Weather Index Insurance and the Pricing of Spatial Basis Risk

Michael Norton,
International Research Institute for Climate and Society
Columbia University

Dan Osgood,
International Research Institute for Climate and Society
Columbia University

Calum G. Turvey
Applied Economics and Management
Cornell University

Selected Paper prepared for presentation at the Agricultural & Applied Economics Association's 2010 AAEA, CAES & WAEA Joint Annual Meeting, Denver, Colorado, July 25-27, 2010.

 This paper was completed in part as Michael Norton's MS thesis at Cornell University. The authors would like to thank contributors to the Myers endowment fund and the Risk Management Agency Partnership Agreement USDA-02IE08310229. *For any questions or comments about this paper, please contact Michael Norton by phone: (845) 680-4435 or email: mnorton@iri.columbia.edu*

Introduction

There have been some rigorous attempts to provide meaningful methods to quantify risk for weather index insurance in a spatial dimension. Part of the complication is that weather index insurance (or weather derivatives) pricing according to burn rate analysis deals not only with a spatial dimension but also a temporal one as historical frequencies are calculated. These calculations are further subjected to potential long-term trends due to climate change, as well as variations due to prevailing weather conditions according to the ENSO (El Nino Southern Oscillation) index.

From an empirical point of view this paper investigates basis risk using a novel approach to space measurement. We have developed a program that is linked to all of the weather stations in a given region (i.e. includes data on all weather stations in the United States). From a randomly selected point, we select all weather stations within a proscribed radius and calculate the particular weather risk at each station to calculate the burn rate insurance premium. This is done for the same specific criteria for all stations within the sphere. First, we compare the premiums at each station to measure heterogeneity over space. Second, we compare on a year by year basis the payouts that would have been made at each pair of locations. Thus, if there are 20 weather stations then there are N(N-1)/2 = 190 pairwise comparisons. Third, we measure the error or basis risk between each station pair and regress the mean differences against a number of spatial variables. These spatial variables include longitude and latitude coordinates as well as elevation difference and distance between weather stations.

This exercise reveals the spatial characteristics which have a persistent effect on temperature and precipitation risk. For temperature/heat risk, the most important independent variable is the difference in elevation between stations. This result is somewhat expected as there will likely be no differences in solar radiation in such a small area, and the primary temperature difference will be due to the elevation of the station. This results in an alternative way to price temperature risk that is highly dependent on differences in elevation and not distance. Rainfall, however, is not strongly correlated over distance as precipitation will often be an extremely localized phenomenon, with the other variables of lesser importance. The difficulty with rainfall is its periodic nature, which will often produce a long series of empty observations followed by a rain event. Many researchers ameliorate this issue by aggregating rainfall measurements into larger periods, such as 10-day periods called dekads. Despite this, the spatial rainfall correlation will likely improve at a larger time scale which will average down total differences between stations. This paper contributes to the literature by offering a pricing strategy for both temperature and precipitation risk tailored to the specific typology of each distinct type of risk. In many cases, the "portfolio" method of selecting a proportion of risk from each nearby station based on similarity in elevation (temperature) or geographic placement (precipitation) offers advantages over an index based more sophisticated spatial statistics algorithms. Also, by compartmentalizing the risk into existing weather stations, the "portfolio" method has the advantage of allowing for easier pricing of policies for insurance and reinsurance companies. In this manner, we will demonstrate a strategy for pricing risk in distributed locations around the U.S.

This paper is structured as follows: first, a discussion of the mathematical and data processing considerations for analyzing weather risk. This is followed by the introduction of a regression equation that attempts to predict differences in risk at existing stations through the use of the geographic characteristics of those stations. The results of the regression equation lead to a discussion of a pricing strategy for both temperature and precipitation risk tailored to the specific typology of each distinct type of risk as revealed by the regression results.

Background

Weather index insurance is a recent financial innovation that has received much attention from academics and implementers alike as a way to smooth risk for farmers in developing countries (Turvey 2001, Skees 2008, Vedenov and Barnett 2004). By using an "index" of weather observations as a proxy for crop loss, the problems of traditional indemnity insurance are reduced or eliminated. This removes the subjective nature of insurance adjustment as well as the problems of adverse selection and moral hazard that are present in the traditional indemnity insurance model. This innovation makes it possible to offer microinsurance to rural farmers in developing countries, which can serve a valuable function in a development intervention and may lead to more interactive benefits, such as improved access to rural credit. In developed countries, several private insurance companies have opened their doors in recent years to offer commercial products for the hedging of weather risk.

