

**Review of**  
**Methodology Analysis for Weighting of Historical Experience**  
**For**  
**Federal Crop Insurance Corporation Board of Directors**

**By**

**Alan Ker**  
**Professor**  
**Department of Agricultural and Resource Economics**  
**University of Arizona**  
**Tucson, Arizona**  
**&**  
**Professor and Chair**  
**Food, Agricultural, and Resource Economics**  
**University of Guelph**  
**Guelph, Ontario**

**Oct 17, 2011**

Personal Biography of  
Alan P. Ker

I am a Professor in the Department of Agricultural and Resource Economics at the University of Arizona, Tucson, Arizona and Chair and Professor in the Department of Food, Agricultural, and Resource Economics at the University of Guelph. I have a B.A. in Economics from the University of Waterloo, a M.Sc. in Agricultural Economics from the University of Guelph, and a joint PhD in Economics and Statistics from North Carolina State University. I received my doctorate in 1996 under the supervision of Barry Goodwin and Ron Gallant. In the academic 2006/07 year I was a visiting Professor in the Department of Statistics and Actuarial Science at the University of Waterloo.

Crop insurance has been an integral part of my research program both as a graduate student at North Carolina State University and as a Professor at the University of Arizona and University of Guelph. I have published numerous refereed journal articles on crop insurance over the past four years. In addition, I have received in excess of \$500,000 to support my research on crop insurance.

I have been a member of the Research Advisory Committee of the RMA when it was in existence and participated regularly at the research meetings. I have given presentations on numerous rating aspects such as detrending yields, Bayesian methods, appropriateness of the Beta distribution, nonparametric methods, etc. I have been intimately involved in the review of revenue rating methodologies and trend estimation methodologies for FCIC. Over the past five years, I have written reports on the rating methodologies of RA, RA-HPO, CRC, and IP. Finally, I have conducted numerous Board reviews.

## Description of Methodology Used

As with all submissions, the approach I undertake to review the packet is the following.

- 1) Review the submission packet.
- 2) Identify areas of potential concern.
- 3) Conduct preliminary analyses on issues of concern.
- 4) Get additional information if required.
- 5) Conduct in-depth analysis of issues that remain points of concern. *Many times there are issues that after a theoretical proof or simulation study I feel were handled appropriately by the developers. These issues are not pointed out in my review.*
- 6) Iterate steps 1-5 as necessary and time permits.
- 7) Write report.

## Executive Summary

Currently, the basis for the yield and combo rates are the historical loss cost experience which is equally weighted through time. That is, last's year historical loss experience is weighted equally with loss experience from past years. While on the one hand this may seem inappropriate given all the changes including technology and program participation, on the other hand weather plays a significant role in crop yields and many realizations are required to form a reliable basis for measuring potential losses. The authors evaluate a number of potential alternatives for weighing historical loss experience in attempts to improve the accuracy of the estimated rates. This is an exceedingly difficult and complex task.

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.

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.

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 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 more accurate rates. My simulations suggest trivial improvements in accuracy which begs the question is this sufficient for what is a very significant change in complexity.

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.

In summary, I would recommend that the Board not adopt the weather weighting scheme in constructing base or catastrophic rates. If the Board does adopt this I would urge that the rates be reproduced (I was not given the data to reproduce the results). Conversely, I am comfortable if the Board adopted the suggestion that base rates be constructed from only the past 20 years with some appropriate accounting for acreage given the significant changes in participation and technology. I feel a necessary but not sufficient condition for adopting a different methodology is that one shows that the alternative is a statistically significant improvement in accuracy. I do not believe the authors showed this for any of their recommendations. The fact that weather significantly influences loss cost (as shown by their models) does not suggest that constructed weights while assuming a one-to-one relationship between weather and loss cost and probability bins will lead to statistically significant improvement in accuracy.

**Review of  
“”**

**Main Body**

The review directly follows the questions outlined in the task order.

**1) Actuarial Soundness**

*A) (i) Is adequate, credible, and reliable rate-making data available?*

The data used are the RMA databases plus the Climate Division Data which comes from the National Climatic Data Center’s Time Bias Corrected Divisional Temperature-Precipitation-Drought Index data. I am not a climatologist but I did discuss the data with a couple of climatologist who informed that the data is credible and reliable. I also believe that the approach the authors used to determine which variables should belong in the model is appropriate as well.

*A) ii) Is it likely that the data will continue to be available?*

Yes, there is no reason to suspect that the data will cease to be available in the future.

*A) iii) Is the data vulnerable to tampering if the proposed policy is approved?*

No, there is no reason to believe that the data is vulnerable to tampering if the proposed policy is approved.

*B) Are the explicit and implicit assumptions used in the rating process reasonable?*

The authors were very good about outlining their assumptions. They have made six explicit assumptions and I will take each in turn.

*(1) Using Fractional Logit Model to Map Weather into a Single Index*

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. As the authors pointed out this is the most appropriate approach but alternative and simpler approaches will have very little effect if any effect on the ordinal ranking for creating the weather bins. I am also comfortable about mapping the weather variables into a single index which is necessary to group into a single dimensional probability bins. I am also comfortable with their variable choosing method and restrictions at less aggregated regions. The authors then map from the probabilities of the weather bins to alter the probabilities of the realized loss costs however these random variables have a joint distribution and thus one can not directly map from weather to lost costs (an implicit assumption in the weather weighted loss costs) as the authors do. That is, there is error in their estimated equation.

