Western Graduate&PostdoctoralStudies

Western University

Scholarship@Western

Electronic Thesis and Dissertation Repository

8-1-2019 2:00 PM

Split credibility: A two-dimensional semi-linear credibility model

Jingbing Qiu, *The University of Western Ontario*

Supervisor: Jiandong Ren, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Statistics and Actuarial Sciences

© Jingbing Qiu 2019

Follow this and additional works at: https://ir.lib.uwo.ca/etd

Part of the Statistical Models Commons

Recommended Citation

Qiu, Jingbing, "Split credibility: A two-dimensional semi-linear credibility model" (2019). *Electronic Thesis and Dissertation Repository*. 6309.
https://ir.lib.uwo.ca/etd/6309

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlswadmin@uwo.ca.

Abstract

In the thesis, we introduce a two-dimensional semi-linear credibility model, which is an extension of the classical credibility or split credibility models used by practicing actuaries. Our model predicts the future expected losses of a policyholder by considering its historical primary and excess losses. The optimal split point is derived based on the mean squared error criterion. We show when and why splitting a policyholder's historical losses into primary and excess parts work analytically. In addition, we derived formulas for estimating our model parameters nonparametrically. Finally, we show the application of our model through three examples.

Keywords: Two-dimensional semi-linear credibility model, split credibility, primary and excess credibility, linear function, mean square error

Summary for Lay Audience

Credibility theory is a set of quantitative tools that allows an insurer to adjust premiums based on policy holders' past loss experience. The theory features the combination of data with other information, such as the mean loss of policyholders in the same rating class.

In this thesis, we introduce a two-dimensional semi-linear credibility model, which considers policyholders' small losses and large losses separately. Our model is an extension of the classical credibility or split credibility models used by practicing actuaries.

Acknowlegements

The author gratefully acknowleges the supervisor, Jiandong Ren for his insights and observations. He gave the author detailed and thoughtful comments and suggestions. The author also would like to acknowlege friends and loved ones for their support and encouragement. These contributions helped to motivate and substantially improve the thesis.

Contents

List of Figures

List of Tables

List of Appendices

Chapter 1

Introduction

Credibility theory is a quantitative tool that enables us to estimate and adjusts the future premium given a policyholder's loss experience. For a detailed introduction of credibility theory, readers are referred to Klugman et al. [8]. Assume that the risks are homogeneous, the manual rate is designed to reflect the past and future expected experience of the entire rating class. Because the policyholders in a rating class are different, the manual rate can not reflect individuals' actual risk. However, it is important that higher risks should have higher rate and lower risks should have lower rate.

Therefore, the insurer is forced to determine how much of the difference between the policyholder's own experience and the expected experience is due to random fluctuations as well as how much is due to the fact that the policyholder's own risk is higher or lower than the average risk. That is, how much credibility does the policyholder's own experience have? The credibility necessarily depends on the amount of data. The more past information we have on a given policyholder, the more credible the policyholder's own experience, all other things being equal. In group insurance, the loss experience of larger groups are more credible than that of small groups.

Another use for credibility is in the setting of rates for classification systems. For example, there may be many occupational classes in workers compensation insurance, some of which may provide very little data. In order to accurately estimate the expected cost for insuring these classes, limited empirical data can be combined with other information, such as past rates and so on.

From a statistical point of view, if loss experience data from an insured or group of insureds is avaiable, we should use the sample mean or some other unbiased estimator to determine the premium. But the results in credibility theory show that it is optimal to give only partial weight to this experience and give the remaining weight to an estimator produced from other information. For details, readers are referred to Klugman et al. [8].

Credibility theory is an approach to combine the manual rate with the policyholder's own loss experience, so that future premium will reflect the future losses accurately. In this chapter, we firstly introduce *limited fluctuation crediblity theory*, a subject developed in the early part of the twentieth century. The theory provides an approach to assign full or partial credibility to a policyholder's experience. Then, we will introduce *greatest accuracy credbility theory*, which was formalized by Bühlmann [1]. The simplest model of Bühlmann [1] is introduced in this section and will be our assumption in the thesis. Besides that, we will introduce an improved model that developed by Bühlmann and Straub [3], after the simplest Bühlmann model.

Then, we will introduce split credibility which is the research object of the thesis. We firstly introduce semi-linear credibility with truncation (referred to Bühlmann et al. [4]). Then we propose our two-dimensional semi-linear credibility model.

The thesis is organised as follows. In chapter 2, we discuss our two-dimensional semi-linear credibility model and also discuss the nonparametric estimation method associated with our model. In chapter 3, we show the application of the model with three examples. Chapter 4 concludes the thesis.

1.1 Limited fluctuation crediblity theory

Limited fluctuation crediblity theory is an approach to determine whether we should assign full credibility on the policyholder's own past experience or not and decide to assign partial credibility if full credibility is inappropriate. Suppose that a policyholder has experienced X_j claims or losses in past experience period j, where $j \in \{1, 2, 3, \cdots, n\}$. Suppose that $E(X_j) = \xi$ and $Var(X_j) = \sigma^2$ for all j. The past experience may be summarized by the average $\bar{X} = \frac{1}{n}\sum_{j=1}^{n} X_j$. Notice that $E(\bar{X}) = \xi$ and if the X_j are independent, $Var(\bar{X}) = \frac{\sigma^2}{n}$.

The insurer's goal is to determine the value of ξ. One way is to ignore the past experience and simply charge M, a value obtained from experience on other similar but not identical policyholders, which is often called the manual premium. Another way is to ignore M and charge $\bar{X}$, which is full credibility. A third way is to choose some combination of M and $\bar{X}$, which is partial credibility.

For full credibility, we use statistical method to decide whether the losses are stable or not. That is, selecting two numbers $r > 0$ and $0 < p < 1$ (with r close to 0 and p close to 1) and assigning full credibility if

$$Pr(-r\xi \leq \bar{X} - \xi \leq r\xi) \geq p. \tag{1.1}$$

Restate (1.1) as

$$Pr\left(\left|\frac{\bar{X} - \xi}{\sigma/\sqrt{n}}\right| \leq \frac{r\xi\sqrt{n}}{\sigma}\right) \geq p. \tag{1.2}$$

Let y_p be defined by

$$y_p = \inf_y \left\{ Pr\left(\left|\frac{\bar{X} - \xi}{\sigma/\sqrt{n}}\right| \leq y\right) \geq p \right\}. \tag{1.3}$$

Then the condition for full credibility is

$$\frac{r\xi\sqrt{n}}{\sigma} \geq y_p. \tag{1.4}$$

Rewrite (1.4) as

$$\frac{\sigma}{\xi} \leq \frac{r}{y_p}\sqrt{n} = \sqrt{\frac{n}{\lambda_0}}, \tag{1.5}$$

where $\lambda_0 = (y_p/r)^2$. If the X_j are independent, we can rewrite (1.5) as

$$Var(\bar{X}) = \frac{\sigma^2}{n} \le \frac{\xi^2}{\lambda_0}. \tag{1.6}$$

Alternatively, solving (1.5) for n gives the number of exposure units required for full credibility, namely,

$$n \ge \lambda_0 \left(\frac{\sigma}{\xi}\right)^2. \tag{1.7}$$

For details, readers are referred to Klugman et al. [8]. Thus, we can get the condition of sample size which meets the standard for full credibility.

If full credibility above is inappropriate, we consider partial credibility which contains the past experience $\bar{X}$ in the net premium as well as the externally obtained mean, M. Then we get the formula of credibility premium,

$$P_c = Z\bar{X} + (1 - Z)M, \tag{1.8}$$

where the credibility factor $Z \in [0, 1]$ needs to be chosen. The theoretical method on the basis of a statistical model to determine the optimal Z will be presented in next section. Another method is based on the same idea as full credibility. We see from (1.6) that there is no assurance that the variance of X will be small enough. However, it is possible to control the variance of the credibility premium, P_c, as follows:

$$\begin{aligned}
\frac{\xi^2}{\lambda_0} &= Var(P_c) \\
&= Var\left[Z\bar{X} + (1 - Z)M\right] \\
&= Z^2 Var(\bar{X}) \\
&= Z^2 \frac{\sigma^2}{n}.
\end{aligned}$$

Thus, $Z = (\xi/\sigma)\sqrt{n/\lambda_0}$. Due to $Z \in [0, 1]$, we have that

$$Z = min\left\{\frac{\xi}{\sigma}\sqrt{\frac{n}{\lambda_0}}, 1\right\}. \tag{1.9}$$

1.2 Greatest accuracy credbility theory

Bühlmann [1] introduced a model-based approach to solve the credibility problem. Suppose that we have the past claims for a particular policyholder with n exposure units, $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)^T$, and a manual rate μ (it is the same as M above) applicable to this policyholder. However, the mean of claims is quite different from the manual rate. This difference raises the question of whether next year's net premium (per exposure unit) should be based on μ, on $\bar{X}$, or on a combination of the two. That is, how credible is the manual rate and how credible is the past experience for this policyholder? For this question, the Bayesian methodology is discussed in Klugman et al. [8]. But here we do not discuss the Bayesian methodology in detail, and focus on the Bühlmann model (the simplest credibility model).

Under the Bühlmann model, for each policyholder, past losses $X_1, \ldots, X_n$ are assumed to have the same mean and variance and are i.i.d. conditional on Θ. Assume that Θ is a random variable which is the risk parameter associated with the policyholder. Define

$$\begin{aligned}
\mu(\theta) &= E(X_j|\Theta = \theta), \\
v(\theta) &= Var(X_j|\Theta = \theta), \\
\mu &= E(\mu(\Theta)), \\
v &= E(v(\Theta)), \\
a &= Var[\mu(\Theta)].
\end{aligned}$$

We are interested in setting a rate to cover X_{n+1}. Define

$$\begin{aligned}
\mu_{n+1}(\theta) &= E(X_{n+1}|\Theta = \theta), \\
\mu_{n+1} &= E(\mu_{n+1}(\Theta)).
\end{aligned}$$

Now, we can calculate the unconditional mean and variance of X_j as well as $\bar{X}$ as follows,

$$E(X_j) = E[E(X_j|\Theta)] = E[\mu(\Theta)] = \mu = E(\bar{X}),$$

and

$$\begin{aligned}
Var(X_j) &= E[Var(X_j|\Theta)] + Var[E(X_j|\Theta)] \\
&= E[v(\Theta)] + Var[\mu(\Theta)] \\
&= v + a,
\end{aligned}$$

and

$$\begin{aligned}
Var(\bar{X}) &= E[Var(\bar{X}|\Theta)] + Var[E(\bar{X}|\Theta)] \\
&= E\left[\frac{v(\Theta)}{n}\right] + Var[\mu(\Theta)] \\
&= \frac{v}{n} + a.
\end{aligned}$$

Because the losses are i.i.d. conditional on Θ, the unconditional mean and variance of X_{n+1} are the same as X_j's. That is,

$$E(X_{n+1}) = E[E(X_{n+1}|\Theta)] = E[\mu_{n+1}(\Theta)] = E[\mu(\Theta)] = \mu = \mu_{n+1},$$

and

$$\begin{aligned}
Var(X_{n+1}) &= E[Var(X_{n+1}|\Theta)] + Var[E(X_{n+1}|\Theta)] \\
&= E[v(\Theta)] + Var[\mu(\Theta)] \\
&= v + a.
\end{aligned}$$

Bühlmann [1] suggested a linear function of past loss data to estimate a policyholder's expected loss next year $\mu_{n+1}(\theta)$. Thus, the credibility premium is

$$P_c = \alpha_0 + \sum_{j=1}^{n} \alpha_j X_j, \tag{1.10}$$

where $\alpha_0, \alpha_1, \cdots, \alpha_n$ are parameters that are needed to be chosen. Hence, we choose the optimal α's to minimize mean squared error(MSE), which is

$$\begin{aligned} Q &= E\left\{\left[\mu_{n+1}(\Theta) - P_c\right]^2\right\} && (1.11) \\ &= E\left\{\left[\mu_{n+1}(\Theta) - \alpha_0 - \sum_{j=1}^{n} \alpha_j X_j\right]^2\right\}. && (1.12) \end{aligned}$$

Thus, by taking derivatives with regard to α's in (1.12), it can be shown that the credibility premium is given by

$$\begin{aligned} P_c = \widehat{\mu_X(\theta)} &= \hat{\alpha}_0 + \sum_{j=1}^{n} \hat{\alpha}_j X_j && (1.13) \\ &= \hat{Z}\bar{X} + (1 - \hat{Z})\mu, && (1.14) \end{aligned}$$

where

$$\hat{Z} = \frac{n}{n+k} \tag{1.15}$$

and

$$k = \frac{v}{a} = \frac{E[Var(X_j|\Theta)]}{Var[E(X_j|\Theta)]} \tag{1.16}$$

and $\hat{\alpha}_0$ is the optimal α_0 and $\hat{\alpha}_j$ is the optimal α_j for all j and $\hat{Z}$ is the optimal credibility factor, Z. For details, readers are referred to Klugman et al. [8].

The minimum mean squared error(MMSE) in the Bühlmann model is

$$\begin{aligned}
\hat{Q} &= E\left\{\left[\mu_{n+1}(\Theta) - \hat{\alpha}_0 - \sum_{j=1}^{n}\hat{\alpha}_j X_j\right]^2\right\} \\
&= E\left\{\left[\mu_{n+1}(\Theta) - \hat{Z}\bar{X} - (1-\hat{Z})\mu\right]^2\right\} \\
&= E\left\{\mu_{n+1}(\Theta)^2 + \left[\hat{Z}\bar{X} + (1-\hat{Z})\mu\right]^2 - 2X_{n+1}\left[\hat{Z}\bar{X} + (1-\hat{Z})\mu\right]\right\} \\
&= E\left\{\mu_{n+1}(\Theta)^2 + \hat{Z}^2\bar{X}^2 + (1-\hat{Z})^2\mu^2 + 2\hat{Z}(1-\hat{X})\mu\bar{X} - 2\hat{Z}X_{n+1}\bar{X} - 2(1-\hat{Z})\mu X_{n+1}\right\} \\
&= E(\mu_{n+1}(\Theta)^2) + \hat{Z}^2E(\bar{X}^2) + (1-\hat{Z})^2\mu^2 + 2\hat{Z}(1-\hat{X})\mu^2 - 2\hat{Z}E(X_{n+1}\bar{X}) - 2(1-\hat{Z})\mu^2 \\
&= Var(\mu_{n+1}(\Theta)) + E(\mu_{n+1}(\Theta))^2 + \hat{Z}^2\left(Var(\bar{X}) + E(\bar{X})^2\right) - 2\hat{Z}E\left[E(X_{n+1}\bar{X}|\Theta)\right] - (1-\hat{Z})^2\mu^2 \\
&= a + \mu^2 + \hat{Z}^2\left(\frac{v}{n} + a + \mu^2\right) - 2\hat{Z}E\left[E(X_{n+1}\Theta)E(\bar{X}|\Theta)\right] - (1-\hat{Z})^2\mu^2 \\
&= a + \mu^2 + \hat{Z}^2\left(\frac{v}{n} + a + \mu^2\right) - 2\hat{Z}E[\mu(\Theta)^2] - (1-\hat{Z})^2\mu^2 \\
&= a + \mu^2 + \hat{Z}^2\left(\frac{v}{n} + a + \mu^2\right) - 2\hat{Z}\left(Var(\mu(\Theta)) + E(\mu(\Theta))^2\right) - (1-\hat{Z})^2\mu^2 \\
&= a + \mu^2 + \hat{Z}^2\left(\frac{v}{n} + a + \mu^2\right) - 2\hat{Z}(a+\mu^2) - (1-\hat{Z})^2\mu^2 \\
&= \hat{Z}^2\frac{v}{n} + (a+\mu^2)(1-\hat{Z})^2 - (1-\hat{Z})^2\mu^2 \\
&= \hat{Z}^2\frac{v}{n} + (1-\hat{Z})^2 a \\
&= \left(\frac{na}{na+v}\right)^2\frac{v}{n} + \left(\frac{v}{na+v}\right)^2 a \\
&= \frac{va(na+v)}{(v+na)^2} \\
&= \frac{va}{v+na}. \qquad (1.17)
\end{aligned}$$

However, the Bühlmann model does not allow for variations in exposure or size. Therefore, the Bühlmann-Straub model is presented in Bühlmann and Straub [3] to correct the problem. The difference between the Bühlmann-Straub model and the Bühlmann model is the conditional variances. The conditional variances is assumed to be

$$Var(X_j|\Theta = \theta) = \frac{v(\theta)}{m_j}, \qquad (1.18)$$

where m_j is a known constant measuring exposure. This model would be appropriate if each X_j were the average of m_j independent (conditional on Θ) random variables each with mean

$\mu(\theta)$ and variance $v(\theta)$. Therefore, the unconditional variance of X_j becomes

$$\begin{aligned} Var(X_j) &= E[Var(X_j|\Theta)] + Var[E(X_j|\Theta)] \\ &= E\left[\frac{v(\Theta)}{m_j}\right] + Var[\mu(\Theta)] \\ &= \frac{v}{m_j} + a. \end{aligned}$$

To obtain the new credibility premium (1.10), we should take derivatives with regard to α's in (1.12) again. Define

$$m = m_1 + m_2 + \cdots + m_n$$

to be the total exposure. Then, the credibility premium (1.10) becomes

$$P_c = \widehat{\mu_X(\theta)} = \hat{\alpha}_0 + \sum_{j=1}^{n} \hat{\alpha}_j X_j = \hat{Z}\bar{X} + (1 - \hat{Z})\mu, \tag{1.19}$$

where

$$\hat{Z} = \frac{m}{m+k} \tag{1.20}$$

where $k = v/a$ from (1.16) and

$$\bar{X} = \sum_{j=1}^{n} \frac{m_j}{m} X_j. \tag{1.21}$$

For details, readers are referred to Klugman et al. [8]. They are a simple introduction for traditional credibility theory. From now on, we will discuss the split credibility.

1.3 Split credibility

National Council on Compensation Insurance (NCCI) [10] introduced that NCCI's Experience Rating Plan Manual for Workers Compensation and Employers Liability Insurance (Plan) was an integral part of determining the cost of workers compensation. This is a way to customize insurance costs based on the characteristics of the employer. It provides employers with the incentive to manage their own expenses through measurable and meaningful cost-saving programs.

However, very large losses including the entire portion of the claim beyond a certain level in the experience period reduces the predictive ability of the Plan. Although very large losses are less likely to occur and are seen as more fortuitous than smaller claims, we should reduce credibility of them for making accurate estimates of future premium. Hence, the split credibility is presented.

NCCI [10] indicates a split rating approach is used to reflect both the frequency and severity of losses. The split point of individual losses is approved as part of each state's rate or loss cost filing. The amount of any individual loss up to the split point is known as primary loss, which reflects frequency. The amount in excess of the split point is known as excess loss, which reflects severity. For individual claims below the split point, the entire amount is primary loss and the excess loss is 0.

Robbin [11] introduced that the Experience Rating Plan for Workers Compensation with a primary-excess split promulgated by the NCCI made the actual losses, denoted by X, divided into primary losses, denoted by X_p, and excess losses, denoted by X_e. That is,

$$X = X_p + X_e, \tag{1.22}$$

where

$$X_p = min(X, K) \quad and \quad X_e = X - X_p. \tag{1.23}$$

This plan estimates the future losses by adding together the credibility-weighted estimates of primary and excess losses separately. That is,

$$P_c = M + Z_p \bar{X}_p + Z_e \bar{X}_e \tag{1.24}$$

where P_c is estimator of the future losses, $M = (1 - Z_p)E(X_p) + (1 - Z_e)E(X_e)$ as well as Z_p and Z_e are constant to be determined, which are credibility factors. The difference between the conventional non-split credibility and two split credibility is clear by comparing (1.8) and (1.24).

Robbin [11] assumed that the distributions of X_p and X_e were dependent on a risk parameter, Θ. Define

$$\begin{aligned}
\mu_p(\Theta) &= E(X_p|\Theta), \quad \mu_e(\Theta) = E(X_e|\Theta), \\
v_p(\Theta) &= Var(X_p|\Theta), \quad v_e(\Theta) = Var(X_e|\Theta), \\
\mu_p &= E(\mu_p(\Theta)), \quad \mu_e = E(\mu_e(\Theta)), \\
v_p &= E(v_p(\Theta)), \quad v_e = E(v_e(\Theta)), \\
a_p &= Var[\mu_p(\Theta)], \quad a_e = Var[\mu_e(\Theta)], \\
C(\Theta) &= Cov(X_p(\Theta), X_e(\Theta)), \quad \rho = E(C(\Theta)), \\
\lambda_p &= v_p + a_p, \quad \lambda_e = v_e + a_e, \\
\pi &= Cov(\mu_p(\Theta), \mu_e(\Theta)), \quad \kappa = \rho + \pi.
\end{aligned}$$

In the above, ρ is the process covariance and π is the parameter covariance.

The MSE is

$$Q = E\left\{\left[Z_p X_p + (1 - Z_p)\mu_p - \mu_p(\Theta) + Z_e X_e + (1 - Z_e)\mu_e - \mu_p(\Theta)\right]^2\right\}. \tag{1.25}$$

where $\mu_{n+1}(\Theta) = \mu(\Theta) = \mu_p(\Theta) + \mu_e(\Theta)$.

By taking derivatives with regard to Z_p and Z_e separately in (1.25) and setting them to zero, the credibility premium is

$$P_c = \widehat{\mu_X(\theta)} = \hat{Z}_p \bar{X}_p + (1 - \hat{Z}_p)\mu_p + \hat{Z}_e \bar{X}_e + (1 - \hat{Z}_e)\mu_e, \tag{1.26}$$

where

$$\hat{Z}_p = \frac{\lambda_e(a_p + \pi) - \kappa(a_e + \pi)}{D}, \tag{1.27}$$

$$\hat{Z}_e = \frac{\lambda_p(a_e + \pi) - \kappa(a_p + \pi)}{D}, \tag{1.28}$$

and

$$D = \lambda_p \lambda_e - \kappa^2. \tag{1.29}$$

Hence, the MMSE is

$$\begin{aligned} \hat{Q} &= E\left\{\left[Z_p X_p + (1 - Z_p)\mu_p - \mu_p(\Theta) + Z_e X_e + (1 - Z_e)\mu_e - \mu_p(\Theta)\right]^2\right\} && (1.30) \\ &= (a_p + \pi)(1 - \hat{Z}_p) + (a_e + \pi)(1 - \hat{Z}_e). && (1.31) \end{aligned}$$

For details, readers are referred to Robbin [11].

In Robbin [11], the MSE was regarded as a criterion for optimization. If the MMSE in split credibility would be less than the MMSE in non-split credibility, then the split method should be better than the non-split way. After considering the correlation between X_p and X_e, Robbin [11] derived the optimal split credibility of each loss by the least squares method and showed when split in effective. However, under collective risk model (CRM), the following situations may make the split ineffective:

1. Inversion: The optimal excess credibility is bigger than the optimal primary credibility.
2. Out of range: One of the optimal split credibilities is beyond 100% or negative.

In fact, Gillam [6][7] showed that there were many empirical evidences for the practicality of split credibility. Why the situations above never happened among those evidences? We find that a pivotal assumption in Gillam [6] is that primary and excess losses are uncorrelated. In addition, the optimal split credibility factors, $\hat{Z}_p$ and $\hat{Z}_e$, given by Gillam [6] are different from Robbin [11]'s. However, we find that they will be same if both the process covariance and parameter covariance are zero in Robbin [11].

As we all know, it is not possible that the process covariance and parameter covariance both become zero because X_p and X_e must exit a correlation because they are from one same X. So the assumption in Gillam [6] should not be satisfied in reality. This means that we should not ignore the covariances between the primary and excess losses.

1.4 Semi-linear credibility with truncation

Bühlmann et al. [2] assumed that the claims(losses) were from two different sources: ordinary claim with density $D_o(X|\Theta)$ with probability $1 - \pi$ and excess claim with density $D_e(X)$ with probability π. Then, the density function of claims, $f_\Theta(X)$, is given by

$$f_\Theta(X) = (1 - \pi)D_o(X|\Theta) + \pi D_e(X) \tag{1.32}$$

To estimate the future premium, we can use Equation (1.13) as before and minimize the MSE (1.11) to get the optimal parameters. However, Bühlmann et al. [2] gave us a new way to estimate the future premium. That is,

$$P_c = a + b\sum_{j=1}^{n}(X_j \wedge M) \tag{1.33}$$

where a, b and M should be chosen by minimizing the MSE,

$$Q = E\left\{\left[E(\mu(\Theta)|\mathbf{X}) - P_c\right]^2\right\} \tag{1.34}$$

$$= E\left\{\left[E(\mu(\Theta)|\mathbf{X}) - a - b\sum_{j=1}^{n}(X_j \wedge M)\right]^2\right\} \tag{1.35}$$

$$= E\left\{\left[\pi\mu_e + (1-\pi)E(\mu_o(\Theta)|\mathbf{X}) - a - b\sum_{j=1}^{n}(X_j \wedge M)\right]^2\right\} \tag{1.36}$$

This method is standard credibility technique combined with data trimming, where the parameter, M, is the trimming point.

Notice that in the model, only the ordinary part of the distribution depends on risk parameter, Θ. What's more, we realize that it is a method that keeps only the primary part of the loss distribution. The excess part is ignored. More details about this credibility estimation technique based on transformed data will be seen in Bühlmann et al. [4]. It is called semi-linear credibility.

In order to avoid the impact of large claims on the overall credibility premium, Bühlmann et al. [4] look for transformations of the data. The credibility estimator is then applied to the transformed data. One approach is to truncate either the aggregate or the individual claims.

In the thesis, the transformed data we assign is given by

$$Y_j = f(X_j) = min(X_j, K) \tag{1.37}$$

where K is the truncation point. The new credibility premium is

$$\widehat{\mu_X(\theta)} = \hat{\alpha}_0 + \sum_{j=1}^{n}\hat{\alpha}_j Y_j \tag{1.38}$$

In Bühlmann et al. [4], the semi-linear credibility estimator of $\mu_X(\theta)$ in the Bühlmann model, based on Y_j above, is given by

$$P_c = \widehat{\widehat{\mu}}_X^{(K)}(\theta) = \mu_X + \frac{n\tau_{XY}}{n\tau_Y^2 + \sigma_Y^2}(\bar{Y} - \mu_Y) \tag{1.39}$$

where

$$\mu_X = E(\mu_X(\Theta)), \tag{1.40}$$

$$\mu_Y = E(\mu_Y(\Theta)), \tag{1.41}$$

$$\tau_{XY} = Cov(\mu_X(\Theta), \mu_Y(\Theta)), \tag{1.42}$$

$$\tau_Y^2 = Var[\mu_Y(\Theta)], \tag{1.43}$$

$$\tau_X^2 = Var[\mu_X(\Theta)], \tag{1.44}$$

$$\sigma_Y^2 = E[Var(Y_j|\Theta)]. \tag{1.45}$$

The MMSE is given by

$$\hat{Q} = E\left\{\left[\widehat{\widehat{\mu}}_X^{(K)}(\Theta) - \mu_X(\Theta)\right]^2\right\} \tag{1.46}$$

$$= \tau_X^2 - \frac{n\tau_{XY}^2}{n\tau_Y^2 + \sigma_Y^2} \tag{1.47}$$

We notice that semi-linear credibility includes non-split credibility model because the credibility premium (1.38) would become (1.13) if the optimal K went to infinity.

