KYOTO UNIVERSITY

NRE Discussion Papers No. 2021-02
November 2021 (Revised in August 2022)
Division of Natural Resource Economics, Graduate School of Agriculture, Kyoto University

Basis Risk and Low Demand for Weather Index Insurance

Wataru Kodama, Ken Miura*, and Takeshi Sakurai

August 15, 2022

Faculty/Graduate School of AGRICULTURE KYOTO UNIVERSITY

Division of Natural Resource Economics
Kitashirakawa Oiwake-cho, Sakyo-ku, Kyoto 606-8502, Japan

Basis risk and low demand for weather index insurance*

Wataru Kodama† Ken Miura‡ Takeshi Sakurai§

First version: November 2021 This version: August 2022

Abstract

Basis risk——an imperfect correlation between an aggregate index and idiosyncratic damage——is a major impediment to index insurance take-up in developing countries. However, empirical evidence is scarce because of the difficulty in its direct measurement. This study uses household panel data from rural Zambia to estimate the impact of spatial basis risk on demand for a rainfall index insurance contract. First, we develop a simple insurance demand model to motivate our econometric specifications. Then, we quantify the spatial basis risk for each household using past rainfall data at the plot level. Exploiting changes in insurance design across years, we use within-household variations in basis risk to identify its impact. The empirical results show that spatial basis risk suppresses insurance demand. Despite its statistical significance, our results also suggest that minimizing the spatial basis risk would not yield sufficient economic benefits to offset the associated costs. Overall, this study offers a general analysis framework for spatial basis risk, product basis risk, and index insurance demand.

Keywords: climate risk, basis risk, index insurance, insurance demand, Zambia.
JEL Classification: D81, O12, O16, Q14.

*We thank Robert Finger, Douglas Gollin, Munenobu Ikegami, Hisaki Kono, Yuma Noritomo, Kazushi Takahashi, Chieko Umetsu, and participants at Kyoto University's Asian Economic Development Seminar, JADE/GRIPS Development Economics Workshop, SEEPS Annual Conference 2021, Oxford CSAE Conference 2022, and JADE Conference 2022 for their valuable comments and suggestions. The authors acknowledge financial support from the "Vulnerability and Resilience of Socio-Ecological Systems" project of the Research Institute of Humanity and Nature, JSPS KAKENHI No. 22223003 and No. 22K14957. Errors, if any, are our own.

†University of Göttingen. E-mail: wataru.kodama@uni-goettingen.de

‡Kyoto University. Corresponding author: Division of Natural Resource Economics, Graduate School of Agriculture, Kyoto University, Kitashirakawa-oiwakecho, Sakyo-ward, Kyoto city, Kyoto 606-8502, Japan; E-mail: miura.ken.3e@kyoto-u.ac.jp

§University of Tokyo. E-mail: atsakura@mail.ecc.u-tokyo.ac.jp

1 Introduction

Giné et al. (2008) and Cole et al. (2013), who conducted pioneering work in the index insurance literature, reported a low uptake rate of innovative weather index insurance products in India and discussed barriers to spurring demand.[1] However, the reported demand for index insurance in developing countries remains low despite the potentially large welfare benefits (e.g., Cai, 2016; Hill et al., 2019; Janzen and Carter, 2018). For example, J-PAL (2016) reports that the uptake rate of unsubsidized insurance products ranges from 6% to 18%. The literature attributes the major constraint to the risk of insufficient or no insurance payout when incurring loss because of an imperfect correlation between an aggregate index and idiosyncratic damage, known as basis risk (Clarke, 2016; Hazell and Hess, 2010).[2] Nevertheless, empirical evidence is scarce because of the difficulties in measuring basis risk (Janzen et al., 2020). This study uses household panel data from a drought-prone region in Zambia to directly test the impact of basis risks on farmers' actual insurance purchases.

Low uptake and cover rates of index insurance matter for both policymakers and insurance providers.[3] From a policy perspective, the extent to which the insurance product would cover the expected loss is a critical question because it affects households' expected utility in the short run and investment and consumption levels. On the supply side, selling an index insurance product in developing countries is a small-profit-and-quick-return business. Thus, scaling up insurance demand is necessary to sustain business.

Basis risk has attracted significant attention in the theoretical and empirical literature because

[1]Farm households in developing countries are subject to substantial weather risks, which leads to fluctuations in their production and consumption levels. Such risks may also prevent farmers from profitable investments (Cai, 2016; Hill et al., 2013; Janzen and Carter, 2018; Karlan et al., 2014). Therefore, an agricultural insurance policy that mitigates these risks is a promising tool for vulnerable smallholders in developing countries. Payouts in the index insurance scheme are conditional only on officially observable indices, such as rainfall. This feature minimizes information asymmetry problems and significant transaction costs incurred by indemnity-based crop insurance (Arnott and Stiglitz, 1991; Hess and Hazell, 2009). Index insurance has attracted considerable attention from researchers and policymakers because it offers protection against weather risks while maintaining lower prices. See Ali et al. (2020) for a recent review of index insurance in developing countries.

[2]Other suggested barriers include a lack of trust in the insurance provider (Cole et al., 2013; Giné and Yang, 2009), low financial literacy (Cai et al., 2020; Cai and Song, 2017; Gaurav et al., 2011), and liquidity constraints (Cai et al., 2020; Cole et al., 2013; Giné et al., 2008).

[3]Throughout the paper, cover rate is defined as the coverage from the insurance payouts divided by the expected loss because of weather disasters.

it potentially constrains insurance demands and is a unique risk specific to index insurance products. As described by Jensen et al. (2018), basis risk is the "Achilles heel of index insurance": conditional payments based on observable indexes enable index insurance products to be affordable and free from classical information asymmetry problems, but spawn another type of uninsured risk. Empirical studies typically relate low insurance demand to either spatial or product basis risks (e.g., Conradt et al., 2015; Hill et al., 2013; Janzen et al., 2020; Jensen et al., 2016). The former risk, our focus in this paper, arises from the poor correlation of the insurance index (e.g., rainfall) between individual plots and the weather station, whereas the latter is due to weak yield–insurance index correlations. Irrespective of the type, basis risk is theoretically the main factor that severely suppresses index insurance demand (Clarke, 2016).

Given theoretical salience, the extent to which basis risk suppresses insurance demand is a central question in the empirical literature. Answering it is meaningful, particularly in rural African settings, where we expect high spatial basis risk given the insufficient number of weather stations. Nevertheless, empirical studies that identify basis risk effects are limited. For example, previous studies often use the distance from a farmer's location to the reference weather station as a proxy of spatial basis risk to estimate the impact (e.g., Hill et al., 2013; Mobarak and Rosenzweig, 2012). However, such empirical proxies may correlate with unobserved factors (e.g., market access) that determine insurance demand, raising concerns about the evidence's credibility. In particular, distance measures could overestimate basis risk impacts if farmers living near the weather station are better off and the income elasticity of insurance products is high. As another example, Jensen et al. (2016, 2018) estimate household-level basis risk. Using panel data from index-based livestock insurance (IBLI) sales in Kenya, they define the "basis error," the deviation of actual loss from the payout in previous years, as an estimate of basis risk. However, the proxy's dependence on past demand for insurance contracts raises endogeneity concerns about their approach (Janzen et al., 2020). For example, when pastoral households experienced substantial livestock loss but did not receive enough payouts to buy back livestock in previous years, their reduced investments in livestock production would constrain the current demand for index insurance. In such cases with high basis errors, wealth effects through past investment decisions contaminate the true basis risk impacts on insurance demand.

This study fills this research gap by directly quantifying the spatial basis risk for each household and relating its empirical measure to actual weather index insurance purchases among small-scale Zambian farmers. During the 2011/12–2013/14 crop seasons, we introduced rainfall index insurance contracts to local farmers in Southern Province, Zambia. Farm households in the study site do not have access to irrigation facilities and thus face substantial drought risks. As a unique feature of this study, we collected plot-level daily rainfall data for a subset of survey households for five consecutive crop years between 2007/08 and 2011/12 and historical rainfall records of reference weather observation points. These rainfall records allow direct estimation of spatial basis risk for each household, defined as the probability of false negatives (i.e., incurring a loss because of drought but receiving no payout). Another essential feature of this research is exploiting within-household variations in spatial basis risk for identification, because we have changed the payout condition specified in the insurance contract over the study years. This contract design change is exogenous to farmers' behavior, providing a setting to estimate the causal impact of spatial basis risk by controlling for unobservable household characteristics that may determine both basis risk and insurance demand. As a result, our household fixed-effect specification finds a significant response of insurance demand to the estimated spatial basis risk.

This study contributes to the literature on index insurance in three essential ways. First, it speaks to the literature on the economic models of index insurance demand by serving as one of the first empirical tests of Clarke's (2016) rational demand model. By constructing a simple model, we characterize the demand function of index insurance contracts and derive theoretical predictions that can be easily translated into an empirical framework. Additionally, the model helps our analysis framework transform the theoretical definitions of basis risk into empirical measures. Overall, the insurance demand observed in the data shows a pattern consistent with theoretical predictions.

Furthermore, we explicitly differentiate between product and spatial basis risk and discuss the theoretical roles of each risk and their interplay in insurance demand. Our results suggest that spatial basis risk alone leaves an unexplained diversion of observed demand from the theoretically optimal one. We also confirm the high explanatory power of the theoretical model for average insurance demand among small-scale farmers in Zambia, especially when accounting for product basis risk. Overall, this study advances our understanding of basis risk and its impact on insurance demand by providing a general analysis framework for spatial basis risk, product basis risk, and index insurance demand.

This study provides another unique contribution by investigating how households' purchase behavior responds to changes in the design of insurance contracts, and hence, spatial basis risk (cf., Janzen et al., 2020). If households are elastic to spatial basis risk, the natural policy implication would be to reduce the spatial basis risk to boost insurance uptake and achieve higher cover rates among policyholders. Thus, confirming whether households react to such an exogenous contract change is more policy-relevant than merely illustrating the cross-sectional correlation between basis risk measures and insurance demand. Our empirical analysis, which exploits within-household variations in estimated spatial basis risk, finds supporting evidence for households' responses to contract design changes. The estimation results indicate that households respond to the changes in spatial basis risk; on average, a five percentage point (p.p.) reduction in the false-negative probability (more than halving its average) would have led to a 43–52% increase in the unit of purchase. However, the evaluation of this impact in terms of the cover rate suggests that the same five p.p. decreases lead to only a 2.25–3.75 p.p. increase in the coverage of expected crop loss with the insurance contract. This welfare gain is smaller than the high costs associated with minimizing basis risk. Our results show that although spatial basis risk adversely affects the demand for weather index insurance, reducing the risks embedded in index insurance cannot be a cost-effective policy instrument given its modest economic impact.

The remainder of this paper is organized as follows. Section 2 presents a simple model of the index insurance demand and econometric specifications. Section 3 describes the study context, survey design, and data. Section 4 decomposes the overall basis risk into spatial and product basis risks and estimates the household-level spatial basis risk and drought-induced crop loss probabilities. Section 5 provides empirical results and discusses the explanatory power of our theoretical model. Finally, Section 6 concludes the paper.

2 Conceptual framework

This section presents a simple model to describe how farmers decide to adopt index insurance contracts in the presence of production and basis risks. Our model setup builds on Clarke's (2016) rational demand model with four states of the world: the farmer either incurs crop loss or not, and the insurer either compensates with an insurance payout or not. We also present our econometric specifications motivated by the model to test the theoretical predictions regarding basis risk impacts on insurance demand.

2.1 Model

The economic agent in the proposed model is farmer i who faces a risk of crop loss due to drought or other idiosyncratic shocks (e.g., crop pests and family diseases). The farmer is strictly risk-averse and their preference is represented by a von Neumann–Morgenstern utility function U that satisfies $U' > 0$ and $U'' < 0$. We define this utility over their consumption level as equal to the sum of the wealth endowments w_i and crop output. Crop output is a random variable subject to productivity shocks. The farmer receives crop output y_i in a normal year. In contrast, the farmer incurs crop loss ($L_i = 1$) as much as l_i with a probability of $p_i = \Pr(L_i = 1) \in (0, 1)$ in a year of drought or other shocks, and thus the crop output is $y_i - l_i$. Given this, the expected utility is

$$p_i U(w_i + y_i - l_i) + (1 - p_i) U(w_i + y_i).$$

We now introduce a weather index insurance product that judges the state as "drought" ($D = 1$) based on a predetermined index (e.g., rainfall amount), and provides an insurance payout with probability $q = \Pr(D = 1) \in (0, 1)$. We assume that the payout probability is constant across farmers.[4] This insurance does not compensate farmers with probability $1 - q = \Pr(D = 0)$. The payout probability (q) is not necessarily equal to the crop loss probability (p_i) because of the rainfall pattern differences between the farmer's plot and reference weather station or other non-climate shocks. Without loss of generality, we further assume that the product compensates for a value of 1 for each

[4] This assumption is valid in our research context as we introduced the same insurance contract to all the survey

Table 1: Probability structure of weather index insurance

	$D = 0$	$D = 1$	Total
$L_i = 1$	r_i	$p_i - r_i$	p_i
$L_i = 0$	$1 - q - r_i$	$q + r_i - p_i$	$1 - p_i$
Total	$1 - q$	q	1

contract unit in the case of "drought." Hence, we define the cost of purchasing α_i insurance units as $\alpha_i mq$, where $m > 0$ is a multiple. Insurance is actuarially fair when $m = 1$, actuarially unfair when $m > 1$, and actuarially favorable when $m < 1$. If this product has no risk of contractual nonperformance (i.e., a perfect match between the index and farmer's crop loss), the farmer's consumption is $w_i + y_i - \alpha_i mq$ in a normal year and $w_i + y_i - l_i + \alpha_i(1 - mq)$ in a bad year.

