Basis Risk in Index Insurance: Lower Tail Dependence and the Demand for Weather Insurance

Digvijay S. Negi[*]   Bharat Ramaswami[†]

May 2018

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.

---

[*]Doctoral student, Indian Statistical Institute, 7 SJS sansanwal Marg, New Delhi 110016, India. Email: digvijaynegi12@gmail.com

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

1

1 Introduction

Agriculture and agriculture-based livelihoods in developing countries are highly prone to weather shocks. Although there exist various informal mechanisms in rural communities that allow farmers to pool their idiosyncratic risks, such insurance is often partial and, moreover, provide limited insurance to individual households when risks are correlated and widespread. [1] Extreme climate events such as droughts, floods and heat waves which affect farming communities in a region simultaneously are instances of correlated and widespread risks. There is substantial evidence that rural households in high risk environment stick to low return subsistence agriculture and cope with a correlated shock by liquidating productive assets to maintain consumption thus remaining trapped in poverty (Rosenzweig and Binswanger, 1993; Carter and Barrett, 2006; Dercon and Christiaensen, 2011).

Even though farmers in developing countries are typically poor and even though they bear the burden of volatile income streams, formal insurance products have had limited success (Mobarak and Rosenzweig, 2013). The difficulties of administering first best insurance programs tailored to production histories of individual farmers have led to index insurance products where payouts are triggered by an index such as rainfall, temperature or local average yields. Premium setting is relatively easier because past data on indices of weather and average yield are more readily available than on individual production histories. As individual farmers have little or no influence on payouts, index-based insurance products are also less likely to fail due to asymmetry in information between the insurer and the insured. Despite the promise of index

---

[1]The literature on risk sharing in communities is large. Overviews include Bardhan and Udry (1999), Fafchamps (2003), Morduch (1999, 2005), Townsend (1994).

insurance, the record is mixed. In particular, the uptake of index insurance is poor, especially when it is not subsidized (Binswanger-Mkhize, 2012; Jensen and Barrett, 2017; Jensen et al., 2016).

The literature has highlighted many reasons for the low uptake. These include the unfamiliarity among farmers of formal insurance, the lack of trust in the insurance provider, the difficulties of communication resulting in poor understanding of the insurance product. Poor farmers also face liquidity constraints and insurance demand is highly sensitive to price (Cole et al., 2013, 2014; Giné et al., 2008).

However, even if the above factors were absent, research has highlighted the fundamental constraint of basis risk which occurs because of imperfect correlation between the index and farmer losses. If the association is weak, then index insurance might not be reliable (Morsink et al., 2016). Research has shown, both theoretically and empirically, that basis risk reduces the demand for insurance (Clarke, 2016; Elabed and Carter, 2015; Giné et al., 2008; Hill et al., 2016). The importance of acknowledging basis risk is stressed in a recent study that states "Discerning the magnitude and distribution of basis risk should be of utmost importance for organizations promoting index insurance products, lest they inadvertently peddle lottery tickets under an insurance label" (Jensen et al., 2016).

Index insurance products are, at best, designed to offer protection against aggregate or covariate risks (Miranda, 1991; Ramaswami and Roe, 2004; Carter et al., 2014). The lack of a perfect association between the index and losses at the farmer level can, therefore, arise either because the index is not accurate or because idiosyncratic losses are substantial. While previous work has established the sensitivity of insurance

demand and farmer welfare to basis risk, there has not been much work on contract design that reduces basis risk. Chantarat et al. (2013) described an index based livestock insurance where the contract was based on a regression of historic mortality rates on an index of vegetative cover and therefore, was designed to minimize basis risk. In a similar vein, this paper examines how rainfall insurance contracts in India can be designed to reduce basis risk. However, we do not use regression-based methods because a least squares fit is based on the idea of linear correlation. Our approach exploits the idea that the joint distribution of rainfall and output might be characterized by tail dependence. This means that the associations between yield losses and index losses are stronger for large deviations than for small deviations. The major implication 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. Or in simpler words, basis risk is least for large deviations of the index. The goal of this paper is to test this hypothesis.

The contribution of this paper is two-fold. First, it adds to the slender work on how contracts can be designed to lower basis risk. Second, it uses general measures of association (rather than the linear concept of correlation) to characterize the dependence between the index and crop losses. Previous work has recognized that lower tail dependence characterizes the joint distribution of spatial yields (Du et al., 2017; Goodwin, 2014; Goodwin and Hungerford, 2015) and also the joint distribution of spatial rainfall (Aghakouchak et al., 2010). The paper argues that these two facts imply that the joint distribution of rainfall and yields will also exhibit lower tail dependence. Testing this hypothesis and examining its implications for the design of insurance is the contribution of this paper.

4

The paper estimates the tail dependence in the joint distribution of weather (i.e., rainfall) and yields using a district level data set for all India and for 9 major crops. Using maximum likelihood methods, the paper estimates a number of copulas from the parametric families of elliptical copulas and the Archimedean copulas. The best-fit copulas are joined to a conceptual model of an insurance purchaser. The simulation of the copulas allows us to estimate the optimal insurance cover for a variety of insurance contracts that vary according to the index threshold value that triggers payout. These results are compared to those obtained from a copula without tail dependence (the Gaussian copula).

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 that the basis risk is least for large deviations of the index. This is also confirmed by simulations that 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. Because of tail dependence, the demand for commercially priced rainfall insurance is more likely to be positive when coverage is restricted to extreme losses.

2 Relation to Literature

There is no universally accepted definition of basis risk. However, it is commonly understood to arise from the imperfect association between farm level losses and the index that triggers insurance payments. As a result, losses that are actually incurred may not always be compensated by insurance. A particularly stark case is when the farmer suffers a loss but receives no payout. Clarke (2016) refers to the probability of such an event as basis risk. Higher is this probability, greater is the basis risk. In these states of high marginal utility, not only does the farmer *not* receive indemnities but actually suffers cash outflow to pay premiums. For this reason, a risk averse farmer would not want to buy 'too much' of insurance. Higher basis risk reduces the demand for insurance.

A simple model is useful to clarify basis risk and to understand the contribution of this paper. An index insurance is offered to farmers in a region R (village, cooperative, or other units of aggregation). Consider the following model of yield risk. [2]

$$y_{ir} = \mu_{iR}\eta_{iR} \tag{1}$$

where μ_{iR} is the expected yield of producer i in region R and η_{iR} is a unit mean random variable capturing the risks of farming. η_{iR} is a product of two independent unit mean shocks - an aggregate or covariate shock θ_R that affects all farmers in the

---

[2]The model is drawn from Ramaswami and Roe (2004).

region and an idiosyncratic shock e_{iR} that affects only producer i and is given by

$$\eta_{iR} = e_{iR}\theta_R \tag{2}$$

Assuming each producer's share of land in the region is w_{iR}, the area yield for the region R is

$$y = \theta_R \sum_{(i \in R)} w_{iR}\mu_{iR}e_{iR} \tag{3}$$

Let $\mu_R = \sum_i w_{iR}\mu_{iR}$ denote the expected area yield. Then the area yield can be approximated as [3]

$$y = \theta_R\mu_R \tag{4}$$

Therefore, in this model, the aggregate shock θ_R is completely captured by area yield. Finally, insurance payouts z to every insured farmer in the region R is a function of the value of an index x_R. While the exact function is unimportant here, a typical insurance contract is of the form

$$z_R = max\{\alpha(x_m - x_R), 0\} \tag{5}$$

where x_m and α are positive parameters of the contract. x_m is a deductible.

---

[3]$\sum_{(i \in R)} w_{iR}\mu_{iR}e_{iR} = \sum_{(i \in R)} w_{iR}(\mu_{iR} - \mu_R)(e_{iR} - \bar{e}) + \mu_R\bar{e}_R$ where $\bar{e}_R = \sum_i w_{iR}e_{iR}$. The first term is approximately zero (independence of idiosyncratic shocks from expected yield) and in the second term the average idiosyncratic shock is approximately equal to its mean, i.e., 1.

If x_m is high, the insurance covers small and large losses. If it is low, the insurance provides only catastrophic cover.

With reference to this model, basis risk can be quantified in various ways. A simple approach is to examine the correlation between farm yield y_{iR} and insurance payments z_R or the index x_R. As this assumes, basis risk is constant for all values for the index, this paper will consider general dependence structures that incorporate non-linear association. In particular, it may be important to consider the association between yield and the index when the index losses are large. Morsink et al. (2016) propose two measures of the reliability of index insurance. The first metric is the probability of not receiving an insurance payout in the event of a catastrophic loss. The second measure is the ratio of expected payout to premium in the event of a catastrophic loss. This paper considers a further nuance: what is the basis risk (or the reliability of index insurance) for different values of the deductible? In particular, is the basis risk appreciably lower for a low x_m?

