Pasture, Rangeland, and Forage – Rainfall Index: Refining the Index

Wilson Collins, North Carolina State University, wfcollin@ncsu.edu

Selected Paper prepared for presentation at the 2022 Agricultural & Applied Economics Association Annual Meeting, Anaheim, CA; July 31-August 2

Pasture, Rangeland, and Forage - Rainfall Index Insurance: Refining the Index

May 18, 2022

Abstract

This paper examines Pasture Rangeland Forage Rainfall Insurance, which is an insurance mechanism that provides coverage for land used for grazing and haying forage crops. The contract is based upon a rainfall index, and this paper looks at improvements to the index through modifications and additional variables using basis risk of the index as a measure of effectiveness. Additionally, it takes an in depth look at the policy implications of basis risk improvement and how the structure of the mechanism may need to evolve with a warming climate.

1 Introduction

Pasture, Rangeland, Forage Rainfall Insurance (PRF-RI) is a USDA insurance program that provides protection for perennial forage used for grazing or haying. Why does PRF land matter? PRF land is used to graze or grow feed for livestock. In the event of crop spoilage, livestock owners must purchase feed on the market, which can be expensive. For this reason, farmers and ranchers value risk management tools to insure their PRF land.[1] PRF-RI insurance has been growing year over year popularity since the program's initiation in 2007, and as is true with most federal crop insurance products, the program is subsidized (about 54% of Total Premium). In 2021, over 53,000 policies were sold covering over 238,000 acres for a total liability of $4.4 billion.[2]

The product operates by insuring a rainfall index based on local historic precipitation on a roughly 17 square mile gird (0.25 degrees), but why an index? The yields of forage crops are inherently difficult to quantify; forage is often grazed by livestock and not counted at harvest or when sold. This nature of forage lands prevents standard crop insurance mechanisms, which rely on quantifiable yields, from being an option.[3]

To provide some form of a risk management product to forage lands, PRF-RI insures against rainfall uncertainty. As an index insurance, indemnities are triggered only by aggregate monthly rainfall being below the threshold, which varies based on location and desired level of coverage. Each grid is allocated and "Expected Grid Index" based on location and time of year. The Expected Grid Index is multiplied by the desired coverage

[1] Matthew Diersen. "Choosing Pasture, Rangeland, Forage Rainfall Index (PRF-RI) Insurance Coverage". In: *AgWeb* (2015). DOI: `https://www.agweb.com/article/choosing-pasture-rangeland-forage-rainfall-index-prf-ri-insurance-coverage-university-news-release`

[2] USDA RMA. *Pasture, Rangeland, and Forage.* 2019. URL: `https://www.rma.usda.gov/en/News-Room/Frequently-Asked-Questions/Pasture-Rangeland-Forage` (visited on 04/11/2019)

[3] Jody Campiche and JJ Jones. "Pasture, Rangeland, Forage Insurance Program". In: *Oklahoma Cooperative Extension Service* (2013)

level (70%, 75% , 80%, 85% or 90%) to determine the threshold. By insuring only a rainfall index, this insurance contract is a more limited form of risk management compared to most crop insurance programs. The highest level of risk management is revenue insurance, which includes price risk (low market price) and yield risks (pests, drought, temperatures etc.). Yield insurance covers a smaller umbrella of risk (yield but not price), but still guarantees production or compensation. As an index insurance, PRF-RI Insurance carries basis risk: the risk of imperfect correlation between the index and the losses. An insured party can incur total crop loss, but if there is a sufficient amount of rain such that the insurance contract does not trigger then they will not be compensated. Additionally, there could be no crop loss, but still a payment due to the level of rainfall. The nature of this mechanism suggest an interesting perception of the risk to PRF land that will be discussed further in the next section.

This insurance product is rated using historic precipitation data collected from the Climate Prediction Center (CPC) of the National Oceanic and Atmospheric Administration (NOAA), as well as local productivity data. Policy owners must indicate the intended use of their land (haying or grazing) and provide documentation. The premium and indemnity of an individual policy depends on expected grid index, productivity factor, and intended use of land.[23] Productivity factor is a multiple of county base production between 60% and 150%. Policy intervals two months in length and are available through the entire year.[32] For example, a farmer may insure at a 150% productivity level and a 70% grid base for a January and February. An indemnity in the first or second month will pay 150% of the county base if rainfall is below 70% of the expected grid index.

The design of this contract, an index based solely on aggregate monthly rainfall, brings another question to mind: what is the relationship between forage yields and the index? This imperfect correlation is knows as basis risk for an index based insurance product. For the purpose of this paper, it can be divided into two ostensibly named categories: spatial basis risk and dependence basis risk. The former represents imperfect correlation between observed rainfall (observed from the perspective of the RMA) and the real rainfall at the field's location. Assigned rainfall values are spatially smoothed from the nearest NOAA observation stations and aggregated at grid level. They can vary from actual rainfall at field level. The latter risk, dependence risk, represents risk derived from imperfect correlation between the rainfall index and forage yields.

The purpose of this study is to evaluate the basis risk of the PRF mechanism and suggest an alternate design of the index. A high basis risk means the product is not useful as a purely risk management tool, which has been shown to reduce demand for the product and could lead to a misallocation of taxpayer funds (through the subsidy).[4,5,6] The reduction in demand due to high basis risk may diminish another goal of the product: to offer benefits to forage producers and livestock owners who reap less rewards of subsidized products available to other crops.[7]

2 The PRF-RI Mechanism and Prior Work

In this section, I describe the PRF-RI mechanism and it's use as a meaningful risk management tool. As previously mentioned, PRF-RI is an index insurance. Index insurances have benefits over conventional insurance products in that they are considered to be free of moral hazard issues.[8] However, moral hazard may become present again if the

[4] Hans Peter Binswanger. "Is There Too Much Hype about Index-based Agricultural Insurance?" In: *Journal of Development Studies* 48.2 (2012), pp. 187–200

[5] Ghada Elabed et al. "Managing basis risk with multiscale index insurance". In: *Agricultural Economics* 44 (2013), pp. 419–431

[6] Daniel J. Clarke. "A Theory of Rational Demand for Index Insurance". In: *American Economic Journal: Microeconomics* 8 (2016), pp. 283–306

[7] Joshua G. Maples, B. Wade Brorsen, and Jon T. Biermacher. "The Rainfall Index Annual Forage Pilot Program as a Risk Management Tool for Cool-Season Forage". In: *Journal of Agricultural and Applied Economics* 48.1 (2016), pp. 29–51

[8] Harold Halcrow. "Actuarial Structures for Crop Insurance". In: *Journal of Farm Economics* 31.3 (1949)

insurance in rated incorrectly. For example, Nadolnyak and Vendevov examine the relationship between PRF-RI and El Nino forecasts. They find intertemporal adverse selection is present when insurers do not account for the forecasts and the insureds do.[9] As mentioned previously, index insurances carry basis risk, which I define components as spatial risk and dependence risk. Both spatial[7,10,11] and dependence[7,10,12] risk have been evaluated at length as important aspects of the effectiveness of the PRF-RI mechanism as a risk management tool.

