**Accounting for Weather Probabilities in Crop Insurance Rating**

Keith H. Coble[*], Mary Frances Miller[†], Roderick M. Rejesus[‡], Ryan Boyles[§], Barry K. Goodwin[**], Thomas O. Knight[††]

*Selected Paper prepared for presentation at the Agricultural & Applied Economics Association's 2013 Crop Insurance and the Farm Bill Symposium, Louisville, KY, October 8-9, 2013.*



[*] W.L. Giles Distinguished Professor, Department of Agricultural Economics, Mississippi State University, Mississippi State, MS 39762, 662-325-6670, coble@agecon.msstate.edu
[†] Senior Consulting Actuary, Select Actuarial Services, 28 White Bridge Road, Suite 205, Nashville, TN 37205, maryfrances.miller@selectactuarial.com
[‡] Corresponding Author and Associate Professor, Department of Agricultural and Resource Economics, North Carolina State University, Raleigh, NC 27695, 919-513-4605, rod_rejesus@ncsu.edu
[§] State Climatologist and Director, State Climate Office of North Carolina and Extension Assistant Professor, Department of Marine, Earth, and Atmospheric Sciences, North Carolina State University, Raleigh, NC 27695, rpboyles@gmail.com
[**] William Neal Reynolds Professor, Departments of Economics and Agricultural and Resource Economics, North Carolina State University. Raleigh, NC 27695, barry_goodwin@ncsu.edu
[††] Horn Professor and Emabeth Thompson Professor of Risk Management, Department of Agricultural and Applied Economics, Texas Tech University, Lubbock, TX 79409, 806-742-1921, ext 255, tom.knight@ttu.edu

**Accounting for Weather Probabilities in Crop Insurance Rating**


ABSTRACT

The US crop insurance program previously used a simple average of equally weighted historical loss cost data to serve as the backbone for estimating crop insurance premium rates. This article develops a procedure for weighting the historical loss cost experience based on longer time-series weather information and improve statistical validity of estimated premium rates. It was determined that the best weather data to account for weather probabilities in crop insurance premium rating is the National Climatic Data Center's Time Bias Corrected Divisional Temperature-Precipitation-Drought Index data, also called the Climate Division Data. Using fractional logit and out-of-sample competitions, weather variables can be selected to construct an index that would allow proper assessment of the relative probability of weather events that drive production losses and to construct proper "weather weights" that can be applied when averaging historical loss cost data to calculate rates. A variable width binning approach with equal probabilities was determined as the best approach for classifying each year in the shorter historical loss cost data used for rating. When the weather weighting approach described above is applied, we find that for apples, barley, cotton, potatoes, rice, and spring/winter wheat, the weather weighted average loss costs at the national level tend to be smaller than the calculated average loss costs without weather weighting. However, for corn, cotton, sorghum, and soybeans, the weather weighted average loss costs at the national level tend to be larger. Around 51% of the counties have weather weighted average loss costs lower than the average loss costs without weather weighting.



*Keywords:* Crop Insurance; Premium Rating; Weather Weighting

JEL Classification: G22, Q10, Q18

**Accounting for Weather Probabilities in Crop Insurance Rating**

## Introduction

The US crop insurance program utilizes historical loss experience as the foundation of its premium rating system. In particular, the Risk Management Agency (RMA) formerly used equally-weighted, adjusted, historical, loss cost data (i.e., the ratio of indemnity payments to total liability) for a crop in a county as the backbone of the premium rating procedure. The former system used county loss cost experience data back to 1975 (where available) and gives each year's experience equal weight. The simple average of the equally weighted historical loss cost data then serves as the main basis for crop insurance premium rates, which is consistent with the fundamental principle of insurance ratemaking where the rate represents an estimate of the expected value of future losses.

The weighting of historical loss data (say, for 38 years from 1975 to 2012) is an important issue because it is directly related to the question of whether this series truly represent the "longer" term weather experience that needs to be captured to accurately estimate premium rates. In many lines of insurance, 38 years of loss history would be considered a very "long" time-series data to use in rate making. However, 38 years may be a relatively short series for accurately reflecting probabilities of weather events that are the most dominant factor in crop losses. For example, in simple averaging of loss cost data to calculate county base rates, the recent 2012 drought year is given a 1/38 weight but the long term frequency of the weather events that drove these losses may be greater or less than 1/38. It could be that the 2012 drought was a 1 in 20 year event rather than a 1 in 38 year event. If so, a larger weight than 1/38 would be appropriate for that year. Alternatively, it could be that the 2012 drought only occurs 1 in 50 years in a longer weather time series and should be given less weight than 1/38. The intent of weather weighting of loss cost data is to bring additional information from a longer series of weather variables to more properly weight the loss cost data used to calculate average county rates.

The objective of this study is to develop a methodology for weighting the historical loss cost experience used in calculating crop insurance premium rates (i.e., specifically for the yield-based Actual Production History (APH) insurance product). A detailed investigation is performed to develop an optimal methodology for weighting, or otherwise adjusting, RMA's historical loss cost data in order to maximize its statistical validity for developing premium rates. Multiple weighting approaches are evaluated based on statistical validity, feasibility, sustainability, and a balance of improvement versus complexity. In particular, we explicitly consider the Palmer Drought Index and other weather variables in the development of the weighting methodology. The RMA has now largely adopted the technique proposed in this study.

## Data Issues and Conceptual Considerations in Weather Weighting Historical Loss Data

### *Historical Loss Cost Data*

The objective of premium rate setting in insurance is to provide an estimate of the expected value of future losses (or costs). In the US crop insurance program, rates are typically set separately for each crop because different crops are subject to different perils and, consequently, varying loss costs. In addition, one of the most important components of the US crop insurance rating system is the use of aggregate county level loss cost data to first estimate a county base rate.[1] County level data is initially used in crop insurance rating because it is rare that an individual insured will have a sufficiently long time series of historical loss data to be able to directly calculate an

[1] The county base rate is then adjusted based on different factors (such as choice of coverage level, crop type, insured's yield relative to the county average, etc.) to provide an individualized rate for each insured. See Coble et al. (2010) for more details of the full RMA rating procedure.

estimate of individual expected losses. Therefore, aggregating experience data of individual farmers in a county, for a particular crop, is the approach followed in US crop insurance rating.

The starting point for constructing the historical county level experience data is to collect observed indemnity and liability data from individual insureds within a county for a particular crop. The individual experience data collected is then "normalized" by removing and/or adjusting individual level insurance data so that the resulting experience data are comparable (i.e., normalized to a common base). For example, prevented planting is not considered to be a production loss since the crop was not planted on the ground (i.e., typically due to weather constraints). Hence, prevented planting indemnities and liabilities are typically excluded when constructing the historical county level data used for estimating county base rates (called the StatPlan database). [2] Revenue insurance experience is also utilized in the construction of the county level data by recomputing indemnity and liability data from these revenue policies as if it were a yield insurance policy.

