Landscape heterogeneity, basis risk and the feasibility of index insurance: An analysis of rice in upland regions of Southeast Asia

DOI:10.1016/j.foodpol.2022.102237

Food Policy 108 (2022) 102237

Contents lists available at ScienceDirect

Food Policy

journal homepage: www.elsevier.com/locate/foodpol

Landscape heterogeneity, basis risk and the feasibility of index insurance: An analysis of rice in upland regions of Southeast Asia

Roberta Rigo [a,*], Paulo Santos [b], Vito Frontuto [a]

[a] Department of Economics and Statistics, University of Torino, Lungo Dora Siena 100, 10124 Torino, Italy
[b] Department of Economic, Monash University, Caufield, VIC, Australia

ARTICLE INFO

Keywords:
Index insurance
Basis risk
NDVI
Landscape heterogeneity
Clustering

ABSTRACT

Despite its promise, the adoption of index insurance has been hindered by the extent of basis risk, the additional variability introduced by its reliance on a signal correlated with losses rather than losses themselves. We examine the feasibility of substantially reducing basis risk by accounting for heterogeneity in production conditions via clustering data into more homogeneous groups. We exemplify this approach using data from a sample of rice producers in northern Laos, using Normalized Difference Vegetation Index (NDVI) data as the index on which the contract is defined. Our results show that accounting for landscape heterogeneity substantially improves the insurance contracts that can be offered to rice producers.

1. Introduction

Risk is ubiquitous in agriculture, influencing a variety of production decisions (Moschini and Hennessy, 2001). In the absence of formal insurance, agricultural households in developing countries adopt a variety of strategies to reduce consumption variability, in the face of a large variation in income (Morduch, 1995). Such strategies come with two limitations.

The first is that, in general, they are not capable of insuring against covariate shocks, even when effective to smooth consumption against idiosyncratic shocks (Townsend, 1994; see, however, Riley (2018) for a recent discussion of the feasibility of insuring against covariate shocks). The second is that these strategies may come at a large cost that may perpetuate poverty (Dercon and Christiaensen, 2011; Hoddinott and Kinsey, 2001; Zimmerman and Carter, 2003). In developing countries, where the importance of agriculture is still large in terms of both growth and poverty reduction (Ligon and Sadoulet, 2018; Ravallion and Datt, 2002), uninsured risk may then contribute to poverty persistence (Barnett et al., 2008).

The recognition that multi-peril insurance, adopted throughout the 20th century in developed countries (Smith and Glauber, 2012) has important practical shortcomings (including problems of moral hazard and adverse selection, compounded by high transaction costs; see Hazell (1992) for an early discussion) led to the relatively recent development of index insurance products. The promise of index insurance is to reduce high transaction costs and informational asymmetries by structuring insurance payments as a function of an easily measurable and objective index correlated with losses rather than as a function of losses themselves. Since the early 2000s several pilot studies have attempted to determine first the feasibility of this approach, and then their uptake and impact. Several recent reviews (for example, Carter et al., 2017; Cole and Xiong, 2017; De Leeuw et al., 2014) converge on two conclusions: insurance can unlock investment and promote growth, but its effectiveness as a poverty reduction strategy is severely diminished by the low demand for insurance by potential beneficiaries.

In understanding the puzzle of low demand, it is important to recognize that the imperfect correlation between index and losses implies that there will be states of the world that should correspond to payments (because insured losses occurred), but in which payments will not occur (because losses were not predicted), and vice-versa. These two states of the world may impact on demand for insurance in different ways. The first, negative basis risk, can explain why a risk averse decision maker who maximizes his/her expected utility and faces an upfront payment may rationally not buy insurance, given that in those states of the world s/he is worse off than without insurance (Clarke, 2016). The second, or positive basis risk, leads to higher premiums and a lower quality of the insurance product which may compound difficulties in understanding its functioning and utility (Carter 2009).

Empirically, the importance of negative basis risk as an explanation for lower demand has been shown by Mobarak and Rosenzweig (2013),

* Corresponding author at: Department of Geography, University of Rennes 2, Place Recteur Henri le Moal, 35000 Rennes, France.
E-mail addresses: roberta.rigo@univ-rennes2.fr (R. Rigo), paulo.santos@monash.edu (P. Santos), vito.frontuto@unito.it (V. Frontuto).

https://doi.org/10.1016/j.foodpol.2022.102237
Received 11 January 2021; Received in revised form 10 December 2021; Accepted 14 February 2022
Available online 4 March 2022
0306-9192/

who experimentally estimate the positive effect of locating weather stations (on which contracts will be based) at village level on demand for index insurance in India. Similarly, using longitudinal data for Index Based Livestock Insurance, Jensen, Mude, and Barrett (2018) found that such risk is negatively related to insurance uptake.[1]

This paper is concerned with one way to improve the design of index insurance, by exploring the value of considering the heterogeneity of the landscape, when it may plausibly shape production conditions and the strength of the correlation between the index and yield. We illustrate this possibility through the definition of different contracts that insure rice producers in northern Laos against production losses, building on recent attempts to explore the use of time and site-specific satellite data on Normalized Difference Vegetation Index (NDVI) to predict output.[2]

The rest of this article is structured as follows. Section 2 presents the geographic context and the data used. Section 3 discusses the approach followed to address the potential importance of landscape heterogeneity and estimate different yield prediction models used to construct a satellite-based index insurance. Using those results, we then discuss the properties of different contracts based on estimates of basis risk and cost. Our results, presented in Section 4, suggest that accounting for heterogeneity in natural conditions allows for substantial improvements in the design of the insurance product, leading to either substantial reductions in premiums paid for the contract, reductions in negative basis risk or both. We conclude with a discussion of the policy implications of these results. We emphasize that, despite these improvements, index insurance is likely to still require substantial subsidies, or other changes in contract design, to be attractive.

2. Rice in northern Laos: Context and data

Rice production and consumption stands out in the context of the economic activity of Laos, with almost half of the total cultivated land allocated to this crop.[3] Following the introduction, in 1986, of economic reforms that aimed to switch from a planned economy to a market-oriented one, overall rice productivity increased since the nineties, with the country experiencing an increasing rice surplus in the 2000s (Goto and Douangngeune, 2017). Such improvements in productivity were achieved through the adoption of modern cultivation techniques including the use of inorganic fertilizer, mechanization and the introduction of improved rice varieties (Goto and Douangngeune, 2017; Newby et al., 2013).

Along with this national picture, it emerged a scenario of regional differentiation: productivity gains were largely concentrated in the central and southern regions of the country, particularly in the plains along the Mekong (Eliste and Santos, 2012; Sacklokham, 2014; World Bank, 2018) while bypassing the northern region, which is still reliant on an integrated rice market to overcome any local production deficit (Newby et al., 2013; Sacklokham, 2014) and where the production system is characterized by manual labor and low use of chemical inputs, leading to relatively low yields (Schiller et al., 2006; World Bank, 2018).

Eliste and Santos (2012) suggest that the main reason for this divergence lies in the region's large heterogeneity in terms of biophysical characteristics. Production technology is largely determined by topographic characteristics, leading to three main agro-ecosystems, namely irrigated lowland, rainfed lowland, and upland. While lowland rice (paddy rice) can grow both during the dry and the wet seasons on submerged plots (either rainfed or irrigated), upland rice grows on rainfed dry soil, and production is concentrated on the wet season, as irrigation is almost never feasible (Goto and Douangngeune, 2017).[4]

The production of rice, in particular rainfed rice, in Laos has always been affected by weather shocks (Eliste and Santos, 2012). Given that, on average, the northern region receives a lower amount of rainfall (1566 mm/year) than the rest of the country (with Luang Prabang, one of the provinces in our study, registering the second lowest mean annual rainfall rate), drought is also the climate shock that most affects rice production in northern Laos (Basnayake et al., 2006; Schiller et al., 2006). The timing of lack of rain also matters (Schiller et al., 2006): at the time of seeding, germination and establishment (between the end of April to end of May), monthly rainfall needs are fairly high ranging from 100 mm/month (upland rice) to 200 mm/month (lowland rice), and water shortages in that period can dramatically reduce production. At later stages of the plant's development, the impact of eventual droughts seems to be less damaging.

