The Retrospective Testing of Stochastic Loss Reserve Models

Glenn Meyers – FCAS, MAAA, Ph.D.
ISO Innovative Analytics
CLRS September 15-16,2001

Background

- Risk based capital proposals, e.g. EU Solvency II and USA SMI rely on stochastic models.
  - VaR@99.5% and TVaR@99%
- There are many stochastic loss reserve models that claim to predict the distribution of ultimate losses.

Are any of these models right?

E-Forum Paper

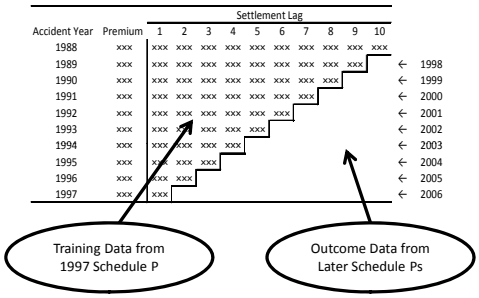
Joint with Peng Shi – Northern Illinois University

- Describes a database
  - Data from several American Insurers
  - Data for six lines of insurance
  - Paid and incurred loss triangles
  - Subsequent outcomes
  - Available online (Free)
- Predicts the distribution of outcomes of two models for several insurers for Commercial Auto Insurance
- Tests the predictions against subsequent reported outcomes.

The CAS Loss Reserve Database

- Schedule P (Data from Parts 1-4) for several US Insurers
  - Private Passenger Auto
  - Commercial Auto
  - Workers' Compensation
  - General Liability
  - Product Liability
  - Medical Malpractice (Claims Made)
- Available on CAS Website – New Version 9/1/2011
  http://www.casact.org/research/index.cfm?fa=loss_reserves_data

The CAS Loss Reserve Database



- Can we predict the distribution of outcomes? Or sums of outcomes?

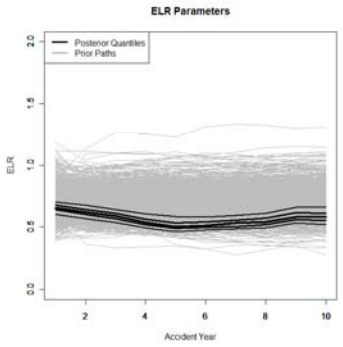
Examples of Tests in This Paper

- Commercial Auto
- 50 Insurers – "Selected" going concern insurers
- Tested two stochastic loss reserve models
  - Bootstrap chain ladder (BCL) model
    - Used the "ChainLadder" package in R
    - Overdispersed Poisson for process risk.
  - Bayesian Autoregressive Tweedie (BAT) model
    - See next slide

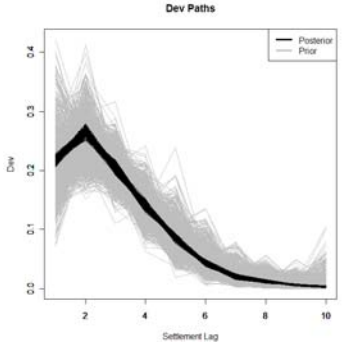
The BAT Model

- Uses earned premium and incremental paid loss data.
- Expected Loss Ratio (ELR) parameters follow an AR(1) process.
- Calendar year trend parameters follow an AR(1) process.
- Generate parameters by a Bayesian MCMC method.
- Process risk described by the Tweedie distribution.
- Prior distribution derived by examining MLE estimates of a similar model on several insurers.

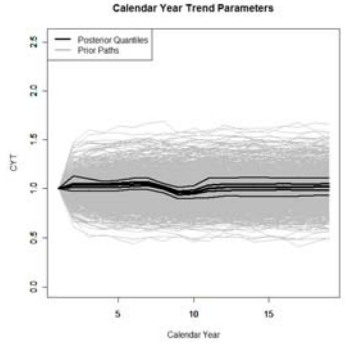
Parameters for Insurer 914



Parameters for Insurer 914



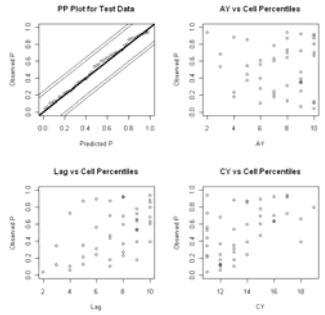
Parameters for Insurer 914



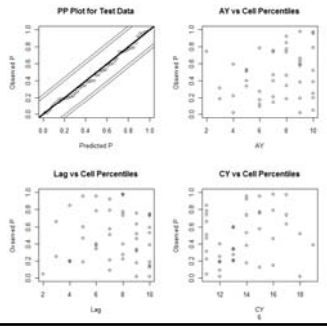
Criteria for a "Good" Stochastic Loss Reserve Model

- Using the upper triangle "training" data, predict the distribution of the outcomes in the lower triangle
  - Can be observations from individual (AY, Lag) cells or sums of observations in different (AY,Lag) cells.
- Using the predictive distributions, find the percentiles of the outcome data.
- The percentiles should be uniformly distributed.
  - Test with PP Plots/KS tests or with histograms.

