Reserving Mixology 201:
Concocting a Reserve Distribution

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

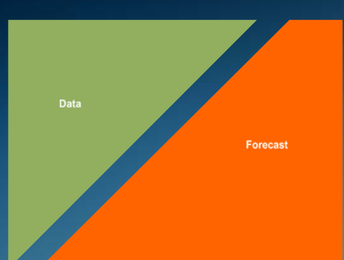
Casualty Loss Reserve Seminar
10 - 11 September 2015
Atlanta, GA

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.

1

Usual Reserve Triangle Problem



2

Usual Reserve Triangle Problem

- Set up some formula or recipe to describe the entire rectangle
  - Chain Ladder: cumulative for year i at age j = factor at age j-1 x cumulative for year i at age j-1
  - Bornhuetter Ferguson: cumulative for year i after age j = a-priori for year i x (1 – percent emerged at age j)
  - Berquist-Sherman: incremental average for year i at age j = incremental average for year 1 at age j x i years of trend
  - Etc.
- Use historical data to derive estimates of unknown amounts (parameters)
- Use estimated parameters to fill in future forecasts or "square the triangle"

3 Milliman

Usual Reserve Triangle Problem

- Some basic observations:
  - Final outcome most likely will be different than any one forecast
  - Each method has strengths and weaknesses
  - In general methods assume future will mirror the past
  - Traditional approaches typically use a variety of approaches to identify differences between method assumptions and actual data
- In fact ASOP No. 43 says:

  The actuary should consider the use of multiple methods or models appropriate to the purpose, nature and scope of the assignment and the characteristics of the claims unless, in the actuary's professional judgment, reliance upon a single method or model is reasonable given the circumstances. If for any material component of the unpaid claim estimate the actuary does not use multiple methods or models, the actuary should disclose and discuss the rationale for this decision in the actuarial communication.

4 Milliman

Enter Probability & Statistics

- Probability is a way to talk about events with uncertain outcomes
- For discrete events that exhaust all possibilities a probability of an event can be thought of as the likelihood of that particular event occurring. As a corollary the sum of all probabilities is 1 (or 100%).
- Simple example – the flip of a fair coin. Here there are only two possible outcomes, often called H or T, each with equal chance, or probability of 0.50 (50%).
- Discrete events do not necessarily have to be finite in number, consider the Poisson distribution. Here events are positive integers k and $\Pr(X=k) = \lambda^k e^{-\lambda} / k!$.

5 Milliman

Enter Probability & Statistics

- Reserves usually concern dollars and cents. Though discrete amounts (you cannot have a claim settle for $0.005), there are so many, and the various events are so “close” together they are often thought of as outcomes of a continuous, not a discrete process
- You can talk about continuous probabilities, but immediately find out that very often the probability of a specific amount is 0, but the probability for an interval containing that amount is positive, no matter how small that interval is.
- Odd, isn’t it. Well, as John von Neumann said “Young man, in mathematics you don't understand things. You just get used to them.”

6 Milliman

Enter Probability & Statistics

- Probability gives us a way to talk about uncertainty
- Many times probability distributions can be expressed in terms of a relatively small number of parameters
  - Poisson in prior example, only one
  - Gaussian, lognormal, Gamma, and many others have two
  - Mixture of exponentials can have as many as you would like
- The trick is to find the right distribution for a problem at hand and then derive estimates of the parameters
- Sometimes the right distribution is pretty obvious, e.g. binomial for tosses of a coin
- Sometimes it is not so obvious as in reserves

7 Milliman

Looking at Uncertainty

- Let’s go back to coin toss problem
- There are only two possible states of the world possible, H or T
- What is the outcome of the next toss?
- Given the information provided here how likely are you to be correct in your answer to the previous question?
- If you answered anything but “I haven’t a clue” to the last question you are fooling yourself
- I have not given you enough information yet
- If I told you it was a fair coin, then your likelihood of being right would be 50%
- In no event, though for a fair coin would you be certain of the outcome of the next single toss (process uncertainty)

8 Milliman

Looking at Uncertainty

- Now I will tell you the coin is possibly biased. In other words the chance of H is *p* with the possibility that *p* is not 0.50
- How would you answer the prior two questions?
- How would you get some comfort with your answer? Experiment
- I'll give you the coin to toss a number of times
- Would that help?
- Here you have a model and some data (the outcomes of a number of experiments)
- Suppose you toss the coin *n* times and come up with *x* heads, what would your guess at *p* be? *x / n?*
- Is this guess any good? Are there better guesses?

