Methodology Analysis for Weighting of Historical Experience

Responses to Reviewer Comments

November 4, 2011

By

Keith H. Coble
Mary Frances Miller
Roderick M. Rejesus
Ryan Boyles
Thomas O. Knight
Barry K. Goodwin

Contact:

George Duffield
Sumaria Systems
618-632-8450
gduffield@sumaria.net

Keith H. Coble
662-352-9491
khcoble11@aol.com

Responses to Reviewer Comments

General Comments

We appreciate the opportunity to respond to the reviews. This project involved combining actuarial, econometric and climatological expertise to address some of the unique attributes of crop insurance. We believe we were able to develop a credible and implementable set of recommendations which will significantly improve the actuarial fairness of the Federal Crop Insurance program. We have written specific responses to each reviewer's comments, but will begin with a few general comments that we believe are essential points to clarify about our study.

<u>Issue 1: Adjusting loss experience for non-stationarity</u>

The proposed adjustments to experience prior to 1995 are significant and impactful on rates. Reviewers generally recognized this and devoted significant attention to the issue. It appears that reviewers generally agree that for various reasons, including program changes, changing technology, and participation, there need to be adjustments to older experience data.[1] We examined various approaches to this issue including trends, alternative weighting schemes, and mixed approaches. In the end, we recommended a conceptually simple and easy to implement approach – a discrete breakpoint adjustment. Of the reviewers who addressed the issue, the main point of disagreement regarded where the breakpoint should be located. This is a fair question. We revisited the issue empirically and evaluated breakpoints of 1989, 1999, and 2004 in addition to the original breakpoint of 1995. We found that several alternatives were statistically significant. The strongest competitor for the 1995 breakpoint was the 1999 alternative suggested by Dr. Babcock. Having compared the two, we conclude the 1995 breakpoint is the more robust but just marginally more so than 1999.

<u>Issue 2: Reducing the number of years of historical data used in calculating the base rate</u>

In insurance ratemaking, there is always a tension between the desire to maximize data volume and the need to ensure that the data used are reasonably free from material limitations. Ratemaking principles require that due consideration be given to trends, changes in policy provisions, operational changes and changes in the mix of business. RMA's program has gone through several changes over time and continues to evolve. In addition, farming techniques and the nature of the crops themselves have undergone continuous change over the period of time for which RMA data are available. The longer the period of time between the observed data and the present, the less likely it is that the data reflect the current program and current conditions. Any adjustment of historical data to attempt to bring it to current levels is dependent on assumptions about the magnitude and direction – and often the interaction – of the effects of the changes. Hence, generally accepted actuarial practice seeks to shorten the length of the time series of data used in the ratemaking process. For the catastrophe load, the extreme infrequency of very large losses leads to the conclusion that all available data must continue to be used, after reasonable

[1] One reviewer did not address the issue.

adjustment for measurable changes in the program. On the other hand, once the data have been censored by the catastrophe procedure and adjusted to reflect the long term weather patterns, it is no longer necessary to include very old, less relevant data.

We considered 15, 20, 25 and 30 year data sets. Given the proposed binning process to account for weather variation, 15 years risks that the sample will be too short to reflect a sufficient range of weather. Twenty years provides sufficiently many observations to make the binning process practical, while the addition of more years does not significantly improve the process.

We acknowledge that reducing the number of years of data in the base rate calculation may in some cases produce more volatility in the rate indication. However, stability should not be confused with predictive accuracy. Continued inclusion of old data of suspect validity perpetuates any bias introduced by such data. Further, weather weighting mitigates the potential small sample problems of using fewer years of data.

<u>Issue 3: Weather weighting and binning of the loss cost data</u>

Probably the most critical comments received were related to the proposed weather weighting and binning procedures. In our responses to individual reviews we provide additional information and explanation of what was proposed. However, to be forthright, we simply disagree with some of the reviewers. We will attempt to provide a clear explanation of our logic.

First, we want to address the assertion of several reviewers that weather weighting does not make much difference in rates. We have included two maps that show that weather weighting does significantly affect rates in many counties. However, as expected, the rate changes are both positive and negative such that at the national level aggregate data masks the disaggregate effects. Given the magnitude of the effects in many locations we must conclude that failing to account for weather will result in adverse selection.

The proposed weather indexing procedure needs to be understood for what it really does. It is used to rank historical years so that a longer history of weather events can be used in assigning probabilities to recent years which go into RMA's loss cost calculations for rate development. In some instances the statistical model selection approach was probably not explained as well as it might have been, but the proposed model selection procedures are widely used and well accepted in econometric modeling. Some reviewers suggest a simpler OLS procedure, while Dr. Sherrick suggests other more complex estimation procedures. We believe our proposed approach strikes a reasonable balance between complexity and simplicity.

The binning procedure we propose is unique to the task at hand, but we believe it is a valid actuarial approach. This simple process of averaging data in bins is a variant of a histogram. We do not make a parametric assumption or generate simulated rates, to avoid imposing structure on the data that is not well justified. We note that the proposed procedure was deemed acceptable by the two actuarial reviewers. We are not surprised as this is the approach actuaries tend to take. Sherrick appears to prefer imposing parametric structure on the problem. In our opinion that might have merit if there was *clear* evidence as to what parametric form generated the data. However, we do not believe such evidence exists. Again, we prefer to follow typical actuarial practices and avoid those assumptions while allowing interaction with other adjustments such as acreage weighting. We note that while there is some concern with the noisiness of the binning procedure, there was not a suggestion that binning failed to properly reflect the probability space. Dr. Ker's review focuses on the noisiness of the binning process. In our opinion, the simulations Ker reports reflect a single distributional scenario and thus the results depend on the data presented. Further it appears that catastrophic loading was not modeled which would have dampened the variance of the data and stabilized the process. Ultimately, we believe the criticism is really directed to the issue of small samples rather than binning itself. Finally, there is no comparison to alternatives such as using a standard histogram as a point of reference.

Specific Responses

1. **Responses to the review by John Pierce Consulting Actuary**

The summary of this review states, "I believe the arguments made in the submission for these changes are convincing."

Response: We appreciate the confirmation of our recommendations.

The John Pierce review does pose several questions which we will attempt to address.

Comment 1
There is no mention of global warming in the submission. Because of the possibility that global warming – or some other long-term trend in weather – does exist, it would appear necessary for the submission to consider the possible impact of this phenomenon on its calculation of weather indices.

