A Flexible Framework for Stochastic Reserving Models

Presented by
Roger M. Hayne, PhD., FCAS, MAAA

CAS Centennial Celebration
10-12 November 2014
New York, NY

Milliman

CAS Antitrust Notice

- The Casualty Actuarial Society is committed to adhering strictly to the letter and spirit of the antitrust laws. Seminars conducted under the auspices of the CAS are designed solely to provide a forum for the expression of various points of view on topics described in the programs or agendas for such meetings.
- Under no circumstances shall CAS seminars be used as a means for competing companies or firms to reach any understanding – expressed or implied – that restricts competition or in any way impairs the ability of members to exercise independent business judgment regarding matters affecting competition.
- It is the responsibility of all seminar participants to be aware of antitrust regulations, to prevent any written or verbal discussions that appear to violate these laws, and to adhere in every respect to the CAS antitrust compliance policy.

2 Milliman

Objectives

- Avoid the need for “tricks” to make models more tractable
- Deal directly with non-linear models
- Establish a common framework for evaluation and study of many candidate models, including many deterministic ones currently in use
- Derive estimates of model parameters as well as of uncertainty in the model parameters
- Selected solution – Maximum Likelihood Estimators (MLEs)
- Many convenient properties of MLEs including
  - Asymptotically unbiased
  - Asymptotically efficient
  - Asymptotically Normal (Gaussian)

3

Framework

- Work with standard development triangles, can be paid or incurred
- Consider incremental amounts, removing induced correlation between amounts at one age and the next
- So as to remove effects of volume change from one accident year to the next focus on average (rather than aggregate) costs, either severity or pure premium
- Assume A_{ij} denoting the incremental average for accident year i and development year j is a random variable
- Assume further that the various incremental amounts are independent across accident and development years

4

Overall Model

- Assume the expected value of the incremental averages can be expressed as functions of a parameter (possibly vector) **θ**

$$\mathrm{E}\left(A_{ij}\right) = g_{ij}\left(\boldsymbol{\theta}\right)$$

- We will also assume that the variance of each A_{ij} is proportional to a power of its mean. Since each A_{ij} represents an average we also adjust the assumed variance to reflect different exposure (claim) levels by accident year, denoted w_i so we assume

$$\mathrm{Var}\left(A_{ij}\right) = e^{\kappa - w_i}\left(g_{ij}\left(\boldsymbol{\theta}\right)^2\right)^p$$

5 Milliman

Overall Model - Continued

- Commonly used development models incorporate this assumed variance structure (omitting the *w* values)
  - Constant variance if $p = 0$
  - Over-Dispersed Poisson if $p = 0.5$
- We allow p to be a parameter to be estimated using the data
- Since the A_{ij} are averages of a number of exposure units, the law of large numbers implies their distribution is asymptotically Gaussian
- Thus we assume the A_{ij} are Gaussian with mean and variance given in the prior slide

6 Milliman

Parameter Estimation

- Under the previous assumptions the negative log likelihood function of the observations given a set of parameters is given by

$$\ell\left(A_{11}, A_{12}, \ldots, A_{n1}; \boldsymbol{\theta}, \kappa, p\right) =$$
$$\sum \frac{\kappa - e_i + \ln\left(2\pi\left(g_{ij}\left(\boldsymbol{\theta}\right)^2\right)^p\right)}{2} + \frac{\left(A_{ij} - g_{ij}\left(\boldsymbol{\theta}\right)\right)^2}{2e^{\kappa - w_i}\left(g_{ij}\left(\boldsymbol{\theta}\right)^2\right)^p}$$

- Except for very simple loss models finding the minimum of this expression not possible
- Numerical methods available in many packages (R, MATLAB, etc.) can handle the job

7 Milliman

Parameter Uncertainty

- MLE theory gives a way to estimate uncertainty in the parameters estimated using MLE, given the data and the underlying model
- For this we need the Fisher Information Matrix
- The Fisher Information Matrix is the Hessian with respect to the parameters of the negative log likelihood function on the previous slide
- The Hessian of a function is the matrix whose ij^{th} element is the second derivative of that function with respect to the i^{th} variable then with respect to the j^{th} variable
- The variance-covariance matrix of the parameters is the inverse of the Fisher Information Matrix evaluated at the MLE