9 Milliman

The Statistics in Probability & Statistics

- In this example the amount *x / n* is a statistic derived from the sample tosses and you think it should be a decent estimate of the unknown parameter *p*
- Notice we have derived this from the frequent repetition of an experiment. This is termed a "frequentist" approach to estimating the parameter *p*
- What is a "good" estimate? What is the characteristic of a "best" or "efficient" estimate.
- Well a "good" estimate will eventually (with a large number of trials) will tend toward what we are trying to estimate
- An "efficient" estimate will have smaller uncertainty (variance) than other estimates.

10 Milliman

Maximum Likelihood

- If we assume all flips are independent, the chance of seeing *x* heads in *N* tosses is the likelihood function

$$\binom{x}{n} p^{x} (1-p)^{n-x}$$

- One estimate we could take for *p* would be that value that gives the largest chance of observing *x* heads in *N* tosses
- In this case that value is *x / n*, the observed proportion of heads in our "experiment"
- The value of a parameter that maximizes the likelihood of the observed outcome of a particular experiment is called the maximum likelihood estimator (MLE) of that parameter

11 Milliman

Some Properties of MLEs

- As the number of observations in an experiment gets large the resulting MLE is
  - Asymptotically unbiased (is expected to converge to the parameter)
  - Asymptotically efficient (no other estimator has lower variance)
  - Asymptotically normal
- Define the Fisher information matrix as the expected value of the Hessian matrix (matrix of second partial derivatives with respect to the parameters) of the negative log-likelihood function
- The variance-covariance matrix of the limiting Gaussian distribution is the inverse of the Fisher information matrix typically evaluated at the parameter estimates

12 Milliman

MLE Example

- In this coin example the negative log likelihood is

$$\ell(p) = -\ln\left(\binom{N}{x} p^x (1-p)^{N-x}\right) = (x-N)\ln(1-p) - x\ln p - \ln\binom{N}{x}$$

- With derivatives

$$\ell'(p) = \frac{N-x}{1-p} - \frac{x}{p}$$

$$\ell''(p) = \frac{N-x}{(1-p)^2} + \frac{x}{p^2}$$

- Thus the MLE for p, $p_0 = x/N$ is asymptotically normal with variance approximately equal to $p_0(1 - p_0)/N$.

13 Milliman

Ultimate Frequentist Method – Bootstrap

- Bootstrap is a non-parametric method to assess volatility in a model fit that essentially assumes that observed values are completely representative of the distribution of interest
- Typical a statistical model assumes the data follows some sort of pattern (often called signal), with some random variation (often called noise)
- Approaches like MLE will use statistical principals to assess the noise given a particular underlying statistical model that usually includes an assumption about the nature of the noise
- Bootstrap assumes that the distribution of noise can be completely modeled by sampling with replacement from the observed errors

14 Milliman

Ultimate Frequentist Method – Bootstrap

- In our coin example we would be interested only in the noise (that is no specific underlying "signal" model)
- If we have 3 observed heads in 3 tosses then the bootstrap simulation of the noise would be approximated by sampling with replacement from the 3 observations
- In the case of 3 out of 3 heads the bootstrap would imply there is no chance of tails
- If the sample size is large and completely representative then bootstrap can be a very powerful tool, particularly if the true error structure is difficult to estimate of simulate

15 Milliman

Are We Frequentists?

- Let's do a thought experiment
- I offer you a game where I will pay you $0.90 if the toss of a coin comes up heads and you pay me $1.10 if it comes up tails
- I let you toss the coin 3 times and it comes up heads all three times
- Will you play the game?
- A true frequentist would only believe the result of the 3 throws, other information is irrelevant.
- With only heads coming out the true frequentist would undoubtedly play (and with a fair coin) lose in the long run.
- Would any of you take this game?

16 Milliman

Frequentist or Other?