However, despite the promise of the technology, it is not always straightforward to apply. In particular, we trade the problems of adverse selection and moral hazard with that of basis risk, which is defined as the risk that payoffs of a hedging instrument do not correspond to the underlying exposures. Basis risk may be reduced through the selection of appropriate weather observations to construct the index, but in reality the prevailing weather conditions are only one variable in crop production and are often considered exogenous to the production function (Turvey and Norton 2008). Basis risk is a major problem when using a risk-smoothing implement such as weather index insurance. The good years should help pay for the bad years, but if the product were not aligned properly, the advantages of smoothing risk could be harmful to the farmers' bottom lines.

The proposed solutions to the problem of basis risk are varied. One approach is to perform spatial analysis techniques on weather data to provide a historical time series in varied geographic locations (Paulson and Hart 2006). Another study intentionally analyzed data from a flat area with consistent elevation (Richards, Manfredo, and Sanders 2004). Other researchers link microinsurance to microcredit and advocate for a central financial institution to aggregate index insurance contracts so as to average out basis risk for all actors (Miranda et al 2010, Woodard and Garcia 2008a). Clearly, if index insurance is to be widely used as a risk mitigation and climate adaptation tool for individual farmers, the problem of basis risk must be overcome.

A traditional sticking point in pricing weather index insurance is that there are only a certain number of weather stations for which historical data exists. A longer time series of data provides more confidence for historical burn rate analysis pricing and often a minimum number of years is needed to understand historical weather patterns, which will take ten or twenty years. Establishing a new station can provide high quality weather data, but there will be no historical record at that new station. This problem is acute in countries with poor infrastructure, which is paradoxically where weather index insurance might do the most good (Morduch 2006). But even in places with many long-established weather stations, the spatial distribution of risk is not yet fully understood. The challenge that is present in the weather index insurance market is how to strategically leverage the information from existing stations at geographic locations where the precise weather observations are unknown.

Some researchers have applied spatial analysis techniques to the weather observations directly, and used that information to construct a surface of historical time series observations for any geographic point. While techniques like kriging have shown to be very accurate in producing a prediction surface for points in space, these spatial analysis techniques are not designed to model for deviations from normal conditions, which is precisely what we are interested in and want to protect against. This paper takes rather the opposite approach to spatial analysis by pricing risk at known locations and analyzing how that risk changes through space. To that effect, we have developed a web-based computer program named Weather Wizard that is able to analyze historical weather observation data for all weather stations in the United States.[1] From a randomly selected point we select all weather stations within a certain radius and calculate the particular weather risk at each station to calculate the burn rate insurance premium.

Mathematical Considerations for Spatial Weather Risk

Table 1: Correlation of average of nearby stations of cumulative weather indexes

Heat	**Base Station**	**Avg. of Surrounding Stations**	**Difference**	**Correlation**
CDD Index (85° F):	68.41	71.51	-4.5%	0.8849
Heat Risk Event (Payout):	$22.37	$27.31	-22.1%	0.7751
Rainfall				
Cumulative Rainfall (in.):	10.65	10.36	2.6%	0.7457
Drought Risk Event (Payout):	$20.95	$22.61	-7.9%	0.6897

For illustration, some summary statistics are presented in Table 1. Listed are the aggregate temperature and rainfall observations for Ithaca, NY for June 1st – August 31st along with the average observation for all stations within a proscribed radius (100 miles for temperatures and 67 miles for rainfall.) The overall averages are similar, but when we examine the yearly variation as measured by the average correlation between the base station (Ithaca) and every other station, we find that heat is highly correlated but rainfall

[1] Weather Wizard is available at: http://www.weatherwizard.us

less so. A familiar pattern is that when we introduce risk events (defined later), the variability increases, not only in the averages but also in the correlation. The information presented here is also for relatively common events over long date ranges; presumably these numbers would weaken if a more specific time frame or risk event were used.

The challenge presented is, very simply, to improve the accuracy of the yearly correlation. Although this may seem somewhat abstract, insurance policies have profound real-world implications for farmers holding a policy and it is crucial to match the years with payments with the actual losses. By taking the payout schedule for all stations and adjusting for geographic variables, we can potentially price insurance contracts for any given point on the map. Because of the vast number of stations located around the country, our hopeful result is a simple equation in which we can build upon this simple methodology and adjust for the differences in distance, altitude, and polar coordinates. What follows is an attempt to provide a universal solution using those readily available geographic variables to arrive at a payout for any unknown location.

Defining the Risk Events

Choosing an event that is sufficiently general yet meaningful for all sites is difficult, because there is no such thing as generality. For example, a heat event in upstate New York is incomparable to a heat event in a warmer climate because temperatures in upstate New York infrequently reach above 90° F, but in Norman, OK this temperature is reached quite frequently in summer months (Turvey and Norton 2008). The sheer variation of climates in America requires us to tailor our heat risk events for each station.