2.1. Data

We use three types of data. Ground data on rice production over a period of three years (2017–2019) was collected as part of a household survey fielded in 72 villages across four districts in northern Laos (Pakxieng and Viengkham in the province of Luang Prabang province, and Kham and Phoukout, in the province of Xiengkhuang province). We interviewed 12 households per village, randomly selected from the village roster. Attrition is relatively low (9% in 2018 and 8% in 2019), as is the percentage of households that do not grow rice (~7.5% per year). In each round of the survey, conducted shortly after the end of the harvest (typically, December or January) to minimize recall bias, respondents provided information on production (in kg) and area (in ha) in the previous wet season. In 2019 we also collected detailed data on rice production technology in the previous wet season.

For each household we have information on rice production (in kg) and land area (in hectares) for the period 2017–19. Yield is estimated based on recall data, which has known measurement problems (see for example the recent discussion in Lobell et al. (2018)) which we address in two ways. Firstly, we define yields above 1.5 times the interquartile range as outliers, and we remove them from the analysis. The number of outliers is relatively small (83 over 2187 observations) and their inclusion in the analysis does not substantially change our conclusions about the design of the insurance product (see Appendix C). Secondly, we use detailed data at the plot level such as the access to a road, the presence of mechanization and the use of inputs during the 2018 wet season, for which we collected information to estimate a production function, both including and excluding outliers. We computed basis risk and premiums on the whole sample by using predicted yields instead of observed ones

[1] Perhaps because negative basis risk undermines the poverty reduction potential of index insurance, it has received relatively more attention than positive basis risk (see Clement et al. 2018 for a recent review). However, and to the extent that overall lower quality on the insurance contract (summarized by the sum of both positive and negative basis risk) transforms the understanding of the insurance product into essentially a lottery, large positive basis risk may reduce the demand for insurance over and above the associated increase in cost. For example, lab-in-the-field evidence suggests that the presence of basis risk turns the decision of buying insurance against risk into a problem of compounded probabilities (the probability of loss and the probability of payment, given a loss). Compound risk can lower demand for insurance (Elabed and Carter, 2015), leading to welfare loss, but specific insurance design can overcome this issue (Harrison et al., 2020). However, we do not know of any study that attempts to separately evaluate the effect of positive basis risk over and above its effect on cost.

[2] See Vroege et al. (2021) for a recent discussion of the promises and limitations of using satellite data for insurance against drought.

[3] As a reflection of the cultural importance of rice in the country, it is perhaps enough to notice that the Laos expression "to eat" also means "to eat rice" (Schiller et al., 2006).

[4] Despite what the name may suggest, it is important to emphasize that there is no association between this distinction and the field's slope or elevation: paddy rice can be produced in mountainous regions and upland rice can be produced in flat fields (Liquist et al., 2006). However, since it is impossible to flood steep plots, most of the upland rice is cultivated in plots with a high slope while paddy rice is cultivated in flat ones.

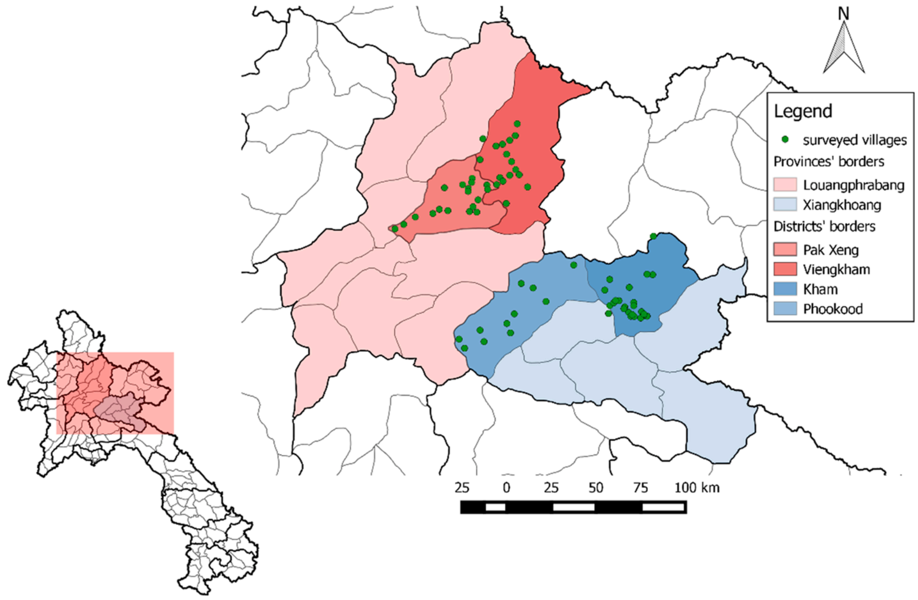

Fig. 1. Location of surveyed villages in northern Laos.

(Appendix B). In both cases, the predictive power of this relation is relatively high, allaying fears that measurement error associated with yield may be driving our conclusions about the importance of basis risk.[5]

Although relatively concentrated in space (see Fig. 1), there is substantial variability in agro-ecological conditions that may influence production in our sample: for example, village elevation ranges from 318 to 1493 m above sea level (masl). We characterize this diversity using a second dataset of village characteristics, described in Field and Odgers (2016). Given the differences in agroecosystems discussed before, we are particularly interested in elevation and the relative importance of the area under different slope ranges, which determines the feasibility of irrigation. We will use these two variables to characterize the heterogeneity of natural production conditions and to determine clusters.

In addition, that same dataset provides estimates, at village level, of soil's physical and chemical properties, including data on soil texture (clay, sand and silt content), acidity (pH) and soil fertility (cation exchange capacity or CEC, and the percentage of organic material in the soil) as well as soil depth. One important characteristic of these variables is that they are potentially related to yield but are unlikely to change significantly over time, hence will not reflect farmers' technological choices in the short-run. We include them as additional covariates in the statistical models used to design the index insurance product.

As an index, we use data on the Normalized Difference Vegetation Index (NDVI) from ORNL DAAC website that provides MOD13Q1 Vegetation Indices from the MODIS satellite every 16 days for each pixel at a maximum resolution of 250 m per side.[6] We use data for the period 2001–2019, and define a village as a polygon with 2.25 km per side centered on the village coordinates, a choice that reflects the fact that

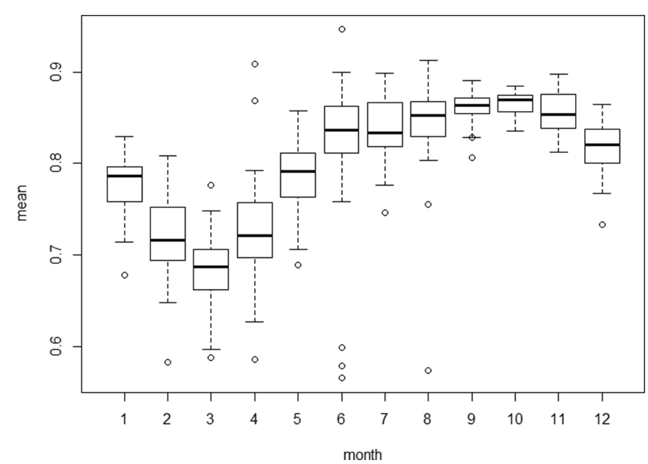

Fig. 2. Boxplot of NDVI values per month in Hadkeo village (Authors' calculation using data from MODIS and VIIRS Land Products Global Subsetting and Visualization Tool (2001–2019)).

most plots are fairly close to the village center.[7] We find no evidence that NDVI would change significantly when we consider larger areas, suggesting that our analysis is robust to this decision.[8]

[5] Despite their relatively high predictive power, those variables have the important limitation that they are either an individual choice and/or too difficult or expensive to audit. Although insurance companies can potentially collect accurate data at household level by conducting surveys on-site, this would lead to high administrative costs and it would typically reintroduce problems of information asymmetry, undermining the utility of index insurance. For these reasons, we omit them from the design of the insurance.

[6] Data available at: https://modis.ornl.gov/globalsubset/. We also use Gross Primary Production (GPP) data. The results are substantially identical and we omit them for the sake of brevity, but they are available from the corresponding author upon request.

[7] Given pixel dimensions (6.25ha) and the average plot size (approximately 1ha) this approach, although perhaps closer to true production shocks, is still not perfect. One way to increase prediction accuracy (and reduce basis risk) would be to use satellite data at a plot level. Although accessing satellite data at that scale is technically feasible, getting data on the exact location of the plot would require an inspection by the insurer (with increased costs) and likely increase the importance of moral hazard and adverse selection problems, and potentially collusion between neighbors. For these reasons we do not use data on exact plot location and explain yield variability through the use of satellite data measured at the village level.

