Perceptions and Participation: Mistaken Beliefs, Encouragement Designs, and Demand for Index Insurance.

Conner Mullally
PhD Candidate
Department of Agricultural and Resource Economics
University of California, Davis
mullally@primal.ucdavis.edu

Steve Boucher
Associate Professor
Department of Agricultural and Resource Economics
University of California, Davis

Michael Carter
Professor
Department of Agricultural and Resource Economics
University of California, Davis

Selected Paper prepared for presentation at the Agricultural & Applied Economics Association 2010
AAEA, CAES, & WAEA Joint Annual Meeting, Denver, Colorado, July 25-27, 2010.

Abstract: Index insurance is an alternative to crop insurance that bases payouts on a variable that is highly correlated with income but beyond the control of individual households, such as rainfall, temperature, or area-yields (i.e., output per hectare in a large area). Index insurance offers risk protection while avoiding the incentive problems that plague traditional crop insurance, and as a result is seen by many as a promising antipoverty tool. However, participation in index insurance to this point has generally been in low. In this paper, we explore one factor that might potentially depress demand for index insurance: mistaken beliefs among farmers with respect to the distribution of the insured risk. Our experiences from an area-yield insurance pilot project for cotton farmers in the Pisco valley of Peru suggest that such errors are common. Lower demand means that not only are the benefits of index insurance not realized by households, but econometric measurement of the these benefits is more difficult due to low precision. One way to counteract these effects is a "randomized encouragement design," i.e., the random assignment of positive economic incentives for the purchase of area-yield insurance, such as discount coupons. We examine the implications of a randomized encouragement design for econometric evaluation.

Introduction

Much of current applied microdevelopment research is concerned with evaluating the impacts of programs on households: e.g., impacts on investment, consumption, or credit demand. A great deal has been written about how to recover consistent estimates of these impacts in the context of the classical evaluation and selection bias problems. The first arises because in any given state of nature, we can only observe households participating or not participating in a program, and therefore cannot directly observe household outcomes when they are both in and out of a program. The selection bias problem arises when households self-select into a program, possibly driven by unobserved factors that are also correlated with outcomes we would like to measure; this results in biased estimates of impacts. Less has been written about the nature of the household participation decision and its implications for the design and evaluation of new programs. In many cases, programs are complex, and the difficulties associated with properly evaluating their benefits can give rise to cognitive errors that depress participation. The implications of low participation are weaker program benefits for the target population as a whole and greater difficulty in estimating impacts, due to imprecise econometric estimates.

Giving individuals better economic incentives for participation or exposure to correct information may be necessary to counteract these problems. In what follows, we explore these issues in the context of index insurance, drawing from our experiences in a pilot program designed to explore the impacts of index insurance on small cotton farmers in coastal Peru. Our view is that while exposure to correct information regarding program benefits can increase participation, economic incentives or "encouragements" may be necessary. Using randomly assigned incentives to predict program participation, or a "randomized encouragement design," can cause program benefits to be more widely realized and make econometric estimates of these benefits possible. We summarize what we have observed in the field with respect to the presence of cognitive errors that affect household participation decisions. We use a simple economic model to demonstrate the effects of one feasible type of error on program takeup and activity choice. We then link our economic model to econometric program evaluation. Randomized encouragement designs (or any instrumental variables strategy) are sometimes viewed as less than ideal from a research perspective, as resulting estimates of impacts may not describe effects on the original population of interest. However, in the absence of encouragement, participation may be so low as to make inference impossible. We demonstrate these tradeoffs by placing our analytical model in an econometric framework.

Index insurance in the Pisco valley of Peru

Agriculture all over the world is subject to risk. What makes the situation in developing economies different is the absence of well-functioning markets to mitigate the effects of risk. Informal mechanisms exist, but available empirical evidence suggests they are either inadequate or very costly (Rosenzweig and Binswanger, 1993; Dercon, 2004, Carter et al., 2007).

Traditional crop insurance, usually based on compensating farmers for losses relative to a farm-specific historical level of output or income, has proven extremely expensive in the countries where it is offered. Covering all sources of risk without the ability to perfectly monitor farmer behavior leads to incentive problems for policyholders, and traditional crop insurance has typically seen high ratios of payouts to premium as a result (Skees et al., 1999). This makes traditional insurance a less than desirable policy option for governments in developing countries. Index insurance is an alternative to traditional crop insurance which ties payouts to a variable that is highly correlated with yields or income, but beyond the control of individual farmers. For example, an index insurance contract based on rainfall might pay policyholders if precipitation were to fall below a certain level, with payouts increasing in the size of the shortfall. Since it does not cover all risk at the individual level, index insurance will not offer protection to the same extent as traditional crop insurance, but it avoids the incentive problems which have made such programs costly and unsustainable.

While the potential of index insurance is strong, empirical evidence of demand for index insurance among poor farmers and impacts of index products on farmer behavior is scarce. In order to address this, our research team began a pilot project in the Pisco valley of coastal Peru in 2008, consisting of the design, marketing, and impact evaluation of an index insurance product for cotton farmers. The project is a cooperative effort between UC Davis, the Instituto de Estudios Peruanos in Lima, Swiss reinsurer Swiss Re, Peruvian insurer La Positiva, and Caja Señor de Luren, the largest formal lender to cotton farmers in Pisco. Cotton is by far the dominant crop in the Pisco valley; Peruvian Ministry of Agriculture data show it accounting for 50-75 percent of sown area over the last decade, while it is planted by an average of 60 percent of Pisco's roughly 5,000 farmers in any given year. Cotton farms tend to be small, with farmers owning an average of 5.15 hectares. While cotton is fairly robust to potential sources of yield risk, output can be strongly affected by the El Niño phenomenon, as well as other shocks. El Niño occurrences can affect the availability of water for local irrigation infrastructure, and create temperature fluctuations that affect levels of pest populations and the rate at which the cotton plant develops. El Niño occurrences in 1992-1993 and 1997-1998 resulted in large drops in cotton productivity in Pisco, as shown in Figure 1:

[INSERT FIGURE 1 HERE]

El Niño is predictable to some extent, as occurrences happen every 3-7 years, but not perfectly so. The co-movement between sown area and productivity in Figure 1 suggests that farmers are able to anticipate covariate shocks to some degree, although the two do not track each perfectly. Other less predictable sources of covariate risk exist as well. The closure of the state agricultural bank by the government of Alberto Fujimori in 1992 created a tightening of access to credit among farmers that was only partly eased over time by expansion of private formal lenders into agriculture (Trivelli and Venero, 2007). In 2007 and 2008, world fertilizer prices reached record highs, sharply increasing the cost of a key input into cotton production. Without detailed time series data on household income and consumption, we cannot know exactly how these shocks translate into drops in consumption or asset levels. But strong covariate shocks do exist, and reliance on cotton in the region suggests that there could be large benefits to insuring farmers against systemic shocks to cotton output. The fact that covariate shocks may come from climatic as well as macroeconomic factors suggests that the best approach would be an index product that can mitigate risk from a variety of factors.

With these concerns in mind, we designed an area-yield insurance (AYI) product for Pisco. If average output per hectare (i.e., or area-yields) falls below the “strike point” of 85 percent of the long term mean in the valley, all policy holders receive a payout equal to the size of the short fall times the previous year's market price in Pisco for cotton. The insurance product was initially priced using a detrended 20 year time series of data from the Peruvian Ministry of Agriculture. This data set and the insurance premium are currently updated every year using our own yield survey, the results of which also determine whether or not a shortfall from the strike point has occurred. Conducting our own yield survey guarantees timely payouts to policyholders, and will eventually allow us to study how yields vary across the valley at a more disaggregated level than what is permitted by ministry data.

Demand for Agropositiva, as the area-yield product is called, has been minimal to this point. A lack of sales in the first year can be attributed in part to administrative difficulties that led to Agropositiva not being made available until late in the planting season. Low demand persisted into the second year, however, despite widespread publicity efforts on the part of La Positiva, the insurer marketing the contract. As part of our impact evaluation efforts, we carried out a survey of farm households in the Pisco valley in 2008 and 2009. Among households in our sample that farmed cotton in 2009, only 3.6 percent purchased Agropositiva (21 of 583). Randomly offered discounts on the premium, in some cases large enough to make the product actuarially fair, had no effect on demand. While we anticipated low initial takeup of AYI, sales have fallen far short of what we had expected.

Whether the benefits of index insurance are realized requires that farmers correctly evaluate its benefits. At a superficial level, one of the good qualities of index insurance is its simplicity. But consider what a farmer must understand in order to make an informed purchase decision. Firstly, there is the indemnity function. Farmers must understand how realizations of the insured risk translate into indemnity payments, and how the size of these payments is affected by the level of coverage purchased. Secondly, there is the distribution of the insured risk, and by extension, the distribution of the indemnity function. For example, purchasing insurance will involve a sacrifice of expected returns in exchange for lower variability of income, unless premiums are highly subsidized; understanding the expected payouts associated with an index insurance contract is necessary to properly evaluate this tradeoff. Lastly, farmers need to understand co-movement of their individual incomes or crop yields and the indemnity function. If a farmer incorrectly perceives the co-movement between her income and the index, she will not be able to properly evaluate the benefits of index insurance, regardless of how well she understands the distribution of outcomes she faces at the individual level.

There are then several potential ways in which farmers can err and improperly value the benefits of index insurance, and potentially depress demand. The possibility for errors is exacerbated by low levels of education and lack of experience with the formal financial sector. We did not fully appreciate these possibilities when we began our own index insurance research program, although we did make strong efforts to educate farmers about Agropositiva and AYI in general. Prior to the first year of Agropositiva's availability, we invited randomly selected farmers to participate in experimental "games" designed to simulate an AYI contract for the Pisco valley; 400 farmers ended up attending, around 50 percent of those invited.[1] The games included real monetary incentives. Farmers played for a total of 12 rounds or "seasons" with average winnings amounting to two days pay for an agricultural worker. Results of questions meant to gauge the level of comprehension of AYI following the games were our first clue that cognitive errors might depress demand for AYI. During the games, farmers were arranged into groups or "valleys." At the start of each round, farmers chose from three options: a low risk activity mean to simulate cotton farming via self-financing, a riskier option mimicking cotton

[1] See Galarza (2009) for a detailed description of the games and results.

farming with a formal loan, and a third option that bundled the loan with AYI. In each round of the game, participants faced two sources of risk to their winnings: an idiosyncratic shock that was specific to each person, and a covariate shock that was shared by every member of his or her group; purchasing AYI mitigated the risk of the covariate shock but forced farmers to sacrifice a portion of expected returns. While it was repeatedly emphasized that AYI was in no way correlated with the idiosyncratic risk, over 25 percent of farmers responded incorrectly when asked if the individual risk affected payouts.

