ON INTRINSIC ULTRACONTRACTIVITY OF PERTURBED LÉVY PROCESSES AND APPLICATIONS OF LÉVY PROCESSES IN ACTUARIAL MATHEMATICS

BY

BINGJI YI

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Mathematics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2017

Urbana, Illinois

Doctoral Committee:

Professor Renming Song, Chair
Associate Professor Runhuan Feng, Director of Research
Professor Richard B. Sowers
Assistant Professor Shu Li

Abstract

In this thesis, we study certain aspects of Lévy processes and their applications. In the first part of this thesis, we study the applications of Lévy processes in actuarial mathematics. Our topics are closely related to the generalized Ornstein-Uhlenbeck processes. We investigate their intimate relationships with the exponential functionals of Lévy processes, which enable us to develop efficient semi-analytical algorithms to solve the pricing and risk management problem of certain exotic variable annuity products. In particular, we consider two variable annuity products with guaranteed benefits, the Guaranteed Minimum Accumulation Benefit (GMAB) and the Guaranteed Minimum Withdrawal Benefit (GMWB). For the first one, we develop efficient semi-analytical algorithms to compute its risk measures and hedging costs to solve the risk management problem of the rider. For the other one, we consider pricing the rider. We identify the Laplace transforms of the GMWB rider's risk-neutral values analytically, which leads to efficient solutions to its pricing problem.

In the second part, we consider the intrinsic ultracontractivity of certain Lévy processes under nonlocal perturbations. More precisely, we establish the intrinsic ultracontractivity of the Laplacian (corresponding to Brownian motions) and the fractional Laplacian (corresponding to symmetric α-stable processes) perturbed by a class of nonlocal operators. Conditions on the nonlocal perturbations are given in order to guarantee that the perturbed operators are intrinsically ultracontractive in general bonded open sets. The methods we use are probabilistic. Essentially, the methods rely on the heat kernel estimates of the fundamental solutions of the operators as well as the Lévy systems of the corresponding processes.

To my beloved parents

Acknowledgments

This thesis would not have been possible without the support of many people. First I wish to express my sincere gratitude to my advisors: Dr. Runhuan Feng and Dr. Renming Song. I thank them for their guidance and support during the last five years. They guided me to do research in the field of applied probability and actuarial mathematics. Many thanks to them for revising my papers and thesis. Moreover, I want to thank Dr. Runhuan Feng for encouraging me to participate in research conferences to present our work. I also learned a lot from him in writing and giving presentations. In addition, I would like to thank Dr. Richard B. Sowers, Dr. Rick Gorvett and Dr. Shu Li for being on my committee and providing valuable suggestions.

I also want to thank all my friends from the University of Illinois, who made my time in graduate school enjoyable and unforgettable.

Finally, I would like to extend my deepest gratitude to my family, for their love and support. During the last five years, their support and encouragement always made me feel they were with me though they were thousands of miles away from me. Also, I would like to thank my girlfriend, Qingleng. No matter what difficulty I meet, she is always with me and provides the best encouragement.

Table of Contents

List of Abbreviations

c.d.f	Cumulative Density Function
GMAB	Guaranteed Minimum Accumulation Benefits
GMWB	Guaranteed Minimum Withdrawal Benefits
PDE	Partial Differential Equation(s)
p.d.f	Probability Density Function
SDE	Stochastic Differential Equation(s)
VA	Variable Annuity

List of Symbols

$\mathbb{C}$	The set of all complex numbers
$\mathbb{N}$	The set of all nonnegative integers
$\mathbb{R}^d$	The d-dimensional real Euclidean space
$\mathbb{R}^+$	The set of all positive real numbers
$\mathbb{Z}$	The set of all integers
$\mathbb{Z}^+$	The set of all positive integers
$C^k(X)$	The space of functions on X which have continuous derivatives up to order k
$C_b^k(X)$	The space of functions on X which have bounded continuous partial derivatives up to order k
$C_c^k(X)$	The space of functions on X which have compactly supported continuous partial derivatives up to order k
$C_0(X)$	The space of the continuous functions on X which vanish at ∞
$L^p(X, dx)$	The L^p space of functions on X with respect to measure dx
$\asymp$	If functions $f \asymp g$, there exists a positive constant c, such that $c^{-1}g \leq f \leq cg$
$:=$	The left-hand side is defined to be the right-hand side
$\stackrel{d}{=}$	The left-hand side is equal in law to the right-hand side
o	If $f(x) = o(g(x))$ as $x \to 0$ (or $x \to \infty$), $\lim_{x\to 0} \frac{f(x)}{g(x)} = 0$ (or $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0$)

Chapter 1

Introduction

Generally speaking, Lévy processes are a class of stochastic processes which have independent and stationary increments. Many stochastic processes that are widely used in Physics, Engineering, and Finance, such as Brownian motions, Poisson processes and stable processes are prototypes of Lévy processes. It can be seen as a random walk in continuous time and is related to many areas of probability theory: Markov processes, potential theory, and stochastic calculus. In this chapter, we present some facts of Lévy processes that will be used in later chapters. We start with a general review of basic properties, path structures and semigroups of Lévy processes. In the next section, we discuss Lévy processes with affine drift and the generalized Ornstein-Uhlenbeck processes. The generalized Ornstein-Uhlenbeck processes will also be discussed in detail in Chapter 2 and Chapter 3. In Section 1.2, we review the concepts of ultracontractive and intrinsically ultracontractive semigroups, which serves as the background and motivation for Chapter 4.

Definition 1.0.1. *Given a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$*, an* $\mathbb{R}^d$*-valued process* $X = \{X_t : t \geq 0\}$ *is said to be a Lévy process if it possesses the following properties:*

1. $X_0 = 0$ *a.s.*

2. *The sample paths of* X *are right continuous with left limits a.s.*

3. X *is stochastically continuous, i.e., for any* $t \geq 0$ *and* $\epsilon > 0$*,*

$$\lim_{s \to t} \mathbb{P}(|X_s - X_t| > \epsilon) = 0.$$

4. *For* $0 \leq s \leq t$*,* $X_t - X_s$ *is independent of* $\{X_u : u \leq s\}$ *and equal in law with* X_{t-s} *(independent and stationary increment).*

Lévy processes have an intimate connection with infinitely divisible distributions, which was discovered by de Finetti in 1929. The celebrated Lévy-Khinchine formula fully characterizes the characteristic functions of infinitely divisible distributions, which in turn characterizes the distribution of Lévy processes. We state the result in the following theorem ([46, Theorem 1.6]):

Theorem 1.0.2 (Lévy-Khintchine formula)**.** *For any $\mathbb{R}^d$-valued Lévy process $\{X_t : t \geq 0\}$ on a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$, define its characteristic exponent $\Psi(z)$ as*

$$e^{\Psi(z)t} := \mathbb{E}[e^{izX_t}] \qquad \text{for any } z \in \mathbb{R}^d.$$

then $\Psi(z)$ has the following representation

$$\psi(z) = i\langle \gamma, z\rangle - \frac{1}{2}\langle z, Az\rangle + \int_{\mathbb{R}^d} (e^{i\langle z,x\rangle} - 1 - i\langle z, x\rangle \mathbb{1}_{(|z|<1)})\pi(dx), \tag{1.1}$$

where $\gamma \in \mathbb{R}^d$, A is a symmetric non-negative-definite $d \times d$ matrix and π is a measure on $\mathbb{R}^d$ satisfying

$$\int_{\mathbb{R}^d} (1 \wedge |x|^2)\pi(dx) < \infty, \qquad \pi(\{0\}) = 0$$

which is called the Lévy measure of X. The triplet (γ, A, π) fully characterize the law of X in the sense that given any triplet (γ, A, π) satisfying the above conditions, we can construct a Lévy process on the canonical Skorokhod space $\mathbb{D}([0, \infty), \mathbb{R}^d)$ with characteristic exponent satisfying (1.1)*.*

A deeper result describing the path structure of Lévy processes is called the Lévy-Itô decomposition, which suggests that any Lévy process could be decomposed into three independent processes: a Brownian motion with drift, a compound Poisson process and a compensated jump process which is a square integrable martingale. The compound Poisson process and the square integrable martingale are driven by a Poisson point process on a certain space. For simplicity, we will just state the result for one-dimensional Lévy processes. We first introduce the concept of Poisson random measures ([46, Chapter III]).

Definition 1.0.3. *Let $(S, \mathcal{S}, \eta)$ be an arbitrary sigma-finite measure space and $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space. Let*

$$N : \Omega \times \mathcal{S} \to \mathbb{N}^+ \cup \{\infty\}$$

be a mapping satisfying the following conditions:

1. *For a.s. $\omega \in \Omega$, $N(\omega, \cdot)$ is a counting measure on $\mathcal{S}$,*
2. *For any $A \in \mathcal{S}$, $N(\cdot, A)$ is a $\mathcal{F}$-measurable Poisson random variable with intensity $\eta(A)$,*
3. *For mutually disjoint $A_1, \dots, A_n$ in $\mathcal{S}$, the variables $N(\cdot, A_1), \dots, N(\cdot, A_n)$ are independent.*

Then we say N is a Poisson random measure on $(S, \mathcal{S})$ with intensity η. From now on, we shall suppress the dependency of N on ω as it is easily understood.

From the definition 1.0.3, if we consider a Poisson random measure N on the product space $([0,\infty)\times\mathbb{R}, \mathcal{B}([0,\infty)\times\mathbb{R}))$ with intensity measure $dt\times\pi(dx)$, where $\pi(dx)$ is a measure concentrated on $\mathbb{R}\setminus\{0\}$. For any $A\in\mathcal{B}(\mathbb{R})$, with $\pi(A)<\infty$, we define

$$X_t := \int_{[0,t]}\int_A xN(ds\times dx).$$

It can be shown that X_t is a compound Poisson process with intensity $\pi(A)$ and jump distribution $\pi(A)^{-1}\pi(dx)|_A$ ([46, Chapter III]). With the concept of Poisson random measures, we are ready to state the main result of the Lévy-Itô decomposition ([46, Theorem 2.1]):

Theorem 1.0.4 (Lévy-Itô decomposition)**.** *For any $\mathbb{R}$-valued Lévy process $\{X_t : t\geq 0\}$ with characteristic exponent $\Psi(z) := \log\mathbb{E}[e^{izX_t}]$ satisfying*

$$\Psi(z) = i\gamma z - \frac{1}{2}\sigma^2 z^2 + \int_{\mathbb{R}}(e^{ixz} - 1 - ixz\mathbb{1}_{\{|z|<1\}})\pi(dx), \tag{1.2}$$

it is equal in law with the sum of three independent Lévy processes on a certain probability space $(\Omega,\mathcal{F},\mathbb{P})$. To be specific, we have

$$X_t \overset{d}{=} X_t^{(1)} + X_t^{(2)} + X_t^{(3)}$$

with $X_t^{(i)}$ having characteristic exponent $\Psi^{(i)}(z)$ ($i=1,2,3$), respectively. Furthermore, we have

- $\Psi^{(1)}(z) = i\gamma z - \frac{1}{2}\sigma^2 z^2$ *and* $X_t^{(1)} = \gamma t + \sigma W_t$, *which is a scaled Brownian motion with drift. W_t is a standard Brownian motion on $(\Omega,\mathcal{F},\mathbb{P})$,*

- $\Psi^{(2)}(z) = \int_{\{|z|>1\}}(e^{ixz}-1)\pi(dx)$ *and* $X_t^{(2)} = \int_{[0,t]}\int_{\{|x|>1\}} xN(ds\times dx)$, *which is a compound Poisson process. N is a Poisson random measure on $([0,\infty)\times\mathbb{R},\mathcal{B}([0,\infty)\times\mathbb{R})$ with intensity $dt\times\pi(dx)$,*

- $\Psi^{(3)}(z) = \int_{\{|z|\leq 1\}}(e^{ixz} - 1 - ixz\mathbb{1}_{(|z|<1)})\pi(dx)$ *and* $X_t^{(3)} = \lim_{\epsilon\to 0}\int_{[0,t]}\int_{\{\epsilon\leq|x|\leq 1\}} xN(ds\times dx)$, *which is the limit of a sequence of compensated compound Poisson processes. The limit $X_t^{(3)}$ is a square integrable martingale which is also a Lévy process. N is the same Poisson random measure in $X_t^{(2)}$.*

The Lévy-Itô decomposition thoroughly describes the jumps of a Lévy process using the underlying Poisson point process. We will use this fact to model the dynamics of a GMWB with the annual high step-up in Chapter 3. Next, we consider Lévy processes as Markov processes and review some properties related to their semigroups. For any given Lévy process $X=\{X_t : t\geq 0\}$ on the canonical Skorokhod space $\Omega = \mathbb{D}([0,\infty),\mathbb{R}^d)$, we consider the filtration $\{\mathcal{F}_t : t\geq 0\}$ associated with the coordinate mapping: $\mathcal{F}_t$ is the $\mathbb{P}$-completed sigma-filed generated by $\{X_s : s\leq t\}$. It is easy to see that the filtration $\{\mathcal{F}_t : t\geq 0\}$ satisfies the usual conditions ([41, Chapter 1]). It is well known that ([7, Chapter I]) $\{X_t, \mathcal{F}_t, t \geq 0\}$ is a strong Markov process on $(\Omega, \mathcal{F}, \mathbb{P})$. Recall the definition of semigroups and their generators ([25, Chapter I]):

Definition 1.0.5. *A family of bounded linear operators $\{P_t : t \geq 0\}$ on a Banach space B is called a (jointly continuous) semigroup if it satisfies the following relations:*

1. $P_0 = 1$.

2. *If $0 \leq s, t < \infty$, then*

$$P_s P_t = P_{s+t}$$

3. *The map*

$$t, f \to P_t f$$

from $[0, \infty) \times B$ to B is jointly continuous.

If we further have $||P_t|| \leq 1$ for all $t \geq 0$, then $\{P_t : t \geq 0\}$ is called a contraction semigroup. If the convergence $P_t f \to f$ as $t \to 0$ is in norm for any fixed $f \in B$, we say that $\{P_t : t \geq 0\}$ is strongly continuous.

Definition 1.0.6. *The (infinitesimal) generator $\mathcal{L}$ of a semigroup P_t is defined by*

$$\mathcal{L}f := \lim_{t \to 0} t^{-1}(P_t f - f).$$

The domain $Dom(\mathcal{L})$ of $\mathcal{L}$ is the set of f for which the limit exists. It is a linear subspace of the Banach space B.

We introduce a family of probability measures $\{\mathbb{P}_x : x \in \mathbb{R}^d\}$ by letting $\mathbb{P}_x$ be the law of $\{X_t + x : t \geq 0\}$ under $\mathbb{P}$. Denote the expectation operator associated with $\mathbb{P}_x$ by $\mathbb{E}_x$. We can define the semigroup associated to the Lévy process X_t by

$$P_t f(x) := \mathbb{E}_x[f(X_t)].$$

Using the Markov property of X_t, we can easily verify P_t is a semigroup on $L^\infty(\mathbb{R}^d)$. More precisely, the semigroup P_t associated with the Lévy process X_t is a Feller semigroup. In other words, we have the following theorem.

Theorem 1.0.7. *The semigroup $\{P_t : t \geq 0\}$ has the Feller property, that is, for any $f \in C_0(\mathbb{R}^d)$:*

1. $P_t f \in C_0(\mathbb{R}^d)$ *for any $t \geq 0$,*

2. $\lim_{t\to 0} ||P_t f - f|| = 0$ *where* $||\cdot||$ *is the supreme norm on* $C_0(\mathbb{R}^d)$*.*

Proof. See [7, Proposition I.5]. □

Next, we characterize the generator $\mathcal{L}$ of the semigroup P_t. Suppose the Lévy process $\{X_t : t \geq 0\}$ in $\mathbb{R}^d$ has the characteristic exponent described in (1.1). Let $\mathscr{S}$ (the Schwartz space) be the set of C^∞ functions f on $\mathbb{R}^d$ such that

$$\lim_{|x|\to\infty} |x|^m |D^\alpha f(x)| = 0,$$

for all integers m and all mixed partial derivatives $D^\alpha f$ of f. $\alpha = (\alpha_1, \ldots, \alpha_n)$ is a multi-index of nonnegative integers. We have the following theorem regarding the generator $\mathcal{L}$ of the Lévy process X_t.

Theorem 1.0.8 (Generators of Lévy processes)**.** *The generator $\mathcal{L}$ of the Feller semigroup $\{P_t : t \geq 0\}$ associated with the Lévy process $\{X_t : t \geq 0\}$ is given by*

$$\mathcal{L}f(x) = \langle \gamma, \nabla f(x) \rangle + \frac{1}{2}\sum_{i,j} A_{i,j} \frac{\partial^2 f(x)}{\partial x_i \partial x_j} + \int_{\mathbb{R}^d} \left(f(x+y) - f(x) - \mathbb{1}_{\{|y|<1\}} \langle y, \nabla f(x) \rangle \right) \pi(dy)$$

for any $f \in \mathscr{S}$. $\mathscr{S} \subset Dom(\mathcal{L})$ is a core (as defined in [25, Chapter I]*) for the generator $\mathcal{L}$ of the Lévy process.*

Proof. Refer to [58, Chapter VII.1]. □

We summarize this part by giving some examples of Lévy processes.

Example 1.0.9 (The Brownian motion)**.** *An $\mathbb{R}^d$-valued Lévy process $W = \{W_t : t \geq 0\}$ is called a Brownian motion if it also satisfies*

$$W_t \stackrel{d}{=} \mathcal{N}(0, t),$$

where $\mathcal{N}(0,t)$ is a normal random variable with mean 0 *and variance t. The characteristic exponent of a Brownian motion is given by*

$$\mathbb{E}e^{i\langle z, W_t \rangle} = e^{-\frac{|z|^2 t}{2}}, \qquad \text{for any } z \in \mathbb{R}^d \ .$$

The Lévy triplet of W_t is given by $(\gamma, A, \pi) = (0, I, 0)$. The generator of $W = \{W_t : t \geq 0\}$ is the Laplacian $\frac{1}{2}\Delta$ by Theorem 1.0.8. (Linear) Brownian motions are the only type of Lévy processes which have continuous sample paths.

Example 1.0.10 (The α-stable process)**.** *An $\mathbb{R}^d$-valued Lévy process $Z = \{Z_t : t \geq 0\}$ is called a (rotation-*

ally) symmetric α-stable process if it has the characteristic exponent

$$\mathbb{E}[e^{i\langle z, Z_t\rangle}] = e^{-t|z|^\alpha}, \qquad \textit{for any } z \in \mathbb{R}^d \textit{ and a certain } \alpha \in (0,2).$$

α is called the stable index of $Z = \{Z_t : t \geq 0\}$, which entails the scaling property

$$\{Z_{\lambda t} : t \geq 0\} \stackrel{d}{=} \{\lambda^{1/\alpha} Z_t : t \geq 0\}, \textit{for any } \lambda > 0.$$

The generator of $Z = \{Z_t : t \geq 0\}$ is given by the fractional Laplacian $\Delta^{\alpha/2}$ on $\mathbb{R}^d$, which is defined as

$$\Delta^{\alpha/2} f(x) := \int_{\mathbb{R}^d} \left(f(x+y) - f(x) - \mathbb{1}_{|y|<1} \langle y, \nabla f(x)\rangle \right) \frac{\mathcal{A}(d,-\alpha)}{|y|^{d+\alpha}} dy, \qquad \textit{for } f \in \mathscr{S},$$

with $\mathcal{A}(d,-\alpha) = \alpha 2^{\alpha-1}\pi^{-d/2}\Gamma(\frac{d+\alpha}{2})\Gamma(1-\frac{\alpha}{2})^{-1}$. Let $\widehat{f}(\xi) := \int_{\mathbb{R}^d} e^{i\langle \xi, x\rangle} f(x)dx$ be the Fourier transform of a function f on $\mathbb{R}^d$, so we have $\widehat{\Delta^{\alpha/2} f}(\xi) = -|\xi|^\alpha \widehat{f}(\xi)$, which justifies the notation $\Delta^{\alpha/2}$.

1.1 Generalized Ornstein-Uhlenbeck processes

In Chapter 2 and 3, our modeling for certain exotic variable annuity products is closely related to a class of stochastic processes which is called the generalized Ornstein-Uhlenbeck processes in the literature. Let $\{(X_t, Y_t) : t \geq 0\}$ be a two-dimensional Lévy process starting from $(0,0)$ and

$$U_t := xe^{X_t} + e^{X_t} \int_0^t e^{-X_{s-}} dY_s.$$

$\{U_t : t \geq 0\}$ is called a generalized Ornstein-Uhlenbeck process associated to the two-dimensional Lévy process $\{(X_t, Y_t), t \geq 0\}$ (For example, see definition in [14, Appendix I]). From [14, Corollary 5.2], we know that $\{U_t, t \geq 0\}$ is a homogeneous Markov process. In this thesis, we only consider the special case $Y_t = t$, which means the generalized Ornstein-Uhlenbeck process $\{U_t : t \geq 0\}$ is determined by the Lévy process X_t. In Chapter 2, we consider the case X_t is a Brownian motion with drift. Using Itô's rule, U_t can be shown to satisfy an SDE. The SDE is the starting point for us to identify the marginal distribution of $\{U_t : t \geq 0\}$ analytically. The analytical results of the marginal distribution of $\{U_t : t \geq\}$ lead to an efficient semi-analytical algorithm to compute the risk measures of a GMAB. In Chapter 3, we consider a more general case when X_t is the independent sum of a linear Brownian motion and a compound Poisson process with mixed-exponential jumps. It was observed in the literature (See, for example, [14]) that there exists an intimate relationship between the generalize Ornstein-Uhlenbeck process $\{U_t : t \geq 0\}$ and the exponential functional of Lévy processes $I_t := \int_0^t e^{X_s} ds$. It turns out that certain functionals of $\{U_t : t \geq 0\}$ are related to the tail distribution of the exponential functional I_t. This fact together with recent developments on exponential functionals of Lévy processes enables us to solve the pricing problem of a GMWB.

1.2 Ultracontractive and intrinsically ultracontractive semigroups

The concepts of ultracontractivity and intrinsic ultracontractivity were first introduced and thoroughly discussed by Davies and Simon in 1984 ([26]) for symmetric semigroups. Generally speaking, ultracontractivity aims to describe the L^∞-properties of contraction semigroups. It characterizes a class of semigroups which has stronger contraction properties that map L^2 to L^∞. To be specific, we recall the definition for ultracontractive semigroups in [26].

Suppose X is a locally compact second countable Hausdorff space and dx is a regular Borel measure on X whose support is also equal to X. Let Z be a nonnegative self-adjoint closed operator on $L^2(X, dx)$. By the convention of denoting semigroups, let e^{-Zt} represent the symmetric contraction semigroup generated by Z. We assume that the semigroup e^{-Zt} has a nonnegative and jointly continuous transition density $p(t, x, y)$, i.e

$$e^{-Zt} f(x) = \int_X p(t, x, y) f(y) dy, \qquad \text{for any } f \in L^2(X, dx).$$

Therefore, e^{-Zt} is also positivity-preserving. Also, we assume

$$\int_X p(t, x, x) dx < \infty, \qquad \text{for any } t > 0.$$

Under these conditions, the operator Z has a purely discrete spectrum (See [26, p339]). We denote the eigenvalues of Z in increasing order as $\lambda_0 < \lambda_1 \leq \lambda_2 \leq \cdots \to \infty$ and let $\phi_0, \phi_1, \phi_2, \ldots$ be the eigenfunctions corresponding to each eigenvalue. Without loss of generality, we can normalize the eigenfunctions such that $||\phi_i||_2 = 1$ for any $i \in \mathbb{N}$. Under the assumption that e^{-Zt} is irreducible ([26, p341]), λ_0 has multiplicity one and ϕ_0 can be chosen to be strictly positive. By the spectrum expansion ([26, Lemma 2.1]), the transition density $p(t, x, y)$ has a series representation

$$p(t, x, y) = \sum_{i=0}^{\infty} e^{-\lambda_i t} \phi_i(x) \phi_i(y),$$

where the series on the right-hand side is locally uniformly convergent on $X \times X$.

Definition 1.2.1. *If the contraction semigroup e^{-Zt} is a bounded operator from $L^2(X, dx)$ to $L^\infty(X, dx)$ for all $t > 0$, e^{-Zt} is ultracontractive.*

Next, we introduce the concept of intrinsic ultracontractivity for symmetric semigroups ([26]). We observe that it is not necessarily true that the semigroup e^{-Zt} is sub-Markovian in the sense that

$$e^{-Zt}1 \le 1.$$

Therefore, we consider a change of measure: $\mu(dx) = \phi_0^2(x)dx$. Define a new semigroup

$$e^{-\overline{Z}t} f(x) := \int_X e^{-\lambda_0 t} \frac{p(t,x,y)\phi_0(y)}{\phi_0(x)} dy.$$

It is easy to see that $e^{-\overline{Z}t}$ has eigenvalues $1 > e^{-(\lambda_1-\lambda_0)t} \ge e^{-(\lambda_2-\lambda_0)t}, \cdots \to 0$ and eigenfunctions $1, \frac{\phi_1}{\phi_0}, \frac{\phi_2}{\phi_0}, \ldots$. Specifically, we have

$$e^{-\overline{Z}t}1 = 1,$$

which shows that $e^{-\overline{Z}t}$ is a Markovian semigroup. Under the reference measure $\mu(dx)$, the transition density of $e^{-\overline{Z}t}$ is given by

$$\overline{p}(t,x,y) := \frac{e^{-\lambda_0 t} p(t,x,y)}{\phi_0(x)\phi_0(y)} \tag{1.3}$$

and it can be shown that ([57, p255]) $e^{-\overline{Z}t}$ is a contraction on all L^p spaces. Then we recall the definition of intrinsic ultracontractivity in [26].

Definition 1.2.2. *The semigroup e^{-Zt} on $L^2(X, dx)$ is called intrinsically ultracontractive if the corresponding semigroup $e^{-\overline{Z}t}$ defined in (1.3) is ultracontractive on $L^2(X, \mu(dx))$.*

Remark 1.2.3. [42, Proposition 2.2] *showed that if a symmetric semigroup $e^{-\overline{Z}_t}$ is sub-Markovian, $e^{-\overline{Z}_t}$ is bounded from $L^2(X, dx)$ to $L^\infty(X, dx)$ is equivalent to its transition density $\overline{p}(t,x,y)$ is bounded by a constant $c_t > 0$ for any t. Therefore, the previous definition of the semigroup e^{-Zt} is intrinsically ultracontractive is equivalent to there exists a constant $c_t > 0$ such that*

$$p(t,x,y) < c_t \phi_0(x)\phi_0(y), \qquad \textit{for any } x, y \in X.$$

It follows from [26, Theorem 3.2] that if e^{-Zt} is intrinsically ultracontractive, it is also true that

$$c'_t\phi_0(x)\phi_x(y) < p(t,x,y)$$

for another constant $c'_t > 0$ depending on t. Therefore, it indicates that if a semigroup is intrinsically ultracontractive, the product of its ground state eigenfunctions $\phi_0(x)\phi_0(y)$ is comparable to its transition density, which provides a good qualitative characterization of the transition density. This is one of the reasons we are interested in intrinsically ultracontractive semigroups.

The concepts of ultracontractivity and intrinsic ultracontractivity were extended to non-symmetric semigroups by Kim and Song ([42]). As stated by Kim and Song, things are much more delicate in the non-symmetric case. In the rest of this section, we review the definition and some basic properties of intrinsically ultracontractive non-symmetric semigroups, which were introduced and obtained by Kim and Song in [42]. Similar to the case of symmetric semigroups, let X be a locally compact second countable Hausdorff space and dx be a finite regular Borel measure whose support is equal to X. Let $\{P_t : t \geq 0\}$ be a (non-symmetric) semigroup on $L^2(X, dx)$. We further assume that $\{P_t : t \geq 0\}$ admits a jointly continuous and positive transition density $p(t,x,y)$, i.e for any $f \in L^2(X, dx)$, we have

$$P_t f(x) = \int_X p(t,x,y)f(y)dy.$$

Suppose there exists a dual semigroup $\{\widehat{P}_t : t \geq 0\}$ of $\{P_t : t \geq 0\}$ with respect to the reference measure dx, which means for any $f, g \in L^2(X, dx)$, we have

$$\int_X g(x)P_t f(x)dx = \int_X f(x)\widehat{P}_t g(x)dx.$$

Therefore, the transition density of semigroup $\{\widehat{P}_t : t \geq 0\}$ is $p(t,y,x)$ since

$$\widehat{P}_t g(x) = \int_X p(t,y,x)g(y)dy. \qquad \text{for any } g \in L^2(X, dx).$$

Recall the definition of ultracontractivity for non-symmetric semigroups in [42],

Definition 1.2.4. *The semigroup $\{P_t : t \geq 0\}$, together with its dual semigroup $\{\widehat{P}_t : t \geq 0\}$, is called ultracontractive if for any $t > 0$, there exists a positive constant c_t so that*

$$p(t,x,y) < c_t, \qquad \text{for any } x, y \in X.$$

Now we also assume $\{P_t : t \geq 0\}$ and $\{\widehat{P}_t : t \geq 0\}$ are sub-Markovian semigroups . As we mentioned in Remark 1.2.3, the equivalence between Definition 1.2.1 and Definition 1.2.4 when $\{P_t : t \geq 0\}$ is symmetric is proved by [42, Proposition 2.2].

To define intrinsic ultracontractivity for non-symmetric semigroups, [42] stated that we also need to assume that $\{P_t : t \geq 0\}$ and $\{\widehat{P}_t : t \geq 0\}$ are strongly continuous and their transition density $p(t, x, y)$ is strictly positive and bounded. Denote the generators of the semigroups $\{P_t : t \geq 0\}$ and $\{\widehat{P}_t : t \geq 0\}$ by Z and $\widehat{Z}$, respectively. Then as stated in [42, p529], we can find a common eigenvalue λ_0 of multiplicity 1 which lies on the top of the spectrums of the operators Z and $\widehat{Z}$ due to Jentzsch's Theorem and the strong continuity. Let ϕ_0 and ψ_0 be the normalized eigenfunctions of the eigenvalue λ_0 with respect to the operators Z and $\widehat{Z}$. By [42, Proposition 2.3], we know that ϕ_0 and ψ_0 can be chosen to be strictly positive and continuous in X. We introduce the following change of measure,

$$\mu(dx) = \phi_0(x)\psi_0(x)dx,$$

which is an analog of the symmetric case. The new transition densities with respect to the Lebesgue measure are given by

$$\begin{aligned} q(t, x, y) &:= \frac{e^{-\lambda_0 t}}{\phi_0(x)} p(t, x, y)\phi_0(y) \\ \widehat{q}(t, x, y) &:= \frac{e^{-\lambda_0 t}}{\psi_0(x)} p(t, y, x)\psi_0(y). \end{aligned} \tag{1.4}$$

Therefore, the new semigroups corresponding to the new transition densities in (1.4) are given by

$$\begin{aligned} Q_t f(x) &:= \int_X q(t, x, y) f(y) dy \\ \widehat{Q}_t f(x) &:= \int_X \widehat{q}(t, x, y) f(y) dy. \end{aligned} \tag{1.5}$$

The reason to consider the change of measure to $\mu(dx)$ is the same as the symmetric case: the ground state eigenvalue λ_0 and eigenfunctions ϕ_0 and ψ_0 are normalized to 1 for the new semigroups. It is obvious that

$$Q_t 1 = \widehat{Q}_t 1 = 1.$$

Moreover, we can also observe that under the new measure $\mu(dx)$, the transition density of the semigroups

$\{Q_t : t \geq 0\}$ and $\{\widehat{Q}_t : t \geq 0\}$ are given by

$$\bar{q}(t,x,y) := e^{-\lambda_0 t} \frac{p(t,x,y)}{\phi_0(x)\psi_0(y)}$$

and

$$\bar{q}(t,y,x) = e^{-\lambda_0 t} \frac{p(t,y,x)}{\phi_0(y)\psi_0(x)},$$

respectively. It is easy to see that $\{Q_t : t \geq 0\}$ and $\{\widehat{Q}_t : t \geq 0\}$ are dual semigroups on $L^2(X, \mu(dx))$, i.e

$$\int_X g(x) Q_t f(x) \mu(dx) = \int_X f(x) \widehat{Q}_t g(x) \mu(dx), \qquad \text{for any } f, g \in L^2(X, \mu(dx)).$$

Letting $g = 1$ (or $f = 1$) in the previous equation, it is easy to see that $\mu(dx)$ is the stationary distribution of $\{Q_t : t \geq 0\}$ (or $\{\widehat{Q}_t : t \geq 0\}$). Now we can introduce the definition of intrinsic ultracontractivity of non-symmetric semigroups ([42, Definition 2.4]).

Definition 1.2.5. *The semigroups $\{P_t : t \geq 0\}$ and $\{\widehat{P}_t : t \geq 0\}$ are called intrinsically ultracontractive if their corresponding semigroups $\{Q_t : t \geq 0\}$ and $\{\widehat{Q}_t : t \geq 0\}$ defined in* (1.5) *are ultracontractive in $L^2(X, \mu(dx))$.*

Remark 1.2.6. *The previous definition is equivalent to there exists $c_t > 0$ such that*

$$\overline{q}(t,x,y) < c_t, \qquad \textit{for any } x, y \in X.$$

In terms of the original transition density $p(t,x,y)$, we have

$$p(t,x,y) < c_t \phi_0(x)\psi_0(y), \qquad \textit{for any } x, y \in X.$$

From [42, Proposition 2.5]*, if $\{P_t, t \geq 0\}$ is intrinsically ultracontractive with*

$$p(t,x,y) < c_t \phi_0(x)\psi_0(y), \qquad \textit{for any } x, y \in X,$$

it is also true that

$$p(t,x,y) > c'_t \phi_0(x)\psi_0(y), \qquad \textit{for any } x, y \in X,$$

for another constant $c'_t > 0$. This shows that the transition densities of non-symmetric intrinsically ultracontractive semigroups are also comparable to the product of ground state eigenfunctions.

