Review of:

"Methodology Analysis for Weighting of Historical Experience"

by *Sumaria Systems*

Technical and Implementation Reports

October 14, 2011

Submitted to:
Board of Directors of
Federal Crop Insurance Corporation

Prepared by:

iFAR

integrated Financial Analytics & Research
1605 S. State Street, Suite 103
Champaign, IL 61820-7264

Contact: Bruce J. Sherrick
sherrick@illinois.edu

Table of Contents

Executive Summary

Sumaria Systems recently completed two related contracted studies entitled "*Methodological Analysis for Weighting Historical Experience*" as part of ongoing efforts by RMA to monitor and improve the ratemaking systems for FCIC insurance products. The findings in their reports can be summarized into five recommendations for changes to the APH ratemaking system, three of which involve reweighting historic data for sample period weather effects, along with a recommendation for modifying CAT load procedures, and a recommendation for altering the effective sample period, and use of historic data. The net effect of the recommendations would be to reduce base rates for major crops across a large portion of the cornbelt, and adjust other base rates on other program crops generally toward the target loss rates from current longer run average rates.

iFAR was contracted to review and evaluate the Technical and Implementation Reports (Reports), and to comment as appropriate on other related items that impact the actuarial performance of the ratings system. Brief (interpreted) summaries of the recommendations and our summary review comments follow, in the general order of their recommendations.

Recommendation 1. RMA should collect and use Climate Division data to generate weather indexes over a long historic period for use in determining the relative severity of the weather and associated loss rates observed in the more recent period in which insurance was offered.

Recommendation 2. RMA should use fractional logit models to relate actual loss experiences to weather index data and form fitted loss rates through a longer history than observed in LC data.

Recommendation 3. RMA should use the fitted values of the losses across the long period of time over which weather is recordable to establish pseudo-probability estimates of the severity of the weather and associated loss rates during the insurance offering history. The process recommended is to identify equal probability bins (variable width pseudo-probability measures) and calculate weighted loss rates in each bin to generate an overall estimate of loss rates.

Recommendation 4. RMA should increase the CAT "cutoff" level to a measure related to the 90^{th} percentile of losses, an increase from the currently used 80^{th} percentile. Additionally, an *ad hoc* procedure for reweighting upper limit observations is included.

Recommendation 5. RMA should first make an adjustment to the loss data prior to 1995 to equalize mean loss rates between older and newer subperiods, then shorten the sample period used to construct base rates to 20 years.

Recommendations 1-3 were investigated through a simulation exercise and through as much replication of the steps indicated in the Reports as possible. In total, the recommended procedures represent little improvement over far simpler re-weighting schemes, and in total, do not make a great deal of difference for major program crops. Even under exactly specified error generating processes, the arbitrariness of the bin interval construction leads to about as much error as simple integration across the fitted in-sample loss rates. While it is conceptually sensible to make the weather adjustments, the *Sumaria* recommendations are difficult to implement and maintain relative to alternative, simpler methods.

Recommendation 4 is generally sensible as the definition of "CAT" has resulted in relatively large loads relative to base rate loads and it seems reasonable to increase the percentile limit used to delineate CAT from base loads. However, the CAT redistribution is intended to redistribute a total loss across a wider range and over time than would be subject to the large losses generated infrequently at catastrophic severity. The suggestion should be modified to define the CAT limit, perhaps using a longer time series to establish the level of division between CAT and base loads, and then redistribute the losses around the cutoff selected, rather than to redistribute the loads from a longer period over a shorter period. Additionally, the minor suggestion to implement an *ad hoc* rule to further adjust the loads above the 97^{th} percentile does not work as described in the recommendation to better reflect loss rate experience and should not be implemented.

Recommendation 5 is the most important and impactful of all the recommendations. There are numerous issues related to data representativeness in the current system and the recommendation generally treats the issue as one of sample selection rather than as a search for other control variables to include in a structural model of loss experience. There are several alternatives investigated, and numerous structural loss models examined with dummy controls for sample period effects – importantly – all of which result in very similar ultimate effects. Due to the confirmation of the magnitude of the effect from several alternate perspectives, we fully support the implementation of this impacts of recommendation. The shortened sample period could have implications in contrast to the possibility of including factors through time that explain changing loss rates, and the longer run evaluation of the impacts should still be conducted. It is our judgment, however, that the choice of method is not as important as the recognition of the effect of recalibrating the rates to actual experience.

Additionally, we recommend that a simple spatial smoothing process be applied to the base rates to help limit the consequence of single location sample variation and to take advantage of spatially correlated loss cost information. This step could substantially improve the equity across similar counties and result in more similar premiums in similar production regions with similar risks.

In summary, we commend the authors and RMA for their sincere ongoing efforts to continue to improve the performance of the federal crop insurance programs. It is clear that the importance of these programs will continue to increase in the future, and the effectiveness of the programs stands to be improved greatly through the recalibration of base rates to more nearly coincide with long run average costs.

Background and Motivation for Reports and Review

In 2009, RMA commissioned *Sumaria Systems* to review methods and procedures used in determining APH and COMBO base (yield) insurance rates.[1] Following the conclusion of that study, *Sumaria Systems* was commissioned to extend their analyses of several issues raised in that report and elsewhere, related to the potential to improve upon the use of equally weighted county/crop based loss cost measures, and to make recommendations targeted at improving the actuarial performance of APH-based insurance rates. They supplied their initial findings and recommendations in what is referred to as their "Technical Report". *Sumaria* was then commissioned to develop implementation recommendations for their recommendations in the Technical Report in the form of what is referred to as the "Implementation Report". ***iFAR*** was asked to review and evaluate the Technical and Implementation Reports (Reports), and to comment as appropriate on related items that impact the actuarial performance of the ratings system. Because the Implementation report largely mirrors the selected suggestions in the Technical report, our review treats them simultaneously as a single effort ordered by the major themes and summary recommendations.

Importantly, the FCIC insurance products are viewed by most as providing the cornerstone of all risk management programs for U.S. agricultural producers, and the prominence of crop insurance programs within the commodity title of the Farm Bill is likely to increase by most accounts of current policy observers. The actuarial performance is therefore of critical concern, and efforts to monitor and improve the programs to reflect the intent in the Federal Crop Insurance Act (FCIA) are likewise crucial. Among the statements in the Act are provisions to insure uniform and equitable treatment of producers, and to undertake efforts to monitor and maintain actuarial soundness. The Reports and reviews thereof represent important efforts by RMA to maintain and improve tier ongoing efforts to monitor and improve crop insurance programs for U.S. agricultural producers.

[1] 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., available at http://www.rma.usda.gov/pubs/2009/comprehensivereview.pdf

The overall summary implications of *Sumaria* reports are that base rates should be revised in general proportion to the divergence between long run loss costs and target loss rates, and that adjustments should be made to the process of using simple averages of long histories of annual loss costs for the formation of base rates. However, as part of a set of efforts to inform the major crop insurance industry participants, RMA presented the intended implementation of the findings to a few key industry groups, and the industry response could be fairly characterized as highly concerned about the impact on profitability. It appears that several industry-led, somewhat informal analyses were conducted immediately following the meetings, and additional requests were provided to evaluate the implications of the reports. We understand that our review to at least some degree is to provide external independent evaluation of the reasonableness of the findings, and that is could be considered during final implementation or rejection of the recommendations of the Reports.

We have carefully reviewed and evaluated the reports, underlying data, and have conducted alternate analyses the portions that are viewed as most critical to accurate rating of major crops in the program. To summarize, the recommendations represent clear improvements that are likely to improve the actuarial characteristics of the programs for the major crops representing the clear majority of the premium and liability. We find the proposed methods for arriving at these recommendations to be a mixture of the practical and obvious with the obtuse and unnecessary, but the debate with how to assess base rate formation should not be confused with judgments about the magnitudes of the findings. The magnitudes of the proposed changes in base rates are generally in line with divergences between loss costs and target rates. In other words, the revised rates will tend to result in more uniform loss rates by region and across crops. Further, basic empirical evidence confirms that the proposed revisions in base rates are reasonable given the actual performance of the programs.

The existing ratings methodology is largely built on a Loss Cost Ratio (LCR) methodology that essentially uses the set of annual loss cost experiences at county/crop exposure units as the "sample" from which to construct forward rates. LCR methods work best in situations where the underlying loss rate distribution is stable across exposure units (which may also involve repeated time periods) and hence, empirical experience (on rates) can be used to form an accurate liability pricing relationship, even if liability changes relative to the underlying exposure units in the sample, as long as the relationship between liability and indemnity is accurately captured by the experience, and extendible if/as changes in liability occur. Departures from target LCRs through time can result from random sample period effects, or from violations in the underlying requirements for construction of accurate expected loss costs from actual loss experiences.

Though somewhat understated, a primary practical motivation to examine the performance of the ratings system in the first place derives from the fact that the loss ratio performance by crop and region do not appear to have resulted in random variations around the target loss rates through time. Questions have been fairly raised about whether the differences through time are simply due to idiosyncratic sample period effects, or whether systematic departures do in fact exist that represent deficiencies in the ratings system. Importantly, as there are limited market-based options to "bid" insurance rates based on such pressures, and the loss performance is thus tightly tied to the performance of the underlying ratings system; and it is therefore important to periodically test the system against its empirical performance through time and recalibrate if needed to achieve the objectives of the program. Though simplistically referred to as the "ratings system", the realities are that the program has survived and been modified through a dizzying set of differential production systems, farm commodity program regimes, seed technologies, product offerings, coverage and subsidy schedules, and through dozens of "recalibration" efforts targeted at keeping the process tractable and implementable. There is not a single most obvious source to investigate as an omitted variable from the ratings system that could be included to "repair" the use of historic loss costs, nor a single omnibus test of the adequacy of ratings against loss experience characteristics that would allow a conclusion of adequacy of rates to be determined. Nonetheless, it is useful to consider plausible sources of persistent divergence pressures on the ratings system, and to determine whether such pressures are simply random features or items whose influences can be used to either adjust the weightings given to historic data in the LCR system, or to modify the LCR system to directly incorporate the impacts of the otherwise omitted influences.