- Would you like to know where I got the coin we use?
- If I said "In change at the grocery store" you might be pretty sure that p = 0.50
- If you said "Joe gave it to me" you might still guess p = 0.50 but not be as certain
- If Joe was well-known to be a prankster your uncertainty would be probably be substantially greater
- In all these situations you do not restrict yourself to simply the observed data but consider prior experience with the situation or similar ones
- If you believe your past experience is worth considering then you might just be a Bayesian

17 Milliman

Frequentist or Bayesian

- The frequentist essentially assumes that the parameter(s) of the distribution of interest are fixed but unknown, to be "divined" from the data observed.
- MLE is often used by frequentists since for large independent samples the MLE for a parameter tends to the "true" value of the parameter
- Note that the MLE is determined only by the observed data and (assumed) model structure
- This is fine for experiments that can be repeated so we can get "large enough" samples
- Also assumes samples are independent and all from the same distribution

18 Milliman

The Bayesian Point of View

- The Bayesian has a different view of the world
- As the Frequentist, the Bayesian assumes that a parameter is unknown. But unlike the Frequentist the Bayesian does not believe there is a single "true" value of the unknown parameter to be "divined" but rather that the parameter itself has its own distribution which is also unknown
- This unknown distribution is usually called the "prior distribution" of the parameter
- The Bayesian will use observed data to modify or "evolve" the assumed distribution of the parameter
- This conditioned distribution is know as the "posterior distribution" conditioned by the data

19 Milliman

Why is it Called Bayesian?

- Bayesian analysis makes extensive use of Bayes Theorem, which for discrete probabilities may be written as

$$\mathrm{P}(B|A) = \frac{\mathrm{P}(A|B)\mathrm{P}(B)}{\mathrm{P}(A)}$$

- Generally our stochastic models will assume a certain distribution of observations given the value of an unknown parameter. What is of interest though is how a set of observations will change the assessment of the distribution of the parameter
- Bayes Theorem gives a way to do this
- Then given a revised assessment of the distribution of the parameter (messy) calculus gives the distribution of values

20 Milliman

Look at the Bayesian Take on the Coin

- If the Bayesian has no prior experience with coins (s)he might have a prior distribution of the parameter *p* that is as likely to be any value on the unit interval as any other, i.e. a "uniform prior"
- After 3 tosses come up heads the Bayesian would modify this distribution to be a Beta distribution, B(4,1). Where the frequentist might say the *p* parameter is 1.00 with certainty, the Bayesian allows for considerable uncertainty, though the expected value is 0.80.
- The Bayesian can easily adapt to the other situations presented as well

21 Milliman

Look at the Bayesian Take on the Coin

- In the case of the coin from the grocery store the prior would be highly concentrated around 0.50 since our experience is that a random coin is very close to fair, say represented by a beta B(1000,1000).
- In this case the three heads changes the expected value of the probability of heads parameter from 0.50 to 0.5007
- With Joe the trickster we might have a different view, still with an expected value of 0.50 but less uncertainty, say a beta B(10,10)
- In this case the 3 heads changes the expected value of the probability of heads parameter to 0.5652.
- Sounds a whole lot like what actuaries do in practice

22 Milliman

What is the Sound Of One Point Clapping?

- Suppose you have only one data point and one parameter
- What can that tell you from a frequentist point of view?
- Not a heck of a lot – pretty much useless
- Suppose you have two data points and two parameters – again not much
- How about 2 data points and 3 parameters – now you really are in trouble from a frequentist point of view
- How about to a Bayesian?
- ANY data is useful
- Sparse data may not change the prior much, but is reflected in the posterior

23 Milliman

Computational Considerations – MLEs

- Many situations result in MLEs that can be expressed in closed form in terms of the data observations
- This was the limit of early use of MLEs to "well behaved" situations including linear and generalized linear models
- Computational advances, particularly in optimization routines in packages like R, MATLAB and others provide a useful tool in moving MLEs beyond (generalized) linear models
- However whether linear or generalized linear, MLEs only have the desirable properties "asymptotically" or for "large samples"

24 Milliman

Computational Considerations – Bayesian

- It turns out that if you have a Bernoulli distribution with probability p which is unknown but assumed to have a Beta distribution $B(\alpha,\beta)$ and then observe x heads in n tosses of the coin then the posterior distribution for p is beta $B(\alpha + x,\beta + n - x)$.
- The number of distributions where we can get a closed form posterior is rather limited
- The calculus required in most other situations gets difficult to impossible very quickly
- As with MLEs technology now has come to the rescue in the form of Markov Chain Monte Carlo (MCMC) methods
- Essentially MCMC uses Monte Carlo sampling to approximate the posterior distribution