To start, evidence indicates that temperatures above 85° F correlate with crop yield losses. (Schlenker and Roberts 2006) Using this as a benchmark, we accumulate a CDD index above 85° F with the mean CDD at the base station serving as the strike value or trigger and a sliding payout for values above that. Payouts are calculated at each station for every year data is available. Figure 1 shows the payout schedule for Ithaca, NY, where mean CDD is 68.41.

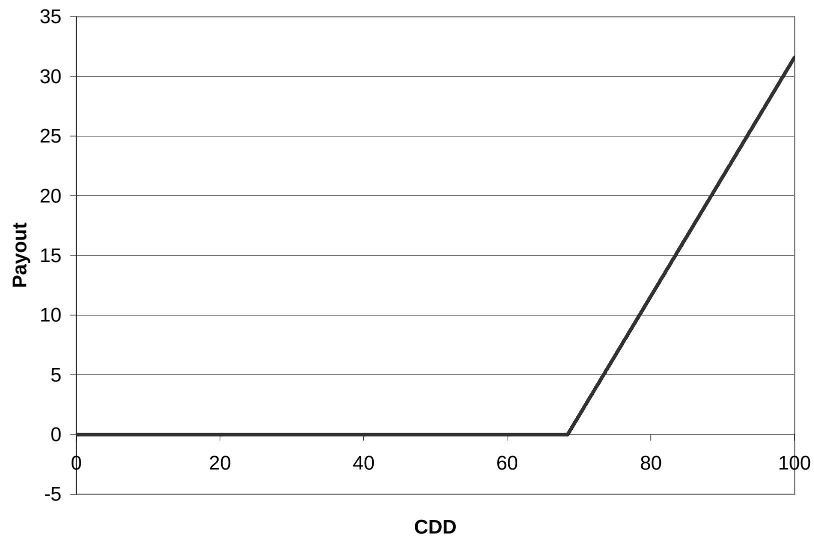

Figure 1: Schedule of payouts for heat risk event

Six weather stations were selected at sites across the country according to quality of data and the absence of geographic features within 50 miles that would prevent weather stations from being placed, such as bodies of water or international borders. Mean CDD for those six stations vary from 68.41 at Ithaca, NY to 720.63 in Davis, CA and are listed in Table 2.

Table 2: Mean CDD (85° F) at each location

Station	**Bridgeport, NE**	**Bethany, MO**	**Greenville, AL**	**Davis, CA**	**Ithaca, NY**	**Mosquero, NM**
Mean CDD (85° F)	442.13	350.74	603.18	720.63	68.41	282.05

For precipitation, the contract is identical for all sites. We use a drought event of less than .1" of precipitation over any 14-day period. The payoff will occur on a sliding scale with $10 accumulating for each hundredth of an inch less than .1", to a maximum of $100 per event if no rainfall was recorded. Up to three non-overlapping events are possible, with an annual maximum liability of $300. Figure 2 shows the payoff schedule due to the observed rainfall in any 14-day period, but yearly payoff amounts range from $0 to $300 because of the possibility of multiple events.

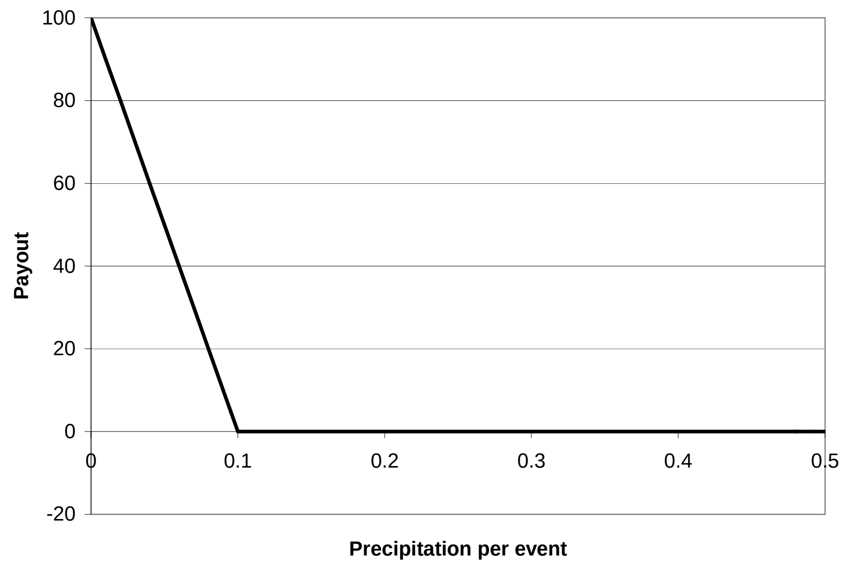

Figure 2: Schedule of payouts for drought risk event

Defining the geographic area

Weather Wizard is flexible as to the distance of the radius extending from the base station, but there are a few requirements that must be considered for a successful trial. A certain number of stations are needed to provide contrast, but there are relatively few stations within a short distance (10 miles) of each other. However, as we increase the radius of the circle, the area of the circle increases exponentially. Barring any obstacles like oceans or international borders, the number of stations increases exponentially as the radius of the circle increases. Because we compare each station against each other in each year, this also dramatically increases the number of comparisons that are made, as given by the following formula:

$$comparisons = \frac{n*(n-1)}{2} * years$$

Where n is the number of stations within the selected geographic radius, and years is the number of years of data at the base station. The total number of comparisons is subject to missing and incomplete data; many stations have only limited data, and with longer time horizons the potential for periods of missing data within years becomes greater.