This concludes the introduction of ultracontractivity and intrinsic ultracontractivity. In Chapter 4, we shall consider the intrinsic ultracontractivity of perturbed Brownian motions and α-stable processes. In fact, our analysis is motivated by perturbing their generators. As we introduced earlier in this chapter, the generator of a standard Brownian motion is Δ and the generator of an α-stable process is Δ^{α} $(0 < \alpha < 2)$. In Chapter 4, we shall consider the semigroups obtained by adding a class of nonlocal perturbations to the Laplacian and the fractional Laplacian. The existence of the Markov processes corresponding to the perturbed operators was shown by [21, 64]. We consider killing these Markov processes upon leaving bounded open sets and establishing the intrinsic ultracontractivity of the killed processes.

Chapter 2

Geometric Brownian Motions with Affine Drift and GMAB Risk Management

2.1 The distribution of geometric Brownian motions with affine drift: a spectral theory result

We consider the SDE

$$dX_t = \left((\mu + \frac{\sigma^2}{2})X_t + w\right) dt + \sigma X_t dW_t, \qquad X_0 = x_0, \tag{2.1}$$

where $\mu, w \in \mathbb{R}$, $\sigma > 0$ and W_t is a standard Brownian motion. A weak solution X_t to this SDE is referred as the geometric Brownian motion with affine drift in the literature. Using Ito's rule, we can find

$$X_t = x_0 e^{\mu t + \sigma W_t} + w e^{\mu t + \sigma W_t} \int_0^t e^{-\mu s - \sigma W_s} ds \tag{2.2}$$

is a solution to (2.1). The solution (2.2) is also called a generalized Ornstein-Uhlenbeck process introduced in Chapter 1. We shall observe that this process has many applications to the quantitative risk management of variable annuity products. By a well-known time reversal argument of Lévy processes ([14, Lemma 2.3]), we can find that

$$x_0 e^{\mu t + \sigma W_t} + w e^{\mu t + \sigma W_t} \int_0^t e^{-\mu s - \sigma W_s} ds \overset{d}{=} x_0 e^{\mu t + \sigma W_t} + w \int_0^t e^{\mu s + \sigma W_s} ds, \tag{2.3}$$

which shows the intimate relationship between generalized Ornstein-Uhlenbeck processes and exponential functionals of Brownian motions. This identity in law plays an important role in solving the risk management problem of GMAB products, which we shall analyze in detail later. The process on the right-hand side of (2.3) is generally not a Markov process, its distribution is not easy to work with directly. The process on the left-hand side of (2.3), however, is a diffusion process (the generalized Ornstein-Uhlenbeck process). We can determine its transition density by solving the Kolmogorov equations. To be noted, we can use the scaling property of Brownian motions to simplify (2.1). To be specific, in (2.2) we do a change of parameters:

$$\begin{cases} v := \frac{2\mu}{\sigma^2}, \\ t' := \frac{\sigma^2 t}{4}, \\ x'_0 = \frac{\sigma^2}{4w}. \end{cases}$$

By the scaling property of Brownian motions ([41, p104, Lemma 9.4]), we have the following identities in law:

$$\begin{cases} e^{\mu t+\sigma B_t} = e^{2(\frac{\mu}{\sigma^2}\frac{\sigma^2 t}{4}+\frac{\sigma}{2}W_t)} \overset{d}{=} e^{2vt'+2W_{t'}}, \\ \int_0^t e^{\mu s+\sigma B_s}ds \overset{d}{=} \frac{4}{\sigma^2}\int_0^{t'} e^{2vs+2W_s}ds. \end{cases}$$

Therefore, if we let $\overline{X}_t := \frac{\sigma^2}{4w}X_t$ where X_t is given in (2.2), $\overline{X}_t$ should satisfy the following SDE

$$d\overline{X}_t = ((2v+1)\overline{X}_t + 1)dt + 2\overline{X}_t dW_t, \qquad \overline{X}_0 = x'_0, \tag{2.4}$$

with $w > 0$. Many previous works were done to characterize the distribution of (2.4). Among them, the spectral expansion results obtained by Linetsky ([49, 50]), Feng and Volkmer ([34]) are efficient and robust for numerical evaluations. Linetsky ([50]) first identified the distribution of (2.4) by integration formulas in terms of the Laguerre polynomials. Later, Feng and Volkmer ([34]) identified the Laplace transforms of various functionals of the process $\overline{X}_t$. The results of the following theorem are obtained by inverting the Laplace transforms obtained by Feng and Volkmer ([34]) using the Bromwich integral. The integration formulas involves Whitaker functions, which can be shown to be equivalent to Linetsky's results in [50]. We summarize the results.

Theorem 2.1.1. *For the diffusion process $\{\overline{X}_t : t \geq 0\}$ satisfying SDE* (2.4)*, we have*

1. *The transition density $p(t, x_0, x) := \mathbb{P}_{x_0}(\overline{X}_t \in dx)/dx$ is given by*

$$\begin{aligned} &p(t, x_0, x) \\ =&\frac{1}{2\pi^2}\left(\frac{x}{x_0}\right)^{\frac{v-1}{2}} e^{\frac{1}{4x_0}-\frac{1}{4x}} \int_0^\infty e^{\frac{-(v^2+y^2)t}{2}} \mathcal{W}_{\frac{1-v}{2},\frac{yi}{2}}\left(\frac{1}{2x_0}\right)\mathcal{W}_{\frac{1-v}{2},\frac{yi}{2}}\left(\frac{1}{2x}\right)\left|\Gamma(\frac{v+yi}{2})\right|^2 \sinh(y\pi)ydy \\ &+\mathbb{1}_{\{v<0\}}\sum_{n=0}^{[\frac{-v}{2}]}(-1)^n e^{2n(v+n)t}\frac{2(-v-2n)}{n!\Gamma(1-v-2n)}(\frac{x}{x_0})^{\frac{v-1}{2}} e^{\frac{1}{4x_0}-\frac{1}{4x}}\mathcal{M}_{\frac{1-v}{2},-n-\frac{v}{2}}(\frac{1}{2x_0})\mathcal{W}_{\frac{-v-1}{2},-n-\frac{v}{2}}(\frac{1}{2x}). \end{aligned} \tag{2.5}$$

2. *The cumulative distribution function* $C(t,x_0,x) := \mathbb{P}_{x_0}(\overline{X}_t \le x)$ *is given by*

$$
\begin{aligned}
&C(t,x_0,x)\\
=&\frac{x_0}{2\pi^2}(\frac{x}{x_0})^{\frac{1+v}{2}}e^{\frac{1}{4x_0}-\frac{1}{4x}}\int_0^\infty e^{\frac{-(v^2+y^2)t}{2}}\mathcal{W}_{\frac{1-v}{2},\frac{yi}{2}}(\frac{1}{2x_0})\mathcal{W}_{\frac{-1-v}{2},\frac{yi}{2}}(\frac{1}{2x})\mid\Gamma(\frac{v+yi}{2})\mid^2\sinh(y\pi)ydy\\
+&\mathbb{1}_{\{v<0\}}\sum_{n=0}^{[\frac{-v}{2}]}(-1)^n e^{2n(v+n)t}\frac{2(-v-2n)x_0}{n!\Gamma(1-v-2n)}(\frac{x}{x_0})^{\frac{1+v}{2}}e^{\frac{1}{4x_0}-\frac{1}{4x}}\mathcal{M}_{\frac{1-v}{2},-n-\frac{v}{2}}(\frac{1}{2x_0})\mathcal{W}_{\frac{-v-1}{2},-n-\frac{v}{2}}(\frac{1}{2x}).
\end{aligned}
\tag{2.6}
$$

3. *The put payoff function* $P(t,x_0,x) := \mathbb{E}_{x_0}\left[(x-\overline{X}_t)_+\right]$ *is given by*

$$
\begin{aligned}
&P(t,x_0,x)\\
=&\frac{x_0^2}{2\pi^2}(\frac{x}{x_0})^{\frac{3+v}{2}}e^{\frac{1}{4x_0}-\frac{1}{4x}}\int_0^\infty e^{\frac{-(v^2+y^2)t}{2}}\mathcal{W}_{\frac{1-v}{2},\frac{yi}{2}}(\frac{1}{2x_0})\mathcal{W}_{\frac{-3-v}{2},\frac{yi}{2}}(\frac{1}{2x})\mid\Gamma(\frac{v+yi}{2})\mid^2\sinh(y\pi)ydy\\
+&\mathbb{1}_{\{v<0\}}\sum_{n=0}^{[\frac{-v}{2}]}(-1)^n e^{2n(v+n)t}\frac{2(-v-2n)x_0^2}{n!\Gamma(1-v-2n)}(\frac{x}{x_0})^{\frac{3+v}{2}}e^{\frac{1}{4x_0}-\frac{1}{4x}}\mathcal{M}_{\frac{1-v}{2},-n-\frac{v}{2}}(\frac{1}{2x_0})\mathcal{W}_{\frac{-v-3}{2},-n-\frac{v}{2}}(\frac{1}{2x}).
\end{aligned}
\tag{2.7}
$$

4. *The truncated expectation* $Z(t,x_0,x) := \mathbb{E}_{x_0}[\overline{X}_t\mathbb{1}_{\{\overline{X}_t<x\}}]$ *is given by*

$$Z(t,x_0,x) = xC(t,x_0,x) - P(t,x_0,x). \tag{2.8}$$

where $\mathcal{W}_{\cdot,\cdot}(\cdot)$, $\mathcal{M}_{\cdot,\cdot}(\cdot)$ *and* $\Gamma(\cdot)$ *are the Whittaker-W, Whittaker-M and Gamma function (See appendix for their definitions), respectively.*

Readers are referred to Feng and Volkmer ([34]) for the Laplace transforms of $C(t,x_0,x)$ and $Z(t,x_0,x)$ defined in the previous theorem and Linetsky ([50]) for equivalent representations. Using these integration formulas, we develop efficient algorithms to determine the risk measures and hedging costs of the GMAB rider in the next section.

2.2 Guaranteed Minimum Accumulation Benefit

A variable annuity is a tax-deferred retirement planning instrument that allows policyholders to choose from a selection of investment options and then pays back various types of benefits determined by the performance of investment portfolio of policyholder's choosing. The guaranteed minimum benefits are in nature similar to payoffs of exotic options in financial markets. For example, the guaranteed minimum maturity benefit offers a policyholder the greater of a guaranteed minimum amount and the balance of their investment accounts, should the policyholder survive to maturity. This can be viewed as a put option contingent on the survival of the investor. Such similarities exist all across the board with many types of guaranteed benefits. There is a tremendous amount of literature on the pricing of guaranteed minimum benefits. For example, Brennan and Schwarz ([11]), Boyle and Schwartz ([10]) were among some pioneering works in this field. Bauer et al. ([5]), Bacinello et al. ([3]), Piscopo and Haberman ([56]) considered the valuation of different guaranteed benefits in a unifying framework using Monte Carlo simulations. Hardy ([39]) systematically exploited the risk-neutral pricing and dynamic hedging of different guaranteed benefits under the Black-Scholes and regime-switching models. Milevsky and Posner ([52]), Milevsky and Salisbury ([53]), Ulm ([63]) considered pricing problems of guaranteed minimum death benefits with various product features and mortality assumptions. In Milevsky and Salisbury ([51]), Dai et al. ([28]), Chen et al. ([20]), Feng and Vecer ([32]), PDE-based numerical schemes were developed for pricing guaranteed minimum withdrawal benefits with various withdrawal strategies. Peng et al. ([55]), Feng and Volkmer ([35]), Feng and Jing ([31]) developed analytical solutions to the pricing of guaranteed minimum withdrawal benefits and guaranteed lifetime withdrawal benefits.

However, there have been few papers on the modeling of risk management problems for guaranteed benefits. Hardy ([39]) compared the traditional actuarial risk management and dynamic hedging strategy of different variable annuity guaranteed benefits. Coleman et al. ([27]) compared delta and dynamic risk minimization hedging strategies under a jump diffusion volatility model. Feng and Volkmer ([33, 34]) used analytical methods to calculate the risk measures of guaranteed minimum withdrawal benefits and guaranteed minimum death benefits. Feng et al. ([30]) used comonotonic approximation technique to estimate risk measures for guaranteed benefits with dynamic policyholder behavior. Feng and Huang ([29]) formulated the statutory financial reporting of variable annuity death benefits.

As guaranteed benefits can largely be viewed as exotic options, the risk management of guaranteed minimum benefits may be naturally considered as option pricing problems, which lead to dynamic hedging strategies. However, one should note that there are at least three technical issues regarding the risk management of variable annuity that are considerably different from that of exchange-traded financial instruments.

1. While most exchange-traded options are short-lived with typical terms arranging from a few months to a year, variable annuities last for many years as an investment vehicle for retirement planning. The long-term nature of the products require projections of cash flows that can be affected by many economic factors.

2. Unlike most financial instruments that require an up-front fee, the costs of variable annuity guaranteed

benefits are compensated by collections of asset-based fees from policyholders' investment accounts. Therefore, the financial risks embedded in the insurer's liability side are also present on the income side. In adverse economic scenarios, high liabilities may be accompanied by low incomes, which exacerbates the severity of potential losses to the insurers.

3. Financial options are typically traded by institutional investors. Guaranteed minimum benefits, however, are sold to individual investors and therefore often involves the interaction of both financial risks and mortality risks, a unique feature to equity-linked insurance products.

In this chapter of the thesis, we propose a quantitative framework for comparing risk management methodologies of variable annuity guaranteed benefits, a problem which has never been analyzed through other methods than Monte Carlo simulations in the literature. There are at least two risk management methodologies both widely used in the insurance industry.

1. **Dynamic hedging: (Risk transference)** The main principle of this risk management approach is to view market-risk-sensitive insurance products as complex financial products and to use modern option pricing theory to develop hedging strategies to cover insurance liabilities from embedded options. Since gains or losses of hedging portfolios are expected to offset those of insurers' product liabilities, the market risks associated with the products are effectively transferred to the capital markets. The success of such a risk management program would largely depend on the insurers' capabilities to accurately measure their exposures and sensitivities to various market conditions in order to develop offsetting hedging portfolios.

2. **Risk measure based reserving: (Risk acceptance)** This risk management approach is largely based on time-honored reserving and capital requirement methods that were originally developed for traditional life insurance. The main strategy is to set aside sufficient liquid assets providing a buffer against unexpected large losses under adverse economic conditions over the long term. An internal or external risk management policy should be in place to ensure that the insurer has the capacity to retain the market risks associated with the product liabilities. The insurer would typically develop an economic scenario generator calibrated by some industry standards in order to project financial conditions in the future. All product liabilities are stress-tested under projected scenarios of all relevant risk factors. Such an exercise would lead to a quantitative assessment of the insurer's risk profile, from which reserves and capital requirements are determined based on certain risk measures.

While both approaches have been adopted in the life insurance industry, it is a trend that most insurance companies offering major market-risk-sensitive products move toward dynamic hedging approach or some hybrid approach. The general perception is that the hedging approach is more cost effective than the traditional reserving approach. However, very few academic research studies have examined the two approaches in the same quantitative framework, let alone the question of whether such a perception reflects the reality. We intend to fill the gap in the literature by taking the guaranteed minimum accumulation benefit (GMAB) as a model example to investigate the similarities and differences of the two approaches.

2.2.1 Modeling and risk management

We consider a two-period GMAB rider with an automatic renewal at the end of the first period. As with any base contract of variable annuities, policyholders make initial purchase payments, or also known as premiums, into certain investment accounts of their choosing, which are usually managed by third-party vendors and grow with the market value of certain equity indices or funds. For simplicity, we consider a base contract with a single fund and a single premium payment at inception. We shall denote the evolution of investment account values over the life of the variable annuity by $\{F_t : t \geq 0\}$. The original maturity is set at time T_1 when the contract is automatically renewed for the second period with no additional underwriting procedure. It is common that most fees and chargers are deducted on a daily basis as a fixed percentage of the policyholder's account value. For convenience, we consider the continuous-time model in this chapter. Assume that the total fees are taken at the rate of m per dollar of the account value. Since the investment account is unit-linked to an equity index/fund, the value process F_t of the policyholder's investment account is proportional to the underlying equity index/fund net of all fees, i.e.

$$\frac{F_t}{F_0} = e^{-mt}\frac{S_t}{S_0}, \qquad 0 < t < T_1,$$

where S_t is the value of the underlying equity index/fund. Note that usually only a portion m_a of the total fees m, which is called the rider charge, is used to fund the guarantees and the rest of fees goes to cover commissions, overheads, and other expenses. At the end of the first contractual period, the policyholder is guaranteed to receive a greater of a pre-determined guarantee amount, say G_0 and the balance of the investment account F_{T_1}. In other words, the combined value of the investment account and investment guarantee is worth

$$M_{T_1} = \max\{G_0, F_{T_1}\}.$$

Let T_2 be the maturity of the renewed contract. The evolution of the policyholder's investment account

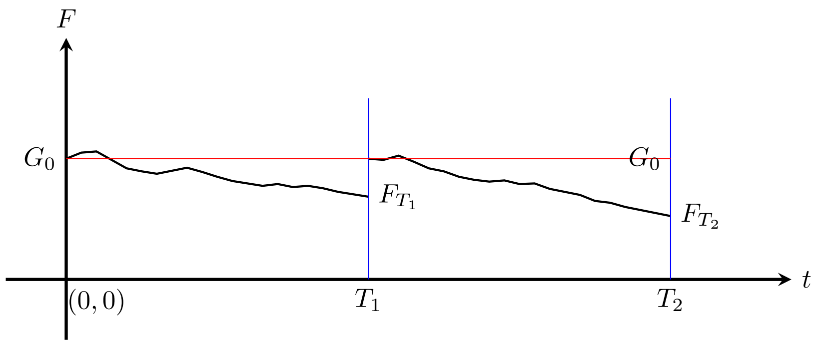

Figure 2.1: GMAB gross liability- Case 1

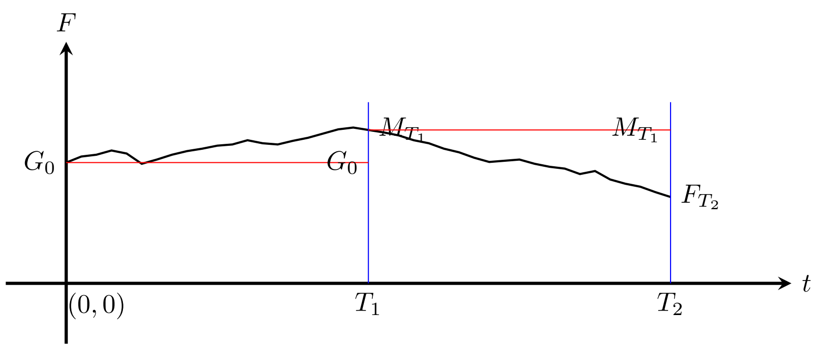

Figure 2.2: GMAB gross liability - Case 2

for the second period is driven by the same equity-linking mechanism as that in the first period, i.e.

$$\frac{F_t}{M_{T_1}} = e^{-m(t-T_1)}\frac{S_t}{S_{T_1}}, \qquad T_1 \leq t < T_2.$$

The contract provides a minimum guarantee G_1 on the investment over a second period, to be determined in two cases.

1. The equity investment performed so poorly in the first period that the policyholder's account investment F_{T_1} falls below G_0 at the renewal date, as shown in Figure 2.1. Then the guarantee is in-the-money and the insurer is responsible for injecting the additional cash $(G_0 - F_{T_1})$ into the policyholder's investment account and the guaranteed amount remains the same over the next period.

2. The equity investment performed so well in the first period that F_{T_1} exceeds G_0 by the end of the first period, as illustrated in Figure 2.2. Then the guarantee is out-of-money and there is no payment from the insurer. However, the guaranteed amount for the second period, G_1, is reset to M_{T_1}. In other words, the policyholder should never lose what has been accumulated at the first maturity.

When the contract reaches its second maturity T_2, the policyholder is guaranteed to receive the greater of the new guaranteed amount and the outstanding balance of the investment account, i.e.

$$\max\{G_1, F_{T_2}\}.$$

Suppose that the rate of return on assets backing up the insurance liability is the continuously compounding rate r per time unit. Let T_x be the future lifetime of a policyholder at age x. Note that the maturity benefits are only payable when the policyholder survives and that the insurer is liable for payments only when guaranteed amounts exceed the policyholder's account values at maturities. Then the present value of the GMAB gross liability is determined by

$$\mathbb{1}_{\{T_1<\tau_x\}}e^{-rT_1}(G_0 - F_{T_1})_+ + \mathbb{1}_{\{T_2<\tau_x\}}e^{-rT_2}(G_1 - F_{T_2})_+. \tag{2.9}$$

While there are possibilities of insurance liabilities for both periods in Case 1, most severe losses would appear in Case 2. Note that the gross liability in Case 1 is essentially bounded by $e^{-rT_1}G_0 + e^{-rT_2}G_0$ whereas the gross liability in Case 2 is unbounded.

Most literature on this subject including practitioners' publications typically only consider the gross liability due to its resemblance to a put option. All dynamic hedging and risk management strategies are subsequently developed in accordance with the "put option". However, such an approach overlooks the financial risks embedded on the income side, as fees are taken as percentages of equity-linked account values. For brevity, we write $(x)_+ = \max(x, 0), x \wedge y = \min(x, y), x \vee y = \max(x, y)$. Observe that rider charges are collected up to the earlier of the maturity and the death of policyholder and hence the present value of all rider charges taken from the first period can be formulated as

$$m_a \int_0^{T_1\wedge\tau_x} e^{-rs}F_s ds,$$

where m_a is the rate of the rider charge per dollar of the account value per period, and similarly, that from the second period is given by

$$m_a \int_{T_1}^{(\tau_x\vee T_1)\wedge T_2} e^{-rs}F_s ds.$$

When equity values are persistently low, not only do insurers face high gross liabilities from the guarantees, they also receive very little income, which exacerbates the overall losses.

In this chapter, we consider the net liabilities, which are defined to be gross liabilities less fee incomes. We denote the present values of net liabilities for the first period and that for the second period by L_1 and L_2, respectively. For the purpose of the risk management, we are interested in the positive net liabilities, the cases when liabilities exceed assets. Hence,

$$L_1 = \left(\mathbb{1}_{\{T_1 < \tau_x\}} e^{-rT_1} (G_0 - F_{T_1})_+ - m_a \int_0^{T_1 \wedge \tau_x} e^{-rs} F_s ds \right)_+ ,$$

and

$$L_2 = \left(\mathbb{1}_{\{T_2 < \tau_x\}} e^{-rT_2} (G_1 - F_{T_2})_+ - m_a \int_{T_1}^{(\tau_x \vee T_1) \wedge T_2} e^{-rs} F_s ds \right)_+ .$$

In other words, the present value of the insurer's total net liabilities is given by

$$\begin{aligned} L =& L_1 + L_2 \\ =& \mathbb{1}_{\{T_1 < \tau_x\}} \left(e^{-rT_1} G_0 - e^{-rT_1} F_{T_1} - m_a \int_0^{T_1} e^{-rs} F_s ds \right)_+ \\ &+ \mathbb{1}_{\{T_2 < \tau_x\}} \left(e^{-rT_2} M_{T_1} - e^{-rT_2} F_{T_2} - e^{-rT_1} m_a \int_0^{T_2 - T_1} e^{-rs} F_{s+T_1} ds \right)_+ \end{aligned} \tag{2.10}$$

2.2.2 Actuarial risk management

The actuarial risk management approach is widely used in North American markets based on the liability runoff projection, which is the classical approach for managing traditional life insurance. As equity-linked insurance involves significant financial risks, it has become an industry practice to run liability projections under a large number of economic scenarios. With certain actuarial assumptions and prudent estimates of model parameters, a set of economic scenarios is generated with regard to equity returns, interest rates, equity volatilities, etc. Under each economic scenario, the insurance liability is determined by the present value of future benefit claims less that of future premium incomes through certain accounting exercises such as income statements. The formulation of the net liability presented in (2.10) is, in essence, a stochastic representation of the liability projection of the GMAB for an individual contract. The resulting surpluses/deficiencies from all scenarios are collected to form an empirical distribution of the insurance liability.

The principle of the actuarial risk management approach is to determine how much reserve or risk capital is necessary for an insurer to set aside on liquid assets in order to cover expected or unexpected losses in adverse economic scenarios. Different reserves serve different audiences and different purposes. For example, Generally Accepted Accounting Principles (GAAP) reserves are determined on a "best estimate" basis, as the primary purpose is to accurately report the value of insurance business for stockholders. Statutory reserves prepared for insurance regulators are often prepared with some degree of conservatism, as the primary objective of statutory accounting is to safeguard the solvency of insurance businesses and the stability of insurance markets. Tax reserves are determined for federal income tax purposes. While these reserves differ

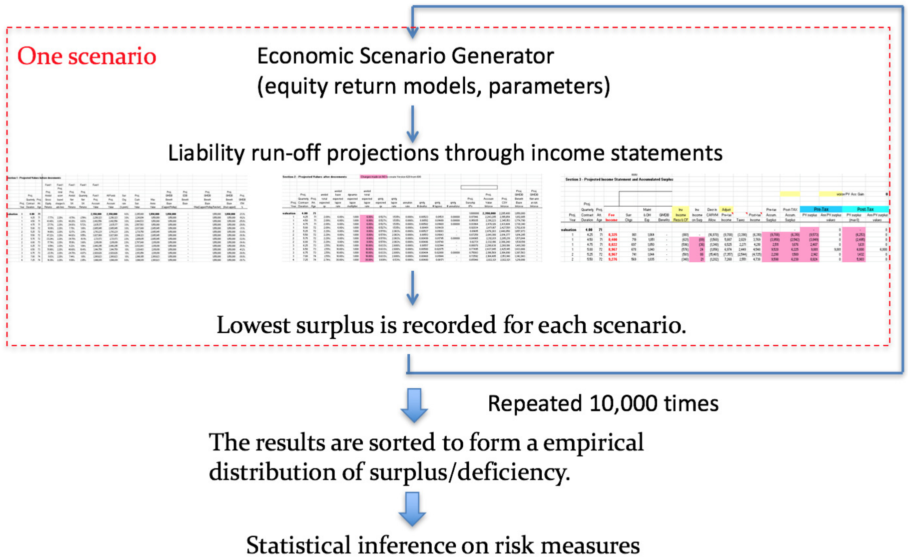

Figure 2.3: Risk measure based reserving

greatly with actuarial assumptions and accounting standards, they all generally follow roughly the same methodology from a modeling point of view, as shown in Figure 2.3. For the purpose of discussion in this work, we do not consider specific accounting rules and take a minimalist approach to only bring out the essence of practical models, as formulated in (2.10).

There are two risk measures of particular interest in practice, namely the Value at Risk (VaR) and the Conditional Tail Expectation (CTE) with respect to the level of confidence $\alpha, 0 < \alpha < 1$. Their definitions are given by

$$\mathrm{VaR}_\alpha = \inf\{y > 0, \mathbb{P}(L > y) < 1 - \alpha\},$$

and

$$\mathrm{CTE}_\alpha = \mathbb{E}[L | L > \mathrm{VaR}_\alpha].$$

In the United States, the insurance market is regulated at the state level by state insurance departments. The National Association of Insurance Commissioners (NAIC) typically develops model standards and regulations. State insurance departments adopt them with no or minor modifications to regulate businesses operating in their jurisdictions. For example, the NAIC specifies the use of 70% CTE and 90% CTE of accumulated surpluses/deficiencies to determine statutory reserves and risk-based capital requirement respectively for equity-linked insurance products.

Under our model assumptions, the net liability L is a continuous random variable. Hence, the quantity

VaR_α can be determined by a root search algorithm such that

$$\mathbb{P}(L > \mathrm{VaR}_\alpha) = 1 - \alpha.$$

Observe that, in order to obtain analytical expressions for VaR and CTE, the key quantities would be the probability of a large loss

$$\mathbb{P}(L > V), \tag{2.11}$$

for a certain level α and the conditional expectation, known as the mean excess function,

$$\mathbb{E}[L|L > V]. \tag{2.12}$$

While there are many ways to model the dynamics of equity returns, we use the most common model of the geometric Brownian motion, also known as the independent lognormal model in the insurance industry, for its mathematical tractability. In other words, we assume that the dynamics of the underlying equity index/fund $\{S_t, t \geq 0\}$ is determined by

$$S_t = S_0 e^{\mu t + \sigma B_t}, \qquad 0 < t < T_2,$$

where $\{B_t, t \geq 0\}$ is a standard Brownian motion, parameters μ and σ can be estimated from historical market data on equity returns.

In the rest of this section, we briefly explain the main methodology for obtaining analytical solutions to the risk measures. While the technical treatments seem irrelevant to the comparison of risk management strategies, one of the novelties of our work is to develop a methodology for analyzing these highly complex risk management strategies, which were only attempted with purely statistical approaches such as Monte Carlo simulations in the literature.

Let us introduce some notation to simplify the upcoming expressions. Recall from (2.10) that

$$L_1 = \mathbb{1}_{\{T_1 < \tau_x\}} F_0 \left(e^{-rT_1} \frac{G_0}{F_0} - \underline{e^{(\mu-m-r)T_1 + \sigma B_{T_1}}} - \underline{m_a \int_0^{T_1} e^{(\mu-m-r)s + \sigma B_s} ds} \right)_+, \tag{2.13}$$

and

$$L_2 = \mathbb{1}_{\{T_1<\tau_x\}} e^{-rT_1} M_{T_1} \times \\ \left(e^{-r(T_2-T_1)} - \underline{e^{(\mu-m-r)(T_2-T_1)+\sigma(B_{T_2}-B_{T_1})}} - \underline{m_a \int_0^{T_2-T_1} e^{(\mu-m-r)s+\sigma(B_{s+T_1}-B_{T_1})} ds} \right)_+ . \tag{2.14}$$

Note that the difficulty with the analysis of the net liability lies in the underlined terms both of which are driven by the underlying Brownian motion. We intend to reformulate the net liability in terms of a diffusion process, for which the analysis can be done in a more efficient manner. Define

$$B_t^v = vt + B_t, \qquad A_t^v = \int_0^t e^{2B_s^v} ds, \tag{2.15}$$

which are a Brownian motion with drift v and the path integral of its exponential functional. Using the scaling property of Brownian motions, we can have the following identities in law.

$$e^{(\mu-m-r)T_1+\sigma B_{T_1}} = e^{2(\frac{2(\mu-m-r)}{\sigma^2}\frac{\sigma^2 T_1}{4}+\frac{\sigma}{2}B_{T_1})} \stackrel{d}{=} e^{2B_{t_1}^v},$$

and

$$\int_0^{T_1} e^{(\mu-m-r)s+\sigma B_s} ds \stackrel{d}{=} \int_0^{T_1} e^{2B^v_{\sigma^2 s/4}} ds = \frac{4}{\sigma^2} A_{t_1}^v,$$

with $v = \frac{2(\mu-m-r)}{\sigma^2}$ and $t_1 = \frac{\sigma^2 T_1}{4}$. For the same reason, we can also get

$$e^{(\mu-m-r)(T_2-T_1)+\sigma(B_{T_2}-B_{T_1})} \stackrel{d}{=} e^{2B_{t_2}^v},$$

and

$$\int_0^{T_2-T_1} e^{(\mu-m-r)s+\sigma(B_{s+T_1}-B_{T_1})} ds \stackrel{d}{=} \frac{4}{\sigma^2} A_{t_2}^v,$$

where $t_2 = \frac{\sigma^2(T_2-T_1)}{4}$. Observe that $(B_{t_1}^v, A_{t_1}^v)$ and $(B_{t_2}^v, A_{t_2}^v)$ are independent by the fact that Brownian motions have independent increments. For simplicity, we shall write $x_0 = \sigma^2/4m_a$. Let

$$X_{t_1} := e^{2B_{t_1}^v} x_0 + A_{t_1}^v,$$

and

$$X_{t_2} := e^{2B_{t_2}^v} x_0 + A_{t_2}^v.$$

Then from Section 2.1, by a time reversal argument, X_{t_1} and X_{t_2} are equal in law to the generalized Ornstein-

Uhlenbeck process satisfying SDE (2.4). Using the notation above, insurance liabilities (2.13) and (2.14) can be simplified to

$$
\begin{aligned}
L &= L_1 + L_2 \\
&\stackrel{d}{=} \mathbb{1}_{\{T_1<\tau_x\}} F_0 \left(e^{-rT_1} \frac{G_0}{F_0} - \frac{X_{t_1}}{x_0} \right)_+ + \mathbb{1}_{\{T_2<\tau_x\}} e^{-rT_1} M_{T_1} \left(e^{-r(T_2-T_1)} - \frac{X_{t_2}}{x_0} \right)_+ ,
\end{aligned} \tag{2.16}
$$

where

$$
M_{T_1} = \max\{G_0, F_{T_1}\} = \max\{G_0, e^{(\mu-m)T_1+\sigma B_{T_1}} F_0\} \stackrel{d}{=} e^{rT_1} F_0 \max\left\{ e^{-rT_1} \frac{G_0}{F_0}, e^{2B^v_{t_1}} \right\}.
$$

In the following, we resort to the known distributions of X_{t_1} and X_{t_2} to obtain analytical solutions to the quantities in (2.11) and (2.12).