To introduce the concept of basis risk, let $r_i = Pr(L_i = 1, D = 0) \in (0, 1)$ be the probability of a false negative; that is, the case when the farmer incurs a loss but receives no insurance payout. This probability represents the overall basis risk, including spatial and product basis risks.[5] Following Clarke (2016), we specify the probability structure of the weather index insurance product defined by (p_i, q, r_i) in Table 1. Because the probability of a false negative is r_i, the joint probability of insurance payment and the farmer's crop loss is discounted by r_i and is $p_i - r_i$. Similarly, the probability of no insurance payouts in a normal year is $1 - q - r_i$. Finally, the probability of a false positive is the difference between q and p_i with the addition of r_i.

This set of probabilities should hold the relationship $p_i - q < r_i < p_i(1 - q)$ to make this structure reasonable (Clarke, 2016). The first inequality is necessary for all the states to have a positive probability of occurrence. As shown in Appendix A.1.1, the second inequality is equivalent to $\Pr(D = 1 \mid L_i = 1) \geq \Pr(D = 1) = q$, indicating a positive correlation between the index and crop loss. We further assume that the farmer fully acknowledges this probability structure when making a decision, and thus, this set of probabilities (p_i, q, r_i) characterizes their insurance purchase.

Table 2 summarizes the payoffs corresponding to four different states that the farmer potentially

farmers in a given agricultural year and other insurance providers were non-existent during the study period.

[5] We decompose overall basis risk into the two types of basis risk in Section 4.1.

Table 2: Pay-off structure of weather index insurance

State	$(1,1)$	$(1,0)$	$(0,0)$	$(0,1)$
Prob.	$p_i - r_i$	r_i	$1-q-r_i$	$q+r_i-p_i$
Pay-off	$w_i + y_i - l_i$	$w_i + y_i - l_i$	$w_i + y_i$	$w_i + y_i$
	$+\alpha_i(1-mq)$	$-\alpha_i mq$	$-\alpha_i mq$	$+\alpha_i(1-mq)$

faces: output loss with insurance payout $(L_i, D) = (1,1)$, no loss without insurance payout $(0,0)$, loss without insurance payout $(1,0)$, and no loss with insurance payout $(0,1)$. The farmer chooses the optimal unit of insurance purchase α_i^* by maximizing the following expected utility:

$$\begin{aligned} EU(\alpha_i) &= (p_i - r_i)U(w_i + y_i - l_i + \alpha_i(1-mq)) + r_i U(w_i + y_i - l_i - \alpha_i mq) \\ &\quad + (1-q-r_i)U(w_i + y_i - \alpha_i mq) + (q + r_i - p_i)U(w_i + y_i + \alpha_i(1-mq)). \end{aligned} \tag{1}$$

Optimal insurance demand can be derived from the first derivative of the expected utility with respect to α_i. However, there is no closed-form solution, even under common preference forms, including constant relative and absolute risk aversion (CRRA and CARA) utility forms. To obtain approximate solutions, we employ a second-order Taylor approximation around the expected payoff $\bar{x}_i = E(x_i)$:

$$\begin{aligned} EU(\alpha_i) &\approx E[U(\bar{x}_i) + U'(\bar{x}_i)(x_i - \bar{x}_i) + \frac{1}{2}U''(\bar{x}_i)(x_i - \bar{x}_i)^2] \\ &= U(\bar{x}_i) + \frac{1}{2}U''(\bar{x}_i)V(x_i). \end{aligned}$$

where $E(x_i) = w_i + y_i - p_i l_i + (1-m)q\alpha_i$ and $V(x_i) = q(1-q)\alpha_i^2 - 2\{p_i(1-q) - r_i\}l_i\alpha_i + p_i(1-p_i)l_i^2$. This Taylor approximation implies that under the concavity assumption of the utility function ($U'' < 0$), the farmer faces a trade-off between decreases in the expected consumption and its variance when purchasing an extra unit of actuarially unfair insurance.

When the insurance product is actuarially fair (i.e., $m = 1$), the expected payoff is independent of the insurance contract units α_i.[6] Therefore, the farmer's optimal behavior is to minimize variance

[6] As we elaborate in the Data section, the actuarially fair assumption is valid in our empirical context.

by purchasing the following unit:

$$\alpha_i^* = \frac{p_i(1-q) - r_i}{q(1-q)} \times l_i. \tag{2}$$

Note that the nonnegativity constraint does not bind under the assumption of $r_i < p_i(1-q)$. Thus, uptake is conditional on whether the insurance payout is correlated with crop loss. If the index is completely independent of crop loss, the farmer's optimal choice is no purchase, which is a case of the corner solution.

The approximated solution has implications for contractural risk and its impact on insurance demand. First, our solution implies linear and adverse effects from basis risk, which justifies our linear specification in the subsequent empirical framework. Second, the magnitude of the basis risk effect is proportional to the expected loss in shock years. Rephrasing this, the optimal cover rate chosen by the farmer and basis risk are linearly related. Third, other probability components (p_i, q) also affect insurance demand. Like basis risk, the crop loss probability p_i has a linear relationship with the insurance demand, and the impact is proportional to the expected loss due to drought or other idiosyncratic productivity shocks. In contrast, the probability of insurance payout q has a theoretically indeterminate effect on demand (see Appendix A.1.2 for the comparative statics). Finally, the optimal demand is independent of the risk preference parameters and initial endowments. Under the above approximation, in which the farmer cares only up to the variance of their payoff, risk attitudes do not affect the optimal demand. However, this does not hold when the farmer considers the higher moments of their pay-off in decision-making. While the third moment plays a significant role in shaping precautionary saving motives in any dynamic economic setting (Kimball, 1990), the farmers in our static model merely maximize their income after the harvest. Given the static nature of the model, incorporating the third moment does not result a significant deviation from the approximated solution. The following subsection discusses the validity of the proposed approximate solution.

2.2 Numerical solutions

The optimal unit of the insurance contract in Equation (2) is an approximated solution. The general solution when insurance is actuarially fair ($m = 1$) should satisfy the following first-order condition based on Equation (1):

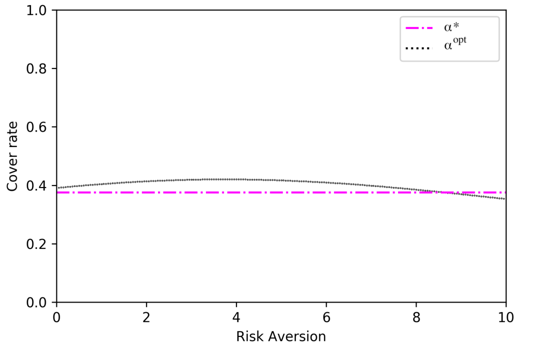
$$\begin{aligned} EU'(\alpha_i) &= (p_i - r_i)(1-q)U'(w_i + y_i - l_i + \alpha_i(1-q)) - r_i q U'(w_i + y_i - l_i - \alpha_i q) \\ &\quad - (1 - q - r_i)qU'(w_i + y_i - \alpha_i q) + (q + r_i - p_i)(1-q)U'(w_i + y_i + \alpha_i(1-q)) = 0. \end{aligned}$$

To check the validity of the approximation, we compare the Taylor-approximated solution α_i^* with a series of numerical solutions, denoted by α_i^{opt}, which satisfies the above condition. In particular, we evaluate the relationship between insurance demand and (i) the degree of risk aversion, and (ii) initial endowments under the CRRA or CARA utility function.

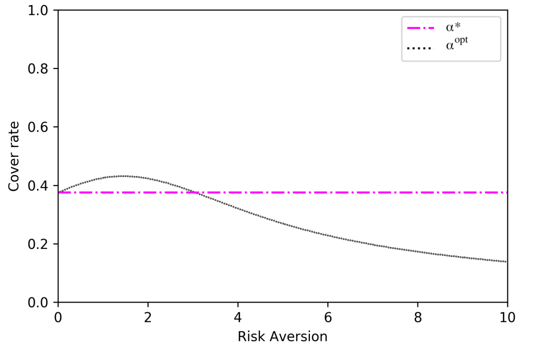
Figure 1 illustrates a pair of approximated and numerical solutions $(\alpha_i^*, \alpha_i^{\mathrm{opt}})$, based on the preference assumption. Panels 1a and 1c indicate that α_i^* provides a good approximation of the numerical solutions over most wealth and risk preference parameter ranges, under the CRRA preference assumption. In contrast, we observe a discrepancy between the approximated and numerical solutions for risk-averse farmers under the CARA preference assumption. The hump-shaped relationship between risk attitude and insurance demand in panel 1b is consistent with Clarke's (2016) general theorem. In the empirical analysis, we test whether the degree of risk aversion affects units of actual insurance uptake.

2.3 Econometric specification

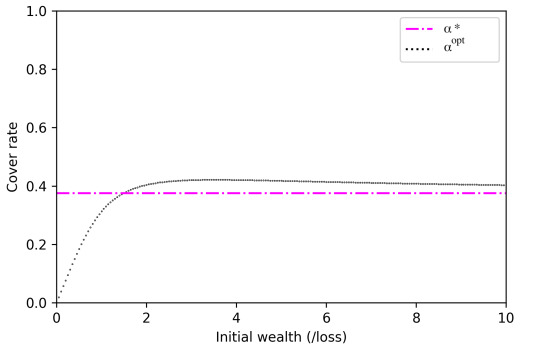
We translate the insurance demand function in Equation (2) into our econometric specification to estimate demand responses to basis risk and examine other theoretical predictions. In the data described in the following section, p_i and r_i are heterogeneous across surveyed households, whereas q is constant in a given year. Moreover, the same household faces a different basis risk r_i across survey years because of the different insurance contract designs regarding payment thresholds and reference locations. These within-household variations allow us to employ the household fixed effects model to test whether farmers respond to changes in basis risk. Specifically, we model



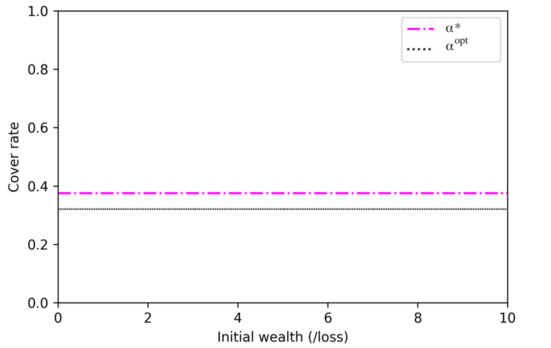
(a) Risk-aversion vs. insurance demand (CRRA)



(b) Risk-aversion vs. insurance demand (CARA)



(c) Initial wealth vs. insurance demand (CRRA)



(d) Initial wealth vs. insurance demand (CARA)

Figure 1: Approximated and numerical solutions of insurance demand under CRRA and CARA

Notes: The cover rate shows the proportion of insurance payout to loss, α_i/l_i. The specified functional forms are $U(x_i) = x_i^{1-\gamma_i}/(1-\gamma_i)$ (CRRA) and $U(x_i) = -\exp(-\gamma_i x_i)/\gamma_i$ (CARA), where γ_i is a risk preference parameter. To compute numerical solutions, we set $(p_i, q_i, r_i) = (0.2, 0.2, 0.1)$. We also assume $w_i = 3l_i$ for panels (a) and (b) and $\gamma_i = 4$ for panels (c) and (d).

insurance demand across multiple years as

$$\alpha_{it} = \beta_0 + \beta_1(r_{it} \times l_i) + \mu_i + \mu_t + \varepsilon_{it}, \tag{3}$$

where α_{it} is the number of insurance contracts that household i purchased in year t; r_{it} represents basis risk or probability of false negatives; l_i is the expected loss in drought years; μ_i and μ_t are fixed effects for household i and year t, respectively; and ε_{it} is an error term possibly correlated within household i. In this specification, the parameter of interest β_1 captures the response of the purchased unit of contracts to a one p.p. increase in basis risk. This fixed-effect specification exploiting within-household variations in basis risk enables us to estimate an unbiased estimator that represents the causal impact of basis risk on farmers' behavior of actual insurance purchases, as long as changes in r_{it} are uncorrelated with changes in ε_{it}. This condition is likely to hold in our study context in which the research team designed insurance contracts for each survey year, and thus r_{it} is beyond household i's control. In particular, when we confirm $\beta_1 < 0$, basis risk is found to decline insurance demand, as predicted by our theoretical model.

In our setting, the probability of household-level crop loss p_i is time invariant, whereas that of insurance payout q is unique among households in year t. The model predicts that these probabilities also determine insurance demand (see Equation (2)). To investigate the relationship between a set of two probabilities (p_i, r_i) and insurance uptake α_i, we run the following regression with the pooled data combining the three survey years. Building on the optimal uptake unit in Equation (2), we specify the regression model as follows:

$$\alpha_{it} = \gamma_0 + \gamma_1(r_{it} \times l_i) + \gamma_2(p_i \times l_i) + X_{it}\gamma_X + \mu_t + \epsilon_{it}, \tag{4}$$

where X_{it} is a vector of household controls, including the head's characteristics and elicited risk preference measure. ϵ_{it} denotes the error term. The empirical evidence is consistent with model predictions when we find $\gamma_1 < 0$ and $\gamma_2 > 0$.

