

# Semiparametric Estimation in the Optimal Dividend Barrier for the Classical Risk Model

Hiroshi Shiraishi\* and Zudi Lu†

## Abstract

In the context of an insurance portfolio which provides dividend income for the insurance company's shareholders, an important problem in risk theory is how the premium income will be paid to the shareholders as dividends according to a barrier strategy until the next claim occurs whenever the surplus attains the level of "barrier". In this paper, we are concerned with the estimation of optimal dividend barrier, defined as the level of the barrier that maximizes the expected discounted dividends until ruin, under the widely used compound Poisson model as the aggregate claims process. We propose a semi-parametric statistical procedure for estimation of the optimal dividend barrier, which is critically needed in applications. We first construct a consistent estimator of the objective function that is complexly related to the expected discounted dividends and then the estimated optimal dividend barrier as the minimizer of the estimated objective function. In theory we show that the constructed estimator of the optimal dividend barrier is statistically consistent. Numerical experiments by both simulated and real data analyses demonstrate that the proposed estimators work reasonably well with an appropriate size of samples.

## 1 Introduction

In classical models of risk theory, an insurance company's surplus is assumed to be able to increase without bounds. This is however often not the case in practice, because the company need to pay back its excess surplus as dividends to the shareholders. In order to make it more realistic, the risk model involving dividend payments has become an important problem in risk theory; see, e.g., [De Finetti, 1957], [Bühlmann, 1970], [Gerber, 1979], [Dickson, 2005], [Gerber et al., 2006] and so on. In such risk models with dividends, whenever the surplus attains the level of "barrier  $b$ ", the premium income will be paid to the shareholders as dividends until next claim occurs. But, if the surplus is below  $b$ , no dividends are paid. [De Finetti, 1957] pioneeringly proposed such a risk model with dividends, considering that the optimal strategy to maximize the expectation of the aggregated discounted dividends is a barrier strategy with an optimal level of the barrier developed under a binomial model. [Bühlmann, 1970] and [Gerber, 1979] made further contribution to the idea of the constant barrier under the classical risk models. In particular they derived the optimal dividend barrier  $b_0$  explicitly in the case of an exponential claim amount distribution. Since then, many publications have dealt with this problem based on different distributions, models and criteria. For example, [Lin et al., 2003] and [Gerber et al., 2006] derived the explicit expression of  $b_0$  when the claim amount follows mixed exponential distributions. [Landriault, 2008] investigated this problem when there exists the dependency between interclaim arrivals and claim sizes. As a

---

\*Department of Mathematics, Keio University, Yokohama, Kanagawa, Japan. E-mail: shiraishi@math.keio.ac.jp.

†S3RI and School of Mathematical Sciences, University of Southampton, Highfield, Southampton, SO17 1BJ, UK. E-mail: Z.Lu@soton.ac.uk.

counterpart of the classical continuous time risk model, the Brownian motion risk model has also been considered, for example, by [Gerber and Shiu, 2004]. Moreover, [Dickson and Waters, 2004] and [Gerber et al., 2006] dealt with finding an optimal dividend strategy in case the shareholders should be liable to cover the deficit at ruin. Although this problem of optimal dividends has been actively studied in the literature, it has been rarely investigated in statistical estimation of the unknown optimal “barrier” that is critically needed in applications.

In this paper, we propose a semiparametric statistical procedure for estimation of the optimal dividend barrier  $b_0$ . Differently from the methods by [Gerber et al., 2008] to approximate  $b_0$  by its first few moments based on the moment estimation, we will propose an estimator of the optimal dividend barrier  $b_0$  as the minimizer of a semiparametric estimator of an objective function,  $\Psi(b)$ , which is derived by using “Beekman’s convolution series” of the expected discounted dividends  $V(u, b)$ . We note that the maximizer of  $V(u, b)$  with respect to  $b$  corresponds to the minimizer of  $\Psi(b)$  (see Section 2 for details). For the semiparametric estimation of  $\Psi(b)$ , we will suggest an approach to estimation as done for ruin probability in the literature. In the case of ruin probability, it is known that the probability can be described as the infinite sum of the  $r$ -fold convolution ( $G^{(r)}$ ) of a distribution function  $G$ , which is the so-called Pollaczek-Khinchin-formula. [Croux and Veraverbeke, 1990] suggested its estimator as the finite sum of the U-statistics ( $U_{nr}$ ), which is expressed as the sum of all combinatorial selections of the size  $r$  from the full set of observations (size  $n$ ). In this paper, similarly to [Croux and Veraverbeke, 1990], we will propose an estimator of  $\Psi(b)$  which can also be expressed as the infinite sum of the (derivative of)  $r$ -fold convolution ( $G^r$ ) of a function  $g$ . We will introduce in Section 3 the estimator  $\hat{\Psi}_{m,s}(b)$  as the finite sum of  $\hat{G}_s^r$ , where  $\hat{G}_s^r$  is the derivative of  $\hat{G}_s^r$  and  $\hat{G}_s^r$  is a Monte Carlo based estimator of  $G^r$  based on the empirical distribution  $\hat{g}$ . Note that  $\hat{G}_s^r$  converges to the U-statistic of  $G^r$  when the sample size of the Monte Carlo estimation ( $s$ ) tends to infinity.

The paper is organized as follows: Section 2 will define the optimal dividend barrier which is the maximizer of the expectation of the aggregated discounted dividends. We show that the optimal dividend barrier can be described as the minimizer of a function  $\Psi(b)$  by use of “Beekman’s convolution series”. In Section 3, we suggest a consistently estimated objective function consisting of Monte Carlo based estimators and derive the consistent estimator of the optimal dividend barrier from it. Section 4 investigates the performance of our proposed estimator by a simulation study and a real data analysis. In the simulation study, the influence of the sample size  $n$  for the optimal dividend barrier estimator is also investigated. For the analysis of a real data set of Danish Fire Insurance Claims in Denmark, the superiority for the estimated optimal dividend barrier is illustrated in terms of approximated expected discounted dividends.

## 2 Optimal Dividend Barrier

In this paper, we are concerned with the optimal dividend barrier in the case of dividends payment with an insurance portfolio. We first introduce the special case of non-dividends that will be needed below. In the absence of dividends, the surplus of an insurance company at time  $t$  can be expressed as

$$U(t) = u + ct - S(t), \quad S(t) = \sum_{i=1}^{N(t)} X_i$$

for  $t \geq 0$ . Here  $u > 0$  is the initial surplus; the premiums are received continuously at a constant rate  $c > 0$ ;  $\{S(t)\}$  is the aggregate claims up to time  $t$ ;  $N(t) \in \mathbb{N}$  is the claim frequency up to time  $t$ ;  $X_i$  is the  $i$ th claim amount.

We impose the following assumption with respect to the  $X_i$ ,  $N(t)$ ,  $c$  and  $u$ .

**Assumption 1.** (A1)  $\{X_i, i \in \mathbb{N}\}$  is a sequence of i.i.d. positively valued random variables with the probability density function  $f(x)$  and the finite mean  $\mu$  and variance  $\sigma^2$ .

(A2)  $\{N(t)\}$  is a Poisson process with the intensity parameter  $\lambda > 0$  (i.e.,  $\{S(t) = \sum_{i=1}^{N(t)} X_i\}$  is a Compound Poisson process).

(A3) The premium rate  $c$  is known and satisfied with  $0 < \frac{\lambda\mu}{c} < 1$ .

(A4)  $\{N(t)\}$  and  $\{X_i\}$  are independent.

Further we define  $\rho$  as follows.

**Definition 1.** (D1) Define

$$\ell(\xi) = \delta + \lambda - c\xi - \lambda\mathcal{L}(\xi), \quad (1)$$

where  $\delta > 0$  is a force of interest, and  $\mathcal{L}(\xi) = \int_0^\infty e^{-\xi x} f(x) dx$  is the Laplace transform of  $f$ . Equation  $\ell(\xi) = 0$  is the so-called ‘‘Lundberg’s fundamental equation’’ or ‘‘generalized Lundberg’s equation’’. It is well known that the equation  $\ell(\xi) = 0$  has a unique positive root  $\rho_\delta > 0$  and a unique negative root  $-R_\delta < 0$  (e.g., see [Gerber and Shiu, 1998]).

In the classical ruin theory, it is concerned with the ruin probability. Denote by  $\psi_0(u)$  a probability of ultimate ruin in the absence of dividend as a function of the initial surplus  $u$ ,

$$\psi_0(u) = P\{T < \infty | U(0) = u\}$$

where  $T = \inf\{t | U(t) < 0\}$  is the time of ruin in the absence of dividends. [Gerber and Shiu, 1998] has introduced its extension for  $\delta > 0$

$$\psi_\delta(u) = E\left\{e^{-\delta T + \rho_\delta U(T)} \mathbb{I}_{\{T < \infty\}} | U(0) = u\right\}.$$

**Remark 1.** When  $\delta \downarrow 0$ , the positive solution of Lundberg’s fundamental equation (1) satisfies  $\rho_\delta \downarrow 0$ , which implies that  $\psi_\delta(u) \Big|_{\delta=0} = E\left\{e^{-0 \cdot T + 0 \cdot U(T)} \mathbb{I}_{\{T < \infty\}} | U(0) = u\right\} = P\{T < \infty | U(0) = u\} = \psi_0(u)$ . Therefore,  $\psi_\delta$  is the extension of  $\psi_0$ .

