

# The Study about Automobile Insurance Based on Linear Empirical Bayesian Estimation

Qiang Yu<sup>1</sup>, Zongjing Yao<sup>1</sup>, Fengyun Zhang<sup>2</sup>, Yujie Zhou<sup>3</sup>

<sup>1</sup>Department of Mathematics, Southwest University of Science and Technology, Mianyang, China

<sup>2</sup>Department of Mathematics, Jining University, Jining, China

<sup>3</sup>Finance Institute, Zhongnan University of Economics and Law, Wuhan, China

Email: yaozongjing@163.com

Received February 10, 2012; revised March 7, 2012; accepted March 14, 2012

## ABSTRACT

Automobile insurance is one of the most popular research areas, and there are a lot of different methods for it. We use linear empirical Bayesian estimation for the study of automobile insurance, giving the estimator of the policy's future claim size. Thus, a new point of view is given on the pricing of automobile insurance.

**Keywords:** Automobile Insurance; Linear Empirical Bayesian Estimation; Claim Size

## 1. Introduction

Automobile insurance is one of the most important insurance in property insurance. Recently, automobile insurance premium income grows steadily in our country. Since 2000, automobile insurance premium income accounted for the proportion of insurance premium income and has been maintained at above 60%. Cao Jing, Li Ping, Gao Yuan (2006) [1] studied the incentive contract in auto Insurance market based on the optimal the strategy game between the insurer and the insured, giving the linear relationship between the optimal indemnity and insurance premium when the insurance company have the biggest benefits. Xu Yun-Bao (2007) [2] established a dynamic theoretical model on the probabilistic moral hazard and the incomplete information between the insurer examination and verification influenced by insured person, studied the optimum games of strategy. Furthermore, insurance fixed price formula such that the insurer's expectation profit was given.

In the current, actuary pricing risk premium in auto insurance usually use the claim number mean and optimal estimation to calculate the claim amount. We commonly use the following methods to study the optimal estimation model of the claim size. For example: Poisson-inverse Gauss distribution, Mixed Pareto distribution Model, Mixed two-parameters Exponential distribution Model, Three parameters of Mixed Gamma distribution Model, Two dimensional Risk Model, Two negative binomial model, compound PNB distribution model, and so on. For the claim amount, there are only a few optimal estimation models, and now the relatively mature model is Pareto model [3]. However, when the Pareto model is

not suitable for the amount of the claim, we can only setting the price according to the number of the history claim pricing times the insured did, and the result may be that the claim amount is different but the amount of insurance premium is the same, which is obviously unfair. Yu Jiamin, Hao Xudong (2008) [4,5] elicited an optimal estimator of the policy's future claim size by constructing the structure function of claim size lognormal distribution parameter. An empirical rating actuarial model with allowance for claim frequency and claim severity was presented. Thus, the price on automobile insurance is more equitable and reasonable. Furthermore, a typical rating actuarial model fit for Chinese automobile insurance reality was put forward, preliminary to the promotion of rates allowance for claim severity. However, there still have parameter  $\mu$  in the optimal estimation of the future policy claim expectation when having the history claim information  $y_{10}, y_{20}, \dots, y_{n0}$ , we have to estimate the parameter  $\mu$ , and it's inconvenience for practical application.

Based on linear empirical Bayesian method (L.E.B method) [6], we study the estimation of the future claim amount, and the parameter estimation is obtained, the calculation is that the minimum mean square error  $\hat{\theta}$  is equal to the weighted average value between the true value  $\ln y$  and the experience value  $\ln \bar{y}$ , and the weight is respectively  $\left(1 - \frac{\sigma^2}{s^2}\right)$  and  $\frac{\sigma^2}{s^2}$ . By statistical theory, we indicate that this estimate is the optimal estimation method, at the same time it avoids the estimation of parameter  $\mu$ , and the result is not only concise but also convenient for practical application.

## 2. Linear Empirical Bayesian Method (L.E.B. Method)

First we introduce two Lemmas [6]:

**Lemma 1:** [6]  $x, y$  are two vectors. Suppose

$E \begin{vmatrix} x \\ p \times 1 \end{vmatrix}^2, E \begin{vmatrix} y \\ q \times 1 \end{vmatrix}^2$  are existed, then the following equation

$$E(y - f(x))'(y - f(x)) \geq E(y - \hat{y})'(y - \hat{y}) \quad (1)$$

is right for all functions  $f(x)$ , where  $f(x)$  is a vector, Representing the functions of  $x$  with the number  $q$ , and the second-order moments of  $f(x)$  are existed with  $\hat{y} = E(y/x)$ .