Among the features identified in past works as plausible omitted ratemaking considerations are: (i) unrepresentative weather in small samples (i.e., a simple equal weighting of an extremely low likelihood event in a short period overstates its appropriate impact on a base rate, while its omission understates its appropriate weight; and the passage of time alters its implied weight if the sample period accumulates - all possible problems), (ii) evolution of yield risk through time (the LCR system requires very specific forms of proportional risk to exist as yields increase through time that are almost surely untenable, and present possible problems in loss rate construction, and results in biased effective coverage levels under trends), (iii) product shares/types and endogenous base rate problems (the producer can self-select coverage and revenue features that result in difficult-to-standardize loss rates; and product mixes and shares have evolved dramatically over the sample period, and observed loss rates are thus related to changing base rates through time and self-selection), (iv) acreage and participation changes (early period participation was very low compared to current acreage covered. It is unlikely that acreage randomly participated at all points in time, and if any self-selection by risk type occurred, low acreage samples are not representative of current exposures. In some cases, historic loss rates in some periods could not occur in present programs even with complete losses), (v) CAT loadings limit the full transmission of information from direct individual loss experiences (the share of losses and the redistribution represent a form of smoothing that could shift loss rates from high to low loss areas, and the areas of application differ from ratings areas).

Review of Recommendations of Reports:

The *Sumaria* reports largely delineate their investigations into (1) weather related effects and the possibility of altering historic weights (away from simple averages) to reflect additional evidence about the relative severity of the conditions at the time points at which loss rates are calculated; and (2) sample period representativeness, characterized primarily as a selection of a discrete adjustment for regimes, and a shortening of the time period over which to average loss rates rather than integration of changing underlying features as control variables through time. (An exception to that treatment was proposed by an earlier reviewer for acre weighting and accepted in the implementation report phase.) Additionally, the reports suggest a sensible redefinition of the CAT load calculation that would help eliminate the tendency to redistribute relatively high CAT loads over arbitrary state divisions rather than like-production regions.

The Reports result in 5 specific recommendations which we evaluated by either recreating the supporting analyses, (usually in limited form against major crops only) to validate, or utilized alternative analyses for confirmation or refutation, and where suitable, compared the general magnitude of the result to that which we believe would result from alternate treatments. Brief summaries by (interpreted) recommendation are provided in the following sections with additional details provided subsequently in a presentation of components of our analyses.

Recommendation 1. (R1) *"....(construct) crop specific weather indexes... (to) allow RMA to compare the weather experience incurred by the modern program to (a more representative sample)."* We interpret the main idea to be for use in reweighting the experience in the sample period to reflect the relative severity of each year's experience modeled relative to a longer period. The general idea is for example, to identify the relative likelihood of the high loss years that might receive too much weight because a 1 in 100 year event happens to fall in the 30+ year observation window, or to increase the empirical rate if the sample period weather happened to be "better than average" expected over a long run period.

R1:Review Conclusions: We strongly endorse the general idea of reweighting yield and or loss experience to reflect sample period weather effects. We have made extensive use of the Palmer related indexes and other weather data for the major crops, and also find that the ability to relate loss experience to weather interval events is plausible and even fairly easy to condition through time (though this feature is not reflected in the *Sumaria* reports -- i.e., the impact of drought of the severity of 1988 would perhaps not have as large a proportional effect with current technology if it recurred today). Though we agree with the intent, the proposed approach is simply untractable for outside replication in many places and unduly complicated in forming the weather indexes. Much of the sample period utilized in the Climate Division Data recommendation is available and mappable directly to CRD and counties in much simpler forms. An extrapolation process is used to generate the weather data used prior to the 1940s in the Climate Division Data (to be clear, there is no obvious need for such a complex representation, though there is also likely little practical cost for doing so other than an overstated set of confidence limits in the logit regressions that result from the first stage fitting conducted then treated as "data", and an overly complex presentation of relatively simple steps to condition loss costs on weather through time). Given the issues raised in later stages of the report with the changing nature of losses through time, we also strongly favor a simplified form for constructing weather indexing variables <u>after</u> the sample period effects and acreage weighting are effected to improve the usefulness of the form relating loss costs to weather severity indexing information.

We also conducted extensive evaluation of the "binning" process across alternative plausible data samples, and find that it is a coarse method for aggregating probabilistic weights that can be very inefficient in controlling for the impact of weather effects. Fortunately, the impacts of this recommendation are minor relative to later conditioning steps that largely obviate the need to weight for weather anyhow.

Recommendation 2. *"...use a fractional logit model... to relate loss cost experience to (constructed weather variables). ...out of sample (selection criterion)... (to create an index) of the growing conditions experienced each year* ". We interpret the idea to be that a selected functional form and estimation framework be used to estimate a structural model of loss cost rates as a function of the constructed weather variables. The primary purpose of this step is to classify the loss cost observations for purposes of reweighting to form a better estimate of the (weather) unconditional mean loss rate (see R3).

R2:Review Conclusions: Given the issues raised in later stages of the report, the weather index could more plausibly have been created by simply regressing yields deviations from trend against the PDSI and CDD items (rather than loss costs over a sample period that is later subdivided and rejected as consistent), and used to define percentile breakpoints in a smaller sample space on loss costs. The sheer number of out-of-sample tests conducted render individual application of significance "sorts" less informative, and thus rather than select based on significance at individual sub samples, it makes more sense to use control variable sets that are more consistent and selected on the basis of preponderance of evidence across the crop/region type. In other words, there is no obvious way to choose the cutoff p-value for use as "significant" or not, and as a control variable set, the null could be argued to be a p-value of .5 for example. It is also not clear why dividing the (continuous) PDSI variables into pairs of censored variables is more useful than using the native scale of the PDSI, but doing so does not appreciably affect the results either in the subsample tested (corn only, Illinois, Iowa and Indiana counties, PDSI May, June, July, and August indexes used in related *iFAR* Trend Adjustment work).

The reports indicate that an “Out of Sample” MSE tournament was run to both select the set of weather variables and apply at a later stage for time period stationarity tests. We assume that the OOS construction included actual regressors (not step ahead versions at mean weather levels). A subtle issue again relates to the intent of the control variables and whether use of type-I significance tests is most appropriate, and the use of single MSE criteria by location to pick state-wide specifications. The results in table 4.5 are somewhat surprising in that the major corn producing states do not have the expected (set 4 from table 4.1) explanatory variables given extensive historic work demonstrating the usefulness of the PDSI. It may be that the high correlation between the CDD measures used and the regular PDSI measures, and the separation of the PDSI variables into “kinked” censored sets resulted in the selection of the CDD measures across the states representing the majority of the corn premium. The reports offer no suggestions why the selected variables differ so dramatically over contiguous states (IA to MN for example) and we doubt that the annual refitting and possible re-specifications contemplated are justified compared to simply using the four monthly PDSI and two CDD interval measures for all locations. A related issue in table 4.6 of the implementation report bears comment as well. The reports indicate that the “significance” of the control variable regressions can be used to flag insignificant relationships. In the present application, the choice of the p-value that best balances Type-I and Type-II errors is unclear and may be better treated as a conditioning step implemented on the basis of the use for calibrating later stage weather (i.e., if an area had uncharacteristically “good” weather in the sample period, there may be no relationship between the sets of essentially zero loss costs and the regressors observed but the use of the control relationship to condition the likelihoods remains valid. Recall that the intent is to improve the estimate of the mean loss rate across all weather, not estimate the loss rate at the mean of each weather effect. The nonlinear relationship between weather and loss rates renders the two dramatically different).

In summary, we agree with the intent to isolate probabilistic information for weighting the loss experiences in the relatively short experience windows but do not see the need for such a complex approach, nor is it clear how the results in the implementation report fully reflect the later implemented acreage weighting (which if useful should also apply prior to this stage) or the discrete separation of the time period of application (i.e., pre 1995 is taken as a different loss distribution, but apparently not separated at this stage of estimation). Interestingly, our use of simpler 4-month interval PSDI, precipitation, and temperature variables appears to work well in corn and soybeans for conditioning yield deviations, and this may be a simpler indexing system to use that provides comparable ordering information without relying on the features of the loss cost series that are later judged as "non-stationary" and thus delineated in time in the reports.

Recommendation 3. *"(use the results of R2 to)... categorize the loss cost experience observed over the period chosen into weather 'probability bins' (for reweighting by bin probability)* ". We interpret this recommendation as the result of the fact that the R2 process creates an index value converted to a percentile estimate for each year's weather index (treating the weather interval as a sufficient probability measure and using the fitted loss costs as the index). The sample period may have observations on the (shorter) loss cost series that are not fully representative of the longer term (unconditional) loss rates, such as if a set of better than average weather years was sampled. Moreover, it is likely the case that the transformation from the weather index to loss costs is non linear and that losses are convex to the weather index. Simple averaging in-sample does not take into account information in the weather experience used to condition the loss rates. Thus, a method for weighting using a system intended to index probabilities is needed. A binning procedure is suggested to "weight" the loss costs based on strata defined by the longer term weather index.

