Loss Simulation Model
Testing and Parameterizing

Presented by Joseph Marker

Marker Actuarial Services, LLC
and
University of Michigan

2013 CAS Spring Meeting
Session C-1

Sources of Information

For more information about LSM and LSMWP, please visit http://www.casact.org/research/lsmwp.

The following provide technical detail on the model.

- "Modeling Loss Emergence and Settlement Processes-CAS Loss Simulation Model Working Party Summary Report," *Casualty Actuarial Society Forum*, Winter 2011, http://www.casact.org/pubs/forum/11wforum/LSMWP.pdf.

- Shang, Kailan, "Loss Simulation Model Testing and Enhancement", *Casualty Actuarial Society E-Forum*, Summer 2011. http://www.casact.org/pubs/forum/11sumforum/Shang.pdf

- "Parameterizing the Loss Simulation Model", Ball State University Research Course, LSMWP http://www.casact.org/research/lsmwp/BSUPaper.pdf

"Parameterizing the Loss Simulation Model"

Task:

- Fit distributions to real data from anonymous source.
- Use this result to design flexibility into LSM.
- Model used Richard Vaughan's previous modeling work done in APL.

Ball State University class did work for this paper in 2007.

- The paper documents the various areas investigated.
- "R" code used to fit models is included in paper.
- Parametric survival models were used to fit censored distributions.

Examples of models developed:

- Univariate modeling of lags and claim size with covariates
- Correlation between Settlement Lag and Claim Size for Auto BI
- Zero modification of the Claim Size distribution.
- Effect of Deductibles on Collision Losses; Pareto Model
- Interaction of Report Lag and Settlement Lag

"Modeling Loss Emergence and Settlement Processes"

Tested distributions of:

- Number of claims (frequency)
- Size of ultimate loss (severity)
- Correlated frequencies between lines using copulas.

Emphasis in Paper:

- Document the LSMWP's work in developing the LSM.
- Document the "R" code used in performing various tests.
- Provide references for those who want to explore the modeling further.
- Provide visual as well as formal tests: QQPlots, histograms, densities, etc.
- Document the model for users and developers wishing to customize the model.

Two types of testing

Q: Is the model output consistent with the specified parametric distribution?

"Non-constructive" test:

Run chi-square test of output distribution vs. theoretical distribution.

"Constructive" test:

1. Fit simulated output to the theoretically correct parametric model.

2. If fit is good, are the input parameter values within the confidence limits for the parameters?

Try to use "constructive" tests:

+ Provides a way for modeler to parameterize model from his/her data.

− More work than simple chi-square test.

"Loss Simulation Model Testing and Enhancement"

The DRMC and Committee on Reserves in 2011 recognized that much additional work was needed and issued a call for papers.

Areas of interest for call papers:

- Applications to reserving problems.
- Enhancements to model.
- Further testing of model in specific areas.

Kailan Shang's paper makes substantial contributions to the areas:

- Further testing of the LSM.
- Enhancement of Model.
- Reserving, through the testing of the case reserve adequacy parameters.

See Shang's presentation at the 2011 CLRS site.

Reserve Adequacy Factors

Shang tested and affirmed many of the LSM distributions.

Reserve adequacy factors allow modeling of reserve values between report and settlement dates. They may depend on claim department philosophy.

Shang found that:

The reserve adequacy factors input were not consistent with simulated output.

This led to simplification of factors in 2012, one of the "**model enhancements**".

Model specifies distribution of factors at "times" 0%, 40%, 70% and 90% of the interval between report date and settlement date.

Hai You's presentation explains these times in more detail.

This enhancement makes it easier to test the distribution from detailed claim data.