Response: *The study team agrees that there are indicators and substantial evidence for trends and long-term changes in climate, especially at global and national scales. However, at the local scales used for this analysis, year-to-year variability generally dominates any background trend. Temporal variance is so high that statistically significant trends are generally not identifiable over the past 20 years used for the analysis. We don't explicitly state this in the report, and we appreciate the reviewer alerting us that this would be of interest to some of the intended audience.*

Another key point that the study team considered is that the purpose of this methodology is to estimate the best rate for the next year, not the best rate for 10, 25 or 50 years from now. We agree that climate change trends could have substantial impacts on Loss-Cost Ratios over the long term, but our analysis strongly suggests no need to adjust for climate change in the year-to-year rate making process given the recommended use of the past 20 years of Loss-Cost data.

We also note that the nationwide effect of weather binning is minimal, supporting the assertion that overall recent weather patterns do not exhibit a dramatic difference from the long-term.

Comment 2
The fractional logit regression procedure is used in the submission. I believe the information provided in the submission leaves a significant part of this procedure as a "black box" for the reviewers.

Response: *The methodology is appropriate to this application and has been vetted through replication of published studies on the topic. The code, though not provided in the reports, has been thoroughly reviewed and consolidated by the RMA and Sumaria.*

Comment 3
Various aspects of the pre-1995 adjustment are unclear. The Technical Report uses ordinary least squares to calculate three methods for calculating this adjustment. However, the technical report does not provide sufficient information for us to actually calculate this adjustment. In

addition, the Implementation Report suggests using the fractional logit regression method for calculating this adjustment, even though the Technical Report used ordinary least squares regression. Finally, the Implementation Report does not reach a conclusion as to whether this adjustment should be on the state, regional or national level.

Response: *We understand the question posed as there was an evolution of thinking as the project progressed. While we originally proposed using OLS it was later decided that fractional logit was easily implemented and consistent with the structure of the data used in the weather models. Thus, in the Implementation Report fractional logit was proposed. You are correct that we did not make a recommendation for a specific level of aggregation. In our opinion this is an empirical question that depends on the crop and region. In general, we recommend estimating the effect at the state level, but in some instances aggregating up to a higher level, such as regional level, will be appropriate when state data are thin.*

Comment 4

The Climate Division Data is available yearly from 1895 to present. However, some data from the pre-1931 years was not available, and had been estimated using regression techniques. The submission does not investigate whether using data from only 1931 to present would be an improvement over their approach of using 1895 to present data.

Response: *Guttman and Quayle (1996) explain that statewide averages of climate data had been calculated since the late 1800s, and these statewide data were used to develop climate division data for each state. This process used regression equations that related each climate division to the statewide average using overlap periods when both state average and climate division data were available from homogenous sources (following Karl, Metcalf, Nocidemus, and Quayle 1983). Guttman and Quayle state that most of the linear regressions produced correlation coefficients above 0.90. However, the pre-1931 estimates have been shown to be a potential source of error, most notably documented by McRoberts and Nielsen-Gammon 2011, Keim et al., 2003; Keim et al., 2005; Allard et al., 2009.*

In the past year, there has been an effort to re-calculate climate division values using new techniques and newly digitized climate observations. This new effort by NOAA will eliminate the need for pre-1931 regressions and should produce more accurate estimates of climate division (and county-scale) climate variables for the 20th century. These data are not yet published, and so are not available for use in this analysis. However, NOAA estimates that these new climate division data will be available with the full history since 1895 sometime in 2012.

The study team gave consideration to the tradeoffs involved in deciding whether to use the pre-1931 climate data, balancing the known issues with pre-1931 estimates with the need for the longest possible record. We decided that information on the relative extremes of climate pre-1931, and the additional 35 years of climate data, outweighed concerns with the regression-based values used prior to 1931. The primary justification is that the raw climate data for any climate division were not critical to this methodology, but instead the relative ranking of climate severity as compared to the remaining record since the ranking is critical for the binning process. In other words, the dry years and wet years are still captured with relative severity, which the authors felt was a requirement 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." While the regression estimates of pre-1931 climate have errors in the precise values for each variable, they still capture the relative climate severity for that period, which we feel warrants their use to gain the additional 35 years of climate history.*

Comment 5

The RMA premium and loss experience by year and by crop provides some evidence that breaking the experience period between pre-2003 years and 2004 and subsequent years might be preferable to the submission's breaking of the experience into pre-1995 years and 1995 and subsequent years.

Response: *Several reviewers suggested alternative breakpoints for adjusting historical data. In addition to the 2004 break that you suggest, 1999 and 1989 were suggested. We understand why casual observation might suggest 2004; however we are unaware of any structural change in policy or participation that might suggest that breakpoint. We conducted additional evaluation of breakpoints under the assumption that a common breakpoint should be imposed in all crops and regions. Based on that analysis we conclude that 1995 is the preferable breakpoint and is supported by programmatic changes occurring at that time. The 1999 breakpoint would appear to be the second choice across crops and also would reflect a break in program participation.*

2. Responses to the review by Oliver Wyman Actuarial Consulting

We appreciate the review by Oliver Wyman which concludes:

Comment
"…the Applicant has provided sufficient actuarial support for the above recommendation. Additionally, the recommendations are in compliance with the referenced actuarial standards. However, if implemented, that there will be significant rate changes for individual insureds. The Applicant does not present a mechanism in the implementation report for modifying, alleviating, or phasing in these changes nor is there a discussion of the potential dislocation in the market due to large changes."

