative to surroundings) but, as discussed earlier, given the low premium payments that may be prohibitively expensive. The first row in the bottom panels of the two tables show that the adverse selection estimates are robust to the sample restriction used. The estimate for typhoons and floods is slightly lower but for pests and crop diseases slightly larger.

The second line of the top panel of Table 7 shows that, even with these extreme outliers included, the moral hazard estimates are statistically significant at the 10% level for both the damage ratio and the log of the damage ratio. If we look at the corresponding line for the second sample the results are robust to this restriction and in fact somewhat more statistically significant. In each case (as we would expect given the randomization) the coefficient estimates are essentially unchanged by including plot level covariates.

7 Moral Hazard and Investment

In this section I examine the impact of insurance on farming decisions using measures on inputs (fertilizer, pesticides, seeds), outputs and damages. In the first set of models I estimate models similar to the models in the last section with an indicator for insurance coverage and farmer-season fixed effects. These models show how farmers treated their insured and uninsured plots differently. But, since the uninsured plots may be affected by insurance coverage on other plots in the portfolio, I also estimate models where I drop the farmer-season fixed effect and include indicators for insurance at the plot level and for getting any insurance (at the farmer-season level). This allows me to test whether decisions on an uninsured plot are affected by insurance coverage on one of the farmers' other plots.

7.1 Investment

An important feature of the insurance contract is that it doesn't provide coverage for yield-enhancing investment such as fertilizer because the payout is based on the share of the harvest lost instead of the absolute loss. Total loss for two farmers, one headed for a bumper crop due to heavy fertilizer use and the other for a lackluster harvest due to low investment, would yield the same payout on a per-hectare basis. As I show in the model section, given the moral hazard incentives in the insurance contract this feature implies that farmers have an incentive to use less variable investment such as fertilizer on insured plots. We might therefore expect to see lower fertilizer use on insured plots and, because of this and because of increased damages due to moral hazard, we might also expect lower output.

I Figure 8 I report results of regressions of the form

$$O_{ij} = \beta_0 + \beta_1 \alpha_{ij} + \beta_2 A_{ij} + \lambda_i + \epsilon_{ij} \tag{8}$$

where O_{ij} is the outcome (or input) on plot j on farm i, α is insurance coverage and λ_i are farm-season fixed effects. For easier reporting I standardized all outcome variables reported in

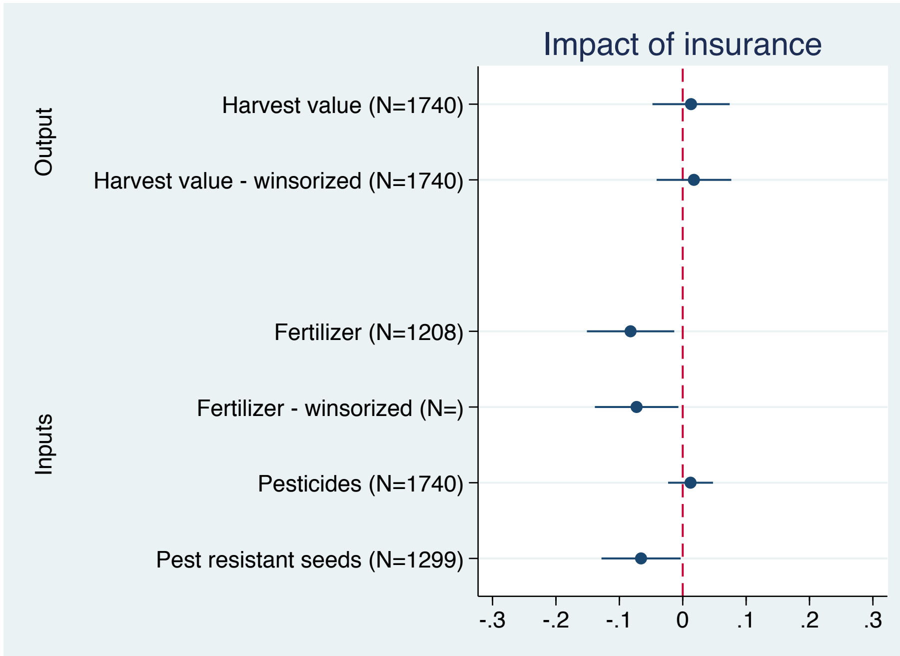


Figure 8: This figure shows the estimated β_1 (insurance) coefficients and 95% confidence intervals from estimating Equation 8 for a series of outcome variables (listed on the left). When outcome variables are marked with "W5" they have been winsorized at the 95th percentile. Each outcome variable is standardized. Standard errors are corrected for spatial dependence using the method developed by Conley (1999).

this figure. For each outcome variable I give the point estimate (in standard deviation units) and 95% confidence intervals computed using spatially clustered standard errors.

I find no effect of insurance on the value of harvest but I find a 5% reduction in average fertilizer use ($p = 0.04$). To temper the effect of possible outliers I also estimated these models after winsorizing the outcome variables. When winsorizing at the 95th (shown) and 90th percentiles I find a reduction of 3.6% and 3%, respectively, both significant at the 10% level. The insurance appears to cause a small reduction in fertilizer use and the data is clearly inconsistent with insurance resulting in an *increase* in variable investment.

7.2 Mechanisms of Moral Hazard

Farmers can manage damages from pests and crop diseases in multiple ways, both individually and together with neighboring farmers. Individually the farmer can choose pest- or disease-resistant seeds, and she can monitor her plots closely, removing infected or infested plants and using pesticides, insecticides, rat poison and other chemicals both as a preventive and as a reaction to a growing outbreak. Collectively neighboring farmers can limit pest and disease damage through coordination, including by synchronizing planting dates. The next to last outcome variable in Figure 8 is total expenditure on chemicals (pesticides, insecticides and rat poison) to prevent insured damages due to pests and crop diseases. Due to moral hazard farmers might be expected reduce these expenditures on insured plots, at least to the degree that these are applied as a preventative (before any outbreak is observed). Once an outbreak is observed, given that the insurance is partial, it is likely that farmers have strong incentives for applying these chemicals whether the plot is insured or not. I find no difference in this expenditure between insured and uninsured plots.[38] This could be either because this is not an important mechanisms for moral hazard in this context or because the measured expenditures are underestimated due to recall bias.[39] The final outcome variable in the figure is an indicator variable that is 1 if the farmer reported in a mid-season survey (right after planting and before realizing damages or harvest amounts) that she chose the seed used on this plot in part because it is resistant to pests, insects or diseases. This measure has some limitations compared to a more objective measure of pest and disease resistance of the seed chosen, but has the advantage of being a measure of the farmers beliefs about the seed type chosen. I find that farmers report choosing pest and disease resistant seeds on 15% of control plots but 13% of insured plots, a statistically significant reduction of 14%.

The reduction in pest- and disease resistant seeds is one additional evidence suggesting that the increase in pest and crop disease damage on insured plots observed in the last section is due to moral hazard and not chance or reporting bias. However, given the small absolute change in the type of seeds used, it is unlikely that this mechanism is the main driver for the observed moral hazard effect.

Perhaps the most important mechanism – the day-to-day care, managing water and fertilizer, monitoring for outbreaks and removing infected and infested plants – is very hard to

[38] I similarly find (but do not show) no difference when limiting this measure to expenses applied as a preventative (before any outbreak is observed).

[39] Some farmers could only give expenditure at the farm rather than plot level and these expenditures are not included in the measure reported in Column 3. The expenses reported by plot are about 0.5% of the average harvest value.

measure. The agricultural surveys included modules for labor allocation and the available data allows estimates of large labor costs such as planting and harvesting costs, costs to applying fertilizer and chemicals and general monitoring costs, but these modules were not sufficient to measure adequately this day-to-day care and I therefore omit testing directly for changes in labor allocation.

