Farmer Crop Choice and Short-Run Weather Expectations

Benjamin M. Miller*

WORK IN PROGRESS: DO NOT CITE

Abstract

A robust literature examines how climate change will affect future crop choices. However, this literature contains mostly long-run estimates of new equilibria with few reliable methods of analyzing how these transitions occur. Prior research relied on long-run cross-sectional comparisons to estimate equilibria, while more recent works use 20-30 year historical "regime switching" weather variation within regions to estimate transition paths in the medium-run. This paper suggests a new and unique method: If individuals are able to predict some of the annual variation in seasonal rainfall, then their response to these expectations can be informative about marginal responses which occur along the path of response to climates change. Short-run sensitivity to seasonal rainfall expectations may further be a key underlying mechanism driving response in medium-run regime change studies.

By combining several county-level panel datasets on agricultural data with historical rainfall data from the Indian state of Madhya Pradesh, this paper presents the first empirical evidence that farmers' crop choices are indeed sensitive to seasonal rainfall expectations. This sensitivity is found to be significant, both statistically and economically. Further, this paper shows irrigation can reduce this sensitivity, although there is important heterogeneity across types of irrigation.

1 Introduction

The issue of climate change has seen a surge of papers estimating long-run changes in crop choices in response to long run changes in climate. However, it remains unclear how this adjustment will occur, particularly in developing countries. Will there be a slow but steady change of crops in response to a slow but steady change in climate, or are crop changes a lumpy response to occasional watershed moments or regime changes? Using extensive panel data on the crop choices of farmers from Madhya Pradesh, this paper provides evidence of the former by showing year-by-year variance in farmers' crop choices is in part a response to short-run seasonal rainfall expectations.

To begin, this paper makes a significant contribution to a much broader literature by significantly re-framing how researchers should think about rainfall and rainfall shocks. Traditionally, researchers have used weather (rainfall, temperature, etc.) as stochastic short-run deviations from these fixed climate-based means. Instead, this paper views weather as the output of an exceptionally complex function, with the implication that short run weather deviations can be expected, to a degree. While new to formal research, this idea has been deeply ingrained in the expectations of individuals for ages (we've all heard says such as "Red sky at night, sailor's delight. Red sky at morning, sailors take warning." Some readers are familiar with less common sayings such as 'When halo rings the

*University of California, San Diego. Email: b5miller@ucsd.edu. I would to thank Craig McIntosh, Jeremy Skog, Igor Vayman, Gordon Dahl, Paul Niehaus, Richard Carson, Chris Wignall, Prashant Bharadwaj, classmates at UCSD, and seminar attendees at the U.S. Census Bureau for their insights and assistance. All errors are my own.

moon or sun, rain's approaching on the run.") This paper posits that occupations which depend on future weather outcomes actively respond to such short-run weather expectations in economically meaningful ways.

A well-established literature has identified many key mechanisms affecting farmers' crop choices[1], such as climate, soil type, and input prices and availability. After accounting for these factors, farmers may still face a variety of potential crops to choose between. It is intuitively reasonable that the final choice of crops should be sensitive to expectations of the upcoming season's weather. Anecdotes suggest that farmers in both developed and developing countries do not plant crops without first consulting the year's weather projections from official sources such as NOAA, and unofficial sources such as the Farmer's Almanac or community sentiments.

The direction and magnitude of the response of crop choice to short-run weather expectations is likely to vary across crops, seasons, locations, and weather variables. Any direction and magnitude outcomes are of potential interest many individuals, from policy makers to those involved in agricultural markets. Policy makers or others involved in the agricultural market may wish to decrease the sensitivity of quantities and hence prices of important agricultural products to short-run weather fluctuations. This will become of increasing importance as global warming is expected to increase short-run weather variation across much of the world. Fortunately, this paper provides evidence that infrastructure investments such as irrigation may help smooth variance in acreage caused by climate change through increased variance of expected rainfall.

This paper proceeds as follows. Section 2 presents a literature review. Section 3 presents a basic profit-maximization model to frame the issue of crop choice and weather expectations. Section 4 describes the data used in this paper. Section 5 presents the empirical strategy for identifying crop choice sensitivity to weather expectations, and discusses limitations and biases. Section 6 presents results. Section 7 concludes.

2 Literature Review

2.1 Response to Climate and Weather

There is a popular literature on how crop choice has and will respond to climate change (ex. Mendelsohn and Dinar (1999), O'Brein et al. (2004), Seo and Mendelsohn (2008), Kurukulasuriya and Mendelsohn (2008)).[2] The idea that farmers might respond to "medium-run" changes is only beginning to emerge (ex. Taraz (2012)), and response to short-run changes have until this paper remained undiscussed. This is perhaps due to the perception that in the short run, deviations from long run means are random and hence cannot elicit ex ante behavioral responses.

In concurrent research, Taraz (2012) finds Indian farmers adjust their crop portfolio and irrigation investments in response to whether they are in a wet or dry rainfall regime (30-40 year regimes). Her focus is on the ability of farmers to recoup loses, finding that this medium-run adaption saves only 15% of profits lost due to climate change. In comparison, this paper uses a distinct source of variation to focus on short-run annual adjustments which create responses to regime changes in the long- or medium-run. An important step in future versions of this paper will be to separate short-run annual adjustments from medium-run regime-change responses. This is important given the large amount of short-run variation even within zonal and meridional rainfall regimes. How much of medium-run

[1] This literature is alternatively referred to as the crop selection literature.

[2] The reader should note that *climate* refers to long-run patterns in rainfall, temperature, and other atmospheric phenomena. *Weather* refers to short-run patterns. So for the purposes of this paper, seasonal weather refers to atmospheric outcomes for that growing season, while climate can be roughly thought of as the average seasonal weather. The overlap between these areas of research depends on the extent to which long-run climate change affects the variance of short-run weather outcomes.

and long-run adjustments can be accounted for by short-run adjustments is an important addition to the literature this paper ultimately hopes to answer.

2.2 Crop Choice in General

There is a wide variety of literature on farmers crop choices; agricultural economists, development economists, historical economists, agronomists, and even agricultural and biomechanical engineers have been writing on the subject for many decades. Most papers fall into one of two broad categories.

The first category attempts to estimate profit maximizing crop choice, usually (but not always) for a risk-neutral farmer. In general these are not studies of human behavior but attempts to calculate a profit-maximizing selection given a set of observable variables. As discussed below, the literature has identified an extensive list of relevant variables including soil type, input prices and availability (such as labor, machinery, seeds, and fertilizer). For meteorological inputs, climate, climate change, long-run climate cycles have all received detailed attention. Short-run weather fluctuations have been generally neglected.

Agriculture and Forestry expects have done much work on optimal crop selection. Mjelde et al. (1996) present a whole farm crop choice model for Eastern Texas. However, the model is unhelpful in extension to developing countries, as it allows for choice only between corn and sorghum, and assumes functioning insurance markets and farmer risk neutrality. Neither of these assumptions tend to hold in developing countries (Rosenzweig and Binswanger (1992)). More recently Cabrera et al. (2007) use a crop choice model based on El Nino and La Nina cycles which is adjusted for risk aversion, but their paper uses a test-farm in northern Florida to estimate optimal crop allocations and planting dates to test the effect of U.S. federal farm policies.

Agricultural and biomechanical engineers have also done work in this field. Whitson et al. (1981) examine optimal selection of farm machinery and how this interacts with crop choice. Due to weather risk farmers have uncertainty about the amount of time available to complete planting and other time-sensitive farm operations. Larger or more machinery can often accomplish tasks quicker, but at greater expense. The authors calculate the optimal selection of machines and crops type for a large farm in Texas, allowing for various levels of risk aversion by adjusting the probability that the farmer can complete tasks within the recommended time frame under the selected mix of machinery and crops. They find in this setting that weather risk "is an important variable to include in a decision model". However, the chosen actions of farmers are not examined.

Another work of interest from the engineering literature is Mohan and Arumugam (1994), who develop a computer program (CROPES) which estimates the optimal crop selection function across Tamilnadu, India. Taking user inputs of location, season, land and soil type, availability of water and labor, and farm size, CROPES selects a crop group. Asking further questions about soil, fertilize, pesticides, and storage facilities, the program suggests an optimal crop which the authors find usually aligns with agronomists' suggestions and farmers' decision.

Most crop selection studies outside development economics are not studies of human behavior but attempts to calculate a profit-maximizing selection given a set of observable variables. A second category of papers seeks to pin down mechanisms which cause farmers to deviate from this optimal choice, such as information asymmetry, salience issues, income constraints, or risk aversion. The philosophical reader correctly notes that the line between whether a factor makes a particular crop choice optimal or a deviation from the optimum is fuzzy if not completely subjective.

Lamb (2002) examines farmer deviation from optimal choice in India to present evidence of risk aversion and poverty traps, showing that poorer households allocate more land to low-risk, low-profit crops. He also notes that "Crop yields are highly susceptible to variations in the timing and duration

of the monsoon." This supports the idea that farmers may place high marginal value on weather expectations.

Another concern might be fixed costs of changing crops. Many crops require specific physical or human capital to grow successfully. These fixed costs may substantially influence optimal crop choice, but are rarely considered in the literature. A recent exception is Holmes and Lee (2009), who use data from the Red River Valley region of North Dakota to show crop choice is influenced by the soil qualities of neighboring fields. The key assumption required is that the soil qualities of neighboring fields are plausibly exogenous after controlling for the soil qualities of the given field. That this affect is diminished by ownership or administration boundaries strongly supports the argument that field owners are indeed planting common crops across several fields to reduce potential fixed costs.

The literature on farmers in developing countries clearly classifies farmers as risk-averse agents. When insurance and credit markets are unavailable ex ante actions are taken to self-insure and ex post actions are taken to smooth consumption. A seminal paper by Rosenzweig and Binswanger (1992) examines the investment choices of farmers through asset accumulation and depletion, rejecting risk neutrality. In the shock response literature, Kochar (1999) examines how farmers smooth consumption following idiosyncratic shocks by adjusting hours of off-farm work. Jayachandran (2006) examines the sensitivity of agricultural wages to TFP shocks. Dercon (1996) examines crop choice as ex ante risk reduction in Tanzania and presenting a useful crop selection model in a developing country context. However, along with the standard literature Dercon models crop selection as a function of households consumption security provided by liquid assets (namely cattle), not as a function of seasonal weather expectations. Response to weather expectations is tangentially discussed as important, but is not empirically examined. Dercon notes that Growing sweet potatoes has traditionally been an important risk management strategy within the farming system. Sweet potatoes have a low yield risk, since they are very drought and locust resistant, compared with the alternatives available. Sweet potatoes are also relatively inexpensive but provide a lower payoff.

3 Models

3.1 Rainfall Expectations

Economists conception of rainfall has remained remarkably constant since the introduction of deviations from mean rainfall as an instrument for income shocks by Wolpin (1982). Here the paper presents a model which extends framework Wolpin introduced.

Wolpin begins by dividing weather, $w_{t,j}$, into two pieces; a "deterministic component" $\overline{w}_j$ and a "random" component $\xi_{t,j}$.[3] Wolpin has t index time[4] and j index households, but we can think of j more generally as different geographic areas. As equation (1) shows,[5] Wolpin clearly had in mind mean weather over the the index of t in mind as a simplified example of the "deterministic component"

$$w_{t,j} = \overline{w}_j + \xi_{t,j} \tag{1}$$

The traditional assumption is that deterministic weather is uncorrelated with random weather, or

$$E[\overline{w}_j \xi_{t,j}] = 0 \tag{2}$$

[3]Note that weather here is more general than rainfall. Wolpin describes a single "summary statistic for weather", but clearly weather could refer to temperature, cloudy days, barometric pressure, any other individual variable or function of variables.

[4]t is a general index of any length of time. In this paper, it is interpreted as seasonal.

[5]Equation (3) in Wolpin (1982).

Of first order is providing convincing evidence that mean seasonal rainfall and expected seasonal rainfall are not necessarily the same. Although meteorologists can forecast the weather of the next several days with general accuracy, Palmer (1993) acknowledges that "At first sight it might appear rather contradictory to suppose that the atmosphere is at all predictable beyond this deterministic limit." However, Palmer and many other meteorologists and physicists have been examining the underlying coupled ocean-atmospheric processes which govern weather outcomes for decades.[6] The El Niño Southern Oscillation (ENSO) has been identified as a particularly important governing force. As examples particular to India, Rasmusson and Carpenter (1983) note correlation between below average monsoon rainfall and warm El Niño events. Most relevantly, Parthasarathy et al. (1988) develop statistical models which account between 70 to 83 percent of inter-annual rainfall variance for all of India, although some variables included ex post information.However, such results are not limited to India. McBride and Nicholls (1983) notes correlations between the Southern Oscillation Index (SOI) and its lags with annual seasonal rainfall across Australia. Rodó et al. (1997) document the correlations between ENSO and the North Atlantic Oscillation (NAO) signals ad the variability of seasonal rainfall on the Iberian Peninsula. A clear summary of the notion presented here can be found in Stockdale et al. (1998) "One conceptual model of weather is that of a series of events which are unconnected. That is, that the weather next week is essentially independent of the weather this week. However, although individual weather systems might be chaotic and unpredictable beyond a week or so, the statistics describing them may be perturbed in a deterministic and predictable way, particularly by the ocean."

Although academic meteorologists and physicists are clearly aware of these seasonal correlations, is it a stretch to expect illiterate farmers rationally expect the upcoming season's rainfall to be different from the mean? If the daily weather forecast with fancy satellites and computer analysis can't accurately predict the weather next weekend, how can illiterate farmers predict seasonal rainfall? The answers is an important distinction between deterministic prediction and simple correlations: Even if that farmer cannot predict exactly how much rain will occur on any given day, they may indeed have a sense that the next few months will have much more, a little more, a little less, or a lot less rain than the mean. Although those farmers may not have access to information on the SOI, they have spent decades carefully observing the outcomes of ocean and atmospheric phenomena at their farm's location. Through this accumulation of human capital, they have acquired the ability to form roughly accurate expectations. It should come as no surprise that their expectations of deviations from mean rainfall influence their weather-sensitive investments.

This suggests that expected rainfall may include information beyond the mean, $\overline{w}_j$. To crystallize this idea, let individuals at time t have access to the information set $\Omega_{t,j}$. Define the "deterministic component", $\tilde{w}_{t,j}$, and random component, $\tilde{\xi}_{t,j}$ as follows

$$\tilde{w}_{t,j} \equiv E[w_{t,j}|\Omega_{t-1,j}] = E[\overline{w}_j + \xi_{t,j}|\Omega_{t-1,j}] \tag{3}$$

$$\tilde{\xi}_{j,t} \equiv w_{t,j} - \tilde{w}_{t,j} \tag{4}$$

Or

$$w_{t,j} = \tilde{w}_{t,j} + \tilde{\xi}_{t,j} \tag{5}$$

The "deterministic component" is now best thought of as "expected" weather, conditional on information about prior weather $\Omega_{t-1,j}$. The random component is now thought of as the unexpected

[6]For those interested or still dubious, the Department of Physics at the University of Oxford has research group devoted to this topic with explanations in far greater detail than are relevant for this paper: see http://www2.physics.ox.ac.uk/research/predictability-of-weather-and-climate

shock, such as the difference between the weather forecast and the true weather. Parallel to equation (1), we now have

$$w_{t,j} = \tilde{w}_{t,j} + \tilde{\xi}_{t,j} \tag{6}$$

We now have the more reasonable (and depending on $\Omega_{t-1,j}$, potentially tautological) assumption that unexpected shocks are uncorrelated with expected weather

$$E[\tilde{w}_{t,j}\tilde{\xi}_{t,j}] = 0 \tag{7}$$

The key assumption behind this ideal version of a rainfall shock is that unexpected weather shocks are uncorrelated with pretty much anything. If they weren't, then the correlation could be expected and incorporated into $\Omega_{t-1,j}$.

3.2 A Basic Crop Choice Model with Shortrun Rainfall Expectations

This section presents a very basic model for intuition on the interaction between farmer crop choice and seasonal weather. Many of the assumptions made here will be relaxed later in later versions of the paper.

Suppose a farmer has some fixed amount of land, A, labor L, and capital K. The farmer can choose between multiple crops to plant on his land (referred to below as crop 1 and crop 2). For the present I assume the farmer has no non-farming options available. For each crop, let the farmer's crop-specific production function depend on the traditional labor and capital, the amount of land A_i planted with crop i, and rainfall, R. For initial simplicity allow the assumption that A_i is homogeneous of degree one, i.e

$$Q_i(L_i, K_i, A_i; R) = Q_i(L_1, K_1, 1; R)A_i \tag{8}$$

This amounts to an assumption that the returns to land use are constant. This may not be the case if land owned by the farmer varies significantly in quality for a given crop, and there is certainly evidence to suggest farmers chose crops based on land quality. Recognizing this, I continue the simple model.

The farmer also has a crop-specific cost function

$$c_i(L_i, K_i, A_i; R) \tag{9}$$

No assumptions will be made about the functional form of this cost function beyond the highly reasonable assumptions that costs associate with any particular crop are increasing in the amount of land planted with that crop (3) and benefit from returns to scale (4).

$$\frac{\partial c_i(\cdot)}{\partial A_i} > 0 \tag{10}$$

$$\frac{\partial^2 c_i(\cdot)}{\partial A_i^2} < 0 \tag{11}$$

A profit maximizing farmer in this situation solves the following maximization problem.

$$\max_{A_1,A_2,L_1,L_2,K_1,K_2} p_1Q_1(L_1, K_1, 1; R)A_1 + p_2Q_2(L_2, K_2, 1; R)A_2 - c_1(L_1, K_1, A_1) - c_1(L_2, K_2, A_2)$$

$$\begin{aligned} A &\geq A_1 + A_2 \\ L &\geq L_1 + L_2 \\ K &\geq K_1 + K_2 \end{aligned}$$

If optimal crop acreage, A_i, changes monotonically with respect to rainfall, R, then farmers should be adjusting their crop mix when the expected value of R changes. Below, this model shows farmers should optimally adjust their crop choices to known changes in rainfall. If we assume the land constraint binds ($A = A_1 + A_2$),[7] then we can identify $\frac{\partial A_i^*}{\partial R}$ by finding the first order condition with respect to A_1

$$p_1 Q_1(L_1, K_1, 1; R) - p_2 Q_2(L_2, K_2, 1; R) = \frac{\partial c_1(L_1, K_1, A_1^*; R)}{\partial A_1} - \frac{\partial c_2(L_2, K_2, A_2^*; R)}{\partial A_1} \tag{12}$$

$$\tag{13}$$

and then taking the derivative with respect to R.

$$\frac{\partial p_1 Q_1(L_1, K_1, 1; R)}{\partial R} - \frac{\partial p_2 Q_2(L_2, K_2, 1; R)}{\partial R} = \frac{\partial^2 c_1(L_1, K_1, A_1^*; R)}{\partial A_1 \partial R} + \frac{\partial^2 c_1(L_1, K_1, A_1^*; R)}{\partial A_1^2} \frac{\partial A_1^*}{\partial R}$$

$$- \frac{\partial^2 c_2(L_2, K_2, A_2^*; R)}{\partial A_1 \partial R} + \frac{\partial^2 c_1(L_2, K_2, A_2^*; R)}{\partial A_1^2} \frac{\partial A_1^*}{\partial R} \tag{14}$$

For conceptual clarity and succinctness, I define profits as $\pi_i = p_i Q_i(\cdot) - c_i(\cdot)$. We can now isolate the desired $\frac{\partial A_i^*}{\partial R}$ to find

$$\frac{\partial A_1^*}{\partial R} = \left(\frac{\partial \pi_1(L_1, K_1, A_1^*; R)}{\partial A_1 \partial R} - \frac{\partial \pi_2(L_2, K_2, A_2^*; R)}{\partial A_1 \partial R} \right) \left(\frac{\partial^2 c_1(L_1, K_1, A_1^*; R)}{\partial A_1^2} + \frac{\partial^2 c_1(L_2, K_2, A_2^*; R)}{\partial A_1^2} \right)^{-1} \tag{15}$$

Recalling assumption (4), we see the second group of parenthesis is strictly positive, and hence the optimal area planted with crop 1 increases in R if

$$\frac{\partial \pi_1(L_1, K_1, A_1^*; R)}{\partial A_1 \partial R} > \frac{\partial \pi_2(L_2, K_2, A_2^*; R)}{\partial A_1 \partial R} \tag{16}$$

Intuitively this tells us farmers plant more of crop 1 if a change in weather increases profits from an additional acre of crop 1 more than profits (or losses) from an addition acre of crop 2..

Of course, farmers face uncertainty from many sources. This profit maximization framework is immensely helpful in organizing the literature presented above on farmer choices under uncertainty. Whitson et al. (1981) examine how investments in K (how nice a tractor farmers purchase) are driven by uncertainty about L (how many days will be available for planting) due to weather. Hence K is a substitute for L and the quantity of L is sensitive to weather, $\frac{\partial L_i^*}{\partial R}$. An excellent paper by Conley and Udry (2010) examines how social learning removes uncertainty about the production function of a new crop, Q_{new}, causing farmers to switch crops. Because the crops examined in this paper are well-established, it is assumed that farmers possess a more detailed knowledge of the various Q_i.

4 Data

This project combines many data sources. The first source of agricultural data is the India Agriculture and Climate Data Set (IAC). This data set contains annual district-level observations on areas planted

[7] With some creativity, this model permits fields to lie fallow. Let the price of the "fallow crop" be the expected increased future profits on that field.

of various crops, as well as yields and prices, over the years 1957-1987[8] and across 271 Indian districts, or over 85% of India. It also includes agricultural inputs (number of workers, fertilizer use, capital stock, etc.) and monthly agricultural wage data, as well as a variety of measures of soil type and land quality, such as aquifer thickness. A major benefit of this dataset is the ability to compare pre- and post-Green Revolution agriculture.

The second source of agricultural data is ICRISAT's Meso Level Data (ICRISAT). Similar in format to the IAC, this data set contains annual district-level observations on areas planted of various crops, as well as yields and prices, over the years 1966-1999. ICRISAT also includes measures of soil type, although not other land quality measures such as soil ph or aquifer thickness. However, ICRISAT includes well-documented measures of irrigation and land use not included in IAC.[9]

Rainfall data is obtained from India Water Portals Meteorological Dataset (Met Data), in the form of monthly observations. [10] This includes monthly district-level observations for agriculturally important weather outcomes such as temperature, precipitation, ground frost frequency, cloud cover, wet day frequency, and other variables over the time period 1901-2002.

At present, analysis is limited to rice and sorghum growth in Madhya Pradesh, a major agricultural state in India. Future work includes expanding this analysis over additional states and crops. This will involve careful considerations, because in a country as large as India even the same crops are planted in different months in different regions. This is largely due to observed differences in climate, mainly the timing of monsoon rainfall and average seasonal temperatures.

5 Background

5.1 Crop Seasons and High Yield Variety (HYV) Crops

India has two major growing seasons: *kharif* and *rabi*, with specific crops grown in each season.[11] For example, rice and sorghum are grown during the *kharif* season, while wheat is grown during the *rabi* season. Because data on acreage for each crop is annual,[12] analysis must be limited to crops which are typically grown only in one specific season. This analysis focuses on the kharif season; the meteorology literature suggests that monsoon seasons may be easier to generate seasonal rainfall expectations than dry seasons (Rodó et al. (1997), Stockdale et al. (1998)), and for the econometrician that season also exhibits greater variance in rainfall for potentially identifying variation (see Figure 1). For crops this paper focuses on rice and sorghum because they offer clear ex-ante predictions for response to rainfall expectations. Both crops grow under similar conditions and require similar amounts of rainfall, and in this sense can be thought of as viable alternatives for any individual farmer to optimize over. However, sorghum is much more drought resistant, so that a farmer expecting a dry season might be more inclined to plant sorghum.