[8] In one tenth of the villages we extended the area in the analysis to polygons with 5km and 7.5km sides. We expected a reduced variability in the NDVI (given that areas of permanent vegetation are more important as we move away from the village center) but the results of this exercise don't confirm this hypothesis leading us to conclude that the selected pixel dimension doesn't seem to substantially influence the corresponding NDVI values.

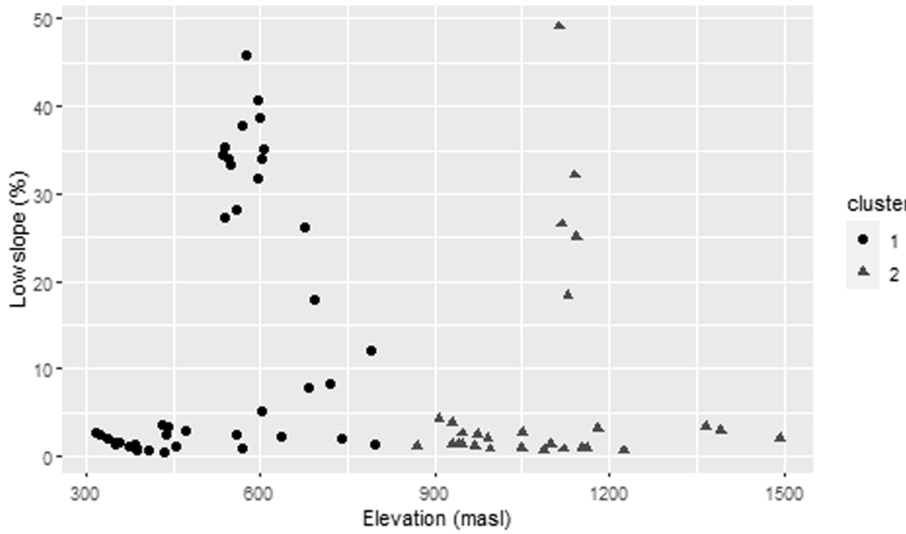

Fig. 3. Clustering sampled villages.

Being a measure of vegetation status, NDVI is expected to be lower in the first part of the year, and to increase after the transplanting stage (April/May, in the case of rice in northern Laos), reaching its peak at the end of reproductive stage (September), and decreasing at the time of harvest (late November-December) (Turvey and McLaurin, 2012, Mosleh et al., 2015). This pattern is shown in Fig. 2 for one village in our sample.

3. Index insurance in the presence of heterogeneity

In our empirical analysis we adopt the standard train-test split approach to inference evaluation (for a detailed description of the procedure see, for example, James et al., 2013), with 80% of the available data randomly allocated to the training set and used to estimate the relation between NDVI and yield, while the remaining 20% were allocated to the test set and used to calculate the differences between expected and observed yield and the associated root mean square error (RSME) of the prediction model.

3.1. *Quantifying the importance of landscape heterogeneity*

We use non-hierarchical cluster analysis with cluster centers to quantify the potential impact of heterogeneity in biophysical conditions, which may determine production technologies, on yield predictions. As environmental characteristics that could impact average yield levels, we consider elevation and topography (measured by the percentage of the village's area with a slope below 3%, which we take as indication of the feasibility of growing paddy rice). We applied both a non-hierarchical (k-means) and a hierarchical aggregation procedure using the Euclidean distance measure and the Ward's linkage method to maximize the internal homogeneity of clusters (Murtagh and Legendre, 2014). The results of the hierarchical clustering are usually presented in dendrograms from which the number of clusters is selected by visual inspection. This criterion can be inefficient and misleading and, instead, we followed Charrad et al. (2014) and used a set of 30 numerical and graphical indices. Applying the majority rule, we identified 2 clusters as optimal grouping (Charrad et al., 2014). These are represented in Fig. 3, suggesting that elevation is the main determinant of clustering, with observations split around the value of 870 masl.

Having defined production conditions that are more homogeneous, a natural follow-up question is whether accounting for such clustering matters when designing an index insurance contract. We address it by estimating the following model:

Table 1
Quantifying the importance of covariate risk.

	Whole sample	Cluster 1	Cluster 2
Village Average Yield	0.935***	0.949***	0.895***
Constant	0.4953*	0.396	0.800.
Adj. R^2	0.372	0.453	0.232
N	1683	964	719

Significance level: ***p ≤ 0.01, ** p ≤ 0.05, *p ≤ 0.10
Notes: Estimates on pooled data over the three years using OLS model (eq. (1)). Estimates are calculated for the whole dataset, as well as for cluster 1 (low elevation villages) and cluster 2 (high elevation village). Sample size (N) and the adjusted R squared are presented in rows 2 and 3, respectively.

$$Y_{qvt} = f(AY_{vt}) + \varepsilon_{qvt} \tag{1}$$

where Y_{vqt} stands for the yield of plot q in village v at time t, which we express as a linear function of average yield in other plots in the village (AY_{vt}) and idiosyncratic risk (ε_{qvt}).

This model allows us to understand if plot yield is correlated with average yield in the village (as a measure of covariate shocks against which index insurance may provide protection). Although useful as a preliminary step, this model is not feasible as a basis to define an insurance contract, given the lack of inexpensive and credible ways to measure average yield in a village. Table 1 presents the empirical pooled data estimates of the importance of covariate risk for the whole sample and for each of the two clusters identified, allowing for two conclusions.[9]

The first conclusion is that average yield (at village level) is a good predictor of yield at household level in our sample, showing that an index insurance product is potentially valuable. The second conclusion is that there are substantial differences in the predictive power of (village level) average yield, as measured by the adjusted R^2 of this regression: the strength of this relation in cluster 2 (higher elevation villages) being approximately half of the value in cluster 1. In short, clustering changes the importance of covariate risk, suggesting that the definition of contracts for each group may potentially improve yield prediction. In the next section, we further quantify the implications of accounting for this heterogeneity when defining a satellite-based index insurance for rice.

[9] An alternative specification of equation (1) would include time fixed effects. All results are robust to this alternative specification.

Table 2
Predicting (ln) rice yield: different statistical models, all observations.

ln(yield)	OLS	STEPWISE	VIF	LASSO	ELASTIC NET	PCA
April – 4th week	0.001 ***	0.001 ***	3.05e−09 ***	−0.109	−0.096	
May – 1st week	0.046 *	0.036 *	0.032 *	0.045	0.040	
May – 4th week	1.72e−11 ***	5.01e−12 ***	1.63e−08 ***	0.145	0.124	
June – 1st week	0.000 ***	0.000 ***	0.001 ***	−0.061	−0.057	
June – 4th week	0.000 ***	0.000 ***	0.006 **	−0.067	−0.057	
July – 1st week	0.9100		0.277	0.001	−0.002	
July – 4th week	0.1560	0.115	0.217	−0.025	−0.025	
August – 1st week	0.0790	0.128	0.015 *	0.0323	0.032	
August – 4th week	0.687		0.544	−0.007	−0.002	
September – 1st week	0.007 **	0.003 **	0.048 *	0.055	0.045	
September – 4th week	0.524		0.090	−0.016	−0.020	
October – 1st week	0.269		0.055	0.032	0.007	
November – 4th week	−2.30e−05 ***	1.15e−05 ***		−0.203	−0.157	
November – 1st week	0.048 *	0.040 *		−0.090	−0.103	
Soil pH	0.812		0.053	0.003	0.003	
Cation exchange capacity	0.000 ***	0.000 ***	0.001 ***	−0.060	−0.056	
Organic concentration 0–50 cm	0.091	0.123	0.083	0.028	0.030	
Clay content 0–30 cm	0.812		0.684	−0.003	−0.004	
Clay content 0–50 cm	0.738		0.993	0.006	0.008	
Sand content 0–15 cm	0.812	0.008 **	0.003 **	0.044	0.042	
Soil Depth	0.509		0.441	−0.011	−0.007	
Zpre	0.343		0.221	0.016	0.013	
Elevation (m)	5.04e−06 ***	4.34e−07 ***	0.001 ***	0.109	0.087	
Low Slope (%)	2.12e−07 ***	1.34e−08 ***	< 2e−16 ***	0.152	0.142	
PC1						< 2e−16 ***
Constant	< 2e−16 ***	< 2e−16 ***	< 2e−16 ***	7.658	7.658	< 2e−+16 ***
N	*1683*	*1683*	*1683*	*1515*	*1515*	*1683*
R^2	*0.309*	*0.308*	*0.291*	*0.304*	*0.304*	*0.230*
Adj.- R^2	*0.299*	*0.301*	*0.282*	*0.293*	*0.307*	*0.228*
RMSE	*0.547*	*0.546*	*0.554*	*0.544*	*0.543*	*0.574*

Significance levels: ***$p \leq 0.01$, ** $p \leq 0.05$, *$p \leq 0.10$.
Notes: Columns present results of different statistical models (OLS, STEPWISE, VIF, LASSO, ELASTIC NET and PCA). Both Elastic Net and Lasso follow a training-test approach that leads to a reduction in sample size (*N*). The first 14 rows refer to the NDVI signal from mid-April to mid-November (2 values per month). Estimates of the effect of environmental conditions (soil conditions, landscape heterogeneity and the state of vegetation before the contract period (Zpre)) are presented in rows 15–24. R2 and R2-adjusted are reported along with RMSE in the last rows.

