AIB Bootstrap Workshop

Four Actuarial Applications of the Bootstrap

₪

Stephen Mildenhall
CNA Re

May 26, 1999

Contents:

1. *Selecting a Unit Rate*
2. *Selecting a Coefficient of Variation and Pricing Excess Programs with Annual Aggregate Deductibles*
3. *A Bayesian-Bootstrap method for updating ultimate loss distributions*
4. *Pricing Weather Derivatives*

312 876 5295
stephen.mildenhall@cna.com

1. Selecting a Unit Rate

Problem

Given a nine year history of unit rates, make a selection. The historical rates were very volatile year-to-year:

Year	Unit Rate
1990	2.863
1991	4.374
1992	0.786
1993	7.047
1994	2.775
1995	20.712
1996	15.725
1997	2.313
1998	10.939
Average	8.171

Solution

Straightforward application of the bootstrap to determine a confidence interval around the mean.

Poor Person's Bootstrap in Excel

If *n* original data items are in a range starting at `A1`, use

```
=offset(A1,n*rand(),0)
```

in *n* cells to resample. Then use a "what-if" table (Alt data/table) with a blank-cell as column input and the required number of rows to automatically iterate the resample. Can use Excel to get histograms and percentiles. Percentiles from 10,000 replications are shown below.

Percentile	Unit Rate
45	7.18
50	7.44
55	7.72
65	8.29
75	8.93
85	9.71

Rich Person's Bootstrap in Matlab

Here is a Matlab program to do the same bootstrap, produce percentiles and the histogram shown in Figure 1.

```
function pctiles = XXXXBootstrap()

lossRate = [2.863
4.374
0.786
7.047
2.775
20.712
15.725
2.313
10.939];

% 10000 replications
bs = lossRate(1+floor(9*rand(10000,9)));

% compute average over 9 years
bs = sum(bs,2)/9;

% output percentiles
pctiles = [(50:100:10000)' bs(100:100:10000)];

% compute histogram
hist(bs,30)
title('Histogram for XXXX loss cost')
```

Comment

I used this method in pricing an account to help determine a comfort with the selected unit rate. We wrote the account.

