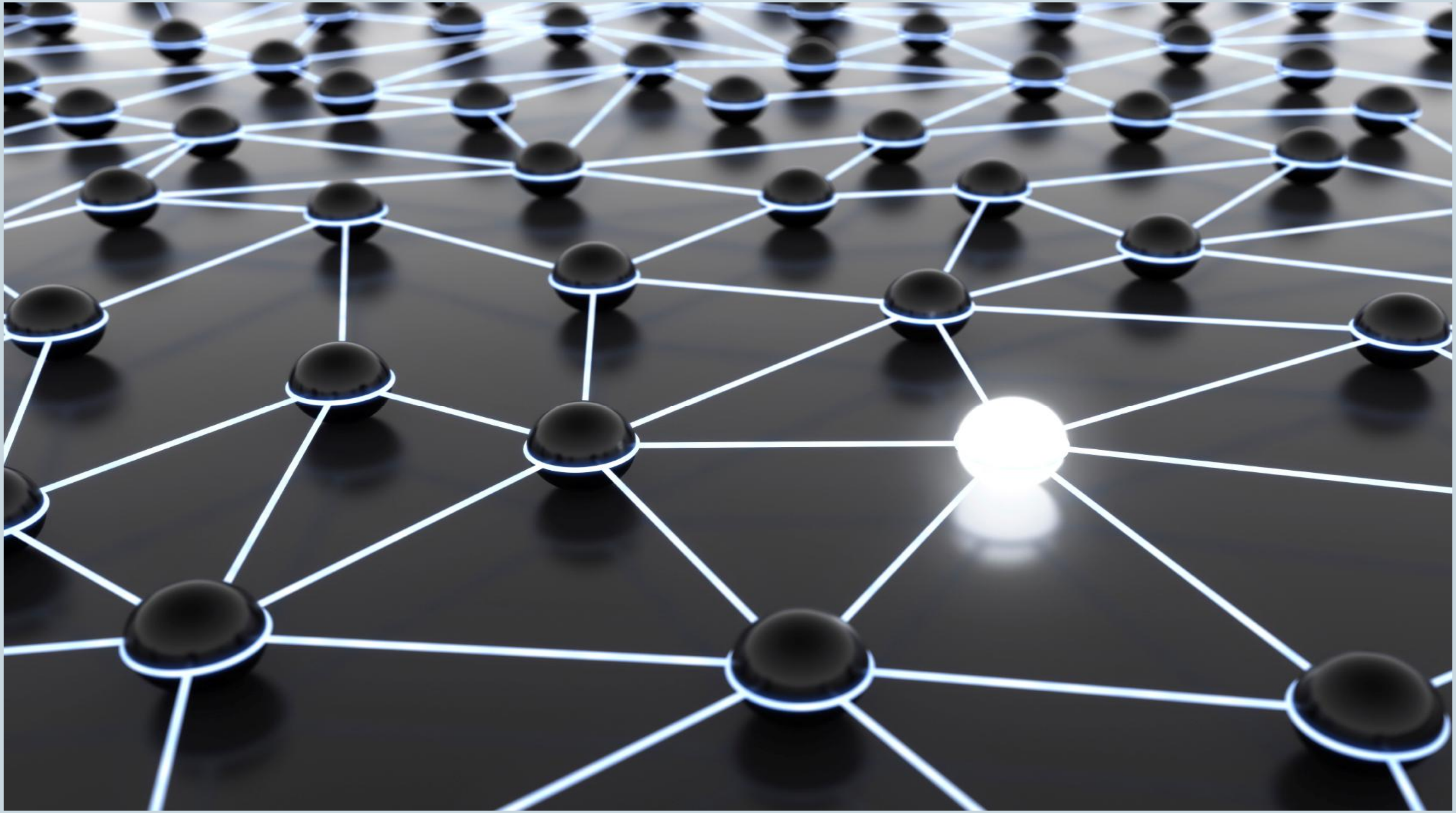


CLRS Meeting — Chicago

September 19-20, 2016

GLMs and Bayesian Models

By: Manolis Bardis

Willis Towers Watson

Overview

Today's presentation will cover the following:

Aggregate Generalized Linear Models

I. General Introduction to Method

II. GLM Basics

III. GLM Reserving Example

IV. Conclusion

Individual Claim Reserving

V. Predictive Modeling Overview

VI. Traditional Reserving Development Methods

VII. Reserving with Predictive Modeling

VIII. Aggregate Reserving Methods

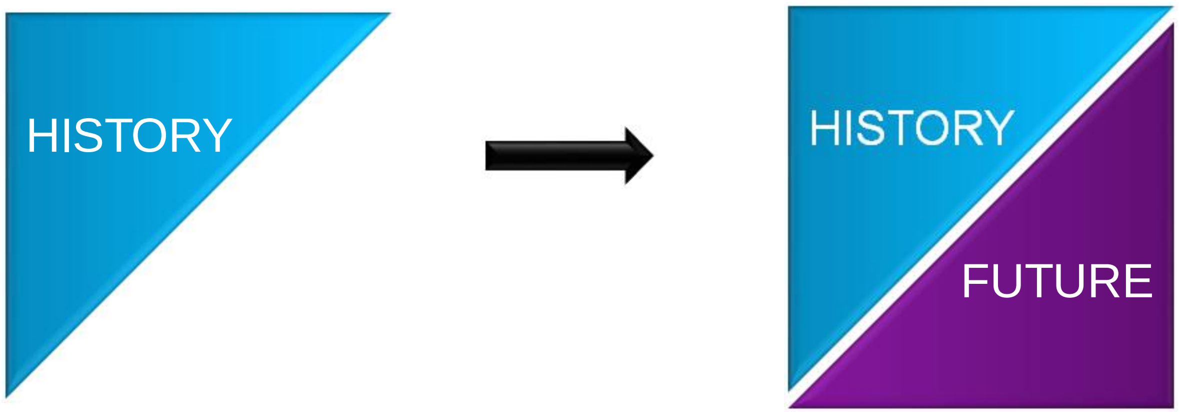
IX. Individual Claim Reserving Methods

Generalized Linear Models

I. General Introduction to Method

Actuarial Reserving in a Nutshell

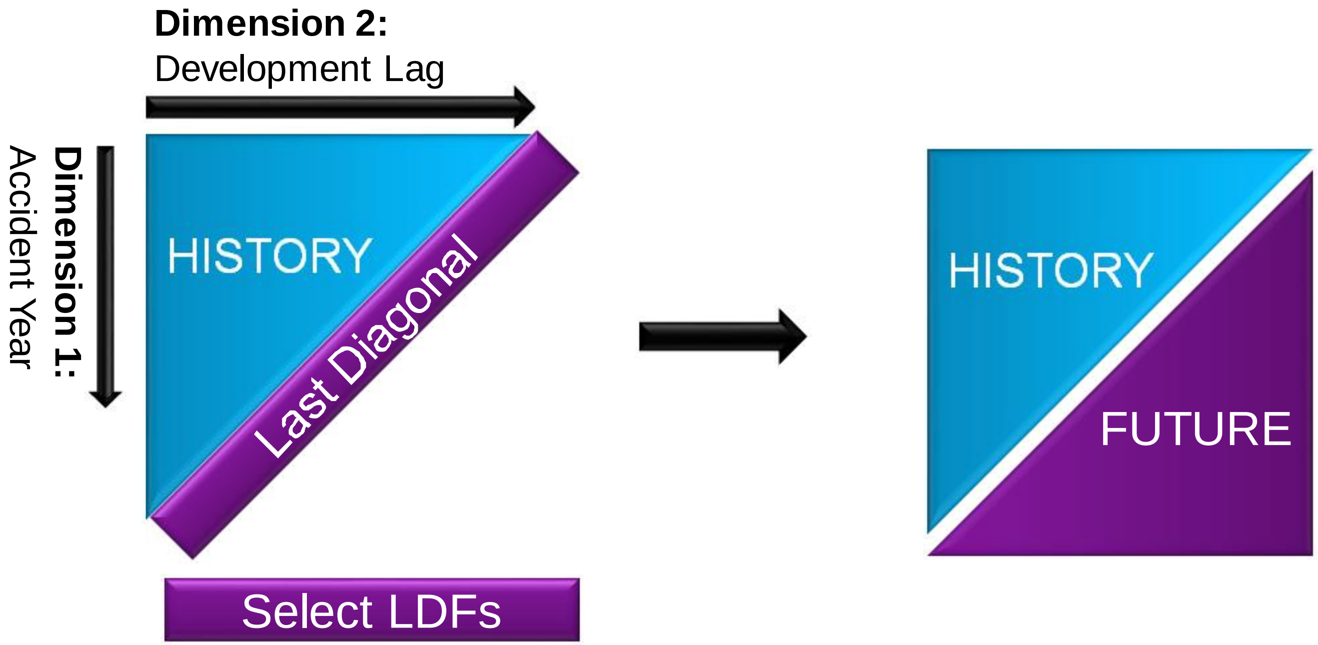
- Traditional actuarial reserving methods has been conceptually described as a process of squaring up a triangle:



- The GLM Reserve method is no different. Estimate future results based on information from historical.

Why GLM ?

- Traditional Chain Ladder method focuses on the development Lag dimension to derive estimates:



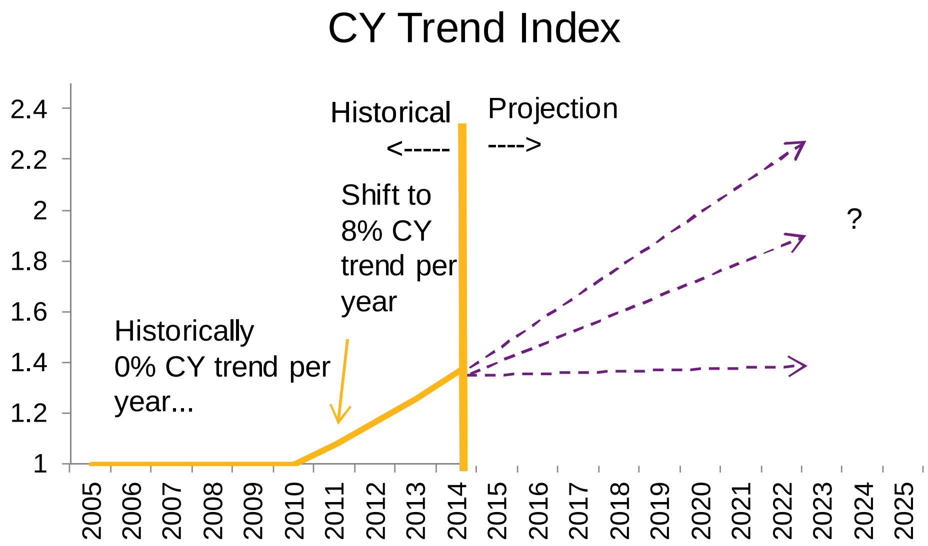
- Each future estimate can be derived based on the selected development factors.

Why GLM ?

- However, one major limitation with chain ladder is that it does not adjust for accident or calendar year effects
- Examples include:
  - New claims handling process
  - Changing settlement pattern
  - Legislative/Regulatory changes
- GLM Reserving allows us to introduce two additional dimensions
  - Dimension 1: Accident Year
  - Dimension 2: Development Lag
  - Dimension 3: Calendar Year