1.5 Our model

As we said before, semi-linear credibility considers only the primary part of the loss distribution. To reflect all parts of the loss distribution, we add the excess part into semi-linear credibility. We calculate the optimal parameters α's given K by minimizing the MSE (1.11) and also calculate the value of MMSE compared with non-split credibility's and semi-linear credibility's to see if split credibility is effective. If the value of MMSE of split credibility is the smallest between them, then the split credibility is effective and is able to make the estimator of future premium more accurate. Finally, we hope to give a way to identify when we should use split credibility and which K should be chosen.

Chapter 2

Two-dimensional semi-linear credibility model

In this chapter, we look for general results of our model, called two-dimensional semi-linear credibility model, including the optimal coefficients, α's, and the value of MSE. Then we will make a summary for general results. Next, we will give specific results in split credibility. We will also discuss some properties of credibility in this model and the method for the optimal split point will be presented. Furthermore, nonparametric estimation will be discussed, which is helpful for us to solve the real problems. Finally, we derive the estimators of credibility premium of the primary part and excess part.

In the thesis, we consider the two-dimensional semi-linear credibility model in the simple Bühlmann model, that is for each policyholder(conditional on Θ), past losses $X_1, \ldots, X_n$ have the same mean and variance and are i.i.d. conditional on Θ. For details, readers are referred to Klugman et al. [8].

2.1 General results

Based on semi-linear credibility, we have only one transformed data $Y_j = f(X_j)$. Now, we add another transformed data $L_j = g(X_j)$ that is different from Y_j into the model. Hence, we get a new credibility premium(or called estimator). That is,

$$\mu_X(\theta) = \alpha_0 + \sum_{j=1}^{n} \alpha_{Yj} Y_j + \sum_{j=1}^{n} \alpha_{Lj} L_j. \tag{2.1}$$

2.1.1 The optimal α's

To choose the optimal α's, we minimize the MSE. That is,

$$Q = E\left\{\left[\mu_{n+1}(\Theta) - \alpha_0 - \sum_{j=1}^{n} \alpha_{Yj} Y_j - \sum_{j=1}^{n} \alpha_{Lj} L_j\right]^2\right\}. \tag{2.2}$$

Taking derivatives with regard to α's and setting them to zero yields for $i = 1, \ldots, n$,

$$\frac{\partial Q}{\partial \hat{\alpha_0}} = 0 = E\left\{2\left[\mu_{n+1}(\Theta) - \hat{\alpha_0} - \sum_{j=1}^{n} \hat{\alpha_{Yj}} Y_j - \sum_{j=1}^{n} \hat{\alpha_{Lj}} L_j\right](-1)\right\}, \tag{2.3}$$

$$\frac{\partial Q}{\partial \hat{\alpha_{Yi}}} = 0 = E\left\{2\left[\mu_{n+1}(\Theta) - \hat{\alpha_0} - \sum_{j=1}^{n} \hat{\alpha_{Yj}} Y_j - \sum_{j=1}^{n} \hat{\alpha_{Lj}} L_j\right](-Y_i)\right\}, \tag{2.4}$$

$$\frac{\partial Q}{\partial \hat{\alpha_{Li}}} = 0 = E\left\{2\left[\mu_{n+1}(\Theta) - \hat{\alpha_0} - \sum_{j=1}^{n} \hat{\alpha_{Yj}} Y_j - \sum_{j=1}^{n} \hat{\alpha_{Lj}} L_j\right](-L_i)\right\}, \tag{2.5}$$

where $\hat{\alpha_0}$ is the optimal α_0 and $\hat{\alpha_{Yi}}$ is the optimal α_{Yi} and $\hat{\alpha_{Li}}$ is the optimal α_{Li} for all i.

Then, expanding them yields

$$E[\mu_{n+1}(\Theta)] = \hat{\alpha_0} + \sum_{j=1}^{n} \hat{\alpha_{Yj}} E(Y_j) + \sum_{j=1}^{n} \hat{\alpha_{Lj}} E(L_j), \tag{2.6}$$

$$E[\mu_{n+1}(\Theta) Y_i] = \hat{\alpha_0} E(Y_i) + \sum_{j=1}^{n} \hat{\alpha_{Yj}} E(Y_j Y_i) + \sum_{j=1}^{n} \hat{\alpha_{Lj}} E(L_j Y_i), \tag{2.7}$$

$$E[\mu_{n+1}(\Theta) L_i] = \hat{\alpha_0} E(L_i) + \sum_{j=1}^{n} \hat{\alpha_{Yj}} E(Y_j L_i) + \sum_{j=1}^{n} \hat{\alpha_{Lj}} E(L_j L_i). \tag{2.8}$$

The left-hand side of Equation (2.7) can be rewritten as

$$\begin{aligned} E[\mu_{n+1}(\Theta) Y_i] &= E\{E[\mu_{n+1}(\Theta) Y_i | \Theta]\} \\ &= E\{\mu_{n+1}(\Theta) E[Y_i | \Theta]\} \\ &= E\{E[X_{n+1} | \Theta] E[Y_i | \Theta]\} \\ &= E\{E[X_{n+1} Y_i | \Theta]\} \\ &= E[X_{n+1} Y_i], \end{aligned}$$

where the second from the last step follows by independence of X_i and X_{n+1} conditional on Θ, which is same as Klugman et al. [8]. Then, Equation (2.7) becomes

$$E[X_{n+1} Y_i] = \hat{\alpha_0} E(Y_i) + \sum_{j=1}^{n} \hat{\alpha_{Yj}} E(Y_j Y_i) + \sum_{j=1}^{n} \hat{\alpha_{Lj}} E(L_j Y_i). \tag{2.9}$$

Hence, we also have

$$E[X_{n+1} L_i] = \hat{\alpha_0} E(L_i) + \sum_{j=1}^{n} \hat{\alpha_{Yj}} E(Y_j L_i) + \sum_{j=1}^{n} \hat{\alpha_{Lj}} E(L_j L_i), \tag{2.10}$$

which is from (2.8) and whose reason is followed by above.

Besides, the left-hand side of Equation (2.6) can be rewritten as

$$E[\mu_{n+1}(\Theta)] = E[E(X_{n+1} | \Theta)] = E[X_{n+1}].$$

Hence, we rewrite Equation (2.6) as

$$E[X_{n+1}] = \hat{\alpha_0} + \sum_{j=1}^{n} \hat{\alpha_{Yj}} E(Y_j) + \sum_{j=1}^{n} \hat{\alpha_{Lj}} E(L_j). \tag{2.11}$$

Next, mutiply (2.11) by $E[Y_i]$ and subtract from (2.9) to obtain

$$Cov(X_{n+1}, Y_i) = \sum_{j=1}^{n} \hat{\alpha_{Yj}} Cov(Y_j, Y_i) + \sum_{j=1}^{n} \hat{\alpha_{Lj}} Cov(L_j, Y_i), \tag{2.12}$$

in the meantime mutiplying (2.11) by $E[L_i]$ and subtract from (2.10), we get

$$Cov(X_{n+1}, L_i) = \sum_{j=1}^{n} \hat{\alpha_{Yj}} Cov(Y_j, L_i) + \sum_{j=1}^{n} \hat{\alpha_{Lj}} Cov(L_j, L_i), \tag{2.13}$$

for $i = 1, \ldots, n$.

For the convenience of expression, we give a table of our notations in the thesis. That is,

Notation	Expression		Notation	Expression		Notation	Expression
$\mu_X(\Theta)$	$E[X\|\Theta]$		$\mu_Y(\Theta)$	$E[Y\|\Theta]$		$\mu_L(\Theta)$	$E[L\|\Theta]$
μ_X	$E[\mu_X(\Theta)]$		μ_Y	$E[\mu_Y(\Theta)]$		μ_L	$E[\mu_L(\Theta)]$
τ_X^2	$Var[\mu_X(\Theta)]$		τ_Y^2	$Var[\mu_Y(\Theta)]$		τ_L^2	$Var[\mu_L(\Theta)]$
σ_X^2	$E[Var(X\|\Theta)]$		σ_Y^2	$E[Var(Y\|\Theta)]$		σ_L^2	$E[Var(L\|\Theta)]$
τ_{XY}	$Cov(\mu_X(\Theta), \mu_Y(\Theta))$		τ_{XL}	$Cov(\mu_X(\Theta), \mu_L(\Theta))$		τ_{YL}	$Cov(\mu_Y(\Theta), \mu_L(\Theta))$
$\delta_{XY}(\Theta)$	$Cov(X\|\Theta, Y\|\Theta)$		$\delta_{XL}(\Theta)$	$Cov(X\|\Theta, L\|\Theta)$		$\delta_{YL}(\Theta)$	$Cov(Y\|\Theta, L\|\Theta)$
δ_{XY}	$E[\delta_{XY}(\Theta)]$		δ_{XL}	$E[\delta_{XL}(\Theta)]$		δ_{YL}	$E[\delta_{YL}(\Theta)]$

Table 2.1: A list of notations

Using the notation above, the left-hand side of Equation (2.12) can be rewritten as

$$\begin{aligned} Cov(X_{n+1}, Y_i) &= E[X_{n+1}Y_i] - E[X_{n+1}]E[Y_i] \\ &= E\{E[X_{n+1}Y_i|\Theta]\} - E[\mu_X(\Theta)]E[\mu_Y(\Theta)] \\ &= E\{E[X_{n+1}|\Theta]E[Y_i|\Theta]\} - E[\mu_X(\Theta)]E[\mu_Y(\Theta)] \\ &= E[\mu_X(\Theta)\mu_Y(\Theta)] - E[\mu_X(\Theta)]E[\mu_Y(\Theta)] \\ &= Cov(\mu_X(\Theta), \mu_Y(\Theta)) \\ &= \tau_{XY}, \end{aligned}$$

for $i = 1, \ldots, n$ and the reason of the third step is independence of X_i and X_{n+1} conditional on Θ.

Under the same reason, we have

$$Cov(Y_j, Y_i) = \tau_Y^2 \quad and \quad Cov(L_j, Y_i) = \tau_{YL},$$

for $j \neq i$ and $i, j = 1, \ldots, n$.

Now, we rewrite Equation (2.12) as

$$\begin{aligned}
\tau_{XY} &= \sum_{\substack{j=1\\ j\neq i}}^{n} \hat{\alpha_{Y j}} Cov(Y_j, Y_i) + \hat{\alpha_{Yi}} Var(Y_i) + \sum_{\substack{j=1\\ j\neq i}}^{n} \hat{\alpha_{Lj}} Cov(L_j, Y_i) + \hat{\alpha_{Li}} Cov(L_i, Y_i) && (2.14)\\
&= \sum_{\substack{j=1\\ j\neq i}}^{n} \hat{\alpha_{Y j}} \tau_Y^2 + \hat{\alpha_{Yi}}(\sigma_Y^2 + \tau_Y^2) + \sum_{\substack{j=1\\ j\neq i}}^{n} \hat{\alpha_{Lj}} \tau_{YL} + \hat{\alpha_{Li}}(\delta_{YL} + \tau_{YL}) && (2.15)\\
&= \sum_{j=1}^{n} \hat{\alpha_{Y j}} \tau_Y^2 + \hat{\alpha_{Yi}}\sigma_Y^2 + \sum_{j=1}^{n} \hat{\alpha_{Lj}} \tau_{YL} + \hat{\alpha_{Li}}\delta_{YL}, && (2.16)
\end{aligned}$$

where

$$\begin{aligned}
Var(Y_i) &= E[Var(Y_i|\Theta)] + Var[E(Y_i|\Theta)] && (2.17)\\
&= \sigma_Y^2 + Var[\mu_Y(\Theta)] && (2.18)\\
&= \sigma_Y^2 + \tau_Y^2, && (2.19)\\
Cov(L_i, Y_i) &= E[Cov(L_i|\Theta, Y_i|\Theta)] + Cov(E[L_i|\Theta], E[Y_i|\Theta]) && (2.20)\\
&= E[Cov(L|\Theta, Y|\Theta)] + Cov(\mu_L(\Theta), \mu_Y(\Theta)) && (2.21)\\
&= \delta_{YL} + \tau_{YL}. && (2.22)
\end{aligned}$$

Compared with (2.16), Equation (2.13) can be rewritten as

$$\tau_{XL} = \sum_{j=1}^{n} \hat{\alpha_{Y j}} \tau_{YL} + \hat{\alpha_{Yi}}\delta_{YL} + \sum_{j=1}^{n} \hat{\alpha_{Lj}} \tau_L^2 + \hat{\alpha_{Li}}\sigma_L^2, \qquad (2.23)$$

for $i = 1, \ldots, n$.

Next, mutiply (2.16) by σ_L^2 and mutiply (2.23) by δ_{YL}, then subtract each other to obtain

$$\hat{\alpha_{Yi}} = \frac{\tau_{XY}\sigma_L^2 - \tau_{XL}\delta_{YL} - (\tau_Y^2\sigma_L^2 - \tau_{YL}\delta_{YL}) \sum_{j=1}^{n} \hat{\alpha_{Y j}} - (\tau_{YL}\sigma_L^2 - \tau_L^2\delta_{YL}) \sum_{j=1}^{n} \hat{\alpha_{Lj}}}{\sigma_Y^2\sigma_L^2 - \delta_{YL}^2}, \qquad (2.24)$$

if $\sigma_Y^2\sigma_L^2 \neq \delta_{YL}^2$ and for $i = 1, \ldots, n$.

We notice that the right-hand side of Equation (2.24) does not depend on i. So we have that $\hat{\alpha_{Y j}} = \hat{\alpha_{Yi}}$ for $i \neq j$ and $i, j = 1, \ldots, n$. We let them be $\hat{\alpha_Y}$. With the same reason, we let all $\hat{\alpha_{Li}}$ be $\hat{\alpha_L}$ for $i = 1, \ldots, n$. Then, we have

$$\begin{aligned}
\tau_{XY} &= \hat{\alpha_Y}(n\tau_Y^2 + \sigma_Y^2) + \hat{\alpha_L}(n\tau_{YL} + \delta_{YL}),\\
\tau_{XL} &= \hat{\alpha_Y}(n\tau_{YL} + \delta_{YL}) + \hat{\alpha_L}(n\tau_L^2 + \sigma_L^2),
\end{aligned}$$

which can be rewritten in matrix form as

$$\begin{pmatrix} n\tau_Y^2 + \sigma_Y^2 & n\tau_{YL} + \delta_{YL} \\ n\tau_{YL} + \delta_{YL} & n\tau_L^2 + \sigma_L^2 \end{pmatrix} \begin{pmatrix} \hat{\alpha_Y} \\ \hat{\alpha_L} \end{pmatrix} = \begin{pmatrix} \tau_{XY} \\ \tau_{XL} \end{pmatrix}. \qquad (2.25)$$

Hence, $\hat{\alpha_Y}$ and $\hat{\alpha_L}$ have a unique solution,

$$\hat{\alpha_Y} = \frac{\tau_{XY}(n\tau_L^2 + \sigma_L^2) - \tau_{XL}(n\tau_{YL} + \delta_{YL})}{(n\tau_Y^2 + \sigma_Y^2)(n\tau_L^2 + \sigma_L^2) - (n\tau_{YL} + \delta_{YL})^2}, \tag{2.26}$$

$$\hat{\alpha_L} = \frac{\tau_{XL}(n\tau_Y^2 + \sigma_Y^2) - \tau_{XY}(n\tau_{YL} + \delta_{YL})}{(n\tau_Y^2 + \sigma_Y^2)(n\tau_L^2 + \sigma_L^2) - (n\tau_{YL} + \delta_{YL})^2}, \tag{2.27}$$

if

$$(n\tau_Y^2 + \sigma_Y^2)(n\tau_L^2 + \sigma_L^2) \neq (n\tau_{YL} + \delta_{YL})^2. \tag{2.28}$$

Let A represent $(n\tau_Y^2 + \sigma_Y^2)$, B represent $n\tau_L^2 + \sigma_L^2$ and C represent $(n\tau_{YL} + \delta_{YL})$. The matrix becomes

$$\begin{pmatrix} n\tau_Y^2 + \sigma_Y^2 & n\tau_{YL} + \delta_{YL} \\ n\tau_{YL} + \delta_{YL} & n\tau_L^2 + \sigma_L^2 \end{pmatrix} = \begin{pmatrix} A & C \\ C & B \end{pmatrix}.$$

If the determinant of this matix is zero, that is $AB = C^2$. Then, there are no solution for Equation (2.25) if $\tau_{XY}/\tau_{XL} \neq C/B$, and there are infinite solutions if $\tau_{XY}/\tau_{XL} = C/B$. For details of matrix with zero determinant, readers are referred to Marcus et al. [9].

Now, we get the value of $\hat{\alpha_Y}$ and $\hat{\alpha_L}$. Then, we can get the value of $\hat{\alpha_0}$ from Equation (2.11). That is

$$\hat{\alpha_0} = \mu_X - n\hat{\alpha_Y}\mu_Y - n\hat{\alpha_L}\mu_L. \tag{2.29}$$

Now, we get the new credibility premium from (2.1). That is

$$\begin{aligned}
\widehat{\mu_X(\theta)} &= \hat{\alpha_0} + \hat{\alpha_Y}\sum_{j=1}^{n} Y_j + \hat{\alpha_L}\sum_{j=1}^{n} L_j && (2.30)\\
&= \mu_X - n\hat{\alpha_Y}\mu_Y - n\hat{\alpha_L}\mu_L + n\hat{\alpha_Y}\bar{Y} + n\hat{\alpha_L}\bar{L} && (2.31)\\
&= \mu_X + n\hat{\alpha_Y}(\bar{Y} - \mu_Y) + n\hat{\alpha_L}(\bar{L} - \mu_L), && (2.32)
\end{aligned}$$

where $(\hat{\alpha_Y}, \hat{\alpha_L})$ are the optimal α's.

2.1.2 The minimum value of MSE

Next, we also need to calculate the minimum value of MSE (2.2) because we will compare it with the MMSEs of non-split credibility and semi-linear credibility. So, we have

$$\begin{aligned}
\hat{Q} &= E\left\{\left[\mu_{n+1}(\Theta) - \widehat{\mu_X(\Theta)}\right]^2\right\} && (2.33)\\
&= E\left\{\left[\mu_{n+1}(\Theta) - \hat{\alpha}_0 - n\hat{\alpha}_Y\bar{Y} - n\hat{\alpha}_L\bar{L}\right]^2\right\} && (2.34)\\
&= E\left[\left(\mu_{n+1}(\Theta) - n\hat{\alpha}_Y\bar{Y}\right)^2 + \left(\hat{\alpha}_0 + n\hat{\alpha}_L\bar{L}\right)^2 - 2\left(\mu_{n+1}(\Theta) - n\hat{\alpha}_Y\bar{Y}\right)\left(\hat{\alpha}_0 + n\hat{\alpha}_L\bar{L}\right)\right] && (2.35)\\
&= E\left(\mu_{n+1}(\Theta)^2\right) + n^2\hat{\alpha}_Y{}^2E\left(\bar{Y}^2\right) - 2n\hat{\alpha}_YE\left(\mu_{n+1}(\Theta)\bar{Y}\right) + \hat{\alpha}_0{}^2 + n^2\hat{\alpha}_L{}^2E\left(\bar{L}^2\right) + 2n\hat{\alpha}_0\hat{\alpha}_LE\left(\bar{L}\right)\\
&\quad -2\hat{\alpha}_0E\left(\mu_{n+1}(\Theta)\right) - 2n\hat{\alpha}_LE\left(\mu_{n+1}(\Theta)\bar{L}\right) + 2n\hat{\alpha}_0\hat{\alpha}_YE\left(\bar{Y}\right) + 2n^2\hat{\alpha}_L\hat{\alpha}_YE\left(\bar{Y}\bar{L}\right) && (2.36)\\
&= \tau_X^2 + \mu_X^2 + n^2\hat{\alpha}_Y{}^2\left(\frac{1}{n}\sigma_Y^2 + \tau_Y^2 + \mu_Y^2\right) - 2n\hat{\alpha}_Y(\tau_{XY} + \mu_X\mu_Y) + \hat{\alpha}_0{}^2\\
&\quad +n^2\hat{\alpha}_L{}^2\left(\frac{1}{n}\sigma_L^2 + \tau_L^2 + \mu_L^2\right) + 2n\hat{\alpha}_0\hat{\alpha}_L\mu_L - 2\hat{\alpha}_0\mu_X - 2n\hat{\alpha}_L(\tau_{XL} + \mu_X\mu_L)\\
&\quad +2n\hat{\alpha}_0\hat{\alpha}_Y\mu_Y + 2n^2\hat{\alpha}_L\hat{\alpha}_Y\left(\frac{1}{n}\delta_{YL} + \tau_{YL} + \mu_Y\mu_L\right) && (2.37)\\
&= \tau_X^2 - \mu_X^2 + n\hat{\alpha}_Y{}^2\sigma_Y^2 + n^2\hat{\alpha}_Y{}^2\tau_Y^2 - n^2\hat{\alpha}_Y{}^2\mu_Y^2 + (\mu_X - n\hat{\alpha}_Y\mu_Y - n\hat{\alpha}_L\mu_L)^2\\
&\quad +n\hat{\alpha}_L{}^2\sigma_L^2 + n^2\hat{\alpha}_L{}^2\tau_L^2 - n^2\hat{\alpha}_L{}^2\mu_L^2 + 2n(\mu_X - n\hat{\alpha}_Y\mu_Y)\hat{\alpha}_L\mu_L + 2n\hat{\alpha}_Y\mu_X\mu_Y\\
&\quad +2n\hat{\alpha}_L\hat{\alpha}_Y\delta_{YL} + 2n^2\hat{\alpha}_L\hat{\alpha}_Y\tau_{YL} - 2n\hat{\alpha}_Y\tau_{XY} - 2n\hat{\alpha}_L\tau_{XL} && (2.38)\\
&= \tau_X^2 + n\hat{\alpha}_Y{}^2\sigma_Y^2 + n^2\hat{\alpha}_Y{}^2\tau_Y^2 - 2n\hat{\alpha}_Y\tau_{XY}\\
&\quad +n\hat{\alpha}_L{}^2\sigma_L^2 + n^2\hat{\alpha}_L{}^2\tau_L^2 - 2n\hat{\alpha}_L\tau_{XL}\\
&\quad +2n\hat{\alpha}_L\hat{\alpha}_Y\delta_{YL} + 2n^2\hat{\alpha}_L\hat{\alpha}_Y\tau_{YL}, && (2.39)
\end{aligned}$$

where $\hat{\alpha}_L$ and $\hat{\alpha}_Y$ meet Equation (2.25).The details of why Equation (2.36) goes to (2.37) will be seen in Appendix A.

Remark The minimum value of MSE above(or called $\hat{Q}$) is the minimum value given by one K in our model. Giving a different K will result in a different MMSE.

2.1.3 Summary

In summary, we give two Theorems as follows.

Theorem 2.1.1 *The two-dimensional semi-linear credibility estimator of $\mu_X(\theta)$ in the simple Bühlmann model, based on two different transformed data $Y_j = f(X_j)$ and $L_j = g(X_j)$, is given by*

$$\widehat{\mu_X(\theta)} = \mu_X + n\hat{\alpha}_Y(\bar{Y} - \mu_Y) + n\hat{\alpha}_L(\bar{L} - \mu_L),$$

where $(\hat{\alpha}_Y, \hat{\alpha}_L)$ satisfies

$$\begin{pmatrix} n\tau_Y^2 + \sigma_Y^2 & n\tau_{YL} + \delta_{YL} \\ n\tau_{YL} + \delta_{YL} & n\tau_L^2 + \sigma_L^2 \end{pmatrix}\begin{pmatrix} \hat{\alpha}_Y \\ \hat{\alpha}_L \end{pmatrix} = \begin{pmatrix} \tau_{XY} \\ \tau_{XL} \end{pmatrix},$$

or

$$\begin{aligned}\hat{\alpha_Y} &= \frac{\tau_{XY}(n\tau_L^2+\sigma_L^2)-\tau_{XL}(n\tau_{YL}+\delta_{YL})}{(n\tau_Y^2+\sigma_Y^2)(n\tau_L^2+\sigma_L^2)-(n\tau_{YL}+\delta_{YL})^2},\\ \hat{\alpha_L} &= \frac{\tau_{XL}(n\tau_Y^2+\sigma_Y^2)-\tau_{XY}(n\tau_{YL}+\delta_{YL})}{(n\tau_Y^2+\sigma_Y^2)(n\tau_L^2+\sigma_L^2)-(n\tau_{YL}+\delta_{YL})^2},\end{aligned}$$

if $(n\tau_Y^2+\sigma_Y^2)(n\tau_L^2+\sigma_L^2) \neq (n\tau_{YL}+\delta_{YL})^2$.

Theorem 2.1.2 *The minimum mean square error of the two-dimensional semi-linear credibility estimator in the simple Bühlmann model is given by*

$$\begin{aligned}\hat{Q_{min}} = E\left\{\left[\mu_{n+1}(\Theta)-\alpha_0-\sum_{j=1}^{n}\alpha_{Yj}Y_j-\sum_{j=1}^{n}\alpha_{Lj}L_j\right]^2\right\} &= \tau_X^2+n\hat{\alpha_Y}^2\sigma_Y^2+n^2\hat{\alpha_Y}^2\tau_Y^2-2n\hat{\alpha_Y}\tau_{XY}\\ &\quad +n\hat{\alpha_L}^2\sigma_L^2+n^2\hat{\alpha_L}^2\tau_L^2-2n\hat{\alpha_L}\tau_{XL}\\ &\quad +2n\hat{\alpha_L}\hat{\alpha_Y}\delta_{YL}+2n^2\hat{\alpha_L}\hat{\alpha_Y}\tau_{YL}.\end{aligned}$$

Remarks The two-dimensional semi-linear credibility model includes non-split credibility model and semi-linear credibility model. It will be semi-linear credibility model if $Y_j = f(X_j) = g(X_j) = L_j$ for $j = 1,\ldots,n$. Furthermore, it will be non-split credibility model if $Y_j = f(X_j) = X_j = g(X_j) = L_j$ for $j = 1,\ldots,n$.