3.2. Defining the NDVI index

As suggested by Chantarat et al. (2013), period-specific observed NDVI data can be compared with its past distribution to get an index of anomalies. This involves obtaining the standardized NDVI by subtracting the mean for the corresponding period over the entire period for which we have data (19 years) and dividing it by the corresponding standard deviation.

$$Z_{vjt} = \frac{NDVI_{vjt} - ENDVI_{vj}}{\sigma_{vj}} \tag{2}$$

where $NDVI_{vjt}$ is the NDVI value for the village v measured in period j of year t, $ENDVI_{vj}$ is its expected value and σ_{vj} is the corresponding standard deviation. The value of Z_{vjt} can be interpreted as a measure of anomaly in the vegetation index, expressed in terms of standard deviations. Positive values of Z_{vjt} will then represent lusher vegetation than the long-term mean for period j, while negative values will represent the opposite (Chantarat et al., 2013).

As the major covariate shock affecting the region is drought, we also follow Peters et al. (2002), and investigate whether a NDVI transformation that can be used as an indicator of this shock improves our estimates of yield. Knowing the distribution of Z, it is possible to calculate the standardized vegetation index (*SVI*) value for each village and each period as:

$$SVI_{vjk} = P(Z < Z_{vjk}) \tag{3}$$

This value can be interpreted as the "probability of occurrence of the present vegetation condition at a given location relative to the possible range of vegetative vigor" (Peters et al. 2002, p.71), which has been proved to be a good proxy for major shocks.

Finally, we use the NDVI data to account for the state of the vegetation before the contract period and losses, following Chantarat et al. (2013) finding a strong negative correlation between that variable and their measure of losses (in their case, livestock mortality). Following their study (and adapting the period range to ours) we calculate these values (labeled Z_{prevt}) for each village v in year t in the following way:

$$Z_{prevt} = \sum_{j \in K_{pre}} Z_{vjt} \tag{4}$$

where, in our case, j is the period and K_{pre} goes from October of the previous year to the first half of April.

In order to obtain the best predictive model for yield, we estimate the following model

$$Y_{qvt} = f(S_{vt}, E_v) + \varepsilon_{qvt} \tag{5}$$

where Y_{qvt} is, as above, the yield of plot q in village v in year t, that we express as a function of the index S_{vt} (Z or SVI), environmental conditions measured at a village level (E_v, including soil properties at village level and *Zprevt*) and a stochastic component (ε_{qvt}) that incorporates idiosyncratic risk and design error.

Finally, we consider different statistical models. The relatively high (biweekly) frequency of the satellite data can lead to problems of multicollinearity, which may result in coefficients with unrealistic magnitude (O'Brien, 2007; Yoo et al., 2014). To reduce the importance of this problem, we estimated the relation between yield and signal (with and without additional controls) using five statistical models: Ordinary Least Squares (OLS), Stepwise regression, OLS considering the Variance Inflation Factor (O'Brien, 2007), LASSO and Elastic Net regressions (Hastie et al., 2015) and OLS using Principal Components (Aguilera et al., 2006).

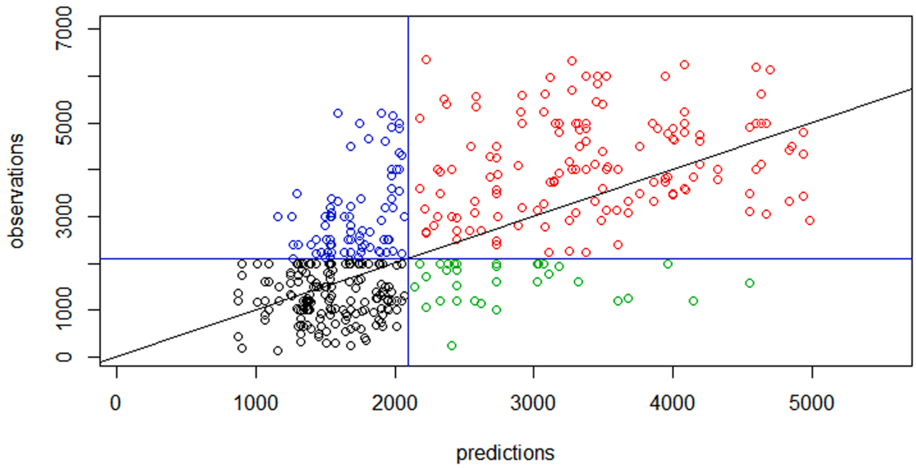

Fig. 4. Predicted vs observed yield (Note: Yield predicted on the entire training sample using elastic net. Trigger level set at 80% of the average yield). Black points represent households who experienced a loss and correctly received indemnities, red points represent households who did not experience a loss and received no indemnity. Blue points represent households who did receive an indemnity even though they did not experience a loss. Finally, green points represent households who did not receive any indemnity even though they experienced a loss. The size of this last group determines the importance of basis risk. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
Basis risk and premiums for different trigger levels.

	All			Cluster 1			Cluster 2		
Trigger	Negative Basis risk (%)	Positive Basis risk (%)	Premium (% average yield)	Negative Basis risk (%)	Positive Basis risk (%)	Premium (% average yield)	Negative Basis risk (%)	Positive Basis risk (%)	Premium (% average yield)
0.70	0.274	0.125	0.059	0.219	0.108	0.080	0.265	0.087	0.007
0.75	0.142	0.137	0.082	0.199	0.095	0.097	0.213	0.126	0.016
0.80	0.107	0.215	0.110	0.182	0.109	0.114	0.189	0.182	0.028
0.85	0.104	0.221	0.138	0.103	0.136	0.132	0.152	0.312	0.045
0.90	0.099	0.250	0.169	0.075	0.127	0.149	0.126	0.311	0.065
0.95	0.074	0.260	0.201	0.072	0.134	0.168	0.127	0.325	0.088
N		1683	964	719					

Notes: Calculations of basis risk and premium are based on elastic net predictions of yield in the test sample at different levels of the trigger. N is the number of observations on which the Basis risk and Premium are calculated. Assuming an actuarially fair contract and price normalized to 1, the premium is the sum of the differences between the predicted yield and the trigger for all observations for which the expected yield is lower than the trigger (see eq. A.1 in the Appendix). Premia are expressed as percentage of the average yield of the sample.

4. Results

4.1. Choice of statistical model

All models were estimated for the whole sample and for each cluster, with the best model selected on the base of the RMSE calculated on the test sample. The results of the different models for the whole sample, when yield is expressed in natural logarithms and the signal is expressed as Z_{vjt}, are presented in Table 2.[10] The best results are obtained when using the Elastic Net regression (RMSE = 0.543). The same statistical model is chosen when splitting the data into clusters (RSME in the first cluster = 0.524; RSME in the second cluster = 0.530).

4.2. Designing index insurance

Using this information, we can then define a different index insurance for both the entire sample and each cluster separately. This requires the definition of the most appropriate trigger level and payout structure, which allows us to quantify the associated Basis risk (both positive and negative).[11] Fig. 4 plots the observed values of yield against predicted yield in the whole sample, assuming (as an illustration) a value of the trigger at 80% of the average yield (represented by the blue line), allowing us to get a visual idea of the importance of prediction error and its implications in terms of the effectiveness of the contract.

Only households to the left of the vertical line (with predicted yield below the trigger) will receive an indemnity. The black points, in the third quadrant in Fig. 4, represent households that experienced a loss and were correctly paid (corresponding to 39.42% of the households in our sample over the three years), while red points (in the first quadrant) represent those that did not experience a loss and were correctly not paid (33.73% of the sample). An ideal index insurance would have all observations in these two quadrants. The blue points, in the fourth quadrant, represent households that did receive an indemnity even though they did not experience a loss (positive basis risk, affecting 19.00% of the sample), while the green points (in the second quadrant) represent households that did not receive any indemnity even though they experienced a loss. The size of this last group determines the importance of negative Basis risk (7.83% of the sample), on which much attention has been placed.