We quantify the key probabilities for our empirical analysis, p_i and r_i, using historical rainfall data from referred weather observation points and plot-level rainfall data. Section 4 presents detailed estimation procedures and their results after describing the data in the following section.

3 Data

3.1 Context

We use household survey data from maize farmers in Southern Province, Zambia. The main cropping season in the survey area is the rainy season (November to April). During the dry season (May

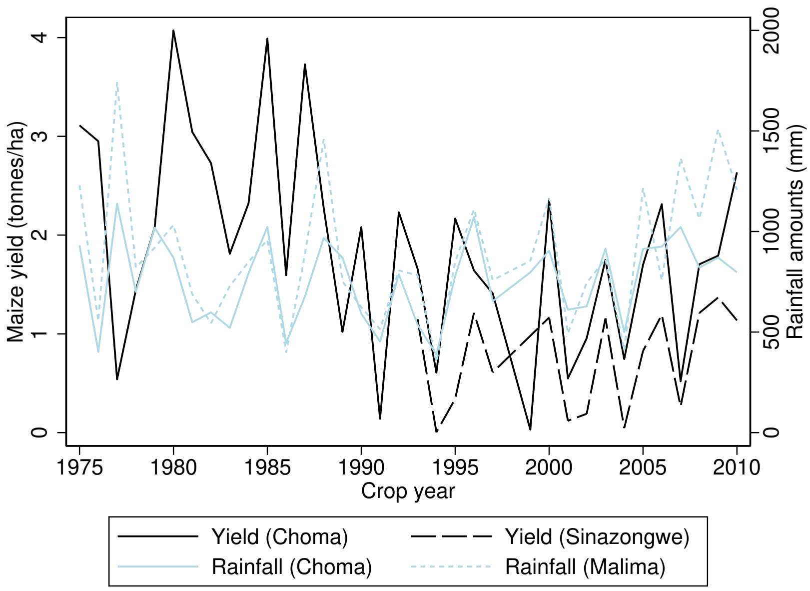


Figure 2: Rainfall and maize yield in Choma and Sinazongwe districts, 1975/76–2010/11.

Source: Rainfall data of Choma and Malima from the Choma Meteorological Station of Zambia Meteorological Department (Mochipapa) and the Malima irrigation site, respectively; Maize yield from the annual reports of Crop Forecast Survey by the Central Statistical Office.

to October), there is no rainfall and agricultural activities are limited to small-scale vegetable cultivation because most local farmers do not have access to irrigation. During the rainy season, local farmers grow crops such as maize, cotton, sweet potato, and groundnut under rainfed conditions. Given this, climate risk is the primary source of threat to farmers' livelihoods. This is especially the case in Southern Province, which is the most drought-prone province in the country.

Figure 2 illustrates the maize yields in our study sites, the Choma and Sinazongwe districts, and the rainfall amounts recorded at the Choma meteorological station and Malima irrigation site (which is in Sinazongwe district) between 1975 and 2010. Figure 2 illustrates that maize production co-moves with observed precipitation patterns. To formally check their statistical relationship, Table 3 presents the regression results of the maize yields on rainfall and other controls. As expected, the total rainfall amount during the rainy season has explanatory power for maize harvests in the region: its impact has an inverted U-shape with a peak at approximately 820 mm (column 1).

Instead of the total rainfall amount, rainfall distributions during the rainy season may affect

Table 3: Rainfall and maize yield in the study area, 1975/76 to 2010/11

	(1)	(2)	(3)	(4)	(5)
Rainfall					
Rainy season (100mm)	1.310**				
	(0.513)				
Rainy season, squared.	-0.080**				
	(0.035)				
Flowering season (100mm)		0.216**	0.747**		
		(0.092)	(0.346)		
Flowering season, squared.			-0.074*		
			(0.043)		
Planting season (100mm)		0.089	0.767		
		(0.119)	(0.501)		
Planting season, squared.			-0.106		
			(0.066)		
"Drought" in 11/12 contract				-0.573**	
				(0.279)	
"Flood" in 11/12 contract				-0.399	
				(0.425)	
"Drought" in 12/13 contract					-0.843***
					(0.218)
"Flood" in 12/13 contract					-0.973**
					(0.390)
Choma district	0.753***	0.807***	0.826***	0.811***	0.769***
	(0.207)	(0.216)	(0.192)	(0.228)	(0.204)
Linear time trend	-0.045***	-0.035**	-0.044**	-0.040***	-0.048***
	(0.014)	(0.016)	(0.018)	(0.015)	(0.012)
Constant	-3.115*	0.688	-0.926	1.977***	2.367***
	(1.738)	(0.506)	(0.877)	(0.450)	(0.395)
R-squared	0.470	0.412	0.477	0.387	0.546
Observations	51	51	51	51	51

Notes: The dependent variable is maize yield (mean = 1.52 tons/ha, std. dev. = 1.04) from the annual Crop Forecast Survey by the Central Statistical Office. Rainfall data come from the Choma Meteorological Station of the Zambia Meteorological Department. The estimation covers the period between 1975/76 and 2010/11 for Choma district and between 1993/94 and 2010/11 for Sinazongwe district, with missing observations. "Drought" ("Flood") in the 2011/12 contract is a dummy variable that equals 1 if the total rainfall amount during the rainy season was below 600 mm (above 1000 mm), and 0 otherwise. "Drought" in the 2012/13 contract is a dummy variable that equals 1 if the total rainfall amount during January and February was below 280 mm, and 0 otherwise. "Flood" in the 2012/13 contract is a variable that equals 1 if the total rainfall amount in December was above 300 mm, and 0 otherwise. After pooling the maize yield data from the Choma and Sinazongwe districts, OLS was used for the estimations. Heteroskedasticity-robust standard errors are in parentheses. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

maize growth. Specifically, local farmers are concerned about rainfall patterns during two periods: the planting season (November–December) and the flowering season (January–February). In this study, we focus more on flowering season rainfall as a determinant of farmers' maize yields for the following reasons. First, historical data support the importance of the flowering season rainfall relative to the planting season rainfall. The linear specification in column 2 of Table 3 demonstrates that rainfall amounts during the flowering season are significantly correlated with maize yield. In contrast, planting season rainfall shows a null association after conditioning on flowering season rainfall. To account for the nonlinear relationship, we add the squared terms of the rainfall variables into the regression equation in column 3. We find that the estimated impacts of rainfall in the planting season are less precise than those of rainfall in the flowering season, although the coefficients are similar in magnitude. According to the regression results, yield predictions based on flowering season rainfall would provide us with more reliable figures. Second, Waldman et al. (2017), who conducted a household survey in the same area as well as our field observations, also supports the above view. Even if drought hits in the early stage of the rainy season, local farmers can re-plant early-maturing seed varieties to offset the loss. In contrast, irregular dry spells during the flowering season significantly limit crop growth, leading to poor maize harvests.

3.2 Insurance sales and data collection

3.2.1 The sample

We now describe our field surveys conducted in Southern Province. Our first survey dates to the 2007/08 crop season. Between 2007/08 and 2011/12, we collected daily rainfall data at the plot level and household information weekly from 48 households in five villages, in collaboration with the Zambia Agriculture Research Institute (ZARI) (Miura and Sakurai, 2021). These five villages were located at three different sites: A, B, and C (Figure 3). Site A is a lower flat lake-side area with an elevation of 500–550 m. Site B is located in a middle escarpment area with an elevation of 750-1000 m. Finally, site C is in an upper terrace area on the Zambian plateau with an elevation of 1050–1100 m. Although these three sites are within a 15-km radius, their rainfall patterns are distinct because of their different attitudes. We randomly selected 16 households from each site for

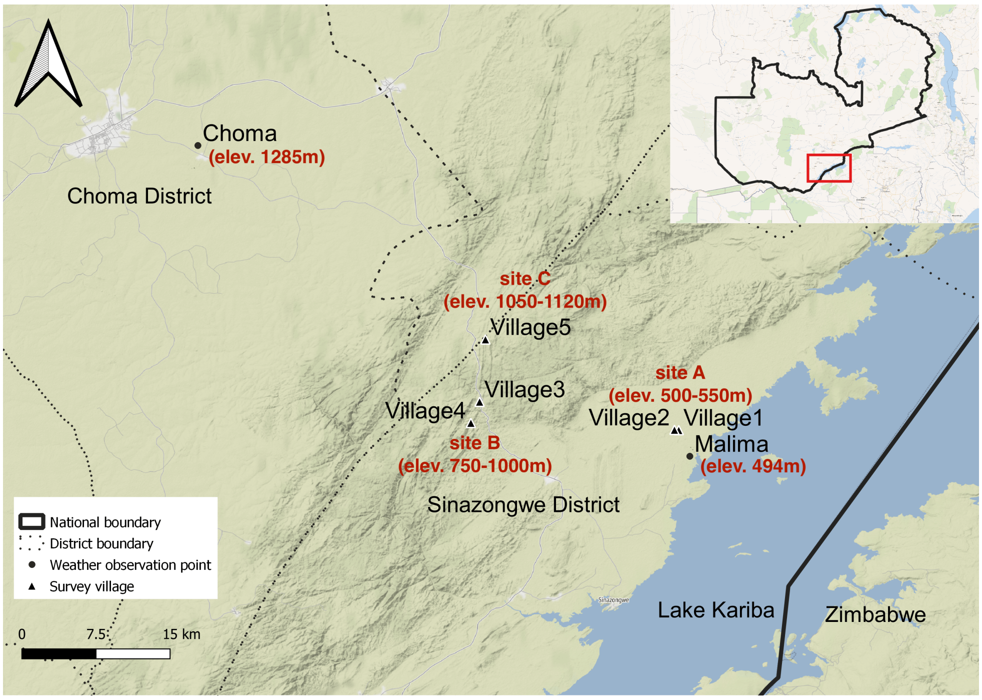


Figure 3: Map of study site and weather observation points

Notes: For Choma, the weather observation point is the Choma weather station of the Zambia Meteorological Department. For Malima, the weather observation point is the Malima irrigation site.

the survey, providing a total sample of 48 households. This sample is called "old sample." After the end of the weekly household survey, we introduced an uncommercial rainfall index insurance as a pilot project with ZARI to 100 households in November 2011 and 160 households in 2012/13 and 2013/14 years. These households include the old sample.

We focus on 48 households in the old sample for two reasons. First, we collected the plot-level rainfall data only from these households. This dataset is crucial for estimating the household-specific spatial basis risk in our empirical analysis. Second, the long-term relationships among the research team, ZARI, and local villagers might alleviate the problems associated with a lack of trust in insurance contracts, which is a potential barrier to insurance demand (Giné and Yang, 2009).[7]

[7]Indeed, we observe higher average units of insurance uptake from the old sample than those of the new sample. See Appendix Table A4 for insurance sales results.

3.2.2 Insurance products

In collaboration with the ZARI and the Zambia Meteorological Department, we introduced a new weather index insurance contract in November 2011. The insurance payout is conditional only on rainfall at a prespecified weather observation point and covers drought and flood events. The premium was set at ZMW 5 per unit.[8] As the average wage for casual agricultural labor was approximately ZMW 10 per day in the survey, the premium was fixed at an affordable level for local farmers.

To set the actuarially fair premium rate, we used the estimation results based on the district-level historical production data in Table 3 and designed weather index insurance contracts for each year. In the 2011/12 crop year, the defined condition under which a farmer would receive an insurance payout was that the total rainfall recorded at the Choma weather station was either less than 600 mm or more than 1000 mm during the rainy season between November and April. Column 4 of Table 3 confirms that the "drought" defined in the first-year contract decreased the maize yield in the corresponding crop year by 0.57 tons/ha.[9] Referring to the historical rainfall data, we set 20% ($q = 1/5$) as the premium rate; thus, the insurance payout per unit was ZMW 25 ($= 5/q$). For the insurance sales in the 2012/13 crop year, we used a more detailed index: farmers would receive the insurance payout if rainfall in the flowering season (January and February) at the Choma weather station was less than 280 mm or rainfall in December was more than 300 mm. Both specified events predict maize yield reductions by 0.84 tons/ha and by 0.97 tons/ha, respectively (column 5 of Table 3). In the final year, we referred to the rainfall recorded at the Malima irrigation site.[10] This new observation site was much closer to the study villages (Figure 3). Reflecting local farmers' repeated requests for the cover of rainfall shocks in the planting season, we used rainfall amounts in the planting season (November and December) as the index in the insurance policy: farmers would be compensated when the total rainfall recorded at the Malima irrigation site was less than 214 mm or exceeded 800 mm in November and December.[11] The historical rainfall data suggest

[8]ZMW = Zambian Kwacha. In November 2011, the exchange rate was 5.2 ZMW/US$.

[9]In contrast, the defined "flood" does not have a significant effect. Nevertheless, we covered flood events in the insurance contract to reflect demand among local farmers. We discuss this point in Section 3.2.5.

[10]In the first two years, we did not realize the availability of rainfall records collected by the local development project at the Malima irrigation site.

[11]The change in the reference index was for this practical reason. Using historical rainfall data from the Malima

$q = 1/3$ for payout conditions in the last two years; thus, the insurance payout per unit was ZMW 15.

Figure 3 depicts the locations of the survey sites, Choma weather station, and Malima irrigation site. Although all study sites were located far from Choma, sites B and C were relatively closer (approximately 35–40 km from the station; whereas site A was 55 km away). Therefore, we expect the spatial basis risk to be higher at site A under the insurance contract in the first two years. In contrast, the Malima irrigation site was close to all the survey villages, especially site A.