R3:Review Conclusions: To use sample percentile values that are not uniformly distributed to reweight the sample and recover the unconditional mean requires a few steps and conditions to result in less biased estimates of the mean loss cost. If one has fully estimated a functional form relating the loss costs to the weather index, it is a straightforward idea to integrate back out the unconditional mean over the weather index distribution (this approach requires much less in the way of assumptions about the sample observed, and possible nonlinear effects by strata would be directly incorporated). It would also be straightforward to parameterize the whole set of fitted percentiles and obviate the need to conduct the second stage re-binning exercise as no more information can be recovered and the same information used to create the logit model is already reflected in the fitted function. The report argues that the binning process avoids the need to parameterize a probability measure and is parsimonious and simple to implement. It should be pointed out that the step of fitting a loss cost function to weather data within the sample period is already a parameterization that defines the probabilities returned indirectly. There is still sample variability, and using the fitted functions over a longer period of RHS variables to capture more complete probability information is useful if the fitted and actual data are very tightly connected. In the data provided in table 4.2, that is not the case. In our own tests, the binning process does not perform superiorly for recovering known means from exactly specified versions of the problem even with perfect correlation between the constructed weather index and the sampled loss data. The performance of the proposed method does not seem sensible in a few additional dimensions as the impact of weather reweighting on corn is to increase the estimated loss rates while for soybeans with essentially the identical region and weather experience reduces the estimated loss rate. We suspect it is an incidental result from the method of reconstructing weighted means and not a real difference. It is a fairly simple method proposed to use in which the percentiles from the fitted loss costs to define probability intervals (argued to be equal probability bins, though conceptually irrelevant how the probability measure is subdivided) and the sets of observed costs within each bin averaged if multiple members and weighted by the bin to create an average that reflects some portion of the relative probability information about each sample period.

While it is not at all clear *why* *Sumaria* recommends this particular approach, it represents a feature that should be reflected in the rates. In other words, the idea of reweighting based on probabilistic evidence about the long term prominence is appropriate, but simplified reweighting approaches are likely preferable in practice. Moreover, the application of the acreage weighting within bins appears to be the more important effect in virtually every case in the Reports.

Recommendation 4. *"... recommend changing the catastrophic load cap to the 90th percentile and reducing the aggregation region for catastrophic load from the state level to a climate division* ". Additionally, procedures are outlined for selecting the CAT division based on a longer data period than used in the base rate calculation, and procedures discussed for reweighting based on fitted percentiles of the loss experiences in the CAT region of losses in the sample.

R4:Review Conclusions: The selection of the division between CAT load and base load in an local insurance context can be complex, but the loss data on CAT clearly favor substantially increasing the "cutoff" percentile deemed to be CAT, and redistributing over more homogenous and smaller geographic regions than a state. Increasing the loss level division between CAT and base rate sets is directionally attractive compared to the current practice of taking 80th percentile and higher losses (in many cases apparently resulting in higher loads than the base rates in specific locations). We also suggest a simple spatial smoothing procedure be applied after the CRD or Division level CAT losses are determined to help avoid stark transitions that are likely sample specific. The weather index re-weighting for the observations contained in the CAT set is an additional *ad hoc* suggestion in the report for associating a fitted percentile value with a probability of observance, but does not work as suggested in the reports, in general, due to a confusion between probability *weights* and percentile levels (the ordered set does not translate to (1-1/ith) probability interval where i is indexed from top down), and the inverse weighting does not add back up across all items in the CAT set to 100%. You could also have sample points with more than 100% weights by approach suggested (if more than 3% of the sample points were greater than the 97th percentile from the first stage), and as an extreme event approached the 100% fitted percentile, it would get approximately a 0% weight which seems to be an odd concept for determining CAT loads. To know how to reweight the CAT loss observations obviates the need to use the reweighting anyhow, as one already must have complete probabilistic information about the loss likelihoods (and the mean loss in the interval in which contained), so the method simply adjusts to a predefined value. Further, the sample size impacts the conditional value of losses expected in excess of the extreme value in the sample, and thus the directional impact of the proposed adjustment will result in downward bias in estimated losses that depends in magnitude on the size of the sample used to associate "exact" cumulative probabilities with the sample points. Finally, there is no clear explanation for the choice of the 97th percentile or any other limit by sample size other than incidental use of 1/sample size.

More directly, for the CAT loading to perform as intended, it should simply represent a reallocation within a region of total losses between a CAT and base component and not change the mean total rate. A higher CAT cutoff results in a higher base rate and a lower CAT load, and an lower CAT cutoff results in higher CAT load and a lower base rate. The delineation is between forms is largely a matter of the degree of pooling across time desired in the particular insurance context. In other words, whatever share of losses is isolated as CAT will reduce the overall average by the same amount on average that it adds to the local weights. In effective applications, CAT loadings represents a form of spatial smoothing and time aggregation without changing the long term total rates paid. There are some additional inconsistencies in using a percentile cut defined over a long sample period to isolate shares of losses in a shorter period deemed CAT and pooled and redistributed as the same effective percentile will not in general result in the shorter sample. It is suitable to define what the value is at a percentile break (say a target of 90th) from a longer history of data, but the actual CAT load in a region should still only be constructed from the set of sample values above that predefined line within the sample period to avoid distortions in total rates, and not based on the actual loss data above that percentile from a different sample period.

Recommendation 5. *"...use adjustments to remove non-stationarity from the loss cost history* ". There are a number of individual recommendations here that we interpret as largely *ad hoc* controls for declining loss costs through time, changing net acreage exposure, and limiting the effect of early period loss rates based on production practices and acreage shares that are viewed as unrepresentative.

R5:Review Conclusions: The term "nonstationarity", though used in a somewhat nonstandard sense, reflects a change in the process relating loss costs from one period to another. The test proposed to identify a "structural break" is to force a breakpoint of 1995 with a dummy variable, and test if the dummy coefficient is significant, (accept null of differ pre- and post-breakpoint loss rates). Other time specific tests were also conducted, but the recommended approach settles on the use of a breakpoint at 1995. It is not totally clear why 1995 is chosen or if time is viewed as the variable influencing loss rates after controlling for other items (like weather) why its effect was not modeled more directly as a control variable (i.e., distance from 1995 is tested as model 2, share of acreage insured, but as dummy sets, or kinked control variables after a given date). Interestingly, when the structural break is found to be significant, the proposed recalibration is to "shift" the average from the pre period loss rate to the post-period loss rate -- apparently resulting in the same average, unconditional on weather, and thus having little effect compared to starting the sample post 1995. We also tested alternate break point dates in a regression of adjusted loss rates on PDSI set, and virtually any date after 1989 is significant as a breakpoint in time for sample differences (statistically significant dummy) all the way down to climate divisions across all three states examined. The conclusion is that the proposal is one acceptable, fairly blunt way, to control for a set of underlying loss rate features that have evolved in time. The acreage changes and evolving yield risk appear most plausible primary features, and each of these effects could be potentially modeled through direct inclusion in an explanatory model of loss rates, but the recommended approach results in a roughly comparable magnitude of loss rate impact (15-25% in corn in most CRDs) and is easy to follow and implement.

There are additional tests in the technical report that are not included in the implementation report; and there are also references to weightings by net acres that would circularly affect the first stage estimates and thus potentially affect the binning, and thus the relative in-bin acre weight, which in turn should be reflected in the first stage estimate.... and so forth (rather than what we view to be a more sensible sample period relative acre weight that can be applied in a simple regression context). It is also unclear how the application of these final sample period recommendations interact with the first stage weather index estimation beyond the acre weighting (if weather sample is related to the structural break effect due to sample period randomness for example, then the order of control matters for treatment).

Other Issues and Observations:

The following materials are provided in response to the charge to consider other relevant issues and features affecting the performance of the ratemaking system, whether included in the report or not, and to highlight items from other experiences that might help identify additional areas to examine for possible improvements to the base rating system that helped to interpret the Reports.

1. A simple empirical review of the crop insurance programs would help focus the efforts, deflect criticism by uniformed parties, clarify purposes of the reports, and could help identify areas for improvement not related to re-weighting historical experience alone. Simple tabulations from the Summary of Business reports including, net acres by crop year and location, premium shares and levels through time, periodic loss ratios, while familiar to most readers of these reports, would have been useful to include for ready context and reference, and would have also helped frame the importance of the question for what are viewed as federal farm bill “program commodities” at minimum. Summary statistics from the Statplan data base would also have helped place context around the results, and give a sense of scale in many key sections. More directly, the ratings concerns are primarily driven by empirical concerns around the performance of corn, soybeans, and to a lesser extent wheat programs, and the sum of the remaining crops deviations in indemnity from total premiums would have been useful to understand. This comment is intended to help pre-empt other criticism as the absence of the empirical context could lead to undue suspicion about lack of context and

possible mistakes in focus especially by outside commercial interests as well and could help to accurately couch any ensuing debates.

2. The process for standardizing loss rates at a 65% coverage level and the related methods for spreading the losses back over coverage choices at the election level is an important part of the ratings system. This step is left as a wholly unexamined premise. However, the four weather data series examined happen to have slightly higher correlation with the unadjusted and acre-weighted unadjusted loss rates than their adjusted analogs in our samples from the Statplan data and it seems that there may be additional information had some reflection of the coverage items at higher elections been available. From our understanding of the original motivations for the adjustment process, and historic data constraints, higher coverage standardization or direct measure by coverage level election would be possible to accurately consider only in the later periods where data are more extensive (perhaps 1996 onward). Simulations conducted incidentally while investigating the impact of the proposed binning procedures indicate that the correlation (and informational content) is substantially lessened from fitting to the shorter segment of the "tail" that results from converting all losses to their 65% coverage analogs (we simply ran a simulation with randomly selected coverages from 60 to 85% and simultaneously captured both "actual" and 65% exact loss values). The necessary mixing at any point across non linear loss rate curves makes the recovery of exact CLDs impossible once pooled in a 65% policy across varying loss conditions). These and related features of the premium system including the UDF, reference yield impacts, and exponent sets should be simultaneously evaluated for their combined total impact on policy premiums to fully understand the impact of the subset of changes proposed in the reports. In other words, the loss experience data were generated under inexact and complicated related situations. Using a fairly agnostic treatment of each issue independently of the impacts on other effects creates additional model risk. More importantly, the ultimate task presented to *Sumaria* was to improve the accuracy of the base rates and the measure of that improvement should be examined more completely against historically realistic cases. For example, if target

loss rates of premium serve as guiding principles, then the impacts should be cast in that context.