The definition of the trigger strongly influences the position of each household in each of the four quadrants just defined as well as the structure of payments and consequently, the conclusions about the value of insurance as a poverty reducing policy and the affordability of the insurance contract. In Table 3 we present the level of Basis risk (positive and negative) for different levels of the trigger ranging from 70% to 95% of the average yield, as well as the respective premium to be paid for an actuarially fair insurance. These results allow us to conclude that cluster-specific contracts can be considered improvements on the contract defined for the whole dataset under minimal assumptions regarding farmers' preferences, at least for some values of trigger.

A more detailed analysis shows that, for trigger values above 0.80, an insurance contract tailored to the conditions of cluster 1 always dominates the insurance contract defined on the basis of data for the whole sample, as it is characterized by lower negative basis risk at lower

[10] Results for SVI are available from the corresponding author upon request.
[11] See the discussion in Appendix A.

premium. The choice is less clear for households in cluster 2: while they are now offered a much less expensive insurance contract, this contract also exhibits higher levels of negative basis risk. However, even in this case, accounting for heterogeneity is potentially beneficial, although only for relatively low values of the trigger (0.70), when the insurance contract has similar values of negative basis risk but is potentially much less expensive. Whether such contracts would provide welfare improvements is, ultimately, an empirical question that will depend on decision-makers' utility function, including how they value risk and the losses associated with negative basis risk. Similar considerations prevent us from attempting to define an "optimal" trigger for those households in cluster 1, when considering trigger values above 0.85.

5. Discussion and policy implications

Index insurance has largely been identified as a promising tool to limit the negative effects of weather shocks on household's income. Nevertheless, basis risk can drastically limit index insurance products' uptake rate, undermining their potential to reduce household's vulnerability and to enhance food security (Barnett et al., 2008; Jensen et al., 2018). Several improvements are potentially possible, and although this contribution is not meant as an exhaustive discussion of ongoing work in this area, several paths that are directly related to our own work merit some discussion.

The first is that, although elevation and slope emerged as natural variables on which to aggregate observations into more homogeneous subsets, given what is known about rice production in our context, they are not the only possibilities (and, in our analysis, slope turned out not to be an important determinant of this heterogeneity). Other variables that determine production risk (for example, importance of irrigation at local level) are likely candidates for a similar approach, in a close parallel with ongoing work that links index insurance with the adoption of (risk reducing) technologies, such as drought resistant seeds (see, for example, Lybbert and Carter (2015) or Ward et al. (2020)): clusters are now homogeneous in terms of adoption of technology that has a direct bearing on production risk, and clearly there is no reason to offer such producers a contract that is not designed to account for such fact. Although the focus of our analysis was reductions in basis risk, to the extent that such technologies reduce production risk, the new contracts should also have lower premiums.

A second set of solutions has attempted to improve the link between signal and yield, by improving the signal used in the design of the contract (for example, Miller et al., 2020, Afshar et al. (2021) and De Oto et al. (2019), perhaps coupled with institutional innovations (such as audits as suggested in Flatnes and Carter (2015)), which may allow insurance contracts to be acceptable in the presence of otherwise low quality of the signal.[12]

A second way to improve the quality of the insurance contract is to use signals with a finer spatial resolution. Although the availability of more spatially detailed satellite data may suggest its use as a possible refinement of the approach adopted in our work (essentially attempting to define finer clusters), the use of such data begs the question of "how low should one go?". In answering this question, it is important to keep in mind that, as we get closer to contracts defined at plot/household level, the importance of asymmetric information problems is likely to increase (see Vroege et al. (2021) for a related discussion).

Although an empirical evaluation of changes in the importance of asymmetric information introduced by accounting for landscape heterogeneity is beyond the scope of this paper, we make two remarks. The first is that the new contract relies on the same signal, NDVI measured at village level. Hence, it is unlikely that there would be large differences in the importance of moral hazard (and, potentially, collusion among insurees to manipulate the signal) compared to a situation when the contract is defined over the whole sample.

The second is that the statistical relation between the index and yield is estimated in different subsamples, each reflecting specific production conditions and its own risk profile. Consequently, the scope for adverse selection is now likely to be different from when insurance contracts are defined over the whole sample.

To understand these changes, we follow an approach similar to Jensen et al. (2018) and we quantify changes in the scope for cross-subsidization by estimating changes in yield variability introduced by clustering the data into different landscape categories. The standard deviation of yield is lower in cluster 1 (lower elevation rice producers) and, although this difference is statistically significant, it is relatively small in economic terms (−49.5 kg, or approximately 3% of the standard deviation of yield estimated in the whole sample; bootstrapped 95% CI = [−79; −23]). The opposite is true in cluster 2 (high elevation rice producers), where variability increases compared with the whole sample (+90 kg, or 6% of the standard deviation in the whole sample; bootstrapped 95% CI = [47; 143]).

Clearly, in this last case, there is a trade-off between lowering basis risk and increasing the scope for adverse selection. Nonetheless, as shown in the previous section, the scope for reducing basis risk for this cluster was itself small. On the contrary, "excising" producers from high variability conditions from the definition of insurance contracts targeted at areas with better production conditions makes insurance both more attractive to households (as basis risk is lower) and to insurers (as the scope for adverse selection is potentially lower).

6. Conclusion

In this article we investigate the intuitive suggestion that accounting for heterogeneity in production conditions, through the use of clustering algorithms to split the whole data into more homogeneous subsamples, may improve the design of satellite based agricultural-index insurance while keeping its advantages (low transaction costs and reduction in informational asymmetries). Working with data from a sample of rice producers in northern Laos, we developed a satellite-based index insurance product using yield data at household level and satellite data (NDVI) at village level, using different transformations of this data as potential signal (NDVI, SVI, NDVI controlling for the value of NDVI at the beginning of the growing season), and a variety of statistical models (OLS, Stepwise regression, OLS considering VIF, LASSO, Elastic Net and OLS using Principal Components) estimated in a training sample (80% of the whole sample) and selected on the basis of their predictive power (judged on the base of the lower RMSE between observed and predicted yield values) in the testing sample. Villages were grouped on the basis of elevation and slope, two geographical conditions that plausibly influence rice production in our context. Our analysis shows that heterogeneity in elevation modifies the importance of covariate risk, a result that we explore to understand whether the definition of more homogeneous clusters allows us to substantially improve on the design of a contract defined for the whole sample.

Using the best models, we then computed the associated Basis risk as the percentage of uncovered losses over the insured loss for different trigger levels (ranging from 70% to 95% of the group's yield average). Afterwards, to get an idea of the product affordability, we compute expected indemnity values to be paid according to the different levels of coverage. In this study we assume a fair contract so that the expected premium would be equal to the indemnity paid.

The results show that accounting for landscape heterogeneity greatly improves the design of the proposed contract, allowing for the definition of contracts with lower basis risk and premiums for different values of

[12] Although such audits would likely increase insurance costs (and premia), technological innovations such as picture-based insurance contracts (Ceballos et al., 2019; Hufkens et al., 2019) may work towards increasing its feasibility, as they have the potential to reduce the costs of such audits. In this design, farmers regularly take pictures of their fields allowing them to assess damage in case of divergence between the claimed yield loss and the estimated loss.

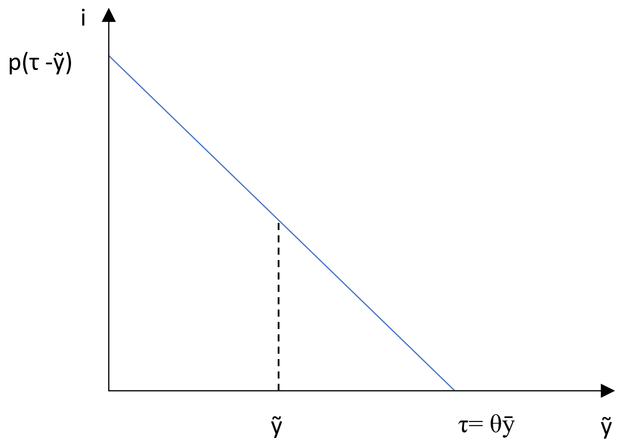

Fig. A1. Indemnity structure.

the insurance trigger. The less encouraging conclusion is that, even with this improvement, the associated costs would likely be too high to be entirely paid by the households. This second result suggests that either technical (e.g. crop masking) or design innovations (e.g. audits) may be further required to improve the effectiveness of insurance products in developing countries.