In 1966, high yield variety (HYV) crops were introduced in India, sparking India's "Green Revolution". Before the introduction of HYV crops, India was a major agricultural importer.[13] Thanks in

[8]For both IAC and ICRISAT, year refers to agricultural year, not calendar year. This means that the 1970-1971 rabi season is included in the 1970 data. As this paper focuses on the kharif season, the distinction is unimportant.

[9]IAC does include variables which suggest a relation to irrigation, but because there is no clear documentation of those particular variables, they are not included in this analysis.

[10]Met Data converts long-lat data to district-level observations. The underlying weather observations come from the Climate Research Unit (CRU) TS2.1 dataset, out of the Tyndall Centre for Climate Change Research, School of Environmental Sciences, University of East Anglia in Norwich, UK.

[11]These can be roughly thought of as summer and winter season, respectively, but are more accurately monsoon season and dry season. Some regions and farmers make use of an intermediary third season, *zaid*, which occurs in the spring. Crops such as gourds are common in this season.

[12]Annual over farm years. Each annual observation begins in June, not January.

[13]A major policy debate of the time was whether to continue researching India-specific high yield varieties, or to

large part to the increased yields of HYV crops, India turned the tables to become one of the worlds leading agricultural exporters. The requirement for HYV crops to realize these greatly increased yields was larger, more consistent supplies of water. HYV crops increased the marginal benefit of irrigation, and became most prevalent in areas which already had or subsequently introduced irrigation. With respect to this paper's research question, one might expect those farmers with access to irrigation to be less responsive to short run weather fluctuations. If farmers with access to irrigation are less responsive to short run weather variation, then the acreage of HYV crops should be less responsive to short run weather variation. Results using both IAC and ICRISAT data confirm this. Further, ICRISAT data allows us to directly view not only irrigation, but also type of irrigation. A more complex picture is clearly painted by this data which is explained in the results.

These result are important to policy makers or anyone involved in India's agricultural markets. Irrigation has been expanding throughout India, suggesting that India's agricultural future will be less sensitive to short-run weather fluctuations. However, much of this irrigation draws on underground aquifers. It is well-known that aquifer levels are falling. Some individuals blame an unsustainable increased demand brought on by the Green Revolution, and news reports abound of farmers have to dig deeper and deeper to access water supplies. Analyzing the sensitivity of traditional crop varieties is important to understand how India might respond to the increased short-run weather variation which accompanies climate change.

6 Methodology

6.1 Empirical Strategy

Empirically, this paper estimates the sensitivity of rice and sorghum acreage to rainfall expectations for the pending season. Specifically, the coefficient of interest is Γ in the following regression

$$\text{Area Crop } j_{it} = \mathbf{\Gamma} \mathrm{R}_{i,t} + \beta \mathbf{X}_{it} + \theta_{\mathbf{i}} \mathbf{Y_t} + \alpha_i + \varepsilon_{it} \tag{17}$$

Area Crop j_{it} is the hectares of crop j planted in district i. $\mathrm{R}_{i,t}$ is the deviation of expected rainfall from a 21-year rolling mean of rainfall for that district. Corresponding with the simple model, $\mathbf{X}_{it}$ is a vector of controls such as soil type, aquifer thickness, and other temporally fixed variables. $\theta_{\mathbf{i}} \mathbf{Y_t}$ controls for district-specific quadratic time trends, and α_i represents district fixed effects.

Unfortunately, direct observation of farmer rainfall expectations is not possible and no data on historical forecasts appears to be available. However, expectations for seasonal rainfall can be estimated using the same prior rainfall information available to farmers. As discussed above farmers may not observe SOI, but do observe their own prior rainfall outcomes. Hence, this paper define the expected deviation from mean rainfall in district i Madhya Pradesh in month t as the fitted values from the OLS regression of deviations from mean rainfall on ex-ante information.

$$E[\mathrm{R}_{i,t}|\Sigma_{i,t}] = \beta + \sum_{k=1}^{24} \gamma_{1,k} \mathrm{R}_{i,t-k} + \gamma_{2,k} \mathrm{R}^2_{i,t-k} + \theta_{\mathbf{i}} \mathbf{Y_t} + \alpha_i + \epsilon_{i,m} \tag{18}$$

Farmers' expectations take account of monthly deviations over the past two years. The functional form allows some non-linearity in how the magnitude of past deviations affect future expectations. While the γ coefficients are required to be constant across districts due to power constraints, each district has its own fixed effect and quadratic time trends. Note that a perfect fit with unobserved expectations is impossible – what is important is correlation. Figure 2 presents the mean realized

more quickly import varieties from other countries. Ultimately high yield wheat and rice were imported from Mexico in 1966. For an interesting discussion of this background, see Abler et al. (1994).

deviations and mean expected deviations across Madhya Pradhesh. Because "true" expectations are unobserved, a researcher can do little more than suggest the expectations look "reasonable". I hope skeptics are appeased by the knowledge that results are robust to a wide variety of functional forms estimating expectations.

Seasonal rainfall expectations are the sum of the corresponding monthly rainfall expectations. For Madhya Pradesh, the rabi season is corresponds to the months of November through February. The kharif season covers the monsoon months of June through October.[14] First stage results for the kharif season are presented in full in Table 3. Due to the two-stage nature of this analysis, standard errors presented in Tables 4 through 8 are bootstrapped.

Finally, Table 8 examines the heterogeneous impact of irrigation. First, the fraction of arable area which is irrigated is measured for each district, $I_{i,t}$. This fraction is further subdivided by type of irrigation (canals, tanks, wells, etc.), which here is subscripted b.

$$\text{Area Crop } j_{it} = \mathbf{\Gamma_1}\mathrm{R}_{i,t} + \Gamma_2(\mathrm{R}_{i,t} \times I_{i,t,b}) + \beta_1 I_{i,t,b} + \beta_2 \mathbf{X}_{it} + \theta_{\mathbf{i}}\mathbf{Y_t} + \alpha_i + \varepsilon_{it} \tag{19}$$

The variables of interest are now not only Γ_1, but also the interaction of the fraction of irrigated arable land with deviation from expected rainfall, Γ_2. If the particular type of irrigation reduces sensitivity to short-run rainfall deviation, then the Γ_2 should have the opposite sign as Γ_1.

6.2 Limitations

A variety of factors will bias the results towards zero. First, the data does not permit observations of substitutions between strains of the same crop ("within" switching). It may be the case that farmers switch between more or less drought-resistant version of rice depending on expected weather. Γ does not account for such activity, and will underestimate total changes in crop choice under this broader definition. It is also acknowledged that the chosen function form directly identifies only changes in "intensity" of use of any single crop. Due to the outside options of leaving fields fallow or changing the rate of farmland expansion, a decrease in the area planted of one crop due to weather expectations does not necessarily imply an increase in other crops.

It is acknowledged that the above specification cannot separate the "indirect effect" of farmers switching crops due to expectations of drought (the desired estimate), from the "direct effect" of farmers always planting the same crops and those farmers who did well buying land from those who did poorly. This is less of a concern as the literature has shown Indian land markets do not to function well.

Classical measurement error should also bias results towards zero. Similarly, imperfect correlation between estimated and unobserved expectations should also decrease the magnitude of the coefficient of interest.

7 Results

Tables 4 through 6 present results from equation (17). Tables 4 and 5 present results using IAC data, while Table 6 shows ICRISAT data produces similar results to IAC's Table 4. As expected Table 4 shows that when greater-than-average rainfall is expected, farmers significantly increase in the acreage of traditional rice varieties and decrease acreage of traditional sorghum varieties. In the ICRISAT data, Table 6 tells a similar story, although the response magnitudes are smaller. This is likely due to the fact that the ICRISAT data spans a later time frame of 1966 - 1999 rather than IAC's 1956-1988. It is likely that improvements in technology (such as irrigation) have decreased

[14] The period of March through May roughly corresponds with the zaid planting season.

over-all responsiveness to short-run rainfall expectations. Table 5 presents twin results, but for HYV crops. Note that the decreased number of observations is due to HYV crops not being introduced until 1966. As expected, acreage for both crops is less sensitive to expected rainfall; for rice in particular this difference is significant. This paper argues this effect is due to HYV crops being much more likely to be irrigated.

With the addition of ICRISAT data, this paper directly tests this assumption in Table 8 and **??**, which shows results from equation (19). In both cases, we see that not all types of irrigation are created equal. Ultimately rain-fed irrigation such as canals and tanks do not decrease farmer's sensitivity to short-run rainfall expectations. However, wells and potentially other sources clearly do decrease sensitivity.

8 Conclusion

This paper presents the first evidence that expectations of short-run weather expectations matter significantly in determining farmers' crop choices. Across the Indian state of Madhya Pradesh, data shows that farmers plant more rice and less sorghum when increased rainfall is expected that season. These effects both are economically and statistically significant.

It is helpful to put the sensitivity estimates in perspective. Given that expected rainfall has a standard deviation of 85 mm, Table 4 suggests roughly 619 more hectares of non-HYV rice are planted in each district for an expected rainfall deviation one standard deviation above 0. This represents an increase of just over 1% of total sorghum acreage in the average district (about 47,000 hectares). Extended across the entire state of Madhya Pradesh, we would see roughly 25,000 more hectares of non-HYV rice planted – equivalent to an additional small district of acreage! For negative expected rainfall deviations, a similar story holds for sorghum acreage.

The decreased sensitivity of HYV crops suggested that investments in infrastructure such as irrigation can potentially mitigate this sensitivity of crop acreage to rainfall expectations. Using ICRISAT data, this paper first replicated the short-run rainfall expectations sensitivity found in IAC data. Then it was shown that irrigation can indeed decrease this sensitivity, although there is important heterogeneity across types of irrigation. This heterogeneity is intuitively reasonable: types of irrigation which require rainwater do not help decrease sensitivity to rainfall expectations, while those which provide water from aquifers or other sources do.

8.1 Future Work

Continued work will examine variation in this sensitivity to rainfall expectations for a wider variety crops, and will be expanded across India. Price instruments may allow a measure of demand elasticity and help separate the price effect from the full effect on crop choice. Normalizing expected rainfall deviations to have a standard deviation of one may make interpretation of coefficients more clear. The simple crop choice model can be extended, and the production function can be estimated to confirm the response of crop yields to changes in rainfall.

Most importantly, future versions of this paper will focus on testing how much of the medium-run sensitivity to rainfall regimes found in other papers can be accounted for by this short-run sensitivity.

References

Abler, David G., George S. Tollow, and G. K. Kripalani, *Technical Change & Income Distribution in Indian Agriculture*, Westview Press, 1994.

Cabrera, Victor E., David Letson, and Guillermo Podest, "The value of climate information when farm programs matter," *Agricultural Systems*, 2007.

Conley, Timothy G. and Christopher R. Udry, "Learning about a New Technology: Pineapple in Ghana," *The American Economic Review*, 2010, *100* (1).

Dercon, Stefan, "Risk, Crop Choice, and Savings: Evidence from Tanzania," *Economic Development and Cultural Change*, 1996.

Holmes, Thomas J. and Sanghoon Lee, "Economies of Density versus Natural Advantage: Crop Choice on the Back Fourty," *NBER Working Paper 14704*, 2009.

Jayachandran, Seema, "Selling Labor Low: Wage Responses to Productivity Shocks in Developing Countries," *Journal of Political Economy*, 2006.

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

Kurukulasuriya, Pradeep and Robert Mendelsohn, "Crop switching as a strategy for adapting to climate change," *African Journal of Agricultural and Resource Economics*, 2008.

Lamb, Russell L., "Weather risk, crop mix, and wealth in the semi-arid Tropics," Technical Report 25, Department of Agricultural and Resource Economics 2002.

McBride, J. L. and N. Nicholls, "Seasonal Relationships between Australian Rainfall and the Southern Oscillation," *Monthly Weather Review*, 1983, *111.*

Mendelsohn, Robert and Ariel Dinar, "Climate Change, Agriculture, and Developing Countries: Does Adaptation Matter?," *World Bank Research Observer*, 1999.

Mjelde, James W., Troy N. Thompson, and Clair J. Nixon, "Government Institutional Effects on the Value of Seasonal Climate Forecasts," *American Journal of Agricultural Economics*, 1996.

Mohan, S. and N. Arumugam, "CROPES: A Rule-Based Expert System for Crop Selection in India," *Transactions of the American Society of Agricultural Engineers*, 1994, *37* (3).

O'Brein, Karen, Robin Leichenko, Ulka Kelkar, Kenry Venema, Guro Aandahl, Heather Tompkins, Akram Javed, Suruchi Bhadwal, Stephan Barg, Lynn Nygaard, and Jennifer West, "Mapping vulnerability to multiple stressors: climate change and globalization in India," *Global Environmental Change*, 2004.

Palmer, T. N., "Extended-Range Atmospheric Prediction and the Lorenz Model," *Bulletin American Meteorological Society*, 1993.

Parthasarathy, B., H. F. Diaz, and J. K. Eischeid, "Prediction of All-India Summer Monsoon Rainfall With Regional and Large-Scale Parameters," *Journal of Geophysical Research*, 1988, *93* (D5).

Rasmusson, Eugene M. and Thomas H. Carpenter, "The Relationship Between Eastern Equatorial Pacific Sea Surface Temperatures and Rainfall over India and Sri Lanka," *Monthly Weather Review*, 1983, *111.*

Rodó, X., E. Baert, and F. A. Comin, "Variations in seasonal rainfall in Southern Europe during the present century: relationships with the North Atlantic Oscillation and the El Niño-Southern Oscillation," 1997.

Rosenzweig, Mark R. and Hans P. Binswanger, "Wealth, Weather Risk and the Composition and Profitability of Agricultural Investments," Policy Research Working Paper Series 1055, The World Bank 1992.

Seo, S. Niggol and Robert Mendelsohn, "An analysis of crop choice: Adapting to climate change in South American farms," *Ecological Economics*, 2008.

Stockdale, T. N., D. L. T. Anderson, and J. O. S. Alves, "Global seasonal rainfall forecasts using a coupled ocean-atmosphere model," *Nature*, 1998, *392*.

Taraz, Vis, "Adaption to Climate Change: Historical Evidence from the Indian Monsoon," 2012. http://www.econ.yale.edu/∼vt48/Vis_Taraz_jmp.pdf.

Whitson, Robert E., Ronald D. Kay, Wayne A. LePori, and Edward M. Rister, "Machinery and Crop Selection with Weather Risk," *Transactions of the American Society of Agricultural Engineers*, 1981, *24* (2).

Wolpin, Kenneth I., "A New Test of the Permanent Income Hypothesis: The Impact of Weather on the Income and Consumption of Farm Households in India," *International Economic Review*, 1982, *23* (3).

9 Figures

Figure 1

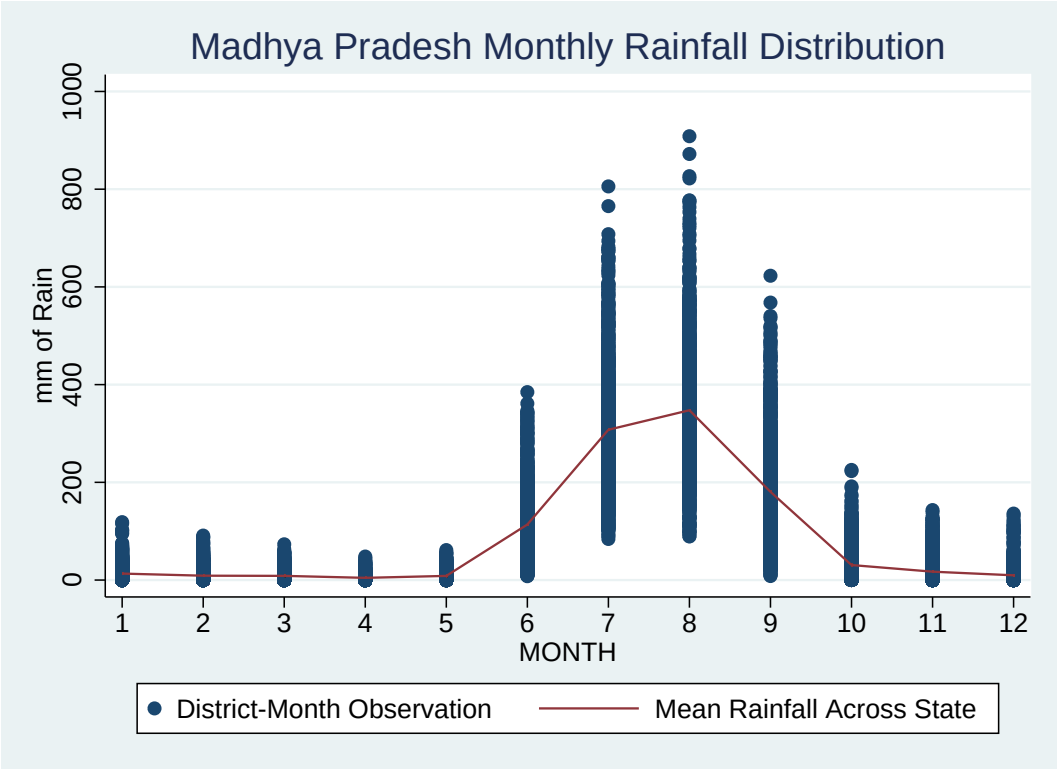
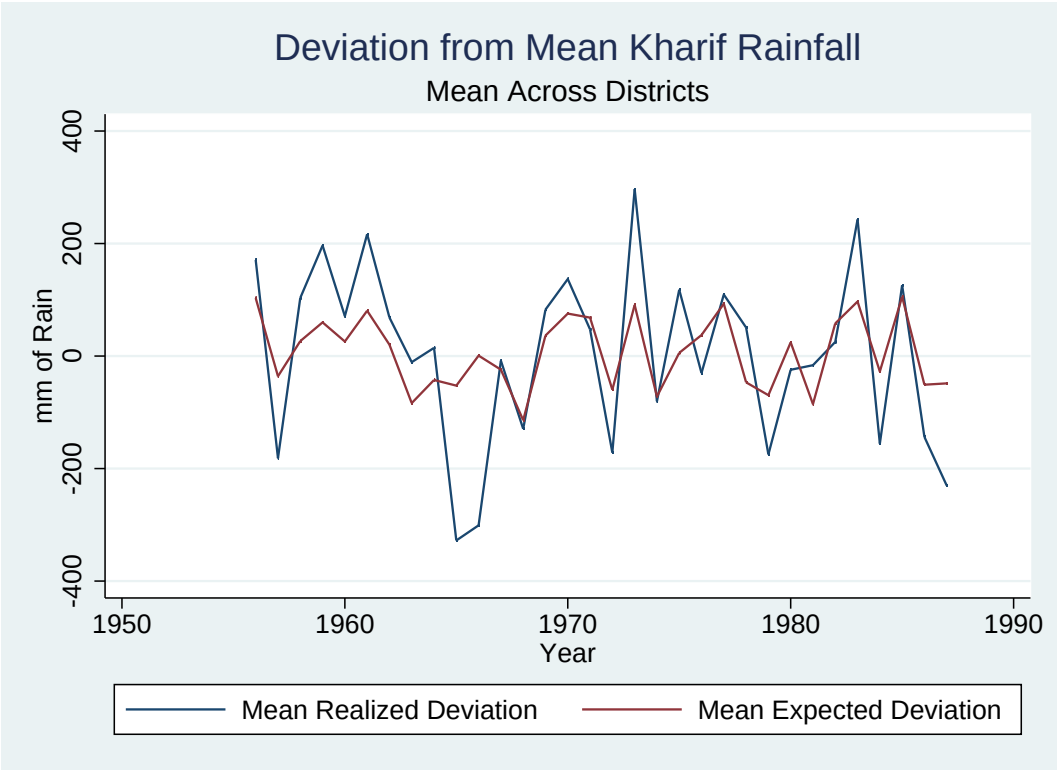


Figure 2



10 Tables

Table 1: IAC Summary Statistics

	1956-1987				
	mean	sd	min	max	count
Hectares of Rice Planted	46961.81	73031.25	100	392200	1147
Hectares of HYV Rice Planted	6775.099	21209.16	0	206509.6	1147
Hectares of Sorghum Planted	52119.35	48592.97	100	198800	1147
Hectares of HYV Sorghum Planted	7098.891	16512.63	0	103119.9	1147
Expected Kharif Rainfall Deviation	-3.020996	72.95241	-221.5834	277.725	1147
Realized Kharif Rainfall Deviation	-2.551528	200.5317	-506.5234	674.5577	1147
Kharif Rain	970.752	249.1783	352.964	1743.005	1147
Rabi Rain	49.70758	42.11443	.212	236.984	1147

Table 2: ICRISAT Summary Statistics

	1966-1999				
	mean	sd	min	max	count
Hectares of Rice Planted	48984.28	76912.67	100	418200	999
Hectares of Kharif Sorghum Planted	50607.51	48847.71	100	198200	999
Hectares of HYV Rice Planted	12054.35	29957.22	0	365000	999
Hectares of HYV Sorghum Planted	14138.24	25845.58	0	161000	999
Expected Rainfall Deviation	-4.856428	92.79364	-304.9528	264.0449	999
Realized Rainfall Deviation	-14.78835	214.7559	-506.5234	919.4176	999
Kharif Rain	930.2505	261.5013	352.964	1956.197	999
Rabi Rain	49.05775	43.31375	.212	236.984	999
USTALF/USTOLLS	.0621622	.208186	0	1	999
VERTISOLS	.1	.1790506	0	.8	999
VERTIC SOILS	.8256757	.2523118	0	1	999
INCEPTISOLS	.0121622	.0512346	0	.25	999
Fraction of Arable Area Irrigated by Canals, standardized	-.1530427	.8422385	-.6362773	4.124685	999
Fraction of Arable Area Irrigated by TANKS, standardized	-.0660405	1.07333	-.4429994	7.362375	999
Fraction of Arable Area Irrigated by TUBEWELL, standardized	-.1958804	.4789778	-.3349452	5.616585	999
Fraction of Arable Area Irrigated by OTHWELL, standardized	-.1238291	.7887361	-.8741401	3.85117	999
Fraction of Arable Area Irrigated by TOTWELL, standardized	-.1871773	.689527	-.8359801	3.187957	999
Fraction of Arable Area Irrigated by OTHSOUR, standardized	-.2681103	.6010353	-.852259	3.063707	999

Table 3: First Stage

	(1) June	(2) July	(3) August	(4) September	(5) October
1 Rainshock	-0.361***	0.000986	1.537***	0.509***	0.122**
	(0.0804)	(0.123)	(0.130)	(0.123)	(0.0420)
L.1 Rainshock	0.0227	0.974***	0.884***	0.0136	0.257***
	(0.0819)	(0.126)	(0.132)	(0.125)	(0.0428)
2 Rainshock	0.201*	0.205	-0.220	-0.0653	0.199***
	(0.0991)	(0.152)	(0.160)	(0.151)	(0.0518)
L.2 Rainshock	0.0305	-0.488**	-0.919***	-0.553***	0.232***
	(0.103)	(0.158)	(0.167)	(0.157)	(0.0540)
3 Rainshock	-0.678***	0.620***	0.511**	-0.120	-0.0771
	(0.105)	(0.161)	(0.169)	(0.160)	(0.0548)
L.3 Rainshock	0.650***	0.180	-0.637***	-0.0654	0.144**
	(0.0988)	(0.152)	(0.160)	(0.151)	(0.0517)
4 Rainshock	-0.471*	0.482	0.612	-0.158	0.192
	(0.222)	(0.340)	(0.358)	(0.338)	(0.116)
L.4 Rainshock	0.639**	-1.423***	-0.236	0.214	0.472***
	(0.215)	(0.330)	(0.347)	(0.328)	(0.112)
5 Rainshock	1.137***	-0.605**	0.316	0.822***	0.369***
	(0.124)	(0.191)	(0.200)	(0.189)	(0.0649)
L.5 Rainshock	0.571***	-1.767***	0.465*	-1.939***	-0.0948
	(0.131)	(0.201)	(0.211)	(0.199)	(0.0683)
L.6 Rainshock	0.00143	-0.153***	0.0893**	0.0873**	-0.0238*
	(0.0189)	(0.0290)	(0.0305)	(0.0288)	(0.00987)
L2.6 Rainshock	-0.112***	0.0410	-0.0262	0.0564*	-0.0511***
	(0.0187)	(0.0288)	(0.0302)	(0.0286)	(0.00980)
L.7 Rainshock	-0.0157	-0.256***	-0.0542**	0.0221	0.00598
	(0.0122)	(0.0187)	(0.0197)	(0.0186)	(0.00638)
L2.7 Rainshock	-0.0381**	-0.0188	0.0556**	0.137***	0.0352***
	(0.0120)	(0.0184)	(0.0193)	(0.0183)	(0.00625)
L.8 Rainshock	-0.0452***	0.0984***	-0.0620**	0.0573**	0.00106
	(0.0117)	(0.0179)	(0.0188)	(0.0178)	(0.00610)
L2.8 Rainshock	0.0169	-0.0853***	-0.0517**	0.0565**	-0.0256***

