Retrospective Test for Loss Reserving Methods - Evidence from Auto Insurers

Peng Shi - Northern Illinois University

joint work with
Glenn Meyers - Insurance Services Office

CAS Annual Meeting, November 8, 2010

Outline

- Introduction
- Loss reserving methods
- Sampling of NAIC Schedule P
- Analysis for the industry
- Analysis for individual insurers
- Concluding remarks

Introduction

- A loss reserving model from a upper triangle (training data), one is interested in whether it is a *good* or *bad* predictive distribution.
- Standard error is commonly used measure of variability, does a small standard error mean a good predictive model?
- Hold-out observations are needed to answer the above question.
- For a run-off triangle of incremental paid losses, suppose we observe all the losses in the lower triangle (hold-out sample), the retrospective test in this study is based on the following well-know result:
  If X is a random variable with distribution F, then the transformation $F(X)$ follows a uniform distribution on (0,1).

Introduction

- *X* : total reserve
  - Use a sample of independent insurers.
  - Test whether the percentiles of total reserves are from uniform (0,1).
  - Informs us whether a predictive model is good for the whole industry.

- *X* : incremental paid losses in each cell of the lower triangle
  - Test for each single insurer.
  - Test whether the percentiles of incremental losses in the lower triangle (hold-out sample) are from uniform (0,1)
  - Informs us whether a predictive model performs well for a particular insurer

Loss reserving methods

- Three methods are considered: Mack chain ladder, bootstrap over-dispersed Poisson, Bayesian log-normal

- An industry benchmark: Chain-Ladder technique
  - Large literature on CL, see England and Verrall (2002), Wüthrich and Merz (2008)
  - Many stochastic models reproduce CL estimates, e.g. Mack (1993,1999), Renshaw and Verrall (1998), Verrall (2000)
  - Modifications of CL, e.g. Barnnett and Zehnwirth (2000)

- Mack CL:
  - Variability can be from recursive relationship, see Mack (1999)
  - Assume normality in the calculation of percentiles
- Bootstrap ODP:
  - Resample residuals of GLM
  - Fit CL to pseudo data
  - Simulate incremental loss for each cell

A Bayesian Log-normal Model

- Previous studies: Alba (2002,2006), Ntzoufras and Dellaportas (2002)
- Calendar year effect has been ignored
- We propose

$$\log(Y_{ij}) \sim N(\mu_{ij}, \sigma^2)$$
$$\mu_{ij} = \alpha_i + \beta_j + \gamma_{t=i+j}, \quad i, j = 1, \cdots N$$

Y_{ij} - normalized incremental paid loss for cell (i,j)

α_i - trend for accident year i

β_j - trend for development lag j

γ_t - trend for calendar year t

- We use accident year premium as exposure variable

A Bayesian Log-normal Model

- Different ways to specify calendar year trend

$$\text{IID Specification}: \gamma_t \sim N(0,\sigma_\gamma^2)$$
$$\text{Autoregressive Model}: \gamma_t = \phi\gamma_{t-1} + \eta_t \quad \eta_t \sim N(0,\sigma_\eta^2)$$
$$\gamma_2 \sim N(\mu_\gamma,\sigma_\gamma^2)$$
$$\text{Random Walk}: \gamma_t = \gamma_{t-1} + \eta_t \quad \eta_t \sim N(0,\sigma_\eta^2)$$
$$\gamma_2 \sim N(0,\sigma_\gamma^2)$$

- Calendar year trend introduce correlation due to calendar year effects
- The state space specification could be used on accident year or development year trend
- We focus on AR and RW specifications in the following analysis