8 Milliman

Distribution of Forecasts, Fixed Parameters

- Since we assume incremental averages are independent once we have the parameter estimates we have estimate of the distribution of future outcomes given the parameters

$$R_i \sim \mathrm{N}\left(E_i \sum_{j=n-i+2}^{n} g_{ij}\left(\hat{\boldsymbol{\theta}}\right), E_i^2 \sum_{j=n-i+2}^{n} e^{\hat{\kappa}-e_i} \left(g_{ij}\left(\hat{\boldsymbol{\theta}}\right)^2 \right)^{\hat{p}} \right)$$

$$R_T \sim \mathrm{N}\left(\sum_{i=1}^{m} E_i \sum_{j=n-i+2}^{n} g_{ij}\left(\hat{\boldsymbol{\theta}}\right), \sum_{i=1}^{m} E_i^2 \sum_{j=n-i+2}^{n} e^{\hat{\kappa}-e_i} \left(g_{ij}\left(\hat{\boldsymbol{\theta}}\right)^2 \right)^{\hat{p}} \right)$$

- This is the estimate for the average future forecast payment per unit of exposure, multiplying by exposures
- This is not a distribution of forecasts under the model since there is no uncertainty in parameter choice

9 Milliman

Distribution of Forecasts

- If we assume sample size sufficient for the MLE to have a distribution that is close to Gaussian then that (multivariate) Gaussian has
  - Mean equal to the MLE
  - Variance-Covariance Matrix equal to the inverse of the Fisher Information Matrix
- With the formula on the prior slide we can estimate the distribution of forecasts from the model given the data
- Generally a closed form solution is either difficult or impossible, but the distribution of forecasts can be simulated
- MCMC is one way, another is a bit more straight forward
  - First randomly pick a parameter vector from the Gaussian
  - Then randomly calculate reserves from distribution on prior slide

10 Milliman

Observations to This Point

- Discussion so far does not specify a model for the incremental averages
- In order for the MLE theorem to work, you only need mild regularity conditions, most usual reserve methods are
- Minimization of the negative log likelihood function can be done numerically, either by bespoke code, or with tools found in available packages such as R or MATLAB
- Usually the optimization routines are more efficient if gradient and Hessian are specified, Hessian also needed for Fisher Information
- All these derivatives can be worked out in terms of derivatives of the g functions assuming only they are twice differentiable

11 Milliman

Parameterization – Cape Cod

- Generally the Bornhuetter-Furgeson method has $A_{ij} = \alpha_i \beta_j$
- Overspecifies the model
- We use the following (similar to England & Verall)

$$g_{ij}(\boldsymbol{\theta}) = \begin{cases} \theta_1 \text{ if } i = j = 1 \\ \theta_1 \theta_i \text{ if } j = 1 \text{ and } i > 1 \\ \theta_1 \theta_{m+j-1} \text{ if } i = 1 \text{ and } j > 1 \\ \theta_1 \theta_i \theta_{m+j-1} \text{ if } i > 1 \text{ and } j > 1 \end{cases}$$

- θ_1 is the upper left corner incremental
- θ_i for $i = 2, \ldots, n$ is change in incremental from accident year $i\text{-}1$ to age i
- θ_i for $i = n+1, \ldots, m+n-1$ is change from age $i - n$ to accident year $i - n + 1$

12 Milliman

Parameterization – Berquist-Sherman

- Cape Cod has *n* + *m* -1 parameters
- Berquist & Sherman recognized that incremental averages may be related to each other from one accident year to the next by trend
- They developed a model that replaces accident year level factors by a simple trend

$$g_{ij}(\boldsymbol{\theta}) = \theta_j e^{i\theta_{n+1}}$$

- θ_j for $j = 1, \ldots, n$ is the accident year 0 average incremental cost at age j
- θ_{n+1} is the natural log of the annual trend in the data

13 Milliman

Parameterization – Wright

- Number of Cape Cod parameters can also be reduced in the development direction
- Wright used a Hoerl curve to accomplish this
- We use a similar approach

$$g_{ij}(\boldsymbol{\theta}) = \exp\left(\theta_i + \theta_{m+1} j + \theta_{m+2} j^2 + \theta_{m+3} \ln(j)\right)$$

- θ_i for $i = 1, \ldots, m$ is the accident year level
- Remaining three parameters determine shape of incremental averages as an accident year ages
- Flexible enough to be either monotonic or have a mode

