

Application of Constrained Bayes Estimation under Balanced Loss Function in Insurance Pricing

Myung Joon Kim^a, Yeong-Hwa Kim^{1,b}

^aDepartment of Business Statistics, Hannam University, Korea

^bDepartment of Applied Statistics, Chung-Ang University, Korea

Abstract

Constrained Bayesian estimates overcome the over shrinkness toward the mean which usual Bayes and empirical Bayes estimates produce by matching first and second empirical moments; subsequently, a constrained Bayes estimate is recommended to use in case the research objective is to produce a histogram of the estimates considering the location and dispersion. The well-known squared error loss function exclusively emphasizes the precision of estimation and may lead to biased estimators. Thus, the balanced loss function is suggested to reflect both goodness of fit and precision of estimation. In insurance pricing, the accurate location estimates of risk and also dispersion estimates of each risk group should be considered under proper loss function. In this paper, by applying these two ideas, the benefit of the constrained Bayes estimates and balanced loss function will be discussed; in addition, application effectiveness will be proved through an analysis of real insurance accident data.

Keywords: Balanced loss function, constrained Bayes estimate, insurance pricing.

1. Introduction

Bayesian and empirical Bayesian methods have become quite popular in the theory and practice of statistics over the last three decades. In particular, hierarchical and empirical Bayesian methods are suitable in the context of simultaneous estimation when there is a genuine need for borrowing strength. However, the goal is often to produce an ensemble of estimates simultaneously for several parameters as well as to produce a set of estimates whose empirical histogram adequately estimates the histogram of population parameters.

The well-known usual Bayes estimates under quadratic loss turn out to be posterior means which are the optimal estimates under any quadratic loss. But the histogram of the posterior means of parameters is underdispersed as an estimate of the histogram of parameters. Accordingly, the histogram of posterior means is clearly inappropriate to estimate the parameter histogram. Thus, there is a need to find a set of suboptimal estimates to compromise between two criteria. Louis (1984) proposed a constrained Bayes method which matches the first two empirical moments of Bayes estimates of the normal means with minimizing the squared distance of the parameters and estimates subject to these constraints. After that Ghosh (1992) generalized the Louis' result for any arbitrary distribution not necessarily normal distribution and also Ghosh and Kim (2002) have extended the result for vector-valued parameters.

This paper has been supported by the 2014 Hannam University Research Fund.

¹ Corresponding author: Department of Applied Statistics, Chung-Ang University, 221 Heuksuk-dong, Dongjak-gu, Seoul 156-756, Korea. E-mail: gogators@cau.ac.kr

The main idea of constrained Bayes estimators is as follows. Consider the situation where the data are denoted by \mathbf{x} and $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)^T$ is the parameter of interest. The basic idea is to find an estimate of parameter $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)^T$ which minimizes

$$E \left\{ \sum_{i=1}^m (\theta_i - t_i)^2 | \mathbf{x} \right\}, \quad (1.1)$$

within the class of all estimates $\mathbf{t}(\mathbf{x}) = \mathbf{t} = (t_1, \dots, t_m)^T$ of $\boldsymbol{\theta}$ that satisfy these two conditions

$$(1) E(\bar{\theta} | \mathbf{x}) = m^{-1} \sum_{i=1}^m t_i(\mathbf{x}) = \bar{t}(\mathbf{x}), \quad (1.2)$$

$$(2) E \left\{ \sum_{i=1}^m (\theta_i - \bar{\theta})^2 | \mathbf{x} \right\} = \sum_{i=1}^m \{t_i(\mathbf{x}) - \bar{t}(\mathbf{x})\}^2. \quad (1.3)$$

Let $\hat{\boldsymbol{\theta}}^B$ denote Bayes estimate of $\boldsymbol{\theta}$ under any quadratic loss, it is easy to see that the Bayes estimate satisfies (1.2); however it does not, satisfy (1.3) and always estimates a smaller value than true variation. That is the limitation of usual Bayes estimates to estimate the true variation of parameter $\boldsymbol{\theta}$ and the reason why a constrained Bayes estimates should be considered.