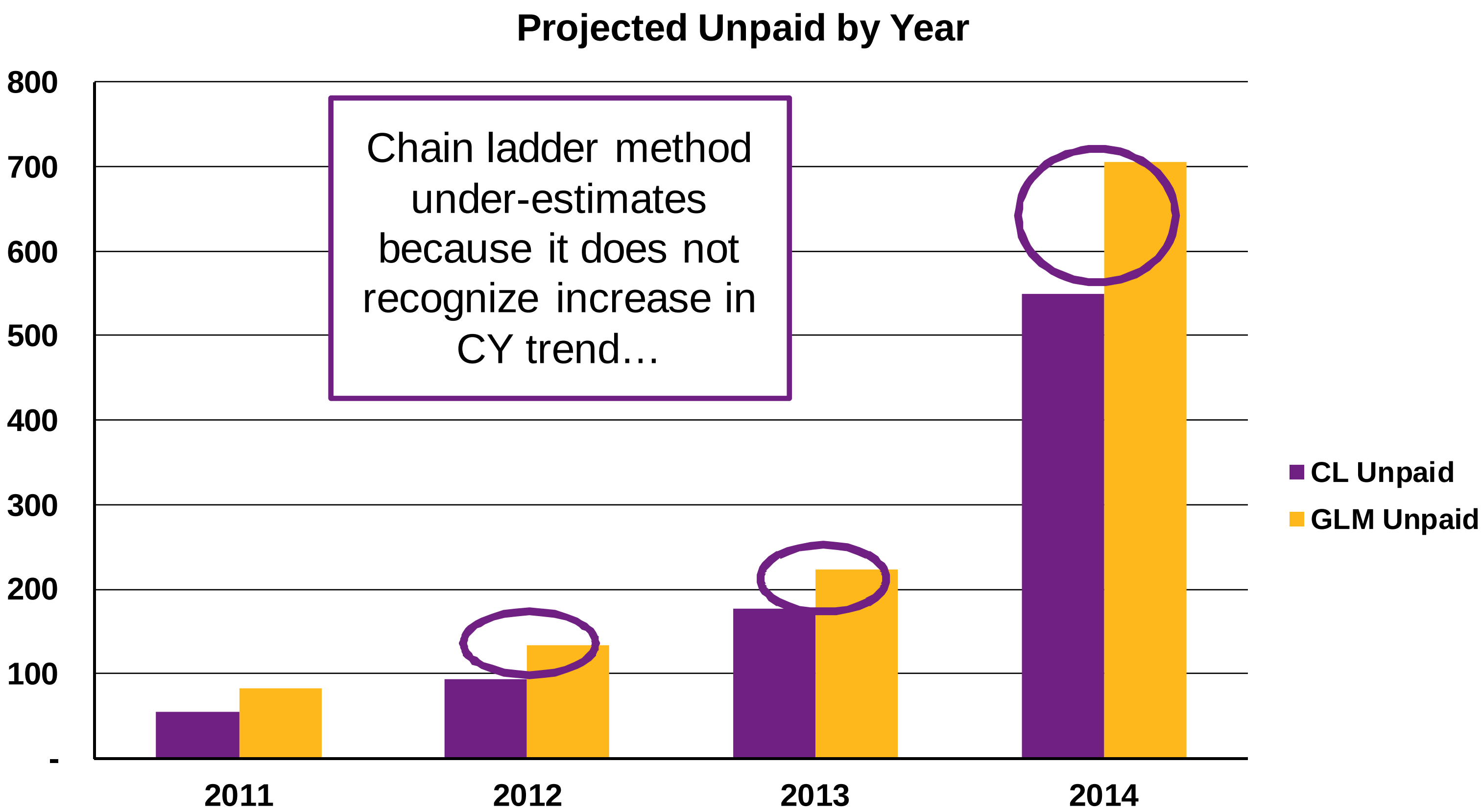
Case Study Example

- Let's quickly go through an illustrative example to demonstrate the impact of calendar year effects using a chain ladder method vs GLM reserving method
- Case Study introduces a calendar year trend in the most recent periods



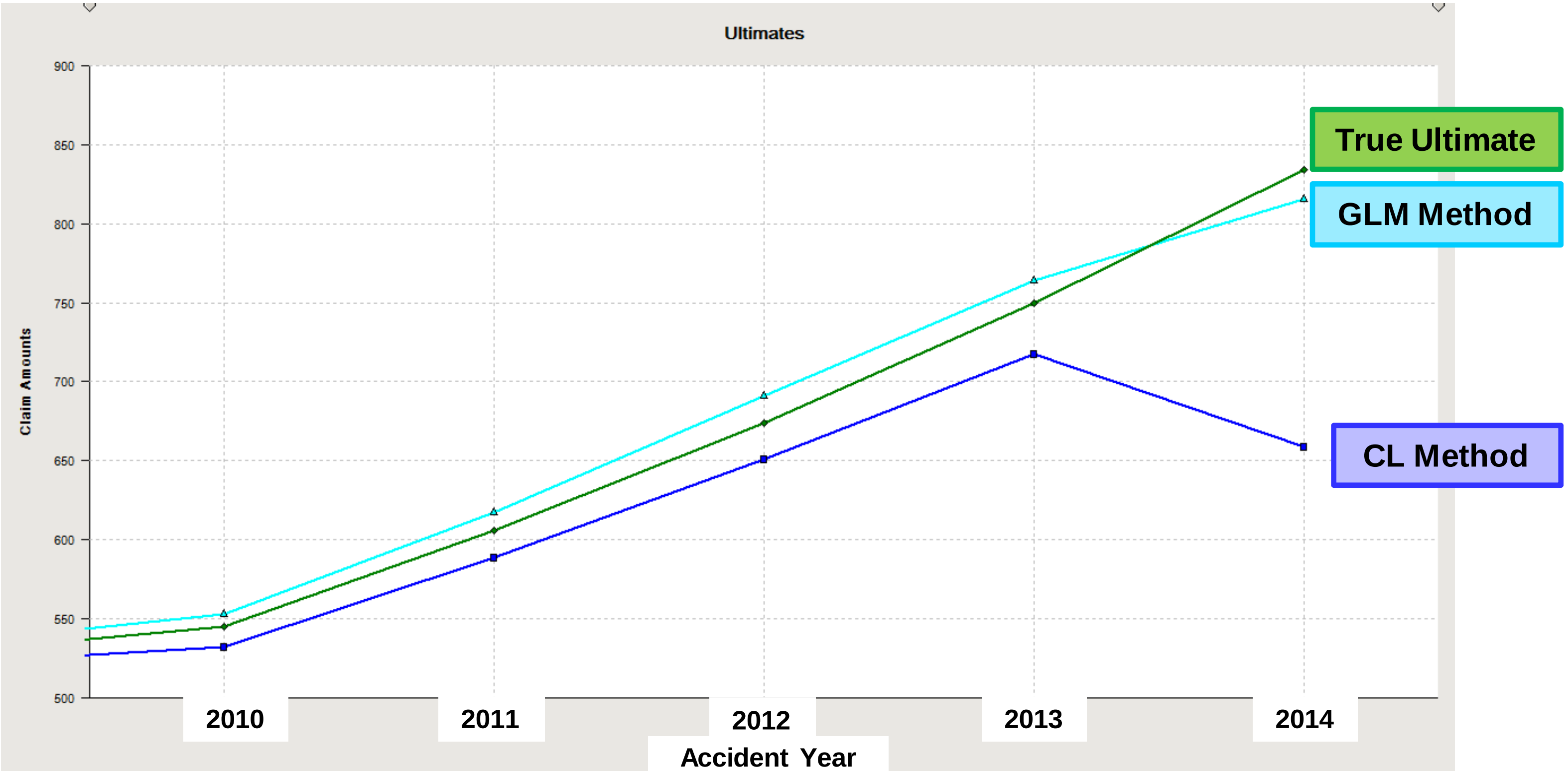
Case Study Example

- Comparing results for GLM Reserving vs. Chain Ladder



Case Study Example

- Impact can be *significant*. In this example, the difference from unpaid is only 4% for GLM Method versus -22% difference for Chain Ladder



- Improved estimates

Generalized Linear Models

II. GLM Basics

Section Introduction

- Overview of Predictive Models
- Explaining the GLM Framework
- Basic GLM Example

