

Article

On Computations in Renewal Risk Models—Analytical and Statistical Aspects

Josef Anton Strini  and Stefan Thonhauser * 

Graz University of Technology, Institute of Statistics, Kopernikusgasse 24/III, A-8010 Graz, Austria

* Correspondence: stefan.thonhauser@math.tugraz.at

Received: 13 December 2019; Accepted: 28 February 2020; Published: 4 March 2020



Abstract: We discuss aspects of numerical methods for the computation of Gerber-Shiu or discounted penalty-functions in renewal risk models. We take an analytical point of view and link this function to a partial-integro-differential equation and propose a numerical method for its solution. We show weak convergence of an approximating sequence of piecewise-deterministic Markov processes (PDMPs) for deriving the convergence of the procedures. We will use estimated PDMP characteristics in a subsequent step from simulated sample data and study its effect on the numerically computed Gerber-Shiu functions. It can be seen that the main source of instability stems from the hazard rate estimator. Interestingly, results obtained using MC methods are hardly affected by estimation.

Keywords: risk theory; renewal model; gerber-shiu functions; PIDEs

1. Introduction

In this article we study the computation of Gerber-Shiu functions. These functions—or more precisely functionals—have been established in [Gerber and Shiu \(1998\)](#) in the context of the classical risk model, with the goal to study ruin relevant quantities in an universal manner. Subsequently, the derived results have been extended to the Sparre-Andersen risk model (see [Sparre Andersen \(1957\)](#)) by the same authors; see [Gerber and Shiu \(2005\)](#). We refer to well-established techniques and results for this particular class of risk models: [Schmidli \(2017\)](#) and [Asmussen and Albrecher \(2010\)](#).

Our contribution considers the Sparre-Andersen (or renewal) risk model as a basis to describe the evolution of an insurance portfolio surplus. But we take the perspective of piecewise-deterministic Markov processes (PDMPs) such that in our theoretical treatment we can also incorporate a state-dependent premium rate, which corresponds to a non-constant drift of the surplus process. Hence, we can exploit the theory of PDMPs to study Gerber-Shiu functions and generalizations of them; see Section 2. The by now classical reference on PDMPs is [Davis \(1993\)](#).

Going one step further means literally going *back* in time, since in our approach we are going to incorporate collected data from the surplus process in order to estimate its respective characteristics. Based on these estimators, we in turn determine the desired functional; e.g., the ruin probability. From this perspective—using estimated characteristics—one could say that we are dealing with *doubly stochastic* random quantities; consult [Brémaud \(1981\)](#). Given the non-parametric estimators in our approach, we use a numerical scheme for the computation of expected values associated with PDMPs, in which we discretize the state and time variable in order to solve the related partial-integro-differential equation (PIDE). In providing the basis for this numerical and statistical procedure, we first of all verify that the Gerber-Shiu functional can be identified with the solution of the associated PIDE. In a next step, we show that our approximation converges to the exact functional by showing that the related generators of the PDMPs also converge in an appropriate sense and applying techniques from the theory of Markov processes in addition to results from [Kritzer et al. \(2019\)](#).

The numerical procedure is necessary, since Gerber-Shiu functions in the renewal risk model admit explicit representations only under restrictive assumptions on the distributions of the inter-jump times and claims. Such assumptions are typically violated if estimated characteristics are used. Certainly, in such situations one can apply specialized simulation techniques, either based on Monte Carlo or quasi-Monte Carlo methods. In a risk theoretic context these methods are analyzed for example in [Preischl et al. \(2018\)](#) and [Kritzer et al. \(2019\)](#). Another method, mentioned before, is based on the numerical solution of associated PIDEs. This is built to a great extent upon properties of PDMPs and a general guideline for the design of a numerical schema is illustrated in [Davis \(1993\)](#). In this a specific implementation in a health insurance application can also be found.

In our contribution we use non-parametric statistical estimators of the risk model's characteristics which rely on the particular feature of *iid* inter-jump times and jump sizes; see Section 5. In this context one can also certainly use methods based on techniques from survival analysis. Such an approach is recommended for quite general PDMPs in [Azaïs et al. \(2013, 2014\)](#). In recent years *phase-type* approximations have also been used in risk theory with a similar purpose. In the classical risk model the claims' distribution can be approximated in such a way. The effect of such an approximation on the resulting ruin probability is discussed by [Vatamidou et al. \(2013, 2014\)](#) on the basis of the Pollaczek-Khintchine formula. A similar approach is also feasible in the Sparre-Andersen model; see [Albrecher and Vatamidou \(2019\)](#). In the last references the approximation is used for theoretical distribution functions, but certainly one can approximate empirical distribution functions based on a sample by means of the EM-algorithm; see [Asmussen et al. \(1996\)](#). Another non-parametric approach in the classical risk model works on the level of Laplace or Fourier transforms of the ruin probability or even the Gerber-Shiu function. Several results in this direction are presented by [Shimizu \(2012\)](#) and [Shimizu and Zhang \(2017\)](#), who apply kernel estimators for the claims' distribution and investigate their properties on the level of the transforms.

2. Model Setup

We fix an underlying complete probability space (Ω, \mathcal{F}, P) on which the following probabilistic ingredients are defined. We consider a renewal process $N = (N_t)_{t \geq 0}$ for determining the number of claims that occurred up to and including time t . The inter-arrival times of N are given by a sequence $\{T_i\}_{i \in \mathbb{N}}$ of positive, independent and identically distributed random variables. We assume that their distribution function F_T is absolutely continuous. The jump times $\{\sigma_i\}_{i \in \mathbb{N}}$ of the process are given by $\sigma_i = \sigma_{i-1} + T_i$ for $i \in \mathbb{N}$ and for the moment we set $\sigma_0 = 0$. Notice, that the existence of a density f_T of the distribution of $T \sim T_i$ ensures that a jump intensity $\lambda(\cdot) = \frac{f_T(\cdot)}{F_T(\cdot)}$ of the to be defined surplus process X exists; see [Rolski et al. \(1999, p. 480\)](#).

The claim sizes are modeled by a sequence $\{Y_i\}_{i \in \mathbb{N}}$ of positive *iid* random variables with continuous distribution function F_Y . We assume independence between $\{Y_i\}_{i \in \mathbb{N}}$ and the process N as is commonly done. For the incoming premiums in a time interval $[0, t]$ we use the relatively general term $\int_0^t c(X_{s-}) ds$, i.e., a state dependent premium rate $c(\cdot)$, which is assumed to be positive, bounded and Lipschitz continuous. Note that this is the driving component keeping the particular line of business of the insurer alive. Denoting the initial capital with $x \in \mathbb{R}^+$, we obtain the surplus process $(X_t)_{t \geq 0}$ as a solution to

$$X_t = x + \int_0^t c(X_{s-}) ds - \sum_{i=1}^{N_t} Y_i. \quad (1)$$

This stochastic process generates the filtration $\{\mathcal{F}_t^X\}_{t \geq 0}$ with the available information at time t . In particular, we set $\mathcal{F}_t^X = \sigma\{\mathcal{F}_t^N, \{Y_1, Y_2, \dots, Y_{N_t}\} \cup \mathcal{N}\}$, where \mathcal{F}^N is the filtration generated by the process N . In addition, \mathcal{N} denotes the sets of measure zero from \mathcal{F} .

Since the process X itself is not a Markov process, we will use the technique of *backward Markovization*, see [Rolski et al. \(1999, p. 480\)](#), in order to modify the process X such that we arrive at a piecewise-deterministic Markov process. For an introduction to processes of this kind, see for instance, [Davis \(1993\)](#) or [Rolski et al. \(1999\)](#). This technique basically demands an enlarged state space. In this, we consider the process $\tilde{X}_t = (X_t, t'_t)$, where $t'_t = t - \sigma_{N_t}$ represents the time since the last jump. \tilde{X} is now a time-homogeneous PDMP. In the PDMP setting we see that the deterministic evolution of the component X between jumps is described via the ordinary differential equation $\frac{\partial}{\partial t}y(t) = c(y(t))$, the solution of which incorporating the boundary condition $y(0) = x$ is denoted by $\phi(t, x)$.

We are now able to write down the infinitesimal generator \mathcal{A} of this process; see [Rolski et al. \(1999, pp. 449, 480\)](#) or [Davis \(1993, p. 70\)](#). For an appropriate function $h \in \mathcal{D}(\mathcal{A})$, we have that the generator of \tilde{X} is given by

$$(\mathcal{A}h)(x, s) = c(x)\frac{\partial}{\partial x}h(x, s) + \frac{\partial}{\partial s}h(x, s) + \lambda(s) \left(\int_0^\infty h(x - v, 0) - h(x, s) dF_Y(v) \right). \tag{2}$$

In this context *appropriate* means that the function h is in the domain $\mathcal{D}(\mathcal{A})$ of the generator. Furthermore, we obtain some useful consequences for the process under consideration. Firstly,

$$Y_t = h(\tilde{X}_t) - h(\tilde{X}_0) - \int_0^t \mathcal{A}h(\tilde{X}_{s-}) ds, \quad Y_0 = h(\tilde{X}_0), \tag{3}$$

is Dynkin’s martingale. Secondly, if we can find a function h with $\mathcal{A}h = 0$, we deduce that $h(\tilde{X}_t)$ itself is a martingale. Conditions which ensure that h belongs to $\mathcal{D}(\mathcal{A})$ are stated in [Rolski et al. \(1999, p. 449, Thm. 11.2.2\)](#) and [Davis \(1993, p. 69, Thm. 26.14\)](#). The following theorem restates these results in our framework.

Theorem 1. *Let \tilde{X} be the PMDP defined above. Let $h : \mathbb{R} \times \mathbb{R}_0^+ \rightarrow \mathbb{R}$ be a measurable function such that*

1. *The function $s \mapsto h(\phi(s, x), s)$ is absolutely continuous on $(0, \infty)$,*
2. *$\forall t \geq 0$ it holds that*

$$\mathbb{E}_{x,t'} \left[\sum_{i:\sigma_i \leq t} |h(X_{\sigma_i}, 0) - h(X_{\sigma_i-}, t - \sigma_{i-1})| \right] < \infty.$$

Then, $h \in \mathcal{D}(\mathcal{A})$, where $\mathcal{A}h$ is given by (2).