Balanced loss functions were introduced by Zellner (1988). Such losses are formulated to reflect two criteria, which is goodness of fit and precision of estimation. As noted by Zellner (1988), least squares estimators reflect goodness of fit consideration, while quadratic losses such as the squared error loss are geared solely towards the precision of estimation.

It is well recognized that sole emphasis on the precision of estimation criterion, for example mean squared error, can lead to biased estimators. In some circumstances bias is not important, but in others, it is critical. Thus there is a need to provide a framework within which the tradeoff between goodness of fit, or lack of bias, and precision of estimation can be considered formally. Zellner (1992) suggested a balanced loss function which meets those needs.

The balanced loss function is defined for any estimator \mathbf{d} of $\boldsymbol{\theta}$ as

$$L(\boldsymbol{\theta}, \mathbf{d}) = w \|\mathbf{x} - \mathbf{d}\|^2 + (1 - w) \|\boldsymbol{\theta} - \mathbf{d}\|^2, \quad (1.4)$$

where $\|\cdot\|$ denotes the Euclidean norm and w is weight.

The first term on the right hand side represents goodness of fit while the second represents precision of estimation. Under balanced loss function, the Bayes estimates are not posterior means of parameters of interest anymore and can be produced as:

$$\tilde{\boldsymbol{\theta}}_* = w \bar{\mathbf{x}} + (1 - w) \hat{\boldsymbol{\theta}}^B, \quad (1.5)$$

where $\hat{\boldsymbol{\theta}}^B$ denotes the Bayes estimator under quadratic loss. If $w = 1$ in (1.5), $\tilde{\boldsymbol{\theta}}_* = \bar{\mathbf{x}}$, while if $w = 0$, $\tilde{\boldsymbol{\theta}}_* = \hat{\boldsymbol{\theta}}^B$, which is usual Bayes estimator under quadratic loss. Since if $w = 0$ in (1.4), balanced loss function is same as squared error loss.

In Section 2, the constrained Bayes estimation under two different loss functions such as squared error loss and balanced loss function is described in detail. In Section 3, those estimators with ANOVA model under both loss functions will be introduced. After that, benefits of insurance pricing application with described estimates and loss function is discussed with a numerical results comparison.

2. Constrained Bayes Estimation

2.1. Constrained Bayes estimation under squared error loss

The main idea of constrained Bayes estimate of parameter θ is to derive a solution that minimizes (1.1) subject to (1.2) and (1.3). That is, let $\mathcal{X}_0 = \{\mathbf{x} : H_2(\mathbf{x}) = \sum_{i=1}^m (\theta_i^B(\mathbf{x}) - \bar{\theta}^B(\mathbf{x}))^2 > 0\}$, then for $\mathbf{x} \in \mathcal{X}_0$, the solution t_i of (1.1) which satisfies (1.2) and (1.3) constraints. Let $\theta^{CB}(\mathbf{x}) = (\theta_1^{CB}(\mathbf{x}), \theta_2^{CB}(\mathbf{x}), \dots, \theta_m^{CB}(\mathbf{x}))^T$ be the constrained Bayes estimate which is satisfying those two conditions, the main result is proved by Ghosh (1992) and can be stated as:

$$\begin{aligned} \theta_i^{CB}(\mathbf{x}) &= a\theta_i^B(\mathbf{x}) + (1-a)\bar{\theta}^B(\mathbf{x}), \quad i = 1, \dots, m, \\ a \equiv a(\mathbf{x}) &= \left[1 + \frac{H_1(\mathbf{x})}{H_2(\mathbf{x})} \right]^{\frac{1}{2}}, \\ H_1(\mathbf{x}) &= \text{tr} \left[V(\theta - \bar{\theta} \mathbf{1}_m | \mathbf{x}) \right] = \text{tr} \left[(\mathbf{I}_m - m^{-1} \mathbf{J}_m) V(\theta | \mathbf{x}) \right], \\ H_2(\mathbf{x}) &= \sum_{i=1}^m (\theta_i^B(\mathbf{x}) - \bar{\theta}^B(\mathbf{x}))^2. \end{aligned} \tag{2.1}$$