Continued on next page

Table 3 – continued from previous page

	(1) June	(2) July	(3) August	(4) September	(5) October
	(0.0116)	(0.0178)	(0.0187)	(0.0176)	(0.00605)
L.9 Rainshock	0.106***	-0.0884***	0.0121	0.0374*	-0.00478
	(0.0123)	(0.0188)	(0.0198)	(0.0187)	(0.00641)
L2.9 Rainshock	0.00536	-0.0496**	-0.0196	-0.0371	0.0451***
	(0.0124)	(0.0190)	(0.0200)	(0.0189)	(0.00648)
L.10 Rainshock	-0.191***	0.422***	-0.0428	-0.222***	0.0359
	(0.0375)	(0.0576)	(0.0605)	(0.0572)	(0.0196)
L2.10 Rainshock	-0.00297	0.193***	0.329***	-0.0826	-0.0302
	(0.0379)	(0.0582)	(0.0611)	(0.0578)	(0.0198)
L.11 Rainshock	-0.00510	0.217**	-0.0172	0.263***	-0.00372
	(0.0446)	(0.0686)	(0.0721)	(0.0681)	(0.0234)
L2.11 Rainshock	0.256***	0.151*	-0.394***	-0.547***	-0.0306
	(0.0442)	(0.0679)	(0.0713)	(0.0674)	(0.0231)
L.12 Rainshock	-0.465***	0.150	-0.529***	-0.660***	-0.271***
	(0.0895)	(0.137)	(0.144)	(0.137)	(0.0468)
L2.12 Rainshock	-0.132	0.865***	0.130	-0.230	-0.211***
	(0.0864)	(0.133)	(0.139)	(0.132)	(0.0452)
Constant	2.536	4.140	2.814	-1.660	0.128
	(10.44)	(16.04)	(16.85)	(15.94)	(5.460)
District FE	X	X	X	X	X
District-Specific Quadratic Year Trend	X	X	X	X	X
N	2960	2960	2960	2960	2960
r2	0.140	0.196	0.138	0.133	0.121

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: Hectares of Crop Area: Non-HYV

	(1) Rice	(2) Rice	(3) Sorghum	(4) Sorghum
Expected Kharif Rainfall Deviation	7.284	7.284	-7.212	-7.212
	(2.474)	(2.807)	(3.005)	(2.792)
District FE	X	X	X	X
District-Specific Quadratic Year Trend	X	X	X	X
Soil Type		X		X
Soil PH		X		X
Aquifer Thickness		X		X
Topsoil Depth		X		X
N	1147	1147	1147	1147
r2	0.997	0.997	0.973	0.973

Dependent variable is the hectares of non-HYV rice or non-HYV Sorghum. Expected Rainfall Deviation estimates total expected deviation of Kharif season rainfall in mm from the mean of the surrounding 20 years. Kharif season is the wet or monsoon season, defined here as June through October. Capital, labor, land quality, and prices are measured at the district level. Capital controls are quantities of bullocks and tractors. Labor controls are quantities of agricultural laborers and cultivators per hectare of all crops. Land controls include indicators of soil quality, soil PH, aquifer depth, and topsoil depth. Bootstrapped standard errors in parenthesis.

Table 5: Hectares of Crop Area: HYV

	(1) Rice	(2) Rice	(3) Sorghum	(4) Sorghum
Expected Kharif	3.174	3.174	0.382	0.382
Rainfall Deviation	(2.166)	(1.824)	(1.794)	(1.542)
District FE	X	X	X	X
District-Specific Quadratic Year Trend	X	X	X	X
Soil Type		X		X
Soil PH		X		X
Aquifer Thickness		X		X
Topsoil Depth		X		X
N	814	814	814	814

Dependent variable is the hectares of HYV rice or HYV Sorghum. Expected Rainfall Deviation estimates total expected deviation of Kharif season rainfall in mm from the mean of the surrounding 20 years. Kharif season is the wet or monsoon season, defined here as June through October. Capital, labor, land quality, and prices are measured at the district level. Capital controls are quantities of bullocks and tractors. Labor controls are quantities of agricultural laborers and cultivators per hectare of all crops. Land controls include indicators of soil quality, soil PH, aquifer depth, and topsoil depth. Bootstrapped standard errors in parenthesis.

Table 6: Hectares of Crop Area: Non-HYV (ICRISAT)

	(1) Rice	(2) Rice	(3) Sorghum	(4) Sorghum
Expected Rainfall	4.258	4.258	-3.678	-3.678
Deviation	(2.008)	(2.007)	(2.555)	(2.384)
District-Specific Quadratic Year Trend	X	X	X	X
District FE	X	X	X	X
Soil Type		X		X
N	1258	1258	1257	1257
r2	0.997	0.997	0.965	0.965

Dependent variable is the hectares of non-HYV rice or non-HYV Sorghum. Expected Rainfall Deviation estimates total expected deviation of Kharif season rainfall in mm from the mean of the surrounding 20 years. Kharif season is the wet or monsoon season, defined here as June through October. Bootstrapped standard errors in parenthesis.

Table 7: Hectares of Non-HYV Rice and Type of Irrigation

	(1) Canals	(2) Tanks	(3) Tube Wells	(4) Other Wells	(5) Total Wells	(6) Other Sources
Expected Rainfall	3.899*	4.104*	3.745*	3.717*	3.426*	4.399*
Deviation	(1.930)	(1.812)	(1.647)	(1.845)	(1.678)	(2.004)
Rain Dev. ×	-0.902	0.280	-1.821	-3.645*	-4.010*	1.102
Frac. Area Irrigated	(1.579)	(1.461)	(1.597)	(1.627)	(1.623)	(1.941)
Fraction of Arable	3285.6* * *					
Area Irrigated by Canals, standardized	(894.7)					
Fraction of Arable		1942.8				
Area Irrigated by TANKS, standardized		(1099.7)				
Fraction of Arable			1189.6*			
Area Irrigated by TUBEWELL, standardized			(599.2)			
Fraction of Arable				-347.5		
Area Irrigated by OTHWELL, standardized				(305.1)		
Fraction of Arable					791.5	
Area Irrigated by TOTWELL, standardized					(407.6)	
Fraction of Arable						979.5**
Area Irrigated by OTHSOUR, standardized						(344.0)
District-Specific Quadratic Year Trend	X	X	X	X	X	X
District FE	X	X	X	X	X	X
Soil Type	X	X	X	X	X	X
N	1258	1258	1258	1258	1258	1258
r2	0.997	0.997	0.997	0.997	0.997	0.997

Dependent variable is hectares of non-HYV rice. Expected Rainfall Deviation estimates total expected deviation of Kharif season rainfall in mm from the mean of the surrounding 20 years. Kharif season is the wet or monsoon season, defined here as June through October. Bootstrapped standard errors in parenthesis.

$^{*}p < 0.05,^{**}p < 0.01,^{***}p < 0.001$. Boostrapped standard errors in parentheses.

Table 8: Hectares of Non-HYV Sorghum and Type of Irrigation

	(1) Canals	(2) Tanks	(3) Tube Wells	(4) Other Wells	(5) Total Wells	(6) Other Sources
Expected Rainfall Deviation	-2.292	-3.479	-0.791	-2.866	-1.048	-3.021
	(2.240)	(2.698)	(2.870)	(2.961)	(2.591)	(2.420)
Rain Dev. × Frac. Area Irrigated	5.697	1.308	9.009	4.917	8.119**	7.518*
	(2.924)	(0.965)	(8.188)	(2.659)	(2.726)	(3.234)
Fraction of Arable Area Irrigated by Canals, standardized	-1884.5*					
	(883.2)					
Fraction of Arable Area Irrigated by TANKS, standardized		-600.1				
		(1703.2)				
Fraction of Arable Area Irrigated by TUBEWELL, standardized			-7632.1*			
			(3524.1)			
Fraction of Arable Area Irrigated by OTHWELL, standardized				-5192.8**		
				(1634.2)		
Fraction of Arable Area Irrigated by TOTWELL, standardized					-10750.6***	
					(3162.4)	
Fraction of Arable Area Irrigated by OTHSOUR, standardized						-3156.3**
						(1143.8)
District-Specific Quadratic Year Trend	X	X	X	X	X	X
District FE	X	X	X	X	X	X
Soil Type	X	X	X	X	X	X
N	1257	1257	1257	1257	1257	1257
r2	0.965	0.965	0.967	0.966	0.967	0.966

Dependent variable is hectares of non-HYV sorghum. Expected Rainfall Deviation estimates total expected deviation of Kharif season rainfall in mm from the mean of the surrounding 20 years. Kharif season is the wet or monsoon season, defined here as June through October. Bootstrapped standard errors in parenthesis.

$^{*}p < 0.05,^{**} p < 0.01,^{***} p < 0.001$. Boostrapped standard errors in parentheses.

Flood-tolerant rice expected to decrease yield variability, especially for socially disadvantaged groups in India

Manzoor H. Dar*, Alain de Janvry†, Kyle Emerick†, David Raitzer+, and Elisabeth Sadoulet†

December 21, 2012

Abstract

Approximately 25% of the cultivated rice area in India is prone to crop damage from prolonged flooding. We use a randomized field experiment in 128 villages of Orissa India to show that Swarna-Sub1, a recently released submergence-tolerant rice variety, has significant positive impacts on rice yield when fields are submerged for 6-14 days. We estimate that Swarna-Sub1 offers an approximate 45% increase in yields over the current popular variety when fields are submerged for 10 days. Further, yield effects are statistically indistinguishable from zero when fields are not flooded. We show additional results suggesting that low-lying areas prone to flooding tend to be more heavily populated by people belonging to lower caste social groups. We estimate that the percentage gain from Swarna-Sub1 during a heavy flood year is larger by 15.2 percentage points for scheduled castes. Thus, a policy relevant implication of our findings is that historically disadvantaged social groups are likely to benefit significantly from the spread of flood-tolerant rice varieties.

*International Rice Research Institute, New Delhi India. †Department of Agricultural and Resource Economics, University of California Berkeley. +Social Sciences Division, International Rice Research Institute, Manila Philippines. We acknowledge financial support from the Agricultural Technology Adoption Initiative and the Stress-Tolerant Rice for Africa and South Asia (STRASA) project. We are also grateful to the staff at the Balasore Social Services Society for outstanding work in the field.

1 Introduction

Given the need to increase world food production by 70% by the year 2050 [1], and rice being one of the three main staple foods, a major concern amongst scientists and policy makers is that rice yields have plateaued during the last two decades. The growth in productivity caused by Green Revolution technologies has stabilized [2]. Given that most of the productivity gains were enjoyed in better endowed areas with access to irrigation, scientists have expanded efforts to bring productivity gains to more marginal areas that are generally rainfed and prone to various environmental stresses [3].

India is one such country where extreme weather events create significant variability in rice production. While rice is India's dominant crop by both land area and value of production, much of the country's rice production takes place in areas prone to flash flooding and prolonged submergence of fields. Approximately 60% of the Kharif season (wet season) rice production during 2010-2011 took place in the particularly flood-prone states of Andhra Pradesh, Bihar, Orissa, Uttar Pradesh, and West Bengal [4].

A significant advancement in rice research is the development of new seed varieties that are more tolerant to abiotic stresses such as drought and flooding [5, 6, 7, 8]. One such advancement is the recent identification of the Sub1 (flood tolerance) gene and its insertion into rice varieties that are commonly grown in flood-prone areas [5]. In this paper we consider Swarna-Sub1, a recently released variety that maintains all the properties of the popular rice variety Swarna except for improved submergence-tolerance. The ability of Swarna-Sub1 to withstand floods of up to two weeks has shown to be promising in agricultural trials [9].

We used a randomized experiment to estimate the productivity impacts of Swarna-Sub1 in 128 villages of Orissa India. Measuring the productivity impacts of Swarna-Sub1 under field conditions is particularly important for two reasons. First, crop productivity is expected to depend on environmental factors other than flood duration [10]. Second, experimentation in farmer's fields allows for impact estimates to include the effects of management changes that will be induced by use of the new technology. Our estimates have policy importance because stress tolerant rice varieties (including Swarna-Sub1) are being widely distributed and demonstrated by the Indian Ministry of Agriculture through the National Food Security Mission (NFSM).

Previous research suggests that the most marginal farmers may be the main beneficiaries from technological innovations such as Bt cotton [11, 12]. We provide empirical evidence suggesting that farmers belonging to scheduled castes (SC) may benefit differentially more from the availability of flood tolerant rice. SC's are low caste groups that are generally less educated, have access to less land, and experience higher levels of poverty [13, 14, 15, 16]. We show that, as a consequence of the way land was allocated historically, SC's are differentially located in villages that are more likely to be flooded. Swarna-Sub1 is thus an example of a technology that will benefit a group that is a constant target of policy attention. It presents a unique opportunity to promote equity without compromising efficiency and without requiring social welfare handouts.

2 Setting

Our experiment took place in the Bhadrak and Balasore districts of Orissa, India. Balasore district is on the northern end of the state, neighboring the state of West Bengal. Bhadrak is neighboring to Balasore on the south. Both districts are made up of primarily coastal lowlands, where flooding during the Kharif season is especially problematic. Rice productivity in both districts is slightly higher than in the other districts of the state. Average annual yields during the period from 1999-2010 were 1.54 tons/ha in Balasore, 1.75 tons/ha in Bhadrak, and 1.43 tons/ha in the remainder of Orissa [17].

Flooding occurs frequently in the coastal districts of northern and central Orissa. Depression over the Bay of Bengal during the monsoon season from June to September often results in heavy rains over a short period, swelling of rivers, and prolonged flooding in low-lying coastal areas. Both Bhadrak and Balasore were affected by flooding during 2007, 2008, 2009, and 2011. The flood season corresponds with the Kharif production season. The Kharif rice crop is generally transplanted in early July and harvested during late November or early December. Approximately 60% of Kharif rice area in the two districts is cultivated with Swarna [18].

The farmers in our sample are mostly small and marginal. Average landholding is 0.83 ha. On average only 17.6% of cultivated area is irrigated. Kharif season rice is the dominant crop. Only 16.6% of farmers cultivate crops during the dry season. An average of approximately 73% of the rice harvest is allocated to home consumption.

3 Experimental Design and Methods

We generated our sample of villages from a list of flood-prone villages in eight different blocks of Balasore and Bhadrak.[1] Figure S1 shows the location of the villages in the experiment. We randomly selected 64 treatment villages and 64 control villages. This village-level randomization was stratified by block. During late May and early June of 2011, enumerators visited each village and met with a local village leader to identify 25 farmers that cultivate Swarna and have land that is prone to flooding. 5 farmers in each of the 64 treatment villages were randomly selected to receive a 5 kgs minikit of Swarna-Sub1 seed.[2] The seeds were delivered to each selected farmer prior to sowing time in mid June. 318 of the 320 selected farmers accepted the seeds and indicated intention to cultivate Swarna-Sub1. Table S1 shows that the randomization succeeded in generating experimental groups that are similar on observable characteristics.

Enumerators returned to all villages following harvest and post-harvest production practices (in March 2012). In each treatment village the five original recipients as well as ten randomly selected non-recipients were sought for surveying. Five randomly selected farmers in each control village were also approached for surveys. Surveys were completed with 1,248 farmers, making a response rate of 97.5%. The main component of the survey used in our analysis is a plot-level record of varietal use, production, and flood duration.

Our first step is to estimate the average productivity impact of Swarna-Sub1 using an OLS regression of the form,

$$Yield_{pivb} = \beta_0 + \beta_1 Sub1_{pivb} + \alpha_b + \varepsilon_{pivb}, \tag{1}$$

where $Yield_{pivb}$ is the rice yield on plot p, cultivated by farmer i, in village v, and block b, $Sub1_{pivb}$ is an indicator for plots cultivated with Swarna-Sub1, α_b is a block fixed effect, and ε_{pivb} is an unobserved error term. Standard errors are clustered at the village level, which is the first tier of randomization. Our estimate of β_1 in (1) represents the average effect of Swarna-Sub1 across different flood durations. We further exploit the spatial variation in flooding intensity to estimate the productivity impact of Swarna-Sub1 under various flooding durations. Heavy rains in September 2011 brought flooding to many parts of central and northern Orissa, including Bhadrak district. We use the duration of flooding for each plot to allow the yield advantage of Swarna-Sub1 to vary according to the length of flooding by estimating,

$$Yield_{pivb} = \beta_0 + \beta_1 Sub1_{pivb} + \beta_2 Daysflood_{pivb} + \beta_3 Sub1_{pivb} * Daysflood_{pivb} + \alpha_b + \varepsilon_{pivb}. \tag{2}$$

[1] Blocks (or tehsils) are administrative units at the level immediately below districts.
[2] 5 kgs of seed is roughly enough to cultivate a plot of approximately 0.1-0.2 ha.

One caveat of these regressions is that they are at the plot level while the random allocation of Swarna-Sub1 was at the individual level. As a robustness check we verify in Table S2 that the estimation results are qualitatively similar when including controls for plot area, ownership, indicators for medium and upland, as well as access to irrigation. All interactions between the duration of flooding and the control variables were also included.

Our second step is to assess whether farmers in lower caste social groups are differentially exposed to flooding. We first use our own survey to investigate the correlation between the caste of the cultivator and the duration of submergence for each plot in Bhadrak district. We estimate

$$Daysflood_{pivb} = \beta_0 + \beta_1 SCorST_{ivb} + x_{ivb}\delta + \alpha_{vb} + \varepsilon_{pivb}, \tag{3}$$

where $SCorST_{ivb}$ is an indicator for whether the farmer belongs to SC or ST, x_{ivb} is a vector of farmer-level covariates, and α_{vb} is a village fixed effect. The estimate of β_1 in (3) measures the expected difference in flood duration for SC/ST farmers within a village, when controlling for factors in x.

We further estimate the correlation between caste and flooding in a broader set of villages using satellite imagery of a large flood event during September 2008 in the Bhadrak, Kendrapara, and Jajpur districts of Orissa. We use RADARSAT images (resolution 100m) from September 18th, September 25th, and September 30th to identify areas that were flooded for 1-7, 8-12, and 13 or more days. Figure S1 helps to visualize these data. We spatially joined these data with geocoded village-level census information for the year 2001 to estimate

$$Flooded_v = \beta_0 + \beta_1 ShareST_v + \beta_2 ShareSC_v + \beta_3 Jajpur_v + \beta_4 Kendrapara_v + \varepsilon_v, \tag{4}$$

where $Flooded_v$ is an indicator equal to 1 if village v was flooded, $ShareSC_v$ and $ShareST_v$ are the shares of the population that are SC and ST, respectively, and $Jajpur_v$ and $Kendrapara_v$ are fixed effects for villages in Jajpur and Kendrapara. Our estimate of β_2 measures the within district expected difference in the probability of flooding for villages that are entirely SC's, when compared to villages that are entirely higher caste people. We observe the centroid of each village and not the approximate boundaries. We account for the fact that farmers cultivate land away from the center of the village by considering a village to be flooded if its center was within 500m of a flooded area.[3]

4 Results

Our main results are first depicted graphically in Figure 1. We focus our attention on the advantages of Swarna-Sub1 relative to Swarna, since Swarna-Sub1 is meant to be identical to Swarna in all ways other than flood tolerance. In Panel A of the figure we estimate nonparametric Fan regressions [19] of yield as a function of the duration of flooding. The regressions are estimated separately for Swarna and Swarna-Sub1 plots. The figure demonstrates that the productivity gain of Swarna-Sub1 is clearly increasing in the duration of flooding. While the estimated yield for Swarna-Sub1 is slightly lower than that of Swarna under non-flood conditions, there is a noticeable yield advantage that increases in the number of days flooded, up to 12-13 days. In Panel B we display the estimated yield advantage of Swarna-Sub1 (vertical difference between the two lines in Panel A) along with its 95% confidence interval. The estimated yield decrease of 180 kg/ha when plots are not flooded (an approximate 5.3% decrease) is not statistically significant from zero. The

[3] Our results are robust to other distance cutoffs. Results are qualitatively similar when using cutoffs of 250m or 750m.

yield advantage of Swarna-Sub1 increases as flood severity worsens, with a maximum advantage of around 725 kg/ha (approximate 66% increase) occurring at approximately 13 days of flooding. In Panel C we show the distribution of days flooded for plots cultivated to Swarna and Swarna-Sub1. A total of 1,960 plots were cultivated with Swarna and a total of 313 plots were cultivated with Swarna-Sub1. There is substantial variation in flood exposure for both varieties, suggesting that the estimated effects at a given flood duration are not generated by an unreasonably small numbers of observations.

In Table 1 we report OLS regression results from estimating equations (1) and (2). We include all variety types for completeness and all regressions include fixed effects for blocks, which were strata for randomization. The estimate in Column 1 for equation (1) indicates that Swarna-Sub1 offers an average yield benefit of around 236 kg/ha, which represents an approximate 10% improvement. The estimate is however only marginally significant at the 10% level (p=0.074). In Column 2 we estimate equation (2) where we allow the yield advantage of each variety type to be a linear function of the duration of flooding. The coefficient for the indicator $Sub1_{pivb}$ indicates that Swarna-Sub1 yields are estimated to be lower by approximately 120 kg/ha when a plot is not flooded. This difference is not statistically distinguishable from zero. Importantly, the statistically significant and positive estimated coefficient on $Sub1_{pivb} * Daysflood_{pivb}$ indicates that the yield advantage of Swarna-Sub1 is increasing in the duration of flooding. For each additional day of flooding, we estimate that the yield gain increases by 65 kg/ha. Following Panel B of Figure 1, in Column 3 of Table 1 we specify the yield advantage of Swarna-Sub1 (relative to Swarna) to be a piecewise linear function of days of flooding with a kink at 12 days. Our estimate of the yield advantage is $-201.8 + 86.5 * Daysflood - 175.4 * (Daysflood - 12) * \mathbb{1}(Daysflood > 12)$, where $\mathbb{1}$ is the indicator function. The magnitudes of these effects are comparable with the nonparametric estimates in Figure 1, and therefore these results represent our preferred specification.

The results demonstrate that Swarna-Sub1 is expected to improve rice yield under flooded conditions. A natural followup question is, who are the most likely beneficiaries of this promising new technology? We first investigate in Table 2 whether plots cultivated by farmers belonging to lower caste groups are exposed to more flooding by estimating equation (3) using OLS. We limit our sample to farmers in Bhadrak district for this analysis since the 2011 floods in our sample region occurred primarily in Bhadrak. In Column 1 we report results without village fixed effects. The interpretation of the result is that the average duration of flooding on plots cultivated by scheduled caste (SC) or scheduled tribe (ST) farmers is longer by 1.83 days, when compared to farmers that belong to the general caste group, which is the highest caste group in our sample.[4] This estimate represents an approximate 21% increase in the length of flooding for plots cultivated by the lowest caste farmers. In Column 2 we include village fixed effects. This changes the interpretation of our estimate to be that *when comparing two plots in the same village*, plots cultivated by SC or ST farmers are expected to be flooded for 1.35 days longer. In Column 3 we add control variables characterizing cultivators with no notable change. Persistence of the correlation with inclusion of village fixed effects and control variables suggests that the land allocation *within flood-prone villages* is likely to result in low-lying plots being cultivated by lower caste farmers.