The findings of spatial dependence risk are relatively consistent. Maples, Brorsen, and Biermacher find a correlation of roughly 0.95 (5% spatial basis risk) between rainfall index and actual index in Burneyville, Oklahoma. The correlation is found to be lowest, around 0.9, from March to May.[7] Yu et al. define locational basis risk as the difference between total basis risk associated with rainfall at the exact location and total basis risk associated with rainfall from the grid value. They consider three locations total, two in Nebraska and one in Kansas, and find the locational risk to be $5-9\%$.[10] Cho and Brorsen provide more comprehensive analysis using the correlation between 131 Oklahoma Mesonet weather stations and their associated grid value. They find an overall correlation of 0.95 that varies slightly by location. Specifically, the correlation is found to be lower in low rainfall areas.[11] Overall, the findings indicate spatial basis risk is between $5-10\%$, but has the potential to be lower in certain locations and times of year.

Prior work investigating dependence related basis risk or even total basis risk is scarce and much less consistent. Maples, Brorsen, and Biermacher estimate the correlation between

[9] Denis Nadolnyak and Dmitry Vedenov. "Information Value of Climate Forecasts for Rainfall Index Insurance for Pasture, Rangeland, and Forage in the Southeast United States". In: *Journal of Agricultural and Applied Economics* 45.1 (2013)

[10] Jisang Yu et al. "Estimating the Basis Risk of Rainfall Index Insurance for Pasture, Rangeland, and Forage ". In: *Journal of Agricultural and Resource Economics* 44.1 (2019), pp. 179–193

[11] Whoi Cho and B. Wade Brorsen. "Design of the Rainfall Index Crop Insurance Program for Pasture Rangeland and Forage". In: *Journal of Agricultural and Resource Economics* 46.1 (2021), pp. 85–100

[12] Ashlee Westerhold et al. "Risk implications from the selection of rainfall index insurance intervals". In: *Agricultural Finance Review* 78.5 (2018), pp. 514–531

forage yields of ryegrass, a common cool season forage, and rainfall in two Oklahoma locations. Winter forages are often planted in the late summer-fall and grow through the late-spring. They find that correlation between rainfall intervals and ryegrass forage yield is only significant in the December-January interval.[7] Using a regression framework, they find some evidence that rainfall from September to February has a positive effect on yields, but results vary greatly by location. They additionally include temperature variables in their model, but find no significance.

Westerhold et al. examine financial outcomes from forage production for two controlled locations in Nebraska and a blend of warm and cool season grasses. They use a gamma curve to model the relation ship between yield and yearly rainfall, recognizing the relationship is likely non-linear. Yearly rainfall is on the forage production year, September through August, rather than the calendar year. They find yearly rainfall accounts for 72% of variation in yields.[12]

Yu et al. evaluate basis risk associated using accurate ranch level data from three university ranches: two Nebraska locations used in Westerhold et al. and one additional Kansas location. They regress annual forage yield on a quadratic time trend, fixed effects, and precipitation (level and squared) from the previous January through July, an extension of the forage production year. They use a variable selection technique to account for their large number of regressors. Their model explains 82.4% of the variability in forage yields in all locations using ranch-level rainfall data; however, only the months March, May, June, and July are significant in aggregate monthly precipitation.[10] They define basis risk as false negative probability or

$$FNP = Prob(\hat{y}_t > \bar{y} | y_t < \bar{y}) \tag{1}$$

where where $\hat{y}_t$ is the predicted yield, $\bar{y}$ is the historical average yield, and y_t is the realized yield in year t. They find that their estimated PRF-RI mechanism that includes months March-July carries a basis risk between 21-43% depending on precipitation data and yield location.[10]

Overall, there is some support of PRF-RI as a risk management tool, but prior work is limited by choice of forage and location. The actual basis risk of the PRF mechanism likely varies greatly by forage type and local weather patterns. In this study, I will use a novel approach, which generalizes across weather regions and forages, to evaluate the basis risk associated with PRF-RI insurance. Then, I will examine index-based mechanisms including alternate variables with the goal of reducing basis risk through increased correlation to forage production. The findings of this study will be of use to policy makers and those in extension who care about the performance of the PRF mechanism as a risk management tool.

3 Data and Conceptual Framework

As seen in the previous section, basis risk of the PRF-RI mechanism has been estimated in multiple ways. I propose a new estimate of basis risk, which is favorable for the type of data used in the rating process. The data used by the RMA are monthly aggregates of precipitation. Therefore, it is natural to use only monthly variables when evaluating the basis risk of the contract. Changing the data interval would question the validity of a basis risk estimate. To measure the correlation between forage yields and the index, a dependent variable to represent forage yields is needed. Forage production is known to vary across location and variety. Temperature also likely plays a large role in forage production, as mentioned by users of the product in their complaints. Additionally, It is reasonable to assume the effect of weather variables on forage yield changes in different climates. To estimate a consistent measure of basis risk of the current PRF mechanism and suggest alternatives, a dataset which includes forage yields with different varieties and in different weather regions would be ideal.

In an effort to provide this consistency, the forage data are taken from the USDA Forest Service's rangeland dataset, which covers the coterminous US. The dataset is derived from satellite imagery, so it is comprehensive, and for inclusion an of land must be 1.0 acre in size and 120.0 feet wide. The categories of land in the dataset are "Rangeland," "Afforested Rangeland" (experiencing encroachment by trees [¿ 25% tree cover]) and "Transitional Rangeland" (currently dominated by herbs or shrubs that will likely become forested without management intervention).

