Gazi University Journal of Science
GU J Sci
26(1):51-56 (2013)

Credibility Using Semiparametric Models With Adaptive Kernel

Mehmet MERT[1], Serdar DEMİR[2,♠]

[1]*Department of Econometrics, Faculty of Economics & Administrative Sciences, Akdeniz University, 07058, Antalya, Turkey*

[2]*Department of Statistics, Faculty of Science, Mugla Sitki Kocman University, 48000, Mugla, Turkey*,

Received: 27/09/2011 *Revised 05/03/2012* *Accepted:30/07/2012*

ABSTRACT

The goal of credibility theory is to estimate the future claim of a given risk. The most accurate estimator is predictive mean. If the conditional mean of losses given the risk parameter and the prior distribution of the risk parameter are known, true predictive mean can be easily obtained. However, risk parameter cannot be observed practically and it can be difficult to estimate its distribution. In this study, the structure function is estimated by using kernel density estimation with several bandwidth selection methods. For comparing the efficiences of these methods, a simulation study performed by using the data from a mixture of a lognormal conditional over a lognormal prior. The results shows that the adaptive bandwidth selection method performs better evidently for low claim severities.

Keywords: *Kernel density, Adaptive bandwidth, Loss distribution, Bayesian estimation*.

1. INTRODUCTION

In an insurance system, risks have different distributions, that is, the losses of each risks have different loss distributions. If the exact loss distribution of a risk is known, the appropriate net premium is determined as the expectation of that loss distribution. On the other hand, if there is no information about a specific policyholder, the net premium is the expectation over the all risks. In order to use Bayesian analysis in insurance system, suppose the insurer has prior claim data for the risk, then the conditional expectation of future claims given the prior claims will be net premium. The insurer usually chooses a parametric conditional loss distribution for each risk and a parametric prior distribution to describe how the conditional distributions vary across the risks. A drawback of this method is that the prior distribution can be difficult to choose and the prior data cannot be observed practically, so the resulting model may not represent the loss data very well. In Bayesian approach in insurance, the prior distribution means the distribution of the parameters for each risk. For example, if a risk is a group policyholder, then x_{ij} (i: risk and j: policy period) may be the average of w_{ij} claims. So X_{ij} have probability distributions with parameters θ_i. As a result of this, we get risk parameter θ which has a probability distribution and its density is called structure function, density of risk parameter. So we cannot observe the values of θ_i easily in real life to determine the distribution of risk parameter.

Young [1] used a semiparametric mixture model to represent the insurance losses of a portfolio of risks. She chose a flexible parametric conditional loss distribution for each risk with unknown conditional mean that varies across the risks. She also applied techniques from nonparametric density estimation,

♠Corresponding author, e-mail: serdardemir@mu.edu.tr

Epanechnikov kernel with fixed bandwidth, to estimate the distribution of conditional means. Young simulated from a mixture of a lognormal conditional over a lognormal prior. Young's simulation study showed that this procedure could lead to good credibility formulas even when she used a gamma conditional instead of a lognormal conditional. Later, Young [2] suggested a penalized square-error loss function to smooth the estimated predictive mean from a semiparametric model. Finally, Huang et al. [3] used a piecewise linear prior and showed their credibility estimator performed better than the estimator of Young [2].

In this study, the focus point is to use nonparametric kernel density estimation with adaptive bandwidths against using fixed bandwidth as in Young [1]. Therefore, it is proposed to use kernel estimations with different bandwidths in credibility formula.