The Explaining of equation (1) is that: conditional expectation  $\hat{y} = E(y/x)$  is the closest function to  $y$  in all functions of  $x$ , and the Close degree can be measured with the mean square error  $E(y - f(x))'(y - f(x))$  between  $y$  and  $f(x)$ . However, it's difficult to calculate the conditional expectation, so if the linear functions are taken into account, the problem will be solved. This is the following conclusion:

**Lemma 2:** [6]  $x, y$  are two vectors. Suppose

$E \begin{vmatrix} x \\ p \times 1 \end{vmatrix}^2, E \begin{vmatrix} y \\ q \times 1 \end{vmatrix}^2$  are existed,  $A$  is a constant matrix,  $b$  is a

Constant vector, then

$$\min_{A,b} E(y - Ax - b)'(y - Ax - b) = E(y - \hat{y})'(y - \hat{y}) \quad (2)$$

in which

$$\hat{y} = E(y) + \text{cov}(y, x) \text{var}^{-1}(x)(x - E(x)) \quad (3)$$

The results show that  $\hat{y}$  is the closest function to  $y$  in all linear functions of  $x$ . with Lemma 2, if we replace  $y$  with the parameter vector  $\theta$  of statistical problems, and regard  $x$  as a sample, then the estimation of parameter  $\theta$  will be

$$\hat{\theta} = E(\theta) + \text{cov}(\theta, x) \text{var}^{-1}(x)(x - E(x)) \quad (4)$$

And it will be the minimum mean square error estimation. From Equation (4), if we can find the values of  $E(\theta)$ ,  $\text{cov}(\theta, x)$ ,  $\text{var}(x)$  and  $E(x)$  in the right equation, the estimation of parameter vector  $\theta$  can be calculated. If we suppose variable  $X$ -normal distribution  $N(\theta, \sigma^2)$  for which the variance  $\sigma^2$  is known, with the above results, we can give the estimation expression of parameter  $\theta$ . With  $p(x/\theta)$  is known in statistical problems, and  $p(x/\theta)$  is a conditional probability density with parameter  $\theta$ , so  $E(x/\theta)$  and  $\text{var}(x/\theta)$  can be evaluated.

**Theorem 1:** If the conditional probability density of normal distribution  $N(\theta, \sigma^2)$  is  $p(x/\theta)$ , then  $p(x/\theta)$

have the following two properties:

$$E(x/\theta) = \theta \quad (5)$$

$$\text{var}(x/\theta) = \sigma^2 \quad (6)$$

And the estimation of parameter vector  $\theta$  will be

$$\hat{\theta} = \left(1 - \frac{\sigma^2}{\text{var}(x)}\right)x + \frac{\sigma^2}{\text{var}(x)}E(x) \quad (7)$$

Proof: With the conditions in the theorem, we can get

$$E(\theta) = E(E(x/\theta)) = E(x) \quad (8)$$

$$\begin{aligned} \text{cov}(\theta, x) &= E\left(E\left\{(\theta - E(x))(\theta - E(x))' / \theta / \theta\right\}\right) \\ &= E\left\{(\theta - E(\theta))(E(x/\theta) - E(x))'\right\} \\ &= E(\theta - E(\theta))(\theta - E(\theta))' \\ &= \text{var}(\theta) \end{aligned} \quad (9)$$

$$\begin{aligned} \text{var}(x) &= E\left(E\left\{xx' - x(Ex)' - (Ex)x' + (Ex)(Ex)' / \theta / \theta\right\}\right) \\ &= E\left\{\text{var}(x/\theta) + E\{x/\theta\}E\{x/\theta\}' - E\{x/\theta\}(E\theta)'\right. \\ &\quad \left. - (E\theta)E\{x/\theta\}' + (E\theta)(E\theta)'\right\} \\ &= \text{var}(\theta) = E \text{var}(E\theta) + \text{var}(\theta) \end{aligned}$$

So the variance

$$\text{var}(\theta) = \text{var}(x) - E \text{var}(x/\theta) = \text{var}(x) - \sigma^2 \quad (10)$$