We point out the clear limitations of usual Bayes estimates to estimate true variation that constrained Bayes estimates focus on. Let \mathbf{I}_m be the identity matrix of order m , $\mathbf{1}_m$ be the m -component column vector with each element equal to 1 and $\mathbf{J}_m = \mathbf{1}_m \mathbf{1}_m^T$. Then variation can be calculated as

$$\begin{aligned} E \left\{ \sum_{i=1}^m (\theta_i - \bar{\theta})^2 | \mathbf{X} \right\} &= \text{tr} \left\{ \left(\mathbf{I}_m - \frac{1}{m} \mathbf{J}_m \right) E(\theta \theta^T | \mathbf{X}) \right\} \\ &= \text{tr} \left\{ \left(\mathbf{I}_m - \frac{1}{m} \mathbf{J}_m \right) \{ V(\theta | \mathbf{X}) + E(\theta | \mathbf{X}) E(\theta | \mathbf{X})^T \} \right\} \\ &= \text{tr} \left\{ \left(\mathbf{I}_m - \frac{1}{m} \mathbf{J}_m \right) V(\theta | \mathbf{X}) \right\} + \sum_{i=1}^m \{ \theta_i^B(\mathbf{X}) - \bar{\theta}^B(\mathbf{X}) \}^2 \\ &= \text{tr} \{ V(\theta - \bar{\theta} \mathbf{1}_m | \mathbf{X}) \} + \sum_{i=1}^m \{ \theta_i^B(\mathbf{X}) - \bar{\theta}^B(\mathbf{X}) \}^2 \\ &> \sum_{i=1}^m \{ \theta_i^B(\mathbf{X}) - \bar{\theta}^B(\mathbf{X}) \}^2. \end{aligned} \tag{2.2}$$

This calculation clearly shows the underdispersion of usual Bayes estimates and the reason why constrained Bayes estimates should be considered in some circumstances. Under the assumption of normal distribution as a prior distribution, where $X_i | \theta_i$ are independent $N(\theta_i, 1)$, $i = 1, \dots, m$ while θ_i are iid $N(\mu, A)$, the constrained Bayes estimator of θ can be expressed as:

$$\begin{aligned} \hat{\theta}^{CB} &= a_B [(1-B)\mathbf{X} + B\mu \mathbf{1}_m] + (1-a_B) [(1-B)\bar{X} + B\mu] \mathbf{1}_m \\ &= (1-B) [a_B \mathbf{X} + (1-a_B) \bar{X} \mathbf{1}_m] + B\mu \mathbf{1}_m, \end{aligned} \tag{2.3}$$

where $B = (1+A)^{-1}$, $a_B^2 = 1 + 1/[(1+B)S]$ and $S = \sum_{i=1}^m (X_i - \bar{X})^2$. However, the constrained Bayes estimators cannot claim any risk improvement over the Bayes estimators, produce to meet the twin objectives as mentioned earlier more satisfactorily. In normal distribution assumption given

as above, typically both μ and A are unknown, and are estimated from the marginal distribution of $\mathbf{X} = (X_1, \dots, X_m)^T$. Marginally, $\mathbf{X} \sim N(\mu \mathbf{1}_m, B^{-1} \mathbf{I}_m)$. Hence, marginally, $S \sim B^{-1} \chi_{m-1}^2 / (m-1)$. We estimate μ by \bar{X} , and B by $\hat{B} = 1/S$. Let $\hat{\theta}^{CEB}$ be the constrained empirical Bayes estimator of parameter θ , then

$$\hat{\theta}^{CEB} = (1 - \hat{B}) [a_{EB} \mathbf{X} + (1 - a_{EB}) \bar{X} \mathbf{1}_m] + \hat{B} \bar{X} \mathbf{1}_m, \quad (2.4)$$

where a_{EB} replaces B by \hat{B} in a_B .