2. SEMIPARAMETRIC MIXTURE MODEL

It is assumed that underlying claim of risk i per unit of exposure is a conditional random variable $Y|\theta_i$, i=1,2,...,r, with probability density function $f(y|\theta_i)$. For each of the r risks, one observes the average claims per unit of exposure $\mathbf{x_i}=(\mathbf{x_{i1}}, \mathbf{x_{i2}},...,\mathbf{x_{ini}})$ with an associated exposure vector $\mathbf{w_i}=(\mathbf{w_{i1}}, \mathbf{w_{i2}},..., \mathbf{w_{ini}})$, i=1,2,...,r. Thus, the observed average claim $\mathbf{x_{ij}}$ is the arithmetic average of w_{ij} claims, each of which is an independent realization of the conditional random variable $Y|\theta_i$. Assume that the parameter θ is the conditional mean, $E[Y|\theta]=\theta$. Assume that parameters, other than conditional mean, are fixed across the risks. The loss distribution of a given risk is, therefore, characterized by its conditional mean, although that mean is generally unknown. Denote the probability density function of θ by $\pi(\theta)$, also called the structure function (Young [1], Young [2]). The structure function characterizes how the conditional mean θ varies from risk to risk.

The goal of credibility theory is to estimate the conditional mean $E[Y|\theta]$ of a risk, given that risk's claim experience is x and exposure is w. As in Young [1], set the credibility formula equal to the predictive mean $E[Y|\bar{x}]$ given the weighted sample average $\bar{x}$ weighted by the exposure w. Also restrict attention to parametric conditional distributions for which $E[Y|\theta]=\theta$ (Thus, the predictive means equals the posterior mean; that is, $E[Y|\bar{x}]=E[\theta|\bar{x}]$), the sample mean is a sufficient statistic for θ, and the functional form of $f(y|\theta)$ is closed under averaging. That is, if $\bar{X}$ is an average of w claims that follow the distribution given by $f(y|\theta)$, then the density of $\bar{X}$ has the same functional forms as $f(y|\theta)$. Many of families of densities satisfy these properties (Young, [1]).

In the Bayesian spirit, for a given loss function L=L(y, $d(\bar{x})$) of the future claim y and the claim predictor d, it is proposed that the credibility estimator d is the function that minimizes the expected loss $E[L(y, d(\bar{x}))]$. This expectation is with respect to the joint density of the sample mean and future claim. In the mixture model, joint density is $\int f(y|\theta)f(\bar{x}|\theta)\pi(\theta)d\theta$.

3. ADAPTIVE KERNEL DENSITY ESTIMATION

Kernel density estimation is a nonparametric method for estimating densities. Young [1] proposed kernel estimation method for $\pi(\theta)$. Kernel estimation of structure function $\pi(\theta)$ was proposed by:

$$\hat{\pi}(\theta)=\sum_{i=1}^{r}\frac{w_i}{w_{tot}h_i}K\left(\frac{\theta-\bar{x}_i}{h_i}\right) \tag{1}$$

Here, $w_{tot}=\sum_{i=1}^{r}w_i=\sum_{i=1}^{r}\sum_{j=1}^{n_i}w_{ij}$. The kernel density estimation is based on kernel function K(u) and bandwidth h_i. Kernel functions are symmetric around 0 and to integrate 1. Gaussian and Epanechnikov are the popular kernel functions. Young [1] used Epanechnikov kernel function because its domain is bounded. The Epanechnikov kernel function can be given as follows:

$$K(t)=\begin{cases}\dfrac{3(1-t^2/5)}{4\sqrt{5}}, & -\sqrt{5}<t<\sqrt{5}\\ 0, & else\end{cases} \tag{2}$$

Epanechnikov kernel function is also used in this study. The bandwidth selection is crucial importance in kernel estimation. There are various methods for selecting the bandwidth; see for example, Hardle [4] and Silverman [5]. Young [1] used a fixed bandwidth selected by reference to a standard distribution by modifying it. Young set the bandwidth h_i equal to $\bar{x}_i/\sqrt{5}$ to guarantee that the support of estimated density of be contained in the nonnegative real numbers.