The dependent variable for this study, forage yields, is taken from the rangeland productivity dataset, which is a subset of the rangeland dataset. It contains annual rangeland productivity in pounds per acre for only the "Rangeland" category of land. The productivity measure is generated using the Normalized Difference Vegetation Index (NDVI) from the Thematic Mapper Suite from 1984 to 2020 at 250 m^2 resolution (one pixel represents 250 m^2). The NVDI operates by measuring near infrared and visible light reflected off vegetation. Figure 1 displays, in yellow, the area included in the rangeland productivity dataset.

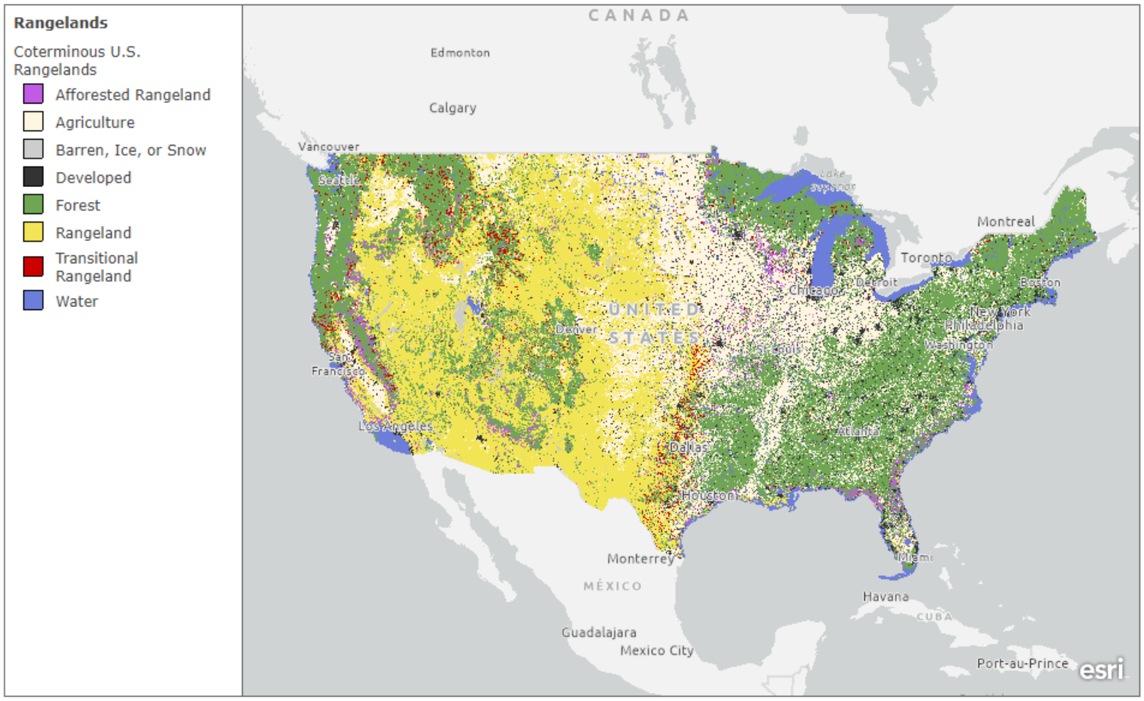

Figure 1: U.S. Rangelands

The regressors used in this study are taken from PRISM Climate Group's monthly time series dataset. The prism climate group uses a network of weather stations in the coterminous U.S. to measure weather data such as precipitation, temperature, and vapor pressure. From 1984 to present, there are at minimum 16,500 weather stations reporting precipitation, temperature maximum, and temperature minimum, the variables used in this study.[13]

The data are aggregated at the PRF grid level as well as the county level. The data are aggregated at the higher county level for multiple reasons. First, there are 3,006 counties in the U.S. The PRISM data are spatially smoothed to cover the entire U.S., but a grid level of aggregation will have observations containing no weather stations. Second, using county level aggregation will help preserve the spatial consistency of the model. Many

[13] PRISM Climate Group. *Descriptions of PRISM Spatial Climate Datasets for the Conterminous United States*. Tech. rep. PRISM Climate Group, Oregon State University, 2021

observations from the rangeland productivity index come from the great plains region with dense amounts of rangeland. Aggregating to a higher level will improve the consistency of the estimates across all rangelands as they are more equally represented. I compare the results from the two different aggregation levels to make conclusions about the mechanisms potential and spatial consistency.

3.1 Framework

Forage yield models are estimated with standardized data as well as non-standardized data. The standardized data accounts for fixed effects of county productivity and conveniently acts in a similar fashion to the current PRF mechanism, which insures rainfall as a percentage of expected (historic mean) rainfall for the area. The non-standardized data provides intuition to the biological needs of forage and tell us how much variation in forage yields can be explained through a weather index. I estimate the basis risk of the current PRF mechanism and of my proposed mechanism. I define basis risk in equation 2 following Yu et al.[10] This definition of basis risk corresponds to the false negative probability (FNP), or, probability of predicting above average yield when yield was below average.

$$FNP = Prob(\hat{y}_{ct} > \overline{y}_c | y_{ct} < \overline{y}_c) \tag{2}$$

$\hat{y}_{ct}$ is the predicted yield for county c in year t, $\overline{y}_c$ is the historic mean yield, and y_{ct} is the realized yield. For the standardized equations, equation 3 defines basis risk.

$$FNP = Prob(\hat{y}_{ct} > 0 | y_{ct} < 0) \tag{3}$$

First, in equation 4, I produce an estimate the basis risk of the current PRF mechanism. Ppt_{cti} is precipitation in county c for year t and month i.

$$y_{ct} = \sum_{i=1}^{12} \beta_i Ppt_{cti} + \epsilon_t \tag{4}$$

Model 1

In reality, this is not quite the actual basis risk of the PRF mechanism. The mechanism is divided into two month intervals throughout the year and does not have month specific factors, which implies identical coefficients on each month. This is a best guess of the basis risk of the current mechanism.

Next, I estimate alternative mechanisms and their associated basis risk. Equation 5 shows model 2. Because this data set covers a large geographic area, effects are likely to vary across regions. For example, precipitation is likely more important in hot dry regions of the country. I use regional fixed effects interacted with monthly precipitation to account for this. I assign a dummy variables for each quartile of yearly precipitation, so each region represents a group that is 25% of the total numbers of counties in size and grouped by similar values of yearly precipitation. I do the same for temperature quartiles to produce four temperature regions, which are independent of precipitation regions. Notice that in the case of standardized data, these mean zero regions would have no effect on mean zero yield; however, when interacted with monthly precipitation, they provide flexibility for precipitation's effect on forage yield to vary by weather region.