The literature has distinguished between two sources of basis risk (Jensen et al., 2016; Morsink et al., 2016). First, if the index is poorly chosen, then aggregate shocks might not be sufficiently sensitive to the index. This has been called insured peril basis risk (Morsink et al., 2016) or design risk (Jensen et al., 2016). In the model described above, area yield is a sufficient statistic for the aggregate shock. However, computation of area yield involves crop cutting experiments or other means of assessing average yield. The greater administrative costs might lead insurance companies to choose an easily measurable weather parameter such as rainfall to approximate the aggregate shock. The problem is that average yield may depend on rainfall as well as

other factors such as hailstorms, or pests that affect the entire region. We can write

$$y_R = f(I_R, \nu_R) \tag{6}$$

where I_R is an index of rainfall and ν_R stands in for all other factors that affect average yield. The absence of a perfect association between rainfall and average yield constitutes the design risk in this model.

Clarke et al. (2012) analyze this source of basis risk in 270 weather insurance contracts in a state of India. They estimate that there is a one-in-three chance of not receiving insurance payout in the event of a total production loss (of area average yield). In a follow-up analysis, Clarke (2016) argued that, if the contracts were priced commercially (i.e., unsubsidized), the basis risk in them was so great as to reduce optimal demand to zero.

The model described above assumed that all producers in the region R face the same aggregate shock. However, even within a small region, rainfall may not occur uniformly. On the other hand, the index of rainfall is computed from one point in the region. Another source of design risk is therefore the imperfect association between rainfall at the farm location and rainfall at the weather station. Previous research has measured such design risk by the distance from the farm to the weather station (that measures the index). This has been shown to reduce insurance demand (Mobarak and Rosenzweig, 2013; Hill et al., 2016).

Even if the index accurately captures aggregate shocks, a second source of basis risk comes from the fact that the aggregate shock is only one component of loss.

In particular, individual specific shocks not captured by the index could also lead to a weak association between the losses in the index and individual farm output losses. Such basis risk has been called production smoothing basis risk (Morsink et al., 2016). Ramaswami and Roe (2004) showed that if individual and aggregate shocks interact multiplicatively (as in above model), then even if index insurance insures aggregate shocks perfectly (i.e., no design risk), the presence of uninsured individual specific risks could reduce the demand for index insurance. [4] Empirically, Jensen et al. (2016), using a unique household level panel, analyze the different sources of basis risk for an index based livestock contract offered in Northern Kenya. They find that the livestock contract did reduce household exposure to aggregate risk, principally, droughts. On average, risk exposure to covariate shocks dropped by about 63%. The failure to reach 100% is reflecting of the design errors in the contract. While the contract was not designed to reduce idiosyncratic risk, such risks were large. Even at the smallest levels of aggregation, idiosyncratic risk accounted for about two-thirds of all risk. Reducing design risk by choosing a better index cannot help in dealing with idiosyncratic risk. The policy imperative would be to keep the aggregation (i.e, region R) as small as possible to minimize idiosyncratic risk.

The fact that index insurance can at best deal with aggregate risk suggests that traditional mechanisms of informal insurance would continue to be important in dealing with idiosyncratic risk. If informal networks provide substantial insurance, it would ameliorate the basis risk in index insurance because of idiosyncratic risk and therefore increase the uptake of index insurance. This hypothesis was tested and con-

---

[4]This is true for all risk averse individuals with convex marginal utility. If individual and aggregate shocks interact additively as in Miranda (1991), then idiosyncratic shocks have no consequence for insurance decision although they do matter to utility (Ramaswami and Roe, 2004).

firmed by Mobarak and Rosenzweig (2012) and Dercon et al. (2014).

This paper is about reducing the design error component of basis risk in rainfall insurance contracts. By considering general dependence structures, the paper opens the door to the possibility that basis risk might vary according to the magnitude of the loss in the index. This possibility is empirically explored by estimating copulas of the distribution of rainfall and yields. While the analysis covers 9 crops across 311 districts from 1966 to 2011, it is limited by the aggregation at the district level. For this reason, the paper cannot throw light on the basis risk due to uninsured idiosyncratic risk.[5] What the research does is to examine the basis risk that arises by using a weather index (rainfall) to measure aggregate or covariate risk. Related papers that share this objective include Clarke (2016), Clarke et al. (2012) and Morsink et al. (2016).

While this literature provides methods to characterize basis risk, this paper advances the research by a formal examination of tail dependence and its implications for redesigning contracts to reduce basis risk. A small literature has begun to explore copula based characterizations of joint distributions to explore the spatial correlations of yield and the implications of pricing premiums (Du et al., 2017; Goodwin, 2014; Goodwin and Hungerford, 2015). The paper extends the application of these methods to a characterization of basis risk in rainfall insurance contracts.

---

[5]This is a limitation shared with much of the literature (e.g., Clarke, 2016) because of the absence of farm level panel data.

3 Background Evidence: Tail Dependence in Rainfall

In the model described in the last section, area average yield was the correct index for local aggregate shocks. More generally, we can let

$$y_{iR} = \mu_{iR}\eta_{iR} \tag{7}$$

where the composite risk is some unspecified function of idiosyncratic and aggregate shock. In other words,

$$\eta_{iR} = g(e_{iR}, \theta_R) \tag{8}$$

The average yield is

$$y = \sum_{(i \in R)} w_{iR}\mu_{iR}\eta_{iR} \tag{9}$$

Once again denoting the expected area yield $\mu_R \equiv \sum_i w_{iR}\mu_{iR}$, we can decompose the right hand side of above as

$$\sum_{(i \in R)} w_{iR}\mu_{iR}\eta_{iR} = \sum_{(i \in R)} w_{iR}(\mu_{iR} - \mu_R)(\eta_{iR} - \bar{\eta}_R) + \mu_R\bar{\eta}_R \tag{10}$$

where $\bar{\eta}_R = \sum_i w_{iR}\eta_{iR}$. If the yield risks are independent of mean yield, the

first term is approximately zero. Hence we can approximate area average yield as

$$y = \mu_R \sum_{(i \in R)} w_{iR} g(e_{iR}, \theta_R) \tag{11}$$

In this more general model, it is no longer sufficient to represent aggregate shocks by average yield. It also depends on the entire distribution of idiosyncratic shocks.

In either model, in so far as rainfall is only one component of aggregate shocks, a rainfall insurance contract would suffer from design basis risk. Ideally, this should be investigated by examining the association between area average yields and the rainfall index that is computed from a weather station within the region. Because of data considerations, we estimate the tail dependence and the copulas of joint distributions of area average yields and area average rainfall.[6] However, this is not a major limitation because tail dependence in the joint distribution of these averages implies tail dependence in the joint distribution of area average yield and a rainfall index.

The reason is as follows. From other parts of the world, it has been found that rainfalls within a region are not only strongly correlated but, in fact, are characterized by tail dependence (e.g., Aghakouchak et al., 2010). Thus, an association of large deviations of area average yield with large deviations of area average rainfall automatically translates to an association of large deviations of area average yield with large deviations of a rainfall index derived from a location within that area.

To confirm the key fact of tail dependence in the distribution of rainfall in

---

[6]Clark's (2016) computations of basis risk in weather insurance products from a state in India is also based on associations of area average yield and area average rainfall.

India, 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.
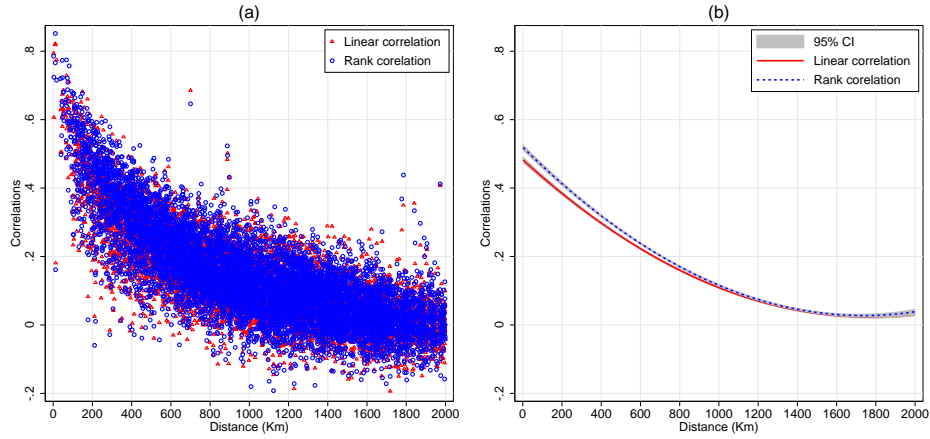
(a) Correlation and Distance
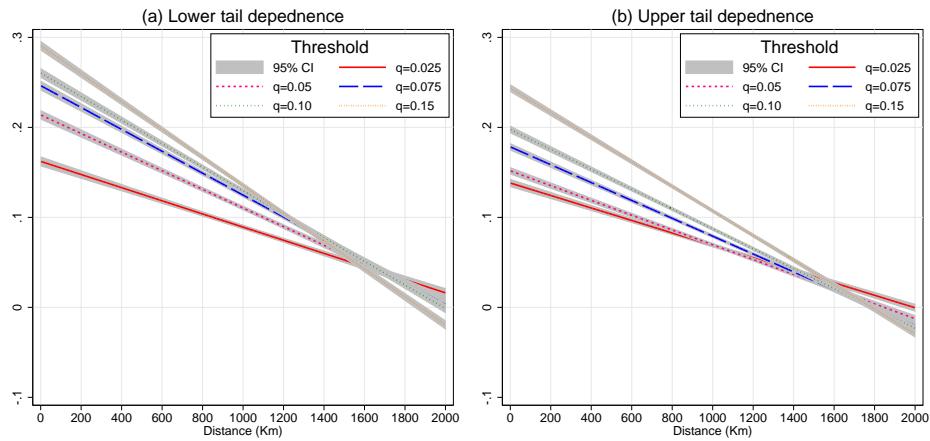
(b) Nonparametric Tail Dependence and Distance

Figure 1: Dependence in Pairwise Station Rainfalls

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 sta-

14

tions 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.[7]

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.[8]

We use a nonparametric estimator of tail dependence (Frahm et al., 2005; Patton, 2013). The estimator is given as:

$$\hat{\lambda}^U = 2 - \frac{log(1 - 2(1-q) + T^{-1}\sum_{t=1}^{T} 1\{G(Y) \le 1-q, F(X) \le 1-q\})}{log(1-q)}, q \approx 0 \quad (12)$$

---

[7]Goodwin (2001) reports a similar finding for spatial correlations between yields.