Funding sources

This study was funded by the Australian Centre for International Agricultural Research (ACIAR), Grant ASEM/2012/073.

CRediT authorship contribution statement

Roberta Rigo: Conceptualization, Software, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing. **Paulo Santos:** Conceptualization, Investigation, Validation, Writing – original draft, Writing – review & editing, Funding acquisition. **Vito Frontuto:** Software, Formal analysis, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Index insurance: a brief summary

The main distinction between conventional and index insurance is that under index insurance payoffs are based on expected losses, predicted by an index correlated with the insured yield, rather than on directly measured losses (Turvey and Mclaurin, 2012; Chantarat et al., 2013; Carter et al., 2014). When the index is objectively measured and cannot be manipulated, the insurance contract ensures that the probability distribution function of the indemnities is independent of farmers' actions (ruling out problems of moral hazard and adverse selection). In addition, payouts can be made automatically and as soon as the index exceeds a pre-determined threshold, avoiding expensive audits and drastically cutting costs and delays in receiving the payments (Carter et al., 2014).

The definition of an index insurance contract can be divided in four steps (Carter et al., 2014). The first, and essential, step is to define a signal that can be related with the outcome to be insured. Data availability clearly shaped how intellectual discussions could be translated to actual financial services, from the earlier development of area-yield insurance to more current attempts to develop satellite-based insurance.

Remotely-sensed index-based insurance, which uses satellite data of vegetation status used to predict yield can be seen as a next step in the direction of moving closer to the location of losses. The biggest advantage of this type of data is their global availability, with a good time and spatial resolutions, overcoming the lack of ground climate data. For example, the MODIS data that we use are available in almost real time, at a spatial resolution of 250 m and relatively high frequency (every 16 days) and can be accessed at low cost.

The Normalized Difference Vegetation Index (NDVI), which reflects photosynthetic activity (which interferes with the red energy reflected), is one of the most widely trialed indexes.[13] Peters et al. (2002) showed how the NDVI can be a good estimator of the vegetation response to climatic shocks so that this kind of data can be potentially used with good results to predict yield, naturally suggesting its use in the definition of index insurance.

Having decided on an index, the second step is to determine the underlying statistical relation between signal and yield: if there is no evidence of correlation between the two variables, the probability that the insurance contract will not pay for a loss when it does occur would be high (negative basis risk) or the index would predict a loss that didn't occur (positive basis risk). Even though policy holders can see this false positive as a favorable index's characteristic, this failure leads to a lower quality of the insurance product and higher premiums (Carter, 2009).

The third step is to determine the payoff structure in order to link each index level with the corresponding indemnity. This includes the definition of the trigger value, from which the payout structure, premium and associated basis risk then follow.

The trigger (τ) is defined as the value of the index below which households will receive an indemnity (i), and is usually defined as a percentage θ of the observed average yield. With indemnities expressed as a linear function of yield (Martin et al., 2001; Vedenov and Barnett, 2004; Collier et al., 2009) we obtain:

$$i_{(\widetilde{y},\tau)} = \{0 \;\; if \;\; \widetilde{y} > \tau p(\tau - \widetilde{y}) \;\; if \;\; \widetilde{y} \leq \tau \tag{A.1}$$

where $\tilde{y}$ is the index value, τ is the trigger, and p is the price of the insured commodity. This relation is represented in Fig. A1.

This definition of indemnity will ensure that every policyholder will receive an income that is at least equal to the one s/he would have got if his yield would have been equal to the trigger. Selecting a proper trigger is crucial since a higher trigger, although it would lead to a lower Basis risk, would come at the cost of higher premiums. The trigger then reflects a compromise between insurance cost and basis risk, and in the analysis, we will

[13] Higher levels of photosynthetic activity lead to less red energy being reflected (given that the high level of photosynthetic activity leads to the absorption of radiation in those wave lengths), while the amount of near-infrared energy reflected will increase (Huete et al., 1999; Chantarat et al., 2013; Turvey and Mclaurin, 2012). Formally, is then defined as:NDVI=NIR + RedNIR – Red (3)leading to an unitless index ranging from -1 to 1 calculated using vegetation spectral reflectance. A high positive value indicates a healthy vegetation, while a high negative value indicates a poor vegetation. Tipically the lowest NDVI values are around zero (and they are normally associated with snow, ice, baresoil and water), while an healty vegetation shows values around 0.8.

Table B1
Estimate of linear production function.

Variables	(1) Harvest	(2) Harvest	(3) Harvest
Area (ha)	1,627***	1,677***	1,594***
	(280.7)	(285.2)	(255.1)
Access wet season (0/1)	38.23	−11.65	22.03
	(115.7)	(121.4)	(131.9)
Seasonal road (0/1)	84.73	80.80	87.31
	(100.5)	(101.7)	(123.0)
Permanent road (0/1)	−76.23	−12.52	−99.53
	(143.0)	(141.7)	(190.6)
Plot borders river	−55.73	−0.919	−10.53
	(105.7)	(105.8)	(117.4)
Plot borders forest	−30.13	10.94	45.48
	(103.0)	(101.4)	(113.4)
Distance (km)	−23.02	−13.45	−13.26
	(23.28)	(23.14)	(26.00)
Soil depth (>50 cm)	−236.6*	−285.6**	−278.3**
	(122.4)	(120.0)	(138.8)
Soil depth (25–50 cm)	−159.7	−169.0	−193.5
	(99.20)	(103.3)	(117.2)
Stony soil	−168.4	−172.2	−76.58
	(133.7)	(127.7)	(132.8)
Slope - moderate	298.3***	270.3***	310.2***
	(95.41)	(93.67)	(106.4)
Slope - flat	397.7***	394.2***	473.1***
	(144.8)	(148.1)	(169.0)
Upland rice	−339.4	278.6	−182.2
	(765.5)	(500.3)	(427.1)
Seed (kg)	6.008*	5.438*	5.806*
	(3.149)	(3.136)	(3.016)
Transplant (0/1)	−276.7	−65.29	7.383
	(316.3)	(275.7)	(281.2)
Continuous cultivation	−1,187***	−989.7**	−1,152**
	(445.2)	(375.4)	(454.3)
Fallow (wet season 2017)	408.7	369.9	559.4*
	(296.6)	(277.3)	(320.5)
Fertilizer (0/1)	495.0***	437.7**	503.1***
	(168.3)	(169.3)	(187.9)
Grazing (0/1)	222.8**	228.6**	362.2***
	(94.38)	(96.93)	(108.2)
Manure (0/1)	371.2*	365.0**	331.3
	(193.3)	(178.6)	(199.0)
Irrigated (0/1)	30.85	−59.61	−85.89
	(182.8)	(194.5)	(254.8)
Slash (0/1)	−317.3	−292.2	−513.6***
	(205.1)	(175.6)	(181.1)
Burn (0/1)	−12.18	−150.0	−34.58
	(374.4)	(343.2)	(470.3)
Weeded on time (0/1)	−150.2	−114.2	−137.7
	(111.8)	(110.6)	(123.1)
Land preparation on time (0/1)	−349.3	−998.1***	−541.1**
	(598.1)	(335.6)	(249.4)
Weed times	−35.86	−28.91	−14.31
	(78.41)	(79.45)	(95.44)
Weed	115.8	148.5	187.8
	(224.4)	(216.2)	(243.4)
Weeding on time (0/1)	223.4*	265.5**	259.2**
	(112.2)	(110.9)	(125.3)
Rodent Control (0/1)	108.2	72.35	1.963
	(114.6)	(116.4)	(128.3)
Two-wheel tractor (0/1)	254.7*	219.1	211.7
	(136.9)	(146.1)	(175.9)
Disease control (0/1)	−99.82	−78.38	−86.95
	(123.8)	(119.7)	(145.9)
Rodent losses	−51.88	−139.6	−223.6
	(152.8)	(148.1)	(139.0)
Flooded plot	−408.0***	−406.3***	−545.8***
	(126.6)	(131.3)	(146.6)
Not enough water	−103.2	−117.8	−338.8*
	(187.9)	(192.5)	(194.4)
Constant	3,152*	3,603**	3,641**
	(1,581)	(1,434)	(1,744)
Observations	700	676	538
R-squared	0.661	0.668	0.682

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

consider values ranging between 70% and 90% of the average yield, as used in previous studies (Elabed and Carter, 2014; Elabed and Carter, 2015; Flatnes and Carter, 2015).