7.3 Farm-level Insurance and Background Risk

Impact of Insurance

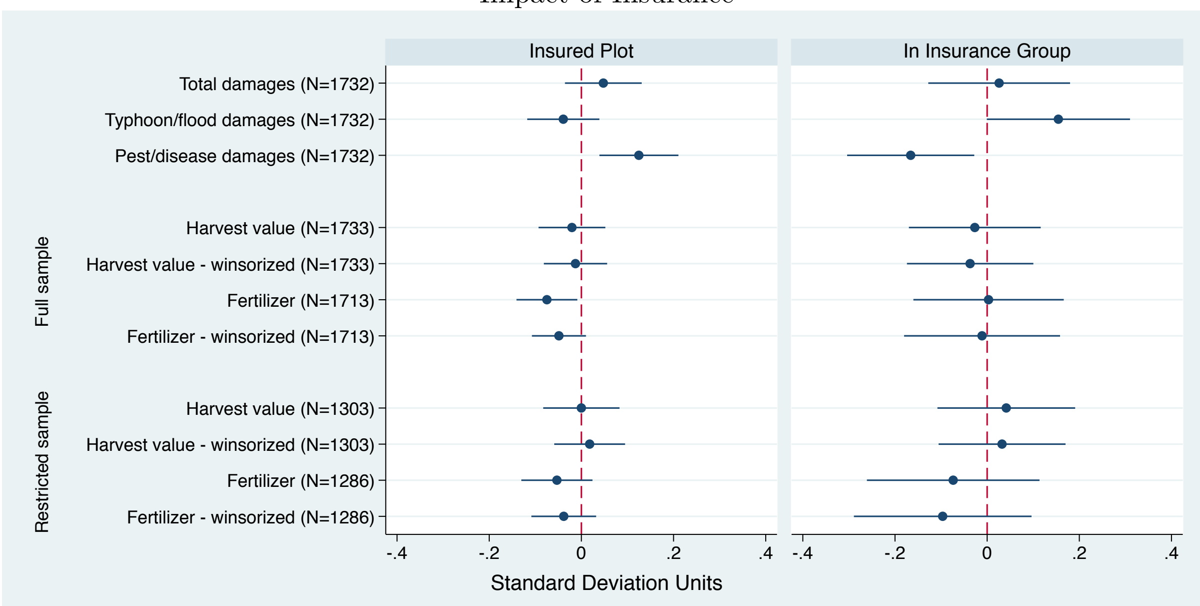


Figure 9: This figure reports estimations of Equation 9. Estimates and 95% confidence intervals for β_1 are reported in the left panel and for β_2 in the right panel. When outcome variables are marked with "W5" they have been winsorized at the 95th percentile. These regressions include 87 (in some cases 86) fixed effects for randomization strata. Standard errors are clustered at the farm-season level.

Insurance was allocated at two levels in the experiments (across farms and across plots within farms). This design provides a test for whether insurance on one plot affects investments on other (uninsured) plots on the farm. Insurance coverage on one plot can affect investment on another through for instance reduction in background risk or scale economies.[40]

[40] One example of the scale economies channel is if there is a fixed cost to going to market and purchasing inputs. If one plot is insured and consequently the farmer does not use pesticides or fertilizer on that plot, then the fixed cost of purchase is no longer shared across the two plots. As a result, in some cases, the farmer might omit these inputs also on the uninsured plot.

I Figure 9 I report results of regressions of the form

$$O_{ijs} = \beta_0 + \beta_1 \alpha_{ijs} + \beta_2 T_{is} + \beta_3 A_{ijs} + \omega_s + \epsilon_{ijs} \tag{9}$$

where O_{ijs} is the outcome (or input) on plot j on farm i in strata s, α_{ijs} is insurance coverage on plot j, T_{is} is an indicator for farmer j in strata s being in the insurance group and ω_s are (farm-level) randomization strata fixed effects. For easier reporting I standardized all outcome variables reported in this figure. Because the variation in T is at the farm-season level I am more concerned about correlated outcomes within farms than spatial correlation so I cluster standard errors at the randomization strata level. The figure reports for each outcome variable the point estimate (in standard deviations) and the 95% confidence interval. In principle we might also want to apply a spatial standard error correction but currently there are no methods to do both at the same time.

Since T is randomized within the strata and α within the farm, β_2 identifies the difference in the outcome of uninsured plots of treatment (insured) farmers and the plots of control (uninsured) farmers. It therefore identifies the effect of any insurance at the farm level on uninsured plots. Likewise, $\beta_2 + \beta_1$ identifies the difference in the outcome of insured plots compared to plots of the control (uninsured) farmers, identifying the combined effect of (some) farm-level insurance coverage and insurance on this specific plot. However, the identification of β_2 is complicated by nontrivial attrition of farmers in the experiments and particularly the fact that the control group had higher attrition (21%) than the treatment group (14%). This attrition is not correlated with farmer demographics (see Table 4) but a key worry is whether it is based on realizations of damages.

The first three models of Figure 9 (from the top) report the results of estimating equation 9 with damages (total, typhoon and flood, pest and crop diseases) as outcome variables. The estimates for β_1 are in the left panel and for β_2 in the right panel. I estimate β_2 to be close to zero for all damages but positive for typhoon and flood damage, and negative for pest and crop disease damage. Both of these are statistically significant. The typhoon and flood effect does not fit with a moral hazard interpretation given the findings of no within-farm differences across insured and uninsured plots, and given that much of this damage is hard or impossible to prevent. The negative effect on pests and crop diseases is also hard to explain based on theory. An important possible interpretation of these findings is that some farmers who suffered high losses from typhoons and floods but were allocated to the control group refused the follow up surveys out of disappointment. This may explain the estimated β_2 coefficient for typhoons and floods. It may also partly explain the pest and crop disease effect since these are negatively correlated (the raw correlation in the control group is -0.15). Another possibility is that the standard errors for β_2 are underestimated since I do not correct for spatial dependence in these models (since methods to do that while also clustering errors within the farm are not currently available).

Because of the above suspicions of differential attrition (at the farm level) based on realized damages I estimate equation 9 for output and fertilizer inputs in two samples. Models four through seven (from the top) in Figure 9 use the full sample while the next four models use a sample that is limited to those farmers that were in the bottom three quarters on the distribution of typhoon and flood damage on their farm (loosing less than 38%). If farmers with very high typhoon and flood damage are attritting due to disappointment then this sample would exclude many farmers in the treatment group that would have attritted had they been allocated to the control group.

For the full sample all estimates for harvest value and fertilizer expenditure of β_2 are close to zero. For the restricted sample none of the estimates is statistically significant compared to zero but the point estimates for fertilizer are negative (about 0.07-0.10SD below zero). Taking these together I do not have evidence to reject the hypothesis that farming decisions on uninsured plots of treatment farmers are unaffected by the treatment. However, this could be due to lack of power, particularly if we look at the fertilizer expenditure in the restricted sample. Interestingly, in the restricted sample $\beta_1 + \beta_2$ is statistically significant (at the 10% level) compared to zero for fertilizer expenditure both using the raw measure and after winsorizing at the 95th percentile ($p = 0.059$ in the former and $p = 0.048$ for the latter). This is consistent with the interpretation that the insurance coverage reduces fertilizer investment. However, these findings must come with the caveat that the β_2 estimates (even in the restricted sample) may be biased due to attrition based on realized damages.

8 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” using measures of predicted damages. The key idea for the decomposition is presented in Figure 1b. I first compute predicted damages based on baseline information separately for insured and uninsured plots (for farmers in the fully random group). Then I identify overall adverse selection by comparing the predicted damages for insured first-choice plots to other insured plots of the same farmer – that is, effect (a) in Figure 1b. 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 – that is, effect (b) in Figure 1b, 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 1b.

In Appendix A I show that plot characteristics observed by me through a baseline survey predict 30% of harvest losses (these plot characteristics are described in Section 5). These characteristics are not observed by the insurance company, and in this section I use them as a 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 plot characteristics, likely because they are too expensive to collect, but they set prices regionally and exclude high risk areas. Although the insurance contract studied is not developed in a competitive market, the fact that they are not collected is suggestive evidence that these characteristics are expensive to collect compared to the premiums that could be sustained in this market. In what follows of this section, I use these variables to construct a measure of predicted damages that I then use to decompose the selection into the two conceptually distinct components.

8.1 Empirical Approach

8.1.1 Adverse Selection without Selection on Moral Hazard

First I consider the case where the farmer does not anticipate changes in effort caused by insurance coverage, either because the cost of effort is very high (low) and the farmer therefore exerts no (full) effort in any scenario or because the farmer is myopic and doesn't take her effort response into account when selecting a plot for insurance. In Section 4.4, I found that in this case 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 $A_j\theta_j$ (area times baseline risk). I will test for adverse selection by comparing this model to the null hypothesis of no adverse selection, where instead farmers simply choose their largest plot. Based on the per-plot utility output derived in the model (Equation 20), the utility of insurance on plot j is: $v_j^* = u_j(\alpha_j = 1) - u_j(\alpha_j = 0) = cA_j\theta_j$ 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:

$$v_j^* = cA_j\bar{\theta} + cA_j(\theta_j - \bar{\theta}) \tag{10}$$

Now, under the assumption that the farmer ignores any effort response to insurance the farmer chooses insurance based on $E\left[D_j|\theta_j\right] = \frac{1}{2}A_j\theta_j$ and I can empirically proxy for the utility by $\hat{u}_j = A_j\hat{E}\left[D_j|\theta_j^{obs}\right]$ where θ_j^{obs} is the portion of risk observable to me based on the baseline characteristics. To empirically estimate 10, I use a conditional logit with the choice conditioned to the portfolio of each farmer (McFadden, 1974). The key assumptions that underlie McFadden's model – that choice probabilities are positive and that choice is independent of irrelevant alternatives – seem reasonable for this context. The estimation equation is

$$Prob(C_{ij} = 1) = \Lambda(\alpha_0 + \alpha_1\hat{E}\left[D|X, I = 1\right] + \alpha_2 A_{ij}) \tag{11}$$

where Λ is the conditional logit function and $C_{ij} = 1$ if farmer i chose plot j as her first choice.[41] To test the model I include a term for area 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.