The theoretical basis has now been prepared. In the remainder of this contribution we are concerned with the analysis of a combination of a Gerber-Shiu and a running reward functional

$$g(x, t') = \mathbb{E}_{x,t'} \left[\int_0^\tau e^{-\delta s} l(X_{s-}) ds + e^{-\delta \tau} \psi(X_{\tau-}, |X_\tau|) I_{\{\tau < \infty\}} \right]. \tag{4}$$

In (4) $\tau = \inf\{t \geq 0 \mid X_t < 0\}$ denotes the time of ruin, $\delta > 0$ is a discount or preference rate, $l : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ represents a running reward function and $\psi : \mathbb{R}^+ \times \mathbb{R}^+ \rightarrow \mathbb{R}^+$ can play the role of a terminal reward function or a penalty function; see [Gerber and Shiu \(2005\)](#). Of course, if we use g as an optimization criterion, i.e., maximization of a running reward with a penalty at ruin, then ψ should assume negative values instead.

3. Analytic Properties

In order to guarantee that our function of interest (4) is the solution to a particular partial-integro-differential equation, we must first verify several statements concerning its regularity.

Since we are going to compute (4) by means of a numerical method which is designed to solve this PIDE, a representation of this kind is essential. Subsequently, we will demonstrate that this approach is also able to incorporate statistically estimated characteristics.

3.1. Feynman-Kac Formulation

We now formulate and prove a Feynman-Kac type equation for our problem. This result states that if a function is smooth enough, it admits a representation by means of a conditional expectation involving the respective function, where in turn we must plug in the underlying stochastic process X as its argument. An analogous result can be found in Fleming and Soner (1993, p. 407), where the process is given as the solution of a particular SDE. A similar result—but in slightly a different form—concerning PDMPs is derived in Davis (1993, p. 92) and in Rolski et al. (1999, p. 454, Thm. 11.2.3).

Theorem 2. *If for a given function $h : \mathbb{R} \times \mathbb{R}_0^+ \rightarrow \mathbb{R}$ it holds that $h \in \mathcal{D}(\mathcal{A})$. Then, we obtain the following representation for h :*

$$h(x, t') = \mathbb{E}_{x,t'} \left[- \int_0^S e^{-\delta s} [(\mathcal{A}h)(X_{s-}, t'_{s-}) - \delta h(X_{s-}, t'_{s-})] ds + e^{-\delta S} h(X_S, t'_S) \right], \tag{5}$$

where S is a bounded $\{\mathcal{F}_t^{\tilde{X}}\}_{t \geq 0}$ stopping time.

Proof. At first we apply the partial integration formula to the process $e^{-\delta S} h(X_S, t'_S)$ to obtain

$$e^{-\delta S} h(X_S, t'_S) = h(x, t') + \int_0^S (-\delta) e^{-\delta s} h(X_{s-}, t'_{s-}) ds + \int_0^S e^{-\delta s} dh(X_s, t'_s).$$

Exploiting the assumption $h \in \mathcal{D}(\mathcal{A})$, which ensures that (3) is a martingale, leads us to the desired result (5). \square

As an immediate consequence we obtain the following lemma.

Lemma 1. *If a function h fulfills the assumptions of Theorem 2 and satisfies*

$$\mathcal{A}h(x, t') - \delta h(x, t') + \tilde{l}(x, t') = 0$$

then

$$h(x, t') = \mathbb{E}_{x,t'} \left[\int_0^S e^{-\delta s} \tilde{l}(X_{s-}, t'_{s-}) ds + e^{-\delta S} h(X_S, t'_S) \right],$$

for a bounded $\{\mathcal{F}_t^{\tilde{X}}\}_{t \geq 0}$ stopping time S .

If, additionally $\tilde{l}(x, t') = l(x) + \lambda(t') \int_x^\infty \psi(x, y - x) dF_Y(y)$, $\psi(x, z) = 0$ for $z < 0$ and $h(x, \cdot) = 0$ for $x < 0$, together with $\lim_{t \rightarrow \infty} \mathbb{E}_{x,t'} [e^{-\delta(t \wedge \tau)} h(X_{t \wedge \tau}, t'_{t \wedge \tau}) I_{\{\tau > t\}}] = 0$ holds, we obtain the result that $h = g$ as in (4).

Remark 1. Note that we have fixed both l and ψ to be positive such that we can formally send $t \rightarrow \infty$ in $t \wedge \tau$ using monotone convergence later. For having $\mathbb{E}_{x,t'} \left[\int_0^\tau e^{-\delta s} l(X_{s-}) ds + e^{-\delta \tau} \psi(X_{\tau-}, |X_\tau|) I_{\{\tau < \infty\}} \right] < \infty$, we certainly need to ask for a growth condition for l and an integrability condition for ψ . Common assumptions in the literature would be to choose l bounded and $\int_0^\infty \int_0^\infty \psi(x, y) f_Y(x + y) dy dx < \infty$ if F_Y admits a density f_Y . But

in specific situations one can relax these assumptions. For example, if c is bounded it suffices for l to fulfill a polynomial growth condition.

Proof. The first statement follows by an application of Theorem 2. In order to prove the specific second statement, we have to use the same line of arguments as in the proof of the previous theorem. In fact, using the result for the bounded stopping time $S = t \wedge \tau$ and the choice of \tilde{l} we get

$$h(x, t') = \mathbb{E}_{x,t'} \left[\int_0^{t \wedge \tau} e^{-\delta s} l(X_{s-}) ds + \int_0^{t \wedge \tau} e^{-\delta s} \lambda(t'_s) \int_{X_{s-}}^{\infty} \psi(X_{s-}, y - X_{s-}) dF_Y(y) ds + e^{-\delta(t \wedge \tau)} h(X_{t \wedge \tau}, t'_{t \wedge \tau}) \right].$$

For the limit $t \rightarrow \infty$ using the assumptions made on the function h , we see that the last part of the sum in the expectation disappears. Hence, it remains to be shown that

$$\lim_{t \rightarrow \infty} \mathbb{E}_{x,t'} \left[\int_0^{(t \wedge \tau)} e^{-\delta s} \int_{X_{s-}}^{\infty} \psi(X_{s-}, y - X_{s-}) dF_Y(y) \lambda(t'_s) ds \right] = \mathbb{E}_{x,t'} \left[e^{-\delta \tau} \psi(X_{\tau-}, |X_{\tau}|) I_{\{\tau < \infty\}} \right].$$

In the following few lines we set $H(x, z) = \psi(x, z) I_{\{z > 0\}}$, and for the sake of clarity we abbreviate the condition $\sigma_0 = t', X_{\sigma_0} = x$ with $\mathcal{I}_{(t',x)}$:

$$\begin{aligned} \mathbb{E}_{x,t'} \left[e^{-\delta \tau} \psi(X_{\tau-}, |X_{\tau}|) I_{\{\tau < \infty\}} \right] &= \mathbb{E}_{x,t'} \left[e^{-\delta \tau} \psi(X_{\tau-}, Y_{N_{\tau}} - X_{\tau-}) I_{\{\tau < \infty\}} \right] = \\ \mathbb{E} \left[\sum_{k=1}^{\infty} \mathbb{E} \left[e^{-\delta \sigma_k} \psi(\underbrace{\phi(\sigma_k - \sigma_{k-1}, X_{\sigma_{k-1}})}_{T_k}, Y_k - \underbrace{\phi(\sigma_k - \sigma_{k-1}, X_{\sigma_{k-1}})}_{T_k}) I_{\{\sigma_k = \tau\}} I_{\{\sigma_k < \infty\}} \middle| \mathcal{F}_{\sigma_{k-1}} \right] \middle| \mathcal{I}_{(t',x)} \right] &= \\ \mathbb{E} \left[\sum_{k=1}^{\infty} e^{-\delta \sigma_{k-1}} \mathbb{E} \left[e^{-\delta T_k} H(\phi(T_k, X_{\sigma_{k-1}}), Y_k - \phi(T_k, X_{\sigma_{k-1}})) I_{\{\sigma_k < \infty\}} \middle| \mathcal{F}_{\sigma_{k-1}} \right] \middle| \mathcal{I}_{(t',x)} \right] &= \\ \lim_{t \rightarrow \infty} \mathbb{E} \left[\sum_{k=1}^{\infty} e^{-\delta \sigma_{k-1}} \mathbb{E} \left[\int_0^{T_k \wedge (t - \sigma_{k-1}) \wedge (\tau - \sigma_{k-1})} e^{-\delta s} \mathbb{E} \left[H(\phi(s, X_{\sigma_{k-1}}), Y_k I_{\{s = T_k\}} - \phi(s, X_{\sigma_{k-1}})) \right] \frac{dI_{\{T_k \leq s\}}}{\lambda(s) ds} \middle| \mathcal{F}_{\sigma_{k-1}} \right] \middle| \mathcal{I}_{(t',x)} \right] &= \\ \lim_{t \rightarrow \infty} \mathbb{E} \left[\sum_{k=1}^{\infty} \mathbb{E} \left[\int_{\sigma_{k-1} \wedge t \wedge \tau}^{\sigma_k \wedge t \wedge \tau} e^{-\delta s} \mathbb{E} \left[H(X_{s-}, Y_k I_{\{s = \sigma_k\}} - X_{s-}) \right] \lambda(t'_s) ds \middle| \mathcal{F}_{\sigma_{k-1}} \right] \middle| \mathcal{I}_{(t',x)} \right] &= \\ \lim_{t \rightarrow \infty} \mathbb{E}_{x,t'} \left[\int_0^{(t \wedge \tau)} e^{-\delta s} \mathbb{E} \left[H(X_{s-}, Y - X_{s-}) \right] \lambda(t'_s) ds \right] &= \\ \lim_{t \rightarrow \infty} \mathbb{E}_{x,t'} \left[\int_0^{(t \wedge \tau)} e^{-\delta s} \int_{X_{s-}}^{\infty} \psi(X_{s-}, y - X_{s-}) dF_Y(y) \lambda(t'_s) ds \right]. \end{aligned}$$

□

3.2. Regularity of Gerber-Shiu Functions

In this paragraph we treat the regularity of g from (4) and demonstrate that it really does fulfill the associated partial-integro-differential equation.

