Tail-dependent Weather Risk and Demand for Index based Crop Insurance

Digvijay S Negi *and Bharat Ramaswami †

September 7, 2017

Abstract

For a variety of reasons, agricultural insurance programs use losses against an index (rainfall, area yield) rather than losses against individual yields to make payouts. While this facilitates the supply of insurance, the resulting basis risk reduces the value of insurance and therefore reduces demand for it. Using district crop yields and rainfall data for India, we find that the association between crop yields and rainfall index is characterized by the statistical property of 'tail-dependence'. This implies that the associations between yield losses and index losses are stronger for large deviations than for small deviations. Or, basis risk is least for large deviations of the index. Using simulation we show that value to a risk averse farmer of index-based insurance relative to actuarial cost is highest for insurance against extreme or catastrophic losses (of the index) than for insurance against all losses.

*Senior Research Fellow, Indian Statistical Institute, 7 SJS sansanwal Marg, New Delhi 110016, India. digvijaynegi12@gmail.com

†Professor, Indian Statistical Institute, 7 SJS sansanwal Marg, New Delhi 110016, India. isid.bharat@gmail.com

1 Introduction

Agriculture and agriculture-based livelihoods in developing countries are highly prone to weather shocks. Although there exist various informal mechanisms [1] in rural communities that allow farmers to pool their idiosyncratic risks, these provide limited insurance to individual households when risks are correlated and widespread. Extreme climate events such as droughts, floods and heat waves which affect farming communities in a region simultaneously, significantly limit the degree of insurance provided by the informal risk sharing mechanisms. There is substantial evidence that these correlated shocks adversely affect crop yields and agricultural production, cause depletion of productive assets, exacerbate rural poverty, force outmigration, and reduce demand for non-agricultural goods (Rosenzweig and Binswanger 1993; Carter and Barrett 2006; Dercon and Christiaensen 2011).

In recent years, there has been a lot of interest in index based crop insurance as an instrument to manage agricultural risks of farmers in developing countries. Unlike the farm insurance contracts, index based insurance payouts are tied to the observed value of the index rather than individual farm crop losses. The index based insurance products specify a threshold establishing a range of values over which indemnity payments are to be made. All farmers covered under the contract are indemnified if the index falls below the pre-determined strike point and individual farmers have little or no influence on payouts. Therefore, the index-based insurance products are less likely to fail due to asymmetry in information between the insurer and the insured. Index insurance based on average yield, cumulative rainfall, and other such variables are claimed to be cost-effective risk management tool in highly fragmented rural communities in developing countries.

The evidence, however, shows unexpectedly poor uptake of the index based insurance products, despite considerable efforts to promote it (for example see Cole et al. 2010). An important feature of the index insurance is that the probability of non-performance of contract is positive. Payouts are based on the index, and as long as there is less than perfect correlation between individual loss and the index, it is possible for a policyholder

[1] For examples of informal risk sharing mechanism in rural communities see Rzosenzweig and Stark (1987); Townsend (1994); Fafchamps and Lund (2003) and DeWeerdt and Dercon (2006).

to experience loss and yet receive no indemnity or vice versa. This is termed as basis risk of the index insurance. Hence, for a policyholder the index based insurance is more of a lottery than insurance. The benefits of index insurance essentially depend on the strength of association between individual loss and index or alternatively the underlying basis risk. It is well-known that if basis risk is large then insurance does not benefit much to the farmers, leading to lower willingness to purchase insurance products. Clarke (2016) theoretically shows that a positive probability of contract non-performance inherent in an index insurance contract substantially suppresses its demand. Using the distance of a farm from the nearest weather station as a proxy for basis risk, Mobarak and Rosenzweig (2013) find higher basis risk adversely affecting the demand for insurance products by Indian farmers. Gine, Townsend and Vickery (2008) while studying rainfall insurance for India find that the farmers who traditionally allocate a high proportion of their land to crops for which index insurance contracts were designed, have a higher probability of purchasing these products.

Besides basis risk, a number of behavioral factors have been found to dampen the demand for index insurance. Lack of trust, poor understating of the product and liquidity constraint (Cole et al. 2013), risk aversion and perception of index insurance as compound lottery (Elabed and Carter 2015) are identified as important factors limiting the demand for index based insurance products. Karlan et al. (2015) from their study in Ghana find that farmers are more likely to purchase insurance if they or other farmers in their social network had received payouts previously. Similar findings have been reported from India (Hill, Robles, and Ceballos 2016; Cole et al. 2014 and Mobarak and Rosenzweig 2013).

The success of an index insurance contract critically hinges on the nature of association between the index and farmers' yields. Traditionally, basis risk has been linked to linear correlation between yields and index. However, stochastic dependence between variables goes beyond linear correlation, and there is no particular reason to assume a linear relationship between farm yields and index. Global weather phenomenon like El Nino can simultaneously affect local weather at multiple locations and generate spatial correlation between them. Literature from hydrology has provided evidence of non-linear association between rainfall at different locations (Kuhn et al. 2007; Liu and Miranda 2010 and Aghakouchak et al. 2010). The particular pattern of association is as follows.

Consider two locations A and B, the evidence indicates that for small deviations there is a weak association between the rainfall deviations at A and B. But, the association is stronger if the deviations are large at both the locations. This pattern of association is called tail-dependence in the joint distribution of rainfalls, i.e. if there is a large deviation at A, then it likely that the deviation in rainfall at B will be also be large.

Suppose crop yield at a farm at location A or B is a function of rainfall at that farm and some other idiosyncratic factors, then tail-dependence in the joint distribution of rainfall implies tail-dependence in the joint distribution of farm yield at location A and B. this implies that average yield of farms at A and B is unlikely to change much if there are small rainfall deviations. But it will change if deviations are large and spatially correlated. Goodwin (2001) has shown that spatial correlation of crop yields is significantly stronger during the years of extreme deviations than during normal years. A second hypothesis is that the association between individual farm yield and average yield will also exhibit tail-dependence. Further, the association between average rainfall and average yield will also exhibit tail-dependence.

We begin the analysis by testing for presence of tail-dependence between rainfalls at two locations in India. This is a primitive requirement for the joint distribution of farm yields and the joint distribution of average yield and average rainfall to exhibit tail-dependence. Then, we examine the tail-dependence between average yields of major crops and average seasonal rainfall. Finally, we analyze implications of these results on farmers willingness to pay, design of contract and viability of the index insurance as an instrument to manage farm risk.

A preview of the findings is as follows. We find that station level rainfall in India do exhibit tail-dependence and the joint distribution of district level crop yields for nine major crops and rainfall index also exhibit tail-dependence. This implies that the associations between yield losses and index losses are stronger for large deviations than for small deviations. Or, basis risk is least for large deviations of the index. Using simulation we show that value to a risk averse farmer of index-based insurance relative to actuarial cost is highest for insurance against extreme or catastrophic losses (of the index) than for insurance against all losses.

The paper is organized as follows. The next section discusses theoretically a typical index insurance contract and the willingness to pay by risk averse farmers. Section 3 describes the data and the empirical methodology used to assess the association between yields and rainfall index. Section 4 presents the results on nature of the joint association between crop yields and rainfall index. In section 5 we simulate the impact of tail-dependence in the joint distribution of yields and rainfall index on the willingness to pay for index insurance and discuss its implication for the design and viability of the index insurance contracts. Conclusions are presented in the last section.