Up to this, constrained Bayes and constrained empirical Bayes estimators have been derived under squared error loss. However Louis (2001) pointed out there is a need to develop such estimators for other losses as well. One such loss considered is the so-called balanced loss function. In the next section, constrained Bayes and constrained empirical Bayes estimators under balanced loss function will be considered.

2.2. Constrained Bayes estimation under balanced loss function

Let $\mathbf{X} = (X_1, X_2, \dots, X_m)^T$ with $E(\mathbf{X}) = \theta = (\theta_1, \theta_2, \dots, \theta_m)^T$. For any estimator $\mathbf{e} = (e_1, e_2, \dots, e_m)^T$ of θ , the balanced loss function as introduced by Zellner (1988, 1992) is

$$L(\theta, \mathbf{e}) = m^{-1} [w \|\mathbf{X} - \mathbf{e}(\mathbf{X})\|^2 + (1 - w) \|\mathbf{e}(\mathbf{X}) - \theta\|^2], \quad (2.5)$$

where $\|\cdot\|$ is the Euclidean norm and w ($0 \leq w \leq 1$) is the known weight. The choice of w reflects the relative weight which the experimenter wants to assign to goodness of fit and precision of estimation. The extreme cases $w = 1$ and $w = 0$ refer solely to the precision of an estimate and goodness of fit respectively.

Under the balanced loss function, constrained Bayes estimator and asymptotic version of Bayes risk is proved by Ghosh *et al.* (2007). We need some notations to state the derived result. Let $H_1(\mathbf{x}) = \sum_{i=1}^m V(\theta_i - \bar{\theta} | \mathbf{x})$, $H_2(\mathbf{x}) = \sum_{i=1}^m (e_i^{PM}(\mathbf{x}) - \bar{e}^{PM}(\mathbf{x}))^2 = (1 - B)^2 \sum_{i=1}^m (x_i - \bar{x})^2$, where e_i^{PM} indicates the posterior mean of parameter and now writing $a(\mathbf{x}) = [1 + H_1(\mathbf{x})/H_2(\mathbf{x})]^{1/2}$, then the constrained Bayes estimator estimators of parameter θ_i is given by

$$\begin{aligned} \hat{\theta}_i^{CB} &= a(\mathbf{x}) (e_i^{PM}(\mathbf{x}) - \bar{e}^{PM}(\mathbf{x})) + \bar{e}^{PM}(\mathbf{x}) \\ &= a(\mathbf{x}) e_i^{PM}(\mathbf{x}) + (1 - a(\mathbf{x})) \bar{e}^{PM}(\mathbf{x}). \end{aligned} \quad (2.6)$$

This is same as the expression obtained assuming squared error loss. In a way, this demonstrates the loss robustness of the constrained Bayes estimator. As pointed out previously, this is not the case with the Bayes estimator, which changes according to the weight values of w .

Now considering same assumption with normal distribution in the previous section, the function $H_1(\mathbf{x})$ is simplified as $H_1(\mathbf{x}) = \sum_{i=1}^m V(\theta_i - \bar{\theta} | \mathbf{x}) = (m-1)(1-B)$ and the posterior mean of θ can be expressed as $e^{PM}(\mathbf{X}) = (1-B)\mathbf{X} + B\mu \mathbf{1}_m$ with Bayes risk $1-B$. we can express (2.6) as follows.

$$\begin{aligned} \hat{\theta}^{CB} &= a(\mathbf{X}) [(1-B)\mathbf{X} + B\mu \mathbf{1}_m] + (1 - a(\mathbf{X})) [(1-B)\bar{X} + B\mu] \mathbf{1}_m \\ &= a(\mathbf{X})(1-B) (\mathbf{X} - \bar{X} \mathbf{1}_m) + [(1-B)\bar{X} + B\mu] \mathbf{1}_m. \end{aligned} \quad (2.7)$$