Theorem 3. The Gerber-Shiu function g from (4) lies in the domain of the generator of (X, t') , i.e., $g \in \mathcal{D}(\mathcal{A})$, and fulfills

$$\mathcal{A}h(x, t') - \delta h(x, t') + \tilde{l}(x, t') = 0, \tag{6}$$

where $\tilde{l}(x, t') = l(x) + \lambda(t') \int_x^\infty \psi(x, y - x) dF_Y(y)$.

Proof. We begin by splitting up the integral associated with the running reward functional and condition on the next jump time, in which $r > 0$ and $\sigma_1 = T_1$ is the upcoming jump of the claims process. We define $v := r \wedge T_1$ and get

$$g(x, t') = \mathbb{E}_{x, t'} \left[\int_0^{r \wedge T_1} e^{-\delta s} l(X_{s-}^{x, t'}) ds + \int_{r \wedge T_1}^\tau e^{-\delta s} l(X_{s-}^{x, t'}) ds + e^{-\delta \tau} \psi(X_{\tau-}^{x, t'}, |X_{\tau}^{x, t'}|) I_{\{\tau < \infty\}} \right]$$

$$= E_{x, t'} \left[\int_0^v e^{-\delta s} l(X_{s-}^{x, t'}) ds + e^{-\delta v} \mathbb{E} \left[\int_v^\tau e^{-\delta(s-v)} l(X_{s-}^{x, t'}) ds + e^{-\delta(\tau-v)} \psi(X_{\tau-}^{x, t'}, |X_{\tau}^{x, t'}|) I_{\{\tau < \infty\}} \middle| \mathcal{F}_v \right] \right].$$

Then, we separate these terms in turn as follows:

$$g(x, t') = e^{-\int_{t'}^{t'+r} \lambda(z) dz} \int_0^r e^{-\delta s} l(X_{s-}^{x, t'}) ds + \int_0^r \lambda(t' + s) e^{-\int_{t'}^{t'+s} \lambda(z) dz} \int_0^s e^{-\delta u} l(X_{u-}^{x, t'}) du ds$$

$$+ e^{-\int_{t'}^{t'+r} \lambda(z) dz} e^{-\delta r} g \left(x + \int_0^r c(X_{u-}^{x, t'}) du, t' + r \right)$$

$$+ \int_0^r \lambda(t' + s) e^{-\int_{t'}^{t'+s} \lambda(z) dz} \left[e^{-\delta s} \int_0^{x + \int_0^s c(X_{u-}^{x, t'}) du} g \left(x + \int_0^s c(X_{u-}^{x, t'}) du - y, 0 \right) dF_Y(y) \right.$$

$$\left. + e^{-\delta s} \int_{x + \int_0^s c(X_{u-}^{x, t'}) du}^\infty \psi \left(x + \int_0^s c(X_{u-}^{x, t'}) du, y - x - \int_0^s c(X_{u-}^{x, t'}) du \right) dF_Y(y) \right] ds.$$

After rearranging some terms we obtain the following equation,

$$0 = e^{-\int_{t'}^{t'+r} \lambda(z) dz} \int_0^r e^{-\delta s} l(X_{s-}^{x, t'}) ds + \left(e^{-\int_{t'}^{t'+r} \lambda(z) dz} e^{-\delta r} - 1 \right) g(x, t')$$

$$+ e^{-\int_{t'}^{t'+r} \lambda(z) dz} e^{-\delta r} \left[g \left(x + \int_0^r c(X_{u-}^{x, t'}) du, t' + r \right) - g(x, t') \right] + \int_0^r H(s) ds. \quad (7)$$

The function $H(s)$ represents the remaining part of the original expression,

$$H(s) = \lambda(t' + s)e^{-\int_{t'}^{t'+s} \lambda(z) dz} \left[\int_0^s e^{-\delta u} l(X_{u-}^{x,t'}) du + e^{-\delta s} \int_0^s c(X_{u-}^{x,t'}) du \int_0^s g\left(x + \int_0^s c(X_{u-}^{x,t'}) du - y, 0\right) dF_Y(y) + \int_{x+\int_0^s c(X_{u-}^{x,t'}) du}^\infty \psi\left(x + \int_0^s c(X_{u-}^{x,t'}) du, y - x - \int_0^s c(X_{u-}^{x,t'}) du\right) dF_Y(y) \right].$$

Since $H(s)$ is continuous in $s \in [0, \infty)$, we have the result that (7) implies continuity of g along the integral curve $(\phi(t, x), t' + t)$ from the right (in $0+$). If we now divide the above equation by r and take the limit $r \searrow 0$, we obtain

$$0 = l(x) - (\lambda(t') + \delta) g(x, t') + \lim_{r \searrow 0} \frac{1}{r} \left[g\left(x + \int_0^r c(X_{u-}^{x,t'}) du, t' + r\right) - g(x, t') \right] + \lambda(t') \left(\int_0^x g(x - y, 0) dF_Y(y) + \int_x^\infty \psi(x, y - x) dF_Y(y) \right).$$

Consequently, $\lim_{r \searrow 0} \frac{1}{r} \left[g\left(x + \int_0^r c(X_{u-}^{x,t'}) du, t' + r\right) - g(x, t') \right]$ exists and equals $\frac{\partial}{\partial t'} g(x, t') + c(x) \frac{\partial}{\partial x} g(x, t')$. The same arguments can be applied for showing left continuity and differentiability for $(x, t') \in \mathbb{R}^+ \times \mathbb{R}^+$ along the deterministically given integral curves.

The integrability of g follows from Remark 1. \square

4. Numerical Procedure

From the above results we obtain that the Gerber-Shiu function g as given in (4) is of adequate regularity, so that it can be represented as a solution to a partial-integro-differential equation. In this section we intend to derive a numerical method for solving such PIDEs, which will later be the basis for a benchmark when discussing statistical effects on Gerber-Shiu functions. We start with the inspection of a particular toy problem. This will subsequently be extended to cover the original problem. The proposed method is universal in the sense that it applies for positive and Lipschitz $c(\cdot)$. But one needs to be carefull with the analysis of the associated boundary conditions. They do heavily depend on the concrete choice of $c(\cdot)$.

4.1. Gambler's Ruin Problem

Despite the fact that this procedure works in greater generality, as can be seen below, we consider as a first illustration the computation of the probability that our process X hits a value $a > 0$ before falling below zero. This is known as the *Gambler's ruin problem*. We therefore denote the first exit time of the interval $[0, a)$ by $\tau_{0,a} = \inf\{t \geq 0 \mid X_t \notin [0, a)\}$. We can now immediately apply Theorem 2

and Lemma 1 with this special exit time. We obtain the result that a function q which satisfies the requirements of Lemma 1, and in addition to solving the equation

$$\mathcal{A}q(x, t') - \delta q(x, t') = 0, \tag{8}$$

obeying the boundary conditions

$$q(x, t') = 0 \text{ for } x < 0 \text{ and } q(x, t') = 1 \text{ for } x \geq a,$$

admits the following representation:

$$q(x, t') = \mathbb{E}_{x,t'} \left[e^{-\delta \tau_a} I_{\{\tau_a < \tau\}} \right] = P(\tau_a = \tau_{0,a}).$$

Here $\tau_a = \inf\{t \geq 0 \mid X_t \geq a\}$, such that $q : \mathbb{R}^+ \times \mathbb{R}^+ \rightarrow [0, 1]$. Note, that the preference δ is set to be zero in order to really obtain the pure probability of the considered event.

Our goal now is to reveal such a function q by solving the associated PIDE, but since this equation contains a non-local term, we have to apply a numerical approximation procedure for general parameter constellations. Namely, we first consider a sequence $\{q^n\}_{n \in \mathbb{N}}$ of solutions, where each q^n is the respective expected value, in case we allow the process X to face at most n jumps. Since inter-arrival times are a.s. positive we have the result that $\lim_{n \rightarrow \infty} q^n = q$ pointwise.

In order to allow c to be non-constant, we use $c(x) = \kappa(b_1 - x)(x - a_1)I_{\{a_1 < x < b_1\}}$, where we assume $0 \leq a_1 < a < b_1$ and $\kappa > 0$. In Remark 2 we give motivation for this particular choice. As mentioned above, this only affects boundary conditions and the initial value of the recursive procedure. To start we set $q^0(x, t') := I_{\{x > a_1\}}$, since if there are no further jumps we arrive at the upper threshold a with probability 1—provided we start above a_1 . We now iterate over the number of remaining jumps $n \in \mathbb{N}$. For every n we discretize the state space $[a_1, a]$ into $N \in \mathbb{N}$ equidistant points $\{x_i\}$ and use finite differences to approximate the state derivative, whereas we do not touch the t' direction. Hence, the Equation (8) transforms to the following discretized counterpart:

$$\frac{\partial}{\partial t'} q^n(x, t') + c(x) \frac{q^n(x+h, t') - q^n(x, t')}{h} + \lambda(t') \int_0^x q^{n-1}(x-y, 0) dF_Y(y) - (\lambda(t') + \delta) q^n(x, t') = 0.$$

Consequently, we have to solve on every grid line (along t') the corresponding ordinary differential equation. We make use of q^{n-1} here by inserting it in the non-local part. Hence, we have to start at $x_N = a$ with $q^n(x_N, t') = 1$, since if the initial surplus is equal to a , then the desired probability is already 1. Further, $q^n(x_i, t')$, where $x_i = a_1 + ih$ for fixed $i \in \{1, \dots, N-1\}$ and $h = \frac{a-a_1}{N}$, solves as a function in $t' \in [0, t_{\text{end}}]$ the ODE:

$$\frac{\partial}{\partial t'} q^n(x_i, t') - q^n(x_i, t') \left(\frac{c(x_i)}{h} + \lambda(t') \right) = H_i(t')$$

$$q^n(x_i, t_{\text{end}}) = \int_0^{x_i} q^{n-1}(x_i - y, 0) f_Y(y) dy.$$

Due to the special choice of our drift function c , we have to fix a time horizon t_{end} , such that we can solve the considered differential equations on a finite time interval. Note that the above equality in t_{end} holds only asymptotically; actually we have $\lim_{t_{\text{end}} \rightarrow \infty} q^n(x_i, t_{\text{end}}) = \int_0^{x_i} q^{n-1}(x_i - y, 0) f_Y(y) dy$.