It is interesting to observe that uncertainties with equity returns on both the liability side and the asset side (reflected in terms underlined in (2.13) and (2.14)) are absorbed in the diffusion process $\{\overline{X}_t : t \geq 0\}$, which is introduced in Section 2.1. Hence, the computation of risk measures boils down to that of the functionals of the diffusion process.

Analytical solutions to VaR **and** CTE

As mentioned previously, the key quantity for computing VaR is the tail probability of the net liability, in other words, $\mathbb{P}(L > V)$ for a certain level $V > 0$. The VaR with the confidence level α is determined by the level V_α such that $\mathbb{P}(L > V_\alpha) = 1 - \alpha$. To solve the inverse problem, a variety of numerical techniques, such as the bisection method and Newton's method, can be used. Here we provide the analytical solutions to (2.11) and (2.12).

It is reasonable to assume that the future lifetimes of policyholders are independent of the performance of equity funds/indices in financial markets. The survival model of τ_x can be either a parametric distribution such as Gompertz-Makeham law of mortality or non-parametric distributions described by life tables. In either case, we denote the survival probability $\mathbb{P}(\tau_x > T) = {}_T p_x$.

For brevity, we introduce functions

$$
K_1(V) := \frac{e^{-rT_1} G_0 - V}{F_0} x_0, \qquad K_2(V) := e^{-r(T_2-T_1)} x_0 - \frac{V}{G_0} e^{rT_1} x_0, \tag{2.17}
$$

and

$$
f(y, V) := \frac{1}{2\sqrt{t_1}} \log \left(\frac{F_0}{x_0} (e^{-r(T_2-T_1)} x_0 - y) \right) - \frac{\log V}{2\sqrt{t_1}}, \tag{2.18}
$$

$$g(y,V) := \frac{G_0}{F_0}\left(e^{-rT_1}x_0 + e^{-rT_2}x_0 - \frac{V}{G_0}x_0 - e^{-rT_1}y\right). \tag{2.19}$$

Theorem 2.2.1 (Tail probability of the net liability)**.** *The tail probability $\mathbb{P}(L > V)$ for any fixed $V > 0$ is given by*

- *If $V < e^{-rT_2}G_0$,*

$$\begin{aligned}\mathbb{P}(L > V) &= \left({}_{T_1}p_x - {}_{T_2}p_x C(t_2, x_0, K_2(0))\right) C\left(t_1, x_0, K_1(V)\right) + {}_{T_2}p_x C(t_2, x_0, K_2(V)) \\ &+ {}_{T_2}p_x \int_{K_2(V)}^{K_2(0)} \mathcal{N}\left(\sqrt{t_1}v + f(y,V)\right) p(t_2, x_0, y)dy + {}_{T_2}p_x \int_{K_2(V)}^{K_2(0)} C\left(t_1, x_0, g(y,V)\right) p(t_2, x_0, y)dy \end{aligned} \tag{2.20}$$

- *If $e^{-rT_2}G_0 < V < e^{-rT_1}G_0$,*

$$\begin{aligned}\mathbb{P}(L > V) &= \left({}_{T_1}p_x - {}_{T_2}p_x C(t_2, x_0, K_2(0))\right) C\left(t_1, x_0, K_1(V)\right) \\ &+ {}_{T_2}p_x \int_{0}^{K_2(0)} \mathcal{N}\left(\sqrt{t_1}v + f(y,V)\right) p(t_2, x_0, y)dy + {}_{T_2}p_x \int_{0}^{K_2(0)} C\left(t_1, x_0, g(y,V)\right) p(t_2, x_0, y)dy \end{aligned} \tag{2.21}$$

- *If $V > e^{-rT_1}G_0$,*

$$\begin{aligned}\mathbb{P}(L > V) =& {}_{T_2}p_x \int_{0}^{K_2(0)} \mathcal{N}\left(\sqrt{t_1}v + f(y,V)\right) p(t_2, x_0, y)dy \\ &+ {}_{T_2}p_x \int_{0}^{x_0e^{-rT_1}+x_0e^{-rT_2}-x_0\frac{V}{G_0}} C\left(t_1, x_0, g(y,V)\right) p(t_2, x_0, y)dy \end{aligned} \tag{2.22}$$

where $C(\cdot)$ and $p(\cdot)$ are defined in Theorem 2.1.1. $\mathcal{N}(\cdot)$ is the standard normal cumulative distribution function and $K_1(\cdot)$, $K_2(\cdot)$, $f(\cdot)$ and $g(\cdot)$ are defined in (2.17), (2.18) *and* (2.19) *respectively.*

Proof. See Subsection 2.2.5 □

Recall that the CTE_α is given by the conditional tail expectation of the liability truncated at the level of VaR_α. Once VaR_α is determined, CTE_α can be computed from

$$\text{CTE}_\alpha = \frac{1}{1-\alpha}\mathbb{E}[L\mathbb{1}_{\{L > \text{VaR}_\alpha\}}].$$

The next theorem gives the analytical solution of the tail expectation $\mathbb{E}[L\mathbb{1}_{\{L>V\}}]$ for any fixed $V > 0$. Like the case of the tail probability of the net liability, it has different analytical formulas corresponding to $V < e^{-rT_2}G_0$, $e^{-rT_2}G_0 < V < e^{-rT_1}G_0$ and $V > e^{-rT_1}G_0$.

Theorem 2.2.2 (Conditional tail expectation of the net liability)**.** *The tail expectation* $\mathbb{E}[L\mathbb{1}_{\{L>V\}}]$ *for any fixed* $V > 0$ *is given by*

- *If* $V < e^{-rT_2}G_0$,

$$\begin{aligned}\mathbb{E}[L\mathbb{1}_{\{L>V\}}] =& \frac{F_0}{x_0}\left({}_{T_1}p_x - {}_{T_2}p_x C(x_0,t_2,K_2(0))\right)\left(K_1(0)C(t_1,x_0,K_1(V)) - Z(t_1,x_0,K_1(V))\right)\\ &+ {}_{T_2}p_x\frac{F_0}{x_0}\left(K_1(0)C(x_0,t_1,K_1(0)) - Z(x_0,t_1,K_1(0))\right)C(x_0,t_2,K_2(V))\\ &+ {}_{T_2}p_x\frac{F_0}{x_0}\int_{K_2(V)}^{K_2(0)}\left(K_1(0)C(x_0,t_1,g(y,V)) - Z(x_0,t_1,g(y,V))\right)p(x_0,t_2,y)dy\\ &+ {}_{T_2}p_x\frac{K_1(0)F_0}{x_0^2}\int_{K_2(V)}^{K_2(0)}C(x_0,t_1,g(y,V))(K_2(0)-y)p(x_0,t_2,y)dy\\ &+ {}_{T_2}p_x\frac{F_0}{x_0}\int_{K_2(V)}^{K_2(0)}e^{2(v+1)t_1}\mathcal{N}\left((v+2)\sqrt{t_1}+f(y,V)\right)(K_2(0)-y)p(x_0,t_2,y)dy\\ &+ {}_{T_2}p_x\frac{F_0}{x_0}\left(K_2(0)C(x_0,t_2,K_2(V)) - Z(x_0,t_2,K_2(V))\right)\times\\ &\left(\frac{K_1(0)}{x_0}\mathcal{N}\left(\frac{1}{2\sqrt{t_1}}\log\left(\frac{K_1(0)}{x_0}\right) - v\sqrt{t_1}\right) + e^{2(v+1)t_1}\mathcal{N}\left((v+2)\sqrt{t_1} - \frac{1}{2\sqrt{t_1}}\log\left(\frac{K_1(0)}{x_0}\right)\right)\right)\end{aligned} \tag{2.23}$$

- *If* $e^{-rT_2}G_0 < V < e^{-rT_1}G_0$,

$$\begin{aligned}\mathbb{E}[L\mathbb{1}_{\{L>V\}}] =& \frac{F_0}{x_0}\left({}_{T_1}p_x - {}_{T_2}p_x C(x_0,t_2,K_2(0))\right)\left(K_1(0)C(t_1,x_0,K_1(V)) - Z(t_1,x_0,K_1(V))\right)\\ &+ {}_{T_2}p_x\frac{F_0}{x_0}\int_0^{K_2(0)}\left(K_1(0)C(x_0,t_1,g(y,V)) - Z(x_0,t_1,g(y,V))\right)p(x_0,t_2,y)dy\\ &+ {}_{T_2}p_x\frac{K_1(0)F_0}{x_0^2}\int_0^{K_2(0)}C(x_0,t_1,g(y,V))(K_2(0)-y)p(x_0,t_2,y)dy\\ &+ {}_{T_2}p_x\frac{F_0}{x_0}\int_0^{K_2(0)}e^{2(v+1)t_1}\mathcal{N}\left((v+2)\sqrt{t_1}+f(y,V)\right)(K_2(0)-y)p(x_0,t_2,y)dy\end{aligned} \tag{2.24}$$

- *If* $V > e^{-rT_1}G_0$,

$$\begin{aligned}&\mathbb{E}[L\mathbb{1}_{\{L>V\}}]\\ =& {}_{T_2}p_x\frac{F_0}{x_0}\int_0^{x_0e^{-rT_1}+x_0e^{-rT_2}-x_0\frac{V}{G_0}}\left(K_1(0)C(x_0,t_1,g(y,V)) - Z(x_0,t_1,g(y,V))\right)p(x_0,t_2,y)dy\\ &+ {}_{T_2}p_x\frac{K_1(0)F_0}{x_0^2}\int_0^{x_0e^{-rT_1}+x_0e^{-rT_2}-x_0\frac{V}{G_0}}C(x_0,t_1,g(y,V))(K_2(0)-y)p(x_0,t_2,y)dy\\ &+ {}_{T_2}p_x\frac{F_0}{x_0}\int_0^{K_2(0)}e^{2(v+1)t_1}\mathcal{N}\left((v+2)\sqrt{t_1}+f(y,V)\right)(K_2(0)-y)p(x_0,t_2,y)dy\end{aligned} \tag{2.25}$$

where $C(\cdot)$ and $p(\cdot)$ are defined in Theorem 2.1.1. $\mathcal{N}(\cdot)$ is the standard normal cumulative distribution function and $K_1(\cdot)$, $K_2(\cdot)$, $f(\cdot)$, $g(\cdot)$ are defined in (2.17), (2.18) *and* (2.19), *respectively.*

Proof. See Subsection 2.2.5 □

Although expressions in Theorems 2.2.1 and 2.2.2 may appear formidable, they can be handled quite efficiently in most computational software platforms, such as Mathematica. Numerical examples and a comparison with Monte Carlo methods are provided in Subsection 2.2.4.

2.2.3 Dynamic hedging risk management

While dynamic hedging techniques have been gradually adopted in the life insurance industry on products with exposure to market risks, most works have been done with a market-consistent valuation of liabilities, which is, in essence, the risk neutral valuation of gross liabilities based on Monte Carlo simulations with economic scenarios generated from models calibrated to market conditions. The principle of market-consistent valuation is used across the board for all equity-linked insurance products. An overview of the industry's adoption of the market consistent valuation and its commercial implications can be found in Sheldon and Smith ([60]), Grosen and Jørgensen ([37]).

Take the two-period GMAB rider for example. The risk neutral valuation of the gross liabilities in (2.9) would be similar to that of European put options. Under the risk neutral measure $\mathbb{Q}$, the non-arbitrage cost of the GMAB gross liability is given by

$$P^g := \mathbb{E}^{\mathbb{Q}}\left[\mathbb{1}_{\{T_1<\tau_x\}}e^{-rT_1}(G_0 - F_{T_1})_+ + \mathbb{1}_{\{T_2<\tau_x\}}e^{-rT_2}(G_1 - F_{T_2})_+\right].$$

While this is a valid approach and commonly used in practice for the purpose of developing a hedging program, there are also a few apparent disadvantages.

1. The same market risks affect both the gross liability and the fee income. A hedging program developed only for the gross liability overlooks the uncertainty from the income side.

2. In most cases, there is either no GMAB liability payment or relatively small liability payments that can be compensated by the accumulated fee incomes, leading to a profit for the insurer. However, a hedging program developed for the gross liability would in theory completely eliminate even these small payouts. In that case, such an offset would be considered excessive. This indicates that a hedging program that does not take fee income into account would be more costly than necessary.

3. The hedging of large liabilities typically requires a high volume of hedging instruments, which incur high transaction costs.

In this work, we propose the notion of *net liability hedging*, which brings fee incomes into consideration for a hedging program. Under the same risk-neutral measure, the non-arbitrage cost of the GMAB net liability is given by

$$P^n := \mathbb{E}^{\mathbb{Q}}[L_1 + L_2],$$

where the net liabilities for the first and second periods are defined in (2.10). It is easy to observe that

$$P^n \leq P^g,$$

where the equality holds if and only if $m_a = 0$. The inequality indicates a saving in the cost of setting up a dynamic hedging program switching from gross liability hedging to net liability hedging.

Analytical solutions to hedging costs

When a hedging program is established, regulators often require insurers to include cash flows from the hedging program in projections of surplus/deficiency from the equity-linked insurance to determine its statutory reserves and risk-based capitals. While the implementation is similar to that without a hedging program as demonstrated in Figure 2.3, the projections involving the dynamics of a hedging program are much more computationally intensive. As shown in Figure 2.4, under each economic scenario (represented by the black sample path), the hedging portfolio is rebalanced at every transaction date according to estimates of Greeks, which are sensitivity measures of the hedging portfolio to various economic risk factors such as equity price, equity volatility, interest rate, etc. In practice, Greeks are often estimated by difference quotients of the average of equity prices with shocks to risk factors, which are themselves determined by further projections of economic scenarios (sample paths of different colors at different time points). The procedure of running stochastic projections in which certain components invoke further stochastic projections is known as the *nested stochastic modeling*. For example, under one single economic scenario, there are approximately $50 \times 20 = 1,000$ transaction dates for a 20-year term GMAB with a weekly hedging program. If $1,000$ economic scenarios are used to determine Greeks at each transaction date, then the single scenario of the hedging program alone would require the simulation of a total of $1,000 \times 1,000 = 1,000,000$ scenarios of all risk factors. Therefore, the computational burden of the nested stochastic modeling grows exponentially when multiple scenarios are required to determine risk measures of insurers' liabilities.

While there are many statistical methods such as variance reduction techniques that can be utilized for more efficient Monte Carlo simulations, one may also utilize approximations by analytical solutions, which

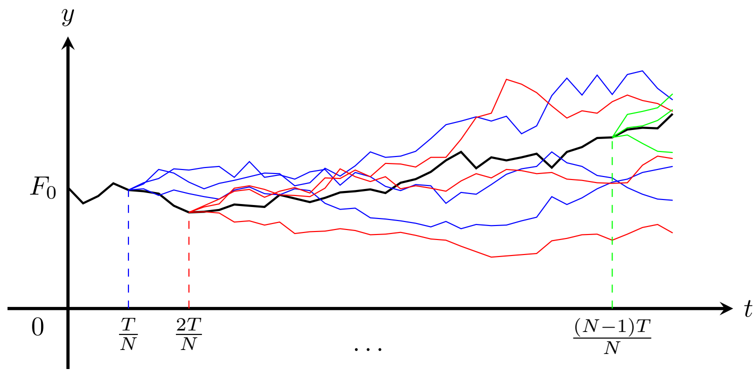

Figure 2.4: Sample paths generation in nested stochastic for dynamic hedging

can be much more efficient and accurate when they are available. In the remainder of this section, we search for analytical solutions to the hedging costs of the GMAB net liability P^n as well as the deltas for hedging.

If we assume the underlying equity index/fund S_t is a traded asset and we can both borrow and lend money at the risk-free interest rate r, then the market with a money market account and a trading account of equity index futures is complete. Therefore, under the risk neutral measure $\mathbb{Q}$, the equity process $\{S_t, t \geq 0\}$ has the dynamics

$$S_t = S_0 e^{(r-\frac{1}{2}\sigma^2)t+\sigma B_t}, \qquad 0 < t < T_2.$$

Comparing with its dynamics under the real world measure $\mathbb{P}$, we find that the only change is that the drift parameter μ has changed to $r - \sigma^2/2$. It means that under the risk measure $\mathbb{Q}$, the net liability L still has the same form as in (2.16), but the only change is that the drift v in definition is now

$$v = -\frac{2m}{\sigma^2} - 1. \tag{2.26}$$

Therefore, the non-arbitrage cost of the net liability is

$$\begin{aligned}\mathbb{E}^{\mathbb{Q}}[L_1 + L_2] &= \frac{F_0}{x_0}\mathbb{E}^{\mathbb{Q}}\left[(K_1^0 - X_{t_1})_+ \mathbb{1}_{\{T_1<\tau_x\}}\right] + \frac{1}{x_0}\mathbb{E}^{\mathbb{Q}}\left[e^{-rT_1}M_{T_1}(K_2^0 - X_{t_2})_+ \mathbb{1}_{(T_2<\tau_x)}\right] \\ &= {}_{T_1}p_x \frac{F_0}{x_0}\mathbb{E}^{\mathbb{Q}}\left[(K_1^0 - X_{t_1})_+\right] + {}_{T_2}p_x \frac{F_0}{x_0}\mathbb{E}^{\mathbb{Q}}\left[\max\left\{e^{-rT_1}\frac{G_0}{F_0}, e^{2B_{t_1}^v}\right\}\right]\mathbb{E}^{\mathbb{Q}}\left[(K_2^0 - X_{t_2})_+\right], \end{aligned} \tag{2.27}$$

where the last equality comes from the independence of equity values and policyholders' mortalities, as well as the fact that $B_{t_1}^v$ and X_{t_1} are independent. Note that

$$\mathbb{E}^{\mathbb{Q}}\left[(K_1^0 - X_{t_1})_+\right] \qquad \text{and} \qquad \mathbb{E}^{\mathbb{Q}}\left[(K_2^0 - X_{t_2})_+\right],$$

in (2.27) can be obtained from

$$\mathbb{E}^{\mathbb{Q}}\left[(K-X_t)_+\right]=K\mathbb{Q}(X_t<K)-\mathbb{E}^{\mathbb{Q}}\left[X_t\mathbb{1}_{\{X_t<K\}}\right]=KC(t,x_0,K)-Z(t,x_0,K),$$

where $C(\cdot)$ and $Z(\cdot)$ are defined in Theorem 2.1.1. We point out here that, the actuarial risk management in the previous subsection is under the real world measure. Therefore v is usually positive and we do not have the additional terms in Theorem 2.1.1. For the dynamic hedging risk management, however, we are under the risk neutral measure. Now the drift $v=-\frac{2m}{\sigma^2}-1$ is negative so that we should have the additional compensation terms in the (2.5), (2.6), (2.7) and (2.8). With the functions defined in Theorem 2.1.1, we can obtain the risk neutral value of the net liability.

Theorem 2.2.3. *Let $v=-\frac{2m}{\sigma^2}-1$, the risk neutral price for the net liability is given by*

$$\begin{aligned}&\mathbb{E}^{\mathbb{Q}}[L_1+L_2]\\ =&{}_{T_1}p_x\frac{F_0}{x_0}P(t_1,x_0,K_1(0))+{}_{T_2}p_x\frac{F_0}{x_0}P(t_2,x_0,K_2(0))\times\\ &\left(\frac{K_1(0)}{x_0}\mathcal{N}\left(\frac{1}{2\sqrt{t_1}}\log(\frac{K_1(0)}{x_0})-v\sqrt{t_1}\right)+e^{2(v+1)t_1}\mathcal{N}\left((v+2)\sqrt{t_1}-\frac{1}{2\sqrt{t_1}}\log\left(\frac{K_1(0)}{x_0}\right)\right)\right),\end{aligned}\tag{2.28}$$

where $P(\cdot)$ is defined in (2.7) *and $\mathcal{N}(\cdot)$ is the cumulative distribution function of standard normal distribution.*

Proof. See Subsection 2.2.5 □

Analytical solutions to Greeks

In theory, if an insurer develops a hedging program for the net liability of the GMAB that rebalances its portfolio continuously according to the delta hedging strategy and there is no transaction cost, the actual hedging costs will be exactly the risk neutral cost computed above. Even though it is impractical to hedge continuously, the theoretical delta results still provide potentially good approximations to those for frequent rebalancing strategies. Note that the total net liability of the GMAB comes from two parts: L_1 and L_2 with different maturity. Therefore, in practice, we can hedge them separately using two accounts. Define

$$\Delta_1:=\frac{\partial(\mathbb{E}^{\mathbb{Q}}[L_1])}{\partial F_0},$$
$$\Delta_2:=\frac{\partial(\mathbb{E}^{\mathbb{Q}}[L_2])}{\partial F_0}.$$

The next theorem provides analytical formulas to the deltas of the GMAB net liabilities.

Theorem 2.2.4. *The deltas of the risk neutral price for the net liability is given by*

$$\Delta_1 = -\frac{{}_{T_1}p_x Z(t_1, x_0, K_1(0))}{x_0},$$
$$\Delta_2 = {}_{T_2}p_x \frac{e^{2(v+1)t_1} P(t_2, x_0, K_2(0))}{x_0} \mathcal{N}\left((v+2)\sqrt{t_1} - \frac{1}{2\sqrt{t_1}} \log\left(\frac{K_1(0)}{x_0}\right)\right).$$

with $v = -\frac{2m}{\sigma^2} - 1$. $P(\cdot)$ *and* $Z(\cdot)$ *are defined in* (2.7) *and* (2.8), $\mathcal{N}(\cdot)$ *is the cumulative distribution function of the standard normal distribution.*

Proof. See Subsection 2.2.5 □

2.2.4 Numerical results

In this subsection, we provide two numerical examples to compare the traditional actuarial and the dynamic hedging risk management of the GMAB. In the first example, we explain how to apply the analytical results to determine the risk measures of the GMAB net liabilities. The analytical results are assessed and benchmarked against traditional Monte Carlo simulations in terms of accuracy and efficiency. In the second example, we develop a program to dynamically hedge the GMAB net liabilities. We compare the hedging costs of net liabilities with the hedging costs of gross liabilities.

Actuarial risk management: risk measures computation

We consider a GMAB designed for policyholders of age 65 with maturity in $T_1 = 10$ years and automatically renewed until $T_2 = 20$ years. To model the future lifetime of policyholders, the period life table for male and calendar year 2010 developed by Social Security Administration is used. Table 2.1 offers an excerpt from Bell et al. ([6]). We also calculate the corresponding survival probability to each age which shall be used later.

x	q_x	k	${}_kp_{65}$	x	q_x	k	${}_kp_{65}$
65	0.01753	0	1.00000	76	0.04715	11	0.72446
66	0.01932	1	0.98247	77	0.05184	12	0.69030
67	0.02122	2	0.96349	78	0.05711	13	0.65451
68	0.02323	3	0.94304	79	0.06305	14	0.61713
69	0.02538	4	0.92113	80	0.06978	15	0.57822
70	0.02785	5	0.89776	81	0.07738	16	0.53787
71	0.03059	6	0.87276	82	0.08596	17	0.49625
72	0.03343	7	0.84606	83	0.09557	18	0.45360
73	0.03633	8	0.81777	84	0.10625	19	0.41025
74	0.03942	9	0.78806	85	0.11800	20	0.36666
75	0.04299	10	0.75700				

Table 2.1: Predicted mortality rates of a male at the age of 65

The following numerical results are obtained by using both the analytical method and Monte Carlo simulations. All results on risk measures are computed up to at least 4 digits of accuracy, which are sufficient for most practical purposes. As we shall see, it is difficult for Monte Carlo simulations to attain accuracy higher than 4 digits because it nearly exhausts all computation resources available to this project. We used Mathematica 10 as our computation tool for both our analytical method and Monte Carlo simulations.

For the analytical method, the first algorithm is developed to obtain the tail probability of the GMAB net liability using analytical solutions in Theorem 2.2.1. While it does require numerical integration of special functions, the computation is easily handled by Mathematica. In the search for VaR_α, the bisection method is employed. The procedure will stop when the interval's length is less than 10^{-5} since we require 4 digits of accuracy. Then the VaR_α is fed into a second algorithm for determining CTE_α based on formulas from Theorem 2.2.2. The efficiency is measured by the total running time used to compute VaR_α and CTE_α.

For Monte Carlo simulations, the first step is to sample paths of the Brownian motion $\{B_t, t \geq 0\}$, which subsequently determine the paths of the underlying equity process $\{S_t, t \geq 0\}$. In practice, this is done by discretizing the time horizon into fine enough subintervals and simulating the increments of the Brownian motion. Note the net liability L defined in (2.16) contains the Yor's process

$$A_t^v = \int_0^t e^{2B_s^v} ds,$$

which is approximated by the discretized Riemann sum

$$\int_0^t e^{2B_s^v} ds \sim \sum_{i=1}^n e^{2B_{t_{i-1}}^v} (t_i - t_{i-1}),$$

where the increasing sequence of time points $\{t_0 = 0, t_1, \cdots, t_n = T\}$ separates the time interval $[0, t]$ into subintervals. In this numerical example, we discretize the time horizon with 2000 segments, which corresponds to about 3 days per segment. An estimator of VaR_α is based on the order statistic. We simulate 300 million sample paths and the $\alpha\%$ sample quantile $\widehat{\mathrm{VaR}}_\alpha$ of the corresponding net liability L is used as the estimator of VaR_α. CTE_α is estimated by the sample average of net liabilities truncated at $\widehat{\mathrm{VaR}}_\alpha$. Parallel computing is used to improve the efficiency: the experiments are simultaneously distributed to all 12 cores of a high-performance computing node of the Illinois Campus Cluster. To construct confidence intervals for the risk measures, we run simulations with 50 nodes and record the estimators $\widehat{\mathrm{VaR}_\alpha}$ and $\widehat{\mathrm{CTE}_\alpha}$ from each node. We denote the mean and standard deviation of the risk measures as $\overline{\mathrm{VaR}_\alpha}$, $\sigma_{\mathrm{VaR}_\alpha}$ and $\overline{\mathrm{CTE}_\alpha}$, $\sigma_{\mathrm{CTE}_\alpha}$, respectively. The confidence intervals of risk measures can be constructed as

$$(\overline{\mathrm{VaR}_\alpha} \pm \sigma_{\mathrm{VaR}_\alpha}) \qquad \textit{and} \qquad (\overline{\mathrm{CTE}_\alpha} \pm \sigma_{\mathrm{CTE}_\alpha}).$$

The efficiency is measured by the total running time of all experiments conducted on each node. It should be pointed out that it requires very intensive computation to run 50 repeated experiments with 300 million of sample paths, which takes several days and nearly exhausts our allotted computation resources.

We consider two sets of parameters leading to two economic conditions with high market volatility and low market volatility. Under each economic condition, risk measures with three confidence levels are calculated.

High volatility environment

α	70%	85%	90%
VaR_α	0.0%	12.160%	25.443%
CTE_α	36.069%	43.353%	58.667%
Time (min)	0.08	3.58	4.00

Table 2.2: Results from the analytical method in high volatility environment

α	70%	85%	90%
$\widehat{\mathrm{VaR}_\alpha}$	0.0%	$12.162 \pm 0.005\%$	$25.443\% \pm 0.006\%$
$\widehat{\mathrm{CTE}_\alpha}$	$36.071\% \pm 0.007\%$	$45.357 \pm 0.009\%$	$58.663\% \pm 0.012\%$
Time (min)	8688.58	7438.35	6863.48

Table 2.3: Results from the Monte Carlo method in high volatility environment

In high volatility environment, we consider risk measures with confidence levels $\alpha = 70\%$, $\alpha = 85\%$ and $\alpha = 90\%$. Other model parameters are given as follows:

1. the annualized risk-free interest rate $r = 0.04$,

2. the annualized mean and standard deviation of the log-return of the underlying equity S_t are $\mu = 0.09$ and $\sigma = 0.3$, respectively,

3. the annualized total fee charges rate is 100 basis points and among which 35 basis points is for the GMAB rider charge, i.e $m = 0.01$ and $m_a = 0.0035$.

We compare the accuracy and efficiency of the two approaches in Tables 2.2 and 2.3. Under the high volatility environment, the GMAB tends to be more costly due to the rollover feature. Therefore, risk measures are considerably higher than those under low volatility environment (See Table 2.4 and 2.5). It is not so surprising that whenever the analytical approach is available, the computation can be much more efficient. In the case of $\alpha = 70\%$ we have determined $\mathrm{VaR}_\alpha = 0$, the bisection search is not called to use which explains why it is considerably faster than the other two cases. The running times are represented in minutes.

Low volatility environment

α	85%	90%	95%
VaR_α	0.0%	4.597%	12.976%
CTE_α	11.667%	15.077%	21.666%
Time (min)	0.09	7.16	7.20

Table 2.4: Results from the analytical method in low volatility environment

α	85%	90%	95%
$\widehat{\mathrm{VaR}_\alpha}$	0.0%	$4.599 \pm 0.002\%$	$12.978\% \pm 0.003\%$
$\widehat{\mathrm{CTE}_\alpha}$	$11.667\% \pm 0.02\%$	$15.079 \pm 0.002\%$	$21.669\% \pm 0.003\%$
Time (min)	7816.50	4906.62	4888.28

Table 2.5: Results from the Monte Carlo method in low volatility environment

In low volatility environment, we consider the risk measures for $\alpha = 85\%$, $\alpha = 90\%$ and $\alpha = 95\%$. Other model parameters are given as follows:

1. the annualized risk-free interest rate $r = 0.02$,
2. the annualized mean and standard deviation of the log-return of the underlying equity S_t are $\mu = 0.045$ and $\sigma = 0.1$, respectively,
3. the annualized total fee charges rate is $m = 0.02$ and among which $m_a = 0.01$ is for the GMAB rider charge.

In the low volatility environment, there is less uncertainty with the product liability. Therefore, risk measures are in general lower than those in the high volatility environment. In addition, we find that the sample deviations of Monte Carlo simulations are in general smaller in the low volatility environment than those in the high volatility environment. The accuracy and efficiency results are also shown in Tables 2.4 and 2.5.

Comparing the results from our analytical method (Tables 2.2 and 2.4) and Monte Carlo simulations (Tables 2.3 and 2.5), we can make the following observations.

1. As expected, the results from the analytical method generally lie in the confidence interval given by the Monte Carlo simulation results. While the analytical method leads to computational errors due to the

numerical integration, its accuracy can be improved by increasing the system precision. In contrast, errors of the Monte Carlo method come from both the sampling procedure and discretization.

2. The analytical method is much more efficient than Monte Carlo simulations. All computations with the analytical method are carried out on a personal computer. In contrast, the simulations occupied 50 computing node of Illinois Campus Cluster, each of which is built on 12 cores. Even with the parallel computing, we are only able to achieve the convergence of simulation results up to 4 digits. Any further improvement can require exponential growth of computational efforts, which shows the limitation of Monte Carlo simulations.

Dynamic hedging of the net liability

L_1	$\mathbb{E}^{\mathbb{Q}}[L_1]$	Mean of $\widehat{\mathbb{E}^{\mathbb{Q}}[L_1]}$	SD of $\widehat{\mathbb{E}^{\mathbb{Q}}[L_1]}$
Net liability	4.82%	4.56%	1.15%
Gross liability	7.79%	6.30%	2.97%

Table 2.6: Hedging costs for L_1

L_2	$\mathbb{E}^{\mathbb{Q}}[L_2]$	Mean of $\widehat{\mathbb{E}^{\mathbb{Q}}[L_2]}$	SD of $\widehat{\mathbb{E}^{\mathbb{Q}}[L_2]}$
Net liability	2.15%	1.87%	0.48%
Gross liability	3.48%	2.60%	1.32%

Table 2.7: Hedging costs for L_2

In this part, we compare the hedging costs of the net liability with that of the gross liability. As expected, Table 2.6 and Table 2.7 show that the hedging costs are significantly reduced when moving from gross liability hedging to net liability hedging. Moreover, we observe a reduction in the hedging error, which is measured by the difference between the theoretical value of hedging costs $\mathbb{E}^{\mathbb{Q}}[L_1]$ ($\mathbb{E}^{\mathbb{Q}}[L_2]$), and the mean of realized hedging costs $\widehat{\mathbb{E}^{\mathbb{Q}}[L_1]}$ ($\widehat{\mathbb{E}^{\mathbb{Q}}[L_2]}$). In this example, separate hedging programs are developed for L_1 and L_2 since they are payable at different times T_1 and T_2. We use the low volatility parameter set as in the previous example. The time horizon from 0 to $T_2 = 20$ is divided into 1000 segments. At the end of each time segment, the portfolio is rebalanced in accordance with the delta estimates obtained from Theorem 2.2.4. This corresponds to a weekly rebalancing. If one can carry out a continuous hedging, the exact hedging costs should in theory match that of the risk-neutral expectation in Theorem 2.2.3. In practice, the actual hedging costs are different from the theoretical value and their differences are considered as the hedging error. We repeat the simulation for 50 times to compute the mean hedging costs and hedging error. One limitation of establishing a net liability dynamic hedging program is that it is very time-consuming. On every trading period, we need to compute the numerical integrals of special functions in Theorem 2.2.4. Since we are doing a weekly rebalancing, the numerical integrals are evaluated for 1,000 times per simulation. It becomes more and more time-consuming to evaluate the numerical integrals as the time to maturity gets shorter and shorter.

Comparison of two risk management strategies

We can also compare the dynamic hedging risk management approach with the traditional actuarial risk management method. As with the actuarial risk management method, we consider the risk measures for the present values of the hedging costs. The VaR and CTE risk measures with various confidence levels α for the two different methods are compared in Figure 2.5 and Figure 2.6, respectively.