3.2.3 Door-to-door insurance marketing

We sold the insurance product every early November (i.e., before the start of the rainy season) during the three agricultural years, 2011/12–2013/14. Our trained enumerators visited the survey households ten days before the insurance sale. They first explained the insurance contract to the household head, left a copy of a leaflet visually explaining it, and informed the head that insurance sales would occur at a designated place—usually the village head's place—approximately ten days later. In addition, the enumerators told the survey participants that they should bring enough money if they were interested in the insurance.

In the door-to-door visits, the enumerators also conducted interviews with household heads to collect information relevant to insurance demand (e.g., subjective perceptions of weather risk, arithmetic skills, and financial literacy) and basic demographics. The interview included the Binswanger (1980)-style lottery game to elicit attitudes toward risk among household heads. The game was incentivized by actual winnings. In the game, the enumerator showed a household head six alternatives with different rewards and explained that winnings were dependent on the result of a coin tossed by the enumerator and would be paid on the insurance sales day. Based on this choice, we categorized respondents into risk aversion classes. Appendix Table A2 summarizes the list of choices and corresponding risk aversion classes.

irrigation site, Appendix Table A1 replicates Table 3. The estimation results confirm the importance of rainfall in both seasons. The design of the final year's contract prioritized farmers' perceptions over statistical evidence.

3.2.4 Insurance sales and result reports

After approximately ten days of door-to-door visits, we held a sales day for each village. We asked each household head the number of insurance units they wanted to purchase. Each participant then paid the money per their demand. At that time, they could use the rewards of the Binswanger game for payment if they had won.

After the end of the contract period every year, the research team and ZARI personnel visited the study villages to inform the survey participants of whether an insurance payout would happen. The rainfall index did not satisfy the payout conditions in all three years; thus, no payouts were made during the research period.[12]

3.2.5 Plot-level rainfall data

The final data used for the empirical analysis are the plot-level rainfall records. Before the start of the weekly survey in November 2007, we installed automatic rainfall gauges and loggers at a representative plot for the 48 households in the old sample. We then recorded the daily rainfall amounts at the plot level for each surveyed household until April 2012. We use the collected field-level rainfall data to compare with the rainfall index observed at the weather reference points when estimating drought and false-negative probabilities for each household.

Table 4 summarizes the rainfall amounts during the flowering season at each site and at the two weather observation points. Rainfall data indicate non-negligible differences in precipitation among locations within a small area, suggesting the salience of spatial basis risk. Although rare, farmers experienced heavy rainfall and floods in the 2009/10 crop year. This is why we covered floods in addition to droughts in our insurance contracts. However, our empirical analysis only considers drought risks, as flood risks are minimal in the local context, and the incorporation complicates the analysis without generating additional implications.

[12] In the first year, the total amount of rainfall during the 2011/12 rainy season was 711.7 mm at the Choma weather station. In the second year, the total rainfall amount at the same station during December 2012 was 200 mm and during the flowering season was 510.7 mm. In the last year, the total rainfall amount in November and December 2013 was 379.5 mm at the Malima irrigation site. At the end of the multi-year research project (May 2014), we returned the premiums to the surveyed farmers based on our past sales records. During the intervention, this reimbursement was kept a secret from the respondents.

Table 4: Rainfall at plots and weather reference points during the flowering season

	2007/08	2008/09	2009/10	2010/11	2011/12
Site A (16 households)	521.8	458.4	912.0	424.5	322.8
	(13.9)	(69.5)	(61.0)	(24.8)	(24.9)
Site B (16 households)	525.7	559.9	843.3	625.2	375.1
	(19.6)	(25.3)	(123.3)	(63.5)	(30.7)
Site C (16 households)	494.4	603.2	721.2	442.8	333.0
	(32.0)	(28.2)	(94.8)	(31.0)	(28.2)
Total (48 households)	514.0	540.5	825.5	497.5	343.6
	(26.6)	(75.9)	(123.5)	(100.9)	(35.7)
Choma meteorological station	394.4	313.3	369.1	334.5	252.4
Malima irrigation site	366.3	459.9	959.1	670.0	286.1

Notes: The numbers represent the average total rainfall in January and February in millimeters. Standard deviations are in parentheses.

3.3 Descriptive statistics

Table 5 presents the descriptive statistics of the insurance sales results and key variables from household interviews by survey year. Appendix Table A4 reports the summary statistics of the units purchased by site and sample. In all years, most farmers purchased more than two insurance contract units. The average insurance uptake was highest in the first year.

The expected maize outputs in the normal and drought years were only obtained during the household interview in November 2012. On average, farmers expect to lose 1.113 tons (67.9% of the normal harvest) of maize harvests in a drought year.[13] We regard this expected loss as time-invariant and use this value as l_i in Equations (3) and (4) for our estimations. The expected loss of 1.113 tons of maize is equivalent to approximately ZMW 835.[14] This implies that one unit of the insurance product covers 2.99% $(25/835)$ in 2011/12 and 1.80% $(15/835)$ in 2012/13–2013/14 of the expected loss (cover rate per unit). Thus, the average farmers covered only 8.69%, 4.70%, and 4.34% of their expected losses with the purchased rainfall index insurance products. These low

[13]The wording of the question is "In this (main) field, what amount of maize do you harvest in a normal year or drought year, respectively?" The average loss ratio was 0.674 with an SD of 0.121. The variation in expected loss is driven by heterogeneity in the field size (mean =1.249 ha and SD=0.791 ha).

[14]According to the 2012/13 household survey, mode and median maize prices for household consumption were both ZMW 0.75 per kg. While only a few farmers sold maize, the same survey showed the range of its output price for sales from ZMW 0.6 to 0.8 per kg. We presume that the economic value of 1 ton of maize is ZMW 750 and approximate the average economic loss in a drought year as ZMW 835 ($\approx 750 \times 1.113$).

Table 5: Descriptive statistics

	2011/12		2012/13		2013/14	
	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.
Units of insurance takeup (#)	2.902	(2.245)	2.614	(1.956)	2.415	(1.245)
Expected output in normal year (ton)	.	(.)	1.640	(1.025)	.	(.)
Expected output in drought year (ton)	.	(.)	0.527	(0.351)	.	(.)
Expected loss (ton)	.	(.)	1.113	(0.744)	.	(.)
RA – Extreme (dummy)	0.390	(0.494)	0.068	(0.255)	0.098	(0.300)
RA – Severe (dummy)	0.122	(0.331)	0.250	(0.438)	0.220	(0.419)
RA – Intermediate (dummy)	0.098	(0.300)	0.205	(0.408)	0.146	(0.358)
RA – Moderate (dummy)	0.073	(0.264)	0.227	(0.424)	0.317	(0.471)
RA – Slight (dummy)	0.146	(0.358)	0.114	(0.321)	0.122	(0.331)
RA – Neutral (dummy)	0.171	(0.381)	0.136	(0.347)	0.098	(0.300)
Cash reward (ZMW)	7.585	(6.156)	11.682	(7.706)	12.878	(7.253)
Observations	41		44		41	

Notes: ZMW 1 = about US$ 0.22. RA denotes the risk aversion class.

cover rates are comparable with the rates reported in the index insurance literature (e.g., Cai and Song, 2017; Dercon et al., 2014; Jensen et al., 2018).

Table 5 also presents the risk game's results. We categorize the household heads into one of the risk attitude groups based on the choice in the Binswanger-style lottery. Despite the innate nature of preferences, we find that the distribution of selected choices was not stable across the surveys.[15] Our empirical analysis treats risk attitude measures as time-variant and controls them in the regression. Finally, as more respondents chose risky options, the average winning amounts were higher in the last two years than in the first year. Our empirical analysis exploits the rewards as positive liquidity shocks to test the role of cash constraints on insurance demand, as proposed in the previous literature (e.g., Cole et al., 2013; Giné et al., 2008).[16]

[15]There are two possible explanations for changes in the distributions of risk attitude measures. The first is learning. The majority chose the safest option only in the first year. Farmers could have learned how the lottery game works and taken risks in the last two years. Second, farmers may have changed their risk attitudes over time according to their experiences (Cai and Song, 2017).

[16]However, we cannot rule out the potential "reciprocity" effect: farmers who received the reward might be more likely to join the contract to reciprocate to the experimenter.

4 Quantifying basis risk

Estimating the equations (3) and (4) requires quantifying the crop loss and false–negative probabilities (p_i, r_i). We begin by decomposing the overall basis risk into spatial and product basis risks. We also express the crop loss probability as a function of the product basis risk. We then estimate the spatial basis risk and drought-induced crop loss probabilities under a few distribution assumptions.

4.1 Decomposing basis risk

Our rational demand model assumes that farmers incur crop loss ($L_i = 1$) due to either drought or other idiosyncratic production shocks (e.g., crop pests and family diseases). Therefore, the theoretical probability of a false negative or basis risk r_i consists of both spatial and product basis risks. In contrast, our rainfall index insurance contract introduced to Zambian farmers does not cover crop loss due to non-rainfall shocks, which source product basis risk. By retrieving the spatial basis risk from r_i, we elaborate on the empirical framework before measuring its empirical counterpart.

As in the model, let (L_i, D_i, D) be indicators of crop loss in household i, drought at household i's plot, and drought at the reference weather station, respectively. Then, we decompose basis risk probability into spatial and product basis factors r_i^s and r_i^p, as follows:

$$\begin{aligned}
r_i &= \Pr(L_i = 1, D = 0) = \Pr(L_i = 1, D = 0, D_i = 1) + \Pr(L_i = 1, D = 0, D_i = 0) \\
&= \Pr(L_i = 1 \mid D = 0, D_i = 1) \Pr(D = 0, D_i = 1) + \Pr(L_i = 1 \mid D = 0, D_i = 0) \Pr(D = 0, D_i = 0) \\
&= \underbrace{\Pr(L_i = 1 \mid D_i = 1)}_{=1 \text{ by assumption}} \underbrace{\Pr(D = 0, D_i = 1)}_{r_i^s\text{: spatial basis risk}} + \underbrace{\Pr(L_i = 1 \mid D_i = 0)}_{r_i^p\text{: product basis risk}} \underbrace{\Pr(D = 0, D_i = 0)}_{=1-q-r_i^s} \\
&= (1 - r_i^p) r_i^s + (1 - q) r_i^p,
\end{aligned}$$

where $Pr(L_i = 1 \mid D, D_i) = Pr(L_i = 1 \mid D_i)$ in the fourth equality follows the natural assumption that weather station-level drought D does not provide additional predictive power for plot-level crop loss L_i after conditioning for plot-level drought D_i. Similarly, we decompose the crop loss probability p_i as

$$\begin{aligned} p_i &= \Pr(L_i = 1) = \Pr(L_i = 1, D_i = 1) + \Pr(L_i = 1, D_i = 0) \\ &= \underbrace{\Pr(L_i = 1 \mid D_i = 1)}_{=1 \text{ by assumption}} \underbrace{\Pr(D_i = 1)}_{p_i^d\text{: drought probability}} + \underbrace{\Pr(L_i = 1 \mid D_i = 0)}_{r_i^p\text{: product basis risk}} \underbrace{\Pr(D_i = 0)}_{=1-p_i^d} \\ &= p_i^d(1 - r_i^p) + r_i^p. \end{aligned}$$

This decomposition indicate that farmers incur crop losses because of either drought shocks or other idiosyncratic sources. Thus, the crop loss probability is composed of the probability of plot-level drought p_i^d and the probability of other production shocks, where we can express the latter as the product of the probability of no plot-level drought $(1 - p_i^d)$ and the product basis risk r_i^p.

The substitution of r_i and p_i into equation (2) alters the theoretical insurance demand to

$$\alpha_i^* = \frac{p_i(1-q) - r_i}{q(1-q)} \times l_i = \frac{p_i^d(1-q) - r_i^s}{q(1-q)} \times (1 - r_i^p) \times l_i. \tag{5}$$

The revised demand function in equation (5) implies that product basis risk attenuates the impact of spatial basis risk and that farmers become less sensitive to spatial basis risk as product basis risk increases. This theoretical prediction sheds light on spatial- and product- basis risk interactions.

Equation (5) provides a theoretical basis for the empirical analysis. As discussed in the following subsection, our plot-level and weather station-level rainfall data allow us to estimate the spatial basis risk r_i^s and household-level drought risk p_i^d under distribution assumptions. However, we cannot quantify each household's product basis risk r_i^p because of the lack of long panels on agricultural production.[17] To identify the parameter in Equation (3), we assume that the product basis risk is constant across households; that is, $r_i^p = r^p$. With this assumption, we refine our benchmark empirical specification as[18]

$$\alpha_{it} = \beta_0 + \beta_1(r_{it}^s \times l_i) + \mu_i + \mu_t + \varepsilon_{it}. \tag{6}$$

[17] Nevertheless, Section 5.3 uses household production data from weekly interviews during the three agricultural years 2008/09-2010/11 to approximate product basis risk for the whole sample and each site. With the rough estimate of product basis risk, Section 5.3 argues how incorporating product basis risk into the analysis framework increases the explanatory power of the theoretical model for the empirical pattern of insurance demand.