Once the indemnities and liabilities are normalized, the historical loss cost data can be calculated and the county base rate can be estimated. More formally, an average of the equally-weighted yearly county level loss cost data is used to calculate a county base rate for which the individual premium rates are derived:

(1) $$\textit{County Base Rate}_i = E(LC_{ij}) = \sum_{j=1}^{n} f(LC_{ij}) \times LC_{ij} = \frac{1}{n}\sum_{j=1}^{n} LC_{ij} ,$$

where $LC_{ij}$ is the loss cost ratio from the StatPlan database for county $i$ and year $j$ = 1, …, $n$, and $f(LC_{ij})$ is the calculated probability based on probability density function (pdf) of $LC_i$ over $n$ discrete time periods. Note that catastrophic loading procedure is also utilized in the estimation

[2] A detailed discussion of the individual adjustments made to indemnity and liability data to construct the StatPlan data base is in Coble et al. (2010).

of county base rates to reduce the influence of outliers in the historical loss experience. Catastrophic loading is intended to remove anomalous experience from the county data while preserving normal loss experience. In general, losses deemed catastrophic are spread across all counties for a crop in a state. The previous catastrophic loading procedure censors the county loss experience at the 80th percentile of the historical county experience. All indemnities above the truncation point are aggregated to the state level. Once the catastrophic portion has been removed from the county experience, the uncensored observations below the 80th percentile and the censored values of the censored observations remain; and these values are used in (1).

Using equal weights in the averaging process in (1) implicitly assumes a uniform pdf where observed loss cost values are equally likely to occur. But given that weather is the major factor that drive agricultural losses the equal weights assumption may be problematic, since weather distributions may not follow a uniform distribution. Conceptually, probability information from a long time series of weather data may help augment the inherently small sample in the StatPlan database and better estimate the weights when averaging loss cost data.

Another related issue is the need to capture a good range of *LC* values typically from a long time series of loss cost data. Ideally one would like a long enough time series such that the range of all possible LC magnitudes is represented. However, as mentioned above, the StatPlan database typically starts from 1975 and data from this time series length may not be able to provide a good range of LC values. Hence, from this perspective it may be desirable to have the longest possible time series of loss data for rating purposes.

But it should be noted that there is an inherent tension between the need to use long time series data to properly capture the probability and magnitudes of infrequent but catastrophic weather events versus the concern that, with a long time series, loss data from the more distant past may not be truly representative of the current production and insurance environment such that the measured losses at that time would not be comparable to more recent losses. For example, contract specifics will have changed over time and, thus, effectively modifying expected losses.  Also, data quality and credibility has improved over time. Finally, agricultural production technology has evolved rapidly such that the effect of a particular weather event on insurance losses may be different from what would have occurred in the past. With a proper weather weighting procedure, it is possible to utilize a shorter series of loss data to assure improved representativeness, but also use a long series of weather data to more properly weight infrequent catastrophic weather events that cause crop loss.

*Weather/Climate Data*

In developing a system to weight loss experience data using longer weather or climate data, one has to consider the following issues: (1) the weather or climate data that will be utilized for weighting (e.g., the length of the data, the degree of coverage and/or level of aggregation, the relationship of such weather to losses, and the availability of weather variables), and (2) the development of a procedure to properly weight each year in the short loss data (e.g., categorizing each loss data year and creating weights for each year in a manner that is consistent with other parts of the rating process).

*Weather Data Considerations*. There is an abundance of weather data available in the US that can be used for weather-based weighting of loss experience data. However, there are several issues to consider in choosing the weather data to be used. First, one has to consider the length of the different climate data series that are available. In the context of weighting insurance data, one would like to have the longest series of historical weather data available. This would help ensure that different weather outcomes, especially the rare extreme weather events that cause losses, would be adequately represented in the longer data series. Information about the probabilities of different weather events will be better captured if one has a very long climate data series.

However, the need for a long data series must be balanced with the second issue to consider – the degree of coverage and level of aggregation. For example, there may be weather data that are available for 200 years, but these data sets may only contain data for a particular part of the country and/or only at the national level. Crop insurance covers a large portion of the US and so weather data covering most or all states are needed. In addition, there is significant heterogeneity of the weather events that drive losses at the county level for a particular year. There is value in having data at a lower level of aggregation (i.e., county level or 5 x 5 mile grids) rather than at the national level only. However, in using weather/climate data at lower levels of aggregation, it may be the case that data interpolation methods were involved in the construction of the data, especially at the sub-county level where there frequently are no weather stations in a particular location.

Another factor to consider in choosing the weather or climate data to use in weighting loss experience is the availability of different weather variables that can be used. Longer series of climate data may be available for some basic variables like temperature or precipitation, but variables like drought indexes may not be available for this longer period of time. Climate data at lower levels of aggregation and with wider coverage may only be available for certain weather variables and may be absent for others. Hence, to have flexibility in determining the weather variables that can help to explain losses, the availability of different weather variables in a particular climate data set is also an important consideration.

Finally, in choosing climate data for weather weighting crop insurance loss cost data, the source of the data and the availability of the data in the future are also important considerations.

The source of the climate data has to be reliable and must have a good reputation in terms of reporting weather/climate data. Moreover, there should be a reasonable expectation that the weather/climate data will continue to be available in the future to support updating of weather weights as more data become available.

*Weather Data Choice*. There are several datasets that partially meet the four weather data considerations above. First, the National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center (CPC) Unified Precipitation Analysis is an interpolation of the available point-based precipitation gauge data collected by both NOAA and USGS. It meets the second and fifth considerations listed above, but provides only information on precipitation and has data only since 1948. Important information on temperature and drought are not provided, and these data do not allow for characterization of the relative frequency of known extreme drought events in the 1920s and 1930s nor hurricane or flooding events prior to 1948.

A national analysis of Palmer Drought Severity Index developed by Dai et al. (2004) meets some of the considerations listed above, but is not updated regularly and provides drought severity information only every 250 kilometers which is insufficient to explain local loss experience.

Another group of data that partially meet the considerations listed above is atmospheric model simulations, including NCEP re-analysis and the North American Regional Reanalysis (NARR). These products meet criteria second, third, and fourth considerations, but NCEP re-analysis (and similar) only provide information since 1948 and NARR only since 1979.

The data collection that best meets all weather data considerations listed above is the National Climatic Data Center’s Time Bias Corrected Divisional Temperature-Precipitation-Drought Index data, also called the Climate Division Data. Climate Division data provide monthly, serially complete information on temperature, precipitation, relative severity of dry and wet periods using drought indexes, and degree day metrics of heat and cold accumulation since 1895 for the continental United States, grouped into 344 divisions. Updates are operationally provided each month by NOAA National Climatic Data Center. A nice description of the history and current status of climate division data is provided by Guttman and Quayle (1996). More technical details on the data and adjustment methods are provided in NCDC (1994) and Karl et al. (1996).

Climate Division data are produced using more than 5,000 National Weather Service Cooperative observer gauge reports. Climate Division boundaries group stations of similar climate into regions that follow state political borders. In most cases, the climate division boundaries also follow county boundaries. However, in regions with more complex geography (including some states with complex topography and/or shorelines), climate division boundaries follow river basins within each state. While climate divisions were originally designed in 1912, boundaries were adjusted in the 1940s to align with crop reporting districts or drainage basins. In some instances, climate divisions cross or split counties. The Climate Division boundaries and consequent assignment of counties to particular Climate Divisions are shown in Figure 1. This allocation is based on relative area, geography and other factors.