[8]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:

$$\lambda^L = \lim_{q \longrightarrow 0} P(G(Y) \le q \mid F(X) \le q)$$

$$\lambda^U = \lim_{q \longrightarrow 1} P(G(Y) > q \mid F(X) > q)$$

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 above equations 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.

$$\hat{\lambda}^U = 2 - \frac{log(T^{-1} \sum_{t=1}^{T} 1\{G(Y) \leq 1 - q, F(X) \leq 1 - q\})}{log(1 - q)}, q \approx 0 \qquad (13)$$

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.

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 the rate of decline is slower for lower values of q. We model this behavior econometrically in the following way.

$$\lambda_{ij} = \beta_1 Ln(Distance)_{ij} + \beta_2 q + \beta_3 Ln(Distance)_{ij} \times q + \alpha_i + \tau_j + \varepsilon_{ij} \qquad (14)$$

where λ_{ij} the estimated tail dependence coefficient between rainfalls measured at two stations i and j, $Ln(Distance)_{ij}$ is the distance in kilometers between the two stations 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 interac-

tion 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 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.

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

Notes: 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.

As a robustness check, we also 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. The appendix A provides details of how bivariate distributions are modeled by a copula. The paper considers the commonly used parametric families of elliptical copulas and the Archimedean copulas. Their statistical properties are also summarized in the appendix A. The copula is estimated by standard methods. Marginal distributions are estimated non-parametrically and substituted in the copula. The dependence parameters are estimated in the second step. These details

17

are also provided in the appendix A.

Table 2: Dependence in Pairwise Station Rainfalls

(a) Copula Models 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	Rotated 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)
	Rotated 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.

The Students t copula is the 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.

18

4   The Joint Distribution of Average Area Yields and Average Area Rainfall

We now turn to the association between average area yield and average area rainfall. District yields are collected from the district database of the International Crops Research Institute for the Semi-Arid Tropics ICRISAT (http://vdsa.icrisat.ac.in/vdsa-database.htm) that is compiled from various official sources. 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.

The database covers 15 major crops across 311 districts in 19 states from the year 1966-67 to 2011-12. India receives 85% of its annual rainfall during the monsoon months of June to September. A rainfall insurance contract is meaningful therefore for crops grown during this period. These are called the *kharif* season crops (June to October). In the data set, these crops are Maize, Cotton, Sorghum, Finger millet, Pigeon pea, Soybean, Pearl millet, Groundnut and Rice. 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.

The high resolution gridded rainfall data from the Indian Meteorological Department is used to construct total *kharif* season rainfall as cumulative rainfall for the months from June to October. The cumulative seasonal rainfall is transformed to standardized deviations from their long term normals.

Table 3 presents coefficients of linear and rank correlation between yield and rainfall deviations. As expected, both measures show a statistically significant posi-

tive association between yield and rainfall deviations, despite some difference in their magnitude. Figure 2 shows the scatter plot of rainfall and yield deviations. Figure 2a shows scatter plots of yield and rainfall deviations along with the linear fit.

Table 3: Linear and Rank Correlation Between Yield and Rainfall Deviations

Crops	(1) Linear correlation	(2) Rank correlation
Maize	0.023	0.004
	(0.009)	(0.01)
Cotton	0.072	0.073
	(0.012)	(0.015)
Sorghum	0.104	0.109
	(0.01)	(0.01)
Finger millet	0.107	0.086
	(0.014)	(0.015)
Pigeonpea	0.145	0.131
	(0.009)	(0.009)
Soybean	0.169	0.122
	(0.018)	(0.017)
Pearl millet	0.183	0.183
	(0.011)	(0.011)
Groundnut	0.177	0.18
	(0.01)	(0.01)
Rice	0.277	0.267
	(0.008)	(0.009)

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

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, 2014). 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:

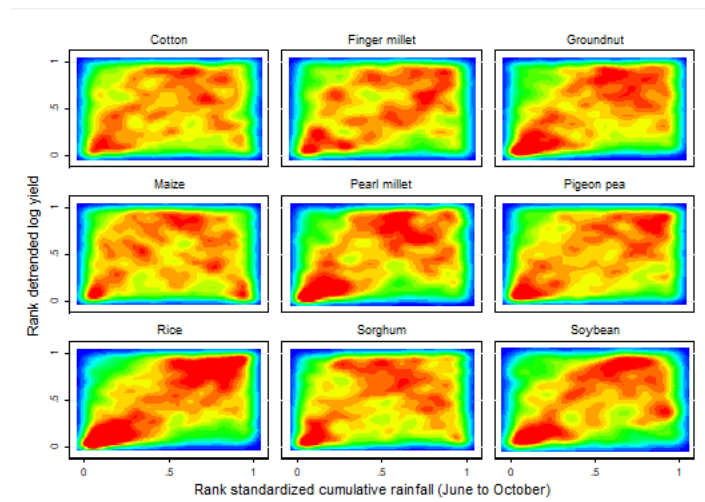$$p^L = \frac{1}{Tq} \sum_{(t=1)}^{n} 1\{U_{Yt} \leq q \mid U_{Xt} \leq q\} \tag{15}$$

$$p^U = \frac{1}{T(1-q)} \sum_{(t=1)}^{n} 1\{U_{Yt} \leq q \mid U_{Xt} \leq q\} \tag{16}$$
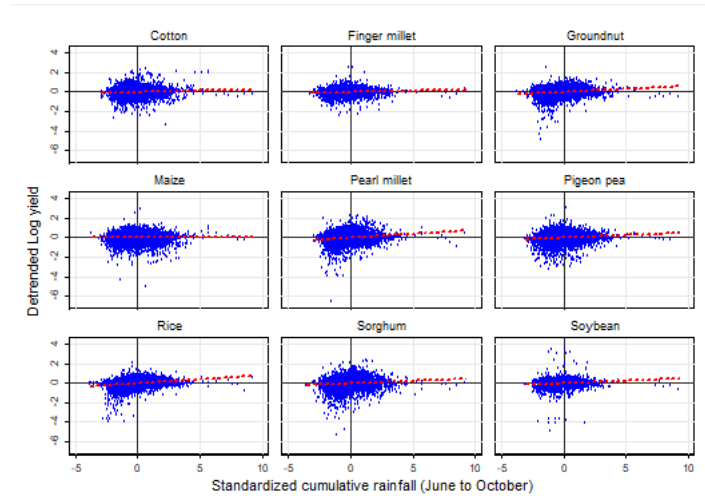
(a) Scatter Plots of Yield and Rainfall Deviations
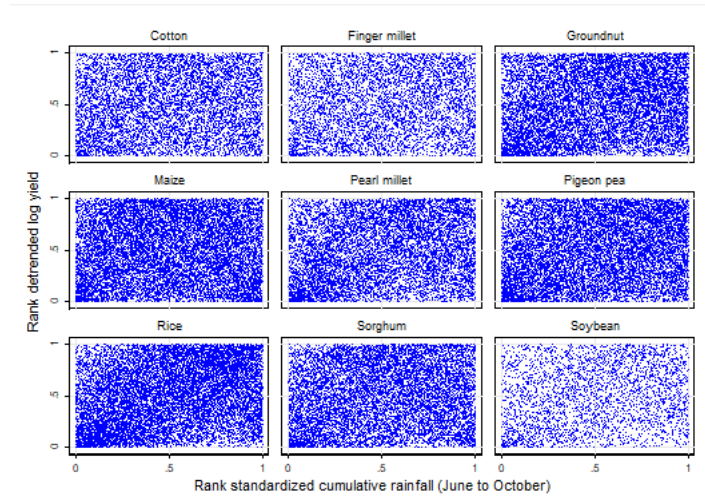
(b) Scatter Plots of Ranks of Yield and Rainfall Deviations

(c) Kernel Density Plots of Ranks of Yield and Rainfall Deviations

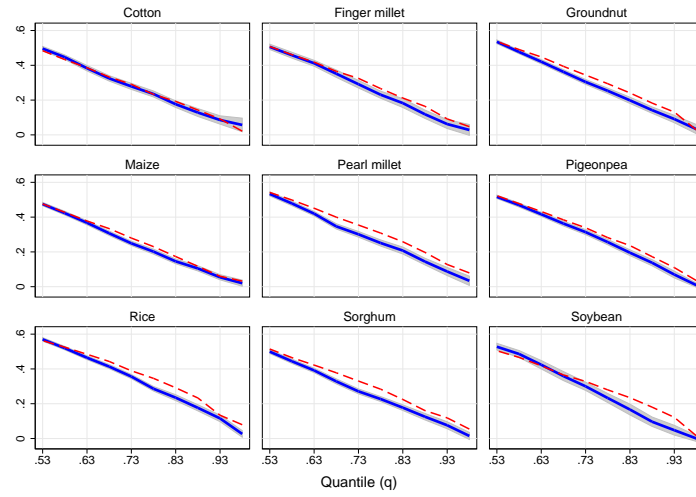Figure 2: Joint Distribution of Yield and Rainfall Deviations

21

Where U_{Yt} and U_{Xt} are the scores of Y and X based on their empirical distribution.