14 Milliman

Parameterization – Hoerl Surface

- The Wright model still has *m* + 3 parameters
- Why not reduce the parameters still further assuming (as in Berquist – Sherman) an expected annual accident year trend?
- This gives

$$g_{ij}(\boldsymbol{\theta}) = \exp\left(\theta_1 + \theta_2 j + \theta_3 j^2 + \theta_4 \ln(j) + \theta_5 i\right)$$

- θ_5 represents the log of the annual trend
- Both this and the Wright model require positive expected incrementals (no such constraint in Cape Cod or Berquist-Sherman)
- Only 5 parameters

15 Milliman

Parameterization – Chain Ladder

- Basic requirements for expected values
  - Ratio of cumulative averages from one age to the next same for all accident years
  - The expected amount to date (on the diagonal) is observed amount to date
- In our parameterization we label the amount to date for accident year *i* as P_i and the age of accident year *i* to date as n_i
- Also in our parameterization we can think of the parameters θ_j as the portion of the total amounts emerging at age *j*
- The incremental percentages can be negative or larger than 1
- We force the percentage for the last age to be the complement of the remainder resulting in *n* – 1 parameters.

16 Milliman

Parameterization – Chain Ladder (Continued)

$$g_{ij}(\boldsymbol{\theta}) = \begin{cases} P_1\theta_j & \text{if } j < n \text{ and } i = 1 \\ P_1\left(1 - \sum_{k=1}^{n-1}\theta_k\right) & \text{if } j = n \text{ and } i = 1 \\ \dfrac{P_i\theta_j}{\sum_{k=1}^{n_i}\theta_k} & \text{if } j < n \text{ and } i \neq 1 \\ \dfrac{P_i}{\sum_{k=1}^{n_i}\theta_k}\left(1 - \sum_{k=1}^{n-1}\theta_k\right) & \text{if } j = n \text{ and } i \neq 1 \end{cases}$$

17

Milliman

Example Commercial Auto Liab. Paid Data

Cumulative Average Paid Loss & Defense & Cost Containment Expenses per Estimated Ultimate Claim

Accident	Months of Development										Count
Year	12	24	36	48	60	72	84	96	108	120	Forecast
2001	670	1,480	1,939	2,466	2,838	3,004	3,055	3,133	3,141	3,160	39,161
2002	768	1,593	2,464	3,020	3,375	3,554	3,602	3,627	3,646		38,672
2003	741	1,616	2,346	2,911	3,202	3,418	3,507	3,529			41,801
2004	862	1,755	2,535	3,271	3,740	4,003	4,125				42,263
2005	841	1,859	2,805	3,445	3,950	4,186					41,481
2006	848	2,053	3,076	3,861	4,352						40,214
2007	902	1,928	3,004	3,881							43,599
2008	935	2,104	3,182								42,118
2009	759	1,585									43,479
2010	723										49,492

18

Milliman

Results

Model	Expected Reserves (000,000)
Cape Cod	$391
Berquist-Sherman	480
Wright	388
Hoerl Surface	474
Chain Ladder	393

- Some difference in expected reserves
- Is the difference random?
- Is the difference significant?
- How do you know?
- Stochastic models help answer these questions

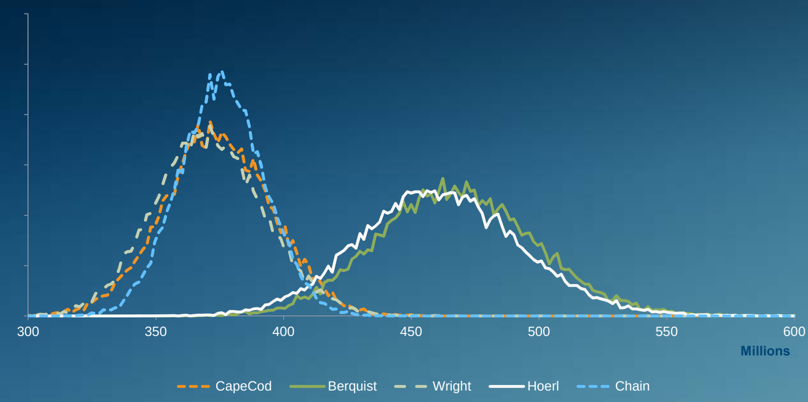
19 Milliman

Process vs. Parameter Uncertainty

Model	Total Reserve Process Std. Dev. (000)	Total Reserve Total Std. Dev. (000)
Cape Cod	$ 9,435	$20,101
Berquist-Sherman	15,997	29,405
Wright	10,029	20,375
Hoerl Surface	16,115	29,454
Chain Ladder	9,447	15,557

20 Milliman

Reserve Forecasts by Model



21

Milliman

What Happened?