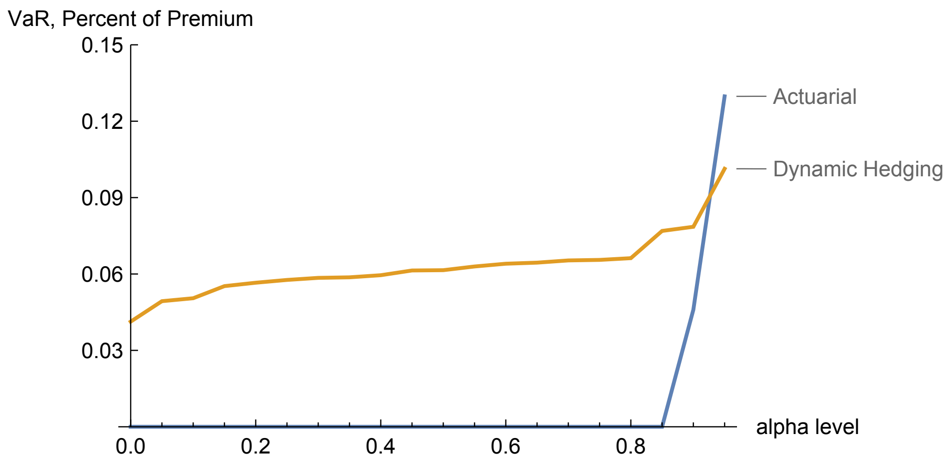

Figure 2.5: Quantile risk measures for the 20-year GMAB

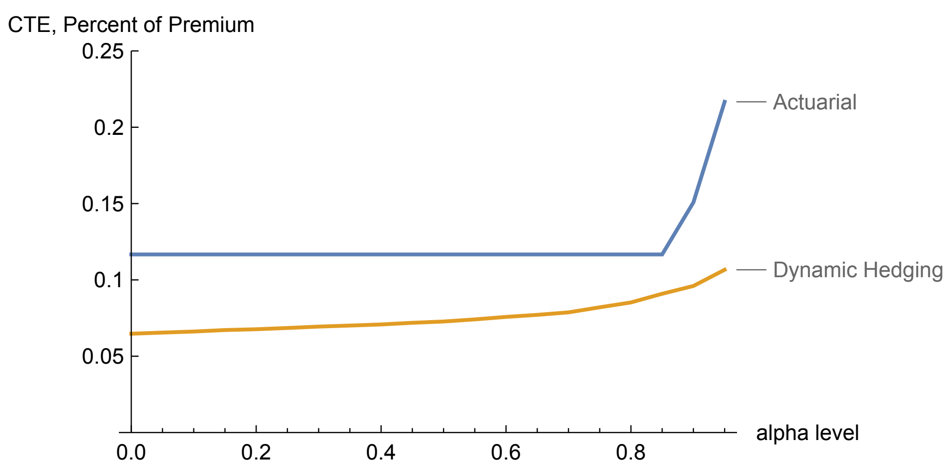

Figure 2.6: CTE risk measures for the 20-year GMAB

In the case of actuarial risk management, the VaR of the net liability stays zero with large probabilities, which means there is no liability most of the time. But risk measures rise sharply when α is very high, which indicates that the distribution of the net liability is in general heavy-tailed. In the case of dynamic hedging risk management, risk measures are always positive and concentrated at modest levels. This is because most of the net liabilities are hedged by the dynamic hedging strategy and the uncertainties mainly come from the hedging errors. One might argue that from an insurer's point of view, the dynamic hedging strategy has more stable outcomes than the actuarial risk management strategy. However, our empirical analysis does not support the common belief by practitioners that dynamic hedging is always less costly than traditional actuarial approach, particularly in the case of quantile risk measures. Note that, in the numerical example, we do not include transaction costs, which can significantly push up the cost of the dynamic hedging program in reality. Moreover, one should bear in mind that in this paper we assume for simplicity that the underlying is a single tradable equity index/fund. In practice, multiple third-party managed equity funds are involved and the mixes of assets in those equity funds are usually not known to insurers. The basis risk can also play a big role in driving up the cost and reduce the effectiveness of a hedging program.

2.2.5 Proofs of main theorems

Proof of Theorem 2.2.1. We first consider the case $V < e^{-rT_2}G_0$. In view of the fact that τ_x is independent of M_{T_1}, X_{t_1} and X_{t_2}, we have

$$\begin{aligned}\mathbb{P}_{x_0}(L > V) =&\mathbb{P}_{x_0}(T_1 < \tau_x < T_2)\mathbb{P}_{x_0}\left(e^{-rT_1}\frac{G_0}{F_0} - \frac{X_{t_1}}{x_0} > \frac{V}{F_0}\right)\\ &+ \mathbb{P}_{x_0}(T_2 < \tau_x)\mathbb{P}_{x_0}\left((e^{-rT_1}\frac{G_0}{F_0} - \frac{X_{t_1}}{x_0})_+ + e^{-rT_1}\frac{M_{T_1}}{F_0}(e^{-r(T_2-T_1)} - \frac{X_{t_2}}{x_0})_+ > \frac{V}{F_0}\right)\\ =&({}_{T_1}p_x - {}_{T_2}p_x)C\left(t_1, x_0, \frac{e^{-rT_1}G_0 - V}{F_0}x_0\right)\\ &+ {}_{T_2}p_x\mathbb{P}_{x_0}\left((e^{-rT_1}\frac{G_0}{F_0} - \frac{X_{t_1}}{x_0})_+ + e^{-rT_1}\frac{M_{T_1}}{F_0}(e^{-r(T_2-T_1)} - \frac{X_{t_2}}{x_0})_+ > \frac{V}{F_0}\right), \end{aligned} \tag{2.29}$$

where $C(\cdot)$ is defined in (2.6) in Theorem 2.1.1. Therefore,

$$\begin{aligned}\mathbb{P}_{x_0}(L > V) =&({}_{T_1}p_x - {}_{T_2}p_x)C(t_1, x_0, K_1(V)) +\\ &{}_{T_2}p_x\mathbb{P}_{x_0}\left((e^{-rT_1}\frac{G_0}{F_0} - \frac{X_{t_1}}{x_0})_+ + e^{-rT_1}\frac{M_{T_1}}{F_0}(e^{-r(T_2-T_1)} - \frac{X_{t_2}}{x_0})_+ > \frac{V}{F_0}\right),\end{aligned}$$

we just need to compute the probability of the last term in the previous equation. Conditioning on X_{t_2}, we can divide the probability into two parts

$$\begin{aligned}
&\mathbb{P}_{x_0}\left((e^{-rT_1}\frac{G_0}{F_0}-\frac{X_{t_1}}{x_0})_+ + e^{-rT_1}\frac{M_{T_1}}{F_0}(e^{-r(T_2-T_1)}-\frac{X_{t_2}}{x_0})_+ > \frac{V}{F_0}\right)\\
=&\mathbb{P}_{x_0}\left(L>V, X_{t_2}<x_0e^{-r(T_2-T_1)}\right)\\
&+\mathbb{P}_{x_0}\left(e^{-rT_1}\frac{G_0}{F_0}-\frac{X_{t_1}}{x_0}>\frac{V}{F_0}, X_{t_2}>x_0e^{-r(T_2-T_1)}\right)\\
=&\mathbb{P}_{x_0}\left(L>V, X_{t_2}<K_2(0)\right)+\mathbb{P}_{x_0}\left(X_{t_1}<K_1(V), X_{t_2}>K_2(0)\right). \qquad (2.30)
\end{aligned}$$

Using the independence of X_{t_1} and X_{t_2}, the last term in the previous equation is given by

$$\mathbb{P}_{x_0}\left(X_{t_1}<K_1(V), X_{t_2}>K_2(0)\right)=C(t_1,x_0,K_1(V))\left(1-C(x_0,t_2,K_2(0))\right). \qquad (2.31)$$

For the first term in (2.30), we further condition on X_{t_1} to have

$$\begin{aligned}
&\mathbb{P}_{x_0}\left(L>V, X_{t_2}<K_2(0)\right)\\
=&\mathbb{P}_{x_0}\left(\max\{e^{-rT_1}\frac{G_0}{F_0}, e^{2B^v_{t_1}}\}>\frac{x_0V}{F_0(x_0e^{-r(T_2-T_1)}-X_{t_2})}, X_{t_1}>x_0e^{-rT_1}\frac{G_0}{F_0}, X_{t_2}<K_2(0)\right)\\
&+\mathbb{P}_{x_0}\left(e^{-rT_1}\frac{G_0}{F_0}+e^{-rT_2}\frac{G_0}{F_0}-\frac{X_{t_1}}{x_0}-e^{-rT_1}\frac{G_0}{F_0}\frac{X_{t_2}}{x_0}>\frac{V}{F_0}, X_{t_1}<x_0e^{-rT_1}\frac{G_0}{F_0}, X_{t_2}<K_2(0)\right)\\
=&\mathbb{P}_{x_0}\left(\max\{e^{-rT_1}\frac{G_0}{F_0}, e^{2B^v_{t_1}}\}>\frac{x_0V}{F_0(K_2(0)-X_{t_2})}, X_{t_1}>K_1(0), X_{t_2}<K_2(0)\right)\\
&+\mathbb{P}_{x_0}\left(X_{t_1}<\min\{g(X_{t_2},V),K_1(0)\}, X_{t_2}<K_2(0)\right). \qquad (2.32)
\end{aligned}$$

For the first term in (2.32), we can use the independence between X_{t_1} and X_{t_2}, as well as (2.5) to get

$$\begin{aligned}
&\mathbb{P}_{x_0}\left(\max\{e^{-rT_1}\frac{G_0}{F_0}, e^{2B^v_{t_1}}\}>\frac{x_0V}{F_0(K_2(0)-X_{t_2})}, X_{t_1}>K_1(0), X_{t_2}<K_2(0)\right)\\
=&\int_0^{K_2(V)}\mathbb{P}_{x_0}\left(\max\{e^{-rT_1}\frac{G_0}{F_0}, e^{2B^v_{t_1}}\}>\frac{x_0V}{F_0(K_2(0)-y)}, X_{t_1}>K_1(0)\right)p(x_0,t_2,y)dy\\
&+\int_{K_2(V)}^{K_2(0)}\mathbb{P}_{x_0}\left(\max\{e^{-rT_1}\frac{G_0}{F_0}, e^{2B^v_{t_1}}\}>\frac{x_0V}{F_0(K_2(0)-y)}, X_{t_1}>K_1(0)\right)p(x_0,t_2,y)dy, \qquad (2.33)
\end{aligned}$$

where we use the fact that when $V<e^{-rT_2}G_0$, $0<K_2(V)<K_2(0)$. Moreover, when $0<y<K_2(V)$, we have

$$0<\frac{x_0V}{F_0(K_2(0)-y)}<e^{-rT_1}\frac{G_0}{F_0}<\max\{e^{-rT_1}\frac{G_0}{F_0}, e^{2B^v_{t_1}}\},$$

so the first term in (2.33) is given by

$$\int_0^{K_2(V)} \mathbb{P}_{x_0}\left(\max\{e^{-rT_1}\frac{G_0}{F_0}, e^{2B^v_{t_1}}\} > \frac{x_0 V}{F_0(K_2(0)-y)}, X_{t_1} > K_1(0)\right) p(x_0,t_2,y)dy$$
$$= (1 - C(x_0,t_1,K_1(0)))\, C(x_0,t_2,K_2(V)). \tag{2.34}$$

Similarly, when $K_2(V) < y < K_2(0)$, we have

$$\frac{x_0 V}{F_0(K_2(0)-y)} > e^{-rT_1}\frac{G_0}{F_0}.$$

In addition, by the fact that

$$X_{t_1} > x_0 e^{2B^v_{t_1}},$$

the second term in (2.33) can be simplified to

$$\int_{K_2(V)}^{K_2(0)} \mathbb{P}_{x_0}\left(\max\{e^{-rT_1}\frac{G_0}{F_0}, e^{2B^v_{t_1}}\} > \frac{x_0 V}{F_0(K_2(0)-y)}, X_{t_1} > K_1(0)\right) p(x_0,t_2,y)dy$$
$$= \int_{K_2(V)}^{K_2(0)} \mathbb{P}_{x_0}\left(e^{2B^v_{t_1}} > \frac{x_0 V}{F_0(K_2(0)-y)}\right) p(x_0,t_2,y)dy$$
$$= \int_{K_2(V)}^{K_2(0)} \mathcal{N}\left(v\sqrt{t_1} + f(y,V)\right) p(x_0,t_2,y)dy. \tag{2.35}$$

Adding up (2.34) and (2.35), we have

$$\mathbb{P}_{x_0}\left(\max\{e^{-rT_1}\frac{G_0}{F_0}, e^{2B^v_{t_1}}\} > \frac{x_0 V}{F_0(K_2(0)-X_{t_2})}, X_{t_1} > K_1(0), X_{t_2} < K_2(0)\right)$$
$$= (1 - C(x_0,t_1,K_1(0)))\, C(x_0,t_2,K_2(V)) + \int_{K_2(V)}^{K_2(0)} \mathcal{N}\left(v\sqrt{t_1} + f(y,V)\right) p(x_0,t_2,y)dy. \tag{2.36}$$

To compute the second term in (2.32), we can also use the independence between X_{t_1} and X_{t_2}. To be noted that when $0 < X_{t_2} < K_2(V)$,

$$g(X_{t_2}, V) > K_1(0),$$

and when $X_{t_2} > K_2(V)$,

$$g(X_{t_2}, V) < K_1(0).$$

Therefore, we have

$$\begin{aligned}
&\mathbb{P}_{x_0}\left(X_{t_1} < \min\{g(X_{t_2}, V), K_1(0)\}, X_{t_2} < K_2(0)\right) \\
=&\int_0^{K_2(V)} \mathbb{P}_{x_0}\left(X_{t_1} < K_1(0)\right) p(x_0, t_2, y) dy + \int_{K_2(V)}^{K_2(0)} \mathbb{P}_{x_0}\left(X_{t_1} < g(X_{t_2}, V)\right) p(x_0, t_2, y) dy \\
=&C(x_0, t_1, K_1(0)) C(x_0, t_2, K_2(V)) + \int_{K_2(V)}^{K_2(0)} C\left(x_0, t_1, g(X_{t_2}, V)\right) p(x_0, t_2, y) dy. \qquad (2.37)
\end{aligned}$$

Collecting the terms in (2.31), (2.36) and (2.37), we have the desired formula in (2.20) .

If $e^{-rT_2}G_0 < V < e^{-rT_1}G_0$, we have $K_2(V) < 0$. It is easy to see that (2.33) becomes

$$\begin{aligned}
&\mathbb{P}_{x_0}\left(\max\{e^{-rT_1}\frac{G_0}{F_0}, e^{2B^v_{t_1}}\} > \frac{x_0 V}{F_0(K_2(0) - X_{t_2})}, X_{t_1} > K_1(0), X_{t_2} < K_2(0)\right) \\
=&\int_0^{K_2(0)} \mathcal{N}\left(v\sqrt{t_1} + f(y, V)\right) p(x_0, t_2, y) dy.
\end{aligned}$$

Similarly, (2.37) becomes

$$\mathbb{P}_{x_0}\left(X_{t_1} < \min\{g(X_{t_2}, V), K_1(0)\}, X_{t_2} < K_2(0)\right) = \int_0^{K_2(0)} C\left(x_0, t_1, g(X_{t_2}, V)\right) p(x_0, t_2, y) dy.$$

Therefore, these facts imply the second formula in (2.21).

Finally, if $V > e^{-rT_1}G_0$, we have $K_1(V) < 0$ and $K_2(V) < 0$. In addition to the changes in (2.33) and (2.37), (2.29) changes to

$$\mathbb{P}_{x_0}(L > V) = {}_{T_2}p_x \mathbb{P}_{x_0}\left((e^{-rT_1}\frac{G_0}{F_0} - \frac{X_{t_1}}{x_0})_+ + e^{-rT_1}\frac{M_{T_1}}{F_0}(e^{-r(T_2 - T_1)} - \frac{X_{t_2}}{x_0})_+ > \frac{V}{F_0}\right).$$

These facts imply the third formula in (2.22).

□

Proof of Theorem 2.2.2. Similar to the proof of Theorem 2.2.1, we first consider the case $V < e^{-rT_2}G_0$. First of all, we have

$$\begin{aligned}
\mathbb{E}_{x_0}[L\mathbb{1}_{\{L>V\}}] &= \mathbb{P}(T_1 < \tau_x < T_2)\mathbb{E}_{x_0}[L\mathbb{1}_{\{L>V\}} | T_1 < \tau_x < T_2] + \mathbb{P}(\tau_x > T_2)\mathbb{E}_{x_0}[L\mathbb{1}_{\{L>V\}} | T_2 < \tau_x] \\
&= ({}_{T_1}p_x - {}_{T_2}p_x)\mathbb{E}_{x_0}[L\mathbb{1}_{\{L>V\}} | T_1 < \tau_x < T_2] + {}_{T_2}p_x \mathbb{E}_{x_0}[L\mathbb{1}_{\{L>V\}} | T_2 < \tau_x]. \qquad (2.38)
\end{aligned}$$

For the first term in the last equation, we have

$$
\begin{aligned}
&\mathbb{E}_{x_0}[L\mathbb{1}_{\{L>V\}}|T_1<\tau_x<T_2] \\
=&\mathbb{E}_{x_0}\left[\frac{F_0}{x_0}(K_1(0)-X_{t_1})_+\mathbb{1}_{\{e^{-rT_1}\frac{G_0}{F_0}-\frac{X_{t_1}}{x_0}>\frac{V}{F_0}\}}\right] \\
=&\frac{F_0}{x_0}K_1(0)C(t_1,x_0,K_1(V))-\frac{F_0}{x_0}Z(t_1,x_0,K_1(V)).
\end{aligned}
$$

Therefore, we only need to compute the second term in (2.38):

$$\mathbb{E}_{x_0}[L\mathbb{1}_{\{L>V\}}|T_2<\tau_x].$$

Using the same technique we employed in the proof of Theorem 2.2.1, we have

$$
\begin{aligned}
&\mathbb{E}_{x_0}[L\mathbb{1}_{\{L>V\}}|T_2<\tau_x] \\
=&\frac{F_0}{x_0}\mathbb{E}_{x_0}\left[(K_1(0)-X_{t_1})_+\mathbb{1}_{\{X_{t_1}<K_1(V),X_{t_2}>K_2(0)\}}\right]+\mathbb{E}_{x_0}\left[L\mathbb{1}_{\{L>V,X_{t_2}<K_2(0)\}}\right] \\
=&\frac{F_0}{x_0}\left(1-C(x_0,t_2,K_2(0))\right)\left(K_1(0)C(x_0,t_1,K_1(V))-Z(x_0,t_1,K_1(V))\right)+\mathbb{E}_{x_0}\left[L\mathbb{1}_{\{L>V,X_{t_2}<K_2(0)\}}\right]. \quad (2.39)
\end{aligned}
$$

For the last term, we further condition on X_{t_1} as we did in the VaR's proof to have

$$
\begin{aligned}
&\mathbb{E}_{x_0}\left[L\mathbb{1}_{\{L>V,X_{t_2}<K_2(0)\}}\right] \\
=&\frac{F_0}{x_0}\mathbb{E}_{x_0}\left[(K_1(0)-X_{t_1})_+\mathbb{1}_{\{X_{t_1}<\min\{g(X_{t_2},V),K_1(0)\},X_{t_2}<K_2(0)\}}\right] \\
&+\frac{F_0}{x_0}\mathbb{E}_{x_0}\left[\max\{e^{-rT_1}\frac{G_0}{F_0},e^{2B^v_{t_1}}\}(K_2(0)-X_{t_2})_+\mathbb{1}_{\{X_{t_1}<\min\{g(X_{t_2},V),K_1(0)\},X_{t_2}<K_2(0)\}}\right] \\
&+\frac{F_0}{x_0}\mathbb{E}_{x_0}\left[\max\{e^{-rT_1}\frac{G_0}{F_0},e^{2B^v_{t_1}}\}(K_2(0)-X_{t_2})_+\mathbb{1}_{\{\max\{e^{-rT_1}\frac{G_0}{F_0},e^{2B^v_{t_1}}\}>\frac{x_0V}{F_0(K_2(0)-X_{t_2})},X_{t_1}>K_1(0),X_{t_2}<K_2(0)\}}\right].
\end{aligned}
\tag{2.40}
$$

Using the independence of X_{t_1} and X_{t_2}, as well as the fact that when $0<X_{t_2}<K_2(V)$,

$$g(X_{t_2},V)>K_1(0),$$

and when $K_2(V)<X_{t_2}<K_2(0)$,

$$g(X_{t_2},V)<K_1(0),$$

the first two terms in (2.40) can be computed as

$$\begin{aligned}
&\mathbb{E}_{x_0}\left[(K_1(0)-X_{t_1})_+\mathbb{1}_{\{X_{t_1}<\min\{g(X_{t_2},V),K_1(0)\},X_{t_2}<K_2(0)\}}\right]\\
=&\mathbb{E}_{x_0}\left[(K_1(0)-X_{t_1})_+\mathbb{1}_{\{X_{t_1}<K_1(0),X_{t_2}<K_2(V)\}}\right]\\
&+\int_{K_2(V)}^{K_2(0)}\mathbb{E}_{x_0}\left[(K_1(0)-X_{t_1})_+\mathbb{1}_{\{X_{t_1}<g(y,V)\}}\right]p(x_0,t_2,y)dy\\
=&\left(K_1(0)C(x_0,t_1,K_1(0))-Z(x_0,t_1,K_1(0))\right)C(x_0,t_2,K_2(V))\\
&+\int_{K_2(V)}^{K_2(0)}\left(K_1(0)C(x_0,t_1,g(y,V))-Z(x_0,t_1,g(y,V))\right)p(x_0,t_2,y)dy, \qquad (2.41)
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{E}_{x_0}\left[\max\{e^{-rT_1}\frac{G_0}{F_0},e^{2B^v_{t_1}}\}(K_2(0)-X_{t_2})_+\mathbb{1}_{\{X_{t_1}<\min\{g(X_{t_2},V),K_1(0)\},X_{t_2}<K_2(0)\}}\right]\\
=&\int_0^{K_2(V)}\mathbb{E}_{x_0}\left[\max\{e^{-rT_1}\frac{G_0}{F_0},e^{2B^v_{t_1}}\}\mathbb{1}_{\{X_{t_1}<K_1(0)\}}\right](K_2(0)-y)p(x_0,t_2,y)dy\\
&+\int_{K_2(V)}^{K_2(0)}\mathbb{E}_{x_0}\left[\max\{e^{-rT_1}\frac{G_0}{F_0},e^{2B^v_{t_1}}\}\mathbb{1}_{\{X_{t_1}<g(y,V)\}}\right](K_2(0)-y)p(x_0,t_2,y)dy\\
=&\int_0^{K_2(V)}\mathbb{E}_{x_0}\left[\max\{e^{-rT_1}\frac{G_0}{F_0},e^{2B^v_{t_1}}\}\mathbb{1}_{\{X_{t_1}<K_1(0)\}}\right](K_2(0)-y)p(x_0,t_2,y)dy\\
&+\frac{K_1(0)}{x_0}\int_{K_2(V)}^{K_2(0)}C(x_0,t_1,g(y,V))(K_2(0)-y)p(x_0,t_2,y)dy. \qquad (2.42)
\end{aligned}$$

For the third term in (2.40), we further condition on X_{t_2} to compute it. To be noted that when $0<X_{t_2}<K_2(V)$,

$$0<\frac{x_0V}{F_0(K_2(0)-X_{t_2})}<e^{-rT_1}\frac{G_0}{F_0}<\max\{e^{-rT_1}\frac{G_0}{F_0},e^{2B^v_{t_1}}\},$$

and when $K_2(V)<X_{t_2}<K_2(0)$,

$$\frac{x_0V}{F_0(K_2(0)-X_{t_2})}>e^{-rT_1}\frac{G_0}{F_0},$$

we have

$$\begin{aligned}
&\mathbb{E}_{x_0}\left[\max\{e^{-rT_1}\frac{G_0}{F_0},e^{2B^v_{t_1}}\}(K_2(0)-X_{t_2})_+\mathbb{1}_{\{\max\{e^{-rT_1}\frac{G_0}{F_0},e^{2B^v_{t_1}}\}>\frac{x_0V}{F_0(K_2(0)-X_{t_2})},X_{t_1}>K_1(0),X_{t_2}<K_2(0)\}}\right]\\
=&\int_0^{K_2(V)}\mathbb{E}_{x_0}\left[\max\{e^{-rT_1}\frac{G_0}{F_0},e^{2B^v_{t_1}}\}\mathbb{1}_{\{X_{t_1}>K_1(0)\}}\right](K_2(0)-y)p(x_0,t_2,y)dy\\
&+\int_{K_2(V)}^{K_2(0)}\mathbb{E}_{x_0}\left[e^{2B^v_{t_1}}\mathbb{1}_{\{e^{2B^v_{t_1}}>\frac{x_0V}{F_0(K_2(0)-y)}\}}\right](K_2(0)-y)p(x_0,t_2,y)dy. \qquad (2.43)
\end{aligned}$$

The first term in (2.42) and (2.43) can be combined to

$$\begin{aligned}&\int_0^{K_2(V)} \mathbb{E}_{x_0}\left[\max\{e^{-rT_1}\frac{G_0}{F_0}, e^{2B^v_{t_1}}\}\right](K_2(0)-y)p(x_0,t_2,y)dy\\ =&\left(K_2(0)C(x_0,t_2,K_2(V)) - Z(x_0,t_2,K_2(V))\right)\mathbb{E}_{x_0}\left[\max\{e^{-rT_1}\frac{G_0}{F_0}, e^{2B^v_{t_1}}\}\right].\end{aligned}$$

Since $B^v_{t_1}$ is a linear Brownian motion, it is easy to compute that

$$\begin{aligned}&\mathbb{E}_{x_0}\left[\max\{e^{-rT_1}\frac{G_0}{F_0}, e^{2B^v_{t_1}}\}\right]\\ =&\frac{K_1(0)}{x_0}\mathcal{N}\left(\frac{1}{2\sqrt{t_1}}\log\left(\frac{K_1(0)}{x_0}\right) - v\sqrt{t_1}\right) + e^{2(v+1)t_1}\mathcal{N}\left((v+2)\sqrt{t_1} - \frac{1}{2\sqrt{t_1}}\log\left(\frac{K_1(0)}{x_0}\right)\right)\end{aligned}$$

and

$$\mathbb{E}_{x_0}\left[e^{2B^v_{t_1}}\mathbb{1}_{\{e^{2B^v_{t_1}} > \frac{x_0 V}{F_0(K_2(0)-y)}\}}\right] = e^{2(v+1)t_1}\mathcal{N}\left((v+2)\sqrt{t_1} + f(y,V)\right).$$

Collecting the terms in (2.39), (2.41), (2.42) and (2.43), we can obtain the first formula in the theorem.

If $e^{-rT_2}G_0 < V < e^{-rT_1}G_0$, we have $K_2(V) < 0$. It is easy to see that (2.41) becomes

$$\begin{aligned}&\mathbb{E}_{x_0}\left[(K_1(0) - X_{t_1})_+\mathbb{1}_{\{X_{t_1} < \min\{g(X_{t_2},V),K_1(0)\}, X_{t_2} < K_2(0)\}}\right]\\ =&\int_0^{K_2(0)}\left(K_1(0)C(x_0,t_1,g(y,V)) - Z(x_0,t_1,g(y,V))\right)p(x_0,t_2,y)dy,\end{aligned} \tag{2.44}$$

and (2.42) becomes

$$\begin{aligned}&\mathbb{E}_{x_0}\left[\max\{e^{-rT_1}\frac{G_0}{F_0}, e^{2B^v_{t_1}}\}(K_2(0)-X_{t_2})_+\mathbb{1}_{\{X_{t_1} < \min\{g(X_{t_2},V),K_1(0)\}, X_{t_2} < K_2(0)\}}\right]\\ =&\frac{K_1(0)}{x_0}\int_0^{K_2(0)} C(x_0,t_1,g(y,V))(K_2(0)-y)p(x_0,t_2,y)dy,\end{aligned}$$

and (2.43) becomes

$$\begin{aligned}&\mathbb{E}_{x_0}\left[\max\{e^{-rT_1}\frac{G_0}{F_0}, e^{2B^v_{t_1}}\}(K_2(0)-X_{t_2})_+\mathbb{1}_{\{\max\{e^{-rT_1}\frac{G_0}{F_0}, e^{2B^v_{t_1}}\} > \frac{x_0 V}{F_0(K_2(0)-X_{t_2})}, X_{t_1} > K_1(0), X_{t_2} < K_2(0)\}}\right]\\ =&\int_0^{K_2(0)}\mathbb{E}_{x_0}\left[e^{2B^v_{t_1}}\mathbb{1}_{\{e^{2B^v_{t_1}} > \frac{x_0 V}{F_0(K_2(0)-y)}\}}\right](K_2(0)-y)p(x_0,t_2,y)dy.\end{aligned}$$

Collecting all the previous terms, we have

$$
\begin{aligned}
\mathbb{E}_{x_0}[L\mathbb{1}_{\{L>V\}}] =& \frac{F_0}{x_0}\left({}_{T_1}p_x - {}_{T_2}p_x C(x_0,t_2,K_2(0))\right)\left(K_1(0)C(t_1,x_0,K_1(V)) - Z(t_1,x_0,K_1(V))\right) \\
&+ {}_{T_2}p_x \frac{F_0}{x_0}\int_0^{K_2(0)} \left(K_1(0)C(x_0,t_1,g(y,V)) - Z(x_0,t_1,g(y,V))\right) p(x_0,t_2,y)dy \\
&+ {}_{T_2}p_x \frac{K_1(0)F_0}{x_0^2}\int_0^{K_2(0)} C(x_0,t_1,g(y,V))(K_2(0)-y)p(x_0,t_2,y)dy \\
&+ {}_{T_2}p_x \frac{F_0}{x_0}\int_0^{K_2(0)} e^{2(v+1)t_1}\mathcal{N}\left((v+2)\sqrt{t_1} + f(y,V)\right)(K_2(0)-y)p(x_0,t_2,y)dy.
\end{aligned}
$$

Finally, if $V > e^{-rT_1}G_0$, we have $K_1(V) < 0$ and $K_2(V) < 0$. In addition to the changes in the previous case, the conditional expectation $\mathbb{E}_{x_0}[L\mathbb{1}_{\{L>V\}}|T_1 < \tau_x < T_2]$ will also vanish. Therefore, we have

$$
\begin{aligned}
&\mathbb{E}_{x_0}[L\mathbb{1}_{\{L>V\}}] \\
=& {}_{T_2}p_x \frac{F_0}{x_0}\int_0^{x_0 e^{-rT_1} + x_0 e^{-rT_2} - x_0\frac{V}{G_0}} \left(K_1(0)C(x_0,t_1,g(y,V)) - Z(x_0,t_1,g(y,V))\right) p(x_0,t_2,y)dy \\
&+ {}_{T_2}p_x \frac{K_1(0)F_0}{x_0^2}\int_0^{x_0 e^{-rT_1} + x_0 e^{-rT_2} - x_0\frac{V}{G_0}} C(x_0,t_1,g(y,V))(K_2(0)-y)p(x_0,t_2,y)dy \\
&+ {}_{T_2}p_x \frac{F_0}{x_0}\int_0^{K_2(0)} e^{2(v+1)t_1}\mathcal{N}\left((v+2)\sqrt{t_1} + f(y,V)\right)(K_2(0)-y)p(x_0,t_2,y)dy,
\end{aligned}
$$

which completes the proof. □

Proof of Theorem 2.2.4. Given the formula in (2.28), we first consider $\frac{\partial\left(\frac{F_0}{x_0}P(t_1,x_0,K_1(0))\right)}{\partial F_0}$. From the definitions in Theorem 2.1.1, we have

$$
\frac{F_0}{x_0}P(t_1,x_0,K_1(0)) = \frac{F_0}{x_0}K_1(0)C(t_1,x_0,K_1(0)) - \frac{F_0}{x_0}Z(t_1,x_0,K_1(0)).
$$

It is easy see that

$$
\begin{aligned}
\frac{\partial C(t_1,x_0,K_1(0))}{\partial K_1(0)} &= p(t_1,x_0,K_1(0)), \\
\frac{\partial Z(t_1,x_0,K_1(0))}{\partial K_1(0)} &= K_1(0)p(t_1,x_0,K_1(0)),
\end{aligned}
$$

so we have

$$\begin{aligned}
&\frac{\partial\left(\frac{F_0}{x_0}P(t_1,x_0,K_1(0))\right)}{\partial F_0}\\
=&\frac{F_0K_1(0)}{x_0}p(t_1,x_0,K_1(0))\frac{\partial K_1(0)}{\partial F_0}-\frac{1}{x_0}Z(t_1,x_0,K_1(0))-\frac{F_0K_1(0)}{x_0}p(t_1,x_0,K_1(0))\frac{\partial K_1(0)}{\partial F_0}\\
=&-\frac{1}{x_0}Z(t_1,x_0,K_1(0)). \qquad (2.45)
\end{aligned}$$

This proves the formula for Δ_1. For the other term, we find that

$$\begin{aligned}
\mathbb{E}^{\mathbb{Q}}_{x_0}[L_2] &= {}_{T_2}p_x\frac{1}{x_0}\mathbb{E}^{\mathbb{Q}}\left[\max\{e^{-rT_1}G_0,F_0e^{2B^v_{t_1}}\}\right]\mathbb{E}^{\mathbb{Q}}\left[(K_2(0)-X_{t_2})_+\right]\\
&= {}_{T_2}p_x\frac{P(t_2,x_0,K_2(0))}{x_0}\mathbb{E}^{\mathbb{Q}}\left[\max\{e^{-rT_1}G_0,F_0e^{2B^v_{t_1}}\}\right].
\end{aligned}$$

The only part of the previous equation that depends on F_0 is the last expectation, which corresponds to the delta of a vanilla call option. We can easily compute that

$$\frac{\partial\mathbb{E}^{\mathbb{Q}}_{x_0}\left(\max\{e^{-rT_1}G_0,F_0e^{2B^v_{t_1}}\}\right)}{\partial F_0}=e^{2(v+1)t_1}\mathcal{N}\left((v+2)\sqrt{t_1}-\frac{1}{2\sqrt{t_1}}\log\left(\frac{K_1(0)}{x_0}\right)\right). \qquad (2.46)$$

Therefore, we obtain the formula for Δ_2 in the theorem.