The "adaptive kernel density estimation" is used for $\pi(\theta)$. Adaptive kernel estimation for $\pi(\theta)$ can be written by

$$\tilde{\pi}(\theta)=\frac{1}{r}\sum_{i=1}^{r}\frac{1}{h(\bar{x}_i)}K\left(\frac{\theta-\bar{x}_i}{h(\bar{x}_i)}\right) \tag{3}$$

The adaptive kernel estimation uses varying bandwidths $h(\bar{x}_i)$. It was first considered by Breiman et al. [6] and Abramson [7]. Usefullness of varying (or local) bandwidths is widely acknowledged to estimate long-tailed or multi-modal density functions with kernel methods. The basic idea of the adaptive kernel estimation is to use a broader bandwidth in the regions of low density to smooth the tail part of the function (Silverman [5]). For details, see Terrell and Scott [8] and Sain [9].

Silverman [5] used a general strategy to obtain the adaptive estimation in (3). The varying bandwidth for each $\bar{x}_i$ is defined as $h(\bar{x}_i)=h\lambda_i$. And then the local bandwidth factors λ_i is defined as

$$\lambda_i = \left[\frac{\tilde{\pi}(\bar{x}_i)}{g}\right]^{-\psi} \qquad (4)$$

where g is the geometric mean of the pilot estimation $\tilde{\pi}(\bar{x}_i)$ over all $\bar{x}_i$ and ψ is the sensitivity parameter, a number satisfying $0 \le \psi \le 1$. Abramson [7] states that $\psi=0.5$ gives good results. The pilot estimation $\tilde{\pi}(\bar{x}_i)$ can be obtain by using any density estimation method with initial fixed bandwidth h. After obtaining each λ_i, varying bandwidths are computed as $h(\bar{x}_i)=h\lambda_i$. Then the adaptive kernel estimation is obtained by replacing $h(\bar{x}_i)$ in (3).

The kernel density estimation method is used for estimating the pilot estimation $\tilde{\pi}(\bar{x}_i)$ in (4). For selecting the initial fixed bandwidth h, the "reference to a standard distribution" is used as bandwidth selection method as in Young [1]. Alternatively, "the least square cross-validation (LSCV)" bandwidth selection method is also used. The LSCV method is an automatic method based on data. In LSCV method, the optimal bandwidth is the bandwidth which minimizes the following cross-validation function CV(h)

$$CV(h) = \int \hat{\pi}^2(\theta)d\theta - \frac{2}{r}\sum_{i=1}^{r}\hat{\pi}_i(\bar{x}_i) \qquad (5)$$

where $\hat{\pi}_i(\bar{x}_i)$ is a leave-one-out kernel estimation. The leave-one-out kernel estimation $\hat{\pi}_i(\bar{x}_i)$ is constructed from all the data points except $\bar{x}_i$ (Hardle [4], Silverman [5]).

4. CREDIBILITY USING SQUARED-ERROR LOSS

The squared-error loss function is used to determine a credibility estimator as given in Young [1]. The squared-error loss function has the form L(y, d($\bar{x}$))^2. It is straightforward to show that the minimizer of the expected loss is the predictive mean, which in this case is the posterior mean of θ given the sample mean $\bar{x}$ which is estimates by $\hat{\mu}(\bar{x}) = \int E(Y|\theta)\hat{\pi}(\theta|\bar{x})d\theta = \hat{E}(\theta|\bar{x})$.

In this study, the fixed bandwidth reference to a standard distribution is defined as ${}^1h$ and the fixed bandwidth using LSCV bandwidth selection method is defined as ${}^2h$. The adaptive bandwidth obtained using ${}^2h$ as the initial fixed bandwidth is called as ${}^3h$ and the adaptive bandwidth obtained using ${}^1h$ as the initial fixed bandwidth is also called as ${}^4h$. As a result, four different estimations of posterior mean of θ are obtained for a general kernel K and bandwidths ${}^k h_i$ (k=1, 2, 3, 4). The posterior mean can be written as

$$\hat{\mu}_k(\bar{x}) = \frac{\sum_{i=1}^{r}\frac{w_i}{{}^k h_i}\int \theta f(\bar{x}|\theta)K\left(\frac{\theta-\bar{x}_i}{{}^k h_i}\right)d\theta}{\sum_{i=1}^{r}\frac{w_i}{{}^k h_i}\int f(\bar{x}|\theta)K\left(\frac{\theta-\bar{x}_i}{{}^k h_i}\right)d\theta}, \quad k=1,2,3,4 \qquad (6)$$