In general, condition (2.28) will not be met if $Y_j = f(X_j) = g(X_j) = L_j$ for $j = 1,\ldots,n$, which means that we will not be able to use Equations (2.26) and (2.27) to get the optimal α's. However, Equation (2.25) will always be established whatever the transformed data are.

2.2 Split results

According to Section 1.3, split credibility model has two different partial losses, primary loss and excess loss. So, generally, we should let Y_j be the primary loss and let L_j be the excess loss. However, in order to simplify the expressions of optimal α's and MSE as well as be more intuitive to compare with non-split credibility model and semi-linear credibility model with truncation, we keep Y being $X_p = min(X, K)$ but set $L = X$. The relationship between their parameters is as follows,

$$\begin{aligned}
\mu_X(\theta) &= \alpha_0 + \alpha_p \sum_{j=1}^{n} X_{pj} + \alpha_e \sum_{j=1}^{n} X_{ej} \\
&= \alpha_0 + \alpha_p \sum_{j=1}^{n} X_{pj} + \alpha_e \sum_{j=1}^{n} (X_j - X_{pj}) \\
&= \alpha_0 + (\alpha_p - \alpha_e) \sum_{j=1}^{n} X_{pj} + \alpha_e \sum_{j=1}^{n} X_j \\
&= \alpha_0 + \alpha_Y \sum_{j=1}^{n} Y_j + \alpha_L \sum_{j=1}^{n} L_j \\
&= \alpha_0 + \alpha_Y \sum_{j=1}^{n} Y_j + \alpha_X \sum_{j=1}^{n} X_j,
\end{aligned}$$

where $X_{pj} = min(X_j, K)$ and $X_{ej} = X_j - X_{pj}$ as well as K is a split point. In the meantime, we change the notation α_L to α_X and change the notation L to X.

Hence, we should have

$$\hat{\alpha}_p = \hat{\alpha}_Y + \hat{\alpha}_L = \hat{\alpha}_Y + \hat{\alpha}_X, \tag{2.40}$$

$$\hat{\alpha}_e = \hat{\alpha}_L = \hat{\alpha}_X. \tag{2.41}$$

In this way, we have the accurate expressions of credibility premium and MMSE as follows,

$$\widehat{\mu_X(\theta)} = \mu_X + n\hat{\alpha}_Y(\bar{Y} - \mu_Y) + n\hat{\alpha}_X(\bar{X} - \mu_X) \tag{2.42}$$

$$= \mu_X + \widehat{Z_Y}(\bar{Y} - \mu_Y) + \widehat{Z_X}(\bar{X} - \mu_X), \tag{2.43}$$

where $\hat{\alpha}_Y$ and $\hat{\alpha}_X$ meet

$$\begin{pmatrix} n\tau_Y^2 + \sigma_Y^2 & n\tau_{XY} + \delta_{XY} \\ n\tau_{XY} + \delta_{XY} & n\tau_X^2 + \sigma_X^2 \end{pmatrix} \begin{pmatrix} \hat{\alpha}_Y \\ \hat{\alpha}_X \end{pmatrix} = \begin{pmatrix} \tau_{XY} \\ \tau_X^2 \end{pmatrix}, \tag{2.44}$$

and

$$\begin{aligned}
\hat{Q} &= \tau_X^2 + n\hat{\alpha}_Y{}^2\sigma_Y^2 + n^2\hat{\alpha}_Y{}^2\tau_Y^2 - 2n\hat{\alpha}_Y\tau_{XY} \\
&\quad + n\hat{\alpha}_X{}^2\sigma_X^2 + n^2\hat{\alpha}_X{}^2\tau_X^2 - 2n\hat{\alpha}_X\tau_X^2 \\
&\quad + 2n\hat{\alpha}_X\hat{\alpha}_Y\delta_{XY} + 2n^2\hat{\alpha}_X\hat{\alpha}_Y\tau_{XY}
\end{aligned} \tag{2.45}$$

$$\begin{aligned}
&= \tau_X^2 + \frac{1}{n}\widehat{Z_Y}^2\sigma_Y^2 + \widehat{Z_Y}^2\tau_Y^2 - 2\widehat{Z_Y}\tau_{XY} \\
&\quad + \frac{1}{n}\widehat{Z_X}^2\sigma_X^2 + \widehat{Z_X}^2\tau_X^2 - 2\widehat{Z_X}\tau_X^2 \\
&\quad + \frac{2}{n}\widehat{Z_X}\widehat{Z_Y}\delta_{XY} + 2\widehat{Z_X}\widehat{Z_Y}\tau_{XY},
\end{aligned} \tag{2.46}$$

where $\widehat{Z_Y} = n\hat{\alpha}_Y$ and $\widehat{Z_X} = n\hat{\alpha}_X$ which are credibility of each part.

Again, we let A represent $(n\tau_Y^2 + \sigma_Y^2)$, B represent $n\tau_X^2 + \sigma_X^2$ and C represent $(n\tau_{XY} + \delta_{XY})$. The matrix becomes

$$\begin{pmatrix} n\tau_Y^2 + \sigma_Y^2 & n\tau_{XY} + \delta_{XY} \\ n\tau_{XY} + \delta_{XY} & n\tau_X^2 + \sigma_X^2 \end{pmatrix} = \begin{pmatrix} A & C \\ C & B \end{pmatrix}. \tag{2.47}$$

If the determinant of this matix is zero, there are no solution for Equation (2.44) if $\tau_{XY}/\tau_X^2 \neq C/B$, and there are infinite solutions if $\tau_{XY}/\tau_X^2 = C/B$. For details of matrix with zero determinant, readers are referred to Marcus et al. [9].

In general, A and B are not equal to 0, but $A = 0$ when $K = 0$. In this situation, $\tau_{XY} = 0$, $C = 0$, $Y = 0$ and our credibility model becomes non-split credibility model.

When we have infinite solutions, the solution of non-split credibility model is also a solution for us in our credibility model, which means that the minimum value of MSE in our credibility model is same as the minimum value of MSE in non-split credibility model. Hence, we choose the solution of non-split credibility model as the solution of our credibility model in this time.

2.2.1 Properties of credibility

Before we study the optimal split point, we present some properties of credibility. This will help us gain intuition and make practical sense of our model.

Rewrite Equation (2.25) as

$$\begin{aligned} \widehat{Z_X} &= \frac{n\tau_{XY}}{n\tau_{XY} + \delta_{XY}} - \frac{n\tau_Y^2 + \sigma_Y^2}{n\tau_{XY} + \delta_{XY}} \widehat{Z_Y}, & (2.48) \\ \widehat{Z_Y} &= \frac{n\tau_X^2}{n\tau_{XY} + \delta_{XY}} - \frac{n\tau_X^2 + \sigma_X^2}{n\tau_{XY} + \delta_{XY}} \widehat{Z_X}, & (2.49) \end{aligned}$$

if $n\tau_{XY} + \delta_{XY} \neq 0$. In general, this condition should be always met because if $X < K$, then $Y = X$ and τ_{XY} would be τ_X^2, which is always bigger than zero except $\mu_X(\Theta)$ is a constant but it should not be happened, as well as δ_{XY} would be σ_X^2, whose property is same as τ_X^2. Otherwise, Y would be always equal to K and we would get $n\tau_{XY} + \delta_{XY} = 0$. In an extreme case which is that all of X are bigger or equal to K, our model would become non-split credibility model and this situation would be the same as $K = 0$.

Let $n \to +\infty$, (2.48) and (2.49) become

$$\begin{aligned} \widehat{Z_X} &= 1 - \frac{\tau_Y^2}{\tau_{XY}} \widehat{Z_Y}, & (2.50) \\ \widehat{Z_Y} &= \frac{\tau_X^2}{\tau_{XY}} - \frac{\tau_X^2}{\tau_{XY}} \widehat{Z_X}, & (2.51) \end{aligned}$$

thus we get $\widehat{Z_X} \to 1$ and $\widehat{Z_Y} \to 0$ if $\tau_{XY}^2 \neq \tau_X^2\tau_Y^2$. In general, this condition is met except $Y = X$ or $K = 0$ or $K \to +\infty$. As we all know, we should give more credibility to the data if there are a lot of losses we get. The properties of credibility above meet our expectation.

For the time being, we don't restrict $Y = min(X, K)$ and let it can be any functions on X for now and go to find the influences of parameters τ_{XY} and δ_{XY}.

Let $\tau_{XY} \to \infty$, we get $\widehat{Z_X} \to 1$ and $\widehat{Z_Y} \to 0$. From the results, we know that this influence is the same as n's if we choose a function on X, $Y = f(X)$, with a large value of τ_{XY}. Both of them will cause more credibility on X. Notice that the value can be either very small or very large.

Similarly, let $\delta_{XY} \to \infty$, we get $\widehat{Z_X} \to 0$ and $\widehat{Z_Y} \to 0$. From the results, we know that we should use the manual rate, μ_X, to estimate the future loss if we choose a function on X, $Y = f(X)$, with a large value of δ_{XY}. Notice that the value can be either very small or very large.

Remarks In the above, we discuss the sensitivity of the credibility to the values of a parameter assuming that other parameters remain unchanged. In fact, other paramters may be changed with τ_{XY} or δ_{XY}, but we don't consider this situation here. So, further research can be whether any other parameters would be changed or not when the parameters we specify change if you are interested in these results and want to explore more.

Notice that the credibility of primary loss, $\widehat{Z_p}$, is equal to $\widehat{Z_Y} + \widehat{Z_X}$ and the credibility of excess loss, $\widehat{Z_e}$, is equal to $\widehat{Z_X}$. We are not able to ensure that $\widehat{Z_p}$ and $\widehat{Z_e}$ are between 0 and 1 as commented in Robbin [11].

2.2.2 The optimal split point

The optimal split point minimizes the mean square error. Let the MMSE of non-split credibility be $\hat{Q}_{nsp}$, the MMSE of semi-linear credibility be $\hat{Q}_{tsl}$ and the MMSE of two-dimensional semi-linear credibility be $\hat{Q}_{tdsl}$. From (1.17), (1.47) and (2.45), we respectively have

$$\hat{Q}_{nsp} = \frac{\sigma_X^2 \tau_X^2}{\sigma_X^2 + n\tau_X^2}, \tag{2.52}$$

$$\hat{Q}_{tsl} = \tau_X^2 - \frac{n\tau_{XY}^2}{\sigma_Y^2 + n\tau_Y^2}, \tag{2.53}$$

$$\begin{aligned} \hat{Q}_{tdsl} = {} & \tau_X^2 + n\hat{\alpha_Y}^2\sigma_Y^2 + n^2\hat{\alpha_Y}^2\tau_Y^2 - 2n\hat{\alpha_Y}\tau_{XY} \\ & + n\hat{\alpha_X}^2\sigma_X^2 + n^2\hat{\alpha_X}^2\tau_X^2 - 2n\hat{\alpha_X}\tau_X^2 \\ & + 2n\hat{\alpha_X}\hat{\alpha_Y}\delta_{XY} + 2n^2\hat{\alpha_X}\hat{\alpha_Y}\tau_{XY}. \end{aligned} \tag{2.54}$$

We notice that $\hat{Q}_{tdsl}$ is a function of the split point, K, because $\hat{\alpha_X}$, $\hat{\alpha_Y}$, μ_Y, σ_Y^2, τ_Y^2, δ_{XY} and τ_{XY} are functions on K. In theory, we take derivatives of $\hat{Q}_{tdsl}$ with respect to K and let the expression equal to zero for getting the optimal split point, $\hat{K}$. In the mathematical formula, it is

$$\frac{\partial \hat{Q}_{tdsl}}{\partial \hat{K}} = 0.$$

However, it is hard to solve this equation because the processes are extremely complex. For practical application, we decide to use another simple method to determine the optimal split point. Before that, we give an inspiring and accurate approach as well as hope that it will be helpful with future research. Now we give the Leibniz's Rule. For details, readers are referred to Border [5].

Lemma 2.2.1 (Leibniz's Rule) *Let $A \subset \boldsymbol{R}^n$ be open, let $I = [a, b] \subset \boldsymbol{R}$ be a compact interval, and let f be a (jointly) continuous mapping of $A \times I$ into $\boldsymbol{R}$. Let α and β be two continuously differentiable mappings of A into I. Then*

$$g(x) = \int_{\alpha(x)}^{\beta(x)} f(x,t)\,dt,$$

is continuous in A. If in addition, the partial derivative $\frac{\partial f}{\partial x}$ exists and is (jointly) continuous on $A \times I$, then g is continuously differentiable on A and

$$g'(x) = \int_{\alpha(x)}^{\beta(x)} \frac{\partial f(x,t)}{\partial x} dt + f(x, \beta(x))\beta'(x) - f(x, \alpha(x))\alpha'(x).$$

Notice that the conditional mean of $Y = X_p$ is the function on X and K. Under the integral expression, the integral interval is only related to K and the function is only integrated on X. Therefore, we can use the Leibniz's Rule to get our partial differential equations. Suppose that the losses, X, have the same distribution with pdf, $f_{X|\Theta}(x)$, and cdf, $F_{X|\Theta}(x)$, conditional on Θ. Let $Y = min(X, K)$, then

$$\mu_Y(\Theta) = \int_0^K x f_{X|\Theta}(x)\,dx + K(1 - F_{X|\Theta}(K)).$$

Under the Leibniz's Rule, we take partial derivative of $\mu_Y(\Theta)$ with respect to K, then

$$\begin{aligned}
\frac{\partial \mu_Y(\Theta)}{\partial K} &= \int_0^K \frac{\partial\,(x f_{X|\Theta}(x))}{\partial K} dx + K f_{X|\Theta}(K) - 0 f_{X|\Theta}(0) \\
&\quad +1 - F_{X|\Theta}(K) - K f_{X|\Theta}(K) \\
&= K f_{X|\Theta}(K) + 1 - F_{X|\Theta}(K) - K f_{X|\Theta}(K) \\
&= 1 - F_{X|\Theta}(K).
\end{aligned}$$

Using the same method, all of the partial differential equations of parameters are derived. In the end, the numerical solution of K should be obtained. However, we guess that the solution would be so complex that it is not easy to be used in reality and we give another way to get it. Let us turn to research for size relationships between $\hat{Q}_{nsp}$, $\hat{Q}_{tsl}$ and $\hat{Q}_{tdsl}$. Then, we get a Theorem as follows,

Theorem 2.2.2 *Let the MMSE of non-split credibility be $\hat{Q}_{nsp}$, the MMSE of semi-linear credibility be $\hat{Q}_{tsl}$ and the MMSE of two-dimensional semi-linear credibility be $\hat{Q}_{tdsl}$. Then, $\hat{Q}_{tdsl}$ will be the minimum value between them, that is,*

$$\hat{Q}_{tdsl} \leq \hat{Q}_{nsp} \quad \textit{and} \quad \hat{Q}_{tdsl} \leq \hat{Q}_{tsl},$$

where $\hat{Q}_{nsp}$ is (2.52), $\hat{Q}_{tsl}$ is (2.53) and $\hat{Q}_{tdsl}$ is (2.54).

Proof: We only proof the first inequality, $\hat{Q}_{tdsl} \leq \hat{Q}_{nsp}$, and the other's proof is similar. Notice that when $\hat{\alpha_Y}$ is equal to 0, we have

$$\begin{aligned}
\hat{Q}_{tdsl} &= \tau_X^2 + n\hat{\alpha_X}^2\sigma_X^2 + n^2\hat{\alpha_X}^2\tau_X^2 - 2n\hat{\alpha_X}\tau_X^2 \\
&= n\hat{\alpha_X}^2\sigma_X^2 + (1 - n\hat{\alpha_X})^2\tau_X^2,
\end{aligned}$$

where $\hat{\alpha_X} = \frac{\tau_X^2}{\sigma_X^2 + n\tau_X^2}$. Then,

$$\begin{aligned}\hat{Q}_{tdsl} &= n\hat{\alpha_X}^2\sigma_X^2 + (1 - n\hat{\alpha_X})^2\tau_X^2 \\ &= \frac{n\tau_X^4\sigma_X^2}{(\sigma_X^2 + n\tau_X^2)^2} + \frac{\sigma_X^4\tau_X^2}{(\sigma_X^2 + n\tau_X^2)^2} \\ &= \frac{\sigma_X^2\tau_X^2}{\sigma_X^2 + n\tau_X^2} \\ &= \hat{Q}_{nsp},\end{aligned}$$

which shows again that the two-dimensional semi-linear credibility includes non-split credibility.

Assume that the MMSE of two-dimensional semi-linear credibility, $\hat{Q}_{tdsl}$, is bigger than $\hat{Q}_{nsp}$ with the optimal $\hat{\alpha_X}$ and $\hat{\alpha_Y}$ where $\hat{\alpha_Y} \neq 0$. However, we know that we could find the value of MMSE, $\hat{\hat{Q}}_{tdsl} = \hat{Q}_{nsp}$ as long as we selected $\hat{\hat{\alpha}}_Y = 0$ and $\hat{\hat{\alpha}}_X = \frac{\tau_X^2}{\sigma_X^2 + n\tau_X^2}$. And we could get a smaller value. Therefore, $\hat{Q}_{tdsl}$ with the optimal $\hat{\alpha_Y} \neq 0$ is not the MMSE and it is not existed.

In total, we have $\hat{Q}_{tdsl} \leq \hat{Q}_{nsp}$ for all times. ∎

As we said before, if the determinant of matix (2.47) is zero, then there are no solution for Equation (2.44) or the solution of non-split credibility model would be chosen as the solution of our credibility model. However, there is a special case that we do not need to split the losses if the determinant of matrix (2.47) is not zero. We give a Theorem as follows.

Theorem 2.2.3 *Assume that $\tau_X^2 \neq 0$, $\tau_{XY} \neq 0$ and $\delta_{XY}/\tau_{XY} = \sigma_X^2/\tau_X^2$, as well as,*

$$\begin{vmatrix} n\tau_Y^2 + \sigma_Y^2 & n\tau_{XY} + \delta_{XY} \\ n\tau_{XY} + \delta_{XY} & n\tau_X^2 + \sigma_X^2 \end{vmatrix} \neq 0.$$

Then, the two-dimensional semi-linear credibility model becomes non-split credibility model whatever the split point is.

Proof: Because the determinant of matrix (2.47) is not equal to zero, $\hat{\alpha_Y}$ and $\hat{\alpha_X}$ have a unique solution. And if $\delta_{XY}/\tau_{XY} = \sigma_X^2/\tau_X^2$, then we know that $(\hat{\alpha_Y}, \hat{\alpha_X}) = (0, \tau_X^2/(n\tau_X^2 + \sigma_X^2))$ is a solution of Equation (2.44) because

$$\hat{\alpha_X} = \frac{\tau_{XY}}{n\tau_{XY} + \delta_{XY}} = \frac{1}{n + \frac{\delta_{XY}}{\tau_{XY}}} = \frac{1}{n + \frac{\sigma_X^2}{\tau_X^2}} = \frac{\tau_X^2}{n\tau_X^2 + \sigma_X^2}.$$

Hence, we know that the solution $(\hat{\alpha_Y}, \hat{\alpha_X}) = (0, \tau_X^2/(n\tau_X^2 + \sigma_X^2))$ is the unique solution of Equation (2.44). And notice that this solution of two-dimensional semi-linear credibility model is also the solution of non-split credibility model, so the two-dimensional semi-linear credibility model becomes non-split credibility model whatever the split point is in this situation. ∎

Under Theorem 2.2.3, the split way is invalid and we do not need to split the losses if the conditions of Theorem 2.2.3 are met.

Now we give our method to get the optimal split point. Firstly, we sort the losses, X, from small to large. And then, we take some percentiles as our split point according to the accuracy we need. The more precise, the more points. From our practical experience, the K value generally is chosen a relatively large value, so we can take more points at larger percentiles to determine the best K value.

For example, we could choose the 0th, 25th, 50th, 75th, 80th, 85th, 90th, 95th and 100th percentiles as our split points. Then, we use our formulas to calculate the parameters and the optimal α's. Next, we calculate the values of MMSE given by different split points. We choose the minimum one and choose the corresponding K value, which is the percentile and would be different in number's value according to different data, as the future optimal percentile in future data from the same source(or the same group).

In this method, we can not guarantee that the K value we choose is the exact optimal K value for each group of data, but the worst result is only the minimum of MMSE of non-split credibility and MMSE of semi-linear credibility.

2.2.3 Nonparametric estimation

In this section, we consider unbiased estimation of our parameters, μ_X, τ_X^2, σ_X^2, τ_{XY} and δ_{XY}. Let us use the simple Bühlmann model as an example.

Suppose that $n_j = n > 1$ for all j and we have policyholders with number of $m > 1$. That is, for policyholder i, we have the loss vector

$$\boldsymbol{X}_i = (X_{i1}, X_{i2}, \cdots, X_{in})^T, \quad i = 1, 2, \cdots, m.$$

Furthermore, conditional on $\Theta_i = \theta_i$, X_{ij} has mean

$$\mu_X(\theta_i) = E(X_{ij}|\Theta_i = \theta_i),$$

and variane

$$V_X(\theta_i) = Var(X_{ij}|\Theta_i = \theta_i),$$

as well as X_{is} and X_{it} are independent if $s \neq t$ conditional on $\Theta_i = \theta_i$. In the meantime, X_{ij} and X_{st} are also independent if $i \neq s$ because of the independence of different policyholders. So we have

$$\begin{aligned} \bar{X}_i &= \frac{1}{n}\sum_{j=1}^{n} X_{ij}, \\ \bar{X} &= \frac{1}{nm}\sum_{i=1}^{m}\sum_{j=1}^{n} X_{ij}. \end{aligned}$$

An unbiased estimator of μ_X is

$$\hat{\mu_X} = \bar{X},$$

because

$$\begin{aligned} E(\hat{\mu_X}) = E(\bar{X}) &= \frac{1}{nm}\sum_{i=1}^{m}\sum_{j=1}^{n} E(X_{ij}) \\ &= \frac{1}{nm}\sum_{i=1}^{m}\sum_{j=1}^{n} E(\mu_X(\Theta_i)) \\ &= \frac{1}{nm}\sum_{i=1}^{m}\sum_{j=1}^{n} \mu_X \\ &= \mu_X. \end{aligned}$$

At the same time, an unbiased estimator of the conditional variance of X_{ij} is

$$\widehat{V_X(\Theta_i)} = \frac{1}{n-1}\sum_{j=1}^{n}(X_{ij} - \bar{X}_i)^2.$$

Hence, an unbiased estimator of σ_X^2 is

$$\begin{aligned} \widehat{\sigma_X^2} = E(\widehat{V_X(\Theta_i)}) &= \frac{1}{m}\sum_{i=1}^{m}\widehat{V_X(\Theta_i)} \\ &= \frac{1}{m(n-1)}\sum_{i=1}^{m}\sum_{j=1}^{n}(X_{ij} - \bar{X}_i)^2. \end{aligned}$$

For details, readers are referred to Klugman et al. [8].

We now turn to determine unbiased estimator of τ_X^2. Since

$$\begin{aligned} Var(\bar{X}_i) &= Var\left[E(\bar{X}_i|\Theta_i)\right] + E\left[Var(\bar{X}_i|\Theta_i)\right] \\ &= Var\left[\mu_X(\Theta_i)\right] + E\left[\frac{V_X(\Theta_i)}{n}\right] \\ &= \tau_X^2 + \frac{\sigma_X^2}{n}, \end{aligned}$$

an unbiased estimator of τ_X^2 is

$$\begin{aligned} \widehat{\tau_X^2} &= \widehat{Var(\bar{X}_i)} - \frac{\widehat{\sigma_X^2}}{n} \\ &= \frac{1}{m-1}\sum_{i=1}^{m}(\bar{X}_i - \bar{X})^2 - \frac{\widehat{\sigma_X^2}}{n} \\ &= \frac{1}{m-1}\sum_{i=1}^{m}(\bar{X}_i - \bar{X})^2 - \frac{1}{mn(n-1)}\sum_{i=1}^{m}\sum_{j=1}^{n}(X_{ij} - \bar{X}_i)^2, \end{aligned}$$

where $\widehat{\sigma_X^2}$ is given and

$$\widehat{Var(\bar{X}_i)} = \frac{1}{m-1}\sum_{i=1}^{m}(\bar{X}_i - \bar{X})^2.$$

For μ_Y, τ_Y^2 and σ_Y^2, we firstly get the data Y from the losses X. In our case, $Y = min(X, K)$ given a split point K. Then, their unbiased estimators are

$$\begin{aligned}
\hat{\mu_Y} &= \bar{Y} = \frac{1}{nm}\sum_{i=1}^{m}\sum_{j=1}^{n} Y_{ij}, \\
\widehat{\sigma_Y^2} &= E(\widehat{V_Y(\Theta_i)}) = \frac{1}{m(n-1)}\sum_{i=1}^{m}\sum_{j=1}^{n}(Y_{ij} - \bar{Y}_i)^2, \\
\widehat{\tau_Y^2} &= \widehat{Var(\bar{Y}_i)} - \frac{\widehat{\sigma_Y^2}}{n} \\
&= \frac{1}{m-1}\sum_{i=1}^{m}(\bar{Y}_i - \bar{Y})^2 - \frac{1}{mn(n-1)}\sum_{i=1}^{m}\sum_{j=1}^{n}(Y_{ij} - \bar{Y}_i)^2.
\end{aligned}$$

For estimation of τ_{XY} and δ_{XY}, we firstly consider an unbiased estimator of the covariance. Suppose that $S_1, S_2, \cdots, S_r$ are independent random variables with same mean $\mu_S = E(S_j)$ and $T_1, T_2, \cdots, T_r$ are independent random variables with same mean $\mu_T = E(T_j)$. And we are only talking about paired data where S_i and T_j are only correlated when $i = j$.