There are limitations to using Climate Division data. Climate division boundaries are not always delineated for climate homogeneity. Especially in the mountainous terrain of the western US, the boundaries follow drainage basins and all locations within those boundaries are not likely to have very similar climate characteristics as climate changes quickly with changes in elevation. Another weakness is that the station network used for each division calculations is not constant. Stations move, cease operation, and new ones are introduced throughout the history of the observing network. This introduces some error with any divisional calculations. Another weakness is the accuracy of division level data prior to 1931, when regression equations are used to estimate division-level data from statewide average data that were standard during that period. Despite these weaknesses, Climate Division data provide the best operationally available climate information for crop loss analysis. They provide serially complete national coverage (with no missing data) at a geographic scale sufficient to characterize local climate extremes with a period of record sufficient to identify the relative frequency of climate events that may be associated with loss experience.

*Merged Loss Cost and Weather Data*

The development of the weather weighting procedure starts by merging the climate data set with the StatPlan loss experience data. Note that the climate data are observed at the climate division level as described above, while the RMA StatPlan data are reported at the county level.[3] Therefore, all counties within a particular climate division have the same weather data and the loss data also must be aggregated to the climate division level. This is done by summing the indemnities and liabilities of all counties within a climate division level and then calculating loss cost ratios based on these summed amounts. The climate division data can then be used to generate a weather index that is needed for classifying loss years, while the county data can be used in averaging the loss cost data to calculate a county base rate.

## Empirical Approach to Weather Weighting Historical Loss Cost Data

*Weather Index Development*

A critical component in the development of a weather weighting approach is the choice of the weather variables that are used to determine the relative weights assigned to each year of loss

[3] The county loss data utilized in this study are typically aggregated for all types/practices (with the exception of wheat, where the data are separated to identify winter and spring wheat). This type of aggregation is consistent with the county level data used in calculating the base county rate (see Coble et al., 2010 p. 38).

data. One can use a single weather variable or a combination of different weather variables. Based on the literature (Wilhemy, Hubbard and Wilhite 2002) and the expert opinion of the climatologist on our study team, we chose to examine a parsimonious set of weather variables – the Palmer Drought Severity Index (PDSI) and Cooling Degree Days (CDD). PDSI is a particularly good weather variable to examine because it subsumes effects of both precipitation and temperature and provides a locally relative scale ranging from very wet to very dry conditions. Wilhemy, Hubbard and Wilhite (2002) show that much loss experience is associated with drought conditions, but PDSI also allows for very wet (flood) conditions that may also be associated with loss. CDD allows for examining heat units for a particular time period that affects crop growth. CDD is equivalent to Growing Degree Days (GDD) at base 65F, and allows exploration of loss experience that may be associated with extended cold or heat that would not be captured in PDSI.

For the PDSI, we created two variables to represent positive PDSI and negative PDSI values. Positive PDSI values represent wet spells (i.e., larger positive numbers indicate more moisture) and negative PDSI values represent drought conditions (i.e., larger negative numbers represent more severe drought conditions). In addition, the positive and negative PDSIs we use are limited to the May-June and July-August periods (i.e., average May-June and average July-August PDSIs are utilized in the study). In summary, four PDSI measures are examined in the development of our weather index – May-June PDSI for positive values (*mj_pdsi_p*), May-June PDSI for negative values (*mj_pdsi_n*), July-August PDSI for positive values (*ja_pdsi_p*), and July-August PDSI for negative values (*ja_pdsi_n*). The CDD variables used in developing the weather index are total season CDD (from May to September) (*total_cdd*) and June-July total

CDD (*jaj_cdd*). The June-July periods are periods in which crop growth is frequently adversely affected by heat units.[4]

Based on these six weather variables, an index is created by estimating a fractional logit regression model (at the climate division level) where the dependent variable is the climate division adjusted loss cost ratio and the independent variables are the six weather variables discussed above (See Papke and Wooldridge 1994). Fractional logit regression is used to account for the proportional nature of the data and censoring of loss costs at zero and one. This approach ensures that predicted values do not fall below zero or above one.[5] Based on our investigation of the degree of censoring of the data at zero, we believe that using the fractional logit is appropriate in this case. The degree of zero censoring in the data ranges from 6-11% for corn and soybeans, to about 30% for barley and potatoes. On the other hand, the degree of censoring at one is significantly lower in the data and it is below 1% for most crops (the exception is apples with censoring at one of about 1.1%.

To have an even more parsimonious model specification, an out-of-sample competition for each state is conducted to determine which combination among the six initial weather variables best predicts losses (i.e., in this case which combination best predicts adjusted loss cost out-of-sample). A minimum mean square error (MSE) criterion is used to evaluate the model with best out-of-sample predictions:

---

[4] These six weather variables apply to all crops except winter and spring wheat. For winter wheat, the following variables are used: Sept./Oct average PDSI (positive and negative), April /May average PDSI (positive and negative), September to May total season CDD, and March to April total CDD. For spring wheat, the following variables are used: April/May average PDS (positive and negative), June/July PDSI (positive and negative), April to August total season CDD, and May to June total CDD. Further note that durum wheat type has been aggregated with spring wheat.

[5] Note that ordinary least squares (OLS) regression can also be used to estimate the index. The disadvantage of OLS is that predictions are not constrained to lie on the [0,1] interval. Nevertheless, one can argue that the predicted loss costs here are only used as a "tool" to rank the years in terms of having "good" vs. "bad" weather (i.e., one could interpret negative values as indicating good weather years). The magnitudes of the predictions are not used 'per se'. Using the OLS model to estimate the model did not result in significantly different classifications of the loss years (relative to the fractional logit model). However, we recommend using the fractional logit given the degree of censoring in the data and the intuitive concept of limiting predicted loss costs between zero and one.

(2) $$MSE = \left( \frac{1}{n} \sum_{i=1}^{n} e_i^2 \right),$$

where $e_i$ is the difference between the actual adjusted loss cost and a predicted adjusted loss cost (out of sample) based on the fractional logit regression model.

A lower MSE means that there is a smaller discrepancy between the actual and predicted adjusted loss cost ratios and one would prefer the combinations of weather variables that produce the lowest MSE values. Note that we run independent regressions for each climate division within the state (i.e., climate divisions do not cross state lines), but undertake the out-of-sample competition to find the best combination of weather variables for the entire state. This implies that each regression model is estimated independently but a common specification, in terms of the weather variables included in the regression model, is applied for all climate divisions within a state for each individual crop. In other words, for a crop in a state, the same weather variables are used in the loss-cost regression though parameters on weather variables may differ across climate divisions.

To facilitate the out-of-sample competition for each state, we limit the number of weather variable combinations to be considered to seven: (1) May-June PDSI positive and May-June PDSI negative, (2) July-August PDSI positive and July-August PDSI negative, (3) total season CDD and June-July total CDD, (4) May-June PDSI positive, May-June PDSI negative, July-August PDSI positive, and July-August PDSI negative, (5) May-June PDSI positive, May-June PDSI negative, total season CDD, and June-July total CDD, (6) July-August PDSI positive, July-August PDSI negative, total season CDD, and June-July total CDD, and (7) May-June PDSI positive, May-June PDSI negative, July-August PDSI positive, July-August PDSI negative, total season CDD, and June-July total CDD. Limiting the combinations to these seven choices and estimating the model for each crop, covering all states allows for less of a computational burden.