2 Conceptual Framework

A typical index based insurance contract triggers payouts based on an index according to following schedule.

$$I(R|R^*) = \bar{Y} \times Max\left\{\frac{(R^* - R)}{R^*}, 0\right\} \tag{1}$$

Where Y and R are yield and rainfall index respectively, and $\bar{Y}$ is the maximum payout. The contract pays out when R falls below the specified trigger R^*, in proportion to the difference between the R and R^*. We set up a simple framework to illustrate how the nature of dependence between yield and rainfall influences a farmer's demand for insurance. We assume that the expected utility of representative farmer is characterized by a constant relative risk aversion (CRRA) utility function.

$$U(\pi) = \frac{\pi^{(1-\gamma)}}{(1-\gamma)} \tag{2}$$

Where π denotes income and γ is the measure of relative risk aversion. For brevity we assume that the farmer cultivates only one crop and the indemnity is paid in terms of crop output. This will not influence indemnity and premiums. The actuarially-fair premium of the contract is the expected value of the indemnity.

$$P^{AF}(R|R^*) = E[I(R|R^*)] = \int I(R|R^*)h(R)dR \tag{3}$$

Where $h(R)$ is the density function of the rainfall index. Given this as the contract offered by the insurer, we wish to know farmer's willingness to pay for insurance. Willingness to pay W is the amount which a farmer is willing to pay for insurance such that the expected utility of net income with insurance is least equal to income without insurance.

$$E[U(Y)] \leq E[U(Y + I(R; R^*) - W)] \tag{4}$$

Where $U(.)$ is the utility function of the farmer with $U^{'}(.) > 0$ and $U^{''}(.) < 0$ and $g(Y)$ is the probability density function of yield. Willingness to pay can also be expressed as the difference between the certainty equivalent with the index insurance contract, CI_I, and the certainty equivalent without insurance, CI_U.

$$WTP = CI_I - CI_U = U^{-1}(E[U(Y + I(R; R^*))]) - U^{-1}(E[U(Y)]) \tag{5}$$

To see how dependence between yield and index influences the demand for insurance we write the above expression as

$$\begin{aligned} WTP = \left(\int \left(\int (Y + I(R; R^*)\right)^{(1-\gamma)} f(Y|R)dY\right) h(R)dR\right)^{1/(1-\gamma)} \\ - \int \left(Y^{(1-\gamma)} g(Y)dY\right)^{1/(1-\gamma)} \end{aligned} \tag{6}$$

Where $f(Y|R)$ is the probability of yield conditional on rainfall. The above expression shows that besides other factors, willingness to pay is also a function of the joint distribution of yield and rainfall index. Hence, the degree of association between yield and rainfall will directly affect the willingness to pay for insurance.

3 Data and empirical methodology

3.1 Data

To test the hypothesis that rainfall at two locations is tail-dependent we use rainfall data from 137 weather stations of the Indian Meteorological Department. The complete data series is available from 1966 to 2007. Rainfall is highly seasonal, and bulk of it is received during June to October. To make rainfall series comparable across stations and months, we standardize rainfall by months.

To test the hypothesis that average district yield and district rainfall shows tail-dependence we utilize data on crop yield from the International Crops Research Institute for the Semi-Arid Tropics ICRISAT (`http://vdsa.icrisat.ac.in/vdsa-database.htm`) compiled from various official sources. The database covers 15 major crops across 311 districts in 19 states from the year 1966-67 to 2011-12. To maintain consistency and comparability of time series across districts, data of the bifurcated districts is returned to the parent district based on the district boundaries in 1966. To generate district level average rainfall we use high resolution gridded rainfall data from the Indian Meteorological Department.

The crops chosen for the analysis are Maize, Cotton, Sorghum, Finger millet, Pigeon pea, Soybean, Pearl millet, Groundnut and Rice. These crops are cultivated in the kharif season (June to October). This is one of the main agricultural season in India and receives around 85% of the annual rainfall. Crop yields typically exhibit significant upward trends overtime due to technological changes. Yield deviations are estimated by fitting a linear trend to log yields of each crop of each district. Similarly district specific cumulative monthly rainfall are transformed to standardized deviations from their long term normals.

3.2 Empirical methods

Our interest in this paper is to examine the nature of association between rainfall and crop yields. We use statistical tools that are flexible enough to allow for complex nonlinear dependence between these variables. Traditionally, the crop insurance literature has used linear correlation to assess basis risk. Using linear correlation is restrictive as it captures only linear dependence with the underlying assumption that the variables are jointly normally distributed. This is a rigid assumption, especially when both rainfall and yield are known to have highly skewed marginal distributions.

Index insurance indemnifies insurers only when there are extreme deviations in the index. Thus, from the point of view of index insurance the nature of dependence at ends or tails of the joint distribution is more important. A statistical concept that measures the joint dependence between two variables at tails of their joint distribution is the coefficient of tail-dependence. The coefficient of tail-dependence measures the strength of dependence at the lower or upper tail of the joint distribution (Joe 2015; Nelsen 2006).

Let X and Y be the continuous random variables with distribution functions F and G, respectively. Then, the lower tail-dependence coefficient, λ^L, is the probability that one variable takes an extremely low value, given that the other variable also takes an extremely low value. Similarly, the upper tail-dependence coefficient, λ^U, is the probability that one variable takes an extremely high value, given that the other variable also takes an extremely high value. Mathematically, these can be expressed as:

$$\lambda^L = \lim_{q \to 0} P(G(Y) \leq q | F(X) \leq q) \tag{7}$$

$$\lambda^U = \lim_{q \to 0} P(G(Y) > q | F(X) > q) \tag{8}$$

Where both $\lambda^L, \lambda^U \in (0, 1]$. For a set of random variables to be tail-dependent the limits of the conditional probabilities in equation (7) and (8) should be non zero. Tail-dependence coefficients are better measures than linear correlation as they provide more detailed information on the joint dependence structure of random variables (Patton, 2013). Since a bivariate normal distribution does not exhibit tail-dependence, the presence of tail-dependence in data goes against the assumption of joint normality. We use nonparametric estimator of tail-dependence as suggested by Frahm, et al. (2005) and Patton (2013) to estimate the tail-dependence in weather station level rainfall. The estimator is given as:

$$\hat{\lambda}^L = \frac{2 - \log\left(1 - 2(1-q) + T^{-1}\sum_{t=1}^{T} 1\{G(Y) \leq 1-q, F(X) \leq 1-q\}\right)}{\log\left(1-q\right)} \tag{9}$$

$$\hat{\lambda}^U = \frac{2 - \log\left(T^{-1}\sum_{t=1}^{T} 1\{G(Y) \leq 1-q, F(X) \leq 1-q\}\right)}{\log\left(1-q\right)} \tag{10}$$

The tail-dependence statistic looks at a specific portion of tail in the joint distribution. Therefore, a threshold q needs to be specified for estimation. This choice of q involves trade off in terms of bias in the estimate and its variance. For small (large) values of q the variance is large (small) and the bias is small (large). Note that the smaller the value of threshold q the more extreme deviations the tail dependence statistic will capture.

A crude test for the presence of tail-dependence in a pair of variables is to examine the scatter plot of these variables (after transforming to uniform scores based on the empirical distribution) for clustering at the extremes (Joe 2015). For different values of q we can also compute conditional quantile dependence probabilities for the lower (p^L) and higher (p^U) extremes of the transformed variables as:

$$p^L = \frac{1}{Tq} \sum_{i=1}^{N} 1\{U_{Yt} \leq q | U_{Xt} \leq q\} \tag{11}$$

$$p^U = \frac{1}{T(1-q)} \sum_{i=1}^{N} 1\{U_{Yt} > q | U_{Xt} > q\} \tag{12}$$

Where U_{Yt} and U_{Xt} are the scores of Y and X based on their empirical distribution. We use copula function approach to estimate the joint distribution of yield and rainfall. The copula function provides a flexible way to bind the univariate marginal distributions of random variables to form a multivariate distribution and can accommodate different marginal distributions of the variables (Nelsen 2006; Trivedi and Zimmer 2007). A two-dimensional copula can be defined as a function $C(u, v) : [0, 1]^2 \rightarrow [0, 1]$ such that

$$F(Y, X) = P[G(Y) \leq G(y), F(X) \leq F(x)] \tag{13}$$

$$F(Y, X) = C(G(Y), F(X); \theta) \tag{14}$$

Where θ represents the strength of dependence. The joint probability density function can be expressed as:

$$c(G(Y), F(X); \theta) = \frac{\partial C(G(Y), F(X); \theta)}{\partial G(Y) \partial F(X)} = C(G(Y), F(X); \theta) g(Y) f(X) \tag{15}$$

Skalar (1959) has shown that for a continuous multivariate distribution, the copula representation holds for a unique copula C. This construction allows us to estimate separately the marginal distributions and the joint dependence of the random variables. There are several parametric families of copula available in the literature. The frequently used ones are the elliptical copulas and the Archimedean copulas. Note that the nature of dependence among the random variables will depend on the copula function chosen for estimation. The statistical properties of the copulas that we use in this paper are given in appendix.

We use two-step maximum likelihood procedure to estimate the copula function wherein the marginals are estimated in the first step, and the dependence in the second step by substituting the estimated marginal distributions in the selected copula function (Trivedi and Zimmer 2007). A non parametric estimator is used to estimate the univariate marginal distribution for crop yield deviations and rainfall deviations. This makes the model semi parametric. Estimation of copula using non parametric distribution does not affect the asymptotic distribution of the estimated copula dependence parameter (Chen and Fan 2006).