8.1.2 Decomposition of Selection on Baseline Characteristics

I allow now that farmers may be sophisticated and take into account their endogenous provision of effort on insured plots. Let $\hat{e}_j^I$ be the farmer's 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 20, the utility of insurance coverage on plot j can in this case be written as:

$$v_j^{**} = u_j(\alpha_j = 1) - u_j(\alpha_j = 0) = \overbrace{\underbrace{\frac{1}{2}A_j\theta_j(1-\hat{e}_j^0)L}_{v_b} - \frac{\rho}{12}\left[(1-L)^2 - 1\right] A_j^2\theta_j^2(1-\hat{e}_j^0)^2}^{\text{Utility from coverage of inherent risk}}$$

$$+ \underbrace{\overbrace{\frac{1}{2}A_j\theta_j(\hat{e}_j^0 - \hat{e}_j^1)L}^{v_m}}_{\substack{\text{Utility of coverage}\\ \text{for moral hazard}}} + \underbrace{\frac{\rho}{12}A_j^2\theta_j^2(1-L)^2\left[(1-\hat{e}_j^0)^2 - (1-\hat{e}_j^1)^2\right]}_{\substack{\text{Utility loss due to higher}\\ \text{variance through lower effort}}} + \underbrace{A_j\psi_j(\hat{e}_j^0 - \hat{e}_j^1)}_{\substack{\text{Utility gain from}\\ \text{saved effort}}} \tag{12}$$

I can proxy for the first and third terms (labelled v_b and v_m above) 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: $\hat{v}_b = A_{ij}\hat{E}\left[D|X, I=0\right]$ and $\hat{v}_m = A_{ij}(\hat{E}\left[D|X, I=1\right] - \hat{E}\left[D|X, I=0\right])$. To test separately for the two types of selection I estimate a conditional logit of the form:

$$\begin{aligned}
\Lambda(C_{ij}) &= \alpha_0 + \alpha_1\hat{v}_b + \alpha_2\hat{v}_m + \alpha_3 A_{ij} + \epsilon_{ij} \\
&= \alpha_0 + \alpha_1\hat{E}\left[D|X, I=0\right] + \alpha_2(\hat{E}\left[D|X, I=1\right] - \hat{E}\left[D|X, I=0\right]) \\
&\quad + \alpha_3 A_{ij} + \epsilon_{ij}.
\end{aligned} \tag{13}$$

[41] Strictly the third term should be multiplied by the expected damages for the average plot, $\hat{E}\left[D|\bar{X}, I=1\right]$, but this does not affect the estimate of α_1

Now $\alpha_1 > 0$ provides a test for selection based on what could be called baseline risk, that is, on $\frac{1}{2}\theta_j(1-\hat{e}_j^0) = E[D|I=0]$. The last term in the Equation 12 is positive if and only if $\alpha_2 > 0$. Therefore, given that the cost of effort is positive ($\psi > 0$), $\alpha_2 > 0$ provides a test for selection based on the plot-specific utility of saved effort.

8.1.3 Predicted Damages Based on Baseline Characteristics

To empirically estimate 11 and 13 I must first obtain empirical estimates of predicted damages $\hat{E}[D|X, I=0]$ (for uninsured plots) and $\hat{E}[D|X, I=1]$ (for insured plots). In Table 9 I estimate models of the form:[42]

$$D_{ij} = \beta_0 + \beta_1 X_{ij} + \beta_2 X_{ij} 1(\text{dry season}) + \lambda_i + \eta_{ij} \tag{14}$$

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 flood index and indicators for medium or high risk of rats or tungro), λ_i are farmer-season fixed effects and η_{ij} is an error term. I then use the predicted values from these two regressions as my measures of $\hat{E}[D|X, I=1]$ and $\hat{E}[D|X, I=0]$.

8.2 Results on Estimated Selection Effects

Using these predicted damages I can now empirically estimate equations 11 and 13. Table 8 presents these results. In each case I estimate a conditional logit model with controls for plot size, land ownership and contractual arrangements. For all estimations in this table I perform 500 bootstrap replications to estimate the predicted damages and the choice models. Although I use only farmers who were in the full randomization group (Group A: Received insurance on half of plots at random) to estimate the predicted damages, I use the full sample when estimating the choice models presented in this table (since the choice is made ahead of the randomization).

In Column 1, I estimate equation 11 and find strong evidence for adverse selection. A one percentage point increase in predicted damages increases the odds of a plot being chosen by

[42]The estimates given in this table are for the full sample but to obtain bootstrapped standard errors this prediction is repeated for each bootstrap sample.

8%. In Column 2 I estimate equation 13 and find strong evidence for adverse selection on baseline risk (suggesting again 8% increased odds). The estimate in the third row of Column 2 is consistent with selection on moral hazard as farmers are estimated to have a 7% increased odds (significant at the 10% level) of choosing a plot for each percentage point increase in the difference of expected damages between insured and uninsured plots.

These findings imply that farmers select not only on the baseline risk of plots (that is, $\theta(1-\hat{e}_0)$ in the model) but also on the cost of effort since a positive α_3 coefficient in Equation 13 implies that farmers also take into account the potential of saved effort when choosing plots for insurance.[43]

9 Probabilistic Choice Experiments

Randomized experiments are powerful tools to evaluate the effects of policies or treatments. The randomized allocation reduces (or eliminates) selection bias when estimating average treatment effects. However, in removing this selection bias randomized experiments often provide no information on the important issue of selection into programs. This is particularly problematic when an individuals' selection into a program or treatment is based in part on her private information about the likely program or treatment effect. Heckman, Urzua and Vytlacil (2006) call heterogeneity in treatment effects of this type "essential heterogeneity." In this section I discuss implications of this paper for experimental design and in particular to the design of experiments when the question of interest involves comparing two or more alternative treatments or programs.

In contrast to the setup of a typical randomized experiment, the experiment in this paper introduced a choice variable, S_i, that signifies the choice that the farmer made at the first stage of the experiment (A or B). The key for identification was that this choice was only probabilistic (but incentive compatible). Chassang, Padró i Miquel and Snowberg (2012) studied a similar mechanism that they term *selective trials* in which agents in an experiment make probabilistic choices. They examined how experimental trials can be designed to maximize informativeness when the objective is to evaluate a policy or treatment compared to the alternative of no policy

[43] In the framework of the model, farmers prefer plots with high $A_j\psi_j(\hat{e}^0_j - \hat{e}^1_j)$.

or treatment.[44] They suggest several ways of designing experiments that allow the estimation of marginal treatment effects (MTE's) that in turn can be used to make a richer set of inferences than can be done in the standard trial.[45]

The experiments in this paper and those studied by Chassang, Padró i Miquel and Snowberg (2012) both rely on a two step procedure. First, choices are obtained from agents in an incentive compatible way. Second, agents are randomly allocated to treatment groups based on their stated preferences. In Chassang, Padró i Miquel and Snowberg (2012) agents make choices between money and the treatment or program (through elicitation of willingness-to-pay) while in this experiment agents choose between two alternative (free) treatments. In this case, rather than allocating agents to treatment based on willingness-to-pay they are allocated based on their relative preference for alternative treatments. This is a new experimental setup that may be useful in other settings. It could be referred to as a modified *selective trial* (although the design of these experiments preceded the paper by Chassang, Padró i Miquel and Snowberg (2012)) or separately as *probabilistic choice experiments.*

This paper identifies essential heterogeneity in the treatment, which in this context manifests as selection on moral hazard, in the following sense. A farmer with two plots is given the choice between two treatments. In treatment A she receives insurance on plot 1 but not on plot 2 and in treatment B she receives insurance on plot 2 but not on plot 1. Because the choice and insurance allocation is orthogonal in the experiments, I can identify treatment effects for programs A and B separately. Then, since I observe a farmers choice between A and B I can estimate these treatment effects conditional on the farmers choice, which allows an identification of essential heterogeneity.