We use satellite data on flooding in September 2008 and geocoded village-level census information to further investigate the correlation between caste and flooding in a broader set of villages. The OLS regression results from estimating equation (4) are in Table 3. The results show that villages where a larger share of

[4] 36.9% of farmers in our sample are OBC, 13.6% are SC, and 4.9% are ST. Only 1.1% of plots in our sample are cultivated by ST's. Tribal groups in Orissa tend to be located in the more mountainous areas which are not included in our sample because those areas are not prone to flooding.

the population are SC's are more likely to be flooded for 1-12 days.[5] In Column 1 the interpretation of our estimate is that within districts, the probability of being flooded for 1-7 days is higher by 0.15 in a village where the entire population is SC's. This increase in probability is relative to a village that is entirely occupied by people belonging to OBC or General castes. The estimate in Column 2 suggests that the probability of being flooded for 8 days and no more than 12 days is higher by 0.094 for a village with entirely SC's. To simplify interpretation, consider a village of entirely OBC/General caste people in Kendrapara district. The estimated probability of being flooded for 8-12 days is 0.194 (=0.036+0.158). The estimated probability increases by 48% to 0.288 if the village was instead entirely SC's. The magnitude of this estimate is therefore large and particularly important since our previous results suggest that areas flooded for approximately one to two weeks are those most likely to benefit the most from Swarna-Sub1. The results in Column 3 demonstrate that the share of the population that is SC is not significantly associated with the village being flooded for more than 12 days. It is in these villages where water likely remains stagnantly and therefore the yield benefits of Swarna-Sub1 are reduced (Figure 1 and Table 1). A final notable result is that villages inhabited by ST's are less likely to be affected by flooding, a result due to their location in upland areas.

As a final exercise, we simulate the impact of replacing Swarna with Swarna-Sub1. This simulation is most policy relevant because an objective of some state governments in India under the National Food Security Mission is to entirely replace Swarna with Swarna-Sub1. We use the regression results from Column 3 in Table 1 to predict production (in kg) for each plot if it was cultivated with Swarna. We aggregate separately for ST/SC's and by district before computing the predicted percentage increase in total production for a scenario when all Swarna land is cultivated with Swarna-Sub1. The results in Panel A of Figure 2 show that universal replacement of Swarna with Swarna-Sub1 during the 2011 floods in Bhadrak would have resulted in a predicted increase in total rice production of 25.4% for OBC/General caste farmers and 40.6% for SC/ST farmers. The difference in predicted impact of 15.2 percentage points is statistically significant from zero (p=0.002). We also vary the intensity of flooding by adding (or subtracting) a day of flooding to the observed 2011 flood duration for each plot, and then repeating the simulation exercise. The projected impacts remain economically and statistically significant for a flood event that is shorter by two days. Projected impacts in Panel B are noticeably smaller and statistically insignificant since the 2011 floods in Balasore district were far less severe.

How severe were the 2011 Orissa floods in relation to flooding during past years? In Figure 3 we use the river discharge estimates from the Global Flood Detection System for a discharge site that is approximately 9 km from our nearest study village [21, 22]. The estimates of daily flooding magnitude in Panel A suggest that the 2011 floods were of similar severity to the previous major flood years. Panel B aggregates the number of heavy flood days by year to show that the flooding in this part of coastal Orissa was just as severe or more severe during 2003, 2007, and 2008. We consider this as an important validation that the types of flooding events where Swarna-Sub1 is expected to have the largest impacts are frequently occurring.

5 Discussion

The experimental evidence reported in this paper suggests that Swarna-Sub1, a recently released flood tolerant rice variety, is expected to have positive impacts on yields when fields are submerged for approximately

[5] These data are entirely cross sectional. The standard errors reported in Table 3 do not adjust for spatial correlation. We estimated the spatial standard errors suggested in [20] using a cutoff of 0.1 decimal degrees (approx 11 km). Statistical significance of our main results is unaffected.

6-14 days. Yields are not noticeably different when fields are not flooded or flooded for only a short period of time. We estimate that wide scale adoption of Swarna-Sub1 prior to the 2011 floods in one of our sample districts would have resulted in an approximate increase in rice production by 27.1%. The high frequency of intense flooding in many rice producing areas of northern and eastern India suggests that the benefits of this technology will be frequently and widely enjoyed.

We provided additional analysis showing that members of the scheduled castes are likely to benefit substantially more from adoption of Swarna-Sub1. Within villages, scheduled caste farmers cultivate plots that are exposed to longer floods. Across villages, villages where a greater share of the population is scheduled caste are more likely to be flooded for 1-12 days. The scheduled castes are historically disadvantaged and continue to be a focus area for poverty alleviation strategies. Our results suggest that land allocation and settlement in flood-prone areas has been unfavorable for scheduled caste people. This observation combined with our result that Swarna-Sub1 outperforms the main popular rice variety under flood conditions indicates that scheduled castes are likely to be a key social group benefiting from the spread of Swarna-Sub1. This new flood tolerant rice variety thus offers the prospect of delivering both efficiency and equity gains for very large segments of the farm population.

References

[1] FAO, How to feed the world in 2050. high level expert forum, issues brief (2009). Rome, Italy.

[2] G. Conway, G. Toenniessen, *Nature* **402**, C55 (1999).

[3] R. Evenson, D. Gollin, *Science* **300**, 758 (2003).

[4] *IndiaStat*, www.indiastat.com (2012).

[5] K. Xu, *et al.*, *Nature* **442**, 705 (2006).

[6] A. Kumar, J. Bernier, S. Verulkar, H. Lafitte, G. Atlin, *Field Crops Research* **107**, 221 (2008).

[7] D. Normile, *Science* **321**, 330 (2008).

[8] S. Verulkar, *et al.*, *Field Crops Research* **117**, 197 (2010).

[9] S. Singh, D. Mackill, A. Ismail, *Field Crops Research* **113**, 12 (2009).

[10] P. Ram, *et al.*, *Field Crops Research* **76**, 131 (2002).

[11] M. Qaim, D. Zilberman, *Science* **299**, 900 (2003).

[12] J. Huang, S. Rozelle, C. Pray, Q. Wang, *Science* **295**, 674 (2002).

[13] A. Deshpande, *The American Economic Review Papers and Proceedings* **90**, 322 (2000).

[14] Y. Kijima, *Economic Development and Cultural Change* **54**, 369 (2006).

[15] A. Banerjee, R. Somanathan, *Journal of Development Economics* **82**, 287 (2007).

[16] V. Hnatkovska, A. Lahiri, S. Paul, *American Economic Journal: Applied Economics* **4**, 274 (2012).

[17] *Indian Ministry of Agriculture*, http://apy.dacnet.nic.in (2012).

[18] D. Behura, W. Jaim, M. Hossain, *Adoption and diffusion of modern rice varieties in Bangladesh and eastern India* (International Rice Research Institute, 2012), pp. 45–57.

[19] J. Fan, *Journal of the American Statistical Association* **87**, 998 (1992).

[20] T. Conley, *Journal of Econometrics* **92**, 1 (1999).

[21] *Darthmouth Flood Observatory*, http://floodobservatory.colorado.edu/India.htm.

[22] G. Brakenridge, S. Nghiem, E. Anderson, R. Mic, *Water Resources Research* **43**, W04405 (2007).

Tables and Figures

Table 1: Impact of flood-tolerant rice (Swarna-Sub1) on yield during 2011 wet season. All three columns present OLS regression results where the dependent variable is rice yield in kg/ha. Sample consists of 4,182 plots cultivated by 1,232 surveyed farmers with complete production data (area and production). 7 plots with recorded yield larger than 8,000 kg/ha are excluded from the analysis.

	Dependent Variable = Yield (kg/ha)		
	(1)	(2)	(3)
Swarna-Sub1	236.24*	-120.34	-201.77
	(130.93)	(161.94)	(168.96)
Other modern variety	-237.78***	-348.86***	-349.32***
	(80.26)	(119.86)	(120.21)
Traditional variety	-572.55***	-973.44***	-962.29***
	(140.07)	(202.80)	(200.72)
Days flood	-69.58***	-92.02***	-94.48***
	(16.68)	(18.22)	(22.82)
Swarna-Sub1*Days flood		65.21***	86.48***
		(22.30)	(26.68)
Traditional variety*Days flood		55.28***	54.20***
		(18.36)	(17.95)
Other modern variety*Days flood		21.18*	21.33*
		(10.90)	(10.96)
(Days flood-12)*1 if Days flood > 12			15.91
			(36.35)
Swarna-Sub1*(Days flood-12)*1 if Days flood > 12			-175.43***
			(62.88)
Block Fixed Effects	Yes	Yes	Yes
Mean of Dep Variable	2214.26	2214.26	2214.26
Number of Observations	4182	4182	4182
R squared	0.446	0.451	0.452

Robust standard errors clustered at the village level are reported in parentheses. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Table 2: Relationship between caste of cultivator and flood duration during 2011 wet season. All columns present OLS regression results where the dependent variable is number of days plot was flooded. Sample consists of 2,169 plots cultivated by 613 surveyed farmers in Bhadrak district. The sample is limited to Bhadrak district because the 2011 floods occurred primarily in Bhadrak.

	Dependent Variable=Days flood		
	(1)	(2)	(3)
Cultivator is ST or SC	1.83**	1.35*	1.38**
	(0.80)	(0.69)	(0.64)
Cultivator is OBC	-0.31	0.14	0.21
	(0.87)	(0.40)	(0.40)
Cultivator has below poverty line card			0.05
			(0.36)
Education of cultivator			0.05
			(0.06)
Cultivator has thatched roof			-0.22
			(0.30)
Cultivator has private tubewell			1.26
			(0.80)
Cultivator has latrine			-0.54
			(0.37)
Village Fixed Effects	No	Yes	Yes
Mean of Dep Variable	8.70	8.70	8.72
Number of Observations	2169	2169	2163
R squared	0.020	0.676	0.679

Robust standard errors clustered at the village level are reported in parentheses. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Table 3: Correlation between village-level caste shares and flooding in September 2008. All columns present OLS regressions for villages in Bhadrak, Jajpur, and Kendrapara districts of Orissa India. Excluded caste group in all regressions is OBC/General. Village assigned a value of flooded=1 if nearest submerged area within 500 meters of village center. Flood data are from RADARSAT and population data are from the 2001 census. Census data and village locations are taken from geodata@UCBerkeley.

	Dependent Variable=1 if flooded		
	(1) September 18	(2) September 18-25	(3) September 18-30
ST Share	-0.570***	-0.178***	-0.144***
	(0.058)	(0.046)	(0.019)
SC Share	0.153***	0.094**	-0.001
	(0.056)	(0.040)	(0.025)
Village in Jajpur district	0.123***	0.067***	0.078***
	(0.024)	(0.015)	(0.010)
Village in Kendrapara district	0.195***	0.158***	0.085***
	(0.025)	(0.018)	(0.010)
Constant	0.235***	0.036***	0.005
	(0.021)	(0.011)	(0.006)
Mean of Dep Variable	0.355	0.125	0.053
Number of Observations	4184	4184	4184
R squared	0.052	0.046	0.033

Heteroskedasticity robust standard errors are in parentheses. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels. All regressions are weighted by number of persons in the village dependent on agriculture (number of cultivators + number of agricultural laborers).

Figure 1: Impact of flood-tolerant rice (Swarna-Sub1) on yield during 2011 wet season. Panel A: Nonparametric fan regressions of yield (kg/ha) on flood duration (days). Fan regressions are estimated separately for each variety. Panel B: Estimated yield advantage of Swarna-Sub1 relative to Swarna, as a function of duration of flooding. Solid black line is treatment effect and shaded area represents 95% confidence interval. Bootstrapped standard errors adjust for village-level clustering by sampling villages. Panel C: Distribution of flood duration for each variety type. It displays the share of plots for each variety that were flooded for the given number of days.

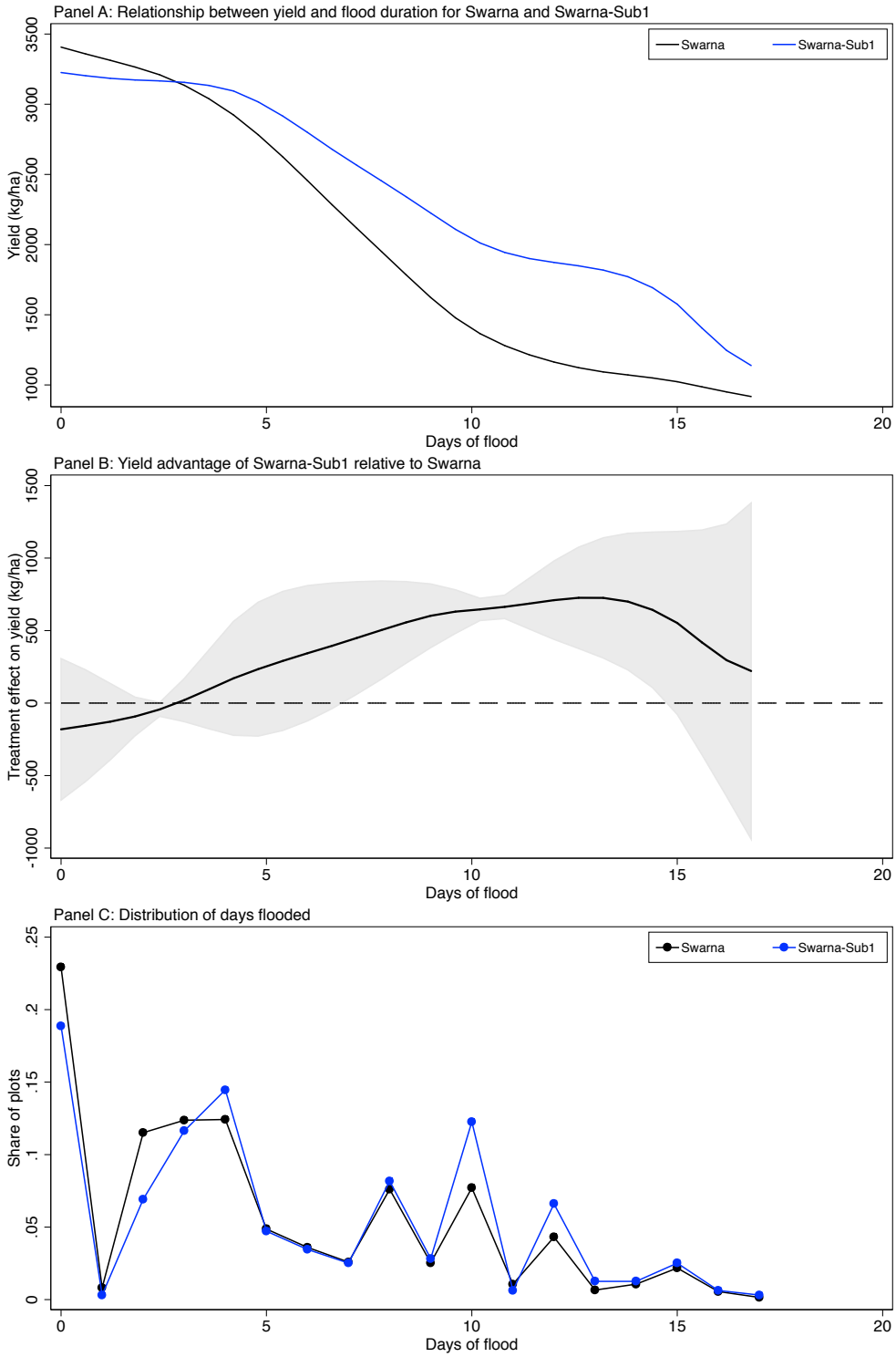


Figure 2: Projected impacts of adoption of flood-tolerant rice (Swarna-Sub1) on rice production. Panel A: Bhadrak district. Panel B: Balasore district. Graphs display forecasted percentage difference in total production between a scenario where all Swarna plots are cultivated with Swarna-Sub1 and a scenario where the plots remain cultivated with Swarna. Differing flood severity is simulated by adding (subtracting) an additional day of submergence for each plot to simulate a flood that is 1 day more (less) severe than 2011 floods. Dots and triangles represent point estimates and whiskers are 95% confidence intervals. Predicted difference in impact between SC/ST and OBC/General farmers is displayed as triangles. Regression estimates from Column 3 of Table 1 used to generate predictions. Standard errors are calculated using the delta method.

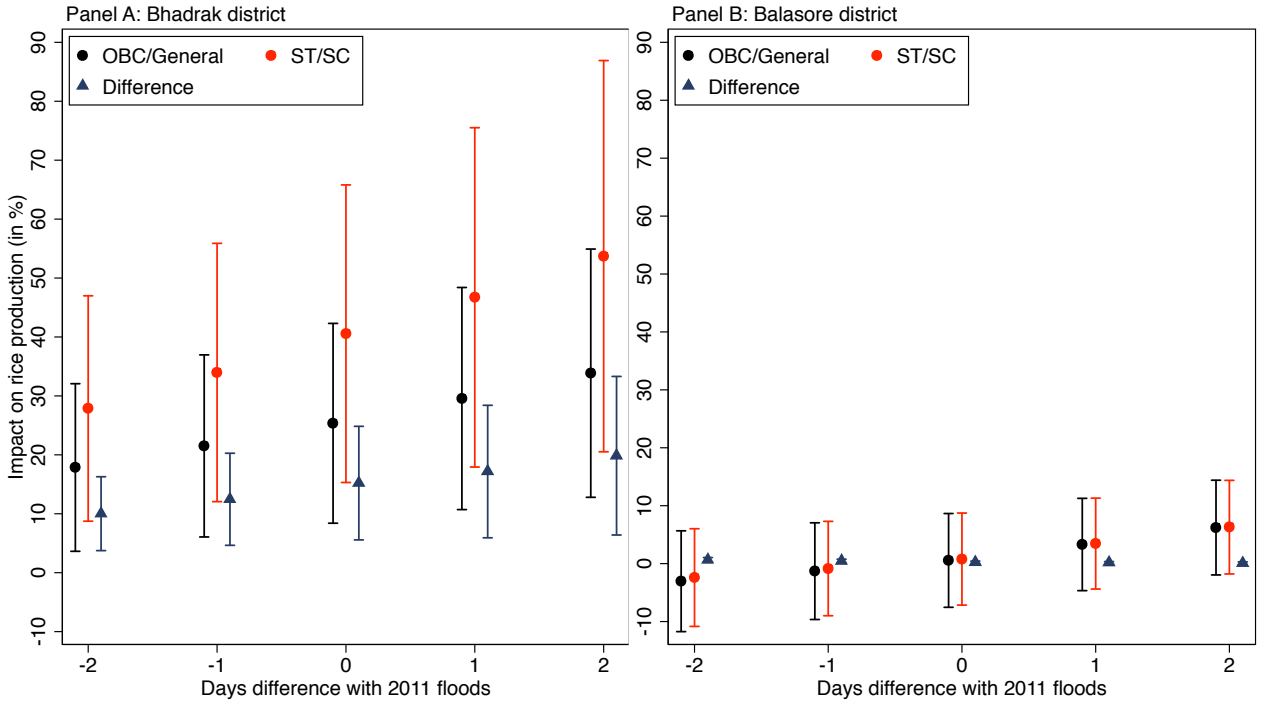
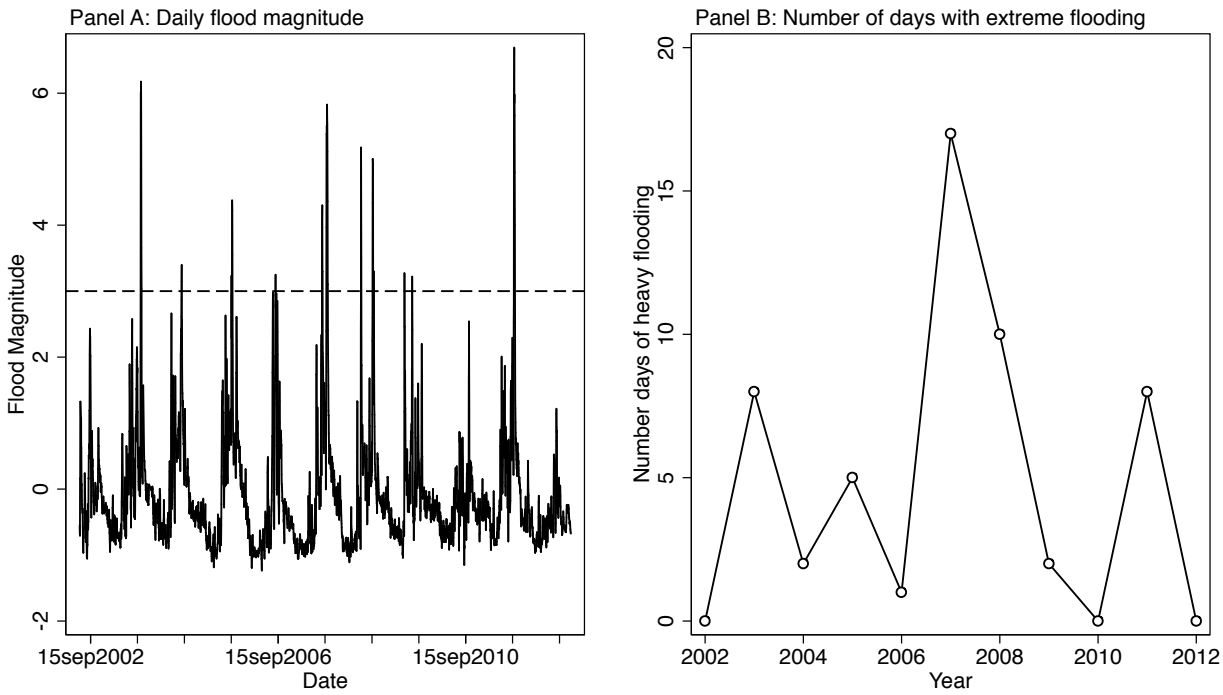


Figure 3: Occurrences of major floods at river discharge site near Brahmani River in Rajkanika block of Kendrapara district. Site is approximately 6 km from Bhadrak boundary and 9 km from nearest study village. Panel A: Daily river discharge used to measure flooding magnitude from June 2002 to December 2012. Flood magnitude is normalized (z-score). Magnitude above 3 considered to be extreme flooding. Panel B: Total number of days with extreme flooding (magnitude>3) for each year. Data come from Global Flood Detection System at Dartmouth Flood Observatory.



Supporting Information

Randomization verification

In Table S1 we show mean values of several variables for treatment and control farmers as well as mean values of village characteristics for treatment and control villages. The similarities between the observable characteristics suggest that both levels of randomization succeeded in generating groups that are similar in characteristics other than access to Swarna-Sub1.

Table S1: Comparison of observable characteristics in treatment and control households/villages. Panel A: Household level statistics for 1,248 households using 2012 survey. Treatment refers to households selected to receive Swarna-Sub1 minikit. Panel B: Village level statistics for 125 villages using 2001 Census of India. 3 of the 128 sample villages were not matched successfully to the 2001 census. Approximate elevation estimated using SRTM digital elevation layer matched to village centroid. Treatment refers to selection of village into treatment group where 5 farmers were selected to receive Swarna-Sub1 minikit.