In this paper, we suppose that the insurance company will refund part of the surplus exceeding a barrier  $b \geq u$ , as the dividends to the shareholder. Let  $D_b(t)$  denote the aggregate dividends up to time  $t$ . Then, we can write

$$dD_b(t) = \begin{cases} 0 & \text{if } U_b(t) < b \\ cdt & \text{if } U_b(t) = b \end{cases},$$

where  $U_b(t)$  is the surplus with the dividend barrier  $b$ , that is,

$$U_b(t) = u + c \int_0^t \mathbb{I}_{\{U_b(s) < b\}} ds - S(t), \quad S(t) = \sum_{i=1}^{N(t)} X_i. \quad (2)$$

[De Finetti, 1957], [Gerber, 1979], [Gerber and Shiu, 1998], [Gerber et al., 2006] and so on treated the problem as maximizing the expected discounted dividends, namely,

$$V(u, b) = E\left[\int_0^{T_b} e^{-\delta t} dD_b(t) | U(0) = u\right], \quad (3)$$

where  $T_b = \inf\{t | U_b(t) < 0\}$  is the ruin time in the existence of dividend with barrier  $b$ . Then, an optimal dividend barrier  $b_0$  can be defined as follows.

**Definition 2.** (D2) Let  $b_0$  denote the optimal dividend barrier as the maximizer of  $V(u, b)$ , i.e.,

$$b_0 = \arg \max_{b>0} V(u, b).$$

[Gerber and Shiu, 1998] showed that  $V(u, b)$  (defined by (3)) can be written as a function of  $\rho_\delta$  and  $\psi_\delta$  (in what follows, for simplicity, we denote  $\rho \equiv \rho_\delta$  and  $\psi \equiv \psi_\delta$ ) as follows.

**Lemma 1.** Suppose Assumptions (A1)-(A4) hold. Then, it follows that

$$V(u, b) = \frac{h(u)}{h'(b)} = \frac{e^{\rho u} - \psi(u)}{\rho e^{\rho b} - \psi'(b)}.$$

From Lemma 1, it follows that the optimal dividend barrier  $b_0$  does not depend on the initial surplus  $u$ . We impose the following assumption.

**Assumption 2.** (A5) There exists uniquely the optimal dividend barrier  $b_0$ .

**Remark 2.** Theorem 9 of [Yin et al., 2015]<sup>1</sup> shows that there exists a unique minimizer  $b_0$  such that  $h'(b_0) = \min_{b>0} h'(b)$  if  $f$  is a completely monotone function<sup>2</sup> on  $(0, \infty)$ .

[Gerber and Shiu, 1997] derived another expression for  $h(u)$  by use of “Beekman’s convolution series” as follows.

**Lemma 2.** Suppose Assumptions (A1)-(A5) hold. Then, we have

$$h(u) = \{1 - \psi(0)\} \left[ e^{\rho u} + \int_0^u e^{\rho z} \sum_{r=1}^{\infty} g^{*r}(u-z) dz \right],$$

where  $g(x) = \frac{\lambda}{c} \int_0^\infty e^{-\rho y} f(x+y) dy$  and  $g^{*r}(x)$  ( $r = 1, 2, \dots$ ) is the  $r$ -fold convolution of  $g$  defined by

$$g^{*r}(x) = \begin{cases} g(x) & r = 1 \\ \int_0^x g(y) g^{*(r-1)}(x-y) dy & r \geq 2 \end{cases}.$$

Summarizing the above, the optimal dividend barrier can be expressed as follows.

**Lemma 3.** Suppose Assumptions (A1)-(A5) hold. Then,  $b_0 \in \mathbb{R}^+$  is a minimizer of  $\Psi(b)$ , i.e.,

$$b_0 = \arg \min_{b>0} \Psi(b), \quad \Psi(b) = \sum_{r=0}^{\infty} \dot{G}^r(b),$$

where  $\dot{G}^r(b) = \frac{d}{db} G^r(b)$  and

$$G^r(b) = \begin{cases} e^{\rho b} & r = 0 \\ \int_0^b e^{\rho z} g^{*r}(b-z) dz & r \geq 1 \end{cases}.$$

<sup>1</sup>The paper [Yin et al., 2015] considers the existence of a unique solution under a Lévy process which includes the compound poisson process.

<sup>2</sup>An infinitely differentiable function  $f : (0, \infty) \mapsto [0, \infty)$  is called completely monotone if  $(-1)^n f^{(n)}(x) \geq 0$  for all  $n = 0, 1, 2, \dots$  and all  $x > 0$ .

### 3 Estimation of Optimal Dividend Barrier

#### 3.1 Estimation

Suppose that  $(X_1, \Delta_1 T), \dots, (X_n, \Delta_n T)$  are observed data, where  $X_i$  is the  $i$ -th claim amount and  $\Delta_i T$  is the interclaim time between the  $(i-1)$ -th and  $i$ -th claims. Note that let  $T_i (i \in \mathbb{N})$  be the time when the  $i$ -th claim occurs. Then,  $T_i = \inf_{t>0} \{N(t) = i\}$ . By using  $T_i$  and  $T_{i-1}$ , the interclaim time is written by  $\Delta_i T = T_i - T_{i-1}$  with  $T_0 = 0$ . If the process  $\{N(t)\}$  follows a homogeneous Poisson process with the intensity  $\lambda$ , the interclaim time  $\Delta_i T$  follows independently an Exponential distribution with  $E(\Delta_i T) = 1/\lambda$ . Since the interclaim time  $\Delta_i T$  satisfies  $E(\Delta_i T) = 1/\lambda$  and the claim amount  $X_i$  has common distribution function  $F(x) := \int_0^x f(y)dy$  for  $x > 0$ , natural estimators of  $\lambda$  and  $F$  can be defined by

$$\hat{\lambda} = \left( \frac{1}{n} \sum_{i=1}^n \Delta_i T \right)^{-1} \quad \text{and} \quad \hat{F}(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{X_i \leq x\}} \quad (4)$$

for  $x > 0$ . By using  $\hat{\lambda}$  and  $\hat{F}$  instead of  $\lambda$  and  $F$ , a natural estimator of  $\rho$  can be defined as a positive solution to the (empirical) Lundberg's fundamental equation:

$$\hat{\ell}(\xi) = \delta + \hat{\lambda} - c\xi - \hat{\lambda} \hat{\mathcal{L}}(\xi) = 0, \quad \hat{\mathcal{L}}(\xi) = \int_0^\infty e^{-\xi x} d\hat{F}(x) = \frac{1}{n} \sum_{i=1}^n e^{-\xi X_i}.$$

In the same way, a natural estimator of  $g$  is defined as follows:

$$\hat{g}(x) = \frac{\hat{\lambda}}{c} \int_0^\infty e^{-\hat{\rho}y} d\hat{F}(x+y) = \frac{\hat{\lambda}}{cn} \sum_{i=1}^n \mathbb{I}_{\{X_i > x\}} e^{-\hat{\rho}(X_i - x)}$$

for  $x > 0$ . Note that let  $Y_1, \dots, Y_r$  be independent with common density function  $\tilde{g}(x) = g(x)/K$ , where  $K = \int_0^\infty g(y)dy$ . Then,  $g^{*r}$ , for  $r \geq 2$  and  $x > 0$ , can be written as

$$\begin{aligned} g^{*r}(x) &= \int_0^x g(y) g^{*(r-1)}(x-y) dy = \int \cdots \int_{\sum_{j=1}^{r-1} y_j \leq x} g \left( x - \sum_{j=1}^{r-1} y_j \right) \prod_{i=1}^{r-1} g(y_i) dy_1 \cdots dy_{r-1} \\ &= K^{r-1} E \left[ \mathbb{I}_{\{\sum_{j=1}^{r-1} Y_j \leq x\}} g \left( x - \sum_{j=1}^{r-1} Y_j \right) \right]. \end{aligned}$$

Therefore, we introduce its Monte Carlo based estimator

$$\widehat{g_s^{*r}}(x) = \begin{cases} \hat{g}(x) & r = 1 \\ \frac{\hat{K}^{r-1}}{s} \sum_{\ell=1}^s \mathbb{I}_{\{\sum_{j=1}^{r-1} Y_j^{(\ell)} \leq x\}} \hat{g} \left( x - \sum_{j=1}^{r-1} Y_j^{(\ell)} \right) & r \geq 2 \end{cases},$$

where  $s \equiv s(n)$  is a sample size of Monte Carlo estimation (in what follows, we call ‘‘Monte Carlo sample size’’) satisfying  $n/s \rightarrow 0$  as  $n \rightarrow \infty$ , and  $\{Y_i^{(\ell)}; i = 1, \dots, r-1, \ell = 1, \dots, s\}$  is a sequence of i.i.d. random variables of common density function  $\hat{g}(y) = \hat{g}(y)/\hat{K}$  with

$$\hat{K} = \int_0^\infty \hat{g}(y) dy = 1 - \frac{\delta}{c\hat{\rho}}.$$

By using  $\{\widehat{g}^{*r}\}$  and  $m \equiv m(n)$  satisfying  $\log n/m \rightarrow 0$  as  $n \rightarrow \infty$ , we can define an estimator of  $b_0$  by

$$\hat{b}_{m,s} = \arg \min_{b>0} \widehat{\Psi}_{m,s}(b), \quad \widehat{\Psi}_{m,s}(b) = \sum_{r=0}^m \dot{\widehat{G}}_s^r(b) \quad (5)$$

where  $\dot{\widehat{G}}_s^r(b) = \frac{d}{db} \widehat{G}_s^r(b)$  and

$$\widehat{G}_s^r(b) = \begin{cases} e^{\hat{\rho}b} & r = 0 \\ \int_0^b e^{\hat{\rho}z} \widehat{g}_s^{*r}(b-z) dz & r \geq 1 \end{cases}.$$