*(2) Using Weather Probability Bins*

I am not comfortable with the use of weather probability bins - the resulting estimate of loss costs can change dramatically as the weather bins change. The main reason is that the ratio of observations to bins is very low. In addition, the decision as to the number of bins seems somewhat arbitrary. I undertook a simulation where I simulated from a joint distribution of a weather index and loss costs. The sample of size 125 was randomly drawn from the joint distribution. The weather index sample was used to delineate the quantiles as dictated by the number of bins assumed (5 to 15). The loss cost sample was assumed to be 20, 25, 30 or 35. The loss cost for a single simulation was calculated using the methodology of weather bins for each scenario for coverage levels 50%, 75%, and 85%.

Table: Lost Cost for 50% Coverage Level

Number of Bins	Sample Size of Loss Cost			
	20	25	30	35
5	100.00%	204.33%	198.98%	110.61%
6	192.94%	209.80%	147.43%	114.09%
7	233.23%	190.36%	136.49%	181.86%
8	218.44%	187.31%	167.60%	203.16%
9	176.11%	156.66%	180.20%	184.06%

10	165.65%	113.61%	106.18%	137.86%
11	130.83%	138.42%	113.78%	144.11%
12	124.58%	246.77%	234.76%	194.07%
13	149.82%	134.60%	244.39%	175.55%
14	122.96%	149.12%	153.44%	104.93%
15	121.70%	145.68%	271.12%	134.79%

All loss costs are standardized by the loss cost estimate for 5 bins and sample size 20. The result are surprising in that the difference between 5 and 6 bins for the same set of data changes the loss cost by close to 100%. The variability is very high across the bin number selection. These results are consistent at higher coverage levels as well as illustrated below.

Table: Lost Cost for 75% Coverage Level

Number of Bins	Sample Size of Loss Cost			
	20	25	30	35
5	100.00%	213.94%	206.43%	120.49%
6	144.62%	239.52%	172.12%	123.01%
7	207.92%	204.99%	164.62%	159.32%
8	179.59%	177.00%	177.77%	184.44%
9	191.47%	157.78%	196.15%	160.53%
10	144.58%	139.15%	126.37%	143.54%
11	126.40%	133.08%	128.81%	125.06%
12	124.50%	181.56%	190.33%	180.94%
13	162.66%	110.88%	218.24%	156.81%
14	109.21%	184.86%	137.10%	102.58%
15	146.50%	179.59%	195.59%	142.71%

Table: Lost Cost for 85% Coverage Level

Number of Bins	Sample Size of Loss Cost			
	20	25	30	35
5	100.00%	217.77%	208.74%	123.58%

6	146.17%	242.13%	173.85%	123.45%
7	210.23%	208.29%	166.82%	159.25%
8	181.57%	180.19%	180.32%	184.79%
9	183.93%	160.55%	186.35%	162.11%
10	145.97%	142.89%	135.08%	145.29%
11	129.74%	133.63%	131.00%	128.52%
12	124.26%	163.23%	169.14%	162.62%
13	147.43%	111.92%	185.29%	143.51%
14	107.63%	159.96%	127.25%	103.65%
15	131.53%	154.14%	164.84%	128.53%

It is clear that the rates are highly variable across the choice of the number of bins. The code for this simulation is in the appendix.

*(C) (i) Are the technical analyses technically correct?*

I have no reason to suspect that they are not correct. The team is first class in terms of their abilities and attention to detail. I would have wished that the raw data was provided to reproduce their results in the tables. While the tables were informative to illustrate the process, the underlying data was not accompanying my packet so that the tables could not be reproduced.

*(C) (ii) Do they provide credible, relevant results?*

I feel it is a necessary but not sufficient condition to show that there is a statistically significant improvement in the accuracy of the rates for RMA to adopt the proposed methodology. This was not shown and needs to be shown.

*(D) Is the data used for the analyses appropriate, reliable, and the best available?*

As stated earlier, yes the data used for the analyses appears to be appropriate, reliable, and the best available.

*(E) Does the experience from prior years and relevant crops and areas support the validity of the proposed rates?*

I was not given the old and new base rates per crop and county with my packet. It would be a worthwhile exercise to match up the proposed rate changes with county-crop loss ratios and determine if in those combinations where the proposed rates are significantly higher than the current rates, loss ratios are significantly lower and of course vice-versa. Similarly, it would be interesting to match up the rate changes with private insurance companies fund allocations to see if the proposed rate changes are directionally consistent

with fund allocations. I would suggest that this be done with both recommendations regarding the use of weather weighting and the truncation of the loss costs more than 20 years in the past.

*(F) Are the proposed premium rates likely to cover anticipated losses and a reasonable reserve?*

I believe so in the aggregate but I am less comfortable with this statement given the uncertainty or standard error surrounding these rates as well as the proposed rates. I am significantly more comfortable with this statement regarding the truncation of historical loss costs at twenty years.

*(J) Is the actuarial method appropriate for the insured risks?*

As previously stated, I believe it is a necessary but not sufficient condition that for the methodology to be adopted one has to show that there is a statistically significant improvement in the accuracy of the resulting rates. The authors have not shown this so while I believe it may be appropriate, it also may not be appropriate. Also, given the complexity of the weather weighting process, I am not convinced that unless there is a statistically significant improvement in rating accuracy that leads to significant economic implications (e.g. SRA) that the additional complexity of weather weighted historical loss costs be incorporated. On the other had, given the simplistic nature of truncating historical loss costs and the strong economic and actuarial rationale, a statistically significant improvement in rating accuracy would suggest that this would be appropriate for the insured risks.