3. Though likely outside the charge of these reports, the existing system could be compared to more direct characterizations of individual exposures and loss experience tested in simulation form, or back-tested through experience data from production report and type 15 record systems. The myriad corrections, controls, and reformulations of the information in the LCR base rates through time, and processed into ultimate premiums hardly represents a true LCR system anymore anyhow, and the basic methods proposed for correction to the rates could be fairly cast as an empirical calibration exercise (albeit with impressive sophistication at certain stages). There are numerous stylized parameterizations already imposed on the loss costs, their responses to weather, and the used of related conditioning controls. It would be useful to also consider a more fundamental alternative rating approach that might actually result in a simpler ratings and calibration scheme based on depictions of the underlying exposure units. The benefits of maintaining the original LCR system are less obvious through time as the coverage of major crops is high and fairly stable, and the use of historic loss rates conditioned on so many intermediate calibration steps.
4. It is likely the case that a partial cause of the declining loss rates observed in the reports, and controlled for primarily through a shortened sample and a pre-1995 correction, is due to declining proportional risk features resulting from modern seed and production technologies, increasing average unit size, and the like. If so, the corrections to the loss rates from shortening the sample period and imposing a break in the loss series around a chosen breakpoint are rather blunt controls for a process whose properties could be incorporated into the ratings system more directly. Taken with point 3, it is appropriate to not only evaluate the characteristics of the existing system, but also to consider a more fundamentally direct incorporation of the features of the loss exposures that translate to loss rates through time such as declining proportional yield risk.
5. There is a fundamental disconnect between the use of small samples within county boundaries and weather controls defined in larger regions (which itself does not change

discretely by county line or Division) in terms of information aggregation. More fundamentally, there are other major features affecting insurance risks that have obvious spatially related components. At minimum, the base rates should be spatially smoothed to reduce the gradient of rate changes between adjacent production regions within CRD and among adjacent counties across CRDs in most cases. The CAT process and simple-circle averaging in cases judged otherwise unreliable, represent coarse forms of spatial smoothing. It is clear from many vantage points that the loss data are spatially related and the informational content of proximate data is not complete within geographically arbitrary regions and devoid across county lines. The simplest forms of mean preserving smoothing at minimum should be considered and tested for their impacts in helping to smooth loss ratio information across space. Numerous possible contiguity criteria exist that are likely to generate virtually identical effects where counties could be conditioned by relative yield risk in CRD, and across CRD, etc.

6. Acre weighting issues may be significantly more complex than simply an in-sample issue as largely treated, as it is likely that initial participation in the program was by those with the highest likelihood for claim and that as participation has grown, even a small share of the current participants equal in size to early participant pools would have a very different loss experience.

Related Analyses:

The subsequent sections present some of the substantiating information and analyses conducted in the course of reviewing the reports. They are ordered roughly in the order encountered in the reports, but with particular emphasis on a few items that might also suggest simplifications to the approaches in the reports for future implementations.[2]

[2] Original simulation files and processed data are available for use by RMA and related interested parties involved in the studies upon request from the authors. In a few cases software specific to the analyses (@Risk and SPSS canned modules) would be required for full replication. All other work was performed in Excel or in standard statistical packages with commonly available features.

In evaluating the performance of the methods proposed to adjust for weather impacts in the sample period, several related issues were considered. First, the weather measures have to be very tightly related to the loss costs to permit the use of a longer sequence of the weather indexes to be used to assign relative probability information to the loss observations in sample. A first visual test of the relatedness is suggested by the data in figure 1 from table 4.2 (page 34 of the implementation report). Ideally, the data would be of the same scale or at least generally form a constant ratio. The simple linear correlation between predicted and actual in the case presented is about .45 and the significance of the fit, even if the two largest actual

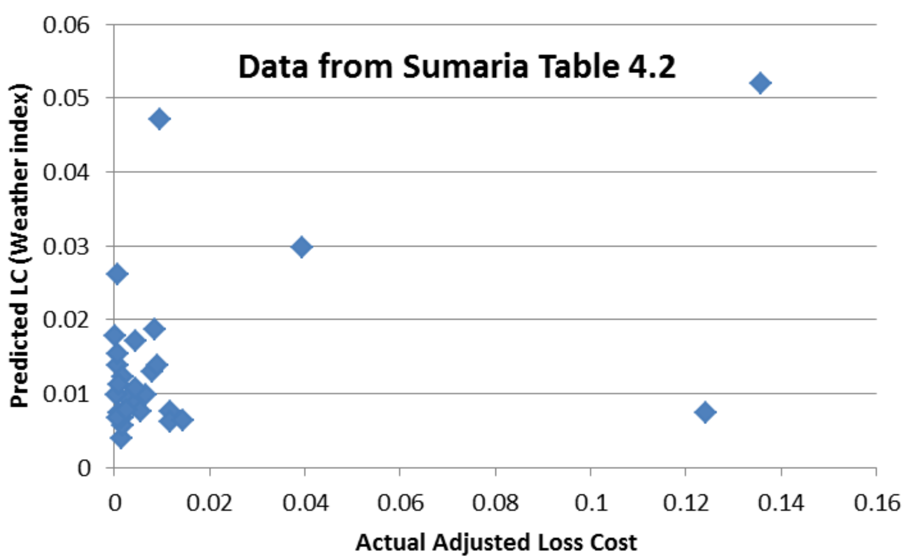


Figure 1. Sumaria data relating predicted and actual loss costs

losses are excluded is marginal at best. A minimum requirement for suitable loss probability assignment would be generally accurate ordering by the predicted of the actual (so that even if it is simply a highly nonlinear relationship between weather and losses (in the present cast the predicted values are used so not the case), at least the order statistics are comparable and a form of probability weighting could be recovered. Table 1 below shows the rankings by severity and the lack of correspondence between the fitted and actual loss costs, raising concern about the ability to use the fitted forms to assign appropriate probability weights to intervals containing actual observations.

Of further concern is the possibility that the zero loss cost observations could be assigned to a bin other than the first, or set of first bins, all containing only zero. Note that this is a separate

Table 1. Loss cost ordering by actual and predicted from *Sumaria* Report.

		--- Loss Costs ---	
Year	Rank Diff	Actual Rank	Predicted Rank
1980	4	9	5
1981	6	20	26
1982	1	16	17
1983	0	3	3
1984	4	11	15
1985	3	15	18
1986	9	12	21
1987	16	26	10
1988	0	1	1
1989	14	6	20
1990	1	10	11
1991	1	8	9
1992	8	22	30
1993	21	2	23
1994	2	24	22
1995	6	13	7
1996	11	18	29
1997	4	21	25
1998	5	7	2
1999	4	28	24
2000	24	30	6
2001	0	14	14
2002	13	29	16
2003	7	19	12
2004	2	17	19
2005	17	25	8
2006	10	23	13
2007	23	27	4
2008	23	5	28
2009	23	4	27

issue from the absence of a loss rate in a binned interval (no assigned Percentile in the sample in the bin interval). Due to the indemnity function features in crop insurance, some (ordered) fraction of the outcome space results in zero losses and the probability mass associated with the zero losses would have to be all contiguous for the reweighting by bin probabilities to make sense (entries in Table 4.3 violate this requirement of any probability measure – suppose a bin with only zero values occurred at more severe weather indexes – the implication violates the requirement of monotonicity between the variable and its CDF).

Bins absent any sample data present the problem indicated in the report (again, zero values of loss present no calculation problem, but empty bins do cause a failure in the weighting scheme proposed). The proposed solution is to start with a large number of bins and reduce the number until each bin is populated. This seems somewhat arbitrary and prone to slowly widening the probability intervals until an observation "slips" into the bin from one edge or the other. This would result in a poor correspondence between the true bin contribution to the calculated mean, but not necessarily bias the mean, just be a low-efficiency use of the data. Further, it is unclear if the variable number of bins across sample aggregates (i.e., the rule stops at say 11 bins in one location and at 6 in a neighboring location) results in assignments of differing effective probability weights to the same weather effects through time. To investigate this issue further and to isolate the degree to which the binning process deviates from a more direct estimation of a consistent probability measure across the fitted percentiles, we built a simple sampling scheme from actual data and then converted it to a simulation against a known probability specification to see how the sample variance effect changed by sample size as well.

Some preliminary items are interesting. First, the simple sample variance effects, when used to convert to percentile levels, are related to sample size, and the likelihood for additional information that is outside sampled range. To get a sense of the importance of this issue on the performance of the proposed weather weighting process (under known conditions) and its likely performance against the LC data used, we began by simply sampling from a (known) percentile distribution and tested sample moment and percentile variation against known theoretical values -- the idea is to help judge whether the 35 or so actual losses form a suitable sample for the scope of the estimation of the effect of weather on losses, that is then extended to a longer range of weather indexing components.