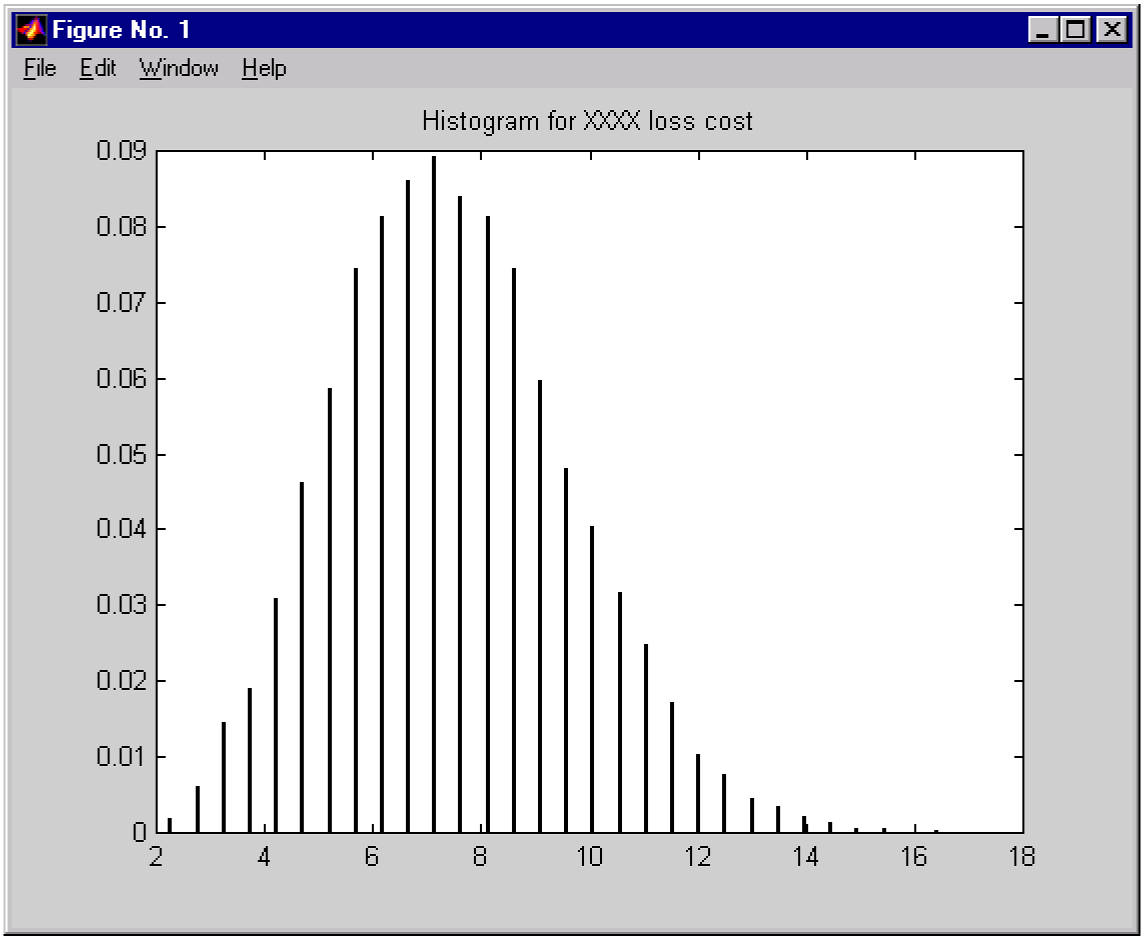


Figure 1: Histogram for Unit Rates

2. Selecting a Coefficient of Variation and Pricing Excess Programs with Annual Aggregate Deductibles

Problem

Price an excess reinsurance contract with an annual aggregate deductible (AAD). Requires a distribution for the aggregate excess losses. Key unknown is the coefficient of variation (CV) of aggregate losses. (Standard practice is to use a lognormal; the mean is the underlying loss pick, so the CV or variance is the only other unknown.)

Approach

In order to investigate the range of possible CV's Frank Bilotti of CNA Re used a bootstrap analysis. This work was completed in Excel and is shown on the next three pages. The underlying dataset was sampled from a reasonable lognormal to give 10 years of loss experience. These losses were then used to perform a bootstrap analysis. Frank computed

- The mean loss before application of the AAD
- The CV of the ten resampled losses
- The mean loss after the application of the AAD
- A modeled mean loss after the AAD, derived from a lognormal distribution fit to each resample.

Results

Happily the actual mean (of the underlying lognormal used to generate the original sample) lay in the 90th percentile of the bootstrap for underlying losses. The indicated range of CVs was very large with a 90% confidence interval of [0.376, 0.835] compared to the actual CV of 0.795. The results from the direct bootstrap and the lognormal fit model were reasonably close.

Bootstrap Method to Form Confidence Intervals for Excess Loss Pure Premiums

Variables Used	
Bootstrap Iterations	1,000
Annual Aggregate Deductible	750,000
Years of Experience	10

On Level Experience				
	Layer Loss	Simulation	Actual after AAD	after AAD Simulation
1998	2,044,243	130,571	1,294,243	-
1997	1,242,872	1,242,872	492,872	492,872
1996	1,904,337	2,753,167	1,154,337	2,003,167
1995	1,156,301	1,223,685	406,301	473,685
1994	803,683	2,753,167	53,683	2,003,167
1993	1,223,685	1,904,337	473,685	1,154,337
1992	2,753,167	1,242,872	2,003,167	492,872
1991	339,852	2,252,254	-	1,502,254
1990	2,252,254	2,252,254	1,502,254	1,502,254
1989	130,571	1,904,337	-	1,154,337
Average	1,385,097	1,765,952	738,054	1,077,895