I undertook a simulation where I sampled from a joint distribution of a weather index and loss cost at the 50%, 75%, and 85% coverage levels. The correlation coefficients were -0.854, -0.877, and -0.878 respectively for the three loss costs with the weather index so there is a very significant relationship between the two random variables. I then took samples of size 125 from the joint distribution and used 25 corresponding to the loss cost distribution and estimated the loss costs and compared them to the true rate. I then used the 125 weather index realizations to create 10 bins of equal probability and assigned each of the 25 loss costs to a bin. I then proceeded to estimate the loss cost by first averaging the loss costs within a bin and then averaging the averages of the loss costs across the bins as the authors are recommending. I then calculated the absolute value of the error for both approaches over 10000 simulations and conducted a paired t-test of the differences in the absolute value of the rating error. There were no statistically significant difference. The p-values were 0.248, 0.250, and 0.249 for the 50%, 75%, and 100% coverage level respectively indicating the null that there is no statistically significant change in the rate accuracy is found using weather weighting. The code for this is in the appendix. I also calculated the difference in the mean squared errors for both approaches and it was trivial.

**5) Questions Specific to this Review**

*None*

## Appendix: R Code written for the analysis

binning\_sim.r

```
library(MASS)
q<-100
cor<-0
dq<-rep(1,q)
dq1<-rep(1,q+1)*cor*ws.1*s.1
dq1[1]<-ws.1^2
mu1<-c(wm.1,m.1*dq)
var.temp1<-matrix(0,q,q)
diag(var.temp1)<-s.1^2
var1<-rbind(dq1[2:(q+1)],var.temp1)
var1<-cbind(t(t(dq1)),var1)
dq2<-rep(1,q+1)*cor*ws.2*s.2
dq2[1]<-ws.2^2
mu2<-c(wm.2,m.2*dq)
var.temp2<-matrix(0,q,q)
diag(var.temp2)<-s.2^2
var2<-rbind(dq2[2:(q+1)],var.temp2)
var2<-cbind(t(t(dq2)),var2)

wsn<-125
ysn<-c(20,25,30,25)
bin_n<-seq(5,15)
d.10<-runif(wsn)
sdata.1<-mvrnorm(wsn,mu1,var1)
sdata.2<-mvrnorm(wsn,mu2,var2)
sdata.all<-rbind(sdata.1[d.10<mix,],sdata.2[d.10>=mix,])
wdata<-sdata.all[,1]
lc50j<-array(0,dim<-c(wsn,1))
lc75j<-array(0,dim<-c(wsn,1))
lc85j<-array(0,dim<-c(wsn,1))
y.data.all<-sdata.all[,2:(q+1)]
for (j in 1:wsn) {
  lc50j[j]<-mean(pmax(0,.50*my-y.data.all[j,]))/(.50*my)
  lc75j[j]<-mean(pmax(0,.75*my-y.data.all[j,]))/(.75*my)
  lc85j[j]<-mean(pmax(0,.85*my-y.data.all[j,]))/(.85*my)
}
d.50<-array(0,dim<-c(11,4))
d.75<-array(0,dim<-c(11,4))
d.85<-array(0,dim<-c(11,4))

for (j in 1:11){
  bin_n<-4+j
  eq<-array(0,dim<-c(bin_n,1))
  for (i in 1:bin_n) {eq[i]<-quantile(wdata,i/bin_n)}
  for (k in 1:4) {
    temp<-sample(seq(1,wsn),ysn[k])
    d50<-lc50j[temp]
    d75<-lc75j[temp]
    d85<-lc85j[temp]
```

```

ww.data<-wdata[temp]
b<-0
  for (h in 1:bin_n) {
    if (h==1 && length(d50[ww.data<=eq[1]])>0) b50<-
mean(d50[ww.data<=eq[1]])
    if (h>1 && length(d50[ww.data>=eq[h-1] &
ww.data<=eq[h]])>0) b50<-b50+mean(d50[ww.data>=eq[h-1] &
ww.data<=eq[h]])
    if (h==1 && length(d75[ww.data<=eq[1]])>0) b75<-
mean(d75[ww.data<=eq[1]])
    if (h>1 && length(d75[ww.data>=eq[h-1] &
ww.data<=eq[h]])>0) b75<-b75+mean(d75[ww.data>=eq[h-1] &
ww.data<=eq[h]])
    if (h==1 && length(d85[ww.data<=eq[1]])>0) b85<-
mean(d85[ww.data<=eq[1]])
    if (h>1 && length(d85[ww.data>=eq[h-1] &
ww.data<=eq[h]])>0) b85<-b85+mean(d75[ww.data>=eq[h-1] &
ww.data<=eq[h]])
  }
  d.50[j,k]<-b50/bin_n
  d.75[j,k]<-b75/bin_n
  d.85[j,k]<-b85/bin_n
}

}

d.50<-cbind(seq(5,15),d.50/d.50[1,1])
d.75<-cbind(seq(5,15),d.75/d.75[1,1])
d.85<-cbind(seq(5,15),d.85/d.85[1,1])

write(t(d.50),file="results_d50_binnin",ncolumns=5)
write(t(d.75),file="results_d75_binnin",ncolumns=5)
write(t(d.85),file="results_d85_binnin",ncolumns=5)

```

weather1\_lc.r

```

library(MASS)

supp<-seq(0,500)
mix<-.15
m.1<-100
s.1<-15
m.2<-175
s.2<-25
y.pdf<-mix*dnorm(supp,m.1,s.1)+(1-mix)*dnorm(supp,m.2,s.2)
plot(supp,y.pdf)
sn<-100000
d.1<-runif(sn)
d.2<-rnorm(sn,m.1,s.1)
d.3<-rnorm(sn,m.2,s.2)