By substituting \hat{B} for B and \bar{X} for μ in constrained Bayes estimators. Accordingly, constrained empirical Bayes estimator $\hat{\theta}^{CEB}$ under balanced loss function is

$$\hat{\theta}^{CEB} = a_{EB} [(1 - \hat{B}) \mathbf{X} + \hat{B} \bar{X} \mathbf{1}_m] + (1 - a_{EB}) \bar{X} \mathbf{1}_m, \quad (2.8)$$

where $a_{EB} = [1 + (1 - \hat{B})^{-1}S^{-1}]^{1/2}$.

Louis (1984) proposed the constrained empirical Bayes estimators in the original James-Stein framework and Ghosh (1992) developed the constrained Bayes estimators in a more general framework.

3. Constrained Bayes Estimation for One-Way ANOVA Model under Balanced Loss Functions

One of the main issues in the insurance pricing is the accurate risk estimation for each segmented group, which can be applied with the ANOVA model. Each group risk could be evaluated properly by trying to minimize the within variance for each group as well as to maximize the between variance. Also estimation of the risk variation should be carefully considered, since the segmented group risk may be changed every policy year. So, we will introduce several Bayes and empirical Bayes estimators with the ANOVA model under two different functions such as squared error and balanced loss function.

Consider the balanced normal ANOVA model with $Y_{ij} = \theta_i + e_{ij}$ and $\theta_i = \mu + \alpha_i$ ($i = 1, \dots, m; j = 1, \dots, k$). Notation i indicates the number of risk group and j indicates the number of observations. Here we are assuming the α_i and the e_{ij} are mutually independent with $\alpha_i \sim iid N(0, \tau^2)$ and $e_{ij} \sim iid N(0, \sigma^2)$. In a Bayesian framework, we can easily show that $Y_{ij}|\theta_i \sim iid N(\theta_i, \sigma^2)$, $i = 1, \dots, m$ and $\theta_i \sim iid N(\mu, \tau^2)$. Minimal sufficiency consideration allows us to restrict to (X_1, \dots, X_m, SSW) , where $X_i = 1/k \sum_{j=1}^k Y_{ij} = \bar{Y}_i$ and $SSW = \sum_{i=1}^m \sum_{j=1}^k (Y_{ij} - \bar{Y}_i)^2$. We may note that marginally X_1, \dots, X_m and SSW are mutually independent with $X_i \sim iid N(\mu, \tau^2 + \sigma^2/k)$, i.e., $N(\mu, \sigma^2/(kB))$, where $B = \{\sigma^2/k\}/\{\sigma^2/k + \tau^2\} = \sigma^2/\{\sigma^2 + k\tau^2\}$ and $SSW \sim \sigma^2\chi_{m(k-1)}^2$.

Usual Bayes estimator under squared error with this assumptions, can be easily derived as a posterior mean and also substituting the μ as \bar{X} and B as empirical Bayes estimator is given by,

$$\begin{aligned} \hat{\theta}^B &= (1 - B)X + B\mu\mathbf{1}_m, \\ \hat{\theta}^{EB} &= (1 - \hat{B})X + \hat{B}\bar{X}\mathbf{1}_m, \end{aligned} \tag{3.1}$$

where $X = (X_1, \dots, X_m)^T$, $\bar{X} = m^{-1} \sum_{i=1}^m X_i$, and $\mathbf{1}_m$ is an m -component column vector with each element equal to 1.

Under balanced loss function, those estimators are changed as:

$$\begin{aligned} \hat{\theta}^B(B) &= wX + (1 - w)\hat{\theta}^B = [1 - (1 - w)B]X + (1 - w)B\mu\mathbf{1}_m, \\ \hat{\theta}^{EB}(B) &= (1 - (1 - w)\hat{B})X + (1 - w)\hat{B}\bar{X}\mathbf{1}_m. \end{aligned} \tag{3.2}$$

Here, $\hat{\theta}^B(B)$ and $\hat{\theta}^{EB}(B)$ indicates the Bayes and empirical Bayes estimators under balanced loss function; in addition, note that it is the same function under the squared error loss function in the case of $w = 0$.