Loss Before AAD			
	Generating LogNorm	Simulated Stats	Fit LogNorm Stats
Mean	1,029,929	1,385,097	1,385,097
StDev	818,982	845,294	845,294
CV	79.5%	61.0%	61.0%
Skew	2.888	0.071	2.058
EPP	**406,863**	**738,054**	**674,190**
95% CI	2,549,473	2,527,756	2,983,130
5% CI	254,894	224,747	468,592
E[X^2]	1.73E+12	2.63E+12	2.63E+12
E[X^3]	4.75E+18	5.36E+18	6.87E+18
E[X^4]	2.13E+25	1.21E+25	2.46E+25
mu	13.600		13.983
sigma	0.700		0.563

Bootstrap Iterations				
	Before AAD		After AAD	
				Lognormal
	Mean	CV	EPP	EPP
	1,765,952	46.1%	1,077,895	1,020,775
1	688,818	17.5%	144,734	132,218
2	717,523	25.2%	220,465	169,222
3	744,463	26.1%	220,626	189,077
4	767,046	26.4%	250,595	217,916
5	775,432	28.0%	256,248	228,111
6	778,366	28.0%	262,908	230,805
7	780,451	28.2%	268,824	234,474
8	783,678	28.4%	270,160	239,762
9	786,905	29.1%	274,032	240,245
10	794,158	29.7%	276,396	240,543
11	800,332	30.2%	281,445	243,557
12	804,040	30.6%	282,852	246,593
13	813,628	30.8%	284,734	247,924
14	818,117	31.1%	293,003	248,765
15	826,653	31.7%	296,225	253,137
16	826,937	32.0%	300,637	254,375
17	830,824	32.2%	300,971	258,225
18	832,996	32.2%	301,537	258,773
19	837,270	32.3%	306,826	261,007
20	839,083	32.8%	308,436	265,362
21	842,887	32.8%	314,113	266,526
22	844,721	33.5%	318,396	266,753
23	845,276	34.5%	319,447	276,440
24	855,014	34.5%	322,931	279,495
25	862,096	34.7%	326,734	280,494
26	863,532	34.7%	328,813	282,381
27	869,833	34.8%	331,016	282,846
28	872,607	35.0%	331,487	287,136
29	874,251	35.1%	334,305	290,494
30	878,276	35.2%	336,855	291,037
31	879,784	35.5%	345,512	291,823
32	883,232	35.5%	346,847	291,876
33	889,597	35.5%	351,455	296,912

Bootstrap Confidence Intervals				
	Losses in Layer Before AAD		**EPP After AAD**	
Percentile	Mean of Losses in Layer Before AAD	CV of losses in layer Before AAD	Bootstrap EPP	LogNormal fit to Mean/CV of Bootstrap
0.99	1,953,819	97.5%	1,244,833	1,204,945
0.95	1,800,870	83.5%	1,085,264	1,053,656
0.5	1,404,235	57.9%	742,052	684,656
0.05	976,472	37.6%	385,123	341,806
0.01	800,332	30.2%	281,445	243,557
Average	1,393,704	58.8%	**745,176**	**690,476**
Stdev	250,541	14.4%	209,445	212,773
CV	17.98%	24.46%	28.1%	30.8%
Skew	-12.1%	43.4%	1.5%	12.2%

3. A Bayesian-Bootstrap method for updating ultimate loss distributions

Problem

Given a prior ultimate loss distribution for a book of business and a paid or incurred loss development triangle, how should the distribution be updated to reflect incurred-to-date or paid-to-date loss information?

Solution

Apply bootstrap resampling to the development triangle to generate a distribution for factors-to-ultimate. Smooth using a low-pass filter, or other technique. Combine with the prior ultimate distribution to produce a bivariate distribution of factors-to-ultimate and ultimate loss. Transform the bivariate distribution to a distribution of observed loss at nth report and ultimate loss, using the fact that the observed loss times the factor-to-ultimate equals the ultimate loss. Then use standard Bayesian techniques to produce a posterior ultimate distribution given the observed loss.