A Bayesian Log-normal Model

- The likelihood function can be derived as follows

$$\text{Let } \mathbf{P}_1 = \{\alpha_i, \beta_j, \gamma_t, \sigma^2\} \text{ and } \mathbf{P}_2 = \{\sigma_\gamma^2, \sigma_\eta^2, \phi\}, \text{ then}$$
$$f(\mathbf{P}_1,\mathbf{P}_2 \mid \mathbf{y}) \propto f(\mathbf{y} \mid \mathbf{P}_1,\mathbf{P}_2) \times f(\mathbf{P}_1,\mathbf{P}_2) = f(\mathbf{y} \mid \mathbf{P}_1) \times f(\mathbf{P}_1 \mid \mathbf{P}_2) \times f(\mathbf{P}_2)$$
$$\bullet\, f(\mathbf{y} \mid \mathbf{P}_1) = \prod_{i=1}^{n}\prod_{j=1}^{n} f(y_{ij} \mid \mathbf{P}_1) \text{ where we use log-normal specification}$$
$$\bullet\, f(\mathbf{P}_1 \mid \mathbf{P}_2) = f(\boldsymbol{\gamma} \mid \mathbf{P}_2)\prod_{i=1}^{n} f(\alpha_i)\prod_{j=1}^{n} f(\beta_j)$$
$$\text{where } f(\boldsymbol{\gamma} \mid \mathbf{P}_2) = f(\gamma_{2n} \mid \gamma_{2n-1}, \mathbf{P}_2) f(\gamma_{2n-1} \mid \gamma_{2n-2}, \mathbf{P}_2) \cdots f(\gamma_3 \mid \gamma_2, \mathbf{P}_2) f(\gamma_2 \mid \mathbf{P}_2)$$
$$= f(\eta_{2n} \mid \mathbf{P}_2) f(\eta_{2n-1} \mid \mathbf{P}_2) \cdots f(\eta_3 \mid \mathbf{P}_2) f(\gamma_2 \mid \mathbf{P}_2)$$
$$\bullet\, f(\mathbf{P}_2) = f(\sigma_\gamma^2) f(\sigma_\eta^2) f(\phi)$$

- We perform the analysis using WinBUGS

Sampling of NAIC Schedule P

SCHEDULE P — PART 3 — SUMMARY

1	CUMULATIVE PAID LOSSES AND ALLOCATED EXPENSES AT YEAR END ($000 OMITTED)										12	13
Years in Which Losses Were Incurred	2 1988	3 1989	4 1990	5 1991	6 1992	7 1993	8 1994	9 1995	10 1996	11 1997	Number of Claims Closed With Loss Payment	Number of Claims Closed Without Loss Payment
1. Prior	0 0 0										XXXX	XXXX
2. 1988											XXXX	XXXX
3. 1989	XXXX										XXXX	XXXX
4. 1990	XXXX	XXXX									XXXX	XXXX
5. 1991	XXXX	XXXX	XXXX								XXXX	XXXX
6. 1992	XXXX	XXXX	XXXX	XXXX							XXXX	XXXX
7. 1993	XXXX	XXXX	XXXX	XXXX	XXXX						XXXX	XXXX
8. 1994	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX					XXXX	XXXX
9. 1995	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX				XXXX	XXXX
10. 1996	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX			XXXX	XXXX
11. 1997	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX		XXXX	XXXX

Sampling of NAIC Schedule P

- Training data is from 1997 schedule P
- Accident year 1988 – 1997
- Hold-out sample is from schedule P of subsequent years
  e.g. actual paid losses for AY 1989 is from 1998 schedule P
  actual paid losses for AY 1990 is from 1999 schedule P
  ……
  actual paid losses for AY 1997 is from 2006 schedule P
- Limit to group insurers or single entities
- Use data for personal auto and commercial auto for our analysis
- Check overlapping periods in training data and hold-out sample
  e.g.

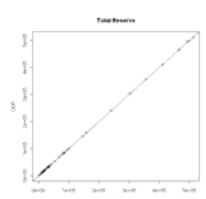
Training	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997
AY 1989										
Hold-out	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
AY 1989										

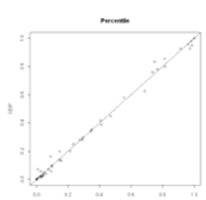
Analysis for the industry

- Consider largest 50 insurers for personal and commercial auto lines
- Use net premiums written to measure size
- For each line of business:
  - derive the predictive distribution of total reserves for insurer *i,* say F_i
  - calculate the percentile of the actual losses $p_i = F_i(\text{loss}_i)$
  - repeat for all 50 firms
- Test if p_i follows uniform (0,1)
- Implications:
  - if a predictive model performs well, percentiles should be a realization from uniform (0,1)
  - an outcome that falls on the lower or higher percentile of the distribution does not suggest a bad model