25 Milliman

Reserves in a Stochastic World

- At a point in time (valuation date) there is a range of possible outcomes for a book of (insurance) liabilities. Some possible outcomes may be more likely than others
- Range of possible outcomes along with their corresponding probabilities are the distribution of outcomes for the book of liabilities – i.e. reserves are a distribution
- The distribution of outcomes may be complex and not completely understood
- Uncertainty in predicting outcomes comes from
  - Process (pure randomness)
  - Parameters (model parameters uncertain)
  - Model (selected model is not perfectly correct)

26 Milliman

Basic Traditional Actuarial Methods

- Traditional actuarial methods are simplifications of reality
  - Chain ladder
  - Bornhuetter-Ferguson or it's close relative Cape Cod
  - Berquist-Sherman Incremental Average
  - Others
- Usually quite simple thereby "easy" to explain
- Traditional reserve approaches rely on a number of methods – model uncertainty explicitly addressed
- Practitioner "selects" an "estimate" based on results of several traditional methods
- No explicit probabilistic component

27 Milliman

Stochastic Models

- In the actuarial context a stochastic model could be considered as a mathematical simplification of an underlying loss process with an explicit statement of underlying probabilities
- Two main features
  - Simplified Statement
  - Explicit probabilistic statement
- In terms of sources of uncertainty two of three sources may be addressed
  - Process
  - Parameter
- Within a single model, the third source (model uncertainty) usually not explicitly addressed

28 Milliman

Reserving Context – Usual Triangles

- In reserving we are faced with the problem of "squaring the triangle"
- Suppose C_{ij} is the amount paid for accident year *i* during year *j*, counting from the start of the accident year
- Keeping things simple, if we have 10 years of experience at annual valuations, we "see" 55 historical points C_{ij}, for *i* running from 1 to 10, and *j* running from 1 to 10 – *i* + 1
- Name of the game is to estimate the remaining 45 values of C_{ij} for *j* running from 10 – *i* to 10

29 Milliman

Traditional Methods

- Traditionally actuaries have relied on a number of methods to "square the triangle"
- Essentially the Bornhuetter-Ferguson method assumes $C_{ij} = \alpha_i \beta_j$ with restrictions on some parameter values to keep the problem well posed, leading to 19 parameters for a 10 x 10 triangle
- The Berquist-Sherman is a special case of the Bornhuetter-Ferguson with a smooth trend, $C_{ij} = \alpha_0 \tau^i \beta_j$ and a total of 11 parameters for a 10 x 10 triangle
- The chain ladder can be seen as another special case of the Bornhuetter-Ferguson, imposing the requirement that expected totals to date match historical total to date which can be parameterized with 9 parameters.

30 Milliman

Stochastic Versions of Traditional Methods

- Note that in each of the traditional methods each of the incremental amounts C_{ij} can be written as a function $g_{ij}(\boldsymbol{\theta})$ of some parameter vector $\boldsymbol{\theta}$
- Other methods can also be written down in a similar fashion, not just the usual simple traditional methods
- This is the first step – a simplified view of reality
- To make this a stochastic problem we need to make some statement about the distribution of the C_{ij} amounts, for example that they have probability density functions that may themselves depend on additional parameters say $f_{ij}(x|\boldsymbol{\theta},\boldsymbol{\gamma})$

31 Milliman

MLE in Reserve Applications

- In this framework, the negative log likelihood function for the values observed in the triangle becomes

$$\ell\left(C_{ij}|\boldsymbol{\theta},\boldsymbol{\gamma}\right) = -\sum_{i=1}^{10}\sum_{j=1}^{10-i+1} \ln f_{ij}\left(C_{ij}|\boldsymbol{\theta},\boldsymbol{\gamma}\right)$$

- If we find values of the parameter vectors $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$ that minimize this negative log likelihood (equivalent to maximizing the likelihood itself) we have estimates for the parameters for the model
- If we are willing to assume we have sufficient "replications" to bring in the asymptotic properties of MLEs we can also say something about the distribution of those parameters

32 Milliman

Forecast Distributions with MLEs

- Once we have (estimates of) parameters the model selected gives us distributions in each cell, past and forecast
- Can use Monte Carlo simulation to estimate process uncertainty in projections
- Assuming asymptotic normality of the MLEs we can also estimate distribution of the parameters, (multivariate) normal with mean (vector) equal to the MLE and variance (-covariance) matrix derived from information matrix
- Can use the latter to simulate parameters and then the parameters to estimate outcome distribution
- As the shampoo label says, “rinse, repeat.”
- In contrast to bootstrap, values outside observed range possible