Now we simultaneously substitute Equation (10) and Equation (8), Equation (9) on Equation (4), and we gain parameter estimation:

$$\begin{aligned} \hat{\theta} &= Ex + \text{var}(\theta) \text{var}^{-1}(x)(x - Ex) \\ &= \left(1 - \frac{\sigma^2}{\text{var}(x)}\right)x + \frac{\sigma^2}{\text{var}(x)}Ex \end{aligned}$$

From Equation (7), as long as we have the history data  $x_1, x_2, \dots, x_n$  of the sample, we can figure out that the

sample mean is  $\bar{x} = \frac{\sum_{i=1}^n \ln x_i}{n}$ , and the sample variance is

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}.$$

With classical statistics methods, we can estimate  $Ex$  with  $\bar{x}$ , while we can estimate  $\text{var}(x)$  with  $S^2$ , then

$$\hat{\theta} = \left(1 - \frac{\sigma^2}{S^2}\right)x + \frac{\sigma^2}{S^2}\bar{x}.$$

### 3. The Optimal Estimation of Insurance Slip's Future Claim Amount

Assuming that the amount claim  $Y$  is subject to log-normal distribution with the parameters  $\theta, \sigma^2$ , with the variance  $\sigma^2$  is known, and its probability density function is

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma y}} e^{-\frac{(\ln y - \theta)^2}{2\sigma^2}} \tag{11}$$

for which  $\theta \in R, \sigma^2 \in R^+, y \in R^+$ . With Equation (11), the expectations of the future claimed amount for a single insurance slip in insurance companies will be

$$\begin{aligned} EY &= \int_{-\infty}^{+\infty} y f_Y(y) dy = \int_{-\infty}^{+\infty} y \frac{1}{\sqrt{2\pi\sigma y}} e^{-\frac{(\ln y - \theta)^2}{2\sigma^2}} dy \\ &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(\ln y - \theta)^2}{2\sigma^2}} dy = e^{\theta + \frac{\sigma^2}{2}} \end{aligned} \tag{12}$$

Thus if we get the optimal estimation of the parameter  $\theta$ , we can calculate the expectation of  $Y$ .

Suppose the Claim history of a single insurance slip in the insurance company is  $y_1, y_2, \dots, y_n$ , and assume the current new observation is a vector  $y$ . Now we Calculate the estimation of parameter  $\theta$  by using the L.E.B method in the classic statistics. Because when a single insured claim occurs, the claim amount of money  $Y$  is subject to log-normal distribution with the parameters  $(\theta, \sigma^2)$ , that is  $\ln Y \sim N(\theta, \sigma^2)$ . So we can get that  $E(\ln Y / \theta) = \theta, \text{var}(\ln Y / \theta) = \sigma^2$ .

When variance  $\sigma^2$  is known, we get the estimation of the parameter  $\theta$  with

$$\begin{aligned} \hat{\theta} &= \overline{\ln y} + \left(1 - \frac{\sigma^2}{s^2}\right) (\ln y - \overline{\ln y}) \\ &= \left(1 - \frac{\sigma^2}{s^2}\right) \ln y + \frac{\sigma^2}{s^2} \overline{\ln y} \end{aligned} \tag{13}$$

For which

$$\overline{\ln y} = \frac{1}{n} \sum_{i=1}^n \ln y_i, s^2 = \frac{1}{n-1} \sum_{i=1}^n (\ln y_i - \overline{\ln y})^2$$

so when we use L.E.B. method to calculate the estimation  $\hat{\theta}$  of the parameter  $\theta$ ,  $\hat{\theta}$  is equal to the weighted average value between the true value  $\ln y$  and the experience value  $\overline{\ln y}$ , and the weight is respectively  $\left(1 - \frac{\sigma^2}{s^2}\right)$  and  $\frac{\sigma^2}{s^2}$ . So when the Claim history of a single insurance slip is  $y_1, y_2, \dots, y_n$ , and assume the current new observation is a vector  $y$ , from Equation (12) we can get that the expected estimation of the future claimed amount for a single insurance slip is

$$EY = e^{\left(1 - \frac{\sigma^2}{s^2}\right) \ln y + \frac{\sigma^2}{s^2} \overline{\ln y} + \frac{\sigma^2}{2}} \tag{14}$$

### 4. Model Application-Automobile Insurance Premium Prediction

Now suppose that automobile insurance claim number  $X$  obey mixed three parameter gamma distribution, and the historic claim frequency information for insurance slip is  $k_1, k_2, \dots, k_n$ , we can calculate the optimal claim frequency estimation for  $\hat{k} = \hat{\beta} + \frac{\hat{\alpha}}{\hat{\gamma}}$  (seeing Reference [7]).