Informal discussions following the games revealed other potential difficulties. Farmers whose average yields were higher than those of the valley as a whole were skeptical of the insurance, stating that their own yields never dropped that low, and thus could not foresee Agropositiva ever paying out. Even farmers with average yields around the valley level mean stated that there was little chance of area-yields ever dropping below the strike point, and viewed the premium as much too high given this low perceived probability. This is despite the fact that all seemed to agree that the bad El Niño years shown in Figure 1 were devastating in terms of production losses. Farmers seemed to hold the contradictory views that El Niño was very harmful to production, yet yields could never fall below the strike point.

In order to delve deeper into these issues, we conducted focus group discussions with farmers following the second year of the project. Farmers continued to insist that the strike point was far too low relative to yields in the valley, with this view most pronounced among individuals with highly productive land. Other issues came to the forefront as well. The Pisco valley is thought of by farmers and agricultural extension workers as consisting of three zones: la parte alta, la parte media, and la parte baja. Farmers in the focus groups insisted that production conditions in the three zones were substantially different, and that la parte baja in particular was characterized by lower productivity and different sources of risk than the other two zones. Given this perceived dissimilarity of production conditions, farmers did not see an index that combined yields in all three zones as useful from a risk management standpoint.

Given that there is only one lengthy time series for cotton yields in Pisco, it is difficult to evaluate claims made by farmers with respect to the size of the strike point and the mean of area-yields. But we can evaluate claims with respect to heterogeneity of production across the valley, as in Figure 2:

[INSERT FIGURE 2]

This figure was drawn using Ministry of Agriculture data through 2006, while remaining years are drawn from data we collected. Data earlier than 1987 are only available at the valley level, and therefore we can only look at this shorter time series. Yields in the three zones have very strong co-movement. All correlation coefficients between and any two of the three zones are higher than 0.91. The lone exception in this pattern of similar outcomes appears to be 2007, where area-yields in the parte baja where substantially lower than elsewhere. But the overall trend of similar productivity levels is clear.

The household survey provides additional evidence against the perception of strong heterogeneity across the three zones. We asked all households their yields per hectare in a normal year, a bad year, and a good year. The results are averaged within each zone in Table 1:

[INSERT TABLE 1]

T-tests show no statistically significant differences across zones for output in normal, bad, and good years. These figures were computed by weighting each household's self-reported figure by the percentage of all land in the zone owned by that household, rather than weighting by a measure of share of sown area. The rental market in Pisco is quite active, and weights based on an average share of sown area might yield different results. In any case, the combination of Figure 2 and Table 1 suggest that there are no glaring differences in productivity or variability of production across the three zones. Good, bad, and normal production conditions appear to occur in the same years for all three zones, and output associated with these conditions seem to be very similar in each region.

Multiple data sources therefore appear to contradict claims made by farmers with respect to risk and cotton production in Pisco, and despite intensive efforts to provide accurate information with respect to the benefits of AYI, these disparities persist. The question is what the implications of these errors might be. In what follows, we model the impacts of one particular type of error on demand for AYI and activity choice, and then later explore how this error might affect econometric analysis of the impacts of AYI on households that purchase it. The cognitive error we use is misperception of the mean of area-yields. Farmers are assumed to believe that the mean is higher than it actually is, and as a result view the AYI strike point as being a smaller percentage of the mean of area-yields than what is actually the case. As described above, farmers have repeatedly told us that area-yields will "never" fall below the strike point of 85 percent of the mean, suggesting that the error used in our model is a feasible one; we do not allow for any other type of cognitive error.

A model of demand for index insurance and beliefs about the distribution of the insured risk

The economy consists of N farmers, each with the option of participating in wage labor or planting cotton on a single hectare of land. Participating in wage labor guarantees a risk free return of w. Cotton is a risky activity, as output in each period is vulnerable to a shock ε_c that is common to all N households, and an idiosyncratic shock ε_i. In this model, we will define the common shock ε_c as deviations of average output per hectare, or area-yields, from their mean:

$$\varepsilon_{ct} = \mu - \overline{q_t} \tag{1}$$

We will hold the price of the crop fixed at unity, making output and income identical. The cotton production function for farmer i at time t is:

$$q_{it} = \mu - \beta_i \varepsilon_{ct} - \varepsilon_{it} \tag{2}$$

This production function is taken from Miranda (1991). The shocks ε_c and ε_i are continuous, jointly independent, and symmetrically distributed about a mean of zero. Both terms are identically distributed across farmers, although for our purposes ε_i need not be homoscedastic. While all households have the same mean level of output μ, the variance of cotton is heteroscedastic. From (2), the variance of output can be written as:

$$\sigma_{q_i}^2 = \beta_i^2 \sigma_{\varepsilon_c}^2 + \sigma_{\varepsilon_i}^2 \tag{3}$$

Total variance of yields can be decomposed into a component due to variation in area-yields, $\beta_i^2 \sigma_{\varepsilon_c}^2$, and another due to all other sources of risk, $\sigma_{\varepsilon_i}^2$.

The heteroscedasticity of q is driven by differences in sensitivity to the covariate shock ε_c, as captured by β_i. The parameter β_i is equal to the covariance of farmer yields q_i with the covariate shock ε_c, divided by the variance of ε_c:

$$\beta_i = \frac{\sigma_{q_i,\varepsilon_c}}{\sigma_{\varepsilon_c}^2} \tag{4}$$

To see this, consider running regressions by household in which the estimating equation is given by (2). Taken together, the individual β_i parameters form the population level distribution of β, which is symmetric and centered at unity with a variance of σ_β^2 . In the long run, the value of β_i for each farmer would likely be a choice variable, as pointed out by Chambers and Quiggin (2002). In the present context we will assume these parameters are fixed.

The decision whether or not to plant cotton will be driven by each farmer's ex-ante evaluation of the benefits of planting cotton versus those of wage labor. Here we will assume that farmer preferences follow a mean-standard deviation specification.[2] Denote by $plant_i$=1 the decision to grown cotton by farmer i, and $plant_i$=0 participation in wage labor. Expected utility of planting cotton is:

$$EU_{i,plant_i=1} = \mu^2 - \gamma\left(\beta_i^2 \sigma_{\varepsilon_c}^2 + \sigma_{\varepsilon_i}^2\right) \tag{5}$$

and for wage labor we have:

$$EU_{i,plant_i=0} = w^2 \tag{6}$$

The functional form used in (5) is taken from Nelson and Escalante (2004). Squaring the mean implies that farmer preference are characterized by constant relative risk aversion, where the coefficient of relative risk aversion is given by γ .

Whether or not a farmer decides to plant cotton will be determined by her β_i . Specifically, there is a range of β_i values for which the expected utility of growing cotton is higher than that of wage labor. Setting the difference of (5) and (6) to zero and solving for β_i results in the following decision rule:

---

[2] Normally this would require making assumptions over the distribution of yields or the nature of the ex-post utility function so as to insure that the two-moment representation of preferences is consistent with expected utility or some generalization thereof, as shown by Meyer (1987), Sinn (1983), and Chamberlain (1983). Here, we are less concerned with following the axioms of expected utility than with creating a simple, coherent model. In any case, two-moment preference functions often yield close approximation of expected utility results, even when the criteria for consistency with expected utility are not satisfied (Kroll and Markowitz, 1984, Garcia et al., 1994).

$$plant_i = 1 \leftrightarrow -\sqrt{\frac{\mu^2 - \gamma\sigma_{\varepsilon_i}^2 - w^2}{\gamma\sigma_{\varepsilon_c}^2}} < \beta_i < \sqrt{\frac{\mu^2 - \gamma\sigma_{\varepsilon_i}^2 - w^2}{\gamma\sigma_{\varepsilon_c}^2}} \tag{7}$$

Farmers with values of β_i outside of this interval will participate in wage labor.

The area-yield insurance contract

Farmers would be more inclined to move out of wage labor if the risk associated with growing cotton could be reduced, and any such shift would increase expected output in the economy. Area-yield insurance is one way in which this might be accomplished. An AYI contract will consist of an indemnity function and a premium. In this model the indemnity pays farmers the covariate shock whenever this shock is positive, i.e., when area-yields fall short of the mean μ , and zero otherwise:

$$I_t = \max[0, \varepsilon_{ct}] \tag{8}$$

The premium, τ , is equal to the expected indemnity, *r*, plus a loading term, *l*:

$$\tau = r + l = E(I) + l = E(\varepsilon_c \mid \varepsilon_c > 0)P(\varepsilon_c > 0) + l \tag{9}$$

where $E(\varepsilon_c \mid \varepsilon_c > 0)$ is the expected value of the covariate shock, conditional on ε_c being positive.

The expected utility of insurance purchasers will be affected through the expected value of the indemnity, as well as its variance and its covariance with the common shock, ε_c . The variance of the indemnity function is:

$$\sigma_I^2 = (\sigma_{\varepsilon_c}^2 \mid \varepsilon_c > 0)P(\varepsilon_c > 0) + E(I)^2 \tag{10}$$

The covariance of the indemnity with ε_c is:

$$\sigma_{I,\varepsilon_c} = \left((\sigma_{\varepsilon_c}^2 \mid \varepsilon_c > 0) + E(\varepsilon_c \mid \varepsilon_c > 0)^2\right)P(\varepsilon_c > 0) \tag{11}$$

Equations (10) and (11) are derived equations (14') and (16') of the appendix. Along with the value of β_i, the covariance term determines the risk reduction potential of AYI for each individual farmer.

