A Stochastic Reserving Today (Beyond Bootstrap)

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

Casualty Loss Reserve Seminar
6-7 September 2012
Denver, CO

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 ï expressed or implied ï 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.

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 ï i.e. reserves are a distribution
- The distribution of outcomes may be complex and not completely understood
- Uncertainty in predicting outcomes comes from
  - ï Process (pure randomness)
  - ï Parameters (model parameters uncertain)
  - ï Model (selected model is not perfectly correct)

Basic Traditional Actuarial Methods

- Traditional actuarial methods are simplifications of reality
  - Chain ladder
  - Bornhuetter-Ferguson or itôs close relative Cape Cod
  - Berquist-Sherman Incremental Average
  - Others
- Usually quite simple thereby ñeasyò to explain
- Traditional reserve approaches rely on a number of methods ï model uncertainty explicitly addressed
- Practitioner ñselectsò an ñestimateò based on results of several traditional methods
- No explicit probabilistic component

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
  - ï Simplified Statement
  - ï Explicit probabilistic statement
- In terms of sources of uncertainty two of three sources may be addressed
  - ï Process
  - ï Parameter
- Within a single model, the third source (model uncertainty) usually not explicitly addressed

Milliman

A Simple Stochastic Model

- Flip of a fair coin
  - ï Simplified view of reality ï a process that generates only two possible outcomes, called H and T.
  - ï Stochastic statement ï the two outcomes are equally likely
  - ï Only source of uncertainty in next trial is random (process)
- Flip of an uncertain coin
  - ï Same simplified view of reality
  - ï Stochastic statement, the chance of H is *p*, and the chance of T is 1 ï *p*. *p* is unknown but fixed
  - ï Here two sources of uncertainty in next trial, random still exists, but we are also uncertain about *p,* so also have parameter uncertainty
  - ï Without additional information nothing more can be said

A Bayesian Digression

- At this point a Bayesian would ask where you got the coin
- If you said ñIn change at the grocery storeò the Bayesian may be pretty sure that p = 0.50
- If you said ñJoe gave it to meò the Bayesian may still guess p = 0.50 but not be as certain
- If Joe was well-known to be a prankster the Bayesianõs uncertainty would be probably be substantially greater
- The Bayesian does not restrict him/herself to simply the observed data but considers prior experience with the situation or similar ones
- The Bayesian, though allows his/her beliefs to be modified in the presence of additional data

Estimating *p*

- How to estimate *p*?
- What can we go on?
- Nothing so far gives any clue.
- Suppose we can actually flip the subject coin a number (*N*) of independent times and observe the number (*x*) of heads and *N* – x tails (that is, the experiment is replicable)
- One estimate of *p* would then be *x*/*N*.
- Is this estimate any good?
- Given this information is there another estimate that is in some way ñbetter?ò
- What is the ñbestò estimate in some sense?

Maximum Likelihood

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

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

- 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 ñexperimentò
- 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

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

MLE Example

- In this coin example the negative log likelihood is

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

- 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 $p_0(1 - p_0)/N$.

Milliman

Reserving Context – Usual Triangles

- In reserving we are faced with the problem of ñsquaring the triangleò
- 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 ñseeò 55 historical points C_{ij}, for i running from 1 to 10, and j running from 1 to 10 ï i + 1
- Name of the game is to estimate the remaining 45 values of C_{ij} for j running from 10 ï i to 10

Traditional Methods

- Traditionally actuaries have relied on a number of methods to ñsquare the triangleò
- 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.

Stochastic Versions of Traditional Methods

- Note that in each of the traditional methods each of the incremental amounts C_{ij} can be written 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 ï 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 $f_{ij}(x|\boldsymbol{\theta},\boldsymbol{\gamma})$

Milliman

MLE in Reserve Applications

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

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

- If we find values of the parameter vectors **θ** and **γ** 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 ñreplicationsòto bring in the asymptotic properties of MLEs we can also say something about the distribution of those parameters

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, ñrinse, repeat.ò
- In contrast to bootstrap, values outside observed range possible

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

Milliman

Results

Model	Expected Reserves (000,000)
Berquist Incremental Severity	$480
Cape Cod	391
Generalized Hoerl Curve	474
Chain Ladder	393