In this model, Ppt_{cti}^2 are monthly precipitation squared, Ppt_Dum_{cj} are fixed effects for each precipitation quartile, $Tmax\_Dum_{cj}$ are fixed effects for each max temperature quartile, PI_{ctij} are the interactions between the precipitation quartiles and monthly precipitation, and TI_{ctij} are the interactions between each temperature quartile and monthly precipitation. I include an additional regressor (for each month) to bring a measure of temperature into the model. I estimate the model using both sum of monthly temperature and temperature days over the threshold as representations of temperature. Temperature days over the threshold, or $Tdays_{cti}$ in equation 5 is drawn from Annan and Schlenker's seminal paper in 2015, which introduces it. Temperature days is a count variable for the number of days within a month with temperature over a threshold.[14] It is intended to capture the direct effect of temperature on precipitation. The model also includes fixed effects, D_c, when applied to non-standardized data too account for county-specific effects such as forage variety, soil quality, and solar radiation.

$$y_{ct} = \sum_{i=1}^{12} \left[\beta_i Ppt_{cti} + \beta_{i+12} Ppt_{cti}^2 + Tdays_{cti}\right] \tag{5}$$

$$+ \sum_{j=1}^{3} \left[\delta_j Ppt\_Dum_{cj} + \delta_{j+3} Tmax\_Dum_{cj} + \sum_{i=1}^{12} \left[\gamma_i PI_{ctij} + \gamma_{i+12} TI_{ctij}\right]\right] + D_c \epsilon_t \tag{6}$$

Model 2

For the standardized case, my expectation is that precipitation's effect is decreasing in precipitation quartile. This represents how above precipitation variability is less important in regions with high expected rainfall. I expect precipitation's effect to be increasing in temperature quartile in summer months. This represents precipitation's increased importance in hotter regions, which is consistent with the agronomy literature.

3.2 Estimation Process

In estimating basis risk, data are split into an estimation group and a test group to observe out of sample performance. To remove inter-year correlation between yields, the test group used is one year of the dataset. The estimation-test process is repeated T times using every

[14] Francis Annan and Wolfram Schlenker. "Federal crop insurance and the disincentive to adapt to extreme heat". In: *American Economic Review* 105.5 (2015), pp. 262–266

year as the test year and basis risk for each year is averaged to get the final measure, known as a k-fold process. This will ensure consistency of the estimates throughout different yearly outcomes. For example, basis risk in the model could be higher in drought years or high precipitation years, but k-fold yearly validation will ensure that all outcomes are taken into account. All models are estimated at the county level as well as the grid level for comparison. They are additionally estimated with standardized data at the county and grid level and non-standardized data at the county level. The standardized data mimics the current PRF mechanism and allows estimation at the grid level that is demeaned: too many fixed effects needed for efficient grid level estimation. The non-standardized county data provides us with understanding of the underlying biological process for forage growth and how much variation can be explained from our regressors.

The current PRF mechanism, model 1, insures all months and insures precipitation at a percent of historic mean rainfall for all months; Therefore, we estimate the basis risk of this model simply by OLS. Due to the high number of regressors in model 2, a variable selection technique, such as the lasso, is a natural choice. However, as Tibshirani[15] documents, if the regressors are highly correlated then the ridge regression outperforms the lasso in minimizing out of sample error. Therefore, I use Zou and Hastie's elastic net regularization method, which still conducts variable selection while performing better in out of sample forecast than the lasso.[16] The elastic net estimator is

[15] Robert Tibshirani. "Regression Shrinkage and Selection via the Lasso". In: *Journal of the Royal Statistical Society* 58.1 (1996), pp. 267–288

[16] Hui Zou and Trevor Hastie. "Regularization and Variable Selection via the Elastic Net". In: *Journal of the Royal Statistical Society* 67.2 (2005), pp. 301–320

$$\hat{\beta} = min_{\beta}\{|Y - X\beta|^2\} \tag{7}$$

$$subjectto(1-\alpha)\sum_{j=1}^{p}|\beta_j| + \alpha\sum_{j=1}^{p}\beta_j \leq s \tag{8}$$

where s and α are tuning parameters. If $\alpha = 1$ this is equivalent to the lasso; if $\alpha = 0$, it is the ridge regression. I use $\alpha = 0.9$ for estimation because it is associated with the lowest RMSE. This method also has the options to select specific variables to be non-penalized, which is used to force regional fixed effects into the standardized model in order for the model to be fully identified. A variable selection technique would not select fixed effects when the dependent variable is standardized, but the interaction terms may be selected.

Before estimating model 2, I perform a first-step estimation of the temperature day threshold for forage. I create yearly count variables for temperature days between one and forty degrees Celsius: variables named $Tdays_i$ for $i = 20 : 40$. For example, $Tdays_2 0$ is the number of days with max temperature between 21 and 22 degrees and $Tdays_{40}$ is the number of days with max temperature greater than 40 degrees. I regress yield on the yearly temperature day count. In this case, I use a yearly count instead of monthly variables for simplicity. Ideally, each month and each region would have a different temperature threshold, but this would surely overfit the data. Equation 9 defines the first step regression including temperature days. I include yearly precipitation and precipitation squared in the regression and estimate it via elastic net.

$$Yield_t = Prec_t + Prec_t^2 + \sum_{j=20}^{40} Tdays_{t,j} + \epsilon_t \tag{9}$$

4 Results

4.1 Current Mechanism

Model 1's parameter estimates can be seed in table 1. Model 1 is estimated at the county and grid level for regionally standardized data. The regressors and dependent variable are standardized by county, so they are interpreted as a one standard deviation increase in monthly precipitation leads to a β standard deviation increase in annual forage yield. Precipitation explains 15.3% of standardized forage yield variation at the county level and 15.6% at the grid level. Basis risk associated with the current PRF mechanism is 34.9% at the county level and 34.8% at the grid level.