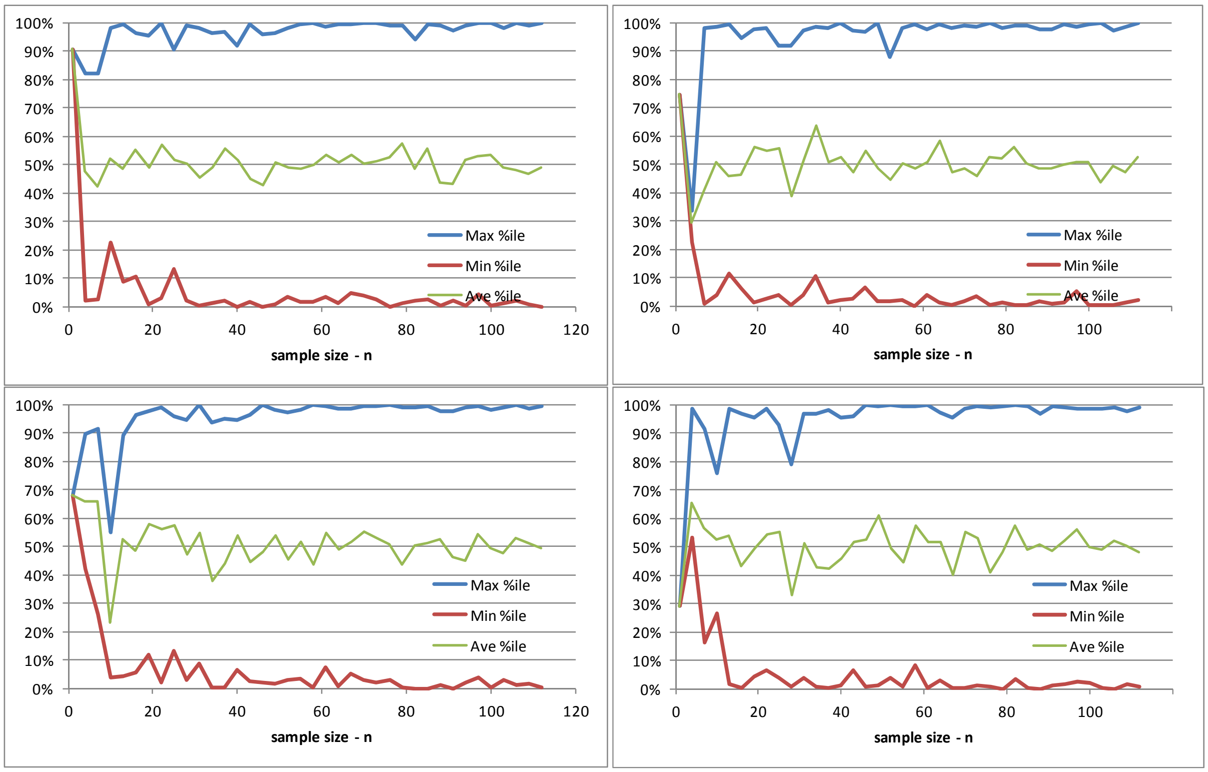


figure 2. Percentile ranges by sample size, four iterations.

Random samples of the weather index analog were first created of each size from 1 to 115 (also compared the actual PDSI total index to its 115 year history, but this sample is fixed and resampling degenerates quickly by sample size). The min, max, median, average, and selected percentiles within each sample are then collected and recorded by sample size. Figure 2 above contains a set of 4 of the iterations in this process. Note that by around n=30, there is still a substantial degree of sample variation in the percentile range observed of the weather index. This feature alone is of no particular concern if the relationship is very stable between the weather and loss rate experienced, but it does make the first stage estimation error important to recognize if used for a calibration of the weather effects later. Ideally, a wide range and uniform coverage of the percentile set observed in sample would be best suited for then using this information to form "bins" of probability that accurately weight the inverse used as the LC function. Gaps near the boundaries of the distribution are the main influence in selecting the number of bins.

Next, a mapping from the weather index to the LC distribution is needed (in the reports, the LC is parameterized through the fractional logit model, the LCs then assigned as predicted or fitted through the inverse weather index distribution), so that it can be extended across the remaining range of the weather variables observed over the longer period. To construct this, we built a similar model to the estimated model, and scaled it to have a loss cost average of a approximately 10% (in figure 3 below, the red line hides the blue outcomes of zero below the first positive red "fitted outcome" as in the report model). Even if weather exactly maps to a given LC (red line, no error), the mixing of weather at a point in time, and simple non-weather based random variation in yields will still result in minor losses even in good weather years. The censoring of losses at zero and the increasing loss severity as the weather index is worse (higher percentile) will result in a convex loss curve as the share of censored observations declines with increases in severity in addition to the natural effect observed in yield-based insurance of the same relationship.

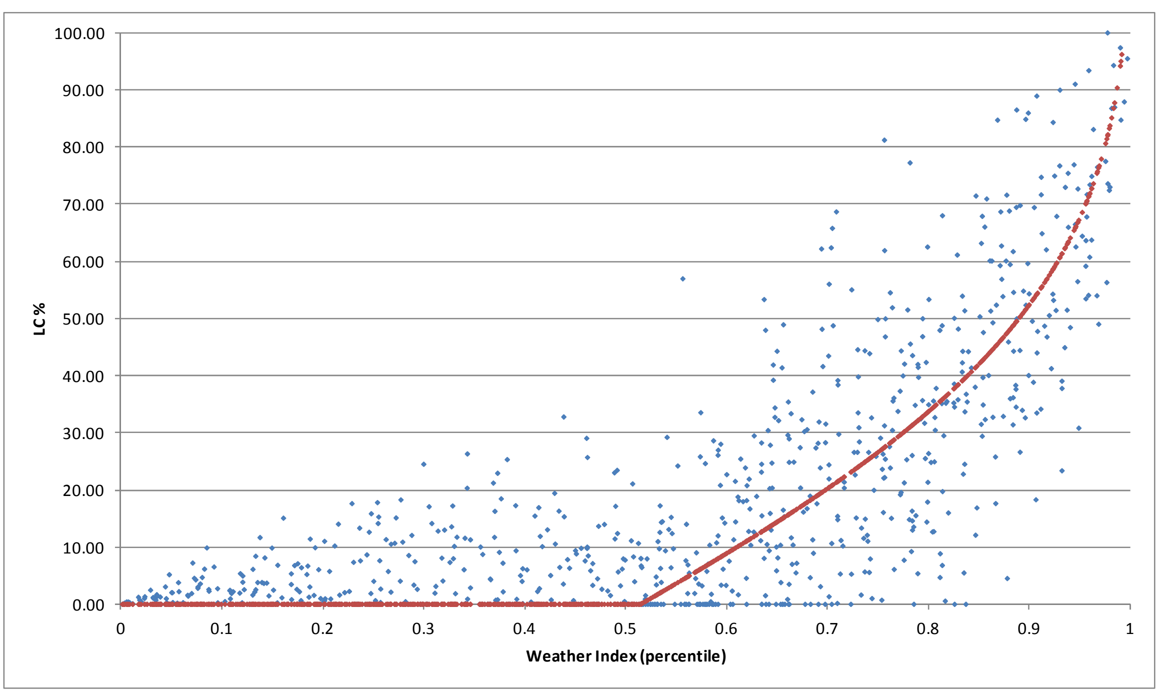
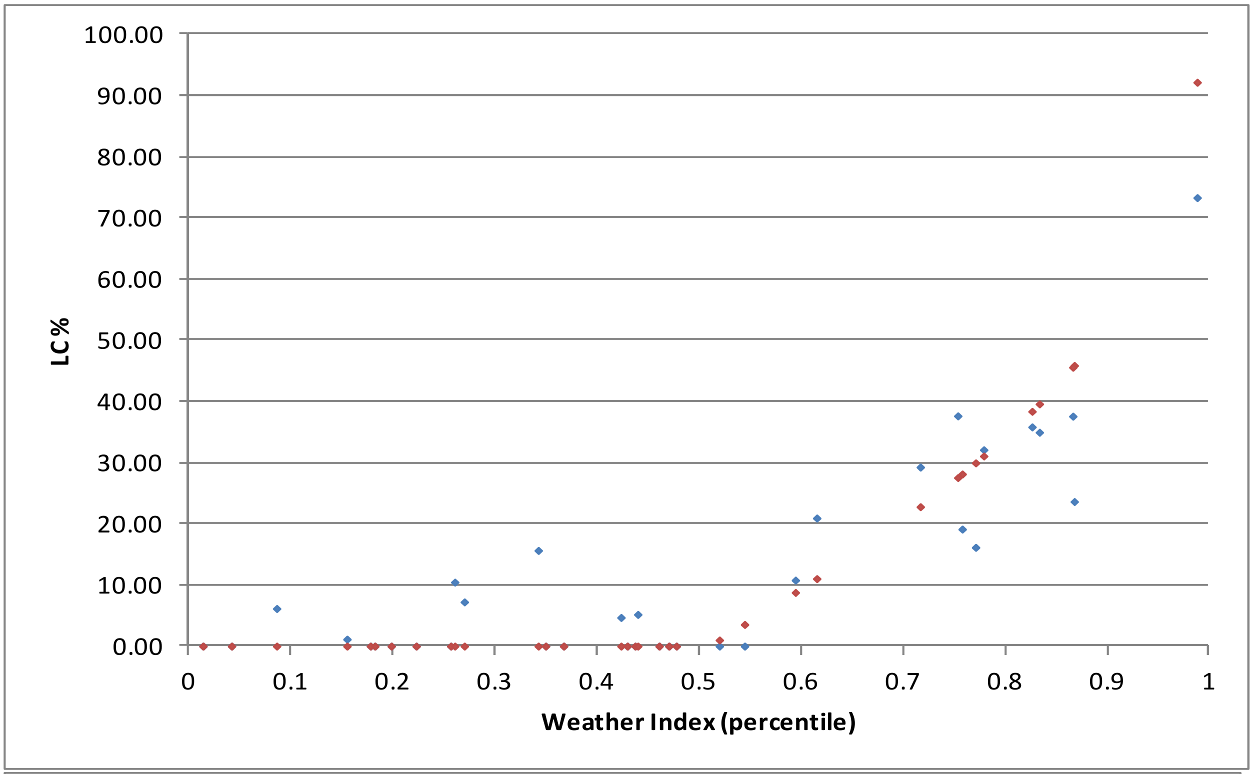


figure 3. Weather index and LC rates, exact (red) and measured with censored errors (blue)

Note that even if individual weather indexing is linearized, the deeper in the loss distribution moved, the greater the delta and hence increasing loss correlation to additional weather severity. This feature suggests that a quantile regression model would probably work better than the fractional logit model as a feature of changing "relatedness" to weather could be an important feature useful in relating the weather indexes to the observed loss costs. In any case, fractional logit models are not widely in use, and there are some meaningful concerns about the sensitivity to censored areas of the data and recreation of the ordering scale.