Developing with the constrained Bayes estimators in the balanced normal ANOVA model under balanced loss function, from the results of the previous section, the constrained Bayes estimator of $\theta = (\theta_1, \dots, \theta_m)^T$ is given by

$$\hat{\theta}^{CB} = a(X)(1 - B)(X - \bar{X}\mathbf{1}_m) + \{(1 - B)\bar{X} + B\mu\}\mathbf{1}_m, \tag{3.3}$$

where $a^2(X) = 1 + H_1(X)/H_2(X)$ and $H_1(X) = (m-1)(1-B)\sigma^2/k$ and $H_2(X) = (1-B)^2 \sum_{i=1}^m (X_i - \bar{X})^2 = (1-B)^2(\text{SSB}/k)$. Then, on simplification we can get,

$$a^2(X) = 1 + \frac{\sigma^2}{(1-B)\text{MSB}}, \quad (3.4)$$

where $\text{MSB} = \text{SSB}/(m-1)$ and $\text{MSW} = \text{SSW}/(k-1)$. However, Bayes and empirical Bayes estimators are differently estimated by the loss function, the results of constrained Bayes estimator are same as the squared error loss due to the robustness of loss function, which is the characteristic that the constrained Bayes estimators have.

At the result of equation (3.4), by substitution of μ by \bar{X} , B by \hat{B} and σ^2 by MSW . Here $\hat{B} = \{(m-3)\text{MSW}\}/\{(m-1)\text{MSB}\}$, the constrained empirical Bayes estimator of θ can be derived as

$$\hat{\theta}^{CEB} = a_{EB}(X) (1 - \hat{B}) (X - \bar{X}\mathbf{1}_m) + \bar{X}\mathbf{1}_m, \quad (3.5)$$

where $a_{EB}(X) = 1 + \text{MSW}/\{(1 - \hat{B})\text{MSB}\}$.

In the next section, with the balanced normal ANOVA model assumption, numerical analysis will be implemented with auto insurance accident data for comparison and will be discussed regarding the benefits of the estimation and loss function application.

4. Data Analysis and Result

Data analysis will be performed based on auto insurance paid loss data. Insurance company usually estimates the premium based on empirically arised accident information. The main goal of estimation to the insurance company is to find out the true risk existed in each risk group.

As discussed previous section, several estimators under two different loss functions with ANOVA model will be estimated for the comparison of the accuracy and properness. For doing this, we need some index to compare. Index suggested by Kim and Kim (2013b) will be used which is considering the two first moments.

$$\text{Index} = w \left[E(\bar{\theta}|\mathbf{x}) - m^{-1} \sum_{i=1}^m t_i(\mathbf{x}) \right]^2 + (1-w) \left[\left\| E \left[\sum_{i=1}^m (\theta_i - \bar{\theta})^2 | \mathbf{x} \right] - \sum_{i=1}^m [t_i(\mathbf{x}) - \bar{t}(\mathbf{x})]^2 \right\| \right], \quad (4.1)$$

The index function in (4.1) is consists of two pair. First term is considering the how well matching the first moment and the second term is how well matching the second moment. The research purpose is to produce the accurate location parameter as well as a proper dispersion parameter, the index is suitable for the comparison.