The only major difference between the level of aggregation is precipitation in December, which is negative at the county level and positive at the grid level. Precipitation in November is also negative. These results imply the design of the PRF mechanism may not be optimal. April through August have by far the largest effect on forage yields, which is consistent with the findings of Yu et al.[10]

Table 1: Model 1

	County Precipitation	Significance
Month 1	0.081	***
Month 2	0.046	***
Month 3	0.07	***
Month 4	0.142	***
Month 5	0.171	***
Month 6	0.148	***
Month 7	0.107	***
Month 8	0.126	***
Month 9	0.046	***
Month 10	0.007	***
Month 11	-0.066	***
Month 12	-0.008	***

	Grid Precipitation	Significance
Month 1	0.103	***
Month 2	0.061	***
Month 3	0.069	***
Month 4	0.168	***
Month 5	0.173	***
Month 6	0.118	***
Month 7	0.082	***
Month 8	0.111	***
Month 9	0.063	***
Month 10	0.016	***
Month 11	-0.051	***
Month 12	0.046	***

I also regress non-standardized forage yields on monthly precipitation and include county level fixed effects to get a measure of real yield variation explained by precipitation. The model explains 45.9% of non-standardized forage yield variation. The precipitation coefficients are displayed in table 2. Using the regression, the basis risk associated with this model is 51.3%, which performers worse than a random guess and much worse than the standardized model.

Table 2: Real Yield Precipitation Coefficients

	Ppt	PV
Month 1	-0.105	
Month 2	0.794	***
Month 3	1.708	***
Month 4	5.319	***
Month 5	3.369	***
Month 6	3.755	***
Month 7	3.638	***
Month 8	3.015	***
Month 9	2.307	***
Month 10	1.513	***
Month 11	1.274	***
Month 12	0.371	***

4.2 Proposed Mechanism

I first estimate model 2 using sum of monthly temperature as the temperature variable. Precipitation is interacted with the weather regions to allow it's effect to vary over regions. When the interaction is selected in one of the quartiles, the net precipitation effect in that region becomes the overall precipitation effect plus the selected quartile effect.

The selected regressors in model 2, seen in table 3, explain 20% of standardized forage yield variation. The basis risk is found to be 0.363%, which is outperformed by the current mechanism. Precipitation is selected in months January-September, with a positive value representing a positive effect from above average precipitation. Above average precipitation has a relatively strong effect on forage yields in April through June regardless of region, and additionally for months July and August, but only in high temperature regions (Tmax quartile 3 and 4). In July and August, the positive coefficient estimates in the higher temperature quartiles represent the importance of above average precipitation in hotter regions. Precipitation quartile estimates are not selected in many cases, but the negative value in the fourth precipitation quartile for May represents the decreased importance of above average precipitation for regions already receiving a high amount of precipitation. The temperature variable, above average sum of monthly temperature, has a negative effect in April and May and a positive effect in January and November. These temperature effects were selected for all regions (they are not interacted with region), but there are likely other temperature effects present throughout the year that vary by region. These will be explored in a robustness check at the end of this section.

Table 3: County Level with Sum Monthly Temperature

	Ppt	Ppt Sq	Ppt Quartile 2	Tmax Quartile 2	Ppt Quartile 3	Tmax Quartile 3	Ppt Quartile 4	Tmax Quartile 4	Tmax
Month 1	0.038								0.122
Month 2	0.028								
Month 3	0.035		0.049						
Month 4	0.090								-0.062
Month 5	0.130						-0.097		-0.131
Month 6	0.136								
Month 7	0.019		0.040			0.080		0.123	
Month 8	0.068					0.101		0.108	
Month 9	0.037							0.035	
Month 10									
Month 11									0.122
Month 12									

4.2.1 Proposed Mechanism with Temperature days

Before estimating model 2 with the temperature threshold, I perform the first-step estimation of the temperature days threshold. The selected parameter estimates are plotted in figure 2. Days over 30 degrees are shown to have a negative impact on forage yields, so I use that as the threshold for model 2. Days between 23 and 29 degrees have a positive effect on forage yield, but that is likely variable by weather region. Days over 35 degrees are not selected, but this is likely due to the relatively low number of those days experienced.

The selected regressors in model 2 with the temperature threshold, seen in table 4, explain 17.2% of standardized forage yield variation. The basis risk associated with model 2 is 0.348% for standardized yields, which performs similarly to the current mechanism in basis risk and outperforms in explained variation. The temperature days threshold model outperforms the sum of temperature model, likely due to temperature days above the threshold's effect being more consistent across different weather regions. This is

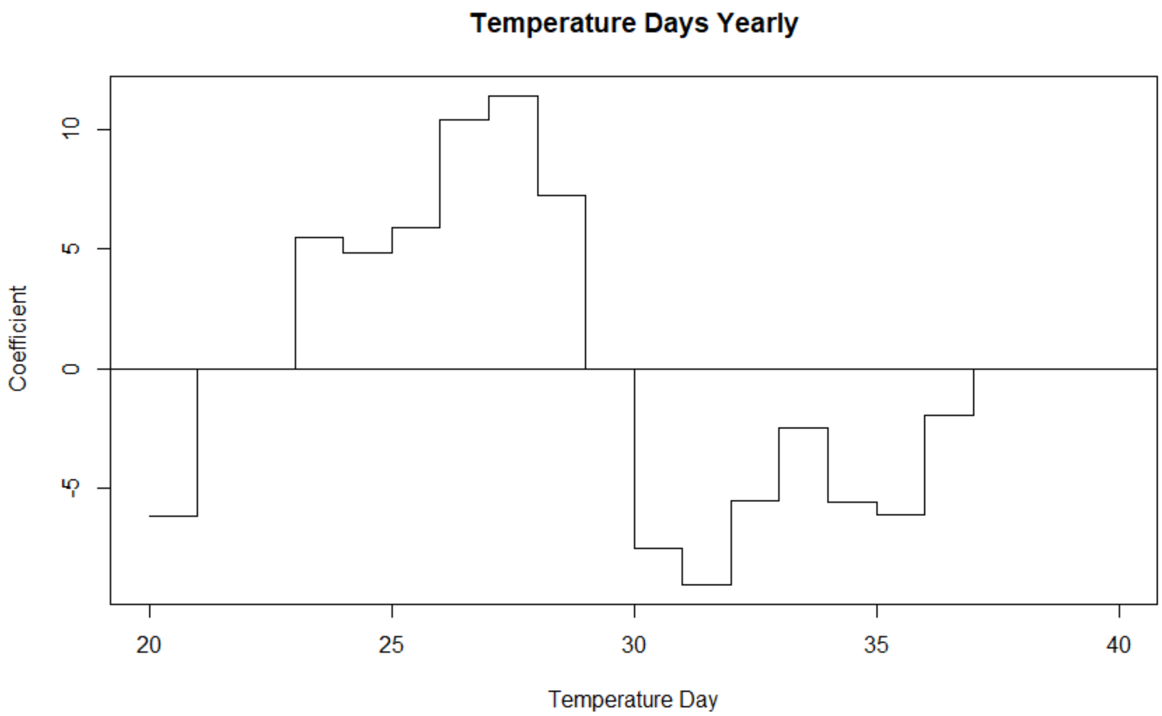

Figure 2: First Step Coefficients

investigated in the robustness check as well.