Commercial Auto

- Consider Mack CL and bootstrap ODP for top 50 insurers
- Compare point estimate of total reserve and prediction error
- 1st figure compares point prediction that confirms two methods provide same estimates
- 2nd figure compares percentiles of actual losses, indicating a similar predictive distribution



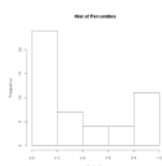


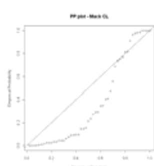
Commercial Auto

- Next two slides present the percentiles p_i (i = 1,…,50) of total reserves for the 50 insurers under different loss reserving methods
- Histogram and uniform pp-plot are produced for four methods
- K-S test is used to test if p_i follows uniform
- We observe:
  - again Mack CL and bootstrap ODP provides similar results
  - pp-plots show both might have overfitting problem
  - among state space modeling, AR1 specification performs better with a high p-value in the K-S test

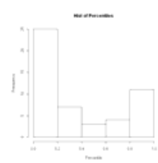
---

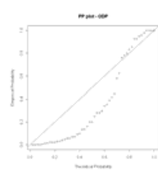
- Mack CL



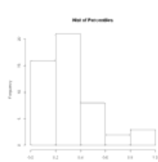


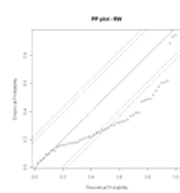
- Bootstrap ODP



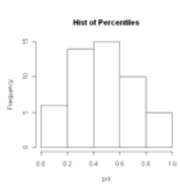


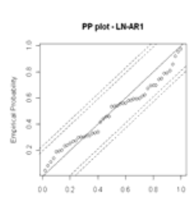
- LN - RW





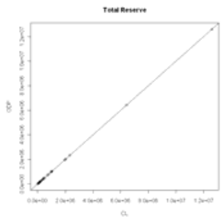
- LN – AR: p-value of K-S test is 0.43

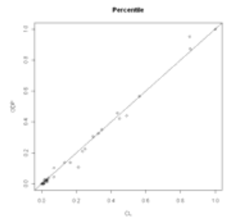




Personal Auto

- Repeat above analysis of total reserves for personal auto
- First we consider Mack CL and bootstrap ODP using data from largest 50 insurers
- Comparison of point prediction and percentile of actual losses confirms the close results from the two chain ladder models

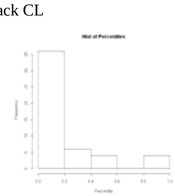


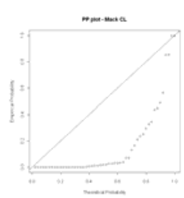


Personal Auto

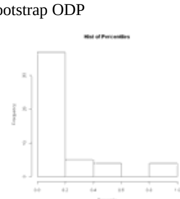
- As done for commercial auto, next two slides present the percentiles p_i (i = 1,…,50) of total reserves for the 50 insurers under different loss reserving methods
- We exhibit both histogram and uniform pp-plot, and K-S test is used to test if p_i follows uniform
- We observe:
  - again Mack CL and bootstrap ODP provides similar results
  - the performance if worse than the commercial auto, since most realized outcomes lie on the lower percentile of the predictive distribution
  - Log-normal model does not suffer like the above two, and a high *p*-value of the K-S test suggests the good performance of the AR1 specification