33 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	2,468	2,838	3,004	3,056	3,133	3,141	3,180	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	3,416	3,507	3,529			41,601
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	3,078	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

34 Milliman

Results

Model	Expected Reserves (000,000)
Berquist Incremental Severity	$480
Cape Cod	391
Wright Model	388
Generalized Hoerl Curve	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

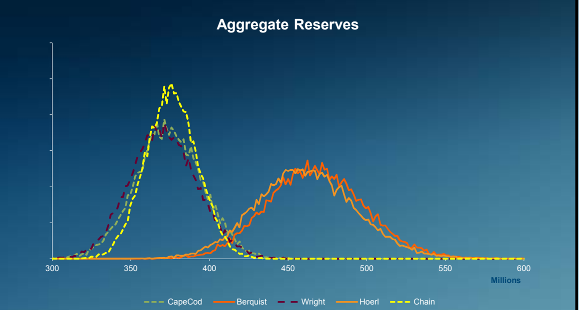
35 Milliman

Process vs. Parameter Uncertainty

Model	Total Reserve Process Std. Dev. (000)	Total Reserve Total Std. Dev. (000)
Berquist Incremental Severity	$15,998	$29,090
Cape Cod	9,435	20,298
Wright Model	10,029	20,375
Generalized Hoerl Curve	16,115	29,728
Chain Ladder	9,448	15,704

36 Milliman

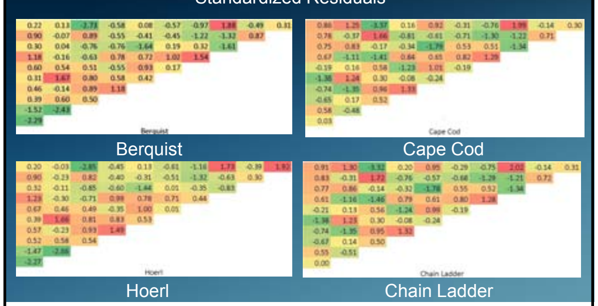
Reserve Forecasts by Model



37 Milliman

What Happened?

Standardized Residuals



38 Milliman

Berquist-Sherman, Bayesian Style

- Now for a different take on things.
- Keep the same basic model that we used for the MLE estimates
- Instead of using a Frequentist approach assume we have do not have much prior information about the parameters and assume
  - The incremental averages by age are uniform on (-5000,10000)
  - The trend is uniform on (-0.25,0.25)
  - The kappa (variance constant) is uniform on (0,100), and
  - The variance power is uniform on (0,5)
- We can now estimate of the distribution of reserves without any additional data (with θ representing all parameters and π the distribution of parameters)

$$g(x) = \int f(x \mid \boldsymbol{\theta}) \pi(\boldsymbol{\theta}) d\boldsymbol{\theta}$$

39 Milliman

Bayesian Berquist-Sherman

- In the prior slice f(x|θ) denotes the distribution of the reserves given values for the parameter (vector) θ.
- Suppose we have observations for the C_{ij} values.
- If π(θ) denotes the prior assumed distribution of the parameter (vector) then the distribution conditioned on observing the C_{ij} values then we can use Bayes Theorem to conclude

$$\pi(\boldsymbol{\theta} \mid C_{ij}) \propto f(C_{ij} \mid \boldsymbol{\theta}) \pi(\boldsymbol{\theta})$$

- We now can use this refined estimate of our parameter uncertainty to get a refined estimate of our reserve distribution

$$g(x \mid C_{ij}) \propto \int f(x \mid \boldsymbol{\theta}) \pi(\boldsymbol{\theta} \mid C_{ij}) d\boldsymbol{\theta}$$

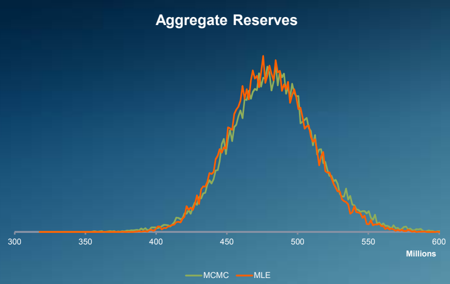
40 Milliman

Enter MCMC

- Once we know the distribution we can get any statistic we want to consider about reserves, their expected value, median, percentiles, etc.
- However we are faced with the problem that this last integral is easy to write down but not that easy to calculate
- This is were Markov Chain Monte Carlo methods can help out
- The integral is the expected value of a function (f) given a probability distribution (π)
- What MCMC does is to sample from the distribution π and use that sample to approximate the value of the integral of interest
- Although complex there are a number of tools to do this in practice (WinBugs, JAGS, etc.)