While it is assumed that the claim amount  $Y$  is subject to lognormal distribution with parameter  $(\theta, \sigma^2)$ , and when the history Claim of a single insurance slip is  $y_1, y_2, \dots, y_n$ , and assume the current new observation is a vector  $y$ , we can get that the expected estimation of the

future claimed amount is  $e^{\left(1 - \frac{\sigma^2}{s^2}\right) \ln y + \frac{\sigma^2}{s^2} \overline{\ln y} + \frac{\sigma^2}{2}}$  based on L.E.B. method, so we can calculate the Optimal estimation of future premiums with

$$P = \left(\hat{\beta} + \frac{\hat{\alpha}}{\hat{\gamma}}\right) \cdot e^{\left(1 - \frac{\sigma^2}{s^2}\right) \ln y + \frac{\sigma^2}{s^2} \overline{\ln y} + \frac{\sigma^2}{2}}$$

### 5. The End

In recent years, with the rapid increase of motor vehicles, automobile insurance premium income accounted for the proportion of insurance income has been gradually improved, thus insurance premium price becomes particularly important. This paper mainly studies the policy's future claim amount estimation value, and on this basis, predicting the premium price through the establishment of model. Through the use of linear empirical Bayesian method (L.E.B method), given the parameter estimation of  $\theta$  in the lognormal distribution with its parameter  $(\theta, \sigma^2)$ . the result is that the minimum mean square error estimation  $\hat{\theta}$  is the weighted average value of the observed value  $\ln y$  and the experience value  $\overline{\ln y}$ , with its weights  $\left(1 - \frac{\sigma^2}{s^2}\right)$  and  $\sigma^2/s^2$ . At the same time,

according to (12), we calculate the policy's future claim amount expected value estimation is

$$\exp\left\{\left(1 - \sigma^2/s^2\right) \ln y + \left(\sigma^2/s^2\right) \overline{\ln y} + \sigma^2/2\right\}$$

Finally, in the hypothesis that claim number  $X$  obey three parameters with mixed gamma distribution, gets the optimal premium predictive value for insurance premium, provides certain theory basis in the pricing problem.

### 6. Acknowledgements

This work was supported by the Sichuan Provincial Office of education projects for Humanities and Social Sciences LY09-10.

### REFERENCES

- [1] J. Cao, P. Li and Y. Gao, "The Strategy Game Based on the Optimal Incentive Contract in Auto Insurance Market," *Statistics and Decision*, Vol. 20, 2006, pp. 52-53.
- [2] Y.-B. Xu, Y.-X. Tian and D. Zhu, "The Property Insurance Pricing Research under Moral Hazard of the Insured Affecting Losing Probability," *The Second International Workshop on Intelligent Finance*, Chengdu, 6-8 July 2007.
- [3] S.-W. Meng and W. Yuan, "Bonusmalus Systems in Automobile Insurance," *Chinese Journal of Applied Probability and Statistics*, Vol. 15, No. 1, 1999, pp. 48-52.
- [4] X.-D. Hao and J.-M. Yu, "An Empirical Rating Actuarial Model with Claim Size Following Lognormal Distribution in Automobile Insurance," *Journal of Harbin University of Commerce (Natural Sciences Edition)*, Vol. 24, No. 3, 2008, pp. 382-384.
- [5] J.-M. Yu and X.-D. Hao, "An Empirical Rating Actuarial Model with Claim Size Following Lognormal Distribution in Automobile Insurance," *Journal of Shanghai Jiaotong University (Science)*, Vol. 42, No. 11, 2008, pp. 1836-1838.
- [6] X.-T. Zhang and H.-F. Chen, "Bayesian Statistical Inference," Science and Technology Press, 1991.
- [7] P. Hu and Z.-H. Yuan, "Claimant Number of Times Essence Studies When Risk Is Not Same Quality," *Statistics and Decision*, Vol. 11, 2007, pp. 21-22.