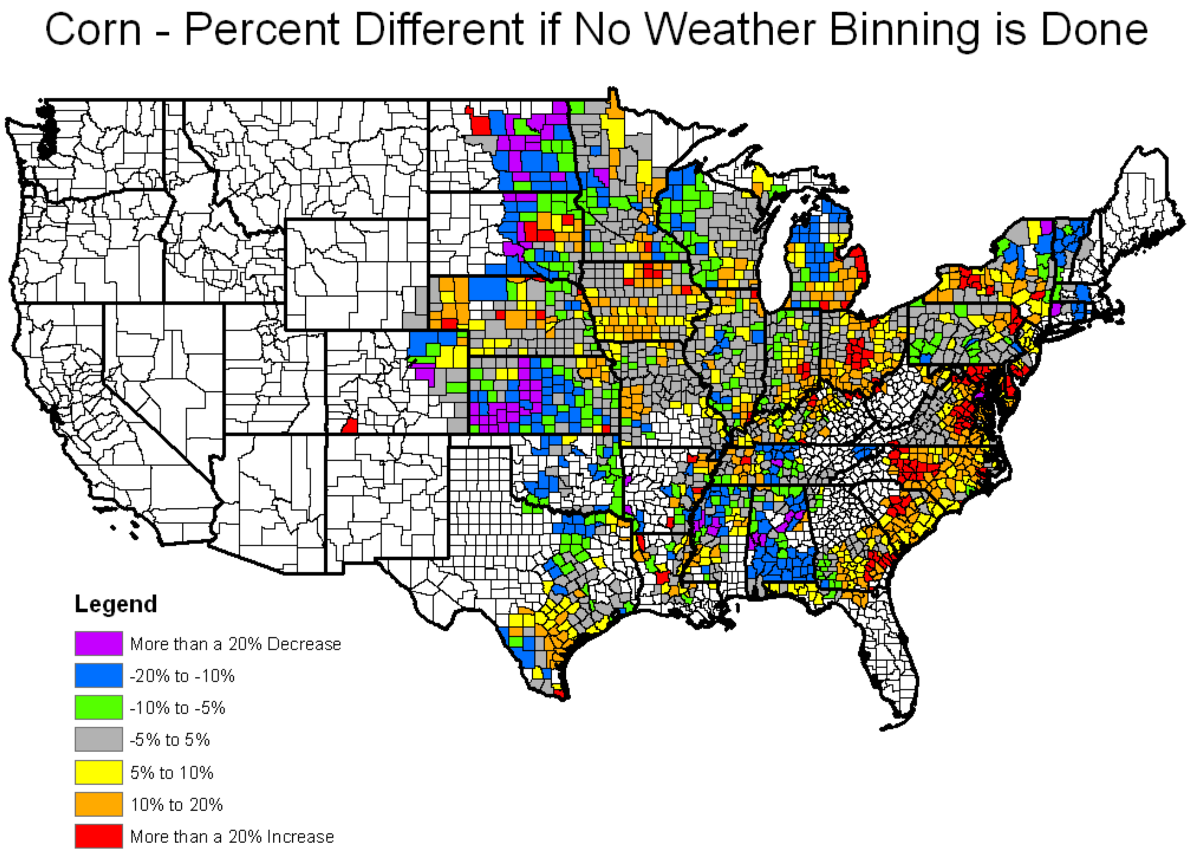
Response: *We appreciate the positive review. Also we generally concur with the suggestion that the proposed rate changes be phased in so that the effect will be gradually felt by those impacted.*

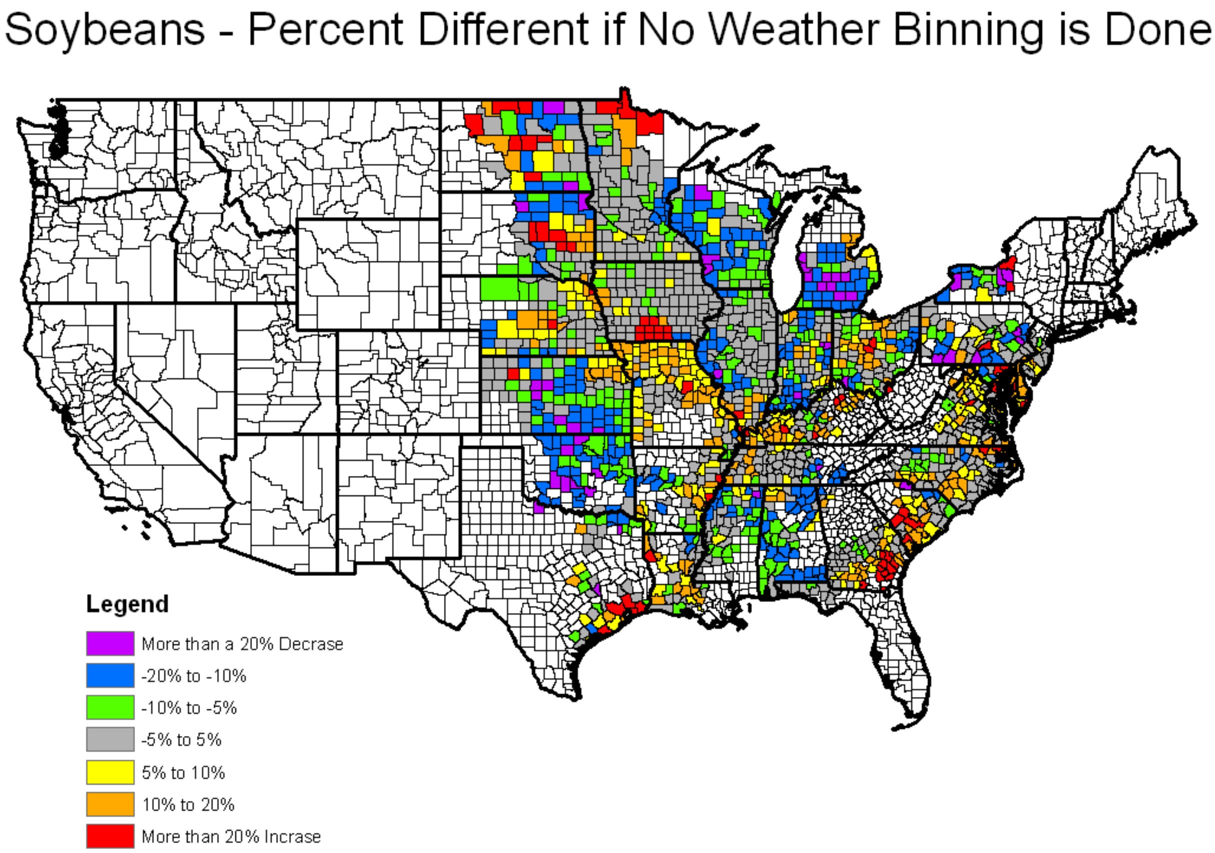
3. **Responses to the review by Bruce Sherrick**

Comment: (page iii)
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.