When sum of monthly temperature is replaced with the temperature threshold variable, there are a few changes in selected variables. There is an additional month in which temperature effect is negative, March, and no months have a positive temperature effect, likely due to the fact there are not many temperature days over the threshold in winter months regardless of region. Additionally, precipitation is selected with a negative value in November, which contradicts the design of the current contract that precipitation has a positive effect in all months. Finally, two more months have a negative precipitation effect in the highest precipitation quartile, April and November, likely representing the decreasing importance of above average precipitation when the region already receives high precipitation in April. However, in November this shows the negative effect of above average precipitation in November is larger in high precipitation regions.

Table 4: County Level with Temperature Threshold

	Ppt	Ppt Sq	Ppt Quartile 2	Tmax Quartile 2	Ppt Quartile 3	Tmax Quartile 3	Ppt Quartile 4	Tmax Quartile 4	30 tmaxday
Month 1	0.039		0.047						
Month 2	0.035								
Month 3	0.054								-0.040
Month 4	0.144						-0.055		-0.020
Month 5	0.184						-0.102		-0.015
Month 6	0.137								
Month 7	0.031		0.032			0.067		0.112	
Month 8	0.069					0.093		0.105	
Month 9	0.025							0.032	
Month 10									
Month 11	-0.046						-0.042		
Month 12									

When estimated with non-standardized yields, the model explains 91.51% of non-standardized forage yield variation, a large increase from 45.9% under the current mechanism. The proposed model carries a basis risk of 38.8%, which is down from 51.3%. In the non-standardized case, the regional dummy variables now have coefficient values which represent the difference in yields for each weather region, as seen in table 5. Yield is increasing in precipitation district, as expected, but is also increasing in temperature district. This may be associated with higher year round yield in regions with warm winters. The monthly coefficient estimates are displayed in table 6. Precipitation is selected in all months and has a positive effect in all except November. Precipitation's effect is decreasing in precipitation region showing the diminishing marginal return of precipitation. Precipitation's effect is higher in temperature quartile 3 and 4, in May through August, representing the increased importance of precipitation in hotter regions. Temperature days over the threshold is positive in December through February and negative in March through July. This represents the benefit of warm winters and the detriment of hot summers.

Table 5: Regional Fixed Effects

	Precipitation	Temperature
Quartile 2	83.47	-685.92
Quartile 3	310.34	-415.42
Quartile 4	566.21	606.7

Model 2 with the temperature threshold is also estimated with data standardized at the grid level. The selected regressors in Model 2, seen in table 7, explain 19.7% of standardized forage yield variation and have a basis risk of 35.5%. Many more variables are selected at the grid level and they explain a higher percent of variation in forage yields.

Table 6: County Level Potential Yields

	Ppt	Ppt Sq	Ppt Quartile 2	Tmax Quartile 2	Ppt Quartile 3	Tmax Quartile 3	Ppt Quartile 4	Tmax Quartile 4	30 tmaxday
Month 1	0.119		0.971		0.252		-0.001		67.099
Month 2	0.400	-0.001		-0.073				0.411	1.458
Month 3	0.285		0.876		0.288	-0.184			-19.356
Month 4	1.673	-0.001			-0.429		-0.798		-8.168
Month 5	2.366	-0.002	-0.072		-0.617		-1.314	0.054	-0.688
Month 6	1.871	-0.001		-0.008	-0.840		-0.525	0.227	-2.680
Month 7	1.097	-0.002	0.512			0.306	-0.215	0.979	-0.489
Month 8	0.972	-0.002	0.356			0.743	-0.110	0.986	
Month 9	0.359	0.000	0.222	-0.093	-0.355	-0.140		0.031	
Month 10	0.252	0.001	-0.076			-0.518	0.108	-0.401	
Month 11	-0.876	0.001	0.460			0.004	-0.141	-0.262	-22.019
Month 12	0.275	0.000	-0.594	0.080	-0.826				47.389

This is likely due to the large number of similar grid observations used. Basis risk is slightly higher than the county level aggregation, reflecting the loss of generalization.

The coefficients weakly confirm that precipitation's effect is increasing in temperature region and decreasing in precipitation region. There are some coefficients that contradict this, the positive precipitation values in precipitation quartile 3 and the negative temperature values in temperature quartile 3.

Table 7: Grid Level with Temperature Threshold

	Ppt	Ppt Sq	Ppt Quartile 2	Tmax Quartile 2	Ppt Quartile 3	Tmax Quartile 3	Ppt Quartile 4	Tmax Quartile 4	30 tday
Month 1	0.120		-0.059				-0.082		0.013
Month 2	0.072				-0.049		-0.061	0.053	0.019
Month 3	0.056				0.038				-0.045
Month 4	0.173			0.047		0.031	-0.104	-0.033	-0.039
Month 5	0.128		0.101	0.073	0.030		-0.043		-0.015
Month 6	0.052		0.145		0.112	-0.017	0.047	-0.033	
Month 7	0.020					0.092		0.122	
Month 8	0.061					0.062		0.118	
Month 9	0.045		0.018			0.026	-0.036	0.055	
Month 10	0.047					-0.079		-0.041	0.013
Month 11	-0.041						-0.044		
Month 12	0.083			0.033	-0.058		-0.055	-0.076	

4.3 Robustness Check by Region

I also split the model up by weather district to perform a sensitivity analysis of precipitation and the temperature. I remove the quartile dummy variables from the model and regress forage yields on precipitation for each quartile region. I then regress yields on precipitation along with a temperature variables for each region, allowing a more clear view of how precipitation and temperature effects vary across regions. I use the same iterative k-fold net elastic procedure for the estimation in each region. These regressions use standardized regressors and yields, similar to model two. To be clear, this means estimated precipitation and temperature effects are for monthly precipitation and temperature compared to the average.