In figures 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 many of the 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 indicates that association between yield and rainfall index may not be 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 (panel 3a) and upper tail (panel 3b) quantile dependence plots; and the difference between the two (panel 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

22
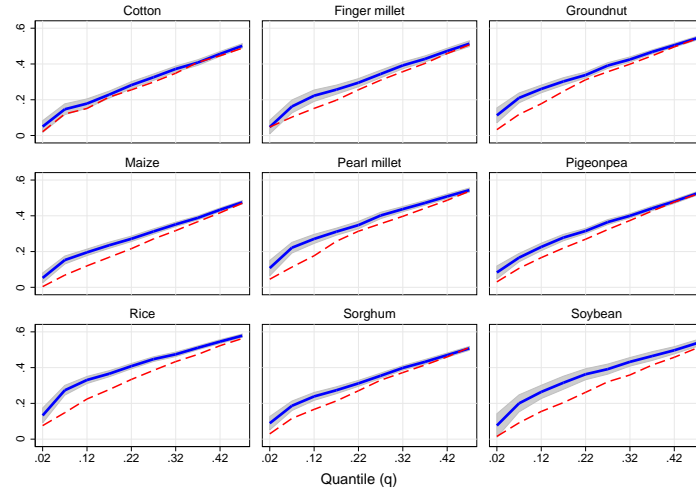
(a) Lower Tail

(b) Upper Tail
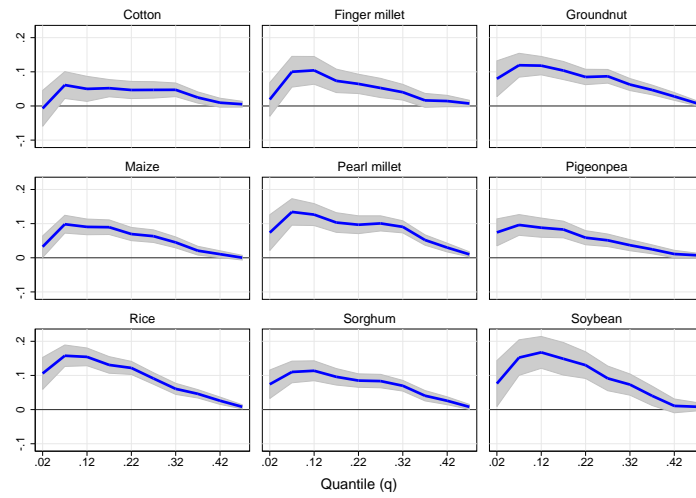
(c) Difference in Upper and Lower Tail Dependence

Figure 3: Tail Dependence at Different Quantiles

is unsuitable to model the joint distribution of yields and rainfalls.

Table 4: Log Likelihood from Different Copula Models

Crops	Gaussian	Clayton	Rotated Clayton	Plackett	Frank	Gumbel	Rotated Gumbel	Student's t
Cotton	20.4	33.7	8.3	18.8	18.4	-12.7	16.0	24.3
Finger millet	27.9	51.5	3.2	31.8	31.1	-9.8	43.3	29.9
Groundnut	183.8	254.9	56.6	175.3	171.8	92.0	235.2	196.9
Maize	3.5	31.4	-0.01	3.6	3.5	-138.5	-31.6	11.9
Pearl millet	165.8	224.7	52.7	154.9	152.0	81.2	214.7	173.6
Pigeon pea	124.8	172.6	29.1	123.9	122.9	39.8	151.9	125.8
Rice	548.3	680.9	204.2	544.4	533.5	334.4	665.8	567.6
Sorghum	68.4	125.9	10.8	56.8	55.7	-8.8	104.0	76.9
Soybean	43.6	68.4	7.5	48.5	48.1	14.0	63.2	45.8

Note: Log likelihood values estimated from copula models.

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.[9] 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.

---

[9]As mentioned earlier, the procedure used to estimate bivariate copulas is explained in Appendix A.

Table 5: Clayton Copula Model Parameter Estimates

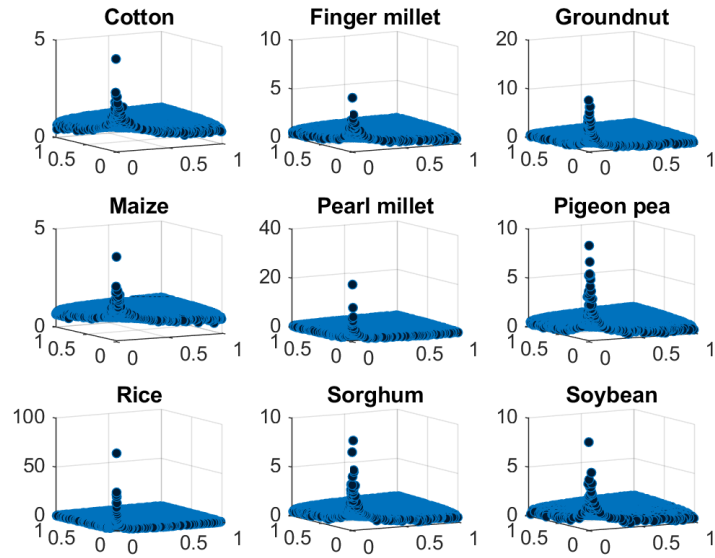| Crops	Parameter Estimates	Standard errors	Tail dependence
Cotton	0.107	0.014	0.0015
Finger millet	0.158	0.018	0.0125
Groundnut	0.260	0.013	0.0695
Maize	0.074	0.011	0.0001
Pearl millet	0.271	0.015	0.0776
Pigeon pea	0.201	0.012	0.0319
Rice	0.415	0.014	0.1878
Sorghum	0.176	0.012	0.0195
Soybean	0.246	0.025	0.0597

Figure 4: Estimated Copula Density for Crops

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 and the AIC criterion, 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 40 percent of the 274 rice growing districts. Student's t copula is the next best. Across all crops about 70% of the cases are accounted by either the Clayton copula or the Student's t copula. These findings clearly indicate nonlinearity in association between weather and yield risk and have implications for the demand for insurance and thus, its design.

Table 6: Percent Districts with Best Fit Copulas

Crops	Gaussian	Clayton	Rotated Clayton	Plackett	Frank	Gumbel	Rotated Gumbel	Student's t	Total
Cotton	12	37	10	7	7	3	2	44	122
	(9.84)	(30.33)	(8.2)	(5.74)	(5.74)	(2.46)	(1.64)	(36.07)	(100)
Finger millet	2	28	5	3	5	0	4	24	71
	(2.82)	(39.44)	(7.04)	(4.23)	(7.04)	(0)	(5.63)	(33.8)	(100)
Groundnut	9	77	8	15	8	3	11	57	188
	(4.79)	(40.96)	(4.26)	(7.98)	(4.26)	(1.6)	(5.85)	(30.32)	(100)
Maize	17	68	13	21	8	4	6	113	250
	(6.8)	(27.2)	(5.2)	(8.4)	(3.2)	(1.6)	(2.4)	(45.2)	(100)
Pearl millet	3	78	7	9	6	3	6	45	157
	(1.91)	(49.68)	(4.46)	(5.73)	(3.82)	(1.91)	(3.82)	(28.66)	(100)
Pigeon pea	12	88	21	15	16	6	7	53	218
	(5.5)	(40.37)	(9.63)	(6.88)	(7.34)	(2.75)	(3.21)	(24.31)	(100)
Rice	13	110	10	12	24	8	34	63	274
	(4.74)	(40.15)	(3.65)	(4.38)	(8.76)	(2.92)	(12.41)	(22.99)	(100)
Sorghum	6	73	8	14	8	2	7	80	198
	(3.03)	(36.87)	(4.04)	(7.07)	(4.04)	(1.01)	(3.54)	(40.4)	(100)
Soybean	0	24	2	2	6	0	2	5	41
	(0)	(58.54)	(4.88)	(4.88)	(14.63)	(0)	(4.88)	(12.2)	(100)
Total	74	583	84	98	88	29	79	484	1519
	(4.87)	(38.38)	(5.53)	(6.45)	(5.79)	(1.91)	(5.2)	(31.86)	(100)

Note: Row percentages in parenthesis.