Standardized Residuals

0.22	0.13	-2.73	-0.58	0.08	-0.57	-0.97	1.88	-0.49	0.31
0.90	-0.07	0.89	-0.55	-0.41	-0.45	-1.22	-1.32	0.87	
0.30	0.04	-0.76	-0.76	-1.64	0.19	0.32	-1.61		
1.18	-0.16	-0.63	0.78	0.72	1.02	1.54			
0.60	0.54	0.51	-0.55	0.93	0.17				
0.31	1.67	0.80	0.58	0.42					
0.46	-0.14	0.89	1.18						
0.39	0.60	0.50							
-1.52	-2.43								
-2.29									

Berquist

Berquist

0.86	1.25	-3.37	0.16	0.92	-0.31	-0.76	1.99	-0.14	0.30
0.78	-0.37	1.66	-0.81	-0.61	-0.71	-1.30	-1.22	0.71	
0.75	0.83	-0.17	-0.34	-1.79	0.53	0.51	-1.34		
0.67	-1.11	-1.41	0.84	0.65	0.82	1.29			
-0.19	0.16	0.58	-1.23	1.01	-0.19				
-1.38	1.24	0.30	-0.08	-0.24					
-0.74	-1.35	0.96	1.33						
-0.65	0.17	0.52							
0.58	-0.48								
0.03									

Cape Cod

Cape Cod

0.20	-0.03	-2.85	-0.45	0.13	-0.61	-1.16	1.73	-0.39	1.92
0.90	-0.23	0.82	-0.40	-0.31	-0.51	-1.32	-0.63	0.30	
0.32	-0.11	-0.85	-0.60	-1.44	0.01	-0.35	-0.83		
1.23	-0.30	-0.71	0.99	0.78	0.71	0.44			
0.67	0.46	0.49	-0.35	1.00	0.01				
0.39	1.66	0.81	0.83	0.53					
0.57	-0.23	0.93	1.49						
0.52	0.58	0.54							
-1.47	-2.66								
-2.27									

Hoerl

Hoerl

0.91	1.30	-3.32	0.20	0.95	-0.29	-0.75	2.02	-0.14	0.31
0.83	-0.31	1.72	-0.76	-0.57	-0.68	-1.29	-1.21	0.72	
0.77	0.86	-0.14	-0.32	-1.78	0.55	0.52	-1.34		
0.61	-1.16	-1.46	0.79	0.61	0.80	1.28			
-0.21	0.13	0.56	-1.24	0.99	-0.19				
-1.38	1.23	0.30	-0.08	-0.24					
-0.74	-1.35	0.95	1.32						
-0.67	0.14	0.50							
0.55	-0.51								
0.00									

Chain Ladder

Chain Ladder

22

Milliman

Some Observations

- The data imply that the variance for payments in a cell are roughly proportional to the mean to the 0.85 power for both Cape Cod and Chain Ladder, roughly to the mean for the Hoerl model and to the mean to the 1.30 power for the Berquist model.
- Total standard deviation well above process, often more than double, meaning parameter uncertainty is significant
- Comparison of forecasts among models underlines the importance of model uncertainty
- Still more work to be done to get a handle on model uncertainty – possibly greater than the other two sources

23 Milliman

More Observations

- We chose a relatively simple models for the expected value
- Nothing in this approach makes special use of the structure of the models
- Models do not need to be linear nor do they need to be transformed to linear by a function with particular properties
- Variance structure is selected to parallel stochastic chain ladder approaches (overdispersed Poisson, etc.) and allow the data to select the power
- The general approach is also applicable to a wide range of models
- This allows us to consider a richer collection of models than simply those that are linear or linearizable

24 Milliman

Some Cautions

- MODEL UNCERATINTY STILL NEEDS TO BE CONSIDERED thus distributions are distributions of outcomes under a specific models and must not be confused with the actual distribution of outcomes for the loss process
- An evolutionary Bayesian approach can help address model uncertainty
  - Apply a collection of models and judgmentally weight (a subjective prior)
  - Observe results for next year and reweight using Bayes Theorem
- We are using asymptotic properties, no guarantee we are far enough in the limit to assure these are close enough
- Actuarial “experiments” not repeatable so frequentist approach (MLE) may not be appropriate

Questions?


Thank You
27
Milliman