A simple maximum likelihood estimator can be used to choose the best fitting copula and estimate the dependence parameter (Patton, 2013). Selection of the copula model can be made based on the Akaike or (Schwarz) Bayesian information criterion. If all the copulas have equal number of parameters, then the choice of model based on these criteria is equivalent to choosing copula with highest log likelihood (Trivedi and Zimmer 2007). The log likelihood function of the copula can be written as:

$$L(\theta) = \sum_{i=1}^{N} \log C(\hat{U}_{Xi}, \hat{U}_{Yi}; \theta) \tag{16}$$

Where $\hat{U}_{Xi} = \hat{G}(Y_i)$ and $\hat{U}_{Yi} = \hat{F}(R_i)$ are the estimated marginal distributions. Copula parameter can be estimated by maximizing the likelihood function using numerical methods. This procedure gives the Inference Functions for Margins (IFM) estimator as θ is conditional on the model that is used to transform the raw data (Joe, 2015; Patton, 2013).

4 Results

4.1 Tail-dependence between station level rainfalls

In this section we investigate the joint association between rainfalls at two stations. Figure 1a shows scatter plot of pair wise linear and rank correlations between all the possible combinations of rainfall stations as a function of the distance between them. The right panel of the figure shows the best fit curve to the rainfall station pair correlations. These clearly show that the joint association between rainfalls at two stations is inversely related to the distance between them. Interestingly the curve for rank correlation is above the curve for linear correlation when two stations are close to each other. But, the difference between the two narrows down as the distance between the stations increases. This is an indication of tail-dependence in rainfall as rank correlation is better suited at capturing nonlinear relationships between the variables.

Correlation is a global measure of association whereas we are interested in the association between random variables when they are at their extremes. To study the behavior of joint distribution of rainfalls at extremes we create a dataset of all possible combinations of rainfall station pairs. Using this, for each station pair, we generate a new dataset of lower and upper tail dependence coefficients for different values of the threshold q. Figure 1b shows the best fitted curves for the lower and upper tail-dependence statistic for pair-wise rainfalls as a function of the distance between the stations. The tail-dependence declines with distance, but rate of decline is less at the lower values of q. We model this behavior econometrically in the following way.

$$\lambda_{ij} = \beta_1 \log(Distance)_{ij} + \beta_2 q + \beta_3 \log(Distance)_{ij} \times q + \alpha_i + \alpha_j + \epsilon_{ij} \tag{17}$$

Where λ_{ij} the estimated tail dependence coefficient between rainfalls measured at two stations i and j, $\log(Distance)_{ij}$ is the distance in kilometers between the two station pair and q is the threshold chosen for the tail dependence statistic. The interaction coefficient captures the interplay between distance and extreme events. Table 1 shows the estimated coefficients from the regressions. The coefficient of the interaction term is negative and statistically significant. Since lower values of q correspond to more extreme deviations in rainfall the analysis reveals that extreme deviations in rainfall are more widespread as compared to the moderate deviations. Hence, extreme rainfall shocks will survive spatial aggregation in comparison to moderate shocks. If yield across farms are dependent on local rainfall then it will also inherit the tail-dependence property. The implication of this finding is that an extreme rainfall anomaly will lead to spatially correlated crop losses.

As a robustness check, we test for tail-dependence between the station-level rainfall by fitting different copula models on station-pairs with distance less than or equal 2000 kilometers. Students t copula appears best fit for almost half of the station-pairs, followed by Plackett and rotated Clayton copula (table 2a). The Students t copula exhibits both upper and lower tail-dependence. This indicates that rainfall in general exhibits a stronger association in case of both extremely low and extremely high deviations from the normal. The mean values of the tail-dependence coefficients based on the copula parameter for all the station-pairs are presented in table 2b and show a declining strength of association when the distance between two stations increases. This is similar to the pattern observed in the non parametric tail-dependence coefficients.

4.2 Tail-dependence in district crop yield and rainfall

Table 3 presents coefficients of linear and rank correlation between yield and rainfall deviations. As expected, both measures show a statistically significant positive association between yield and rainfall deviations, despite some difference in their magnitude.

Figure 2a shows the scatter plots of yield and rainfall deviations along with the linear fit. In figure 2b and 2c we present the scatter and bivariate kernel density plots of the rank-based empirical marginal distribution of yield and rainfall deviation. We observe clustering of rank scores (for yield and rainfall deviations) in the lower-left corner of scatter plots for most crops. Such a clustering corresponds to extreme shortfalls in yield and rainfall, and implies greater probability of simultaneous occurrence of these events.

The scatter plots of rank-based empirical distributions gives us strong indication that association between yield and rainfall index is not linear. Therefore, we test for the presence of tail-dependence in their joint distribution using the conditional quantile dependence probabilities. Figure 3 shows estimated lower tail (figure 3a) and upper tail (figure 3b) quantile dependence plots; and the difference between the two (figure 3c). For comparison we also present the quantile dependence from the moments matched bivariate normal distribution as dashed line in this figure. For all crops the quantile dependence probability at the lower tail of the joint distribution is greater than the same exhibited by normal distribution. This again is evidence of lower tail-dependence in crop yield and rainfall deviations. The quantile dependence plots for the upper tail don't show any evidence of tail-dependence in the joint distribution of yield and rainfall distribution. We also find strong evidence that the joint distribution of crop yield and rainfall deviations exhibit asymmetric tail-dependence. The difference between the upper and lower quantile dependence is statistically significant and is greater at lower quantiles (figure 3c). These results clearly reveal that the bivariate normal distribution is unsuitable to model the joint distribution of yields and rainfalls.

We use copula functions to capture the asymmetric dependence between yield and rainfall deviations by fitting copulas to rank-based empirical marginal distributions of yield and rainfall deviations. Based on the log likelihood values, the Clayton copula is the best model to describe the dependence between yield and rainfall deviations (Table 4). This is not surprising as Clayton copula exhibits only lower tail-dependence and no upper tail-dependence. The worst performing copula models are one with zero lower tail-dependence and allow only upper tail-dependence like Gumbel and rotated Clayton. Table 5 presents the parameters of the Clayton copula with bootstrapped standard errors and lower tail-dependence based on the fitted copula parameter.

The estimated copula density for different crops is presented in Figure 4. As expected, all crops show significantly higher density at the lower tail. This further confirms that the association between yield and rainfall deviations is stronger at the lower tail. This means when rainfall is abnormally low, yield losses are widespread. Therefore, the basis risk is low for extreme shortfall in rainfall.

As a robustness check, we fit all the selected eight copula models to each district that has at least 40 data observations. Based on the log likelihood values, we choose the one that best describes the dependence. Table 6 summarizes the results. For example, in the case of rice Clayton copula gives best fit for 41 percent of the 274 rice growing districts. Students t copula is the next best. Clayton and Students t copula models best describe the joint dependence between yield and rainfall deviations. These findings clearly indicate nonlinearity in association between weather and yield risk and has implications for designing of the insurance products and demand thereof.

5 Implications for Index based crop insurance

For a variety of reasons, agricultural insurance programs use losses against an index (rainfall, area yield) rather than losses against individual yields to make payouts. While this facilitates the supply of insurance, the resulting basis risk reduces the value of insurance and therefore reduces demand for it. The ideal index should be as close as possible to farm-specific yields but it should not be amenable to manipulation by the actions of the insured. Because of heterogeneity in farm-specific yields, it is clear that an index that applies to a group of farmers need not necessarily be an ideal index. In other words, the basis risk is inevitable. Our findings show that the joint density of yield and rainfall exhibit lower tail-dependence, i.e. a stronger association between yield and rainfall when rainfall is abnormally low. This implies that the basis risk varies across the joint distribution of yield and index. This opens up the possibility of designing insurance such that it covers the losses with the least basis risk. Here, we analyze the implications of these findings for the demand and design of index insurance.