Adding the parameters given in (9), (10), and (11) to the expected utility of planting cotton gives us the expected utility of planting cotton with AYI. Denote by d_i=1 the decision to buy AYI and d_i=0 the decision not to do so. Expected utility conditional on planting cotton and buying AYI is:

$$EU_{i,plant_i=1,d_i=1} = (\mu - l)^2 - \gamma\left(\beta_i^2\sigma_{\varepsilon_c}^2 + \sigma_{\varepsilon_i}^2 + \sigma_I^2 - 2\beta_i\sigma_{I,\varepsilon_c}\right) \qquad (12)$$

Farmers will compare between (5), (6), and (12) when making decisions about activity choice and insurance.

The decision to buy area-yield insurance

In order to identify decision rules for the purchase of AYI, and the possibility of new decision rules with respect to activity choice following the introduction of AYI, we must examine the difference in expected utility between planting cotton with insurance and planting cotton without insurance, as well as the difference in expected utility between planting cotton with insurance and wage labor. The former is:

$$\begin{aligned}\Delta_1 &= EU_{d_i=1,plant_i=1} - EU_{d_i=0,plant_i=1} = \\ &(\mu - l)^2 - \gamma\left(\beta_i^2\sigma_{\varepsilon_c}^2 + \sigma_{\varepsilon_i}^2 + \sigma_I^2 - 2\beta_i\sigma_{I,\varepsilon_c}\right) - \left(\mu^2 - \gamma\left(\beta_i^2\sigma_{\varepsilon_c}^2 + \sigma_{\varepsilon_i}^2\right)\right)\end{aligned} \qquad (13)$$

Ignoring the possibility of wage labor for the moment, for a farmer who was already planting cotton to purchase AYI, the following would have to be satisfied:

$$\beta_i > \frac{(2\mu - l)l + \gamma\sigma_I^2}{2\gamma\sigma_{I,\varepsilon_c}} \qquad (14)$$

Satisfying this inequality guarantees that equation (13) is strictly positive, and therefore that the farmer gains in terms of expected utility from the purchase of AYI.

Now consider the difference between expected utility with AYI and wage labor. This difference is:

$$\Delta_2 = EU_{d_i=1,plant_i=1} - EU_{d_i=0,plant_i=0} = (\mu - l)^2 - \gamma\left(\beta_i^2\sigma_{\varepsilon_c}^2 + \sigma_{\varepsilon_i}^2 + \sigma_I^2 - 2\beta_i\sigma_{I,\varepsilon_c}\right) - w^2 \qquad (15)$$

This is a concave function with respect to β_i with a global maximum. If there any farmers are better off planting cotton and purchasing AYI than participating in wage labor, then there exists at least one value of β_i for which Δ_2 >0. Assume that this is the case. This implies that there is a finite interval of β_i values within which the difference given in (15) is strictly positive. This interval is bounded by the values of β_i that set (15) equal to zero. Solving for these values yields:

$$\Delta_2 > 0 \leftrightarrow$$
$$\frac{\sigma_{I,\varepsilon_c} - x}{\sigma_{\varepsilon_c}^2} < \beta_i < \frac{\sigma_{I,\varepsilon_c} + x}{\sigma_{\varepsilon_c}^2} \qquad (16)$$
$$\text{where } x = \sqrt{\gamma\left(\sigma_{I,\varepsilon_c}\right)^2 + \sigma_{\varepsilon_c}^2\left((\mu - l)^2 - w^2 - \gamma\left(\sigma_{\varepsilon_i}^2 + \sigma_I^2\right)\right)}$$

In order for a farmer to plant cotton and purchase AYI, her β_i values must satisfy (14) and lie within the interval given in (16). In other words, the set of β_i values for which planting cotton and purchasing AYI is optimal is defined by the intersection of (14) and (16). This set will be empty and AYI demand zero if the lower bound given in (14) is greater than the upper bound given in (16). That is, if the maximum value of β_i for which buying AYI and planting cotton gives higher expected utility than wage labor is lower than the smallest value of β_i for which buying AYI and planting cotton gives higher expected utility than cotton alone, no one will buy AYI. All farmers with values of β_i such that cotton and insurance yield higher expected utility than cotton alone would be better off working in wage labor. If this is not the case and the set is non-empty, the β_i values for all farmers buying AYI will obey the following rule:

$$d_i = 1, plant_i = 1 \leftrightarrow \max\left[\frac{(2\mu - l)l + \gamma\sigma_I^2}{2\gamma\sigma_{I,\varepsilon_c}}, \frac{\sigma_{I,\varepsilon_c} - x}{\sigma_{\varepsilon_c}^2}\right] < \beta_i < \frac{\sigma_{I,\varepsilon_c} + x}{\sigma_{\varepsilon_c}^2} \quad (17)$$

where x is defined as in equation (16). As shown in equations (1') through (10') in the appendix, the size of this interval is increasing with respect to the covariance of the indemnity and the covariate shock, σ_{I,ε_c} , decreasing with respect to the loading term, l, and decreasing with respect to the variance of the indemnity, σ_I^2 . Greater risk reduction via a larger covariance term leads to higher participation in AYI, as we would expect. A higher loading term makes AYI more costly, driving demand down, while greater variation in the indemnity lowers demand for insurance by offsetting its risk reduction benefits.

Area-yield insurance and activity choice

Prior to the introduction of AYI, farmers choosing to plant cotton rather than participate in wage labor had values of β_i falling within the bounds given in (7). Once AYI is available, the new set of β_i values which make planting cotton optimal will be given by the union of the intervals in (17) (i.e., the set of β_i values for which planting cotton and purchasing AYI is optimal) and (7). The lower bound of (7) is negative, and since all farmers who benefit from AYI must have β_i>0, it follows that the lower bound of (7) is less than the lower bound of (17). Therefore, the set of β_i values for which planting cotton is optimal is defined as:

$$plant_i = 1 \leftrightarrow -\sqrt{\frac{\mu^2 - \gamma\sigma_{\varepsilon_i}^2 - w^2}{\gamma\sigma_{\varepsilon_c}^2}} < \beta_i < \max\left[\frac{\sigma_{I,\varepsilon_c} + x}{\sigma_{\varepsilon_c}^2}, \sqrt{\frac{\mu^2 - \gamma\sigma_{\varepsilon_i}^2 - w^2}{\gamma\sigma_{\varepsilon_c}^2}}\right] \quad (18)$$

As we would expect, all increases in the share of farmers planting cotton will come from farmers who previously deemed cotton as too risky relative to wage labor, but whose output levels covary positively with area-yields. Whether there is any increase in cotton planting will depend on the upper bound of (18), which cannot be determined without assigning parameter values. AYI must offer enough risk protection to induce farmers with large β_i values to switch to planting cotton. The same changes in the parameters of the AYI contract that increased the size of the set of β_i values for which purchasing AYI and planting cotton were optimal will increase the size of the interval in (18) if its upper bound corresponds to the upper bound given in (17). Otherwise, changes in the AYI contract will have no effect on the number of farmers planting cotton.

Farmer errors and demand for area-yield insurance

If farmers know the different means, variance, and covariance terms detailed above, then any farmer who is better off with AYI will purchase it. Our field experience in Peru suggests that at least in some circumstances, common perceptions may strongly differ from reality. To introduce the possibility that farmers make mistakes with respect to evaluating the insurance contract, we will assume that farmers shift the location of the distribution of $\bar{q}$ to the right, resulting in a higher mean for area-yields:

$$E\left(\bar{q}^*\right) = E\left(\bar{q} + v\right) = \mu^* \quad (19)$$

where v is the size of the location shift and:

$$\mu^* - \mu = g > 0 \quad (20)$$

No other errors are introduced into the model. The * indicates a parameter or random variable as perceived by the farmer rather than its true value, i.e., the random variable according to the subjective distribution held by farmers. This mistake is common to all farmers in the economy. The error shifts the distribution of area-yields to the right, and as a result the definition of the subjective covariate shock ε_c^* is different from that of ε_c, but their distributions are identical, so that:

$$E(\varepsilon_c^*) = E(\mu^* - \bar{q}^*) = 0 \tag{21}$$

and

$$E((\varepsilon_c^*)^2) = \sigma_{\varepsilon_c}^2 \tag{22}$$

Since the distributions of ε_c and ε_c^* are identical, we can use the moments of the former when evaluating the AYI contract under farmer misperceptions with respect to the mean of area-yields. Note that a change in the location of the distribution of $\bar{q}$ will not affect the covariance between output for farmer *i* and the covariate shock ε_c. Individual β_i parameters are therefore unaffected by the error.

It might appear that since the distributions of the true covariate shock ε_c and the subjective covariate ε_c^* shock are identical, the impact of the change in definition of the covariate shock on the AYI contract will be trivial. This is not the case, however, as the size of the error *g* can strongly affect how farmers view the strike point for the AYI contract. Suppose farmers are told that the AYI contract will pay out whenever area-yields fall short of μ, but believe that the mean of area-yields is equal to μ^*. Noting again that the distributions of ε_c and ε_c^* are identical, the perceived indemnity function given farmer errors is:

$$I_t^* = \max[0, \varepsilon_{ct} - g] \tag{23}$$

Mistaken beliefs regarding the mean of area-yields cause farmers to view the AYI contract as having a higher strike point than is actually the case. This leads farmers to perceive a higher loading term. The expected indemnity under the correct distribution of area-yields is:

$$r = E(I) = E(\max[0, \varepsilon_{ct}]) = E(\varepsilon_c \mid \varepsilon_c > 0)P(\varepsilon_c > 0) = \frac{\int_0^{\varepsilon_c^{\max}} \varepsilon_c f(\varepsilon_c) d\varepsilon_c}{P(\varepsilon_c > 0)} P(\varepsilon_c > 0) = \int_0^{\varepsilon_c^{\max}} \varepsilon_c f(\varepsilon_c) d\varepsilon_c \quad (24)$$

where $\varepsilon_c^{\max}$ is the upper bound of the support of ε_c , and $f(\varepsilon_c)$ is its probability density function. In this case we have assumed that ε_c is continuous, but this does not affect our results. Under the subjective distribution of area-yields, (24) becomes:

$$r^* = E(I^*) = E(\max[0, \varepsilon_{ct} - g]) = E(\varepsilon_c \mid \varepsilon_c > g)P(\varepsilon_c > g) - gP(\varepsilon_c > g) = \frac{\int_g^{\varepsilon_c^{\max}} \varepsilon_c f(\varepsilon_c) d\varepsilon_c}{P(\varepsilon_c > g)} P(\varepsilon_c > g) - gP(\varepsilon_c > g) = \int_g^{\varepsilon_c^{\max}} \varepsilon_c f(\varepsilon_c) d\varepsilon_c - gP(\varepsilon_c > g) \quad (25)$$

The difference between (25) and (24) is:

$$\int_0^g \varepsilon_c f(\varepsilon_c) d\varepsilon_c + gP(\varepsilon_c > g) = \int_0^g f(\varepsilon_c)\varepsilon_c d\varepsilon_c + g\int_g^0 f(\varepsilon_c) d\varepsilon_c = h > 0 \quad (26)$$

with

$$\frac{\partial h}{\partial g} = P(\varepsilon_c > g) > 0 \quad (27)$$

Farmer errors lead to a larger perceived loading term, (*l*+*h*), which is an increasing function of the size of the error, *g*.