To illustrate the possible benefits of this approach, let potential outcome for individual i be denoted by Y_{iA}, Y_{iB} and Y_{iC} if individual i got treatment A, B, or got no treatment, respectively.[46] Let D_i denote the treatment received by individual i. The observed outcome can

[44]Chassang, Padró i Miquel and Snowberg (2012) define informativeness by saying that a mechanism G is at least as informative as mechanisms G' if the data generated by G' can be simulated from the data generated by G.

[45]In the recent literature the concept of the MTE was introduced and developed in Heckman and Vytlacil (2005). According to Heckman (2010), this concept was first introduced by Bjorklund and Moffitt (1987) and further developed in Heckman and Vytlacil (2007).

[46]For discussion on the potential outcome framework, see for example Angrist and Pischke (2008) from which

therefore be written as:

$$Y_i = Y_{iA}1_{\{D_i=A\}} + Y_{iB}1_{\{D_i=B\}} + Y_{iC}1_{\{D_i=C\}} \tag{15}$$

A typical randomized experiment with two treatment groups and a control group would yield one of three outcomes for each individual: Y_{iA} or Y_{iB} if she is in one of the two programs, or Y_{iC} if she is in the control group. Each individual is only in one group but given the random allocation of treatment we can compute the average treatment effects by:

$$\begin{aligned} ATE_A &:= E\left[Y_i|D_i = A\right] - E\left[Y_i|D_i = C\right] \\ ATE_B &:= E\left[Y_i|D_i = B\right] - E\left[Y_i|D_i = C\right] \end{aligned}$$

and the difference in treatment effects by $ATE_{(A-B)} := E\left[Y_i|D_i = A\right] - E\left[Y_i|D_i = B\right]$.

Given a choice variable S_i, in addition to being able to calculate ATE_A, ATE_B and $ATE_{(A-B)}$, four additional treatment effects can be calculated by

$$\begin{aligned} ATE_{(A|A)} &:= E\left[Y_i|D_i = A, S_i = A\right] - E\left[Y_i|D_i = C, S_i = A\right] \\ &= \text{the average treatment effect for those who chose A and received A,} \\ ATE_{(A|B)} &:= E\left[Y_i|D_i = A, S_i = B\right] - E\left[Y_i|D_i = C, S_i = B\right] \\ &= \text{the average treatment effect for those who chose B and received A,} \end{aligned}$$

and similarly for $ATE_{(B|A)}$ and $ATE_{(B|B)}$. Each of the four treatment effects is potentially of interest, as well as the interaction terms, defined as $I_A := T_{A|A} - T_{A|B}$ and $I_B := T_{B|B} - T_{B|A}$. I_A identifies the extent to which those who choose program A benefit more from program A than those who chose program B. That is, I_A identifies essential heterogeneity in the treatment effect of program A when the choice is between A and B (the degree of this essential heterogeneity of course depends on both program A and B since changing either program would change the composition of agents choosing program A).

Some recent work has applied designs that have some of the above features. In a recent article,

I also borrow notation.

Bloom et al. (2015) study the effect of working from home (WFH) on productivity, retention, work satisfaction and promotions among call center workers at a Chinese travel agency. The firm conducted an experiment where workers who volunteered to WFH were randomized to do so or continue working from the office. The choice was incentive compatible (workers would not volunteer to be assigned to WFH if they strongly preferred to be in the office) and those choosing WFH were randomized between the two groups. However, since those choosing to stay at the office were not randomized to treatment groups, this experiment is not maximally informative in the sense defined in Chassang, Padró i Miquel and Snowberg (2012). Bloom et al. (2015) find that those assigned to WFH increased their productivity, had greater work satisfaction and higher retention. They were, however, less likely to be promoted at work. Interestingly, many workers who initially volunteered to WFH decided to return to the office after the experiment had concluded, highlighting the learning process involved. One key strength of the designs outlined above is to be able to identify the extent to which people make mistakes in their choice between treatments or programs.

Experimental designs that incorporate probabilistic choice in this way have the additional benefit of allowing a computation of the "gains from choice." For a simple illustration, take as an example an experiment that obtains the participants preferences between two treatments, A and B, and given the participants choice, assigns R share of participants to that choice and $1-R$ to the other treatment (for incentive compatibility $R > 0.5$). Let's assume the share of people choosing program A is γ_A. The expected outcomes under "full choice" and "no choice" regimes, $\bar{Y}_{\text{full choice}}$ and $\bar{Y}_{\text{no choice}}$, where each person get's their choice versus each person being randomly allocated to treatment, can be recovered from the observed outcomes by:

$$\bar{Y}_{\text{full choice}} = \frac{R\gamma_A E\left[Y_i|D_i = A, S_i = A\right] + R(1-\gamma_A)E\left[Y_i|D_i = B, S_i = B\right]}{R} \tag{16}$$

and

$$\begin{aligned}\bar{Y}_{\text{no choice}} =& R\gamma_A E\left[Y_i|D_i = A, S_i = A\right] + (1-R)\gamma_A E\left[Y_i|D_i = B, S_i = A\right]\\ &+ (1-R)(1-\gamma_A)E\left[Y_i|D_i = A, S_i = B\right] + R(1-\gamma_A)E\left[Y_i|D_i = B, S_i = B\right].\end{aligned}$$

We can then recover the average gain from choice, which is equal to $\bar{Y}_{\text{full choice}} - \bar{Y}_{\text{no choice}}$. For instance, in the crop insurance experiment the per-hectare average payout was 372 pesos among those who (randomly) got their first choice versus 258 pesos for all farmers in the full randomization group, a gain of 44% (this comparison is not statistically significant; however, the comparison of those who got their choice and those who did not is statistically significant with T=2.57).

This approach has several limitations. In many settings it will be hard to randomize people to programs that were not their first choice. A practical problem can also arise if a large majority prefers one program over another. This problem could potentially be solved by implementing the procedures in Chassang, Padró i Miquel and Snowberg (2012) and obtain willingness-to-pay for treatment A versus B. Nevertheless, in certain situations these limitations will be manageable. Applying and improving on designs of this type is a potentially fruitful path for future work.

10 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. The empirical results show that farmers are able to take advantage of a great deal of private information when engaging with crop insurance. This private information includes land features, actions and anticipated actions, giving rise to adverse selection, moral hazard and selection on moral hazard.

I find strong evidence for adverse selection both in the insurance for typhoons and floods and in the insurance for pests and crop diseases. I also find that the pest and crop disease coverage leads to moral hazard, selection on moral hazard and provides (small) disincentives for investment. Interestingly I find none of these effects among the subsample of risk averse farmers suggesting that these may be limited among paying customers who would tend to be the more risk averse. However, this may not be the case for the customers in PCIC's portfolio who often carry insurance coverage through combined credit and insurance contracts, which may draw in the less risk averse.

For a well functioning market (even if it includes some subsidies) the insurance contracts need to either separate the low and high risk types into different contracts or the low risk types must be willing to enter into a pooling contract. The large differences in risk that are evident from the data make the latter hard. This is particularly true since the contract is covering risk that is relatively routine (13.6% of insured plots get some payout) and often nowhere close to catastrophic (23% of payouts are for less than 33% total loss). Assuming that farmers are substantially risk averse, such a contract is less valuable (dollar-for-dollar) than one that exclusively covers catastrophic losses. Adjusting this contract towards covering only the catastrophic losses would reduce verification costs (per unit of premium) and may, through a lower price and greater coverage of events with high utility cost, improve demand. But, based on the empirical findings, this is likely to be at a cost of even greater adverse selection.

Finally, I considered how the experimental design could be generalized for use in other contexts. As discussed in the last section, introducing a two step procedure where agents first make incentive compatible choices between two (or more) treatments followed by randomized allocation of treatment according to these choices, produces data that provides much more information than data from a typical randomized trial. It will depend on the context whether this additional information is of use or whether a design of this type is feasible. But when they are possible, designs of this type would substantially enhance the learning from randomized experiments.

References

Abbring, J, J Heckman, PA Chiappori, and J Pinquet. 2003. "Adverse selection and moral hazard in insurance: Can dynamic data help to distinguish?" *Journal of the European Economic Association*, 1(2-3): 512–521.

Angrist, Joshua D, and Jörn-Steffen Pischke. 2008. *Mostly Harmless Econometrics. An Empiricist's Companion*, Princeton University Press.

Attanasio, Orazio, and Steven J Davis. 1996. "Relative Wage Movements and the Distribution of Consumption." *Journal of Political Economy*, 104(6): 1227–1262.