5   Implications for Rainfall Insurance

5.1   Basis Risk

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.

Assume that a farmer's yield q is a random variable with distribution function $g(q)$. The payout from one unit of rainfall based insurance contract is given by

$$I = Max\{\hat{R} - R, 0\} \tag{17}$$

where R is the rainfall index with distribution function $h(R)$ and $\hat{R}$ is the rainfall threshold set by the insurance selling agency. Lower is the threshold, greater is the deductible in the insurance payouts. The contract trigger's payouts only if actual rainfall falls below $\hat{R}$. The implicit assumption in offering such a contract is that farmers' yield and the rainfall index are correlated such that in periods of low rainfall crop yields will also be lower. The actuarially fair price P of such a contract is just the expectation

of I.

$$P = \int_0^{\hat{R}} (\hat{R} - R)h(R)dR \tag{18}$$

The net profits of a farmer purchasing a rainfall insurance contract can be written as

$$\pi = q + \alpha(I - mP) \tag{19}$$

where α is the number of insurance units purchased and m is the mark-up over actuarially fair insurance. We want to find the optimal value of α that maximizes the expected indirect utility.

$$Max_\alpha \eta(\alpha) \equiv Eu(q + \alpha(I - mP)) \tag{20}$$

where $u(.)$ is the utility function of the farmer with $u'(.) > 0$ and $u''(.) < 0$. Starting from no insurance, the increment to expected utility because of insurance is given by

$$\eta'(\alpha)\mid_{\alpha=0} = Eu'(q)(I - mP) \tag{21}$$

Or,

$$\eta'(\alpha)\mid_{\alpha=0} = \iint u'(q)(I - mP)h(R \mid q)g(q)dRdq \tag{22}$$

where $h(R \mid q)$ is the density of rainfall conditional on yield. This can be rewritten as

$$\eta'(\alpha)\mid_{\alpha=0} = \int u'(q)[\int (I - mP)h(R \mid q)dR]g(q)dq \tag{23}$$

The term inside the square brackets is nothing but $E(I \mid q) - mP$. Hence we have

$$\eta'(\alpha)\mid_{\alpha=0} = Eu'(q)(E(I \mid q) - mP) \tag{24}$$

From the above it can be seen that the insurance demand is zero if $E(I \mid q) \leq mP$, for all values of q. This result is a restatement of a theorem in Clarke (2016). Clarke defines

$$\kappa(q) = \frac{E(I \mid q)}{mP} = \frac{\text{Expected claim payment over yield distribution}}{\text{Commercial premium}} \tag{25}$$

The ratio basically reflects the average amount a farmer gets back as claims per dollar paid as commercial premium. He shows that if $\kappa(q) \leq 1$ over the entire yield

distribution then $\alpha = 0$, for a risk-averse individual. In our model, this result follows from (24).

Clarke et al. (2012) use the payout structure of 270 weather based crop insurance products sold to Indian farmers in one state in one year and combine it with historical data to simulate payouts over the period 1999-2007. Their work finds the ratio $\kappa(q)$ to be almost flat over the entire yield distribution. The ratio is below 1 for most values of q and is barely above 1 for very low levels of q. It follows then that the basis risk in these contracts is so large that it would be optimal not to purchase them. Morsink et al. (2016) proposed that the ratio defined in (25) should be used as a measure of reliability of weather insurance contracts and called it the catastrophic performance ratio. They further suggested that the ratio could be used to "improve the quality of products, protect consumers, and reduce reputational risk".

We use the catastrophe performance ratio to examine how tail dependence matters to basis risk. A hypothetical rainfall insurance contract of the form in (17) is considered. The payoffs are simulated using 10,000 draws of rainfall and yield from a Gaussian copula and from a copula exhibiting lower tail dependence. The correlation between the two variables is held constant across the two copulas. The comparison of the performance ratio across the two copulas is, then, revealing about the effect of tail dependence.

The exact procedure is as follows. For both these copulas, the marginal distribution of yield and rainfall are assumed to be normal with a mean of 2000 and standard deviation of 300. In the last section, the best fit copula to the joint distribution of rice yields and rainfall was found to be the Clayton copula with a parameter of 0.42.

The marginal distributions are combined in a Clayton copula with a parameter of 0.42 to generate 10,000 observations of yield and rainfall. These observations are used to compute the insurance and payoffs. The linear correlation between rainfall and yield draws from the Clayton copula is combined with the assumed marginal distributions to generate another 10,000 observations from a bivariate normal distribution.

Thus, we have two empirical joint distributions such that they share the same marginal distributions and the same correlation between rainfall and yield. The only difference is that yield and rainfall index simulated from Clayton copula exhibit lower tail dependence, while the other does not.
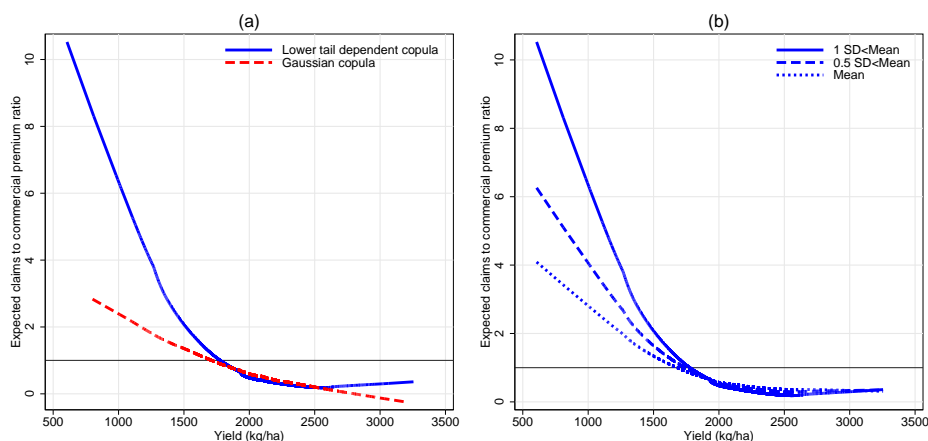
Figure 5: Expected Claims to Commercial Premium Ratio: All India

Figure 5a plots the nonparametrically estimated relationship between claims to commercial premium ratio and yield from the simulated data, i.e.

$$I(q) = E\left(\frac{Max\{\hat{R} - R, 0\}}{mP} \mid q\right) \tag{26}$$

where the insurance contract parameter $\hat{R}$ is assumed to be one standard deviation below the mean rainfall and m is assumed to be 1.56 times the actuarially fair

premium.[10] At this premium level, the catastrophic performance ratio is below 1 for the rainfall insurance contracts considered by Clarke et al. (2012). This is not true, however, for the payouts from rainfall contracts in Figure 5a The ratio from the normal distribution and from Clayton Copula are above 1 for low output levels. There is, however a substantial divergence between the normal distribution and the Clayton copula at these low output levels. The catastrophic performance ratio is substantially higher for the Clayton copula. Thus, by the measures proposed by Morsink et al. (2016), accounting for tail dependence markedly reduces basis risk.

Figure 5b plots the Clayton copula based catastrophic performance ratio for different levels of the deductible. $\hat{R}$ is chosen to be either the mean, or 0.5 standard deviation below the mean or 1 standard deviation below the mean. It can be seen that as the deductible rises (i.e., $\hat{R}$ falls) so does the basis risk. Catastrophic insurance carries the least basis risk.

5.2 Optimal Insurance

In figures 5a and 5b, the Clarke condition that is sufficient to ensure zero insurance demand is not met. $(E(I \mid q) - mP)$ is above 1 for low realizations of output but below 1 for high realizations of output. This does not mean that insurance demand is necessarily positive. That depends on the evaluation of equation (24) which depends

---

[10]Clarke (2016) based on 270 weather based crop insurance products sold to Indian farmers report's that a markup greater than 1.56 times the fair premium will lead to no demand for rainfall insurance.

on the extent of risk aversion. (24) can also be written as

$$\eta^{'}(\alpha)\mid_{\alpha=0}= Cov(u^{'}(q), E(I\mid q)) - (m-1)PEu^{'}(q) \qquad (27)$$

Risk aversion and the expected shape of the regression $E(I\mid q)$ guarantees the first term to be positive. When insurance is actuarially fair, the second term is zero and it is optimal for farmers to buy some insurance. When $m > 1$, the answer would depend on risk aversion and the mark-up over the fair premium.

To investigate these issues, we consider data from two districts in India, Mahabubnagar and Anantapur, that have been heavily researched for the extent of local risk sharing (e.g., Townsend, 1994). These districts are characterized by dependence on rainfed agriculture and vulnerability to droughts. Households in these districts have also been recently surveyed for their risk aversion using Binswanger type lotteries (Binswanger, 1980; Cole et al., 2013) and we use those estimates.

Using the procedures in appendix A, a best fit copula model is selected for rice yields and rainfall in each of the two districts. Table 7 displays the results. Unlike the exercise that generated figures 5a and 5b, we do not assume marginal distributions of rainfall and yield to be normal. Instead, we consider various parametric form and choose the best fit functional form (Table 7).[11] Rainfall is log-normal in both districts. Yield follows a Weibull distribution in Anantapur and follows a gamma distribution in Mahabubnagar. Plots of estimated parametric distributions against the observations are presented in appendix B.