Table 3 displays the number of stations for each type of weather data within a certain number of miles. Using the number of stations as the value of n in the equation above as well as the number of years of data at the base station, the number of potential comparisons between years is calculated.

Table 3: Number of comparisons in Ithaca, NY for a given number of miles

Rainfall

Miles	Stations	Potential Comparisons	Actual Comparisons	
10	2	225	93	41.33%
15	4	750	223	29.73%
20	12	5850	801	13.69%
25	17	11,475	1628	14.19%
30	25	24,375	3,304	13.55%
35	35	47,250	6,921	14.65%

Heat

Miles	Stations	Potential Comparisons	Actual Comparisons	
10	2	225	23	10.22%
15	2	225	23	10.22%
20	4	750	105	14.00%
25	6	1575	220	13.97%
30	10	4125	525	12.73%
35	16	10,200	1,366	13.39%

Such factors as length of contract and number of years selected will also affect the percentage of available data as presented in

Table 3. These percentages are somewhat low because of a relatively long date range. In this case, a 92-day window encompassing June-August was selected, which offers more opportunities for data to be missing than a more carefully targeted risk event. Also, more importantly, very few stations have data continuously to 1926 as Ithaca does; most stations date to just after World War II, and it's not uncommon for a station to have as little as one or two years of data for the entire 75 year period. If we selected a shorter contract length (say, 15 days instead of 92) fewer stations would be disqualified for missing data; likewise, if we only considered years after 1949, the percentage of actual comparisons would improve markedly. This discussion is intended to underscore the fact that even though we might define an identical geographic area, there is often a very different spatial distribution of data within that area depending on the parameters we select.

Also, perhaps in acknowledgement of the periodic, unpredictable nature of rainfall, precipitation observation stations are more densely placed and often contain more years of data. In the case presented here, there are more than twice as many precipitation gauges in a given radius than temperature stations, even if the percentage of data which is usable is roughly similar. It is very likely that temperature observations are placed more sparsely to reflect that temperatures are considered to be more continuous over a geographic area.

The advantage of this comparison-based model is that it treats all weather stations equally and is able to include otherwise useless data. In this model, the data will be compared on a year by year basis, regardless of how many years of data are at a particular station. The weather stations that only have a few years of data help provide contrast for spatial distributions of risk even though it is impossible to accurately price a contract for that station individually.

Also of pertinent interest is what these details entail for selecting a radius to study. As the radius increases, the area of study increases exponentially (according to the area of a circle – πr^2). The number of stations increases accordingly, which has vast ramifications for the number of potential comparisons according to the equation above. Since Weather Wizard is hosted on a web platform, there are limitations to the amount of data that it can process in a single iteration - selecting a radius requires the user to select a value large enough to offer meaningful results that will also fit within technical possibilities. For this paper, we are using a radius of 50 miles, which is large enough to allow the inclusion of sufficient stations for both heat and precipitation, but small enough to run properly on the Weather Wizard website.

The Regression Equation

The goal when formulating this regression equation was to try and predict the difference in payouts in any given year between any two locations using simple geographic variables.

$$(P_1 - P_2) = \beta_1 \varphi + \beta_2 (\alpha_1 - \alpha_2) + \beta_3 (\omega_1 - \omega_2) + \beta_4 (\lambda_1 - \lambda_2) + \beta_0 + \varepsilon$$

Where P_x are payouts at station 1 and 2, φ is the distance between the two stations, α_x is the altitude at each station, ω_x is the latitude at each station, and λ_x is the absolute value of the longitude of each station (as longitudes in the western hemisphere are traditionally negative.)

This equation is primarily a difference equation, where we are attempting to explain the difference in payouts by the difference in altitude and geographic coordinates. At first blush, it seems as if the φ variable, distance, is ill-suited for inclusion because distance is strictly positive, and the differences in any part of the equation can easily be negative. However, by imposing a condition of $P_1 >= P_2$ we may ensure symmetry between the left and right sides of the equation; only if $(P_1 - P_2)$ is strictly positive will it reflect a potential linear relationship with φ. Furthermore, distance is a trigonometric function of the individual latitude and longitude variables but is highly correlated to neither. This is because it is a joint function of latitude and longitude, and a degree of longitude is not a constant surface measurement but varies according to distance from the pole. It is more useful to think of the latitude/longitude coordinates as reflecting directionality, and distance as an adjustment for increasing variability at increased distances.