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

Bjorklund, Anders, and Robert Moffitt. 1987. "The Estimation of Wage Gains and Welfare Gains in Self-Selection Models." *The Review of Economics and Statistics*, 69(1): 42.

Bloom, N, J Liang, J Roberts, and Z J Ying. 2015. "Does Working from Home Work? Evidence from a Chinese Experiment." *The Quarterly Journal of Economics*, 130(1): 165–218.

Burbidge, J B, L Magee, and A L Robb. 1988. "Alternative transformations to handle extreme values of the dependent variable." *Journal of the American Statistical Association*, 83(401): 123–127.

Cai, Hongbin, Yuyu Chen, Hanming Fang, and Li-An Zhou. 2015. "The Effect of Microinsurance on Economic Activities: Evidence from a Randomized Field Experiment." *Review of Economics and Statistics*, 97(2): 287–300.

Cai, J, A de Janvry, and E Sadoulet. 2015. "Social networks and the decision to insure." *American Economic Journal Applied Economics*, 7(2): 81–108.

Cai, Jing. 2016. "The Impact of Insurance Provision on Households' Production and Financial Decisions." *Forthcoming in American Economic Journal Economic Policy.*

Cardak, Buly A, and Roger Wilkins. 2009. "The determinants of household risky asset holdings: Australian evidence on background risk and other factors." *Journal of banking & Finance*, 33(5): 850–860.

Cardon, J H, and I Hendel. 2001. "Asymmetric information in health insurance: evidence from the National Medical Expenditure Survey." *The RAND Journal of Economics*, 32(3): 408–427.

Carter, M, A de Janvry, E Sadoulet, and A Sarris. 2014. "Index-based weather insurance for developing countries: A review of evidence and a set of propositions for up-scaling." *Fondation Pour les Études et Recherches sur le Développement International: Working Paper No. 111.*

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

Chassang, Sylvain, Gerard Padró i Miquel, and Erik Snowberg. 2012. "Selective Trials: A Principal-Agent Approach to Randomized Controlled Experiments." *American Economic Review*, 102(4): 1279–1309.

Chiappori, PA, and B Salanie. 2000. "Testing for asymmetric information in insurance markets." *Journal of Political Economy*, 108(1): 56–78.

Chiappori, PA, B Jullien, B Salanie, and F Salanie. 2006. "Asymmetric information in insurance: General testable implications." *Rand Journal of Economics*, 37(4): 783–798.

Christian, G. 2006. "Increased Risk-Bearing with Background Risk." *The BE Journal of Theoretical Economics.*

Clarke, D J. 2016. "A theory of rational demand for index insurance." *American Economic Journal: Microeconomics.*

Cohen, Alma, and Liran Einav. 2005. "Estimating Risk Preferences from Deductible Choice." *The American Economic Review*, 97(3): 745–788.

Cole, Shawn Allen, Xavier Giné, and James I Vickery. 2013. "How Does Risk Management Influence Production Decisions? Evidence from a Field Experiment." *World Bank Policy Research Working Paper.*

Cole, Shawn, Daniel Stein, and Jeremy Tobacman. 2014. "Dynamics of Demand for Index Insurance: Evidence from a Long-Run Field Experiment." *American Economic Review*, 104(5): 284–290.

Cole, Shawn, Xavier Giné, Jeremy Tobacman, Petia Topalova, Robert Townsend, and James Vickery. 2013. "Barriers to Household Risk Management: Evidence from India." *American Economic Journal Applied Economics*, 5(1): 104–135.

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

Conning, Jonathan, and Christopher Udry. 2007. *Handbook of Agricultural Economics.* Vol. 3 of *Rural Financial Markets in Developing Countries*, Elsevier.

Currie, Janet, and Tom Vogl. 2013. "Early-Life Health and Adult Circumstance in Developing Countries." *Annual Review of Economics*, 5(1): 1–36.

Cutler, D, Amy Finkelstein, and K McGarry. 2008. "Preference heterogeneity and insurance markets: Explaining a puzzle of insurance." *American Economic Review*, 98(2): 157–162.

Deaton, Angus. 1992. "Saving and Income Smoothing in Côte d'Ivoire." *Journal of African Economies*, 1(1): 1–24.

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

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

Dercon, Stefan, Ruth Vargas Hill, Daniel Clarke, Ingo Outes-Leon, and Alemayehu Seyoum Taffesse. 2014. "Offering rainfall insurance to informal insurance groups: Evidence from a field experiment in Ethiopia." *Journal of Development Economics*, 106: 132–143.

Eeckhoudt, Louis, Christian Gollier, and Harris Schlesinger. 1996. "Changes in background risk and risk taking behavior." *Econometrica: Journal of the Econometric Society*, 64(3): 683–689.

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

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

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

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

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

Finkelstein, Amy, Kathleen McGarry, and Amir Sufi. 2005. "Dynamic Inefficiencies in Insurance Markets: Evidence from Long-Term Care Insurance." *The American Economic Review*, 95(2): 224–228.

Gine, X, and D Yang. 2009. "Insurance, credit, and technology adoption: Field experimental evidencefrom Malawi." *Journal of Development Economics*.

Gollier, Christian, and John W Pratt. 1996. "Risk vulnerability and the tempering effect of background risk." *Econometrica: Journal of the Econometric Society*, 1109–1123.

Harrison, Glenn W, John A List, and Charles Towe. 2007. "Naturally Occurring Preferences and Exogenous Laboratory Experiments: A Case Study of Risk Aversion." *Econometrica*, 75(2): 433–458.

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

Heckman, James J. 2010. "Building Bridges Between Structural and Program Evaluation Approaches to Evaluating Policy." *Journal of Economic Literature*, 48(2): 356–398.

Heckman, James J, and Edward J Vytlacil. 2007. "Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Econometric Estimators to Evaluate Social Programs, and to Forecast their Effects in New Environments." In *Handbook of Econometrics*. 4875–5143. Elsevier.

Heckman, James J, and Edward Vytlacil. 2005. "Structural Equations, Treatment Effects, and Econometric Policy Evaluation." *Econometrica*, 73(3): 669–738.

Heckman, J J, S Urzua, and E Vytlacil. 2006. "Understanding instrumental variables in models with essential heterogeneity." *The Review of Economics and Statistics*, 88(3): 389–432.

Herberich, David H, and John A List. 2012. "Digging into Background Risk: Experiments with Farmers and Students." *American Journal of Agricultural Economics*, 94(2): 457–463.

Hill, Ruth Vargas, Miguel Robles, and Francisco Ceballos. 2016. "Demand for a Simple Weather Insurance Product in India: Theory and Evidence." *American Journal of Agricultural Economics*, 98(4): 1250–1270.

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

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

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

Karlan, Dean, Robert Osei, Isaac Osei-Akoto, and Christopher Udry. 2014. "Agricultural decisions after relaxing credit and risk constraints." *The Quarterly Journal of Economics*, 129(2): 597–652.

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

Kochar, Anjini. 1995. "Explaining Household Vulnerability to Idiosyncratic Income Shocks." *The American Economic Review*, 85(2): 159–164.

Kochar, Anjini. 1999. "Smoothing Consumption by Smoothing Income: Hours-of-Work Responses to Idiosyncratic Agricultural Shocks in Rural India." *The Review of Economics and Statistics*, 81(1): 50–61.

Lee, Jinkwon. 2008. "The effect of the background risk in a simple chance improving decision model." *Journal of risk and uncertainty*, 36(1): 19–41.

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

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

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

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

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

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

Mobarak, Ahmed Mushfiq, and Mark R Rosenzweig. 2013. "Informal Risk Sharing, Index Insurance, and Risk Taking in Developing Countries." *American Economic Review*, 103(3): 375–380.

Mobarak, Ahmed Mushfiq, and M Rosenzweig. 2012. "Selling Formal Insurance to the Informally Insured." *Yale Economic Growth Center Discussion Paper*, , (1007).

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

Newhouse, J, and Rand Corporation. Insurance Experiments group (eds.). 1993. *Free for all?: lessons from the RAND Health Insurance Experiment.* Harvard University Press.

Porter, Catherine. 2012. “Shocks, Consumption and Income Diversification in Rural Ethiopia.” *Journal of Development Studies*, 48(9): 1209–1222.

Reyes, Celia M, and Sonny N Domingo. 2009. “Crop Insurance: Security for Farmers and Agricultural Stakeholders in the Face of Seasonal Climate Variability.” *Philippine Institute for Development Studies Discussion Paper Series*, 2009(12).

Rose, E. 1999. “Consumption smoothing and excess female mortality in rural India.” *Review of Economics and Statistics*, 81(1): 41–49.