5.1 Simulation

To see how tail-dependence influences willingness to pay for index insurance we undertake a simple simulation exercise. We generate 10 million observations of yield and rainfall from Clayton copula and from bivariate normal distribution. For both the marginal distribution of yield and rainfall are assumed to be normal with an assumed mean of 2000 and standard deviation of 300. The copula parameter is set at 0.39, same as that estimated for rice yield and rainfall. The linear correlation between the simulated yield and rainfall from Clayton copula and bivariate normal distribution is assumed to be the same. The only difference is that yield and rainfall index simulated from Clayton copula exhibit lower tail-dependence, while the other does not. To calculate willingness to pay, the coefficient of risk aversion is assumed to be 0.9.

Figure 5 plots the simulated insurance contract schedules for different triggers (Equation 1) superimposed over marginal density of rainfall. As the contract provides indemnity in proportion to shortfall in the index from trigger, the contract schedule is steeper at lower triggers of rainfall. The simulation results are shown in figure 6. In panel (a) of the figure 6 we present willingness to pay for insurance from Clayton copula and bivariate normal distribution normalized by the willingness to pay without basis risk. The willingness to pay for insurance with tail dependent risk is much higher than the willingness to pay from normal distribution. The difference between the two is greater at lower rainfall triggers, which reflects lower basis risk at the tails of Clayton copula.

Figure 6 panel (b) plots the ratio of willingness to pay in relation to actuarially fair premium at different triggers for lower tail of the rainfall distribution. In general, the willingness to pay relative to the actuarially fair premium declines at higher trigger levels. With tail-dependent risk, a risk-averse farmer is willing to pay much higher for insurance that triggers payout at extremely low rainfall. This is again due to lower basis risk at the lower tail of Clayton copula. In comparison to willingness to pay with basis risk the willingness to pay in absence of basis risk remains higher at all trigger levels.

Figure 7 plots certainty equivalent when an agent pays actuarially fair premium rate. In comparison to the certainty equivalent under bivariate normal distribution, the certainty equivalent for Clayton copula reaches its maximum at a higher trigger level. This indicates that the choice of trigger is contingent upon the underlying dependency between yields and index. In figure 8 we show how the joint dependence between yield and rainfall will influence the demand for insurance with increasing risk-aversion. In general, the demand for insurance with tail dependent risk relative to bivariate normal dependence increases with increase in risk aversion, but the rate of increase is higher at extremely low rainfall trigger. This implies that highly risk-averse farmers will gain the most from insurance against extreme shortfall in rainfall. The gains are even more in case of tail dependent risk.

The main finding from the simulation exercise is that the value (to farmers) of index-based insurance relative to actuarial cost is highest for insurance against extreme or catastrophic losses (of the index) than for insurance against all losses.

5.2 Implications

The importance of crop insurance programs stems from the fact that farm incomes are volatile and the farmers are not capable to cope with extremely high risk. This is the equity objective of crop insurance. The efficiency objective of insurance is that by creating a market for output risk, the producer behavior and decisions should move closer to socially optimal levels. The experience from the implementation of crop insurance schemes in different countries shows that the crop insurance programs are safety nets primarily meant to support the farmers and farm sector. Having assessed the costs and benefits of US federal crop insurance program, Goodwin (2015) observes that for each dollar paid as premium, the program on average pays back $1.88 as indemnity payments.

India has a long history in crop insurance schemes. In the Comprehensive Crop Insurance Scheme which was operational over the 1985-1998 period, the average claims to premium ratio for was 5.75 and the same for National Agricultural Insurance Scheme which was under operation form 1999-2012 was 3.31. These schemes were primarily meant to provide economic support to the agricultural sector as no private insurance programs can profitably operate under such conditions. The recently launched Pradhan Mantri Fasal Bima Yojana (PMFBY) is the Indian government's recent attempt to expand insurance coverage to majority of Indian farmers by heavily subsidizing insurance premiums.

The idea behind heavily subsidizing insurance premium is that subsidies are essential for widespread uptake of insurance products. If so, the question is: What is the best way to provide subsidy? Our analysis shows that crop losses are widespread during extreme climatic events such as droughts. This implies that a considerable proportion of farmers would benefit from a program that covers their risks during an extreme weather event. In other words, any form of insurance that protects from extreme losses is likely to be favored by a majority of the farmers. The actuarial cost of such an insurance scheme will be lower compared to a normal insurance; hence less burden on government exchequer. Indeed, a policy that completely subsidizes extreme loss insurance could possibly be revenue neutral relative to an insurance program that covers crop losses based on rainfall-deficit.

Extreme loss insurance programs are likely to be more useful to local aggregators of risk such as banks, producer companies, cooperatives, agri-business firms and local governments. There is a very established protocol for drought relief expenditures by the government. However, its timeliness is often questioned because of many layers of permissions required for such expenditures. On the other hand, an extreme loss insurance program offers the benefits of drought relief but in a timely manner.

We note that farmers may not purchase insurance for other reasons as well including poor understanding of the product, credit constraints, low trust of the insurance seller, and optimism about yields. If these are binding constraints, then again a reduction in basis risk may not impact the demand for insurance.

6 Conclusions

Although cost effective and free from moral hazard and adverse selection, the index based crop insurance products have seen poor uptake because of imperfect association between index and crop loss that reduces the value of insurance and therefore its demand.

We find the association between crop yield and rainfall index characterized by the statistical property of tail-dependence. This implies that the associations between yield losses and index are stronger for large deviations than for small deviations. The most important implication of our findings is that for farmers the utility of index-based insurance relative to actuarial cost is more during extreme or catastrophic losses than for insurance against all losses. This opens up the issue of evaluating the cost effectiveness of an insurance product that limits itself to compensation against extreme events. Our findings also generates a need to systematically evaluate the basis risk and uptake for index insurance products that differ with respect to the contract threshold.

Finally, we wish to point out that tail-dependence is unlikely to be India specific since it flows from the nature of spatial associations of weather. Therefore, although our results are based on Indian data, the general lessons are available for other countries too.

References

[1] Aghakouchak, A., Ciach, G., & Habib, E. (2010). Estimation of tail dependence coefficient in rainfall accumulation fields. *Advances in water Resources*, 33(9), 1142-1149.

[2] Carter, M. R., & Barrett, C. B. (2006). The economics of poverty traps and persistent poverty: An asset-based approach. *The Journal of Development Studies*, 42(2), 178-

199.

[3] Chen, X., & Fan, Y. (2006). Estimation of copula-based semiparametric time series models. *Journal of Econometrics*, 130(2), 307-335.

[4] Cole, S., Gine, X., Tobacman, J., Topalova, P., Townsend, R., & Vickery, J. (2013). Barriers to household risk management: Evidence from India. *American Economic Journal: Applied Economics*, 5(1), 104-135.

[5] Cole, S., Stein, D., & Tobacman, J. (2014). Dynamics of demand for index insurance: Evidence from a long-run field experiment. *The American Economic Review*, 104(5), 284-290.

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

[7] De Weerdt, J., & Dercon, S. (2006). Risk-sharing networks and insurance against illness. *Journal of development Economics, 81(2), 337-356.*

[8] Dercon, S., & Christiaensen, L. (2011). Consumption risk, technology adoption and poverty traps: Evidence from Ethiopia. *Journal of development economics*, 96(2), 159-173.

[9] Dercon, S., Hill, R. V., Clarke, D., Outes-Leon, I., & 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.

[10] Elabed, G., & Carter, M. R. (2015). Compound-risk aversion, ambiguity and the willingness to pay for microinsurance. *Journal of Economic Behavior & Organization*, 118, 150-166.

[11] Fafchamps, M., & Lund, S. (2003). Risk-sharing networks in rural Philippines. *Journal of development Economics*, 71(2), 261-287.

[12] Frahm, G., Junker, M., & Schmidt, R. (2005). Estimating the tail-dependence coefficient: properties and pitfalls. *Insurance: mathematics and Economics*, 37(1), 80-100.

[13] Gine, X., Townsend, R., & Vickery, J. (2008). Patterns of rainfall insurance participation in rural India. *The World Bank Economic Review*, 22(3), 539-566.

[14] Goodwin, B. K. (2001). Problems with market insurance in agriculture. *American Journal of Agricultural Economics*, 83(3), 643-649.

[15] Hill, R. V., Robles, M., & Ceballos, F. (2016). Demand for a Simple Weather Insurance Product in India: Theory and Evidence. *American Journal of Agricultural Economics*, 98(4), 1250-1270.

[16] Joe, H. (2015) *Dependence Modelling with Copulas*, Chapman and Hall, London.