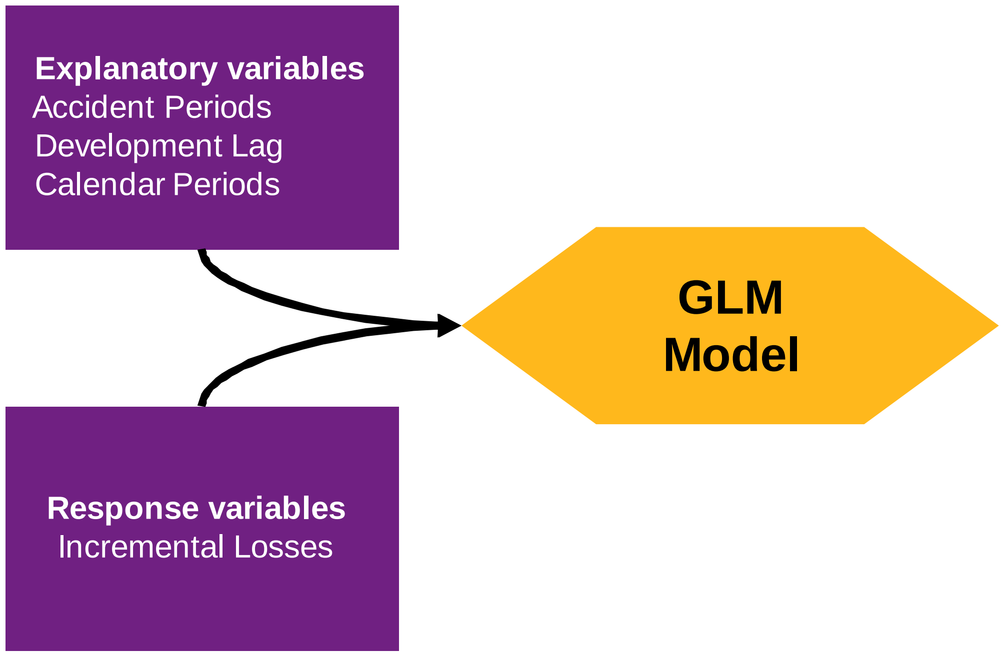
Before going into the GLM Reserve Method, we will cover some basic GLM concepts that will help us down the road…

Predictive Models

- Multivariate statistical model to predict a response variable using a series of explanatory variables

Explanatory variables
- Variables that help **explain** what we are seeing

Response variables
- Variables that **respond** to explanatory variables and we are trying to predict



- We will use the explanatory variables to try and explain the behavior of incremental losses

Practical User Considerations
Selecting a Link Function & Error Structure

Options for Error Structure

Normal or Gamma

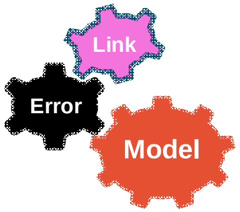
- Normal distribution assumes that all observations have the same fixed variance
- Gamma distribution assumes that the variance increases with the square power of the expected value of each observation

Poisson Scale Free

- A.k.a. "Over-dispersed Poisson" Distribution
- Mean = λ
- Variance = λ x Scale factor
- Allows variance to be lesser/greater than the mean

Poisson – Scale = 1

- Strict definition of Poisson distribution is applied, mean must equal the variance
- It assumes that the variance increases with the expected value of each observation



GLM Building Blocks

y = h(**Linear Combination of Parameters**) + Error

2

Linear Combination of Parameters

Accident Year Parameters

$\beta_{14}, \beta_{13}, \beta_{12}, \ldots, \beta_{05}$

Development Lag Parameters

$\beta_{12m}, \beta_{24m}, \beta_{36m}, \ldots, \beta_{120m}$

Calendar Year Parameters

$\beta_{CY14}, \beta_{CY13}, \beta_{CY12}, \ldots, \beta_{CY05}$

Generalized Linear Models

III. GLM Reserving Example

Section Introduction

In this section, we will cover the following:

- Start with 2-dimensional approach
- Show all years volume weighted average vs GLM
- Show how any cell in the historical triangle is linear combination of beta parameters

A simple example

- In order to “demystify” the GLM reserve model, we will walk through a basic example and show how future estimates are calculated:
  - Start with building a 2 dimensional GLM reserve model:
    - Dimension 1 = Accident Year
    - Dimension 2 = Development Lag
  - Show that results are comparable to Chain Ladder Method using all years volume weighted average

A simple example

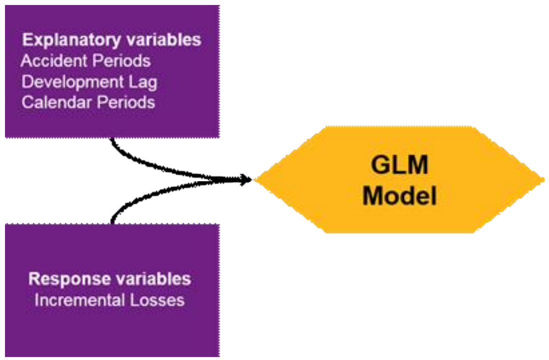
Incremental Paid Loss Triangle

Accident Year	12m	24m	36m	48m	60m	72m	84m	96m	108m	120m
2005	92	265	47	24	14	7	5	5	6	3
2006	95	273	49	25	12	8	6	6	7	
2007	98	281	50	22	14	9	7	7		
2008	100	290	46	24	15	10	8			
2009	103	288	51	27	17	11				
2010	72	321	57	30	19					
2011	80	357	64	33						
2012	89	397	71							
2013	98	441								
2014	110									

- GLM reserve method is based on predicting the response variable, incremental losses.

$Y_{AY,DL}$ = Incremental loss
AY = Accident Year AY,
DL = Development Lag, DL

Example:
$Y_{2011,12m} = 80$



A simple example

Incremental Paid Loss Triangle

Accident Year	12m	24m	36m	48m	60m	72m	84m	96m	108m	120m
2005	92	265	47	24	14	7	5	5	6	3
2006	95	273	49	25	12	8	6	6	7	
2007	98	281	50	22	14	9	7	7		
2008	100	290	46	24	15	10	8			
2009	103	288	51	27	17	11				
2010	72	321	57	30	19					
2011	80	357	64	33						
2012	89	397	71							
2013	98	441								
2014	110									

- Any cell in the historical triangle is linear combination of “beta” parameters
- Incremental losses are related to explanatory variables multiplicatively
- Resulting model gives exactly the same forecast as the chain ladder model

$$Y_{AY,DL} = EXP\ (\beta_0 + \beta_{AY} + \beta_{DL}) + \varepsilon$$

Log link function

Linear combination of explanatory variables predicts incremental losses, based on AY and DL

A simple example

		β_{12}	β_{24}	β_{36}	β_{48}	β_{60}	β_{72}	β_{84}	β_{96}	β_{108}	β_{120}
	Accident Year	**12m**	**24m**	**36m**	**48m**	**60m**	**72m**	**84m**	**96m**	**108m**	**120m**
β_{05}	2005	92	265	47	24	14	7	5	5	6	3
β_{06}	2006	95	273	49	25	12	8	6	6	7	
	2007	98	281	50	22	14	9	7	7		
	2008	100	290	46	24	15	10	8			
β_{09}	2009	103	288	51	27	17	11				
β_{10}	2010	72	321	57	30	19					
β_{11}	2011	80	357	64	33						
	2012	89	397	71							
β_{13}	2013	98	441								
β_{14}	2014	110									

Begin with a Base Parameter, β_0

We will choose Accident Year 2005, Development Lag 12 months as the base parameter

Why use a Base Parameter?