The next step is to investigate how well the binning process allows the reassignment of probability to in-bin averages of loss costs. To help mitigate concern about the accuracy of our reconstruction of the first stage error distribution, we started with an exact specification (red line) and varied the coverage assumed (to limit the potential impact of the 65% adjustment in the Statplan data), to focus solely on the ability to create accurate probability bins from the observed LC sample and an exact recovery of the weather index sample that created the LC rates from a known data generating process. In other words, we stacked the deck considerably in favor of the process supported in the reports.



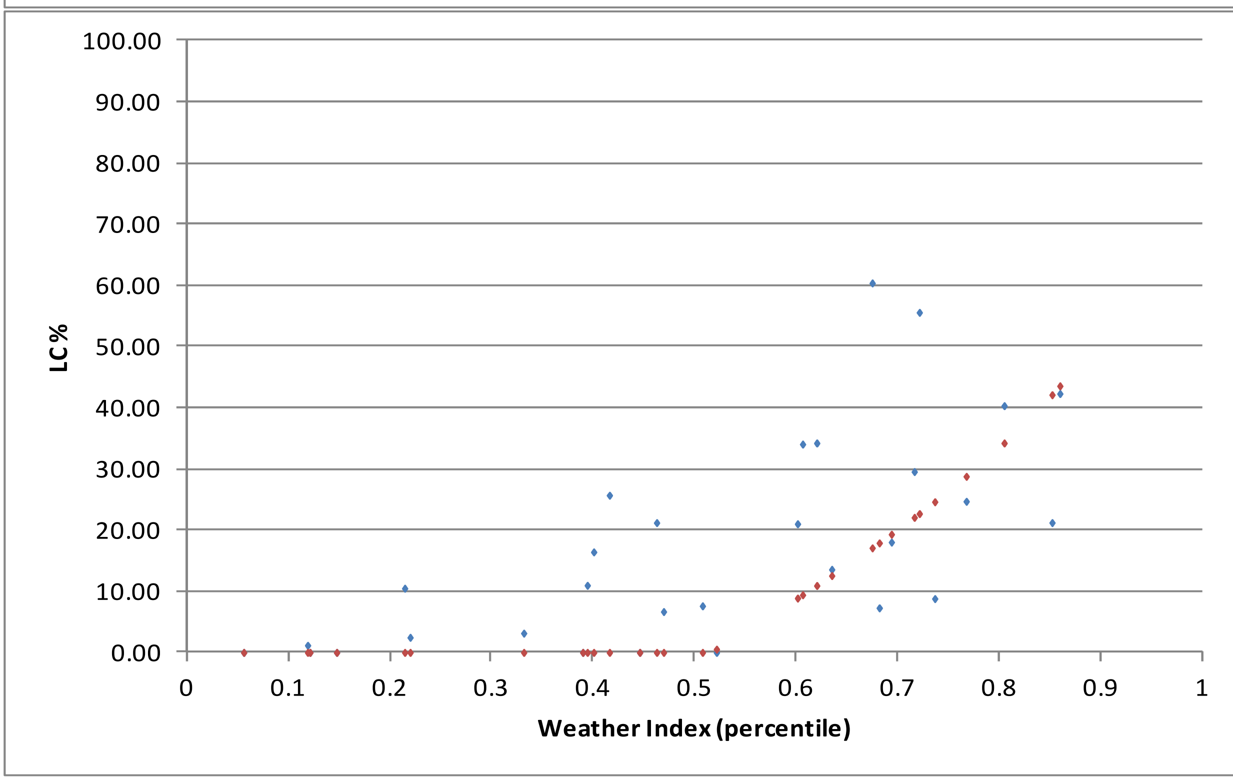


figure 4. Two samples of weather and associated loss costs.

In smaller samples, the process is even more demanding (35 shown above in figure 4, two iterations), so our treatment of the first stage as "exact" should result in a fairly good ability to recover the red line from the red dots above in the smaller sample. The minimum bin width process divides the probability into equal shares (1/bin number) until each bin contains at least some observations. In the first sample shown , the bin width could be rather narrow as there are samples near each end of the range of the maximum percentile gap of about 10% is not located where the 1/n rule has natural breakpoints near each end (i.e., a gap from the 2nd to 12th percentile is occupied all the way up to 16 bins, but the second sample with approximately a 13% gap on the results in a 7 bin minimum so the last bin contains more than 13% mass). Conceptually, middle gaps of the same width should not be more problematic than end gaps, but the process used essentially fixes the end points and varies the internal breakpoints until at least some data fall into each probability interval and then use the average of the data in the interval (acre weighted in the end) and the interval width as a weight to re-estimate a mean value.

The data in figure 5 below show the process graphically. Each of the series shows the probability breakpoints for different numbers of bins from 1 to 15. Series 16 is a sample of thirty weather index percentiles from the reports and the process of subdividing the probability space until all bins are occupied. From 15 down to 9, at least one bin is not filled. The first fully occupied binning occurs at widths of 12.5%.

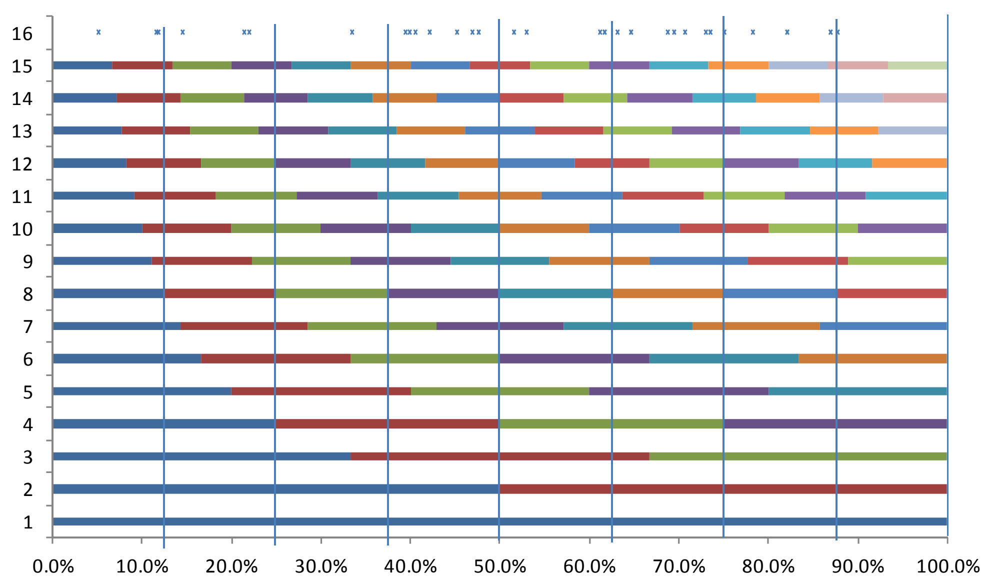


figure 5.

The contents of each bin are then averaged and the average of the averages taken as the new mean loss rate. In this case, the average weather index is slightly worse than an average sample, so the reweighting would be expected to bring the re-weighted sample average down.

The properties of this "reweighting process" are next investigated against known data generating conditions before testing against the PDSI data and the loss rate data in the Statplan sample provided. The graphic below shows one iteration where a set of 1000 at a time (35 shown) percentile draws from the weather index (WeatherDex) are used to generate exact losses that would occur using the red line function from figure 3. The relative loss scale is shown in the red bars with the width of the "0" loss bars being the relative frequency in the sample that would have zero losses (the analog blue measured LCs are also recovered). The loss scale is on a fraction of liability scaled to 150, roughly to keep analogous to corn yields, but only for stylized convenience in presentation). The "true" average loss is 15.06 and the true loss cost from the exact process is .10. The sample loss cost is in this case only 9.58, and the average weather index is about 2% "better than average". Given the sample, the maximum number of occupied bins happens to be 7 given the sample locations. The minimum bias re-binning can also be constructed since the true process is known, and the re-weighted estimate compared to a direct reintegration over the piecewise assembled percentile EDF.

index	WeatherDex	Losses
1		0.00
2		44.41
3		17.94
4		42.97
5		20.17
6		0.00
7		0.00
8		25.48
9		9.73
10		0.00
11		0.00
12		18.75
13		0.00
14		0.00
15		29.63
16		23.54
17		0.00
18		35.08
19		0.00
20		0.00
21		0.00
22		0.00
23		11.79
24		1.39
25		0.00
26		10.27
27		13.41
28		0.00
29		22.92
30		0.00
31		2.93
32		0.00
33		23.60
34		0.00
35		0.00

Again, it is important to understand the properties of the proposed reweighting process under known conditions before using it to reweight actual LC data from the far more demanding "real" process represented in the insurance program data. To test the reweighting system, the sets of samples were regenerated 5000 times per scenario and the process in the Reports undertaken whereby the weather percentiles were used to order the loss data, and the binning process worked downward from 15 until all bins had at least one member. The bins were then averaged within bin and then averaged across bin-averages for an estimate of the "true" loss cost. This value is then compared to the true values and alternative weighting schemes. During each iteration, it is also possible to examine "optimal binning" in terms of minimum bias to the known mean as well. Further, the simple piecewise integrated mean from the sample is recovered, along with the number of bins, size of bias, and other sample information. Sensitivity is then conducted around the correlation measure between the percentile and the loss function, and to the coverage equivalent.