[18] While our limited sample size prohibits the application, incorporating heterogeneity in product basis risk into the

Similarly, we update the pooled OLS specification as follows:

$$\alpha_{it} = \gamma_0 + \gamma_1(r_{it}^s \times l_i) + \gamma_2(p_i^d \times l_i) + X_{it}\gamma_X + \mu_t + \epsilon_{it}.$$

4.2 Estimating spatial basis risk

Our estimation of the spatial basis risk r_i^s and probability of drought-related crop loss p_i^d builds upon key assumptions on maize production and rainfall probability distributions. Let R_i, R_c, and R_m be the observed rainfall at household i's plot, Choma weather station, and Malima irrigation site, respectively. The following assumptions are introduced:

- Assumption 1: $R_i^{\text{fl}} \geq 280 \Leftrightarrow D_i = 1 \Rightarrow L_i = 1$
  Farmers incur crop loss because of plot-level "drought" if rainfall during the flowering season (January–February) is 280 mm or less.
- Assumption 2: R_i^{fl} follows a truncated normal distribution.
- Assumption 3: (R_i^{fl}, R_c) and (R_i^{fl}, R_m) follow a joint truncated normal distribution.

Assumption 1 specifies the flowering season rainfall as a key determinant of the maize harvests. The statistically significant results in Table 3 justify Assumption 1 in the research context (see Section 3.1 for details). The historical rainfall data at the two weather reference points motivate Assumption 2 regarding the probability distribution of rainfall. Appendix Figure A2 presents the historical distribution of the flowering season rainfall at the Choma weather station during 1949/50–2012/13, which provides credence to Assumption 2. The statistical tests also validate the normality assumption (see Appendix Table A3). We also confirm a similar rainfall pattern in rainfall data from the Malima irrigation site. As precipitation is non-negative, we additionally assume that the distribution is truncated at zero. Given these, we extend the assumption of truncated normality to the joint distribution of field-level and indexed rainfall (Assumption 3).

Under Assumptions 1 and 2, we can derive the probability of crop loss due to plot-level drought

empirical framework by estimating the random coefficient model is a promising avenue for future research.

for each household as

$$p_i^d = \int_0^{280} f_{R_i^{\text{fl}}}(s)ds,$$

where $f_{R_i^{\text{fl}}}$ is the probability density function for the truncated normal distribution of plot-level rainfall during the flowering season (January–February). The estimated p_i^d is constant across the survey years, 2011/12–2013/14. In contrast, the spatial basis risk differs by survey year because of changes in insurance payout conditions:

$$\begin{aligned}
2011: r_i^s &= \int_0^{280}\int_{600}^{\infty} \zeta_{R_i^{\text{fl}},R_c^{\text{ra}}}(s,t)dsdt,\\
2012: r_i^s &= \int_0^{280}\int_{280}^{\infty} \zeta_{R_i^{\text{fl}},R_c^{\text{fl}}}(s,t)dsdt,\\
2013: r_i^s &= \int_0^{280}\int_{214}^{\infty} \zeta_{R_i^{\text{fl}},R_m^{\text{pl}}}(s,t)dsdt.
\end{aligned}$$

For the 2011/12 (2012/13) year, $\zeta_{R_i^{\text{fl}},R_c^{\text{ra}}}$ ($\zeta_{R_i^{\text{fl}},R_c^{\text{fl}}}$) is a probability density function of the joint truncated normal distribution of plot-level rainfall in the flowering season and the total rainfall during the rainy season (flowering) at the Choma weather station. For the 2013/14 year, $\zeta_{R_i^{\text{fl}},R_m^{\text{ra}}}$ is the probability density function that jointly draws plot-level flowering season rainfall and planting season (November–December) rainfall at the Malima irrigation site. We calibrate the moments and correlations for the truncated normal and binormal probability density functions based on rainfall data from household plots and reference sites. Appendix A.2 describes the procedures in detail.

Table 6 the presents descriptive statistics of a set of estimated probabilities and Figure 4 depicts their distributions. The estimated spatial basis risk is higher in 2011/12 than that in 2012/13. This difference reflects a higher risk of contractual nonperformance because of the simpler contract design in 2011/12. Although we expected the spatial basis risk to be smaller because of the closer reference location in the final year's contract, the estimates were relatively high for 2013/14. This unexpected result may be driven by weak rainfall correlations between planting and flowering seasons. While site A is farthest from Choma and closest to Malima, farmers at site A faced a higher spatial basis risk in all years than their counterparts at other sites. This counterintuitive result would reflect higher exposure to drought risks at site A. Overall, the estimated spatial basis risk is household-specific and changes across the years. This within-household variation in spatial

Table 6: Descriptive statistics of estimated probabilities

	2011/12		2012/13		2013/14	
	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.
p_i^d, probability of drought	0.130	(0.043)	0.130	(0.043)	0.130	(0.043)
Site A (16 households)	0.173	(0.016)	0.173	(0.016)	0.173	(0.016)
Site B (16 households)	0.095	(0.042)	0.095	(0.042)	0.095	(0.042)
Site C (16 households)	0.123	(0.022)	0.123	(0.022)	0.123	(0.022)
r_i^s, spatial basis risk	0.078	(0.040)	0.060	(0.036)	0.094	(0.029)
Site A (16 households)	0.124	(0.014)	0.103	(0.013)	0.121	(0.010)
Site B (16 households)	0.050	(0.030)	0.036	(0.024)	0.072	(0.031)
Site C (16 households)	0.060	(0.019)	0.041	(0.016)	0.091	(0.018)
Observations	48		48		48	

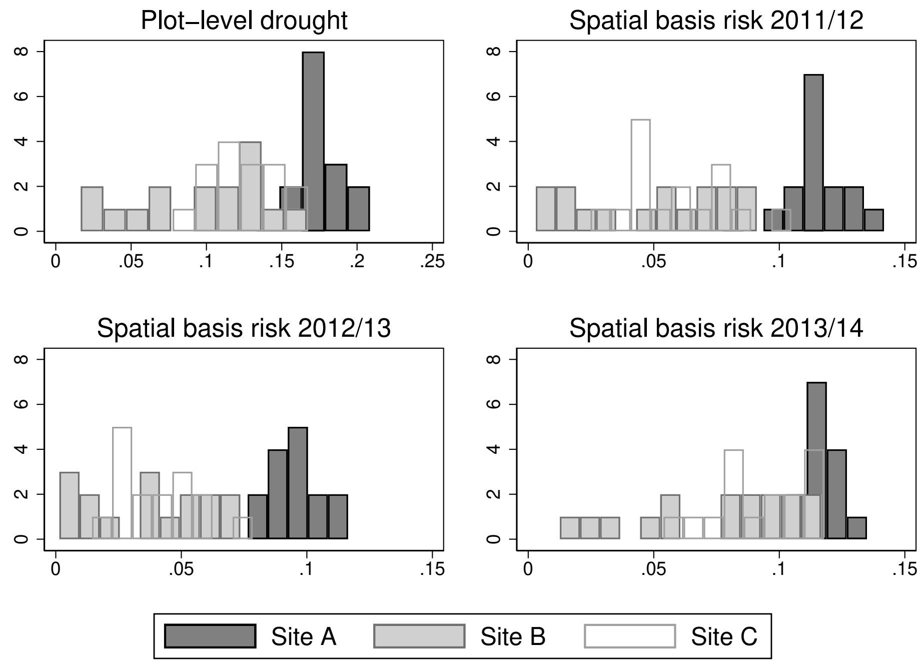


Figure 4: Distribution of plot-level drought and spatial basis risk

basis risk motivates our fixed-effect specification in Equation (6).

5 Results

This section presents our preferred fixed-effect model estimation results, followed by the model predictions test. We also discuss the economic impact of spatial basis risk based on the empirical results. Finally, we offer possible explanations for the observed low spatial basis risk effects and discuss the explanatory power of the theoretical model by comparing optimal and observed cover rates. We also consider the role of product basis risk in household insurance demand.

5.1 Baseline results

Table 7 presents the regression results for Equation (6). The dependent variable is the purchased insurance contract unit. Our estimation results show that farmers adjust their purchased units in response to a change in spatial basis risk. As the model predicts, we detect the significant effect of within-household variation in basis risk only when interacting with expected loss (columns (1) and (2)). Columns (3) and (4) confirm the robustness of this empirical pattern to the inclusion of dummies for the risk aversion classes. While one concern is the small sample size, statistical significance does not change with bootstrap standard errors across specifications (Appendix Table A5). Overall, our data suggest a causal effect of the spatial basis risk on household demand for rainfall index insurance.

In addition, we do not find that the cash reward, the byproduct of the Binswanger-style lottery, is a significant determinant of the purchase. The null effect of immediate cash would be natural given that we set the premium to ZMW 5, an affordable amount even for worse-off farmers. Finally, the change in elicited risk attitudes has no significant effects (not shown).

According to the result in column (3), the average household with 1.113 tons of expected maize loss due to drought shocks increases insurance uptake by 1.25 ($= 22.524 \times 0.05 \times 1.113$) units in response to a five p.p. decrease in spatial basis risk, equivalent to halving the spatial basis risk in our setting. Because the farmers purchased 2.41–2.90 units on average, this increment equals a 43-52% increase in unit purchase. Thus, the impact of spatial basis risk on insurance demand is not negligible.

However, when this impact is evaluated with regard to the cover rate, we find a modest welfare

Table 7: Effect of spatial basis risk on insurance demand: Fixed effect estimates

	(1)	(2)	(3)	(4)	(5)
$r^s \times l$	-20.718*	-21.248*	-22.524*	-23.435*	
	(10.344)	(11.165)	(11.823)	(12.621)	
r^s		3.989		7.193	-20.619
		(19.798)		(17.108)	(17.487)
Cash reward			0.023	0.024	
			(0.029)	(0.029)	
Year = 2012	-1.102*	-1.050*	-1.402**	-1.304*	-0.699
	(0.565)	(0.612)	(0.656)	(0.666)	(0.532)
Year = 2013	-0.095	-0.157	-0.254	-0.366	-0.210
	(0.392)	(0.416)	(0.452)	(0.412)	(0.425)
Constant	4.669***	4.392***	4.743***	4.202***	4.609***
	(0.937)	(1.516)	(1.152)	(1.429)	(1.482)
Risk aversion class	No	No	Yes	Yes	No
R squared	0.678	0.678	0.736	0.736	0.632
Observations	126	126	126	126	126

Notes: The dependent variable is the number of insurance contracts purchased by a household. Robust standard errors clustered by household are in parentheses. $*p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

gain from reducing spatial basis risk. Each year, the average farmer covered 8.69% (2011/12), 4.70% (2012/13), and 4.34% (2013/14) of their expected losses with our index insurance product. For example, a five p.p. decrease in basis risk could have increased their cover rate by 3.75 p.p. or 43.19% (2011/12), 2.25 p.p. or 47.96% (2012/13), and 2.25 p.p or 51.91% (2013/14). However, even after this increase, the expected cover rate remains low at 12.44% (2011/12), 6.95% (2012/13), and 6.59% (2013/14). As a policy intervention to decrease spatial basis risk by five p.p. would be costly in practice, our results imply that minimizing basis risk would not yield sufficient economic benefits to offset the associated costs.

5.2 Testing the model predictions

To test the theoretical predictions from Equation (5), we run an OLS regression with the pooled data. Table 8 shows the estimation results. Across all specifications, the coefficients on the interaction terms between a pair of probabilities (p_i^d, r_i^s) and the expected loss are statistically significant

Table 8: Test of the model predictions: Pooled OLS estimates

	(1)	(2)	(3)	(4)	(5)	(6)
$r^s \times l$	-33.180***	-35.255***	-33.502***	-36.368***	-33.287***	-36.343***
	(11.118)	(11.807)	(10.307)	(13.471)	(10.860)	(13.473)
$p^d \times l$	26.314***	37.397***	26.457***	34.691**	26.075***	34.361**
	(8.069)	(13.062)	(6.937)	(14.599)	(7.237)	(14.003)
r^s		18.628		17.192		18.340
		(22.268)		(21.844)		(21.025)
p^d		-31.610		-24.700		-26.497
		(24.044)		(24.075)		(22.438)
l		-1.439		-0.943		-0.901
		(1.704)		(1.606)		(1.538)
RA – Extreme/Severe			-1.348***	-1.314**	-1.354***	-1.292***
			(0.499)	(0.497)	(0.485)	(0.471)
RA – Intermediate			0.172	0.185	0.136	0.215
			(0.477)	(0.453)	(0.425)	(0.395)
RA – Slight			-0.074	-0.098	-0.095	-0.099
			(0.651)	(0.652)	(0.674)	(0.677)
RA – Neutral			-0.173	-0.152	-0.148	-0.085
			(0.476)	(0.472)	(0.450)	(0.433)
Cash reward			0.030	0.027	0.029	0.023
			(0.027)	(0.026)	(0.028)	(0.028)
Year = 2012	-1.001**	-0.713	-1.418***	-1.153**	-1.426***	-1.123**
	(0.418)	(0.535)	(0.478)	(0.562)	(0.476)	(0.540)
Year = 2013	0.227	-0.080	-0.208	-0.448	-0.220	-0.447
	(0.319)	(0.406)	(0.382)	(0.426)	(0.391)	(0.433)
Constant	1.947***	4.865**	2.441***	4.482**	2.814***	4.776**
	(0.370)	(2.102)	(0.535)	(1.955)	(0.611)	(2.034)
HH characteristics	No	No	No	No	Yes	Yes
R squared	0.220	0.247	0.375	0.388	0.381	0.396
Observations	126	126	126	126	126	126

Notes: The dependent variable is the number of insurance contracts purchased by a household. RA denotes risk aversion class and RA–Moderate is used as a reference. Robust standard errors clustered by household are in parentheses. Household characteristics include the household head's sex, age, and years of schooling. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

in the expected direction. These significant results hold even after controlling for risk preferences and other household characteristics.