Needed to allow for model convergence

Setting a base parameter reduces the number of variables by 1

A simple example

	Accident Year	12m	24m	36m	48m	60m	72m	84m	96m	108m	120m
	2005	92	265	47	24	14	7	5	5	6	3
β_{06}	2006	95	273	49	25	12	8	6	6	7	
	2007	98	281	50	22	14	9	7	7		
	2008	100	290	46	24	15	10	8			
β_{09}	2009	103	288	51	27	17	11				
β_{10}	2010	72	321	57	30	19					
β_{11}	2011	80	357	64	33						
	2012	89	397	71							
β_{13}	2013	98	441								
β_{14}	2014	110									

Explanatory Variables

Dimension 1 = Accident Year

β_{11} = Multiplicative parameter that describes accident year 2011

$Y_{11,DL} = EXP(\beta_0 + \beta_{11} + \beta_{DL}) + \varepsilon$

A simple example

		β_{24}	β_{36}	β_{48}	β_{60}	β_{72}	β_{84}	β_{96}	β_{108}	β_{120}
Accident Year	**12m**	**24m**	**36m**	**48m**	**60m**	**72m**	**84m**	**96m**	**108m**	**120m**
2005	92	265	47	24	14	7	5	5	6	3
2006	95	273	49	25	12	8	6	6	7	
2007	98	281	50	22	14	9	7	7		
2008	100	290	46	24	15	10	8			
2009	103	288	51	27	17	11				
2010	72	321	57	30	19					
2011	80	357	64	33						
2012	89	397	71							
2013	98	441								
2014	110									

Explanatory Variables

Dimension 2 = Development Lag

β_{48m} = Multiplicative parameter that describes development lag 48 months

$Y_{AY,48m} = EXP(\beta_0 + \beta_{AY} + \beta_{48m}) + \varepsilon$

A simple example

	Accident Year	12m	24m	36m	48m	60m	72m	84m	96m	108m	120m
	2005	β_0 92	265	47	24	14	7	5	5	6	3
β_{06}	2006	95	273	49	25	12	8	6	6	7	
	2007	98	281	50	22	14	9	7	7		
	2008	100	290	46	24	15	10	8			
β_{09}	2009	103	288	51	27	17	11				
β_{10}	2010	72	321	57	30	19					
β_{11}	2011	80	357	64	33						
	2012	89	397	71	??						
β_{13}	2013	98	441								
β_{14}	2014	110									

Here's another example.

Example 1. $Y_{12,36m} = EXP(\beta_0 + \beta_{12} + \beta_{36m}) + \varepsilon$

Example 2. $Y_{12,48m} = EXP(\beta_0 + \beta_{12} + \beta_{48m}) + \varepsilon$

A simple example

Accident Year Parameter	Value
β_{2005}	n/a
β_{2006}	0.029
β_{2007}	0.056
β_{2008}	0.082
β_{2009}	0.105
β_{2010}	0.124
β_{2011}	0.225
β_{2012}	0.325
β_{2013}	0.424
β_{2014}	0.338

Development Lag Parameter	Value
β_{12m}	n/a
β_{24m}	1.260
β_{36m}	(0.485)
β_{48m}	(1.177)
β_{60m}	(1.704)
β_{72m}	(2.244)
β_{84m}	(2.533)
β_{96m}	(2.612)
β_{108m}	(2.470)
β_{120m}	(3.143)

Base Parameter	Value
β_0	4.358

Example 1:

$Y_{12,36m} = EXP(\beta_0 + \beta_{12} + \beta_{36m})$

$= EXP\ (4.358 + 0.325 - 0.485)$

= 67 (vs actual 71)

Example 2:

$Y_{12,48m} = EXP(\beta_0 + \beta_{12} + \beta_{48m})$

$= EXP\ (4.358 + 0.325 – 1.177)$

= 33

A simple example

Accident Year	2-D GLM Unpaid	Chain Ladder Unpaid	Difference
Prior	470	470	0
2008	484	484	0
2009	497	497	0
2010	510	510	0
2011	522	522	0
2012	532	532	0
2013	589	589	0
2014	651	651	0
Total	**5,632**	**5,632**	**0**

- When excluding the calendar year dimension, as we did in this example, the results are the same as chain ladder method using all year volume weighted average

Incorporating the Calendar year effect

$$\log(\mu_{ij}) = \eta_{ij}$$

← Log "link" function

$$\eta_{ij} = c + \sum_i a_i + \sum_j b_j + r\tau$$

← Linear predictor

$$\tau = i + j - 2$$

← Calendar time

Problem:

The model is now over-parameterised – there is a relationship between origin, development and calendar time, one dimension is a linear combination of the other two. A unique solution is not identifiable.

The Optimal Model

- Use stepwise procedures to reduce the number of parameters and find the optimal model
- Several optimisation schemes could be proposed
  - Optimise backward – iteratively tests each parameter and removes the ones that are not statistically significant
  - Optimise forward – Iteratively tests each parameter and adds in the ones that are statistically significant
  - Optimise backward/forward – Optimise backward first and Optimise forward second

Generalized Linear Models

IV. Conclusion

Conclusion

- Model Limitations
  - Still working with limited set of data points; i.e. a 10 x 10 triangle only has 55 data points
    - Run the risk of "Overfitting" if too many parameters included – Model explains historical experience but poor future predictive value
- Origin, development and calendar period effects are interlinked, so it can be very difficult to interpret the parameters
- When calendar period effects are included, it is always necessary to extrapolate in the calendar period direction
  - The results will be sensitive to the assumptions regarding extrapolation
  - A model that fits the observed data well may not be good for forecasting!