Table 8 shows the selected precipitation months by region. In October-December, precipitation is selected at a low rate with mixed effects in October and December and negative effects in November. This is consistent with the findings of Yu et al. 10 that precipitation in these months does not significantly impact forage growth. In January, the findings are also mixed, but the lower half of the precipitation counties, quartiles one and two, show some evidence of a positive impact on temperature. Additionally, there is mixed support for precipitation in high-precipitation and high-temperature regions, with the exception being region 4-4, with a negative impact on forage yields. In February, some support is also shown for low-precipitation regions, with no apparent impact of temperature region. In March, there is additional support for precipitation in low-precipitation regions and high-precipitation and high-temperature regions. In the months April-June, precipitation is selected in almost ever low-precipitation region with a positive effect. Additionally, the high-precipitation and high-temperature regions show support for precipitation's benefit on forage yields. This effect is clearly emphasized in July and August, where precipitation is selected in every high-temperature regions except one. This result presents convincing evidence that temperature should be considered in the design of the PRF mechanism, even if just in select months. September also shows some support for precipitation in lower-half-precipitation and highest-temperature regions.

A few overall trends from this analysis: above average precipitation is far more important in low-precipitation regions; it is still important in high-precipitation regions, but only in the hottest regions (temperature quartiles 3-4) and the hottest months (June-August). Even if a temperature variable is not directly included, there is evidence that the design of the PRF mechanism should consider temperature region as well as precipitation region.

Additionally, with a warming climate, more counties will enter the higher-temperature category, where precipitation is of increased importance.

Table 8: Regional Precipitation Effects

	Ppt 1	Ppt 2	Ppt 3	Ppt 4	Ppt 5	Ppt 6	Ppt 7	Ppt 8	Ppt 9	Ppt 10	Ppt 11	Ppt 12	n	Rsq	Basis
Ppt 1 Temp 1	0.073	0.062	0.041	0.203	0.352	0.278	0.038			0.045		0.046	3723	0.383	0.209
Ppt 1 Temp 2	0.037	0.030	0.042	0.258	0.296	0.162			0.028	0.041	-0.045	0.097	4066	0.348	0.273
Ppt 1 Temp 3		0.033		0.215	0.198	0.131	0.148	0.129					3778	0.256	0.252
Ppt 1 Temp 4	0.061	0.088		0.115	0.159	0.050	0.177	0.246	0.102				2873	0.292	0.252
Ppt 2 Temp 1	0.110		0.062	0.096		0.060							5176	0.065	0.389
Ppt 2 Temp 2	0.056				0.203	0.256	0.057						3363	0.201	0.309
Ppt 2 Temp 3	-0.109	0.071		0.194	0.288	0.231	0.115	0.176		-0.109			2764	0.385	0.300
Ppt 2 Temp 4	0.063		0.168	0.234	0.258	0.090	0.187	0.240	0.064			-0.090	3136	0.459	0.336
Ppt 3 Temp 1											-0.060		4155	0.013	0.912
Ppt 3 Temp 2													4874	0.000	0.433
Ppt 3 Temp 3	0.080			0.061	0.159	0.097	0.054	0.147					3003	0.173	0.318
Ppt 3 Temp 4	0.067		0.097	0.122	0.242	0.160	0.227	0.221					2407	0.330	0.341
Ppt 4 Temp 1													1395	0.003	0.967
Ppt 4 Temp 2													2128	0.000	1.000
Ppt 4 Temp 3						0.058		0.049					4921	0.023	0.787
Ppt 4 Temp 4	-0.122					0.140	0.055	0.053			-0.122		5995	0.089	0.496

Table 9 shows the selected temperature months by region. The first clear observations are that temperature carries a negative effect in April through July and a positive effect in October through January regardless of region. These effects are consistent with the variables selected by model two. The positive effects of temperature in December and January are highest in the high-temperature regions, implying that these hotter regions have some forage growth even in the winter months. Some evidence is shows for temperature having a negative effect on forage yield, specifically in April and May. June and July show limited evidence for a negative temperature effect in high-temperature regions. Temperature does not appear to significantly impact forage yields in February, March, August, and September.

Table 9: Regional Temperature Effects

	Tsum 1	Tsum 2	Tsum 3	Tsum 4	Tsum 5	Tsum 6	Tsum 7	Tsum 8	Tsum 9	Tsum 10	Tsum 11	Tsum 12	n	Rsq	Basis
Ppt 1 Temp 1				-0.122	-0.255						0.096	0.100	3723	0.501	0.269
Ppt 1 Temp 2				-0.186	-0.126					0.056	0.095	0.092	4066	0.435	0.299
Ppt 1 Temp 3										0.084	0.187	0.099	3778	0.360	0.294
Ppt 1 Temp 4	0.107			-0.077	-0.072		-0.025		-0.048			0.107	2873	0.326	0.273
Ppt 2 Temp 1				-0.167	-0.138						0.088	0.021	5176	0.173	0.413
Ppt 2 Temp 2	0.173			-0.163	-0.100		-0.064				0.173		3363	0.309	0.389
Ppt 2 Temp 3						-0.188	-0.082			0.092	0.247	0.038	2764	0.476	0.321
Ppt 2 Temp 4	0.146				-0.150			0.077				0.146	3136	0.508	0.351
Ppt 3 Temp 1	0.020				-0.104						0.020		4155	0.047	0.691
Ppt 3 Temp 2					-0.007								4874	0.019	0.580
Ppt 3 Temp 3	0.086						-0.152				0.086		3003	0.186	0.342
Ppt 3 Temp 4	0.109		-0.034	-0.183		-0.150					0.118	0.182	2407	0.453	0.390
Ppt 4 Temp 1									0.047				1395	0.064	0.643
Ppt 4 Temp 2	0.025				-0.082						0.025		2128	0.040	0.720
Ppt 4 Temp 3					-0.094								4921	0.061	0.613
Ppt 4 Temp 4		-0.102					-0.134						5995	0.151	0.484

I also perform the robustness check with the temperature days threshold as the temperature variable, as seen in table 10. There is much less evidence of the positive temperature effect in October through January, likely due to the low number of days that actually exceed the threshold. Temperature continues to show a negative effect in March through July, but in less region. Specifically, the negative effects of temperature in high-temperature regions are emphasized. While we want the model to explain as much variability as possible, it is important to note the trade off in basis risk, which we also saw in model 2. Basis risk was overall higher when using sum of monthly temperature instead of temperature days threshold. In this breakdown by region, the out-performance of temperature days becomes less clear.