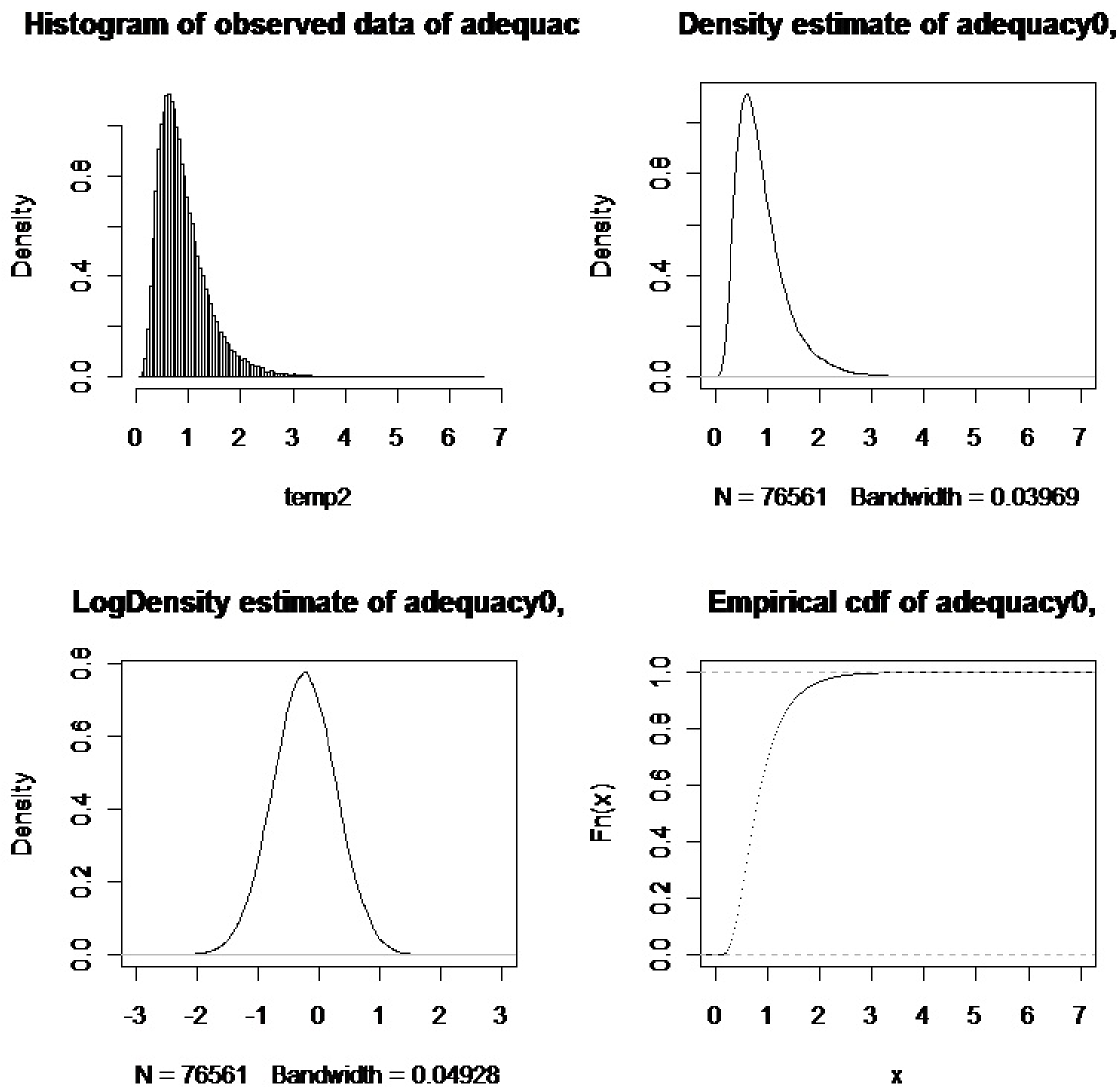
Determine and Test Adequacy Factors:

The test is "constructive" (defined earlier).
We discuss only the factor "adequacy0" at the 0% time. Others are similar.

Ran model with adequacy0 distributions specified as:
lognormal ($\mu = -0.239855$, $\sigma = 0.518642$), implying mean = 0.9, s.dev. = 0.5
Then examined adequacy0 output distribution and fit to lognormal.

"R" code:

```
hist(adeqauacy0,main="Histogram of observed adequacy0,",
  freq=FALSE,breaks=100,xlim=c(0,7))
plot(density(adequacy0),
  main="Density estimate of adequacy0,",   xlim=c(0,7))
plot(density(log(adequacy0)),
  main="LogDensity estimate of adequacy0,", xlim=c(-3,3))
plot(ecdf(adequacy0),  ## ecdf is empirical distribution function
  main="Empirical cdf of adequacy0,", xlim=c(0,7))
```



Descriptive statistics for adequacy0.