	Control	Treatment	P-value of difference
Panel A: Farmer Level Statistics			
Land owned in acres	2.002	2.144	0.22
Share of cultivated land w/ irrigation	0.173	0.186	0.57
Sharecropper	0.152	0.178	0.27
HH has private tubewell	0.332	0.325	0.82
Education of farmer	6.896	6.946	0.83
Age of farmer	51.191	51.783	0.44
HH has thatched roof	0.557	0.548	0.78
HH has latrine	0.289	0.354	0.03**
HH has electricity	0.843	0.822	0.38
HH has below poverty line card	0.574	0.559	0.64
ST or SC	0.189	0.176	0.61
Panel B: Village Level Statistics			
Number of households	175.803	179.547	0.91
Population	927.049	974.656	0.79
Number of cultivators	106.443	112.203	0.78
Number of agricultural laborers	53.754	53.031	0.95
Share of village that is SC	0.178	0.229	0.16
Share of village that is ST	0.088	0.091	0.93
Literacy	0.597	0.596	0.98
Approximate elevation (m)	8.393	9.703	0.40

Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Location of villages

In Figure S1 we show a map of the approximate centroids of the 128 villages in the experiment. The villages (dots in the map) are spread across eight blocks. This figure also displays a visualization of the RADARSAT imagery of flooding that is used to estimation Equation (4) in the main text.

Robustness to Plot Controls

In Table S2 we show that our main regression results are robust to controlling for some observable plot characteristics. In Columns 1-3 we control for plot area, an indicator for medium land, an indicator for upland, an indicator for irrigation access, and indicator for an owned plot. We also interact these controls with the duration of flooding. Our results with the addition of these controls remain similar to those in Table 1 of the main text.

One important consideration is that the minikit provided is only enough to cultivate a plot of up to approximately 0.2 ha. Figure S2 shows the estimated density of plot size for Swarna and Swarna-Sub1 plots. The overlap between the two distributions occurs mostly in the range of 0-0.5 acres (0-0.2 ha). In Columns 4 and 5 of Table S2 we limit our sample to plots of 0.5 acres or less. The results are similar in this smaller sample where there is the most overlap between the distributions of plot size.

Description of Impact Simulation

The regression equation used to generate aggregate impact estimates is,

$$\begin{aligned} Yield_{pivb} &= \beta_0 + \beta_1 Sub1_{pivb} + \beta_2 Daysflood_{pivb} + \beta_3 Sub1_{pivb} * Daysflood_{pivb} + \beta_4 (Daysflood_{pivb} - 12) * \\ &\mathbb{1}(Daysflood_{pivb} > 12) + \beta_5 Sub1_{pivb} * (Daysflood_{pivb} - 12) * \mathbb{1}(Daysflood_{pivb} > 12) \\ &+ \beta_6 OMV_{pivb} + \beta_7 TV_{pivb} + \beta_8 OMV_{pivb} * Daysflood_{pivb} + \beta_9 TV_{pivb} * Daysflood_{pivb} + \alpha_b + \varepsilon_{ivbp}. \end{aligned} \tag{5}$$

The impact on aggregate production is therefore,

$$\sum Hectares_{pivb} * \left(\hat{\beta}_1 + \hat{\beta}_3 * Daysflood_{pivb} + \hat{\beta}_5 * (Daysflood_{pivb} - 12) * \mathbb{1}(Daysflood_{pivb} > 12)\right) \tag{6}$$

The estimated percentage increase in total production due to Swarna-Sub1 is the estimate in (6) divided by the predicted total production when all Swarna plots continue to be cultivated with Swarna. That is,

$$\frac{\sum Hectares_{pivb} * \left(\hat{\beta}_1 + \hat{\beta}_3 * Daysflood_{pivb} + \hat{\beta}_5 * (Daysflood_{pivb} - 12) * \mathbb{1}(Daysflood_{pivb} > 12)\right)}{\sum Hectares_{pivb} * \left(\hat{\beta}_0 + \hat{\beta}_2 * Daysflood_{pivb} + \hat{\beta}_4 * (Daysflood_{pivb} - 12) * \mathbb{1}(Daysflood_{pivb} > 12) + \hat{\alpha}_b\right)} \tag{7}$$

The impact estimate in (7) is estimated separately for each district / caste group pair and displayed in Figure 2 of the main text.

Table S2: Impact of flood-tolerant rice (Swarna-Sub1) on yield during 2011 wet season. All columns present OLS regression results where the dependent variable is rice yield in kg/ha. Plot controls are area, indicator for medium land, indicator for upland, indicator for irrigation access, and indicator for an owned plot. Columns 4 and 5 limit to plots of 0.5 acres or less (see Figure S2).

	All plots			Plot Area $\leq$ 0.5 Acres	
	(1)	(2)	(3)	(4)	(5)
Swarna-Sub1	171.07	-208.43	-303.03*	-118.03	-203.08
	(131.07)	(162.90)	(170.68)	(166.07)	(161.32)
Other modern variety	-201.66***	-297.39***	-297.47***	-560.38***	-508.07***
	(73.62)	(112.59)	(112.58)	(160.50)	(150.12)
Traditional variety	-472.47***	-863.43***	-860.17***	-1038.93***	-978.08***
	(125.20)	(173.69)	(173.80)	(255.23)	(229.45)
Days flood	-63.33***	-71.62***	-67.99**	-92.55***	-65.71
	(16.27)	(23.79)	(30.64)	(26.41)	(45.92)
Swarna-Sub1*Days flood		70.36***	94.93***	71.54**	79.41***
		(23.17)	(27.76)	(28.61)	(29.64)
Traditional variety*Days flood		54.99***	54.85***	45.53*	57.03**
		(16.64)	(16.55)	(24.82)	(24.96)
Other modern variety*Days flood		19.70*	19.75*	44.46***	44.69***
		(11.06)	(11.05)	(16.89)	(16.75)
(Days flood-12)*1 if Days flood > 12			-2.03	27.97	-19.64
			(37.76)	(45.41)	(51.55)
Swarna-Sub1*(Days flood-12)*1 if Days flood > 12			-205.07***	-171.32***	-177.68***
			(65.04)	(60.21)	(63.02)
Block Fixed Effects	Yes	Yes	Yes	Yes	Yes
Plot Controls	Yes	Yes	Yes	No	Yes
Plot Controls*Days Flood	No	Yes	Yes	No	Yes
Mean of Dep Variable	2221.76	2221.76	2221.76	2440.47	2450.03
Number of Observations	4138	4138	4138	2255	2226
R squared	0.465	0.472	0.473	0.373	0.398

Robust standard errors clustered at the village level are reported in parentheses. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Figure S1: Map of experimental villages and flooding during September 2008. All flood data are from RADARSAT. The flood data are only available for Bhadrak, Kendrapara, and Jajpur districts. River discharge station is the location where flood measurements are taken (see Figure 3 in main text)

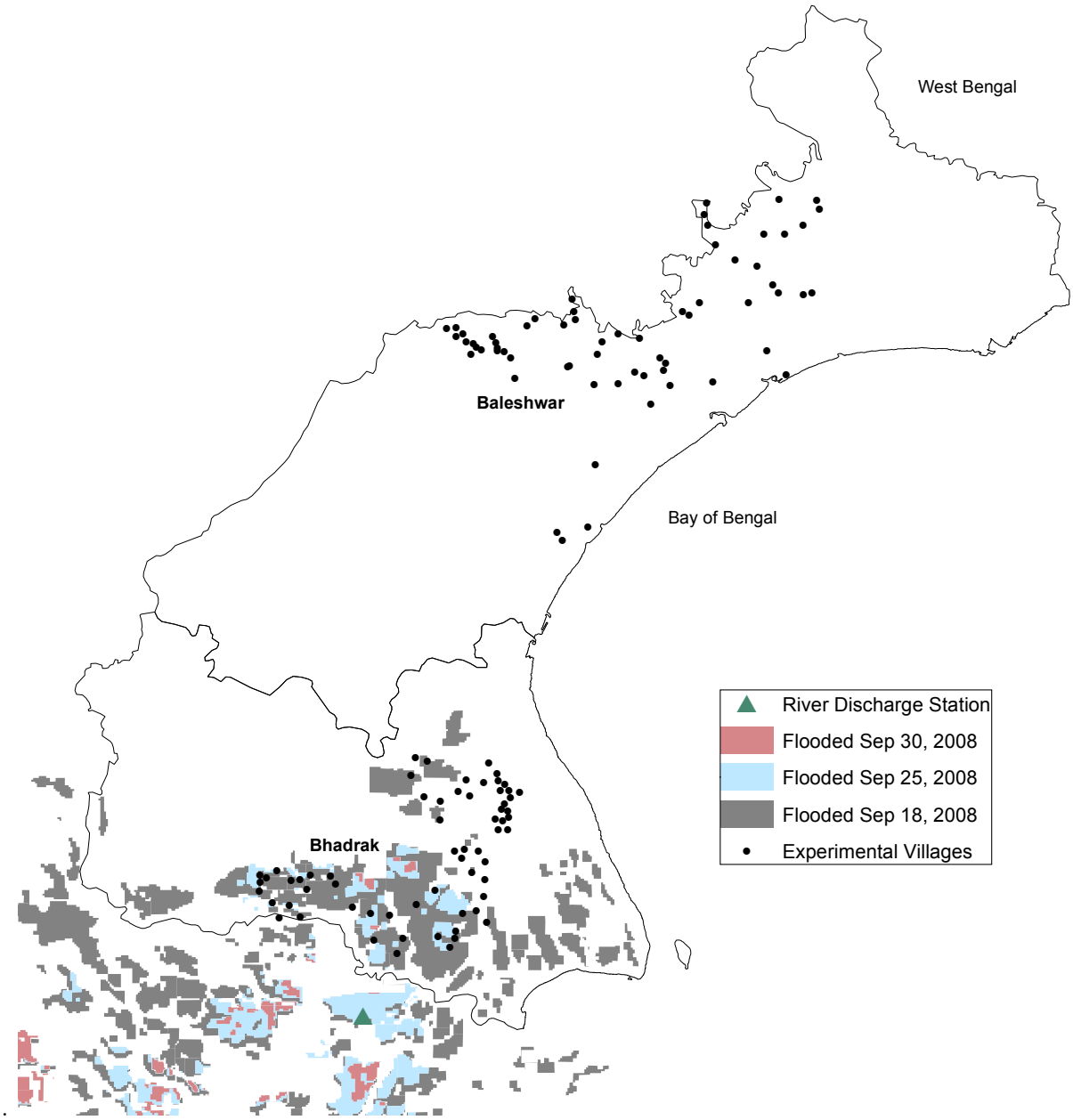
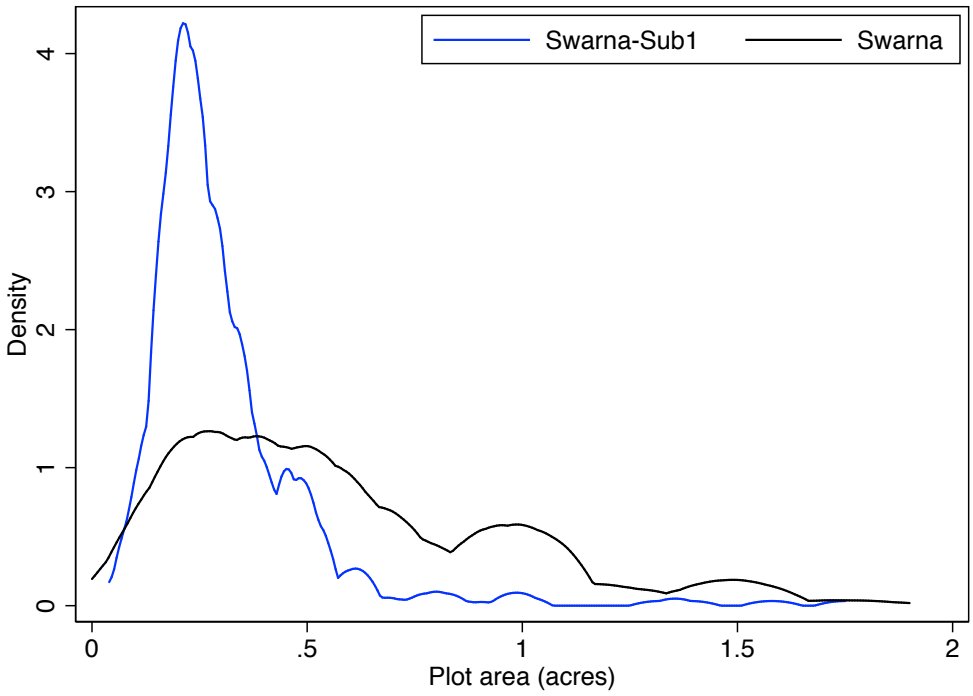


Figure S2: Estimated kernel density of plot area (in acres) for plots cultivated with Swarna and Swarna-Sub1. 102 plots of 2 acres or more are not included in the density estimates.



The Impact of Insurance Provision on Households' Production and Financial Decisions*

Jing Cai†
University of Michigan

October 21, 2012

Abstract

Taking advantage of a natural experiment and a rich household-level panel dataset, this paper tests the impact of an agricultural insurance program on household level production, borrowing, and saving. The empirical strategy includes both difference-in-difference and triple difference estimations. I find that, first, introducing insurance increases the production area of insured crops by around 20% and decreases production diversification; second, provision of insurance raises the credit demand by 25%; third, it decreases household saving by more than 30%; fourth, the effect of insurance on borrowing persists in the long-run, while the effect on saving is significant only in the medium-run; and fifth, the impact of insurance is greater on larger farmers and on households with lower migration remittance.

Keywords: Insurance; Production; Borrowing; Saving
JEL Codes: D14, G21, G22, O16, Q12

1 Introduction

Poor households in rural areas are exposed to substantial negative shocks such as weather disasters, which can generate large fluctuations in income and consumption if insurance markets are incomplete. To protect themselves from these risks, rural households undertake risk management and coping strategies such as informal insurance, avoiding high risk-high return agricultural activities, holding precautionary savings, and reducing investment in

*I am grateful to Alain de Janvry, Elisabeth Sadoulet, Michael Anderson, David Levine, Ethan Ligon, Jeremy Magruder, Craig McIntosh, Edward Miguel, and participants at the UC Berkeley Development Workshop for helpful suggestions and comments. I thank the Rural Credit Cooperative of Jiangxi province for providing the data. Financial support from the Center for Chinese Studies of UC Berkeley and the Institute of Business and Economic Research is greatly appreciated. All errors are my own.

†caijing@umich.edu

production (Morduch (1995), Rosenzweig and Stark (1989)). However, existing evidence shows that informal insurance mechanisms cannot effectively reduce negative impacts of regional weather shocks (Townsend (1994)). In the absence of formal insurance markets, the negative shocks and forgone profitable opportunities can lead to highly variable household income and persistent poverty (Dercon and Christiaensen (2011), Jensen (2000), Rosenzweig and Wolpin (1993)).

Although many developing countries have started to develop and market formal insurance products to shield farmers from risks, take-up is usually surprisingly low, even with heavy government subsidies[1]. While there is a growing literature studying ways to improve insurance demand (Cole et al. (2011), Cai (2012), Cai and Song (2011), Bryan (2010)), rigorous evaluations of the impacts of insurance provision are quite rare. With a rich household level panel data (2000-2008) from the Rural Credit Cooperative (RCC)[2] of China, this paper studies the effect of insurance provision on household's production, borrowing, and saving decisions. The program I am studying is a weather insurance policy for tobacco farmers offered by the People's Insurance Company of China (PICC), starting from 2003 in selected counties of Jiangxi province. It was expanded to more areas afterward and was implemented province-wide at the beginning of 2010. Purchase of insurance was made compulsory for tobacco farmers in treatment regions. I take advantage of the variation in insurance provision across both regions and household types (tobacco households vs. other households) to estimate the effect of insurance provision on household behavior, focusing on the initial stage of the policy in 2003.

The empirical strategy includes both difference-in-difference (DD) and triple difference (DDD) estimations. Because purchase of insurance in treatment regions was compulsory, household take-up decisions are not endogenous here. I use tobacco households outside of the treatment region to control for industry-specific trends in outcomes, and use non-tobacco households both within and outside the treatment region to control for region-specific trends in the absence of the policy intervention. Thus the extra changes in household behavior for tobacco households in treatment regions can be attributed to the insurance policy implementation. I find the following. First, insurance provision has a significantly positive effect on the production of the insured crop: it raises tobacco production by around 22% and decreases production diversification by around 29%. Second, insured households tend to borrow more from the rural bank for investment in tobacco production, and the

[1]For example, Giné et al. (2008) found a low take-up (4.6%) of a rainfall insurance policy among farmers in rural India in 2004, while Cole et al. (2011) found an adoption rate of 5% - 10% of a similar insurance policy in two regions of India in 2006

[2]RCC is the most important financial institution in rural China. It is the main provider of microcredit, and most farmers have saving accounts there.

magnitude of effect is about 25%. Third, the insurance policy decreases the household saving rate by more than 30%. Fourth, estimation of dynamic effects shows that, while the effect of insurance policy on both borrowing and saving became significant shortly after the policy was implemented, the impact on borrowing is persistent through the end of the sample period, while the effect on saving became significant several years after the intervention and decreased toward the end of the sample period. Finally, the impact of having insurance is greater on larger farmers and on households with lower migration remittance.

This paper contributes to the existing literature in the following ways. First, it provides insights on the literature about insurance take-up and impact. Estimating the causal effect of insurance policy on household behavior is made challenging by the endogenous insurance purchase decisions. There are a few papers studying the effects of insurance markets on household behavior using different estimation strategies. For example, Cole et al. (2011) use a randomized experiment which provided free rainfall insurance for selected farmers in India, and find that the insurance induced farmers to shift production towards higher-return but higher-risk cash crops. Karlan et al. (2012) use experimental methods and also find strong responses of investment in agriculture from insurance provision in Ghana. Gine and Yang (2009) implemented an experiment in Malawi which randomly bundled insurance with loans for selected farmers, and they found a negative effect of insurance on borrowing. Carter et al. (2007) use simulation method to show that insurance provision significantly improved producers' welfare, credit supply, and loan repayment in Peru. In contrast, Rosenzweig and Wolpin (1993) show by simulation that the gain from weather insurance for Indian farmers was minimal due to the existence of informal insurance mechanisms. This paper complements the existing literature by using rigorous estimation strategy to test both short-term and long-term effects of insurance provision on households' production, borrowing, and saving behavior in China, taking advantage of administrative borrowing and saving data from the rural bank. Because large and significant impacts of insurance policy are found in this paper, it supports the proposition that studying ways to improve voluntary insurance take-up is important.

Second, the paper contributes to the literature explaining low investment and technology adoption in developing countries. Credit constraints and the lack of information or knowledge are often proposed as explanations (Feder et al. (1985)). Duflo et al. (2011) argue that behavioral biases limit profitable agricultural investments. This paper shows that the riskiness of such investments is an important barrier, and therefore reducing risk can persistently improve investments.

The rest of the paper is organized as follows. Section 2 describes the background for the study and the insurance contract. Section 3 explains the data and summary statistics.

Section 4 presents estimation strategies and results, and section 5 concludes.

2 Background

Tobacco is an important cash crop in China. There are more than 2,000,000 rural households that live on tobacco production. The net profit of tobacco production is around 2000 RMB per mu[3], which is 3 to 5 times that of food crops such as rice.

In China, most tobacco producing counties are poor and mountainous areas. In the province that I study, there are 12 main tobacco production counties. Those counties are in two agricultural cities, Fuzhou and Ganzhou. Nearly half of those 12 counties are national poverty-stricken counties. To reduce poverty, in the late 1990s, these counties started to develop highly profitable tobacco industries by encouraging farmers to cultivate tobacco, organizing tobacco associations to teach farmers production techniques, etc. Taxes on tobacco production are now the main source of government revenue in these counties.

However, as other crops, tobacco production can be greatly influenced by weather risks. For example, in 2002, a flood destroyed most tobacco production in some of those 12 counties, which caused huge losses in household income and government revenue. The vice-head of Guangchang County, who is in charge of finance matters was previously a manager of an insurance company. He proposed to cooperate with insurance companies to shield tobacco farmers from frequent weather disasters in order to give them more incentives to continue tobacco production. In 2003, the People's Insurance Company of China (PICC) designed and offered the first tobacco production insurance program to households in four tobacco production counties, including Guangchang, Yihuang, Lean, and Zixi. The policy was extended to some other counties afterwards.

The insurance contract is as follows. The actuarially fair price estimated by the insurance company is 12 RMB per mu. The county and town level government gives a 50% subsidy on the premium, so farmers only pay the remaining half, around 6 RMB per mu. All households whose main source of income is tobacco production were required to buy the insurance for all their tobacco areas. The insurance covers natural disasters including heavy rain, flood, windstorm, extremely high or low temperature, and drought. If any of the above natural disasters happened and led to a 30% or more loss in yield, farmers were eligible to receive payouts from the insurance company. The amount of payout increases linearly with the loss rate in yield, with a maximum payout of 420 RMB. The loss rate in yield is investigated and determined by a group of insurance agents and agricultural experts[4]. The average net

[3] 1 RMB = 0.15 USD; 1 mu = 0.067 hectare

[4] For example, consider a farmer who has 5 mu in tobacco production. If the normal yield per mu is 500kg

income from cultivating tobacco is around 2000 RMB per mu, and the production cost is around 400 RMB to 600 RMB per mu (not including labor cost). Thus, this insurance program provides partial insurance that covers around 20% of the gross income or most of the production cost.

3 Theoretical Model

Here I provide a two period, two state model to show how the provision of insurance influences farmers' investment and financial decisions[5]. Intuitively, in the first period, insurance provision increases farmers' investment in production because the expected income from production is higher in that case. As a result, insurance has a negative effect on saving and a positive effect on borrowing. However, saving can be affected in two other ways. Because income uncertainty is reduced by insurance, people have less precautionary incentive to save, in this sense, saving tends to decrease. At the same time, if we assume that people have rational expectations, if they expect to become richer in future periods, they will smooth consumption across periods by increasing consumption and reducing saving in the current period. Furthermore, if the purchase of insurance is subsidized, this has a positive effect on farmers' wealth, which has a positive effect on saving.

Consider a representative farmer who lives for two periods with initial wealth W_0. In the first period, the farmer consumes C_1 and uses the remaining wealth for investment. There are two ways to invest this money: one is to save it in the bank with a saving interest rate R_f, the other is to invest it in a risky project like crop production which has a return function $F(\cdot)$. The farmer can borrow from a local bank for investment in a risky project with interest rate R_B. So the total investment I on the risky project includes the initial wealth less consumption and saving, and a loan equal to B from the bank. The return of the risky project is uncertain because it depends on whether a disaster happens in period one. In this simple model I assume that there are two states: a good state (no disaster) and a bad state (disaster). In the good state, the farmer gets $F(I)$, while in bad state he gets nothing. Assume that there is no strategic default and that farmers have limited liability, then in the good state, the farmer will repay fully in the second period; under a bad state, the farmer default on the loan if he does not have money to repay.

Suppose that for a farmer who invests I on the risky project (production), in order to buy

and because of a windstorm, the farmer's yield decreased to 250kg per mu, then the loss rate is 50% and he will receive 420*50% = 210 RMB per mu from the insurance company.

[5]Throughout the model I assume that farmers who are provided with insurance buy it in every period, because it is compulsory, while those who are not provided with insurance cannot buy it in any period.

an insurance which covers all his production[6], he needs to pay a premium which equals δI[7]. The production insurance works as follows: in the bad state, the farmer will be reimbursed by the insurance company by an amount equals to part of the cost invested in the risky project, γI. As a result, even in the bad state, the farmer who purchased insurance will be able to repay part or all of the loan.

In order to compare farmers' financial and investment behavior depending on whether they have insurance or not, I will solve the two-period model separately for insured and uninsured farmers because in the second period, their consumptions are different in the bad state. Throughout the model I assume that farmers are price takers: they don't think their behavior can influence either the premium charged by the insurance company or the saving and borrowing interest rate set by the bank.