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

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

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

Townsend, Robert. 1994. “Risk and Insurance in Village India.” *Econometrica*, 62(3): 539–591.

Udry, Christopher. 1994. “Risk and Insurance in a Rural Credit Market: An Empirical Investigation in Northern Nigeria.” *The Review of Economic Studies*, 61(3): 495–526.

Table 3: Summary statistics of baseline and outcome variables

Variable	Mean	(Std. Dev.)	N
Damages			
All cause harvest loss			
Damage ratio	23.93	(26.9)	1774
Value of damages per hectare	15.5	(19.38)	1774
Harvest loss to typhoons and floods			
Damage ratio	15.94	(23.57)	1774
Value of damages per hectare	10.65	(17.2)	1774
Harvest loss to pests and crop diseases			
Damage ratio	7.98	(17.47)	1774
Value of damages per hectare	4.85	(10.78)	1774
Value of Harvest and Payouts			
Value of harvest per hectare	47.32	(23.08)	1774
Payouts per hectare			
For the sample of insured plots	0.68	(1.9)	690
For the sample of plots with any payout	4.88	(2.32)	96
Payouts as share of average harvest			
For the sample of insured plots	1.44	(4.01)	690
For the sample of plots with any payout	10.32	(4.91)	96
Inputs			
Plot-level expenses for chemicals per hectare (missing set to zero)	0.21	(0.46)	1774
Used pest or disease resistant seeds on plot	0.15	(0.35)	1311
Plot-level fertilizer expenses per hectare	5.23	(3.29)	1238
Log plot-level fertilizer expenses per hectare	2.16	(0.72)	1238
Plot Risk Characteristics			
Flooding index	-0.03	(0.99)	1667
Medium rat risk	0.41	(0.49)	1774
High rat risk	0.2	(0.4)	1774
Medium tungro risk	0.38	(0.49)	1774
High tungro risk	0.14	(0.35)	1774
Other			
Plot size (hectares)	0.59	(0.39)	1774
Owner of plot	0.23	(0.42)	1453
Number of plots (one observation per farmer-season)	2.63	(1.29)	641

The table presents summary statistics of key outcome variables.
All variables that indicate value are in 1000's of pesos.

Table 4: 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	1.72	1.58	0.13
			(0.09)			(0.18)
Number of enrolled plots	2.89	2.84	0.05	2.89	2.82	0.06
			(0.66)			(0.60)
Education (years)	10.21	10.47	−0.26	10.16	10.45	−0.29
			(0.40)			(0.39)
Age (years)	53.89	52.95	0.94	53.82	53.14	0.67
			(0.34)			(0.54)
Gender (1 = female)	0.17	0.16	0.00	0.17	0.15	0.01
			(0.87)			(0.65)
Observations	606	233		518	179	

B. Randomization of plots
(excludes plots not randomized (Group C and first-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	0.00
			(0.84)			(0.86)
Area (hectares)	0.59	0.60	−0.02	0.59	0.60	−0.01
			(0.39)			(0.60)
Owns plot (1 = yes)	0.23	0.25	−0.02	0.20	0.22	−0.02
			(0.36)			(0.47)
Flooding index (unit SD)	0.02	−0.03	0.05	−0.02	−0.05	0.04
			(0.37)			(0.49)
High Rat Risk (1 = yes)	0.19	0.22	−0.02	0.19	0.21	−0.02
			(0.21)			(0.44)
High Tungro Risk (1 = yes)	0.15	0.14	0.01	0.15	0.14	0.01
			(0.68)			(0.64)
High Wind Risk (1 = yes)	0.04	0.04	0.00	0.05	0.05	0.00
			(0.91)			(0.83)
Observations	852	852		662	656	

This table shows summary statistics by treatment condition and tests for treatment balance. At the farmer level the right hand part of the table (the analysis sample) is those farmers that did not drop out. At the plot level the right hand part is those plots that have a non-missing value for total damages. Observations are given for the full sample. Some rows are based on a smaller sample due to missing values of the specific variable.

Table 5: Empirical Estimates of Adverse Selection and Moral Hazard

	Harvest loss (percent) due to:						
	All causes		Typhoons and floods		Pests and diseases		Payout (1000's pesos)
First-choice	4.58 * **	3.63 * **	2.45 * **	1.67	2.13 * **	1.96 * *	0.97 * *
	(0.92)	(1.32)	(0.82)	(1.17)	(0.67)	(0.92)	(0.39)
Insurance	1.06	0.18	−0.60	−1.33	1.66 * **	1.51*	
	(0.92)	(1.21)	(0.78)	(1.08)	(0.63)	(0.84)	
First-choice X Insurance		2.57		2.12		0.45	
		(2.38)		(2.10)		(1.82)	
Area (hectares, centered)	1.54	1.54	0.84	0.85	0.70	0.70	−0.84
	(1.54)	(1.54)	(1.12)	(1.12)	(1.22)	(1.22)	(0.59)
Farmer-season FE	Yes		Yes		Yes		Yes
Sample:	Excluding first choice plots of farmers in the choice group						Insured plots
Mean of non-first choice plots	22.2	22.2	15.0	15.0	7.3	7.3	1.4
Mean of non-insured plots	23.5	23.5	15.8	15.8	7.7	7.7	
Num FE's	695	695	695	695	695	695	492
Observations	1739	1739	1739	1739	1739	1739	691

This table shows separate estimation adverse selection and moral hazard using estimation equation 6. In the first 6 columns I exclude first choice plots of farmers in the choice group, while the last column includes only insured plots. Damages reported in the first 6 columns are based on self reports from a follow-up survey, while payouts in the last column are based on administrative data.. The identification of the selection coefficient in the last column is based on the 92 farmers who had both their first choice plot and at least one other plot insured. Standard errors are corrected for spatial dependence using the method developed by Conley (1999).

Table 6: Robustness of the Adverse Selection and Moral Hazard Estimates on Typhoon and Flood Damage

	Harvest losses due to typhoons and floods measured as:					
	Raw Proportion		Proportion Winsorized at top 2.5%		Log of the Proportion	
	Sample: Full					
First-choice	2.46 ∗ ∗∗	0.82	2.42 ∗ ∗∗	0.79	0.19 ∗ ∗∗	0.057
	(0.82)	(0.82)	(0.81)	(0.82)	(0.068)	(0.069)
Insurance	−0.53	−0.81	−0.58	−0.86	−0.054	−0.077
	(0.78)	(0.77)	(0.78)	(0.77)	(0.065)	(0.065)
Flooding index		5.00 ∗ ∗∗		4.98 ∗ ∗∗		0.35 ∗ ∗∗
		(0.69)		(0.69)		(0.051)
High risk from winds		2.47		2.44		−0.053
		(2.63)		(2.63)		(0.18)
High risk of rats		−3.57 ∗ ∗∗		−3.60 ∗ ∗∗		−0.32 ∗ ∗∗
		(1.29)		(1.28)		(0.11)
High risk of tungro		0.083		0.078		0.19
		(1.87)		(1.86)		(0.15)
Area (hectares, centered)	0.83	1.33	0.83	1.33	0.033	0.070
	(1.12)	(1.08)	(1.12)	(1.07)	(0.11)	(0.10)
Mean of control plots	14.9	14.9	14.9	14.9	1.7	1.7
Num FE's	696	696	696	696	696	696
Observations	1744	1744	1744	1744	1744	1744
	Sample: Omitting top and bottom 2.5% of the potential harvest distribution					
First-choice	2.01 ∗ ∗	0.43	2.00 ∗ ∗	0.43	0.16 ∗ ∗	0.027
	(0.84)	(0.83)	(0.83)	(0.83)	(0.069)	(0.070)
Insurance	−0.26	−0.53	−0.25	−0.53	−0.041	−0.063
	(0.81)	(0.79)	(0.80)	(0.79)	(0.067)	(0.067)
Flooding index		4.82 ∗ ∗∗		4.80 ∗ ∗∗		0.35 ∗ ∗∗
		(0.68)		(0.67)		(0.051)
High risk from winds		0.40		0.37		−0.14
		(2.16)		(2.15)		(0.17)
High risk of rats		−4.50 ∗ ∗∗		−4.48 ∗ ∗∗		−0.39 ∗ ∗∗
		(1.38)		(1.36)		(0.12)
High risk of tungro		−0.019		−0.053		0.19
		(1.87)		(1.86)		(0.16)
Area (hectares, centered)	1.29	1.75∗	1.28	1.74∗	0.100	0.13
	(1.10)	(1.04)	(1.09)	(1.04)	(0.11)	(0.10)
Mean of control plots	14.6	14.6	14.6	14.6	1.7	1.7
Num FE's	685	685	685	685	685	685
Observations	1659	1659	1659	1659	1659	1659

The table explores the robustness of the adverse selection and moral hazard findings on typhoon and flood damage. In this table I use all available data, including the outliers discussed in Section 5.2. The three outcome variables are: (1) the raw damage ratio; (2) the damage ratio constructed by winsorizing the damages and harvest the 97.5th percentile before constructing the ratio; and (3) the inverse hyperbolic sine transformation of the damage ratio. The inverse hyperbolic sine transformation ($f(x) = log(x + \sqrt{(x^2+1)})$; see e.g., Burbidge, Magee and Robb (1988)) can be interpreted in a similar way as a log transformation but has the advantage of being defined and differentiable at zero. In the lower panel I exclude plots in the top and bottom 2.5% of the counterfactual harvest distribution (harvest plus total damages) per hectare. The regressions include farmer-season fixed effects. Standard errors are corrected for spatial dependence using the method developed by Conley (1999).