Recall that $\bar{x}$ is an average of w iid claims, each of which follows the density $f(y|\theta)$. If the estimator d is linear, it is well known that the least-squares linear estimator of $E(Y|\bar{x}) = E(\theta|\bar{x})$ is d($\bar{x}$)=(1-Z)E(Y)+Z $\bar{x}$, in which Z=w/(w+v) with v= $\hat{\mathrm{E}}(\mathrm{Var}(\mathrm{Y}|\theta))/\hat{\mathrm{Var}}(\theta)$ and E(Y)= $\bar{\bar{\mathrm{X}}}$ average of sample means $\bar{x}_i$, i=1, 2, ...,r (Bühlmann [10]).

It is shown that as w approaches infinity, $\hat{\mu}(\bar{x})$ approaches the true expected value θ_0, for the given risk. Thus, if an actuary gets more claim information for a given policyholder (that's w becomes large), the estimated expected claim approaches the true expected claim with probability 1 (Young [1]).

5. SIMULATION

In this section, it is assumed that having individual claim data, that is, w_{ij}=1, for all risks i and policy periods j. Therefore, X=Y. The lognormal-lognormal mixture is modelled as follows:

$$f(x|\theta) = \frac{1}{\sigma x\sqrt{2\pi}}\exp\left\{-\frac{1}{2\sigma^2}\left(\ln\left(\frac{x}{\theta}\right)\right)^2\right\}, \qquad x>0,$$

in which $\sigma > 0$ is known parameter, and

$$\pi(\theta) = \frac{1}{\tau\theta\sqrt{2\pi}}\exp\left\{-\frac{1}{2\tau^2}\left(\ln\left(\frac{\theta}{\mu}\right)\right)^2\right\}, \qquad \theta>0$$

in which $\mu>0$ and $\tau>0$ are known parameters. That is, $(\ln X)|\theta \sim$N(ln θ,σ^2), and $\ln\theta \sim$N(ln μ ,

τ^2). The marginal distribution of X is lognormal; $\ln X \sim N(\ln\mu, \sigma^2+\tau^2)$.

Given claim data for a specific policyholder, $X=x=<x_1, x_2, \ldots, x_n> \in [0,\infty]^n$, the posterior distribution of $\theta|x$ is lognormal; $(\ln\theta)|x \sim N(\ln\mu^*, \tau^{*2})$ in which

$$\mu^* = \exp\left(\frac{\sigma^2\ln\mu+\tau^2\upsilon}{\sigma^2+n\tau^2}\right),$$

$$\upsilon=\sum_{i=1}^{n}\ln(x_i) \text{ and } \tau^{*2}=\frac{\sigma^2\tau^2}{\sigma^2+n\tau^2}.$$

Thus, the predictive distribution of $X_{n+1}|x$ is lognormal; $(\ln X_{n+1})|x \sim N(\ln\mu^*, \sigma^2+\tau^{*2})$. It follows that the true predictive mean is a function of the statistic υ

$$\mu(x)=E(X_{n+1}|x)=\exp\left(\frac{\sigma^2\ln\mu+\tau^2\upsilon}{\sigma^2+n\tau^2}+\frac{\sigma^2(\sigma^2+(n+1)\tau^2)}{2(\sigma^2+n\tau^2)}\right)$$