3.1 Two-period model when insurance is not provided

The optimization problem as follows:

$$\max_{C_1,I,B} U(C_1) + E\beta U(C_2)$$
$$\iff \max_{C_1,I,B} U(C_1) + \beta p U\left[F(I) - (1+R_B)B + (1+R_f)S\right] + \beta(1-p)U\left[(1+R_f)S\right]$$
$$\text{s.t. } I = W_0 - C_1 - S + B$$

Assume that the return function and the utility function are:

$$F(I) = I^{\alpha}, \alpha < 1^{8}$$
$$U(C) = \log C$$

Then the first order conditions are:

$$U'(C_1) \;=\; \beta p U'\left[F(I) - (1+R_B)B + (1+R_f)S\right] F'(I) \;=\; \beta p U'(C_g) F'(I) \tag{3.1}$$

$$\beta p U'(C_g)\left[(1+R_f) - F'(I)\right] + \beta(1-p)U'\left[(1+R_f)S\right](1+R_f) = 0 \tag{3.2}$$
$$\beta p U'(C_g)\left[F'(I) - (1+R_B)\right] = 0 \tag{3.3}$$

$$\Rightarrow F'(I^*) = 1 + R_B{}^{9} \tag{3.4}$$

According to the return function form, I can rewrite equation (3.4) as:

[6]An assumption here is that to reduce the average risk and to prevent adverse selection, the insurance company requires the farmer to buy insurance for all his production area.

[7]In my data, δ should be quite low because farmers only need to pay 6 RMB per mu to buy the insurance, but the production cost (I) is around 400-600 RMB per mu.

[8]This return function form can exclude the case of infinite investment.

[9]This makes sense since project has return only in good states and it is the only time repayment is required.

$$F'(I^*) = \alpha I^{*\alpha-1} = 1 + R_B$$
$$\Rightarrow I^* = \left(\frac{1+R_B}{\alpha}\right)^{\frac{1}{\alpha-1}} \qquad (3.5)$$

So the optimal level of investment is decreasing in the borrowing interest rate R_B, or in other words, people tend to investment more on the risky project when the cost of borrowing is lower. Part 1 in Appendix A gives the solution of the above optimization problem.

3.2 Two-period model when insurance is provided

If a farmer has production insurance, the framework is as follows:

$$\max_{C_1,B,S} U(C_1) + \beta p U[C_g] + \beta(1-p)U[C_b]$$
$$s.t. I = B + [W_0 - C_1 - S - \delta I]$$
$$\Rightarrow I = \frac{W_0 - C_1 - S}{1+\delta} + \frac{B}{1+\delta}$$

Where C_g and C_b are the farmer's consumption in period two under good and bad state, respectively. The biggest difference in this model is that under bad state, the farmer receives a reimbursement from the insurance company which covers part of their cost, which equals $\gamma I = \gamma\frac{W_0-C_1-S}{1+\delta} + \gamma\frac{B}{1+\delta}$, so I can write the return of production under bad state as γI. Since I have assumed there's no strategical default, the farmer will repay the bank $\gamma\frac{B}{1+\delta}$, which is the return that is generated by a loan with size B. Given this, the consumption in period two under two states is defined as follows, respectively:

$$C_g = F(I) - (1+R_B)B + (1+R_f)S$$
$$C_b = \frac{\gamma}{1+\delta}(W_0 - C_1 - S + B) - \frac{\gamma}{1+\delta}B + (1+R_f)S$$

The three first order conditions are:

$$U'(C_1) - \beta p U'(C_g)F'(I)\frac{1}{1+\delta} - \beta(1-p)U'(C_b)\frac{\gamma}{1+\delta} = 0 \qquad (3.12)$$
$$\beta p U'(C_g)\left[-(1+R_B) + F'(I)\frac{1}{1+\delta}\right] = 0 \qquad (3.13)$$
$$\beta p U'(C_g)\left[(1+R_f) - F'(I)\frac{1}{1+\delta}\right] + \beta(1-p)U'(C_b)\left[-\frac{\gamma}{1+\delta} + 1 + R_f\right] = 0 \qquad (3.14)$$

The utility and return function forms are the same as that in previous sections:

$$U(C) = \log C$$
$$F(I) = I^\alpha, 0 < \alpha < 1$$

Part 2 in Appendix A gives the solution of the above optimization problem.

3.3 Combine the two models

The expressions of the optimal investment, consumption, saving and borrowing for insured and uninsured farmers are as follows:

$$I^*(insured) = \left(\frac{(1+R_B)(1+\delta)}{\alpha}\right)^{\frac{1}{\alpha-1}}$$
$$I^*(unisured) = \left(\frac{1+R_B}{\alpha}\right)^{\frac{1}{\alpha-1}}$$
$$C_1^*(insured) = \frac{1}{D+E}\left[\frac{(R_B-R_f)\gamma+(1+R_B)\left[(1+\delta)(1+R_f)-\gamma\right]}{(1+R_B)(1+\delta)\left[(1+\delta)(1+R_f)-\gamma\right]}W_0 + (\alpha^{-1}-1)\left(\frac{1+R_B}{\alpha}\right)^{\frac{1}{\alpha-1}}(1+\delta)^{\frac{1}{\alpha-1}}\right]$$
$$C_1^*(uninsured) = \frac{1}{1+\beta}\left[W_0 + (\alpha^{-1}-1)\left(\frac{1+R_B}{\alpha}\right)^{\frac{1}{\alpha-1}}\right]$$
$$S^*(insured) = \frac{(1+R_B)(1+\delta)}{R_B-R_f}\frac{E}{D+E}\frac{(R_B-R_f)\gamma+(1+R_B)\left[(1+\delta)(1+R_f)-\gamma\right]}{(1+R_B)(1+\delta)\left[(1+\delta)(1+R_f)-\gamma\right]}W_0$$
$$+\frac{(1+R_B)(1+\delta)}{R_B-R_f}\frac{E}{D+E}(\alpha^{-1}-1)\left(\frac{1+R_B}{\alpha}\right)^{\frac{1}{\alpha-1}}(1+\delta)^{\frac{1}{\alpha-1}} - \frac{\gamma W_0}{(1+R_f)(1+\delta)-\gamma}$$
$$S^*(unisured) = \frac{(1+R_B)(1-p)\beta}{(1+\beta)(R_B-R_f)}\left[W_0 + (\alpha^{-1}-1)\left(\frac{1+R_B}{\alpha}\right)^{\frac{1}{\alpha-1}}\right]$$
$$B^* = (1+R_B)^{\frac{1}{\alpha-1}}(1+\delta)^{\frac{\alpha}{\alpha-1}}\alpha^{-\frac{\alpha}{\alpha-1}} - \frac{D}{1+R_B}C_1^* + \frac{1+R_f}{1+R_B}S^*$$
$$B^*(unisured) = (1+R_B)^{\frac{1}{\alpha-1}}\alpha^{-\frac{\alpha}{\alpha-1}} - \frac{\beta\left[p(R_B+1)-(1+R_f)\right]}{R_B-R_f}C_1^*$$

3.4 Break-even conditions of the bank

Now I have solved farmers' optimization problem, the next step is to consider the break-even conditions of the bank[10].

If the bank's client does not have insurance, he gets nothing in bad state, so the break-even condition is:

$$B(1+R_f) = p(1+R_B)B$$
$$\Rightarrow R_B = [1+R_f]\frac{1}{p} - 1$$

If insurance is purchased, the break-even condition becomes:

$$(1+R_f)B = p(1+R_B)B + (1-p)\frac{\gamma}{1+\delta}B$$
$$\Rightarrow R_B = \left[1+R_f-\frac{(1-p)\gamma}{1+\delta}\right]\frac{1}{p} - 1.$$

In summary:

$$R_B = \begin{array}{ll} [1+R_f]\frac{1}{p} - 1, & \text{if not insured} \\ \left[1+R_f-\frac{(1-p)\gamma}{1+\delta}\right]\frac{1}{p} - 1, & \text{if insured} \end{array}$$

We can see that the bank will set a lower interest rate for people who have insurance because their repayments are better guaranteed.

[10] Here I assume that the institution's objective is to break-even for simplicity.

3.5 Conclusion of the model

Now I plug the interest rate into optimal decisions in 3.3 and compare the magnitude of investment, consumption, saving and borrowing between insured and uninsured farmers.

- Investment: Farmers will invest more when they have insurance

$$I^*(insured) = \left(\frac{(1+R_B)(1+\delta)}{\alpha}\right)^{\frac{1}{\alpha-1}} = \left(\frac{\frac{1+R_f}{p} - \frac{(1-p)\gamma}{(1+\delta)p}}{\alpha}\right)^{\frac{1}{\alpha-1}}$$

$$I^*(unisured) = \left(\frac{1+R_B}{\alpha}\right)^{\frac{1}{\alpha-1}} = \left(\frac{\frac{1+R_f}{p}}{\alpha}\right)^{\frac{1}{\alpha-1}}$$

Because $\alpha - 1 < 0$, so if $\frac{(1-p)\gamma}{(1+\delta)p} > 0$, the investment increase as a result of insurance provision. Intuitively, when insurance is provided, borrowing becomes cheaper and the expected return of the risky project will increase, so investing in the risky project becomes more attractive.

- Consumption: The first period consumption is higher when the farmer have insurance.

$$C_1^*(insured) = C_1^*$$

$$= \frac{1}{D+E}\left[\frac{(R_B-R_f)\gamma+(1+R_B)\left[(1+\delta)(1+R_f)-\gamma\right]}{(1+R_B)(1+\delta)\left[(1+\delta)(1+R_f)-\gamma\right]}W_0 + (\alpha^{-1}-1)(\frac{1+R_B}{\alpha})^{\frac{1}{\alpha-1}}(1+\delta)^{\frac{1}{\alpha-1}}\right]$$

$$=$$

$$\frac{1}{1+\beta}\left\{\left[\frac{1+R_B}{1+R_f} + \frac{(R_B-R_f)\gamma}{(1+R_f)[(1+R_B)(1+\delta)-\gamma]}\right]W_0 + \frac{(1+\delta)(1+R_B)\left[(1+R_f)(1+\delta)-\gamma\right]}{(1+R_f)[(1+R_B)(1+\delta)-\gamma]}(\alpha^{-1}-1)(\frac{R_f/p+1/p-(1-p)\gamma/p(1+\delta)}{\alpha})^{\frac{1}{\alpha-1}}(1+\delta\right.$$

$$C_1^*(unisured) = \frac{1}{1+\beta}\left[W_0 + (\alpha^{-1}-1)\left(\frac{R_f/p+1/p}{\alpha}\right)^{\frac{1}{\alpha-1}}\right]$$

Because $\frac{1+R_B}{1+R_f} + \frac{(R_B-R_f)\gamma}{(1+R_f)[(1+R_B)(1+\delta)-\gamma]} > 1$, $(\frac{R_f/p+1/p-(1-p)\gamma/p(1+\delta)}{\alpha})^{\frac{1}{\alpha-1}} > \left(\frac{R_f/p+1/p}{\alpha}\right)^{\frac{1}{\alpha-1}}$
and $(1+R_f)(1-p)(1+\delta-\delta\eta) > R_f\delta\eta$[11]
then $C_1^*(insured) > C_1^*(unisured)$

So the second message from the model is that, people who bought insurance will consume more in the first period. This is because if a farmer has insurance, he expect himself to be richer in the second period compared to the condition when he does not have insurance, so he will smooth the consumption between periods by increasing the consumption in period one.

- Saving: The provision of insurance can decrease farmers' total saving and saving rate in period one.

[11] This condition holds in my data.

$$S^*(insured) = \frac{(1+R_B)(1+\delta)}{R_B-R_f}\frac{E}{D+E}\frac{(R_B-R_f)\gamma+(1+R_B)\left[(1+\delta)(1+R_f)-\gamma\right]}{(1+R_B)(1+\delta)\left[(1+\delta)(1+R_f)-\gamma\right]}W_0+$$
$$\frac{(1+R_B)(1+\delta)}{R_B-R_f}\frac{E}{D+E}(\alpha^{-1}-1)(\frac{1+R_B}{\alpha})^{\frac{1}{\alpha-1}}(1+\delta)^{\frac{1}{\alpha-1}} - \frac{\gamma W_0}{(1+R_f)(1+\delta)-\gamma}$$
$$= \left[\frac{\beta}{1+\beta} - \frac{\beta\gamma p}{(1+\beta)\left[(1+\delta)(1+R_f)-\gamma\right]}\right]W_0 + (\alpha^{-1}-1)(\frac{1+R_B}{\alpha})^{\frac{1}{\alpha-1}}(1+\delta)^{\frac{1}{\alpha-1}}\frac{(1+R_B)(1+\delta)}{R_B-R_f}\frac{E}{D+E}$$
$$S^*(unisured) = \frac{\beta}{(1+\beta)}\left[W_0 + (\alpha^{-1}-1)\left(\frac{1/p+R_f/p}{\alpha}\right)^{\frac{1}{\alpha-1}}\right]$$
$$= \frac{\beta}{1+\beta}W_0 + (\alpha^{-1}-1)\alpha^{-\frac{1}{\alpha-1}}(\frac{1}{p}+\frac{R_f}{p})^{\frac{1}{\alpha-1}}\frac{\beta}{1+\beta}$$

Because $\left[\frac{\beta}{1+\beta} - \frac{\beta\gamma p}{(1+\beta)\left[(1+\delta)(1+R_f)-\gamma\right]}\right] < \frac{\beta}{(1+\beta)}$, so if W_0 is large enough, $S^*(insured) < S^*(unisured)$ and $Savingrate^*(insured) < Savingrate^*(unisured)$. This result is consistent with the precautionary saving story: farmers' future income uncertainty is decreased by introducing insurance, so people have less precautionary incentive to save in the first period for smoothing future consumption.

- Borrowing: The effect of insurance provision on borrowing is ambiguous.

The total investment on risky project is $I = B + [W_0 - C_1 - S]$, I have proved that the provision of insurance will increase C_1 and I, and decrease S, so the effect on B is ambiguous.

In summary, the conclusion from this two-period model is that insurance has a positive effect on investment in risky projects and consumption, and it reduces farmers' total saving and saving rate. As a result, its effect on borrowing is not determined.

4 Data and Summary Statistics

As shown in Table 1, the empirical analysis is based on data from 12 tobacco production counties in Jiangxi province of China: Guangchang, Yihuang, Lean, Zixi, Shicheng, Ningdu, Ganxian, Huichang, Xinfeng, Xinguo, Ruijin, and Quannan. Among these twelve counties, only tobacco farmers in Guangchang, Yihuang, Lean, and Zixi were eligible to buy the tobacco insurance policy after 2002. In eligible counties, only tobacco households whose main source of income is from tobacco production were offered insurance, while households working in other activities were not eligible to buy similar products.

The primary data source is the household level panel dataset, ranging from 2000 to 2008, provided by the Rural Credit Cooperatives (RCC). The whole sample includes around 6500 households. The data is composed of two parts. The first part is the administrative data of RCC, including their clients' saving and borrowing information[12]. Specifically, it has

[12]Because more than 90% of farmers in Jiangxi province are RCC clients, this data is representative of the whole sample of farmers

variables such as loan certification number, total borrowing during the year, interest rate, use of loan, repayment, total annual saving, savings in the deposit account, savings in the current account, and annual growth in savings[13]. The second part is RCC annual survey data[14], which contains two broad categories of information. The first is family background information: age, national ID, gender, occupation and education of household heads, primary and secondary source of household income, family address, and household size. The second is household income and production, including total annual income, household income from different sources, remittance income, area of land for cultivation, and production areas of different crops.

The data includes 6548 households in total, of which 3580 households are tobacco households, and 2968 households are other households whose main source of income is not tobacco production[15]. For tobacco households, 1429 of them are in the treatment region where the insurance policy was available, and 2151 of them are in control regions.

The summary statistics of key variables before the insurance policy was implemented (2000-2002) are provided in Table 2. Household heads are almost exclusively male and the average age is around 40. The average household size is around five people, and household heads have an average education level of between primary and secondary school. The above household characteristics are very similar across different household groups. The average annual household income of tobacco households in treatment regions equals 10,650 RMB, while that of tobacco households in control regions is a bit higher, around 12,000 RMB. Annual income of non-tobacco households is much lower, with only 7,270 RMB. Considering households' borrowing behavior, the average borrowing of non-tobacco households is the highest (4,980 RMB), followed by tobacco households in control regions (4,560 RMB), and tobacco households in treatment regions (3,900 RMB). The household saving rate is defined as the ratio between net annual saving and household income. For tobacco households in treatment regions, the saving rate is around 3.6%, which is lower than that of tobacco households in control regions (4.5%). Saving rate of non-tobacco households is similar as that of tobacco households in treatment regions, of around 3.4%[16]. This table suggests that, as treatment and control tobacco households behave statistically differently in pre-policy

[13]While RCC is the main place for farmers to make deposits, households may have saving accounts in other institutions. As a result, the amount of saving in RCC does not represent a household's total saving. To account for this factor, RCC reported the village-level ratio of RCC saving to total household saving. I adjusted the RCC saving data by this ratio in all of the empirical analyses

[14]RCC implements a household survey every year in order to adjust the lending interest rate and loan ceiling for each household

[15]These households work in agricultural activities such as rice production, cultivation, etc. or in non-agricultural activities

[16]Households with outliers (the lowest or highest 1%) in income, loan size, and savings were deleted from the sample for analysis

periods, I cannot study the policy impact by taking a simple difference.

In order to check whether Difference-in-Difference (DD) estimation can be a convincing strategy in this context, I test the common-trend assumption in Table 3, using the following regression:

$$Y_{irt} = \eta_0 + \eta_1 Year_t + \eta_2 Insurance_{ir} + \eta_3 Year_t * Insurance_{ir} + \epsilon_{irt} \tag{1}$$

Where i, r, t are household, region, and year indices respectively. $Insurance_{ir}$ is the treatment indicator equal to 1 for treatment regions and 0 for control regions. The common-trend assumption does not hold if the coefficient of the interaction term, η_3, is statistically significant. Results show that the common trend assumption is not valid for all outcomes in which I am interested in, so using only DD estimation is not sufficient.

To get a basic sense of how insurance provision impacts production, borrowing, and saving, I plot the evolution of these variables in Figures 1 to 5. Figure 1 shows that, while tobacco production was similar for tobacco households in treatment and control regions before insurance was in place, production increased greatly in treatment regions after 2002. Referring to Figure 2, we can see that, while tobacco households in treatment regions borrowed less than those in control regions before 2002, the pattern reversed after 2003. However, Figure 3 shows that the borrowing pattern is different across the sample period between non-tobacco households in treatment and control regions, which suggests that there might be some regional-specific trend for which we should control when estimating the policy effect. In Figure 4, I show that, for tobacco households, while the saving rate is higher in control regions than in treatment regions, the trend reversed slightly after 2004. The difference in saving rates of non-tobacco households between treatment and control regions is much larger, as shown in Figure 5.

Table 4 reports the average area of tobacco production, size of loans, and saving rate by time period, region, and sector eligibility. Consider loan size for example, for each region-sector category, the average loan size increases from the period 2000-2002 to the period 2003-2008, reflecting the aggregate economic trend. For tobacco households, the average loan size in treatment regions increases by 1,450 RMB more than that of households in control regions. This could be a result of both the implementation of the insurance policy and other region-specific changes. For example, for non-tobacco households, the average loan size also grows faster in treatment regions than in control regions, by 480 RMB. Taking into account this regional difference in the absence of the insurance policy, the loan size for tobacco households in treatment regions increases by 970 RMB more than that for tobacco households in control regions. The regression analysis in the next section demonstrates that this effect is robust to controlling for other confounding factors. These results suggest that

triple difference estimation can be a more convincing empirical strategy than DD in this case.

5 Estimation Strategies and Results

5.1 Empirical Strategies

The implementation of the tobacco insurance policy introduced variations in insurance provision in three dimensions: years before and after the policy was introduced, regions with and without the policy, and eligible and ineligible households (tobacco households v.s. non-tobacco households). These variations allow me to use both difference-in-difference (DD) and difference-in-difference-in-difference (DDD) estimation as the empirical strategy. First, the DD analysis compares the change in tobacco households' behavior in treatment regions before and after 2002 with that of tobacco households in control regions, assuming that tobacco households in treatment and control regions follow the same trend in the absence of the provision of insurance policy. The estimation equation is as follows:

$$Y_{irt} = \alpha_0 + \alpha_1 After_{it} + \alpha_2 Insurance_{ir} + \alpha_3 After_{it} * Insurance_{ir} + \epsilon_{irt} \tag{2}$$

Where i, r, t are household, region, and year indices respectively. This framework is based on tobacco households only. Y represents outcome variables including tobacco production area, size of loan borrowed from the rural bank, and saving rate. $After$ is a dummy variable equal to 1 for the 2000-2002 period and 0 for years 2003-2008, which reflects the impact on outcomes of time-varying aggregate economic environment and policies. $Insurance_{ir}$ is the treatment indicator equal to 1 for treatment regions and 0 for control regions. The coefficient of interest is the one before the interaction term, between $After$ and $Insurance_{ir}$, α_3.

However, the DD estimation cannot remove all confounding factors. For example, there may be some other contemporary changes in the economic environment or other policies specific to the treatment region that can influence households' production and financial decisions. This can be captured by taking another DD analysis, which compares behavior of non-tobacco households in treatment regions before and after 2002 with that of non-tobacco households in control regions. As a result, the DDD framework, which takes the difference between the two differences from the first two steps, can further control for region-specific trends. Under the DDD framework, we don't need to assume that behaviors of tobacco households in both treatment and control regions evolve similarly in expectation, but only need to assume that the difference affects tobacco households and other households similarly (in other words, there are no other region-sector specific policy changes). I will test this

assumption later. The DDD regression is as follows:

$$\begin{aligned} Y_{ijrt} = \beta_0 + & \beta_1 After_{it} + \beta_2 Insurance_{ir} + \beta_3 Tobacco_{ij} + \beta_4 After_{it} * Insurance_{ir} \\ & + \beta_5 After_{it} * Tobacco_{ij} + \beta_6 Tobacco_{ij} * Insurance_{ir} \\ & + \beta_7 After_{it} * Insurance_{ir} * Tobacco_{ij} + \epsilon_{ijrt} \end{aligned} \tag{3}$$

Where j is sector indicator, and $Tobacco_{ij}$ is a dummy variable equal to 1 for tobacco households and 0 otherwise. The coefficient of the time, region, and sector interaction (β_7) captures the average effect of insurance provision on household behavior, after other confounding factors are removed.

Significant influences of insurance provision on households' production and investment decisions may take place either shortly after the policy was introduced or several years later, and the magnitude of the effect may change over time. Consequently, it would be interesting to test the dynamic effect of insurance provision on household behavior. The estimation equation is as follows:

$$\begin{aligned} Y_{ijrt} = \rho_0 + & \rho_1 Year_t + \rho_2 Insurance_{ir} + \rho_3 Tobacco_{ij} + \rho_4 Year_t * Insurance_{ir} \\ & + \rho_5 Year_t * Tobacco_{ij} + \rho_6 Tobacco_{ij} * Insurance_{ir} \\ & + \rho_7 Year_t * Insurance_{ir} * Tobacco_{ij} + \epsilon_{ijrt} \end{aligned} \tag{4}$$

Where $Year_t$ includes a set of year dummies. Estimating the above equation not only allows me to test the dynamic effect, but also to test the crucial assumption that validates the DDD estimation: in the absence of the insurance policy, the production and financial behaviors of tobacco households and non-tobacco households should evolve similarly.