---

[11]The distributions that were considered were Gamma, Weibull, log-normal and Gumbel. All of these are two-parameter distributions and the parameters were estimated by maximum likelihood procedures. The distribution that maximizes the log likelihood is picked as the marginal distribution.

Table 7: Best Fit Parametric Marginal Distributions and Copula Models

		Parameter estimates	
(a) *Fitted marginal distribution of cumulative rainfall*			
Anantapur	Log normal	6.06	0.28
Mahabubnagar	Log normal	6.38	0.24
(b) *Fitted marginal distribution of de-trended recentered yield*			
Anantapur	Weibull	2961.8	15.0
Mahabubnagar	Gamma	126.7	21.3
(c) *Copula model of joint distribution of yield and rainfall*			
Anantapur	Rotated Gumbel	1.187	
Mahabubnagar	Clayton	1.127	

These marginal distributions are combined in the appropriate copula (as in Table 7) to generate 10,000 observations of yield and rainfall. These observations are used to compute the insurance and payoffs. The linear correlation between these rainfall and yield draws is combined with the selected marginal distributions to generate another 10,000 observations from a Gaussian copula.

Figures 6a and 6b show the catastrophe performance ratios for these districts. These pictures are very much like Figures 5a and 5b. Once again, basis risk is much lower relative to a Gaussian copula. Further, basis risk falls with a larger deductible.

Next we move to an evaluation of equation (24). For a constant risk aversion utility function with parameter γ (24) becomes

$$\eta^{'}(\alpha) \mid_{\alpha=0} = Eq^{-\gamma}(E(I \mid q) - mP) \tag{28}$$

Based on the work of Cole et al. (2013), the risk aversion parameter is assumed to be 0.57. The above equation can be used to compute the mark-up over the actuarially fair premium for which insurance demand is positive. From the results displayed in Figure 7, it can be seen that the m that extinguishes insurance demand is higher for a
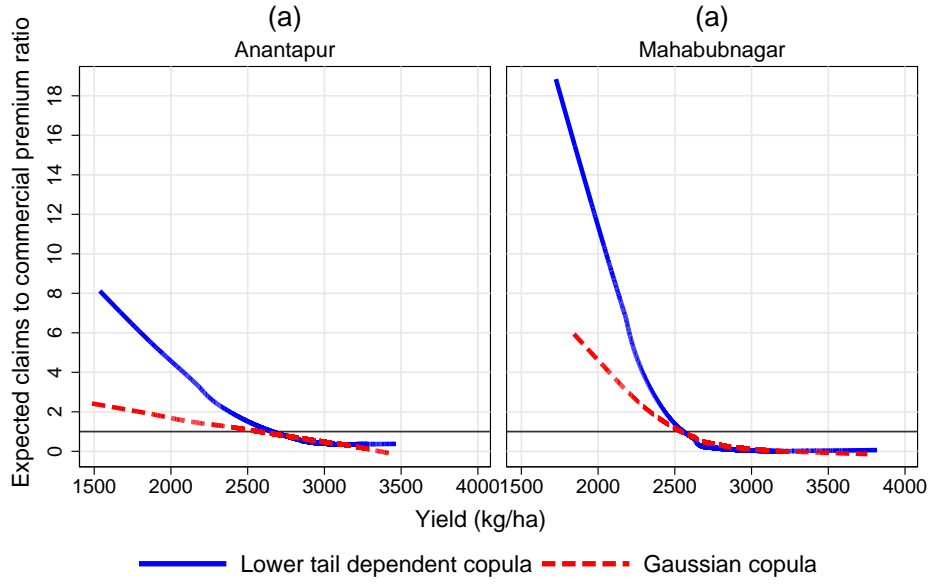
(a) Expected Claims to Premium Ratio With and Without Tail Dependence
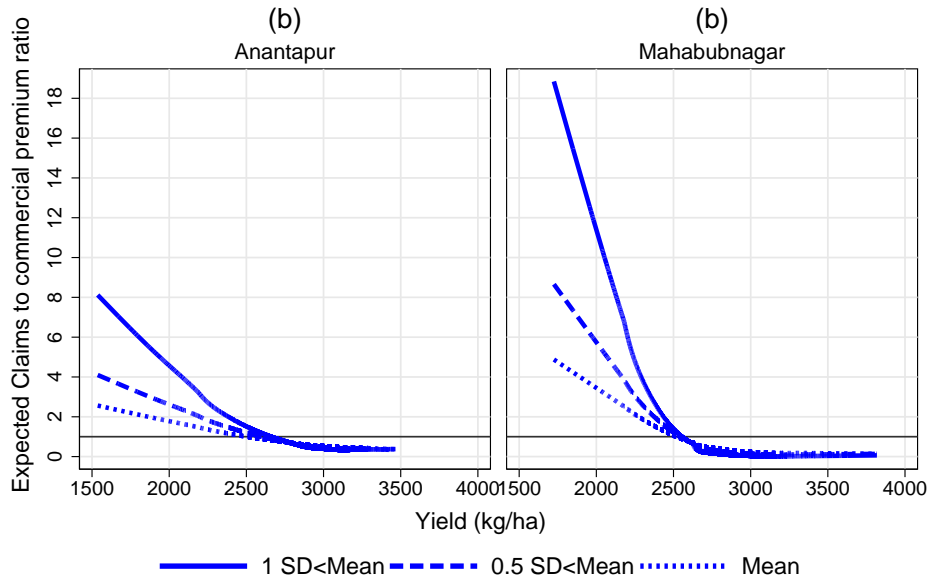
(b) Expected Claims to Premium Ratio With Different Trigger Thresholds

Figure 6: Expected Claims to Premium Ratio for Two Districts of Andhra Pradesh

tail-dependent copula as compared to a Gaussian copula. This is simply a reflection of the lower basis risk that comes with lower tail dependence. A second finding of Figure 7 is that the maximum mark-up for which insurance demand is positive is higher when the deductible is larger. This again is a reflection of the earlier figure 6b that showed the basis risk is lowest in contracts with the smallest rainfall threshold.
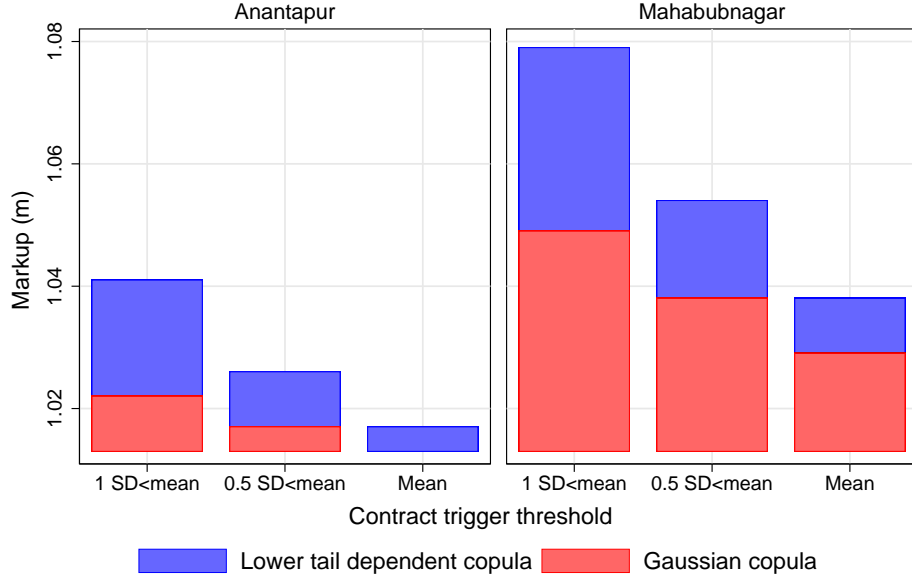

Figure 7: Markups at Which Demand for Insurance Cover is Zero

For the constant relative risk aversion utility function, the optimal insurance units can be solved from

$$\eta^{'}(\alpha) = E(q + \alpha(I - mP))^{-\gamma}(I - mP) = 0 \qquad (29)$$

The payouts and the premium that were simulated to compute the catastrophe performance ratios can also be used to evaluate (29). We continue to use $\gamma = 0.57$. Optimal insurance cover is computed with and without tail dependent yield and rainfall distribution and for insurance contracts that vary according to the index threshold