The equation for distance is given thusly:

$$\varphi = R * Cos^{-1}(Sin(\omega_1) * Sin(\omega_2) + Cos(\omega_1) * Cos(\omega_2) * Cos(\lambda_2 - \lambda_1))$$

Where R is a constant reflecting the radius of the sphere we can use to normalize to standard units; the constant for miles is 3963.1.

What we are left with is a description of how each station compares to each other in three-dimensional space, not only in distance (φ) but with x and y coordinates given by the latitude (ω_x) and longitude (λ_x), and z coordinate given by altitude (α_x). The initial hypothesis is that distance (φ) should be positively correlated in both heat and precipitation, meaning that as distance increases, so do the differences in premiums. For rainfall, the rest of the geographic variables are indeterminate, given that coordinates and/or altitude would seemingly have no effect on the sporadic nature of rainfall. For heat, however, we might expect that altitude and latitude have a negative effect on risk; or, in other words, heat risk is decreased by either an increase in elevation or more northerly locations.

Regression Results

Table 4: Regression results for heat risk event

Station	**Bridgeport, NE**	**Bethany, MO**	**Greenville, AL**	**Davis, CA**	**Ithaca, NY**	**Mosquero, NM**
# of Years	104	75	66	83	74	71
Mean CDD	442.13	350.74	603.18	720.63	68.41	282.05
Stations within 50 miles	19	25	21	44	35	16
N	4300	4417	3052	7255	5953	1831

R^2	0.0103	0.0419	0.0010	0.0157	0.0525	0.4921
Distance	-.215 (-2.22)**	.118 (1.46)	.094 (0.64)	.285 (1.99)**	.035 (1.32)	.394 (1.86)*
Alt. Diff.	-.011 (-2.45)**	-.138 (-8.86)**	-.020 (-1.03)	-.020 (-6.86)**	-.008 (-4.67)**	-.210 (-23.79)**
Lat. Diff.	.428 (0.13)	14.897 (3.03)**	-7.660 (-1.42)	31.278 (5.28)**	-15.926 (-14.41)**	60.886 (3.90)**
Long. Diff.	-9.286 (-2.59)**	15.125 (3.52)**	-2.933 (-0.49)	18.269 (4.33)**	7.852 (9.88)**	84.168 (9.84)**
Constant	52.298 (11.42)**	35.226 (9.48)**	66.999 (8.74)**	80.507 (11.99)**	20.724 (15.17)**	69.946 (8.27)**

Table 5: Regression results for precipitation risk event

Station	Bridgeport, NE	Bethany, MO	Greenville, AL	Davis, CA	Ithaca, NY	Mosquero, NM
# of Years	104	75	66	83	74	71
Stations within 50 miles	27	33	41	70	79	35
N	7515	8693	10895	22393	34846	5770
R^2	0.0104	0.0082	0.0131	0.0236	0.0074	0.028
Distance	.074 (2.34)**	.177 (4.94)**	.0001 (0.00)	.049 (8.66)**	.127 (8.91)**	.216 (5.12)**
Alt. Diff.	-.014 (-7.23)**	-.027 (-3.63)**	.017 (4.13)**	-.003 (-18.51)**	.009 (13.36)**	.002 (0.86)
Lat. Diff.	.308 (0.25)	1.821 (0.77)	2.345 (1.84)*	-1.616 (-6.58)**	3.090 (5.26)**	-16.058 (-4.73)**
Long. Diff.	10.054 (7.99)**	10.929 (6.03)**	13.975 (11.21)**	-1.234 (-6.74)**	-.741 (-1.68)*	9.48 (4.26)**
Constant	57.898 (34.48)**	54.71 (31.94)**	63.995 (41.25)**	14.614 (48.68)**	32.353 (43.74)**	66.293 (30.71)**

The first thing to notice when looking at these numbers is that with one exception the R^2 values are usually quite near to zero, which is to say that these geographic variables provide a very poor fit for predicting differences in payout amounts between stations. In and of itself this is scant evidence for the predictive power of the geographic variables on the differences in payouts, but most of the difference is accounted for in the constant term. The coefficients for the geographic variables are quite often significant, but unpredictably so. This may reflect the fact that localized conditions can be expected to have effects on the latitude/longitude coefficients, as directionality within different locations might reflect different geographic characteristics.