**Remark 3.** Note that the indifferentiable points for  $\widehat{G}_s^r$  (i.e.,  $\sum_{j=1}^{r-1} Y_i^{(\ell)}$  for  $r \geq 2, \ell = 1, \dots, s$ ) are countable, which implies that the Lebesgue measure of the set

$$\left\{ b \in \mathbb{R}^+ \equiv (0, \infty) \mid \widehat{G}_s^r(b) \text{ is indifferentiable under the given } \{(X_i, \Delta_i T); i = 1, \dots, n\} \right\}$$

converges to 0 with probability 1. Therefore, there exists  $\hat{b}_{m,s}$  almost surely.

Theoretically, we can show that the proposed estimators are consistent as the sample size  $n$  tends to infinity in next section.

### 3.2 Consistency

We first introduce some notation. Let  $\Theta = \{(\lambda, F) : \lambda > 0, F \in \mathcal{F}\}$  be a infinite dimal parameter space where

$$\mathcal{F} = \left\{ F \mid F : \mathbb{R}^+ \mapsto [0, 1], \text{ there exists } f = \frac{dF}{dx}, F(0) = 1, \text{ and } f \text{ is satisfied with Assumptions (A1) - (A5)} \right\}.$$

Denote the distribution function of the claim amount process  $\{X_i\}$  by  $F_0$  and the intensity parameter of the interclaim time process  $\{\Delta_i T\}$  by  $\lambda_0$  (i.e.,  $\{\Delta_i T\} \stackrel{i.i.d.}{\sim} \text{Exp}(1/\lambda_0)$ ). We then suppose that the true parameter  $\theta_0 := (\lambda_0, F_0)$  is included in the parameter space (i.e.,  $\theta_0 \in \Theta$ ). For a fixed  $\theta = (\lambda, F) \in \Theta$  and given  $\delta, c > 0$ , let  $\rho \equiv \rho(\theta)$  be a unique positive root of  $\ell(\xi) \equiv \ell(\xi; \theta)$  defined in (1). We write  $g(x; \theta) = \frac{\lambda}{c} \int_0^\infty e^{-\rho(\theta)y} dF(x+y)$  and  $g^{*r}(\cdot; \theta)$  is the  $r$ -fold convolution of  $g(\cdot; \theta)$ . Denoting

$$\Psi(b; \theta) = \sum_{r=0}^{\infty} \dot{G}^r(b; \theta)$$

where  $\dot{G}^r(b; \theta) = \frac{\partial}{\partial b} G^r(b; \theta)$  and

$$G^r(b; \theta) = \begin{cases} e^{\rho(\theta)b} & r = 0 \\ \int_0^b e^{\rho(\theta)z} g^{*r}(b-z; \theta) dz & r \geq 1 \end{cases},$$

we can define  $b_0$  as a minimizer of  $\Psi(b) \equiv \Psi(b; \theta_0)$ , i.e.,

$$b_0 = \arg \min_{b>0} \Psi(b; \theta_0).$$

Based on  $(X_1, \Delta_1 T), \dots, (X_n, \Delta_n T)$ , we can construct an estimator  $\hat{\theta} \equiv (\hat{\lambda}, \hat{F})$  by (4). Then, we can identify  $\hat{b}_{m,s} \equiv b_{m,s}(\hat{\theta})$  and  $\hat{\Psi}_{m,s}(b) \equiv \Psi_{m,s}(b; \hat{\theta})$  in (5) where  $b_{m,s}(\cdot) = \arg \min_{b>0} \Psi(b; \cdot)$ . Similarly to

[Lu et al., 2007], in order to prove the consistency of  $\hat{b}_{m,s}$ , we need to check the following conditions under Assumptions (A1)-(A5).

**Lemma 4.** *Under Assumptions (A1)-(A5), we have the followings:*

$$(i) \quad \Psi(b_0; \theta_0) = \inf_{b>0} \Psi(b; \theta_0).$$

$$(ii) \quad \Psi_{m,s}(\hat{b}_{m,s}; \hat{\theta}) = \inf_{b>0} \Psi_{m,s}(b; \hat{\theta}) + o_p(1).$$

(iii) *For all  $\delta > 0$ , there exists  $\epsilon(\delta) > 0$  such that*

$$\inf_{|b-b_0|>\delta} \Psi(b; \theta_0) \geq \Psi(b_0; \theta_0) + \epsilon(\delta).$$

(iv) *Uniformly for all  $b > 0$ ,  $\Psi(b; \theta) \equiv \Psi(b; \lambda, F)$  is continuous with respect to  $\lambda$  and the metric  $\|\cdot\|_\infty$  in  $\Theta$  at  $\theta_0 = (\lambda_0, F_0)$  where*

$$\|F - F_0\|_\infty = \sup_{x \in \mathbb{R}^+} |F(x) - F_0(x)|.$$

$$(v) \quad |\hat{\lambda} - \lambda_0| + \|\hat{F} - F_0\|_\infty = o_p(1).$$

(vi) *For all  $\{\delta_n\}$  with  $\delta_n = o(1)$ ,*

$$\sup_{b>0} \sup_{|\lambda-\lambda_0|+\|F-F_0\|_\infty \leq \delta_n} |\Psi_{m,s}(b; \theta) - \Psi(b; \theta)| = o_p(1).$$

**Proof of Lemma 4** Since  $\Psi(b; \theta_0) \propto h'(b)$  and  $h'$  is strictly convex on the neighbourhood of  $b_0$  under Assumption 2 (see Remark 2), (i) and (iii) are satisfied. As  $\hat{b}_{m,s}$  is the minimizer of  $\Psi_{m,s}(b; \hat{\theta})$ , (ii) hold obviously. As for (iv), it is easy to see that  $\Psi(b; \theta) \equiv \Psi(b; \lambda, F)$  is differentiable with respect to  $\lambda$ , which implies that  $\Psi$  is continuous with respect to  $\lambda$ . Put a sufficiently small  $\epsilon > 0$ . Let  $F_\epsilon \in \mathcal{F}$  denotes a distribution function satisfying  $\|F_\epsilon - F_0\|_\infty = \epsilon$ . Based on the  $F_\epsilon$ , we obtain  $\rho_\epsilon$  as a positive root of  $\ell_\epsilon(\xi) \equiv \delta + \lambda(1 - \mathcal{L}_\epsilon(\xi)) - c\xi = 0$  defined in (1), where  $\mathcal{L}_\epsilon(\xi) = \int_0^\infty e^{-\xi x} dF_\epsilon(x)$ . Then, there exists  $m_1$  ( $0 < m_1 < \infty$ ) such that  $|\rho_\epsilon - \rho_0| \leq m_1 \epsilon$  where  $\rho_0 \equiv \rho(\theta_0)$  (see Appendix A.1.). Let  $g_\epsilon^{*r}$  and  $G_\epsilon^r$  denote  $g^{*r}$  and  $G^r$  defined in Lemmas 2 and 3 under  $\theta_\epsilon = (\lambda, F_\epsilon)$ . The above result implies that there exist  $\{m_{2r}\}$  and  $\{m_{3r}\}$  such that  $|g_\epsilon^{*r} - g_0^{*r}| \leq m_{2r} \epsilon$  with  $\sum_r m_{2r} < \infty$  (see Appendix A.2.) and  $|G_\epsilon^r - G_0^r| \leq m_{3r} \epsilon$  with  $\sum_r m_{3r} < \infty$  (see Appendix A.3.) . Therefore, there exists  $0 < M < \infty$  such that

$$\begin{aligned} |\Psi(b; \lambda, F_\epsilon) - \Psi(b; \lambda, F_0)| &\leq \sum_{r=0}^{\infty} \left| \dot{G}^r(b; \lambda, F_\epsilon) - \dot{G}^r(b; \lambda, F_0) \right| \\ &= \sum_{r=0}^{\infty} |\rho_\epsilon G^r(b; \lambda, F_\epsilon) - \rho_0 G^r(b; \lambda, F_0)| + \sum_{r=1}^{\infty} |g^{*r}(b; \lambda, F_\epsilon) - g^{*r}(b; \lambda, F_0)| \\ &\leq M \epsilon \end{aligned}$$

which implies that  $\Psi$  is continuous with respect to the metric  $\|\cdot\|_\infty$ . As for (v), from the law of large number (LLN), it can be seen  $|\hat{\lambda} - \lambda_0| = o_p(1)$  and the Glivenko-Cantelli theorem gives