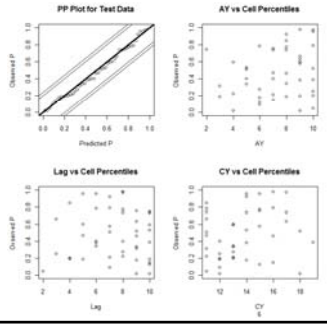
Testing the Distributions of (AY,Lag) Outcome Percentiles for a Single Insurer BCL - Insurer 914



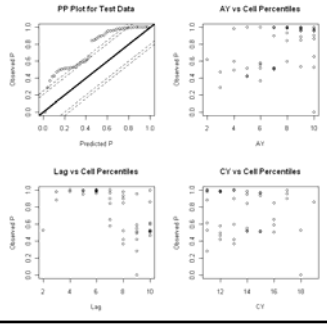
Testing the Distributions of (AY,Lag)
Outcome Percentiles for a Single Insurer
BAT - Insurer 914



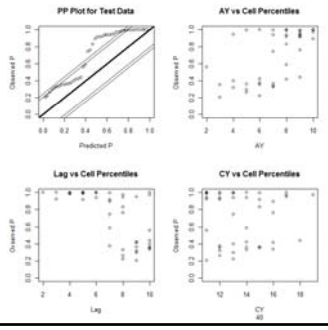
Testing the Distributions of (AY,Lag)
Outcome Percentiles for a Single Insurer
BAT - Insurer 914



Testing the Distributions of (AY,Lag)
Outcome Percentiles for a Single Insurer
BCL - Insurer 310



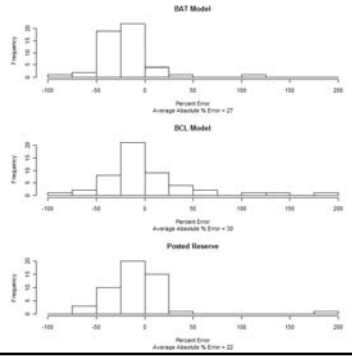
Testing the Distributions of (AY,Lag) Outcome Percentiles for a Single Insurer BAT - Insurer 310



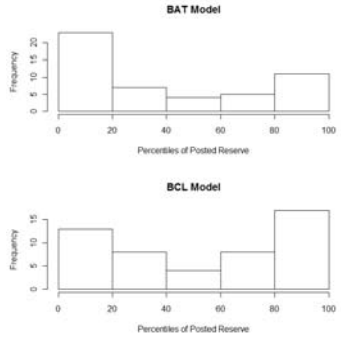
Testing the Model on Multiple Insurers

- Each model can predict the distribution of the sum of all outcomes in the lower triangle.
- Compare the mean of the predicted distribution with the sum of all outcomes.
  - For each model
  - For the posted reserve

% Error



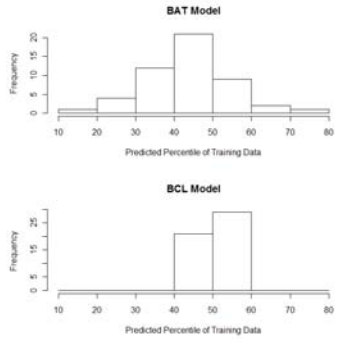
Percentile of Posted Reserve for Each Model



Testing the Model on Multiple Insurers

- Each model can predict the distribution of the sum of all outcomes in the lower triangle.
- Find the percentile of the actual sum of outcomes for each insurer.
- These percentiles should be uniformly distributed.
- This is a test of the model.

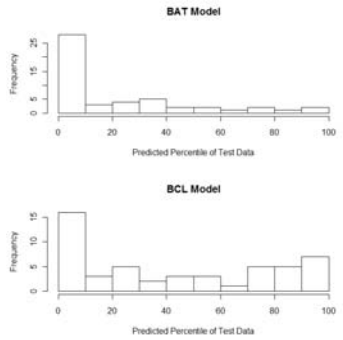
Predicted Percentiles of Outcomes



Should be Uniformly Distributed

Overfitting!

Predicted Percentiles of Outcomes



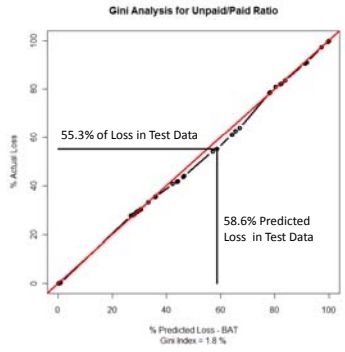
Conclusions

- Neither the BAT or the BCL does a good job at predicting the distribution of outcomes.
- Two possible reasons
  - We don't have the right model
  - Changes in the claim settlement environment make the outcomes unpredictable.

Finding the Right Model

- These models used only paid data. Could we do a better job by including incurred loss data?
- BAT used earned premium data. Does this help or hinder the prediction?
- Is there other external data available?
- Work with other lines of insurance.

A Hint – Use Unpaid Loss Information



Unpredictable Environmental Changes

- If so, how do we manage insurer risk?
- Self correcting over time? Can we make adjustments as additional data come in?
- Challenge – Our new proposed solvency regulations (i.e. EU Solvency II and American SMI) depend on our ability to predict the distribution of outcomes. What happens if we cannot accurately predict the distributions?