Response: *The proposed weather reweighting procedure has a simple premise – utilize actual historical loss experience while allowing the weight given a particular year to reflect additional data regarding the probability of weather events. The proposed binning procedure specifically avoids making distributional assumptions because we do not know the true data generating process for losses. Thus, we find the evaluation conducted by the reviewer to be largely irrelevant. We are familiar with the alternative approach suggested and do not consider it simpler. The procedure proposed by our team takes weighted averages of data in each bin and averages those – a very simple process that was confirmed by the actuarial reviewers.*

We also appreciate the opportunity to respond to the assertion that weather weighting does not make much difference. This is simply a factually incorrect assertion. As one would expect weather is random and rate inaccuracies resulting from ignoring weather events tend to average out across regions. However, we include two maps in this response that show how much weather weighting affects rates in isolation from the other effects proposed in the report. Each map shows the difference in rates without weather binning as compared to the rates developed under the proposed methodology. It is clear that on a local scale there is significant variation in recent weather as compared to long-term patterns.


Corn - Percent Different if No Weather Binning is Done
Legend
More than a 20% Decrease
-20% to -10%
-10% to -5%
-5% to 5%
5% to 10%
10% to 20%
More than a 20% Increase


Soybeans - Percent Different if No Weather Binning is Done
Legend
More than a 20% Decrase
-20% to -10%
-10% to -5%
-5% to 5%
5% to 10%
10% to 20%
More than 20% Incrase

Comment: (page iii)
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 97th percentile does not work as described in the recommendation to better reflect loss rate experience and should not be implemented.

Response: *We agree with the reviewer's comment that the longer experience period be used to calculate both the division point between the CAT and base loads and the magnitude of the CAT load. In the implementation report example, the full 31 years of experience are used to calculate the 90th percentile loss and to cap individual years' losses in the base rate calculation. The 20 year period, after capping and adjustment for variation in the weather pattern, is sufficient to calculate the base rate. However, the extreme rarity of catastrophic events precludes the exclusion of any relevant loss experience from the calculation of the catastrophic load. Other than the shifts in cutoff from 80th to 90th percentiles and from a statewide load to one based on climate division, the procedure for calculating the CAT load is essentially unchanged. The only additional recommendation that we have made is to adjust for years in the data where the observed weather would be expected to occur less frequently than once in the number of years of observation based on the longer term weather index. Since data are available for at least 30 years in almost all cases, only weather indexes above the 97th percentile would be candidates for adjustment. Hence, the implementation report recommends only examining years above the 97th percentile index. The 97th percentile is therefore neither ad hoc nor arbitrary.*

Comment (page iii)
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 the 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.

Response: *In general we appreciate the confirmation of the proposed approach.*

Comment: (page iv)
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.

Response: *While not a focus of this project the issue of spatial smoothing was addressed in our previous report.[2] RMA does use credibility weighting to smooth rates across contiguous counties. While it appears RMA's approach differs from the proposed approach it essentially accomplishes the same thing.*

Comment (page 5)
"Proposed approach is simply untractable *[sic]* for outside replication … and unduly complicated."

Response: *The approach is entirely appropriate, is straightforward in concept, design, and implementation, and has already been automated through SAS coding. These facts are counter to the reviewer's assertions.*

Comment: (page 6)
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.

Response: *We do not agree with the suggestion to use yield deviations, which are correlated with but do not completely equate to crop insurance losses. A simulation using yield data would introduce error into the process. Actuarial principles lead us to use actual loss experience. Thus, the yield deviation modeling is actually less plausible and inappropriate.*

Comment (page 7)
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.

---

[2] Coble, K.H., T.O. Knight, B. J. Goodwin, M.F. Miller, and R. Rejesus, "A Comprehensive Review of APH and COMBO Rating, A report prepared for the USDA Risk Management Agency, March 17, 2010.

Response: *A wide range of alternative specifications were considered and the final forms applied reflect the outcomes of this extensive evaluation. Suggestions to adopt specific regressors based on the reviewer's own priors are simply not supported by the out-of-sample forecast evaluation. Further to the point, this analysis was designed to minimize the potential for over-fitting in the relatively small samples that are available. This is also the rationale for determining the specification at the state-crop level—to minimize the potential for over-fitting by basing the final specification on the larger set of pooled data.*

Comment (page 9)

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.

Response: *It strikes us that this is largely an argument about relative merits of parametric and non-parametric approaches. We have chosen to 'let the data speak' rather than impose structure. We elected to take that approach as there is relatively little guidance on the appropriate structure to impose.*

Comment (page 12)
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.

Response: *Our use of the term non-stationarity simply implies we assessed whether there are trends or discrete breaks in the loss cost time series. Various tests were conducted assuming differing functional forms. As clearly stated in the report, we chose 1995 based on statistical significance and because of a clear structural change in program participation and Federal policy. Based on the useful suggestion of another reviewer we also evaluated 1999 as an alternative breakpoint. It is our conclusion that 1999 is a close second, but we still find 1995 is a more statistically significant breakpoint.*

Comment (beginning on page 13)
The reviewer states "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.

Response: *Our first response to this entire section is that it appears to bring in a set of issues that can be simply summarized as "Throw out the entire rating system and do it another way (i.e., my way)." We will address some key points from this section.*

- *The purpose of our report was well defined and many of these comments do not pertain to our analysis.*

- *The point about providing empirical context of data was essentially accomplished in our previous program review.*
- *Adjusting data to a base coverage level was also addressed in our comprehensive program review and found to be consistent with actuarial principles by our team including Mary Frances Miller, FCAS, MAAA, and CPCU.*
- *The suggestion to back-test through experience data from production reports and type 15 record systems is one that we reject out of hand. Members of our team have 20 years of experience using the type 15 records in other studies. We would not suggest using these data to validate actual loss data because actual losses are not fully captured by the APH data.*
- *The suggestion to spatially smooth data appears to ignore the fact that RMA already smoothes rates across spatially similar counties.*

Comment (page 22)
"…fractional logit models are not widely in use."