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

[18] Kuhn, G., Khan, S., Ganguly, A. R., & Branstetter, M. L. (2007). Geospatialtemporal dependence among weekly precipitation extremes with applications to observations and climate model simulations in South America. *Advances in Water Resources*, 30(12), 2401-2423.

[19] Liu, P., & Miranda, M. J. (2010, July). Tail dependence among agricultural insurance indices: the case of Iowa county-level rainfalls. In *AAEA annual meeting*, Denver CO.

[20] Mobarak, A. M., & Rosenzweig, M. R. (2013). Informal risk sharing, index insurance, and risk taking in developing countries. *The American Economic Review*, 103(3), 375-380.

[21] Mobarak, A. M., & Rosenzweig, M. R. (2013). Selling formal insurance to the informally insured. Working Paper No. 97, Economic Growth Center Discussion Paper No. 1007, Yale University.

[22] Nelsen, R. B. (2006). An introduction to copulas, ser. *Lecture Notes in Statistics. New York: Springer.*

[23] Patton, A. J. (2012). Copula methods for forecasting multivariate time series. *Handbook of economic forecasting*, 2, 899-960.

[24] Rosenzweig, M. R., and Binswanger, H. P. (1993). Wealth, weather risk and the profitability of agricultural investment. *Economic Journal*, 103(416), 56-78.

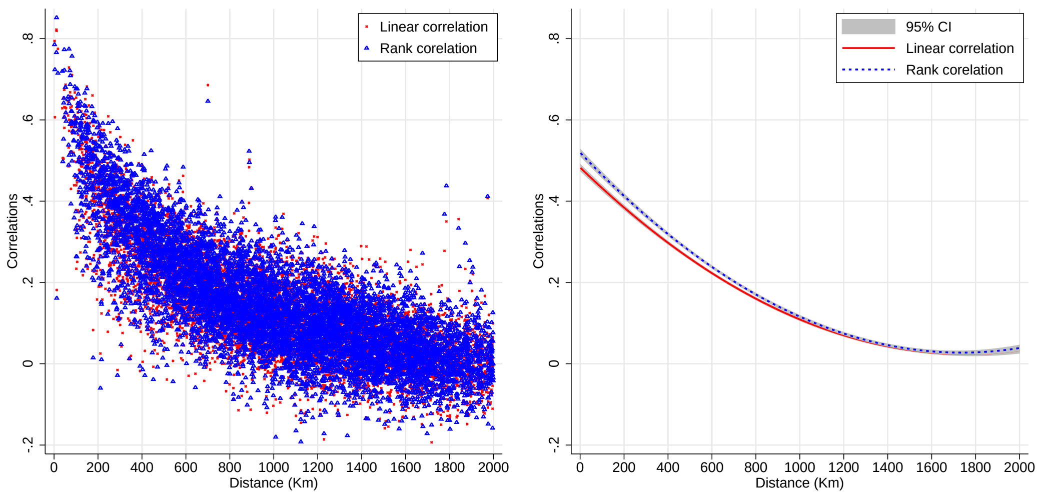
[25] Rosenzweig, M. R., & Stark, O. (1989). Consumption smoothing, migration, and marriage: Evidence from rural India. *Journal of political Economy*, 97(4), 905-926.

[26] Sklar, M. (1959). Fonctions de repartition an dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris, 8*, 229-231.

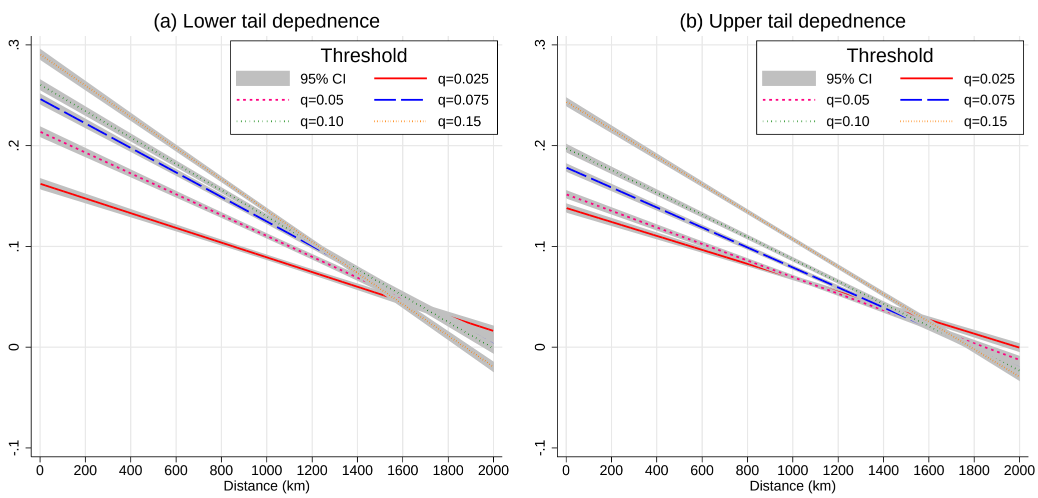
[27] Townsend, R. M. (1994). Risk and insurance in village India. *Econometrica: Journal of the Econometric Society*, 539-591.

[28] Trivedi, P. K., & Zimmer, D. M. (2007). Copula modeling: an introduction for practitioners. *Foundations and Trends in Econometrics*, 1(1), 1-111.

Figures

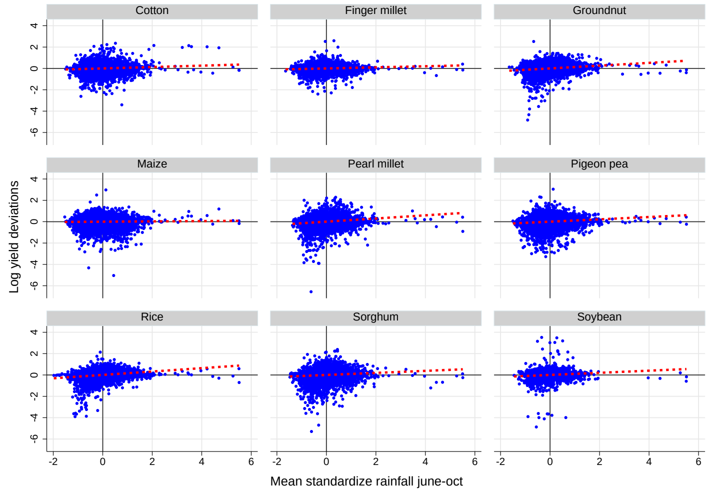


(a) Correlation and distance

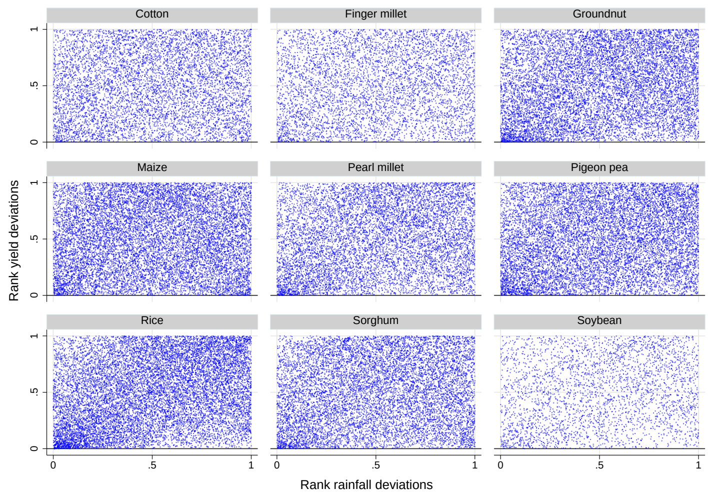


(b) Nonparametric tail-dependence and distance

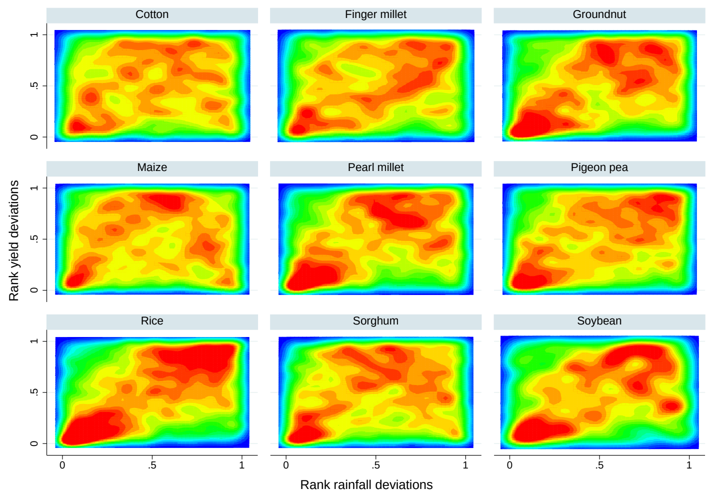
Figure 1: Dependence in pairwise station rainfalls