Table 7: Robustness of the Adverse Selection and Moral Hazard Estimates on Pest and Crop Disease Damage

	Harvest losses due to pests and crop diseases measured as:					
	Raw Proportion		Proportion Winsorized at top 2.5%		Log of the Proportion	
			Sample: Full			
First-choice	$1.98***$	$1.98***$	$1.98***$	$1.99***$	$0.25***$	$0.27***$
	(0.67)	(0.69)	(0.66)	(0.67)	(0.065)	(0.066)
Insurance	$1.26*$	$1.24*$	$1.30**$	$1.29**$	$0.12*$	$0.12*$
	(0.66)	(0.67)	(0.65)	(0.65)	(0.063)	(0.064)
Flooding index		0.15		0.10		−0.022
		(0.49)		(0.46)		(0.044)
High risk from winds		2.25		2.27		$0.30*$
		(1.44)		(1.45)		(0.17)
High risk of rats		0.47		0.49		0.13
		(1.08)		(1.03)		(0.099)
High risk of tungro		0.42		0.46		0.092
		(1.64)		(1.62)		(0.14)
Area (hectares, centered)	0.64	0.66	0.74	0.76	0.10	0.10
	(1.23)	(1.23)	(1.21)	(1.21)	(0.11)	(0.11)
Mean of control plots	7.5	7.5	7.3	7.3	1.0	1.0
Num FE's	696	696	696	696	696	696
Observations	1744	1744	1744	1744	1744	1744
			Sample: Omitting top and bottom 2.5% of the potential harvest distribution			
First-choice	$2.13***$	$2.19***$	$2.07***$	$2.14***$	$0.26***$	$0.28***$
	(0.67)	(0.69)	(0.66)	(0.67)	(0.067)	(0.068)
Insurance	$1.40**$	$1.40**$	$1.41**$	$1.41**$	$0.14**$	$0.14**$
	(0.68)	(0.68)	(0.66)	(0.66)	(0.065)	(0.066)
Flooding index		0.037		−0.032		−0.041
		(0.50)		(0.48)		(0.045)
High risk from winds		1.97		1.87		0.25
		(1.57)		(1.57)		(0.18)
High risk of rats		1.34		1.22		$0.19*$
		(1.07)		(1.02)		(0.100)
High risk of tungro		0.58		0.64		0.12
		(1.52)		(1.49)		(0.14)
Area (hectares, centered)	−0.10	−0.12	−0.029	−0.050	0.059	0.052
	(1.24)	(1.24)	(1.22)	(1.22)	(0.12)	(0.12)
Mean of control plots	7.5	7.5	7.4	7.4	1.0	1.0
Num FE's	685	685	685	685	685	685
Observations	1659	1659	1659	1659	1659	1659

The table explores the robustness of the adverse selection and moral hazard findings on pest and crop disease damage. In this table I use all available data, including the outliers discussed in Section 5.2. The three outcome variables are: (1) the raw damage ratio; (2) the damage ratio constructed by winsorizing the damages and harvest the 97.5th percentile before constructing the ratio; and (3) the inverse hyperbolic sine transformation of the damage ratio. The inverse hyperbolic sine transformation ($f(x) = log(x + \sqrt{(x^2+1)})$; see e.g., Burbidge, Magee and Robb (1988)) can be interpreted in a similar way as a log transformation but has the advantage of being defined and differentiable at zero. In the lower panel I exclude plots in the top and bottom 2.5% of the counterfactual harvest distribution (harvest plus total damages) per hectare. The regressions include farmer-season fixed effects. Standard errors are corrected for spatial dependence using the method developed by Conley (1999).

Table 8: Empirical Estimates of a Selection Model Based on Predicted Damages

	Outcome: Indicator for first choice plot	
	Odds-ratio [90% CI] (p-value)	Odds-ratio [90% CI] (p-value)
Predicted Total Damages:		
If Insured	1.08 * **	
	[1.04, 1.12]	
	0.00026	
If Not Insured		1.08 * **
		[1.04, 1.12]
		0.00031
Difference (Insured - Not Insured)		1.07*
		[1.01, 1.13]
		0.073
Land Arrangement (ref. = Sharecropping)		
Owner of plot	2.09 * *	2.10 * *
	[1.18, 3.67]	[1.20, 3.66]
Land under fixed rent	0.84	0.85
	[0.47, 1.51]	[0.45, 1.62]
Land mortaged in	0.61	0.62
	[0.31, 1.22]	[0.31, 1.22]
Land lent for free	1.53	1.56
	[0.56, 4.22]	[0.55, 4.42]
Other predictors of choice		
Plot size	7.55 * **	7.48 * **
	[4.17, 13.7]	[4.15, 13.5]
Number of farmer-seasons	484	484
Observations	1259	1259

Columns 1 and 3 present empirical estimates of Equation (11). The coefficients reported are odds ratios and in the brackets below I report 90% confidence intervals. I use the full sample for the choice models but predict damages only using the sample of farmers in the full randomization group (except that I exclude the data from the first season since the plot characteristics were not collected in that season). All models are calculated using a cluster bootstrap procedure (clustering at the farm-season level) with 500 repetitions of the prediction and choice estimation (to account for the fact that the predicted damages are computed values).

Table 9: Estimation of Predicted Damages by Treatment Group

	Loss Due to All Causes	
	Estimate	(SE)
Season 2 X Flooding index	6.97	(2.37)
Season 2 X Medium risk of rats	−3.99	(5.41)
Season 2 X High risk of rats	−12.5	(5.58)
Season 2 X Medium risk of tungro	0.14	(4.39)
Season 2 X High risk of tungro	1.45	(8.71)
Season 3 X Flooding index	3.21	(2.09)
Season 3 X Medium risk of rats	0.94	(5.02)
Season 3 X High risk of rats	−0.30	(5.28)
Season 3 X Medium risk of tungro	−2.71	(4.35)
Season 3 X High risk of tungro	−0.70	(6.81)
Insurance X Season 2 X Flooding index	−2.61	(2.47)
Insurance X Season 2 X Medium risk of rats	6.70	(5.18)
Insurance X Season 2 X High risk of rats	9.99	(6.48)
Insurance X Season 2 X Medium risk of tungro	−4.15	(4.51)
Insurance X Season 2 X High risk of tungro	1.47	(9.54)
Insurance X Season 3 X Flooding index	0.59	(2.79)
Insurance X Season 3 X Medium risk of rats	1.40	(5.14)
Insurance X Season 3 X High risk of rats	2.20	(5.71)
Insurance X Season 3 X Medium risk of tungro	3.96	(4.59)
Insurance X Season 3 X High risk of tungro	1.47	(7.59)
Insurance	−4.86	(3.69)
Constant	20.2	(3.71)
Observations	1259	

This table reports coefficient estimates for a prediction model that predicts damages based on plot characteristics and insurance status. The predicted values from this regression are used in the selection model that I report on in Table 8 (these are estimates for the sample used to estimate the selection models; the bootstrap procedure used to estimate those models re-estimates this prediction model for each bootstrap sample). The outcome variable is total damages divided by the sum of total damages and harvest (I use this outcome variable instead of the damage measures used before to avoid using plot area as a part of the prediction model). The plot characteristics used are described in Section 5.2. The standard errors in Column 2 are estimated using OLS (this is only an illustration, as mentioned before, the prediction model is re-estimated for each bootstrap sample to estimate the selection model).

A Appendix Tables

Table 10: Farmer level intent-to-treat sample

ITT Status	Season 1	Season 2	Season 3	Total
Informed: In ITT sample	107	279	443	829
Not informed because already insured	1	3	0	4
Sick/dead/moved before he/she was informed	0	2	0	2
Mistake in enrollement / Not eligible	0	1	3	4
Total	108	285	446	839

This table reports on the construction of the ITT sample (number of farmers in each cell). A few farmers were randomized but never informed of their randomization allocation and didn't receive insurance through the experiment.