Determine and Test Adequacy Factors:

Logdensity graph shows that lognormal is likely to fit well.

R code to fit the data:

```
> dist0<-fitdistr(adequacy0,densfun='lognormal',
   control=list(trace=1,fnscale=10000));
> dist0$estimate
   meanlog      sdlog
-0.2417236  0.5191148
> -dist0$loglik
[1] 39933.05
```

Note closeness of fitted values to the inputs $\mu = -0.239855$, $\sigma = 0.518642$.

The fitting was done on "training data" = the first 80 of the 100 LSM simulations.
The model fit was then tested on the "test data" = last 20 simulations.

Chi-square tests (not shown here) shows model fits well.

Enhancement: Two-state Regime Switching Model

Shang suggested this enhancement, which has been implemented.

Frequency and severity distribution may not be stable over time. Some reasons:

- Structural changes
- Cyclical patterns
- Idiosyncratic nature

To handle this, we allow the model to be in one of two "states" for each month.
The "state" over time follows a stationary Markov Chain:

If X_n is the state at time n, then $\Pr[X_{n+1} = j \mid X_n = i] = p_{ij}$, for $i, j = 1,2$.
The conditional probability of $X_{n+1} \mid X_1, X_2 \ldots X_n$ depends only on X_n

There are two states and time is discrete for this Markov Chain.

The modeler specifies the transition probabilities p_{ij}.

Testing the Two-state Regime Switching:

We did not find many sources discussing this topic.
The best is Anderson and Goodman's "Statistical Inference about Markov Chains"[1]

We ran the LSM with 12 accident months and 10,000 simulations.
The programmer modified the model to output the value of the "state" .
Transition probabilities were $p_{11} = 0.80$ and $p_{22} = 0.40$.

Resultant transition matrix is $P = \begin{pmatrix} 0.80 & 0.20 \\ 0.60 & 0.40 \end{pmatrix}$

Ran several tests based on the Anderson-Goodman ("A-G") paper.

---

[1] Anderson, T.W and Goodman, Leo A, "Statistical Inference about Markov Chains". Annals of Mathematical Statistics (1957), available on line.

Testing the Two-state Regime Switching:

<u>Test 1: Calculate Estimated Transition Probabilities:</u>

A-G shows that the maximum likelihood estimate of p_{ij} is:

$$\hat{p}_{ij} = \frac{n_{ij}}{n_i^*} \text{ with } n_{ij} = \sum_{t=1}^{11} n_{ij}(t), \text{ and } n_i^* = \sum_{t=1}^{11} n_i^*(t) .$$

Here $n_{ij}(t)$ is the number of observations in state i at time t and state j at time t+1, while $n_i^*(t)$ is the number in state i at time t.

We can add across t if the transition probabilities are stationary.

The fitted transition matrix is

$$\hat{P} = \begin{pmatrix} 0.8008729 & 0.1991271 \\ 0.5968471 & 0.4030529 \end{pmatrix}, \quad \text{very close to the input matrix.}$$

Testing the Two-state Regime Switching:

Test 2: Test hypothesis that fitted probabilities match the inputs.

A-G shows that, under the null hypothesis, $\sum_{j=1}^{2} n_i^* \frac{\left(\hat{p}_{ij} - p_{ij}^0\right)^2}{p_{ij}^0}$, for $i = 1, 2$,

has the asymptotic chi-square distribution with 1 degree of freedom.

Here the p_{ij}^0 are the transition probabilities input to the model.

Chi-square values are not significant.

Thus, we cannot reject the null hypothesis that p_{ij}^0 are the correct probabilities.

Test 3: Test the stationary distribution.

The stationary distribution for matrix P^0 is $\pi_1 = 0.75$ and $\pi_2 = 0.25$.

The model uses the stationary distribution as the initial distribution.

Therefore, the distribution at each time should have stationary distribution.

A chi-square test confirms this.