The corresponding imprecision stems from the truncation of the tail jump distribution; namely, we use above $\bar{F}_T^{\text{end}}(t') = \bar{F}_T(t')I_{\{t' < t_{\text{end}}\}}$. The inhomogeneity for every i has the form

$$H_i(t') = -\frac{c(x_i)}{h}q^n(x_{i+1}, t') - \lambda(t') \int_0^{x_i} q^{n-1}(x_i - y, 0)f_Y(y)dy.$$

This term is known at step n and state x_i . Note that due to the features of the function c , if the process arrives at a state smaller or equal to a_1 , then the process remains at this state. Hence, we have the boundary condition $q^n(a_1, t') = 0 \forall n \in \mathbb{N}$. Moreover, one can show that $q(x, t')$ is rightly-continuous in a_1 . Finally, we interpolate across the grid points $\{x_i\}$ the numerically determined functions in t' to obtain a function $q^n(x, t')$.

Remark 2. In our numerical examples we use the drift function $c(x) = \kappa(b_1 - x)(x - a_1)I_{\{a_1 < x < b_1\}}$, where $0 \leq a_1 < b_1$ and $\kappa > 0$. Despite the fact that this function appears to be quite specific, it turns out that—modifying the parameters and without the indicator—it is able to cover various practical situations.

First of all, this function can be used to approximate a reflecting barrier at level b_1 in a continuous way. Such a feature is of interest, in case the insurance company is willing to pay out dividends, or otherwise if large cash holdings are penalized. Those situations are nowadays quite reasonable, since negative interest rates are more and more common.

Furthermore, if we increase κ , then the deterministic path approximates an indicator function. Hence, it can mimic deterministic jumps of the surplus process which arise in problems with capital injections. In case we allow c —canceling the indicator—to be negative, then the resulting decreasing paths approach either levels b_1 or zero from above. This corresponds to a post-dividend surplus approaching the dividend barrier b_1 (especially if κ is chosen to be large, this would approximate a jump downwards; i.e., a lump sum dividend), or to a liquidation of the portfolio due to inefficiency of the insurance line.

Overall, if we combine such functions to a piecewise drift with an additional positive constant we are able to reproduce continuous versions of a variety of common dividend strategies (barrier and band type). Another nice application arises for a small choice of κ . We can fix $b_1 > 0$ and $a_1 < 0$ such that

$$-a_1b_1\kappa + (a_1 + b_1)\kappa x - \kappa x^2 = c + rx - \kappa x^2,$$

and choose κ appropriately small enough to get a local approximation of the classical drift rate with investment return $r \in \mathbb{R}$. Here $a_1 = -((-r + \sqrt{r^2 + 4\kappa}) / (2\kappa)) < 0$ and $b_1 = (r + \sqrt{r^2 + 4\kappa}) / (2\kappa)$ tend to zero and infinity as $\kappa \searrow 0$, thereby capturing the natural boundaries of the surplus.

Beyond the insurance context, such drift functions are frequently used to describe the growth of a population; see [Alvarez and Shepp \(1998\)](#).

4.2. Extended Gerber-Shiu Functional

As another example, we want to compute (4) in a more general setting including running reward and a Gerber-Shiu function. Here $q_{GS}^n(x, t')$ denotes the functional comprising at most $n \in \mathbb{N}$ jumps. We proceed in a manner analogous to that above. For the sake of clarity we assume that $l(x) \equiv L$, whereas the function c remains the same as in the previous case, namely, $c(x) = \kappa(b_1 - x)(x - a_1)I_{\{a_1 < x < b_1\}}$, where $0 \leq a_1 < a < b_1$ and $\kappa > 0$. Note that we have chosen a_1 to be zero in our subsequent example. For bounding the state space we choose a cut-off value a . This ensures that the computations are feasible; i.e., we have given boundary values and do not need to solve integral equations to obtain these. We denote with $t^*(a, x)$ the point in time when we reach the value a if we start in x by following the deterministic ODE path. In fact this function is just the inverse in t' of the deterministic path function $\phi(x, t')$.

The initial value of the approach is

$$q_{GS}^0(x, t') := I_{\{0 \leq x \leq a_1\}} \frac{L}{\delta} + I_{\{a_1 < x < a\}} e^{-t^*(a,x)\delta} \frac{L}{\delta} + I_{\{a \leq x\}} \frac{L}{\delta}.$$

For the further iterative procedure we have at $x_N = a$ that $q_{GS}^n(x_N, t') = \frac{L}{\delta}$, since if we start at the cut-off value we just obtain the discounted reward continuously and forever. Analogously as above, $q_{GS}^n(x_i, t')$, where $x_i = a_1 + ih$ for fixed $i \in \{1, \dots, N - 1\}$ and $h = \frac{a-a_1}{N}$, solves for $t' \in [0, t_{\text{end}}]$ the ODE:

$$\frac{\partial}{\partial t'} q_{GS}^n(x_i, t') - q_{GS}^n(x_i, t') \left(\frac{c(x_i)}{h} + \lambda(t') + \delta \right) = H_i^{GS}(t')$$

$$q_{GS}^n(x_i, t_{\text{end}}) = \int_0^{x_i} q_{GS}^{n-1}(x_i - y, 0) f_Y(y) dy + \int_{x_i}^{\infty} \psi(x_i, y - x_i) f_Y(y) dy.$$

As above we consider a finite time interval and therefore make use of t_{end} . In this case the inhomogeneity admits the following form for every i

$$H_i^{GS}(t') = -L - \frac{c(x_i)}{h} q_{GS}^n(x_{i+1}, t') - \lambda(t') \left(\int_0^{x_i} q_{GS}^{n-1}(x_i - y, 0) f_Y(y) dy + \int_{x_i}^{\infty} \psi(x_i, y - x_i) f_Y(y) dy \right).$$

Doing this for every point x_i results in a discretized approximation of q_{GS}^n , which one may denote by $q_{GS}^{n,h}$ for highlighting the dependence on the step width $h > 0$ (here $h = \frac{a-a_1}{N}$). In contrast to the previous problem, we have in this case that at a_1 the function $q_{GS}^n(x, t')$ must be determined. In our numerical example we assume that $a_1 = 0$; hence, we obtain the boundary condition

$$q_{GS}^n(0, t') = \int_0^{\infty} \int_0^{\infty} \left(\int_0^{t_1} e^{-\delta s} L ds + e^{-\delta t_1} \psi(0, y) \right) \frac{f_T(t_1 + t')}{F_T(t')} dt_1 f_Y(y) dy,$$

which can be computed explicitly.

Again, interpolation leads to a function $q_{GS}^n(x, t')$ on the whole domain which approximates (4). Note, that in the case of a non-constant reward l the boundary values need to be replaced by $\int_0^{\infty} e^{-\delta t} l(\phi(t, x)) dt$.

4.3. Convergence of Numerical Scheme

Consider a PDMP $\tilde{X}^h = (X^h, t')$ with state space $E^h = \{kh : k \in \mathbb{Z}\} \times \mathbb{R}_0^+ \subset E = \mathbb{R} \times \mathbb{R}_0^+$ for some $h > 0$. We identify $X_i^h = k$ with the actual position kh ; i.e., the first component of \tilde{X}^h describes an external discrete state. For suitable $g : E^h \rightarrow \mathbb{R}$, this process is described by its generator

$$\mathcal{A}^h g(k, t') = \frac{\partial}{\partial t'} g(k, t') + \sum_{l=-\infty}^k \lambda(t') p_{kl}^h g(l, 0) + \frac{c(k)}{h} g(k+1, t') - \left(\lambda(t') + \frac{c(k)}{h} \right) g(k, t'),$$

where $p_{kl}^h = F_Y((k-l)h) - F_Y((k-l-1)h) = P[kh - Y \in [lh, (l+1)h]]$. Note that this process has its origins in the numerical procedure presented in the section above. In a next step we will apply Theorem 5.16 from [Kritzer et al. \(2019\)](#) or directly Theorem 19.25 from [Kallenberg \(2002\)](#) to show that $\tilde{X}^h \xrightarrow{d} \tilde{X} = (X, t')$, our original process. As a consequence, we get that expected values of certain functionals of the underlying processes, \tilde{X}^h, \tilde{X} , converge against each other. Lemma 5.14 of [Kritzer et al. \(2019\)](#) tells us that the relevant ingredients of q^n and q_{GS}^n are appropriately continuous if ψ and l are bounded.