36

value that triggers payout. The results are displayed in Figure 8 where the computations assume $m = 1$. What is noteworthy about the results is that the optimal insurance cover is much larger with a tail dependent copula than with a Gaussian copula. This is consistent with the lower basis risk with a tail dependent copula.
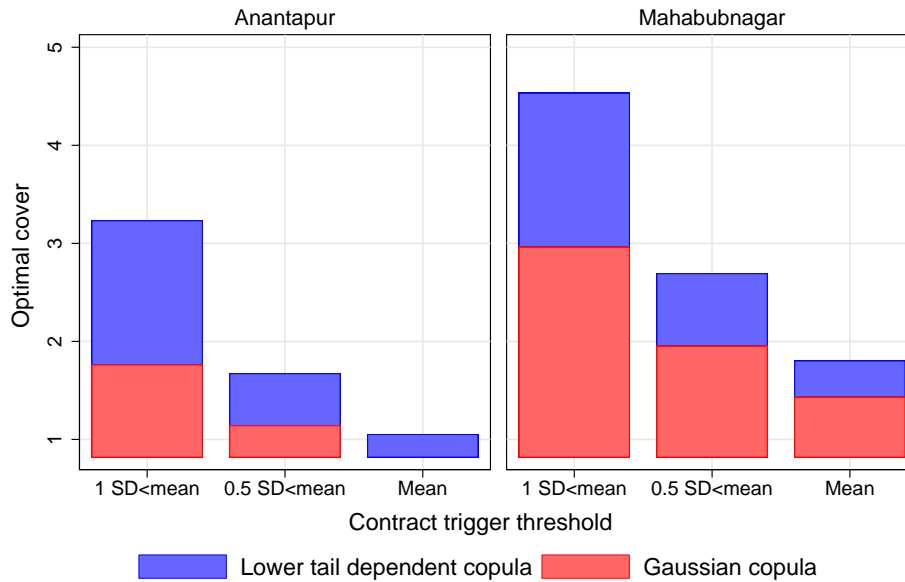

Figure 8: Optimal Cover for Actuarially Fair Contract Under Different Thresholds

The fact that contracts with the lowest threshold (highest deductible) have the lowest basis risk and the greatest demand for insurance, does not, however, mean that farmers necessarily prefer these contracts to all others. Figure 9 evaluates the expected utility for the optimal levels of insurance for actuarially fair premiums. This shows that the optimal threshold is 0.5 standard deviation below the mean for Anantapur while it is the mean yield for Mahabubnagar. For the given risk aversion parameter, it is optimal to accept higher basis risk in exchange for a greater insurance protection. However, if insurance is actuarially unfair, then as Figure 8 showed, it is more likely that the contracts with the least basis risk are favored by farmers.
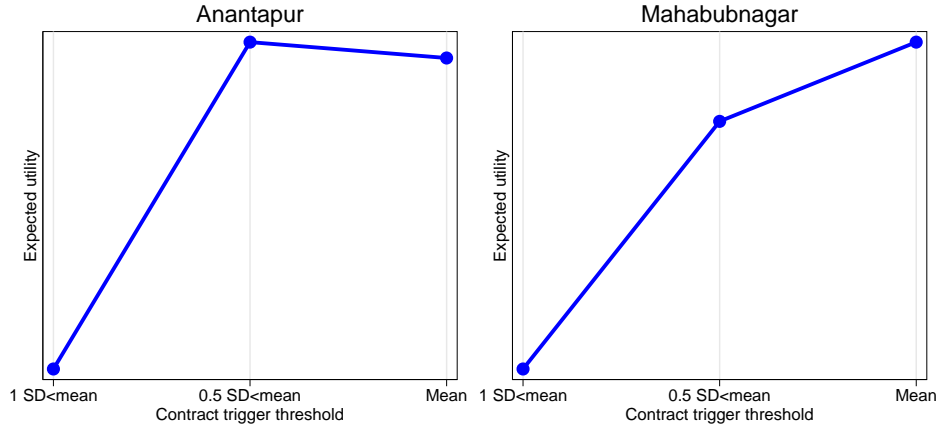
Figure 9: Expected Utility with Optimal Cover for Different Thresholds

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. The 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.

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.

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.

Appendix

A   Copula Estimation

We use copula functions 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 $C(u,v) : [\theta, 1]^2 \longrightarrow [0,1]$ such that

$$F(Y, X) = P[G(Y) \leq G(y), F(X) \leq F(x)] \tag{30}$$

$$F(Y, X) = C(G(Y), F(X); \theta) \tag{31}$$

Where θ represents the strength of dependence and $G(.)$ and $F(.)$ are the marginal distribution functions of Y and X respectively. The joint probability density function can be expressed as:

$$c(G(Y), F(X); \theta) = \frac{\partial C(G(Y), F(X); \theta)}{\partial G(Y) \partial F(X)} = C(G(Y), F(X); \theta)g(Y)f(X) \tag{32}$$

Sklar (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 table 8.

Table 8: Some Common Copula Models

Copula models	Functional forms	Dependence parameter	Parameter space	Lower tail dependence	Upper tail dependence
Gaussian	$\Phi_{\Sigma}(\Phi^{-1}(u), \Phi^{-1}(v); \rho)$	ρ	$(-1, 1)$	0	0
Clayton	$(u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}}$	θ	$(0, \infty)$	$2^{-\frac{1}{\theta}}$	0
Rotated Clayton	Same as Clayton with $1-u$ and $1-v$	θ	$(0, \infty)$	0	$2^{-\frac{1}{\theta}}$
Plackett	$\frac{1+(\theta-1)(u+v)-\sqrt{[1+(\theta-1)(u+v)]^2-4\theta(\theta-1)uv}}{2(\theta-1)}$	θ	$(0, \infty)$	0	0
Frank	$-\frac{1}{\theta}log\left(1 + \frac{(exp^{-\theta u}-1)(exp^{-\theta u}-1)}{(exp^{-\theta}-1)}\right)$	θ	$(-\infty, \infty)$	0	0
Gumbel	$\exp\left\{-(-logu^{\theta} - logv^{\theta})^{\frac{1}{\theta}}\right\}$	θ	$(1, \infty)$	0	$2 - 2^{-\frac{1}{\theta}}$
Rotated Gumbel	Same as Gumbel with $1-u$ and $1-v$	θ	$(1, \infty)$	$2 - 2^{-\frac{1}{\theta}}$	0
Student's t	$t_{\nu,\Sigma}(t_{\nu}^{-1}(u), t_{\nu}^{-1}(v); \rho)$	ρ, ν	$(-1, 1) \times (2, \infty)$	$2 \times t_{\nu+1}\left(-\sqrt{(\nu+1)}\frac{\sqrt{(1-\rho)}}{\sqrt{(1+\rho)}}\right)$	$2 \times t_{\nu+1}\left(-\sqrt{(\nu+1)}\frac{\sqrt{(1-\rho)}}{\sqrt{(1+\rho)}}\right)$

Note: Table presents some common parametric copula models with their functional forms, parameter spaces and the expression for tail dependence coefficient implied by the specific copula model.

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 (AIC). 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:

$$L(\theta) = \sum_{i=1}^{N} LnC(\hat{U}_{X_i}, \hat{U}_{Y_i}; \theta) \qquad (33)$$