Implementation

This is a very down to earth and practical method which is easy to implement on a computer and which gives reasonable results. It has minimal data input requirements: a loss development triangle and a reasonable prior ultimate. It can be used in DFA, in setting confidence intervals for reserves, and in profitability analysis.

Figures 2, 3 and 4 are from a Matlab tool I have built to implement the method. I am currently working on a Visual C++ (free/cheap) version of the model with 3-d graphics. The underlying data is paid losses for WC at 12 and 24 months---note the very high factors-to-ultimate! The six graphs in Figures 2 and 3 show the following.

- Top left: distribution of factor-to-ultimate obtained by applying bootstrap method to input triangle (spiky line) and a smoothing of the same distribution (bold line).
- Middle left: contour plot of the bivariate density of ultimate loss and factor-to-ultimate. Figure 3 assumes the two are independent. Figure 2 uses a Frank copula to induce correlation with a Kendall's tau of 0.25, which explains the northeast slant of the contours. (See Wang's PCAS paper on Aggregating Correlated Distributions for more details on copulas.) The observed loss corresponds to the diagonal dashed line.
- Lower left: prior ultimate (thin line) and posterior ultimate given observed loss (thick line). The indicated means are also shown.
- Upper right: derived distribution of the observed loss at 12 and 24 months.
- Middle right: transformed bivariate distribution, showing the density of loss at 12 and 24 months (Figures 2 and 3 respv.) vs. ultimate loss. Note the scales are different and that the dashed line has slope 1. Also the x-axis is on a much larger scale than for the other contour plot which explains why the contours appear patchy. There is more y-axis resolution than x-axis resolution.
- Lower right: comparison of the chain-ladder, B.F. and mean of posterior distribution (bold) estimates of ultimate losses for different observed losses (x-axis), showing that

the latter provides a very nice balance between the two more traditional methods. (Clearer on Figure 3.)

Figure 4 shows the underlying loss development triangle and some of the other user inputs to the model.

Next Steps

This model is still under development. I hope to roll out a practical version of the model at CNA Re to produce confidence intervals around profitability estimates, which will be used as part of results analysis. I will also be talking about the model at the CAS DFA Seminar in Chicago.