$\|\hat{F} - F_0\|_\infty \xrightarrow{a.s.} 0$  (see, e.g. Theorem 19.1 of [van der Vaart, 1998]). As for (vi), letting  $\delta_n = o(1)$  and  $|\lambda - \lambda_0| + \|F - F_0\|_\infty \leq \delta_n$ , we notice that

$$\begin{aligned} |\Psi_{m,s}(b; \theta) - \Psi(b; \theta)| &= \Psi_{m,s}(b; \theta) - \Psi_{m,s}(b; \theta_0) \\ &\quad + \Psi_{m,s}(b; \theta_0) - \Psi(b; \theta_0) \\ &\quad + \Psi(b; \theta_0) - \Psi(b; \theta) \\ &= I + II + III \quad (\text{say}) \end{aligned}$$

where by condition (iv)  $III = o(1)$ , uniformly for  $b > 0$  and with  $\theta = (\lambda, F)$  satisfying  $|\lambda - \lambda_0| + \|F - F_0\|_\infty \leq \delta_n$ . That  $I = o_p(1)$ , uniformly for  $b > 0$  and  $\theta = (\lambda, F)$  with  $|\lambda - \lambda_0| + \|F - F_0\|_\infty \leq \delta_n$ , can be proved in the same way as for  $III$ , because in fact  $E[I] = E[\Psi_{\infty,s}(b; \theta) - \Psi_{\infty,s}(b; \theta_0)] + o(1) = III + o(1)$ . In addition,  $II = o_p(1)$  can be also proved easily.  $\square$

Under the above conditions, we have the following result.

**Theorem 1.** *Under Assumptions (A1)-(A5), it follows*

$$\hat{b}_{m,s} - b_0 = o_p(1).$$

**Proof of Theorem 1** The proof is similar to that of Lemma 4.1 in [Lu et al., 2007]. By condition (iii) of Lemma 4, for all  $\delta > 0$ ,

$$P(|\hat{b}_{m,s} - b_0| > \delta) \leq P\left\{\Psi(\hat{b}_{m,s}; \theta_0) - \Psi(b_0; \theta_0) \geq \epsilon(\delta)\right\}, \quad (6)$$

hence it suffices to show that

$$\Psi(\hat{b}_{m,s}; \theta_0) - \Psi(b_0; \theta_0) = o_p(1).$$

Note that

$$\begin{aligned} \Psi(\hat{b}_{m,s}; \theta_0) - \Psi(b_0; \theta_0) &= \Psi(\hat{b}_{m,s}; \theta_0) - \Psi(\hat{b}_{m,s}; \hat{\theta}) \\ &\quad + \Psi(\hat{b}_{m,s}; \hat{\theta}) - \Psi_{m,s}(\hat{b}_{m,s}; \hat{\theta}) \\ &\quad + \Psi_{m,s}(\hat{b}_{m,s}; \hat{\theta}) - \Psi(b_0; \theta_0) \\ &= L_1 + L_2 + L_3 \quad (\text{say}). \end{aligned}$$

From conditions (iv) and (v) of Lemma 4, we have  $L_1 = o_p(1)$ . Conditions (v) and (vi) of Lemma 4 imply

$$|L_2| \leq \sup_{b>0} |\Psi(b; \hat{\theta}) - \Psi_{m,s}(b; \hat{\theta})| = o_p(1),$$

because for any  $b > 0$  and  $\epsilon > 0$  there exists  $\{\delta'_n\}$  with  $\delta'_n/\delta_n = o(1)$ ,

$$\begin{aligned} P\left\{|\Psi(b; \hat{\theta}) - \Psi_{m,s}(b; \hat{\theta})| > \epsilon\right\} &\leq P\left\{\sup_{\theta=(\lambda, F) \in \{|\lambda - \lambda_0| + \|F - F_0\|_\infty \leq \delta_n\}} |\Psi(b; \hat{\theta}) - \Psi_{m,s}(b; \hat{\theta})| > \epsilon\right\} \\ &\quad + P\left\{|\hat{\lambda} - \lambda_0| + \|\hat{F} - F_0\|_\infty > \delta_n\right\} \leq \frac{\delta'_n}{\delta_n} = o(1). \end{aligned}$$

Finally, we have to show  $L_3 = o_p(1)$ . As  $\Psi_{m,s}(\hat{b}_{m,s}; \hat{\theta}) = \inf_{b>0} \Psi_{m,s}(b; \hat{\theta})$ , note that

$$\begin{aligned} \Psi_{m,s}(b; \hat{\theta}) &= \left\{\Psi_{m,s}(b; \hat{\theta}) - \Psi(b; \hat{\theta})\right\} + \left\{\Psi(b; \hat{\theta}) - \Psi(b; \theta_0)\right\} + \Psi(b; \theta_0) \\ &\leq \sup_{b>0} |\Psi_{m,s}(b; \hat{\theta}) - \Psi(b; \hat{\theta})| + \sup_{b>0} |\Psi(b; \hat{\theta}) - \Psi(b; \theta_0)| + \Psi(b; \theta_0) \\ &= R_1 + R_2 + \Psi(b; \theta_0) \quad (\text{say}). \end{aligned}$$



Then we have

$$\Psi_{m,s}(\hat{b}_{m,s}; \hat{\theta}) \leq R_1 + R_2 + \inf_{b>0} \Psi(b; \theta_0) = R_1 + R_2 + \Psi(b_0; \theta_0).$$

It follows, from conditions (v) and (vi) of Lemma 4 that  $R_1 = o_p(1)$ , and from conditions (iv) and (v) of Lemma 4 that  $R_2 = o_p(1)$ , and thus we deduce for any  $\epsilon > 0$

$$\mathbb{P} \left\{ \Psi_{m,s}(\hat{b}_{m,s}; \hat{\theta}) \leq \epsilon + \Psi(b_0; \theta_0) \right\} = o(1),$$

and by exchanging  $\Psi_{m,s}(\hat{b}_{m,s}; \hat{\theta})$  and  $\Psi(b_0; \theta_0)$

$$\mathbb{P} \left\{ \Psi(b_0; \theta_0) \leq \epsilon + \Psi_{m,s}(\hat{b}_{m,s}; \hat{\theta}) \right\} = o(1),$$

which implies  $L_3 = o_p(1)$ . Therefore, it can be seen that  $\Psi(\hat{b}_{m,s}; \theta_0) - \Psi(b_0; \theta_0) = o_p(1)$  and the proof is complete by (6).  $\square$

## 4 Numerical Study

In this section, we first discuss the finite-sample accuracy of our proposed estimator by some simulation studies. Then, we will analyze the optimal dividend barrier estimation using a real data set.

### 4.1 Simulation Study

Suppose that  $\{X_i\}$  and  $\{\Delta_i T\}$  are sequences of i.i.d. positively valued random variables with the probability density functions  $f(x) = e^{-x}$  and  $p(x) = e^{-x}$  (i.e., Exponential distribution with mean 1), respectively. We set the premium rate as  $c = 1.3$  and the interest rate as  $\delta = 0.01$ . In our estimation method, an estimator ( $\hat{g}$ ) of  $g$  is constructed based on the sample  $\{(X_i, \Delta_i T), i = 1, \dots, n\}$ , which implies that the distribution function  $F_Y(y) := \int_0^y \tilde{g}(z) dz$  is replaced by the empirical distribution function  $\hat{F}_Y(y) := \int_0^y \hat{\tilde{g}}(z) dz$  for any  $y > 0$ . Furthermore, the empirical distribution function  $\hat{F}_Y(y)$  is replaced by the Monte Carlo based distribution function  $\hat{F}_Y^{(j,s)}(y) := \frac{1}{s} \sum_{\ell=1}^s \mathbb{I}_{\{Y_j^{(\ell)} \leq y\}}$  for any  $y > 0$ . Therefore, we first investigate the differences between  $F_Y(y)$ ,  $\hat{F}_Y$  and  $\hat{F}_Y^{(j,s)}$  for sample size  $n = 10, 1000$  and Monte Carlo sample size  $s = 100, 10000$  in Figure 1.

**Figure 1 is about here.**

It is easy to see that the graph of empirical distribution function  $\hat{F}_Y$  approaches the graph of true distribution function  $F_Y$  as the sample size  $n$  increases. On the other hand, the graph of the Monte Carlo based distribution function  $\hat{F}_Y^{(j,s)}$  approaches the graph of  $\hat{F}_Y$  as the Monte Carlo sample size  $s$  increases. Summarizing the above, if both the sample size  $n$  and the Monte Carlo sample size  $s$  are sufficiently large, it can be seen that the graph of  $\hat{F}_Y^{(j,s)}$  approaches the graph of  $F_Y$ .