```

```

y.samp<-c(d.2[d.1<mix],d.3[d.1>=mix])
my<-mean(y.samp)
r75<-(sum((pmax(0,.75*my-y.samp)))/sn)/(.75*my)
r85<-(sum((pmax(0,.85*my-y.samp)))/sn)/(.85*my)
lc75<-pmax(0,.75*my-y.samp)/(.75*my)
lc85<-pmax(0,.85*my-y.samp)/(.85*my)

# get mse from only the yield data

ysn<-25
lc75<-array(0,dim<-c(sn,1))
lc85<-array(0,dim<-c(sn,1))
for (j in 1:sn) {
  d.4<-runif(ysn)
  d.5<-rnorm(ysn,m.1,s.1)
  d.6<-rnorm(ysn,m.2,s.2)
  y.samp<-c(d.5[d.4<mix],d.6[d.4>=mix])
  my<-mean(y.samp)
  lc75[j]<-(sum((pmax(0,.75*my-y.samp)))/ysn)/(.75*my)
  lc85[j]<-(sum((pmax(0,.85*my-y.samp)))/ysn)/(.85*my)
}
m.rs75<-mean(rs75)
m.rs85<-mean(rs85)
mse.75<-sum((rs75-r75)^2)/sn
mse.85<-sum((rs85-r85)^2)/sn

# weather bins
mix<-.15
wm.1<-10
ws.1<-1.5
wm.2<-20
ws.2<-2
w.pdf<-mix*dnorm(supp,m.1,s.1)+(1-mix)*dnorm(supp,m.2,s.2)
plot(supp,y.pdf)
#sn<-10000
d.7<-runif(sn)
d.8<-rnorm(sn,wm.1,ws.1)
d.9<-rnorm(sn,wm.2,ws.2)
w.samp<-c(d.8[d.7<mix],d.9[d.7>=mix])
q.1<-quantile(w.samp,.1)
q.2<-quantile(w.samp,.2)
q.3<-quantile(w.samp,.3)
q.4<-quantile(w.samp,.4)
q.5<-quantile(w.samp,.5)
q.6<-quantile(w.samp,.6)
q.7<-quantile(w.samp,.7)
q.8<-quantile(w.samp,.8)
q.9<-quantile(w.samp,.9)

# joint sample of yields and weather
#sn<-10000

```

```

cor<-.15
mu1<-c(wm.1,m.1)
var1<-rbind(c(ws.1^2,cor*ws.1*s.1),c(cor*s.1*ws.1,s.1^2))
mu2<-c(wm.2,m.2)
var2<-rbind(c(ws.2^2,cor*ws.2*s.2),c(cor*s.2*ws.2,s.2^2))
data.1<-mvrnorm(sn,mu1,var1)
data.2<-mvrnorm(sn,mu2,var2)
d.9<-runif(sn)
data.all<-rbind(data.1[d.9<mix,],data.2[d.9>=mix,])
b.my<-mean(data.all[,2])
b.r75<-(sum((pmax(0,.75*b.my-data.all[,2])))/sn)/(.75*b.my)
b.r85<-(sum((pmax(0,.85*b.my-data.all[,2])))/sn)/(.85*b.my)
w.samp<-data.all[,1]
q.1<-quantile(w.samp,.1)
q.2<-quantile(w.samp,.2)
q.3<-quantile(w.samp,.3)
q.4<-quantile(w.samp,.4)
q.5<-quantile(w.samp,.5)
q.6<-quantile(w.samp,.6)
q.7<-quantile(w.samp,.7)
q.8<-quantile(w.samp,.8)
q.9<-quantile(w.samp,.9)

```

```

# using the weather distribution
#sn<-10000
ysn<-25
wsn<-100
w.rs75<-array(0,dim<-c(sn,1))
w.rs85<-array(0,dim<-c(sn,1))
el75<-array(0,dim<-c(10,1))
el85<-array(0,dim<-c(10,1))

```

```

for (j in 1:sn) {
  d.10<-runif(wsn)
  sdata.1<-mvrnorm(wsn,mu1,var1)
  sdata.2<-mvrnorm(wsn,mu2,var2)
  sdata.all<-rbind(sdata.1[d.10<mix,],sdata.2[d.10>=mix,])
  ydata<-sdata.all[sample(seq(1,wsn),ysn),]
  mydata<-mean(ydata[,2])
  b1<-ydata[ydata[,1]<q.1,2]
  b2<-ydata[ydata[,1]>=q.1 & ydata[,1]<q.2,2]
  b3<-ydata[ydata[,1]>=q.2 & ydata[,1]<q.3,2]
  b4<-ydata[ydata[,1]>=q.3 & ydata[,1]<q.4,2]
  b5<-ydata[ydata[,1]>=q.4 & ydata[,1]<q.5,2]
  b6<-ydata[ydata[,1]>=q.5 & ydata[,1]<q.6,2]
  b7<-ydata[ydata[,1]>=q.6 & ydata[,1]<q.7,2]
  b8<-ydata[ydata[,1]>=q.7 & ydata[,1]<q.8,2]
  b9<-ydata[ydata[,1]>=q.8 & ydata[,1]<q.9,2]
  b10<-ydata[ydata[,1]>=q.9,2]
}