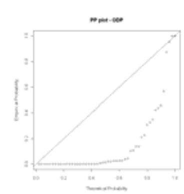
- Mack CL



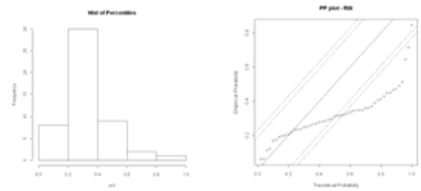


- Bootstrap ODP

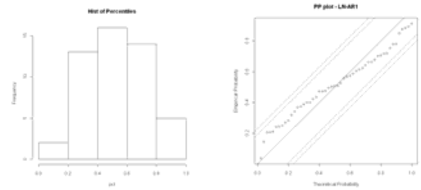




- LN - RW



- LN – AR: *p*-value of K-S test is 0.12



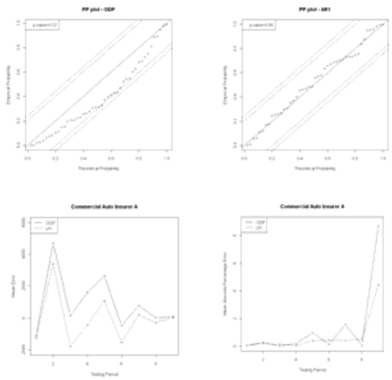
Analysis for individual insurers

- Consider individual insurers
- For illustrative purposes, we pick out 2 insurers for each line
- Compare ODP and LN-AR model
- Out of the two individual insurers for each line, we show that ODP is better for one firm and LN model is better for the other one
- Though the analysis, we hope to explain why a certain method outperforms the other one

Commercial Auto – Insurer A

- For insurer A, we derive the predictive distribution for each cell in the lower part of the triangle
- Then calculate the percentiles for actual incremental paid losses in the hold-out sample
- Uniform pp-plots of percentiles with the *p*-value of K-S tests are shown in next slide
- LN model outperforms ODP slightly
- We also compare mean error and mean absolute percentage error of the two methods over the 9 testing periods
- The result, to a great extent, agrees with K-S test

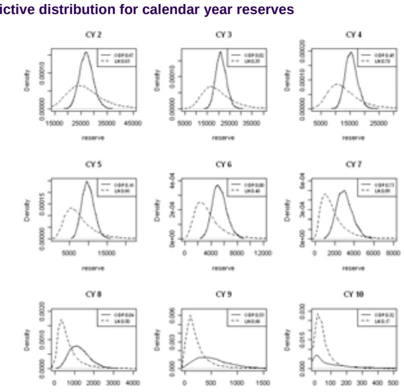
Commercial Auto – Insurer A



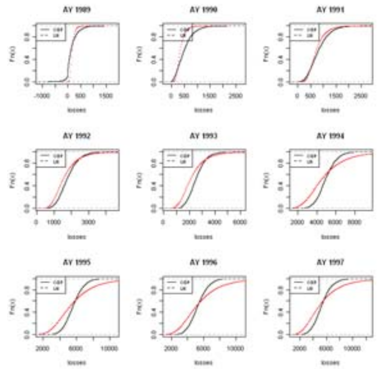
Commercial Auto – Insurer A

- In the next two slides, we analyze the predictive distributions from the two methods
- 1st slide shows the predictive distributions for calendar year reserves
  - for early calendar years, LN provides wider distribution, as one moves to the bottom right of the lower triangle, LN provides narrow distribution
  - recall calendar year reserve is the sum of losses from cells in the same diagonal
- 2nd slide shows the predictive distribution of each cell in calendar year CY=2, that is calendar year 1998
  - for top right cells on the diagonal, LN provides narrower distribution, and for bottom left cells, LN provides wider distribution
  - LN provides higher volatility for early development year

Predictive distribution for calendar year reserves

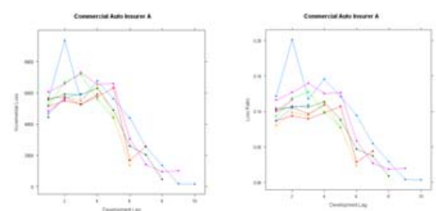


Predictive distribution for each cell of first calendar year



Commercial Auto – Insurer A

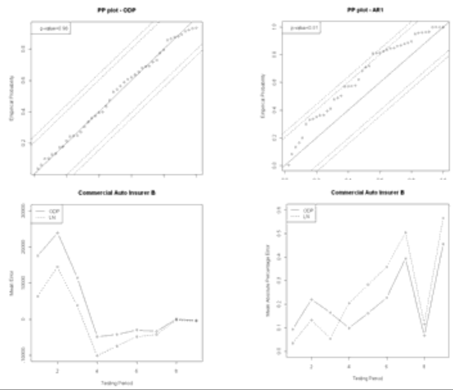
- We look into the pattern of the training data
- We show time-series plot of incremental losses for each accident year and over development lag
- Left panel shows losses and right panel shows loss ratio
- We observe high volatility in early development lag that might explain the better performance of LN model