Next, we investigate the difference between the graphs of true objective function  $\Psi$  and its estimator  $\hat{\Psi}_{m,s}$  for some  $m$  in Figure 2. The figure plots the graph of  $\Psi$  under  $f(x) = e^{-x}, \lambda = 1, c = 1.3, \delta = 0.01$ , and the graph of  $\hat{\Psi}_{m,s}$  for  $m = 1, 3, 5$ , and 10, under the sample size  $n = 1000$  and the Monte Carlo sample size  $s = 2000$ . Furthermore, the smallest  $b$ 's of  $\Psi$  and  $\hat{\Psi}_{m,s}$  for  $m = 1, 3, 5$ , and 10 are plotted with dotted lines.

**Figure 2 is about here.**

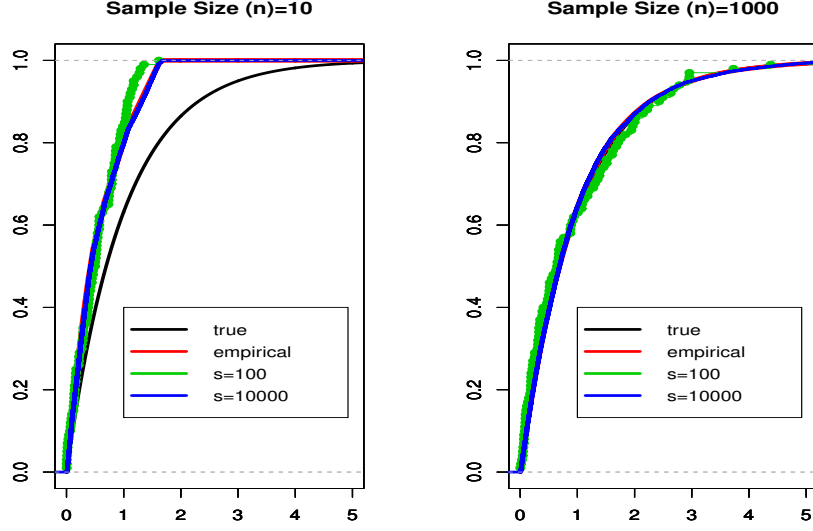


Figure 1: Graphs for the true distribution function  $F_Y(y)$  (black), the empirical distribution function  $\hat{F}_Y$  (red) and the Monte Carlo based distribution function  $\hat{F}_Y^{(j,s)}$  (green, blue) for sample size  $n = 10$  (left) and  $n = 1000$  (right)

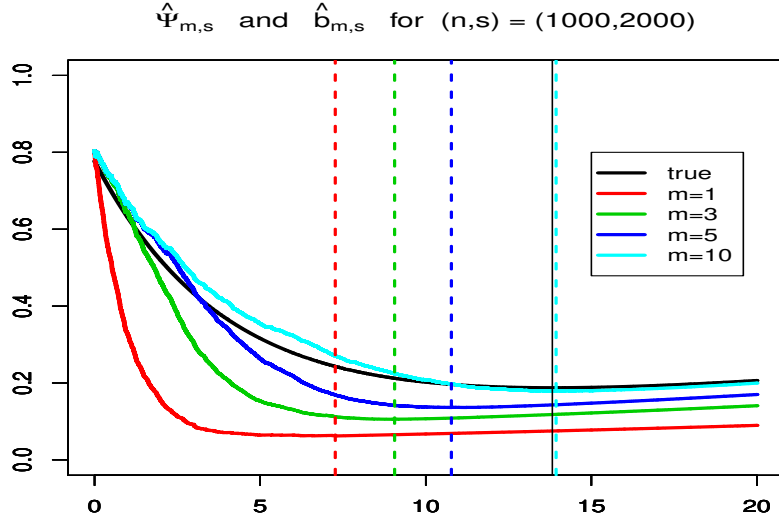


Figure 2: Graphs for the true objective function  $\Psi$  (black) and its estimator  $\hat{\Psi}_{m,s}$  for  $m = 1$  (red),  $m = 3$  (green),  $m = 5$  (blue), and  $m = 10$  (light blue), under the sample size  $n = 1000$  and the Monte Carlo sample size  $s = 2000$ . Dotted lines show their minimizer, respectively.

In the case of  $m = 1$ , there is a large gap between the true objective function  $\Psi$  and  $\hat{\Psi}_{m,s}$ , but as  $m$  increases, the estimated objective function  $\hat{\Psi}_{m,s}$  approximates to  $\Psi$  uniformly. Also along with this, the minimum value (i.e.,  $\hat{b}_{m,s} = \arg \min_{b>0} \hat{\Psi}_{m,s}(b)$ ) also converges to true  $b_0 = \arg \min_{b>0} \Psi(b)$  as  $m$  increases. From this numerical study, under the situation that the sample size is sufficiently large, it can be seen that ignoring the influence of  $m$  above a certain size does not affect the behavior of the objective function. Theoretically, it is known that  $m$  should satisfy a condition that  $\log n/m$  converges to 0. Note that the estimated objective function  $\hat{\Psi}_{m,s}$  is not continuous and the jump exists in some places, and the number of the jumps depends on the Monte Carlo sample size  $s \equiv s(n)$ , so that it increases as the sample size  $n$  increases, but conversely the jump size decreases as the sample size  $n$  increases. As a result,  $\hat{\Psi}_{m,s}$  tends to be a continuous function.

Finally, we calculate the histogram of the optimal dividend barrier estimator  $\hat{b}_{m,s}$  in Figure 3. In this figure, in the case that the sample size  $n = 10, 100, 1000$ ,  $s = 1000$  and  $m = 10$ , the estimated value is repeatedly calculated 100 times, and a histogram is calculated.

**Figure 3 is about here.**

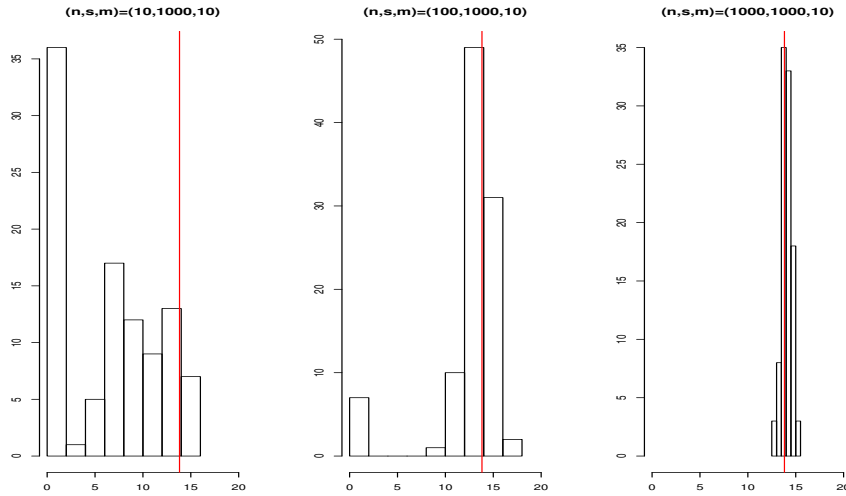


Figure 3: Histograms of  $\hat{b}_{m,s}$  for the sample size  $n = 10$  (left),  $n = 100$  (middle),  $n = 1000$  (right),  $s = 1000$  and  $m = 10$  with 100 times iteration. Red lines show the magnitude of true optimal dividend barrier.

Obviously, as the sample size increases, the optimal dividend barrier estimator  $\hat{b}_{m,s}$  converges to the true optimal dividend barrier  $b_0$ . In this study, in order to see the influence of the sample size  $n$ ,  $s$  and  $m$  are fixed, but both are set so as to be larger depending on the sample size  $n$ , and it can be expected to appear similar phenomenon even if  $s$  and  $m$  depend on  $n$ . Clearly, under this setting, our proposed estimation method works well.

## 4.2 Real Data Analysis

We next investigate the empirical performance of our proposed estimator by using a real data set. The data is a set of Danish Fire Insurance Claims in Denmark from Thursday 3rd January 1980 until Monday 31st December 1990 (This data set is contained in R-Package ‘evir’). In this

data set, numeric vectors of “claim amount” and “claim reported date” involve 2167 observations. Since the maximum number of the claim within 1 day equals 6, we divide 1 day into 6 intervals as  $((j-1)/6, j/6]$  for  $j = 1, 2, 3, 4, 5, 6$  and we assume that the claim reported hours is uniformly distributed in the each intervals without overlapping. We then write each claim reported date and hours as  $T_i$  and the interclaim time as  $\Delta_i T = T_i - T_{i-1}$ . For instance, suppose that there exist two claims in a date and let the claim reported times (i.e., date and hours) be  $T_{i_1}, T_{i_1+1}$  ( $T_{i_1} < T_{i_1+1}$ ), respectively. Then, the interclaim time ( $\Delta_{i_1} T = T_{i_1+1} - T_{i_1}$ ) is within  $\{1/6, 2/6, 3/6, 4/6, 5/6\}$  and  $P(\Delta_{i_1} T = j/6) = \binom{6-j}{1} / \binom{6}{2}$  for  $j = 1, 2, 3, 4, 5$ . We plot the histograms for the sequence of claim amount  $\{X_i\}$  and  $\{\Delta_i T\}$  in Figure 4.