The magnitude of the impact of insurance provision on household behavior can be different for different groups of households. I consider two types of heterogeneity here, depending on farming size and the importance of migration remittance in household income. The regression is as follows:

$$\begin{aligned} Y_{ijrt} = \gamma_0 + & \gamma_1 After_{it} + \gamma_2 Insurance_{ir} + \gamma_3 Tobacco_{ij} + \gamma_4 After_{it} * Insurance_{ir} \\ & + \gamma_5 After_{it} * Tobacco_{ij} + \gamma_6 Tobacco_{ij} * Insurance_{ir} \\ & + \gamma_7 After_{it} * Insurance_{ir} * Tobacco_{ij} + \gamma_8 Index_{it} + \gamma_9 Index_{it} * After_{it} \\ & + \gamma_{10} Index_{it} * Insurance_{ir} + \gamma_{11} Index_{it} * Tobacco_{ij} + \gamma_{12} Index_{it} * After_{it} * Insurance_{ir} \\ & + \gamma_{13} Index_{it} * After_{it} * Tobacco_{ij} + \gamma_{14} Index_{it} * Insurance_{ir} * Tobacco_{ij} \\ & + \gamma_{15} Index_{it} * After_{it} * Insurance_{ir} * Tobacco_{ij} + \epsilon_{ijrt} \end{aligned} \tag{5}$$

Where $Index_{it}$ is an indicator equal to 1 if, in the pre-policy period (2000-2002), the households' total production area or the percentage of migration remittance in total income is higher than the sample median, and 0 otherwise. The coefficient of interest is γ_{15}.

5.2 Estimation Results

Tables 5 - 8 report DD and DDD estimation results on the effect of insurance provision on households' production, borrowing, and saving decisions, respectively[17]. Look first at the effect on production. Refer to Column (1) in Table 5, the increase in tobacco production post of 2002 is 1.161 mu larger for households in treatment regions compared with households in control regions. Because the pre-policy mean of tobacco production in treatment regions is about 5.25 mu (refer to Table 2), this result means that insurance provision can raise tobacco production by 22%. This is consistent with the story that, as the expected return of tobacco production increases once insurance is provided, insurance gives households greater incentives to invest more heavily in tobacco production. Column (2) includes year dummies in addition, and the magnitude of the effect increased slightly. In Column (3), I further control for household characteristics, including household size, annual household income, age, and education of household head. The magnitude of the treatment effect remains similar, at around 1.2 mu (23%). Column (3) also shows that households with higher annual income tend to produce more tobacco, as the production cost of tobacco cultivation is high relative to that of other crops. Moreover, larger households, and those with more well-educated and younger household heads, are likely to have a larger production scale. This can be explained by the fact that tobacco production not only requires more labor than other production, but also thorough knowledge of the techniques necessary to have high yield and good quality tobacco.

In Table 6, I look at the impact of insurance provision on households' choice of production diversification, which is defined as one minus the Herfindahl index of agricultural production. The results show that agricultural production became less diversified after the insurance was provided, by around 29%. This means that households tend to focus more on producing the insured crop after the intervention.

Second, Table 7 reports the DDD estimation results on the effect of insurance on borrowing. Results suggest a significant insurance treatment effect on borrowing, of around 972 RMB. Comparing this result to the average loan size of tobacco households in treatment regions before 2003 (shown in Table 2) tells us that tobacco households borrow 25% more once their production is insured.

[17]Please note that the DDD framework is not applicable to estimating the effect on tobacco production area, because there is almost no tobacco production for non-tobacco households

Third, the effect of insurance provision on household saving is reported in Table 8. According to Columns (1) and (2), after the tobacco insurance policy was introduced, the increase in the average saving rate of tobacco households in treatment regions is around 1.24 percentage points lower than that of tobacco households in control regions. This means that providing insurance can decrease a household's saving rate by more than 30%. In Columns (3) and (4), I consider the level of net saving rather than the saving rate. The estimation results show that, while the insurance policy has a significant impact on saving rate, it does not significantly influence the level of saving. Finally, in Columns (5) and (6), I estimate the effect of insurance on the composition of saving. In China, households can have two types of saving accounts: fixed-term saving or flexible-term saving (like checking accounts in the United states). I show that the insurance policy does not have any statistically significant impact on the composition of saving.

The dynamic impact of insurance provision on households' borrowing and saving behavior is illustrated in Table 9. The result shows that first, before the insurance policy was introduced, there is no significant difference between households with or without tobacco production, because interactions of 2001-2002 year dummies, region, and sector are insignificant. Second, according to Column (1), the effect of insurance provision on borrowing is insignificant until two years after the intervention. However, both the magnitude and significance of the effect persists through the end of our sample period. In contrast, according to Column (2), insurance impact on household saving become significant three years after the policy was introduced, but the magnitude and significance decrease and become insignificant toward the end of the sample period.

In Table 10, I report the heterogeneity in the impact of insurance, depending on how large the farming size is, and how important is the migration remittance to the household's income. Columns (1) - (3) shows that insurance provision has a larger effect on borrowing for large farmers, while the effect on production and saving is not statistically different for farmers with different farming sizes. In Columns (4) - (6), I show that the effect of insurance policy has a smaller impact on the production and borrowing decisions of households who depend more on migration remittance.

Once the insurance policy was implemented for tobacco farmers, we may expect an endogenous switch of non-tobacco households to tobacco households. If a significant number of households do so, the effect might be overestimated. In Table 11, I report the percentage of households that stay in the same sector, switch from tobacco to the non-tobacco sector, and switch from the non-tobacco sector to the tobacco sector between the previous and current year, for treatment and control regions. This table shows that only a very small fraction of households changed sectors during the sample period. I did a robustness check by excluding

all households that had ever switched sectors and it does not change the effect much.

6 Conclusions

Household incomes in developing rural economies are subject to great uncertainty. As a result, many developing countries are making efforts to improve the quality and coverage of agricultural insurance products. Taking advantage of a natural experiment of insurance provision in rural China, this paper uses both DD and DDD estimations to study the effect of insurance provision on households' production and financial decisions. I find that households tend to increase tobacco production once it is insured. Moreover, insurance not only makes households borrow more from the bank, but also decrease the household saving rate. However, while the impact of insurance on borrowing persists in the long-run, the impact on saving is only significant in the medium-run and vanishes in the long-run.

References

Bryan, Gharad, "Ambiguity and Insurance," *Working Paper*, 2010.

Cai, Jing, "Social Networks and the Decision to Insure: Evidence from Randomized Experiments in China," *Working Paper*, 2012.

_ **and Changcheng Song**, "Insurance Take-up in Rural China: Learning from Hypothetical Experience," *Working Paper*, 2011.

Carter, Michael R., Francisco Galarza, and Stephen Boucher, "Underwriting area-based yield insurance to crowd-in credit supply and demand," MPRA Paper 24326, University Library of Munich, Germany November 2007.

Cole, Shawn, Petia Topalova, , Xavier Gene, Jeremy Tobacman, Robert Townsend, and James Vickery, "Barriers to Household Risk Management: Evidence from India," Working Papers id:4293, eSocialSciences July 2011.

Dercon, Stefan and Luc Christiaensen, "Consumption risk, technology adoption and poverty traps: Evidence from Ethiopia," *Journal of Development Economics*, November 2011, *96* (2), 159–173.

Duflo, Esther, Michael Kremer, and Jonathan Robinson, "Nudging Farmers to Use Fertilizer: Theory and Experimental Evidence from Kenya," *American Economic Review*, October 2011, *101* (6), 2350–90.

Feder, Gershon, Richard E Just, and David Zilberman, "Adoption of Agricultural Innovations in Developing Countries: A Survey," *Economic Development and Cultural Change*, January 1985, *33* (2), 255–98.

Gine, Xavier and Dean Yang, "Insurance, credit, and technology adoption: Field experimental evidencefrom Malawi," *Journal of Development Economics*, May 2009, *89* (1), 1–11.

Giné, Xavier, Robert Townsend, and James Vickery, "Patterns of Rainfall Insurance Participation in Rural India," *World Bank Economic Review*, October 2008, *22* (3), 539–566.

Jensen, Robert, "Agricultural Volatility and Investments in Children," *American Economic Review*, May 2000, *90* (2), 399–404.

Karlan, Dean, Robert Osei, Isaac Osei-Akoto, and Christopher Udry, "Agricultural Decisions after Relaxing Credit and Risk Constraints," *Working Paper, Yale University*, 2012.

Morduch, Jonathan, "Income Smoothing and Consumption Smoothing," *Journal of Economic Perspectives*, Summer 1995, *9* (3), 103–14.

Rosenzweig, Mark R and Kenneth I Wolpin, "Credit Market Constraints, Consumption Smoothing, and the Accumulation of Durable Production Assets in Low-Income Countries: Investment in Bullocks in India," *Journal of Political Economy*, April 1993, *101* (2), 223–44.

_ **and Oded Stark**, "Consumption Smoothing, Migration, and Marriage: Evidence from Rural India," *Journal of Political Economy*, August 1989, *97* (4), 905–26.

Townsend, Robert M, "Risk and Insurance in Village India," *Econometrica*, May 1994, *62* (3), 539–91.

Figure 1. Evolution of Tobacco Production, by Treatment

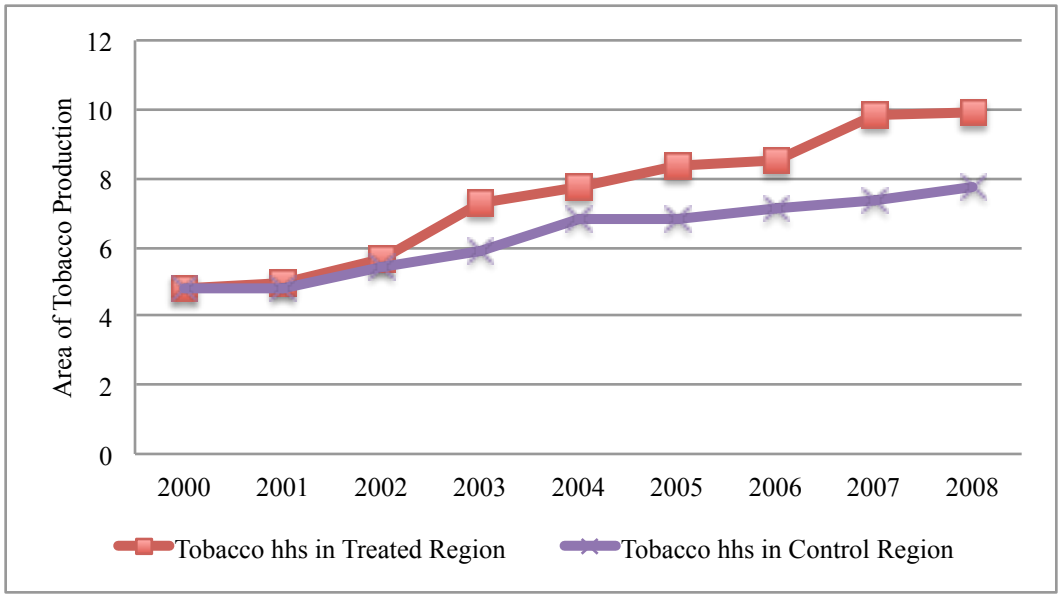


Figure 2. Evolution of Loan Size for Tobacco Households, by Treatment

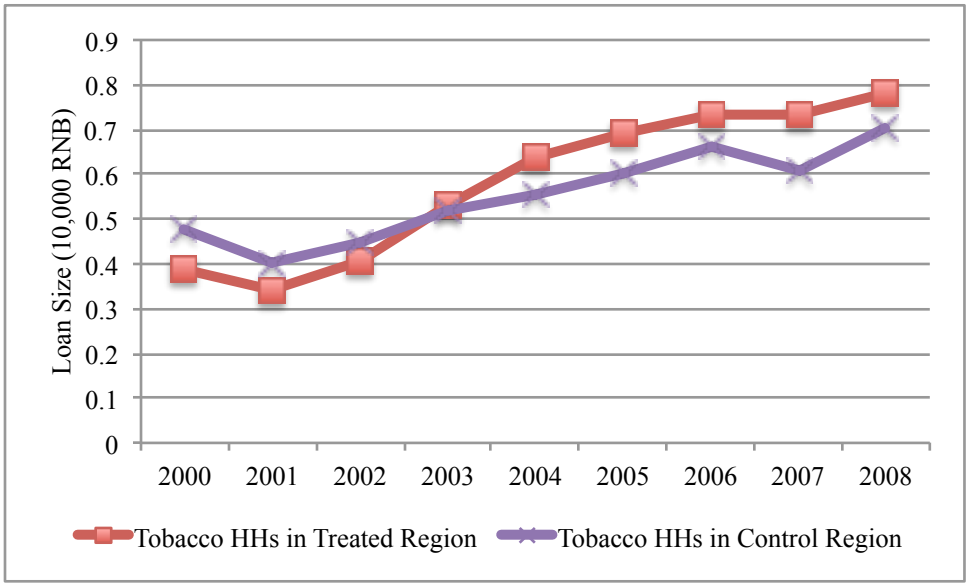


Figure 3. Evolution of Loan Size for Other Households, by Treatment

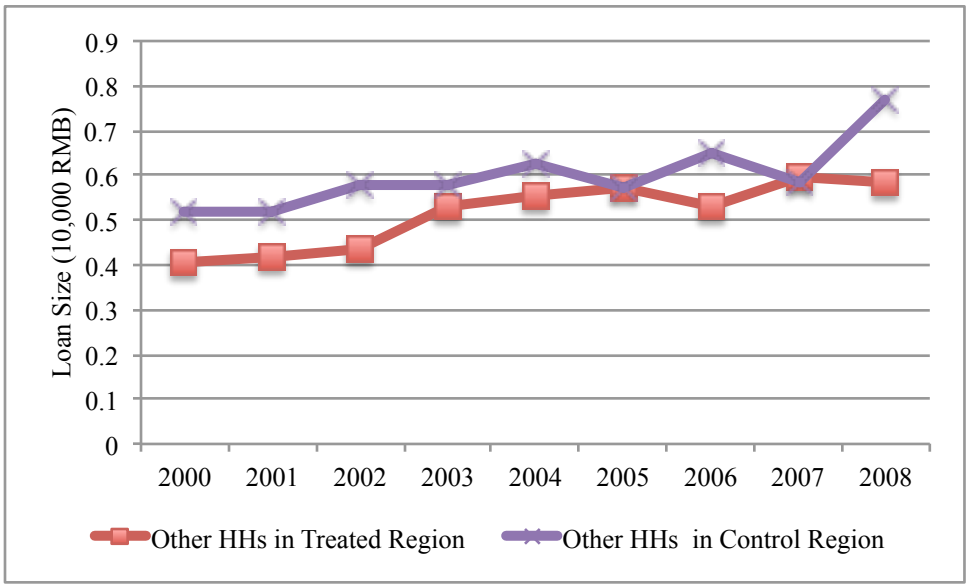


Figure 4. Evolution of Saving for Tobacco Households, by Treatment

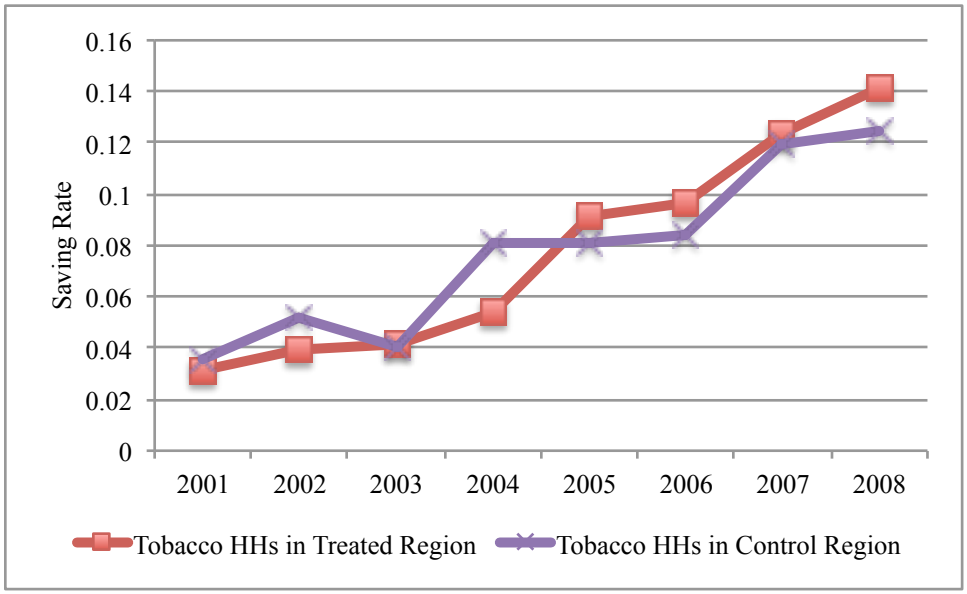


Figure 5. Evolution of Saving for Other Households, by Treatment

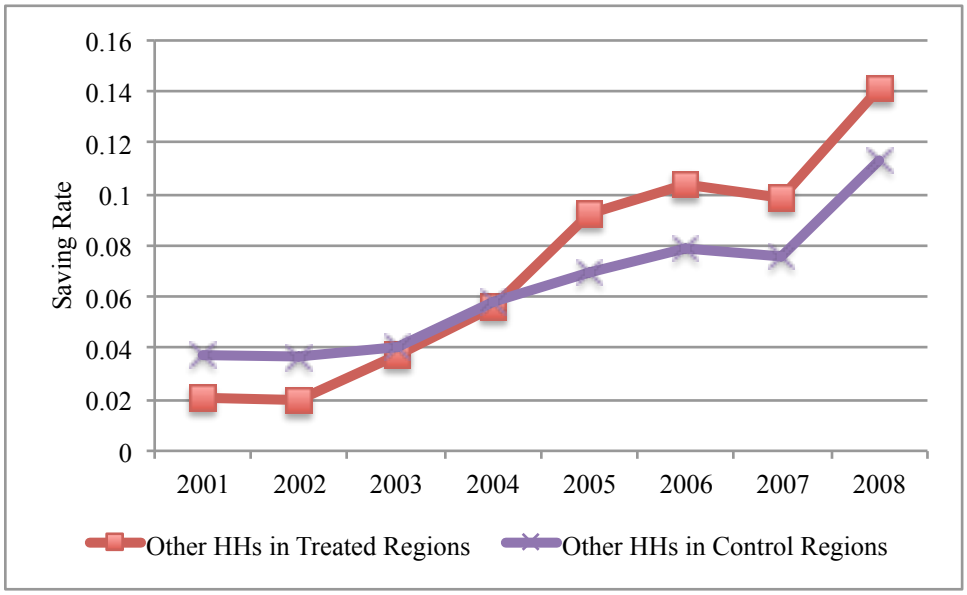


Table 1. Insurance Provision in Tobacco Production Counties of Jiangxi Province

County	Insurance Provision	Start Year	Premium	Subsidy of Premium	Maximum Payout
Guangchang	Yes	2003	12	50%	420
Yihuang	Yes	2003	12	50%	420
Lean	Yes	2003	12	50%	420
Zixi	Yes	2003	12	50%	420
Shicheng	No				
Ningdu	No				
Ganxian	No				
Huichang	No				
Xinfeng	No				
Xingguo	No				
Quannan	No				
Ruijin	No				

Notes: The unit of premium and payouts is RMB per mu (1 mu = 1/15 hectare). The exchange rate between US dollars and RMB is around 6.3.

Table 2. Summary Statistics

	Tobacco Households			Other Households	All Sample
	Treated	Control	Difference		
	(1)	(2)	(3)	(4)	(5)
Number of Households	*1429*	*2151*		*2968*	*6548*
Gender of Household Head (1 = Male, 0 = Female)	0.996	0.982	0.014***	0. 978	0.983
	(0.062)	(0.134)	(0.000)	(0.146)	(0.131)
Age	40.418	40.731	-0.313*	40.205	40.429
	(8.959)	(8.124)	(0.091)	(8.645)	(8.526)
Household Size	4.781	4.728	0.053*	4.930	4.832
	(1.022)	(1.355)	(0.054)	(1.312)	(1.284)
Education (0 = illiteracy, 1 = primary,	1.626	1.759	-0.133***	1.813	1.760
2 = secondary, 3 = high school, 4 = college)	(0.54)	(0.929)	(0.000)	(0.644)	(0.746)
Area of Tobacco Production (mu)	5.249	4.999	0.249***	0.307	2.857
	(2.119)	(2.874)	(0.000)	(1.194)	(3.175)
Production Diversification Index (0-1)	0.389	0.275	0.114***	0.119	0.237
	(0.229)	(0.261)	(0.000)	(0.203)	-0.256
Annual Household Income (10,000 RMB)	1.065	1.202	-0.137***	0.727	0.956
	(0.477)	(1.402)	(0.000)	(0.941)	(1.094)
Loan Size (10,000 RMB)	0.390	0.456	-0.066***	0.498	0.483
	(0.203)	(0.189)	(0.003)	(0.089)	(0.13)
Saving Rate (Net Saving Divided by Income)	0.036	0.045	-0.009***	0.034	0.038
	(0.079)	(0.12)	(0.003)	(0.093)	(0.101)

Notes: This table reports the mean of key variables in pre-treatment periods (2000-2002). For columns (1), (2), (4) and (5), standard deviations are in brackets. For column (3), P-value for F test of equal means of two groups are in brackets.

Table 3. Test Common Trend in Key Outcome Variables Before Policy Intervention

VARIABLES	Area of Tobacco Production (mu)	Loan Size (10,000 RMB)	Saving Rate (Net Saving Divided by Income)
	(1)	(2)	(3)
Year	0.322	-0.0543**	0.0163**
	(0.305)	(0.0233)	(0.00776)
Insurance	0.123	-0.110***	0.00429
(= 0 if control region, = 1 if treatment region)	(0.305)	(0.0161)	(0.0383)
Year*Insurance	0.160	0.0708***	-0.00860
	(0.308)	(0.0233)	(0.00776)
Observations	9,201	659	5,761
R-squared	0.080	0.034	0.006

Notes: Bootstrap clustered standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 4. Area of Tobacco Production, Loan Size, and Saving Rate by Region, Sector, and Year

	Tobacco Households			Other Households		
	2000-2002	2003-2008	Difference	2000-2002	2003-2008	Difference
	(1)	(2)	(3)	(4)	(5)	(6)
I. Area of Tobacco Production (mu)						
Treatment	5.249	8.464	3.215***			
	(0.038)	(0.043)	(0.000)			
Control	4.999	7.054	2.054***			
	(0.037)	(0.027)	(0.000)			
DD			1.161***			
			(0.000)			
II. Loan Size (10,000 RMB)						
Treatment	0.390	0.724	0.004***	0.412	0.568	0.156***
	(0.022)	(0.016)	(0.000)	(0.006)	(0.011)	(0.000)
Control	0.456	0.645	0.189***	0.523	0.630	0.108***
	(0.008)	(0.013)	(0.000)	(0.002)	(0.032)	(0.000)
DD			0.145***			0.048
			(0.000)			(0.163)
DDD			0.097**			
			(0.036)			
III. Saving Rate (Net Saving Divided by Income)						
Treatment	0.036	0.086	0.049***	0.020	0.098	0.078***
	(0.002)	(0.002)	(0.000)	(0.003)	(0.003)	(0.000)
Control	0.045	0.093	0.049***	0.037	0.064	0.028***
	(0.002)	(0.001)	(0.000)	(0.002)	(0.002)	(0.000)
DD			0.001			0.05***
			(0.804)			(0.000)
DDD			-0.05***			
			(0.000)			

Notes: For columns (1), (2), (4) and (5), standard deviations are in brackets. For columns (3) and (6), P-value are in brackets.