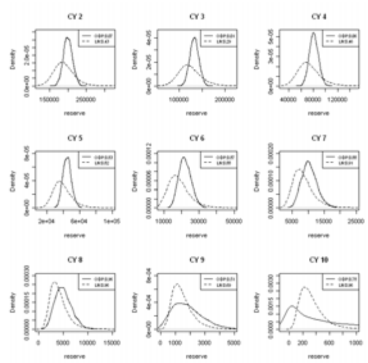
Commercial Auto – Insurer B

- We did similar analysis for insurer B and the results are summarized in the following three slides
- For insurer B, ODP outperforms LN model slightly
- Again we observe that the wider predictive distribution for early calendar years from LN model is explained by the wider distribution for early development year

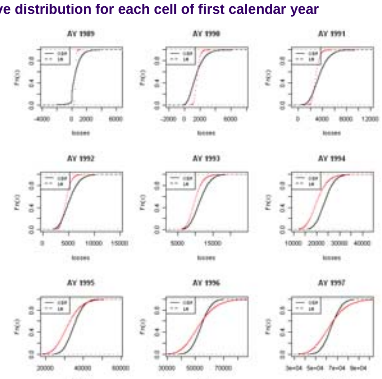
Commercial Auto – Insurer B



Predictive distribution for calendar year reserves

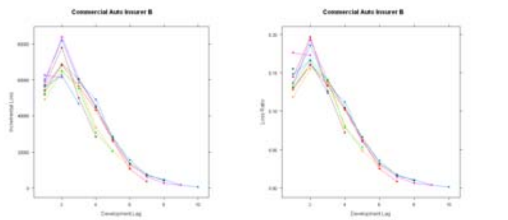


Predictive distribution for each cell of first calendar year



Commercial Auto – Insurer B

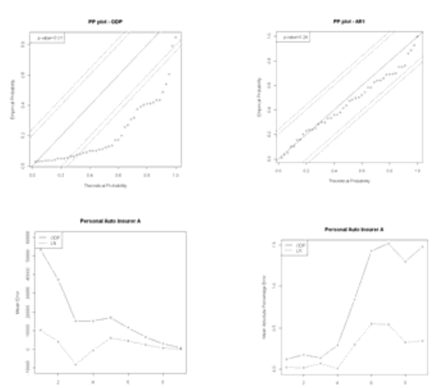
- Figures below show time-series plot of incremental losses for each accident year and over development lag
- Left panel shows losses and right panel shows loss ratio
- Different with insurer A, there is less volatility in early development years
- LN model might "underfit" the data



Personal Auto – Insurer A

- For insurer A, we derive the predictive distribution for each cell in the lower part of the triangle
- Then calculate the percentiles for actual incremental paid losses in the hold-out sample
- Uniform pp-plots of percentiles with the *p*-value of K-S tests are shown in next slide
- LN model outperforms ODP
- We also compare mean error and mean absolute percentage error of the two methods over the 9 testing periods
- For each testing period, LN performs better than ODP