Figure 2: Bayesian bootstrap, first evaluation period

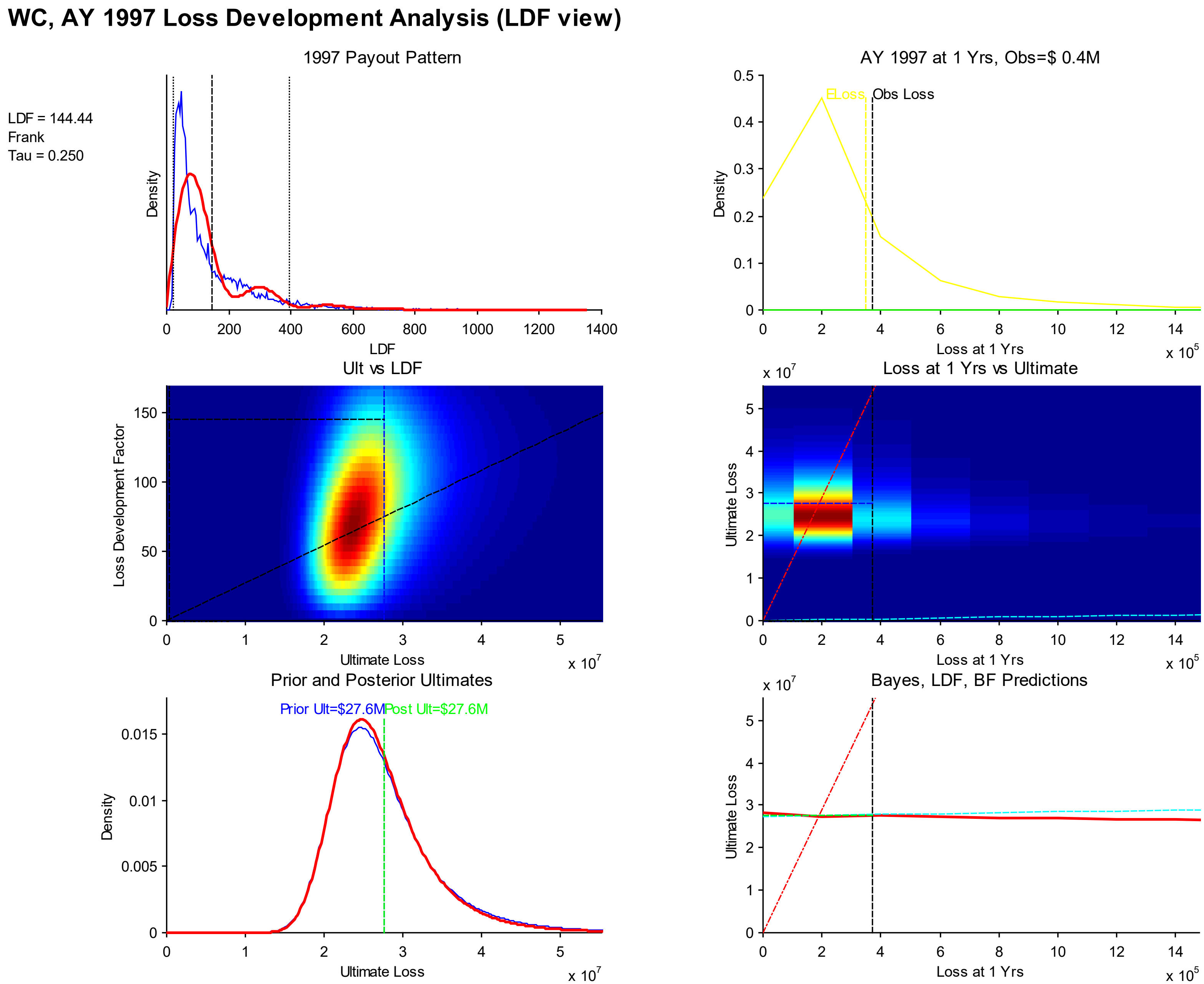
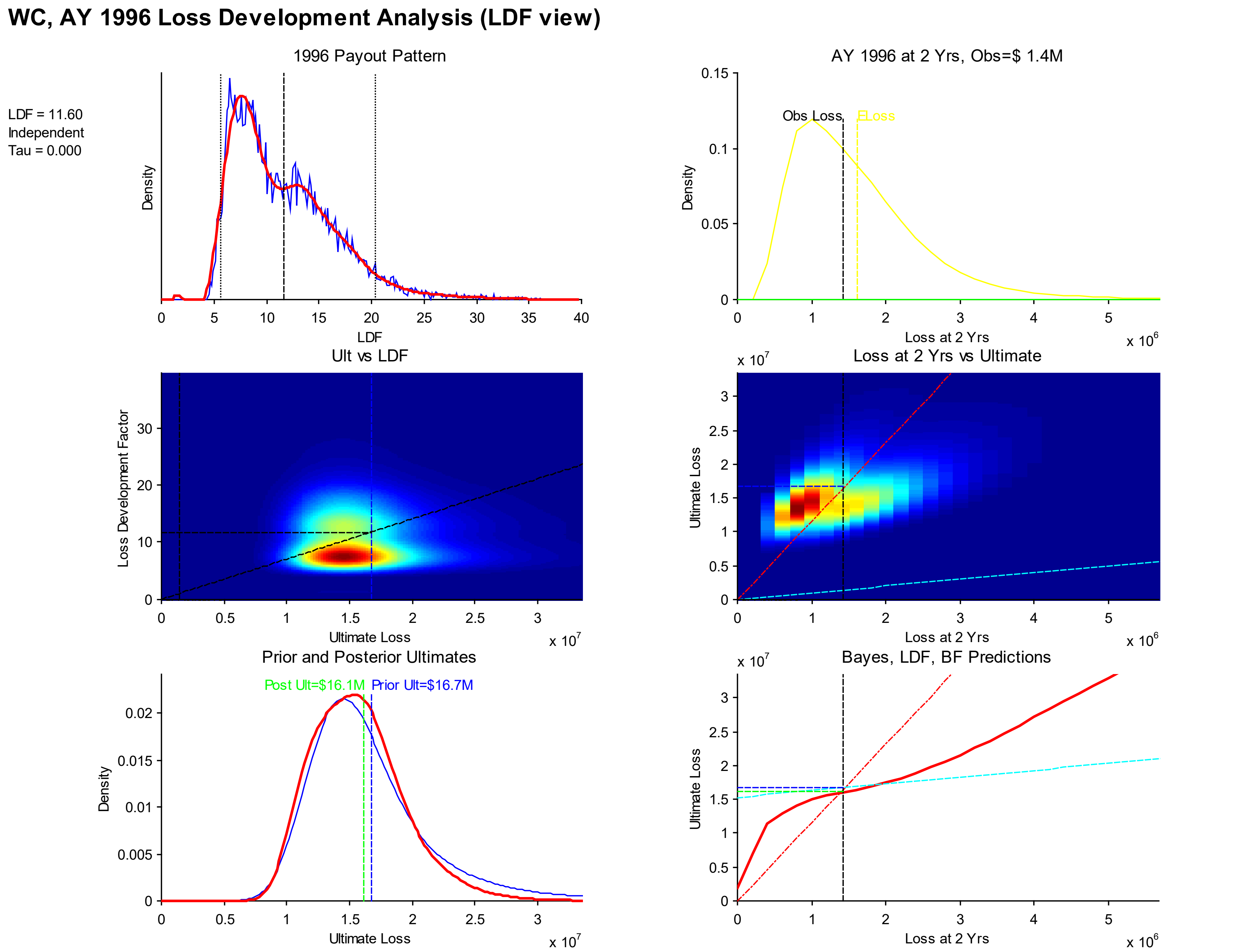