Table 11: Farmer level attrition in the experiment

	Season 1		Season 2		Season 3		
Dropout Status	CT	TX	CT	TX	CT	TX	Total
Not dropped: in final analysis sample	24	62	61	150	94	307	698
Died/sick/moved	0	0	4	2	4	4	14
Refused	5	3	7	12	6	11	44
Did not plant this season	1	1	2	5	1	6	16
Could not give output on any plot	0	0	5	11	5	5	26
Could not give any data on damages	0	2	2	14	0	0	18
Unknown	5	4	2	2	0	0	13
Total	35	72	83	196	110	333	829
Comparison	$p = 0.53$		$p = 0.92$		$p = 0.69$		

This table reports attrition (number of farmers) by season and (farmer level) treatment status (CT = Control, TX = Treatment). The last row presents p-values from a Chi-square test of the difference in attrition rates across treatment and control groups (for each phase).

Table 12: Plot-level attrition in the experiment

Dropout Status	Season 1 CT	Season 1 TX	Season 2 CT	Season 2 TX	Season 3 CT	Season 3 TX	Total
In sample	63	70	188	184	402	403	1,310
Could not report output or damages	7	9	25	18	37	35	131
Is worker on this plot, not farmer	0	0	0	1	3	1	5
Other	0	0	5	5	0	0	10
Total	70	79	218	208	442	439	1,456
Comparison	p = 1.00		p = .89		p = .96		

This table reports attrition (number of plots) by season and plot-level treatment status for plots included in the plot randomization (that is, excluding the pure control group and excluding first choice plots of farmers in the choice group), excluding plots of farmers that dropped out completely. The last row presents p-values from a Chi-square test of the difference in attrition rates across treatment and control groups (for each phase).

B Share of Adverse Selection Predicted by Baseline Characteristics

In Table 13 I estimate equations of the form

$$D_{ij} = \beta_0 + \beta_1 C_{ij} + \beta_2 X_{ij} + \beta_3 X_{ij} 1(\text{dry season}) + \lambda_i + \eta_{ij}. \tag{17}$$

I only use seasons 2 and 3 for this estimation since the relevant baseline characteristics were not collected in the first season. The estimated selection effect in Column 2 is 30% lower than in Column 1, where β_2 and β_3 are constrained to zero suggesting the observables explain a significant part of the observed selection effect.

Table 13: Estimated Share of Adverse Selection Explained by Baseline Characteristics

	Loss (%) Due to: All causes			
	Eq. (1)		Eq. (2)	
	Estimate	(SE)	Estimate	(SE)
First-choice	4.03 * **	[1.24]	2.84 * **	[1.24]
Plot size	21.3 * **	[3.68]	21.5 * **	[3.58]
Season 2 X Flooding index			2.55 * **	[1.22]
Season 2 X Medium risk of rats			4.44*	[4.36]
Season 2 X High risk of rats			1.28	[3.38]
Season 2 X Medium risk of tungro			4.64 * **	[2.66]
Season 2 X High risk of tungro			3.56	[5.31]
Season 3 X Flooding index			2.95 * **	[1.20]
Season 3 X Medium risk of rats			−1.09	[3.19]
Season 3 X High risk of rats			−0.63	[3.45]
Season 3 X Medium risk of tungro			1.30	[2.76]
Season 3 X High risk of tungro			3.35	[4.38]
Constant	13.3 * **	[0.46]	11.6 * **	[2.32]
F-test All Risk Characteristics				p = 0.00
Mean of dependent variable for non-first choice plots	12.4		12.4	
Num FE's	483		483	
Observations	1259		1259	

The table only includes data from Season 2 and 3 since these baseline characteristics were not collected in Season 1.

C Additional Results on the Model

Given the uniform distribution for the share of harvest lost, S, we have: $E\left[S_j|e_j\right] = \frac{1}{2}\theta_j(1-e_j)$ and $Var\left[S_j|e_j\right] = \frac{1}{12}\theta_j^2(1-e_j)^2$. This implies that $E\left[\Pi|(\alpha,\mathbf{e})\right] = \sum_{j=1}^{N} A_j\left(1-\frac{1}{2}(1-\alpha_j L)\theta_j(1-e_j)\right) - C(\mathbf{e})$ and $Var\left[\Pi|(\alpha,\mathbf{e})\right] = \frac{1}{12}\sum_{j=1}^{N} A_j^2(1-\alpha_j L)^2\theta_j^2(1-e_j)^2$. 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:

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

C.1 Optimal Effort

Optimal effort given insurance coverage is given by the solution to:

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

The first order condition for effort is

$$\begin{aligned}\frac{\partial C}{\partial e_j} &= A_j\left[\frac{(1-\alpha_j L)\theta_j}{2} - \rho(1-\tau)A_j(1-\alpha_j L)^2\theta_j^2\frac{-2(1-e_j)}{12}\right] \\ &= A_j(1-\alpha_j L)\theta_j\left[\frac{1}{2} + \rho(1-\tau)A_j(1-\alpha_j L)\theta_j\frac{1-e_j}{6}\right] \\ &= W_j\left[1+\rho(1-\tau)W_j\frac{2(1-e_j)}{3}\right] = 0\end{aligned}$$

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).[47] This amount will be prominent in the calculations below and I use it as a convenient shorthand.

[47] The second order condition for a local maximum is $-\rho(1-\tau)\frac{W_j^2}{6} < 0$, which is always satisfied.

The first order condition for effort implies that:

$$
\begin{aligned}
& A_j\psi_j = W_j\left[1+\rho(1-\tau)W_j\frac{2(1-e_j)}{3}\right] \\
\Leftrightarrow \quad & \rho(1-\tau)W_j\frac{2(1-e_j)}{3} = \frac{A_j\psi_j}{W_j} - 1 \\
\Leftrightarrow \quad & 1-e_j = \frac{3A_j\psi_j}{2\rho(1-\tau)W_j^2} - \frac{3}{2\rho(1-\tau)W_j} \\
\Leftrightarrow \quad & e_j = 1 - \frac{3A_j\psi_j}{2\rho(1-\tau)W_j^2} + \frac{3}{2\rho(1-\tau)W_j} \\
\Leftrightarrow \quad & e_j = 1 - \frac{3}{2}\frac{\psi_j - w_j}{\rho(1-\tau)A_j w_j^2}
\end{aligned} \tag{18}
$$

For an interior solution we must have $e_j \in (0,1)$. This implies for an interior solution we must have $e_j > 0 \Leftrightarrow \psi_j < w_j + \frac{2}{3}\rho(1-\tau)A_j^2 w_j^2$ and $e_j < 1 \Leftrightarrow w_j < \psi_j$. This implies that optimal effort is given by:

$$
\hat{e}_j(\alpha_j,\theta_j,\psi_j,A_j,\rho,\tau) = \begin{cases} 0 & \text{if } \psi_j \geq w_j + \frac{2}{3}\rho(1-\tau)A_j w_j^2 \\ 1 - \frac{3}{2}\frac{\psi_j - w_j}{\rho(1-\tau)A_j w_j^2} & \text{if } w_j < \psi_j < w_j + \frac{2}{3}\rho(1-\tau)A_j w_j^2 \\ 1 & \text{if } \psi_j \leq w_j \end{cases} \tag{19}
$$

Figure 2 depicts this function for a given plot, both when it is insured and when it is not insured.

C.2 Value Function for Insurance Choice

Given the optimal effort $\hat{e}_j(\alpha_j,\theta_j,\psi_j,A_j,\rho,\tau)$, the utility output on plot j is:

$$
\begin{aligned}
u_j(\alpha_j,\theta_j,\psi_j,A_j,\rho) = A_j & - A_j w_j(1-\hat{e}(\alpha_j,\psi_j,\rho,A_j,w_j)) \\
& - \frac{\rho(1-\tau)}{3}A_j^2 w_j^2 (1-\hat{e}(\alpha_j,\psi_j,\rho,A_j,w_j))^2 \\
& - A_j\psi_j\hat{e}(\alpha_j,\psi_j,\rho,A_j,w_j)
\end{aligned} \tag{20}
$$

Therefore, the correct value function for insurance choice used by the fully sophisticated farmer is $V(\alpha) = \sum_{j=1}^{N} u_j(\alpha_j,\theta_j,\psi_j,A_j,\rho)$ and her 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 the utility output (equation 20) and substituting u_j^{myopic} for u_j in the value function.