Testing the First Order Markov Chain Assumption:

Defn: A Markov chain is of order *n* if the distribution of X_t depends on $\{X_{t\text{-}1}, X_{t\text{-}2} \ldots \quad X_{t\text{-}n}\}$, but not on any observations before time *t-n*.

The traditional term "Markov chain" means a first-order Markov Chain.

We test whether our state process is first-order, assuming it is second-order.

<u>Test 4 (non-constructive):</u>

If the chain is first-order with the given input parameters, then the two-step transition matrix $P^{(2)} = \left(P^0\right)^2 = \begin{pmatrix} 0.76 & 0.24 \\ 0.72 & 0.28 \end{pmatrix}$.

Chi-square tests on the observed two-step transition probabilities confirm this.

Testing the First Order Markov Chain Assumption:

<u>Test 5 (constructive):</u>

Terminology: $p_{ijk} = \Pr[X_{t+2}=k \mid X_t = i, X_{t+1}=j]$

$n_{ijk}(t)$ = number of observations with $X_t = i$, $X_{t+1}=j$, and $X_{t+2}=k$

The second-order chain is a first-order chain if p_{ijk} is independent of i.

Testing a second-order chain is not complex because a second-order chain is a first-order chain as follows:

Define the "state" at time t as i, j if $X_{t-1} = i$ and $X_t = j$ (2^2 states in all).

The transition probabilities are $p_{ij,mk} = \begin{cases} p_{ijk} & \text{if } m = j \\ 0 & \text{otherwise} \end{cases}$

The original chain is first order if, for all j and k, p_{ijk} are identical for all i.

The transition matrix for the modified chain is $P = \begin{pmatrix} p_{111} & p_{112} & 0 & 0 \\ 0 & 0 & p_{121} & p_{122} \\ p_{211} & p_{212} & 0 & 0 \\ 0 & 0 & p_{221} & p_{222} \end{pmatrix}$.

Testing the First Order Markov Chain Assumption:

Test 5 continued:

A-G shows that the maximum likelihood estimate of p_{ijk} is:

$$\hat{p}_{ijk} = \frac{n_{ijk}}{n_{ij}^*} \text{ with } n_{ijk} = \sum_{t=1}^{10} n_{ijk}(t), \text{ and } n_{ij}^* = \sum_{t=1}^{10} n_{ij}^*(t) .$$

Using the model's output, we obtain

$$\hat{P} = \begin{pmatrix} \underline{i} & \underline{j} & | & & & & \\ 1 & 1 & | & 0.8013 & 0.1987 & 0 & 0 \\ 1 & 2 & | & 0 & 0 & 0.5934 & 0.4066 \\ 2 & 1 & | & 0.7981 & 0.2019 & 0 & 0 \\ 2 & 2 & | & 0 & 0 & 0.5996 & 0.4004 \end{pmatrix}.$$

Compare to the null hypothesis that

$$P = \begin{pmatrix} 0.80 & 0.20 & 0 & 0 \\ 0 & 0 & 0.60 & 0.40 \\ 0.80 & 0.20 & 0 & 0 \\ 0 & 0 & 0.60 & 0.40 \end{pmatrix}$$

For each ij, the statistic $\sum_{k=1}^{2} n_{ij}^* \frac{\left(\hat{p}_{ijk} - p_{jk}^0\right)^2}{p_{jk}^0}$ is asymptotically chi-square with one d.f.

Each chi-square result confirms that the null hypothesis should not be rejected.

Summary

Reserve adequacy factors and the two-state switching appear to work as intended. We ran the LSM with known parameter values and fit the output to the theoretical models, using chi-square tests to verify the "null hypothesis."

The work is intended to help modelers in several ways:

- A paper showing containing the R code will be furnished later.
- The code and paper show how to organize "user data" into arrays and to fit the data to parametric distributions.
- The fitting procedures described enable the modeler to set LSM parameters for reserve adequacy and for the two-stage switching.
- The presentation starts with an outline of all the previous work done to design and test parameters for other distributions, such as frequency and severity.

Acknowledgement: **Wenwen Ying**, a student at the University of Michigan, developed the testing procedures and wrote most of the "R" code.