The two enduring relationships that can be deduced are the effect of altitude on heat risk and the effect of distance on precipitation risk. The signs are consistent and significant for all stations except Greenville, AL. The coefficient for altitude for the heat risk regressions is consistently negative and significant, which makes sense – we would expect heat risk to decrease as elevation increases. The effect of altitude on rainfall payoffs is unclear, as one might expect – rain likely doesn't consider the altitude of the land on which it is falling. The coefficient attached to distance for rainfall is, with one exception, significant and positive, meaning that as distance increases the difference in the payoffs does too. Or in other words, as distance increases, the payoffs become less accurate. We might expect a similar result for heat, as stations further apart produce more differentiated results, but it seems that temperatures vary continuously throughout a geographic region and the directionality measures are often of more interest.

These relationships may have interesting implications for future efforts to model spatial variability. In effect, the relationship between rainfall and distance is shown to be strong, which indicates that spatial prediction models could have some success. Heat risk, however, is shown to be heavily influenced by altitude and spatial prediction models would do well to account for that effect above and beyond the effect of distance.

These results may seem to be providing little beyond the very obvious – heat risk decreases with altitude because of lower temperatures at higher elevations; likewise, rainfall correlations decrease with distance because of the unpredictable, periodic nature of rainfall. However, there is little evidence for other seemingly obvious implications, like the relationship between latitude and heat risk – we would expect that heat risk would decrease with increased latitudes, but in fact only one of the six coefficients is negative and significant. In fact, it is somewhat remarkable how little we can say about the relationship between simple geographic variables and differences in downside risk. It has been assumed by many researchers that it would be possible to provide a statistical solution to the problem of geographic basis risk; these results belie the fact that weather risk may indeed defeat the ability of statistical methods to predict.

Improving the Fit

There are a few transformations that we can do to improve the fit, which is not a purely academic exercise if our goal is to make out-of-sample predictions for unknown locations. The easiest way to improve the fit of the regression is to include dummy variables for the weather stations and years.

The justification for including dummy variables is thus: it is easy to postulate that each station is to some degree idiosyncratic; these dummy variables are intended to catch the effects of nearby lakes or valleys, or anything else that can't be captured by the simple geographic variables that we use. The dummy variables for each year isolate the amount of variability in any given year because the dependent variable is strictly positive. This will account for any years in which payout differences were more pronounced. Both of these dummy variable types may also be included in a pricing algorithm as well, although if we are pricing a premium for an unknown location for which there have never been weather observations, we cannot use the variables which account for station idiosyncrasies.

In addition, the geographic variables don't explain the *difference* in payouts very well, but there is some evidence that the problem is one of scale. More specifically, this difference equation has no way of distinguishing between payouts which are $0/$200 and $1000/$1200. In both cases the dependent variable will be $200, even though they are quite different on a percentage basis.

There are several potential ways to modify the equation to account for this. One method is to move the P_2 variable to the right side of the equation, where it may be fit with a regression coefficient. This approach improves the fit markedly but necessitates difficult interpretations of the equation. First, if the coefficient attached to the P_2 variable is significantly different than one, it is difficult to interpret what that means, because P_1 and P_2 are identical in nature and the matter of which one is written first depends only on the ($P_1 >= P_2$) condition. Second, if we're trying to make an out of sample prediction, we can't assume that the P_1 variable will be larger than P_2, which may bias the results.

Table 6: Results of transformations in the regression equation

	Original	Incl. Station	Incl. Year	Incl. Station & Year	P1 as Y	All Effects
Rainfall						
DF	34841	34690	34768	34617	34840	34616
R²	**0.0074**	**0.1788**	**0.1056**	**0.2412**	**0.2644**	**0.4435**
Distance	.127 (8.91)**	.104 (7.05)**	.122 (9.02)**	.102 (7.13)**	.133 (9.37)**	.088 (6.22)**
Alt. Diff.	.009 (13.36)**	-.027 (-4.35)**	.007 (10.50)**	-.0243 (-3.98)**	.009 (12.87)**	-.028 (-4.56)**
Lat. Diff.	3.090 (5.26)**	8.155 (0.81)	3.049 (5.37)**	10.343 (1.06)	2.999 (5.12)**	17.119 (1.76)*
Long. Diff.	-.741 (-1.68)*	12.824 (1.73)*	-.534 (-1.26)	19.805 (2.75)**	-.607 (-1.38)	17.933 (2.50)**
Constant	32.353 (43.74)**	23.671 (8.27)**	25.335 (2.50)**	10.075 (1.03)	30.445 (40.35)**	10.707 (1.11)
P2	--	--	--	--	1.413 (110.38)**	.419 (26.27)**
Heat						
DF	5948	5889	5875	5816	5947	5815
R²	**0.0525**	**0.2294**	**0.5404**	**0.6134**	**0.6771**	**0.8720**
Distance	.035 (1.32)	.064 (2.21)**	.008 (0.46)	.086 (4.11)**	.037 (1.53)	.070 (3.69)**
Alt. Diff.	-.008 (-4.67)**	-.028 (-3.35)**	-.008 (-6.60)**	.022 (3.59)**	-.0122 (-8.14)**	.035 (6.14)**
Lat. Diff.	-15.926 (-14.41)**	-7.074 (-0.65)	-11.570 (-14.69)**	35.801 (4.55)**	-17.146 (-16.81)**	35.523 (4.97)**
Long. Diff.	7.852 (9.88)**	-24.862 (-2.52)**	5.436 (9.62)**	5.653 (0.78)	7.853 (10.71)**	12.452 (1.88)*
Constant	20.724 (15.17)**	22.048 (5.34)**	8.27 (1.51)	-28.215 (-4.82)**	13.025 (10.16)**	-37.615 (-7.06)**
P2	--	--	--	--	1.124 (110.29)**	.717 (57.20)**