While the impact on the price of the AYI contract is clear, a higher subjective mean of area-yields does not unequivocally reduce the risk-reduction potential of AYI. As shown in the appendix this ambiguity stems from the fact that while the higher strike point reduces the covariance of the indemnity and the covariate shock, it may also reduce the variance of indemnity. The net effect of the increase in loading and the change in total variance due to purchasing AYI given the subjective mean of area-yields cannot be signed without choosing explicit parameter values, as shown in the appendix.

Since the incorrect subjective mean of area-yields affects how farmers perceive the expected return and risk reduction offered by AYI, it may also affect the size of the interval of β_i values for which it is optimal to purchase AYI, and as a result, diminish any possible increases in the number of farmers planting cotton. Incorporating the subjective distribution of area-yields into the definition of this interval given in (17) gives us:

$$d_i = 1, plant_i = 1 \leftrightarrow \max\left[\frac{\left(2\mu - (l+h)\right)(l+h) + \gamma\sigma_{I^*}^2}{2\gamma\sigma_{I^*,\varepsilon_c}}, \frac{\sigma_{I^*,\varepsilon_c} - x^*}{\sigma_{\varepsilon_c}^2}\right] < \beta_i < \frac{\sigma_{I^*,\varepsilon_c} + x^*}{\sigma_{\varepsilon_c}^2} \tag{28}$$

where $x^* = \sqrt{\gamma\left(\sigma_{I^*,\varepsilon_c}\right)^2 + \sigma_{\varepsilon_c}^2\left(\left(\mu - (l+h)\right)^2 - w^2 - \gamma\left(\sigma_{\varepsilon_i}^2 + \sigma_{I^*}^2\right)\right)}$.

As shown in the appendix, the ambiguity mentioned above with respect to the effects of farmer errors on the risk reduction of AYI also means that we cannot determine how the size of g affects the width of the interval given in (28). In what follows, we assign values to the parameters of our model in order to examine this question further.

Assigning parameter values to the model

To parameterize the model, we use two data sources for cotton farming in the coastal region of Pisco in Peru. The mean of area-yields, μ , and the variance of the covariate shock,

$\sigma^2_{\varepsilon_c}$, are estimated using the 20 year time series of area-yields for the Pisco valley from the Peruvian Ministry of Agriculture shown in Figure 1, combined with two years of yield data collected by our research team. The area-yields time series was detrended using a linear trend, and μ and $\sigma^2_{\varepsilon_c}$ are estimated using their sample counterparts from this detrended series. The pure risk premium, *r*, and the covariance of the indemnity with the covariate shock, σ_{I,ε_c}, are estimated assuming $\varepsilon_c \sim N\left(0, \sigma^2_{\varepsilon_c}\right)$ where $\sigma^2_{\varepsilon_c}$ is estimated as described above. The loading term *l* is set to 10 percent of the pure risk premium.

The return to the safe activity, *w*, is set equal to 490. Based on our household survey, this is approximately one tenth of what a person in the Pisco valley could earn as a day laborer during the three month planting season and the three month harvest season. In the household survey data, average farm size is around 5 hectares and the average price received per kilogram of cotton is 2 Peruvian soles. Since the model uses a farm size of 1 hectare and a price of unity for cotton, we divide wage laborer income as indicated by the data by 10 to put cotton and labor income as used in the simulation in the same units. The use of a constant relative risk aversion specification for the expected utility function guarantees that this rescaling will not affect our results.

To estimate the idiosyncratic variance component, $\sigma^2_{\varepsilon_i}$, and the variance of β, σ^2_{β}, the linear production function given in (2) was estimated via maximum likelihood using a four years of panel data on cotton parcels in the Pisco valley, covering the years from 2002 to 2005. These data are taken from the National Survey of Production and Sales survey collected by the Ministry of Agriculture. This requires assuming that all random components of the right hand side of (2) are normally distributed. The data are for parcels rather than individual farmers, and thus estimates may differ from the underlying farmer-level parameters. But we do not need exact values of these parameters for the purposes of this paper, just reasonable values to guide our simulations. Our parameter values are given in Table 2 below.

[INSERT TABLE 2 HERE]

The values of these parameters will determine the portion of farmers that decide to purchase AYI as well as those that participate in cotton farming.

Figure 3 shows how the set of β_i values for which it is optimal to purchase AYI changes with respect *g*, along with the concomitant fall in demand for insurance:

[INSERT FIGURE 3 HERE]

The bounds on β_i for AYI purchasers shrink rapidly with the size of the error, and the interval closes completely for errors larger than 5 percent of μ. This is paralleled by the rapid drop in demand for AYI, from 43 percent participation when there is no error down to zero when the error passes 5 percent.

Figure 4 shows how the share of farmers planting cotton changes with respect to the size of the error made by farmers, *g*:

[INSERT FIGURE 4 HERE]

As the size of *g* increases, participation in cotton farming drops from 82 percent to 73.6 percent. When farmers correctly perceive the mean of area-yields μ, participation in cotton farming is 8.4 percent higher than it would be if AYI were not available. Increasing the size of the error *g* steadily erodes this increase until it is reduced to zero.

As argued in the introduction, correctly evaluating the benefits of participation in AYI or any other development program may require detailed knowledge and complex reasoning on the part of eligible individuals. The analytical model and simulation presented here show that one type of error suggested via conversations with small farmers in Peru may strongly depress demand for AYI. For the sake of simplicity, it was assumed that all farmers make the same error. The result is that farmers do not enjoy the benefits of a program that might be a powerful tool against a source of risk that deepens poverty and discourages productive investment. Impacts on demand for AYI might be less extreme if this assumption were adjusted, but farmers who incorrectly believe the mean to be larger than it is would still undervalue AYI, and therefore be less likely to buy it.

If the benefits of index insurance are being foregone due to cognitive errors, two potential solutions immediately come to mind. The first is information campaigns or targeted marketing to try and correct mistaken perceptions that decrease demand. The second is changing economic incentives for participation in index insurance via subsidizing the premium. While our research team was not successful promoting demand for AYI through either of these mechanisms, others have had more success. In their study of rainfall insurance in India, Cole et al. (2008) study a variety of randomized treatments meant to increase demand for index insurance, or "randomized encouragements." These included price and informational treatments, as well as other manipulations. The treatments that included informational content might correct cognitive errors of the sort explored above, while reducing the price would offset the perceived increase in loading due to cognitive errors. Informational treatments included education about how rainfall in millimeters (the basis of the payout mechanism in the contract) translated into soil moisture on individual farms, and how this varied according to soil type. All treatments had significant effects, and in the expected directions. The price treatment was administered to households that also watched an informational video, and estimated impacts of this treatment capture than just the effect of price. Resulting estimates of the price elasticity vary by study region, ranging from

0.66 on the low end to 0.875. It appears possible to identify what sorts of information can increase demand for index insurance, and to what degree incentives must be manipulated to convince farmers to purchase it. Identifying what these mechanisms might be is something that ought to be done prior to econometric analysis. Economic theory, preliminary data collection, and focus group discussions such as those carried out by our research team in Peru can be helpful in this capacity.

Increasing demand for index insurance in the context of program evaluation

If farmers can benefit from index insurance but are not doing so, then any tool that can raise demand should be explored. But as pointed out in the introduction, empirical evidence of these benefits is scant. Measuring the impacts of index insurance on outcomes of interest should therefore be a goal of program evaluation. If this is the case, then the question of what tools we should use to raise demand in the context of research design becomes more nuanced. Consider program evaluation in the context of index insurance. Ideally, for each individual we could measure the level of some outcome of interest *y* with and without index insurance. The difference between these two outcomes is the individual level "treatment effect," where the treatment is having index insurance:

$$\Delta y_i = y_i^1 - y_i^0 \tag{29}$$

Here y_i^1 denotes the level of the outcome *y* for person *i* with index insurance and y_i^0 is the outcome without insurance. Assuming away the existence of spillover effects, (29) captures all relevant individual-level impacts of index insurance. If we could observe or consistently estimate these individual level effects, we could estimate the distribution of outcomes with and without index insurance, as well as the distribution of individual level treatment effects. This is impossible, however, due to what is known as the “evaluation problem”: we can only observe individuals at any given point in time with or without the treatment.

As an alternative, we could estimate average treatment effects via a comparison of individuals with and without index insurance. A direct comparison of this sort will yield inconsistent estimates, however, due to the selection problem: the decision to purchase index insurance may be driven at least in part by unobserved heterogeneity that may also be correlated with outcomes of interest. For example, in our model of area-yield insurance, the individual level β_i parameters are unobserved and correlated with the decision to purchase AYI. If these parameters are also correlated with the outcome variable in an econometric specification, estimated impacts will be inconsistent.

Another method that can yield consistent estimates of average treatment effects is directly randomizing individuals into and out of the index insurance program. Compelling people to purchase or not purchase insurance is probably not feasible, so this strategy will not be discussed further. A fourth method consists of randomizing a variable that affects demand for index insurance without affecting the variable to be used as the outcome. Comparisons of the groups formed by these randomizations can yield consistent estimates of average treatment effects. However, the average treatment effect that is estimated may vary depending on the randomization strategy used. This is because the randomization strategy chosen will affect the composition of the comparison groups. Unless impacts are the same for all individuals or vary randomly in a way that is not correlated with the decision to buy insurance, the average outcome in each group will depend on group composition (Heckman and Vytlacil, 2007).