**Figure 4 is about here.**

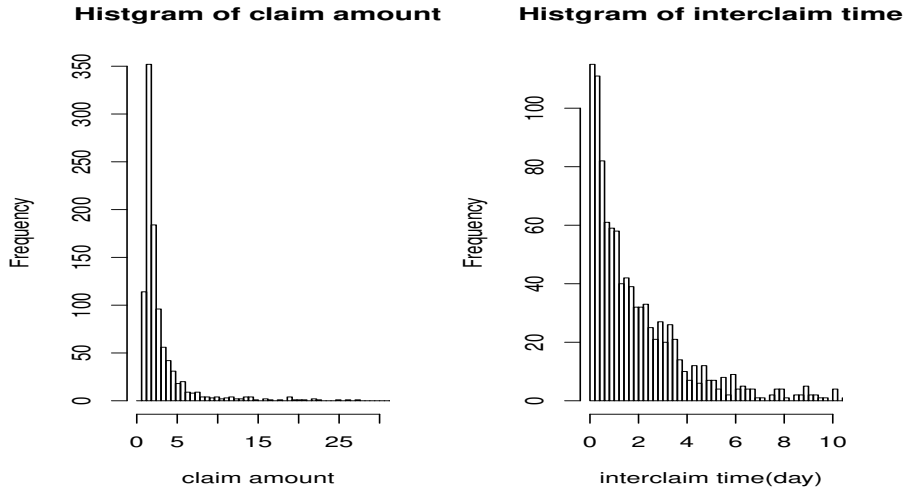


Figure 4: Histogram of claim amount ( $X_i$ ) and interclaim time ( $\Delta_i T$ )

As for the claim amount ( $X_i$ ), one may see that the volume of most accidents is less than 5. However, although it is rare, there seem some outliers such as 10 and 20. In view of this, it could be danger to simply assume a parametric model such as exponential distribution. But on the other hand, the frequency of interclaim time ( $\Delta_i T$ ) decreases smoothly in the range of almost less than 10. Therefore, it seems not a problem if applying the exponential distribution.

Next, we divide this data set into the “Training Data” for the first 1000 observations and the “Test Data” for the remaining 1167 observations. By using the “Training Data”, we construct the optimal dividend barrier estimator  $\hat{b}_{m,s}$ , under the premium rate as  $c = 2 \times \bar{X} \times \hat{\lambda} = 2.354472$ , the interest rate as  $\delta = 0.01$  and the initial surplus as  $u = 10$ . In order to construct the estimator, we set  $s = 2000$  and  $m = 10$ . As a result, we obtain  $\hat{b}_{m,s} = 33.429$  with  $\hat{\Psi}_{m,s}(b)$  displayed in Figure 5.

**Figure 5 is about here.**

Please note that even if we do not know the probability density function  $f$  and its Laplace transform  $g$ , we can generate  $\{Y_i^{(\ell)}\}$  from the empirical version  $\hat{g}$ . Since the  $\hat{g}$  can be constructed based on the sample claim amount  $\{X_1, \dots, X_n\}$  and sample interclaim time  $\{\Delta_1 T, \dots, \Delta_n T\}$ , the optimal dividend barrier estimator  $\hat{b}_{m,s}$  can be constructed without any other information

about functions  $f$  and  $g$  except for the information from the observation. In addition, since the computation time of  $\hat{b}_{m,s}$  based on the “Training Data” is 77.88 second by **R**, anyone could be implement it. The above results show that our estimation can be used in practical point of view.

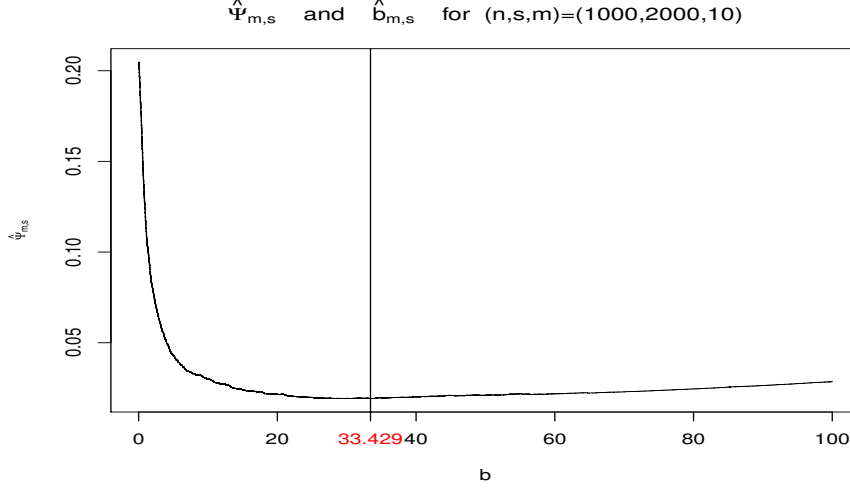


Figure 5: Graph of  $\hat{\Psi}_{m,s}(b)$  and its minimizer  $\hat{b}_{m,s} = 33.429$

The graph shows the convexity of  $\hat{\Psi}_{m,s}(b)$  and the uniqueness of the minimizer. Note that the objective function  $\hat{\Psi}_{m,s}(b)$  is not continuous but tends to be a continuous function as the  $s \equiv s(n)$  is increasing.

Based on the “Test Data”, we can evaluate the accuracy of the optimal dividend barrier estimator in view of the magnitude of the expected discounted dividends

$$V(u, b) = E[AD_b], \quad AD_b = \int_0^{T_b} e^{-\delta t} dD_b(t).$$

Since the optimal dividend barrier  $b_0$  maximizes  $V$ , the estimator  $\hat{b}_{m,s}$  should be close to the maximizer of  $V$  (i.e.,  $V(u, \hat{b}_{m,s}) \approx \max_b V(u, b) = V(u, b_0)$ ). Before the calculation of  $V(u, b)$ , we first need to consider how to calculate  $AD_b$  under the fixed  $u, b$  and the given  $\{(X_i, \Delta_i T); i = 1, \dots, n\}$ . For that purpose, construction of a sample path of  $\{U_b(t)\}$  defined by (2) is needed and  $U_b(t)$  can be rewritten as follows:

$$U_b(t) = \begin{cases} \min \{U_{b,0}^+ + ct, b\} & \text{if } 0 \leq t < T_1 \\ \min \{U_{b,1}^+ + c(t - T_1), b\} & \text{if } T_1 \leq t < T_2 \\ \vdots & \vdots \\ \min \{U_{b,i-1}^+ + c(t - T_{i-1}), b\} & \text{if } T_{i-1} \leq t < T_i \\ \vdots & \vdots \\ \min \{U_{b,N(T_b)-1}^+ + c(t - T_{N(T_b)-1}), b\} & \text{if } T_{N(T_b)-1} \leq t < T_{N(T_b)} \equiv T_b \end{cases}$$

where  $U_{b,i}^+$  (and  $U_{b,i}^-$ ) is the surplus immediately after (and before) the  $i$ th claim has occurred, defined by

$$U_{b,i}^+ = \begin{cases} u & \text{if } i = 0 \\ U_{b,i}^- - X_i & \text{if } i = 1, 2, \dots, N(T_b) \end{cases}, \quad U_{b,i}^- = \lim_{t \uparrow T_i} U_b(t),$$

and  $T_i$  is the aggregated time ( $T_i = \sum_{k=1}^i \Delta_k T$ ) with  $T_0 = 0$  and  $N(T_b)$  is the number of claims up to the ruin time  $T_b = \inf\{t | U_b(t) < 0\}$ . The sample path of  $\{U_b(t)\}$  is shown in Figure 6 based on  $b = 33.429 (\equiv \hat{b}_{m,s})$  and the “Test Data”.

Figure 6 is about here.

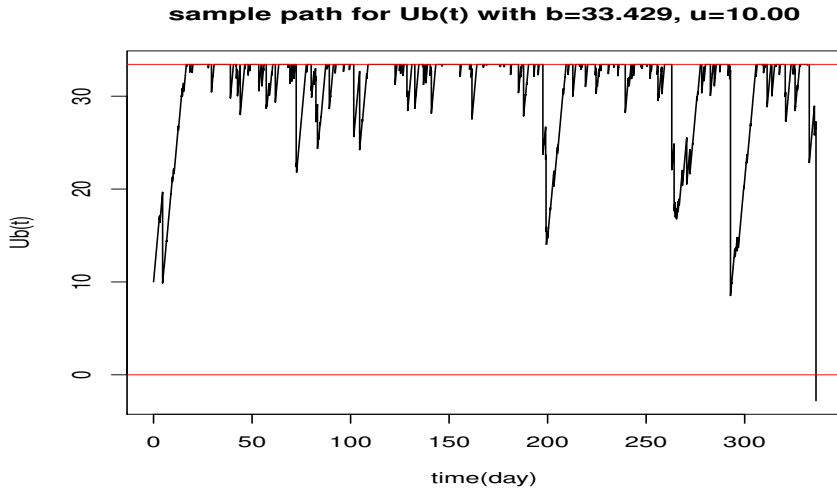


Figure 6: Sample path of  $\{U_b(t)\}$  based on  $b = 33.429 (\equiv \hat{b}_{m,s})$

At time  $t = 0$ , we have  $U_b(0) = u = 10$  and the ruin time (i.e.,  $T_b = \inf\{t | U_b(t) < 0\}$ ) is  $T_b = 336.2$ . Based on the sample path of  $\{U_b(t)\}$ , we can obtain a sequence of dividend payable terms  $\{(\tau_i, T_i); i = 1, 2, \dots, N(T_b)\}$ , where

$$\tau_i = \min \left\{ T_{i-1} + \frac{b - U_{b,i-1}^+}{c}, T_i \right\},$$

which implies that the discounted aggregated dividends

$$AD_b = \sum_{i=1}^{N(T_b)} \int_{T_{i-1}}^{T_i} e^{-\delta t} dD_b(t) = \sum_{i=1}^{N(T_b)} \int_{\tau_i}^{T_i} e^{-\delta t} (c dt) = \frac{c}{\delta} \sum_{i=1}^{N(T_b)} (e^{\tau_i} - e^{T_i}).$$

In the case of the “Test Data”, we obtain the discounted aggregated dividends  $AD_b = 116.1224$ .