Fix $k_x(h) := \lfloor \frac{x}{h} \rfloor$ for $x \in \mathbb{R}$ such that $X_0^h = k_x(h)h \rightarrow x = X_0$ as $h \rightarrow 0$. Furthermore, let $f \in C_b^\infty(E, \mathbb{R})$ which is certainly an element of $\mathcal{D}(\mathcal{A})$ and $\mathcal{D}(\mathcal{A}^h)$. We need to focus on

$$\begin{aligned} & \left| \mathcal{A}f(x, t') - \mathcal{A}^h f(k_x(h)h, t') \right| \leq |f_{t'}(x, t') - f_{t'}(k_x(h)h, t')| \\ & + |\lambda(t')| \left| \int_0^\infty f(x - y, 0) dF_Y(y) - \sum_{l=-\infty}^{k_x(h)-1} f(lh, 0) p_{k_x(h)l}^h \right| + |\lambda(t')| |f(x, t') - f(k_x(h)h, t')| \\ & + \left| c(x)f_x(x, t') - c(k_x(h)h) \left(\frac{f((k_x(h) + 1)h, t') - f(k_x(h)h, t')}{h} \right) \right| \\ & \leq \|f_{t'x}\|_\infty h + |\lambda(t')| \int_0^\infty \sum_{l=-\infty}^{k_x(h)-1} I_{\{(k_x(h)-l-1)h < y \leq (k_x(h)-l)h\}} |f(x - y, 0) - f(lh, 0)| dF_Y(y) \\ & + (|\lambda(t')| + L_c) \|f_x\|_\infty h + \|c\|_\infty \frac{3}{2} \|f_{xx}\|_\infty h \\ & \leq \|f_{t'x}\|_\infty h + |\lambda(t')| \|f_x\|_\infty 2h + (|\lambda(t')| + L_c) \|f_x\|_\infty h + \|c\|_\infty \frac{3}{2} \|f_{xx}\|_\infty h, \end{aligned}$$

where L_c denotes the Lipschitz constant of $c(\cdot)$. The remaining terms, which bound the difference of the two generators, converge to zero uniformly in (x, t') if we assume a bounded jump intensity λ and a differentiable, bounded and Lipschitz function c . Therefore, [Kritzer et al. \(2019, Theorem 5.16\)](#) tells us that \tilde{X}^h and \tilde{X} converge weakly against each other as $h \rightarrow 0$ and the associated Gerber-Shiu and reward functions do as well, if ψ and l are bounded—as previously mentioned.

This is certainly a qualitative and not a quantitative statement (involving convergence rates depending on the discretization h), but it shows that the schema are correctly designed.

Remark 3. *Moreover, compare [Fleming and Soner \(1993\)\[ch. IX\]](#) where they use techniques based on viscosity solutions in order to verify the convergence of the numerical state and time discretization scheme. The basis for such results, as well as for our own, is certainly provided in [Kushner and Dupuis \(2001\)](#) where Markov chain approximations of continuous time stochastic processes are extensively discussed.*

5. Statistical Complement

In this part of the manuscript we will discuss the effect of estimated parameters on the aforementioned numerical methodology for the computation of functionals of \tilde{X} . We place our focus on the jump intensity λ and jump size distribution F_Y . Naturally, the estimators used are based on a sample $\{(Y_i, T_i)\}_{i \in \mathbb{N}}$ of (iid) claims and inter-arrival times (or—practically correct—a finite sub-set is used).

5.1. Kernel Estimator

Our main goal is to compute (4), which will allow us to make a quantitative valuation of the underlying insurance line. In order to determine the function (4) according to the specifics of a given insurance line, we have to use statistical methods to incorporate the behavior of important characteristics; i.e., the distribution of the inter-arrival time F_T and the distribution of the claim height F_Y .

For this purpose we apply a non-parametric approach; namely, we use the kernel method to estimate the respective densities f_T and f_Y . This is necessary, since we want to apply our approximation procedure for the construction of PDMPs later. Therefore, we need to have *smooth* estimates for these densities and for the associated jump rate $\lambda(t') = \frac{f_T(t')}{\int_{t'}^\infty f_T(s) ds}$ in particular.

Given the respective data we use the kernel method—after a logarithmic transformation of it, as described in the monograph [Silverman \(1986, p. 30\)](#)—to estimate the required probability densities.

The approach works as follows: since we want to estimate densities with support on the positive half line, we take the logarithms of the data and apply the common kernel estimator for the transformed data to obtain a function $\hat{f}^{\log}(z)$, where we denote the given data points by $\text{data}_1, \dots, \text{data}_\nu$, the sample size by ν , the bandwidth by h and the kernel by K . We obtain that

$$\hat{f}^{\log}(z) = \frac{1}{\nu h} \sum_{i=1}^{\nu} K\left(\frac{z - \log(\text{data}_i)}{h}\right).$$

The estimator of the actual data is then given by $\hat{f}(z) = \frac{1}{z} \hat{f}^{\log}(\log(z))$ for $z > 0$. For our numerical considerations we use *Wolfram Mathematica*, which in turn uses for the density at point z a linearly interpolated version of the kernel estimator $\frac{1}{\nu h} \sum_{i=1}^{\nu} K\left(\frac{z - \text{data}_i}{h}\right)$. Added to this, the used kernel K is the density of the standard normal distribution and in order to determine the bandwidth h , the *Silverman rule* is applied. Note that it is mentioned in [Silverman \(1986, p. 45 et seq.\)](#) that using this rule for the bandwidth can lead to overly smooth results, if the underlying distribution is multimodal.

5.2. Uniform Consistency

Of course we have to make sure that an increase of the number ν of sample points will lead to a decrease of the estimation error. Following again the lines of [Silverman \(1986, pp. 71–74\)](#) we obtain that

$$\sup_x |\hat{f}(x) - f(x)| \rightarrow 0 \quad \text{as } \nu \rightarrow \infty$$

with probability one; if the kernel K is bounded, or has a bounded variation, it holds that $\int |K(t)| dt < \infty$ and $\int K(t) dt = 1$, in addition to the property that the set of discontinuities of K is a Lebesgue null set. Furthermore, the true density f has to be uniformly continuous on $(-\infty, \infty)$ and the bandwidth has to fulfill $h_\nu \rightarrow 0$ and $\frac{\nu h_\nu}{\log(\nu)} \rightarrow \infty$ as $\nu \rightarrow \infty$.

Fulfilling the above assumptions ensures that the estimators of the density converge uniformly against the true density. In order to make sure that we can apply Theorem 5.16 from [Kritzer et al. \(2019\)](#) we have to estimate the jump rate $\lambda(t)$ for $F_T(t) < 1$ in such a way that the estimator converges uniformly. With regard to the special form of $\lambda(t) = \frac{f_T(t)}{1 - F_T(t)}$ we consider the estimation on a compact interval; compare to ([Liu and Van Ryzin 1985, p. 600, Thm. 3.4](#)).

For that purpose, we choose an analogous starting point as in [Antoniadis et al. \(1999, pp. 65–66\)](#). First of all, we set $\Theta_{F_T} = \sup\{t : F_T(t) < 1\}$ and restrict our estimation procedure to the finite interval $[0, \bar{\Theta}]$, where $\bar{\Theta} < \Theta_{F_T}$ holds. Let now \hat{F}_T^ν be the empirical distribution function of T using ν data points; we obtain with $\Theta_{\hat{F}_T^\nu} = \sup\{t : \hat{F}_T^\nu(t) < 1\}$ that $\Theta_{\hat{F}_T^\nu} = T_{(\nu)}$. Here $T_{(\nu)}$ denotes the ν th order statistic of a sequence of length ν , and thus, the largest element. In the implementation of the statistical procedure one can take $\bar{\Theta} = T_{(\nu)}$, since it can be shown that $\Theta_{\hat{F}_T^\nu} = T_{(\nu)} \rightarrow \Theta_{F_T}$ almost surely if $\nu \rightarrow \infty$. The uniform convergence of the jump rate estimator can then be verified analogously as described in [Liu and Van Ryzin \(1985, p. 600\)](#). For further results on uniform consistency, see for example, [Einmahl and Mason \(2005\)](#).

Remark 4. Note that the above requirements for uniform convergence are fulfilled using for example the Gaussian kernel in combination with the Silverman rule, which yields a bandwidth proportional to $\nu^{-\frac{1}{5}}$.

5.3. Convergence of Estimated Gerber-Shiu Functions

Complementing the discussion about the convergence above, we now investigate the respective behaviors of the generator \mathcal{A}^h when we replace λ , F_Y and c by their estimated counterparts $\hat{\lambda}$, \hat{F}_Y and \hat{c} in \mathcal{A}^h . Note that concerning the estimation of the function c we demand that the respective estimator \hat{c} has to converge uniformly to the true c . In our numerical examples we try to mirror real-life applications. This leads to the fact that the premium rate c underlies the estimation procedure in such

a way that it depends on the first moments of Y and T . Hence, we assume that $c = \mathbb{E}[Y]\mathbb{E}[T](1 - 0.1)$, which is a quite standard approach. Another established option for a common premium principle would be the variance premium principle.

We can directly tie in with the line of arguments provided in Section 4.3. This will yield the convergence of our computational procedure. Hence, let again $E^h = \{kh : k \in \mathbb{Z}\} \times \mathbb{R}_0^+ \subset E$ denote the state space of the associated Markov chain. For the starting value $X_0^h = k_x(h)h$, where $k_x(h) := \lfloor \frac{x}{h} \rfloor$ we have that $X_0^h \rightarrow X_0$ if $h \rightarrow 0$ as above. Here h again denotes the step size of the discretization, and moreover, ν denotes the sample size. Finally, let $f \in C_b^\infty(E, \mathbb{R})$; then, we have to show that the following difference tends to zero:

$$\begin{aligned} & \left| \mathcal{A}f(x, t') - \hat{\mathcal{A}}^h f(k_x(h)h, t') \right| \\ & \leq \left| \mathcal{A}f(x, t') - \mathcal{A}^h f(k_x(h)h, t') \right| + \left| \mathcal{A}^h f(k_x(h)h, t') - \hat{\mathcal{A}}^h f(k_x(h)h, t') \right|. \end{aligned} \tag{9}$$

Since we know from Section 4.3 that the first summand tends to zero uniformly in (x, t') , the same remains to be shown for the second one. Hence, we obtain for the second term

$$\begin{aligned} & \left| \mathcal{A}^h f(k_x(h)h, t') - \hat{\mathcal{A}}^h f(k_x(h)h, t') \right| \leq \left| \lambda(t') \sum_{l=-\infty}^{k_x(h)-1} f(lh, 0) p_{k_x(h)l}^h - \hat{\lambda}(t') \sum_{l=-\infty}^{k_x(h)-1} f(lh, 0) \hat{p}_{k_x(h)l}^h \right| \\ & + \left| \lambda(t') - \hat{\lambda}(t') \right| \left| f(k_x(h)h, t') \right| + \left| c(k_x(h)h) - \hat{c}(k_x(h)h) \right| \left| \frac{f((k_x(h) + 1)h, t') - f(k_x(h)h, t')}{h} \right| \\ & \leq \left| \lambda(t') - \hat{\lambda}(t') \right| \left| \sum_{l=-\infty}^{k_x(h)-1} f(lh, 0) p_{k_x(h)l}^h \right| + \left| \hat{\lambda}(t') \right| \left| \sum_{l=-\infty}^{k_x(h)-1} f(lh, 0) p_{k_x(h)l}^h - \sum_{l=-\infty}^{k_x(h)-1} f(lh, 0) \hat{p}_{k_x(h)l}^h \right| \\ & + \left| \lambda(t') - \hat{\lambda}(t') \right| \|f\|_\infty + \left| c(k_x(h)h) - \hat{c}(k_x(h)h) \right| \|f_x\|_\infty. \end{aligned}$$

In fact, the above terms tend to zero uniformly in (x, t') on compacts for t' a.s., because of the aforementioned convergence of the estimators $\hat{\lambda}$, \hat{F}_Y and \hat{c} as $\nu \rightarrow \infty$.