Once the optimal combination of weather variables is chosen for a particular crop and state, this combination of weather variables is used to produce a weather index for all of the climate divisions within the state producing the crop. Essentially, the predicted values of the "best" regression model specification are used as the weather index for each year of weather data. Using predicted values (i.e., predicted loss costs in this case), makes it possible to "backcast" a weather index for each year in which weather data are available (e.g., from 1895 onwards) even when there are no available loss experience data for the pre-crop insurance years. The relative probability of an extreme weather event (or an extreme loss event) can therefore be assessed over a 116 year time span (1895-2010) based on the predicted values. For example, the weather index for 1988 can be compared to other years from 1895 onward to determine the relative probability of this weather event occurring in the larger sample.

A concern with using the predicted values is that there may be cases when even the "best" combination of weather variables does not produce a statistically significant model that explains losses over time. For example, in some climate divisions, the Pearson chi-square test of overall model fit for the preferred model specification is not statistically significant and the correlation of the predicted values with the actual loss costs is actually negative. This means that the weather variables we considered do not have enough power to explain the pattern of losses observed over time and that there is no significant positive correlation between the model predictions and the actual loss costs. We flag these cases, and the weighting methods based on the weather index developed are not applied.

*Approaches to Loss Year Classification and Weight Assignment*

Using the predicted loss cost values from the regression model, each year needs to be classified and assigned a weight that represents its likelihood as indicated by the longer weather series.

There are a number of ways to classify a year and assign a weight. One approach is to generate a histogram with equal bin widths and variable probabilities (or frequencies) (see Coble et al., 2010, p. 85 and Figure 2).[6] The bins or groupings with equal widths can then be used to classify each year of the loss experience (i.e. which bin does the loss year belong to given the actual experience) and the probability associated with the bin assigned to the year will serve as the weather weight. An alternative to the histogram approach is to develop variable bin (or grouping) widths with equal probabilities associated with each bin (See Figure 3). The bins or groupings will again be used to classify each year, but since these are variable width bins with equal probability, there is no need to have differential weights for each actual year of experience.

In both of these loss classification procedures, one has to evaluate the number of bins to be used and make sure that all bins are represented in the shorter loss data used in averaging the loss costs (i.e., the empty bin problem). If not, the weighted average may not fully reflect the available historical experience. In addition, the complexity of the procedure and the ease of implementation should also be a considered in choosing the approach to classify and assign weights to the actual loss years. With these considerations in mind, we believe this variable bin width approach may be better than a standard histogram approach because this mitigates the "empty bin" issue described above. That is, the likelihood of having empty bins for the years with loss data (1980-2009) is smaller under this approach as compared to a histogram approach with equal bin widths and variable probabilities. The number of bins in the variable bin width with an equal probability approach tends to be greater than if we used the histogram approach. Moreover, the variable bin width with equal probability approach is a fairly straightforward

[6] Alternative methods such as generating kernel densities or fitting parametric distributions can also be used instead of histograms. However, one should recognize that these more complex procedures may have implications for implementation. One has to weigh the relative benefits of more complex approaches against the efficiency and ease of more simple approaches (like using a histogram).

method compared to the approach of using kernel densities or parametric distributions. This "simplicity" facilitates the practical implementation of this procedure for multiple crops and for nationwide coverage.

*Description of Variable Bin Width Approach to Weather Weighting*

The variable bin width approach to weather weighting is implemented by first determining the number of bins or percentiles and assigning the predicted loss costs to the appropriate bin or percentile cut-off. For example, assuming that we are interested in 10 bins we would like to find the predicted loss costs in the long history of weather data that correspond to the 10th, 20th, 30th, 40th, 50th, 60th, 70th, 80th, 90th percentile, in addition to the minimum and maximum values. In this case, we have variable width bins, since the ranges of the loss cost values used to delineate the bins are not equal across bins, but the probability of falling into each bin is always equal to 10%. If the predicted values are normally distributed, the tails (at both ends of the distribution) tend to have wider bin ranges since only a few observations fall in these areas, but the middle bins tend to have smaller widths because a lot of observations fall in these middle bins.

Once the variable width bins are delineated, the predicted loss cost value for each year (from 1895 onward) can be classified and assigned to the bin in which it falls. Using the above example, if the bin width for the 10th bin (from the 90th percentile to the maximum) is, say, from 0.09 to 0.15 and the year 1988 predicted loss cost is 0.13 (i.e., one of the high loss years), then year 1988 is in the 10th bin. Each year is similarly classified using predictions from the fractional logit regression models. Since the probability of each bin is equal in this approach, there is no need to assign a specific differential weight to each bin.

As mentioned above, one issue that needs to be addressed is the number of bins to assume and the possible existence of empty bins during the years with loss cost data (from 1980 till 2009). Once the years from 1895 onward are classified based on the weather index, the RMA's actual adjusted loss cost data from 1980 till 2009 are utilized to calculate the average loss cost for a county. Hence, it is possible that years from 1980 to 2009 do not contain a dispersion of data such that each bin has one or more loss costs (i.e., not all bins are represented in the 1980-2009 period). For example, it may be that no year in the 1980 to 2009 period is classified as falling into bin 9. This will have adverse implications for the calculation of the average loss costs if not all bins are represented in the 1980-2009 period (i.e., not all bin probabilities are represented). In particular, a range of observed weather history is not being captured in the weighting of loss costs. Therefore, to address the issue of empty bins and, at the same time, determine the appropriate number of bins, the approach we pursue is to first look at 15 bins and then move down one bin at a time (i.e., from 15 till 2 bins) to establish the largest number of bins for which there are no cases of empty bins in the years with loss data (1980-2009). This is done for each climate division, and so the number of bins may vary for each climate division within a state.

A hypothetical example of bin classification results for soybeans in Mississippi is presented in Table 1. In this example, the number of bins is 10 and this assures that there are no "empty bins" from 1980-2009. All bin classifications are represented in the 1980-2009 data. We also show in this table that the model insignificance flag and state proxy flag are both equal to zero, which means that the model fit results for this climate division is significant and the number of observations used in the estimation is at least 10.

**Loss Cost Averaging Procedure**

After each year is classified into a particular bin at the climate division level (for all 116 years), the classified data for each year and the insignificance flags (based on regression model) are then merged back with the original county level loss data. Since the regressions and year classifications based on the weather indexes are done at the climate division level, all counties within a particular climate division will have the same year classification and insignificance flags.

The average loss costs are next calculated using the 1980-2009 data where there are available actual loss cost values in the StatPlan data. We first calculate the aggregate loss cost for each county, which is the current procedure used for computing the county base rate. Then we do a “weather weighting” average of loss costs for each county. This weather weighting is done by first taking the average loss cost within each of the defined bins and then taking the “average of the average loss costs” across the bins. For example, if there are 9 bins within a county, we first calculate a simple average of the loss costs within each of these 9 bins (i.e., one average loss cost for each bin that results in 9 “average” observations). Then, we take the average of the 9 average loss costs for the 9 bins (i.e., “average of the average loss costs”). Since the bins are constructed to have equal probabilities, there is no need for taking a “weighted average of the average loss costs”.