41 Milliman

Berquist-Sherman – MLE vs. MCMC



42

Berquist-Sherman – MLE vs. MCMC

Percentile	MLE	MCMC
0.5%	409	397
1.0%	416	407
2.5%	425	422
5.0%	434	432
10.0%	443	444
25.0%	460	462
50.0%	480	482
75.0%	499	502
90.0%	518	522
95.0%	529	536
97.5%	540	547
99.0%	551	562
99.5%	559	575

43

Berquist-Sherman Trend - MCMC

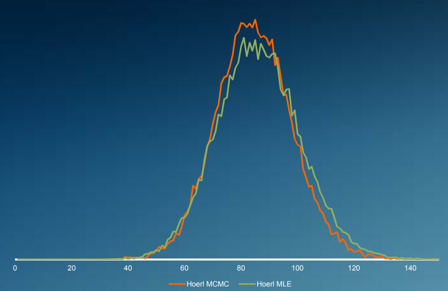
44

Berquist-Sherman – MLE vs. MCMC

- The two methods are reasonably close
- In general MCMC seems to have a slightly broader distribution of results than the MLE approach
- Some possible explanations
  - MLE results assume sufficient data for asymptotic properties to come into play. There is no assurance that is the case
  - MLE assumes the parameters are jointly normal
  - MCMC starts with a pretty broad diffuse prior
  - Closer comparison of parameter estimates show more variation in the "tail" where data are sparse. Bayesian methods are not as strongly influenced by sparsity of data as MLE methods.
- Thanks to Glenn Meyers for the MCMC results here.

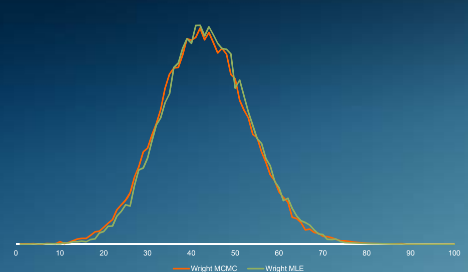
45 Milliman

Hoerl – MLE vs. MCMC



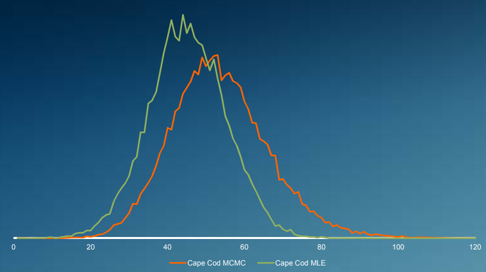
46 Milliman

Wright – MLE vs. MCMC



47 Milliman

Cape Cod – MLE vs. MCMC



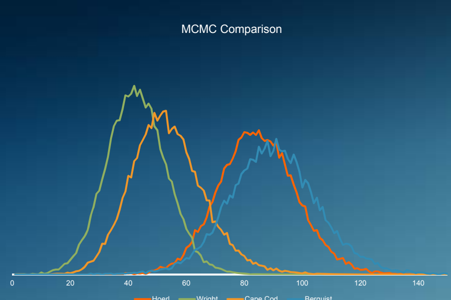
48

Other Models – MLE vs. MCMC

- The indicated distributions for both MLE and MCMC for the Hoerl Curve and Wright models are both reasonably close
- There is a marked difference in the Cape Cod model
- This is the model with the most parameters (21 in the case of the 10 x 10 triangle) and many estimated using very few points
- Here the question of whether or not there is sufficiently large sample size for the MLE’s asymptotic properties to come into play is probably greatest
- As one might expect the MCMC distribution is noticeably wider

49

MCMC Summary



50

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

51 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

52 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

53 Milliman

APPENDIX

- The following slides, not formally presented provide details behind the models used in this presentation.
- The models used here will be presented in greater detail in "A Flexible Framework for Stochastic Reserving Models," in Volume 7, Issue 2 of *Variance* scheduled to be published late 2013 or early 2014.
- See also the CAS Monograph *STOCHASTIC LOSS RESERVING USING BAYESIAN MCMC MODELS* by Glenn Meyers

54

A Stochastic Framework

- Instead of incremental paid, consider incremental average $A_{ij} = C_{ij}/E_{ij}$
- The amounts are averages of a (large?) sample, assumed from the same population
- Law of large numbers would imply, if variance is finite, that distribution of the average is asymptotically normal
- Thus assume the averages have Gaussian distributions (next step in stochastic framework)
- Note here we have not specified which of the above traditional methods we are considering

55

A Stochastic Incremental Model – Cont.