Response: *The extent to which any specific econometric procedure is used is necessarily dependent on the specific circumstances which the procedure is intended to address. The appropriateness of a procedure is not defined by how frequently it is used—it is either appropriate to a particular application or it is not. In this case, the fractional logit model is the most appropriate methodology. Although the results are not sensitive to the use of a fractional logit model (as compared to an OLS regression), the methodology remains the most appropriate for such fractional data. Further, the fractional logit was adopted based on concurrence between our study team and RMA analysts as to the appropriateness and practicality of the procedure.*

4. Response to the review by Bruce Babcock

Comment on Recommendation 1. RMA should use Climate Division Data for calculating crop-specific weather indexes. This recommendation consists of two parts. The first is that RMA should construct weather indexes. The second is that RMA should use a particular data set to construct the weather indexes. There are sound reasons why RMA should use weather indexes to help it set premium rates for yield insurance. Construction of weather indexes potentially allows for better estimates of the likelihood of future weather events because weather records cover a much longer time span than is covered by RMA's loss cost data. Use of a 100-year weather history can provide better insight into whether the probability of recurrence of a 1993-type weather event (lack of heat in parts of the Corn Belt and excess rain) is a 1-in-30 year event or a 1-in-100 year event. If it is a 1-in-100 year event then the loss cost for 1993 should be given 70 percent less weight in rate making than it now receives. Coble, et al do a good job justifying why the Climate Division Data is the most appropriate data set to use to construct the weather indexes. The only significant weakness of the Climate Division Data for use in rating the most important crops in the crop insurance program is the use of regression analysis prior to 1931 for allocating state average growing conditions to climate divisions. However, with the possible exception of geographically large and diverse states, such as Texas and Montana, there is a high correlation between state average weather and climate division weather, so the regression estimates should adequately capture significant departures from normal weather.

Response: *We appreciate the positive review of this step in the process.*

Comment on Recommendation 2. RMA should use fractional logit models to estimate weather indexes with Palmer Drought Severity Index (PDSI) and Cooling Degree Days (CDD) in the regressions. Because loss-cost data are by definition limited to between zero and one it makes sense to use an estimation technique that explicitly accounts for these limits. But, as the authors point out the results of the weather index regressions are only used to rank years in terms of severity of losses. It would be surprising if OLS rankings would differ in a meaningful way from rankings obtained from fractional logit regressions. If the computational burden of fractional logit is small relative to OLS, then I agree with the recommendation. If the computational burden is high, then I recommend using OLS instead. Use of PDSI and CDD to construct the weather indexes makes sense. PDSI captures lack of precipitation, excess precipitation, and excessive heat, all of which can lead to crop losses. CDD captures the amount of heat during the growing season. I make some suggestions in the main part of my review for improvement in how the weather indexes should be estimated. My judgment is that the recommended approach is sound, but RMA should verify the reliability and performance of the regression equations for each climate division before implementing the procedures. A simple plot of predicted loss cost rank against actual loss cost rank provides insight into this reliability.

Response: *It is the case that the rankings are generally not sensitive to the application of the fractional logit versus OLS. However, there is no significant additional computational burden imposed by the logit model.*

Comment (page iii): A discrete adjustment should be made to pre-1995 losses and a 20 year loss history should be used for base rates.
There is ample justification for making a discrete adjustment to losses in the early period of the crop insurance program. However, Coble et al's justification for choosing 1995 as the year to make this adjustment rather than 1998 or 1999 is not adequately documented. Use of a 20-year history for base rates combined with a longer time period for a catastrophic load is a simple, straightforward change that would result in current premium rates that reflect both modern production practices and all available observations of high loss years.

Response: *Several reviewers suggested alternative breakpoints for adjusting historical data. In addition to the 1999 break that you suggest, 2004 and 1989 were suggested. We understand why buy-up participation data would lead to the suggestion of using 1999 as a break point. We conducted additional evaluation of break points under the assumption that a common break point should be imposed in all crops and regions. Based on that analysis we conclude that 1995 is the preferable break point and is supported by programmatic changes occurring at that time. The 1999 break point would appear to be the second choice across crops and also would reflect a break in program participation.*

Comment (page 6 and following) Use of the PDSI and CDD in the weather index regressions makes sense. PDSI captures lack of precipitation, excess precipitation, inadequate soil moisture and excessive heat, all of which can lead to crop losses. CDD captures the amount of heat during the growing season. Coble et al improve on some of their earlier work by including separate regression coefficients for both positive values of PDSI and negative values of PDSI. As shown in Table 4.4 for Indian soybeans, increases in PDSI when PDSI is negative (dryer and hotter than normal conditions) has a marginal effect on loss costs of -0.8383 whereas increases in PDSI when it is positive have a marginal effect on loss costs of 0.2246. This means that when conditions are hot and dry, reductions in hot and dry conditions (increases in the PDSI) decrease loss costs. When conditions are wet and cool, then increases in PDSI increase loss costs. This illustrates the importance of allowing for differential marginal response to changes in the PDSI variable depending on whether it is negative or positive.

I am surprised that the same logic was not applied to CDD. It appears that only one marginal response to changes in CDD was allowed. But for many crops, increases in CDD are beneficial up to some point at which further increases cause losses. For example, Yu and Babcock (2011b) estimate that in the Northern Corn Belt, corn yields increase modestly with increases in CDD when CDD is low. But when CDD is high, increases in CDD cause sharp yield losses.

Response: *Regarding the use of separate CDD variables: The study team did not consider possible use of the differential response of CDD because CDD is accumulated only from base 65, and does not have a universal break point for where CDD accumulation becomes helpful or hurtful to crop production. CDD (also known as growing degree day base 65) measures heat accumulation as mean daily temperatures above 65F. Days with mean temperature below 65F (which would be negative) are not added into the accumulation, and therefore do not negate the values. CDD can therefore not be negative.*

CDD measures all accumulated heat above 65F, not just excessive heat accumulation. Excessive heat becomes a location and crop specific threshold, and therefore difficult to estimate for RMA's ratemaking process.

Because PDSI is an index that by definition separates relative wet and dry conditions at zero, it is possible to allow for differential response in the variable. Equivalent treatment of CDD is not possible, as it is never negative.

We did consider using Heating Degree Days (HDD) at base 65, which is also available in the NOAA Climate Division Data. HDD represent chilling accumulation, and could be added into the weather indexing process.

The study team was very concerned with the use of too many variables and the potential for model over-fitting given the limited LCR data history. Therefore we chose to use CDD which we considered a better indicator than HDD of more typical loss during the growing season.

Comment (page 7)
"If cross-validation was used by fitting the models many times using subsets of the entire sample and predicting the rest of the sample, then the out-of-sample data used for model selection contains all the data that is available."