Let

$$\begin{aligned}
\bar{S} &= \frac{1}{r}\sum_{j=1}^{r} S_j, \\
\bar{T} &= \frac{1}{r}\sum_{j=1}^{r} T_j.
\end{aligned}$$

Then, consider the statistic $\sum_{j=1}^{r}(S_j - \bar{S})(T_j - \bar{T})$. It can be rewritten as

$$\begin{aligned}
\sum_{j=1}^{r}(S_j - \bar{S})(T_j - \bar{T}) &= \sum_{j=1}^{r}(S_jT_j - S_j\bar{T} - T_j\bar{S} + \bar{S}\bar{T}) \\
&= \sum_{j=1}^{r} S_jT_j - r\bar{S}\bar{T} - r\bar{S}\bar{T} + r\bar{S}\bar{T}) \\
&= \sum_{j=1}^{r} S_jT_j - r\bar{S}\bar{T}.
\end{aligned}$$

Taking expectation of both sides yields

$$\begin{aligned}
E\left[\sum_{j=1}^{r}(S_j-\bar{S})(T_j-\bar{T})\right] &= E\left[\sum_{j=1}^{r}S_jT_j\right]-rE(\bar{S}\bar{T}) \\
&= E\left[\sum_{j=1}^{r}S_jT_j\right]-r\left[Cov(\bar{S},\bar{T})+\mu_S\mu_T\right] \\
&= E\left[\sum_{j=1}^{r}S_jT_j\right]-r\left[Cov(\frac{1}{r}\sum_{j=1}^{r}S_j,\frac{1}{r}\sum_{j=1}^{r}T_j)+\mu_S\mu_T\right] \\
&= E\left[\sum_{j=1}^{r}S_jT_j\right]-r\left[\frac{1}{r^2}Cov(\sum_{j=1}^{r}S_j,\sum_{j=1}^{r}T_j)+\mu_S\mu_T\right] \\
&= E\left[\sum_{j=1}^{r}S_jT_j\right]-\frac{1}{r}\sum_{j=1}^{r}Cov(S_j,T_j)-r\mu_S\mu_T \\
&= E\left[\sum_{j=1}^{r}S_jT_j\right]-Cov(S,T)-r\mu_S\mu_T \\
&= rE(ST)-Cov(S,T)-r\mu_S\mu_T \\
&= r\left[E(ST)-\mu_S\mu_T\right]-Cov(S,T) \\
&= (r-1)Cov(S,T).
\end{aligned}$$

Therefore, an unbiased estimator of the covariance is $\frac{1}{r-1}\sum_{j=1}^{r}(S_j-\bar{S})(T_j-\bar{T})$.
To estimate δ_{XY}, notice that we have an unbiased estimator of $\delta_{XY}(\Theta_i)$ as follows,

$$\widehat{\delta_{XY}(\Theta_i)}=\frac{1}{n-1}\sum_{j=1}^{n}(X_{ij}-\bar{X}_i)(Y_{ij}-\bar{Y}_i).$$

Hence, an unbiased estimator of δ_{XY} is

$$\begin{aligned}
\hat{\delta_{XY}}=E(\widehat{\delta_{XY}(\Theta_i)}) &= \frac{1}{m}\sum_{i=1}^{m}\widehat{\delta_{XY}(\Theta_i)} \\
&= \frac{1}{m(n-1)}\sum_{i=1}^{m}\sum_{j=1}^{n}(X_{ij}-\bar{X}_i)(Y_{ij}-\bar{Y}_i).
\end{aligned}$$

Now we consider the value of $Cov(\bar{X}_i,\bar{Y}_i)$. Recall that $X_{i1},X_{i2},\cdots,X_{in}$ are independent conditional on $\Theta_i=\theta_i$ and so are Ys. So we also know that X_{is} and Y_{it} are independent if $s\neq t$

conditional on $\Theta_i = \theta_i$. Then

$$
\begin{aligned}
Cov(\bar{X}_i, \bar{Y}_i) &= E\left[Cov(\bar{X}_i|\Theta_i, \bar{Y}_i|\Theta_i)\right] + Cov\left[E(\bar{X}_i|\Theta_i), E(\bar{Y}_i|\Theta_i)\right] \\
&= E\left[Cov(\frac{1}{n}\sum_{j=1}^{n} X_{ij}|\Theta_i, \frac{1}{n}\sum_{j=1}^{n} Y_{ij}|\Theta_i)\right] + Cov(\mu_X(\Theta), \mu_Y(\Theta)) \\
&= E\left[\frac{1}{n^2}Cov(\sum_{j=1}^{n} X_{ij}|\Theta_i, \sum_{j=1}^{n} Y_{ij}|\Theta_i)\right] + \tau_{XY} \\
&= E\left[\frac{1}{n^2}\sum_{j=1}^{n} Cov(X_{ij}|\Theta_i, Y_{ij}|\Theta_i)\right] + \tau_{XY} \\
&= E\left[\frac{1}{n}Cov(X|\Theta, Y|\Theta)\right] + \tau_{XY} \\
&= \frac{\delta_{XY}}{n} + \tau_{XY}.
\end{aligned}
$$

We also have an unbiased estimator of $Cov(\bar{X}_i, \bar{Y}_i)$ as follows,

$$
\widehat{Cov(\bar{X}_i, \bar{Y}_i)} = \frac{1}{m-1}\sum_{i=1}^{m}(\bar{X}_i - \bar{X})(\bar{Y}_i - \bar{Y}).
$$

Thus, we get an unbiased estimator of τ_{XY}. That is,

$$
\begin{aligned}
\hat{\tau_{XY}} &= \widehat{Cov(\bar{X}_i, \bar{Y}_i)} - \frac{\hat{\delta_{XY}}}{n} \\
&= \frac{1}{m-1}\sum_{i=1}^{m}(\bar{X}_i - \bar{X})(\bar{Y}_i - \bar{Y}) - \frac{1}{mn(n-1)}\sum_{i=1}^{m}\sum_{j=1}^{n}(X_{ij} - \bar{X}_i)(Y_{ij} - \bar{Y}_i).
\end{aligned}
$$

Now, we give all unbiased estimation of our needed parameters and you can review the whole results of this section in Appendix B.

2.2.4 The estimators of $\mu_p(\theta)$ and $\mu_e(\theta)$

Besides the estimator of $\mu_X(\theta)$, we also concern about the estimators of primary part($\mu_p(\theta)$) and excess part($\mu_e(\theta)$). In reality, the large claims generally are large relative to all of data and we do not want that they occupy so large proportion that it will influence our future income as well as increase risk and instability. In the other words, we hope that the mean of the excess part is very small to reduce the overall impact of that part. Therefore, we could control the influence and risk if we could estimate $\mu_p(\theta)$ and $\mu_e(\theta)$.

We discuss $\mu_e(\theta)$ only and the other is similar. Let Y be $X_p = min(X, K)$ and L be $X_e = X - X_p$. Then $\mu_p(\Theta) = \mu_Y(\Theta)$ and $\mu_e(\Theta) = \mu_L(\Theta)$. And we let the estimator of $\mu_p(\theta)$ be $\widehat{\mu_p(\theta)}$ and the estimator of $\mu_e(\theta)$ be $\widehat{\mu_e(\theta)}$. We change $\hat{\alpha_0}$ to be $\hat{\beta_0}$, $\hat{\alpha_Y}$ to be $\hat{\beta_Y}$ and change $\hat{\alpha_X}$ to be $\hat{\beta_X}$.

We still use Ys and Xs as data to estimate $\mu_e(\theta)$. Because the independence of L is the same as X, the processes are the same as the processes of the estimator of $\mu_X(\theta)$. Therefore, we have

$$\begin{aligned}\widehat{\mu_e(\theta)} = \widehat{\mu_L(\theta)} &= \hat{\beta_0} + \hat{\beta_Y}\sum_{j=1}^{n} Y_j + \hat{\beta_X}\sum_{j=1}^{n} X_j \\ &= \mu_L - n\hat{\beta_Y}\mu_Y - n\hat{\beta_X}\mu_X + n\hat{\beta_Y}\bar{Y} + n\hat{\beta_X}\bar{X} \\ &= \mu_L + n\hat{\beta_Y}(\bar{Y} - \mu_Y) + n\hat{\beta_X}(\bar{X} - \mu_X),\end{aligned}$$

compared with (2.32) as well as where $\hat{\beta_X}$ and $\hat{\beta_Y}$ meet

$$\begin{pmatrix} n\tau_Y^2 + \sigma_Y^2 & n\tau_{XY} + \delta_{XY} \\ n\tau_{XY} + \delta_{XY} & n\tau_X^2 + \sigma_X^2 \end{pmatrix}\begin{pmatrix} \hat{\beta_Y} \\ \hat{\beta_X} \end{pmatrix} = \begin{pmatrix} \tau_{YL} \\ \tau_{XL} \end{pmatrix}, \tag{2.55}$$

compared with (2.44). Because we only want to use the parameters we have used and calculated, we rewrite τ_{YL} and τ_{XL} as

$$\begin{aligned}\tau_{YL} &= Cov(\mu_Y(\Theta), \mu_L(\Theta)) \\ &= Cov(\mu_Y(\Theta), \mu_X(\Theta) - \mu_Y(\Theta)) \\ &= Cov(\mu_Y(\Theta), \mu_X(\Theta)) - Cov(\mu_Y(\Theta), \mu_Y(\Theta)) \\ &= \tau_{XY} - Var[\mu_Y(\Theta)] \\ &= \tau_{XY} - \tau_Y^2, \\ \tau_{XL} &= Cov(\mu_X(\Theta), \mu_L(\Theta)) \\ &= Cov(\mu_X(\Theta), \mu_X(\Theta) - \mu_Y(\Theta)) \\ &= Cov(\mu_X(\Theta), \mu_X(\Theta)) - Cov(\mu_X(\Theta), \mu_Y(\Theta)) \\ &= Var[\mu_X(\Theta)] - \tau_{XY} \\ &= \tau_X^2 - \tau_{XY},\end{aligned}$$

because

$$\begin{aligned}\mu_X(\Theta) &= E[X|\Theta] = E[X_p + X_e|\Theta] = E[Y + L|\Theta] = E[Y|\Theta] + E[L\Theta] \\ &= \mu_Y(\Theta) + \mu_L(\Theta) = \mu_p(\Theta) + \mu_e(\Theta), \\ \mu_X &= \mu_Y + \mu_L = \mu_p + \mu_e.\end{aligned}$$

For $\mu_p(\theta)$, we change $\hat{\alpha_0}$ to be $\hat{\gamma_0}$, $\hat{\alpha_Y}$ to be $\hat{\gamma_Y}$ and change $\hat{\alpha_X}$ to be $\hat{\gamma_X}$. The estimator is

$$\begin{aligned}\widehat{\mu_p(\theta)} = \widehat{\mu_Y(\theta)} &= \hat{\gamma_0} + \hat{\gamma_Y}\sum_{j=1}^{n} Y_j + \hat{\gamma_X}\sum_{j=1}^{n} X_j \\ &= \mu_Y - n\hat{\gamma_Y}\mu_Y - n\hat{\gamma_X}\mu_X + n\hat{\gamma_Y}\bar{Y} + n\hat{\gamma_X}\bar{X} \\ &= \mu_Y + n\hat{\gamma_Y}(\bar{Y} - \mu_Y) + n\hat{\gamma_X}(\bar{X} - \mu_X),\end{aligned}$$

where

$$\begin{pmatrix} n\tau_Y^2 + \sigma_Y^2 & n\tau_{XY} + \delta_{XY} \\ n\tau_{XY} + \delta_{XY} & n\tau_X^2 + \sigma_X^2 \end{pmatrix}\begin{pmatrix} \hat{\gamma_Y} \\ \hat{\gamma_X} \end{pmatrix} = \begin{pmatrix} \tau_Y^2 \\ \tau_{XY} \end{pmatrix}. \tag{2.56}$$

Notice that the right-hand side of Equation (2.44) is the sum of the right-hand side of Equation (2.55) and the right-hand side of Equation (2.56). That is,

$$\widehat{\mu_X(\theta)} = \widehat{\mu_p(\theta)} + \widehat{\mu_e(\theta)},$$

which meets our expectation. Now, we can estimate $\mu_p(\theta)$ and $\mu_e(\theta)$. More elegantly, using matrix notation we can write

$$\begin{pmatrix} \widehat{\mu_X(\theta)} \\ \widehat{\mu_p(\theta)} \\ \widehat{\mu_e(\theta)} \end{pmatrix} = \begin{pmatrix} \mu_X \\ \mu_Y \\ \mu_L \end{pmatrix} + \begin{pmatrix} n\hat{\alpha}_Y & n\hat{\alpha}_X \\ n\hat{\gamma}_Y & n\hat{\gamma}_X \\ n\hat{\beta}_Y & n\hat{\beta}_X \end{pmatrix} \begin{pmatrix} \bar{Y} - \mu_Y \\ \bar{X} - \mu_X \end{pmatrix}, \tag{2.57}$$

where $\mu_Y = \mu_p$ and $\mu_L = \mu_e$.

Chapter 3

Examples

In this chapter, we discuss three examples. The first one is that the losses follow an Exponential distribution conditional on risk parameters. The second one is that the losses follow a Poisson distribution conditional on risk parameters. The final one is that the losses follow a mixture of two Exponential distributions where one distribution is conditional on risk parameters. All of risk parameters, Θ, follow a Gamma distribution.

3.1 Exponential distribution conditional on Θ

We consider an example much like the collective risk model (CRM) in Robbin [11]. Suppose that the losses $X_{i1}, X_{i2}, X_{i3}, \ldots, X_{in}$ conditional on $\Theta_i = \theta_i$ follow an Exponential distribution with the same mean and same variance as well as Θ follows a Gamma distribution with shape parameter, α, and rate parameter, β. Let $f_{X|\Theta}(x)$ be the probability density function(pdf) of the losses conditional on Θ and $f_\Theta(\theta)$ be the pdf of the Θ. Let $F_{X|\Theta}(x)$ be the cumulative distribution function(cdf) of the losses conditional on Θ. Then, we have

$$f_{X|\Theta=\theta}(x) = \theta e^{-\theta x}, \tag{3.1}$$
$$F_{X|\Theta=\theta}(x) = 1 - e^{-\theta x}, \tag{3.2}$$
$$f_\Theta(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}. \tag{3.3}$$

Hence, the mean and variance of losses conditional on Θ as well as the mean and variance of Θ are

$$E(X|\Theta = \theta) = \frac{1}{\theta}, \tag{3.4}$$

$$Var(X|\Theta = \theta) = \frac{1}{\theta^2}, \tag{3.5}$$

$$E(\Theta) = \frac{\alpha}{\beta}, \tag{3.6}$$

$$Var(\Theta) = \frac{\alpha}{\beta^2}. \tag{3.7}$$

Assume that there are $m = 6$ policyholders. Each policyholder's risk parameter is represented by a random variable $\Theta = (\Theta_1 = \theta_1, \Theta_2 = \theta_2, \ldots, \Theta_6 = \theta_6)$, which follows a Gamma distribution with shape parameter, $\alpha = 6$, and rate parameter, $\beta = 50$. For each policyholder, conditional on $\Theta_i = \theta_i$, $n = 45$ past losses are observed. Then, we use R code to randomly generate risk parameters and losses. For details of data, readers are referred to Table 3.1 and for details of R code, readers are referred to Appendix C.

3.1.1 Parametric estimation

By (3.4), $\mu_X(\theta) = E(X|\Theta = \theta) = 1/\theta$. Thus, the unconditional mean of X is

$$\begin{aligned}
\mu_X = E[\mu_X(\Theta)] = E\left(\frac{1}{\Theta}\right) &= \int \frac{1}{\theta} f_\Theta(\theta)\, d\theta, \\
&= \int \frac{1}{\theta}\frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}\, d\theta, \\
&= \int \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-2}e^{-\beta\theta}\, d\theta, \\
&= \int \frac{\beta}{\alpha-1}\frac{\beta^{\alpha-1}}{\Gamma(\alpha-1)}\theta^{\alpha-2}e^{-\beta\theta}\, d\theta, \\
&= \frac{\beta}{\alpha-1}\int \frac{\beta^{\alpha-1}}{\Gamma(\alpha-1)}\theta^{\alpha-2}e^{-\beta\theta}\, d\theta, \\
&= \frac{\beta}{\alpha-1}. \qquad (3.8)
\end{aligned}$$

The unconditional variance of X is

$$\begin{aligned}
\sigma_X^2 = E[Var(X|\Theta)] = E\left(\frac{1}{\Theta^2}\right) &= \int \frac{1}{\theta^2}\frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}\, d\theta, \\
&= \int \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-3}e^{-\beta\theta}\, d\theta, \\
&= \frac{\beta^2}{(\alpha-1)(\alpha-2)}\int \frac{\beta^{\alpha-2}}{\Gamma(\alpha-2)}\theta^{\alpha-3}e^{-\beta\theta}\, d\theta, \\
&= \frac{\beta^2}{(\alpha-1)(\alpha-2)}, \qquad (3.9)
\end{aligned}$$

where $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$.

Hence, we can calculate τ_X^2 as follows,

$$\begin{aligned}
\tau_X^2 = Var[\mu_X(\Theta)] = Var\left(\frac{1}{\Theta}\right) &= E\left(\frac{1}{\Theta^2}\right) - E\left(\frac{1}{\Theta}\right)^2, \\
&= \frac{\beta^2}{(\alpha-1)(\alpha-2)} - \frac{\beta^2}{(\alpha-1)^2}, \\
&= \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}. \qquad (3.10)
\end{aligned}$$

		Losses														
		X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}
Thetas	θ_1	4.53	21.14	50.51	5.26	3.86	1.28	2.35	18.87	10.45	11.82	4.76	5.64	3.21	11.65	13.16
	θ_2	13.53	15.83	1.78	7.90	22.76	5.52	1.48	14.16	5.10	21.12	0.52	1.23	12.38	3.06	2.56
	θ_3	3.34	2.54	2.75	5.49	24.52	4.40	64.47	3.04	35.21	11.83	0.41	0.44	28.31	4.43	2.01
	θ_4	2.13	11.63	3.44	2.03	2.88	3.70	4.26	5.60	4.62	5.76	4.56	2.04	3.38	4.16	4.31
	θ_5	10.29	4.12	2.70	26.45	2.93	1.26	7.45	13.68	16.31	7.96	2.06	4.94	4.25	3.85	4.64
	θ_6	14.86	5.21	0.05	7.13	0.72	1.28	14.50	8.93	7.24	24.37	2.94	8.49	8.35	7.05	2.20
		X_{16}	X_{17}	X_{18}	X_{19}	X_{20}	X_{21}	X_{22}	X_{23}	X_{24}	X_{25}	X_{26}	X_{27}	X_{28}	X_{29}	X_{30}
Thetas	θ_1	26.11	3.85	5.74	25.92	27.18	22.13	18.49	17.27	34.22	8.32	3.04	40.83	24.75	5.29	21.25
	θ_2	4.94	20.70	5.12	17.89	0.04	32.38	11.64	11.51	67.39	7.20	3.55	4.85	2.20	3.67	6.30
	θ_3	0.54	29.82	30.88	6.27	32.47	11.14	3.87	0.81	7.86	1.27	5.79	1.79	9.52	24.13	6.08
	θ_4	18.60	5.12	0.73	3.29	0.81	3.41	1.02	1.78	10.40	1.76	0.77	10.53	3.64	5.69	16.22
	θ_5	12.31	19.20	0.65	2.92	5.09	16.06	2.90	1.98	10.09	10.89	5.78	7.92	7.27	1.03	14.10
	θ_6	10.34	20.91	0.85	56.40	6.41	15.89	15.53	39.81	8.84	1.88	0.57	0.21	17.69	14.24	7.95
		X_{31}	X_{32}	X_{33}	X_{34}	X_{35}	X_{36}	X_{37}	X_{38}	X_{39}	X_{40}	X_{41}	X_{42}	X_{43}	X_{44}	X_{45}
Thetas	θ_1	2.72	3.64	5.95	3.28	19.70	7.60	9.90	2.83	10.60	2.00	1.02	14.88	23.92	17.92	1.29
	θ_2	5.39	17.01	0.80	14.77	2.80	30.22	1.34	12.86	7.01	1.87	9.79	10.52	0.34	3.82	2.05
	θ_3	2.31	27.64	8.24	39.81	2.81	27.04	3.44	7.28	0.89	6.20	4.18	20.60	1.54	0.40	36.15
	θ_4	14.13	0.67	0.74	0.73	5.76	7.10	1.47	10.25	2.24	1.29	0.04	5.60	3.87	0.36	0.26
	θ_5	7.57	3.78	4.23	1.27	11.66	15.63	1.29	3.67	3.57	13.74	0.22	0.81	1.29	4.83	6.04
	θ_6	16.09	0.15	15.02	6.46	17.15	6.53	11.05	2.16	0.76	58.36	8.00	1.27	0.63	14.78	12.40

Remark: $\Theta = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6) = (0.08385895, 0.08750489, 0.07935393, 0.17249887, 0.15398474, 0.09140351)$.

Table 3.1: A set of data from R code in example1 (only two decimals)

Then, we calculate the parameters of $Y = min(X, K)$ given by split point, K. The mean of Y conditional on θ is

$$\begin{aligned}
\mu_Y(\theta) &= \int_0^K x f_{X|\Theta}(x)\, dx + K(1 - F_{X|\Theta}(K)), \\
&= \int_0^K x\theta e^{-\theta x}\, dx + K e^{-\theta K}, \\
&= \int_0^K -x\, d\left(e^{-\theta x}\right) + K e^{-\theta K}, \\
&= \left(-x e^{-\theta x}\right)\Big|_0^K - \int_0^K -e^{-\theta x}\, dx + K e^{-\theta K}, \\
&= -K e^{-\theta K} + K e^{-\theta K} - \left(\frac{1}{\theta} e^{-\theta x}\right)\Bigg|_0^K, \\
&= \frac{1}{\theta} - \frac{1}{\theta} e^{-\theta K}. \qquad (3.11)
\end{aligned}$$

Now, the unconditional mean of Y is

$$\begin{aligned}
\mu_Y = E[\mu_Y(\Theta)] = E\left(\frac{1}{\Theta} - \frac{1}{\Theta} e^{-\Theta K}\right) &= E\left(\frac{1}{\Theta}\right) - E\left(\frac{1}{\Theta} e^{-\Theta K}\right), \\
&= \frac{\beta}{\alpha - 1} - \int \frac{1}{\theta} e^{-\theta K} f_\Theta(\theta)\, d\theta, \\
&= \frac{\beta}{\alpha - 1} - \int \frac{1}{\theta} e^{-\theta K} \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha - 1} e^{-\beta\theta}\, d\theta, \\
&= \frac{\beta}{\alpha - 1} - \int \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha - 2} e^{-(\beta + K)\theta}\, d\theta, \\
&= \frac{\beta}{\alpha - 1} - \frac{\beta^\alpha}{(\beta + K)^{\alpha - 1}(\alpha - 1)} \int \frac{(\beta + K)^{\alpha - 1}}{\Gamma(\alpha - 1)} \theta^{\alpha - 2} e^{-(\beta + K)\theta}\, d\theta, \\
&= \frac{\beta}{\alpha - 1} - \frac{\beta^\alpha}{(\beta + K)^{\alpha - 1}(\alpha - 1)}, \\
&= \frac{\beta}{\alpha - 1}\left[1 - \left(\frac{\beta}{\beta + K}\right)^{\alpha - 1}\right]. \qquad (3.12)
\end{aligned}$$

According to the calculation of (3.8), (3.9) and (3.12), we can easily calculate the following equations.

$$E\left(\frac{1}{\Theta} e^{-2\Theta K}\right) = \frac{\beta}{\alpha - 1}\left(\frac{\beta}{\beta + 2K}\right)^{\alpha - 1}, \qquad (3.13)$$

$$E\left(\frac{1}{\Theta^2} e^{-\Theta K}\right) = \frac{\beta^2}{(\alpha - 1)(\alpha - 2)}\left(\frac{\beta}{\beta + K}\right)^{\alpha - 2}, \qquad (3.14)$$

$$E\left(\frac{1}{\Theta^2} e^{-2\Theta K}\right) = \frac{\beta^2}{(\alpha - 1)(\alpha - 2)}\left(\frac{\beta}{\beta + 2K}\right)^{\alpha - 2}. \qquad (3.15)$$

Then, the value of τ_Y^2 is

$$
\begin{aligned}
\tau_Y^2 &= Var[\mu_Y(\Theta)], \\
&= E[\mu_Y(\Theta)^2] - E[\mu_Y(\Theta)]^2, \\
&= E\left[\left(\frac{1}{\Theta} - \frac{1}{\Theta}e^{-\Theta K}\right)^2\right] - \left\{\frac{\beta}{\alpha-1}\left[1 - \left(\frac{\beta}{\beta+K}\right)^{\alpha-1}\right]\right\}^2, \\
&= E\left(\frac{1}{\Theta^2} + \frac{1}{\Theta^2}e^{-2\Theta K} - 2\frac{1}{\Theta^2}e^{-\Theta K}\right) - \frac{\beta^2}{(\alpha-1)^2}\left[1 - \left(\frac{\beta}{\beta+K}\right)^{\alpha-1}\right]^2, \\
&= E\left(\frac{1}{\Theta^2} + \frac{1}{\Theta^2}e^{-2\Theta K} - 2\frac{1}{\Theta^2}e^{-\Theta K}\right) - \frac{\beta^2}{(\alpha-1)^2}\left[1 - \left(\frac{\beta}{\beta+K}\right)^{\alpha-1}\right]^2, \\
&= \frac{\beta^2}{(\alpha-1)(\alpha-2)} + \frac{\beta^2}{(\alpha-1)(\alpha-2)}\left(\frac{\beta}{\beta+2K}\right)^{\alpha-2} - \\
&\quad \frac{2\beta^2}{(\alpha-1)(\alpha-2)}\left(\frac{\beta}{\beta+K}\right)^{\alpha-2} - \frac{\beta^2}{(\alpha-1)^2}\left[1 - \left(\frac{\beta}{\beta+K}\right)^{\alpha-1}\right]^2, \\
&= \frac{\beta^2}{(\alpha-1)(\alpha-2)}\left[1 + \left(\frac{\beta}{\beta+2K}\right)^{\alpha-2} - 2\left(\frac{\beta}{\beta+K}\right)^{\alpha-2}\right] - \\
&\quad \frac{\beta^2}{(\alpha-1)^2}\left[1 - \left(\frac{\beta}{\beta+K}\right)^{\alpha-1}\right]^2.
\end{aligned} \tag{3.16}
$$