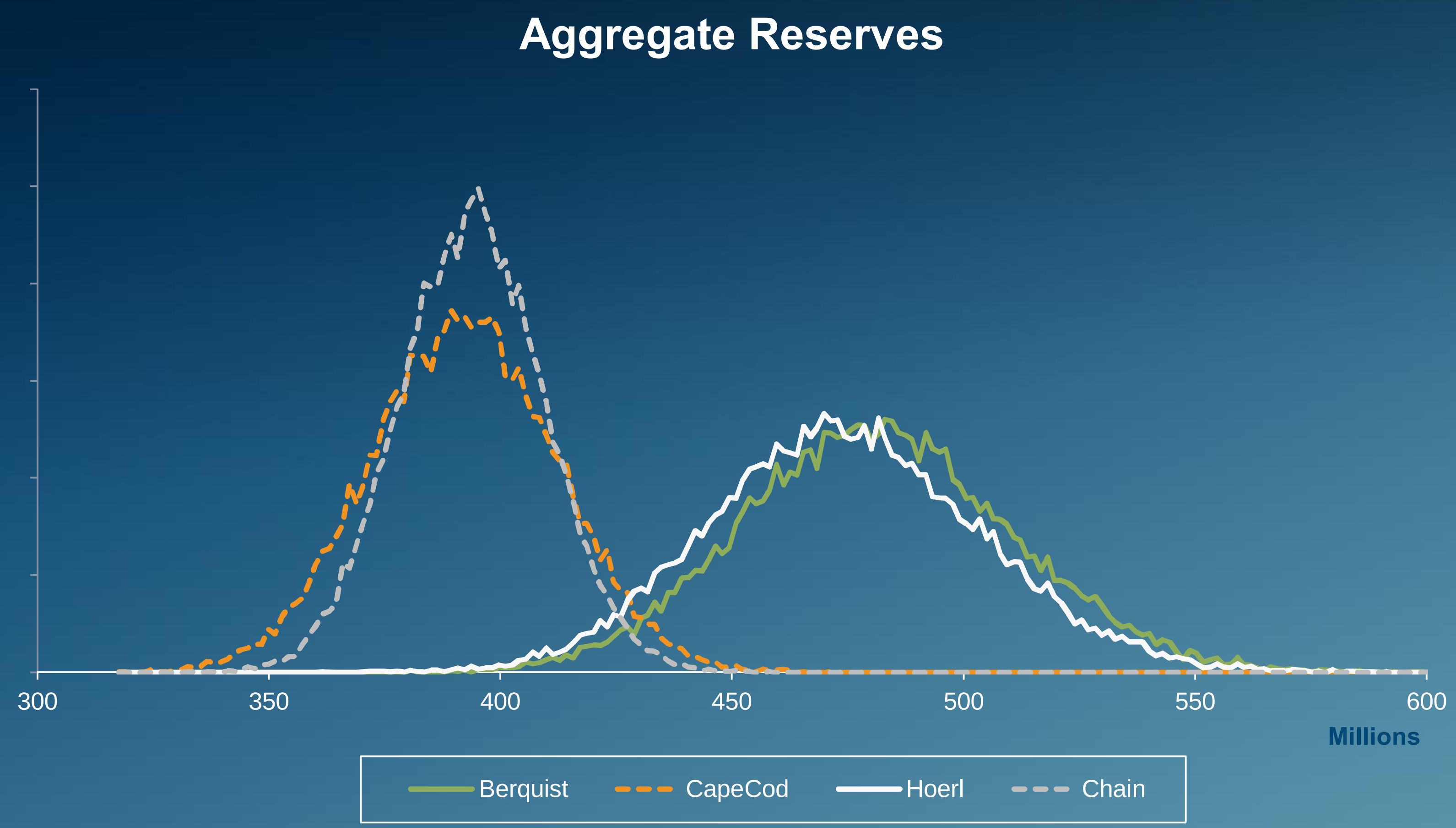
- Some difference in expected reserves
- Is the difference random?
- Is the difference significant?
- How do you know?
- Stochastic models help answer these questions

Milliman

Process vs. Parameter Uncertainty

Model	Total Reserve Process Std. Dev. (000)	Total Reserve Total Std. Dev. (000)
Berquist Incremental Severity	$15,997	$29,405
Cape Cod	9,435	20,101
Generalized Hoerl Curve	16,115	29,454
Chain Ladder	9,447	15,557

Reserve Forecasts by Model



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

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 ï possibly greater than the other two sources

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

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 ñexperimentsò not repeatable so frequentist approach (MLE) may not be appropriate

Milliman

APPENDIX

- The following slides, not formally presented provide details behind the models used in this presentation.

Milliman

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

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

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

- 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

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

Milliman

An Observation on the Methods

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

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

- Here $\boldsymbol{\theta}$ 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

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

Milliman

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

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

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

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

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

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

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

Milliman

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 **θ**, *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}

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

Parameterization – Cape Cod

- Simple parameterization for the Cape Cod above 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, é , n is change in incremental from accident year *i-1* to age *i*
- θ_i for i = n+1, é , m+n-1is change from age *i* ï *n* to accident year *i* – *n* +1

Milliman

Parameterization – Berquist-Sherman & Surface Models

- 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, é , 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
- Parameterization of surface model is unchanged from above

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

Milliman

Parameterization – Chain Ladder

- Basic requirements for expected values
  - ï Ratio of cumulative averages from one age to the next same for all accident years
  - ï 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.

Parameterization – Chain Ladder (Continued)

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