(a) Scatter plot of yield and rainfall deviations

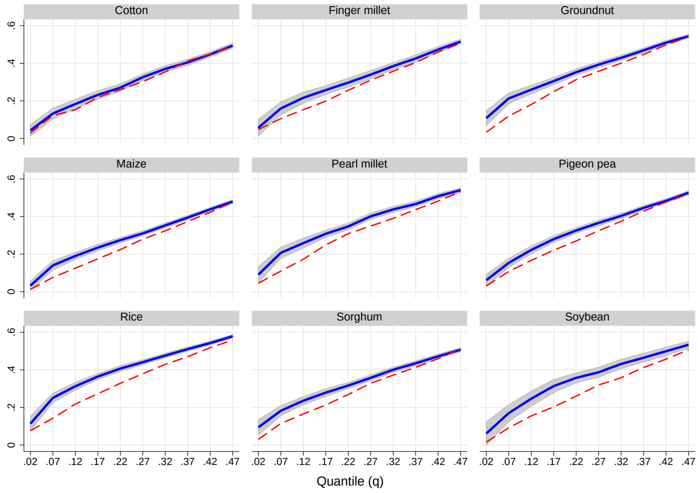


(b) Scatter plot of ranks of yield and rainfall deviations

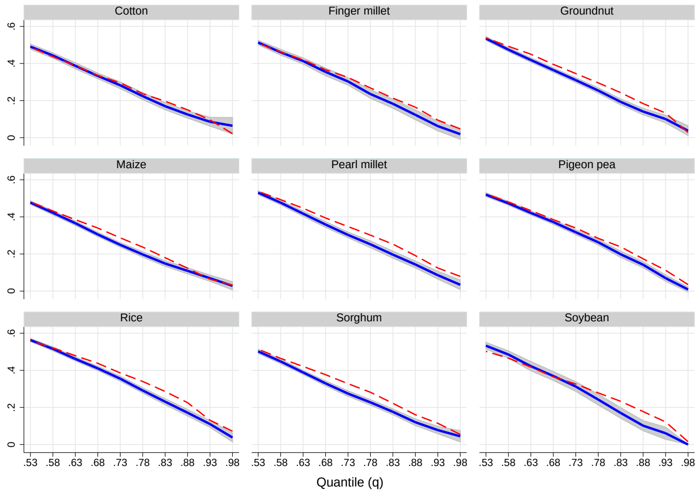


(c) Kernel density plots of ranks of yield and rainfall deviations

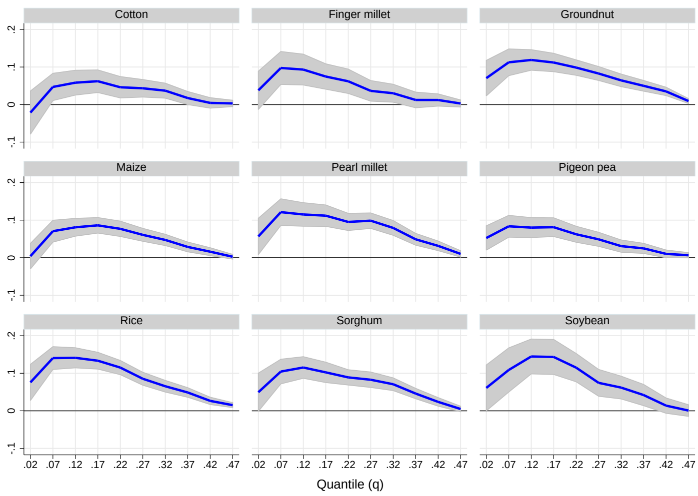
Figure 2: Joint distribution of yield and rainfall deviations