Response: *Yes, the method was traditional cross-validation (or "leave-one-out" validation) and was performed for every observation, conditional on all the others. The review does make reasonable suggestions for alternative model selection criteria, including the use of ranks and weighted errors. We agree such methods have merit but at the same time note that the possible alternatives are unlimited and any specific example of a county or district is not likely to be generalizable to the entire portfolio of insurance contracts. That is, it will always be possible to find examples where the proposed method does not perform well and to recommended alternatives that are better in such cases.*

Comment on Recommendation 3. RMA should place each year of loss cost experience into discrete probability categories that are defined by a long-term history of weather. Some method must be used to determine how likely a set of growing conditions that occurred in a particular year in RMA's loss experience history will occur again in the future. The proposed method of using probability bins is a robust method that when combined with Recommendations 4 and 5, should result in a much more reliable rating system. My only reservation with this approach is that it could lead to poor rate making if prediction errors from the weather index equations are large.

Response: *We concur with this comment and believe that the statistical criteria used for selecting the weather index should be sufficient to avoid predictive errors. Otherwise, we suggest weather weighting be omitted.*

Comment on Recommendation 4. RMA should change its method of calculating catastrophic loads by adopting a 90th percentile load cap, by spreading the load to the climate division instead of the state, and by dampening the weight given to the most extreme weather years. Spreading catastrophic loads to the state level subsidizes farmers who live in climate regions within a state that are prone to large losses and penalizes farmers who live in climate regions that are not. This proposal is consistent with the sound proposal to use make climate division as the basis for categorizing weather as it impacts crop insurance losses. Coble et al do not clearly justify a 90th percentile cap versus an 80th percentile cap. Either would work with the rest of their proposal.

Response: *A 90th percentile threshold for catastrophe ratemaking is more consistent with catastrophe procedures in other lines of insurance. Milliman and Robertson performed a comprehensive evaluation of the RMA cat loading process[3], concluding that there are several potential changes to the process that would improve the efficiency ratio, and in particular recommending that RMA change the catastrophe threshold to the 90th percentile. Our 2010 study[4]concurred with the recommendation. We agree with the reviewer's observation that either cap would work within the rest of the proposal.*

[3] Milliman, Inc. 2008. "Fixed Rate Load Review."

[4] Coble, K.H., T.O. Knight, B. J. Goodwin, M.F. Miller, and R. Rejesus, "A Comprehensive Review of APH and COMBO Rating, A report prepared for the USDA Risk Management Agency, March 17, 2010.

5. Response to the review by Alan Ker

Comment
The authors have made 5 recommendations. The first four refer to the use of weather variables to augment historical loss cost experience. Recommendations 1-3 illustrate steps the authors took in developing a methodology to augment the historical loss cost experience while recommendation 4 suggests adopting the methodology. The fifth recommendation refers to the nonstationarity of loss cost.

With respect to the methodology employed by the authors to incorporate the use of weather variables to weight the historical loss costs I have a few major concerns. The use of the Climate Division Data (National Climatic Data Center's Time Bias Corrected Divisional Temperature-Precipitation-Drought Index data) appears to be an appropriate choice with respect to weather variables.

I am comfortable using the fractional logit model to map from the various weather variables to an index (loss cost) although as the authors point out more simplistic approaches would lead to almost identical ordinal rankings and thus lead to very similar results.

Response: *This comment supports one of our significant recommendations. We agree that simpler estimation procedures would likely provide a substantially similar ordering of years.*

Comment
I am significantly less comfortable with the use of the probability bins. As the number of bins change the resulting rates can change very dramatically. It appears that the authors choose the number of bins somewhat arbitrarily in the sense that they start from a fixed number (15) and reduce the number of bins such that at least every bin has one realization. Currently, the rates assume a single weather bin. My simulation results suggest that the rates may vary quite significantly with the assumed number of bins.

Response: *This review was taken seriously and we have revisited this issue in response to the comments and simulation results. We note that the binning procedure was constructed to decrease variability of the estimates relative to a standard histogram. It is not clear to us how robust the simulations results are since they depend on the assumed data generating process.*

We do think the evaluation conducted failed to account for the fact that the procedure would be applied to data that have had the catastrophic load removed. Removing the catastrophic portion dampens the variance of the data and stabilizes the process. Finally, there is no comparison to alternatives such as using a standard histogram as a point of reference.

Comment
The fourth recommendation is that RMA use this methodology to estimate base rates. This is perhaps where I have my biggest concern. First, the authors have not shown that these methodologies actually lead to more accurate rates, different rates yes, but not necessarily more accurate rates. While they have shown that there is a statistically significant relationship between weather and historical loss costs for some/most crop-region combinations, how that (estimated and assumed) relationship is exploited through construction (with assumptions) of weather bins

for increased rate accuracy is not obvious nor is it proved. This was brought up by a previous is not obvious nor is it proved. This was brought up by a previous reviewer and while the relationship may be statistically significant as the authors note in their response, that fact does not suggest that the methodology leads to statistically different rates nor a more accurate rate. My simulations suggest trivial improvements in accuracy which begs the question is this sufficient for what is a very significant change in complexity.

Response: *Accuracy of the rates is an extremely difficult concept to measure – or even to attempt to measure. Certainly, very long term deviation in results from the target would support an assertion that the rates are biased, but especially given the large variance in RMA experience, large deviations in experience relative to targets over a short or even medium term might not indicate that the rates are inaccurate. On the other hand, significant criticisms of RMA's current ratemaking methodology include (a) that it implicitly assumes that the frequency of adverse weather in the data set used for ratemaking reflects the expected frequency of such events and (b) that all available data are equally relevant. Given the significant changes over time in programs, practices and types, it is clear that data far removed in time from the present are not representative of current conditions. Because it is difficult (or impossible) to measure the effect of such changes and adjust old data accordingly, generally accepted actuarial principles suggest the exclusion of data of suspect relevance. The proposed process allows RMA to drop very old data years while still ensuring that the data used are adjusted for potential distortion introduced by recent local weather inconsistent with the long term pattern.*

Comment

With respect to the fifth recommendation it is not surprising that the authors find non-stationarity in the historical loss cost given the significant participation changes, technological improvements including biotechnology, and even farmer attrition. While I do not have the raw data to reproduce their results, I am comfortable with shortening the loss cost history to minimize these impacts. I am also comfortable with using acreage data as weights in the historical loss costs but again would have concerns, although less so, about using bins.

Response: *We appreciate this confirmation of a significant recommendation.*

Comment (page 7)

I am comfortable with the use of the fractional Logit model but I do wonder whether RMA will have the expertise to estimate these models in the future.