Note that we use the following estimate for the series term

$$\left| \sum_{l=-\infty}^{k_x(h)-1} f(lh, 0) p_{k_x(h)l}^h - \sum_{l=-\infty}^{k_x(h)-1} f(lh, 0) \hat{p}_{k_x(h)l}^h \right| \leq 2K \|f\|_\infty \|F_Y - \hat{F}_Y\|_\infty + \|f\|_\infty \left[\bar{F}_Y(K) + \hat{F}_Y(K) \right],$$

where K has to be chosen such that $\max \left(\bar{F}_Y(K), \hat{F}_Y(K) \right) < \frac{\epsilon}{2\|f\|_\infty}$ for a given $\epsilon > 0$.

Overall, if we take the limits $h \searrow 0$ and $\nu \rightarrow \infty$, then we obtain that the left-hand side of (9) converges a.s. to zero uniformly on compacts in t' and uniformly in x .

Remark 5. In order to meet the requirements of the above statistical procedure, we denote with $(\Theta^\mathbb{N}, \mathcal{G}^\mathbb{N}, \mathcal{P})$ the sample space such that for $\omega \in \Theta^\mathbb{N}$ we have that $\{Y_i(\omega), T_i(\omega)\}_{i \in \mathbb{N}}$ denotes one sample. Given such an $\omega \in \Theta^\mathbb{N}$ and a sample size $\nu \in \mathbb{N}$ we have that $\hat{c} = \hat{c}(\omega, \nu)$, $\hat{\lambda} = \hat{\lambda}(\omega, \nu)$ and $\hat{F}_Y = \hat{F}_Y(\omega, \nu)$. Note that the dependence on ν expresses that the estimators only use the first ν claim heights and inter-arrival times, respectively, for the estimation. Hence, following the previous lines, this yields that $\exists N \in \mathcal{G}^\mathbb{N}$ with $\mathcal{P}(N) = 0$ such that $\forall \omega \in \Theta^\mathbb{N} \setminus N$

$$\begin{aligned} \hat{c}(\omega, \nu) & \rightarrow c && \text{uniformly,} \\ \hat{\lambda}(\omega, \nu) & \rightarrow \lambda && \text{uniformly on compacts in } t' \text{ and} \\ \hat{F}_Y(\omega, \nu) & \rightarrow F_Y && \text{uniformly} \end{aligned}$$

for $\nu \rightarrow \infty$. Furthermore, for each $\omega \in \Theta^\mathbb{N} \setminus N$ we have that

$$\hat{\mathcal{A}}(\omega, \nu)f = \frac{\partial}{\partial t'} f(x, t') + \hat{c}(x, \omega, \nu) \frac{\partial}{\partial x} f(x, t') - \hat{\lambda}(t', \omega, \nu) f(x, t') + \hat{\lambda}(t', \omega, \nu) \int_0^\infty f(x - y, 0) d\hat{F}_Y(y, \omega, \nu)$$

is generator of a sequence of PDMPs on E with index $v \in \mathbb{N}$. These converge as $v \rightarrow \infty$ to $\mathcal{A}f$ uniformly in (x, t') on compacts for t' . Notice that incorporating the statistical procedure we use $\mathcal{G}_0 = \sigma(\{Y_i, T_i : i \in \mathbb{N}\}) \subseteq \mathcal{F}_0$ in order to face fixed characteristics as an underlying basis of our considered renewal risk model. Above we have shown that for each such ω , and the discrete version of $\hat{\mathcal{A}}, \hat{\mathcal{A}}^h$ converges in the appropriate sense to \mathcal{A} .

At this point we are again in the situation to use Kritzer et al. (2019, Theorem 5.16), which yields a.s. weak convergence of the approximated and estimated process against its underlying counterpart.

6. Numerical Illustrations

Complementing the theoretical part, we illustrate in the subsequent paragraphs numerical results for the two considered problems. In order to outline the exemplifications we want to point out that there are four different approaches which result in four functions $V(x, t), \hat{V}(x, t), MC(x)$ and $\widehat{MC}(x)$. First of all $V(x, t)$ denotes the solution we obtain following the lines in Section 4. On the other hand, if we replace in this setup the respective characteristics by their estimated counterparts we end up with the solution $\hat{V}(x, t)$. Furthermore, we computed via Monte-Carlo simulations the functions $MC(x)$ and $\widehat{MC}(x)$ where we fixed $t = 0$, and for the latter we again made use of the estimated ingredients. The simulated results are primarily needed for reasons of comparability and hence serve as a benchmark.

6.1. Hitting Probabilities

In the following graphical illustrations we have applied the above described approach to compute the hitting probability, here denoted by $V(x, t)$. Note that the cut-off value a is natural for our choice of non-constant drift $c(x)$ for which ruin is certain. In the general situation—constant c or unbounded ODE paths—the probability of hitting a finite level a before getting ruined serves as an approximation for the ruin probability. Certainly, as $a \rightarrow \infty$ we would face a rare event problem. In such a situation, one needs to fix an appropriate finite a . This is in order to not lose the focus of the method, since otherwise there would be need to consider *variance reduction methods* or the like for dealing with this additional feature.

We have used the set of parameters listed in Table 1.

Table 1. Set of parameters used in the first example.

κ	a_1	b_1	h	a	n	Y	T	v	t_{end}
0.2	0	10	0.025	6	10	$\Gamma(3,3)$	$\Gamma(2,1)$	1000	20

Furthermore, as mentioned above we assumed that $c(x) = \kappa(b_1 - x)(x - a_1)I_{\{a_1 < x < b_1\}}$. A probabilistic reference solution is based on Monte Carlo simulation, where we used 10000 sample paths of the renewal process, where we used 2000 random points in turn for every path to simulate the inter-arrival time and the claim height, respectively; i.e., 1000 jump events. The solution after the 10th jump iteration is shown in Figure 1. Moreover, the associated improvement, namely, the difference of the last two iterations illustrated in the subsequent Figure 2, is already relatively small.

Figure 3 depicts this hitting probability for $x = 1$ as a function in t . The black curve $V_{10}(1, t)$ corresponds to the numerical solution using the above mentioned distribution functions, and the red curve $\hat{V}_{10}(1, t)$ depicts the solution we obtain using their estimated counterparts with $v = 1000$ given data points generated from the respective distribution. What we immediately observe is that the accuracy of the latter method depends to an enormous extent on the quality of the estimator for the hazard rate. This is explained by the fact that the ODE methods rely on *smooth coefficients* with preferably low variation. In our approach we need to estimate the density f_T in such a manner that the associated hazard rate function is sufficiently regular; see Figure 4.

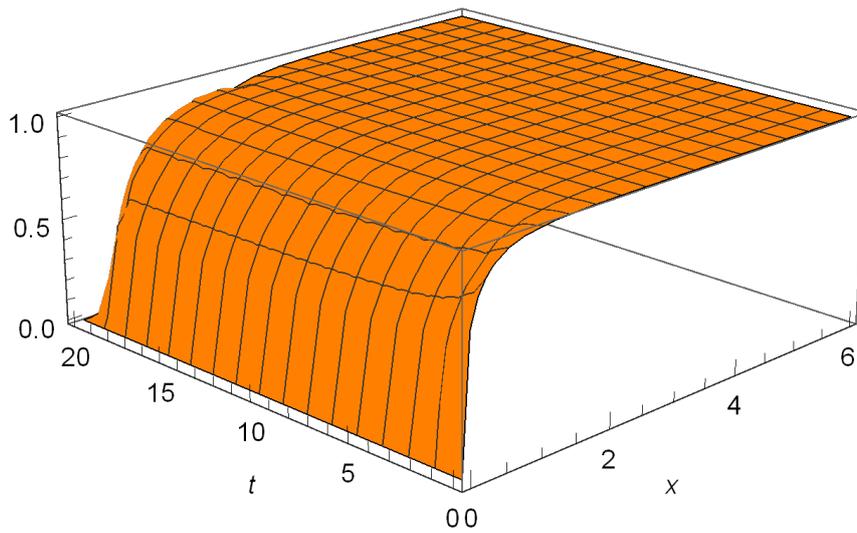


Figure 1. Hitting probability as a function of (x, t) .

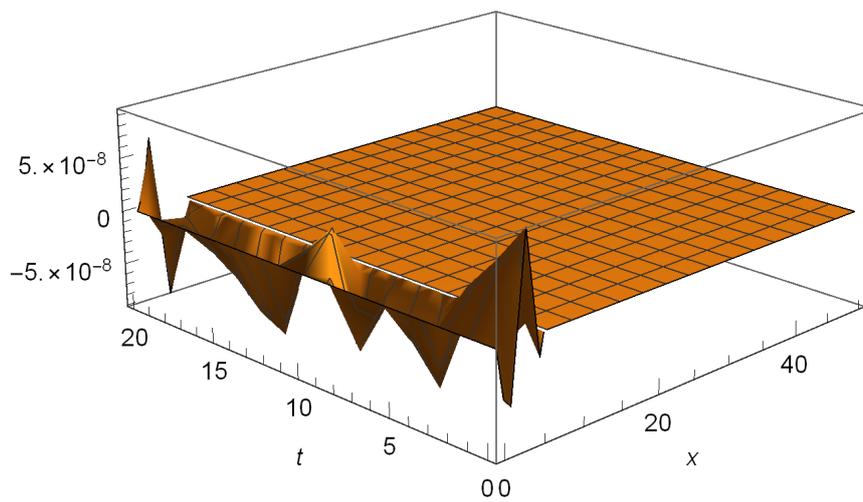


Figure 2. Difference of 9th and 10th iterations.