Bayesian Models

V. Statistics 101

Statistics 101 – Bayes Theorem

- Bayes theorem indicates how prior subjective belief changes based on evidence
- $P(A/B) = \frac{P(B/A)P(A)}{P(B)}$, where
- P(A) is the prior belief
- P(A/B) is the posterior belief accounting for B
- P(B/A)/P(B) represents the support B provides to A

Statistics 101- Likelihood function

- Usually we think in terms of probabilities, i.e., the probability of an outcome X given a parameter Θ; P(X| Θ)

- The Likelihood instead is a function of Θ given an outcome, i.e. L(Θ|X)

- With an observed outcome X the maximum likelihood principle chooses the parameter Θ that maximizes the P(X| Θ)

- The GLM model in ResQ produces a maximum likelihood function

Statistics 101 – Posterior Probability

- Given prior belief p(θ) and observation x with likelihood P(x / θ) the posterior is:

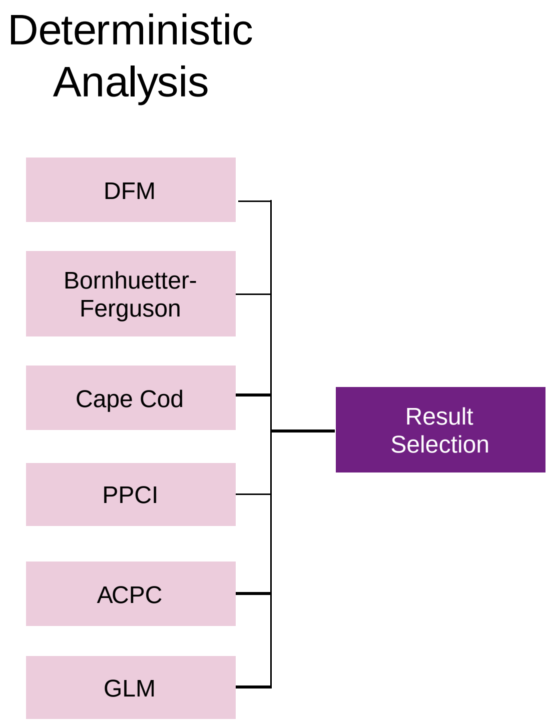
$$P(\theta/x) = \frac{P(x/\theta)P(\theta)}{P(x)}, \text{ i.e}$$

- Posterior probability ∝ Likelihood x Prior Probability

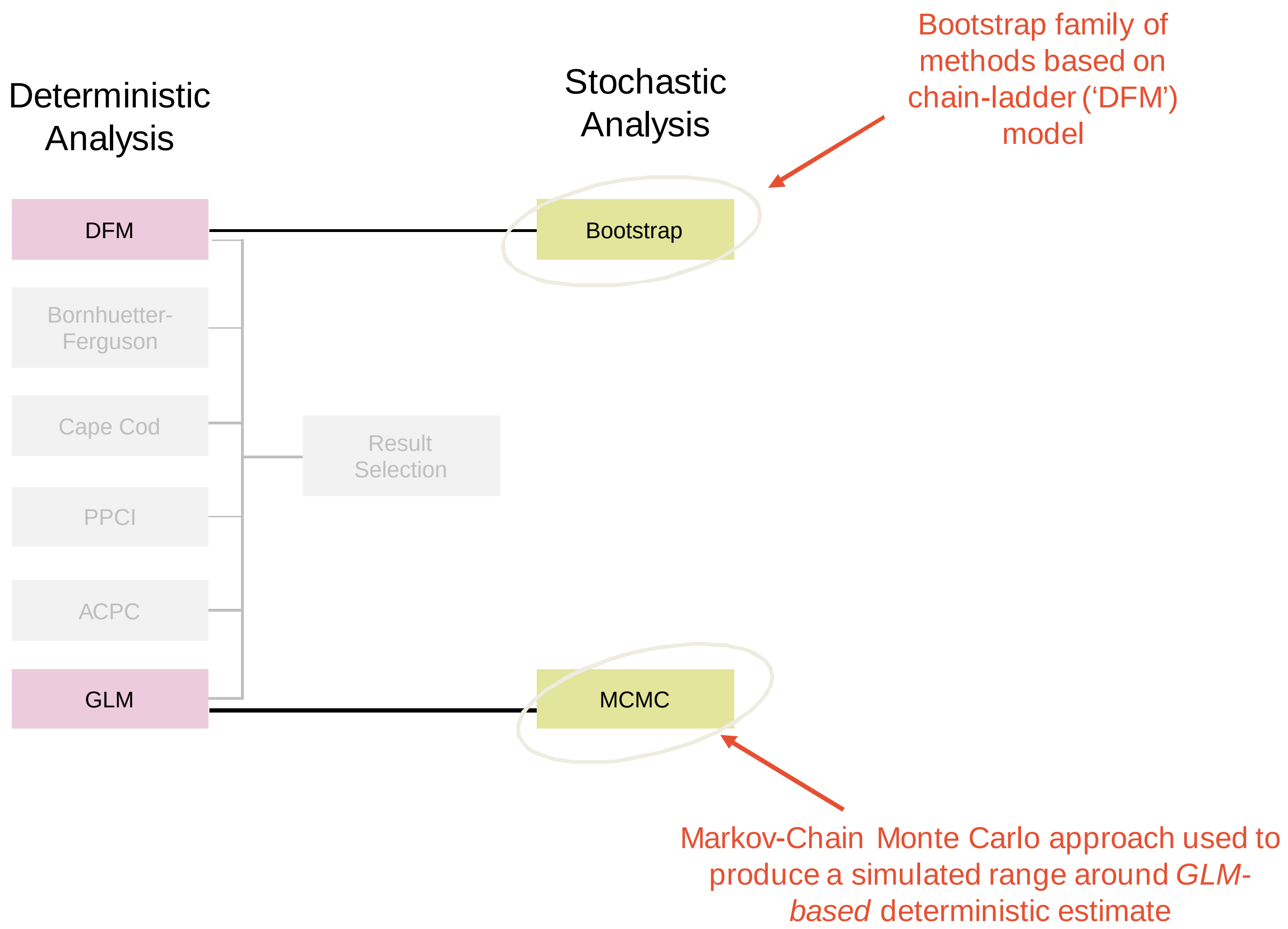
Statistics 101 – MCMC stochastic methods

- MCMC methods are a class of algorithms for sampling from a probability distribution
  - This distribution is usually difficult to approximate with analytical functions
- MCMC constructs a random process that undergoes transition from one state to another, called *Markov Chain*
  - This process is *memoryless*, i.e. the next state is based only on the current state but not the sequence of the preceding states
  - The quality of the convergence to an *equilibrium* distribution improves with the number of steps employed in the process
  - The first few draws are usually thrown away (called *burn-in*) to ensure target is independent of starting point and improve convergence