□

Chapter 3

Jump Diffusions with Affine Drift and GMWB Pricing

3.1 Exponential functionals of Lévy processes with mixed-exponential jumps

In this section, we focus on the exponential functionals of a class of Lévy processes and we shall review some distributional properties of the exponential functionals. The analytical results in this section are used in solving the pricing problem of a GMWB rider. With recent developments in this field ([12, 45]), it is possible to analytically determine the exponential functionals for a large class of Lévy processes. Originally, Cai and Kou ([12]) obtained an analytical expression for the Mellin transform of the exponential functional of Lévy processes with hyper-exponential jumps (See [12] for definition). Later, Kuznetsov ([45]) extended their results to Lévy processes with jumps of rational transform. These results enable us to develop an efficient semi-analytical algorithm to price the GMWB rider. Readers are referred to [45] for the detailed proofs of the results reviewed in this section.

Let X_t be a Lévy process which has a linear drift coefficient μ, a diffusion coefficient σ and a Lévy measure $\pi(dx)$ satisfying

$$\pi(dx) = \sum_{i=1}^{J} \alpha_i \rho_i e^{-\rho_i x} \mathbb{1}_{\{x \geq 0\}} dx + \sum_{i=1}^{\hat{J}} \hat{\alpha}_i \hat{\rho}_i e^{\hat{\rho}_i x} \mathbb{1}_{\{x < 0\}} dx$$

where $0 < \rho_1 < \rho_2 < \cdots < \rho_J$, $0 < \hat{\rho}_1 < \hat{\rho}_2 < \cdots < \hat{\rho}_{\hat{J}}$ and $\alpha_i, \hat{\alpha}_j \in \mathbb{R}$ for $1 \leq i \leq J$, $1 \leq j \leq \hat{J}$. Equivalently, we can write the Lévy process X_t as

$$X_t = \mu t + \sigma W_t + \sum_{i=1}^{N_t} \xi_i, \tag{3.1}$$

where $\{N_t, t \geq 0\}$ is a Poisson process with intensity

$$\lambda = \sum_{i=1}^{J} \alpha_i + \sum_{i=1}^{\hat{J}} \hat{\alpha}_i > 0,$$

and $\{\xi_i, i \in \mathbb{N}\}$ are i.i.d random variables with the p.d.f

$$\frac{\pi(dx)}{\lambda}.$$

We assume that $\frac{\pi(dx)}{\lambda}$ is a proper p.d.f (For some necessary and sufficient conditions, see [13] and [4]). This implies that the Lévy process X_t is the sum of a linear Brownian motion and a compound Poisson process (independent of the linear Brownian motion) whose jump distribution is a mixture of exponential distributions. By the Lévy-Khinchine formula, the Laplace exponent $\psi(z) := \ln(\mathbb{E}[e^{zX_1}])$ is given by

$$\psi(z) = \frac{\sigma^2}{2}z^2 + \mu z + \left(\sum_{i=1}^{J} \frac{\alpha_i \rho_i}{\rho_i - z} + \sum_{i=1}^{\hat{J}} \frac{\hat{\alpha}_i \hat{\rho}_i}{\hat{\rho}_i + z} \right) - \lambda,$$

for $z \in (-\hat{\rho}_1, \rho_1)$. We can see that the Laplace exponent $\psi(z)$ is a rational function and it has J $\{\hat{J}\}$ poles in the half-plane $\Re(z) > 0$ $\{\Re(z) < 0\}$. To keep the same notations as [45], let $P = J + \hat{J}$ be the total number of poles of $\psi(z)$ and let Q be the total number of roots of $\psi(z)$, where

$$Q = \begin{cases} P+2, & \text{if } \sigma > 0, \\ P+1, & \text{if } \sigma = 0 \text{ and } \mu \neq 0, \\ P, & \text{if } \sigma = \mu = 0. \end{cases}$$

Moreover for $q \geq 0$, the zeros of $\psi(z) - q$ in the half-plane $\Re(z) > 0$ $\{\Re(z) < 0\}$ are denoted by $\zeta_1, \zeta_2, \ldots, \zeta_K$ $\{-\hat{\zeta}_1, -\hat{\zeta}_2, \ldots, -\hat{\zeta}_{\hat{K}}\}$. Therefore, we have $Q = K + \hat{K}$ by definition. Let e_q be an independent exponential random variable (of process X_t) with rate q and define the exponential functional

$$I_q := \int_0^{e_q} e^{X_s} ds.$$

Our main focus in this section is to characterize the law of I_q. [45] obtained an analytical expression of the p.d.f of the exponential functional I_q for a more general class of Lévy processes, which includes the Lévy processes with the jump part being a mixture of Gamma distributions. Hence, our situation here is a special case of [45]. The reason we do not consider the most general case in [45] is that: there is no need to increase model complexity. In practice, we use a mixture of distributions to approximate the jump distribution we want. It is well known that the mixed-exponential distribution is dense in the space of all distribution functions (See [9]). Therefore, mixed-exponential distributions can approximate any jump distribution. We do not need to consider more complicated mixed-Gamma distributions since they introduce more parameters into the model. We state the main result concerning the distribution of the exponential functional I_q as follows, which was obtained by [45].

Define vectors $\boldsymbol{a} \in \mathbb{C}^{P+1}$ and $\boldsymbol{b} \in \mathbb{C}^{Q}$ as

$$\begin{aligned}\boldsymbol{a} &= [1, 1-\hat{\rho}_1, 1-\hat{\rho}_2, \ldots, 1-\hat{\rho}_{\hat{J}}, 1+\rho_1, 1+\rho_2, \ldots, 1+\rho_J] \\ \boldsymbol{b} &= [1+\zeta_1, 1+\zeta_2, \ldots, 1+\zeta_K, 1-\hat{\zeta}_1, 1-\hat{\zeta}_2, \ldots, 1-\hat{\zeta}_{\hat{K}}], \end{aligned} \tag{3.2}$$

and constants A and B as

$$A = \begin{cases} \frac{\sigma^2}{2} & \text{if } \sigma > 0, \\ |\mu| & \text{if } \sigma = 0 \text{ and } \mu \neq 0, \\ q+\lambda & \text{if } \sigma = \mu = 0, \end{cases}$$

$$B = \frac{\Pi_{j=1}^{K}\Gamma(\zeta_j)}{\Pi_{j=1}^{J}\Gamma(\rho_j)} \times \frac{\Pi_{j=1}^{\hat{J}}\Gamma(1+\hat{\rho}_j)}{\Pi_{j=1}^{\hat{K}}\Gamma(1+\hat{\zeta}_j)},$$

where $\Gamma(x)$ is the Gamma function (See appendix for its definition). Furthermore, we assume that **Assumption A** in [45] are satisfied:

Assumption A

A.1 $\psi(z) - q$ has no multiple zeros in the half-plane $\Re(z) > 0$

A.2 For $1 \leq i \leq \hat{J}$, $\hat{\rho}_i \notin \mathbb{N}$

A.3 For $1 \leq i < j \leq \hat{J}$, $\hat{\rho}_j - \hat{\rho}_i \notin \mathbb{N}$

A.4 For $1 \leq i < j \leq K$, $\zeta_j - \zeta_i \notin \mathbb{N}$

The following theorem characterizes the p.d.f of the exponential functional I_q.

Theorem 3.1.1 (Kuznetsov (2012))**.** *For $q > 0$, the probability density function $p(x) := \mathbb{P}(I_q \in dx)/dx$ of I_q exists and it has the following representation in terms of the Meijer G-function:*

$$p(x) = \frac{A}{B} G_{P+1,Q}^{K,\hat{J}+1}\left(\frac{1}{Ax} \middle| \begin{matrix} \boldsymbol{a} \\ \boldsymbol{b} \end{matrix} \right).$$

G is the Meijer G-function (See appendix for its definition).

Proof. See [45, Proposition 3]. □

Theorem 3.1.1 characterizes the p.d.f of the exponential functional of Lévy processes with mixed-exponential jumps analytically in terms of a special function. Using certain identities of the Meijer G-function, we can obtain a similar representation of the tail probability of the exponential functional I_q.

Corollary 3.1.2. *For $q > 0$, the tail probability function of the exponential functional $\bar{F}(x) := \mathbb{P}(I_q > x)$ is given by*

$$\bar{F}(x) = \frac{Ax}{B} G^{K,\hat{J}+2}_{P+2,Q+1}\left(\frac{1}{Ax} \middle| \begin{matrix} 2, \boldsymbol{a} \\ \boldsymbol{b}, 1 \end{matrix} \right). \tag{3.3}$$

.

Proof. Observe that

$$\begin{aligned} \bar{F}(x) &= \int_x^\infty p(y)dy \\ &= \frac{A}{B} \int_x^\infty G^{K,\hat{J}+1}_{P+1,Q}\left(\frac{1}{Ay} \middle| \begin{matrix} \boldsymbol{a} \\ \boldsymbol{b} \end{matrix} \right) dy \\ &= \frac{Ax}{B} \int_0^1 y^{-2} G^{K,\hat{J}+1}_{P+1,Q}\left(\frac{y}{Ax} \middle| \begin{matrix} \boldsymbol{a} \\ \boldsymbol{b} \end{matrix} \right) dy. \end{aligned}$$

Using identity [54, p416, (16.19.6)], we can directly obtain the desired expression in (3.3). □

As stated in [45], **Assumption A** is often satisfied in practice. For modelings in financial markets, we can always avoid using integer rates $\hat{\rho}_j$ and avoid having integer differences between rates $\hat{\rho}_j$ and $\hat{\rho}_i$. In fact, in numerical evaluations, the difference between roots ζ_j and ζ_i can hardly be an integer. In regard to numerical implementations, numerical computations of the Meijer G-function and the Gamma function are available in most numerical or symbolic computing software, such as MATLAB, Mathematica, and Maple. These special functions can be efficiently and accurately computed. As we shall see, these special functions often arise when we take Laplace transforms of certain quantities in the pricing problem of a GMWB. To obtain the original solution, numerical Laplace inversion algorithm is needed. There are a variety of easy-to-implement and fast algorithms based on the Fourier-series method or the Bromwich inversion integral (For example, [1, 2]), all of which give efficient numerical solutions to the pricing problem of the GMWB.

3.2 Guaranteed Minimum Withdrawal Benefit

Like the GMAB rider in Chapter 2, a GMWB rider is another type of variable annuity products embedded with exotic options. It guarantees that the policyholder can withdraw up to a fixed amount each year until the entire initial investment is returned regardless of the market performance. Moreover, the contract will not expire until the initial payment is fully returned. If the policyholder's account value is driven to zero before the total withdrawals add up to his initial investment, the insurer is responsible for paying the remaining guaranteed annual withdrawals. After the entire initial investment is withdrawn, the GMWB expires. The policyholder is left with the remaining value of his account. The work of Milevsky and Salisbury ([51]) was among the first works to provide a valuation framework for the GMWB. They introduced two approaches to price the GMWB: the static approach and the dynamic approach. In the static approach, policyholders are assumed to behave passively in making use of their guarantees. In dynamic approach, however, it is assumed that policyholders electing GMWB riders dynamically seek to maximize their contract values by surrendering the product at an optimal time, which leads to an optimal control problem. In this section, we focus on the static approach in which policyholders do not surrender the GMWB contract in relation to the contract value. Assumptions on time-varying surrenders can be included, however, in the mortality assumptions. To offer an example, suppose that a policyholder makes an initial investment of 100 dollars in his variable annuity sub account. If the guaranteed withdrawal rate is 7% (according to [51], this is what most insurance companies offer) per year, then the maturity of the GMWB should be $100/7 \approx 14.3$ years. Say, the policyholder takes withdrawals of 7 dollars per year until the end of 14.3 years. Then he is entitled to a terminal payment of the remaining value on his account (if there is any) when the GMWB matures. To fund the guarantee rider, the insurer charges fees on a daily basis from the policyholder's account. Usually, the fee charges are a fixed percentage of the current account value. In [51], Milevsky and Salisbury used geometric Brownian motions to model the dynamics of the investment fund and decomposed the product into a Quanto Asian Put option and a generic term-certain annuity. They solved the pricing problem of the Quanto Asian Put option by numerical PDE methods. Later, Feng and Volkmer ([35]) used a different approach to solve the pricing problem of the GMWB in the same model. That approach utilized certain analytical representations of the density of integrated geometric Brownian motions and made their algorithm much more efficient than numerical PDE methods.

This work presents an extension to [35], in which the dynamics of the equity-linked investment can be driven by an exponential Lévy process with mixed-exponential jumps. One motivation for introducing jump diffusion is to generalize the geometric Brownian motion model used by [35] and [51]. Also, empirical evidence shows that asset prices have heavier tails than the tail of a log-normal distribution. This is also referred as the "volatility smile" of the classic Black-Scholes model by practitioners (See [38, Chapter 19]). Jump models are potential alternatives to address this problem (See [38, 13]). Moreover, Lévy processes with mixed-exponential jumps are a large class of Lévy processes as we mentioned in the first section of this chapter: mixed-exponential distributions are dense in the space of all distribution functions so that they able to approximate any jump distribution (with finite intensity). Finally, we shall see that the analytical results in the first section provide more tractability of our models which make numerical computations efficient. We do not need to use such numerical methods as Monte Carlo simulations or solving integro-differential equations, which are not efficient.

3.2.1 Modeling and pricing of the GMWB

Assume the underlying investment fund S_t of a GMWB rider is driven by an exponential Lévy process under the physical measure:

$$S_t = xe^{X_t},$$

where X_t is a Lévy process with mixed-exponential jumps defined in (3.1). Since we consider the pricing problem of the GMWB in a non-arbitrage setting, we are interested in the dynamics of S_t under the risk neutral measure instead of the physical measure. We can choose a risk neutral measure $\mathbb{Q}$ such that the dynamics of S_t under the risk neutral measure is governed by the following SDE

$$\frac{dS_t}{S_{t-}} = \left(r - \sum_{j=1}^{J} \frac{\alpha_j \rho_j}{\rho_j - 1} - \sum_{j=1}^{\hat{J}} \frac{\hat{\alpha}_j \hat{\rho}_j}{\hat{\rho}_j - 1} + \lambda \right) dt + \sigma dW_t + d\left(\sum_{i=1}^{N_t} e^{\xi_i} - 1 \right),$$

where N_t and ξ_i are defined in (3.1), r is the risk-free interest rate and σ is the volatility of the market. Apparently, we have $\mathbb{E}^{\mathbb{Q}}[S_t] = xe^{rt}$. From now on, we suppress the superscript $\mathbb{Q}$ for the risk neutral expectation operator $\mathbb{E}^{\mathbb{Q}}$ since the dynamics we consider in this chapter are all under the risk neutral measure. The insurer charges fees from the investor's account. We assume that the rate at which the total fees are continuously deducted from the account to be m per dollar per year. Then the investor's account value process F_t should satisfy

$$\frac{dF_t}{F_{t-}} = \frac{dS_t}{S_{t-}} - mdt = \left(r - m - \sum_{j=1}^{J} \frac{\alpha_j \rho_j}{\rho_j - 1} - \sum_{j=1}^{\hat{J}} \frac{\hat{\alpha}_j \hat{\rho}_j}{\hat{\rho}_j - 1} + \lambda \right) t + \sigma dW_t + d\left(\sum_{i=1}^{N_t} e^{\xi_i} - 1 \right),$$

under the risk neutral measure. Equivalently, we have

$$F_t = S_t e^{-mt} = xe^{-mt+X_t} = xe^{X_t^{(m)}},$$

among which

$$X_t^{(m)} := (\mu - m)t + \sigma W_t + \sum_{i=1}^{N_t} \xi_i,$$

with

$$\mu = r - \frac{\sigma^2}{2} - \sum_{j=1}^{J} \frac{\alpha_j \rho_j}{\rho_j - 1} - \sum_{j=1}^{\hat{J}} \frac{\hat{\alpha}_j \hat{\rho}_j}{\hat{\rho}_j - 1} + \lambda.$$

$X_t^{(m)}$ is also a Lévy process with mixed-exponential jumps. Consider a GMWB rider which allows withdrawing at a constant rate w from the account. The remaining account value process U_t should satisfy the following SDE

$$dU_t = \left((\mu + \frac{\sigma^2}{2} - m) U_{t-} - w \right) dt + \sigma U_{t-} dW_t + U_{t-} d \left(\sum_{i=1}^{N_t} e^{\zeta_i} - 1 \right).$$

Using Ito's rule, the solution to the SDE above can be explicitly identified as

$$U_t = F_t - w \int_0^t \frac{F_t}{F_s} ds = x e^{X_t^{(m)}} - w e^{X_t^{(m)}} \int_0^t e^{-X_{s-}^{(m)}} ds. \tag{3.4}$$

The process U_t here is the generalized Ornstein-Uhlenbeck process we introduced in Chapter 1. We are interested in the first time τ that the process U_t hits zero, which is also called the time of ruin

$$\tau := \inf\{t > 0 | U_t < 0\}.$$

It is easy to see that τ is the same as the hitting time of level $\frac{x}{w}$ for the exponential functional of $-X_t^{(m)}$, i.e

$$\tau = h_{\frac{x}{w}}^{(m)} := \inf \left\{ t > 0 : \int_0^t e^{-X_s^{(m)}} ds > \frac{x}{w} \right\}.$$

We can see from the above equations that the time of ruin of the generalized Ornstein-Uhlenbeck process is related to the exponential functional of Lévy processes. In the next section, we can see that some of the quantities of our interests in pricing the GMWB can be determined by the distribution of the exponential functional. Similar to the last section, the Laplace exponent $\psi_m(z) := \ln \mathbb{E}[e^{zX_1^{(m)}}]$ of $X_t^{(m)}$ under the risk neutral measure is given by

$$\psi_m(z) = \frac{\sigma^2}{2} z^2 + (\mu - m) z + \left(\sum_{i=1}^{J} \frac{\alpha_i \rho_i}{\rho_i - z} + \sum_{i=1}^{\hat{J}} \frac{\hat{\alpha}_i \hat{\rho}_i}{\hat{\rho}_i + z} \right) - \lambda,$$

for $z \in (-\hat{\rho}_1, \rho_1)$. Consider the exponential functional of process $X_t^{(m)}$:

$$I_{m,q}^{\pm} = \int_0^{e_q} e^{\pm X_t^{(m)}} dt, \tag{3.5}$$

where e_q is an independent exponential random variable with mean $1/q$. From Theorem 3.1.1 and Corollary 3.1.2, its distribution can be characterized by the Meijer G-function. In the rest of this section, we use $\mathbb{E}_x$ to denote the expectation operator corresponding to the Markov process U_t starting from x. In other situations, we use the notation $\mathbb{E}$ as the expectation operator. First of all, we have the following result regarding the expectation of U_t.

Lemma 3.2.1. *The expectation of U_t is given by*

$$\mathbb{E}_x[U_t] = xe^{\psi_m(1)t} + w\frac{1 - e^{\psi_m(1)t}}{\psi_m(1)}. \tag{3.6}$$

For $q > \psi_m(1)$, the expectation $\mathbb{E}_x(U_{e_q})$ exists and is given by

$$\mathbb{E}_x[U_{e_q}] = \frac{xq - w}{q - \psi_m(1)}, \tag{3.7}$$

where e_q is an independent exponential random variable (of U_t).

Proof. By the duality of Lévy processes ([14, Lemma 2.3]), we have

$$(e^{X^{(m)}} \int_0^t e^{-X_s^{(m)}} ds, X_t^{(m)}) \stackrel{d}{=} (\int_0^t e^{X_s^{(m)}} ds, X_t^{(m)}),$$

which implies

$$U_t = xe^{X_t^{(m)}} - we^{X_t^{(m)}} \int_0^t e^{-X_s^{(m)}} ds \stackrel{d}{=} xe^{X_t^{(m)}} - w \int_0^t e^{X_s^{(m)}} ds.$$

Therefore, we have

$$\begin{aligned}
\mathbb{E}_x[U_t] =& x\mathbb{E}e^{X_t^{(m)}} - w \int_0^t \mathbb{E}e^{X_s^{(m)}} ds \\
=& xe^{\psi_m(1)t} - w \int_0^t e^{\psi_m(1)s} ds \\
=& xe^{\psi_m(1)t} + w\frac{1 - e^{\psi_m(1)t}}{\psi_m(1)}.
\end{aligned}$$

We observe from the last equality that $\mathbb{E}_x[U_t]$ diverges exponentially at the rate $\psi_m(1)$ (or converges at the rate $-\psi_m(1)$ if $\psi_m(1) < 0$) . Therefore, its Laplace transform $\int_0^\infty \mathbb{E}_x[U_t] e^{-qt} dt$ exists when $q > \psi_m(1)$. Then the other statement in this lemma follows readily from straightforward calculation. □

From now on, we focus on the pricing a GMWB rider. Similar to Feng and Volkmer ([35]), this can be done from two perspectives: pricing from a policyholder's perspective and pricing from an insurer's perspective. These two approaches are equivalent under certain conditions, which shall be explained later. We propose a semi-analytical algorithm to price the GMWB from both perspectives.

Policyholder's perspective

Suppose the initial investment of a policyholder is x and the annualized withdrawal rate is w. Then $t = x/w$ is the maturity of the GMWB rider. From the policyholder's view, his income comes from two parts. The first part is due to the continuous withdrawal from time 0 to t, which is referred as a generic term-certain annuity in [51]. The present value of this part is straightforward to calculate:

$$w\int_0^t e^{-rs}ds = \frac{w}{r}(1-e^{-rt}).$$

The second part comes from the policyholder's account, if the market is good and the policyholder's investment account is not exhausted due to withdrawal at maturity, the policyholder is entitled to the remaining balance. Therefore, the present value of what the policyholder receives at maturity is

$$e^{-rt}\max\{U_t, 0\}.$$

Since the policyholder's initial deposit is x, the non-arbitrage net present value of his investment in the variable annuity contract with a GMWB rider is

$$e^{-rt}\mathbb{E}_x[\max\{U_t,0\}] + \frac{w}{r}(1-e^{-rt}) - x. \tag{3.8}$$

To compute the first term, we have

$$\begin{aligned}
&\mathbb{E}_x\left[\max\{U_t,0\}\right] \\
=&\mathbb{E}_x[U_t\mathbb{1}_{\{t<\tau\}}] \\
=&\mathbb{E}_x[U_t] - \mathbb{E}_x[U_t\mathbb{1}_{\{t>\tau\}}] \\
=&\left(x-\frac{w}{\psi_m(1)}\right)e^{\psi_m(1)t} + \frac{w}{\psi_m(1)} - \mathbb{E}_x\left[\mathbb{E}[U_t\mathbb{1}_{\{\tau<t\}}|\mathcal{F}_\tau]\right] && \text{(Lemma 3.2.1)} \\
=&\left(x-\frac{w}{\psi_m(1)}\right)e^{\psi_m(1)t} + \frac{w}{\psi_m(1)} - \mathbb{E}_x\left[\mathbb{1}_{\{\tau<t\}}\mathbb{E}_0[U_{t-\tau}]\right] && \text{(strong Markov property)} \qquad (3.9)
\end{aligned}$$

So we only need to compute the last term in the previous equality, i.e

$$a(t,x) := \mathbb{E}_x[U_t \mathbb{1}_{\{t>\tau\}}] = \mathbb{E}_x\left[\mathbb{1}_{\{\tau<t\}}\mathbb{E}_0[U_{t-\tau}]\right].$$

Direct computing of $a(t,x)$ can be difficult: it might be impossible to get a neat analytical representation of $a(t,x)$. We saw in [35] that $a(t,x)$ has a long and complicated integration formula under the geometric Brownian motion model. Since the exponential Lévy process model is more complex, the expression of $a(t,x)$ under the exponential Lévy process model could be more complicated. Therefore, it might not be optimal to directly compute $a(t,x)$. We found that, however, the Laplace transform of $a(t,x)$ (with respect to t) can be analytically identified. We have the following theorem:

Theorem 3.2.2. *The Laplace transform $\hat{a}(q,x) := \int_0^\infty a(t,x)e^{-qt}dt$ exists for any $q > \psi_m(1)$. Under this condition,*

$$\hat{a}(q,x) = -\frac{w}{q(q-\psi_m(1))}\mathbb{P}(I^-_{m,q} > \frac{x}{w}), \tag{3.10}$$

where $I^-_{m,q}$ is the exponential functional defined in (3.5)*. By Corollary 3.1.2, the tail probability in* (3.10) *is given by* (3.3) *in terms of the Meijer G-function.*

Proof. From (3.6), we have

$$\mathbb{E}_0[U_t] = \frac{w}{\psi_m(1)}(1-e^{\psi_m(1)t}),$$

so

$$a(t,x) = \frac{w}{\psi_m(1)}\mathbb{E}_x\left[\mathbb{1}_{\{\tau<t\}}(1-e^{\psi_m(1)(t-\tau)})\right].$$

Therefore, we have

$$\begin{aligned}
&\hat{a}(q,x)\\
=&\int_0^\infty a(t,x)e^{-qt}dt\\
=&\frac{w}{\psi_m(1)}\int_0^\infty e^{-qt}dt\int_0^t(1-e^{\psi_m(1)(t-s)})\mathbb{P}_x(\tau\in ds)\\
=&\frac{w}{\psi_m(1)}\int_0^\infty \mathbb{P}_x(\tau\in ds)\int_s^\infty(e^{-qt}-e^{-\psi_m(1)s}e^{-(q-\psi_m(1))t})dt \qquad \text{(Fubini's Theorem)}\\
=&\frac{w}{\psi_m(1)}\int_0^\infty \mathbb{P}_x(\tau\in ds)(\frac{e^{-qs}}{q}-\frac{e^{-qs}}{q-\psi_m(1)})\\
=&-\frac{w}{q(q-\psi_m(1))}\int_0^\infty e^{-qs}\mathbb{P}_x(\tau\in ds)\\
=&-\frac{w}{q(q-\psi_m(1))}\mathbb{P}_x(\tau<e_q)\\
=&-\frac{w}{q(q-\psi_m(1))}\mathbb{P}(I^-_{m,q}>\frac{x}{w}),
\end{aligned}$$

which completes the proof. Alternatively, we have a simpler probabilistic proof as follows:

$$\begin{aligned}
\hat{a}(q,x)=&\frac{1}{q}\mathbb{E}_x[U_{e_q}\mathbb{1}_{\{\tau<e_q\}}]\\
=&\frac{1}{q}\mathbb{E}_x\left[\mathbb{E}_x[\mathbb{1}_{\{\tau<e_q\}}U_{e_q}|\mathcal{F}_\tau]\right]\\
=&\frac{1}{q}\mathbb{E}_x\left[\mathbb{1}_{\{\tau<e_q\}}\mathbb{E}_x[U_{e_q}|\mathcal{F}_\tau]\right]\\
=&\frac{1}{q}\mathbb{P}_x(\tau<e_q)\mathbb{E}_0[U_{e_q}] \qquad (3.11)\\
=&-\frac{w}{q(q-\psi_m(1))}\mathbb{P}(I^-_{m,q}>\frac{x}{w}), \qquad \text{(use (3.7))}
\end{aligned}$$

where e_q here is an independent (of U_t) exponential random variable. In (3.11) we used the strong Markov property of U_t and the lack of memory property of e_q. □

From Theorem 3.2.2, we identifies the Laplace transform of $a(t,x)$, which completely determines $a(t,x)$. To get the original value of $a(t,x)$, numerical Laplace inversion algorithms can be used. There are a number of such algorithms, such as Euler's method, the Gaver-Stehfest algorithm, and Talbot's method. In the last section of this chapter, we employ the Euler algorithm which is proven to be fast and accurate. After computing $\mathbb{E}_x[\max\{U_t,0\}]$, the GMWB's net present value from a policyholder's perspective in (3.8) can be easily determined.

Insurer's perspective

From an insurer's point of view, the net present value of a GMWB can be decomposed into the income and the liability. The insurer's income comes from continuous deduction of management fees from the investment fund. Its net present value is hence given by

$$m_w \int_0^{\tau\wedge t} e^{-rs} U_s ds,$$

where m_w is the portion among the total fees that the insurer uses to cover the liability of the GMWB rider. In reality, $m_w < m$ because the insurance company need to use part of the total fees to cover other expenses. On the other hand, the insurer's liability is due to the guaranteed policyholder's withdrawals after the investment account is depleted. The out-of-pocket cost for the insurer is given by

$$w \mathbb{1}_{\{\tau<t\}} \int_\tau^t e^{-rs} ds.$$

Therefore, the non-arbitrage cost of the GMWB rider from the insurer's perspective is given by

$$m_w \mathbb{E}_x \left[\int_0^{\tau\wedge t} e^{-rs} U_s ds \right] - w \mathbb{E}_x \left[\mathbb{1}_{\{\tau<t\}} \int_\tau^t e^{-rs} ds \right]. \tag{3.12}$$

We first focus on the income part. We have

$$\mathbb{E}_x \left[\int_0^{\tau\wedge t} e^{-rs} U_s ds \right] = \mathbb{E}_x \left[\int_0^{t} e^{-rs} U_s ds \right] - \mathbb{E}_x \left[\mathbb{1}_{\{\tau<t\}} \int_\tau^t e^{-rs} U_s ds \right], \tag{3.13}$$

where the first term on the right-hand side of the previous equality can be calculated by

$$\begin{aligned} \mathbb{E}_x \left[\int_0^{t} e^{-rs} U_s ds \right] &= \int_0^t \left((x - \frac{w}{\psi_m(1)}) e^{(\psi_m(1)-r)s} + \frac{w e^{-rs}}{\psi_m(1)} \right) ds \qquad \text{(use (3.6))} \\ &= \frac{x\psi_m(1) - w}{\psi_m(1)(\psi_m(1)-r)} (e^{(\psi_m(1)-r)t} - 1) + \frac{w(1-e^{-rt})}{\psi_m(1) r} \\ &= \frac{x\psi_m(1) - w}{\psi_m(1)(\psi_m(1)-r)} e^{(\psi_m(1)-r)t} - \frac{w}{\psi_m(1) r} e^{-rt} + \frac{w - rx}{r(\psi_m(1)-r)}, \end{aligned}$$

and the second term is yet to be determined. Similarly, we try to identify the second term on the right-hand side of (3.13). We found that its Laplace transform can also be characterized by the exponential functional. To be precise, we have the following theorem:

Theorem 3.2.3. *Let* $b(t,x) := \mathbb{E}_x \left[\mathbb{1}_{\{\tau<t\}} \int_\tau^t e^{-rs} U_s ds \right]$. *Its Laplace transform* $\hat{b}(q,x) := \int_0^\infty e^{-qs} b(s,x) ds$ *exists when* $q > \psi_m(1) - r$. *Moreover, we have*

$$\hat{b}(q,x) = \left(-\frac{w}{rq(r-\psi_m(1))} - \frac{w}{r(q+r)\psi_m(1)} + \frac{w}{\psi_m(1)(r-\psi_m(1))(q-\psi_m(1)+r)}\right)\mathbb{P}\left(I^-_{m,r+q} > \frac{x}{w}\right), \tag{3.14}$$

where $I^-_{m,q}$ *is the exponential functional defined in* (3.5).