At high relative loss scales, and with an exact mapping between the weather percentile and the loss rate, the (1) report binning process works about like the alternatives of (2) simple weighting and (3) integration across the EDF with biases per $10 of liability of (1) -.17, (2) .0017, and(3) .04 respectively. The re-weighting scheme has slightly higher bias, but slightly lower standard deviations than the alternatives ((1) .13, (2) .24,(3) .33 respectively) but none seems to be a dominant method when the exact percentiles are mapped to exact losses. If the error structure is made more realistic (say rho = .65 between the percentile and actual losses), then significant "accidental" re-ordering or losses tends to occur among the bins, but the relative behaviors are unchanged -- they all just perform worse as the problem becomes more realistic, as expected. The following screen shot in figure 7 show one iteration with the error structure used similar to that implied in the reports (7a.), followed by one iteration under the exact order generating analog (7b). Both iterations happen to have 9-bin solutions to the minimum populated test, and both have similar distributions of weather and losses. The left panel in each shows the outputs in random (sampled order), the central section ordered by predicted severity, and the right three columns are simply part of the calculation sequences to track the report's process.

Interestingly, as the coverages and loss rates get higher, all the systems perform better and vice versa. In cases with lower coverage (and lower losses), all three alternative examined perform worse in a relative sense measured by percentage bias, and the Reports' proposed system tends to have slightly higher bias but lower variance in each case tested.

Another interesting feature is that the minimum bias estimator is only equal to the report's proposed bin number about 11% of the time, often with a smaller number of bins performing better. We do not find an obvious pattern or explanation, but repeatedly found about this proportion regardless of the other features of the simulation examined, except that at larger sample numbers (say 50 years) the correspondence is even weaker, and around 20 years, the correspondence is slightly higher. In any case, there does not seem to be a good theoretical or empirical reason to select bin numbers on the basis of a rule that indicates first fully occupied strata over any other rule. That said, there is also not much impact from that rule. Figure 8 below shows the distribution of the report's suggested process without the limitation on minimums to insure 100% sum.

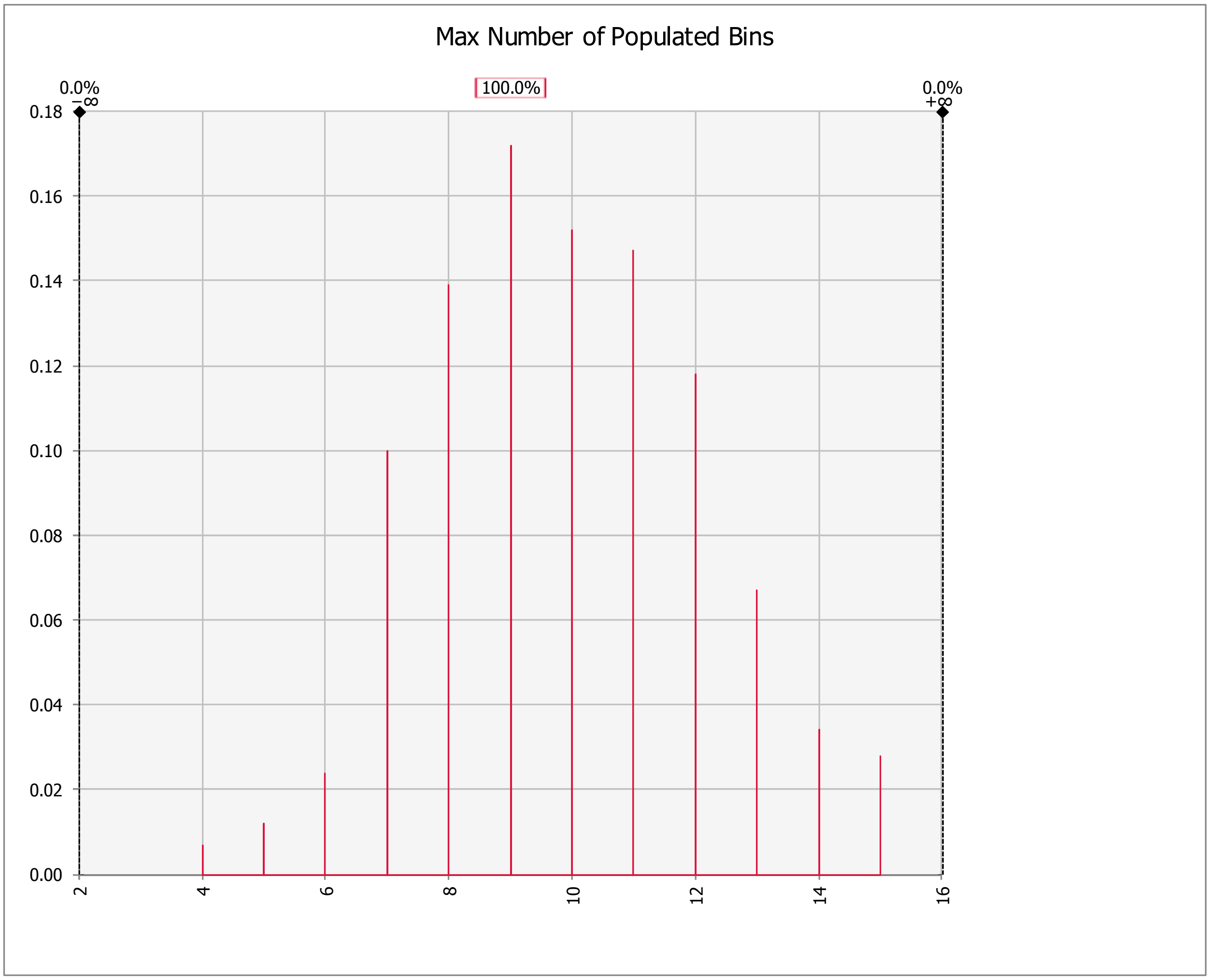
Figure 7a. Single sample from process with assignment error

Rank	Losses	Wdex	Bin	Sequenced	Ordered	Wdex	Interval	EvenBin	BinNum	BinAve
6	7.435901	0.717684	7	1	0.952155		0.013801	1.070406	1	0.80
29	0.36799	0.016972	1	2	0.36799		0.00317		2	0.15
23	0	0.233097	3	3	1.891072		0.034021		3	0.43
28	1.891072	0.050993	1	4	0		0.044287	0.146551	4	5.23
19	0	0.365325	4	5	0.439653		0.025394		5	0.00
13	0	0.487981	5	6	0		0.012444		6	0.00
21	1.303635	0.305211	3	7	0		0.086707	0	7	23.37
17	4.518796	0.378526	4	8	0		0.013272		8	16.73
15	0	0.431344	4	9	0		0.063458		9	60.07
20	5.182221	0.337467	4	10	1.303635		0.008656	2.161952		
18	7.637614	0.370498	4	11	5.182221		0.032256			
4	0.434483	0.757677	7	12	0		0.027858			
27	0	0.09528	1	13	7.637614		0.005172	8.733049		
10	0	0.524157	5	14	4.518796		0.008028			
30	0.952155	0.013801	1	15	14.04274		0.007112			
14	0	0.447475	5	16	0		0.045706	0		
26	0.439653	0.120674	2	17	0		0.016131			
11	0	0.523161	5	18	0		0.040507			
3	16.72809	0.807858	8	19	0		0.022778	0		
1	72.17621	0.98789	9	20	0		0.012402			
25	0	0.133118	2	21	0		0.000996			
16	14.04274	0.385638	4	22	0		0.104171	14.88033		
9	0	0.628328	6	23	0		0.024075			
24	0	0.219825	2	24	44.64099		0.01713			
7	44.64099	0.669533	7	25	8.676716		0.048151	16.28068		
2	47.9547	0.944041	9	26	39.73085		0.016529			
8	0	0.652403	6	27	0.434483		0.023463			
5	39.73085	0.734214	7	28	16.72809		0.050181	45.61967		
22	0	0.296555	3	29	47.9547		0.136184			
12	0	0.510759	5	30	72.17621		0.043848			

Figure 7b. Single sample from process with exact ordering process

Rank	Losses	Wdex	Bin	Sequenced	Ordered	Wdex	Interval	EvenBin	BinNum	BinAve
26	0	0.102291	1	1	0		0.020411	0	1	0.00
8	3.473561	0.679085	7	2	0		0.011556		2	0.00
24	0	0.193531	2	3	0		0.032617		3	0.00
13	0	0.511583	5	4	0		0.023016	0	4	0.00
5	27.08983	0.846491	8	5	0		0.01469		5	0.00
29	0	0.031966	1	6	0		0.002005		6	0.00
10	0	0.550009	5	7	0		0.089236	0	7	8.02
11	0	0.534432	5	8	0		0.004712		8	24.55
28	0	0.064584	1	9	0		0.003248		9	62.53
16	0	0.304893	3	10	0		0.013353	0		
25	0	0.104295	1	11	0		0.004067			
4	37.76537	0.898466	9	12	0		0.022794			
6	22.00523	0.816524	8	13	0		0.007273	0		
27	0	0.0876	1	14	0		0.041771			
18	0	0.248978	3	15	0		0.014143			
23	0	0.198243	2	16	0		0.021207	0		
30	0	0.020411	1	17	0		0.036912			
15	0	0.3261	3	18	0		0.148571			
14	0	0.363012	4	19	0		0.014165	0		
19	0	0.241706	3	20	0		0.008684			
1	90.34678	0.995956	9	21	0		0.015577			
17	0	0.29075	3	22	0		0.044035	5.344813		
2	70.27083	0.981012	9	23	3.473561		0.085041			
7	12.56088	0.751819	7	24	12.56088		0.072734			
12	0	0.525748	5	25	22.00523		0.064704	28.95348		
9	0	0.594044	6	26	27.08983		0.029968			
20	0	0.218911	2	27	37.76537		0.051974			
22	0	0.201492	2	28	51.75007		0.047599	70.78923		
3	51.75007	0.946065	9	29	70.27083		0.034947			
21	0	0.214844	2	30	90.34678		0.014944			

Figure 8. Distribution of maximum populated bin numbers



In total, there does not seem to be any advantage to the particular weather re-weighting methodology proposed over other simpler methods, and the strength of the relationship between the LC and the weather data in sample does not seem incontrovertibly strong enough to use as a component in the longer-term weather indexing system proposed. Of particular concern is that there is not a clear accommodation of the later (and much stronger) findings in the reports that there has been a systematic change in the loss cost data through time and that the early period (when much of the indexing information is effectively calibrated) is not suitable to use for other reasons. In any case, it appears arbitrary to pick equal probability (variable width) intervals over equal bin widths, or equal numbers of elements per bin, or to simply use the empirical distribution function directly as they are all about the same in terms of performance as a means to assign relative probability information within sample. Other credibility theory approaches should also be investigated before investing in such an elaborate and apparently low impact weather re-weighting scheme, at least for the major program crops (we could not replicate the effects for other crops for which we did not have the Statplan data).