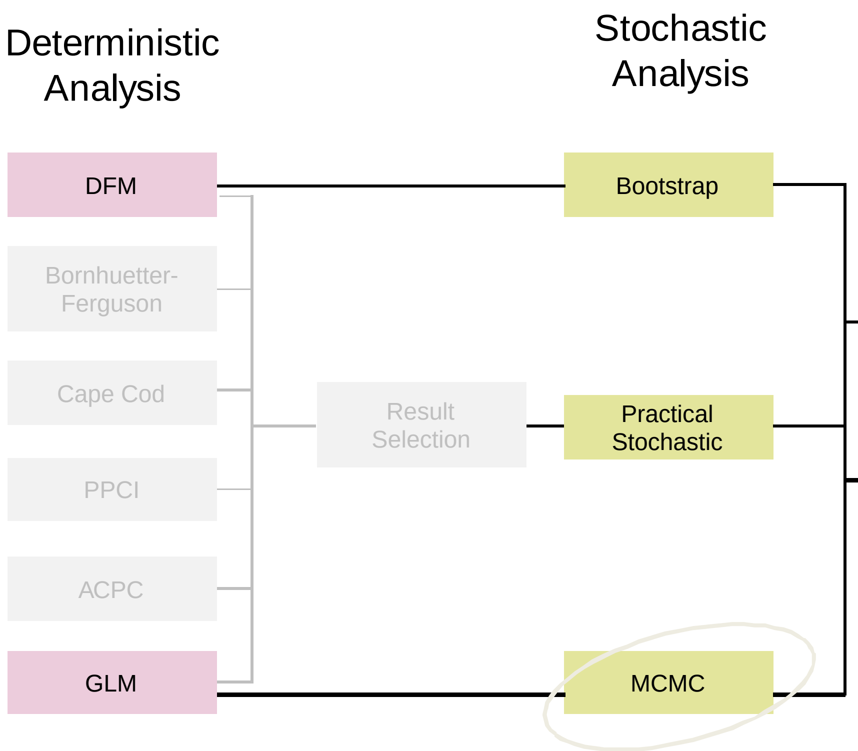
Statistics 101 – MCMC compared to Bootstrapping



Statistics 101 – MCMC compared to Bootstrapping (cont'd)



Statistics 101 – MCMC compared to Bootstrapping (cont'd)



As with Bootstrapping, and Practical Stochastic methods, produces full predictive distribution of outcomes - as well as cashflows – that can be:

...incorporated into Run-off Result analysis (re-reserving)

...consolidated to reflect uncertainty across multiple LoBs

Statistics 101 – MCMC compared to Bootstrapping (cont'd)

- **Similarities**
- Both are simulation methods
- They produce a full predictive distribution of outcomes including
  - Parameter risk
  - Process risk

- **Differences**
- Bootstrapping
  - Samples with replacement the residuals from an actual versus expected comparison of historical development
  - Simulations are independent from one another
  - There is no convergence in the simulations
- MCMC
  - Samples the parameters of the resulting GLM likelihood function
  - Simulations are built through a Markov chain
  - The simulations converge into an *equilibrium* state

Bayesian Models

VI. Bayesian Modeling Steps

Bayesian modeling steps

Step 1: Specify probability distribution for the data given some unknown parameters (data distribution)

Step 2: Specify prior probability distribution for the parameters of the data distribution (prior distribution)

Step 3: Derive the likelihood function of the parameters, given the data (likelihood function)

Step 4: Combine prior distribution and likelihood function to derive posterior joint distribution of parameters(posterior distribution)

Step 5: Obtain parameters for posterior distribution

Step 6: Combine data distribution and posterior distributions to obtain forecast of predictive distribution

Bayesian modeling steps (cont'd)

Step 1 & 2: Example is the ODP model.
Prior distributions could assume some distributional shape (i.e. Lognormal, Gamma etc.)
Informative priors with small variance that could affect shape of posterior

Step 4: Posterior is based on Bayesian theory and is proportional to prior and the likelihood function

Step 5: It is easy to obtain the parameters of posterior when the shape of distribution is known. Otherwise special statistical algorithms, like Gibbs MCMC, are needed

Step 6: Like step 5 the complexity of the forecasting depends on whether predictive distribution is recognizable. Generic sampling algorithms such as Adaptive Rejection Sampling (ARS) might be needed

Bayesian Models

VII. Bayesian Modeling within the Reserving Context

Bayesian modeling within the reserving context

- Reserving example

Origin Period	Development Period 1	2	3	…	n
1	C_{11}	C_{12}	C_{13}	…	C_{1n}
2	C_{21}	C_{22}	C_{23}	…	C_{2n}
3	C_{31}	C_{32}	C_{33}	…	C_{3n}
…	…	…	…	…	…
n	C_{n1}	C_{n2}	C_{n3}	…	C_{nn}

- C={Cij: i+j<n+1} is the upper-left triangle of observed payments, and the reserving problem attempts to estimate the unobserved values in the lower-right triangle

Bayesian modeling within the reserving context (cont'd)

- Step 1&2: Assumes C_{ij} follows a probability density distribution of $f(C_{ij} / \theta)$, where θ denotes parameters describing a particular claims generating process and $\pi(\theta)$ is the prior distribution function
- Step 3: The likelihood function $L(\theta/\underline{c})$ for the parameters given observe data is:

$$L(\theta / \underline{c}) = \prod_{i+j \leq n+1} f(C_{ij} / \theta)$$

Bayesian modeling within the reserving context (cont'd)

- **Step 4**: Given the data distribution and the prior distribution, the posterior distribution $f(\theta/\underline{c})$ is proportional to the product of the likelihood and the prior:

  $$f(\theta\ /\underline{c}) \propto L(\theta\ /\underline{c})\ \pi(\theta)$$

- **Step 5**: Parameters θ are obtained from the posterior distribution $f(\theta\ /\underline{c})$ and are used in **Step 6**

Bayesian modeling within the reserving context (cont'd)

- **Step 6**: The known data C_{ij} for $i + j \leq n + 1$ is used to predict unobserved values in the lower right triangle C_{ij} for $i + j > n + 1$ by means of the predictive distribution:

$$f(C_{ij} / \underline{c}) = \int f(C_{ij} / \theta) \, f(\theta / \underline{c}) \, d\theta$$

- Predictive distribution can either be obtained in a closed form analytically or through a generic sampling algorithm instead

Bayesian Models

VIII. Simple Example – No Simulations Needed

Simple example – no simulations needed

- **Step 1&2**: Assume the loss generating process follows a Poisson distribution with parameter and the parameter follows a Gamma distribution with some known parameters a and b