For example, consider a randomization of eligibility to purchase index insurance. At the individual level, this would require some sort of mechanism by which randomly chosen persons are not allowed to purchase index insurance, while others are left to purchase it or not purchase as they please. Denote by z_i=1 random assignment to the eligible group whose members can purchase AYI at the market price $\tau = (r + l)$, and z_i=0 random assignment to the ineligible group. Note that since eligibility is assigned randomly, the share of eligible farmers buying index insurance is an unbiased estimate of the proportion of farmers who would buy index insurance if it were made available to all farmers in the population. Suppose that a proportion p of the eligible group would buy index insurance. We could then estimate the following average treatment effect, using data on the outcome and demand for insurance among the eligible and ineligible groups:

$$\frac{\Delta E(y)}{\Delta P(d_i = 1)} = \frac{E(y \mid z_i = 1) - E(y \mid z_i = 0)}{P(d_i = 1 \mid z_i = 1) - P(d_i = 1 \mid z_i = 0)} = \frac{E(y \mid P(d_i = 1) = p) - E(y \mid P(d_i = 1) = 0)}{p} \quad (30)$$

where d_i=1 denotes having purchased index insurance and $P(d_i = 1 \mid z_i = 1)$ is the proportion of eligible farmers purchasing index insurance. This is the change in the expected outcome (or, alternatively, the average change in the outcome) with respect to a change in the share of farmers with AYI from 0 to p. Note that since eligibility is assigned randomly, $E(y \mid z_i = 1)$ is equivalent to the expected value of y in the population when a proportion p of farmers buys AYI, with $E(y \mid z_i = 0)$ defined similarly for the case of zero participation in AYI among the population of farmers. Differences between these two expected values are due to changes in y among farmers who would buy AYI at the market price if they were eligible to do so. Equation (30) can be estimated by replacing each component with its sample counterpart.

As stated above, different randomization strategies yield different estimated average treatment effects. The randomization of eligibility described here makes it possible to estimate the average treatment effect of index insurance on farmers who would buy it if it were made available to all farmers at the premium τ . From the perspective of informing policymaking, this is arguably the most interesting average treatment effect, as it captures the average change in the outcome that would occur due to the introduction of index insurance. For this reason it will be referred to in what follows as the "Policy Relevant Treatment Effect," or PRTE, using the nomenclature of Heckman and Vytlacil (2007).

Unfortunately, when participation in the program being evaluated is low, it may not be feasible to estimate significant impacts based on a randomization of eligibility, even if impacts are large in magnitude. The reason is that low participation sharply increases the size of the standard error of the estimate of (30), which can be estimated as:

$$\left[\sqrt{\frac{\sum_{i=1}^{n}\sigma_y^2\left(1-R^2\right)}{n\pi_{z_i=1}\left(1-\pi_{z_i=1}\right)}}\right]\frac{1}{p} \qquad (31)$$

where n is the size of the sample used in the analysis, σ_y^2 is the variance of the observed outcome y, R^2 is the goodness of fit from the regression of the y on z, and $\pi_{z_i=1}$ is the proportion of the sample chosen to be eligible to buy index insurance. The standard error increases linearly with the participation rate p, and low participation may cause (31) to explode as a result.

Now consider a "randomized encouragement design," or randomly varying the incentives for the purchase of index insurance faced by individuals. Possibilities include randomizing the insurance premium via discounts or "coupons" for the purchase of index insurance, or the sorts of randomized information treatments used by Cole et al described above. Denote by c_i=1 assignment to the encouraged group, and c_i=0 assignment to control, and suppose that a proportion *p"* of the former group participates, while a share *p'* of the latter buy index insurance. This type of design will allow us to estimate the following:

$$\frac{E(y \mid c_i = 1) - E(y \mid c_i = 0)}{P(d_i = 1 \mid c_i = 1) - P(d_i = 1 \mid c_i = 0)} = \frac{E(y \mid P(d_i = 1) = p'') - E(y \mid P(d_i = 1) = p')}{p'' - p'} \tag{32}$$

In addition to having no direct effect on the outcome, the randomized encouragement *c* must satisfy the following assumption for equation (32) to represent an average treatment effect:

$$\begin{aligned} &d_i = 1 \mid c_i = 0 \rightarrow d_i = 1 \mid c_i = 1 \text{ and} \\ &d_i = 0 \mid c_i = 1 \rightarrow d_i = 0 \mid c_i = 0 \text{ for all } i \end{aligned} \tag{33}$$

In other words, if a farmer would purchase AYI without a coupon, then she would also do so with a coupon, and if she would not participate with a coupon, she would also not do so if she were not to receive a coupon. This is known as the monotonicity assumption, and was first introduced by Imbens and Angrist (1994). It implies that the randomized encouragement can never induce farmers to not buy index insurance. It may have no effect, or induce them to buy insurance, but cannot push them out of the program.[3]

If monotonicity is satisfied, then equation (32) is a "Local Average Treatment Effect" (LATE). This LATE represents the change in the expected outcome due to increasing the rate of participation in index insurance from *p'* to *p"*. This increase in participation is caused by the encouragement *c*, which induces some farmers to buy AYI who would otherwise not do so. For individuals whose purchase decision is unaffected by the encouragement, the outcome *y* is identical when the proportion of farmers with AYI is *p'* or *p"*. As a result, the LATE given in (32) captures the expected impact of index insurance on farmers who are induced to buy index insurance by the encouragement.

[3] See page 434 of Angrist and Imbens (1995) for a simple explanation of why this assumption is needed.

Note that whether or not monotonicity is satisfied will vary from treatment to treatment, and cannot be directly tested using data. Consider the random distribution of discount coupons for purchasing index insurance. It seems reasonable to assume that a lower price will only serve to encourage farmers to purchase index insurance, and therefore will satisfy monotonicity. This may not be the case, however, with information treatments. Clarifying the benefits of index insurance may push some farmers towards purchasing index insurance, but have the opposite effect on others. One could imagine that this tendency might be correlated with unobserved heterogeneity such as the β_i used in the model above, and could therefore not be controlled. However, other sorts of informational treatments may indeed satisfy monotonicity. The point is that this may be hard to justify and cannot be taken for granted, whereas with a variable such as price the argument is much simpler. This must be taken into account when designing an encouragement.

Even if monotonicity is satisfied, the LATE has other flaws. As stated above, the LATE captures average impacts on individuals induced to buy index insurance by the encouragement. The composition of this group will not in general be identical to that of the group of farmers who would buy index insurance were it made available, i.e., the farmers for whom impacts are captured by the PRTE. These two groups may be quite different, as the former needs some sort of extra incentive to buy the insurance, while the latter group would buy it at the market price. As a result, the LATE will not in general coincide with the PRTE. What a randomized encouragement designs can do is make it easier to estimate statistically significant effects, by increasing the share of farmers purchasing index insurance. The formula for the standard error of the LATE is:

$$\left[\sqrt{\frac{\sum_{i=1}^{n}\sigma_y^2\left(1-R^2\right)}{n\omega_{c_i=1}\left(1-\omega_{c_i=1}\right)}}\right]\frac{1}{p''-p'} \tag{34}$$

where $\omega_{c_i=1}$ is the proportion of farmers randomly selected to receive a coupon. A strong enough encouragement could make the difference *p"- p'* small enough to yield significant estimates, where the rate participation in the randomization of eligibility is always fixed at *p*. The estimated LATE can tell us something about impacts of index insurance on farmers. Although this may not correspond exactly to what we would like to learn, increasing what we know about the effects of index insurance on small farmers seems preferable to being inflexible about what parameters we choose to estimate, and learning less as a result.

Randomizing eligibility and randomized encouragements: simulating the tradeoffs

To simulate the tradeoff between the bias of the LATE with respect to the PRTE and the gains in precision offered by randomized encouragements, we return to the model economy of cotton farmers used above, and assume that farmers mistakenly believe the mean of area-yields to be $\mu^* > \mu$. Suppose we are to econometrically evaluate the impact of buying index insurance on farm profits; we will ignore production costs for simplicity. Impacts of index insurance on profits will be driven by increases in the number of farmers planting cotton rather than participating in wage labor. Assume that in the year in which we will carry out our evaluation, area-yields will be 10 percent higher than the historical mean. Profits for farmer *i* are:

$$q_{it} = \mu - \beta_i(.10) - \varepsilon_{it} \tag{35}$$

if she chooses to plant cotton, and:

$$w = 490 \tag{36}$$

if she participates in wage labor. Equation (35) is a realization of the production function given in equation (2) above. For each farmer the shock ε_{it} is a draw from a mean-zero normally distributed random variable, with a variance of 242,068 taken from the results of the estimates listed in Table 2. The parameter β_i and the shock ε_{it} are unobserved. The outcome of interest is therefore correlated with unobserved heterogeneity (β_i) that will also drive the decision to buy area-yield insurance. A research design is needed that generates variation in demand for AYI while not affecting the outcome, and creates a large enough difference in participation between the treatment and control groups so as to yield precise estimates of average impacts.

In this case, the PRTE is the average change in farm profits due to buying AYI among those who would purchase it at the market price. This parameter can be estimated by randomly assigning eligibility for the purchase of index insurance, e.g., by randomly disallowing farmers who want to purchase the insurance from doing so. It would then be calculated using average outcomes in the treatment and control groups, and the participation rate in the treatment group, as shown in equation (30). Figure 5 maps the PRTE and the bounds of its 95 percent confidence interval as a function of the size of the error made by farmers, *g*:

[INSERT FIGURE 5 HERE]

This is the PRTE for a randomization of eligibility that assigns half of a sample of *n*=1,000 farmers to treatment and half to control, in order to maximize precision. The sample is drawn from the economy described by the parameters in Table 2. The PRTE grows with the size of the error; since area-yields are higher than average, profits are positively correlated with β_i , and increasing *g* drives farmers with low values of β_i out of the AYI program, making the impact of AYI on profits larger as a result. When the error passes 3.5 percent of μ , the PRTE is no longer significant, despite little change in the magnitude of the effect.