Figure 3: Bayesian bootstrap, second evaluation period



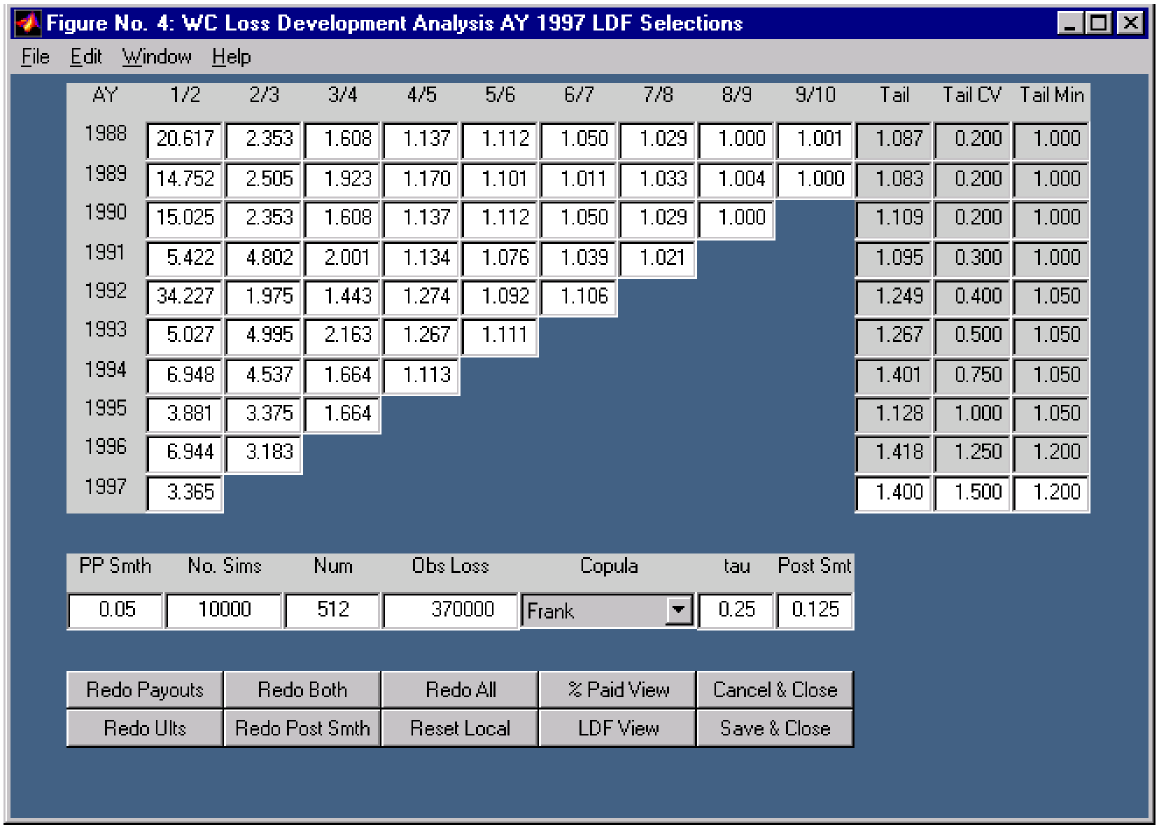


Figure 4: Bayesian Bootstap, underlying options

4. Pricing Weather Derivatives

Problem

Price a weather derivative product. Weather derivatives are typically options on a Heating Degree Day (winter) or Cooling Degree Day (summer) index. A Cooling Degree Day is

$$\text{CDD} = max(\text{daily temperature} - 65^{o}\text{F}, 0).$$

The temperature can be any agreed measure such as daily max, noon temperature or average temperature. The options are based on the sum of CDDs over a specified contract period. Each CDD is given a nominal value. For example, the contract may pay $5,000 for each CDD in excess of 350CDDs over a one month period. If the average high temperature is about 75 ºF then 300CDDs would be expected. CDD options can be used to hedge electricity generation costs. Clearly the distribution of CDDs is needed to price the product.

Data

Underlying data on weather is widely available on the Web. My underlying data source was a 20 year daily temperature time series. I approached the problem somewhat academically and assumed that the past is an indicator of the future. Actual players in the market combine similar statistical models with forecasts and opinions about global warming. (I have heard the less sophisticated use Black-Scholes!)

Model

The model broke the problem into several stages.

- Pick out the seasonal effect by smoothing the daily time series using a low-pass filter. This produces the sinusoidal graph shown in the top panel of Figure 5. The contract period is shown in bold.
- The year-to-year variability in sinusoids is shown in the middle panel, which collapses all the sine waves into one year.
- Fit a linear model to the original time-series minus the seasonal sinusoid. The model was simply today's temperature on yesterday's. The model fit initially looked extremely good, though on closer inspection the errors were not quite normal, slightly heteroscedastic (more variable in the spring) and they had a slight seasonal component.
- To produce the distribution of CDDs first sample from the annual sinusoids. Then sample separately from the linear model residuals for the correct time period and use the linear model to reconstruct the difference between the daily temperature and the sinusoid. Combine to produce a sample of the daily temperature over the required time period.
- Produce histograms and statistics as required from the resulting CDD distribution. In the lower panel of Figure 5, the tick marks show the historical observations.

This example shows how it is possible to combine statistical models with the bootstrap in order to retain correlations in the original data.