```

```

    if (sum(b1)>0) e175[1]<-mean(pmax(0,.75*mydata-b1)) else
e175[1]=0
    if (sum(b2)>0) e175[2]<-mean(pmax(0,.75*mydata-b2)) else
e175[2]=0
    if (sum(b3)>0) e175[3]<-mean(pmax(0,.75*mydata-b3)) else
e175[3]=0
    if (sum(b4)>0) e175[4]<-mean(pmax(0,.75*mydata-b4)) else
e175[4]=0
    if (sum(b5)>0) e175[5]<-mean(pmax(0,.75*mydata-b5)) else
e175[5]=0
    if (sum(b6)>0) e175[6]<-mean(pmax(0,.75*mydata-b6)) else
e175[6]=0
    if (sum(b7)>0) e175[7]<-mean(pmax(0,.75*mydata-b7)) else
e175[7]=0
    if (sum(b8)>0) e175[8]<-mean(pmax(0,.75*mydata-b8)) else
e175[8]=0
    if (sum(b9)>0) e175[9]<-mean(pmax(0,.75*mydata-b9)) else
e175[9]=0
    if (sum(b10)>0) e175[10]<-mean(pmax(0,.75*mydata-b10)) else
e175[10]=0
    w.rs75[j]<-mean(e175)/(.75*mydata)
    if (sum(b1)>0) e185[1]<-mean(pmax(0,.85*mydata-b1)) else
e185[1]=0
    if (sum(b2)>0) e185[2]<-mean(pmax(0,.85*mydata-b2)) else
e185[2]=0
    if (sum(b3)>0) e185[3]<-mean(pmax(0,.85*mydata-b3)) else
e185[3]=0
    if (sum(b4)>0) e185[4]<-mean(pmax(0,.85*mydata-b4)) else
e185[4]=0
    if (sum(b5)>0) e185[5]<-mean(pmax(0,.85*mydata-b5)) else
e185[5]=0
    if (sum(b6)>0) e185[6]<-mean(pmax(0,.85*mydata-b6)) else
e185[6]=0
    if (sum(b7)>0) e185[7]<-mean(pmax(0,.85*mydata-b7)) else
e185[7]=0
    if (sum(b8)>0) e185[8]<-mean(pmax(0,.85*mydata-b8)) else
e185[8]=0
    if (sum(b9)>0) e185[9]<-mean(pmax(0,.85*mydata-b9)) else
e185[9]=0
    if (sum(b10)>0) e185[10]<-mean(pmax(0,.85*mydata-b10)) else
e185[10]=0
    w.rs85[j]<-mean(e185)/(.85*mydata)
}
wm.rs75<-mean(w.rs75)
wm.rs85<-mean(w.rs85)
wmse.75<-sum((w.rs75-r75)^2)/sn
wmse.85<-sum((w.rs85-r85)^2)/sn

#using estimate of weather distribution
#sn<-10000

```

```

ysn<-25
wsn<-100
ew.rs75<-array(0,dim<-c(sn,1))
ew.rs85<-array(0,dim<-c(sn,1))
eel75<-array(0,dim<-c(10,1))
eel85<-array(0,dim<-c(10,1))

for (j in 1:sn) {
  d.11<-runif(wsn)
  sdata.1<-mvrnorm(wsn,mu1,var1)
  sdata.2<-mvrnorm(wsn,mu2,var2)
  sdata.all<-rbind(sdata.1[d.10<mix,],sdata.2[d.10>=mix,])
  wdata<-sdata.all[,1]
  eq.1<-quantile(wdata,.1)
  eq.2<-quantile(wdata,.2)
  eq.3<-quantile(wdata,.3)
  eq.4<-quantile(wdata,.4)
  eq.5<-quantile(wdata,.5)
  eq.6<-quantile(wdata,.6)
  eq.7<-quantile(wdata,.7)
  eq.8<-quantile(wdata,.8)
  eq.9<-quantile(wdata,.9)
  eq.10<-quantile(wdata,.10)
  ydata<-sdata.all[sample(seq(1,wsn),ysn),]
  mydata<-mean(ydata[,2])
  b1<-ydata[ydata[,1]<eq.1,2]
  b2<-ydata[ydata[,1]>=eq.1 & ydata[,1]<q.2,2]
  b3<-ydata[ydata[,1]>=eq.2 & ydata[,1]<q.3,2]
  b4<-ydata[ydata[,1]>=eq.3 & ydata[,1]<q.4,2]
  b5<-ydata[ydata[,1]>=eq.4 & ydata[,1]<q.5,2]
  b6<-ydata[ydata[,1]>=eq.5 & ydata[,1]<q.6,2]
  b7<-ydata[ydata[,1]>=eq.6 & ydata[,1]<q.7,2]
  b8<-ydata[ydata[,1]>=eq.7 & ydata[,1]<q.8,2]
  b9<-ydata[ydata[,1]>=eq.8 & ydata[,1]<q.9,2]
  b10<-ydata[ydata[,1]>=eq.9,2]
  if (sum(b1)>0) eel75[1]<-mean(pmax(0,.75*mydata-b1)) else
eel75[1]=0
  if (sum(b2)>0) eel75[2]<-mean(pmax(0,.75*mydata-b2)) else
eel75[2]=0
  if (sum(b3)>0) eel75[3]<-mean(pmax(0,.75*mydata-b3)) else
eel75[3]=0
  if (sum(b4)>0) eel75[4]<-mean(pmax(0,.75*mydata-b4)) else
eel75[4]=0
  if (sum(b5)>0) eel75[5]<-mean(pmax(0,.75*mydata-b5)) else
eel75[5]=0
  if (sum(b6)>0) eel75[6]<-mean(pmax(0,.75*mydata-b6)) else
eel75[6]=0
  if (sum(b7)>0) eel75[7]<-mean(pmax(0,.75*mydata-b7)) else
eel75[7]=0
  if (sum(b8)>0) eel75[8]<-mean(pmax(0,.75*mydata-b8)) else
eel75[8]=0
}