(a) Lower tail-dependence



(b) Upper tail-dependence



(c) Difference in upper and lower tail-dependence

Figure 3: Tail-dependence at different quantiles

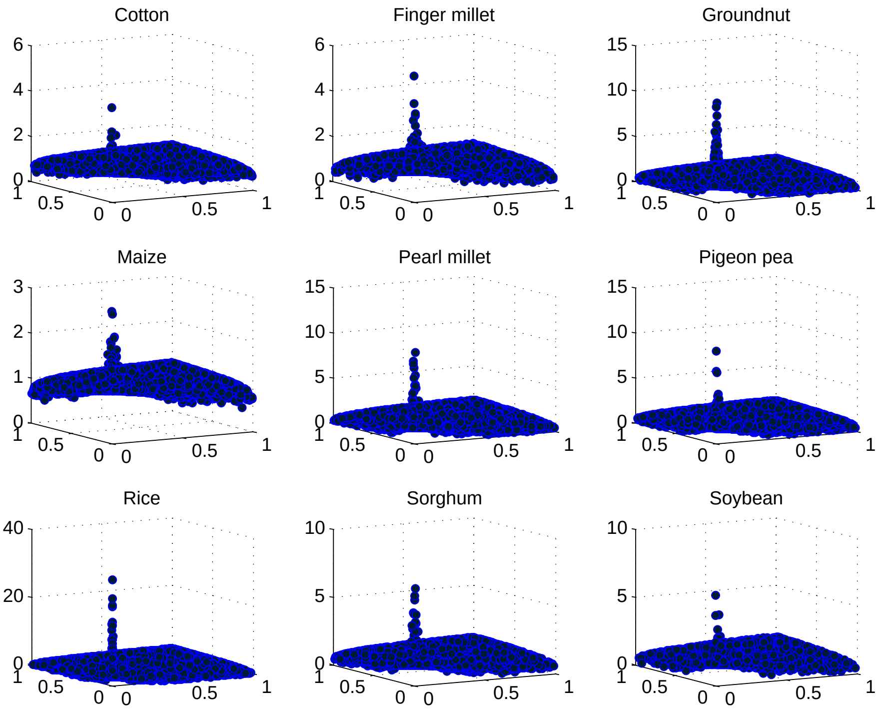


Figure 4: Estimated copula density by crops

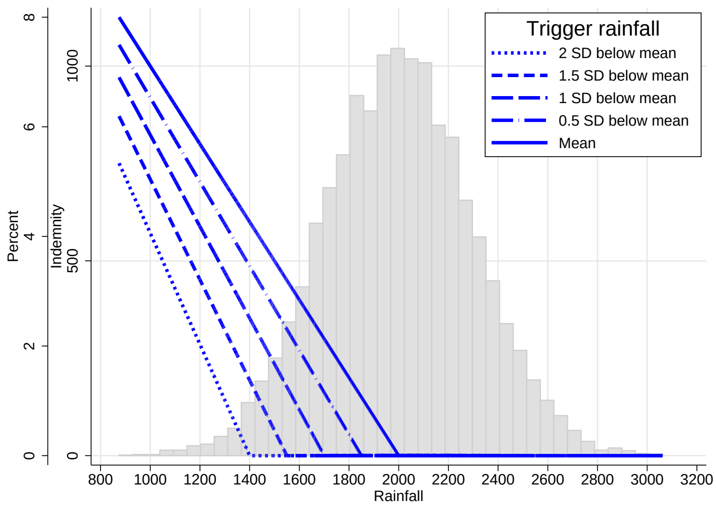


Figure 5: Simulated contract schedule and rainfall index density

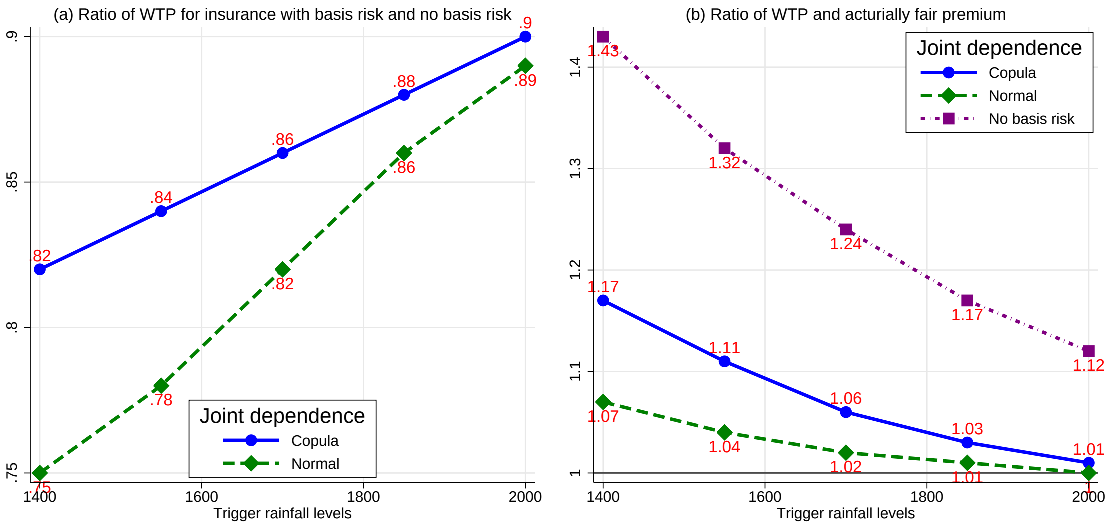


Figure 6: Trigger levels and willingness to pay

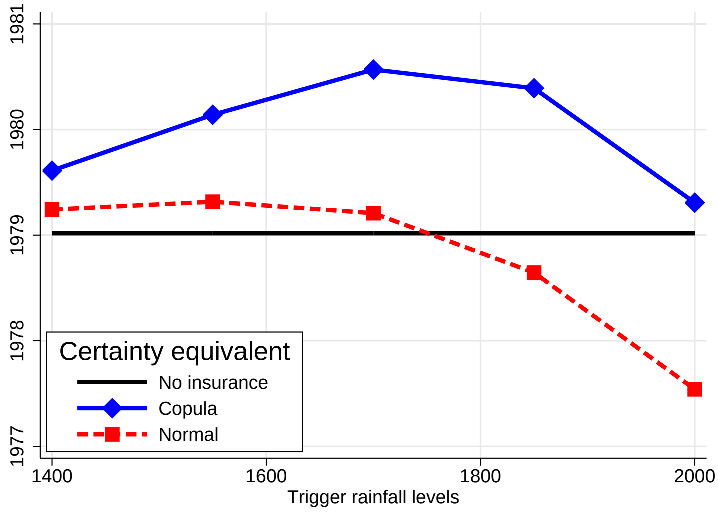


Figure 7: Trigger levels and certainty equivalent

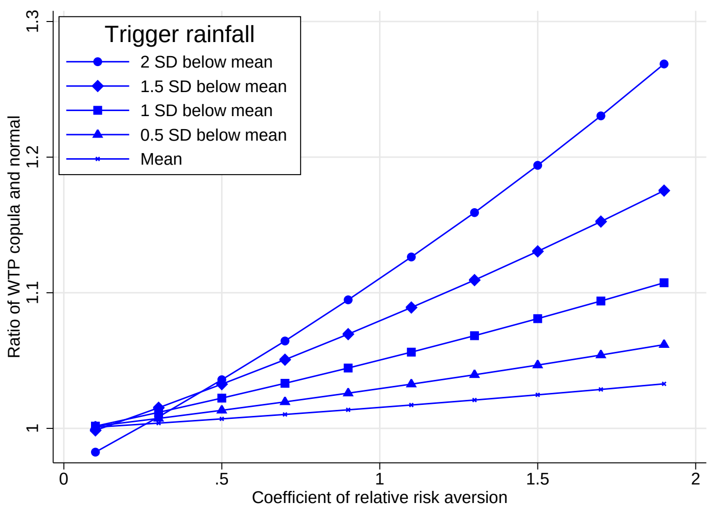


Figure 8: Willingness to pay and risk aversion

Tables

Table 1: Extreme events, tail dependence and distance

	(a) Weather station data		(b) Gridded data	
	Upper $\hat{\lambda}^L$	Lower $\hat{\lambda}^U$	Upper $\hat{\lambda}^L$	Lower $\hat{\lambda}^U$
$\log(Distance)$	-0.06***	-0.06***	-0.10***	-0.09***
	(0.004)	(0.003)	(0.003)	(0.002)
q	2.49***	2.32***	3.96***	3.89***
	(0.240)	(0.205)	(0.094)	(0.062)
$\log(Distance) \times q$	-0.31***	-0.29***	-0.50***	-0.49***
	(0.035)	(0.030)	(0.013)	(0.008)
Constant	0.53***	0.50***	0.81***	0.65***
	(0.027)	(0.021)	(0.023)	(0.016)
Observations	55896	55896	381276	381276
Adjusted R^2	0.48	0.47	0.67	0.68

Note: The dependent variable are the estimated nonparametric tail dependence coefficients. The tail dependence statistic varies between 0 and 1. The regressions include station (grid point) fixed effects. Figure in parenthesis are standard errors clustered at rainfall station level. Panel (a) shows results from the data on actual rainfalls measured at 137 weather stations spread all over India. Panel (b) shows results from the Indian meteorology department's high resolution gridded rainfall data based on rainfall records from 6995 rain gauge stations in India. ***, ** and * indicate statistical significance at the 1%, 5% and 10% levels, respectively.

Table 2: Dependence in pairwise station rainfalls

(a) Copula fitted to pairwise rainfalls

Copula model	Station pairs	Percent
Gaussian	354	4.43
Clayton	437	5.46
Rotated Clayton	950	11.88
Plackett	1204	15.05
Frank	318	3.98
Gumbel	188	2.35
Rotated Gumbel	698	8.73
Student's t	3849	48.12
Total	7998	100

(b) Estimated tail-dependence based on fitted copula and distance

		Distance between pair of stations in kilometers					
	Copula	2-479	498-776	777-1033	1033-1287	1287-1572	1573-1999
Lower	Clayton	0.183	0.019	0.01	0.009	0.006	0.003
		(0.236)	(0.039)	(0.02)	(0.025)	(0.019)	(0.013)
	Rotated Gumbel	0.284	0.209	0.163	0.146	0.133	0.129
		(0.09)	(0.06)	(0.043)	(0.032)	(0.02)	(0.019)
	Student's t	0.573	0.523	0.503	0.493	0.49	0.482
		(0.051)	(0.034)	(0.031)	(0.026)	(0.026)	(0.024)
	Total	0.353	0.336	0.292	0.237	0.194	0.172
		(0.274)	(0.238)	(0.238)	(0.236)	(0.231)	(0.226)
Upper	Rotated Clayton	0.126	0.064	0.026	0.017	0.006	0.003
		(0.094)	(0.058)	(0.04)	(0.031)	(0.014)	(0.008)
	Gumbel	0.294	0.185	0.149	0.149	0.133	0.122
		(0.091)	(0.054)	(0.032)	(0.037)	(0.014)	-
	Student's t	0.573	0.523	0.503	0.493	0.49	0.482
		(0.051)	(0.034)	(0.031)	(0.026)	(0.026)	(0.024)
	Total	0.348	0.324	0.28	0.228	0.185	0.165
		(0.276)	(0.247)	(0.247)	(0.241)	(0.235)	(0.229)

Note: Standard deviation in parenthesis.

Table 3: Linear and rank correlation between yield and rainfall deviations

Crops	Pearson linear correlation	Spearman rank correlation
Rice	0.255(0.008)	0.272(0.008)
Groundnut	0.184(0.009)	0.184(0.009)
Pearl millet	0.169(0.011)	0.177(0.010)
Pigeon pea	0.135(0.009)	0.156(0.009)
Soybean	0.119(0.019)	0.171(0.016)
Sorghum	0.107(0.011)	0.108(0.010)
Finger millet	0.089(0.015)	0.110(0.014)
Cotton	0.082(0.017)	0.067(0.012)
Maize	0.020(0.011)	0.034(0.009)

Note: Bootstrapped (200 replications) standard errors in parenthesis.