Now consider a randomized coupon scheme. Figure 6 graphs the LATE as calculated using equation (32), its 95 percent confidence interval, and the bias of the LATE relative to the PRTE as a function of the size of the coupon. The error g is fixed at 5 percent of μ :

[INSERT FIGURE 6 HERE]

Low values of the coupon result in extremely wide bounds on the confidence interval; hence only coupon values ranging from 10 to 50 are included in the figure. When the coupon reaches a value of 33, or approximately 21 percent of the total premium of 159, the estimated LATE becomes statistically significant. Over this range of the coupon, the difference between the LATE and PRTE remains fairly steady, although it is substantial. It is negative, as we would expect, given that β_i is positively correlated with profits. Farmers who only participate if they receive a coupon tend to have lower values of β_i than farmers who would buy AYI at the market price, and therefore will also have lower profits in a year with high area-yields. In this particular example, the LATE and PRTE do not correspond, but only using the former enables us to estimate significant impacts of AYI on an outcome of interest.

This simulation could have used an information treatment as the encouragement, as correcting misperceptions with respect to μ could only make farmers more likely to purchase AYI. If we had allowed for both positive and negative errors, or an error associated with the values of β_i, this would have no longer been the case. The treatment would have encouraged some while discouraging others from purchasing AYI. The resulting estimates of how information can affect demand for AYI would be interesting and valuable, but would not be suitable as part of an analysis of impacts of index insurance on outcomes.

Conclusion

This paper explored the implications of cognitive errors for willingness to participate in an index insurance program. Experiences in Peru lead us to believe that these sorts of errors are widespread. This must be taken into account at every step of research design, from the structure of the index contract itself, to the way in which variation in demand is generated for the purpose of program evaluation. The analytical model presented here and the simulation based upon it demonstrate that one type of feasible error, mistakes with regard to the mean of area-yields, can exert a strong effect on participation. On one hand, if we believe that area-yield insurance and index insurance in general are valuable risk management tools, seemingly correctable misconceptions are leading farmers to miss out on large potential benefits. On the other, little is known about the size and nature of the benefits stemming from the purchase of index insurance, and anything that depresses demand makes it more difficult to enhance our knowledge of these effects. Randomized encouragement designs may address both of these concerns by stimulating demand in a way that allows for econometrically sound program evaluation. Resulting estimates may be less than ideal, as they may not describe impacts on the population of greatest interest. But they can allow for estimation of significant impacts, something that might not be possible under a randomization of eligibility, and thus increase what we know about the benefits of index insurance. Future research might examine how randomized encouragements can aid in the learning process. Information treatments may be helpful for stimulating demand, but perhaps nothing is more powerful than creating economic incentives strong enough to push individuals to buy index insurance and learn about its benefits through their own experiences. This would likely require a lengthy commitment on the part of researchers, as learning about the distribution of payouts is a process that cannot be accomplished in a single growing season. But it may be necessary if the promise of index insurance is to be realized.

References

Angrist, J., Imbens, G., 1995. "Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity." *Journal of the American Statistical Association*, 90(430), 431-442.

Carter, M., Little, P., Mogues, T., Negatu, W. 2007. "Poverty traps and natural disasters in Ethiopia and Honduras." *World Development*, 35(5), 835-856.

Chamberlain, G. 1983. "A characterization of the distributions that imply mean-variance utility functions." *Journal of Economic Theory*, 29(1), 185-201.

Chambers, R., Quiggin, J. 2002. "Optimal producer behavior in the presence of area-yield crop insurance." *American Journal of Agricultural Economics*, 84(2), 320-334.

Cole, S., Giné, X., Tobacman, J., Topalova, P., Townsend, R., Vickery, J. 2009. "Barriers to household risk management: evidence from India." Harvard Business School Finance Working Paper No. 09-116.

Dercon, S. 2004. "Growth and shocks: evidence from rural Ethiopia." *Journal of Development Economics*, 74(2), 309-329.

Galarza, F. 2009. "Risk, credit, and insurance in Peru: field experimental evidence." Unpublished working paper. http://mpra.ub.uni-muenchen.de/17833/1/MPRA_paper_17833.pdf.

Garcia, P., Adam, B.D., Hauser, R.J. 1994. "The use of mean-variance for commodity futures and options hedging decisions." *Journal of Agricultural and Resource Economics*, 19(1), 32-45.

Hanson, S.D., Ladd, G.W. 1991. "Robustness of the mean-variance model with truncated probability distributions." *American Journal of Agricultural Economics*, 73(2), 436-445.

Heckman, J., Vytlacil, E. 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 Heckman, J., Leamer, E. (Eds.), Handbook of Econometrics, vol. 6B, 4875-5143. Elsevier, Amsterdam.

Imbens, G., Angrist, J. 1994. "Identification and estimation of local average treatment effects." *Econometrica*, 62(2), 467-475.

Kroll, Y., Levy, H., Markowitz, H.M. 1984. "Mean-variance versus direct utility maximization." *The Journal of Finance*, 39(1), 47-61.

Meyer, J. 1987. "Two-moment decision models and expected utility maximization." *American Economic Review*, 77(3), 421-430.

Miranda, M. 1991. "Area-yield crop insurance reconsidered." *American Journal of Agricultural Economics*, 73(2), 233-242.

Nelson, C.H., Escalante, C. 2004. "Toward exploring the location-scale condition: a constant relative risk aversion location-scale objective function." *European Review of Agricultural Economics*, 31(3), 273-287.

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

Sinn, H.W. 1983. *Economic Decisions under Uncertainty*. New York: North Holland Publishing.

Skees, J., Hazell, P., Miranda, M. 1999. "New approaches to crop yield insurance in developing countries." International Food Policy Research Institute EPTD Discussion Paper 55. http://www.ifpri.org/sites/default/files/publications/eptdp55.pdf.

Trivelli, C., Venero, H. 2007. *Banca de desarrollo para el agro: experiencias en curso en América Latina*. Lima: Instituto de Estudios Peruanos.

Appendix

Demand for AYI and the parameters of the AYI contract

Recall that farmers who purchase AYI must have values of β_i lying within the following interval:

$$\max\left[\frac{(2\mu - l)l + \gamma\sigma_I^2}{2\gamma\sigma_{I,\varepsilon_c}}, \frac{\sigma_{I,\varepsilon_c} - x}{\sigma_{\varepsilon_c}^2}\right] < \beta_i < \frac{\sigma_{I,\varepsilon_c} + x}{\sigma_{\varepsilon_c}^2} \tag{1'}$$

Where $x = \sqrt{\gamma\left(\sigma_{I,\varepsilon_c}\right)^2 + \sigma_{\varepsilon_c}^2\left((\mu - l)^2 - w^2 - \gamma\left(\sigma_{\varepsilon_c}^2 + \sigma_I^2\right)\right)}$. The derivative of the upper bound of (1') with respect to σ_{I,ε_c} is:

$$\frac{1 + \frac{\gamma\sigma_{I,\varepsilon_c}}{x}}{\sigma_{\varepsilon_c}^2} > 0 \tag{2'}$$

Depending on which lower bound is utilized, its derivative is either:

$$\frac{1 - \frac{\gamma\sigma_{I,\varepsilon_c}}{x}}{\sigma_{\varepsilon_c}^2} \tag{3'}$$

which has an indeterminate sign, or

$$\frac{-\left((2\mu - l)l + \gamma\sigma_I^2\right)}{2\gamma\left(\sigma_{I,\varepsilon_c}\right)^2} < 0 \tag{4'}$$

If the derivative is (3'), the lower bound may be increasing, but at a rate slower than that of the upper bound; the difference between (18) and (19) is always negative. Increasing the risk reduction potential of AYI via a higher value of σ_{I,ε_c} will weakly increase demand for insurance.

The derivative of the upper bound with respect to the variance of the indemnity, σ_I^2, is:

$$\frac{\gamma}{-2x} < 0 \tag{5'}$$

The derivative of the lower bound with respect to σ_I^2 is then:

$$\frac{\gamma}{2x} > 0 \tag{6'}$$

or

$$\frac{1}{2\sigma_{I,\varepsilon_c}} > 0 \tag{7'}$$

depending on which lower bound is used. The size interval is therefore decreasing with respect to σ_I^2.

Lastly, we have the loading term *l* and its impact on the interval given in (1'). The derivative of the upper bound with respect to *l* is:

$$\frac{-(\mu - l)}{x} < 0 \tag{8'}$$

While the derivative of the lower bound with respect to *l* is:

$$\frac{(\mu - l)}{x} > 0 \tag{9'}$$

or

$$\frac{\mu - l}{\gamma\sigma_{I,\varepsilon_c}} > 0 \tag{10'}$$

depending on which lower bound is used. The size of the interval is a decreasing function of the loading term, *l*.

Farmer errors and the AYI contract

Suppose that farmers believe the mean of area-yields to be μ^*, where $\mu^* - \mu = g$. Given this error, the perceived covariate shock is $\varepsilon_c^* = \mu^* - \overline{q}_t^*$, where $E\left(\overline{q}^*\right) = \mu^*$. Given this error with respect to the mean of area-yields, the AYI contract that replaces shortfalls in area-yields from μ is perceived as:

$$I_t^* = \max\left[0, \varepsilon_{ct}^* - g\right] \tag{11'}$$

while the true contract is:

$$I_t = \max\left[0, \varepsilon_{ct}\right] \tag{12'}$$

The variance of area-yields given farmer errors is assumed to be identical to that of true area-yields. This implies that the distributions of ε_c and ε_c^* are identical. When we evaluate the moments of ε_c^* as they pertain to the AYI contract, we can use the distribution of ε_c, as long as we account for the fact that the incorrectly perceiving the mean of area yields leads farmers to view the strike point of the contract as higher than it actually is.