Note that in the approach described above, a “recency weighting” procedure can be applied when taking the average loss cost within a bin. That is, more recent years of data can be given more weight relative to older years within each bin. Alternatively, the procedure above can be easily implemented with less than the 30 years in the 1980-2009 data series (say, for example, using data from 1990-2009 only, if the older loss data’s loss environment is sufficiently different than the more recent time period.  The procedure above also allows for consistency with the current catastrophic loading procedure. In this case, we also calculate the unweighted and weather weighted average loss costs where the adjusted loss cost data are censored at the 80th percentile. A similar calculation can be done where the censoring is done at the 90th percentile (since there was a recent recommendation to increase the censoring for catastrophic loading to this level).

**Results and Discussion**

An example case where county level loss costs are merged with the bin classification data can be seen in Table 2 for corn in Dewitt County, IL. The unweighted and weather weighted average loss costs at the county level can be calculated using the data presented in Table 4.3. The bin classification column allows us to conduct the weather weighting procedure described above. If the insignificance flag for model fit is equal to one in any county, we do not recommend using weather weighting for the county (i.e., we do not report a weather weighted average in this case).

Examples of unweighted and weather weighted average loss costs for several counties in Iowa are presented in Table 3. Note that we calculate six loss costs averages (i.e. six weighting types) per county where: Weighting type = 1 if the average loss cost is calculated with no weather weighting; Weighting type =2 if the average loss cost is calculated with weather weighting; Weighting type = 3 if the average loss cost is calculated with censoring at the 80th percentile and no weather weighting; Weighting type = 4 if the average loss cost is calculated with censoring at the 80th percentile and with weather weighting; Weighting type = 5 if the average loss cost is calculated with censoring at the 90th percentile and no weather weighting; Weighting type = 6 if the average loss cost is calculated with censoring at the 90th percentile and with weather weighting. In the example in Table 4.8, it can be seen that the weather weighted average loss cost tends to be smaller than the unweighted average loss cost. However, this is not a pattern observed in every county-crop combination. There are cases where the weather weighted average loss costs are higher than the unweighted average loss costs.

Table 4 presents the national average of the calculated unweighted and weighted loss costs for all crops we examined. This is the liability weighted average across counties (i.e., the liability weighted average (not simple average) of the average county level loss costs based on the 2009 liability of each county). For apples, barley, cotton, potatoes, rice, and spring/winter wheat, the weather weighted average loss costs (at the national level) tend to be smaller than the unweighted loss costs. However, for corn, cotton, sorghum, and soybeans the weather weighted average loss costs (at the national level) tend to be larger. A map showing the pattern of the difference between unweighted and weighted average loss costs for corn is presented in Figure 4.5. Around 51% of the counties have weather weighted average loss costs lower than the unweighted loss costs.

**Conclusions and Implications**

This article develops a procedure for weighting the historical loss cost experience used in calculating crop insurance premium rates. The idea is to utilize longer time-series information about weather variables to augment the shorter historical county loss cost data used for crop insurance rating, thereby improving statistical validity of premium rates. In developing the weighting methodology, the following factors were explicitly considered: statistical validity, feasibility, sustainability, and a balance of improvement versus complexity.

Our evaluation suggests that the National Climatic Data Center's Time Bias Corrected Divisional Temperature-Precipitation-Drought Index data, also call the Climate Division Data, is the most appropriate data set to use in weather weighting historical loss cost data. Fractional logit models can be used to relate loss cost experience to weather variables like the Palmer Drought Severity Index (PDSI) and Cooling Degree Days (CDD). Out-of-sample forecasting competition can then be used to select the specific weather variables the best explains weather experience for each particular climate division. This process creates a weather index from 1895 to the present, which allows one to better assess the likelihood of the loss experienced in each year using a longer series of weather data. A variable width, equal probability "binning" approach can then be implemented on the historical, county level loss experience data to more properly calculate the expected (or average) loss cost used in estimating a county base rate.

Results of this study showed that a weather weighting approach is indeed feasible within the context of US crop insurance rating and the approach developed in this study provides a way to capture "longer" term weather experience to augment "shorter" historical loss cost data used in estimating premium rates. Given that previous studies have provided evidence that asymmetric information problems are prevalent in the US crop insurance program due to the inability to estimate premium rates commensurate with the actual level of risk (Goodwin and Smith, 1995), the weather weighting approach developed here may be viewed as another step in more accurately estimating premiums and hopefully reducing asymmetric information problems such as adverse selection and moral hazard. In this case, "hidden" weather probability information, which is not captured in the shorter historical loss cost data used for calculating base rates, can now be utilized to better assess county level weather risk and improve the estimation of county base rates, which in turn can reduce asymmetric information problems in the long-run.

**References:**

Coble, K.H., T.O. Knight, B.K. Goodwin, M.F. Miller and R.M. Rejesus. 2010. "A Comprehensive Review of the RMA APH and COMBO Rating Methodology: Final Report." Report prepared for the USDA – Risk Management Agency-USDA by Sumaria Systems, Inc.

Dai, A., K. E. Trenberth, and T. Qian. 2004. A global data set of Palmer Drought Severity Index for 1870-2002: Relationship with soil moisture and effects of surface warming. J. Hydrometeorology, vol. 5, 1117-1130, American Meteorological Society, Boston, MA.

Guttman, N. and R. Quayle. 1996. A historical perspective of U.S. climate divisions. Bulletin of the American Meteorological Society, vol. 77(2), pp. 293-303, American Meteorological Society, Boston, MA.

Goodwin, B.K., and V.H. Smith. 1995. *The Economics of Crop Insurance and Disaster Aid*. Washington DC: American Enterprise Institute.

Karl, T.R., C. N. Williams, P. J. Young, and W. Wendland. 1986. A model to estimate the time of observation bias associated with monthly mean maximum, minimum, and mean temperatures for the United States. Journal of Climate and Applied Meteorology, January 1986, vol. 25, pp. 145-160, American Meteorological Society, Boston, MA.

NCDC. 1994. Time Bias Corrected Divisional Temperature-Precipitation-Drought Index. Documentation for dataset TD-9640. Available from DBMB, NCDC, NOAA, Federal Building, 37 Battery Park Ave. Asheville, NC 28801-2733. 12 pp.

Papke, L. E. and J. Wooldridge. 1996. "Econometric methods for fractional response variables with an application to 401(k) plan participation rates." *Journal of Applied Econometrics* 11: 619–632.

Wilhelmy, O., K. Hubbard, and D. Wilhite, D. 2002. "Spatial representation of agroclimatology in a study of agricultural drought." *International Journal of Climatology*, vol 22(11), pp. 1399–1414.