Unfortunately, at this stage, we can not evaluate the goodness of our estimator, because the optimal dividend barrier  $b$  is a maximizer of the expected value of  $AD_b$ , not  $AD_b$  itself. In order to calculate (an approximate value of) the expected  $AD_b$ , it is necessary to generate a number of sample paths. In what follows, we consider generating some sample paths from the “Test Data”

using a resampling method. Denote the “Test Data” by  $\{(X_i, \Delta_i T); i = 1, \dots, 1167\}$ . Then, we generate 1000 sample paths  $\{\chi_\ell; \ell = 1, \dots, 1000\}$  by

$$\chi_\ell = \left\{ (X_{i_j^{(\ell)}}, \Delta_{i_j^{(\ell)}} T); i_j^{(\ell)} \in \{1, \dots, 1167\}, j = 1, \dots, 1167 \text{ without overlapping} \right\}.$$

Calculating  $\{U_b(t)\}$  for each sample path  $\chi_\ell$  generated above, we obtain the values of the ruin time  $T_b$  and the discounted aggregated dividends  $AD_b$  in the same way as above. In Figure 7, the histograms of  $T_b$  and  $AD_b$  derived from 1000 sample paths are plotted.

**Figure 7 is about here.**

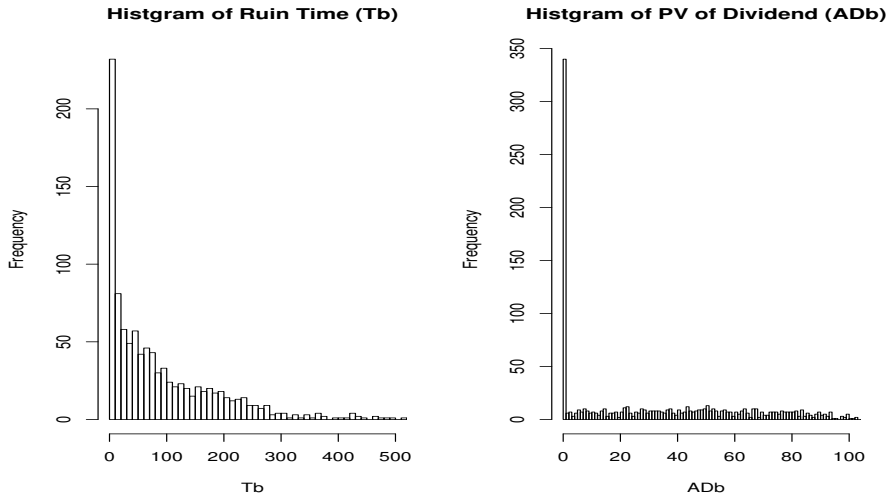


Figure 7: Histogram of  $T_b$  and  $AD_b$  for 1000 sample paths under  $b = 33.429(\equiv \hat{b}_{m,s})$

From the histogram of the ruin time  $T_b$ , we can see that in some cases, the ruin occurs at around time zero. On the other sample paths, the frequency of the ruin time slowly decreases, and there also exist some cases that the ruin time exceeds 500 days. As a distribution, it follows a heavy-tailed distribution. As for the histogram of  $AD_b$ , in most cases, the ruin occurs before the dividend payout (that is,  $AD_b = 0$ ). When  $AD_b \neq 0$ , it seems to be distributed uniformly over a wide range although it is very infrequent.

Finally, 1000 values of  $T_b$  and  $AD_b$  are derived for various  $b$  (such as  $b = 33.429(\equiv \hat{b}_{m,s})$  and  $b = 10, 15, \dots, 85, 90$ ) in the same way as above, see Table 1 for the comparison of their means.

Obviously, the mean of  $T_b$  increases as  $b$  increases. On the other hand, when looking at the mean of  $AD_b$ , it can be seen that the function (denote  $\overline{AD}_b$ ) is a concave function with respect to  $b$  and the maximizer is close to the optimal dividend barrier estimator  $b = 33.429$ , which implies that

$$\overline{AD}_{\hat{b}_{m,s}} \approx V(u, \hat{b}_{m,s}) \approx \max V(u, b) = V(u, b_0).$$

This phenomenon illustrates our proposed estimator works well.

**Acknowledgments** We thank the anonymous referees for constructive comments. This work was supported by JSPS KAKENHI Grant Number JP16K00036.

Table 1: Comparison of  $T_b$  and  $AD_b$

b	mean of $T_b$	mean of $AD_b$	b	mean of $T_b$	mean of $AD_b$
33.429	120.6574	58.12917	50.00	318.6832	56.38832
10.00	32.2352	40.89303	55.00	355.2182	54.47957
15.00	46.6110	45.38146	60.00	386.3034	52.65691
20.00	69.9586	51.45603	65.00	417.4244	50.55746
25.00	91.1134	53.36816	70.00	439.3756	48.46212
30.00	124.7152	56.04076	75.00	455.5426	46.33621
35.00	171.2750	58.02137	80.00	465.0066	44.26150
40.00	205.2788	57.74939	85.00	469.6726	42.21538
45.00	250.1554	57.29334	90.00	470.9046	40.24690

## References

- [Bühlmann, 1970] Bühlmann, H. (1970). *Mathematical methods in risk theory*. Die Grundlehren der mathematischen Wissenschaften, Band 172. Springer-Verlag, New York-Berlin.
- [Croux and Veraverbeke, 1990] Croux, K. and Veraverbeke, N. (1990). Nonparametric estimators for the probability of ruin. *Insurance: Mathematics and Economics*, 9(2):127–130.
- [De Finetti, 1957] De Finetti, B. (1957). Su unfimpostazione alternativa della teoria collettiva del rischio. In *Transactions of the XVth international congress of Actuaries*, volume 2, pages 433–443.
- [Dickson, 2005] Dickson, D. C. M. (2005). *Insurance risk and ruin*. International Series on Actuarial Science. Cambridge University Press, Cambridge.
- [Dickson and Waters, 2004] Dickson, D. C. M. and Waters, H. R. (2004). Some optimal dividends problems. *Astin Bull.*, 34(1):49–74.
- [Gerber, 1979] Gerber, H. U. (1979). *An introduction to mathematical risk theory*, volume 8 of *S.S. Heubner Foundation Monograph Series*. University of Pennsylvania, Wharton School, S.S. Huebner Foundation for Insurance Education, Philadelphia, Pa.; Distributed by Richard D. Irwin, Inc., Homewood, Ill. With a foreword by James C. Hickman.
- [Gerber and Shiu, 1997] Gerber, H. U. and Shiu, E. S. (1997). The joint distribution of the time of ruin, the surplus immediately before ruin, and the deficit at ruin. *Insurance: Mathematics and Economics*, 21(2):129–137.
- [Gerber and Shiu, 1998] Gerber, H. U. and Shiu, E. S. W. (1998). On the time value of ruin. *N. Am. Actuar. J.*, 2(1):48–78. With discussion and a reply by the authors.
- [Gerber and Shiu, 2004] Gerber, H. U. and Shiu, E. S. W. (2004). Optimal dividends: analysis with Brownian motion. *N. Am. Actuar. J.*, 8(1):1–20.
- [Gerber et al., 2006] Gerber, H. U., Shiu, E. S. W., and Smith, N. (2006). Maximizing dividends without bankruptcy. *Astin Bull.*, 36(1):5–23.
- [Gerber et al., 2008] Gerber, H. U., Shiu, E. S. W., and Smith, N. (2008). Methods for estimating the optimal dividend barrier and the probability of ruin. *Insurance Math. Econom.*, 42(1):243–254.



- [Landriault, 2008] Landriault, D. (2008). Constant dividend barrier in a risk model with interclaim-dependent claim sizes. *Insurance Math. Econom.*, 42(1):31–38.
- [Lin et al., 2003] Lin, X. S., Willmot, G. E., and Dreikic, S. (2003). The classical risk model with a constant dividend barrier: analysis of the Gerber-Shiu discounted penalty function. *Insurance Math. Econom.*, 33(3):551–566.
- [Lu et al., 2007] Lu, Z., Tjøstheim, D., and Yao, Q. (2007). Adaptive varying-coefficient linear models for stochastic processes: Asymptotic theory. *Statistica Sinica*, 17(1):177–198.
- [van der Vaart, 1998] van der Vaart, A. W. (1998). *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge.
- [Yin et al., 2015] Yin, C., Yuen, K. C., and Shen, Y. (2015). Convexity of ruin probability and optimal dividend strategies for a general lévy process. *The Scientific World Journal*, 2015.