- Now that we have an assumption about the distribution (Gaussian) and expected value all needed to specify the model is the variance in each cell
- In stochastic chain ladder frameworks the variance is assumed to be a fixed (known) power of the mean

$$\mathrm{Var}\left(C_{ij}\right) = \sigma \mathrm{E}\left(C_{ij}\right)^{k}$$

- We will follow this general structure, however allowing the averages to be negative and the power to be a parameter fit from the data, reflecting the sample size for the various sums

$$\mathrm{Var}\left(A_{ij}\right) = e^{\kappa - e_i}\left(\mathrm{E}\left(A_{ij}\right)^{2}\right)^{p}$$

56

An Observation on the Methods

- Each of the five traditional methods can be expressed as a function of a number of parameters

$$C_{ij} = g_{ij}(\boldsymbol{\theta})$$

- Here **θ** represents a vector of the parameters with different lengths for different models
- Instead of specifying a particular method now we will talk in terms of a general method where the incremental amounts can be expressed as a function of a vector of parameters
- For the stochastic version we assume

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

57

Parameter Estimation

- Number of approaches possible
- If we have an a-priori estimate of the distribution of the parameters we could use Bayes Theorem to refine those estimates given the data
- Maximum likelihood is another approach
- In this case the negative log likelihood function of the observations given a set of parameters is given by

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

58

Distribution of Outcomes Under Model

- 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

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

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

- This is the estimate for the average future forecast payment per unit of exposure, multiplying by exposures
- This assumes parameter estimates are correct – does not account for parameter uncertainty

59

The Information Matrix

- Key to calculating the variance-covariance matrix for the parameter estimates is calculating the Fisher Information Matrix
- Recall the negative log likelihood function is a function of the parameters $\boldsymbol{\theta}$, κ, and p.

$$\ell\left(A_{11}, A_{12}, \ldots, A_{n1}; \boldsymbol{\theta}, \kappa, p\right) =$$

$$\sum \frac{\kappa - e_i + \ln\left(2\pi\left(g_{ij}\left(\boldsymbol{\theta}\right)^2\right)^p\right)}{2} + \frac{\left(A_{ij} - g_{ij}\left(\boldsymbol{\theta}\right)\right)^2}{2e^{\kappa - e_i}\left(g_{ij}\left(\boldsymbol{\theta}\right)^2\right)^p}$$

- So the Hessian and hence its expected value is a function of the parameters κ and p, as well as the partial derivatives of g_{ij} with respect to the θ parameters otherwise independent of g_{ij}

60 Milliman

Incorporating Parameter Uncertainty

- If we assume
  - The parameters have a multi-variate Gaussian distribution with mean equal to the maximum likelihood estimators and variance-covariance matrix equal to the inverse of the Fisher information matrix
  - For fixed parameters the losses have a Gaussian distribution with the mean and variance the given functions of the parameters
- The posterior distribution of outcomes is rather complex
- Can be easily simulated:
  - First randomly select parameters from a multi-variate Gaussian Distribution
  - For these parameters simulate losses from the appropriate Gaussian distributions

61 Milliman

Parameterization – Cape Cod

- Simple parameterization for the Cape Cod above overspecifies the model
- We use the following (similar to England & Verall)

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

- θ_1 is the upper left corner incremental
- θ_i for i = 2, …, n is change in incremental from accident year i-1 to age i
- θ_i for i = n+1, …, m+n-1is change from age i – n to accident year i – n +1

62 Milliman

Parameterization – Berquist-Sherman

- Actually a special case of the Cape Cod
- Replace the accident year change parameters by trend

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

- θ_j for j = 1, …, 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

63

Parameterization – Wright & Hoerl Models

- Actually a special cases of the Cape Cod
- For Wright replace the development year parameters by a curve

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

- θ_i for i = 1, …, m sets the accident year loss level
- Emergence defined by 3-parameter curve
- Hoerl model replaces separate accident year levels with trend from above

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

64

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.

65

Parameterization – Chain Ladder (Continued)

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

66 Milliman