With this information, we can then compute insurance premiums, which allow us to discuss insurance affordability. Assuming an actuarially fair contract, so that the expected premium would be equal to the indemnity paid, the premium to be paid to insure one hectare would be equal to:

$$E(i) = p^* \int_0^{\tau} (\tau - \tilde{y}) h(\tilde{y}) d\tilde{y} \tag{A.2}$$

where *h(y)* is the probability of obtaining the estimated yield.

Finally, for each value of trigger (expressed in percentage of average yield), we can quantify negative basis risk as the percentage of uncovered loss over the insured loss:

$$negative\ \ basis\ \ risk = \frac{\sum_q (\tau - y) I}{\sum_q (\tau - y) F} \tag{A.3}$$

while positive basis risk as the percentage of the reimbursed non-losses over the non-losses

$$positive\ \ basis\ \ risk = \frac{\sum_q (\tau - y) G}{\sum_q (\tau - y) H} \tag{A.4}$$

where q is quantity, I is an indicator function that takes value 1 if $y < \tau$ and $\tilde{y} > \tau$ (and *0* otherwise); F is an indicator that takes value 1 if $y < \tau$ (and 0 otherwise); G is an indicator function that takes value 1 if $y > \tau$ and $\tilde{y} < \tau$ (and *0* otherwise); H is an indicator that takes value 1 if $y > \tau$ (and 0 otherwise). Clearly, if everyone who experienced a loss also received a corresponding indemnity, the numerator of equation A.4 (and hence the negative basis risk) would be equal to zero.

Appendix B. Quantifying the importance of measurement error

As mentioned, yield data was self-reported raising concerns that measurement error may be a major determinant of Basis risk (ie, of low correlation between declared rice yield and an objective measure of vegetation status which, as argued, would mostly reflect rice production during wet season). In order to investigate how important such measurement error can be, we use data from the 2018 wet season, for which we collected detailed data on rice production technology, to estimate a linear production function that relates harvested quantity of rice (Y, in kg) with inputs (land A, labor L and capital K) and technological choices (subsumed under X: use of fertilizer, rotation & fallow, weeding, as well as plot characteristics).

$$\mathrm{Y} = \mathrm{a} + \mathrm{bA} + \mathrm{cL} + \mathrm{dK} + \mathrm{eX} \tag{B.1}$$

The results are presented in Table B1, for the entire sample (column 1), and when we exclude outliers, as defined in the text (column 2) and when we restrict the analysis to the trial sample (column 3).

These results suggest two comments. The first is that there are no surprising results in terms of individual coefficients, which are plausible in terms of magnitude and statistical significance given what is known about rice production in the region. The second is the relatively high R2 (>0.66). Taken together, they suggest that self-reported values of harvested rice are a plausible approximation of the true values of rice production.

Using the estimated parameters of Table B1 we used predicted yields to estimate basis risk and premiums in the absence of measurement error. Table B2 reports these results on the whole sample computed, contrasting observed and predicted yields. The results are quite similar, with basis risks slightly higher and premiums lower when using predicted yields, suggesting that, at least in this application, measurement

Table B2
Basis risk and premiums for Observed and predicted yields

	Observed Yields			Predicted Yields		
Trigger	Negative Basis risk (%)	Positive Basis risk (%)	Premium (% average yield)	Negative Basis risk (%)	Positive Basis risk (%)	Premium (% average yield)
0.70	0.274	0.125	0.059	0.302	0.118	0.043
0.75	0.142	0.137	0.082	0.251	0.156	0.061
0.80	0.107	0.215	0.110	0.202	0.180	0.084
0.85	0.104	0.221	0.138	0.171	0.211	0.108
0.90	0.099	0.250	0.169	0.144	0.320	0.136
0.95	0.074	0.260	0.201	0.088	0.374	0.166

Table C1

	All			Cluster 1			Cluster 2		
Trigger	Negative Basis risk (%)	Positive Basis risk (%)	Premium (% average yield)	Negative Basis risk (%)	Positive Basis risk (%)	Premium (% average yield)	Negative Basis risk (%)	Positive Basis risk (%)	Premium (% average yield)
0.70	0.226	0.099	0.055	0.143	0.060	0.088	0.386	0.128	0.036
0.75	0.181	0.133	0.077	0.078	0.093	0.110	0.385	0.117	0.052
0.80	0.143	0.183	0.101	0.071	0.096	0.133	0.354	0.127	0.069
0.85	0.112	0.204	0.129	0.071	0.097	0.157	0.209	0.180	0.091
0.90	0.106	0.229	0.157	0.059	0.091	0.181	0.212	0.166	0.116
0.95	0.082	0.234	0.188	0.048	0.165	0.207	0.114	0.202	0.143
N	1737			988			749		

error does not affect our conclusions.

Appendix C. Additional tables and figures

As mentioned in the main text, we eliminate from the sample the observations reporting zero yields or above 1.5 times the interquartile range. Table C1 reports the results with the inclusion of all the data, but the observations reporting zero yields.

In general, the results confirm that taking into account heterogeneity, by clustering the data, helps to define the index insurance. Nevertheless, we notice that in the case of cluster 2 the basis risks are higher with respect to the ones of the whole sample. This outcome is associated with a lower predictive capability of the models applied to Cluster 2 (lower R2). In these cases it may not be easy to define the index insurance and further investigations are needed.

References

Afshar, M.H., Foster, T., Higginbottom, T.P., Parkes, B., Hufkens, K., Mansabdar, S., Ceballos, F., et al., 2021. Improving the Performance of Index Insurance Using Crop Models and Phenological Monitoring. Remote Sens., MDPI AG 13 (5), 924.

Aguilera, A.M., Escabias, M., Valderrama, M.J., 2006. Using principal components for estimating logistic regression with high-dimensional multicollinear data. Comput. Stat. Data Anal. 50 (8), 1905–1924.

Barnett, B.J., Barrett, C.B., Skees, J.R., 2008. Poverty Traps and Index-Based Risk Transfer Products. *World Dev.*, Elsevier 36 (10), 1766–1785.

Basnayake, J., Fukai, S., Sipaseuth, Schiller, J., Monthathip, C., 2006, Advances in agronomic research in rainfed lowland rice environments of Lao PDR. In: Schiller, J., Chanphengxay, M., Linquist, B., Appa, S. (Eds.), *Rice in Laos*, International Rice Research Institute, Los Baños.

Carter, M., de Janvry, A., Sadoulet, E., Sarris, A., 2017. Index Insurance for Developing Country Agriculture: A Reassessment. Annu. Rev. Resour. Econ. 9 (1), 421–438.

Carter, M.R., de Janvry, A., Sadoulet, E., Alexandros, S., 2014. "Index-based weather insurance for developing countries: A review of evidence and a set of propositions for up-scaling", Working Papers n. 111, FERDI.

Carter, M.R., 2009. Intelligent design of index insurance for smallholder farmers and pastoralists. In: Vargas-Hill, R., Torrero, M. (Eds.), *Innovations in Insuring the Poor*, Food Policy Research Institute, Washington, D.C.

Ceballos, F., Kramer, B., Robles, M., 2019. The feasibility of picture-based insurance (PBI): Smartphone pictures for affordable crop insurance. Dev. Eng. 4, 100042. https://doi.org/10.1016/j.deveng.2019.100042.

Chantarat, S., Mude, A.G., Barrett, C.B., Carter, M.R., 2013. Designing Index-Based Livestock Insurance for Managing Asset Risk in Northern Kenya. J. Risk Insurance, John Wiley & Sons Ltd 80 (1), 205–237.

Charrad, M., Ghazzali, N., Boiteau, V., Niknafs, A., 2014. Nbclust: An R package for determining the relevant number of clusters in a data set. *J. Stat. Softw.*, American Statistical Association 61 (6), 1–36.

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

Clement, K.Y., Wouter Botzen, W.J., Brouwer, R., Aerts, J.C.J.H., 2018. A global review of the impact of basis risk on the functioning of and demand for index insurance. *Int. J. Disaster Risk Reduct.*, Elsevier Ltd 28, 845–853.

Cole, S.A., Xiong, W., 2017. Agricultural Insurance and Economic Development. Annu. Rev. Econ. 9 (1), 235–262.

Collier, B., Skees, J., Barnett, B., 2009. Weather Index Insurance and Climate Change: Opportunities and Challenges in Lower Income Countries. Geneva Papers Risk Insurance - Issues Practice 34 (3), 401–424.