## Appendix

### A.1 Existence $\{m_1\}$

Suppose that we fix a sufficiently small  $\xi_0 > 0$ . For any  $\xi > \xi_0$ , it follows that

$$|\mathcal{L}_\epsilon(\xi) - \mathcal{L}_0(\xi)| = \left| \int_0^\infty e^{-\xi x} dF_\epsilon(x) - \int_0^\infty e^{-\xi x} dF_0(x) \right| \leq 2\epsilon \int_0^\infty e^{-\xi x} dx \leq \left( \frac{2}{\xi_0} \right) \epsilon, \quad (7)$$

and for any  $\xi_1, \xi_2 > \xi_0$ , it follows that from Tayler expansion

$$\begin{aligned} |\mathcal{L}_0(\xi_1) - \mathcal{L}_0(\xi_2)| &= \left| \int_0^\infty e^{-\xi_1 x} dF_0(x) - \int_0^\infty e^{-\xi_2 x} dF_0(x) \right| = \left| (\xi_1 - \xi_2) \int_0^\infty x e^{-\xi^* x} dF_0(x) \right| \\ &\leq |(\xi_1 - \xi_2)| \int_0^\infty x dF_0(x) = |\xi_1 - \xi_2| \mu, \end{aligned} \quad (8)$$

where  $\xi_1 < \xi^* < \xi_2$ . On that other hand, for any  $\epsilon > 0$  we have from (1)

$$\ell_\epsilon(\rho_\epsilon) = \delta - c\rho_\epsilon + \lambda \{1 - \mathcal{L}_\epsilon(\rho_\epsilon)\} = 0, \quad \ell_0(\rho_0) = \delta - c\rho_0 + \lambda \{1 - \mathcal{L}_0(\rho_0)\} = 0. \quad (9)$$

From (7), (8) and (9), we can write

$$|\rho_\epsilon - \rho_0| = \frac{\lambda}{c} |\mathcal{L}_\epsilon(\rho_\epsilon) - \mathcal{L}_0(\rho_0)| \leq \frac{\lambda}{c} \left\{ \left( \frac{2}{\xi_0} \right) \epsilon + |\rho_\epsilon - \rho_0| \mu \right\},$$

for a sufficiently small  $\epsilon > 0$ , which implies that

$$|\rho_\epsilon - \rho_0| \leq \left( 1 - \frac{\lambda\mu}{c} \right)^{-1} \frac{\lambda}{c} \left( \frac{2}{\xi_0} \right) \epsilon := m_1 \epsilon.$$

Note that under Assumption (A3),  $\frac{\lambda\mu}{c} < 1$  is satisfied and since  $\mathcal{L}_\epsilon$  and  $\mathcal{L}_0$  are continuous, there exists  $\xi_0$  such that  $\min(\rho_0, \rho_\epsilon) > \xi_0$ .  $\square$

## A.2 Existense $\{m_{2r}\}$

Similary to Appendix A.1., for any  $x > 0$ , there exist  $0 < m_{21} < \infty$  such that

$$|g_\epsilon^{*1}(x) - g_0^{*1}(x)| = |g_\epsilon(x) - g_0(x)| = \frac{\lambda}{c} \left| \int_0^\infty e^{-\rho_\epsilon y} dF_\epsilon(x+y) - \int_0^\infty e^{-\rho_0 y} dF_0(x+y) \right| \leq m_{21}\epsilon,$$

and

$$\begin{aligned} |g_\epsilon^{*2}(x) - g_0^{*2}(x)| &= \left| \int_0^x g_\epsilon(y)g_\epsilon(x-y)dy - \int_0^x g_0(y)g_0(x-y)dy \right| \\ &\leq \left| \int_0^x g_\epsilon(y) \{g_\epsilon(x-y) - g_0(x-y)\} dy \right| + \left| \int_0^x \{g_\epsilon(y) - g_0(y)\} g_0(x-y)dy \right| \\ &\leq \left\{ \int_0^\infty g_\epsilon(y)dy + \int_0^\infty g_0(y)dy \right\} m_{21}\epsilon := m_{22}\epsilon. \end{aligned}$$

Here it can be seen that

$$\begin{aligned} \int_0^\infty g_\epsilon(y)dy &= \int_0^\infty \frac{\lambda}{c} \int_0^\infty e^{-\rho_\epsilon y} f_\epsilon(x+y) dx dy \\ &= \frac{\lambda}{c} \int_0^\infty \left( \int_0^t e^{-\rho_\epsilon s} ds \right) f_\epsilon(t) dt \quad (t = x+y, s=y) \\ &= \frac{\lambda}{c} \int_0^\infty \left( \frac{1 - e^{-\rho_\epsilon t}}{\rho_\epsilon} \right) f_\epsilon(t) dt \\ &= \frac{\lambda}{c\rho_\epsilon} \{1 - \mathcal{L}(\rho_\epsilon)\} \\ &= \frac{\lambda}{c\rho_\epsilon} \left( 1 - \frac{\delta + \lambda - c\rho_\epsilon}{\lambda} \right) \quad \text{by (1)} \\ &= 1 - \frac{\delta}{c\rho_\epsilon} < 1, \end{aligned}$$

which implies that  $0 < m_{22} < \infty$ . In the same way, it follows that for any  $r \geq 3$

$$\begin{aligned} |g_\epsilon^{*r}(x) - g_0^{*r}(x)| &= \left| \int_0^x g_\epsilon(y)g_\epsilon^{*r-1}(x-y)dy - \int_0^x g_0(y)g_0^{*r-1}(x-y)dy \right| \\ &\leq \left\{ m_{2,r-1} \int_0^\infty g_\epsilon(y)dy + m_{21} \int_0^\infty g_0^{*r-1}(y)dy \right\} \epsilon := m_{2r}\epsilon \end{aligned}$$

with  $0 < m_{2r} < \infty$ . Lemma 2 implies that  $0 < \sum_{r=1}^\infty g_\epsilon^{*r}(b) < \infty, 0 < \sum_{r=1}^\infty g_0^{*r}(b) < \infty$  and there exists  $r_\epsilon \in \mathbb{N}$  such that  $\sum_{r=r_\epsilon}^\infty \{g_\epsilon^{*r}(b) + g_0^{*r}(b)\} < \epsilon$ . Threfore, denoting  $m_{2r} := \{g_\epsilon^{*r}(b) + g_0^{*r}(b)\} / \sum_{r=r_\epsilon}^\infty \{g_\epsilon^{*r}(b) + g_0^{*r}(b)\}$  for  $r \geq r_\epsilon$ , instead of the above definition, we have

$$|g_\epsilon^{*r}(x) - g_0^{*r}(x)| \leq m_{2r}\epsilon$$

and  $0 < \sum_{r=1}^\infty m_{2r} = \sum_{r=1}^{r_\epsilon-1} m_{2r} + \sum_{r=r_\epsilon}^\infty m_{2r} = \sum_{r=1}^{r_\epsilon-1} m_{2r} + 1 < \infty$ .  $\square$

### A.3 Existense $\{m_{3r}\}$

From Appendix A.1., for a fixed  $b > 0$ , there exist  $0 < m_{30} < \infty$  such that  $|G_\epsilon^0(b) - G_0^0(b)| = |e^{\rho_\epsilon b} - e^{\rho_0 b}| \leq m_{30}\epsilon$ . For  $r \geq 1$ , there exists  $0 < m_{3r} < \infty$  such that

$$\begin{aligned} |G_\epsilon^r(b) - G_0^r(b)| &= \left| \int_0^b e^{\rho_\epsilon z} g_\epsilon^{*r}(b-z) dz - \int_0^b e^{\rho_0 z} g_0^{*r}(b-z) dz \right| \\ &\leq \left| \int_0^b (e^{\rho_\epsilon z} - e^{\rho_0 z}) g_\epsilon^{*r}(b-z) dz \right| + \left| \int_0^b e^{\rho_0 z} \{g_\epsilon^{*r}(b-z) - g_0^{*r}(b-z)\} dz \right| \leq m_{3r}\epsilon. \end{aligned}$$

In the same way to Appendix A.2., there exists  $\tilde{r}_\epsilon \in \mathbb{N}$  such that  $\sum_{r=\tilde{r}_\epsilon}^\infty \{G_\epsilon^r(b) + G_0^r(b)\} < \epsilon$ . Therefore, denoting  $m_{3r} := \{G_\epsilon^r(b) + G_0^r(b)\} / \sum_{r=\tilde{r}_\epsilon}^\infty \{G_\epsilon^r(b) + G_0^r(b)\}$  for  $r \geq \tilde{r}_\epsilon$ , instead of the above definition, we have

$$|G_\epsilon^r(b) - G_0^r(b)| \leq m_{3r}\epsilon$$

with  $0 < \sum_{r=0}^\infty m_{3r} < \infty$ . □