Table 1. Hypothetical Example of Bin Classification: Soybeans in Mississippi (State=28) climate division 1 (1980-2009).

| State | Climate Division | Year | State proxy flag=1 if used state predicted values | Bin Classification | No of Bins for the Climate Division | Flag =1 if insignificant |
|---|---|---|---|---|---|---|
| 28 | 1 | 1980 | 0 | 4 | 10 | 0 |
| 28 | 1 | 1981 | 0 | 8 | 10 | 0 |
| 28 | 1 | 1982 | 0 | 2 | 10 | 0 |
| 28 | 1 | 1983 | 0 | 5 | 10 | 0 |
| 28 | 1 | 1984 | 0 | 4 | 10 | 0 |
| 28 | 1 | 1985 | 0 | 8 | 10 | 0 |
| 28 | 1 | 1986 | 0 | 9 | 10 | 0 |
| 28 | 1 | 1987 | 0 | 1 | 10 | 0 |
| 28 | 1 | 1988 | 0 | 10 | 10 | 0 |
| 28 | 1 | 1989 | 0 | 8 | 10 | 0 |
| 28 | 1 | 1990 | 0 | 4 | 10 | 0 |
| 28 | 1 | 1991 | 0 | 6 | 10 | 0 |
| 28 | 1 | 1992 | 0 | 5 | 10 | 0 |
| 28 | 1 | 1993 | 0 | 2 | 10 | 0 |
| 28 | 1 | 1994 | 0 | 4 | 10 | 0 |
| 28 | 1 | 1995 | 0 | 1 | 10 | 0 |
| 28 | 1 | 1996 | 0 | 1 | 10 | 0 |
| 28 | 1 | 1997 | 0 | 5 | 10 | 0 |
| 28 | 1 | 1998 | 0 | 10 | 10 | 0 |
| 28 | 1 | 1999 | 0 | 5 | 10 | 0 |
| 28 | 1 | 2000 | 0 | 8 | 10 | 0 |
| 28 | 1 | 2001 | 0 | 5 | 10 | 0 |
| 28 | 1 | 2002 | 0 | 4 | 10 | 0 |
| 28 | 1 | 2003 | 0 | 5 | 10 | 0 |
| 28 | 1 | 2004 | 0 | 3 | 10 | 0 |
| 28 | 1 | 2005 | 0 | 7 | 10 | 0 |
| 28 | 1 | 2006 | 0 | 9 | 10 | 0 |
| 28 | 1 | 2007 | 0 | 8 | 10 | 0 |
| 28 | 1 | 2008 | 0 | 6 | 10 | 0 |
| 28 | 1 | 2009 | 0 | 4 | 10 | 0 |
| 28 | 1 | 2010 | 0 | 10 | 10 | 0 |

Note: The state proxy flag is equal to 1 if there are not enough observations (n>10) in the climate divisions to run a credible fractional regression model and calculate a predicted loss cost (weather index).

Table 2. Hypothetical Example County-level Data used for Calculating Weather Weighted Average Loss Costs for De Witt county (County=39), IL (State=17): Corn.

| State | County | Climate Division | Year | Actual Adjusted loss costs | Bin Classification | No. of Bins | Flag =1 if insignificant |
|---|---|---|---|---|---|---|---|
| 17 | 39 | 4 | 1980 | 0.1237103 | 10 | 11 | 0 |
| 17 | 39 | 4 | 1981 | 0.0083081 | 3 | 11 | 0 |
| 17 | 39 | 4 | 1982 | 0.0040853 | 2 | 11 | 0 |
| 17 | 39 | 4 | 1983 | 0.1285333 | 11 | 11 | 0 |
| 17 | 39 | 4 | 1984 | 0.0081736 | 5 | 11 | 0 |
| 17 | 39 | 4 | 1985 | 0 | 2 | 11 | 0 |
| 17 | 39 | 4 | 1986 | 0 | 5 | 11 | 0 |
| 17 | 39 | 4 | 1987 | 0 | 9 | 11 | 0 |
| 17 | 39 | 4 | 1988 | 0.1321881 | 10 | 11 | 0 |
| 17 | 39 | 4 | 1989 | 0.0007658 | 2 | 11 | 0 |
| 17 | 39 | 4 | 1990 | 0.0031037 | 3 | 11 | 0 |
| 17 | 39 | 4 | 1991 | 0.0008012 | 10 | 11 | 0 |
| 17 | 39 | 4 | 1992 | 0.0006445 | 1 | 11 | 0 |
| 17 | 39 | 4 | 1993 | 0.0004054 | 3 | 11 | 0 |
| 17 | 39 | 4 | 1994 | 0 | 3 | 11 | 0 |
| 17 | 39 | 4 | 1995 | 0.0185295 | 8 | 11 | 0 |
| 17 | 39 | 4 | 1996 | 0 | 2 | 11 | 0 |
| 17 | 39 | 4 | 1997 | 4.105E-05 | 2 | 11 | 0 |
| 17 | 39 | 4 | 1998 | 0.0009253 | 8 | 11 | 0 |
| 17 | 39 | 4 | 1999 | 0.0004244 | 6 | 11 | 0 |
| 17 | 39 | 4 | 2000 | 0 | 4 | 11 | 0 |
| 17 | 39 | 4 | 2001 | 0.0007537 | 4 | 11 | 0 |
| 17 | 39 | 4 | 2002 | 0.0125182 | 9 | 11 | 0 |
| 17 | 39 | 4 | 2003 | 9.802E-05 | 3 | 11 | 0 |
| 17 | 39 | 4 | 2004 | 0.0011999 | 1 | 11 | 0 |
| 17 | 39 | 4 | 2005 | 0.0031927 | 10 | 11 | 0 |
| 17 | 39 | 4 | 2006 | 0.0006764 | 7 | 11 | 0 |
| 17 | 39 | 4 | 2007 | 0.0020617 | 9 | 11 | 0 |
| 17 | 39 | 4 | 2008 | 0.0008186 | 3 | 11 | 0 |
| 17 | 39 | 4 | 2009 | 0.0026792 | 1 | 11 | 0 |

Table 3. Hypothetical Example of Unweighted and Weather Weighted Loss Costs at the County-level for Boone County (county=15), Dallas County (county=49), and Grundy County (county=75), IA (State=19).

| State | Climate Division | County | County Average loss costs | Flag =1 if insignificant | Weighting Type |
|---|---|---|---|---|---|
| 19 | 5 | 15 | 0.0096378 | 0 | 1 |
| 19 | 5 | 15 | 0.0076921 | 0 | 2 |
| 19 | 5 | 15 | 0.0028386 | 0 | 3 |
| 19 | 5 | 15 | 0.0027737 | 0 | 4 |
| 19 | 5 | 15 | 0.0035587 | 0 | 5 |
| 19 | 5 | 15 | 0.0033862 | 0 | 6 |
| 19 | 5 | 49 | 0.0100697 | 0 | 1 |
| 19 | 5 | 49 | 0.0097928 | 0 | 2 |
| 19 | 5 | 49 | 0.0058953 | 0 | 3 |
| 19 | 5 | 49 | 0.0058029 | 0 | 4 |
| 19 | 5 | 49 | 0.007514 | 0 | 5 |
| 19 | 5 | 49 | 0.0075715 | 0 | 6 |
| 19 | 5 | 75 | 0.0091694 | 0 | 1 |
| 19 | 5 | 75 | 0.0051299 | 0 | 2 |
| 19 | 5 | 75 | 0.001323 | 0 | 3 |
| 19 | 5 | 75 | 0.0010593 | 0 | 4 |
| 19 | 5 | 75 | 0.0044935 | 0 | 5 |
| 19 | 5 | 75 | 0.0032308 | 0 | 6 |

Note: Weighting type = 1 if the average loss cost is calculated with no weather weighting and no censoring; Weighting type =2 if the average loss cost is calculated with weather weighting but no censoring; Weighting type = 3 if the average loss cost is calculated with censoring at the 80$^{th}$ percentile and no weather weighting; Weighting type = 4 if the average loss cost is calculated with censoring at the 80$^{th}$ percentile and with weather weighting; Weighting type = 5 if the average loss cost is calculated with censoring at the 90$^{th}$ percentile and no weather weighting; Weighting type = 6 if the average loss cost is calculated with censoring at the 90$^{th}$ percentile and with weather weighting.