One may be concerned about the assumption of drought in the flowering season as a determinant of poor harvest (Assumption 1 in Section 4.2). For example, as the relationship between maize yield and rainfall at the Choma weather station provides empirical support (Table 3), this assumption may not hold for agricultural production in places distant from Choma (e.g., villages at site A). To address this concern, Appendix Table A6 presents the same result except for using the sample restricted to sites B and C. These additional regressions produce an empirical pattern similar to that in Table 8, although the result is less precise for some specifications because of the small sample size. Overall, the results in Tables 8 and A6 validate our proposed theoretical model and further justify the benchmark fixed-effect specification.

It would be worthwhile to discuss the coefficients on other empirical variables. As predicted in the approximated solution of insurance demand, risk preferences do not predict the purchased units of the contract, except for the extreme/severe risk aversion class. This result is in line with the predictions from our numerical exercise, which demonstrates a drop in insurance demand specifically for the most risk-averse farmers because of basis risk (Panels 1a and 1b). Finally, Tables 7 and 8 indicate that insurance uptake is significantly lower in the 2012/13 agricultural year than that in the 2011/12 agricultural year. The first year had no insurance payout, although some farmers complained that they had experienced drought. Farmers may have updated the expected insurance product benefits in a negative direction (Jensen et al., 2018). Such behavioral responses may explain the decrease in insurance demand in the second year.

5.3 Discussions

Low spatial basis risk effect

Our empirical results find significant adverse impacts of spatial basis risk on insurance demand. However, farmers' responses were insufficient for our insurance contracts to to protect against weather risks in their livelihoods. We offer three possible explanations for the farmers' limited reactions.

First, the insurance product learning process might affect farmers' reactions to changes in spatial basis risk. A few previous studies have reported the effectiveness of experimental games in helping individuals with limited education understand the general concept of insurance (e.g., Cai and Song, 2017; Jensen et al., 2018; Janzen et al., 2020). Among them, Janzen et al. (2020) randomly provided farmers with opportunities to play experimental games and, after the game, elicited their willingness to pay for index insurance products with different basis risks. They find that participation in experimental games does not alter farmers' preference sensitivity to basis risk, indicating the ineffectiveness of short-run experiences in boosting their understanding of basis risks in index insurance. However, another possible interpretation is farmers' rational insensitivity to basis risk. Even in the first year, two-thirds of the survey participants correctly understood the risk of no payout, even when they experienced crop failure.[19] Considering the good understanding level among farmers, learning through experiences is unlikely to spur insurance demand in the future.

Another explanation for the reported low spatial basis risk effect is the attenuation bias. Our estimated spatial basis risks may contain measurement errors because their measurements rely on plot-level rainfall data from only five consecutive crop years. The imprecise estimation of the spatial basis risk would be a concern because the rainfall data collection period of five years is too short to observe some extreme rainfall events. Even if the noise is uncorrelated with the true spatial basis risk, such classical measurement errors will bias the estimated coefficient toward zero. The fixed-effect specification may ease concerns by absorbing a large part of measurement errors if the noise is relatively constant across time. This condition is likely in our context because plot-level rainfall data are used only for the distribution of the flowering season rainfall. Thus, the errors embedded in the estimated spatial basis risk can be stable across the three years.[20]

The last but most convincing interpretation is that there is no incorporation of interplay with product basis risk. Finally, the revised demand function in Equation (5) suggests that the spatial basis risk effect is restrictive when product basis risk is salient. To see this, consider a high product basis risk case with $r_i^p = 0.4$. Halving this risk to $r_i^p = 0.2$, the marginal effect of spatial basis risk increases by 33% $(= (1 - 0.2)/(1 - 0.4) - 1)$.[21] Moreover, although we assume that the product

[19] In November 2011, we asked the following True or False question: "Assume that you experience crop failure due to drought and it rains 700 mm at the Choma weather station. Will you get an insurance payout?" Of the respondents, 66% answered False correctly.

[20] When the true spatial basis risk is serially correlated and noises in estimated spatial basis risk are random, fixed effects estimation may magnify the problem of measurement errors. Considering the empirical definition of spatial basis risk and the relative importance of constant error components, we argue that the advantage of fixed effect estimation outweighs the cost.

[21] In this case, our estimates imply that a five p.p. reduction in spatial basis risk would bring a 58-69% increase in

basis risk is constant and orthogonal to the spatial basis risk, this assumption may be too restrictive. In such a case, the spatial basis risk effect may be underestimated because our specifications do not allow the spatial basis risk measure to interact with the product basis risk. In this regard, we further discuss the explanatory power of the proposed model.

Explanatory power of rational demand model

Our pooled OLS regression results found coefficient signs and significance consistent with the theory, providing evidence for the rational demand model for weather index insurance. Therefore, the natural question is: To what extent does our proposed model explain the low demand observed in our rainfall index insurance sales? To evaluate its explanatory power, we compare the optimal cover rate predicted by our rational demand model (CR^o_{it}) to the actual cover rate observed in our data (CR^a_{it}). We compute the optimal and actual cover rates as

$$\mathrm{CR}^o_{it} = \frac{p^d_i(1-q_t) - r^s_{it}}{q_t(1-q_t)} \times (1 - r^p_i) \;\text{ and }\; \mathrm{CR}^a_{it} = \frac{\alpha_{it} P_t}{l^m_i},$$

where (p^d_i, r^s_{it}) is a set of probabilities estimated in Section 4; $q_t = 1/5$ for the 2011/12 contract and $q_t = 1/3$ for the other two years' contracts; l^m_i is the expected loss in the drought year reported by farmers (in monetary term); and P_t is the insurance payout per unit where $P_t = 25$ for the 2011/12 contract and $P_t = 15$ for the last two years' contracts.

We consider the following three scenarios with different assumptions of product basis risk r^p_i. First, we assume no product basis risk as an extreme case; that is, $r^p_i = 0$. Second, assuming that all farmers face the same product basis risk, we provide a rough estimate based on plot-level rainfall data and households' maize harvest records during the three agricultural years 2008/09–2010/11.[22] Assuming a constant product basis risk among the whole sample, the estimated product basis risk is 0.20. In the third scenario, we allow the product basis risk to differ across sites A, B, and C, while

insurance demand and 3.00–5.00 p.p. higher cover rate. Nonetheless, the estimated spatial basis risk effect is much lower than the value predicted by the theory. When the product basis risk is as high as $r^p_i = 0.4$, the demand function predicts that the five p.p. reduction in spatial basis risk should lead to a 13.50–18.75 p.p. increase in cover rate.

[22]We compute the rough estimate of product basis risk as follows. First, we judge whether each household experienced "drought" (i.e., flowering season rainfall $<$ 280 mm) based on plot-level rainfall data from 2008/09 to 2010/11. Second, we judge whether each household incurred crop loss for each year by comparing their reported maize harvest from weekly survey data during 2008/09–2010/11 with the self-reported expected output in the drought year ($y_i - l_i$)

Table 9: Descriptive statistics of actual and optimal cover rates

	Description	2011/12	2012/13
CR_i^a	Actual cover rate (%)	10.222	5.573
		(8.331)	(7.050)
CR_i^o	Optimal cover rate (%), no product basis risk	16.955	12.352
		(6.895)	(5.698)
$\text{CR}_i^{o,p1}$	Optimal cover rate (%), constant product basis risk	13.854	9.942
	across all samples ($r_i^p = 0.195$)	(5.451)	(4.586)
$\text{CR}_i^{o,p2}$	Optimal cover rate (%), constant product basis risk	14.275	10.323
	across villages ($r_i^p = 0.351, 0.105, 0.146$)	(6.622)	(5.378)
$\text{CR}_i^o - \text{CR}_i^a$	Differences in cover rates, no spatial basis risk	6.991	6.779
		(10.124)	(9.875)
$\text{CR}_i^{o,p1} - \text{CR}_i^a$	Differences in cover rates, constant product basis risk	3.632	4.369
	across all samples	(9.424)	(9.116)
$\text{CR}_i^{o,p2} - \text{CR}_i^a$	Differences in cover rates, constant product basis risk	4.053	4.750
	across villages	(10.099)	(9.664)
Observations		41	44

villagers face the same product basis risk within the site. When we estimate site-specific product basis risk, the results find higher product basis risk at site A (0.35), followed by sites C (0.15) and B (0.10).

Table 9 presents the results of a comparison between CR_{it}^o and CR_{it}^a.[23] We restrict our attention to only the 2011/12 and 2012/13 insurance contracts because we find that the 2013/14 contract does not satisfy the non-negativity condition for the interior solution (i.e., $r_i^s < p_i^d(1-q)$) for most households.[24] As shown in Table 9, when we assume no product basis risk, the optimal cover rate is higher than the actual cover rate. When we account for the product basis risk suggested by crop output data, the optimal cover rate is closer to the actual cover rate. Nevertheless, the gaps between

from the 2012/13 household survey. Third, we define an indicator variable "basis error" that equals one when the household incurred loss in a year and did not experience plot-level drought and zero otherwise. We set missing values for this variable when the household experienced a plot-level drought. This indicator represents a realized state of product basis risk, where farmers incurred crop loss conditional on not having "drought." Finally, we compute the average of the indicator—the ratio of farmers who had basis error—for the whole sample and each site as an estimate for constant product basis risk. This exercise excludes production data from the 2007/08 agricultural year in which unusual flood events occurred.

[23] Appendix Figure A3 displays the distribution of the cover rates for the 2011/12 and 2012/13 insurance contracts.

[24] Given that most purchased at least one unit of the insurance in 2013/14, estimated spatial basis risk might contain serious measurement errors. While fixed effects relying on within-HH variations work even in this case, the cross-

the optimal and actual cover rates remained at approximately 4 p.p. for both scenarios.[25]

Overall, our empirical results suggest that the demand function from the rational demand model predicts the average demand for our actuarially fair rainfall index insurance well, albeit not perfectly. Moreover, the rational model's explanatory power is high, especially when accounting for product basis risk. Thus, incorporating product basis risk into the empirical framework warrants further research.

6 Conclusion

This study tested whether basis risk is a barrier to demand for index insurance in rural Zambia. Using unique rainfall data collected from the households' plots for five years, we estimated the theory-based spatial basis risk as a joint probability of plot-level drought and no insurance payout for each household. We then related this direct measure of basis risk to the actual demand for the rainfall index insurance contract introduced in the survey area. Our theoretical model of farmers' optimal insurance purchase behavior motivated the econometric specifications. Exploiting contract design changes during the survey period and within-household variations in spatial basis risk, we provided empirical evidence of the negative impact of spatial basis risk on household demand for rainfall index insurance contracts. Moreover, the cross-sectional estimation results validated our model predictions. The provision of such direct evidence on the relationship between basis risk and insurance demand is our first contribution to the literature on index insurance.

Another important lesson from this study is that, despite its statistical significance, the economic impact of spatial basis risk on insurance demand is modest. For example, our estimation results revealed that a five p.p. reduction in the probability of false negatives (more than half of the average) would increase the cover rate of expected crop loss through insurance by 2.25–3.75 p.p. (31–38%) at most. This result implies that even if we introduce an insurance product to minimize spatial basis risk, farm households will cover a small fraction of their potential loss due to weather

section comparison of cover rates would not produce meaningful results in the 2013/14 data.

[25]Constraints other than basis risk would explain the remaining gap. While the extensive discussion of other preventing forces is beyond the scope of this study, Appendix Section A.3 briefly discusses the roles of factors that this study does not explore in insurance demand.

shocks. Our results also suggest that incorporating product basis risk into the empirical framework enhances the explanatory power of the proposed demand model for actual demand among Zambian farmers.

Finally, we acknowledge that this study has some limitations that warrant future work. First, we measured only the spatial basis risk at the household level. As a result, this study does not fully uncover the impacts of household-specific product basis risk and its interaction with spatial basis risk. Second, our results were based on a small sample from a particular area in Africa. To check the external validity of our findings, further data collection that provides an opportunity to quantify household-specific basis risk is desirable. Finally, this study does not identify cost-effective interventions or insurance designs that boost demand for weather index insurance products. An attempt to address these outstanding issues is a promising avenue for future research.

References

Ali, W., Abdulai, A., and Mishra, A. K. (2020). Recent advances in the analyses of demand for agricultural insurance in developing and emerging countries. *Annual Review of Resource Economics*, 12:411–430.

Arnott, R. and Stiglitz, J. E. (1991). Moral hazard and nonmarket institutions: Dysfunctional crowding out or peer monitoring? *The American Economic Review*, 91(1):179–190.

Binswanger, H. P. (1980). Attitudes toward risk: Experimental measurement in rural india. *American Journal of Agricultural Economics*, 62(3):395–407.

Cai, J. (2016). The impact of insurance provision on household production and financial decisions. *American Economic Journal: Economic Policy*, 8(2):44–88.

Cai, J., de Janvry, A., and Sadoulet, E. (2020). Subsidy policies and insurance demand. *American Economic Review*, 110(8):2422–2453.

Cai, J. and Song, C. (2017). Do disaster experience and knowledge affect insurance take-up decisions? *Journal of Development Economics*, 124:83–94.

Clarke, D. J. (2016). A theory of rational demand for index insurance. *American Economic Journal: Microeconomics*, 8(1):283–306.

Cole, S., Giné, X., Tobacman, J., Topalova, P., Townsend, R., and Vickery, J. (2013). Barriers to household risk management: Evidence from india. *American Economic Journal: Applied Economics*, 5(1):104–135.