Table 10: Regional Temperature Threshold Effects

	Tdays 1	Tdays 2	Tdays 3	Tdays 4	Tdays 5	Tdays 6	Tdays 7	Tdays 8	Tdays 9	Tdays 10	Tdays 11	Tdays 12	n	Rsq	Basis
Ppt 1 Temp 1													3723	0.373	0.201
Ppt 1 Temp 2						-0.074							4066	0.358	0.287
Ppt 1 Temp 3													3778	0.255	0.249
Ppt 1 Temp 4		0.066	-0.061	-0.081	-0.045	-0.044			-0.076				2873	0.329	0.271
Ppt 2 Temp 1													5176	0.077	0.368
Ppt 2 Temp 2				-0.153					-0.061				3363	0.253	0.378
Ppt 2 Temp 3	0.018		-0.062			-0.112	-0.102			0.018			2764	0.419	0.327
Ppt 2 Temp 4			-0.122	-0.086	-0.100								3136	0.487	0.322
Ppt 3 Temp 1													4155	0.027	0.856
Ppt 3 Temp 2													4874	0.012	0.484
Ppt 3 Temp 3							-0.149						3003	0.200	0.347
Ppt 3 Temp 4				-0.172									2407	0.345	0.401
Ppt 4 Temp 1													1395	0.003	0.985
Ppt 4 Temp 2													2128	0.000	1.000
Ppt 4 Temp 3					-0.037								4921	0.041	0.654
Ppt 4 Temp 4							-0.130	-0.047					5995	0.119	0.457

An interesting result regardless of temperature variable choice is that basis risk is increasing in precipitation regions. I confirm this effect by regressing precipitation and a time trend on basis risk for each year in model two. Again, basis risk was averaged over the years to remove inter-year correlation that would cause an endogeneity issue. I look at the basis risk from each year in the k-fold process and its relation to yearly precipitation. The time trend is included to make sure the regression is not spurious due to the trend in rainfall. I find that basis risk is significantly increasing in yearly precipitation and that a one standard deviation increase in yearly precipitation leads to a 11.7% increase in basis risk. This shows the models do not due well at predicting FNP in high precipitation years.

5 Conclusions

I find that basis risk under the current PRF mechanism is roughly 35% with basis risk increasing in high precipitation years. Unlike other studies, the unique dataset I use combined with my analysis by weather region ensures these results should extrapolate well across all forage producing regions. This measure of basis risk should correspondingly extrapolate across regions. My proposed mechanism has a lower basis risk than the current PRF contract, especially in dryer years, but still does not perform well when under high precipitation conditions.

I find temperature should be included in the index product for ideal yield risk management. I also find that weather region should be accounted for in the index design and that precipitation and temperature effects vary significantly by weather region.

While the inclusion of a temperature variable in the index does not lead to a significant reduction in basis risk, the effects of precipitation are shown to vary by temperature region. As temperature continues to rise in a warming climate, consideration of temperature region will become more important to correlation between precipitation and forage yield.

References

Annan, Francis and Wolfram Schlenker. “Federal crop insurance and the disincentive to adapt to extreme heat”. In: *American Economic Review* 105.5 (2015), pp. 262–266.

Binswanger, Hans Peter. “Is There Too Much Hype about Index-based Agricultural Insurance?” In: *Journal of Development Studies* 48.2 (2012), pp. 187–200.

Campiche, Jody and JJ Jones. “Pasture, Rangeland, Forage Insurance Program”. In: *Oklahoma Cooperative Extension Service* (2013).

Cho, Whoi and B. Wade Brorsen. “Design of the Rainfall Index Crop Insurance Program for Pasture Rangeland and Forage”. In: *Journal of Agricultural and Resource Economics* 46.1 (2021), pp. 85–100.

Clarke, Daniel J. “A Theory of Rational Demand for Index Insurance”. In: *American Economic Journal: Microeconomics* 8 (2016), pp. 283–306.

Diersen, Matthew. “Choosing Pasture, Rangeland, Forage Rainfall Index (PRF-RI) Insurance Coverage”. In: *AgWeb* (2015). DOI: `https://www.agweb.com/article/choosing-pasture-rangeland-forage-rainfall-index-prf-ri-insurance-coverage-university-news-release`.

Elabed, Ghada et al. “Managing basis risk with multiscale index insurance”. In: *Agricultural Economics* 44 (2013), pp. 419–431.

Group, PRISM Climate. *Descriptions of PRISM Spatial Climate Datasets for the Conterminous United States*. Tech. rep. PRISM Climate Group, Oregon State University, 2021.

Halcrow, Harold. “Actuarial Structures for Crop Insurance”. In: *Journal of Farm Economics* 31.3 (1949).

Maples, Joshua G., B. Wade Brorsen, and Jon T. Biermacher. “The Rainfall Index Annual Forage Pilot Program as a Risk Management Tool for Cool-Season Forage”. In: *Journal of Agricultural and Applied Economics* 48.1 (2016), pp. 29–51.

Nadolnyak, Denis and Dmitry Vedenov. “Information Value of Climate Forecasts for Rainfall Index Insurance for Pasture, Rangeland, and Forage in the Southeast United States”. In: *Journal of Agricultural and Applied Economics* 45.1 (2013).

RMA, USDA. *Pasture, Rangeland, and Forage*. 2019. URL: `https://www.rma.usda.gov/en/News-Room/Frequently-Asked-Questions/Pasture-Rangeland-Forage` (visited on 04/11/2019).

Tibshirani, Robert. “Regression Shrinkage and Selection via the Lasso”. In: *Journal of the Royal Statistical Society* 58.1 (1996), pp. 267–288.

Westerhold, Ashlee et al. “Risk implications from the selection of rainfall index insurance intervals”. In: *Agricultural Finance Review* 78.5 (2018), pp. 514–531.

Yu, Jisang et al. “Estimating the Basis Risk of Rainfall Index Insurance for Pasture, Rangeland, and Forage ”. In: *Journal of Agricultural and Resource Economics* 44.1 (2019), pp. 179–193.

Zou, Hui and Trevor Hastie. “Regularization and Variable Selection via the Elastic Net”. In: *Journal of the Royal Statistical Society* 67.2 (2005), pp. 301–320.