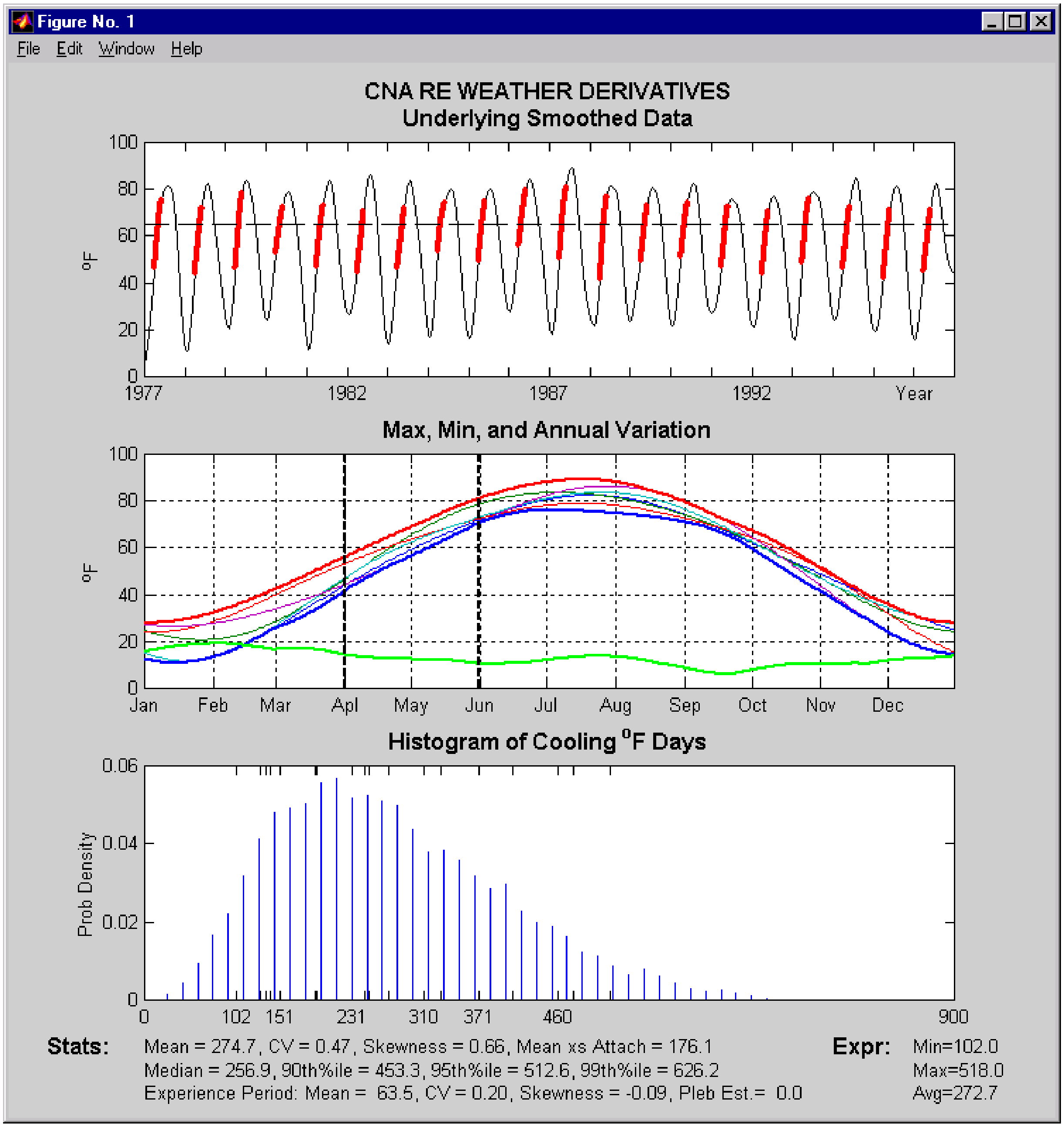


Figure 5: Weather Derivatives Pricing Tool