Response: *The fractional logit approach was chosen in consultation with the RMA. It is appropriate, as most reviewers note, and can be incorporated into rating by RMA personnel (as has already been shown to be the case).*

6. **Response to the review by NCIS**

<u>Part 1 – The 20 Year Experience Period</u>

Comment (page 8)
In evaluating the ratemaking methodology, it is important to recognize that crop insurance rates are established from a bottom-up manner, separately by crop and county. This differs from more traditional actuarial ratemaking methods that operate on a top-down basis, starting with a statewide rate indication and using an allocation process to spread the statewide rate change in a seemingly reasonable manner to all territories within the state. The inherent conflict of actuarial techniques with actuarial principles is not readily apparent when using the traditional top-down approach, but is a critical issue when rates are established using a bottom-up approach as in crop insurance. The distinction arises out of the high variability of county data in relation to the relatively stable and predictable experience observed on a statewide basis for most types of Property and Casualty insurance. Even capping losses at the 80th or 90th percentile can be ineffective at resolving the problem in that the capped loss costs are still highly variable (as measured by their coefficient of variation, for example).

Response: *On the contrary, it can be shown that top down and bottom-up approaches are algebraically equivalent. Regardless of the approach taken, the selection of data is always dependent on the need to balance responsiveness to changes in conditions with the need for a sufficient volume of credible data. We note that the recommended 20 year experience period to be used in calculating the base rate far exceeds the time span used to establish base rates in other property/casualty coverages. At the base rate level, after capping of catastrophic losses and adjustment for unrepresentative weather patterns, we believe that the most recent 20 years of experience are likely to be a better predictor of the next year's expected experience. RMA's procedures for credibility weighting and capping changes will also mitigate against large year to year swings in the rates.*

Comment (page 8-continued)
The problem can be easily illustrated by considering an insurance program having low likelihood of a severe loss. While this is a simplified example, the concept applies equally well to the proposed ratemaking methodology. Loss capping is disregarded in order to simplify the discussion. Assume that the loss cost is 10 in normal years (years 1 through 9) and 50 in the tenth year, and similarly in future years. In this example, the expected loss cost is equal to the long-term average loss cost of 14. Now consider an insurer that begins to offer coverage starting in year 6. Suppose that the insurer employs a ratemaking procedure that uses a five year experience period so that each year's rates are equal to the average loss cost over the prior five years. Over the long term, this results in a rate of 10 for the first five years, followed by a rate of 18 over the next five years. This pattern continues to repeat over time. However, at no time does this method establish a rate equal to the long-term expected loss cost of 14. Instead, it alternates between an inadequate rate of 10 and an excessive rate of 18. In other words, the insurer experiences a significant loss in one year out of 10 and attempts to recoup the loss by overcharging its policyholders over the following five years. Not only does this ratemaking approach clearly fail to set the rate at the expected value of future costs, it also fails to meet the

actuarial requirement that rates be established on a prospective basis. This methodology can be more accurately described as a recoupment process.

Response: *While the example clearly results in inappropriate rate projections, rather than illustrate issues with the proposed methodology, it illustrates why it would not be appropriate to adopt a shortened experience period without both capping the base rate experience and continuing to use a longer period for the CAT load. It is also inappropriate to characterize this example as "recoupment." We note that, although the charged rate is never the correct long-term loss cost, in fact the rates charged over time would equal the losses.*

Comment (page 8-continued)
This should be apparent to anyone working in the Property and Casualty insurance industry. The Florida Homeowners insurance market provides a clear example of what may occur if rates are based on insufficient history. Prior to Hurricane Andrew in 1992, insurers failed to include an appropriate loading for hurricane losses in their rates due the lengthy period since the previous landfall along the Florida coast, leading to severe market disruptions that continue to the present day.

Response: *On the contrary, property rates prior to Hurricane Andrew were based on very long time series of observed hurricane experience, including very substantial losses from the 1960's. The historical experience, however, failed to reflect dramatic changes in the mix of property experience in the state, particularly huge growth in expensive coastal developments. Without appropriate restatement to reflect the change in the mix of business, even a very long time series of historical experience failed to capture the true risk. The experience of Hurricanes Hugo and Andrew led to significant changes in the methodology used to set hurricane rates, from one that was primarily based on historical experience to a fully prospective, engineering-based model. Properly reflecting actual prospective windstorm losses, not recoupment for past claims, has indeed resulted in very large increases in rates, especially for coastal properties, but there is no dispute that current rates are based on prospective costs.*

Comment (page 9)
The Property and Casualty industry differs from crop insurance in that P&C insurers have the ability to recoup underwriting losses reasonably quickly due to their control over their own rates, prices, and underwriting decisions. Crop insurers, on the other hand, have no ability to recoup their prior losses. Recognizing this, Congress established a requirement that rates be set at the anticipated level of losses rather than through the use of a recoupment mechanism. This helps to ensure that participating insurers are able to build adequate financial reserves in advance of a catastrophic year in recognition of their inability to recoup those losses in future years. This also serves to ensure the financial stability of participating insurers, particularly in light of the catastrophic nature of the risk. However, due its use of a 20 year experience period, the proposed methodology operates as a recoupment method rather than a prospective pricing method. This inability to set rates at the anticipated level of losses means that the methodology is fundamentally incompatible with the Congressional objective. As a result, participating insurers would be incapable of building adequate reserves in advance of a catastrophic year, which could threaten the financial stability of the delivery system.

Response: *This is a mischaracterization of the laws applying to property/casualty ratemaking. Every US jurisdiction has a law requiring that rates be neither excessive nor inadequate for the exposure being written. The proposed change in methodology leaves the full experience period intact for the CAT load. We note also that Crop Insurance legislation explicitly provides for a margin in the rates in excess of the prospective loss cost.*

Part 2 – Exclusion of significant costs from the rates

The Review cites A&O expense, surplus requirements and reinsurance expense as potential sources of rate inadequacy.

Response: *The scope of the study was limited to improvements in estimates of the loss cost portion of the rate, hence these considerations are outside the scope of the study.*

Part 3 – Impact on the Financial Soundness of the Program

Comment (page 11)
Promoting the financial soundness of the insurance system is a fundamental principle of the actuarial profession. In light of that, a decision to drastically modify rates for the two largest crops without correcting for deficiencies throughout the rest of the program will materially affect the overall financial soundness of the program. We find it troubling that large reductions are being proposed for immediate implementation while any increases are being delayed until a future date.