```

```

    if (sum(b9)>0) eel75[9]<-mean(pmax(0,.75*mydata-b9)) else
eel75[9]=0
    if (sum(b10)>0) eel75[10]<-mean(pmax(0,.75*mydata-b10)) else
eel75[10]=0
    ew.rs75[j]<-mean(eel75)/(.75*mydata)
    if (sum(b1)>0) eel85[1]<-mean(pmax(0,.85*mydata-b1)) else
eel85[1]=0
    if (sum(b2)>0) eel85[2]<-mean(pmax(0,.85*mydata-b2)) else
eel85[2]=0
    if (sum(b3)>0) eel85[3]<-mean(pmax(0,.85*mydata-b3)) else
eel85[3]=0
    if (sum(b4)>0) eel85[4]<-mean(pmax(0,.85*mydata-b4)) else
eel85[4]=0
    if (sum(b5)>0) eel85[5]<-mean(pmax(0,.85*mydata-b5)) else
eel85[5]=0
    if (sum(b6)>0) eel85[6]<-mean(pmax(0,.85*mydata-b6)) else
eel85[6]=0
    if (sum(b7)>0) eel85[7]<-mean(pmax(0,.85*mydata-b7)) else
eel85[7]=0
    if (sum(b8)>0) eel85[8]<-mean(pmax(0,.85*mydata-b8)) else
eel85[8]=0
    if (sum(b9)>0) eel85[9]<-mean(pmax(0,.85*mydata-b9)) else
eel85[9]=0
    if (sum(b10)>0) eel85[10]<-mean(pmax(0,.85*mydata-b10)) else
eel85[10]=0
    ew.rs85[j]<-mean(eel85)/(.85*mydata)
}
ewm.rs75<-mean(ew.rs75)
ewm.rs85<-mean(ew.rs85)
ewmse.75<-sum((ew.rs75-r75)^2)/sn
ewmse.85<-sum((ew.rs85-r85)^2)/sn

results<-array(0,dim<-c(7,2))
results[1,1]<-r75
results[1,2]<-r85
results[2,1]<-m.rs75
results[2,2]<-m.rs85
results[3,1]<-mse.75
results[3,2]<-mse.85
results[4,1]<-wm.rs75
results[4,2]<-wm.rs85
results[5,1]<-wmse.75
results[5,2]<-wmse.85
results[6,1]<-ewm.rs75
results[6,2]<-ewm.rs85
results[7,1]<-ewmse.75
results[7,2]<-ewmse.85
results

```

```

weather1_yield

library(MASS)

supp<-seq(0,500)
mix<-.15
m.1<-100
s.1<-15
m.2<-175
s.2<-25
y.pdf<-mix*dnorm(supp,m.1,s.1)+(1-mix)*dnorm(supp,m.2,s.2)
plot(supp,y.pdf)
sn<-100000
d.1<-runif(sn)
d.2<-rnorm(sn,m.1,s.1)
d.3<-rnorm(sn,m.2,s.2)
y.samp<-c(d.2[d.1<mix],d.3[d.1>=mix])
my<-mean(y.samp)
r75<-(sum((pmax(0,.75*my-y.samp)))/sn)/(.75*my)
r85<-(sum((pmax(0,.85*my-y.samp)))/sn)/(.85*my)

# get mse from only the yield data

ysn<-25
rs75<-array(0,dim<-c(sn,1))
rs85<-array(0,dim<-c(sn,1))
for (j in 1:sn) {
  d.4<-runif(ysn)
  d.5<-rnorm(ysn,m.1,s.1)
  d.6<-rnorm(ysn,m.2,s.2)
  y.samp<-c(d.5[d.4<mix],d.6[d.4>=mix])
  my<-mean(y.samp)
  rs75[j]<-(sum((pmax(0,.75*my-y.samp)))/ysn)/(.75*my)
  rs85[j]<-(sum((pmax(0,.85*my-y.samp)))/ysn)/(.85*my)
}
m.rs75<-mean(rs75)
m.rs85<-mean(rs85)
mse.75<-sum((rs75-r75)^2)/sn
mse.85<-sum((rs85-r85)^2)/sn

# weather bins
mix<-.15
wm.1<-10
ws.1<-1.5
wm.2<-20
ws.2<-2
w.pdf<-mix*dnorm(supp,m.1,s.1)+(1-mix)*dnorm(supp,m.2,s.2)
plot(supp,y.pdf)
#sn<-10000
d.7<-runif(sn)
d.8<-rnorm(sn,wm.1,ws.1)
d.9<-rnorm(sn,wm.2,ws.2)