Conradt, S., Finger, R., and Spörri, M. (2015). Flexible weather index-based insurance design. *Climate Risk Management*, 10:106–117.

Dercon, S., Hill, R. V., Clarke, D., Outes-Leon, I., and Taffesse, A. S. (2014). Offering rainfall insurance to informal insurance groups: Evidence from a field experiment in ethiopia. *Journal of Development Economics*, 106:132–143.

Fafchamps, M., Udry, C., and Czukas, K. (1998). Drought and saving in west africa: are livestock a buffer stock? *Journal of Development Economics*, 55(2):273–305.

Gaurav, S., Cole, S., and Tobacman, J. (2011). Marketing complex financial products in emerging markets: Evidence from rainfall insurance in india. *Journal of Marketing Research*, 48(SPL):S150–S162.

Giné, X., Townsend, R., and Vickery, J. (2008). Patterns of rainfall insurance participation in rural india. *The World Bank Economic Review*, 22(3):539–566.

Giné, X. and Yang, D. (2009). Insurance, credit, and technology adoption: Field experimental evidence from malawi. *Journal of Development Economics*, 89(1):1–11.

Hazell, P. B. R. and Hess, U. (2010). Drought insurance for agricultural development and food security in dryland areas. *Food Security*, 2(4):395–405.

Hess, U. and Hazell, P. (2009). Sustainability and scalability of index-based insurance for agriculture and rural livelihoods. Working paper, IFPRI, Washington, DC.

Hill, R. V., Hoddinott, J., and Kumar, N. (2013). Adoption of weather-index insurance: learning from willingness to pay among a panel of households in rural ethiopia. *Agricultural Economics*, 44(4-5):385–398.

Hill, R. V., Kumar, N., Magnan, N., Makhija, S., de Nicola, F., Spielman, D. J., and Ward, P. S. (2019). Ex ante and ex post effects of hybrid index insurance in bangladesh. *Journal of Development Economics*, 136:1–17.

J-PAL (2016). Make it rain. In *J-PAL Bulletin*. Abdul Latif Jameel Poverty Action Lab, Center for Effective Global Action, and Agricultural Technology Adoption Initiative.

Janzen, S., Magnan, N., Mullally, C., Shin, S., Palmer, I. B., Oduol, J., and Hughes, K. (2020). Can experiential games and improved risk coverage raise demand for index insurance? evidence from kenya. *American Journal of Agricultural Economics*, 103(1):338–361.

Janzen, S. A. and Carter, M. R. (2018). After the drought: The impact of microinsurance on consumption smoothing and asset protection. *American Journal of Agricultural Economics*, 101(3):651–671.

Jensen, N. D., Barrett, C. B., and Mude, A. G. (2016). Index insurance quality and basis risk: Evidence from northern kenya. *American Journal of Agricultural Economics*, 98(5):1450–1469.

Jensen, N. D., Mude, A. G., and Barrett, C. B. (2018). How basis risk and spatiotemporal adverse selection influence demand for index insurance: Evidence from northern kenya. *Food Policy*, 74:172–198.

Karlan, D., Osei, R., Osei-Akoto, I., and Udry, C. (2014). Agricultural decisions after relaxing credit and risk constraints. *The Quarterly Journal of Economics*, 129(2):597–652.

Kimball, M. S. (1990). Precautionary saving in the small and in the large. *Econometrica*, 58(1):53.

Miura, K., Kanno, H., and Sakurai, T. (2012). Shock and livestock transactions in rural zambia: a re-examination of the buffer stock hypothesis. *The Japanese Journal of Rural Economics*, 14(0):20–34.

Miura, K. and Sakurai, T. (2021). News from the sky: An empirical test of forward-looking behavior among zambian farmers. In *International Association of Agricultural Economists (IAAE) Conference, Virtual, August 17-31, 2021*. International Association of Agricultural Economists.

Mobarak, A. M. and Rosenzweig, M. (2012). Selling formal insurance to the informally insured. Working Paper 97, Yale University, New Haven, CT.

Waldman, K. B., Blekking, J. P., Attari, S. Z., and Evans, T. P. (2017). Maize seed choice and perceptions of climate variability among smallholder farmers. *Global Environmental Change*, 47:51–63.

A Appendix

A.1 Mathematical appendix

A.1.1 Condition for insurance uptake

We show that the condition for insurance uptake, $r_i < p_i(1 - q_i)$, is equivalent to insurance payout being positively correlated with crop loss:

$$\begin{aligned} r_i < p_i(1 - q_i) &\Leftrightarrow \ \Pr(D = 0 \mid L_i = 1) \times p_i < p_i(1 - q_i) \\ &\Leftrightarrow \ \Pr(D = 0 \mid L_i = 1) < \Pr(D = 0) \\ &\Leftrightarrow \ \Pr(D = 1 \mid L_i = 1) > \Pr(D = 1). \end{aligned}$$

A.1.2 Comparative statics of insurance demand

The optimal unit of the insurance contract is $\alpha_i^* = \frac{p_i(1-q)-r_i}{q(1-q)} l_i$ when $p_i(1 - q) - r_i \geq 0$ holds. The partial derivatives of the optimal demand with respect to p_i and r_i are:

$$\frac{\partial \alpha_i^*}{\partial p_i} = \frac{l_i}{q} > 0 \ \text{ and } \ \frac{\partial \alpha_i^*}{\partial r_i} = -\frac{l_i}{q(1 - q)} < 0.$$

In contrast, the insurance payout probability q has an ambiguous effect on the farmer's insurance uptake. This is because a larger q increases not only the additional cost of insurance purchases (i.e., premium), but also the expected pay-off. To demonstrate this, the partial derivative with respect to q is:

$$\frac{\partial \alpha_i^*}{\partial q} = \frac{r_i(1 - 2q) - p_i(1 - q)^2}{q(1 - q)^2} l_i.$$

When $q < 1/2$, the marginal effect can be positive or negative. (p_i, q, r_i) determine the probabilities of false positives, that is, $q - p_i + r_i$. Given the pay-off structure and the effect of the false–positive state on the expected utility, the marginal impact of q is positive when r_i is sufficiently large and negative when p_i is sufficiently small:

$$\frac{\partial \alpha_i^*}{\partial q} \geq 0 \leftrightarrow r_i \geq \frac{p_i(1 - q)}{1 - 2q} \leftrightarrow p_i \leq \frac{r_i(1 - 2q)}{1 - q}.$$

A.2 Moments and correlations for the probability distribution

The specification of truncated normal and binormal distributions requires means and standard deviations of plot-level and weather station-level rainfall amounts and the correlation coefficient between these two rainfall variables.

1. Plot-level mean and standard deviation

The plot-level mean and standard deviation of plot i's rainfall during the flowering season were estimated by adjusting the five-year plot-level rainfall data with historical rainfall data from the Choma weather station, 1949/50–2011/12. Specifically, we estimated the population mean using the following estimator:

$$\hat{E}(R_i^{\text{fl}}) = \left(\sum_{t=2007}^{2011} R_{it}^{\text{fl}} / \sum_{t=2007}^{2011} R_{ct}^{\text{fl}} \right) \times \frac{1}{63} \sum_{t=1949}^{2011} R_{ct}^{\text{fl}}$$

The standard deviation was estimated in the same manner.

2. Weather station-level mean and standard deviation

We estimate the means and standard deviations for rainfall at the Choma weather station and Malima irrigation site using historical data.

3. The correlation coefficient between the plot-level and rainfall index for each contract was estimated by comparing plot-level flowering season rainfall and weather observation point-level rainfall based on a specific season and specific referred point per the insurance contract design. For example, we estimate the correlation coefficient for the 2011/12 rainfall index insurance using the following estimator:

$$C\hat{orr}(R_i^{\text{fl}}, R_c^{\text{ra}}) = \frac{1}{5} \sum_{t=2007}^{2011} (R_{it}^{\text{fl}} - \bar{R}_i^{\text{fl}})(R_{ct}^{\text{ra}} - \bar{R}_c^{\text{ra}}) / \left(\hat{SD}(R_i^{\text{f}}) \hat{SD}(R_c^{\text{ra}}) \right).$$

A.3 Other potential factors behind low demand for index insurance

In this section, we briefly discuss the role of other factors that might affect insurance demand. To facilitate the discussion, we regress the deviation of the optimal cover rate under the assumption of no product basis risk (CR_{it}^{o}) from the actual cover rate (CR_{it}^{a}) on potential determinants other than spatial basis risk. We interpret this deviation as an unexplained part of the low demand in our rational demand model. As other potential barriers, we consider those discussed in the literature: financial literacy (e.g., Gaurav et al., 2011; Janzen et al., 2020), past drought experience to test the recency bias (e.g., Cai and Song, 2017; Jensen et al., 2016), the presence of informal risk-sharing as existing mutual insurance (e.g., Arnott and Stiglitz, 1991; Dercon et al., 2014; Mobarak and Rosenzweig, 2012), and livestock assets as self-insurance tools (e.g., Fafchamps et al., 1998; Miura et al., 2012). As empirical proxies for these factors are collected only for the first two years, our estimation uses observations from these two years.

Table A7 presents the estimation results of the auxiliary regression analysis. First, the proxies for insurance literacy do not significantly predict insurance demand that is unexplained by our rational demand model. This result is in line with Cai and Song (2017) who report that knowledge of insurance does not affect insurance uptake per se, but that the experience acquired in insurance games significantly increases insurance demand. The combination of our results with their findings suggests that experiences may matter more for insurance demand than financial literacy.

Second, the perceived experience of drought in the previous year boosts insurance demand irrespective of the model specifications. Because the experience narrows the deviation in the cover rate, those who perceived drought purchased more insurance units. These robust results imply that previous experience may update farmers' perceived probability of drought. Another explanation is that farmers make insurance purchase decisions heuristically by assigning more weights to recent weather shocks. This behavioral recency bias is also essential for updating the belief on basis risk. In this light, Jensen et al. (2016, 2018) report the negative impact of experiencing the "basis error" on IBLI demand, implying that the dynamics of the perceived basis risk affect the insurance demand. Because little is known about how farmers update their understandings of basis risk across time, the role of perceptions of extreme weather events and basis risk and their dynamics needs further empirical scrutiny in future work.

Third, Table A7 shows that the extent of informal risk-sharing, proxied by the number of people who can call on in time of need, is not significantly related to insurance demand. This result is inconsistent with previous work suggesting that pre-existing informal risk-sharing may alleviate the discouraging effect of basis risk (Dercon et al., 2014; Mobarak and Rosenzweig, 2012). In our context, most farm households mutually share idiosyncratic risks among relatives and neighbors in geographically confined areas, which may not work well against aggregate shocks. This nature of village economies in the local area may prevent farmers from hedging against basis risk by leveraging the existing informal risk-sharing group. Finally, the estimation results show no significant correlations between livestock asset holdings and insurance demand.

Overall, our estimation results identify drought experience in the previous year as a predictor of insurance demand. However, we acknowledge that the empirical variables used in this exercise may be endogenous to the household insurance demand. Thus, our results do not reflect causal relationships.

A.4 Appendix figure and table

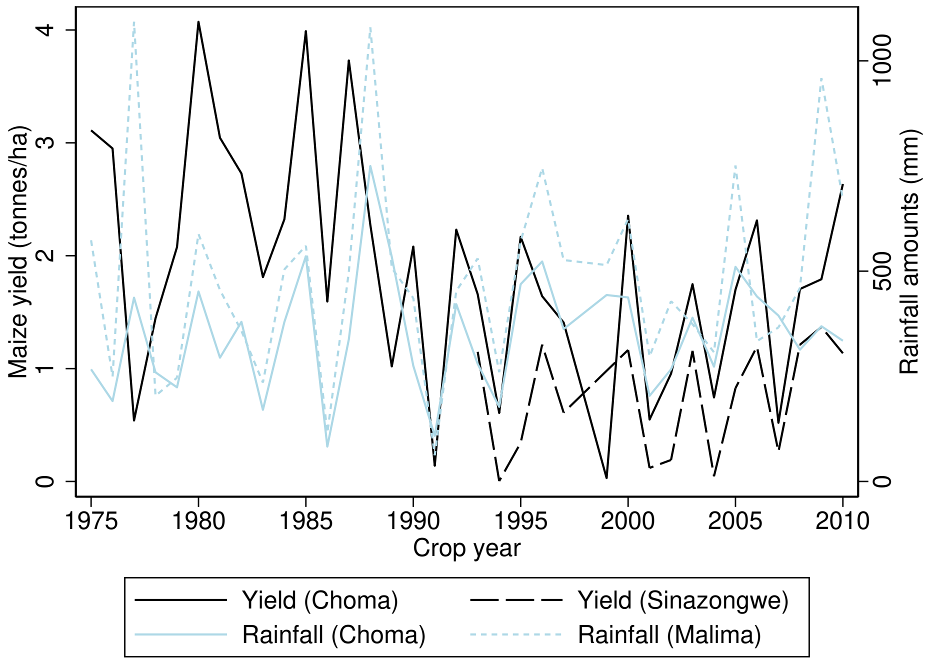


Figure A1: Flowering season's rainfall and maize yield in Choma and Sinazongwe districts, 1975/76–2010/11.

Source: Crop forecast survey data from the Central Statistical Office; Rainfall data of Choma and Malima from the Choma Meteorological Station of Zambia Meteorological Department (Mochipapa) and the Malima irrigation site, respectively.