Table 4. Liability Weighted National Average (across counties) of Unweighted and Weather Weighted Average Loss Costs for Apples, Barley, Corn, Cotton, Potatoes, Rice, Sorghum, Soybeans, Spring Wheat and Winter Wheat. – table 4.9

| Crop | No. of Counties | Unweighted loss costs (no censoring) | Weather weighted loss costs (no censoring) | Unweighted loss costs (censoring at 80th) | Weather weighted loss costs (censoring at 80th) | Unweighted loss costs (censoring at 90th) | Weather weighted loss costs (censoring at 90th) |
|---|---|---|---|---|---|---|---|
| apples | 140 | 0.1839529 | 0.1756118 | 0.1509251 | 0.1458255 | 0.1722479 | 0.1649113 |
| barley | 646 | 0.1033683 | 0.0952631 | 0.071994 | 0.0677116 | 0.088203 | 0.0820236 |
| corn | 1930 | 0.0505333 | 0.0525652 | 0.028726 | 0.0293841 | 0.0394102 | 0.0409063 |
| cotton | 437 | 0.143511 | 0.1459077 | 0.1103868 | 0.1110684 | 0.1292813 | 0.1305584 |
| potatoes | 128 | 0.083174 | 0.0807186 | 0.0659818 | 0.0646853 | 0.0752233 | 0.0730846 |
| rice | 84 | 0.0263574 | 0.0251909 | 0.015527 | 0.0148564 | 0.0203618 | 0.0193536 |
| sorghum | 750 | 0.1208383 | 0.1317581 | 0.0887164 | 0.09226 | 0.1079448 | 0.1140653 |
| soybeans | 1523 | 0.0542112 | 0.0538458 | 0.0384229 | 0.0379807 | 0.0467105 | 0.0460899 |
| spring wheat | 244 | 0.1218715 | 0.1171909 | 0.0887732 | 0.0872793 | 0.1094074 | 0.1063092 |
| winter wheat | 951 | 0.0982152 | 0.0852073 | 0.0719574 | 0.065563 | 0.0851164 | 0.0759965 |

Note: These are the national average loss costs across all counties (i.e., liability weighted average) where the insignificance flags and state proxy flags are not equal to one. All weighted and unweighted loss costs for each county is available from the authors upon request.

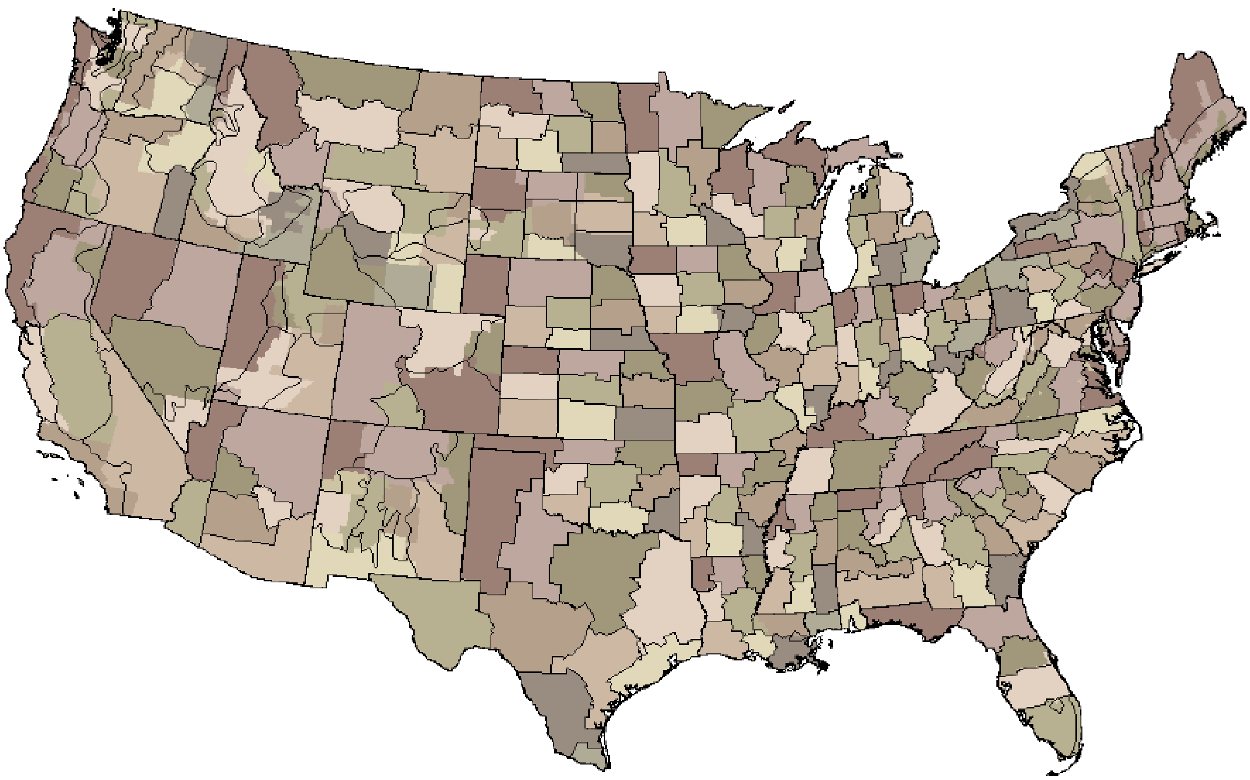

Figure 1. Climate Division Boundaries and County Assignment within Climate Divisions