```

```

w.samp<-c(d.8[d.7<mix],d.9[d.7>=mix])
q.1<-quantile(w.samp,.1)
q.2<-quantile(w.samp,.2)
q.3<-quantile(w.samp,.3)
q.4<-quantile(w.samp,.4)
q.5<-quantile(w.samp,.5)
q.6<-quantile(w.samp,.6)
q.7<-quantile(w.samp,.7)
q.8<-quantile(w.samp,.8)
q.9<-quantile(w.samp,.9)

# joint sample of yields and weather
#sn<-10000
cor<-.15
mu1<-c(wm.1,m.1)
var1<-rbind(c(ws.1^2,cor*ws.1*s.1),c(cor*s.1*ws.1,s.1^2))
mu2<-c(wm.2,m.2)
var2<-rbind(c(ws.2^2,cor*ws.2*s.2),c(cor*s.2*ws.2,s.2^2))
data.1<-mvrnorm(sn,mu1,var1)
data.2<-mvrnorm(sn,mu2,var2)
d.9<-runif(sn)
data.all<-rbind(data.1[d.9<mix,],data.2[d.9>=mix,])
b.my<-mean(data.all[,2])
b.r75<-(sum((pmax(0,.75*b.my-data.all[,2])))/sn)/(.75*b.my)
b.r85<-(sum((pmax(0,.85*b.my-data.all[,2])))/sn)/(.85*b.my)
w.samp<-data.all[,1]
q.1<-quantile(w.samp,.1)
q.2<-quantile(w.samp,.2)
q.3<-quantile(w.samp,.3)
q.4<-quantile(w.samp,.4)
q.5<-quantile(w.samp,.5)
q.6<-quantile(w.samp,.6)
q.7<-quantile(w.samp,.7)
q.8<-quantile(w.samp,.8)
q.9<-quantile(w.samp,.9)

# using the weather distribution
#sn<-10000
ysn<-25
wsn<-100
w.rs75<-array(0,dim<-c(sn,1))
w.rs85<-array(0,dim<-c(sn,1))
el75<-array(0,dim<-c(10,1))
el85<-array(0,dim<-c(10,1))

for (j in 1:sn) {
  d.10<-runif(wsn)
  sdata.1<-mvrnorm(wsn,mu1,var1)
  sdata.2<-mvrnorm(wsn,mu2,var2)
}

```

```

sdata.all<-rbind(sdata.1[d.10<mix,],sdata.2[d.10>=mix,])
ydata<-sdata.all[sample(seq(1,wsn),ysn),]
mydata<-mean(ydata[,2])
b1<-ydata[ydata[,1]<q.1,2]
b2<-ydata[ydata[,1]>=q.1 & ydata[,1]<q.2,2]
b3<-ydata[ydata[,1]>=q.2 & ydata[,1]<q.3,2]
b4<-ydata[ydata[,1]>=q.3 & ydata[,1]<q.4,2]
b5<-ydata[ydata[,1]>=q.4 & ydata[,1]<q.5,2]
b6<-ydata[ydata[,1]>=q.5 & ydata[,1]<q.6,2]
b7<-ydata[ydata[,1]>=q.6 & ydata[,1]<q.7,2]
b8<-ydata[ydata[,1]>=q.7 & ydata[,1]<q.8,2]
b9<-ydata[ydata[,1]>=q.8 & ydata[,1]<q.9,2]
b10<-ydata[ydata[,1]>=q.9,2]
  if (sum(b1)>0) e175[1]<-mean(pmax(0,.75*mydata-b1)) else
e175[1]=0
  if (sum(b2)>0) e175[2]<-mean(pmax(0,.75*mydata-b2)) else
e175[2]=0
  if (sum(b3)>0) e175[3]<-mean(pmax(0,.75*mydata-b3)) else
e175[3]=0
  if (sum(b4)>0) e175[4]<-mean(pmax(0,.75*mydata-b4)) else
e175[4]=0
  if (sum(b5)>0) e175[5]<-mean(pmax(0,.75*mydata-b5)) else
e175[5]=0
  if (sum(b6)>0) e175[6]<-mean(pmax(0,.75*mydata-b6)) else
e175[6]=0
  if (sum(b7)>0) e175[7]<-mean(pmax(0,.75*mydata-b7)) else
e175[7]=0
  if (sum(b8)>0) e175[8]<-mean(pmax(0,.75*mydata-b8)) else
e175[8]=0
  if (sum(b9)>0) e175[9]<-mean(pmax(0,.75*mydata-b9)) else
e175[9]=0
  if (sum(b10)>0) e175[10]<-mean(pmax(0,.75*mydata-b10)) else
e175[10]=0
  w.rs75[j]<-mean(e175)/(.75*mydata)
  if (sum(b1)>0) e185[1]<-mean(pmax(0,.85*mydata-b1)) else
e185[1]=0
  if (sum(b2)>0) e185[2]<-mean(pmax(0,.85*mydata-b2)) else
e185[2]=0
  if (sum(b3)>0) e185[3]<-mean(pmax(0,.85*mydata-b3)) else
e185[3]=0
  if (sum(b4)>0) e185[4]<-mean(pmax(0,.85*mydata-b4)) else
e185[4]=0
  if (sum(b5)>0) e185[5]<-mean(pmax(0,.85*mydata-b5)) else
e185[5]=0
  if (sum(b6)>0) e185[6]<-mean(pmax(0,.85*mydata-b6)) else
e185[6]=0
  if (sum(b7)>0) e185[7]<-mean(pmax(0,.85*mydata-b7)) else
e185[7]=0
  if (sum(b8)>0) e185[8]<-mean(pmax(0,.85*mydata-b8)) else
e185[8]=0
  if (sum(b9)>0) e185[9]<-mean(pmax(0,.85*mydata-b9)) else
e185[9]=0