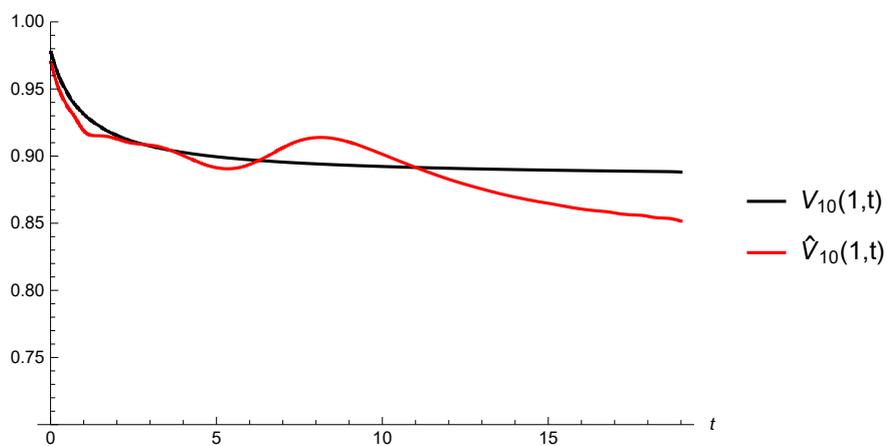


Figure 3. The hitting probabilities in $x = 1$ —original and estimated.

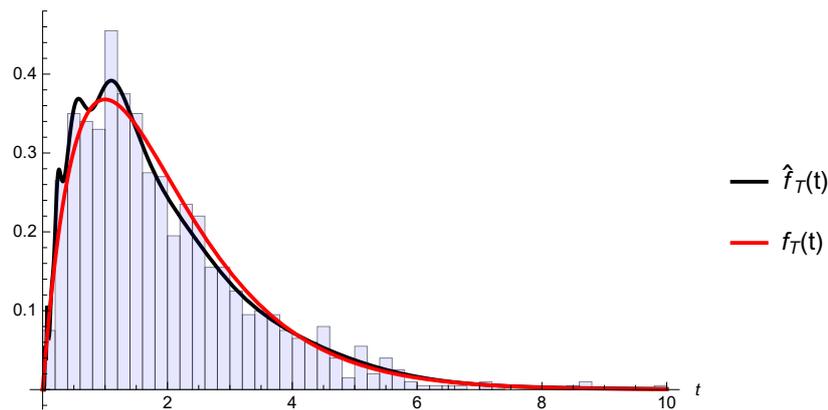


Figure 4. Comparison of the jump time density and its estimated analogue together with the used data depicted in a histogram.

This is for the applicability of the above mentioned theoretical results. Using smooth kernel estimators for the density leads to an appropriate estimator of the hazard rate—from an analytic point of view. Nevertheless, the hazard’s estimation error explains the deviation of $V_{10}(x, t)$ and $\hat{V}_{10}(x, t)$ for large t values. This is an immediate consequence of our way of computing the desired quantities. The difference of $V_{10}(x, t)$ and $\hat{V}_{10}(x, t)$ as a function of x and t is shown in Figure 5. One can see clearly that for increasing x the hazard’s impact declines.

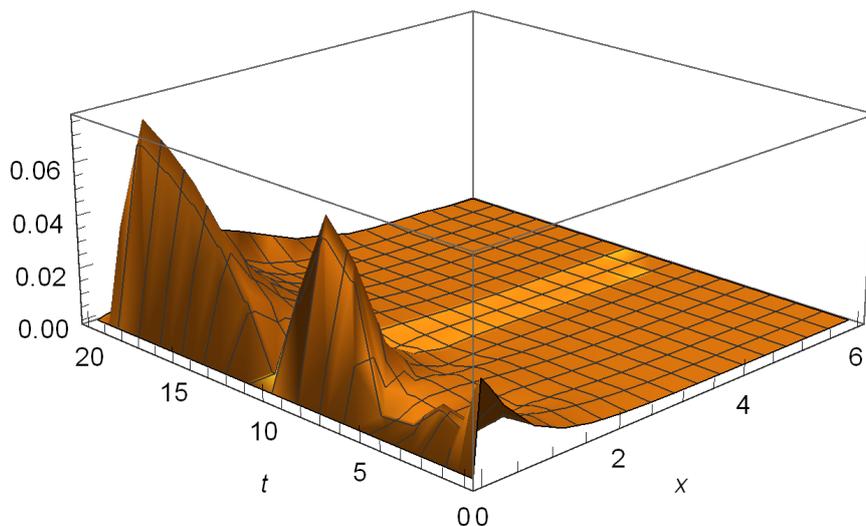


Figure 5. Difference of $V_{10}(x, t)$ and $\hat{V}_{10}(x, t)$.

The convergence regarding the discretization method described above is illustrated in Figure 6, where we can see a sequence of solutions for decreasing step size h . The use of a step size $h = 0.025$ applied for the computation of $V_{10}(x, 0)$ results in a solution which is reasonably close to the one obtained via a comparative MC-simulation (depicted as a dashed line together with its 95% confidence intervals).

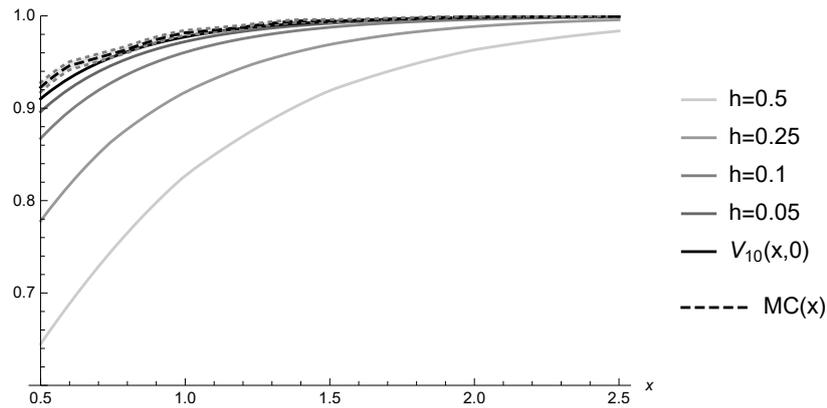


Figure 6. Improvement of the solution due to decrease of step size h .

In order to illustrate the development concerning the estimation as ν increases, we list in Table 2 the values of the function $V_{10}(x, t)$ and $\hat{V}_{10}(x, t)$ for $t = 0$ and different values of x , in addition to the results from the MC-simulation.

Table 2. Development of the solution with estimation due to variation of the sample size ν .

	$x = 0.5$	$x = 1$	$x = 1.5$	$x = 2$	$x = 2.5$
$MC(x)$	0.9226	0.9817	0.9946	0.9991	0.9997
$\widehat{MC}(x)$	0.9228	0.9820	0.9961	0.9991	0.9998
$V_{10}(x, 0)$	0.9102	0.9773	0.9938	0.9982	0.9995
$\hat{V}_{10}^{\nu=1000}(x, 0)$	0.8917	0.9698	0.9907	0.9969	0.9989
$\hat{V}_{10}^{\nu=750}(x, 0)$	0.8961	0.9712	0.9917	0.9977	0.9994
$\hat{V}_{10}^{\nu=500}(x, 0)$	0.9001	0.9733	0.9925	0.9980	0.9995
$\hat{V}_{10}^{\nu=250}(x, 0)$	0.9036	0.9732	0.9915	0.9972	0.9991
$\hat{V}_{10}^{\nu=50}(x, 0)$	0.9159	0.9778	0.9929	0.9982	0.9997
$\hat{V}_{10}^{\nu=25}(x, 0)$	0.8701	0.9620	0.9884	0.9973	0.9994

6.2. Gerber-Shiu Functions

As above, we provide some illustrations of the numerical computations concerning the determination of the expectation (4) in this passage in a more general setting. We denote the solution in this framework by $V^{GS}(x, t)$. We have made the following assumptions and used the parameters given in Table 3.

Table 3. Set of parameters used in the second example.

κ	a_1	b_1	$l(x)$	δ	$\psi(x, z)$	β	h	a	n	Y	T	ν	t_{end}
0.2	0	10	2	0.05	$e^{-\beta z}$	2	0.025	9	6	$\Gamma(3,3)$	$\Gamma(2,1)$	1000	20

Moreover, we have again $c(x) = \kappa(b_1 - x)(x - a_1)I_{\{a_1 < x < b_1\}}$. In Table 4 we display values obtained using the four different approaches described above for $t = 0$ and for several values of x . An immediate observation is that the deviation declines while x increases, which is due to the decreasing influences of the errors of estimation affecting the coefficients and the boundary conditions.

Table 4. Comparison of results obtained via different approaches.

	$x = 0.5$	$x = 1$	$x = 1.5$	$x = 2$	$x = 2.5$
$MC^{GS}(x)$	37.0267	39.2955	39.7925	39.9658	39.9869
$\widehat{MC}^{GS}(x)$	37.0351	39.3071	39.8494	39.9649	39.9903
$V_6^{GS}(x, 0)$	36.5443	39.1244	39.7592	39.9317	39.9802
$\hat{V}_6^{GS}(x, 0)$	35.8309	38.8363	39.6413	39.8808	39.9577

Figure 7 illustrates the improvement of the solution of the GS function for different n .