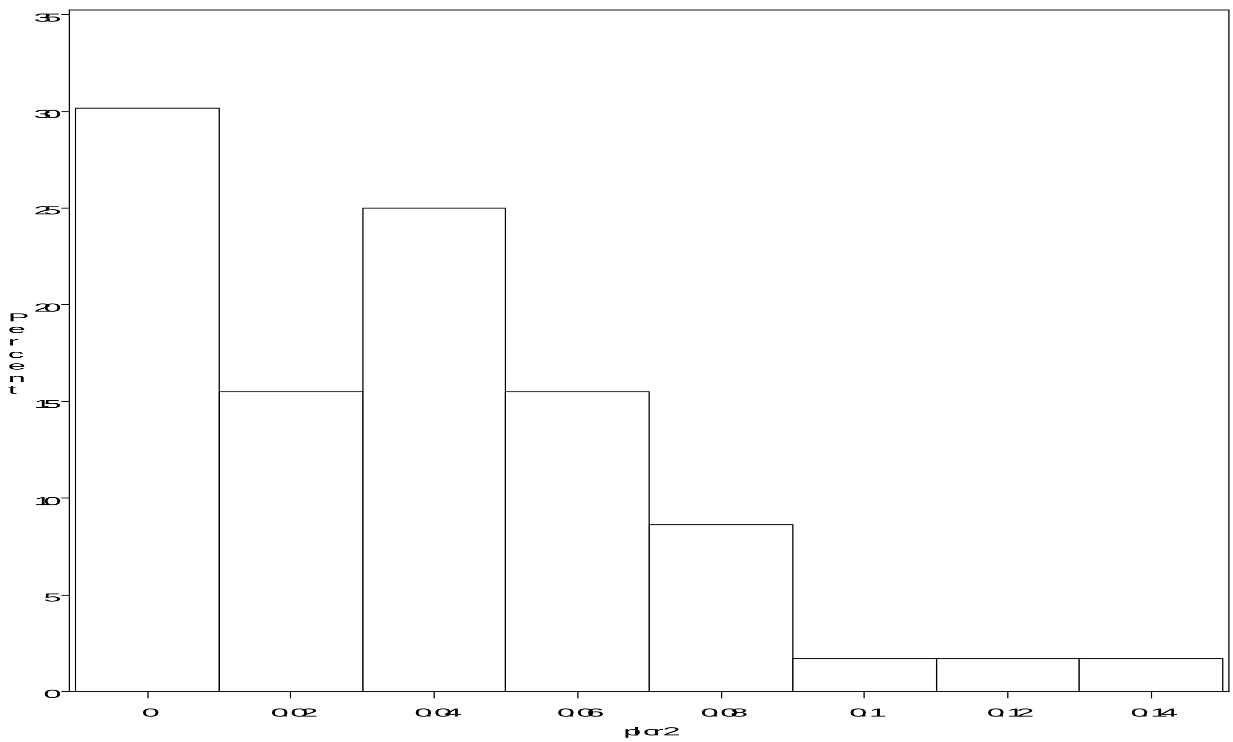


Figure 2. Example Histogram with Equal Bin Widths and Variable Probabilities for Each Bin

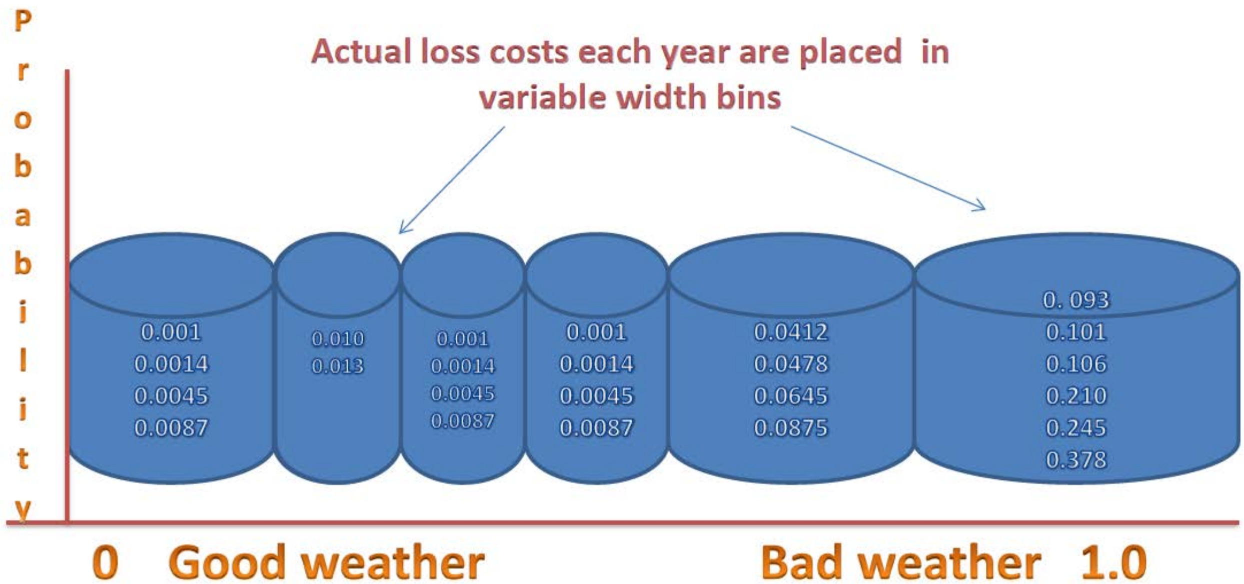


Figure 3. Example of Variable Width Bins with Equal Probability for each Bin

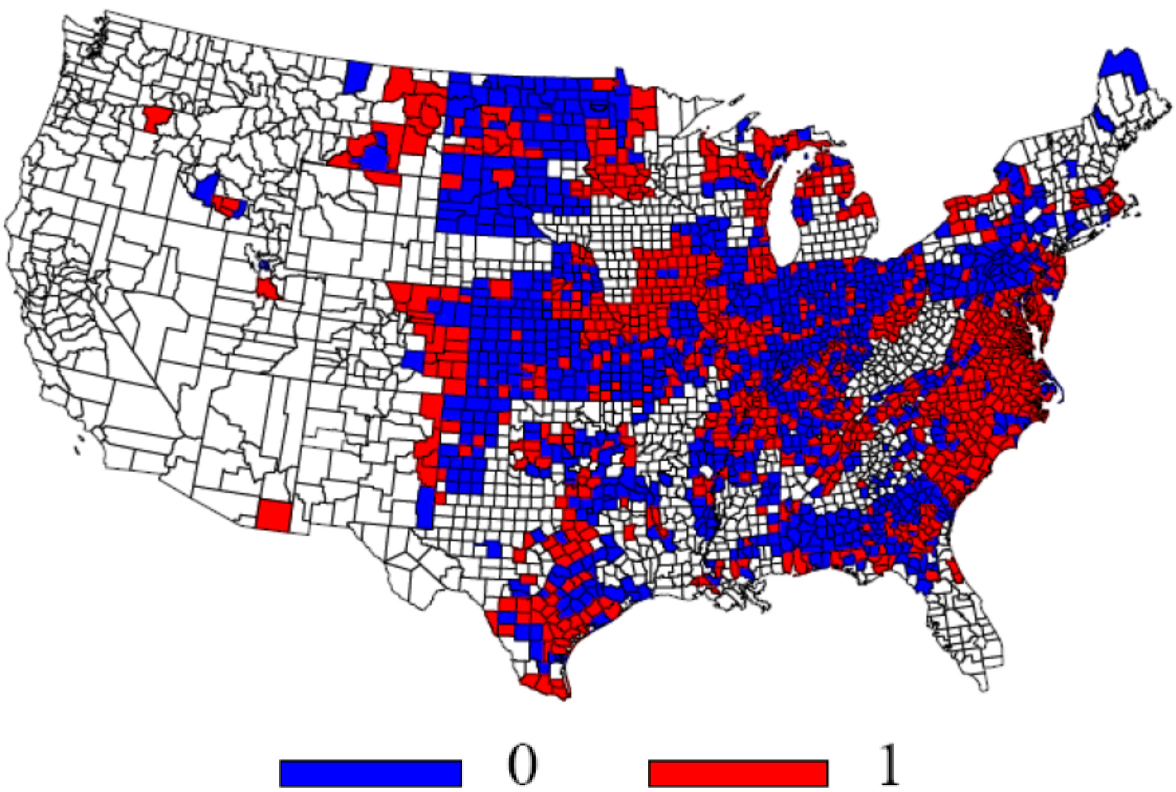


Figure 4. Map of the Difference between the Unweighted Average Loss Cost and the Weather Weighted Loss Costs for Corn [Note: negative difference (e.g., weather weighted < unweighted) is in blue (0) and positive difference (e.g., weather weighted > unweighted) is in red (1).]