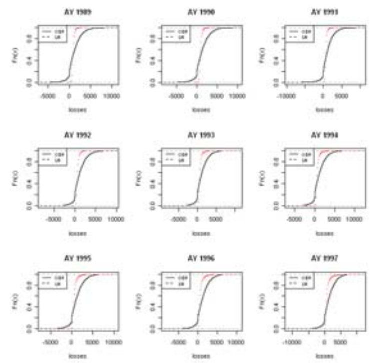
Personal Auto – Insurer A



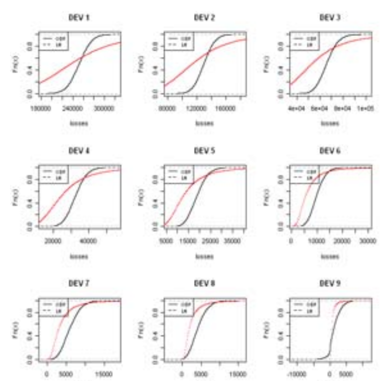
Personal Auto – Insurer A

- In the next two slides, we analyze the predictive distributions from the two methods
- 1st slide shows the predictive distributions for cells with development year 10
  - the predictive distribution for all accident year are similar
  - LN provides narrower distributions
- 2nd slide shows the predictive distribution for cells in accident year 1997
  - we want to see the effects over development lags
  - LN provides wider distribution for early development years and narrower distribution for later development years

Predictive distributions for cells with development year 10

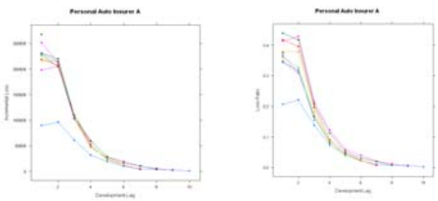


Predictive distributions for cells with accident year 1997



Personal Auto – Insurer A

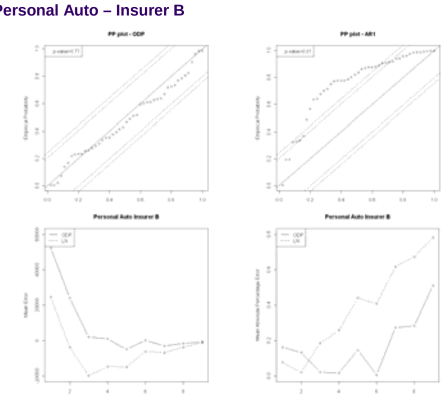
- Again look into the pattern of the training data
- We show time-series plot of incremental losses for each accident year and over development lag
- Left panel shows losses and right panel shows loss ratio
- Again the high volatility in early development lag that might explain the better performance of LN model



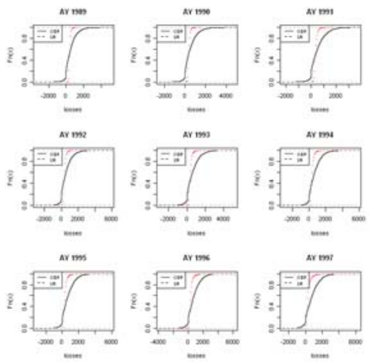
Personal Auto – Insurer B

- We did similar analysis for insurer B and the results are summarized in the following three slides
- For insurer B, ODP outperforms LN model
- From the predictive distributions, we observe again LN provides wider distributions for early development years, while the distributions across accident years are similar under two methods

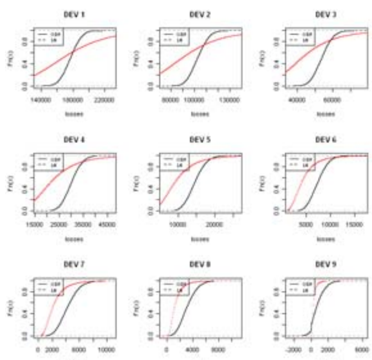
Personal Auto – Insurer B



Predictive distributions for cells with development year 10

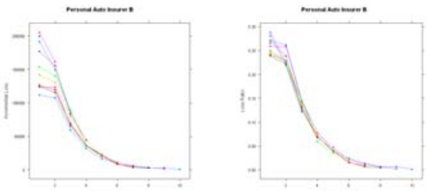


Predictive distributions for cells with accident year 1997



Personal Auto – Insurer B

- Figures below show time-series plot of incremental losses for each accident year and over development lag
- Left panel shows losses and right panel shows loss ratio
- Different with insurer A, there is less volatility in early development years, especially for the loss ratio
- Thus LN model introduces more volatility and does not work well for this insurer



Concluding Remarks

- We use simple test to examine the performance of loss reserving methods
- Our analysis is based on hold-out sample
- We find the current industry standard over-estimate reserves for the industry
- We compare chain ladder and LN model on individual insurers
- Chain ladder fails in case of higher volatility
- Bayesian methods mitigates the potential overfitting problem