Two hundred simulations of a lognormal-lognormal mixture of claims are performed. For the simulations, it is assumed that $\sigma^2=0.25$, $\tau^2=0.50$ and $\mu=2000e^{-0.25}$. For each simulation run, claim data from this lognormal-lognormal mixture are simulated for r=100 risks (values of θ). For each of the 100 risks, $n_i=w_i=5$ claims are simulated. To estimate the distribution of the conditional means, the kernel density estimation with Epanechnikov kernel is used as given by (2). The fixed bandwidths which are the bandwidth reference to a standart normal distribution and LSCV bandwidth are also used. Then the adaptive bandwidths by using these fixed bandwidths are obtained as described in Section 2. The bandwidths ${}^k h$ (k=1,2,3,4) are modified, for a given risk if, by otherwise using it, the prior density would have a negative support. Specifically, if ${}^k h > \bar{x}_i/\sqrt{5}$ then the bandwidths ${}^k h_i$ are taken as being equal to $\bar{x}_i/\sqrt{5}$ to guarantee that the support of the estimated density of θ be contained in the nonnegative real numbers.

As Young (1997), instead of assuming that the conditional is lognormal, it is assumed that the coefficient of variation is constant from risk to risk and, therefore, fit a gamma conditional with mean $\theta=\alpha/\beta$ and fixed shape parameter α to each risk. The parameter α is estimated by the median of the sample statistic $\bar{x}_i^2 \Big/ \left[\frac{1}{5-1}\sum_{j=1}^{5}(x_{ij}-\bar{x}_i)^2\right]$.

The estimated mixture model is used to estimate the predictive mean of X_{n+1} given claim data x and the estimators $\hat{\mu}_k(x)$ are obtained. Also, for each risk, the linear Bühlmann credibility estimator *lin*(x), defined as below, is computed.

$$lin(x) = Z\bar{X}_i + (1+Z)\bar{\bar{X}} \quad \text{(Bühlmann, [10])}.$$

Here, Z is credibility factor and $\bar{\bar{X}}$ is overall mean. We get these values from the equations below (Bühlmann, [10]).

$$\bar{\bar{X}} = \frac{1}{r}\sum_{i=1}^{r}\bar{X}_i$$

$$Z = \frac{1}{1+\dfrac{E\hat{P}V}{n\,V\hat{H}M}}$$

$$E\hat{P}V = \frac{1}{r(n-1)}\sum_{i=1}^{r}\sum_{j=1}^{n}(x_{ij}-\bar{x}_i)^2$$

$$V\hat{H}M = \frac{1}{r-1}\sum_{i=1}^{r}(\bar{x}_i-\bar{\bar{x}})^2 - \frac{E\hat{P}V}{n}$$

Therefore, the expected process variance is estimated by

$$E\hat{P}V = \frac{1}{100(5-1)}\sum_{i=1}^{100}\sum_{j=1}^{5}(x_{ij}-\bar{x}_i)^2$$

and the variance of the hypothetical means by

$$V\hat{H}M = \frac{1}{100-1}\sum_{i=1}^{100}(\bar{x}_i-\bar{\bar{x}})^2 - \frac{E\hat{P}V}{5} \quad \text{(Young, [1])}.$$

For the simulation study, the computer programs were written in Delphi and the estimated predictive means $\hat{\mu}_k(x)$ and the linear Bühlmann credibility estimator *lin*(x) are compared to the true predictive mean $\mu(x)$. To compare these credibility estimators numerically for small claims which can be called as low claim severities, for each of the 200 simulation runs, the mean squared errors are calculated up to the 10th percentile of X. See Table 1 for the descriptive statistics of the mean squared errors of the credibility estimators for small claims.