In addition, τ_{XY} is

$$
\begin{aligned}
\tau_{XY} &= Cov(\mu_X(\Theta), \mu_Y(\Theta)), \\
&= E[\mu_X(\Theta)\mu_Y(\Theta)] - \mu_X\mu_Y, \\
&= E\left(\frac{1}{\Theta^2} - \frac{1}{\Theta^2}e^{-\Theta K}\right) - \frac{\beta^2}{(\alpha-1)^2}\left[1 - \left(\frac{\beta}{\beta+K}\right)^{\alpha-1}\right], \\
&= \frac{\beta^2}{(\alpha-1)(\alpha-2)} - \frac{\beta^2}{(\alpha-1)(\alpha-2)}\left(\frac{\beta}{\beta+K}\right)^{\alpha-2} - \\
&\quad \frac{\beta^2}{(\alpha-1)^2}\left[1 - \left(\frac{\beta}{\beta+K}\right)^{\alpha-1}\right], \\
&= \frac{\beta^2}{(\alpha-1)(\alpha-2)}\left[1 - \left(\frac{\beta}{\beta+K}\right)^{\alpha-2}\right] - \\
&\quad \frac{\beta^2}{(\alpha-1)^2}\left[1 - \left(\frac{\beta}{\beta+K}\right)^{\alpha-1}\right].
\end{aligned} \tag{3.17}
$$

To calculate σ_Y^2 and δ_{XY}, we need to calculate the following equations.

$$
\begin{aligned}
E[Y^2|\Theta = \theta] &= \int_0^K x^2\theta e^{-\theta x}\,dx + K^2e^{-\theta K}, \\
&= \left(-x^2e^{-\theta x}\right)\Big|_0^K - \int_0^K -2xe^{-\theta x}\,dx + K^2e^{-\theta K}, \\
&= -K^2e^{-\theta K} + 2\int_0^K xe^{-\theta x}\,dx + K^2e^{-\theta K}, \\
&= 2\int_0^K xe^{-\theta x}\,dx, \\
&= 2\int_0^K -\frac{x}{\theta}d\left(e^{-\theta x}\right), \\
&= \left(-\frac{2x}{\theta}e^{-\theta x}\right)\Bigg|_0^K - \frac{2}{\theta}\int_0^K -e^{-\theta x}\,dx, \\
&= -\frac{2K}{\theta}e^{-\theta K} - \frac{2}{\theta}\left(\frac{1}{\theta}e^{-\theta x}\right)\Bigg|_0^K, \\
&= -\frac{2K}{\theta}e^{-\theta K} + \frac{2}{\theta^2} - \frac{2}{\theta^2}e^{-\theta K}. \qquad (3.18)
\end{aligned}
$$

In addition,

$$
\begin{aligned}
E[XY|\Theta = \theta] &= \int_0^K x^2\theta e^{-\theta x}\,dx + \int_K^{+\infty} Kx\theta e^{-\theta x}\,dx, \\
&= -\frac{2K}{\theta}e^{-\theta K} + \frac{2}{\theta^2} - \frac{2}{\theta^2}e^{-\theta K} - K^2e^{-\theta K} + \\
&\quad K\left(\int_0^{+\infty} x\theta e^{-\theta x}\,dx - \int_0^K x\theta e^{-\theta x}\,dx\right), \\
&= -\frac{2K}{\theta}e^{-\theta K} + \frac{2}{\theta^2} - \frac{2}{\theta^2}e^{-\theta K} - K^2e^{-\theta K} + \\
&\quad K\left(\frac{1}{\theta} - \frac{1}{\theta} + \frac{1}{\theta}e^{-\theta K} + Ke^{-\theta K}\right), \\
&= -\frac{K}{\theta}e^{-\theta K} + \frac{2}{\theta^2} - \frac{2}{\theta^2}e^{-\theta K}. \qquad (3.19)
\end{aligned}
$$

Therefore, the value of σ_Y^2 is

$$
\begin{aligned}
\sigma_Y^2 = E[Var(Y|\Theta)] &= E\left[E(Y^2|\Theta) - E(Y|\Theta)^2\right], \\
&= E\left[-\frac{2K}{\Theta}e^{-\Theta K} + \frac{2}{\Theta^2} - \frac{2}{\Theta^2}e^{-\Theta K} - \left(\frac{1}{\Theta} - \frac{1}{\Theta}e^{-\Theta K}\right)^2\right], \\
&= E\left[-\frac{2K}{\Theta}e^{-\Theta K} + \frac{2}{\Theta^2} - \frac{2}{\Theta^2}e^{-\Theta K} - \frac{1}{\Theta^2} - \frac{1}{\Theta^2}e^{-2\Theta K} + \frac{2}{\Theta^2}e^{-\Theta K}\right], \\
&= E\left[-\frac{2K}{\Theta}e^{-\Theta K} + \frac{1}{\Theta^2} - \frac{1}{\Theta^2}e^{-2\Theta K}\right], \\
&= \frac{\beta^2}{(\alpha-1)(\alpha-2)} - 2K\frac{\beta}{\alpha-1}\left(\frac{\beta}{\beta+K}\right)^{\alpha-1} - \\
&\quad \frac{\beta^2}{(\alpha-1)(\alpha-2)}\left(\frac{\beta}{\beta+2K}\right)^{\alpha-2}, \\
&= \frac{\beta^2}{(\alpha-1)(\alpha-2)}\left[1 - \left(\frac{\beta}{\beta+2K}\right)^{\alpha-2}\right] - \frac{2K\beta}{\alpha-1}\left(\frac{\beta}{\beta+K}\right)^{\alpha-1}.
\end{aligned} \tag{3.20}
$$

In addition, δ_{XY} is

$$
\begin{aligned}
\delta_{XY} = E[Cov(X|\Theta, Y|\Theta)] &= E\left[E(XY|\Theta) - \mu_X(\Theta)\mu_Y(\Theta)\right], \\
&= E\left[-\frac{K}{\Theta}e^{-\Theta K} + \frac{2}{\Theta^2} - \frac{2}{\Theta^2}e^{-\Theta K} - \frac{1}{\Theta^2} + \frac{1}{\Theta^2}e^{-\Theta K}\right], \\
&= E\left[-\frac{K}{\Theta}e^{-\Theta K} + \frac{1}{\Theta^2} - \frac{1}{\Theta^2}e^{-\Theta K}\right], \\
&= \frac{\beta^2}{(\alpha-1)(\alpha-2)} - K\frac{\beta}{\alpha-1}\left(\frac{\beta}{\beta+K}\right)^{\alpha-1} - \\
&\quad \frac{\beta^2}{(\alpha-1)(\alpha-2)}\left(\frac{\beta}{\beta+K}\right)^{\alpha-2}, \\
&= \frac{\beta^2}{(\alpha-1)(\alpha-2)}\left[1 - \left(\frac{\beta}{\beta+K}\right)^{\alpha-2}\right] - \\
&\quad \frac{K\beta}{\alpha-1}\left(\frac{\beta}{\beta+K}\right)^{\alpha-1}.
\end{aligned} \tag{3.21}
$$

Remarks According to the formulas above, the shape parameter, α, of Gamma distribution must be bigger than 2. That is, $\alpha > 2$.

If $\alpha \to +\infty$ and $\beta = M$ is a constant, then all of the parameters mentioned above go to zero whatever the split point K is. In this situation, the MMSE of two-dimensional semi-linear credibility model goes to zero, which is the same as the MMSE of non-split credibility model. Furthermore, the credibility premium goes to zero too in this situation.

However, it is worth noting that when the determinant of matrix (2.47) is not equal to zero, this example is a very special case because regardless of the split point, the MMSE of our credibility model is always the same as the MMSE of non-split credibility model, which means that we don't need to split the losses. Because this example under the situation meets the conditions of Theorem 2.2.3. That is,

$$\frac{\sigma_X^2}{\tau_X^2} = \frac{\dfrac{\beta^2}{(\alpha-1)(\alpha-2)}}{\dfrac{\beta^2}{(\alpha-1)^2(\alpha-2)}} = \alpha - 1.$$

Now we turn to calculate δ_{XY}/τ_{XY} and we have

$$\begin{aligned}
\frac{\delta_{XY}}{\tau_{XY}} &= \frac{\dfrac{\beta^2}{(\alpha-1)(\alpha-2)}\left[1-\left(\dfrac{\beta}{\beta+K}\right)^{\alpha-2}\right] - \dfrac{K\beta}{\alpha-1}\left(\dfrac{\beta}{\beta+K}\right)^{\alpha-1}}{\dfrac{\beta^2}{(\alpha-1)(\alpha-2)}\left[1-\left(\dfrac{\beta}{\beta+K}\right)^{\alpha-2}\right] - \dfrac{\beta^2}{(\alpha-1)^2}\left[1-\left(\dfrac{\beta}{\beta+K}\right)^{\alpha-1}\right]}, \\
&= \frac{\dfrac{\beta^2}{(\alpha-1)(\alpha-2)}\left[1-\left(\dfrac{\beta}{\beta+K}\right)^{\alpha-2}\right] - \dfrac{K\beta^2}{(\alpha-1)(\beta+K)}\left(\dfrac{\beta}{\beta+K}\right)^{\alpha-2}}{\dfrac{\beta^2}{(\alpha-1)(\alpha-2)}\left[1-\left(\dfrac{\beta}{\beta+K}\right)^{\alpha-2}\right] - \dfrac{\beta^2}{(\alpha-1)^2}\left[1-\left(\dfrac{\beta}{\beta+K}\right)\left(\dfrac{\beta}{\beta+K}\right)^{\alpha-2}\right]},
\end{aligned}$$

which can be simplified to

$$\frac{\delta_{XY}}{\tau_{XY}} = \frac{\dfrac{1}{\alpha-2}\left[1-\left(\dfrac{\beta}{\beta+K}\right)^{\alpha-2}\right] - \dfrac{K}{\beta+K}\left(\dfrac{\beta}{\beta+K}\right)^{\alpha-2}}{\dfrac{1}{\alpha-2}\left[1-\left(\dfrac{\beta}{\beta+K}\right)^{\alpha-2}\right] - \dfrac{1}{\alpha-1}\left[1-\left(\dfrac{\beta}{\beta+K}\right)\left(\dfrac{\beta}{\beta+K}\right)^{\alpha-2}\right]}.$$

Let D be the denominator of the fraction above. We have

$$\begin{aligned}
D &= \frac{1}{\alpha-2}\left[1-\left(\frac{\beta}{\beta+K}\right)^{\alpha-2}\right] - \frac{1}{\alpha-1}\left[1-\left(\frac{\beta}{\beta+K}\right)\left(\frac{\beta}{\beta+K}\right)^{\alpha-2}\right] + \\
&\quad \frac{1}{\alpha-1}\left(\frac{\beta}{\beta+K}\right)^{\alpha-2} - \frac{1}{\alpha-1}\left(\frac{\beta}{\beta+K}\right)^{\alpha-2}, \\
&= \frac{1}{(\alpha-1)(\alpha-2)}\left[1-\left(\frac{\beta}{\beta+K}\right)^{\alpha-2}\right] + \frac{1}{\alpha-1}\left(\frac{\beta}{\beta+K}\right)\left(\frac{\beta}{\beta+K}\right)^{\alpha-2} - \frac{1}{\alpha-1}\left(\frac{\beta}{\beta+K}\right)^{\alpha-2}, \\
&= \frac{1}{(\alpha-1)(\alpha-2)}\left[1-\left(\frac{\beta}{\beta+K}\right)^{\alpha-2}\right] - \frac{1}{\alpha-1}\left(\frac{K}{\beta+K}\right)\left(\frac{\beta}{\beta+K}\right)^{\alpha-2}.
\end{aligned}$$

Now, we easily get that

$$\frac{\delta_{XY}}{\tau_{XY}} = \alpha - 1.$$

Therefore, the conditions of Theorem 2.2.3 are met and we do not need to split the losses in this example under the situation.

Under the assumptions and data of Table 3.1, we can use R code to calculate the parameters we need by parametric estimation. For details of R code, readers are referred to Appendix C. Suppose that $K = 5$ and let $\hat{Z}$ be the optimal credibility of non-split credibility model. Let the MMSE of non-split credibility be $\hat{Q}_{nsp}$, the MMSE of semi-linear credibility be $\hat{Q}_{tsl}$ and the MMSE of two-dimensional semi-linear credibility be $\hat{Q}_{tdsl}$. For details of their formulas, readers are referred to Equations (2.52), (2.53) and (2.54). Then we have Table 3.2.

Parameter	Value		Parameter	Value		Parameter	Value
μ_X	10		σ_X^2	125		τ_X^2	25
μ_Y	3.790787		σ_Y^2	2.626232		τ_Y^2	0.1582076
δ_{XY}	8.577252		τ_{XY}	1.71545		$\hat{Z}$	0.9
$\hat{\alpha_Y}$	-2.49E-16		$\hat{\alpha_X}$	0.02		$\hat{\alpha_0}$	1
$\hat{Q}_{nsp}$	2.5		$\hat{Q}_{tsl}$	11.41182		$\hat{Q}_{tdsl}$	2.5

Table 3.2: The value of parameters by parametric estimation calculated using R in example1 (when $K = 5$)

From Table 3.2, we know that $\hat{Q}_{nsp} = \hat{Q}_{tdsl}$, $\hat{Z} = n\hat{\alpha_X}$ and $\hat{\alpha_0} = (1 - \hat{Z})\mu_X$. These above mean that the two-dimensional semi-linear credibility model is the same as non-split credibility model in this time. You will see all of the MMSEs of each model for all of the split points, K, which are from 0 to the maximum value of the losses, X, in Figure 3.1. Notice that the line of MMSE of the two-dimensional semi-linear credibility model coincides with the line of MMSE of non-split credibility model in Figure 3.1.

3.1.2 Nonparametric estimation

The results of nonparametric estimation are different from the results of parametric estimation because our parameters are estimated from data and our method is sensitive to estimation error. In addition, because this example does not have a fixed optimal split point, the optimal split point given by nonparametric estimation has a large variance. For details of the formulas of nonparametric estimation, readers are referred to Appendix B.

Again, under the assumptions and data of Table 3.1, we use R code to calculate the parameters we need by nonparametric estimation. For details of R code, readers are referred to Appendix C. Suppose that $K = 5$ and let $\hat{Z}$ be the optimal credibility of non-split credibility model. Let the MMSE of non-split credibility be $\hat{Q}_{nsp}$, the MMSE of semi-linear credibility be $\hat{Q}_{tsl}$ and the MMSE of two-dimensional semi-linear credibility be $\hat{Q}_{tdsl}$. For details of their formulas, readers are referred to Equations (2.52), (2.53) and (2.54). Then we have Table 3.3.

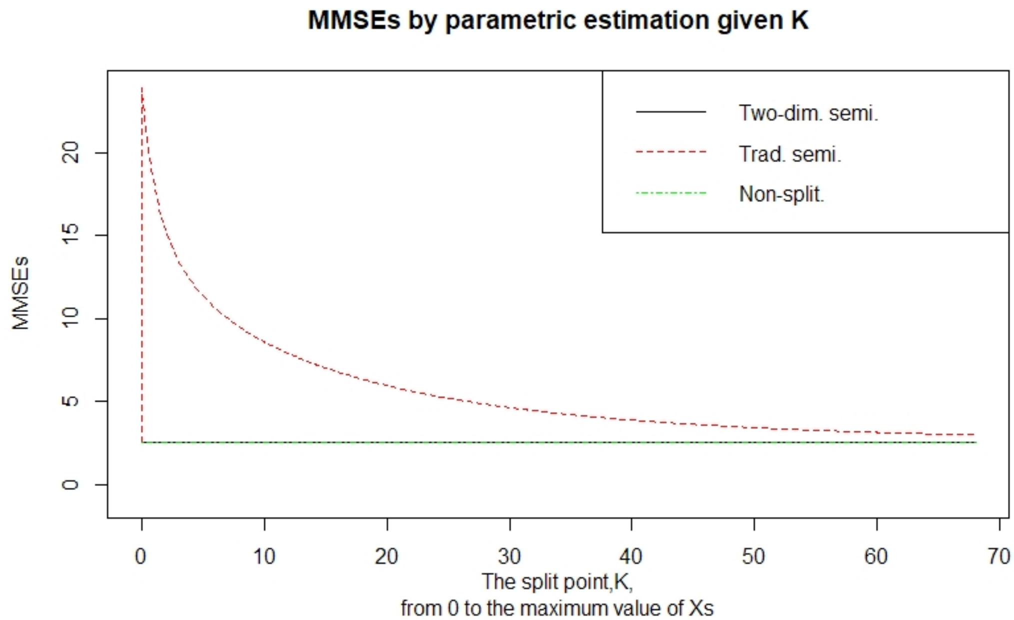

Figure 3.1: MMSEs by parametric estimation given K in example1

We can see the estimation error in Table 3.3 and they are large in the variance and covariance. You will see all of the MMSEs of each model by nonparametric estimation for all of the split points, K, which are from 0 to the maximum value of the losses, X, in Figure 3.2.

Thus, we get that the optimal split point is the 97th percentile (38.2). For details, readers are referred to Appendix C.

Accoring to the method in Section 2.2.2, we choose the 0th, 25th, 50th, 75th, 80th, 85th, 90th, 95th and 100th percentiles as our split points. Then, we use R code to calculate the parameters and the optimal α's. We calculate the values of MMSE given by different Ks. You will see all of the MMSEs of each model by nonparametric estimation for specific split points in Figure 3.3.

We choose the minimum one and choose the corresponding K value as the future optimal percentile in future data from the same source(or the same group). In this case, the optimal split point is the 95th percentile.

3.2 Poisson distribution conditional on Θ

Suppose that the losses $X_{i1}, X_{i2}, X_{i3}, \ldots, X_{in}$ conditional on $\Theta_i = \theta_i$ follow a Poisson distribution. Θ follows a Gamma distribution with shape parameter, α, and rate parameter, β. Let $f_{X|\Theta}(k)$ be the probability mass function(pmf) of the losses conditional on Θ and $f_\Theta(\theta)$ be the pdf of the

Parameter	Value		Parameter	Value		Parameter	Value
μ_X	9.622611		σ_X^2	115.353		τ_X^2	8.223654
μ_Y	3.759468		σ_Y^2	2.673828		τ_Y^2	0.08377466
δ_{XY}	9.389524		τ_{XY}	0.7828881		$\hat{Z}$	0.7623633
$\hat{\alpha_Y}$	0.01151496		$\hat{\alpha_X}$	0.01588295		$\hat{\alpha_0}$	0.7969588
$\hat{Q}_{nsp}$	1.954242		$\hat{Q}_{tsl}$	3.943323		$\hat{Q}_{tdsl}$	1.940266

Table 3.3: The value of parameters by nonparametric estimation calculated using R in example1 (when $K = 5$)

Θ. Then, we have

$$f_{X|\Theta=\theta}(k) = \frac{\theta^k e^{-\theta}}{k!}, \tag{3.22}$$

$$f_\Theta(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}. \tag{3.23}$$

The mean and variance of X conditional on Θ are

$$E(X|\Theta=\theta) = \theta, \tag{3.24}$$

$$Var(X|\Theta=\theta) = \theta. \tag{3.25}$$

Hence, we easily get that

$$\mu_X = E(\Theta) = \frac{\alpha}{\beta}, \tag{3.26}$$

$$\sigma_X^2 = E(\Theta) = \frac{\alpha}{\beta}, \tag{3.27}$$

$$\tau_X^2 = Var(\Theta) = \frac{\alpha}{\beta^2}. \tag{3.28}$$

Let $Y = min(X, K)$, then the mean of Y conditional on Θ is

$$\begin{aligned}\mu_Y(\Theta=\theta) &= \sum_{k=0}^{K-1} k\frac{\theta^k e^{-\theta}}{k!} + \sum_{k=K}^{+\infty} K\frac{\theta^k e^{-\theta}}{k!},\\ &= \sum_{k=0}^{K-1}(k-K)\frac{\theta^k e^{-\theta}}{k!} + K. \end{aligned} \tag{3.29}$$

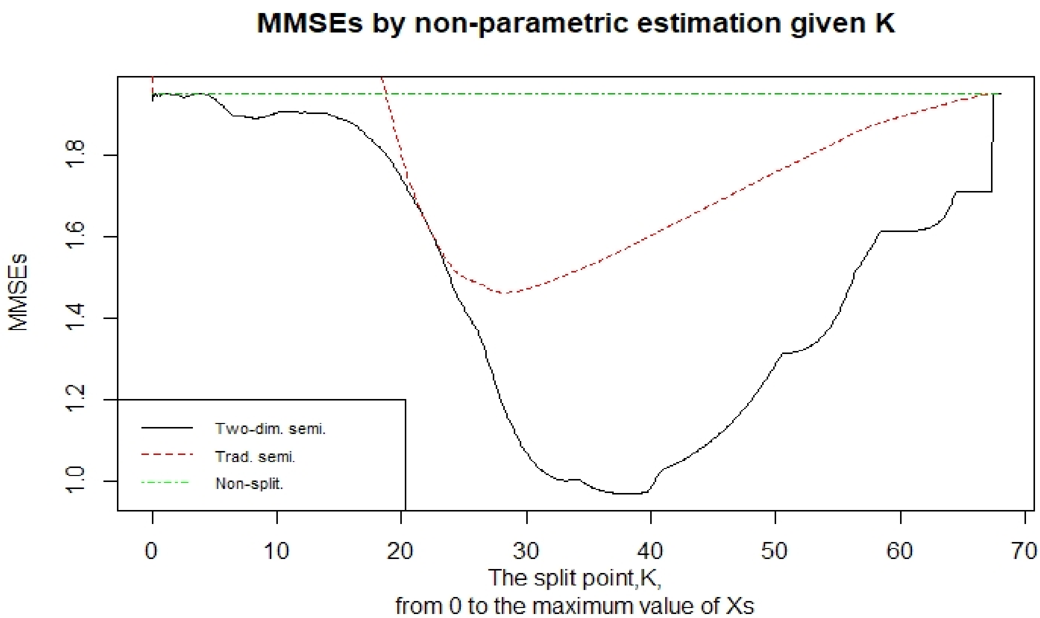

Figure 3.2: MMSEs by nonparametric estimation given K in example1

We calculate $E(\Theta^k e^{-\Theta})$ as follows,

$$
\begin{aligned}
E(\Theta^k e^{-\Theta}) &= \int \theta^k e^{-\theta} \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta}\, d\theta, \\
&= \int \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha+k-1} e^{-(\beta+1)\theta}\, d\theta, \\
&= \int \frac{\Gamma(\alpha+k)}{\Gamma(\alpha)} \frac{\beta^\alpha}{(\beta+1)^{\alpha+k}} \frac{(\beta+1)^{\alpha+k}}{\Gamma(\alpha+k)} \theta^{\alpha+k-1} e^{-(\beta+1)\theta}\, d\theta, \\
&= \frac{(\alpha+k-1)!}{(\alpha-1)!} \frac{\beta^\alpha}{(\beta+1)^{\alpha+k}}.
\end{aligned} \tag{3.30}
$$

Hence, the unconditional mean of Y is

$$
\begin{aligned}
\mu_Y &= \sum_{k=0}^{K-1} (k-K) \frac{E\left(\Theta^k e^{-\Theta}\right)}{k!} + K, \\
&= \sum_{k=0}^{K-1} \frac{k-K}{k!} \frac{(\alpha+k-1)!}{(\alpha-1)!} \frac{\beta^\alpha}{(\beta+1)^{\alpha+k}} + K.
\end{aligned} \tag{3.31}
$$

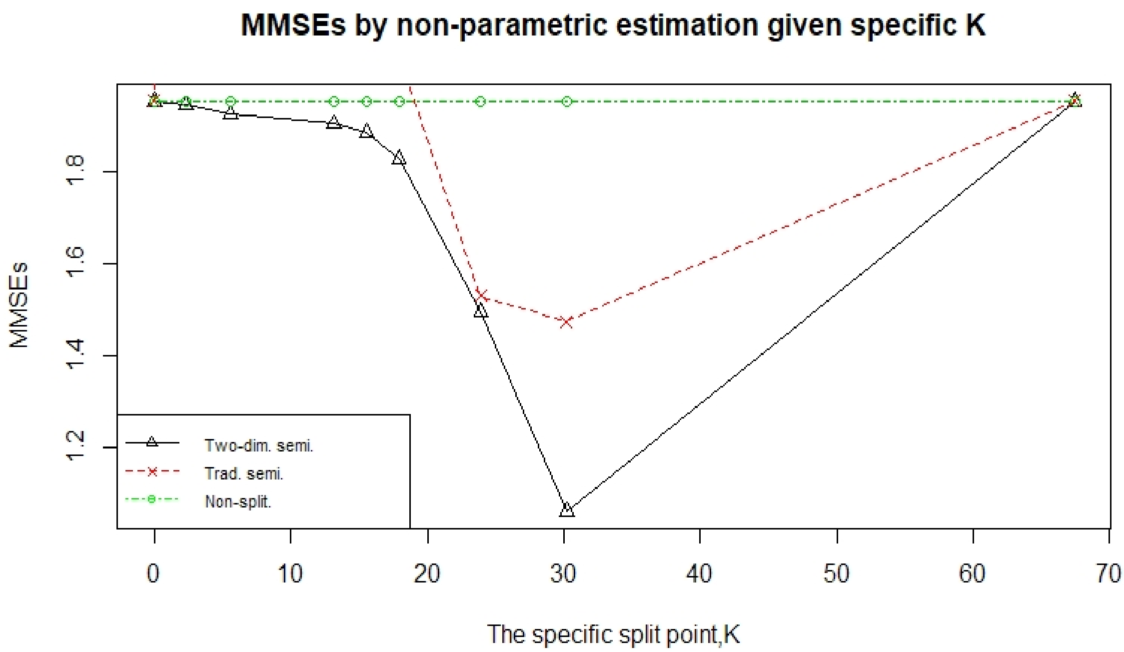

Figure 3.3: MMSEs by nonparametric estimation given specific K in example1

To calculate τ_{XY} and δ_{XY}, we need to calculate the following equations.

$$
\begin{aligned}
E(XY|\Theta=\theta) &= \sum_{k=0}^{K-1} k^2 \frac{\theta^k e^{-\theta}}{k!} + \sum_{k=K}^{+\infty} Kk \frac{\theta^k e^{-\theta}}{k!}, \\
&= \sum_{k=0}^{K-1} k^2 \frac{\theta^k e^{-\theta}}{k!} + K\left[\theta - \sum_{k=0}^{K-1} k \frac{\theta^k e^{-\theta}}{k!}\right], \\
&= \sum_{k=0}^{K-1} k(k-K) \frac{\theta^k e^{-\theta}}{k!} + K\theta, \qquad (3.32) \\
E(\mu_X(\Theta)\mu_Y(\Theta)) &= \sum_{k=0}^{K-1} (k-K) \frac{E\left(\Theta^{k+1} e^{-\Theta}\right)}{k!} + KE(\Theta), \\
&= \sum_{k=0}^{K-1} \frac{k-K}{k!} \frac{(\alpha+k)!}{(\alpha-1)!} \frac{\beta^\alpha}{(\beta+1)^{\alpha+k+1}} + K\frac{\alpha}{\beta}. \qquad (3.33)
\end{aligned}
$$

Hence, we have

$$\begin{aligned}
\tau_{XY} &= E(\mu_X(\Theta)\mu_Y(\Theta)) - \mu_X\mu_Y, \\
&= \sum_{k=0}^{K-1} \frac{k-K}{k!}\frac{(\alpha+k)!}{(\alpha-1)!}\frac{\beta^\alpha}{(\beta+1)^{\alpha+k+1}} + K\frac{\alpha}{\beta} - \\
&\quad \sum_{k=0}^{K-1} \frac{k-K}{k!}\frac{\alpha}{\beta}\frac{(\alpha+k-1)!}{(\alpha-1)!}\frac{\beta^\alpha}{(\beta+1)^{\alpha+k}} - K\frac{\alpha}{\beta}, \\
&= \sum_{k=0}^{K-1} \frac{k-K}{k!}\frac{(\alpha+k-1)!}{(\alpha-1)!}\frac{\beta^\alpha}{(\beta+1)^{\alpha+k}}\left[\frac{\alpha+k}{\beta+1} - \frac{\alpha}{\beta}\right], \qquad (3.34) \\
\delta_{XY} &= E[E(XY|\Theta)] - E(\mu_X(\Theta)\mu_Y(\Theta)), \\
&= \sum_{k=0}^{K-1} k(k-K)\frac{E\left(\Theta^k e^{-\Theta}\right)}{k!} + KE(\Theta) - \\
&\quad \sum_{k=0}^{K-1} \frac{k-K}{k!}\frac{(\alpha+k)!}{(\alpha-1)!}\frac{\beta^\alpha}{(\beta+1)^{\alpha+k+1}} - K\frac{\alpha}{\beta}, \\
&= \sum_{k=0}^{K-1} \frac{k-K}{k!}\frac{(\alpha+k-1)!}{(\alpha-1)!}\frac{\beta^\alpha}{(\beta+1)^{\alpha+k}}\left[k - \frac{\alpha+k}{\beta+1}\right]. \qquad (3.35)
\end{aligned}$$

Notice that

$$\begin{aligned}
\beta\left[\frac{\alpha+k}{\beta+1} - \frac{\alpha}{\beta}\right] &= \frac{k\beta-\alpha}{\beta+1}, \\
\left[k - \frac{\alpha+k}{\beta+1}\right] &= \frac{k\beta-\alpha}{\beta+1}.
\end{aligned}$$

Hence, we easily get that

$$\frac{\delta_{XY}}{\tau_{XY}} = \beta = \frac{\sigma_X^2}{\tau_X^2}.$$

From now on, when the determinant of matrix (2.47) is not equal to zero, the conditions of Theorem 2.2.3 are met and we do not need to split the losses in this example under the situation. This results are the same as the first example's.