De Leeuw, J., Vrieling, A., Shee, A., Atzberger, C., Hadgu, K.M., Biradar, C.M., Keah, H., Turvey, C., 2014. The potential and uptake of remote sensing in insurance: A review. Remote Sens. 6 (11), 10888–10912.

De Oto, L., Vrieling, A., Fava, F., de Bie, K.C., 2019. Exploring improvements to the design of an operational seasonal forage scarcity index from NDVI time series for livestock insurance in East Africa. Int. J. Appl. Earth Observ. Geoinform. 82, 101885.

Dercon, S., Christiaensen, L., 2011. Consumption risk, technology adoption and poverty traps: Evidence from Ethiopia. *J. Dev. Econ.*, Elsevier 96 (2), 159–173.

Elabed, G., Carter, M.R., 2014. Basis Risk and Compound-risk Aversion: Evidence from a WTP Experiment in Mali. University of California, Davis.

Elabed, G., Carter, M.R., 2015. Compound-risk aversion, ambiguity and the willingness to pay for microinsurance. J. Econ. Behav. Org. 118, 150–166.

Eliste, P., Santos, N., 2012. *Lao People's Democratic Republic Rice Policy Study*.

Field, D., Odgers, N., 2016. *Soil Capability and Land Suitability, Food Security in the Northern Uplands Discussion Paper 1*, Melbourne.

Flatnes, J.E., Carter, M.R., 2015. *Fail-Safe Index Insurance without the Cost: A Satellite Based Conditional Audit Approach.* Working Paper, University of California, Davis.

Goto, K., Douangngeune, B., 2017. Agricultural Modernization and Rural Livelihoods in Rice Farming Villages of Laos. Can. J. Dev. Stud. 38 (4), 467–486.

Harrison, G., Martínez-Correa, J., Morsink, K., Min Ng, J., Swarthout, T., 2020. Compound Risk and the Welfare Consequences of Insurance. Working paper 2020-10, Georgia State University, Center for the economic analysis of risk.

Hastie, T., Tibshirani, R., Wainwright, M., 2015. *Statistical Learning with Sparsity: The Lasso and Generalizations*, edited by CRC Press, Boca Raton.

Hazell, P.B.R., 1992. The appropriate role of agricultural insurance in developing countries, *J. Int. Dev.*, John Wiley & Sons, Ltd, vol. 4, no. 6, pp. 567–581.

Hoddinott, J., Kinsey, B., 2001, Child Growth in the Time of Drought. *Oxford Bulletin of Economics and Statistics*, John Wiley & Sons, Ltd, Vol. 63 No. 4, pp. 409–436.

Hufkens, K., Melaas, E.K., Mann, M.L., Foster, T., Ceballos, F., Robles, M., Kramer, B., 2019. Monitoring crop phenology using a smartphone based near-surface remote sensing approach. *Agric. Forest Meteorol.* 265, 327-337.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. *An Introduction to Statistical Learning with Applications in R*, Springer, New York.

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

Ligon, E., Sadoulet, E., 2018. Estimating the Relative Benefits of Agricultural Growth on the Distribution of Expenditures. World Dev. 109, 417–428.

Liquist, B.A., Keoboualapha, B., Sipaseuth, Inthapanya, P., 2006. Rice production systems of Laos. In: Schiller, J.M., Chanphengxay, M.B., Linquist, B., Appa, S. (Eds.), *Rice in Laos*, International Rice Research Institute, Los Baños.

Lybbert, T.J., Carter, M.R., 2015. Bundling drought tolerance and index insurance to reduce rural household vulnerability. In: Balisacan, A., C.U. and R.M. (Eds.), *Sustainable Economic Development: Resources, Environment, and Institutions*, Academic Press, Oxford, pp. 401–414.

Lobell, D.B., Azzari, G., Burke, M., Gourlay, S., Jin, Z., Kilic, T., Murray, S., 2018. Eyes in the Sky, Boots on the Ground: Assessing Satellite and Ground-Based Approaches to Crop Yield Measurement and Analysis in Uganda (2018). Working Paper No. 8374, World Bank Policy Research.

Martin, S.W., Barnett, B.J., Coble, K.H., 2001. Developing and pricing precipitation insurance. J. Agric. Resource Econ. 26 (1), 261–274.

Miller, S.E., Adams, E.C., Markert, K.N., Ndungu, L., Ellenburg, W.L., Anderson, E.R., Kyuma, R., Limaye, A., Griffin, R., Irwin, D., 2020. Assessment of a spatially and temporally consistent MODIS derived NDVI product for application in index-based drought insurance. Remote Sens, 12 (18), 3031.

Mobarak, A.M., Rosenzweig, M.R., 2013. Informal risk sharing, index insurance, and risk taking in developing countries. Am. Econ. Rev. 103 (3), 375–380.

Morduch, J., 1995. Income Smoothing and Consumption Smoothing. *J. Econ. Perspect.*, American Economic Association 9 (3), 103–114.

Moschini, G., Hennessy, D.A., 2001. Uncertainty, risk aversion, and risk management for agricultural producers. *Handbook of Agricultural Economics*, Elsevier, Vol. 1, Part 1, pp. 88–153.

Mosleh, M.K., Hassan, Q.K., Chowdhury, E.H., 2015. Application of remote sensors in mapping rice area and forecasting its production: A review. Sensors 15 (1), 769–791.

Murtagh, F., Legendre, P., 2014. Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? J. Classif. 31 (3), 274–295.

Newby, J.C., Manivong, V., Cramb, R.A., 2013. Intensification of lowland rice-based farming systems in Laos in the context of diversified rural livelihoods. *57th AARES Annual Conference 5–8 February 2013*, Sydney, Australia.

O'brien, R.M., 2007. A caution regarding rules of thumb for variance inflation factors. *Qual. Quant.*, Springer 41 (5), 673–690.

Peters, A., Walter-Shea, E., Ji, L., Viña, A., Hayes, M., Svoboda, M., 2002. Drought Monitoring with NDVI-Based Standardized Vegetation Index. Photogramm. Eng. Remote Sens. 68 (2), 71–75.

Ravallion, M., Datt, G., 2002. Why has economic growth been more pro-poor in some states of India than others? *J. Dev. Econ.*, Elsevier 68 (2), 381–400.

Riley, E., 2018. Mobile money and risk sharing against village shocks. *J. Dev. Econ.* 135, 43-58.

Sacklokham, S., 2014. Rice-based farming systems in Lao PDR— opportunities and challenges for food security. *A Policy Dialogue on Rice Futures: Rice-Based Farming Systems Research in the Mekong Region*, Phnom Penh, Cambodia.

Schiller, J., Hatsadong, A., Doungsila, K., 2006. A history of rice in Laos. In: Schiller, J. M., Chanphengxay, M.B., Linquist, B., R.S.A. (Ed.), *Rice in Laos*, International Rice Research Institute, Los Baños, pp. 9–28.

Smith, V.H., Glauber, J.W., 2012. Agricultural Insurance in Developed Countries: Where Have We Been and Where Are We Going? Appl. Econ. Perspect. Policy 34 (3), 363–390.

Townsend, R.M., 1994. Risk and Insurance in Village India. Econometrica 62 (3), 539.

Turvey, C.G., McLaurin, M.K., 2012. Applicability of the normalized difference vegetation index (NDVI) In index-based crop insurance design. *Weather Clim. Soc.*, American Meteorological Society 4 (4), 271–284.

Vedenov, D.V., Barnett, B.J., 2004. Efficiency of weather derivatives as primary crop insurance instruments. J. Agric. Resource Econ. 29 (3), 387–403.

Vroege, W., Vrieling, A., Finger, R., 2021. Satellite support to insure farmers against extreme droughts. Nature Food 2 (4), 215–217.

Ward, P.S., Makhija, S., Spielman, D.J., 2020. Drought-tolerant rice, weather index insurance, and comprehensive risk management for smallholders: evidence from a multi-year field experiment in India. Aust. J. Agric. Resour. Econ. 64 (2), 421–454.

World Bank, 2018. *Multi-Sectoral Support To The Nutrition Convergence Agenda In Lao PDR*, Washington, D.C.

Yoo, W., Mayberry, R., Bae, S., Singh, K., He, Q.P., Lillard, J.W., 2014. A Study of Effects of MultiCollinearity in the Multivariable Analysis. Int. J. Appl. Sci. Technol. 4 (5), 9–19.

Zimmerman, F.J., Carter, M.R., 2003. Asset smoothing, consumption smoothing and the reproduction of inequality under risk and subsistence constraints. *J. Dev. Econ.*, Elsevier 71 (2), 233–260.