Proof. Using the strong Markov property of U_t and Lemma 3.2.1, we have

$$\begin{aligned}
&b(t,x)\\
=&\mathbb{E}_x\left[\mathbb{1}_{\{\tau<t\}}\int_\tau^t e^{-rs}U_s ds\right]\\
=&\mathbb{E}_x\left[\mathbb{1}_{\{\tau<t\}}\mathbb{E}_x\left[\int_\tau^t e^{-rs}U_s ds|\mathcal{F}_\tau\right]\right]\\
=&\mathbb{E}_x\left[\mathbb{1}_{\{\tau<t\}}e^{-r\tau}\mathbb{E}_0\left[\int_0^{t-\tau} e^{-rs}U_s ds\right]\right]\\
=&\mathbb{E}_x\left[\mathbb{1}_{\{\tau<t\}}e^{-r\tau}(-\frac{w}{r(r-\psi_m(1))} - \frac{we^{-r(t-\tau)}}{r\psi_m(1)} + \frac{we^{-(r-\psi_m(1))(t-\tau)}}{\psi_m(1)(r-\psi_m(1))})\right]\\
=&-\frac{w}{r(r-\psi_m(1))}\mathbb{E}_x[\mathbb{1}_{\{\tau<t\}}e^{-r\tau}] - \frac{we^{-rt}}{r\psi_m(1)}\mathbb{P}_x(\tau<t) + \frac{we^{-(r-\psi_m(1))t}}{\psi_m(1)(r-\psi_m(1))}\mathbb{E}_x[\mathbb{1}_{\{\tau<t\}}e^{-\psi(1)\tau}].
\end{aligned} \tag{3.15}$$

Then for $q > -(r-\psi_m(1))$, we take the Laplace transform of each of the three terms in the last equality,

$$\begin{aligned}
&\int_0^\infty e^{-qt}\mathbb{E}_x[\mathbb{1}_{\{\tau<t\}}e^{-r\tau}]dt\\
=&\int_0^\infty e^{-qt}dt\int_0^t e^{-rs}\mathbb{P}(\tau\in ds)\\
=&\int_0^\infty e^{-rs}\mathbb{P}_x(\tau\in ds)\int_s^\infty e^{-qt}dt \qquad \text{(Fubini's Theorem)}\\
=&\frac{1}{q}\int_0^\infty e^{-(r+q)s}\mathbb{P}_x(\tau\in ds)\\
=&\frac{1}{q}\mathbb{P}_x(\tau<e_{q+r})\\
=&\frac{1}{q}\mathbb{P}\left(I^-_{m,q+r} > \frac{x}{w}\right).
\end{aligned} \tag{3.16}$$

Similarly, we can show that

$$\int_0^\infty e^{-qt}e^{-rt}\mathbb{P}_x(\tau<t)dt = \frac{1}{q+r}\mathbb{P}\left(I^-_{m,q+r} > \frac{x}{w}\right),$$

and

$$\int_0^\infty e^{-qt}e^{-(r-\psi_m(1))t}\mathbb{E}_x(\mathbb{1}_{(\tau<t)}e^{-\psi_m(1)\tau})dt = \frac{1}{q-\psi_m(1)+r}\mathbb{P}_x\left(I^-_{m,q+r} > \frac{x}{w}\right).$$

Collect these three terms will give us (3.14). Similar to Theorem 3.2.2, we have a probabilistic proof for (3.14):

$$\begin{aligned}
\hat{b}(q,x) =& \frac{1}{q}\mathbb{E}_x\left[\mathbb{1}_{\{\tau<e_q\}}\int_\tau^{e_q} e^{-rs}U_s ds\right]\\
=& \frac{1}{q}\mathbb{E}_x\left[\mathbb{1}_{\{\tau<e_q\}}e^{-r\tau}\mathbb{E}\left[\int_\tau^{e_q} e^{-r(s-\tau)}U_s ds|\mathcal{F}_\tau\right]\right]\\
=& \frac{1}{q}\mathbb{E}_x[\mathbb{1}_{\{\tau<e_q\}}e^{-r\tau}]\mathbb{E}_0[\int_0^{e_q} e^{-rs}U_s ds] \qquad (3.17)\\
=& \frac{1}{q}\mathbb{E}_x[\mathbb{1}_{\{\tau<e_q\}}e^{-r\tau}]\left(-\frac{w}{r(r-\psi_m(1))} - \frac{wq}{r(q+r)\psi_m(1)} + \frac{wq}{\psi_m(1)(r-\psi_m(1))(q-\psi_m(1)+r)}\right)\\
=& \mathbb{P}(I^-_{m,r+q} > \frac{x}{w})\left(-\frac{w}{rq(r-\psi_m(1))} - \frac{w}{r(q+r)\psi_m(1)} + \frac{w}{\psi_m(1)(r-\psi_m(1))(q-\psi_m(1)+r)}\right).
\end{aligned}$$

In (3.17), we used the strong Markov property of U_t, the lack of memory property of e_q and Lemma 3.2.1. □

Next, we consider the non-arbitrage cost of the insurer's liability. Let

$$c(t,x) := \mathbb{E}_x\left[\mathbb{1}_{\{\tau<t\}}\int_\tau^t e^{-rs}ds\right].$$

Theorem 3.2.4. *The Laplace transform $\hat{c}(q,x) := \int_0^\infty e^{-qt}c(t,x)dt$ exists for $q>0$ and it is given by*

$$\hat{c}(q,x) = \frac{1}{q(q+r)}\mathbb{P}(I^-_{m,q+r} > \frac{x}{w}), \qquad (3.18)$$

where $I^-_{m,q}$ is the exponential functional defined in (3.5).

Proof. Similar to the proof of Theorem 3.2.3, we have

$$\begin{aligned}
c(t,x) =& \mathbb{E}_x\left[\mathbb{1}_{\{\tau<t\}}\int_\tau^t e^{-rs}ds\right]\\
=& \frac{1}{r}\mathbb{E}_x\left[\mathbb{1}_{\{\tau<t\}}(e^{-r\tau}-e^{-rt})\right]\\
=& \frac{1}{r}\mathbb{E}_x[\mathbb{1}_{\{\tau<t\}}e^{-r\tau}] - \frac{1}{r}e^{-rt}\mathbb{P}_x(\tau<t), \qquad (3.19)
\end{aligned}$$

From the proof of Theorem 3.2.3, we know that for $q > 0$,

$$\int_0^\infty e^{-qt}\mathbb{E}_x[\mathbb{1}_{\{\tau<t\}}e^{-r\tau}] = \frac{1}{q}\mathbb{P}_x(I^-_{m,q+r} > \frac{x}{w}),$$

and

$$\int_0^\infty e^{-qt}e^{-rt}\mathbb{P}_x(\tau < t)dt = \frac{1}{q+r}\mathbb{P}_x(I^-_{m,q+r} > \frac{x}{w}),$$

which implies $\hat{c}(q,x) = \frac{1}{q(q+r)}\mathbb{P}(I^-_{m,q+r} > \frac{x}{w})$. Again, we provide a probabilistic proof for (3.18):

$$\begin{aligned}
\hat{c}(q,x) =& \frac{1}{q}\mathbb{E}_x\left[\mathbb{1}_{\{\tau<e_q\}}\int_\tau^{e_q} e^{-rs}ds\right] \\
=& \frac{1}{q}\mathbb{E}_x\left[\mathbb{1}_{\{\tau<e_q\}}e^{-r\tau}\mathbb{E}\left[\int_\tau^{e_q} e^{-r(s-\tau)}ds|\mathcal{F}_\tau\right]\right] \\
=& \frac{1}{q}\mathbb{E}_x[\mathbb{1}_{\{\tau<e_q\}}e^{-r\tau}]\mathbb{E}_0\left[\int_0^{e_q} e^{-rs}ds\right] \qquad (3.20) \\
=& \frac{1}{q(q+r)}\mathbb{E}_x[\mathbb{1}_{\{\tau<e_q\}}e^{-r\tau}] \\
=& \frac{1}{q(q+r)}\int_0^\infty qe^{-qt}\int_0^t e^{-rs}\mathbb{P}_x(\tau \in ds) \\
=& \frac{1}{q(q+r)}\mathbb{P}_x(\tau < e_{q+r}) \\
=& \frac{1}{q(q+r)}\mathbb{P}\left(I^-_{m,q+r} > \frac{x}{w}\right),
\end{aligned}$$

where in (3.20) the strong Markov property of U_t, the lack of memory property of e_q and Lemma 3.2.1 are used. □

From Theorem 3.2.3 and 3.2.4, we obtain the Laplace transforms of the net present value of a GMWB rider from an insurer's perspective. To get the numerical value of the net present value, we can use the numerical Laplace inversion algorithms mentioned previously. Moreover, we point out that all the quantities we encounter here are related to the tail probability $\mathbb{P}(I^-_{m,q+r} > x/w)$, which is given by the Meijer G-function. This makes our semi-analytical method for pricing the GMWB rider more efficient since we only need to compute this value for one time and store it. We illustrate our algorithm in detail with some examples in the last section of this chapter.

3.3 GMWB with the annual high step-up

In this section, we consider a more sophisticated type of rider which is derived from the traditional GMWB rider we considered in the previous section. The option embedded in the rider is called the ratchet option.

As we saw in the previous section, during the withdrawal period of a traditional GMWB rider, a policyholder is allowed to withdraw in proportion to a fixed guarantee base per year, which typically matches the initial account value. Nowadays, however, insurers in the market tend to offer more generous options in order to attract investors. The most common type of these options is the ratchet option, which is also known as the step-up option. The step-up option allows the originally fixed guarantee base of a traditional GMWB rider to possibly increase in accordance with the performance of the investment account. The amount of the guarantee base step-up is determined at the end of each contractual period (usually 1 year or 6 months) until the rider matures. There are two typical step-up options in the market: the lifetime high step-up and the annual high step-up. The former is based on the historical high of a policyholder's account value. If the current account value exceeds the guarantee base from the previous period, the guarantee base resets to the current account value for the next period. If the current account value does not exceed the previous period guarantee base, the guarantee base remains the same. The latter is based on the performance of a policyholder's account in the most recent period only. If the current account value exceeds its previous contractual period ending value, then the guarantee base increases by the same percentage amount as the account value increases. If the current account value does not exceed its previous period ending value, the guarantee base stays the same. Huang et al. ([40]) considered the case of the lifetime high step-up option for the Guaranteed Lifetime Withdrawal Benefit (GLWB). In this section, we are interested in the latter case. We can see that for a GMWB rider with the annual high set-up, we must keep track of two balances: the remaining value process U_t of the policyholder's account and the guarantee base process G_t. This makes a GMWB rider with the annual high step-up more complicated since for a traditional GMWB rider we only need to consider the remaining account value process U_t. Intuitively, the dynamics of an annual high step-up option can be described as

$$\frac{G_{i+1}}{G_i} = \max\left(1, \frac{U_{i+1}}{U_i}\right), \tag{3.21}$$

for $i = 0, 1, \dots$ until maturity. Figure 3.1 shows a simulation example of one sample path of the account value process U_t and the corresponding guarantee base process G_t. From Figure 3.1 we can see that the annual high step-up option is very generous since the guarantee base process G_t increases quickly. If the step-up event of the guarantee base process G_t is discretized, it may not be convenient to work with since usually continuous-time models are used for modeling variable annuity products. It is natural to consider providing a continuous-time step-up model for the GMWB rider with the annual high step-up. Therefore, our goal in this section is two-fold: On one hand, we want to extend the intuition (3.21) to a continuous-time step-up model to describe the dynamics of a GMWB rider with the annual high step-up. On the other

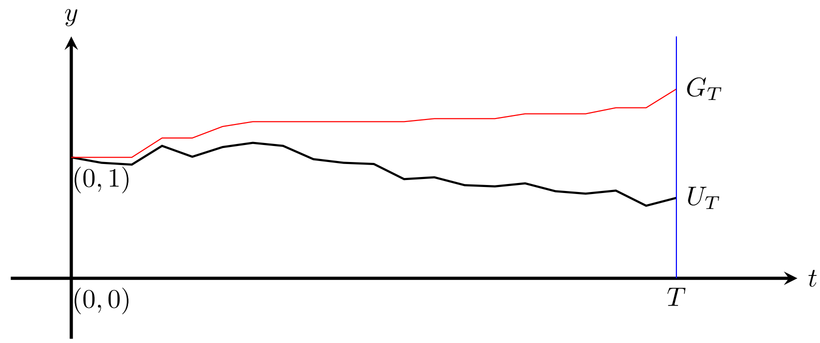

Figure 3.1: GMWB with the annual high step-up

hand, we want to use appropriate stochastic processes to model the dynamics so that we can solve the pricing problem of this more complex product. As we shall see, Lévy processes of bounded variation can be used for modeling the dynamics of the rider. Moreover, with the help of the distributional properties of the exponential functionals of Lévy processes (See Section 3.1), we have analytical tractability of the model under the assumption that the jump distribution of the Lévy processes is a mixture of exponential distributions. This results in an efficient semi-analytical pricing algorithm which solves the pricing problem of the GMWB rider with the annual high step-up.

3.3.1 Modeling the annual high step-up

First, we explain the intuition behind extending (3.21) to a continuous-time step-up model and the reason for considering Lévy processes of bounded variation. When the step-up period shrinks from 1 year to dt years, (3.21) becomes

$$\frac{G_{t+dt}}{G_t} = \max\left(1, \frac{U_{t+dt}}{U_t}\right),$$

which is equivalent to

$$\frac{G_{t+dt} - G_t}{G_t} = \max\left(0, \frac{U_{t+dt} - U_t}{U_t}\right). \tag{3.22}$$

The continuous-time version of (3.22) is given by

$$\frac{dG_t}{G_t} = \max\left(0, \frac{dU_t}{U_t}\right). \tag{3.23}$$

(3.23) intends to describe a continuous-time step-up dynamics. Intuitively, it shows that the return of the guarantee process G_t is the positive part of the return of the account value process U_t. Since the account value and the guarantee base is always positive, we can assume $U_t = U_0 e^{X_t}$ for a certain stochastic process

X_t and $G_t = G_0 e^{\overline{X}_t}$ for another stochastic process $\overline{X}_t$. At this time, we can assume $G_0 = U_0$. By (3.23), the relationship between $\overline{X}_t$ and X_t should be $d\overline{X}_t = \max(dX_t, 0)$, intuitively. This implies that: $\overline{X}_t$ should increase whenever X_t increases and the increment of $\overline{X}_t$ should be the same as X_t; when X_t decreases, $\overline{X}_t$ stays unchanged. From the definition of G_t, we know that it is an increasing process and therefore $\overline{X}_t$ is also an increasing process. It is not hard to see that $\overline{X}_t$, in fact, is the positive variation process of X_t:

$$\overline{X}_t = \sup_{\Pi} \sum_i (X_{t_{i+1}} - X_{t_i})_+,$$

where $\Pi = \{t_i : 0 = t_0 < t_1 \cdots < t_i < \cdots < t_n = t\}$ is an arbitrary partition of interval $[0, t]$. This indicates X_t should be a process of bounded variation, otherwise $\overline{X}_t$ is infinity. As a result, we have to model the dynamics of the underlying equity as a process of bounded variation if we want to provide a continuous-time model for the continuous-time step-up. It is not hard to imagine that if X_t is a Brownian motion, the continuous-time step-up option will be so generous that the guarantee base G_t will explode in any finite time. Since any continuous martingale is of unbounded variation except a constant, it is reasonable to consider X_t to be a jump process. Specifically, we want to model the dynamics of the underlying equity by a Lévy process of bounded variation. By the Lévy-Itô decomposition (Theorem 1.0.4) , we only have two possibilities:

$$\sigma = 0 \quad \text{and} \quad \pi(\mathbb{R}) < \infty,$$

and

$$\sigma = 0, \quad \pi(\mathbb{R}) = \infty \quad \text{and} \quad \int_{\mathbb{R}} (1 \wedge |x|)\pi(dx) < \infty,$$

where σ is the diffusion coefficient and $\pi(dx)$ is the Lévy measure of the Lévy process defined in the first section of this chapter. In this section, we focus on the first case where we model the dynamics of the underlying equity by a compound Poisson process. As we mentioned in the first section, the Lévy processes with mixed-exponential jumps have two advantages from a financial modeling perspective: it is general enough so that it can approximate any jump distribution; it provides analytical solutions to the pricing problem. Therefore, in this section, we consider modeling the dynamics of the underlying equity by a compound Poisson process with mixed-exponential jump distribution, which is a special case of the Lévy processes we introduced in the first section of this chapter with $\sigma = 0$. We shall see that the quantities involved in the pricing problem of the GMWB rider with the annual high step-up are also closely related to the exponential functionals of Lévy processes.

Assuming the underlying investment fund is driven by an exponential Lévy process under the physical measure

$$S_t = xe^{X_t},$$

where X_t is a compound Poisson process with mixed-exponential jumps. It is easier to write X_t into the form of the Poisson integral in Theorem 1.0.4, i.e

$$X_t = \delta t + \int_{[0,t]} \int_{\mathbb{R}} xN(dt \times dx),$$

where $N(\cdot)$ is a Poisson random measure on $[0,\infty) \times \mathbb{R}$ with intensity $dt \times \pi(dx)$. $\pi(dx)$ is the Lévy measure of X_t satisfying

$$\pi(dx) = \sum_{i=1}^{J} \alpha_i \rho_i e^{-\rho_i x} \mathbb{1}_{\{x>0\}} dx + \sum_{i=1}^{\hat{J}} \hat{\alpha}_i \hat{\rho}_i e^{\hat{\rho}_i x} \mathbb{1}_{\{x<0\}} dx,$$

where $0 < \rho_1 < \rho_2 < \cdots < \rho_J$, $0 < \hat{\rho}_1 < \hat{\rho}_2 < \cdots < \hat{\rho}_{\hat{J}}$ and $\alpha_i, \hat{\alpha}_j \in \mathbb{R}$ for $1 \leq i \leq J$, $1 \leq j \leq \hat{J}$. By the Lévy-Itô decomposition, we can rewrite X_t as the difference of two independent compound Poisson subordinators, i.e

$$\begin{aligned} X_t &= \delta t + \int_{[0,t]} \int_{(0,\infty)} xN(dt \times dx) - \int_{[0,t]} \int_{(-\infty,0)} (-x)N(dt \times dx) \\ &:= \delta t + X_t^+ - X_t^-. \end{aligned} \tag{3.24}$$

Let

$$\begin{aligned} Y_t^+ &:= \int_{[0,t]} \int_{(0,\infty)} (e^x - 1)N(dt \times dx), \\ Y_t^- &:= \int_{[0,t]} \int_{(-\infty,0)} (1 - e^x)N(dt \times dx), \end{aligned} \tag{3.25}$$

then we know Y_t^+ and Y_t^- are the compound Poisson subordinators corresponding to $e^{X_t^+}$ and $e^{X_t^-}$:

$$\frac{de^{X_t^+}}{e^{X_{t-}^+}} = dY_t^+, \qquad \frac{de^{X_t^-}}{e^{X_{t-}^-}} = dY_t^-. \tag{3.26}$$

Like pricing the traditional GMWB, we are interested in the dynamics of S_t under the risk neutral measure. we can choose a risk neutral measure such that the dynamics of S_t under the risk neutral measure is described by the following SDE

$$\frac{dS_t}{S_{t-}} = \left(r - \sum_{j=1}^{J} \frac{\alpha_j \rho_j}{\rho_j - 1} - \sum_{j=1}^{\hat{J}} \frac{\hat{\alpha}_j \hat{\rho}_j}{\hat{\rho}_j + 1} + \lambda \right) dt + dY_t^+ - dY_t^-,$$

where r is the risk-free interest rate and $\lambda = \sum_{j=1}^{J} \alpha_j + \sum_{j=1}^{\hat{J}} \hat{\alpha}_j$ is the intensity of the compound Poisson process. The investor's account value F_t under the risk neutral measure should satisfy

$$\frac{dF_t}{F_{t-}} = \frac{dS_t}{S_{t-}} - mdt = \left(r - m - \sum_{j=1}^{J} \frac{\alpha_j \rho_j}{\rho_j - 1} - \sum_{j=1}^{\hat{J}} \frac{\hat{\alpha}_j \hat{\rho}_j}{\hat{\rho}_j + 1} + \lambda \right) dt + dY_t^+ - dY_t^-,$$

since the insurer charges fees at the rate m from the investor's account. Let

$$\mu = r - \sum_{j=1}^{J} \frac{\alpha_j \rho_j}{\rho_j - 1} - \sum_{j=1}^{\hat{J}} \frac{\hat{\alpha}_j \hat{\rho}_j}{\hat{\rho}_j + 1} + \lambda,$$

be the drift of the underlying investment fund under the risk neutral measure. Consider a GMWB rider with the annual high step-up. A policyholder takes withdrawals at a constant rate w per year based on the guarantee base. To fund the step-up option, the insurer charges the investor's account a rate m_r of the guarantee base. Therefore, the remaining value process of the investor's account U_t and the guarantee base process G_t should satisfy the following system of SDEs:

$$\begin{aligned} \frac{dU_t}{U_{t-}} &= \left((\mu - m) - (w + m_r) \frac{G_{t-}}{U_{t-}} \right) dt + dY_t^+ - dY_t^-, \\ \frac{dG_t}{G_{t-}} &= \left((\mu - m) - (w + m_r) \frac{G_{t-}}{U_{t-}} \right)_+ dt + dY_t^+. \end{aligned} \tag{3.27}$$

Note that the previous system characterizes the dynamics of the account value and the guarantee base before the time of ruin:

$$\tau := \inf\{t > 0 | U_t < 0\}.$$

When $t > \tau$, U_t is absorbed at 0 and G_t is fixed at G_τ. This system, in general, can be not easy to deal with. But under certain special conditions, we can explicitly solve the system. We point out here that the process G_t/U_t is an increasing process which starts at 1. In the market, we usually observe that the withdrawal rate w is $4-8\%$ and the rider charge m_r is around 1%. In view of the fact that μ is close to the risk free interest rate, we know that in many situations, the drift coefficient of the SDE of G_t is already 0 at the inception of the contract. In that case, (3.27) can be simplified to

$$\frac{dU_t}{U_{t-}} = \left((\mu - m) - (w + m_r)\frac{G_{t-}}{U_{t-}}\right) dt + dY_t^+ - dY_t^-,$$
$$\frac{dG_t}{G_{t-}} = dY_t^+. \tag{3.28}$$

The second equation in (3.28) implies that $G_t = xe^{X_t^+}$. Using Ito's rule for semimartingales, the solution of the SDE of U_t can be explicitly identified as

$$U_t = xe^{X_t^+}\left(e^{(\mu-m)t-X_t^-} - (w + m_r)e^{(\mu-m)t-X_t^-}\int_0^t e^{-(\mu-m)s+X_s^-} ds\right). \tag{3.29}$$

The process U_t is very similar to the generalized Ornstein-Uhlenbeck process we encountered in (3.4). If we define

$$U_t^- = e^{(\mu-m)t-X_t^-} - (w + m_r)e^{(\mu-m)t-X_t^-}\int_0^t e^{-(\mu-m)s+X_s^-} ds, \tag{3.30}$$

U_t^- is a generalized Ornstein-Uhlenbeck process driven by $(\mu - m)t - X_t^-$. Since U_t^- is a Markov process which starts at 1, we use $\mathbb{E}_x$ to denote the expectation operator corresponding to U_t^- starts from x. In other situations, we use the notation $\mathbb{E}$ as the expectation operator. Due to the independence of X_t^+ and X_t^-, we observe that τ has the same distribution as the hitting time of the exponential functional of $(\mu-m)t+X_t^-$, i.e

$$\tau = h^{(m)}_{\frac{1}{w+m_r}} := \inf\left\{t > 0 : \int_0^t e^{-(\mu-m)_s+X_s^-} ds > \frac{1}{w + m_r}\right\}.$$

It is easy to see that some of the quantities of our interest in pricing the GMWB rider with the annual high step-up are also related to the distribution of the exponential functionals. We introduce the Laplace exponent of X_t^+ and $(\mu - m)t - X_t^-$ under the risk neutral measure $\psi^+(z) := \ln \mathbb{E}[e^{zX_1^+}]$ and $\psi_m^-(z) := \ln \mathbb{E}[e^{z(\mu-m)-zX_1^-}]$, which are given by

$$\psi^+(z) = \sum_{i=1}^{J} \frac{\alpha_i \rho_i}{\rho_i - z} - \sum_{i=1}^{J} \alpha_j,$$
$$\psi_m^-(z) = (\mu - m)z + \sum_{i=1}^{\hat{J}} \frac{\hat{\alpha}_i \hat{\rho}_i}{\hat{\rho}_i + z} - \sum_{i=1}^{\hat{J}} \hat{\alpha}_j,$$

for $z \in (-\infty, \rho_1)$ and $z \in (-\hat{\rho_1}, \infty)$, respectively. Consider the exponential functional of the process $(\mu - m)t - X_t^-$,

$$I_{m,q}^- = \int_0^{e_q} e^{-(\mu-m)t+X_t^-} dt, \tag{3.31}$$

where e_q is an independent exponential random variable with mean $1/q$. From Theorem 3.1.1 and Corollary 3.1.2, its distribution can be characterized by the Meijer G-function. Using the independence of X_t^+ and U_t^- and Lemma 3.2.1, it is straightforward to see that

$$\mathbb{E}[U_t] = xe^{\psi^+(1)t+\psi_m^-(1)t} + x(w+m_r)e^{\psi^+(1)t}\frac{1-e^{\psi_m^-(1)t}}{\psi_m^-(1)}. \tag{3.32}$$

3.3.2 Pricing the GMWB with the annual high step-up

Next, we consider the pricing problem of the GMWB rider with the annual high step-up. The framework is similar to the pricing of the traditional GMWB we considered in the previous section: we consider the non-arbitrage value of the GMWB rider with the annual high step-up from a policyholder's perspective and from an insurer's perspective. We obtain the Laplace transforms of the quantities related to pricing analytically. Again, we shall see many of the quantities are closely related to the exponential functionals of Lévy processes.

Policyholder's perspective

We start by considering a policyholder's income. Assume the maturity of the rider is t years. If the investment fund performs so well that it is not exhausted at the maturity, the first part of the policyholder's income is generated by the outstanding balance of the account at the maturity. Therefore, the present value of what the policyholder receives at the maturity is given by

$$e^{-rt}\max\{U_t, 0\}.$$

The second part of the policyholder's income is due to the continuous withdrawals from time 0 to t. In contrast to the traditional GMWB rider case, the withdrawal part should be further divided into two parts: the withdrawals before ruin and the withdrawals after ruin (if the policyholder's account value becomes 0 before the maturity). The present value of the total withdrawals is given by

$$w\int_0^{t\wedge\tau} e^{-rs}G_s ds + w\mathbb{1}_{\{\tau<t\}}G_\tau\int_\tau^t e^{-rs}ds.$$

Since the policyholder's initial payment is x, the non-arbitrage net present value of the GMWB rider with the annual high step-up is given by

$$e^{-rt}\mathbb{E}\left[\max\{U_t, 0\}\right] + w\mathbb{E}\left[\int_0^{t\wedge\tau} e^{-rs}G_s ds\right] + w\mathbb{E}\left[\mathbb{1}_{\{\tau<t\}}G_\tau\int_\tau^t e^{-rs}ds\right] - x, \tag{3.33}$$

from the policyholder's perspective. The expectation operator $\mathbb{E}$ is under the risk neutral measure. The computation for the first term is similar to that of (3.9): we have

$$\begin{aligned}
&\mathbb{E}\left[\max\{U_t,0\}\right]\\
=&\mathbb{E}[U_t\mathbb{1}_{\{t<\tau\}}]\\
=&\mathbb{E}[U_t]-\mathbb{E}_x[U_t\mathbb{1}_{\{t>\tau\}}]\\
=&xe^{\psi^+(1)t+\psi_m^-(1)t}+x(w+m_r)e^{\psi^+(1)t}\frac{1-e^{\psi_m^-(1)t}}{\psi_m^-(1)}-xe^{\psi^+(1)t}\mathbb{E}_1[U_t^-\mathbb{1}_{\{\tau<t\}}].
\end{aligned}$$

Therefore, we need to determine the last term in the previous equality, i.e

$$d(t):=\mathbb{E}_1[U_t^-\mathbb{1}_{\{t>\tau\}}].$$

Recalling Theorem 3.2.2, we already identified the Laplace transform of $d(t)$ since U_t^- is also a generalized Ornstein-Uhlenbeck process driven by a Lévy process with mixed-exponential jumps. We state the result in the following theorem without proof.

Theorem 3.3.1. *The Laplace transform* $\hat{d}(q):=\int_0^\infty d(t)e^{-qt}dt$ *exists for any* $q>\psi_m^-(1)$. *Under this condition, we have*

$$\hat{d}(q)=-\frac{w+m_r}{q(q-\psi_m^-(1))}\mathbb{P}(I_{m,q}^->\frac{1}{w+m_r}),\tag{3.34}$$

where $I_{m,q}^-$ *is the exponential functional defined in* (3.31). *By Corollary 3.1.2, the tail probability in* (3.34) *is given by* (3.3) *in terms of the Meijer G-function.*

As mentioned in the previous section, numerical Laplace inversion algorithms such as Euler's method can be used to determine the numerical value of $d(t)$. We shall see examples of the Laplace inversion in the next subsection. To continue pricing the rider from the policyholder's perspective, we consider the second term in (3.33). It is easy to derive that

$$\begin{aligned}
&\mathbb{E}\left[\int_0^{t\wedge\tau}e^{-rs}G_sds\right]\\
=&\mathbb{E}\left[\int_0^{t}e^{-rs}G_sds\right]-\mathbb{E}\left[\mathbb{1}_{\{\tau<t\}}\int_\tau^{t}e^{-rs}G_sds\right]\\
=&x\int_0^{t}e^{-(r-\psi^+(1))s}ds-x\mathbb{E}\left[\mathbb{1}_{\{\tau<t\}}\int_\tau^{t}e^{-(r-\psi^+(1))s}ds\right]\\
=&x\frac{1-e^{-(r-\psi^+(1))t}}{r-\psi^+(1)}-x\mathbb{E}\left[\mathbb{1}_{\{\tau<t\}}\int_\tau^{t}e^{-(r-\psi^+(1))s}ds\right],
\end{aligned}$$

so we need to compute the last term in the previous equality: $e(t) := \mathbb{E}\left[\mathbb{1}_{\{\tau<t\}}\int_\tau^t e^{-(r-\psi^+(1))s}ds\right]$. The Laplace transform of $e(t)$ has already been identified in Theorem 3.2.4. For our current situation, we only need to change the parameter r to $r-\psi^+(1)$ in Theorem 3.2.4. Therefore, we have the following theorem.

Theorem 3.3.2. *The Laplace transform $\hat{e}(q) := \int_0^\infty e^{-qt}e(t)dt$ exists for $q > \psi^+(1) - r$, it is given by*

$$\hat{e}(q) = \frac{1}{q(q+r-\psi^+(1))}\mathbb{P}(I^-_{m,q+r-\psi^+(1)} > \frac{1}{w+m_r}) \tag{3.35}$$

where $I^-_{m,q}$ is the exponential functional defined in (3.31).

For the third term in (3.33), we also consider its Laplace transform. Let $f(t,x) := \mathbb{E}[\mathbb{1}_{\{\tau<t\}}G_\tau\int_\tau^t e^{-rs}ds]$, then we have the following theorem:

Theorem 3.3.3. *The Laplace transform $\hat{f}(q,x) := \int_0^\infty e^{-qt}f(t,x)dt$ exists for $q > \psi^+(1) - r$. It is given by*

$$\hat{f}(q,x) = \frac{x}{q(q+r)}\mathbb{P}(I^-_{m,q+r-\psi^+(1)} > \frac{1}{w+m_r}), \tag{3.36}$$

where $I^-_{m,q}$ is the exponential functional defined in (3.31).

Proof. By the definition of $f(t,x)$, we have

$$\begin{aligned}
f(t,x) &= \mathbb{E}\left[\mathbb{1}_{\{\tau<t\}}G_\tau\int_\tau^t e^{-rs}ds\right] \\
&= x\mathbb{E}\left[\mathbb{1}_{\{\tau<t\}}e^{X^+_\tau}\frac{e^{-r\tau}-e^{-rt}}{r}\right] \\
&= \frac{x}{r}\int_0^t (e^{-rs}-e^{-rt})\mathbb{E}[e^{X^+_s}]\mathbb{P}(\tau\in ds) \\
&= \frac{x}{r}\int_0^t (e^{-rs}-e^{-rt})e^{\psi^+(1)s}\mathbb{P}(\tau\in ds).
\end{aligned}$$

The Laplace transform of $f(t,x)$ is given by

$$\begin{aligned}
\hat{f}(q,x) &= \frac{x}{r}\int_0^\infty e^{-qt}dt\int_0^t (e^{-rs}-e^{-rt})e^{\psi^+(1)s}\mathbb{P}(\tau\in ds) \\
&= \frac{x}{r}\int_0^\infty e^{\psi^+(1)s}\mathbb{P}(\tau\in ds)\int_s^\infty e^{-qt}(e^{-rs}-e^{-rt})dt \\
&= \frac{x}{q(q+r)}\int_0^\infty e^{-(q+r-\psi^+(1))s}\mathbb{P}(\tau\in ds) \\
&= \frac{x}{q(q+r)}\mathbb{P}(I^-_{m,q+r-\psi^+(1)} > \frac{1}{w+m_r}),
\end{aligned}$$

which completes the proof. ∎

Since we have determined the Laplace transform of every term in (3.33), we can compute the non-arbitrage value of the GMWB rider with the annual high step-up to a policyholder using numerical Laplace inversion methods. Therefore, the pricing problem of the rider is solved from a policyholder's perspective and we can consider pricing the rider from an insurer's perspective.