Next we investigated the proposed methods for evaluating whether there are structural breakpoints in the loss cost series. Additionally, we relied somewhat on work by Woodard et. al (JARE 2011) relating losses to weather indexes through time (rather than to identify structural breaks).

As we were given the Statplan elements about midway in review period, we have only performed major confirming tests, concentrating on the central corn belt corn and soybean intense states first as these are both the most important, and most impacted against which the most summary information in the reports was provided. We first assigned counties to our CRD (same as Climate Division in states of Illinois, Indiana, Iowa) and matched weather variables. Because of the collinearity between the set of PDSI variables and CDD data, and minor omissions in counties, we focus on PDSI (4 separate monthly) measures. For future work, we recommend including precipitation in addition to PDSI, and basic soil productivity information as additional control variables. We replicated where possible, both net acre weighted and unweighted forms of regressions of adjusted and unadjusted LCs, with and without dummy variables for pre and post 1995 data, at combined and at state and at state and CRD levels of aggregation (and only for corn in Illinois, Indiana, and Iowa). We also constructed a separate acre weighting variables equal to the in-year county acres divided by the total acres through time in the County and CRD, to test relative local and relative crop representation effects, but the form does not appear to matter much as the results are virtually indistinguishable from net acres in each location as originally given. The CRD level results are most closely tied to the information in the report and are used as the focus of the remaining discussion.

The climate information is easily related to LCs at any level of aggregation tested, partly due the sheer sample sizes, whether the data pivots are constructed at county level within state or at CRDs, the results are incredibly stable and highly significant. Interestingly, the in-sample fit is slightly better using unadjusted loss costs and the average coverage level over adjusted loss costs, but it is not likely useful to consider that point further given the use of adjusted LCs elsewhere in the report and actual ratings process. The structural breakpoint proposed is also easily verified as strongly significant- as is each of the other years we tested from 1993-1997. It doesn't really indicate that there is a single breakpoint at 1995 in any sense, only that the two periods pre- and post-95 or any other split point that tends to separate 1988 and 1993 into one group will have substantially different members. There are numerous other methods to either detect or control for simply intercept shifts, though it seems rather off point to worry about technique when the coarse results all agree. it is not clear however, why a time control was not considered instead of a sample truncation response. It may be the case that there is still useful information in the variation in loss rates pre-1995 and pre 20 years that could be maintained with a more direct model of how loss rates have changed through time. Moreover, if there is a trend for discernable reasons in the loss costs, then the rolling average of 20 years eventually proposed in the reports will still suffer the same consequence, it will just be that the average bias will be made constant by the rolling time period rule.

Another observation is that, despite strong individual significance, counties are not best treated as though independent and only linked through the weather correction in CRD or climate division (for example, Champaign has a LC roughly twice as high as Piatt county in the report, yet Piatt would be considered virtually identical or even slightly riskier due to soils differences). The sample effects at individual counties in the annual loss ratios bear additional attention for ways to smooth the resulting average loss rates more effectively across space (long results file by county supplied electronically - the county results within CRD are puzzlingly variable in some cases. In the case of CRD 5 in Illinois (results on page 57 of implementation report), the ratio of the county to their CRD average LC is greater than 1 in the raw data and less than 1 in the reported final values for Champaign and vice versa for Livingston County) it seems that the adjustments are highly sensitive to own-county small sample loss experience and could be weighted by a credibility measure of their difference from the average in their region, or other simple smoothing rule to help avoid what might appear to otherwise be too stark a set of rate changes across county lines.

Overall Review Conclusions:

It is clear that the final proposed effects closely mirror the inverse of the aggregated historic loss performance by crop county for major crops, and our related analyses and own separate experiences strongly support the proposed adjustments to base rates. It is stunning to compare the final maps in the report with comparable loss ratio maps from roughly the past 20 years as they are virtually identical in shading pattern where the lowest historic loss ratios correspond by crop to the largest proposed reductions in base rates. Our sense is that roughly half of the difference between actual loss rates and the targets are represented in the proposed changes and that seems to be a reasonable rule for evolving toward targets and for allowing absorptions of large systemic losses. We find, however, the methodologies used by *Sumaria* to be obtuse, indirect, and involving procedures that are in certain cases completely *ad hoc*, in particular with regard to the re-weighting by weather experience. In our opinion, methodologies for conducting the rate revision should be relatively direct, involving simple historical loss experience compared to targeted levels with conditioning variables such as easily replicated weather indicators and insured acres. A broadly defensible conclusion is that if the features required for use of a LCR system exist, then divergences in performance from targets through time provides the best estimate of the needed correction to existing base rates. The variation through time, and extreme event risk provide limits on the degree to which one can move directly toward target loss rates though base rate revisions, but the point is that if empirical loss experience is used to set rates, then there must be a reconciliation between expected rates formed from historic experience, and actual loss rates experienced. Whether omitted or indexing variables can be found and incorporated directly, or whether other data-use rules suffice requires both careful assessments of the data, and considerations of practical implementation features that follow. In the present case, the primary consequential re-weightings proposed involve (1) complex and relatively benign final weather adjustments; and (2) simple and highly consequential sample period choices. The third consequential effect discussed is to redefine the CAT division of experience, but if properly done is a simple redistribution, not a total rate consequence, so will be treated as a mechanical exercise on load divisions.

For the weather effects, we suggest that a more complete set of alternative designs be investigated and thoroughly vetted before implementation. While there is nothing conceptually wrong with binning by variable width, we have not frequently encountered that practice in any other field of loss loading and do not see its clear advantage. The development of long series of weather cases likewise could have appeal as an indexing system, but the actual losses are not actually that tightly tied, and the accidental re-ordering of loss values actually observed is disturbing for the system to be relied upon to make finely tuned adjustments to data prone to meaningful sample variation due to other (non-weather) effects. Our own efforts to investigate the properties of the weighting process led to nonplussed outcomes -- sometimes better, sometimes worse, never clearly different -- from far simpler in-sample methods which can be related to credibility reweighting more directly. And, it may be the case that we were simply unable to follow the details provided in the report, but there is no clear way to decide which adjustment to make first, and if the later sample data differ in structure from the early period, it is not clear how to use a weather index across all the periods anyhow. If the weather adjustment is made independently of the sample period alterations, then it is not clear how to link the effects the opposite direction. Finally, other more obvious adjustments to the historical loss rates due to simple liability/indemnity changes in the current programs (i.e., current program total losses would generate lower loss ratios than observed in at least two prior periods with partial losses) should be investigated first before committing to such an opaque weather conversion process.

The second issue broadly treats the question of data representativeness as one of data period selection rather than to search for other control variables to include in a structural model of loss experience. In total, we cannot argue with the magnitude of the effect. It is in particular striking how closely the proposed directional changes mimic deviations from target loss ratios. There are numerous compelling arguments for and against methods for representing base rates from historic LC rates, and the present proposal tends to take highly stylized arguments for representativeness from the perspective of a P&C loss curve stability standpoint. That is not incompatible with other features, but may be less responsive through time than a model that, say, included direct controls for things like the relative risk in yields, percentages of acreage planted by crop, the average soil quality underlying insured acreage, weather controls, and insurance product features (coverage, revenue elections, etc.). The quality of information about the insured and their experience data have also improved through time fundamentally altering the grouping risk losses endemic to any insurance program that has to rely on some forms of average experience to insure individual exposures. Taken together, it may simply be better or easier, or a combination of both, to search for limitations against the data as observed to effectively stratify them for representation of a different experience (future expected losses). To that effect, the present primary recommendations are to create a shift for losses prior to 1995 to equate the mean rates, ceteris paribus, and to shorten the rolling window from which loss rate averages are constructed. All other approaches we have investigated reach similar qualitative conclusions. There is some meaningful insight in the sentiment that it is better to have an approximate answer to the right question than an exact answer to the wrong question.[3] We would also point out that most alternative methods lead to very similar conclusions - one could periodically take a fixed share of the difference between long run loss cost ratios and the target as the percentage change in base rates and get roughly a calibration rule that would lead to the same outcomes. One could also use methods in the literature to recalibrate based on estimated underlying yield function evolution and get virtually the same implied rate reductions.[4] Each of the alternatives investigated in regression frameworks led us to similar conclusions regardless of the weather variables used, the aggregation from county to CRD or to state, and regardless of the particular division in time identified.

In summary, we commend the authors and RMA for their sincere ongoing efforts to continue to improve the performance of the federal crop insurance programs. It is clear that the importance of these programs will continue to increase in the future, and the perception of the effectiveness of the programs stands to be improved greatly through the recalibration of base rates to more nearly coincide with long run average costs.

[3] Quote generally attributed to John Tukey, "Far better an approximate answer to the right question, which is often vague, than the exact answer to the wrong question, which can always be made precise."

[4] See Woodard, et. al, *Journal of Agricultural and Resource Economics* 36(1):211–228, and references therein for some recent examples.