Table 4: Log likelihood from different copula models

Crops	Gaussian	Clayton	Rotated Clayton	Plackett	Frank	Gumbel	Rotated Gumbel	Student's t
Cotton	20.4	30.8	9.5	16.4	16.1	-11.7	10.2	23.5
Finger millet	30.4	51.2	4.6	33.5	32.8	-6.6	44.5	31.9
Groundnut	201.6	266.4	66.2	189.0	185.8	103.3	247.9	210.1
Maize	8.5	31.4	0.1	7.5	7.4	-121.0	-29.4	13.7
Pearl millet	154.2	204.9	50.7	145.1	142.4	76.3	194.4	161.4
Pigeon pea	139.5	172.0	41.0	142.7	141.8	57.2	154.8	139.9
Rice	518.5	615.5	200.1	519.9	512.6	318.3	602.1	529.2
Sorghum	73.9	127.8	13.3	61.6	60.3	-2.4	109.8	79.6
Soybean	42.3	61.0	8.9	49.6	49.0	14.1	56.5	43.9

Note: Negative of log likelihood values.

Table 5: Clayton copula model parameter estimates

Crops	Parameter Estimates	Standard errors	Tail dependence
Cotton	0.102	0.014	0.0011
Finger millet	0.158	0.018	0.0124
Groundnut	0.267	0.013	0.0747
Maize	0.075	0.011	0.0001
Pearl millet	0.257	0.014	0.0676
Pigeon pea	0.204	0.013	0.0333
Rice	0.391	0.013	0.1702
Sorghum	0.178	0.012	0.0205
Soybean	0.233	0.025	0.0512

Table 6: Percent districts with best fit copula

Crops	Gaussian	Clayton	Rotated Clayton	Plackett	Frank	Gumbel	Rotated Gumbel	Student's t	Total
Cotton	10	28	9	7	7	2	5	33	100
	(12)	(34)	(11)	(8)	(8)	(3)	(6)	(40)	(122)
Finger millet	4	44	6	1	7	0	3	35	100
	(3)	(31)	(4)	(1)	(5)	(0)	(2)	(25)	(71)
Groundnut	4	45	4	6	6	3	6	26	100
	(8)	(84)	(8)	(12)	(12)	(5)	(11)	(48)	(188)
Maize	7	28	5	9	4	2	3	43	100
	(17)	(70)	(12)	(22)	(10)	(4)	(7)	(108)	(250)
Pearl millet	2	50	3	6	5	1	3	31	100
	(3)	(78)	(4)	(10)	(8)	(1)	(5)	(48)	(157)
Pigeon pea	4	42	8	5	7	3	3	28	100
	(9)	(91)	(18)	(10)	(16)	(7)	(7)	(60)	(218)
Rice	5	40	3	7	10	2	9	26	100
	(14)	(109)	(7)	(18)	(27)	(5)	(24)	(70)	(274)
Sorghum	4	41	4	7	4	1	5	35	100
	(7)	(81)	(7)	(14)	(8)	(2)	(10)	(69)	(198)
Total	5	39	5	6	6	2	5	32	100
	(73)	(578)	(71)	(95)	(94)	(27)	(72)	(468)	(1478)

Note: Number of districts in parenthesis.

Appendix

There are a variety of copula models available in the literature. Table 1 which is taken from Patton (2012) shows the properties of some of the popular copula models which we use to find the best fitting model for our data.

	Parameter(s)	Parameter space	Indep	Pos & Neg dep?	Rank correlation	Kendall's τ	Lower tail dep	Upper tail dep
Normal	ρ	$(-1,1)$	0	*Yes*	$\frac{6}{\pi}\arcsin\frac{\rho}{2}$	$\frac{2}{\pi}\arcsin\rho$	0	0
Clayton	γ	$(0,\infty)$	0	*No*†	*n.a.*	$\frac{\gamma}{\gamma+2}$	$2^{-1/\gamma}$	0
Rotated Clayton	γ	$(0,\infty)$	0	*No*†	*n.a.*	$\frac{\gamma}{\gamma+2}$	0	$2^{-1/\gamma}$
Plackett	γ	$(0,\infty)$	1	*Yes*	$\frac{\gamma^2-2\gamma\log\gamma-1}{(\gamma-1)^2}$	*n.a.*	0	0
Frank	γ	$(-\infty,\infty)$	0	*Yes*	$g_\rho(\gamma)$	$g_\tau(\gamma)$	0	0
Gumbel	γ	$(1,\infty)$	1	*No*	*n.a.*	$\frac{\gamma-1}{\gamma}$	0	$2-2^{1/\gamma}$
Rotated Gumbel	γ	$(1,\infty)$	1	*No*	*n.a.*	$\frac{\gamma-1}{\gamma}$	$2-2^{1/\gamma}$	0
Sym Joe-Clayton	$\tau^L,\ \tau^U$	$[0,1)\times[0,1)$	$(0,0)$	*No*	*n.a.*	*n.a.*	τ^L	τ^U
Student's t	$\rho,\ \nu$	$(-1,1)\times(2,\infty)$	$(0,\infty)$	*Yes*	*n.a.*	$\frac{2}{\pi}\arcsin(\rho)$	$g_T(\rho,\nu)$	$g_T(\rho,\nu)$

Notes: This table presents some common parametric copula models, along with their parameter spaces, and analytical forms for some common measures of dependence, if available. . For more details on these copulas see Joe (1997, Chapter 5) or Nelsen (2006, Chapters 4-5). Measures that are not available in closed form are denoted "*n.a.*". Parameter values that lead to the independence copula are given in the column titled "Indep". Frank copula rank correlation: $g_\rho(\gamma)=1-12\left(D_1(\gamma)-D_2(\gamma)\right)/\gamma$ and Frank copula Kendall's tau: $g_\tau(\gamma)=1-4\left(1-D_1(\gamma)\right)/\gamma$, where $D_k(x)=kx^{-k}\int_0^x t^k\left(e^t-1\right)^{-1}dt$ is the "Debye" function, see Nelsen (2006). Student's t copula lower and upper tail dependence: $g_T(\rho,\nu)=2\times F_{Studt}\left(-\sqrt{(\nu+1)\frac{\rho-1}{\rho+1}},\nu+1\right)$, see Demarta and McNeil (2005). †The Clayton (and rotated Clayton) copula allows for negative dependence for $\gamma\in(-1,0)$, however the form of this dependence is different from the positive dependence case $(\gamma>0)$, and is not generally used in empirical work.

Figure 1: Estimated copula density by crops

The copula model which best fits our data and we extensively use in our analysis is the Clayton copula. The functional form of the Clayton copula function is given as.

$$C(u,v)=(u^{-\theta}+v^{-\theta}-1)^{-\frac{1}{\theta}} \tag{18}$$

Where $u=G(Y)$ and $v=H(R)$. This copula shows strong lower tail-dependence but no upper tail-dependence. The probability density function of the Clayton copula is given by the following expression.

$$c(u,v)=\frac{\partial C(u,v)}{\partial u\partial v} \tag{19}$$

$$c(u,v)=(\theta+1)uv^{-(\theta+1)}(u^{-\theta}+v^{-\theta}-1)^{-2-\frac{1}{\theta}} \tag{20}$$

Figure 2 gives the scatter plot of data simulated from Clayton copula and bivariate Normal distribution where the linear correlation between the two simulated datasets is kept the same.

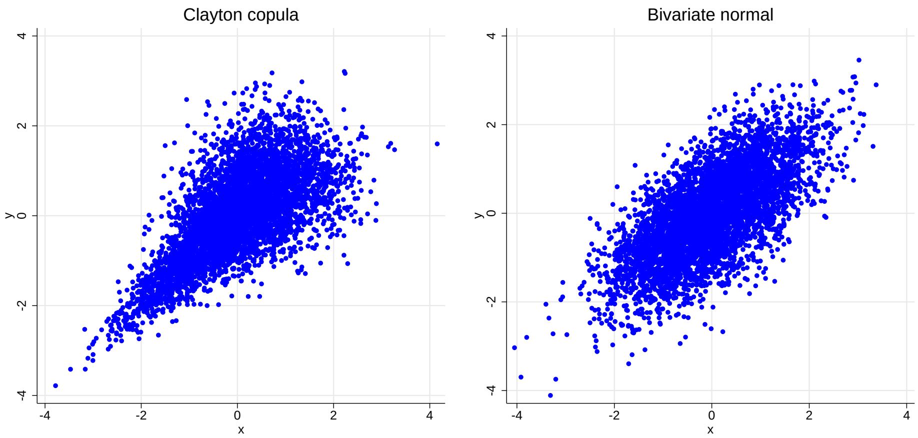


Figure 2: Joint distribution simulated from Clayton copula and bivariate normal distribution