Insurer's perspective

From an insurer's point of view, the net present value of the GMWB rider with the annual high step-up can be decomposed into the income from the fees and the liability from the guaranteed payments. On the liability side, the insurer is responsible for the policyholder's guaranteed payments after the investment account is depleted, which is given by

$$w\mathbb{1}_{\{\tau<t\}}G_\tau \int_\tau^t e^{-rs}ds.$$

We have already encountered this quantity when we considered pricing the rider from a policyholder's perspective. The risk neutral value of it is given by $f(t,x)$ and its Laplace transform is given by Theorem 3.3.3. On the income side, the insurer collects the management fees, which is asset-value-based, and the rider charges, which is guarantee-based. The total net present value of these two parts is given by

$$m_w \int_0^{\tau\wedge t} e^{-rs}U_s ds + m_r \int_0^{\tau\wedge t} e^{-rs}G_s ds,$$

where m_w is portion among the asset management fees that is used to fund the GMWB rider with the annual high step-up. In practice, $m_w < m$ since insurance companies need to use part of the total fees they collect to cover other expenses. Therefore, the non-arbitrage net present value of the GMWB rider with the annual high step-up from an insurer's perspective is given by

$$m_w\mathbb{E}\left[\int_0^{\tau\wedge t} e^{-rs}U_s ds\right] + m_r\mathbb{E}\left[\int_0^{\tau\wedge t} e^{-rs}G_s ds\right] - w\mathbb{E}\left[\mathbb{1}_{\{\tau<t\}}G_\tau \int_\tau^t e^{-rs}ds\right]. \tag{3.37}$$

The last two terms in (3.37) have already been computed when we considered pricing the rider from a policyholder's perspective. Therefore, we focus on the first term in (3.37). We have

$$\mathbb{E}\left[\int_0^{\tau\wedge t} e^{-rs}U_s ds\right] = \mathbb{E}\left[\int_0^t e^{-rs}U_s ds\right] - \mathbb{E}\left[\mathbb{1}_{\{\tau<t\}} \int_\tau^t e^{-rs}U_s ds\right]. \tag{3.38}$$

For the first term on the right-hand side of the previous equation, we can directly calculate

$$
\begin{aligned}
&\mathbb{E}\left[\int_0^t e^{-rs}U_s ds\right]\\
=&\int_0^t \left((x-\frac{(w+m_r)x}{\psi_m^-(1)})e^{(\psi^+(1)+\psi_m^-(1)-r)s}+\frac{x(w+m_r)}{\psi_m^-(1)}e^{(\psi^+(1)-r)s}\right)ds\\
=&\frac{x(\psi_m^-(1)-w-m_r)}{\psi_m^-(1)(\psi^+(1)+\psi_m^-(1)-r)}(e^{(\psi^+(1)+\psi_m^-(1)-r)t}-1)+\frac{x(w+m_r)}{\psi_m^-(1)(\psi^+1(1)-r)}(e^{(\psi^+(1)-r)t}-1)\\
=&\frac{x(\psi_m^-(1)-w-m_r)}{\psi_m^-(1)(\psi^+(1)+\psi_m^-(1)-r)}e^{(\psi^+(1)+\psi_m^-(1)-r)t}+\frac{x(w+m_r)}{\psi_m^-(1)(\psi^+1(1)-r)}e^{(\psi^+(1)-r)t}\\
&-\frac{x(w+mr)\psi^+(1)+x(\psi^+(1)-r)\psi_m^+(1)}{\psi_m^-(1)(\psi^+(1)+\psi_m^-(1)-r)(\psi^+(1)-r)}.
\end{aligned}
$$

Similarly to the previous section, we can identify the Laplace transform of the second term in (3.38). Using the independence of $e^{X_t^+}$ and U_t^-, we have

$$
\mathbb{E}\left[\mathbb{1}_{\{\tau<t\}}\int_\tau^t e^{-rs}U_s ds\right]=x\mathbb{E}_1\left[\mathbb{1}_{\{\tau<t\}}\int_\tau^t e^{-(r-\psi^+(1))s}U_s^- ds\right]:=h(t,x).
$$

For $h(t,x)$, we have already identified its Laplace transform in Theorem 3.2.3 so we summarize the result in the next theorem without proof.

Theorem 3.3.4. *The Laplace transform $\hat{g}(q,x):=\int_0^\infty e^{-qt}g(t,x)dt$ exists for $q>\psi^+(1)+\psi_m^-(1)-r$, it is given by*

$$
\begin{aligned}
\hat{g}(q,x)=&\Big(-\frac{(w+m_r)x}{q(r-\psi^+(1))(r-\psi^+(1)-\psi_m^-(1))}-\frac{(w+m_r)x}{(q+r-\psi^+(1))(r-\psi^+(1))\psi_m^-(1)}\\
&+\frac{(w+m_r)x}{\psi_m^-(1)(r-\psi^+(1)-\psi_m^-(1))(q-\psi^+(1)-\psi_m^-(1)+r)}\Big)\mathbb{P}(I^-_{m,q+r-\psi^+(1)}>\frac{1}{w+m_r}),
\end{aligned}
\tag{3.39}
$$

where $I_{m,q}^-$ is the exponential functional defined in (3.31).

Now we have computed every term in (3.37), which determines the non-arbitrage net present value of the GMWB rider from an insurer's perspective. To obtain the numerical value of the price, the numerical Laplace inversion approach we mentioned in the previous section can be used. In conclusion, the pricing problem of the more complex product, the GMWB rider with the annual high step-up, can also be solved by a non-arbitrage approach. It can be seen that the nature of pricing the GMWB rider with the annual high step-up is similar to that of pricing the traditional GMWB rider: the distributional properties of the exponential functionals of Lévy processes. The pricing algorithms rely on the analytical solutions of the Laplace transforms of the quantities we encountered in previous sections. The analytical solutions enable us to find the fair value of fee charges accurately and efficiently. In the next section, we shall see some numerical examples showing the accuracy and the efficiency of the pricing algorithms we developed based the results we obtained in this section and in the previous section.

3.4 Numerical examples

In this section, we provide some numerical examples to explain how the results obtained in previous sections can be applied to solve the pricing problem of the GMWB rider. The pricing scheme for the traditional GMWB rider and the GMWB rider with the annual high step-up will be very similar, so we only provide examples of the pricing algorithms for the traditional GMWB rider. Our semi-analytical algorithms based on numerical Laplace inversion and previous sections' results are implemented in WOLFRAM MATHEMATICA 11. Monte Carlo simulations, which are the benchmarks we use for accuracy tests, are implemented in MATLAB 2015a.

The pricing of the GMWB rider can be done from two perspectives: from a policyholder's perspective and from an insurer's perspective. The nature of pricing GMWB from a policyholder's perspective (an insurer's perspective) is to find the fair fee rate m such that the GMWB rider's non-arbitrage net present value in (3.8) (in (3.12)) is 0. It was shown in [35] that, if there is no friction cost ($m = m_w$), pricing from the two perspectives are equivalent under the geometric Brownian motion model. As we shall see, the following examples will numerically confirm that this equivalence also holds under the jump diffusion model. To explain our pricing algorithm, we take the policyholder's perspective as an example and the pricing from an insurer's perspective is similar. To find the fair fee charges rate m, the bisection method can be used: we begin with two different fee charges rate m_l and m_r such that the non-arbitrage net present value in (3.8) have opposite signs. Within the interval $[m_l, m_r]$, a search for the value m which makes the value of (3.8) equals 0 is conducted until the pre-determined accuracy is reached. For each m, we need to compute the numerical value of (3.8). By Theorem 3.2.2, we obtained the Laplace transform of $\mathbb{E}[\max\{U_t, 0\}]$. To get its numerical value, numerical Laplace transform inversion algorithms are employed. We use the one-dimensional Euler's method described in [2]. We provide two sets of numerical examples as follows: in the first set, we provide accuracy tests for our algorithm in which we compare the accuracy of it with the Monte Carlo simulation for computing the value of $\mathbb{E}[\max\{U_t, 0\}]$. In the second set, we provide some examples showing the bisection method to find the fair level of fee charges. We shall see the equivalence between pricing from a policyholder's and an insurer's perspective under the condition $m = m_w$.

Accuracy test

Since the model we propose in this chapter includes a large class of exponential Lévy processes, it is more general than the geometric Brownian motion model introduced in [51, 35]. Therefore, our Laplace inversion method can also solve the pricing problem of the GMWB rider under the geometric Brownian motion model. Numerical PDE methods were used to find the fair fee charges rate in [51]. In [35], analytical solutions were identified and used to find the fair fee charges. As stated by Feng and Volkmer in [35], the accuracy of the numerical PDE method is not as high as the integration formulas in terms of Whittaker functions they obtained. Therefore, in our first example, we compare the accuracy and efficiency of our Laplace transform inversion method with Feng and Volkmer's method. Also, we provide the results obtained from Monte Carlo simulations as benchmarks. We use the same valuation basis described in [35, 51]:

- The annualized risk-free interest rate is $r = 0.05$ and the volatility coefficient is $\sigma = 0.2$;
- The annualized withdrawal rate is $w = 0.04$ and the fee charges rate is $m = 0.001782$.

For simplicity, we set $x = 1$ and hence the maturity of the GMWB rider is given by $t = x/w = 25$ years. We consider the remaining value of the policyholder's account at the maturity $\mathbb{E}\left[\max\{U_t, 0\}\right]$ and compare the calculation of this quantity using our Laplace inversion method (E.I), Feng and Volkmer's integration formula (I.F) as well as Monte Carlo simulations (M.C). For the Monte Carlo simulation, we generate the account value process by discretizing the maturity t into 25,000 segments, which corresponds to a time step of $\Delta t = 0.001$. For each simulation, the estimate of $\mathbb{E}[\max\{U_t, 0\}]$ is determined by averaging $\max\{U_t, 0\}$ for $N = 10,000$ sample paths. We repeat the simulation for $M = 50$ times to compute the average and the standard deviation (s.d) of the estimator. Table 3.1 summarizes these numerical results, from which we can observe that:

$\mathbb{E}[\max\{U_t, 0\}]$	E.I	I.F	M.C
Value	1.49809	1.49809	1.5075 ± 0.0256
Time (s)	0.29	3.04	982.10

Table 3.1: Accuracy and efficiency comparison under geometric Brownian motion

- The result obtained by using our semi-analytical method, which is based on numerical Laplace inversion, agrees with the result obtained by Feng and Volkmer ([35]). The value lies within the confidence interval constructed from Monte Carlo simulations. This verifies the accuracy of our semi-analytical method.
- The efficiency of our semi-analytical method is the highest among the three methods. Our method is even faster than the analytical method proposed by Feng and Volkmer ([35]), which is based on integration formulas of Whittaker functions. Though we are using different computing device and

software with Feng and Volkmer ([35]), which may affect the fairness of comparison for efficiency, the accuracy and efficiency of our semi-analytical method is high enough for most practical purposes.

Next, we provide an accuracy test for our semi-analytical method under the exponential Lévy processes model. We still choose to verify the quantity $\mathbb{E}[\max\{U_t, 0\}]$ and compare the efficiency and accuracy of our method with the Monte Carlo simulation. For the valuation basis, we set:

- The annualized risk-free interest rate is $r = 0.05$ and volatility coefficient is $\sigma = 0.2$;
- For the jump part of the Lévy process, we set $(\rho_1 = 30.5, \rho_2 = 50.1, \hat{\rho}_1 = 30.2, \hat{\rho}_2 = 40.8)$, $(\alpha_1 = 4.8, \alpha_2 = -0.8, \hat{\alpha}_1 = 7.8, \hat{\alpha}_2 = -1.8)$. This generates a mixed-exponential jump process with a jump rate $\lambda = 10$ and a two-sided jump distribution.
- The annualized withdrawal rate is $w = 0.04$ and fee charges rate is $m = 0.005$.

For the Monte Carlo simulation, we use the same discretizing settings as the previous geometric Brownian motion model. To simulate the mixed-exponential jumps, rejection sampling is used: we first generate a mixed-exponential jump distribution with only positive weights in α_i and $\hat{\alpha}_j$ $(i = 1, \ldots, J, j = 1, \ldots, \hat{J})$, then use rejection method to take negative weights into consideration. Numerical results are summarized in Table 3.2.

$\mathbb{E}[\max\{U_t, 0\}]$	E.I	M.C
Value	1.42466	1.4170 ± 0.0456
Time (s)	1.81	14469.89

Table 3.2: Accuracy and efficiency comparison under jump diffusion

From Table 3.2 we can observe that, the value obtained by our semi-analytical method is within one standard deviation from the sample mean of the Monte Carlo simulation, which verifies its accuracy. The running time (1.81 s) of the algorithm is longer than that of the geometric Brownian motion model (0.29 s). This is due to computing the tail probability $\mathbb{P}(I^-_{m,q} > x/w)$ of the exponential functional. Because of the jump component, the equation $\psi_m(z) = q$ has 6 roots under the jump diffusion model while it only has 2 roots under the geometric Brownian motion model. Therefore, it takes more time to find all roots and to compute the Meijer-G function. In addition, we can also see that simulating a jump diffusion process is much more time-consuming than simulating a Brownian motion. For each discretized time interval, we need to simulate one more Poisson random variable as well as the mixed-exponential jump distribution. The rejection sampling takes more time. It is not hard to imagine that if the jump part has more exponential components, the Monte Carlo method will be extremely time-consuming. Our semi-analytical method becomes preferable in this situation.

Bisection pricing

In this part, we provide numerical examples to illustrate how to determine the fair fee charges rate m. We consider the two models introduced in the last part: the underlying equity process is a geometric Brownian motion (G.B.M) and a jump diffusion (J.D). The evaluation basis is the same as the last part. We consider pricing the rider from both a policyholder's perspective and an insurer's perspective under the condition $m = m_w$, which means no friction cost. We start with initial guesses of fair fee charges rate $[m_l, m_r]$, where the net values in (3.8) and (3.12) evaluated at m_l and m_r have opposite signs. We use the bisection method on the interval $[m_l, m_r]$ to find the root m and the algorithm will stop if the error is less than 10^{-6}. The results are summarized in Table 3.3, from which we can conclude that:

- If there is no friction cost ($m = m_w$), pricing from a policyholder's perspective and an insurer's perspective are equivalent. we can see the pricing results agree with each other.
- Pricing the rider under the jump diffusion model is more time-consuming than pricing under the geometric Brownian motion model. This is due to the increase of the model complexity we mentioned previously. The pricing algorithm under both models is still very efficient and accurate.

It is not hard to imagine that Monte Carlo simulations can hardly be employed in the pricing of the GMWB rider. Since the bisection method requires high precision in computing the net present value, the simulation error will prevent Monte Carlo methods from successfully conducting the binary search. In addition, under the more general exponential Lévy processes model, no analytical result regarding the quantities related to the pricing of the GMWB rider is currently known. Therefore, our method based on numerical Laplace inversions provides an efficient and accurate way to solve the pricing problem the GMWB rider.

	G.B.M-investor	G.B.M-insurer	J.D-investor	J.D-insurer
m (basis points)	17.8238	17.8238	34.8636	34.8636
Time (s)	3.66	2.96	29.41	29.09

Table 3.3: Bisection pricing under different models and from different perspectives, with initial guesses $[5, 50]$ base points

Chapter 4

Intrinsic Ultracontractivity of the (Fractional) Laplacian Under nonlocal Perturbations

4.1 Introduction

For $d \geq 1$ and $0 < \beta < \alpha \leq 2$. We consider the operator $\mathcal{L}^b = \Delta^{\alpha/2} + \mathcal{S}^b$, where

$$\mathcal{S}^b f(x) := \int_{\mathbb{R}^d} (f(x+z) - f(x) - \mathbb{1}_{\{|z|\leq 1\}} \langle \nabla f(x), z \rangle) \frac{b(x,z)}{|z|^{d+\beta}} dz$$

and $b(x,z)$ is a bounded measurable function on $\mathbb{R}^d \times \mathbb{R}^d$ with $b(x,z) = b(x,-z)$ for $x, z \in \mathbb{R}^d$. This class of operators can be seen as a Laplace operator ($\alpha = 2$) or a fractional Laplace operator ($0 < \alpha < 2$) under a lower order perturbation $\mathcal{S}^b$. They were introduced in [21, 64], where the authors considered the existence and uniqueness of the fundamental solution $p^b(t,x,y)$ corresponding to these operators. It was shown in [21, 64] that if $b(x,z)$ satisfies certain conditions, the fundamental solution $p^b(t,x,y)$ is a strictly positive and continuous function. $p^b(t,x,y)$ determines a conservative Feller process X^b which has the strong Feller property. Various forms of sharp two-sided estimates of $p^b(t,x,y)$ on $\mathbb{R}^d$ were obtained in [21, 64]. Later in [22], the authors considered the killed process of X^b upon leaving a $C^{1,1}$ open set (see [22] for the definition). Sharp two-sided estimates of the transition density of the killed process were also obtained when $0 < \alpha < 2$ and $d \geq 2$. In this section, we consider the intrinsic ultracontractivity of the killed process $X^{b,D}$ of X^b upon leaving a general bounded open set $D \subset \mathbb{R}^d$. Two-sided estimates of the transition density of $X^{b,D}$ in terms of ground state eigenfunctions will be obtained.

Our method of establishing intrinsic ultracontractivity of $X^{b,D}$ is probabilistic, which relies on the heat kernel estimates of the original process X^b. The heat kernel estimates of the Laplacian and the fractional Laplacian, however, are different. For this reason, we separate our analyses with respect to the case of $\alpha = 2$ and the case of $0 < \alpha < 2$. We establish the intrinsic ultracontractivity for $\Delta + \mathcal{S}^b$ in a bounded open set first, then apply the same method to $\Delta^{\alpha/2} + \mathcal{S}^b$. From Subsection 4.2.1 to Subsection 4.2.4, we begin with reviewing some preliminaries about the Laplacian under the perturbation $\mathcal{S}^b$. Then we consider the killed process $X^{b,D}$ of X^b in a bounded open set and its properties. In order to establish intrinsic ultracontractivity, the existence of the dual process $\widehat{X}^b_D$ of $X^{b,D}$ must be guaranteed. This is achieved by constructing an appropriate reference measure which guarantees the existence of $\widehat{X}^b_D$. We finally prove the intrinsic ultracontractivity of $\Delta + \mathcal{S}^b$ in Subsection 4.2.4. The case of the fractional Laplacian under perturbation is discussed in Section 4.3. The idea of the proof is very similar to the case of the Laplacian. Due to different heat kernel estimates and jump behaviors, different assumptions on $b(x,z)$ are needed and some properties from Subsection 4.2.1 to Subsection 4.2.4 need to be re-established. The result in this part of the thesis on intrinsic ultracontractivity is a joint work with Yinghui Shi.

4.2 Intrinsic ultracontractivity of $\Delta + \mathcal{S}^b$ in bounded open sets

4.2.1 Preliminaries

First, we review some facts about α-stable processes and Brownian motions perturbed by α-stable processes. Let $d \geq 1$ be an integer and $0 < \alpha < 2$. Let Y be a d-dimensional Brownian motion and Z be a symmetric α-stable process. Recall the definition of an α-stable process in Example 1.0.10. The operator $\mathcal{S}^b$ can be considered as a generalization of the generator of an α-stable process. Previously, [19, 62] considered the heat kernel estimates of the independent mixture of a Brownian motion and an α-stable process: $Y^a_t := Y_t + a^{1/\alpha} Z_t$. The infinitesimal generator of Y^a is given by $\Delta + a\Delta^{\alpha/2}$. We denote the transition density of Y^a by $p_a(t,x,y)$. It was shown in [19, 62, 64] that

$$\begin{aligned} &c_1(t^{-d/2} \wedge (at)^{-d/\beta}) \wedge (p_0(t, c_2 x, c_2 y) + \frac{at}{|x-y|^{d+\alpha}}) \leq p_a(t,x,y) \\ \leq\ & c_3(t^{-d/2} \wedge (at)^{-d/\alpha}) \wedge (p_0(t, c_4 x, c_4 y) + \frac{at}{|x-y|^{d+\alpha}}), \qquad \text{for any } (t,x,y) \in (0,\infty) \times \mathbb{R}^d \times \mathbb{R}^d. \end{aligned} \tag{4.1}$$

$c_i, i = 1, \dots, 4$ are positive constants depending on α and d. $p_0(t,x,y) = (4\pi t)^{-d/2} e^{-|x-y|^2/4t}$ is the heat kernel of a standard Brownian motion. The process Y^a_t can be seen as a Brownian motion perturbed by a lower order α-stable process. The class of operators $\Delta + \mathcal{S}^b$ here, which were introduced in [64], can be seen as a generalization of the work of [19, 62]. It was shown in [64] that if $b(x,z)$ is bounded and satisfies

$$b(x,z) = b(x,-z), \qquad \text{for any } x, z \in \mathbb{R}^d \tag{4.2}$$

and

$$b(x,z) \geq 0, \qquad \text{for any } x \in \mathbb{R}^d \text{ and for a.e. } z \in \mathbb{R}^d, \tag{4.3}$$

then the fundamental solution $p^b(t,x,y)$ of $\mathcal{L}^b$ is a strictly positive and continuous function. $p^b(t,x,y)$ determines a conservative Feller process X^b which has the strong Feller property. Different forms of sharp two-sided estimates of $p^b(t,x,y)$ on $\mathbb{R}^d$ were obtained under different assumptions on $b(x,z)$ in [64]. In this section, we consider the killed process $X^{b,D}$ of X^b upon leaving a bounded open set $D \subset \mathbb{R}^d$. We consider the two-sided estimates of the heat kernel $p^b_D(t,x,y)$ of $X^{b,D}$ by establishing the intrinsic ultracontractivity of the semigroup generated by $X^{b,D}$. Here, we do not impose any additional assumptions on the bounded open set D. As we shall see later, we only need to make slightly stronger assumptions on $b(x,z)$, the intrinsic ultracontractivity of the semigroup of $X^{b,D}$ will be guaranteed. The next theorem summarizes some of the key facts about the existence and the estimates of heat kernel $p^b(t,x,y)$, where we only assume that $b(x,z)$ satisfies the most general conditions in [64]. These conditions guarantee the existence and the positivity of the fundamental solution of $\Delta + \mathcal{S}^b$. Define

$$m_b := \inf_x \text{ess inf}_z b(x,z) \qquad \text{and} \qquad M_b := \text{ess sup}_{x,z} b(x,z).$$

The following result is from [64] and the heat kernel estimates in it are crucial for our analysis.

Theorem 4.2.1 (Wang (2015)). *For any $A > 0$, there are positive constants $C_i = C_i(d,\beta,A) > 0, i = 1,\dots,7$, so that for any bounded function $b(x,z)$ with $||b||_\infty < A$ and satisfying condition (4.2) and (4.3), we have*

$$C_1^{-1} p_{m_b}(t, C_2x, C_2y) \le p^b(t,x,y) \le C_1 p_{M_b}(t, C_3x, C_3y), \qquad \text{for } t \in (0,1] \text{ and } x, y \in \mathbb{R}^d. \tag{4.4}$$

Moreover,

$$C_4^{-1} e^{-C_5t} p_{m_b}(t, C_6x, C_6y) \le p^b(t,x,y) \le C_4 e^{C_5t} p_{M_b}(t, C_7x, C_7y), \qquad \text{for } t \in (0,\infty) \text{ and } x, y \in \mathbb{R}^d. \tag{4.5}$$

The heat kernel $p^b(t,x,y)$ uniquely determines a Feller process $X^b = \{X^b, t \ge 0, \mathbb{P}_x, x \in \mathbb{R}^d\}$ on the canonical Skorokhod space $\mathbb{D}([0,\infty), \mathbb{R}^d)$ such that

$$\mathbb{E}_x[f(X^b_t)] = \int_{\mathbb{R}^d} p^b(t,x,y) f(y) dy$$

for every bounded continuous function f on $\mathbb{R}^d$. The Feller process X^b is conservative and has a Lévy system $(J^b(x,y)dy, t)$ with

$$J^b(x,y) = \frac{b(x, x-y)}{|x-y|^{d+\beta}},$$

which means for any stopping time T and any nonnegative measurable f on $\mathbb{R}^+ \times \mathbb{R}^d \times \mathbb{R}^d$ satisfying $f(s,y,y) = 0$ for any $s \in \mathbb{R}_+$ and $y \in \mathbb{R}^d$, we have the following identity

$$\mathbb{E}_x \left[\sum_{s \leq T} f(s, X^b_{s-}, X^b_s) \right] = \mathbb{E}_x \left[\int_0^T \left(\int_{\mathbb{R}^d} f(s, X^b_s, y) J(X^b_s, y) dy \right) ds \right]. \tag{4.6}$$

For the rest of this section, the upper case constants $C_1, C_2, \cdots$ are fixed, the lower case constants $c_0, c_1, c_2 \cdots$ can change from one appearance to another. We use dx to denote the Lebesgue measure in $\mathbb{R}^d$ and use $|A|$ to denote the Lebesgue measure of a Borel set $A \subset \mathbb{R}^d$.

To obtain the intrinsic ultracontractivity of $\Delta + \mathcal{S}^b$ for any bounded open set, stronger assumptions on $b(x,z)$ are needed in addition to (4.2) and (4.3). Consider the special case $b(x,z) \equiv 0$, $X^{b,D}$ degenerates to a killed Brownian motion on an arbitrary bounded open set $D \subset \mathbb{R}^d$. If the open set D has multiple connected components, the Brownian motion can not travel freely between the connected components due to its continuous sample path. Hence the transition density of the killed process will be 0 on some of the connected components of D. In that case, intrinsic ultracontractivity can not be established. This shows the necessity of keeping $b(x,z)$ strictly positive for any compact sets. In this case, the process X^b_D can jump between any two connected components in D. So we impose the following stronger assumption on $b(x,z)$: for any compact subset $K, L \subset \mathbb{R}^d$,

$$\inf_{x \in K} \inf_{z \in L} b(x,z) > 0. \tag{4.7}$$

This condition will only be used in the proof of Theorem 4.2.9 and Lemma 4.2.17. In Theorem 4.2.9, it is used to show the strict positivity of the heat kernel of the killed process $X^{b,D}$ on any open set. In Lemma 4.2.17, it is used as an intermediate step to establish intrinsic ultracontractivity.

4.2.2 Properties of the killed process

For any open subset $D \subset \mathbb{R}^d$, we define $\tau^b_D = \inf\{t > 0 : X^b \notin D\}$, which is known as the first exit time of D for X^b. The subprocess $X^{b,D}_t$ of X^b in D is defined as

$$X^{b,D}_t = \begin{cases} X^b_t, & \text{if } t < \tau^b_D \\ \partial, & \text{if } t \geq \tau^b_D \end{cases} \tag{4.8}$$

where ∂ is a cemetery state. $X^{b,D}$ is also referred as the killed process X^b upon exiting from D. Throughout this section, we use the convention that $f(\partial) = 0$ for any function f. We use $\{P^{b,D}_t : t \geq 0\}$ and $\mathcal{L}^b|_D$ to denote the semigroup and the infinitesimal generator of $X^{b,D}$, respectively. Because of the strong Markov property of X^b, for any $t > 0$ and Borel set $A \subset \mathbb{R}^d$, we have

$$\begin{aligned}
\mathbb{P}_x(X_t^{b,D} \in A) &= \mathbb{P}_x(X_t^b \in A, t < \tau_D^b) \\
&= \mathbb{P}_x(X_t^b \in A) - \mathbb{P}_x(X_t^b \in A, \tau_D^b < t) \\
&= \mathbb{P}_x(X_t^b \in A) - \mathbb{E}_x[\mathbb{P}_{X^b_{t-\tau_D^b}}(X^b_{t-\tau_D^b} \in A); \tau_D^b < t] \\
&= \mathbb{P}_x(X_t^b \in A) - \mathbb{E}_x[\int_A p^b(t-\tau_D^b, X^b_{t-\tau_D^b}, y)dy; \tau_D^b < t] \\
&= \mathbb{P}_x(X_t^b \in A) - \int_A \mathbb{E}_x[p^b(t-\tau_D^b, X^b_{t-\tau_D^b}, y); \tau_D^b < t]dy
\end{aligned}$$

Therefore, the transition density $p_D^b(t,x,y)$ of the process $X^{b,D}$ with respect to the Lebesgue measure is given by

$$p_D^b(t,x,y) = p^b(t,x,y) - k_D^b(t,x,y), \tag{4.9}$$

with

$$k_D^b(t,x,y) := \mathbb{E}_x[p^b(t-\tau_D^b, X^b_{\tau_D^b}, y); \tau_D^b < t]. \tag{4.10}$$

We would like to show that the transition density $p_D^b(t,x,y)$ in (4.9) is continuous and strictly positive in D. Firstly, we focus on the continuity. We begin with several basic lemmas. The first lemma's proof is standard and could be found, for example, in [64, Lemma 2.1] and [62, Lemma 2.1], therefore we omit it.

Lemma 4.2.2. *There exists a constant $C_1 = C_1(d) > 0$, such that*

$$p_0(t,x,y) \le C_1 \left(t^{-d/2} \wedge \frac{t}{|x-y|^{d+2}} \right), \qquad \text{for any } (t,x,y) \in (0,\infty) \times \mathbb{R}^d \times \mathbb{R}^d. \tag{4.11}$$

Lemma 4.2.3. *For any $\delta > 0$,*

$$\lim_{s \downarrow 0} \sup_{x \in \mathbb{R}^d} \mathbb{P}_x(\tau^b_{B(x,\delta)} \le s) = 0.$$

Proof. By the strong Markov property of the Feller process X^b, we have

$$\begin{aligned}
&\mathbb{P}_x(\tau^b_{B(x,\delta)} \le s) \\
\le &\mathbb{P}_x(\tau^b_{B(x,\delta)} \le s, X_s^b \in B(x,\delta/2)) + \mathbb{P}_x(X_x^b \in B(x,\delta/2)^c) \\
\le &\mathbb{E}_x[\mathbb{P}_{X_{\tau^b_{B(x,\delta)}}}(|X^b_{s-\tau^b_{B(x,\delta)}} - X_0^b| \ge \delta/2); \tau^b_{B(x,\delta)} \le s] + \mathbb{P}_x(X_x^b \in B(x,\delta/2)^c) \\
\le &2 \sup_{t \le s} \sup_{x \in \mathbb{R}^d} \mathbb{P}_x(|X_t^b - X_0^b| \ge \delta/2).
\end{aligned}$$

For the last term in the previous inequality, using heat kernel estimates (4.4), (4.1) and Lemma 4.2.2, we have

$$
\begin{aligned}
&\sup_{t\leq s}\sup_{x\in\mathbb{R}^d}\mathbb{P}_x(|X_t^b-X_0^b|\geq\delta/2)\\
=&\sup_{t\leq s}\sup_{x\in\mathbb{R}^d}\int_{|y-x|>\delta/2}p^b(t,x,y)dy\\
\leq&\sup_{t\leq s}\sup_{x\in\mathbb{R}^d}c_1\int_{|y-x|>\delta/2}(t^{-d/2}\wedge(at)^{-d/\beta})\wedge(p_0(t,c_2x,c_2y)+\frac{M_bt}{c_3|x-y|^{d+\beta}})dy\\
\leq&\sup_{t\leq s}\sup_{x\in\mathbb{R}^d}c_1\int_{|y-x|>\delta/2}(\frac{t}{|x-y|^{d+2}}+\frac{M_bt}{|x-y|^{d+\beta}})dy\\
\leq&c_1s\int_{\delta/2}^{\infty}(\frac{1}{r^3}+\frac{M_b}{r^{1+\beta}})dr\\
\leq&c_1s(\frac{1}{\delta^2}+\frac{1}{\delta^\beta})\to 0
\end{aligned}
$$

as $s\to 0$. We complete the proof of the lemma. □

With the help of Lemma 4.2.3, we can show the continuity of the transition density $p_D^b(t,x,y)$. The method of the proof is similar to [17, Theorem 3.4] and [24, p33, Theorem 2.4].