```

```

    if (sum(b10)>0) el85[10]<-mean(pmax(0,.85*mydata-b10)) else
el85[10]=0
    w.rs85[j]<-mean(el85)/(.85*mydata)
}
wm.rs75<-mean(w.rs75)
wm.rs85<-mean(w.rs85)
wmse.75<-sum((w.rs75-r75)^2)/sn
wmse.85<-sum((w.rs85-r85)^2)/sn

#using estimate of weather distribution
#sn<-10000
ysn<-25
wsn<-100
ew.rs75<-array(0,dim<-c(sn,1))
ew.rs85<-array(0,dim<-c(sn,1))
eel75<-array(0,dim<-c(10,1))
eel85<-array(0,dim<-c(10,1))

for (j in 1:sn) {
  d.11<-runif(wsn)
  sdata.1<-mvrnorm(wsn,mu1,var1)
  sdata.2<-mvrnorm(wsn,mu2,var2)
  sdata.all<-rbind(sdata.1[d.10<mix,],sdata.2[d.10>=mix,])
  wdata<-sdata.all[,1]
  eq.1<-quantile(wdata,.1)
  eq.2<-quantile(wdata,.2)
  eq.3<-quantile(wdata,.3)
  eq.4<-quantile(wdata,.4)
  eq.5<-quantile(wdata,.5)
  eq.6<-quantile(wdata,.6)
  eq.7<-quantile(wdata,.7)
  eq.8<-quantile(wdata,.8)
  eq.9<-quantile(wdata,.9)
  eq.10<-quantile(wdata,.10)
  ydata<-sdata.all[sample(seq(1,wsn),ysn),]
  mydata<-mean(ydata[,2])
  b1<-ydata[ydata[,1]<eq.1,2]
  b2<-ydata[ydata[,1]>=eq.1 & ydata[,1]<q.2,2]
  b3<-ydata[ydata[,1]>=eq.2 & ydata[,1]<q.3,2]
  b4<-ydata[ydata[,1]>=eq.3 & ydata[,1]<q.4,2]
  b5<-ydata[ydata[,1]>=eq.4 & ydata[,1]<q.5,2]
  b6<-ydata[ydata[,1]>=eq.5 & ydata[,1]<q.6,2]
  b7<-ydata[ydata[,1]>=eq.6 & ydata[,1]<q.7,2]
  b8<-ydata[ydata[,1]>=eq.7 & ydata[,1]<q.8,2]
  b9<-ydata[ydata[,1]>=eq.8 & ydata[,1]<q.9,2]
  b10<-ydata[ydata[,1]>=eq.9,2]
  if (sum(b1)>0) eel75[1]<-mean(pmax(0,.75*mydata-b1)) else
eel75[1]=0
  if (sum(b2)>0) eel75[2]<-mean(pmax(0,.75*mydata-b2)) else

```

```

eel75[2]=0
  if (sum(b3)>0) eel75[3]<-mean(pmax(0,.75*mydata-b3)) else
eel75[3]=0
  if (sum(b4)>0) eel75[4]<-mean(pmax(0,.75*mydata-b4)) else
eel75[4]=0
  if (sum(b5)>0) eel75[5]<-mean(pmax(0,.75*mydata-b5)) else
eel75[5]=0
  if (sum(b6)>0) eel75[6]<-mean(pmax(0,.75*mydata-b6)) else
eel75[6]=0
  if (sum(b7)>0) eel75[7]<-mean(pmax(0,.75*mydata-b7)) else
eel75[7]=0
  if (sum(b8)>0) eel75[8]<-mean(pmax(0,.75*mydata-b8)) else
eel75[8]=0
  if (sum(b9)>0) eel75[9]<-mean(pmax(0,.75*mydata-b9)) else
eel75[9]=0
  if (sum(b10)>0) eel75[10]<-mean(pmax(0,.75*mydata-b10)) else
eel75[10]=0
  ew.rs75[j]<-mean(eel75)/(.75*mydata)
  if (sum(b1)>0) eel85[1]<-mean(pmax(0,.85*mydata-b1)) else
eel85[1]=0
  if (sum(b2)>0) eel85[2]<-mean(pmax(0,.85*mydata-b2)) else
eel85[2]=0
  if (sum(b3)>0) eel85[3]<-mean(pmax(0,.85*mydata-b3)) else
eel85[3]=0
  if (sum(b4)>0) eel85[4]<-mean(pmax(0,.85*mydata-b4)) else
eel85[4]=0
  if (sum(b5)>0) eel85[5]<-mean(pmax(0,.85*mydata-b5)) else
eel85[5]=0
  if (sum(b6)>0) eel85[6]<-mean(pmax(0,.85*mydata-b6)) else
eel85[6]=0
  if (sum(b7)>0) eel85[7]<-mean(pmax(0,.85*mydata-b7)) else
eel85[7]=0
  if (sum(b8)>0) eel85[8]<-mean(pmax(0,.85*mydata-b8)) else
eel85[8]=0
  if (sum(b9)>0) eel85[9]<-mean(pmax(0,.85*mydata-b9)) else
eel85[9]=0
  if (sum(b10)>0) eel85[10]<-mean(pmax(0,.85*mydata-b10)) else
eel85[10]=0
  ew.rs85[j]<-mean(eel85)/(.85*mydata)
}
ewm.rs75<-mean(ew.rs75)
ewm.rs85<-mean(ew.rs85)
ewmse.75<-sum((ew.rs75-r75)^2)/sn
ewmse.85<-sum((ew.rs85-r85)^2)/sn

results<-array(0,dim<-c(7,2))
results[1,1]<-r75
results[1,2]<-r85
results[2,1]<-m.rs75
results[2,2]<-m.rs85
results[3,1]<-mse.75

```

```
results[3,2]<-mse.85
results[4,1]<-wm.rs75
results[4,2]<-wm.rs85
results[5,1]<-wmse.75
results[5,2]<-wmse.85
results[6,1]<-ewm.rs75
results[6,2]<-ewm.rs85
results[7,1]<-ewmse.75
results[7,2]<-ewmse.85
results
```