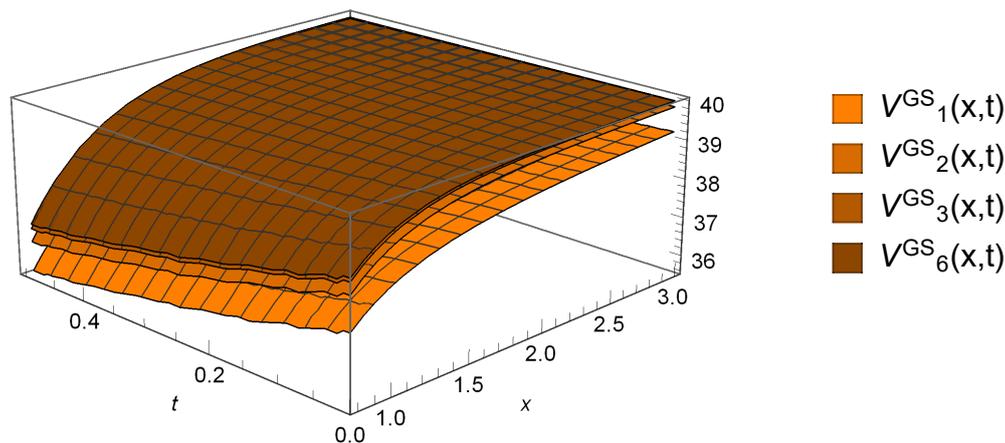


Figure 7. Sequence of solutions for $n = 1, 2, 3, 6$.

Figure 8 shows the solution for the GS function compared to the version obtained using the estimated parameters. In addition, the result of the MC-simulation and the respective confidence intervals for both of the computations are included. Note that for the computation of the MC-simulated counterpart $MC^{GS}(x)$ of $V^{GS}(x, 0)$, we have used the true distributions F_T and F_Y , whereas for the quantity $\widehat{MC}^{GS}(x)$ we generated the sample points from the estimated distributions given $\nu = 1000$ data points each. In fact this corresponds to a bootstrap procedure. We can observe that the simulated results are not very sensitive with respect to estimation, which can also be seen in Table 4 and that the numerically calculated $V^{GS}(x, 0)$, with a grid size of $h = 0.025$, lies partly within the 95% confidence band. In contrast to this, the numerically determined $V^{GS}(x, 0)$, with estimated coefficients, is not able to reach the confidence band. This is because of the fact that the approximation error of the estimated hazard rate is too significant here—even for a relatively large sample size of $\nu = 1000$. This nicely illustrates that the stated convergence results have a qualitative nature and hold asymptotically for $\nu \nearrow \infty$ and $h \searrow 0$. At the same time one observes that this initial error becomes less influential for increasing x .

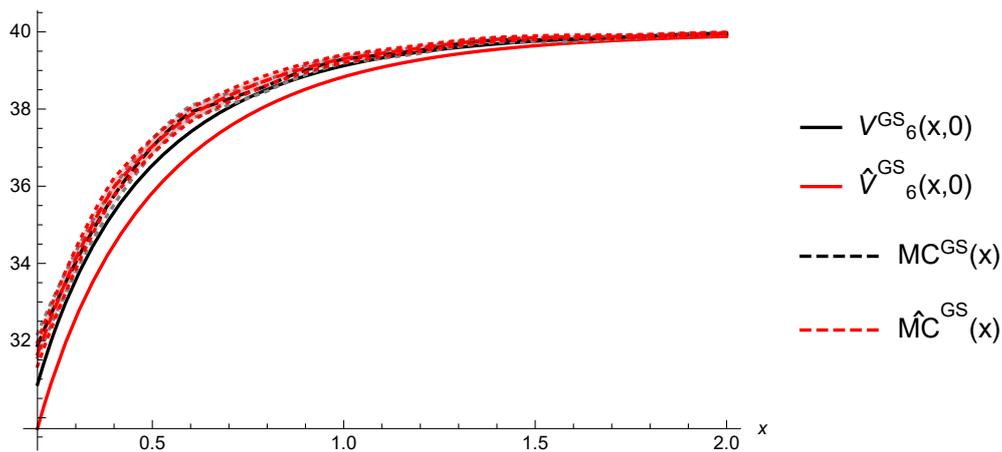


Figure 8. Comparison of solution with and without estimation together with their MC simulated counterparts.

7. Conclusions and Discussions

In this contribution we present numerical and statistical methods for the computation of Gerber-Shiu functions in general renewal risk models. The numerical method we provide incorporates the fact that the underlying model itself faces stochastic ingredients, i.e., drift, estimated intensity and distribution of claims, and thus takes the statistical aspect into account. In particular, we have shown that the proposed schemata converge and illustrate these findings by means of two informative examples. For practical applications we wish to point out that the solutions obtained via MC methods are relatively robust with respect to estimation, whereas by nature of the scheme, methods from numerical analysis are more sensitive. The main reason is typically the high variation of the estimated hazard rate. Here, the fact that estimating derivatives is a relatively complex task impacts on our approach. On the other hand, the computations are less affected by estimation of the claims' distribution F_Y .

Several directions are open for future research. Namely, the incorporation of estimation methods such as the use of parametrized families or the direct usage of phase-type approximations for the empirical distribution functions of observed inter-jump times and claims.

Author Contributions: J.A.S. and S.T. contributed equally to this work. Both authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors would like to thank the anonymous referees for their remarks which clearly helped to improve this contribution.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Albrecher, Hansjörg, and Eleni Vatamidou. 2019. Ruin Probability Approximations in Sparre Andersen Models with Completely Monotone Claims. *Risks* 7: 104. [\[CrossRef\]](#)
- Alvarez, Luis, and Larry Shepp. 1998. Optimal harvesting of stochastically fluctuating populations. *Journal of Mathematical Biology* 37: 155–77. [\[CrossRef\]](#)
- Antoniadis, Anestis, Gérard Grégoire, and Guy Nason. 1999. Density and hazard rate estimation for right-censored data by using wavelet methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 61: 63–84. [\[CrossRef\]](#)
- Asmussen, Søren, and Hansjörg Albrecher. 2010. *Ruin Probabilities*, 2nd ed. River Edge: World Scientific.
- Asmussen, Søren, Olle Nerman, and Marita Olsson. 1996. Fitting phase-type distributions via the em algorithm. *Scandinavian Journal of Statistics* 23: 419–41.

- Azaïs, Romain, François Dufour, and Anne Gégout-Petit. 2014. Non-parametric estimation of the conditional distribution of the interjumping times for piecewise-deterministic Markov processes. *Scandinavian Journal of Statistics* 41: 950–69. [\[CrossRef\]](#)
- Azaïs, Romain, François Dufour, and Anne Gégout-Petit. 2013. Non-parametric estimation of the jump rate for non-homogeneous marked renewal processes. *Annales de l'IHP Probabilités et Statistiques* 49: 1204–31. [\[CrossRef\]](#)
- Brémaud, Pierre. 1981. *Point Processes and Queues*. New York and Berlin: Springer.
- Davis, Mark. 1993. *Markov Models and Optimization*. London: Chapman & Hall.
- Einmahl, Uwe, and David Mason. 2005. Uniform in bandwidth consistency of kernel-type function estimators. *The Annals of Statistics* 33: 1380–403. [\[CrossRef\]](#)
- Fleming, Wendell Helms, and Halil Mete Soner. 1993. *Controlled Markov Processes and Viscosity Solutions*. New York: Springer.
- Gerber, Hans-Ulrich, and Elias Shiu. 1998. On the time value of ruin. *North American Actuarial Journal* 2: 48–78. [\[CrossRef\]](#)
- Gerber, Hans-Ulrich, and Elias Shiu. 2005. The time value of ruin in a Sparre Andersen model. *North American Actuarial Journal* 9: 49–84. [\[CrossRef\]](#)
- Kallenberg, Olav. 2002. *Foundations of Modern Probability. (Probability and its Applications)*, 2nd ed. New York: Springer.
- Kritzer, Peter, Gunther Leobacher, Michaela Szölgényi, and Stefan Thonhauser. 2019. Approximation methods for piecewise deterministic Markov processes and their costs. *Scandinavian Actuarial Journal* 2019: 308–35. [\[CrossRef\]](#) [\[PubMed\]](#)
- Kushner, Harold, and Paul Dupuis. 2001. *Numerical Methods for Stochastic Control Problems in Continuous Time*, 2nd ed. New York: Springer.
- Liu, Regina, and John Van Ryzin. 1985. A histogram estimator of the hazard rate with censored data. *The Annals of Statistics* 13: 592–605. [\[CrossRef\]](#)
- Preischi, Michael, Stefan Thonhauser, and Robert Tichy. 2018. Integral equations, quasi-Monte Carlo methods and risk modeling. In *Contemporary Computational Mathematics—A Celebration of the 80th birthday of Ian Sloan*. Cham: Springer, pp. 1051–74.
- Rolski, Tomasz, Hanspeter Schmidli, Volker Schmidt, and Jozef Teugels. 1999. *Stochastic Processes for Insurance and Finance*. New York: John Wiley & Sons.
- Schmidli, Hanspeter. 2017. *Risk Theory*. Cham: Springer.
- Shimizu, Yasutaka. 2012. Non-parametric estimation of the Gerber-Shiu function for the Wiener-Poisson risk model. *Scandinavian Actuarial Journal* 2012: 56–69. [\[CrossRef\]](#)
- Shimizu, Yasutaka, and Zhimin Zhang. 2017. Estimating Gerber-Shiu functions from discretely observed Lévy driven surplus. *Insurance: Mathematics and Economics* 74: 84–98. [\[CrossRef\]](#)
- Silverman, Bernard Walter. 1986. *Density Estimation for Statistics and Data Analysis*. Monographs on Statistics and Applied Probability. London: Chapman & Hall.
- Sparre Andersen, Erik. 1957. On the collective theory of risk in the case of contagion between the claims. *Bulletin of the Institute of Mathematics and its Applications* 2: 212–19.
- Vatamidou, Eleni, Ivo Jean Baptiste François Adan, Maria Vlasiou, and Bert Zwart. 2013. Corrected phase-type approximations of heavy-tailed risk models using perturbation analysis. *Insurance: Mathematics and Economics* 53: 366–78. [\[CrossRef\]](#)
- Vatamidou, Eleni, Ivo Jean Baptiste François Adan, Maria Vlasiou, and Bert Zwart. 2014. On the accuracy of phase-type approximations of heavy-tailed risk models. *Scandinavian Actuarial Journal* 2014: 510–34. [\[CrossRef\]](#)