While it is proved in the main text that incorrectly perceiving the mean of area-yields leads farmers to view AYI as more costly, we cannot unequivocally say that it leads AYI to be viewed as less risk-reducing. The true change in the variance of output due to purchasing AYI is:

$$\sigma_I^2 - 2\beta_i \sigma_{I,\varepsilon_c} \tag{13'}$$

Replacing I and ε_c with I^* and ε_c^* yields the perceived change in variance. Consider the covariance between the indemnity and the covariate shock. The true value of this parameter is:

$$\sigma_{I,\varepsilon_c} = E\left(\max[0,\varepsilon_{ct}]\varepsilon_c\right) - E\left(\max[0,\varepsilon_{ct}]\right)E\left(\varepsilon_c\right) =$$

$$E\left(\varepsilon_c^2 \mid \varepsilon_c > 0\right) * P\left(\varepsilon_c > 0\right) =$$

$$\left(\frac{\int_0^{\varepsilon_c^{\max}} \varepsilon_c^2 f\left(\varepsilon_c\right) d\varepsilon_c}{P\left(\varepsilon_c > 0\right)}\right) P\left(\varepsilon_c > 0\right) = \tag{14'}$$

$$\int_0^{\varepsilon_c^{\max}} \varepsilon_c^2 f\left(\varepsilon_c\right) d\varepsilon_c$$

Under the subjective distribution of area-yields, the covariance term is:

$$\sigma_{I^*,\varepsilon_c} = E(\max[0,\varepsilon_{ct}-g]\varepsilon_c) - E(\max[0,\varepsilon_{ct}-g])E(\varepsilon_c) =$$
$$\left(E(\varepsilon_c^2 \mid \varepsilon_c > g) - gE(\varepsilon_c \mid \varepsilon_c > g)\right)P(\varepsilon_c > 0) =$$
$$\left(\frac{\int_g^{\varepsilon_c^{\max}} \varepsilon_c^2 f(\varepsilon_c)d\varepsilon_c}{P(\varepsilon_c > g)}\right)P(\varepsilon_c > g) - g\left(\frac{\int_g^{\varepsilon_c^{\max}} \varepsilon_c f(\varepsilon_c)d\varepsilon_c}{P(\varepsilon_c > g)}\right)P(\varepsilon_c > g) = \qquad (15')$$
$$\int_g^{\varepsilon_c^{\max}} \varepsilon_c^2 f(\varepsilon_c)d\varepsilon_c - g\int_g^{\varepsilon_c^{\max}} \varepsilon_c f(\varepsilon_c)d\varepsilon_c$$

The last line of (14') is larger than the last line of (15'); i.e., the true covariance is larger than the covariance given the incorrectly perceived mean.

Ambiguity with respect to the risk reduction potential of AYI comes from the impact of the misperceived mean of area-yields on the variance of the indemnity. The true variance of the indemnity function is equal to:

$$\sigma_I^2 = E(I^2) - E(I)^2 = E(I^2) - r^2 =$$
$$E(\max[0,\varepsilon_{ct}]\max[0,\varepsilon_{ct}]) - r^2 =$$
$$E(\varepsilon_c^2 \mid \varepsilon_c > 0)P(\varepsilon_c > 0) - E(\varepsilon_c \mid \varepsilon_c > 0)^2 P(\varepsilon_c > 0)^2 =$$
$$\left((\sigma_{\varepsilon_c}^2 \mid \varepsilon_c > 0) + E(\varepsilon_c \mid \varepsilon_c > 0)^2\right)P(\varepsilon_c > 0) - E(\varepsilon_c \mid \varepsilon_c > 0)^2 P(\varepsilon_c > 0)^2 =$$
$$(\sigma_{\varepsilon_c}^2 \mid \varepsilon_c > 0)P(\varepsilon_c > 0) + (1 - P(\varepsilon_c > 0))E(\varepsilon_c \mid \varepsilon_c > 0)^2 P(\varepsilon_c > 0) = \qquad (16')$$
$$(\sigma_{\varepsilon_c}^2 \mid \varepsilon_c > 0)P(\varepsilon_c > 0) + \frac{P(\varepsilon_c < 0)E(\varepsilon_c \mid \varepsilon_c > 0)^2 P(\varepsilon_c > 0)^2}{P(\varepsilon_c > 0)} =$$
$$(\sigma_{\varepsilon_c}^2 \mid \varepsilon_c > 0)P(\varepsilon_c > 0) + E(\varepsilon_c \mid \varepsilon_c > 0)^2 P(\varepsilon_c > 0)^2$$
$$(\sigma_{\varepsilon_c}^2 \mid \varepsilon_c > 0)P(\varepsilon_c > 0) + r^2$$

The second to last line follows from the symmetry of ε_c , i.e., $P(\varepsilon_c > 0) = P(\varepsilon_c < 0)$.

Using similar reasoning, it can be shown that under the subjective mean of area-yields μ^*, the variance of the indemnity is:

$$\left(\sigma_{\varepsilon_c}^2 \mid \varepsilon_c > g\right)P\left(\varepsilon_c > g\right) + \frac{r^{*2} P\left(\varepsilon_c < g\right)}{P\left(\varepsilon_c > g\right)} \tag{17'}$$

Since $P(\varepsilon_c < g) > P(\varepsilon_c > g)$, the second term in the right hand side of (17') may increase the variance of the indemnity, even though $\tau^{*2} < \tau^2$. However, raising the truncation point of a distribution decreases the truncated variance, and as a result $\left(\sigma_{\varepsilon_c}^2 \mid \varepsilon_c > 0\right)P\left(\varepsilon_c > 0\right) >$ $\left(\sigma_{\varepsilon_c}^2 \mid \varepsilon_c > g\right)P\left(\varepsilon_c > g\right)$. Incorrectly perceiving the mean of area-yields unequivocally lowers the covariance between the indemnity and the covariate shock, but it may also decrease the variance of the indemnity.

Farmer errors and demand for AYI

The impact of farmer errors on demand will depend on the derivatives of the bounds of the set of β_i values given in (1') for which purchasing AYI is optimal with respect to *g*. For convenience, the bounds on this set are restated here:

$$\max\left[\frac{(2\mu - l)l + \gamma\sigma_I^2}{2\gamma\sigma_{I,\varepsilon_c}}, \frac{\sigma_{I,\varepsilon_c} - x}{\sigma_{\varepsilon_c}^2}\right] < \beta_i < \frac{\sigma_{I,\varepsilon_c} + x}{\sigma_{\varepsilon_c}^2}$$

The derivative of the first possible lower bound of this set with respect to *g* is:

$$\frac{\sigma_{I^*,\varepsilon_c}\left(2\frac{\partial h}{\partial g}\left(\mu - (l+h)\right) + \gamma\frac{\partial \sigma_{I^*}^2}{\partial g}\right) - \frac{\partial \sigma_{I^*,\varepsilon_c}}{\partial g}\left(\left(2\mu - (l+h)\right)(l+h) + \gamma\sigma_{I^*}^2\right)}{2\gamma\left(\sigma_{I^*,\varepsilon_c}\right)^2} \tag{18'}$$

The sign of the left hand bracketed term in the numerator is ambiguous. If it is positive, then this possible lower bound is increasing with respect to *g*.

Now we consider the other potential lower bound of the set, as well as the upper bound. Since $\sigma_{\varepsilon_c}^2 > 0$ and $\partial \sigma_{I^*,\varepsilon_c} / \partial g < 0$, the sign of derivatives of the other two bounds with respect to *g* will depend on $\partial x^* / \partial g$. This derivative is:

$$\frac{2\gamma\sigma_{I^*,\varepsilon_c}\frac{\partial \sigma_{I^*,\varepsilon_c}}{\partial g} - \sigma_{\varepsilon_c}^2\left(2\left(\mu - (l+h)\right)\frac{\partial h}{\partial g} + \gamma\frac{\partial \sigma_{I^*}^2}{\partial g}\right)}{2x^*} \tag{19'}$$

The sign of the bracketed term in the numerator is also ambiguous. Comparing the different derivatives upon which the signs of (18') and (19') depend (i.e., $\partial h / \partial g$, $\partial \sigma_{I^*,\varepsilon_c} / \partial g$, and $\partial \sigma_{I^*}^2 / \partial g$) yields no further information without explicitly assigning values to parameters.