The results of these transformations for all stations and years of data are presented in Table 6. The R^2 calculations are highlighted and improve considerably, but remember standard caveats on the effects of R^2 values when adding dozens, if not hundreds, of variables to the equation. Of course, the greatest effect on the R^2 value comes from the addition of a single variable (and the manipulation of the Y variable) that accompanies moving P_2 to the right side of the equation.

Remarkably, the P_2 variable is highly statistically significant in all places that it is introduced. The two variables that we identified as causal in the regression equation maintain their significance and sign through all modifications even as all other variables experience widely ranging results.

Out of Sample Predictions

Table 7: Out-of-sample predictions

Heat	**Prediction**	**Obs. Average**	**Difference**	**Correlation**
Geo. Variables Only	$25.27	$64.39	-154.8%	-0.4422
With Station & Year Effects	$32.20	$64.39	-100.0%	0.4623
And moving P_2 to right side	$67.30	$64.39	4.3%	0.8956
Rainfall				
Geo. Variables Only	$37.34	$64.17	-71.8%	0.4294
With Station & Year Effects	$51.33	$64.17	-25.0%	0.5217
And moving P_2 to right side	$66.81	$64.17	4.0%	0.6201

This leads us to Table 7, which shows the results of out-of-sample predictions of payoffs in Ithaca, NY for heat and rainfall using several different types of effects for illustration – first with the simple geographic variables, then including the station and year dummy variables, and finally when moving P_2 to the right side of the equation. Also of note is that this prediction was only performed when Ithaca was the station listed first (i.e. the P_1 variable), the consequence of which is that the payouts are significantly higher ($64.39 and $64.17) than the long-term averages as presented in Table 1 ($20.76 and $20.87). Whether this has implications for the end results is an important consideration.

What this shows is that the predictions with geographic variables are not very accurate, but improve with the addition of the station and year effects. The strongest effect is obtained by moving P_2 to the right side of the equations, which may make sense – the weather observations are the strongest piece of information we have about prevailing conditions in any given year and by taking the difference we often censor that important piece of information. In any case, it must be said that the geographic variables seem to be useful only in the optimization of an already robust distribution – in any successful prediction presented herein, the "heavy lifting" is done by the station, year, and P_2 effects. And in the case of rainfall, this entire exercise has resulted in payouts that are in fact slightly worse than the very simplistic approach taken in Table 1 of simply averaging payouts for each station within 67 miles of Ithaca.

The Next Step

While it may be difficult to propose an index insurance contract based on an average of every weather station within 67 miles, even if it is most accurate, the concept of doing so reveals a larger principle. Woodard and Garcia (2008b) write that "portfolio" of derivatives from established derivatives markets in large cities could be a solution to the problem of spatial basis risk. While the authors of this paper would not advocate for using information from stations that are very far away, it may be possible to extend this concept by offering an index insurance contract which is a configured portfolio of local stations. This is because our results strongly indicate that the strongest predictor of a payout at any given station is in fact whether or not there is a payout in a nearby station. Based on the discussion presented in this paper, the portfolio would use weather stations as close as possible and explicitly include elevation and distance into the portfolio selection criteria, if not the other variables which may or may not be significant in any geographic location.

The next step for this research would be to construct a function which could serve to provide guidelines for a "portfolio" of weather index insurance priced to the nearby stations. The amount of index insurance purchased for the portfolio from nearby stations would depend on the similarity of the pertinent geographic characteristics. For example, in a topologically diverse area such as the Finger Lakes region, this could mean a station that is quite removed in distance but similar in altitude. Particularly in regard to the relationship between heat and elevation, the closest station in a geographically diverse region may not always be the most similar.

A portfolio of index insurance chosen according to pertinent spatial variables would offer several advantages. First, since the pricing would be done at a known location according to historical burn rates, the price calculated would be straightforward and done in a manner consistent with previous applications of index insurance. This would likely lead to easier adoption of the technology, as insurance companies would not need to build additional safeguards for risk into the model than those that already exist in established methodology. Second, by pricing only at known locations, the concepts behind this method a more transparent to the layperson than a mathematically rigorous treatment designed to predict risk at any given point. It is too true that as the complexity of mathematical instruments increases, the basis risk inherent in the equation decreases, as does comprehension. The portfolio approach is a simple extension of existing methodology and has the advantage of being easier to comprehend and price.