Assume that there are $m = 6$ policyholders. Each policyholder's risk parameter is represented by a random variable Θ, which follows a Gamma distribution with shape parameter, $\alpha = 50$, and rate parameter, $\beta = 5$. For each policyholder, conditional on $\Theta_i = \theta_i$, $n = 45$ past losses are observed. Then, we use R code to randomly generate risk parameters and losses. For details of data, readers are referred to Table 3.4 and for details of R code, readers are referred to Appendix C.

You will see all of the MMSEs of each model by nonparametric estimation for all of the split points, K, which are from 0 to the maximum value of the losses, X, in Figure 3.4.

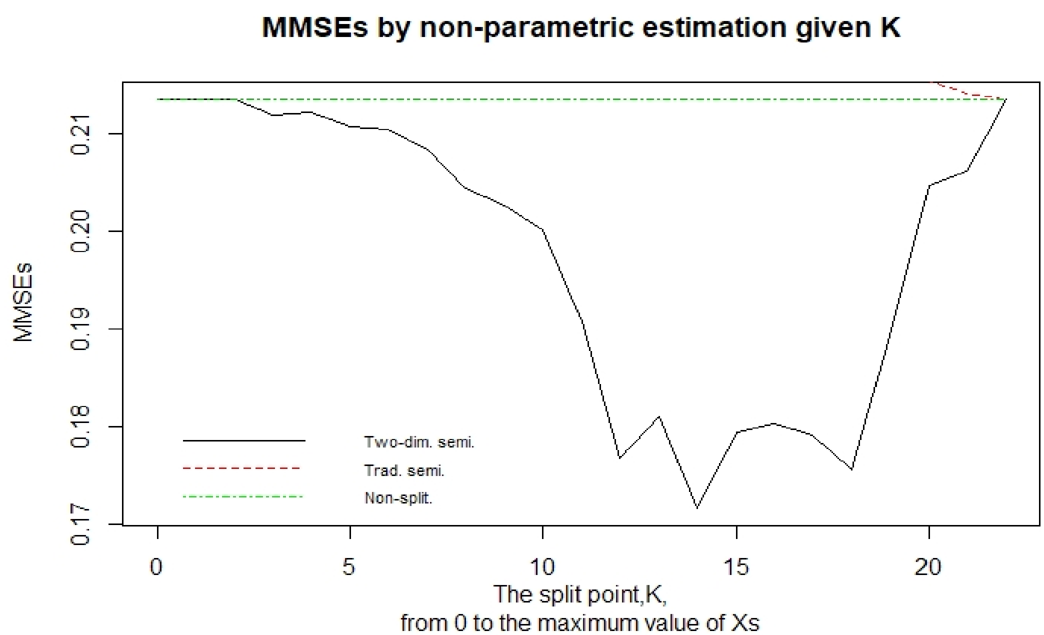

Figure 3.4: MMSEs by nonparametric estimation given K in example2

		Losses														
		X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}
Thetas	θ_1	7	12	5	11	12	9	6	9	4	8	7	11	14	13	8
	θ_2	10	7	14	10	16	10	15	6	18	17	7	9	9	14	12
	θ_3	11	13	11	14	19	20	9	13	12	19	13	13	17	18	14
	θ_4	6	12	6	13	10	9	9	14	11	12	8	9	11	6	6
	θ_5	8	11	5	7	12	5	8	3	5	8	14	12	5	12	12
	θ_6	5	14	8	15	17	9	6	11	5	8	8	14	7	9	5
		X_{16}	X_{17}	X_{18}	X_{19}	X_{20}	X_{21}	X_{22}	X_{23}	X_{24}	X_{25}	X_{26}	X_{27}	X_{28}	X_{29}	X_{30}
Thetas	θ_1	8	9	10	10	8	10	11	8	10	7	8	9	12	10	8
	θ_2	11	13	8	10	14	13	11	10	7	8	8	7	11	12	16
	θ_3	15	15	13	18	16	14	13	20	21	17	17	19	19	12	10
	θ_4	15	11	14	9	9	8	9	10	9	6	12	10	7	14	14
	θ_5	8	6	12	13	4	8	9	2	18	16	7	13	13	11	6
	θ_6	6	10	8	16	10	8	12	6	9	8	9	10	11	10	13
		X_{31}	X_{32}	X_{33}	X_{34}	X_{35}	X_{36}	X_{37}	X_{38}	X_{39}	X_{40}	X_{41}	X_{42}	X_{43}	X_{44}	X_{45}
Thetas	θ_1	14	7	5	10	8	10	7	8	8	9	9	14	10	10	9
	θ_2	4	10	10	10	10	15	11	9	15	7	20	12	15	20	9
	θ_3	9	15	22	13	16	11	10	12	13	15	16	8	16	13	13
	θ_4	14	5	8	11	7	7	13	8	6	12	16	13	8	11	10
	θ_5	14	7	8	14	9	8	11	13	5	14	9	10	10	12	9
	θ_6	5	10	9	9	15	7	11	8	10	8	10	11	10	7	9

Remark: $\Theta = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6) = (9.356824, 11.129673, 14.282631, 9.581132, 9.390054, 9.060537)$.

Table 3.4: A set of data from R code in example2

Thus, we get that the optimal split point is the 78th percentile. For details, readers are referred to Appendix C.

Accoring to the method in Section 2.2.2, we choose the 0th, 25th, 50th, 75th, 80th, 85th, 90th, 95th and 100th percentiles as our split points. Then, we use R code to calculate the parameters and the optimal α's. We calculate the values of MMSE given by different Ks. You will see all of the MMSEs of each model by nonparametric estimation for specific split points in Figure 3.5.

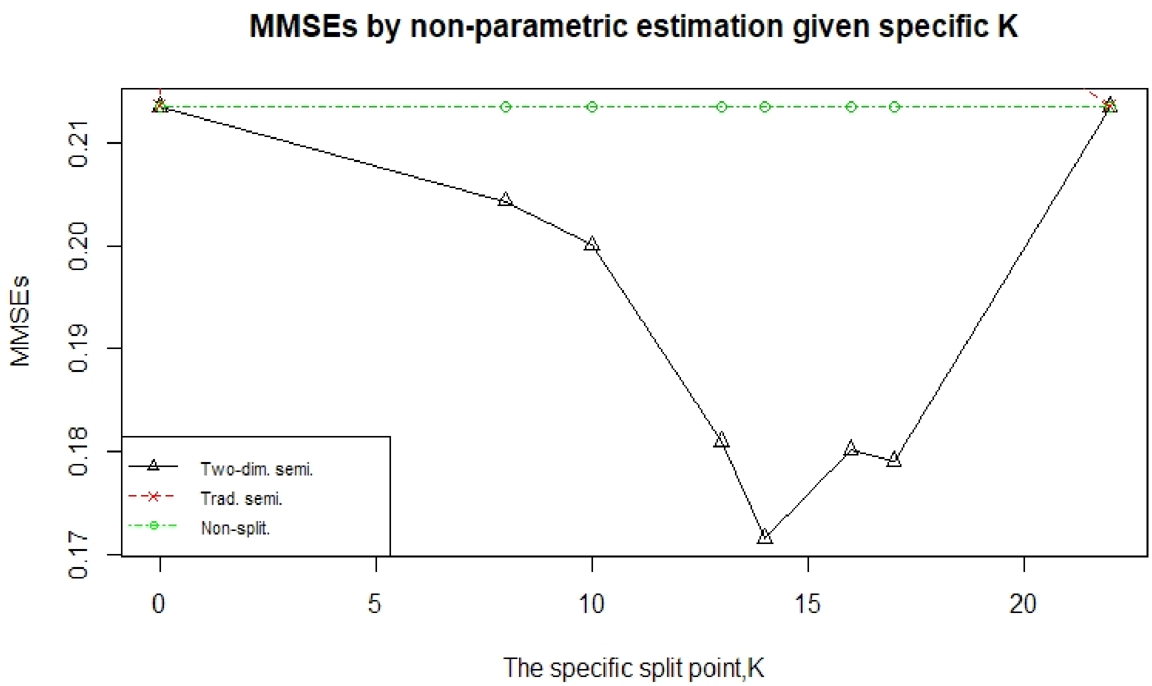

Figure 3.5: MMSEs by nonparametric estimation given specific K in example2

We choose the minimum one and choose the corresponding K value as the future optimal percentile in future data from the same source(or the same group). In this case, the optimal split point is the 80th percentile.

3.3 Mixture of two Exponential distributions conditional on Θ

Suppose that the losses $X_{i1}, X_{i2}, X_{i3}, \ldots, X_{in}$ follow a mixture of two Exponential distribution where one distribution has mean $1/\Theta_i = 1/\theta_i$ with weight ω and the other has mean $1/\lambda$ with weight $1-\omega$. Θ is a Gamma distribution with shape parameter, α, and rate parameter, β.

Let $f_{X|\Theta}(x)$ be the probability density function(pdf) of the losses conditional on Θ and $f_\Theta(\theta)$ be the pdf of the Θ. Let $F_{X|\Theta}(x)$ be the cumulative distribution function(cdf) of the losses conditional on Θ. Then, we have

$$
\begin{aligned}
f_{X|\Theta=\theta}(x) &= \omega\theta e^{-\theta x} + (1-\omega)\lambda e^{-\lambda x}, && (3.36)\\
F_{X|\Theta=\theta}(x) &= \omega(1-e^{-\theta x}) + (1-\omega)(1-e^{-\lambda x}) = 1 - \omega e^{-\theta x} - (1-\omega)e^{-\lambda x}, && (3.37)\\
f_\Theta(\theta) &= \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}. && (3.38)
\end{aligned}
$$

Hence, the mean and variance of losses conditional on Θ are

$$
\begin{aligned}
E(X|\Theta) &= \omega\frac{1}{\Theta} + (1-\omega)\frac{1}{\lambda}, && (3.39)\\
Var(X|\Theta) &= E(X^2|\Theta) - E(X|\Theta)^2, \\
&= \omega\frac{2}{\Theta^2} + (1-\omega)\frac{2}{\lambda^2} - \left[\omega\frac{1}{\Theta} + (1-\omega)\frac{1}{\lambda}\right]^2, \\
&= (2\omega-\omega^2)\frac{1}{\Theta^2} - \frac{2\omega(1-\omega)}{\lambda}\frac{1}{\Theta} + \frac{1-\omega^2}{\lambda^2}. && (3.40)
\end{aligned}
$$

Assume that there are $m = 6$ and $\lambda = 0.1$ and $\omega = 0.75$. Each policyholder's risk parameter is represented by a random variable $\Theta = (\Theta_1 = \theta_1, \Theta_2 = \theta_2, \ldots, \Theta_6 = \theta_6)$, which follows a Gamma distribution with shape parameter, $\alpha = 6$, and rate parameter, $\beta = 25$. For each policyholder, conditional on $\Theta_i = \theta_i$, $n = 45$ past losses are observed.

Then, we use R code to randomly generate risk parameters and losses. For details of data, readers are referred to Table 3.5 and for details of R code, readers are referred to Appendix C.

3.3.1 Parametric estimation

Because this example is an extension of Example1, the derivation of the following formulas can be referred to Example1's. The unconditional mean of X is

$$
\begin{aligned}
\mu_X = E[\mu_X(\Theta)] &= E\left[\omega\frac{1}{\Theta} + (1-\omega)\frac{1}{\lambda}\right], \\
&= \omega\frac{\beta}{\alpha-1} + (1-\omega)\frac{1}{\lambda}. && (3.41)
\end{aligned}
$$

The unconditional variance of X is

$$
\begin{aligned}
\sigma_X^2 = E[Var(X|\Theta)] &= E\left[(2\omega-\omega^2)\frac{1}{\Theta^2} - \frac{2\omega(1-\omega)}{\lambda}\frac{1}{\Theta} + \frac{1-\omega^2}{\lambda^2}\right], \\
&= (2\omega-\omega^2)\frac{\beta^2}{(\alpha-1)(\alpha-2)} - \frac{2\omega(1-\omega)}{\lambda}\frac{\beta}{\alpha-1} + \frac{1-\omega^2}{\lambda^2}. && (3.42)
\end{aligned}
$$

		Losses														
		X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}
Thetas	θ_1	1.50	2.01	0.30	0.66	0.55	0.65	1.07	2.36	1.57	3.29	4.69	5.69	0.51	6.37	9.25
	θ_2	7.11	4.46	3.93	11.18	8.90	1.03	0.51	7.94	8.06	1.24	1.77	34.62	1.97	17.23	7.00
	θ_3	0.02	2.39	3.24	1.75	4.25	0.49	1.73	0.65	0.47	0.68	1.70	5.55	3.74	10.18	0.07
	θ_4	0.42	1.39	0.62	1.82	1.28	0.65	1.24	0.33	1.81	1.72	5.24	2.99	26.48	14.83	3.97
	θ_5	9.80	10.09	6.80	1.37	11.43	2.52	18.84	9.41	3.37	0.78	1.39	3.60	11.37	0.41	6.75
	θ_6	0.80	10.06	11.42	6.20	0.79	6.32	2.68	12.43	7.27	0.80	0.31	1.27	11.41	4.24	4.02
		X_{16}	X_{17}	X_{18}	X_{19}	X_{20}	X_{21}	X_{22}	X_{23}	X_{24}	X_{25}	X_{26}	X_{27}	X_{28}	X_{29}	X_{30}
Thetas	θ_1	2.05	2.41	6.01	3.73	28.20	33.05	4.60	0.22	5.95	0.60	7.17	0.07	1.76	0.45	0.30
	θ_2	8.02	9.73	7.87	4.59	5.10	3.72	3.55	8.48	3.73	3.57	1.37	12.03	0.41	1.93	15.64
	θ_3	1.06	1.77	10.77	0.31	5.68	0.82	18.40	0.58	0.90	2.45	1.25	1.45	4.22	22.99	2.43
	θ_4	1.27	1.69	0.87	1.21	0.89	24.05	1.98	8.88	3.77	0.39	2.02	1.21	0.25	0.45	1.49
	θ_5	10.30	3.03	12.74	0.36	4.73	6.47	1.06	4.90	0.77	8.96	7.34	0.80	0.52	6.39	11.39
	θ_6	26.12	10.48	11.60	24.44	6.30	19.49	8.90	4.33	5.13	1.24	17.12	0.22	0.08	0.60	1.80
		X_{31}	X_{32}	X_{33}	X_{34}	X_{35}	X_{36}	X_{37}	X_{38}	X_{39}	X_{40}	X_{41}	X_{42}	X_{43}	X_{44}	X_{45}
Thetas	θ_1	3.67	0.09	0.42	6.95	0.09	2.55	13.37	11.49	2.22	2.86	0.18	1.43	2.82	1.91	3.94
	θ_2	2.67	1.43	13.74	0.47	0.44	0.89	16.40	7.60	12.08	1.79	3.11	4.16	0.53	15.78	0.29
	θ_3	10.12	5.54	11.86	14.99	10.55	5.59	1.46	6.75	6.33	1.16	3.51	2.19	0.38	22.73	2.15
	θ_4	1.93	13.44	2.72	1.12	0.24	0.32	1.14	1.66	2.36	1.28	2.08	4.27	0.86	2.95	16.80
	θ_5	4.25	7.04	1.25	1.45	12.87	4.97	0.03	22.78	3.91	8.53	0.02	0.79	1.40	7.65	3.37
	θ_6	0.18	0.10	29.59	11.46	10.25	6.46	8.72	0.62	14.99	2.68	1.30	2.36	7.26	2.89	6.27

Remark: $\Theta = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6) = (0.4623506, 0.1851715, 0.2718307, 0.4881058, 0.1714109, 0.1638119)$.

Table 3.5: A set of data from R code in example3 (only two decimals)

Hence, we can calculate τ_X^2 as follows,

$$\begin{aligned}
\tau_X^2 = Var[\mu_X(\Theta)] &= Var\left[\omega\frac{1}{\Theta} + (1-\omega)\frac{1}{\lambda}\right], \\
&= Var\left[\omega\frac{1}{\Theta}\right], \\
&= \omega^2\frac{\beta^2}{(\alpha-1)^2(\alpha-2)}. \qquad (3.43)
\end{aligned}$$

Then, we calculate the parameters of $Y = min(X, K)$ given by split point, K. The mean of Y conditional on θ is

$$\mu_Y(\theta) = \omega\left(\frac{1}{\theta} - \frac{1}{\theta}e^{-\theta K}\right) + (1-\omega)\left(\frac{1}{\lambda} - \frac{1}{\lambda}e^{-\lambda K}\right). \qquad (3.44)$$

Now, the unconditional mean of Y is

$$\mu_Y = E[\mu_Y(\Theta)] = \omega\left\{\frac{\beta}{\alpha-1}\left[1 - \left(\frac{\beta}{\beta+K}\right)^{\alpha-1}\right]\right\} + (1-\omega)\left(\frac{1}{\lambda} - \frac{1}{\lambda}e^{-\lambda K}\right). \qquad (3.45)$$

Then, the value of τ_Y^2 is

$$\begin{aligned}
\tau_Y^2 &= Var[\mu_Y(\Theta)], \\
&= Var\left[\omega\left(\frac{1}{\Theta} - \frac{1}{\Theta}e^{-\Theta K}\right) + (1-\omega)\left(\frac{1}{\lambda} - \frac{1}{\lambda}e^{-\lambda K}\right)\right], \\
&= Var\left[\omega\left(\frac{1}{\Theta} - \frac{1}{\Theta}e^{-\Theta K}\right)\right], \\
&= \omega^2\left\{\frac{\beta^2}{(\alpha-1)(\alpha-2)}\left[1 + \left(\frac{\beta}{\beta+2K}\right)^{\alpha-2} - 2\left(\frac{\beta}{\beta+K}\right)^{\alpha-2}\right] - \right. \\
&\quad \left. \frac{\beta^2}{(\alpha-1)^2}\left[1 - \left(\frac{\beta}{\beta+K}\right)^{\alpha-1}\right]^2\right\}. \qquad (3.46)
\end{aligned}$$

In addition, τ_{XY} is

$$\begin{aligned}
\tau_{XY} &= Cov(\mu_X(\Theta), \mu_Y(\Theta)), \\
&= E[\mu_X(\Theta)\mu_Y(\Theta)] - \mu_X\mu_Y, \\
&= E\left[\omega^2\left(\frac{1}{\Theta^2} - \frac{1}{\Theta^2}e^{-\Theta K}\right)\right] - \omega^2\frac{\beta^2}{(\alpha-1)^2}\left[1 - \left(\frac{\beta}{\beta+K}\right)^{\alpha-1}\right], \\
&= \omega^2\left\{\frac{\beta^2}{(\alpha-1)(\alpha-2)}\left[1 - \left(\frac{\beta}{\beta+K}\right)^{\alpha-2}\right] - \frac{\beta^2}{(\alpha-1)^2}\left[1 - \left(\frac{\beta}{\beta+K}\right)^{\alpha-1}\right]\right\}. \qquad (3.47)
\end{aligned}$$

To calculate σ_Y^2 and δ_{XY}, we firstly calculate the following equations.

$$E[Y^2|\Theta=\theta] = \omega\left(-\frac{2K}{\theta}e^{-\theta K} + \frac{2}{\theta^2} - \frac{2}{\theta^2}e^{-\theta K}\right) + (1-\omega)\left(-\frac{2K}{\lambda}e^{-\lambda K} + \frac{2}{\lambda^2} - \frac{2}{\lambda^2}e^{-\lambda K}\right). \quad (3.48)$$

In addition,

$$E[XY|\Theta=\theta] = \omega\left(-\frac{K}{\theta}e^{-\theta K} + \frac{2}{\theta^2} - \frac{2}{\theta^2}e^{-\theta K}\right) + (1-\omega)\left(-\frac{K}{\lambda}e^{-\lambda K} + \frac{2}{\lambda^2} - \frac{2}{\lambda^2}e^{-\lambda K}\right). \quad (3.49)$$

Therefore, the value of σ_Y^2 is

$$\begin{aligned}
\sigma_Y^2 &= E[Var(Y|\Theta)] = E\left[E(Y^2|\Theta) - E(Y|\Theta)^2\right],\\
&= E\Bigg[\omega\left(-\frac{2K}{\Theta}e^{-\Theta K} + \frac{2-\omega}{\Theta^2} + \frac{2\omega-2}{\Theta^2}e^{-\Theta K} - \frac{\omega}{\Theta^2}e^{-2\Theta K}\right) + (1-\omega)\bigg(-\frac{2K}{\lambda}e^{-\lambda K} + \frac{1+\omega}{\lambda^2}\\
&\quad -\frac{2\omega}{\lambda^2}e^{-\lambda K} - \frac{1-\omega}{\lambda^2}e^{-2\lambda K}\bigg) - 2\omega(1-\omega)\left(\frac{1}{\lambda} - \frac{1}{\lambda}e^{-\lambda K}\right)\left(\frac{1}{\Theta} - \frac{1}{\Theta}e^{-\Theta K}\right)\Bigg],\\
&= \frac{-2K\omega\beta}{\alpha-1}\left(\frac{\beta}{\beta+K}\right)^{\alpha-1} + \frac{(2-\omega)\omega\beta^2}{(\alpha-1)(\alpha-2)} + \frac{2\omega(\omega-1)\beta^2}{(\alpha-1)(\alpha-2)}\left(\frac{\beta}{\beta+K}\right)^{\alpha-2} -\\
&\quad \frac{\omega^2\beta^2}{(\alpha-1)(\alpha-2)}\left(\frac{\beta}{\beta+2K}\right)^{\alpha-2} + (1-\omega)\left(-\frac{2K}{\lambda}e^{-\lambda K} + \frac{1+\omega}{\lambda^2} - \frac{2\omega}{\lambda^2}e^{-\lambda K} - \frac{1-\omega}{\lambda^2}e^{-2\lambda K}\right) -\\
&\quad 2\omega(1-\omega)\left(\frac{1}{\lambda} - \frac{1}{\lambda}e^{-\lambda K}\right)\frac{\beta}{\alpha-1}\left[1 - \left(\frac{\beta}{\beta+K}\right)^{\alpha-1}\right]. \qquad (3.50)
\end{aligned}$$

In addition, δ_{XY} is

$$\begin{aligned}
\delta_{XY} &= E[Cov(X|\Theta, Y|\Theta)] = E\left[E(XY|\Theta) - \mu_X(\Theta)\mu_Y(\Theta)\right],\\
&= E\Bigg[\omega\left(-\frac{K}{\Theta}e^{-\Theta K} + \frac{2-\omega}{\Theta^2} - \frac{2-\omega}{\Theta^2}e^{-\Theta K}\right) +\\
&\quad (1-\omega)\left(-\frac{K}{\lambda}e^{-\lambda K} + \frac{1+\omega}{\lambda^2} - \frac{1+\omega}{\lambda^2}e^{-\lambda K}\right) -\\
&\quad \omega(1-\omega)\left(\frac{1}{\lambda} - \frac{1}{\lambda}e^{-\lambda K}\right)\frac{1}{\Theta} - \frac{\omega(1-\omega)}{\lambda}\left(\frac{1}{\Theta} - \frac{1}{\Theta}e^{-\Theta K}\right)\Bigg],\\
&= \frac{-K\omega\beta}{\alpha-1}\left(\frac{\beta}{\beta+K}\right)^{\alpha-1} + \frac{(2-\omega)\omega\beta^2}{(\alpha-1)(\alpha-2)} - \frac{\omega(2-\omega)\beta^2}{(\alpha-1)(\alpha-2)}\left(\frac{\beta}{\beta+K}\right)^{\alpha-2} +\\
&\quad (1-\omega)\left(-\frac{K}{\lambda}e^{-\lambda K} + \frac{1+\omega}{\lambda^2} - \frac{1+\omega}{\lambda^2}e^{-\lambda K}\right) -\\
&\quad \omega(1-\omega)\left(\frac{1}{\lambda} - \frac{1}{\lambda}e^{-\lambda K}\right)\frac{\beta}{\alpha-1} - \frac{\omega(1-\omega)}{\lambda}\frac{\beta}{\alpha-1}\left[1 - \left(\frac{\beta}{\beta+K}\right)^{\alpha-1}\right]. \qquad (3.51)
\end{aligned}$$

Under the assumptions and data of Table 3.5, we can use R code to calculate the parameters we need by parametric estimation. For details of R code, readers are referred to Appendix C. You will see all of the MMSEs of each model for all of the split points, K, which are from 0 to the maximum value of the losses, X, in Figure 3.6.