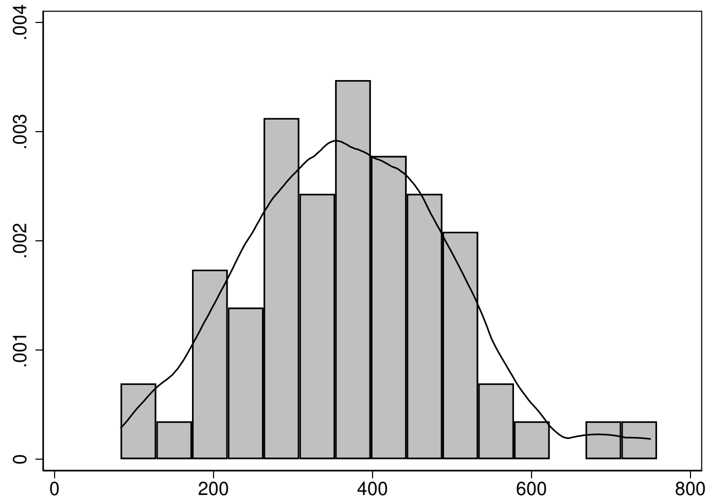


Figure A2: Distribution of rainfall in the flowering season, Choma 1949/50–2011/12

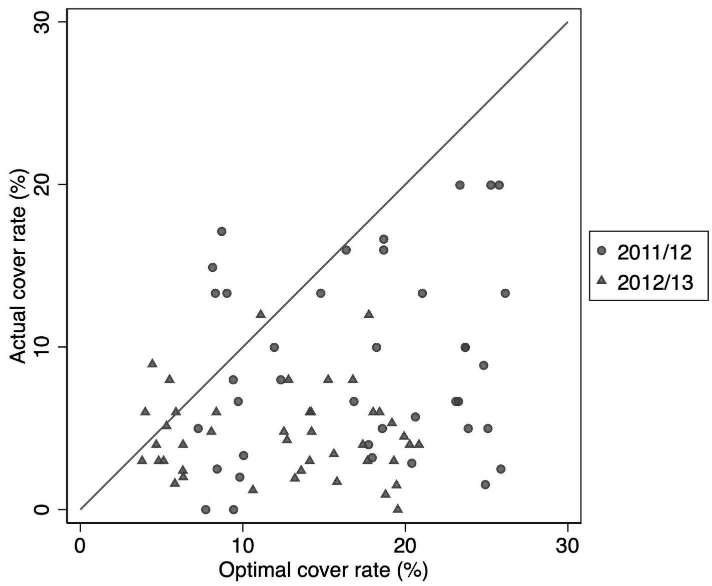


Figure A3: Distribution of actual cover rate and optimal cover rate

Table A1: Rainfall in Malima and maize yield in Choma and Sinazongwe districts, 1975/76–2010/11

	(1)	(2)	(3)	(4)
Rainfall				
Rainy season (100mm)	0.644***			
	(0.156)			
Rainy season, squared.	-0.030***			
	(0.009)			
Flowering season (100mm)		0.108	0.718***	
		(0.082)	(0.195)	
Flowering season, squared.			-0.055***	
			(0.019)	
Planting season (100mm)		0.035	0.591***	
		(0.073)	(0.151)	
Planting season, squared.			-0.062***	
			(0.015)	
"Drought" in 13/14 contract				-0.699***
				(0.232)
"Flood" in 13/14 contract				-0.933***
				(0.250)
Choma district	0.803***	0.830***	0.846***	0.812***
	(0.202)	(0.210)	(0.186)	(0.158)
Linear time trend	-0.037***	-0.035*	-0.038***	-0.029**
	(0.013)	(0.017)	(0.013)	(0.014)
Constant	-1.349*	1.030**	-1.198**	1.862***
	(0.712)	(0.429)	(0.580)	(0.426)
R-squared	0.527	0.388	0.587	0.454
Observations	51	51	51	51

Notes: The data sources were crop forecast survey data from the Central Statistical Office and rainfall data from the Malima irrigation site. The dependent variable is maize yield (mean = 1.52 tons/ha, std. dev. = 1.04). The sample covered the period between 1975/76 and 2010/11 for Choma district and between 1993/94 and 2010/11 for Sinazongwe district, with missing observations. "Drought" is a dummy variable that takes the value of 1 if the total rainfall amount during November and December (planting season) was below 214 mm, and 0 otherwise. "Flood" is a dummy variable that takes the value of 1 if the total rainfall amount during November and December (planting season) was above 800 mm, and 0 otherwise. After pooling maize yield data from the Choma and Sinazongwe districts, OLS was used for the estimations. Heteroskedasticity-robust standard errors are in parentheses. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A2: Design of Binswanger-style lottery

Option	Heads Low payoff	Tails High payoff	Risk aversion class	Coefficient of partial risk aversion
a	ZMW 5	ZMW 5	Extreme	$3.93 \leq \sigma \leq \infty$
b	ZMW 4	ZMW 12	Severe	$1.00 \leq \sigma \leq 3.93$
c	ZMW 3	ZMW 16	Intermediate	$0.56 \leq \sigma \leq 1.00$
d	ZMW 2	ZMW 19	Moderate	$0.27 \leq \sigma \leq 0.56$
e	ZMW 1	ZMW 21	Slight-to-neutral	$0.00 \leq \sigma \leq 0.27$
f	ZMW 0	ZMW 22	Neutral-to-negative	$-\infty \leq \sigma \leq 0.00$

Notes: ZMW 1 = about US$ 0.22

Table A3: Results of the normality test for rainfall in the flowering season at weather observation points

Test	Statistics	p–value
Choma meteorological station		
Shapiro–Wilk test	$W = 0.986$	$p = 0.688$
Skewness and kurtosis tests	$\chi^2 = 2.320$	$p = 0.313$
Kolmogorov–Smirnov test		
Normal distribution	$D = 0.055$	$p = 0.681$
Log-normal distribution	$D = 0.997$	$p = 0.000$
Observation	64	
Malima irrigation site		
Shapiro–Wilk test	$W = 0.973$	$p = 0.474$
Skewness and kurtosis tests	$\chi^2 = 2.250$	$p = 0.324$
Kolmogorov–Smirnov test		
Normal distribution	$D = 0.131$	$p = 0.246$
Log-normal distribution	$D = 0.971$	$p = 0.000$
Observation	41	

Notes: The table reports statistical test results for the normality of total rainfall in January and February for both the Choma weather station and the Malima irrigation site. For the Choma weather station, the rainfall data period used for the tests is between 1949/50 and 2011/12. For the Malima irrigation site, the rainfall data period used for the tests is between 1972/73 and 2012/13. The Kolmogorov–Smirnov test compares the observed distribution and the normal (log-normal) distribution using the sample mean and standard deviation.

Table A4: Purchased units of insurance contracts, 2011/12–2013/14

	2011/12		2012/13		2013/14	
	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.
Old sample	2.929	(2.224)	2.614	(1.956)	2.405	(1.231)
Site A	1.857	(1.460)	1.933	(1.032)	2.200	(0.775)
Site B	3.615	(2.219)	3.077	(1.706)	2.231	(1.013)
Site C	3.333	(2.554)	2.875	(2.630)	2.786	(1.717)
New sample	2.589	(2.357)	2.259	(1.884)	2.182	(1.272)
Combined	2.735	(2.295)	2.356	(1.904)	2.243	(1.260)
Total observations	100		160		152	

Table A5: Effect of basis risk on insurance demand: Fixed-effect estimates–bootstrap standard errors

	(1)	(2)	(3)	(4)	(5)
$r^s \times l$	-20.718**	-21.248*	-22.524*	-23.435*	
	(10.392)	(12.420)	(12.347)	(14.236)	
r^s		3.989		7.193	-19.765
		(22.618)		(21.414)	(18.844)
Cash reward			0.023	0.024	0.023
			(0.030)	(0.030)	(0.035)
Year = 2012	-1.102**	-1.050*	-1.402**	-1.304**	-0.894
	(0.520)	(0.559)	(0.621)	(0.636)	(0.568)
Year = 2013	-0.095	-0.157	-0.254	-0.366	-0.422
	(0.387)	(0.423)	(0.464)	(0.446)	(0.463)
RA – Extreme/Severe			-0.604	-0.570	-0.518
			(0.448)	(0.464)	(0.461)
RA – Intermediate			0.433	0.462	0.349
			(0.573)	(0.579)	(0.531)
RA – Slight			0.878	0.911	0.871
			(0.546)	(0.571)	(0.537)
RA – Neutral			0.230	0.261	0.151
			(0.605)	(0.618)	(0.543)
Constant	4.669***	4.392***	4.743***	4.202**	4.442**
	(0.974)	(1.589)	(1.340)	(1.688)	(1.871)
Risk aversion class	No	No	Yes	Yes	No
R squared	0.678	0.678	0.736	0.736	0.680
Observations	126	126	126	126	126

Notes: The dependent variable is the number of insurance contracts purchased by a household. RA denotes risk aversion class and RA–Moderate is used as a reference. Robust standard errors computed using 1000 bootstrap replications are in parentheses. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A6: Test of the model predictions with a restricted sample

	(1)	(2)	(3)	(4)	(5)	(6)
$r^s \times l$	-25.221	-32.843**	-27.947	-34.281**	-25.904	-30.882**
	(17.505)	(14.929)	(17.100)	(16.248)	(17.763)	(14.682)
$p^d \times l$	22.316*	50.080***	23.499**	47.877***	21.978**	60.082***
	(11.403)	(14.776)	(9.951)	(16.775)	(10.156)	(15.974)
r^s		56.014		51.621		102.063***
		(42.651)		(32.164)		(33.629)
p^d		-75.406**		-68.810**		-128.611***
		(36.419)		(29.687)		(40.199)
l		-3.190		-2.803		-4.515**
		(2.234)		(2.283)		(2.120)
RA – Extreme/Severe			-2.120***	-1.989**	-1.983***	-1.664***
			(0.767)	(0.776)	(0.652)	(0.587)
RA – Intermediate			0.030	0.259	0.412	0.874
			(0.706)	(0.748)	(0.626)	(0.632)
RA – Slight			-0.325	-0.213	-0.169	-0.053
			(0.716)	(0.761)	(0.722)	(0.685)
RA – Neutral			-0.752	-0.591	-0.475	-0.319
			(0.731)	(0.795)	(0.676)	(0.606)
Cash reward			0.014	0.011	-0.001	-0.026
			(0.036)	(0.033)	(0.041)	(0.039)
Year = 2012	-1.112*	-0.385	-1.808**	-1.115	-1.740**	0.003
	(0.599)	(0.974)	(0.675)	(0.892)	(0.680)	(0.996)
Year = 2013	-0.137	-1.583	-0.703	-2.018**	-0.665	-3.375***
	(0.633)	(1.248)	(0.644)	(0.943)	(0.660)	(0.986)
Constant	2.198***	8.139**	3.422***	8.668***	3.791***	10.854***
	(0.502)	(3.064)	(0.907)	(2.778)	(0.840)	(2.718)
HH characteristics	No	No	No	No	Yes	Yes
R squared	0.209	0.256	0.408	0.446	0.450	0.524
Observations	82	82	82	82	82	82

Notes: The estimation sample is restricted to observations in sites B and C. The dependent variable is the number of insurance contracts purchased by a household. RA denotes risk aversion class and RA–Moderate is used as a reference. Heteroskedasticity-robust standard errors are in parentheses. Household characteristics include the household head's sex, age, and years of schooling. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A7: Other potential constraints to insurance demand

	(1)	(2)	(3)	(4)	Descriptive statistics	
					2011/12	2012/13
Literacy	0.787	0.909	0.325	0.158	3.524	3.773
	(1.419)	(1.478)	(1.326)	(1.303)	(1.215)	(1.054)
Math	-0.745	-0.614	0.408	1.477	1.929	1.932
	(1.306)	(1.385)	(1.262)	(1.197)	(1.156)	(1.129)
Drought	-6.331***	-6.312***	-7.314**	-8.134**	0.0952	0.818
	(1.934)	(1.992)	(2.770)	(3.390)	(0.297)	(0.390)
Risk sharing	0.064	0.050	0.209	0.209	2.643	3.886
	(0.356)	(0.365)	(0.392)	(0.417)	(3.051)	(2.990)
Small livestock, asinh		-0.141	-0.174	-0.239	4.292	5.576
		(0.303)	(0.328)	(0.349)	(4.604)	(3.278)
Large livestock, asinh		0.020	-0.060	-0.207	5.867	6.118
		(0.418)	(0.398)	(0.408)	(4.525)	(2.515)
Cash reward			-0.055	-0.065		
			(0.176)	(0.173)		
Constant	6.129	5.868	5.047	5.613		
	(4.161)	(4.924)	(4.975)	(7.189)		
Year fixed effect	Yes	Yes	Yes	Yes		
Risk aversion class	No	No	Yes	Yes		
HH characteristics	No	No	No	Yes		
R squared	0.059	0.062	0.158	0.232		
Observations	85	85	85	85		

Notes: The dependent variable is the deviation in cover rates (%). The estimation sample is restricted to the 2011/12 and 2012/13 crop years. OLS was used for the estimation. Heteroskedasticity-robust standard errors are in parentheses. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Variable descriptions: Literacy—a total score of five questions assessing the comprehension of the insurance product. Math—a total score of three basic arithmetic questions. Drought—the perceived experience of drought in the previous crop year (dummy). Risk sharing—the number of people who can be called in times of need. Small livestock—total value of goats, pigs, and chickens the household owns. Large livestock—total value of cows and oxen the household owns.