Lastly, this approach would likely take the form of a set of suggestions based on the best available evidence that would be easily modified to include additional information. In other words, this approach would allow the farmer to individually tailor their index insurance portfolio according to their perceived risk. The farmer is likely to have the most detailed local knowledge as to prevailing weather patterns and will even have information as to the effects of other geographic characteristics that do not show up in the regression equation, such as mountains, bodies of water, or even the general pattern by which rain falls on their fields. The ultimate source of information as to the minimization of basis risk is with the farmer and the portfolio approach would allow the flexibility for farmers to hedge their weather risk in the manner they best see fit.

The one major disadvantage of this method that must be mentioned is that it would benefit from a rich series of historical data in the area surrounding the point of interest. Unfortunately, this may preclude a portfolio method from being used in developing countries, which are areas of the world that could benefit greatly from the adoption of this technology. Further research should be done to test the applicability of this so-called "portfolio" approach to pricing weather risk.

Conclusion

This paper tackles the problem of spatial basis risk for weather index insurance, which is a basic and fundamental problem in the widespread adoption of the technology, and shows that there is no easy solution at present. Even when putting aside considerations of the weather/yield relationship and the predictive power of the selected index on crop yields, geographic basis risk will likely remain a persistent problem due to complex interactions of weather, space, and geography. Our search for a general principle for pricing risk at unknown locations has all but failed, which is of interest in and of itself.

The two enduring relationships presented in this research are the relationship of altitude on heat risk and distance on precipitation risk, which may have interesting implications for future efforts at spatial prediction models. Rainfall is proven to be heavily influenced by distance measures, which holds promise for future efforts to model spatial variability, but future efforts to model heat risk should explicitly account for altitude as the predominant variable of interest.

Our results also strongly show that the best predictor of a payout at any given location is the presence of a payout at another location, which entails that a portfolio method for buying index insurance may have advantages above and beyond a strategy which predicts risk at unknown locations. These advantages include transparent pricing and ease of understanding, as well as a simple way for the consumer to configure the product for their own needs.

Although our search for a general principle was unsuccessful, the concepts produced in this research provide insight into the possibilities and challenges present in pricing spatial basis risk. It is expected that future researchers will be able to carry on the material presented in this paper by building upon the relationships that we did find to be evident in the struggle to overcome a persisting problem in the adoption of weather index insurance.

References

Miranda , M., C. Gonzalez-Vega, and M. Villafani-Ibarnegaray. 2010. "Index Insurance and the Optimal Management of Loan Portfolios in Rural Microfinance Institutions in the Presence of Systematic Risk." AEA Annual Meeting Paper, Allied Social Sciences Associations Meeting, Atlanta, GA, January 3-5, 2010.

Morduch, J. "Micro-Insurance: The Next Revolution?" In: Banerjee Abhijit, Benabou Roland, Mookherjee Dilip., editors. *What Have We Learned About Poverty*? Oxford University Press; 2006.

Paulson, N.D., and C. E. Hart, (2006), "A Spatial Approach to Addressing Weather Derivative Basis Risk: A Drought Insurance Example," Selected paper, American Agricultural Economics Association Meeting, Long Beach, CA, July 23-26, 2006.

Richards, T.J., M. Manfredo, and D. Sanders, (2004), "Pricing Weather Derivatives," *American Journal of Agricultural Economics* 86(4): 1005-1017.

Schlenker, W., and M.J. Roberts. (2006). "Nonlinear Effects of Weather on Corn Yields." *Review of Agricultural Economics*, 28(3): 391-398.

Skees, J.R. (2008). "Innovations in Index Insurance for the Poor in Lower Income Countries." *Agricultural and Resource Economics Review* 37/1:1-15.

Turvey, C.G. (2001). "Weather Insurance and Specific Event Risks in Agriculture." *Review of Agricultural Economics* 23(2):333-351.

Turvey, C.G., and M. Norton. (2008). "An Internet-Based Tool for Weather Risk Management." *Agricultural and Resource Economics Review* 37/1:63-78.

Vedenov, D.V., and B.J. Barnett. (2004). "Efficiency of Weather Derivatives as Primary Crop Insurance Instruments." *Journal of Agricultural and Resource Economics* 29(3): 387-403.

Woodard, J. D., and P. Garcia. (2008a). "Weather Derivatives, Spatial Aggregation, and Systemic Risk: Implications for Reinsurance Hedging." *Journal of Agriculture and Resource Economics* 33(1).

Woodward, J.D., and P. Garcia. (2008b). "Basis Risk and Weather Hedging Effectiveness." *Agricultural Finance Review* 68(1):99-118.