Table 1. Descriptive Statistics of Mean Squared Errors for Low Claim Severities

	Mean	Median	St.Dev.	Q1	Q3
MSE_B	165565.783	145620.896	101139.856	93940.147	223475.060
MSE_1	15251.914	12825.675	10089.356	7602.218	22776.495
MSE_2	15860.229	13978.373	9919.903	8125.618	23392.517
MSE_3	12755.359	10666.630	9351.071	5325.465	19136.605
MSE_4	11168.086	7786.156	9314.397	3892.948	17783.491

MSE_B: mean squared error of lin(x), MSE_k: mean squared error of $\hat{\mu}_k(x)$, k=1,2,3,4

Table 1 shows that estimated predictive means perform much better than the linear Bühlmann credibility estimator for small claims. In addition, the estimated predictive means obtained by using adaptive bandwidths, especially $\hat{\mu}_4(x)$, perform better than the other credibility estimators. This result can be seen graphically in Figure 1. An example graphic for one of the simulation samples is showed in Figure 1, compared $\hat{\mu}_1(x)$, $\hat{\mu}_4(x)$ and $\mu(x)$. In this simulation run, [1]h=549, MSE_1=8278,15, MSE_4=2379,47 and 10[th] percentile of X is 700.

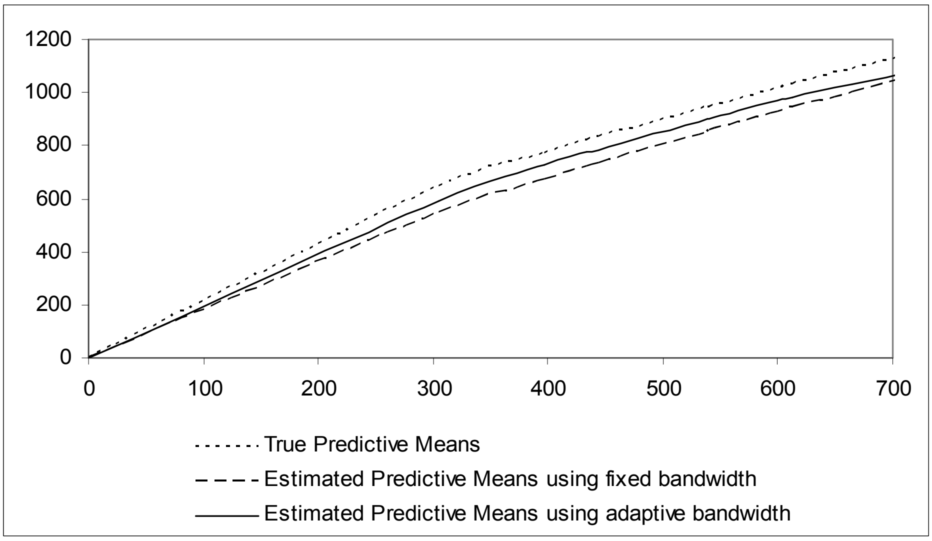

Figure 1. Credibility estimators for low claim severities

The estimated predictive means $\hat{\mu}_k(x)$ and the linear Bühlmann credibility estimator *lin*(x) are also compared to the true predictive mean $\mu(x)$ for medium values of X which can be called as medium claim severities. To compare these credibility estimators numerically for medium claims, for each of the 200 simulation runs, the mean squared errors are calculated between the 10[th] percentile of X and the 95[th] percentile of X. See Table 2 for the descriptive statistics of the mean squared errors of the credibility estimators for medium claims.

Table 2. Descriptive Statistics of Mean Squared Errors for Medium Claim Severities

	Mean	**Median**	**St.Dev.**	**Q1**	**Q3**
$\mathbf{MSE_B}$	49680.569	45301.274	28875.875	25769.089	66553.026
$\mathbf{MSE_1}$	23357.223	13479.231	42536.487	5280.421	27401.022
$\mathbf{MSE_2}$	25939.005	15646.002	44189.330	6123.811	31441.819
$\mathbf{MSE_3}$	27304.428	16770.205	44375.150	7283.329	32252.554
$\mathbf{MSE_4}$	27176.974	15707.343	44547.945	6699.398	30506.988

MSE_B: mean squared error of lin(x), MSE_k: mean squared error of $\hat{\mu}_k(x)$, k=1,2,3,4

Table 2 shows that the estimated predictive means, on average, perform much better than the linear Bühlmann credibility estimator for medium claim severities. In addition, the estimated predictive means obtained by using fixed bandwidths, especially $\hat{\mu}_1(x)$, perform better than the other credibility estimators. However, the mean squared errors of the linear Bühlmann estimator have smaller standard deviation than the other estimators. Despite having smaller average of mean squared errors, the estimated predictive means show more variations than the linear Bühlmann credibility estimator.