Regarding the balanced loss function, the value of estimators can be changed according to the weight in the loss function. In this paper, we consider same weight for the goodness of fit and precision of estimation. So the loss function used for the numerical analysis implementation can be given by,

$$L(\theta, \mathbf{d}) = \frac{1}{2} \|\mathbf{x} - \mathbf{d}\|^2 + \frac{1}{2} \|\theta - \mathbf{d}\|^2. \quad (4.2)$$

4.1. Data descriptions

Auto insurance coverage usually consists of five components (bodily injury liability, property damage liability, medical expense, uninsured and collision coverage). Property damage liability coverage

Table 1: Data structure (n denotes the sample size of each group and here $n = 5,000$)

N	Group			
	Small size car	Medium size car	Large size car	SUV & Van type
1	y_{11}	y_{21}	y_{31}	y_{41}
2	y_{12}	y_{22}	y_{32}	y_{42}
\vdots	\vdots	\vdots	\vdots	\vdots
n	y_{1n}	y_{2n}	y_{3n}	y_{4n}

Table 2: Numerical analysis result.

Estimates	Group				Index
	Small size car	Medium size car	Large size car	SUV & Van type	
θ	488,940	540,360	597,200	660,000	
$\hat{\theta}^B(S)$	494,970	534,020	592,010	576,030	1.0161
$\hat{\theta}^{EB}(S)$	495,480	534,160	591,550	575,740	1.0163
$\hat{\theta}^B(B)$	494,960	534,020	592,010	576,040	1.0161
$\hat{\theta}^{EB}(B)$	495,220	534,090	591,780	575,890	1.0162
$\hat{\theta}^{CB}$	494,320	533,840	592,600	576,410	1.0159
$\hat{\theta}^{CEB}$	494,750	533,960	592,210	576,160	1.0161

data will be used in this paper. The reason for the selection is characteristics of the coverage having short term settlement and mandatory policy. Number of rating factors, auto insurance companies are adopting currently more than thirty, in this study, among them, the one of main and basic factor in auto insurance, car type-small size car, medium size car incidental factor will be considered.

The data used for the analysis comes from property damage liability coverage of auto insurance accident at the ‘S’ fire insurance company. The data is based on the accident year basis and; in addition, 5,000 corresponding samples are randomly selected from real accident company data at each group divided by car characteristics for the consideration of the model used in this study. Table 1 illustrates the data structures. In the Table 1, y_{ij} indicates the loss amount during the policy period by accident of each policy and i indicates kinds of car type, and j indicates the loss observations in each car type. So in our simulation study, m is equal to 4 and n is equal to 5,000. As many sample observations as possible are considered under the restriction of the balancing ANOVA model and a log transformation is implemented for the normality assumption.

From the data structure described above, by assuming the parameter empirically, numerical analysis regarding the ANOVA model will be implemented and the results are the estimated risks for the each car type. Based on the estimated risks, the first and second moment will be derived and then a suggested index function will be calculated for comparison.

4.2. Analysis result and comparison

Six estimators are derived under two different loss functions for the comparison of effectiveness. In the Table 2, $\hat{\theta}^B(S)$ means Bayes estimator under squared loss function and $\hat{\theta}^B(B)$ means Bayes estimator under balanced loss function. The same notation is applied for the empirical Bayes estimators. For the constrained Bayes estimators, same results comes from two different loss functions because of the robustness for the loss function as stated in the previous section; consequently, the results are not separated according to loss functions.

Index function in (4.1), we will use $w = 1/2$ to assign the equal weight for each moments, that is the same consideration for the location and dispersion evenly.

Table 2 provides the summary of the data analysis. The result shows that the considerations of both balanced loss function and constrained Bayes estimation are effective to estimate true risk. The consideration of balanced loss function improves the effectiveness regarding the distance of two parameters; in addition, constrained Bayes estimation is closer to the true distance than usual Bayes estimation. The results prove the properness of the balanced loss function in insurance pricing application by simultaneously considering both precision and goodness of fit. Even though a sample of each group is considered, some advantages are proved numerically; in addition, the effects are expected to be expanded in the case of small group estimation. The benefits of the constrained Bayes estimators comes from two aspects, one is robustness of loss function where the same estimator is derived with two different loss functions and the other is by trying to match the first and second moments to more closely estimate location parameter and the dispersion parameter simultaneously. In the table the advantage of the constrained Bayes estimation coming from the second term of the index which calculates the distance of the second moments and approximately 1% by blocking the over-shrinkness usual Bayes and empirical Bayes estimators have. Regarding to loss amount aspect, usual Bayes and empirical Bayes estimators shrink toward the mean from 4000 to 7000 KRW compared to the constrained Bayes estimators and empirical Bayes estimators.