Response: *The proposed changes result in both increases and decreases to the base rates. The decision to cap changes as a part of the implementation, while potentially reasonable, is beyond the scope of the analysis.*

Technical Considerations

Comment (page 13)
The use of acre weighting is in direct conflict with the fundamental basis of the study itself. The study focuses on weather as a primary driver of the loss experience. Since weather is a systematic risk affecting large areas within a county simultaneously, the average loss cost based on a large insured area should be reasonably similar to the loss cost for a smaller area. The implication of the weather model is that the size of the area insured is unrelated to the credibility of the loss experience.

Response: *The reviewers suggest that, based on credibility considerations, net acreage should not be used in projecting future loss costs because, theoretically, acreage insured should not affect county experience. However, our analysis showed that the use of net acreage as a variable improved prediction over simple averages, and that other potential variables, such as time, were not better predictors. We concur with the continued use of spatial credibility procedures in the ratemaking process.*

We also note that, while county losses would be expected to be correlated, not all locations in a county would be expected to produce identical results. Inclusion of a wider sample of acreage in *the data would be expected to increase the accuracy of the estimate of the expected average loss cost for the entire county.*

Comment (page 15)
It is standard practice in econometric analysis or statistical estimation to properly specify the model being estimated and to assign a priori expectations to parameter estimation. Our review finds the technical report conspicuously lacking in this regard. Improper model specification will result in biased estimators, which in turn result in an over or under estimation of the predicted values, depending upon the nature of the bias.

Due to ill-defined specification, we cannot ascertain if the empirical output of the model is reliable. We believe our discussion in this section of our review more than adequately demonstrates our concerns.

Response: *The review argues that the modeling approach applied in the analysis does not follow "standard practice in econometrics." This is not the case. We are well aware of model specification and a priori expectations. One member of our team is a fellow of the Agricultural and Applied Economics Association, in large part due to his contributions in applied econometric analysis. The comment seems to suggest that a well defined structural model is necessary for all econometrics. That is simply not universally true. We must disagree with the characterization of the models used. Well accepted statistical tests were applied. In many instances hypotheses were tested, and out-of-sample forecasting competition is a well accepted procedure. Our analysis is based upon the most appropriate econometric approach for such fractional data. Further, the model specification is based upon an out-of-sample cross-validation evaluation. Appropriate criteria are used (e.g., the correlation between actual and predicted loss-costs) to determine whether sufficient evidence exists to merit the reweighting.*

Comment: (page 15) Regarding the use of CDD
The contractor uses Cooling Degree Days (CDD) variables and Palmer Draught Severity Index (PDSI) as the explanatory variables. However, there are significant issues arising from: (1) inadequate specification of the included variables; and (2) omitted variables…

Response: *We chose to use July-July total CDD (also known as Growing Degree Day base=65) as that variable was thought to best represent accumulated heat impact in the cornbelt. However, we also allowed for this to vary by region relative to RMA data on typical planting dates.*

The NOAA data used for this study do not allow for lower and upper bounds to be defined by users. The only CDD variable provided by the NOAA Climate Division Data is CDD Base=65. Because CDD values are accumulated over a month, CDD cannot be directly calculated from the other temperature variables provided in the NOAA Climate Division Data (such as average monthly maximum temperature and average monthly minimum temperature). Therefore, use of other thresholds, such as those described by Schlenker and Robert (2008) and Vado and Goodwin (2010) is not possible. We are familiar with other data sets such as used by Schlenker and Roberts, but recognize that they are either not available for *broad geographic regions or they limit the number of years available.*

We considered use of raw daily temperature and precipitation data, but this would have required RMA to appropriately aggregate and accurately calculate the drought indices and degree day, which would add substantial complexity as compared to relying on the widely used NOAA Climate Division Data.

We also considered using Heating Degree Days (HDD) at base 65, which is available in the NOAA Climate Division Data. HDD represent chilling accumulation, and could be added into the weather indexing process. However, as for CDD in the NOAA data set, there is no option to use limits other than base 65.

In the end, we were concerned with the use of too many variables and the potential for model over-fitting given the limited Loss-Cost data history. We believe that CDD is a better indicator than HDD of more typical loss during the growing season.

Econometric Model Specification:

Comment: Omitted Variables (page 17)
The contractor omits some important variables despite these variables being identified as significant in the literature. Omitted significant variables would make the predicted loss costs biased and the estimation may have large errors, therefore, the resulting bases rates could be biased, inefficient and less accurate.

Response: *We considered ways to incorporate other variables, including freeze and flooding impacts. However, both of these are very geographically limited and often do not impact whole counties (or even whole units). Flooding generally occurs only within the floodplain, which is a highly complex geography with impacts often on the sub-field scale. Similarly, many freeze events (especially those due to radiative cooling on calm, clear nights) are localized with impact variations from one side of a field to another. These events occur on spatial scales too fine to be captured with currently available long-term climate data, and are therefore extremely difficult to insert into a re-weighting process.*

Estimation of flood events would require substantial floodplain modeling and a long-term history of flood events based on USGS stream gauges. There are only a handful of USGS gauges which provide data beyond 100 years. The lack of available observations and the need for complex flood frequency mapping at field scales make the inclusion of flooding for re-weighting extremely difficult. However, our recommendation to use the positive side of PDSI allows for an appropriate measure of excessive moisture. While excessive moisture does not equate to flooding, it does provide a measure of local climate impacts from excessive precipitation.

Analysis of freeze risk would require the availability of climate data for each valley as freezes can be very localized. Indeed, observations of freezing temperatures are not available on this spatial scale. There are models that can be used to estimate freeze event extent and duration at *these spatial scales, but there is no standard model or routinely produced data that captures this. The authors agree that both freeze and flood are sources of risk for crop loss, but the lack of readily available and routinely produced climate data on events like these, that have impacts at small spatial scales, prevent their inclusion in a re-weighting process.*

Comment: (page 18)
"It appears that the contractor ignores the statistical insignificance of certain explanatory variables."

Response: *On the contrary, the specification is based on out-of-sample predictive power rather than in-sample fit or significance. In such applications, where over-fitting is a significant concern, this approach has advantages. The review notes other applications of out-of-sample model evaluations, an example of which uses 85% of the data to estimate and then predict the other 15% of the sample. Our approach is superior in that it reestimates the model while omitting each observation and predicting it based on all remaining observations. Such an approach is not sensitive to an ad-hoc division of the sample.*