Lastly, the estimated predictive means $\hat{\mu}_k(x)$ and the linear Bühlmann credibility estimator *lin*(x) are compared to the true predictive mean $\mu(x)$ for large values of X which can be called as high claim severities. To compare these credibility estimators numerically for large claims, for each of the 200 simulation runs, the mean squared errors are calculated for the values of X which is higher than 95th percentile of X. See Table 3 for the descriptive statistics of the mean squared errors of the credibility estimators for large claims.

Table 3: Descriptive Statistics of Mean Squared Errors for High Claim Severities

	Mean	Median	St.Dev.	Q1	Q3
MSE_B	849970.760	219901.934	3125017.155	90891.435	543060.670
MSE_1	3340718.926	519094.110	10293101.362	93636.087	2020368.320
MSE_2	3296257.094	509038.702	10251891.506	91894.688	2013041.923
MSE_3	3390720.507	854543.107	9603905.624	186230.414	2628670.382
MSE_4	3321895.984	776052.502	9957207.598	178585.155	2532854.997

MSE_B: mean squared error of lin(x), MSE_k: mean squared error of $\hat{\mu}_k(x)$, k=1,2,3,4

Table 3 shows that the linear Bühlmann credibility estimator performs much better than the estimated predictive means for high claim severities. In addition, the mean squared errors of the estimated predictive means get higher for large values of X. From the simulation results, it can be seen that the estimated predictive means diverge upward for claims larger than the 95th percentile of X. So, the mean squared errors of the estimated predictive means become higher values for large claims. Young [1] used a constancy penalty which prevents the estimated predictive mean from diverging too greatly from the true predictive mean.

6. CONCLUSION

In this paper, the adaptive bandwidth selection method applied to semiparametric credibility and compared credibility estimators in different classes of claims. Simulation results show that the estimators obtained by using adaptive bandwidths give better results evidently in low claim severities. For medium claim severities, the estimators obtained by using fixed bandwidths performs well. For high claim severities, Bühlmann credibility estimator have the best performance.

REFERENCES

[1] Young, V.R., Credibility using semiparametric models, ***ASTIN Bulletin***, 27: 273-285 (1997).

[2] Young, V.R., Credibility using semiparametric models and a loss function with a constancy penalty, ***Insurance: Mathematics & Economics,*** 26: 151-156 (2000).

[3] Huang, X., Song, L., Liang, Y., Semiparametric credibility ratemaking using a piecewise linear prior, ***Insurance: Mathematics & Economics***, 33: 585-593 (2003).

[4] Hardle, W.,. Smoothing techniques with implementations in S., ***Springer-Verlag***, New York (1990).

[5] Silverman, B.W., Density Estimation for Statistics and Data Analysis, ***Chapman and Hall,*** New York (1986).

[6] Breiman, L., Meisel, W., Purcell, E., Variable kernel estimates of multivariate densities, ***Technometrics,*** 19: 135-144 (1977).

[7] Abramson, I., On bandwidth variation in kernel estimates-a square root law, ***Annals of Statistics,*** 10: 1217-1223 (1982)

[8] Terrell, G.R. and Scott, D.W.,. Variable kernel density estimation, ***Annals of Statistics***, 20:1236-1265 (1992).

[9] Sain, S.R.,. Adaptive kernel density estimation. Ph.D. Thesis, Department of Statistics, ***Rice University***, Houston, Texas (1994).

[10] Bühlmann, H.,. Experience rating and credibility, ***ASTIN Bulletin***, 4: 199-207 (1967).