Table 5. Effect of Insurance Provision on Production

VARIABLES	Area of Tobacco Production (mu)		
	(1)	(2)	(3)
After	2.054***	7.938***	7.279***
(= 0 if 2000-2002, = 1 if 2003-2008)	(0.320)	(2.261)	(1.938)
Insurance	0.249	0.175	0.338***
(= 0 if control region, = 1 if treatment region)	(0.370)	(0.195)	(0.112)
After * Insurance	1.161***	1.450***	1.223***
	(0.320)	(0.167)	(0.116)
Household Size			0.0727***
			(0.00790)
Annual Household Income (10,000 RMB)			0.787***
			(0.257)
Education (0 = illiteracy, 1 = primary,			0.343*
2 = secondary, 3 = high school, 4 = college)			(0.178)
Age			-0.0252**
			(0.0112)
No. of Observation	31,207	31,207	31,207
Year Fixed Effects	No	Yes	Yes
R-squared	0.105	0.131	0.226

Notes: Bootstrap clustered standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 6. Effect of Insurance Provision on Production Diversification

VARIABLES	Production Diversification (0-1)		
	(1)	(2)	(3)
After	0.0543	0.0364	0.0268
(= 0 if 2000-2002, = 1 if 2003-2008)	(0.0536)	(0.063)	(0.0550)
Insurance	-0.0492	-0.0525	-0.0470
(= 0 if control region, = 1 if treatment region)	(0.0753)	(0.0524)	(0.0759)
Tobacco Household	0.144**	0.144***	0.145***
(= 0 if No, = 1 if Yes)	(0.0585)	(0.0214)	(0.0491)
After * Insurance	0.0755	0.0670	0.0657
	(0.0536)	(0.0607)	(0.0612)
After * Tobacco Household	-0.0320	-0.0433	-0.0482
	(0.0320)	(0.0386)	(0.0336)
Tobacco Household * Insurance	0.164***	0.168***	0.171***
	(0.0585)	(0.0237)	(0.0496)
After * Insurance * Tobacco Household	-0.129***	-0.116***	-0.113***
	(0.0320)	(0.0440)	(0.0366)
Household Size			0.00382
			(0.00256)
Education (0 = illiteracy, 1 = primary,			0.0510***
2 = secondary, 3 = high school, 4 = college)			(0.0195)
Age			0.00312***
			(0.000371)
No. of Observation	47951	47951	47951
Year Fixed Effects	No	Yes	Yes
R-squared	0.106	0.112	0.141

Notes: Bootstrap clustered standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 7. Effect of Insurance Provision on Borrowing

VARIABLES	Loan Size (10,000 RMB)		
	(1)	(2)	(3)
After	0.108	1.889***	1.679***
(= 0 if 2000-2002, = 1 if 2003-2008)	(0.0769)	(0.659)	(0.328)
Insurance	-0.111***	-0.111***	-0.133***
(= 0 if control region, = 1 if treatment region)	(0.0109)	(0.0119)	(0.0185)
After * Insurance	0.0481	0.0505	0.0923**
	(0.0769)	(0.0615)	(0.0368)
Tobacco Household	-0.0665**	-0.0669***	-0.147***
(= 0 if No, = 1 if Yes)	(0.0272)	(0.0240)	(0.0172)
After * Tobacco Household	0.0810	0.0636	0.146***
	(0.0583)	(0.0649)	(0.0389)
Tobacco Household * Insurance	0.0441	0.0183	0.0267
	(0.0272)	(0.0309)	(0.0336)
After * Insurance * Tobacco Household	0.0972*	0.134*	0.115**
	(0.0583)	(0.0757)	(0.0556)
Household Size			0.00125
			(0.00656)
Annual Household Income (10,000 RMB)			0.0911***
			(0.0137)
Area of Tobacco Production (mu)			0.00208
			(0.00292)
Education (0 = illiteracy, 1 = primary,			0.0642***
2 = secondary, 3 = high school, 4 = college)			(0.0163)
Age			-0.00103
			(0.000739)
No. of Observation	8,382	8,382	8,382
Year Fixed Effects	No	Yes	Yes
R-squared	0.017	0.029	0.081

Notes: Bootstrap clustered standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 8. Effect of Insurance Provision on Saving

VARIABLES	Saving Rate (Net Saving Divided by Income)		Net Saving (10,000 RMB)		Ratio of Net Checking to Net Total Saving	
	(1)	(2)	(3)	(4)	(5)	(6)
After	0.0275***	0.0906***	0.0561***	0.190	-0.0212	-0.173***
(= 0 if 2000-2002, = 1 if 2003-2008)	(0.00576)	(0.0311)	(0.00372)	(0.129)	(0.0791)	(0.0545)
Insurance	-0.0166	-0.0149***	-0.00219	0.0101	-0.472***	-0.487***
(= 0 if control region, = 1 if treatment region)	(0.0218)	(0.00391)	(0.00797)	(0.00865)	(0.166)	(0.114)
After * Insurance	0.0504***	0.0294***	0.0834***	0.00267	0.272***	0.269***
	(0.00576)	(0.00509)	(0.00372)	(0.0142)	(0.0791)	(0.0857)
Tobacco Household	0.00781	0.0132***	0.0360	0.0347	-0.352	-0.349
(= 0 if No, = 1 if Yes)	(0.0244)	(0.00263)	(0.0456)	(0.0340)	(0.294)	(0.243)
After * Tobacco Household	0.0211*	-0.00141	0.0562	-0.000966	-0.0420	-0.0534
	(0.0109)	(0.00332)	(0.0583)	(0.0304)	(0.290)	(0.296)
Tobacco Household * Insurance	0.00801	0.00343	-0.00975	-0.0194	0.121	0.120
	(0.0244)	(0.00471)	(0.0456)	(0.0422)	(0.294)	(0.244)
After * Insurance * Tobacco Household	-0.0495***	-0.0124**	-0.0708	0.0414	0.169	0.190
	(0.0109)	(0.00617)	(0.0583)	(0.0359)	(0.290)	(0.299)
Household Size		0.00285***		0.00599*		0.00570
		(0.000544)		(0.00338)		(0.00464)
Annual Household Income (10,000 RMB)		-0.0117***		0.00731		-0.0512
		(0.000448)		(0.0115)		(0.0329)
Education (0 = illiteracy, 1 = primary,		0.0235***		0.0322***		-0.0208
2 = secondary, 3 = high school, 4 = college)		(0.000951)		(0.00796)		(0.0186)
Age		0.000540***		0.000542		-0.00173***
		(8.34e-05)		(0.000429)		(0.000620)
No. of Observation	40,561	40,559	40,561	40,559	20,975	20,975
Year Fixed Effects	No	Yes	No	Yes	No	Yes
R-squared	0.027	0.077	0.012	0.043	0.12	0.13

Notes: Bootstrap clustered standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 9. Dynamic Effects of Insurance Provision on Borrowing and Saving

VARIABLES	Loan Size (10,000 RMB)	Saving Rate (Net Saving Divided by Income)
	(1)	(2)
Insurance	-0.122***	-0.00960**
(= 0 if control region, = 1 if treatment region)	(0.0136)	(0.00428)
Tobacco Household	-0.0224	0.0127***
(= 0 if No, = 1 if Yes)	(0.0261)	(0.00167)
Tobacco Household * Insurance	0.00750	-0.00168
	(0.0309)	(0.00475)
2001 * Insurance * Tobacco Household	-0.0591	
	(0.0364)	
2002 * Insurance * Tobacco Household	-0.0176	0.00822
	(0.0395)	(0.00860)
2003 * Insurance * Tobacco Household	0.0185	-0.00740
	(0.107)	(0.00919)
2004 * Insurance * Tobacco Household	0.0970	-0.0127
	(0.107)	(0.00903)
2005 * Insurance * Tobacco Household	0.138***	-0.0115
	(0.0395)	(0.0108)
2006 * Insurance * Tobacco Household	0.217***	-0.0187**
	(0.0287)	(0.00955)
2007 * Insurance * Tobacco Household	0.150***	-0.0147**
	(0.0561)	(0.00713)
2008 * Insurance * Tobacco Household	0.210***	-0.0111
	(0.0568)	(0.00902)
No. of Observation	8,382	40,561
Year Fixed Effects	Yes	Yes
Year Dummies * Insurance	Yes	Yes
Year Dummies * Tobacco Household	Yes	Yes
R-squared	0.021	0.057

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 10. Heterogeneity of the Insurance Effect: Production Size and Migration Income

	Production Size			Migration Income		
VARIABLES	Area of Tobacco Production (mu)	Loan Size (10,000 RMB)	Saving Rate (Net Saving Divided by Income)	Area of Tobacco Production (mu)	Loan Size (10,000 RMB)	Saving Rate (Net Saving Divided by Income)
	(1)	(2)	(3)	(4)	(5)	(6)
After	8.180***	1.875***	0.0588***	7.616***	2.018***	0.0655***
(= 0 if 2000-2002, = 1 if 2003-2008)	(3.141)	(0.114)	(0.0137)	(1.807)	(0.0676)	(0.0251)
Insurance	0.764***	-0.0924***	0.0358***	0.381***	-0.104***	-0.00684
(= 0 if control region, = 1 if treatment region)	(0.283)	(0.00290)	(0.00182)	(0.132)	(2.47e-05)	(0.0508)
Tobacco Household		-0.0834***	-0.0172**		-0.102***	-0.00981
(= 0 if No, = 1 if Yes)		(0.00157)	(0.00718)		(0.00645)	(0.0305)
After * Insurance	1.225***	0.0631***	0.0361***	1.493***	-0.0323	0.0318*
	(0.175)	(0.0220)	(0.00197)	(0.0816)	(0.0490)	(0.0187)
After * Tobacco Household		0.113***	0.0222***		-0.00452	0.0247
		(0.00939)	(0.000263)		(0.0288)	(0.0209)
Tobacco Household * Insurance		0.0180***	-0.0233***		0.0437***	0.0263
		(0.00125)	(0.00785)		(0.0101)	(0.0310)
After * Insurance * Tobacco Household		0.0294***	0.00355***		0.258***	-0.0396*
		(0.0108)	(0.000700)		(0.0527)	(0.0206)
Pre-treatment Total Production Area	2.178***	-0.0102***	-0.0288***			
(= 0 if < Median, = 1 if > Median)	(0.637)	(0.00211)	(0.00213)			
Pre-treatment Total Production Area * After	-0.163*	0.203***	0.00760***			
	(0.0873)	(0.0649)	(0.00164)			
Pre-treatment Total Production Area * Insurance	-1.374**	-0.0126***	-0.0327***			
	(0.676)	(0.00234)	(0.00218)			
Pre-treatment Total Production Area * Tobacco Household		0.0295	0.0606***			
		(0.0289)	(0.0195)			
Pre-treatment Total Production Area	0.0177	-0.178***	-0.0187***			
*After * Insurance	(0.112)	(0.0658)	(0.00198)			
Pre-treatment Total Production Area		-0.224***	-0.0309**			
* After * Tobacco Household		(0.0286)	(0.0148)			
Pre-treatment Total Production Area		-0.0107	-0.00221			
* Tobacco Household * Insurance		(0.0289)	(0.0199)			
Pre-treatment Total Production Area		0.291***	-0.00808			
* After * Insurance * Tobacco Household		(0.0290)	(0.0153)			
Pre-treatment Share of Migration Income in Total Income				-0.291**	-0.00283	0.0180
				(0.117)	(0.00211)	(0.0316)
Pre-treatment Share of Migration Income in Total Income				0.239*	-0.146***	-0.00792
* After				(0.124)	(0.00939)	(0.0141)
Pre-treatment Share of Migration Income in Total Income				-1.092***	-0.0279***	-0.0254
* Insurance				(0.0873)	(0.00407)	(0.0315)
Pre-treatment Share of Migration Income in Total Income				-0.280***	0.218***	-0.00658
* Tobacco Households				(0.0948)	(0.0158)	(0.0141)
Pre-treatment Share of Migration Income in Total Income					0.0132	0.0281
*After * Insurance					(0.0608)	(0.0351)
Pre-treatment Share of Migration Income in Total Income					0.194*	-0.0347
* After * Tobacco Household					(0.115)	(0.0236)
Pre-treatment Share of Migration Income in Total Income					-0.0458	-0.0393
* Tobacco Household * Insurance					(0.0768)	(0.0351)
Pre-treatment Share of Migration Income in Total Income					-0.278**	0.0319
* After * Insurance * Tobacco Household					(0.129)	(0.0237)
Observations	34,207	8,382	40,561	34,207	8,382	40,561
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Household Characteristics	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.208	0.030	0.074	0.157	0.036	0.070

Notes: Bootstrap clustered standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 11. Percentage of Households Changing Sector by Region and Year

	Treatment Regions			Control Regions		
Year	Tobacco to Non-Tobacco	No Change	Non-Tobacco to Tobacco	Tobacco to Non-Tobacco	No Change	Non-Tobacco to Tobacco
2000	N/A	N/A	N/A	N/A	N/A	N/A
2001	0	100	0	0.08	99.92	0
2002	0	98.72	1.28	0	94.97	5.03
2003	0.32	99.68	0	0.53	99.42	0.05
2004	0.63	99.37	0	0.32	99.17	0.51
2005	0	99.81	0.19	0	99.63	0.37
2006	0.23	99.53	0.23	0	99.9	0.1
2007	0.4	99.6	0	0.54	99.29	0.17
2008	0.42	99.43	0.14	0.11	99.33	0.56

Appendices

A Two-period model when insurance is not provided

Combine equation (3.1) and (3.4) we can get:

$$U'(C_1) = \beta p U'(C_g) F'(I) = \beta p U'(C_g)(1+R_B) \tag{3.6}$$
$$\Rightarrow \frac{C_g}{C_1} = \frac{F(I)-(1+R_B)B+(1+R_f)S}{C_1} = \beta p(1+R_B)$$
$$\Rightarrow C_1 = \frac{F(I)-(1+R_B)B+(1+R_f)S}{\beta p(1+R_B)}$$
$$= \frac{\left(\frac{1+R_B}{\alpha}\right)^{\frac{\alpha}{\alpha-1}} - (1+R_B)B+(1+R_f)S}{\beta p(1+R_B)} \tag{3.7}$$

Rewrite equation (3.3) as:

$$\beta p U'(C_g) F'(I) = \beta p U'(C_g)(1+R_f) + \beta(1-p)U'\left[(1+R_f)S\right](1+R_f) \tag{3.3'}$$

Then combine (3.3)' with equation (3.7) we have:

$$\frac{1}{C_1} = \frac{\beta p(1+R_f)}{F(I)-(1+R_B)B+(1+R_f)S} + \frac{\beta(1-p)}{S} = \frac{\beta p(1+R_B)}{\left(\frac{1+R_B}{\alpha}\right)^{\frac{\alpha}{\alpha-1}} - (1+R_B)B+(1+R_f)S}$$
$$\Rightarrow \frac{\beta p(R_B-R_f)}{F(I)-(1+R_B)B+(1+R_f)S} = \frac{\beta(1-p)}{S}$$
$$\Rightarrow \beta p(R_B-R_f)S = \beta(1-p)[F(I)-(1+R_B)B+(1+R_f)S]$$
$$\Rightarrow (1+R_B)B = F(I) - \frac{p}{1-p}(R_B-R_f)S + (1+R_f)S$$
$$\Rightarrow B = \alpha^{-\frac{\alpha}{\alpha-1}}(1+R_B)^{\frac{1}{\alpha-1}} - S\left[\frac{p}{1-p}\frac{R_B-R_f}{R_B+1} - \frac{1+R_f}{1+R_B}\right] \tag{3.8}$$

Plug equation (3.8) into (3.7)

$$\Rightarrow C_1 = \frac{1}{1-p}\frac{R_B-R_f}{\beta(1+R_B)}S \tag{3.9}$$

We know that the total investment is:

$$I = W_0 - C_1 + B - S$$

Replace C_1 and B by (3.9) and (3.8), respectively, we have:

$$I = W_0 - \frac{1}{1-p}\frac{R_B-R_f}{\beta(1+R_B)}S - S + \alpha^{-\frac{\alpha}{\alpha-1}}(1+R_B)^{\frac{1}{\alpha-1}} - S\left[\frac{p}{1-p}\frac{R_B-R_f}{R_B+1} - \frac{1+R_f}{1+R_B}\right]$$
$$\Rightarrow (1-\alpha^{-1})I = (1-\alpha^{-1})\left(\frac{1+R_B}{\alpha}\right)^{\frac{1}{\alpha-1}}$$
$$= W_0 - \frac{1+\beta}{\beta(1-p)}\frac{R_B-R_f}{R_B+1}S$$
$$\Rightarrow S^* = \frac{(1+R_B)(1-p)\beta}{(1+\beta)(R_B-R_f)}\left[W_0 + (\alpha^{-1}-1)\left(\frac{1+R_B}{\alpha}\right)^{\frac{1}{\alpha-1}}\right]$$
$$= A * \left[W_0 + (\alpha^{-1}-1)\left(\frac{1+R_B}{\alpha}\right)^{\frac{1}{\alpha-1}}\right] \tag{3.10}$$

Now let's consider consumption. Plug the expression of S into equation (3.9):

$$\begin{aligned} C_1 &= \tfrac{1}{1-p}\tfrac{R_B-R_f}{\beta(1+R_B)}\tfrac{(1+R_B)(1-p)\beta}{(1+\beta)(R_B-R_f)}\left[W_0+(\alpha^{-1}-1)\left(\tfrac{1+R_B}{\alpha}\right)^{\frac{1}{\alpha-1}}\right] \\ &= \tfrac{1}{1+\beta}\left[W_0+(\alpha^{-1}-1)\left(\tfrac{1+R_B}{\alpha}\right)^{\frac{1}{\alpha-1}}\right] \qquad (3.11) \end{aligned}$$

The last variable that we are interested in is the borrowing. According to equation (3.8):

$$B = \alpha^{-\frac{\alpha}{\alpha-1}}(1+R_B)^{\frac{1}{\alpha-1}} - S\left[\tfrac{p}{1-p}\tfrac{R_B-R_f}{R_B+1} - \tfrac{1+R_f}{1+R_B}\right] = D + S*E$$
$$\text{where } D = \alpha^{-\frac{\alpha}{\alpha-1}}(1+R_B)^{\frac{1}{\alpha-1}} \text{ and } E = \tfrac{1+R_f}{1+R_B} - \tfrac{p}{1-p}\tfrac{R_B-R_f}{R_B+1}$$

B Two-period model when insurance is provided

From equation (3.13), we can see that the expression of optimal investment is:

$$F'(I) = (1+R_B)(1+\delta) \Rightarrow I^* = \left(\tfrac{(1+R_B)(1+\delta)}{\alpha}\right)^{\frac{1}{\alpha-1}}$$

Rewrite equations (3.12) and (3.14) as:

$$\begin{aligned} \tfrac{1}{C_1} &= \tfrac{\beta p(1+R_B)}{C_g} + \tfrac{\beta(1-p)\gamma}{C_b(1+\delta)} \qquad (3.15) \\ \tfrac{\beta p(R_B-R_f)}{C_g} &+ \tfrac{\beta(1-p)\gamma}{C_b(1+\delta)} = \tfrac{\beta(1-p)(1+R_f)}{C_b} \\ \Rightarrow C_g &= AC_b, A = \tfrac{(R_B-R_f)p}{(1-p)\left[(1+R_f)(1+\delta)-\gamma\right]} \qquad (3.16) \end{aligned}$$

Plug expression (3.16) into (3.15):

$$\begin{aligned} \tfrac{1}{C_1} &= \tfrac{\beta p(1+R_B)}{AC_b(1+\delta)} + \tfrac{\beta(1-p)\gamma}{C_b(1+\delta)} \\ \Rightarrow C_b = \tfrac{\beta p(1+R_B)+\beta(1-p)\gamma A}{A(1+\delta)}C_1 &= \tfrac{\gamma}{1+\delta}(W_0 - C_1 - S + B) - \tfrac{\gamma}{1+\delta}B + (1+R_f)S \\ \Rightarrow \tfrac{\beta p(1+R_B)+\beta(1-p)\gamma A}{A(1+\delta)}C_1 &= \tfrac{\gamma}{1+\delta}W_0 - \tfrac{\gamma}{1+\delta}C_1 - \tfrac{\gamma}{1+\delta}S + (1+R_f)S \qquad (3.17) \\ \Rightarrow S = \tfrac{1}{1+R_f-\gamma/(1+\delta)}&\left[\tfrac{\gamma}{1+\delta} + \tfrac{\beta p(1+R_B)+\beta(1-p)\gamma A}{A(1+\delta)}\right]C_1 - \tfrac{\gamma W_0}{(1+R_f)(1+\delta)-\gamma} \qquad (3.18) \end{aligned}$$

Combining (3.16) and (3.17) we can get:

$$\begin{aligned} C_g &= \left[\beta p(1+R_B) + \beta(1-p)\gamma A\right]C_1 \\ \Rightarrow \left[\beta p(1+R_B) + \beta(1-p)\gamma A\right]C_1 &= f(R_B) - (1+R_B)B + (1+R_f)S \\ \Rightarrow B = (1+R_B)^{\frac{1}{\alpha-1}}(1+\delta)^{\frac{\alpha}{\alpha-1}}\alpha^{-\frac{\alpha}{\alpha-1}} &- \tfrac{D}{1+R_B}C_1 + \tfrac{1+R_f}{1+R_B}S \qquad (3.19) \end{aligned}$$

Becasue the total investment is $I = \frac{B+[W_0-C_1-S]}{1+\delta}$, according to equation (3.18) and (3.19) we have:

$$\left(\frac{(1+R_B)(1+\delta)}{\alpha}\right)^{\frac{1}{\alpha-1}} =$$

$$\left[\frac{(R_B-R_f)\gamma+(1+R_B)\left[(1+\delta)(1+R_f)-\gamma\right]}{(1+R_B)(1+\delta)\left[(1+\delta)(1+R_f)-\gamma\right]}W_0 + (1+R_B)^{\frac{1}{\alpha-1}}(1+\delta)^{\frac{1}{\alpha-1}}\alpha^{-\frac{\alpha}{\alpha-1}}\right] - [D+E]\,C_1$$

$$\Rightarrow C_1^* = \frac{1}{D+E}\left[\frac{(R_B-R_f)\gamma+(1+R_B)\left[(1+\delta)(1+R_f)-\gamma\right]}{(1+R_B)(1+\delta)\left[(1+\delta)(1+R_f)-\gamma\right]}W_0 + (\alpha^{-1}-1)(\tfrac{1+R_B}{\alpha})^{\frac{1}{\alpha-1}}(1+\delta)^{\frac{1}{\alpha-1}}\right] \quad (3.20)$$

$$\text{Where } D = \frac{(1+\beta p)(1+R_B)+\beta(1-p)A}{(1+R_B)(1+\delta)}$$

$$E = \frac{R_B-R_f}{(1+R_B)\left[(1+\delta)(1+R_f)-\gamma\right]}\frac{A\gamma+\beta p(1+R_B)+\beta(1-p)A\gamma}{A(1+\delta)}$$

$$\Rightarrow S^* = \frac{(1+R_B)(1+\delta)}{R_B-R_f}\frac{E}{D+E}\frac{(R_B-R_f)\gamma+(1+R_B)\left[(1+\delta)(1+R_f)-\gamma\right]}{(1+R_B)(1+\delta)\left[(1+\delta)(1+R_f)-\gamma\right]}W_0$$

$$+\frac{(1+R_B)(1+\delta)}{R_B-R_f}\frac{E}{D+E}(\alpha^{-1}-1)(\tfrac{1+R_B}{\alpha})^{\frac{1}{\alpha-1}}(1+\delta)^{\frac{1}{\alpha-1}} - \frac{\gamma W_0}{(1+R_f)(1+\delta)-\gamma} \quad (21)$$

$$B^* = (1+R_B)^{\frac{1}{\alpha-1}}(1+\delta)^{\frac{\alpha}{\alpha-1}}\alpha^{-\frac{\alpha}{\alpha-1}} - \frac{D}{1+R_B}C_1^* + \frac{1+R_f}{1+R_B}S^* \quad (3.22)$$