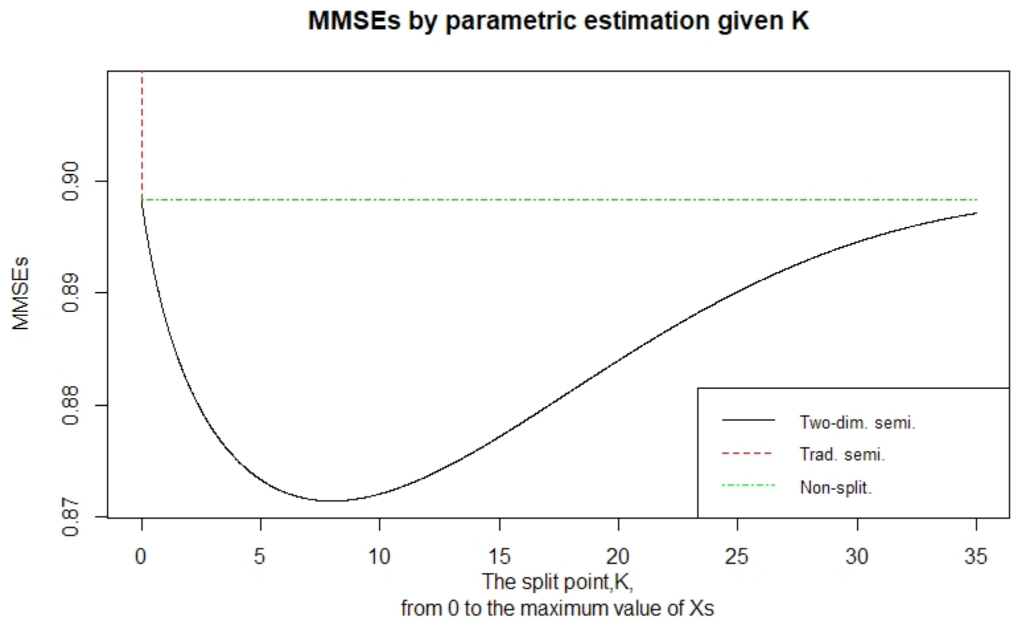

Figure 3.6: MMSEs by parametric estimation given K in example3

From Figure 3.6, we get that the optimal split point is the 77th percentile (7.99) under parametric estimation. For details, readers are referred to Appendix C.

Suppose that $K = 7.99$ and let $\hat{Z}$ be the optimal credibility of non-split credibility model. Let the MMSE of non-split credibility be $\hat{Q}_{nsp}$, the MMSE of semi-linear credibility be $\hat{Q}_{tsl}$ and the MMSE of two-dimensional semi-linear credibility be $\hat{Q}_{tdsl}$. For details of their formulas, readers are referred to Equations (2.52), (2.53) and (2.54). Then we have Table 3.6.

Parameter	Value		Parameter	Value		Parameter	Value
μ_X	6.25		σ_X^2	54.29687		τ_X^2	3.515625
μ_Y	4.188381		σ_Y^2	7.933374		τ_Y^2	0.5068244
δ_{XY}	15.04467		τ_{XY}	1.233023		$\hat{Z}$	0.7444853
$\hat{\alpha_Y}$	0.009024379		$\hat{\alpha_X}$	0.013548843		$\hat{\alpha_0}$	0.7384988
$\hat{Q}_{nsp}$	0.8982939		$\hat{Q}_{tsl}$	1.290039		$\hat{Q}_{tdsl}$	0.8714286

Table 3.6: The value of parameters by parametric estimation calculated using R in example 3 (when $K = 7.99$)

From Table 3.6, we know that $\hat{Q}_{tdsl} \leq \hat{Q}_{nsp}$, which means that our model improves non-split credibility model. In this time, we have that $\widehat{Z_Y} = n\widehat{\alpha_Y} = 0.406097$ and $\widehat{Z_X} = n\widehat{\alpha_X} = 0.6096979$, so the credibility permium under our model is that $\widehat{\mu_X(\theta)} = \mu_X + \widehat{Z_Y}(\bar{Y} - \mu_Y) + \widehat{Z_X}(\bar{X} - \mu_X) = 5.548993$.

3.3.2 Nonparametric estimation

The results of nonparametric estimation are different from the results of parametric estimation because our parameters are estimated from data and our method is sensitive to estimation error. For details of the formulas of nonparametric estimation, readers are referred to Appendix B.

Again, under the assumptions and data of Table 3.1, we use R code to calculate the parameters we need by nonparametric estimation. For details of R code, readers are referred to Appendix C. You will see all of the MMSEs of each model by nonparametric estimation for all of the split points, K, which are from 0 to the maximum value of the losses, X, in Figure 3.7.

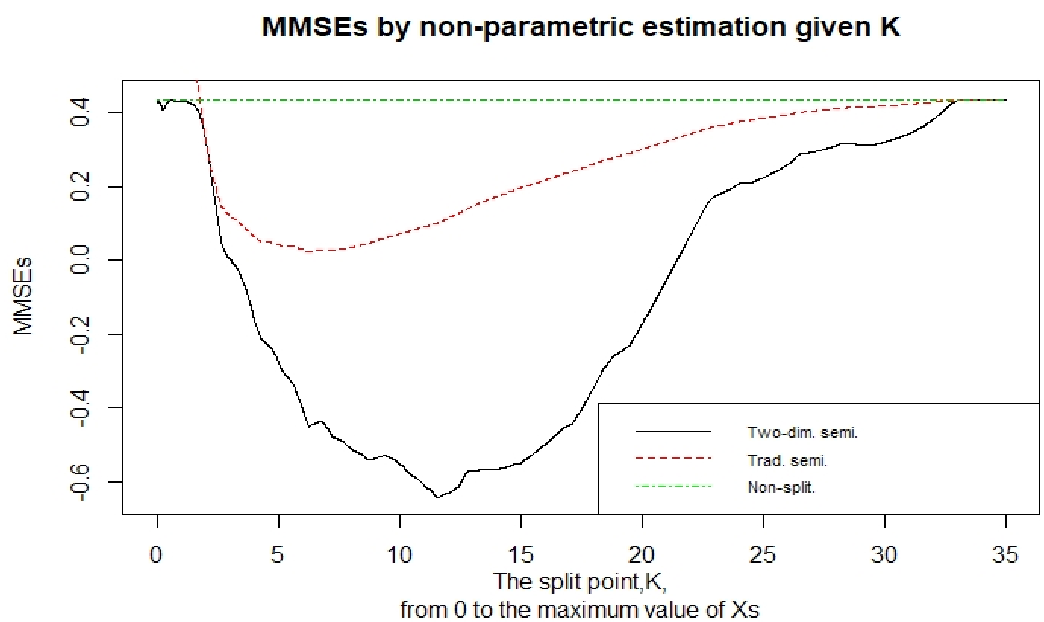

Figure 3.7: MMSEs by nonparametric estimation given K in example3

Thus, we get that the optimal split point is the 88th percentile (11.6). For details, readers are referred to Appendix C. However, due to our method being sensitive to estimation error, the minimum value of MMSE is smaller than zero, which should not be happened.

Accoring to the method in Section 2.2.2, we choose the 0th, 25th, 50th, 75th, 80th, 85th, 90th, 95th and 100th percentiles as our split points. Then, we use R code to calculate the parameters and the optimal α's. We calculate the values of MMSE given by different Ks. You will see all of the MMSEs of each model by nonparametric estimation for specific split points in Figure 3.8.

We choose the minimum one and choose the corresponding K value as the future optimal percentile in future data from the same source(or the same group). In this case, the optimal split point is the 85th percentile.

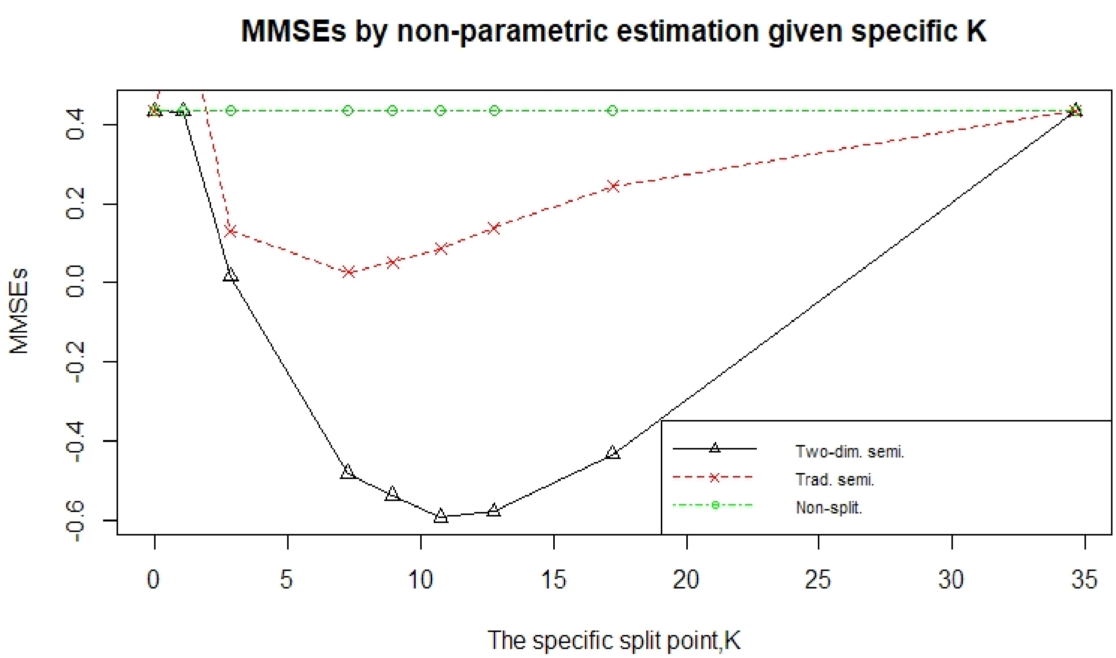

Figure 3.8: MMSEs by nonparametric estimation given specific K in example3

Chapter 4

Conclusion

In this thesis, we proposed a two-dimensional semi-linear credibility model. This model extended the commonly used semi-linear credibility model with truncation and the split credulity model by explicitly considering the covariance between the primary and excess losses.

For the two-dimensional semi-linear credibility model, we derived the parameter values by minimizing the mean squared errors. The formula for the value of MSE at its minimum is also derived. The key results are repeated in the following:

1. No solution if the determinant of matix (2.47) is zero and $\tau_{XY}/\tau_X^2 \neq (n\tau_{XY} + \delta_{XY})/(n\tau_X^2 + \sigma_X^2)$.

2. Non-split solutions if the determinant of matix (2.47) is zero and $\tau_{XY}/\tau_X^2 = (n\tau_{XY} + \delta_{XY})/(n\tau_X^2 + \sigma_X^2)$ or if Theorem 2.2.3 is accepted.

3. Split solutions if the determinant of matix (2.47) is not zero and Theorem 2.2.3 is not accepted.

We then suggested a simple method to determine the optimal split point. Furthermore, the formulas of nonparametric estimation have been derived.

We showed the application of our model through three examples: an Exponential distribution with the rate parameter following a Gamma distribution and a Poisson distribution with the rate parameter following a Gamma distribution and a mixture of two Exponential distributions where one distribution is conditional on the rate parameter following a Gamma distribution.

In the first two examples, we showed that, in theory, splitting the loss into primary and the excess did not reduce the MSE. However, in our numerical examples, since parameters are estimated from data, our method is sensitive to estimation error and one may find that the splitting does help reducing MSE.

In the third example, we showded that splitting the loss into primary and the excess did reduce the MSE. However, since parameters are estimated from data, our method is sensitive to estimation error and the minimum value of MMSE under nonparametric estimation is smaller than zero.

In reality, whether one should split the losses or not is an important but complicated issue. The classical split credibility model used by NCCI assumes that the primary and the excess losses are independent, which in our view is inaccurate. However, the method is used by actuaries for many years. Our model is an attempt to improve the NCCI model by considering the dependence between the primary and excess losses, but more theoretical and empirical investigations are needed.

Bibliography

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

[2] H Bühlmann, A Gisler, and WS Jewell. Excess claims and data trimming in the context of credibility rating procedures. *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker 82*, (p. 117-147), 1981.

[3] H Bühlmann and E Straub. Glaubwürdigkeit für schadensätze (credibility for loss ratios). *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker 70*, (p. 111-133), 1967.

[4] Hans Bühlmann and Alois Gisler. *A Course in Credibility Theory and its Applications*. Springer-Verlag Berlin Heidelberg, 2005.

[5] KC Border. Differentiating an integral: Leibniz' rule. *California Institute of Technology, Division of the Humanities and Social Sciences*, Spring 1-8, 2002.

[6] William R. Gillam. Parametrizing the workers compensation experience rating plan. *Proceedings of the Casualty Actuarial Society*, (p. 857-911), 1992.

[7] William R. Gillam. Workers compensation experience rating: What every actuary should know. *PCAS LXXIX*, (p. 215-239), 1992.

[8] Stuart A. Klugman, Harry H. Panjer, and Gordon E. Willmot. *Loss Models: From Data to Decisions, 4th Edition*. Wiley Series in Probability and Statistics, 2012.

[9] Marcus M. and Minc H. *Introduction to linear algebra*. Dover, 1988.

[10] NCCI (National Council on Compensation Insurance). Abcs of experience rating. *NCCI*, (p. 3-4), 2019.

[11] Ira Robbin. The theory of split credibility. *Variance 7:1*, (p. 29-60), 2013.

Appendix A

Propositions

We continue using Table 2.1 as the notations here.

Proposition A.0.1

$$E\left(\mu_{n+1}(\Theta)^2\right) = \tau_X^2 + \mu_X^2 \tag{A.1}$$

Proof:

$$\begin{aligned} E\left(\mu_{n+1}(\Theta)^2\right) &= Var(\mu_{n+1}(\Theta)) + E(\mu_{n+1}(\Theta))^2 && \text{(A.2)} \\ &= \tau_X^2 + \mu_X^2 && \text{(A.3)} \end{aligned}$$

■

Proposition A.0.2

$$E\left(\bar{Y}^2\right) = \frac{1}{n}\sigma_Y^2 + \tau_Y^2 + \mu_Y^2 \tag{A.4}$$

Proof:

$$\begin{aligned} E\left(\bar{Y}^2\right) &= Var(\bar{Y}) + E(\bar{Y})^2 && \text{(A.5)} \\ &= E\left[Var(\bar{Y}|\Theta)\right] + Var\left[E(\bar{Y}|\Theta)\right] + \mu_Y^2 && \text{(A.6)} \\ &= E\left[\frac{1}{n^2}Var(Y_1 + \cdots + Y_n|\Theta)\right] + Var\left[\frac{1}{n}E(Y_1 + \cdots + Y_n|\Theta)\right] + \mu_Y^2 && \text{(A.7)} \\ &= E\left[\frac{1}{n}Var(Y|\Theta)\right] + Var\left[E(Y|\Theta)\right] + \mu_Y^2 && \text{(A.8)} \\ &= \frac{1}{n}\sigma_Y^2 + \tau_Y^2 + \mu_Y^2 && \text{(A.9)} \end{aligned}$$

■

With the same proof, we have

$$E\left(\bar{L}^2\right) = \frac{1}{n}\sigma_L^2 + \tau_L^2 + \mu_L^2 \tag{A.10}$$

Proposition A.0.3

$$E\left(\mu_{n+1}(\Theta)\bar{Y}\right) = \tau_{XY} + \mu_X\mu_Y \tag{A.11}$$

Proof:

$$\begin{aligned}
E\left(\mu_{n+1}(\Theta)\bar{Y}\right) &= E\left(E\left(\mu_{n+1}(\Theta)\bar{Y}|\Theta\right)\right) && \text{(A.12)}\\
&= E\left(\mu_{n+1}(\Theta)E\left(\bar{Y}|\Theta\right)\right) && \text{(A.13)}\\
&= E\left(\mu_{n+1}(\Theta)\mu_Y(\Theta)\right) && \text{(A.14)}\\
&= Cov(\mu_{n+1}(\Theta), \mu_Y(\Theta)) + E(\mu_{n+1}(\Theta))E(\mu_Y(\Theta)) && \text{(A.15)}\\
&= \tau_{XY} + \mu_X\mu_Y && \text{(A.16)}
\end{aligned}$$

∎

With the same proof, we have

$$E\left(\mu_{n+1}(\Theta)\bar{L}\right) = \tau_{XL} + \mu_X\mu_L \tag{A.17}$$

Proposition A.0.4

$$E\left(\bar{Y}\bar{L}\right) = \frac{1}{n}\delta_{YL} + \tau_{YL} + \mu_Y\mu_L \tag{A.18}$$

Proof:

$$\begin{aligned}
E\left(\bar{Y}\bar{L}\right) &= E\left(E\left(\bar{Y}\bar{L}|\Theta\right)\right) && \text{(A.19)}\\
&= E\left\{\frac{1}{n^2}E\Big[(Y_1 + \cdots + Y_n)(L_1 + \cdots + L_n)|\Theta\Big]\right\} && \text{(A.20)}\\
&= \frac{1}{n^2}E\left[nE(Y_iL_i|\Theta) + (n^2 - n)E(Y_iL_j|\Theta)\right] && \text{(A.21)}\\
&= \frac{1}{n}E\left[E(Y_iL_i|\Theta) + (n-1)E(Y_i|\Theta)E(L_j|\Theta)\right] && \text{(A.22)}\\
&= \frac{1}{n}E(Y_iL_i) + (1 - \frac{1}{n})E\left(\mu_Y(\Theta)\mu_L(\Theta)\right) && \text{(A.23)}\\
&= \frac{1}{n}\Big[Cov(Y_i, L_i) + \mu_Y\mu_L\Big] + (1 - \frac{1}{n})\Big[Cov(\mu_Y(\Theta), \mu_L(\Theta)) + \mu_Y\mu_L\Big] && \text{(A.24)}\\
&= \frac{1}{n}\Big(\delta_{YL} + \tau_{YL} + \mu_Y\mu_L\Big) + (1 - \frac{1}{n})\Big(\tau_{YL} + \mu_Y\mu_L\Big) && \text{(A.25)}\\
&= \frac{1}{n}\delta_{YL} + \tau_{YL} + \mu_Y\mu_L && \text{(A.26)}
\end{aligned}$$

for $i \neq j$ and the reason of the fourth step is independence of X_i and X_j conditional on Θ. The second from the last step follows by Equation (2.22). ∎

Appendix B

Formulas of nonparametric estimation

We continue using Table 2.1 as the notations here. Suppose that we have the data as follows,

Losses / Theta's	$\boldsymbol{X}$						$\boldsymbol{Y} = min(\boldsymbol{X}, K)$				
$\Theta_1 = \theta_1$	X_{11}	X_{12}	X_{13}	$\cdots$	X_{1n}		Y_{11}	Y_{12}	Y_{13}	$\cdots$	Y_{1n}
$\Theta_2 = \theta_2$	X_{21}	X_{22}	X_{23}	$\cdots$	X_{2n}		Y_{21}	Y_{22}	Y_{23}	$\cdots$	Y_{2n}
$\Theta_3 = \theta_3$	X_{31}	X_{32}	X_{33}	$\cdots$	X_{3n}		Y_{31}	Y_{32}	Y_{33}	$\cdots$	Y_{3n}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$\Theta_m = \theta_m$	X_{m1}	X_{m2}	X_{m3}	$\cdots$	X_{mn}		Y_{m1}	Y_{m2}	Y_{m3}	$\cdots$	Y_{mn}

Table B.1: A list of data

And the unbiased estimation of our needed parameters are

$$\hat{\mu_X} = \bar{X} = \frac{1}{nm}\sum_{i=1}^{m}\sum_{j=1}^{n} X_{ij}, \tag{B.1}$$

$$\hat{\mu_Y} = \bar{Y} = \frac{1}{nm}\sum_{i=1}^{m}\sum_{j=1}^{n} Y_{ij}, \tag{B.2}$$

and

$$\widehat{\sigma_X^2} = E(\widehat{V_X(\Theta_i)}) = \frac{1}{m(n-1)}\sum_{i=1}^{m}\sum_{j=1}^{n}(X_{ij} - \bar{X}_i)^2, \tag{B.3}$$

$$\widehat{\sigma_Y^2} = E(\widehat{V_Y(\Theta_i)}) = \frac{1}{m(n-1)}\sum_{i=1}^{m}\sum_{j=1}^{n}(Y_{ij} - \bar{Y}_i)^2, \tag{B.4}$$

and

$$\widehat{\tau_X^2} = \widehat{Var(\bar{X}_i)} - \frac{\widehat{\sigma_X^2}}{n} \quad \text{(B.5)}$$
$$= \frac{1}{m-1}\sum_{i=1}^{m}(\bar{X}_i - \bar{X})^2 - \frac{1}{mn(n-1)}\sum_{i=1}^{m}\sum_{j=1}^{n}(X_{ij} - \bar{X}_i)^2, \quad \text{(B.6)}$$

$$\widehat{\tau_Y^2} = \widehat{Var(\bar{Y}_i)} - \frac{\widehat{\sigma_Y^2}}{n} \quad \text{(B.7)}$$
$$= \frac{1}{m-1}\sum_{i=1}^{m}(\bar{Y}_i - \bar{Y})^2 - \frac{1}{mn(n-1)}\sum_{i=1}^{m}\sum_{j=1}^{n}(Y_{ij} - \bar{Y}_i)^2, \quad \text{(B.8)}$$

and

$$\hat{\delta_{XY}} = E(\delta_{XY}\widehat{(\Theta_i)}) = \frac{1}{m}\sum_{i=1}^{m}\delta_{XY}\widehat{(\Theta_i)} \quad \text{(B.9)}$$
$$= \frac{1}{m(n-1)}\sum_{i=1}^{m}\sum_{j=1}^{n}(X_{ij} - \bar{X}_i)(Y_{ij} - \bar{Y}_i), \quad \text{(B.10)}$$

and

$$\hat{\tau_{XY}} = \widehat{Cov(\bar{X}_i, \bar{Y}_i)} - \frac{\hat{\delta_{XY}}}{n} \quad \text{(B.11)}$$
$$= \frac{1}{m-1}\sum_{i=1}^{m}(\bar{X}_i - \bar{X})(\bar{Y}_i - \bar{Y}) - \frac{1}{mn(n-1)}\sum_{i=1}^{m}\sum_{j=1}^{n}(X_{ij} - \bar{X}_i)(Y_{ij} - \bar{Y}_i). \quad \text{(B.12)}$$