- $C_{ij} / \theta \sim \rho(\theta)$

- $\theta / a,b \sim Gamma(a,b)$

Simple example – no simulations needed (cont'd)

- **Step 3**: The likelihood function is given by $L(\theta / \underline{c}) = \prod_{i=1}^{n} \theta^{x_i} e^{-\theta} / x_i!$

- **Step 4**: The posterior distribution is proportional to the product of the likelihood and the prior:

  $f(\theta/\underline{c},a,b) = \prod_{i=1}^{n} \theta^{x_i} e^{-\theta} / x_i!\ b^a/\Gamma(a)\ \theta^{a-1} e^{-b\theta} =$

  ~ Gamma $(a + \sum_{i=1}^{n} x_i\ , b + n)$, i.e. posterior follows a Gamma distribution

Simple example – no simulations needed (cont'd)

- **Step 6**: The product of the posterior and the data distribution, i.e., the product of a Gamma and a Poisson distribution results in a negative binomial distribution:

$$NB(a + \sum_{i=1}^{n} x_i \,, \frac{1}{1+b_1})$$

- No need for complicating sampling here !

Bayesian Models

IX. Examples of Popular Sampling Techniques

Examples of popular sampling techniques – Gibbs sampler

- Gibbs sampler avoids sampling from a complicated bivariate distribution f(x,y) by making random draws instead from univariate conditional distributions (f(x/y) and f(y/x))
- For two parameters and n iterations it produces an

  n x 2 table where xo is the initial value – Next steps:

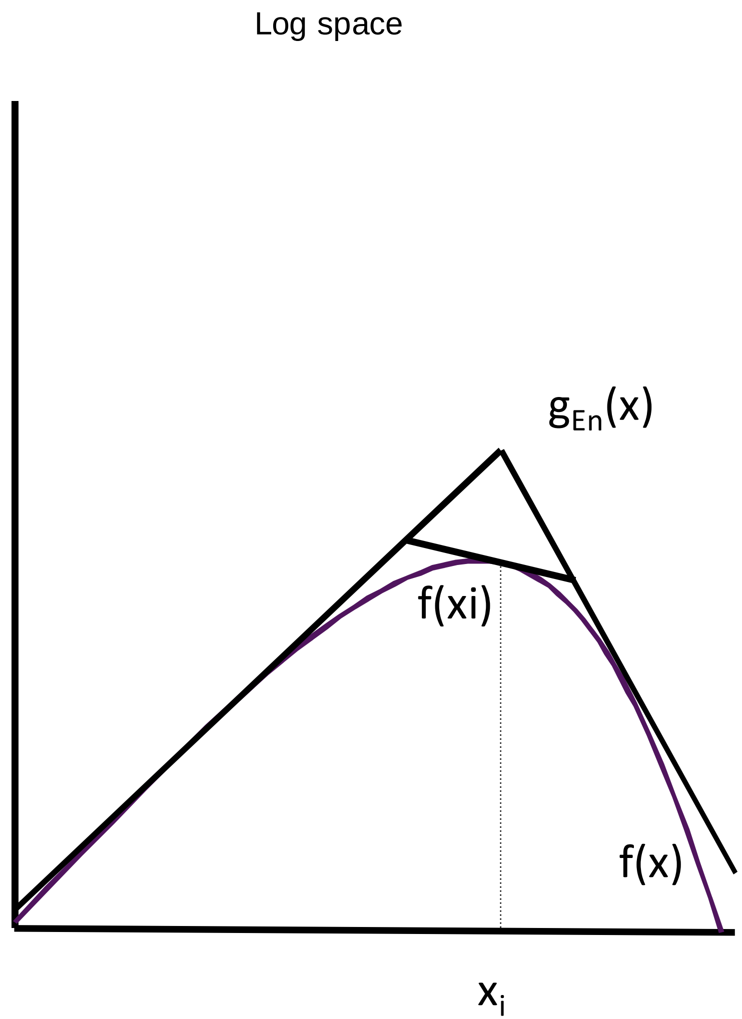
$$y_1 \sim f(y/x = x_0)$$

…………………

$$x_i \sim f(x/y = y_{i-1})$$
$$y_i \sim f(y/x = x_i)$$

- Eventually $(x_i,y_i) \rightarrow (x,y) \sim f(x,y)$ for sufficient large number of iterations (so called burn-in sample)
- After burn-in, it is common to define a spacing between accepted points, maybe every m draws, to ensure independence of random draws

Examples of popular sampling techniques – Adaptive Rejection Sampling



- ARS works with log concave densities f(x)
- An envelope function $g_{En}(x)$ as an upper bound of the log density function is employed
- A random draw x_i from the x-axis is then sampled
- When the resulting $g_{En}(x_i)$ is close to $f(x_i)$ the envelope function remains unchanged
- When the resulting $g_{En}(x_i)$ is much larger to $f(x_i)$ the envelope function changes to incorporate a line that is tangent to $f(x_i)$

Examples of popular sampling techniques – Metropolis-Hastings Algorithm

1. Create GLM with Error distribution f(x|μ) (typically Poisson)

Data Ultimate

Parameters

Selected Parameters

2. GLM produces Parameter estimates with uncertainty

Parameters, Error

3. Set Initial Markov Chain μ_0 equal to parameter estimates from GLM

4. Sample a Candidate Parameter μ*

p ~ Multivariate Normal

5. Calculate Markov Transition Probability R

Based on a ratio of Likelihood Functions, where the fit of μ* is compared to the fit of μ_{t-1}

6. Draw U from Uniform(0,1) Distribution

U ~ Uniform(0,1)

7. Accept or Reject the Candidate from step 4

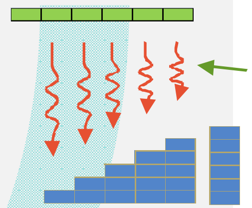
set $\mu_t = \mu^*$ if U<R

Otherwise set $\mu_t = \mu_{t-1}$

8. Calculate Reserves based on the Markov Chain ending value μ_t

Data

Selected μ_t



Simulated Ultimate

9. Repeat steps 4-8 10,000 times for burn-in period

10. Discard the burn-in steps and Repeat steps 4-8 10,000 times for final result

Bayesian Models

X. Conclusions

MCMC Bayesian stochastic reserving method has both

Advantages:

- Flexible not constrained by any type of model
- Allows the incorporation of user's judgment
- Provides a full distribution of outcomes

Disadvantages:

- More sophisticated mathematics
- Can be influenced by judgment
- Actuaries are "scared" of it

Questions and Discussion