To prevent misleading the facts from the full dependency of data affected by unusual observations, credibility theory is used to estimate parameters. For those reasons, credibility theory was developed by Bühlmann (1967, 1969) and further study was implemented to suggest more reasonable application methodologies by Kim *et al.* (2012), Kim and Kim (2013a). Balanced loss function by considering both the distance from parameter and also data itself could be one of the alternatives to reflect the idea of credibility issues.

The auto insurance policy period is one year in Korea and six month in United States; therefore, accurate estimation as well as consistent estimation techniques are required to suggest proper premiums. That is the reason why both the location and dispersion of each group should be considered and also errors coming from the parameter estimation and data itself happen every policy year. With those reasons, applying the constrained Bayes estimates and balanced loss function consideration with auto insurance pricing could be an alternative way to reach company goals in real business settings. That aspect, this research is meaningful to suggest new accurate and proper tools for the risk estimation, which are essential for insurance companies to compete with each other.

5. Further Study

We compare various Bayes estimators assuming different loss function such as squared error and balanced error loss function. We focus on risk estimation solely in this paper, however, it is meaningful to compare the Bayes risk for each estimator. Asymptotic version of Bayes risk is proved with a specific condition by Ghosh *et al.* (2007, 2008), since exact calculation is impossible to derive. Comparing the Bayes risk of estimators using the derived results could be a further research topic; in addition, more study on the weight in the balanced loss function could represent a new direction for future research.

References

- Bühlmann, H. (1967). Experience rating and credibility I, *JASTIN Bulletin*, **4**, 199–207.
- Bühlmann, H. (1969). Experience rating and credibility II, *ASTIN Bulletin*, **5**, 157–165.
- Ghosh, M. (1992). Constrained Bayes estimation with applications, *Journal of the American Statistical Association*, **87**, 533–540.

- Ghosh, M. and Kim, D. (2002). Multivariate constrained Bayes estimation, *Pakistan Journal of Statistics*, **18**, 143–148.
- Ghosh, M., Kim, M. and Kim, D. (2007). Constrained Bayes and empirical Bayes estimation with balanced loss function, *Communications in Statistics - Theory and Methods*, **36**, 1527–1542.
- Ghosh, M., Kim, M. and Kim, D. (2008). Constrained Bayes and empirical Bayes estimation under random effects normal ANOVA model with balanced loss function, *Journal of Statistical Planning and Inference*, **138**, 2017–2028.
- Kim, Y., Kim, M. H. and Kim, M. J. (2012). Estimating the automobile insurance premium based on credibilities, *The Korean Journal of Applied Statistics*, **14**, 1–14.
- Kim, Y. and Kim, M. (2013a). Suggestions of partial credibilities for non-life insurance premium, *The Korean Journal of Applied Statistics*, **26**, 321–333.
- Kim, M. and Kim, Y. (2013b). Constrained Bayes and empirical Bayes estimator application in insurance pricing, *Communications for Statistical Applications and Methods*, **20**, 321–327.
- Louis, T. A. (1984). Estimating a population of parameter values using Bayes and empirical Bayes method, *Journal of the American Statistical Association*, **79**, 393–398.
- Louis, T. A. (2001). *Bayes/EB Ranking, Histogram and Parameter Estimation: Issues and Research Agenda*, Springer-Verlag, New York.
- Zellner A. (1988). Bayesian analysis in econometrics, *Journal of Econometrics*, **37**, 27–50.
- Zellner, A. (1992). Bayesian and non-Bayesian estimation using balanced loss functions, *Statistical Decision Theory and Related Topics V*, Springer-Verlag, New York, 377–390.

Received February 28, 2014; Revised April 7, 2014; Accepted April 11, 2014