Appendix C

R code of examples

Example1 and Example2 are referred to the following R code.

```
memory.limit(102400)
#### n represents the numer of losses and m represents the number
#### of thetas.
n <- 45
m <- 6
#### Randomly generate thetas with gamma distribution.
alpha <- 6
beta <- 50
theta <- rgamma(m,alpha,rate = beta)
#### Randomly generate losses with exponential distribution
#### conditional on thetas.
X <- matrix(nrow=0,ncol=n)
for(value in theta){
  random_exp <- rexp(n,rate = value)
  X <- rbind(X,random_exp)
  # ------------------------------------------------------------
  #### use the code when we generate losses with poisson distribution
  #### conditional on thetas.
  #random_pois <- rpois(n,lambda = value)
  #X <- rbind(X,random_pois)
}
# ------------------------------------------------------------
#### use the code when we collect the value of parameters.
#hat_total <- matrix(nrow=0,ncol=15,dimnames = list(c(),c(
#                "mu_{X}","sigma_{X}^{2}","tau_{X}^2",
#                "mu_{Y}","sigma_{Y}^{2}","tau_{X}^2",
#                "delta_{XY}","tau_{XY}","hat{Z}",
#                "hat{alpha_{Y}}","hat{alpha_{X}}","hat{alpha_{0}}",
#                "hat{Q}_{nsp}","hat{Q}_{tsl}","hat{Q}_{tdsl}")))
# ------------------------------------------------------------
#### Calculate the mean, variance and covariance of losses by
#### nonparametric estimation.
X_mu_theta <- rep(0,m)
```

```
X_v_theta <- rep(0,m)
for(index in 1:m){
  X_mu_theta[index] <- mean(X[index,])
  X_v_theta[index] <- var(X[index,])
}
X_mu <- mean(X_mu_theta) #### mu_X
X_v <- mean(X_v_theta) #### sigma^2_X
X_a <- var(X_mu_theta)-X_v/n #### tau^2_X
#### Calculate the mean, variance and covariance of losses by
#### parametric estimation under exponential distribution
#### conditional on thetas.
X_mu_para <- beta/(alpha-1)
X_v_para <- beta^2/(alpha-1)/(alpha-2)
X_a_para <- X_v_para-X_mu_para^2
#### Z_hat is the credibility solution of non-split credibility
#### model by nonparametric estimation.
Z_hat <- n/(n+X_v/X_a)
#### Z_hat_para is the credibility solution of non-split credibility
#### model by parametric estimation under exponential distribution
#### conditional on thetas.
Z_hat_para <- n/(n+X_v_para/X_a_para)
# -------------------------------------------------------------
#### MMSE_split represents the value of MMSE in our credibility model
#### by nonparametric estimation.
#### MMSE_split_Y represents the value of MMSE in
#### semi-linear credibility model by nonparametric estimation.
#### MMSE_nonsplit represents the value of MMSE in non-split
#### credibility model by nonparametric estimation.
# -------------------------------------------------------------
#### MMSE_split_para represents the value of MMSE in our credibility
#### model by parametric estimation.
#### MMSE_split_Y_para represents the value of MMSE in
#### semi-linear credibility model by parametric estimation.
#### MMSE_nonsplit_para represents the value of MMSE in non-split
#### credibility model by parametric estimation.
# -------------------------------------------------------------
#### kk is the chosen split point.
# -------------------------------------------------------------
#### use this code when we choose the 0th,25th,50th,75th,80th,85th,
#### 90th,95th,100th percentiles split point
#PP <- sort(X)
#RESULT <- c("0th","25th","50th","75th","80th","85th","90th","95th",
#                    "100th")
#kk <- c(0,PP[0.25*n*m],PP[0.5*n*m],PP[0.75*n*m],PP[0.8*n*m],
#         PP[0.85*n*m],PP[0.9*n*m],PP[0.95*n*m],PP[n*m])
# -------------------------------------------------------------
#### Under poisson distribution conditional on thetas, we change
```

```
#### 'by=0.1' to 'by=1'.
kk<-seq(from=0,to=max(ceiling(X)),by=0.1)
MMSE_split <- rep(0,length(kk))
MMSE_split_Y <- rep(0,length(kk))
MMSE_nonsplit <- rep(0,length(kk))
MMSE_split_para <- rep(0,length(kk))
MMSE_split_Y_para <- rep(0,length(kk))
MMSE_nonsplit_para <- rep(0,length(kk))
#### Begin to calculate all of the MMSEs given by split points
#### and choose the optimal split point.
number <- 0
for(k in kk){
  number <- number + 1
  # --------------------------------------------------------------
  #### use the code when we need to specify K.
  # k <- 5
  # --------------------------------------------------------------
  #### Generate Y=min(X,K) given by split point, K.
  Y <- matrix(nrow=0,ncol=n)
  for(index in 1:m){
    random_function <- ifelse(X[index,]<k,X[index,],k)
    Y <- rbind(Y,random_function)
  }
  #### Calculate the mean, variance and covariance of Y by
  #### nonparametric estimation.
  Y_mu_theta <- rep(0,m)
  Y_v_theta <- rep(0,m)
  for(index in 1:m){
    Y_mu_theta[index] <- mean(Y[index,])
    Y_v_theta[index] <- var(Y[index,])
  }
  Y_mu <- mean(Y_mu_theta) #### mu_Y
  Y_v <- mean(Y_v_theta) #### sigma^2_Y
  Y_a <- var(Y_mu_theta)-Y_v/n #### tau^2_Y
  #### Calculate the mean, variance and covariance of Y by
  #### parametric estimation under exponential distribution
  #### conditional on thetas.
  Y_mu_para <- beta/(alpha-1)*(1-(beta/(beta+k))^(alpha-1))
  Y_v_para <- beta^2/(alpha-1)/(alpha-2)*(1-(beta/(beta+2*k))^
              (alpha-2))-2*k*beta/(alpha-1)*(beta/(beta+k))^(alpha-1)
  Y_a_para <- beta^2/(alpha-1)/(alpha-2)*(1+(beta/(beta+2*k))^
              (alpha-2)-2*(beta/(beta+k))^(alpha-2))-Y_mu_para^2
  #### Calculate tau_XY and delta_XY by nonparametric estimation.
  pi_theta <- rep(0,m)
  for(index in 1:m){
    pi_theta[index] <- cov(X[index,],Y[index,])
  }
```

```
delta <- mean(pi_theta) #### delta_XY
tau <- cov(X_mu_theta,Y_mu_theta)-delta/n #### tau_XY
#### Calculate tau_XY and delta_XY by parametric estimation under
#### exponential distribution conditional on thetas.
delta_para <- beta^2/(alpha-1)/(alpha-2)*(1-(beta/(beta+k))^
              (alpha-2))-k*beta/(alpha-1)*(beta/(beta+k))^(alpha-1)
tau_para <- beta^2/(alpha-1)/(alpha-2)*(1-(beta/(beta+k))^
              (alpha-2))-beta^2/(alpha-1)^2*(1-(beta/(beta+k))^
              (alpha-1))
# ---------------------------------------------------------------
#### use the code when we generate losses with poisson distribution
#### conditional on thetas.
#### Calculate tau_XY and delta_XY by parametric estimation
#### under poisson distribution conditional on thetas.
#delta_para <- 0
#tau_para <- 0
#for (i in 0:(k-1)) {
#  delta_para <- delta_para+(i-k)*beta^alpha/(beta+1)^(alpha+k)*
#    gamma(alpha+k)/gamma(alpha)*(k-(alpha+k)/(beta+1))
#  tau_para <- tau_para+(i-k)*beta^alpha/(beta+1)^(alpha+k)*
#    gamma(alpha+k)/gamma(alpha)*((alpha+k)/(beta+1)-alpha/beta)
#}
# ---------------------------------------------------------------
#### Calculate the optimal alphas by nonparametric estimation.
left <- matrix(c(n*Y_a+Y_v,n*tau+delta,n*tau+delta,n*X_a+X_v),
              nrow=2,ncol=2)
right <- matrix(c(tau,X_a))
b_hat <- matrix(c(0,0)) #### (alpha_Y, alpha_X)^T
if(det(left)>1e-5){
  b_hat <- solve(left,right)
}else{
  if(tau/X_a-(n*tau+delta)/(n*X_a+X_v)<1e-5){
    b_hat[1,1] <- 0
    b_hat[2,1] <- X_a/(n*X_a+X_v)
  }else{
    cat("No solution when K=",k,"\n",sep="")
    next
  }
}
a_hat <- X_mu-n*Y_mu*b_hat[1,1]-n*X_mu*b_hat[2,1] #### alpha_0
#### Calculate the optimal alphas by parametric estimation under
#### exponential distribution conditional on thetas.
left_para <- matrix(c(n*Y_a_para+Y_v_para,n*tau_para+delta_para,
                      n*tau_para+delta_para,n*X_a_para+X_v_para),
                      nrow=2,ncol=2)
right_para <- matrix(c(tau_para,X_a_para))
b_hat_para <- matrix(c(0,0))
```

```
if(det(left_para)>1e-5){
  b_hat_para <- solve(left_para,right_para)
}else{
  if(tau_para/X_a_para-
     (n*tau_para+delta_para)/(n*X_a_para+X_v_para)<1e-5){
    b_hat_para[1,1] <- 0
    b_hat_para[2,1] <- X_a_para/(n*X_a_para+X_v_para)
  }else{
    cat("No solution when K=",k,"\n",sep="")
    next
  }
}
a_hat_para <- X_mu_para-n*Y_mu_para*b_hat_para[1,1]-
              n*X_mu_para*b_hat_para[2,1]
#### Calculate the MMSEs by nonparametric estimation.
MMSE_split[number] <- X_a + n*b_hat[1,1]^2*Y_v +
  n^2*b_hat[1,1]^2*Y_a - 2*n*b_hat[1,1]*tau +
  n*b_hat[2,1]^2*X_v + n^2*b_hat[2,1]^2*X_a - 2*n*b_hat[2,1]*X_a +
  2*n*b_hat[1,1]*b_hat[2,1]*delta + 2*n^2*b_hat[1,1]*b_hat[2,1]*tau
MMSE_nonsplit[number] <- Z_hat^2*X_v/n+(1-Z_hat)^2*X_a
if(k==0){
  MMSE_split_Y[number] <- MMSE_nonsplit[number]
}else{
  MMSE_split_Y[number] <- X_a-n*tau^2/(n*Y_a+Y_v)
}
#### Calculate the MMSEs by parametric estimation under
#### exponential distribution conditional on thetas.
MMSE_split_para[number] <- X_a_para + n*b_hat_para[1,1]^2*Y_v_para+
  n^2*b_hat_para[1,1]^2*Y_a_para - 2*n*b_hat_para[1,1]*tau_para +
  n*b_hat_para[2,1]^2*X_v_para + n^2*b_hat_para[2,1]^2*X_a_para -
  2*n*b_hat_para[2,1]*X_a_para +
  2*n*b_hat_para[1,1]*b_hat_para[2,1]*delta_para +
  2*n^2*b_hat_para[1,1]*b_hat_para[2,1]*tau_para
MMSE_nonsplit_para[number] <- Z_hat_para^2*X_v_para/n+
                             (1-Z_hat_para)^2*X_a_para
if(k==0){
  MMSE_split_Y_para[number] <- MMSE_nonsplit_para[number]
}else{
  MMSE_split_Y_para[number] <- X_a_para-n*tau_para^2/
                              (n*Y_a_para+Y_v_para)
}
# ---------------------------------------------------------------
#### use the code when we collect the value of parameters.
#hat_total <- rbind(hat_total,c(X_mu,X_v,X_a,
#                               Y_mu,Y_v,Y_a,
#                               delta,tau,Z_hat,
#                               b_hat[1,1],b_hat[2,1],a_hat,
```

```
  #                               MMSE_nonsplit[number],
  #                               MMSE_split_Y[number],
  #                               MMSE_split[number]))
  # ------------------------------------------------------------
}
# ------------------------------------------------------------
#### use this code when we choose the 0th,25th,50th,75th,80th,85th,
#### 90th,95th,100th percentiles split point
#### Draw a picture for comparing with each other by nonparametric
#### estimation with specific K.
#plot(kk,MMSE_split,ylim = c(min(min(MMSE_split)),max(MMSE_split)),
#     main="MMSEs by nonparametric estimation given specific K",
#     xlab="The specific split point,K",
#     ylab="MMSEs",lty=1,pch=2,type='o')
#lines(kk,MMSE_split_Y,lty=2,pch=4,type='o',col='2')
#lines(kk,MMSE_nonsplit,lty=4,pch=1,type='o',col='3')
#legend("bottomleft", cex=0.7, pch=c(2,4,1),lty=c(1,2,4),
#       col=c("black", "red", "green"),
#       legend=c("Two-dim. semi.", "Trad. semi.", "Non-split."))
#### K is the optimal specific split point in percentage mode by
#### nonparametric estimation.
#K <- RESULT[which(MMSE_split==min(MMSE_split))]
# ------------------------------------------------------------
#### Draw a picture for comparing with each other by
#### nonparametric estimation.
plot(kk,MMSE_split,ylim = c(min(min(MMSE_split)),max(MMSE_split)),
     main="MMSEs␣by␣nonparametric␣estimation␣given␣K",
     xlab="The␣split␣point,K,\nfrom␣0␣to␣the␣maximum␣value␣of␣Xs",
     ylab="MMSEs", lty=1,type = 'l')
lines(kk,MMSE_split_Y,lty=2,col='2')
lines(kk,MMSE_nonsplit,lty=4,col='3')
legend("bottomleft", cex=0.7, lty=c(1,2,4),
       col=c("black", "red", "green"),
       legend=c("Two-dim.␣semi.", "Trad.␣semi.", "Non-split."))
#### K is the optimal split point in percentage mode by
#### nonparametric estimation.
K <- kk[which(MMSE_split==min(MMSE_split))]
K <- sum(X<K)/n/m
# ------------------------------------------------------------
#### Draw a picture for comparing with each other by parametric
#### estimation.
plot(kk,MMSE_split_para,ylim = c(min(-1,min(MMSE_split_para)),
     max(MMSE_split_Y_para)),
     main="MMSEs␣by␣parametric␣estimation␣given␣K",
     xlab="The␣split␣point,K,\nfrom␣0␣to␣the␣maximum␣value␣of␣Xs",
     ylab="MMSEs", lty=1,type = 'l')
lines(kk,MMSE_split_Y_para,lty=2, col='2')
```

```
lines(kk,MMSE_nonsplit_para,lty=4, col='3')
legend("topright", lty=c(1,2,4), col=c("black", "red", "green"),
       legend=c("Two-dim.␣semi.", "Trad.␣semi.", "Non-split."))
#### K_para is the optimal split point in percentage mode by
#### parametric estimation.
K_para <- kk[which(MMSE_split_para==min(MMSE_split_para))]
K_para <- sum(X<K_para)/n/m
```

Example3 is referred to the following R code.

```
memory.limit(102400)
#### n represents the numer of losses.
#### m represents the number of thetas.
#### omega is the weight of the first exponential distribution.
#### 1/lambda is the mean of the second exponential distribution.
n <- 45
m <- 6
omega <- 0.75
lambda <- 0.1
#### Randomly generate thetas with gamma distribution.
alpha <- 6
beta <- 25
theta <- rgamma(m,alpha,rate = beta)
#### Randomly generate losses with a mixture of two exponential
#### distribution conditonal on thetas.
X <- matrix(nrow=0,ncol=n)
for(value in theta){
  #### Sample n random uniforms U.
  U =runif(n)
  #### Variable to store the samples from the mixture distribution
  rand.samples = rep(NA,n)
  #### Sampling from the mixture
  for(i in 1:n){
    if(U[i] < omega){
      rand.samples[i] = rexp(1,rate = value)
    }else{
      rand.samples[i] = rexp(1,rate = lambda)
    }
  }
  X <- rbind(X,rand.samples)
}
# --------------------------------------------------------------
#### use the code when we collect the value of parameters.
#hat_total <- matrix(nrow=0,ncol=15,dimnames = list(c(),c(
#                 "mu_{X}","sigma_{X}^{2}","tau_{X}^2",
#                 "mu_{Y}","sigma_{Y}^{2}","tau_{X}^2",
#                 "delta_{XY}","tau_{XY}","hat{Z}",
#                 "hat{alpha_{Y}}","hat{alpha_{X}}","hat{alpha_{0}}",
```

```
#                  "hat{Q}_{nsp}","hat{Q}_{tsl}","hat{Q}_{tdsl}")))
# ---------------------------------------------------------------
#### Calculate the mean, variance and covariance of losses by
#### non-parametric estimation.
X_mu_theta <- rep(0,m)
X_v_theta <- rep(0,m)
for(index in 1:m){
  X_mu_theta[index] <- mean(X[index,])
  X_v_theta[index] <- var(X[index,])
}
X_mu <- mean(X_mu_theta) #### mu_X
X_v <- mean(X_v_theta) #### sigma^2_X
X_a <- var(X_mu_theta)-X_v/n #### tau^2_X
#### Calculate the mean, variance and covariance of losses by
#### parametric estimation.
X_mu_para <- omega*beta/(alpha-1)+(1-omega)/lambda
X_v_para <- (2*omega-omega^2)*beta^2/(alpha-1)/(alpha-2)-
  2*omega*(1-omega)/lambda*beta/(alpha-1)+(1-omega^2)/lambda^2
X_a_para <- omega^2*beta^2/(alpha-1)^2/(alpha-2)
#### Z_hat's are the credibility solution of non-split credibility
#### model.
Z_hat <- n/(n+X_v/X_a)
Z_hat_para <- n/(n+X_v_para/X_a_para)
# ---------------------------------------------------------------
#### use this code when we choose the 0th,25th,50th,75th,80th,85th,
#### 90th,95th,100th percentiles split point
PP <- sort(X)
RESULT <- c("0th","25th","50th","75th","80th","85th","90th","95th",
                    "100th")
kk <- c(0,PP[0.25*n*m],PP[0.5*n*m],PP[0.75*n*m],PP[0.8*n*m],
        PP[0.85*n*m],PP[0.9*n*m],PP[0.95*n*m],PP[n*m])
# ---------------------------------------------------------------
kk<-seq(from=0,to=max(ceiling(X)),by=0.01)
MMSE_split <- rep(0,length(kk))
MMSE_split_Y <- rep(0,length(kk))
MMSE_nonsplit <- rep(0,length(kk))
MMSE_split_para <- rep(0,length(kk))
MMSE_split_Y_para <- rep(0,length(kk))
MMSE_nonsplit_para <- rep(0,length(kk))
#### Begin to calculate all of the MMSEs given by split points and
#### choose the optimal split point.
number <- 0
for(k in kk){
  number <- number + 1
  # ---------------------------------------------------------------
  #### use the code when we need to specify K.
  # k <- 7.99
```

```
# -------------------------------------------------------------
#### Generate Y=min(X,K) given by split point, K.
Y <- matrix(nrow=0,ncol=n)
for(index in 1:m){
  random_function <- ifelse(X[index,]<k,X[index,],k)
  Y <- rbind(Y,random_function)
}
#### Calculate the mean, variance and covariance of Y by
#### non-parametric estimation.
Y_mu_theta <- rep(0,m)
Y_v_theta <- rep(0,m)
for(index in 1:m){
  Y_mu_theta[index] <- mean(Y[index,])
  Y_v_theta[index] <- var(Y[index,])
}
Y_mu <- mean(Y_mu_theta) #### mu_Y
Y_v <- mean(Y_v_theta) #### sigma^2_Y
Y_a <- var(Y_mu_theta)-Y_v/n #### tau^2_Y
#### Calculate the mean, variance and covariance of Y by
#### parametric estimation.
Y_mu_para <- omega*beta/(alpha-1)*(1-(beta/(beta+k))^(alpha-1))+
  (1-omega)*(1/lambda-1/lambda*exp(-lambda*k))
Y_v_para <- -2*k*omega*beta/(alpha-1)*(beta/(beta+k))^(alpha-1)+
  (2-omega)*omega*beta^2/(alpha-1)/(alpha-2)+
  2*omega*(omega-1)*beta^2/(alpha-1)/(alpha-2)*
               (beta/(beta+k))^(alpha-2)-
  omega^2*beta^2/(alpha-1)/(alpha-2)*(beta/(beta+2*k))^(alpha-2)+
  (1-omega)*(-2*k/lambda*exp(-lambda*k)+(1+omega)/lambda^2-
               2*omega/lambda^2*exp(-lambda*k)-(1-omega)/
               lambda^2*exp(-2*lambda*k))-
  2*omega*(1-omega)*(1/lambda-1/lambda*exp(-lambda*k))*
               beta/(alpha-1)*(1-(beta/(beta+k))^(alpha-1))
Y_a_para <- omega^2*(beta^2/(alpha-1)/(alpha-2)*
               (1+(beta/(beta+2*k))^(alpha-2)-
  2*(beta/(beta+k))^(alpha-2))-beta^2/(alpha-1)^2*
               (1-(beta/(beta+k))^(alpha-1))^2)
#### Calculate tau_XY and delta_XY by non-parametric estimation.
pi_theta <- rep(0,m)
for(index in 1:m){
  pi_theta[index] <- cov(X[index,],Y[index,])
}
delta <- mean(pi_theta) #### delta_XY
tau <- cov(X_mu_theta,Y_mu_theta)-delta/n #### tau_XY
#### Calculate tau_XY and delta_XY by parametric estimation.
delta_para <- -k*omega*beta/(alpha-1)*(beta/(beta+k))^(alpha-1)+
  (2-omega)*omega*beta^2/(alpha-1)/(alpha-2)-
  omega*(2-omega)*beta^2/(alpha-1)/(alpha-2)*
```

```
                  (beta/(beta+k))^(alpha-2)+
    (1-omega)*(-k/lambda*exp(-lambda*k)+(1+omega)/lambda^2-
                  (1+omega)/lambda^2*exp(-lambda*k))-
    omega*(1-omega)*(1/lambda-1/lambda*exp(-lambda*k))*
                  beta/(alpha-1)-
    omega*(1-omega)/lambda*beta/(alpha-1)*
                  (1-(beta/(beta+k))^(alpha-1))
tau_para <- omega^2*(beta^2/(alpha-1)/(alpha-2)*
                  (1-(beta/(beta+k))^(alpha-2))-
    beta^2/(alpha-1)^2*(1-(beta/(beta+k))^(alpha-1)))
#### Calculate the optimal alphas by non-parametric estimation.
left <- matrix(c(n*Y_a+Y_v,n*tau+delta,n*tau+delta,n*X_a+X_v),
                  nrow=2,ncol=2)
right <- matrix(c(tau,X_a))
b_hat <- matrix(c(0,0)) #### (alpha_Y, alpha_X)^T
if(det(left)>1e-5){
  b_hat <- solve(left,right)
}else{
  if(tau/X_a-(n*tau+delta)/(n*X_a+X_v)<1e-5){
    b_hat[1,1] <- 0
    b_hat[2,1] <- X_a/(n*X_a+X_v)
  }else{
    cat("No solution when K=",k,"\n",sep="")
    next
  }
}
a_hat <- X_mu-n*Y_mu*b_hat[1,1]-n*X_mu*b_hat[2,1] #### alpha_0
#### Calculate the optimal alphas by parametric estimation.
left_para <- matrix(c(n*Y_a_para+Y_v_para,n*tau_para+delta_para,
                       n*tau_para+delta_para,n*X_a_para+X_v_para),
                       nrow=2,ncol=2)
right_para <- matrix(c(tau_para,X_a_para))
b_hat_para <- matrix(c(0,0))
if(det(left_para)>1e-5){
  b_hat_para <- solve(left_para,right_para)
}else{
  if(tau_para/X_a_para-
     (n*tau_para+delta_para)/(n*X_a_para+X_v_para)<1e-5){
    b_hat_para[1,1] <- 0
    b_hat_para[2,1] <- X_a_para/(n*X_a_para+X_v_para)
  }else{
    cat("No solution when K=",k,"\n",sep="")
    next
  }
}
a_hat_para <- X_mu_para-n*Y_mu_para*b_hat_para[1,1]-
              n*X_mu_para*b_hat_para[2,1]
```

```
  #### Calculate the MMSEs by nonparametric estimation.
  MMSE_split[number] <- X_a + n*b_hat[1,1]^2*Y_v +
    n^2*b_hat[1,1]^2*Y_a - 2*n*b_hat[1,1]*tau +
    n*b_hat[2,1]^2*X_v + n^2*b_hat[2,1]^2*X_a - 2*n*b_hat[2,1]*X_a +
    2*n*b_hat[1,1]*b_hat[2,1]*delta + 2*n^2*b_hat[1,1]*b_hat[2,1]*tau
  MMSE_nonsplit[number] <- Z_hat^2*X_v/n+(1-Z_hat)^2*X_a
  if(k==0){
    MMSE_split_Y[number] <- MMSE_nonsplit[number]
  }else{
    MMSE_split_Y[number] <- X_a-n*tau^2/(n*Y_a+Y_v)
  }
  #### Calculate the MMSEs by parametric estimation.
  MMSE_split_para[number] <- X_a_para + n*b_hat_para[1,1]^2*Y_v_para+
    n^2*b_hat_para[1,1]^2*Y_a_para - 2*n*b_hat_para[1,1]*tau_para +
    n*b_hat_para[2,1]^2*X_v_para + n^2*b_hat_para[2,1]^2*X_a_para -
    2*n*b_hat_para[2,1]*X_a_para +
    2*n*b_hat_para[1,1]*b_hat_para[2,1]*delta_para +
    2*n^2*b_hat_para[1,1]*b_hat_para[2,1]*tau_para
  MMSE_nonsplit_para[number] <- Z_hat_para^2*X_v_para/n+
                               (1-Z_hat_para)^2*X_a_para
  if(k==0){
    MMSE_split_Y_para[number] <- MMSE_nonsplit_para[number]
  }else{
    MMSE_split_Y_para[number] <- X_a_para-n*tau_para^2/
                                (n*Y_a_para+Y_v_para)
  }
  # --------------------------------------------------------------
  #### The credibility permium given by the optimal split point
  #### under parametric estimation.
  # X_mu_para+n*b_hat_para[1,1]*(Y_mu-Y_mu_para)+
  #      n*b_hat_para[2,1]*(X_mu-X_mu_para)
  # --------------------------------------------------------------
  #### use the code when we collect the value of parameters.
  #hat_total <- rbind(hat_total,c(X_mu,X_v,X_a,
  #                               Y_mu,Y_v,Y_a,
  #                               delta,tau,Z_hat,
  #                               b_hat[1,1],b_hat[2,1],a_hat,
  #                               MMSE_nonsplit[number],
  #                               MMSE_split_Y[number],
  #                               MMSE_split[number]))
  # --------------------------------------------------------------
}
# --------------------------------------------------------------
#### use this code when we choose the 0th,25th,50th,75th,80th,85th,
#### 90th,95th,100th percentiles split point
#### Draw a picture for comparing with each other by
#### non-parametric estimation with specific K.
```

```
plot(kk,MMSE_split,ylim = c(min(MMSE_split),max(MMSE_split+0.01)),
     main="MMSEs␣by␣non-parametric␣estimation␣given␣specific␣K",
     xlab="The␣specific␣split␣point,K",ylab="MMSEs",
     lty=1,pch=2,type='o')
lines(kk,MMSE_split_Y,lty=2,pch=4,type='o',col='2')
lines(kk,MMSE_nonsplit,lty=4,pch=1,type='o',col='3')
legend("bottomright", cex=0.7, pch=c(2,4,1),lty=c(1,2,4),
       col=c("black", "red", "green"),
       legend=c("Two-dim.␣semi.", "Trad.␣semi.", "Non-split."))
#### K is the optimal specific split point in percentage mode by
#### non-parametric estimation.
K <- RESULT[which(MMSE_split==min(MMSE_split))]
# -------------------------------------------------------------
#### Draw a picture for comparing with each other by
#### non-parametric estimation.
plot(kk,MMSE_split,ylim = c(min(MMSE_split),max(MMSE_split+0.01)),
     main="MMSEs␣by␣non-parametric␣estimation␣given␣K",
     xlab="The␣split␣point,K,\nfrom␣0␣to␣the␣maximum␣value␣of␣Xs",
     ylab="MMSEs", lty=1,type = 'l')
lines(kk,MMSE_split_Y,lty=2,col='2')
lines(kk,MMSE_nonsplit,lty=4,col='3')
legend("bottomright", cex=0.7, lty=c(1,2,4),
        col=c("black", "red", "green"),
       legend=c("Two-dim.␣semi.", "Trad.␣semi.", "Non-split."))
#### K is the optimal split point in percentage mode by
#### non-parametric estimation.
K <- kk[which(MMSE_split==min(MMSE_split))]
K <- sum(X<K)/n/m
# -------------------------------------------------------------
#### Draw a picture for comparing with each other by
#### parametric estimation.
plot(kk,MMSE_split_para,ylim = c(min(MMSE_split_para),
                                 max(MMSE_split_para+0.01)),
     main="MMSEs␣by␣parametric␣estimation␣given␣K",
     xlab="The␣split␣point,K,\nfrom␣0␣to␣the␣maximum␣value␣of␣Xs",
     ylab="MMSEs", lty=1,type = 'l')
lines(kk,MMSE_split_Y_para,lty=2, col='2')
lines(kk,MMSE_nonsplit_para,lty=4, col='3')
legend("bottomright", cex=0.8,lty=c(1,2,4),
       col=c("black", "red", "green"),
       legend=c("Two-dim.␣semi.", "Trad.␣semi.", "Non-split."))
#### K_para is the optimal split point in percentage mode by
#### parametric estimation.
K_para <- kk[which(MMSE_split_para==min(MMSE_split_para))]
K_para <- sum(X<K_para)/n/m
```

Curriculum Vitae

Name:	Jingbing Qiu
Post-Secondary Education and Degrees:	South China University of Technology Guangzhou, China 2013 - 2017 B.Sc.
	University of Western Ontario London, ON 2017 - 2019 M.Sc.
Honours and Awards:	Western Graduate Research Scholarship 2018 - 2019
Related Work Experience:	Teaching Assistant The University of Western Ontario 2017 - 2019
	Remote actuarial assistant Shanghai, China 2018 - 2019

Publications:

None