Where $\hat{U}_{X_i}$ and $\hat{U}_{Y_i}$ are the nonparametrically 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 (Trivedi and Zimmer, 2007; Patton, 2013). All copula models and tail dependence statistics are estimated using Patton's (2013) procedure and MATLAB codes.
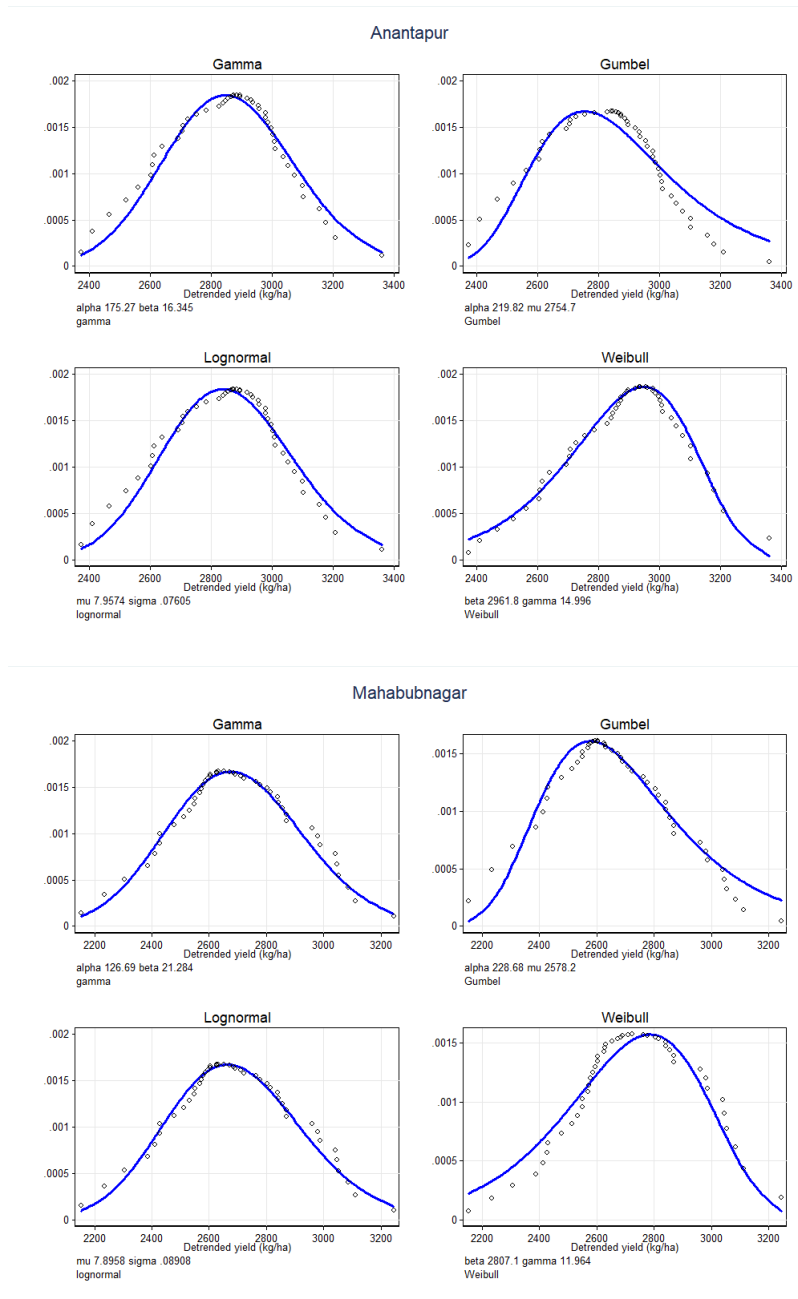
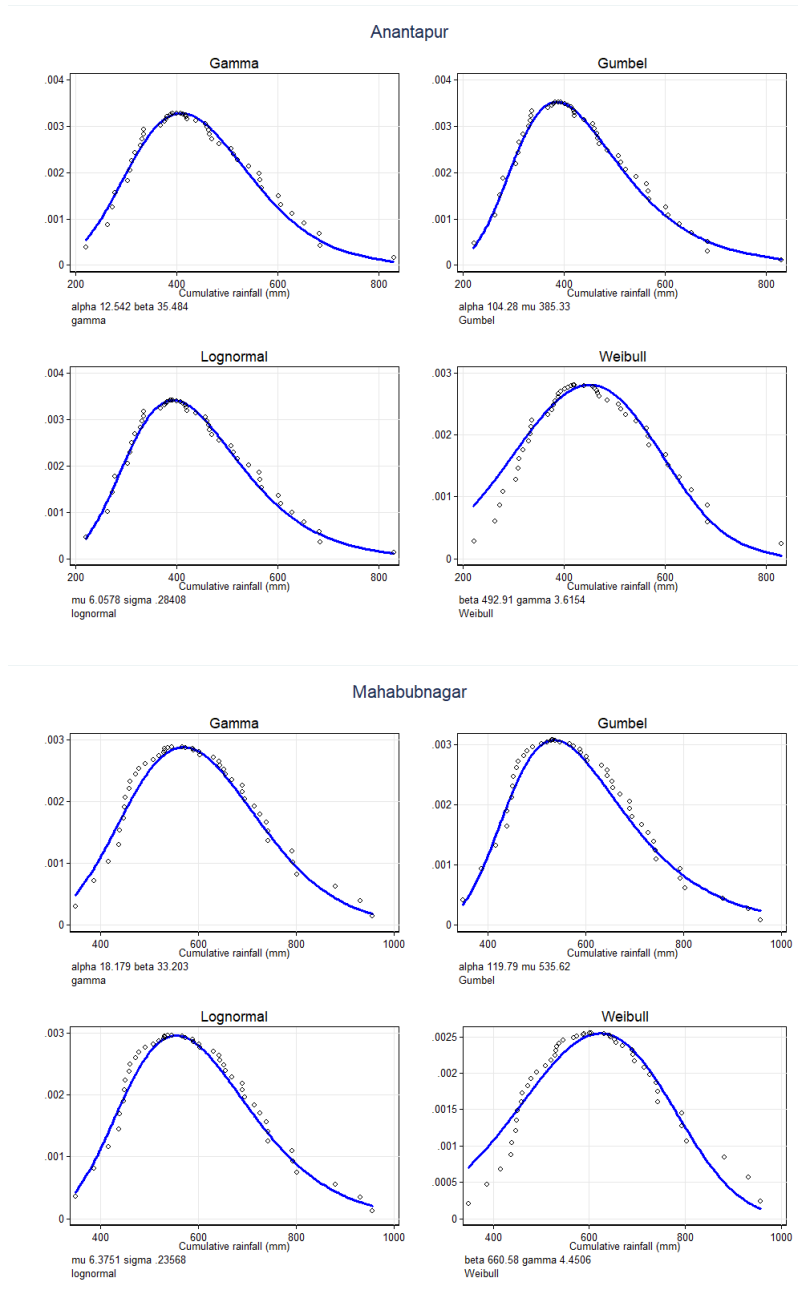# B Estimated Marginal Densities

Figure 10: De-trended yield

Figure 11: Cumulative Seasonal Rainfall

References

Aghakouchak, A., Ciach, G., and Habib, E. Estimation of tail dependence coefficient in rainfall accumulation fields. *Advances in water Resources*, 33(9):1142–1149, 2010.

Bardhan, P. and Udry, C. *Development microeconomics*. OUP Oxford, 1999.

Binswanger, H. P. Attitudes toward risk: Experimental measurement in rural India. *American journal of agricultural economics*, 62(3):395–407, 1980.

Binswanger-Mkhize, H. P. Is there too much hype about index-based agricultural insurance? *Journal of Development studies*, 48(2):187–200, 2012.

Carter, M., de Janvry, A., Sadoulet, E., and Sarris, A. Index-based weather insurance for developing countries: A review of evidence and a set of propositions for up-scaling. *Development Policies Working Paper*, 111, 2014.

Carter, M. R. and Barrett, C. B. The economics of poverty traps and persistent poverty: An asset-based approach. *The Journal of Development Studies*, 42(2):178–199, 2006.

Chantarat, S., Mude, A. G., Barrett, C. B., and Carter, M. R. Designing index-based livestock insurance for managing asset risk in northern Kenya. *Journal of Risk and Insurance*, 80(1):205–237, 2013.

Chen, X. and Fan, Y. Estimation of copula-based semiparametric time series models. *Journal of Econometrics*, 130(2):307–335, 2006.

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

Clarke, D. J., Clarke, D., Mahul, O., Rao, K. N., and Verma, N. Weather based crop insurance in India. 2012.

Cole, S., Giné, X., Tobacman, J., Topalova, P., Townsend, R., and Vickery, J. Barriers to household risk management: Evidence from India. *American Economic Journal: Applied Economics*, 5(1):104–35, 2013.

Cole, S., Stein, D., and Tobacman, J. Dynamics of demand for index insurance: Evidence from a long-run field experiment. *American Economic Review*, 104(5):284–90, 2014.

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

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

Du, X., Hennessy, D. A., Feng, H., and Arora, G. Land resilience and tail dependence among crop yield distributions. *American Journal of Agricultural Economics*, 100(3): 809–828, 2017.

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

Fafchamps, M. *Rural poverty, risk and development*. Edward Elgar Publishing, 2003.

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

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

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

Goodwin, B. K. Agricultural policy analysis: The good, the bad, and the ugly. *American Journal of Agricultural Economics*, 97(2):353–373, 2014.

Goodwin, B. K. and Hungerford, A. Copula-based models of systemic risk in US agriculture: Implications for crop insurance and reinsurance contracts. *American Journal of Agricultural Economics*, 97(3):879–896, 2015.

Hill, R. V., Robles, M., and Ceballos, F. Demand for a simple weather insurance product in India: Theory and evidence. *American Journal of Agricultural Economics*, 98(4):1250–1270, 2016.

Jensen, N. and Barrett, C. Agricultural index insurance for development. *Applied Economic Perspectives and Policy*, 39(2):199–219, 2017.

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

Joe, H. *Dependence modeling with copulas*. CRC Press, 2014.

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

Mobarak, A. and Rosenzweig, M. Selling formal insurance to the informally insured. 2012.

Mobarak, A. M. and Rosenzweig, M. R. Informal risk sharing, index insurance, and risk taking in developing countries. *American Economic Review*, 103(3):375–80, 2013.

Morduch, J. Between the state and the market: Can informal insurance patch the safety net? *The World Bank Research Observer*, 14(2):187–207, 1999.

Morduch, J. *Consumption smoothing across space: Testing theories of risk-sharing in the ICRISAT study region of South India*. Oxford: Oxford University Press, 2005.

Morsink, K., Clarke, D., and Mapfumo, S. How to measure whether index insurance provides reliable protection. 2016.

Nelsen, R. An introduction to copulas. *Lecture Notes in Statistics. New York: Springer*, 2006.

Patton, A. Copula methods for forecasting multivariate time series. In *Handbook of economic forecasting*, volume 2, pages 899–960. Elsevier, 2013.

Ramaswami, B. and Roe, T. L. Aggregation in area-yield crop insurance: The linear additive model. *American Journal of Agricultural Economics*, 86(2):420–431, 2004.

Rosenzweig, M. R. and Binswanger, H. P. Wealth, weather risk, and the composition and profitability of agricultural investments. *American Journal of Agricultural Economics*, 103(416):56–78, 1993.

Sklar, A. Fonctions de repartition an dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8:229–231, 1959.

Townsend, R. M. Risk and insurance in village India. *Econometrica: Journal of the Econometric Society*, pages 539–591, 1994.

Trivedi, P. K. and Zimmer, D. M. Copula modeling: An introduction for practitioners. *Foundations and Trends in Econometrics*, 1(1):1–111, 2007.