Tables and Figures

Figure 1: Cotton Area and Productivity in Pisco, 1986-2006

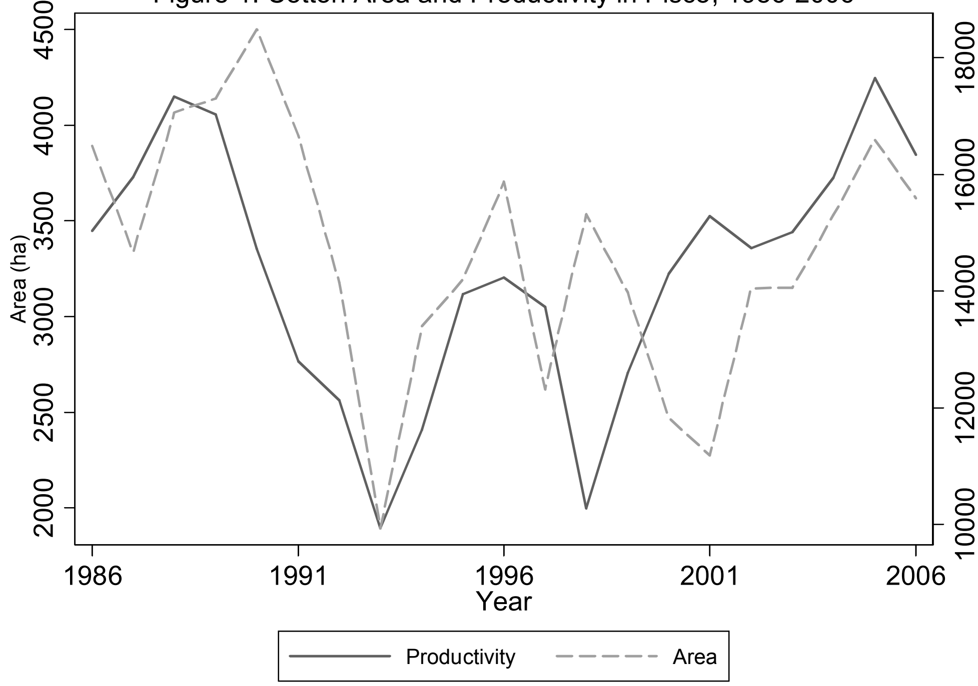

Figure 2: Area-yields in Pisco by Zone, 1997-2009

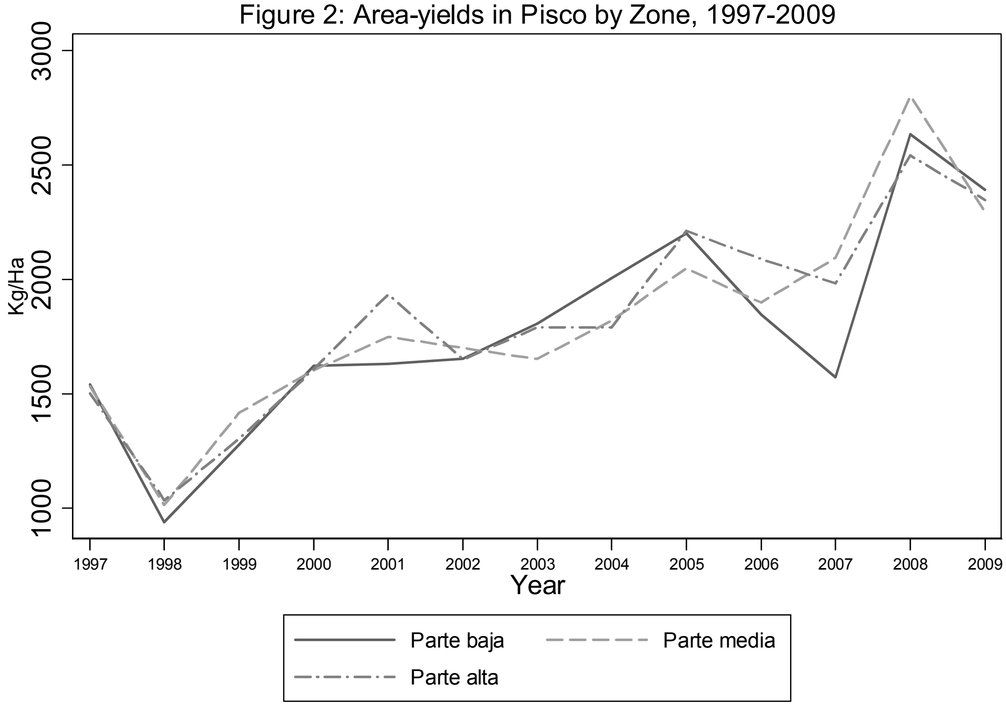

Table 1: Self-reported yields by zone in normal, good, and bad years (kg/ha).			
	Normal	Good	Bad
Parte Alta	2,253.76	2,808.11	1,277.55

Parte Media	2,108.59	2,668.18	1,187.05
Parte Baja	2,276.74	2,826.07	1,276.12

Table 2: Parameter Values for Economic Model.		
μ =1,852 kg	γ =1	σ_{β}^{2} =1.57
$\sigma_{\varepsilon_c}^{2}$ =160,528 kg^2	σ_{I}^{2} =54,711 kg^2	r=159 kg
$\sigma_{\varepsilon_i}^{2}$ =242,068 kg^2	σ_{I,ε_c} =80,065 kg^2	l=15.90 kg
w=490		

Figure 3: Farmer Errors and Demand for AYI

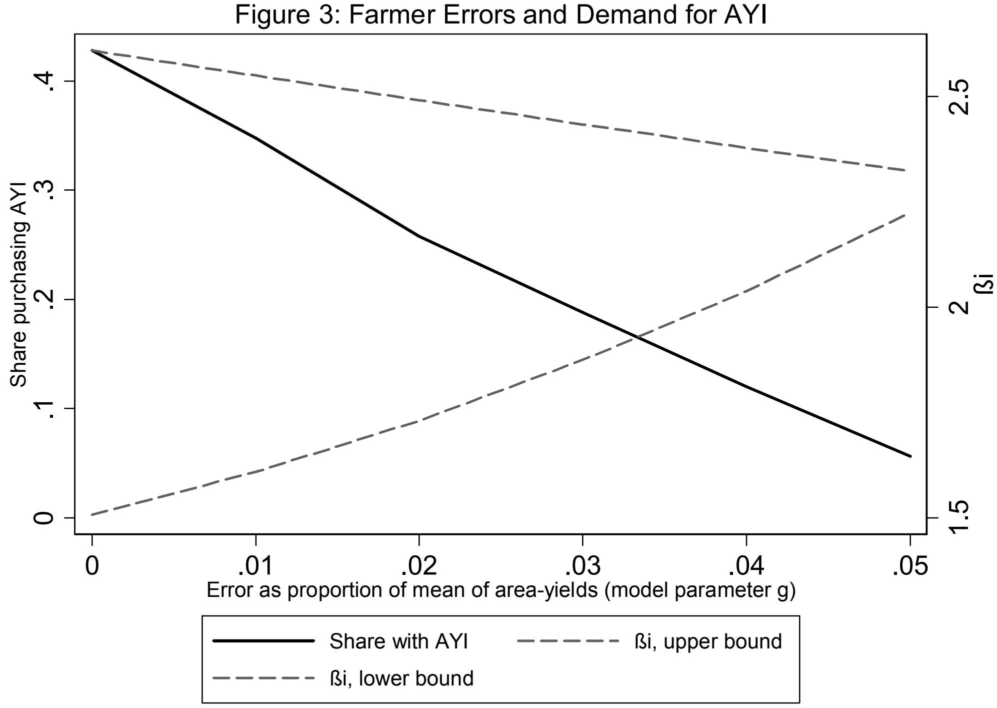

Figure 4: Farmer Errors and Participation in Cotton Farming

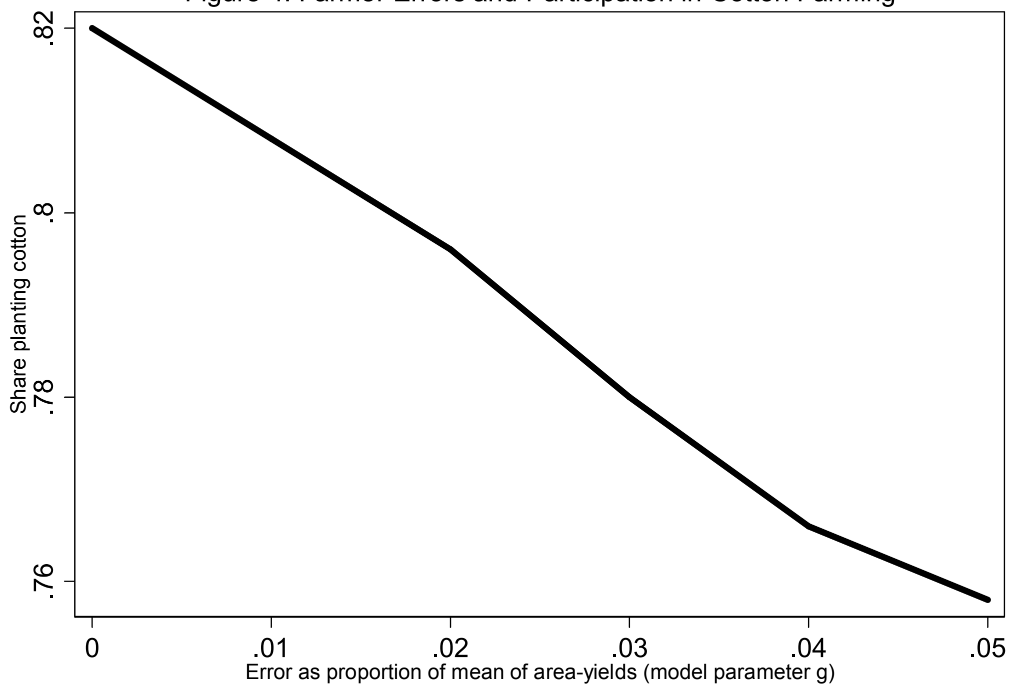

Figure 5: Magnitude and Precision of the PRTE with Farmer Errors

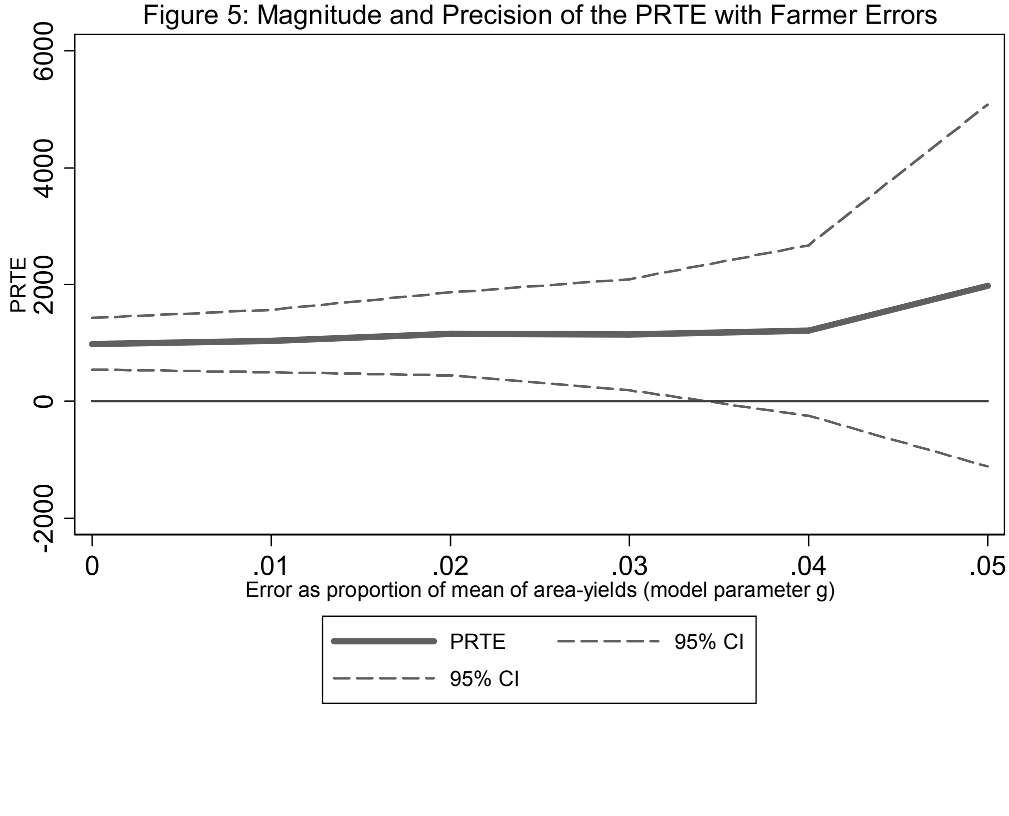

Figure 6: Precision and Bias of the LATE.