Theorem 4.2.4. *The transition density $p_D^b(t,x,y)$ of the killed process $X^{b,D}$ on any open set D is jointly continuous in (t,x,y). For every $t,s>0$, it satisfies the semigroup property*

$$
p_D^b(t+s,x,y)=\int_D p_D^b(t,x,z)p_D^b(s,z,y)dz. \tag{4.12}
$$

Proof. Since $p^b(t,x,y)$ is jointly continuous, we just need to show that $k_D^b(t,x,y)$ is jointly continuous in (t,x,y). By (4.5), (4.1) and Lemma 4.2.2, there exist $c_1,c_2>0$ so that

$$
\begin{aligned}
&\sup_{t\leq t_0}\sup_{|x-y|\geq\delta}p^b(t,x,y)\\
\leq&c_1e^{c_2t_0}\sup_{t\leq t_0}\sup_{|x-y|\geq\delta}(t^{-d/2}\wedge(at)^{-d/\beta})\wedge(p_0(t,c_4x,c_4y)+\frac{M_bt}{|x-y|^{d+\beta}})\\
\leq&c_1e^{c_2t_0}\sup_{t\leq t_0}\sup_{|x-y|\geq\delta}(p_0(t,c_4x,c_4y)+\frac{M_bt}{|x-y|^{d+\beta}})\\
\leq&c_1e^{c_2t_0}t_0(\frac{1}{\delta^{d+2}}+\frac{1}{\delta^{d+\beta}})<\infty,
\end{aligned}
\tag{4.13}
$$

for all $t_0>0$ and $\delta>0$. For any sufficiently small $\delta>0$, define set $D_\delta=\{x\in D:\rho(x,D^c)>\delta\}$. Define a function

$$
h(s,r,x,y):=\mathbb{E}_x[p^b(r-\tau_D^b,X^b_{\tau_D^b},y);s\leq\tau_D^b<r],
$$

for any $0 \le s < r$ and $x, y \in D_\delta$. Then by the strong Markov property of X^b, we have

$$\begin{aligned}
k_D^b(t,x,y) =& h(0,t,x,y) \\
=& h(s,t,x,y) + \mathbb{E}_x[p^b(t-\tau_D^b, X^b_{\tau_D^b}, y); \tau_D^b < s] \\
=& \mathbb{E}_x[h(0,t-s,X_s^b,y)] - \mathbb{E}_x[h(0,t-s,X_s^b,y); \tau_D^b < s] \\
& + \mathbb{E}_x[p^b(t-\tau_D^b, X^b_{\tau_D^b}, y); \tau_D^b < s].
\end{aligned} \tag{4.14}$$

Note that for any given $t_0 > 0$, $p^b(t,x,y)$ is uniformly bounded by certain constant $c_1 > 0$ on $[0,t_0] \times D^c \times D_\delta$ by (4.13). This indicates that $p^b(t-\tau_D^b, X^b_{\tau_D^b}, y)$ is uniformly bounded in (t,y) by c_1 and $h(0,t-s,X_s^b,y)$ is uniformly bounded in (t,s,y) by c_1. Therefore, for any $x, y \in D_\delta$ we have

$$\begin{aligned}
& |k_D^b(t,x,y) - \mathbb{E}_x[h(0,t-s,X_s^b,y)]| \\
\le& \mathbb{E}_x[p^b(t-\tau_D^b, X^b_{\tau_D^b}, y); \tau_D^b < s] + \mathbb{E}_x[h(0,t-s,X_s^b,y); \tau_D^b < s] \\
\le& 2c_1 \mathbb{P}_x(\tau_D^b < s) \\
\le& 2c_1 \sup_{x \in \mathbb{R}^d} \mathbb{P}_z(\tau^b_{B(x,\delta)} < s).
\end{aligned}$$

By Lemma 4.2.3, the last term in previous inequality goes to 0 as $s \to 0$. So $\mathbb{E}_x[h(0,t-s,X_s^b,y)]$ converges to $k_D^b(t,x,y)$ as $s \to 0$ uniformly in $(t,x,y) \in [0,t_0] \times D_\delta \times D_\delta$. We only need to show $\mathbb{E}_x[h(0,t-s,X_s^b,y)]$ is jointly continuous in (s,t,x,y) due to uniform convergence. Since $p^b(t,x,y)$ is jointly continuous and uniformly bounded on $(t,x,y) \in [0,t_0] \times D_\delta \times D_\delta$, by the bounded convergence theorem, $\mathbb{E}_x[h(0,t-s,X_s^b,y)]$ is jointly continuous in (s,t,y) on $\{(s,t,y) : s \le t, t \le t_0, y \in D_\delta\}$. Moreover,

$$\mathbb{E}_x[h(0,t-s,X_s^b,y)] = \int_{\mathbb{R}^d} p(s,x,z) h(0,t-s,z,y) dz$$

is equicontinuous in x for any locally compact subset of $\{(s,t,y) : s \le t, t \le t_0, y \in D_\delta\}$. Therefore, $\mathbb{E}_x[h(0,t-s,X_s^b,y)]$ is jointly continuous in $(s,t,x,y) \in \{(s,t,x,y) : s \le t, t \le t_0, x \in D_\delta, y \in D_\delta\}$. As a result, $k^b(t,x,y)$ is jointly continuous in (t,x,y) on $\{(t,x,y) : t \le t_0, x \in D_\delta, y \in D_\delta\}$ for any $\delta > 0$ and $t_0 > 0$. Consequently, $k_D^b(t,x,y)$ is jointly continuous on $(0,\infty) \times D \times D$. For the semi-group property, we have

$$\begin{aligned}\int_A p_D^b(t+s,x,y)dy &= \mathbb{P}_x(X_{t+s}^b \in A, t+s<\tau_D^b)\\ &= \mathbb{E}_x[t<\tau_D^b; \mathbb{P}_{X_t^b}(X_s^b\in A; \tau_D^b<s)]\\ &= \int_A \left(\int_D p_D^b(t,s,z)p_D^b(s,z,y)dz\right)dy, \qquad \text{for any } A\in\mathcal{B}(D).\end{aligned}$$

(4.12) follows from the continuity of $p_D^b(t,x,y)$ and the proof is completed. □

Next, we would like to show the strict positivity of $p_D^b(t,x,y)$. The main facts we use in the following proofs are the heat kernel estimates of $p^b(t,x,y)$ in (4.4). We point out here that if $m_b>0$, the upper and lower bound in (4.4) would be of the same form. In this case, the proof for strict positivity is easier. For the case $m_b=0$, however, we need to use a different method.

Lemma 4.2.5. *Under the condition $m_b>0$, for any given sufficiently small $r_0>0$, there exists*

$$t_1=t_1(r_0,b,d,\beta)>0 \qquad \textit{and} \qquad r_1=r_1(r_0,b,d,\beta)>0,$$

such that for any $x,y\in D$ and $t<t_1$, satisfying $r_0<\rho(x,\partial D)\wedge\rho(y,\partial D)$ and $\rho(x,y)<r_1$, we have $p_D^b(t,x,y)>0$.

Proof. For any x,y satisfying $\rho(x,\partial D)>r_0$ and $\rho(y,\partial D)>r_0$, we first recall the definition of $p_D^b(t,x,y)$ in (4.9) and (4.10). Using the estimates in (4.4) and (4.1), there exist constants $c_i:=c_i(d,A,\beta)$, $i=1,\dots,6$ so that for $t<1$,

$$\begin{aligned}p_D^b(t,x,y) &= p^b(t,x,y)-k_D^b(t,x,y)\\ &= p^b(t,x,y)-\mathbb{E}_x[p^b(t-\tau_D^b, X_{\tau_D^b}^b, y); \tau_D^b<t],\end{aligned}$$

where

$$p^b(t,x,y)>c_1\left(t^{-d/2}\wedge(m_bt)^{-d/\beta}\right)\wedge\left((4\pi t)^{-d/2}e^{-\frac{c_2^2|x-y|^2}{4t}}+\frac{m_bt}{c_5|x-y|^{d+\beta}}\right),$$

and

$$p^b(t-\tau_D^b, X_{\tau_D^b}^b, y)<c_3\left((t-\tau_D^b)^{-d/2}\wedge(M_b(t-\tau_D^b))^{-d/\beta}\right)\wedge\left((4\pi(t-\tau_D^b))^{-d/2}e^{-\frac{c_4^2r_0^2}{4(t-\tau_D^b)}}+\frac{M_b(t-\tau_D^b)}{c_6r_0^{d+\beta}}\right).$$

It is easy to see that the function

$$f(t, r_0, c_4, c_6, M_b, d) := (4\pi t)^{-d/2} e^{-\frac{c_4^2 r_0^2}{4t}} + \frac{M_b t}{c_6 r_0^{d+\beta}},$$

is increasing for $t \in [0, \frac{c_2^4 r_0^2}{2d}]$ and $\lim_{t\to 0} f(t, r_0, c_4, c_6, M_b, d) = 0$. Therefore, there exists $t_0(r_0, d, A, \beta, b) > 0$, such that the following holds for any $t < t_0$,

$$\begin{cases} f(t, r_0, c_4, c_6, M_b, d) < \frac{c_1}{c_3} t^{-d/2} \wedge (m_b t)^{-d/\beta} \\ f(t, r_0, c_4, c_6, M_b, d) < t^{-d/2} \wedge (M_b t)^{-d/\beta} \\ \frac{\partial f}{\partial t}(t, r_0, c_4, c_6, M_b, d) > 0. \end{cases}$$

So we have

$$\begin{aligned} k_D^b(t, x, y) &= \mathbb{E}_x[p^b(t - \tau_D^b, X^b_{\tau_D^b}, y); \tau_D^b < t] \\ &< c_3 \mathbb{E}_x[(4\pi t)^{-d/2} e^{-\frac{c_4^2 r_0^2}{4t}} + \frac{M_b t}{c_6 r_0^{d+\beta}}; \tau_D^b < t] \\ &< c_3 (4\pi t)^{-d/2} e^{-\frac{c_4^2 r_0^2}{4t}} + \frac{c_3 M_b t}{c_6 r_0^{d+\beta}}. \end{aligned}$$

Now we only need to prove that there exists $r_1 = r_1(r_0, b, d, \beta) > 0$, so that for any $|x - y| < r_1$, we have

$$c_1 (4\pi t)^{-d/2} e^{-\frac{c_2^2 |x-y|^2}{4t}} + \frac{c_1 m_b t}{c_5 |x-y|^{d+\beta}} > c_3 (4\pi t)^{-d/2} e^{-\frac{c_4^2 r_0^2}{4t}} + \frac{c_3 M_b t}{c_6 r_0^{d+\beta}}.$$

It will be sufficient to prove the previous inequality if the following conditions are satisfies:

$$\begin{cases} \frac{c_1}{c_3} > e^{-\frac{c_4^2 r_0^2 - c_2^2 |x-y|^2}{4t}}, \\ (\frac{c_1 m_b}{c_3 M_b} \frac{c_6}{c_5})^{\frac{1}{d+\beta}} r_0 > |x - y|. \end{cases}$$

We define

$$\begin{cases} r_1 := (\frac{c_1 m_b}{c_3 M_b} \frac{c_6}{c_5})^{\frac{1}{d+\beta}} r_0 \wedge \frac{c_4}{\sqrt{2} c_2} r_0, \\ t_1 := t_0 \wedge \frac{c_4^2 r_0^2}{8 \log(\frac{c_3}{c_1})}. \end{cases}$$

Then for any $|x-y|<r_1$ and $t<t_0$, we have

$$\frac{c_1 m_b t}{|x-y|^{d+\beta}} > \frac{c_1 m_b t}{r^{d+\beta}} > \frac{c_3 M_b t}{r_0^{d+\beta}}$$
$$\frac{c_1}{c_3} > e^{-\frac{c_4^2 r_0^2}{8t}} > e^{-\frac{c_4^2 r_0^2 - c_2^2|x-y|^2}{4t}},$$

which implies

$$p_D^b(t,x,y) > c_1\left(t^{-d/2}\wedge(m_b t)^{-d/\beta}\right)\wedge\left((4\pi t)^{-d/2}e^{-\frac{c_2^2 r^2}{4t}} + \frac{m_b t}{c_5 r^{d+\beta}}\right) - c_3(4\pi t)^{-d/2}e^{-\frac{c_4^2 r_0^2}{4t}} - \frac{c_3 M_b t}{c_6 r_0^{d+\beta}}$$
$$> 0.$$

The proof is completed. □

The previous lemma shows the strict positivity of $p_D^b(t,x,y)$ locally. To show the positivity of $p_D^b(t,x,y)$ globally, we use a chaining argument whose idea comes from [24].

Theorem 4.2.6. *Let D be any domain (a connected open set) in $\mathbb{R}^d$. If $m_b>0$, then for any $x,y\in D$, $p_D^b(t,x,y)>0$.*

Proof. Since D is a domain, for any $x,y\in D$, there exists a curve Γ in D connecting x,y. we can find r_0 sufficiently small such that

$$\rho(\Gamma,\partial D)=2r_0>0.$$

By Lemma 4.2.5, there exists $r_1>0$ and $t_1>0$, such that for any $x,y\in D$, if $|x-y|<r_1$, $\rho(x,\partial D)>r_0$, $\rho(y,\partial D)>r_0$ and $t<t_1$, $p_D^b(t,x,y)>0$. We choose a sufficiently large integer n such that $\frac{t}{n}<t_0$ and there exist points $a_0=x, a_1,\dots,a_n,a_{n+1}=y$ on Γ with $a_i\in B(a_{i-1}\frac{r}{3}\wedge r_0)$ for any $i=1,\dots,n+1$. Then for any $x_i\in B(a_i,\frac{r_1}{3}\wedge r_0)$, we have $|x_i-x_{i-1}|<|x_i-a_i|+|a_i-a_{i-1}|+|a_{i-1}-x_{i-1}|<r_1$ and $\rho(x_i,\partial D)\geq 2r_0-r_0>r_0$. Finally, using the semigroup property (4.12), we have

$$p_D^b(t,x,y)=\int_D\cdots\int_D p_D^b(\frac{t}{n},x,x_1)p_D^b(\frac{t}{n},x,x_1)\dots p_D^b(\frac{t}{n},x_n,y)dx_1\dots dx_n$$
$$>\int_{B(a_1,\frac{r}{3}\wedge r_0)}\cdots\int_{B(a_n,\frac{r}{3}\wedge r_0)} p_D^b(\frac{t}{n},x,x_1)p_D^b(\frac{t}{n},x,x_1)\dots p_D^b(\frac{t}{n},x_n,y)dx_1\dots dx_n.$$
$$>0.$$

In the last inequality, we used the continuity of $p_D^b(t,x,y)$ and Lemma 4.2.5. □

Then we would like to prove the strict positivity of $p_D^b(t,x,y)$ on any domain under the weaker condition

$m_b \geq 0$. Different from the approach we used for the case of $m_b > 0$, we would like to prove the strict positivity of $p_D^b(t,x,y)$ when D is a ball.

Lemma 4.2.7. *Let D be a ball with radius R. i.e $D = B(z_0, R)$ for some $z_0 \in \mathbb{R}^d$. Under the condition $m_b = 0$, we have $p_D^b(t, x_0, y_0) > 0$ for any $x_0, y_0 \in D$ and $t > 0$.*

Proof. Similar to Lemma 4.2.5, we can show that for any given $r_0 > 0$ and $x, y \in D$ with $\rho(x, \partial D) > r_0$ and $\rho(y, \partial D) > r_0$, there exists $t_0(r_0, b, d, \beta) > 0$, such that for any $t < t_0$, we have

$$\begin{cases} (4\pi t)^{-d/2} e^{-\frac{c_4^2 r_0^2}{4t}} + \frac{M_b t}{c_6 r_0^{d+\beta}} < \frac{c_1}{c_3} t^{-d/2} \wedge (m_b t)^{-d/\beta} \\ (4\pi t)^{-d/2} e^{-\frac{c_4^2 r_0^2}{4t}} + \frac{M_b t}{c_6 r_0^{d+\beta}} < t^{-d/2} \wedge (M_b t)^{-d/\beta}. \end{cases} \tag{4.15}$$

Note that $(4\pi t)^{-d/2} e^{-\frac{c_4^2 r_0^2}{4t}} + \frac{M_b t}{c_6 r_0^{d+\beta}}$ is increasing in t, therefore we have

$$k_D^b(t,x,y) < c_3 (4\pi t)^{-d/2} e^{-\frac{c_4^2 r_0^2}{4t}} + \frac{c_3 M_b t}{c_6 r_0^{d+\beta}}. \tag{4.16}$$

Let Γ be the straight line in D connecting x_0, y_0 with

$$\rho(\Gamma, \partial D) := 2r_0 > 0.$$

We divide Γ into n segments $a_0 = x_0, a_1, \ldots, a_n, a_{n+1} = y_0$ with $a_i = \frac{n-i}{n} x_0 + \frac{i}{n} y_0$. So we have $|a_i - a_{i-1}| = \frac{|x_0 - y_0|}{n} < \frac{2R}{n}$, $i = 1, 2, \ldots, n$. We choose a sufficiently large integer n such that $\frac{t}{n} < t_0$ and $\frac{2R}{n} < r_0$. Then for any $x_i \in B(a_i, \frac{2R}{n})$, we have $|x_i - x_{i-1}| < |x_i - a_i| + |a_i - a_{i-1}| + |a_{i-1} - x_{i-1}| < \frac{6R}{n}$ and $\rho(x_i, \partial D) \geq 2r_0 - \frac{2R}{n} > 2r_0 - r_0 = r_0$. Next we would like to show $p_D^b(\frac{t}{n}, x_{i-1}, x_i) > 0$ for any $i = 1, 2, \ldots, n$ with n sufficiently large. By (4.4), (4.15) and (4.16), we have

$$\begin{aligned} p_D^b(\frac{t}{n}, x_{i-1}, x_i) &> c_1 (\frac{4\pi t}{n})^{-d/2} e^{-\frac{n c_2^2 |x_{i-1} - x_i|^2}{4t}} - c_3 (\frac{4\pi t}{n})^{-d/2} e^{-\frac{n c_4^2 r_0^2}{4t}} - \frac{c_3 M_b t}{n c_6 r_0^{d+\beta}} \\ &> c_1 (\frac{4\pi t}{n})^{-d/2} e^{-\frac{9 c_2^2 R^2}{nt}} - c_3 (\frac{4\pi t}{n})^{-d/2} e^{-\frac{n c_4^2 r_0^2}{4t}} - \frac{c_3 M_b t}{n c_6 r_0^{d+\beta}} \\ &= (\frac{4\pi t}{n})^{-d/2} \left(c_1 e^{-\frac{9 c_2^2 R^2}{nt}} - c_3 e^{-\frac{n c_4^2 r_0^2}{4t}} - c_3 \frac{M_b t^{1+d/2}}{c_6 r_0^{d+\beta} n^{1+d/2}} \right). \end{aligned} \tag{4.17}$$

Obviously, the last term in (4.17) will be positive as $n \to \infty$. Therefore, we can show that there exists $N_0(r_0, t, R, b, d, \beta) \in \mathbb{N}^+$, such that for any $n > N_0$ we have

$$p_D^b(\frac{t}{n}, x_{i-1}, x_i) > 0, \qquad \text{for any } i = 1, 2, \ldots, n.$$

Then by the semigroup property and continuity of $p_D^b(t,x,y)$, we have

$$\begin{aligned}
p_D^b(t,x_0,y_0) &= \int_D \cdots \int_D p_D^b(\frac{t}{n},x_0,x_1)p_D^b(\frac{t}{n},x,x_1)\dots p_D^b(\frac{t}{n},x_n,y_0)dx_1\dots dx_n \\
&> \int_{B(a_1,\frac{2R}{n})} \cdots \int_{B(a_n,\frac{2R}{n})} p_D^b(\frac{t}{n},x_0,x_1)p_D^b(\frac{t}{n},x,x_1)\dots p_D^b(\frac{t}{n},x_n,y_0)dx_1\dots dx_n \\
&> 0,
\end{aligned}$$

which finishes the proof. □

Then we use a similar chaining argument as in Theorem 4.2.6, the positivity of $p_D^b(t,x,y)$ can be extended to any domain.

Theorem 4.2.8. *Let D be any domain in $\mathbb{R}^d$, then $p_D^b(t,x,y)>0$ for any $x,y\in D$ and $t>0$.*

Proof. Similar to Theorem 4.2.8, there exists a curve Γ in D connecting x,y with

$$\rho(\Gamma,\partial D)=2r_0>0.$$

We can find $n+1$ points, $a_0=x,a_1,\dots,a_n,a_{n+1}=y$ on Γ such that $|a_i-a_{i-1}|<r_0$. For any $x_i\in B(a_i,r_0)$, we have $\rho(x_i,\partial D)\geq 2r_0-r_0>r_0$, which indicates $x_i\in D$. Moreover, we have $|x_i-a_{i-1}|<2r_0$ and $B(a_{i-1},2r_0)\subset D$. By Lemma 4.2.7 and the domain monotonicity of $p_D^b(t,x,y)$, we have $p_D^b(\frac{t}{n},x_{i-1},x_i)>0$. Finally, using the semigroup property and the continuity of $p_D^b(t,x,y)$, we have

$$\begin{aligned}
p_D^b(t,x,y) &= \int_D \cdots \int_D p_D^b(\frac{t}{n},x,x_1)p_D^b(\frac{t}{n},x,x_1)\dots p_D^b(\frac{t}{n},x_n,y)dx_1\dots dx_n \\
&> \int_{B(a_1,r_0)} \cdots \int_{B(a_n,r_0)} p_D^b(\frac{t}{n},x,x_1)p_D^b(\frac{t}{n},x,x_1)\dots p_D^b(\frac{t}{n},x_n,y)dx_1\dots dx_n \\
&> 0.
\end{aligned}$$

The proof is completed. □

Finally, by either Theorem 4.2.6 or Theorem 4.2.8, we showed that $p_D^b(t,x,y)$ is strictly positive in any domain. To show its positivity on any open set, we need to use the Lévy system of the process X^b and condition (4.7).

Theorem 4.2.9. *If condition (4.7) holds, $p_D^b(t,x,y)$ is strictly positive on any open subset $D\subset\mathbb{R}^d$.*

Proof. The idea of the proof is almost the same as [17, Corollary 3.6]. For $x \in D$, let $D(x)$ be the connected component of D that contains x. If $y \in D(x)$, then by Theorem 4.2.6 and the domain monotonicity of the transition density, we have

$$p^b_D(t,x,y) \geq p^b_{D(x)}(t,x,y) > 0.$$

If $y \notin D(x)$, the Lévy system $J^b(x,y) = \frac{b(x,x-y)}{|x-y|^{d+\beta}}$ is strictly positive for any $x, y \in \mathbb{R}^d$ by (4.7). Therefore, using the strong Markov property of X^b, we have

$$\begin{aligned}
p^b_D(t,x,y) &= \mathbb{E}_x[p^b_D(t-\tau^b_{D(x)}, X^b_{\tau^b_{D(x)}}, y); \tau^b_{D(x)} < t] \\
&\geq \mathbb{E}_x[p^b_{D(x)}(t-\tau^b_{D(x)}, X^b_{\tau^b_{D(x)}}, y); \tau^b_{D(x)} < t, X^b_{\tau^b_{D(x)}} \in D(y)] \\
&> \int_0^t \int_{D(x)} p^b_{D(x)}(s,x,z) \left(\int_{D(y)} J^b(z,w) p^b_{D(y)}(t-s,w,y) dw \right) dz ds \\
&> 0.
\end{aligned}$$

The proof is completed. □

The next lemma shows the exponential decay (with respect to t) of the killed heat kernel $p^b_D(t,x,y)$.

Lemma 4.2.10. *There exist positive constants $C_i(d,\beta,diam(D)) > 0$, $i = 1,2$, such that*

$$p^b_D(t,x,y) \leq C_1 e^{-C_2 t}, \qquad \text{for any } (t,x,y) \in (1,\infty) \times D \times D.$$

Proof. Let $L := diam(D)$. It follows from the lower bound of (4.4) and (4.1) that

$$\begin{aligned}
\mathbb{P}_x(\tau^b_D \leq 1) &\geq \mathbb{P}_x(X^b_1 \notin D) \\
&= \int_{D^c} p^b(1,x,y) dy \\
&\geq c_1 \int_{D^c} \left[\left(1 \wedge (m_b)^{-d/\beta}\right) \wedge \left((4\pi)^{-d/2} e^{-\frac{c_2^2 |x-y|^2}{4}} + \frac{m_b}{c_6 |x-y|^{d+\beta}} \right) \right] dy,
\end{aligned}$$

for any $x \in D$, $D \subset B(x,L)$. When $m_b > 0$, we have

$$\mathbb{P}_x(\tau^b_D \leq 1) \geq c_1 \int_{|z|>L} \left[\left(1 \wedge (m_b)^{-d/\beta}\right) \wedge \frac{m_b}{c_5 |z|^{d+\beta}} \right] dz > 0.$$

When $m_b = 0$, we have

$$\mathbb{P}_x(\tau_D^b \le 1) \ge c_1 \int_{|z|>L} \left(1 \wedge (4\pi)^{-d/2} e^{-\frac{c_2^2|z|^2}{4}}\right) dz > 0.$$

In either case, we can conclude

$$\sup_{x\in D} \int_D p_D^b(1,x,y)dy = \sup_{x\in D} \mathbb{P}_x(\tau_D^b > 1) < 1.$$

We would like to prove that there exist positive constants $c_3, c_4 > 0$ such that

$$\int_D p_D^b(t,x,y)dy \le c_3 e^{-c_4 t}, \qquad \text{for } (t,x) \in (0,\infty) \times D.$$

To prove this, set $e^{-c_4} := \sup_{x\in D} \mathbb{P}_x(\tau_D^b > 1) < 1$ and $c_3 := e^{c_4}$. Assuming $t \in \mathbb{N}^+$, using the Chapman-Kolmogorov equation (4.12) for $(t-1)$ times, we have

$$\begin{aligned}\int_D p_D^b(t,x,y)dy &= \int_D \cdots \int_D p_D^b(1,x,x_1)p_D^b(1,x_1,x_2)\cdots p_D^b(1,x_{t-1},y)dx_1dx_2\cdots dx_{t-1}dy \\ &\le e^{-c_4 t}.\end{aligned}$$

If t is not a positive integer, using the previous inequality, we have

$$\begin{aligned}\int_D p_D^b(t,x,y)dy &= \int_D \int_D p_D^b(\lfloor t\rfloor,x,z)p_D^b(t-\lfloor t\rfloor,z,y)dzdy \\ &\le \int_D p_D^b(\lfloor t\rfloor,x,z)dz \le e^{-c_4\lfloor t\rfloor} \\ &= e^{c_4(t-\lfloor t\rfloor)}e^{-c_4 t} \le c_3 e^{-c_4 t}.\end{aligned}$$

By the upper bound of p^b in (4.4), there exists $c_5 > 0$ such that $p_D^b(1,x,y) \le p^b(1,x,y) \le c_5$ for any $(x,y) \in D \times D$. Therefore, for any $(t,x,y) \in (0,\infty) \times D \times D$, we have

$$\begin{aligned}p_D^b(t,x,y) &= \int_D p_D^b(t-1,x,z)p_D^b(1,z,y)dz \\ &\le c_5 \int_d p_D^b(t-1,x,z)dz \le c_3c_5e^{-c_4(t-1)},\end{aligned}$$

which completes the proof.

□

To conclude this subsection, we prove the existence and continuity theorem of the Green function for the killed process and provide upper bounds for the Green function.

Theorem 4.2.11. *The Green function* $G_D^b(x,y) := \int_0^\infty p_D^b(t,x,y)dt$ *exists and it is continuous off the diagonal. Moreover, there exists a constant* $C_1 = C_1(d, A, \beta, diam(D)) > 0$*, such that*

- *If* $d \geq 3$, $G_D^b(x,y) < C_1 \frac{1}{|x-y|^{d-2}}$.
- *If* $d = 2$, $G_D^b(x,y) < C_1 \ln(\frac{1}{|x-y|}) \vee 1$.
- *If* $d = 1$, $G_D^b(x,y) < C_1$.

Proof. By Lemma 4.2.2, (4.4) and the domain monotonicity of the heat kernel, we have for $0 < t < 1$,

$$\begin{aligned}
p_D^b(t,x,y) &\leq p^b(t,x,y) \\
&\leq c_3 t^{-\frac{d}{2}} \wedge (M_b t)^{-\frac{d}{\beta}} \wedge \left(\frac{1}{(4\pi t)^{d/2}} e^{-\frac{c_4^2|x-y|^2}{4t}} + \frac{M_b t}{c_6|x-y|^{d+\beta}} \right) \\
&\leq c_3 t^{-\frac{d}{2}} \wedge (M_b t)^{-\frac{d}{\beta}} \wedge \left(c_0 t^{-\frac{d}{2}} \wedge \frac{t}{|x-y|^{d+2}} + \frac{M_b t}{c_6|x-y|^{d+\beta}} \right) \\
&\leq c_3 t^{-\frac{d}{2}} \wedge (M_b t)^{-\frac{d}{\beta}} \wedge \left(\frac{t}{|x-y|^{d+2}} (c_0 + \frac{M_b}{c_6}|x-y|^{2-\beta}) \right) \\
&\leq c_3 t^{-\frac{d}{2}} \wedge (M_b t)^{-\frac{d}{\beta}} \wedge \left(\frac{t}{|x-y|^{d+2}} (c_0 + \frac{M_b}{c_6} diam(D)^{2-\beta}) \right) \\
&\leq c_3 t^{-\frac{d}{2}} \wedge (M_b t)^{-\frac{d}{\beta}} \wedge \frac{c_0 t}{|x-y|^{d+2}} \\
&\leq c_3 t^{-\frac{d}{2}} \wedge \frac{t}{|x-y|^{d+2}}.
\end{aligned}$$

Together with Lemma 4.2.10, we have for $d \geq 3$,

$$
\begin{aligned}
G_D^b(x,y) &= \int_0^\infty p_D^b(t,x,y)dt \\
&< c_3 \int_0^1 t^{-\frac{d}{2}} \wedge \frac{t}{|x-y|^{d+2}} dt + c_1 \int_1^\infty e^{-c_2 t} dt \\
&= c_3 \int_0^{|x-y|^2 \wedge 1} \frac{t}{|x-y|^{d+2}} dt + c_3 \int_{|x-y|^2 \wedge 1}^1 t^{-\frac{d}{2}} dt + \frac{c_1}{c_2} \\
&= \frac{c_1}{c_2} - c_3(\frac{d}{2} - 1) + \frac{c_3}{2} \frac{1}{|x-y|^{d-2}} \wedge \frac{1}{|x-y|^{d+2}} + c_3(\frac{d}{2} - 1) \frac{1}{|x-y|^{d-2}} \vee 1 \\
&< c_1 + c_2 \frac{1}{|x-y|^{d-2}} \\
&= \frac{1}{|x-y|^{d-2}} (c_2 + |x-y|^{d-2}) \\
&< \frac{1}{|x-y|^{d-2}} (c_2 + diam(D)^{d-2}) \\
&< \frac{c_1}{|x-y|^{d-2}}.
\end{aligned}
$$

Similarly, for $d = 2$, we have

$$
\begin{aligned}
G_D^b(x,y) &= \int_0^\infty p_D^b(t,x,y)dt \\
&\leq c_3 \int_0^1 t^{-\frac{d}{2}} \wedge \frac{t}{|x-y|^{d+2}} dt + c_1 \int_1^\infty e^{-c_2 t} dt \\
&= \frac{c_1}{c_2} + \frac{c_3}{2} 1 \wedge \frac{1}{|x-y|^{d+2}} - c_3 \ln(|x-y|^2 \wedge 1) \\
&\leq c_1 + c_2 \ln(\frac{1}{|x-y|} \vee 1) \\
&\leq c_1 \ln(\frac{1}{|x-y|}) \vee 1,
\end{aligned}
\tag{4.18}
$$

and for $d = 1$, we have

$$
\begin{aligned}
G_D^b(x,y) &= \int_0^\infty p_D^b(t,x,y)dt \\
&\leq c_3 \int_0^1 t^{-\frac{d}{2}} \wedge \frac{t}{|x-y|^{d+2}} dt + c_1 \int_1^\infty e^{-c_2 t} dt \\
&= \frac{c_1}{c_2} + \frac{c_3}{2} |x-y| \wedge \frac{1}{|x-y|^3} + 2c_3 |x-y| \wedge 1 \\
&\leq c_1.
\end{aligned}
$$

The above inequalities imply the existence and finiteness of the Green function. The continuity of $G_D^b(x,y)$ follows from the continuity of $p_D^b(t,x,y)$ and the dominated convergence theorem.

□

4.2.3 The dual process under the reference measure

From now on, let E be an arbitrary bounded open set in $\mathbb{R}^d$. As we have seen in the last subsection, the killed process $X^{b,E}$ of X^b in E has nice properties: its transition density is continuous and strictly positive; the Green function of $X^{b,E}$ exists. It is, however, not guaranteed that the dual process of $X^{b,E}$ exists. We have seen in the first chapter, one prerequisite of the intrinsic ultracontractivity of a semigroup is the existence of its dual semigroup. To resolve this issue, a change of measure is considered: even though $X^{b,E}$ does not necessarily have a dual process under the Lebesgue measure, we are able to find its dual process $\widehat{X}^{b,E}$ under a certain reference measure. To construct the reference measure, the Green function of $X^{b,E}$ is used. One way of finding such a reference measure was introduced by Chen, Kim, and Song ([17]). Under the reference measure, we are able to consider the dual process and the intrinsic ultracontractivity of the semigroup of $X^{b,E}$. In the rest of this section, we discuss some basic properties of $X^{b,E}$ and its dual process under the reference measure. As in [17, Proposition 5.1], we can see that the killed process $X^{b,E}$ is a Hunt process which satisfies the strong Feller property. Similar to [17], we introduce the following change of measure using the Green function of X^b_E:

$$h_E(x) := \int_E G^b_E(y,x)dy \qquad \text{and} \qquad \xi_E(dx) := h_E(x)dx.$$

Note that the existence of $h_E(x)$ is guaranteed by the local integrability of the Green function $G^b_E(y,x)$, which comes from Theorem 4.2.11.

Theorem 4.2.12. *h_E is a strictly positive, bounded and continuous function on E. For any nonnegative Borel function f, we have*

$$\int_E f(x)\xi_E(dx) \geq \int_E \mathbb{E}_x[f(X^{b,E}_t)]\xi_E(dx).$$

Proof. The strict positivity of $h_E(x)$ directly follows from the strict positivity of the heat kernel and the Green function. For the continuity, we would like to show

$$\lim_{x_n\to x}\int_E G^b_E(x_n,y)dy = \int_E G^b_E(x,y)dy.$$

For any sufficiently small $\delta > 0$, let $B(x, \delta)$ be a ball centered any x which also belongs to E. Then we have

$$
\begin{aligned}
&|\int_E G_E^b(x_n, y)dy - \int_E G_E^b(x, y)dy| \\
&< \int_{E\setminus B(x,\delta)} |G_E^b(x_n, y) - G_E^b(x, y)|dy + \int_{B(x,\delta)} G_E^b(x, y)dy + \int_{b(x,\delta)} G_E^b(x_n, y)dy. \qquad (4.19)
\end{aligned}
$$

We first consider the first in the last inequality. Since $\lim_{n\to\infty} x_n = x$, we have $x_n \in B(x, \frac{\delta}{2})$ for sufficiently large n. By Theorem 4.2.11, we have

$$
G_E^b(x_n, y) < \begin{cases} c_1 \frac{1}{(\delta/2)^{d-2}}, & \text{if } d \geq 3 \\ c_1 \ln(\frac{2}{\delta}), & \text{if } d = 2 \\ c_1, & \text{if } d = 1 \end{cases}
$$

and

$$
G_E^b(x, y) < \begin{cases} c_1 \frac{1}{(\delta)^{d-2}}, & \text{if } d \geq 3 \\ c_1 \ln(\frac{1}{\delta}), & \text{if } d = 2 \\ c_1. & \text{if } d = 1 \end{cases}
$$

Therefore, by the dominated convergence theorem and the continuity of $G_E^b(x, y)$, we have

$$
\limsup_{n\to\infty} \int_{E\setminus B(x,\delta)} |G_E^b(x_n, y) - G_E^b(x, y)|dy = 0. \qquad (4.20)
$$

For the second term in (4.19), also by Theorem 4.2.11, we have

$$
\int_{B(x,\delta)} G_E^b(x, y)dy < \begin{cases} c_1 \int_{B(x,\delta)} \frac{1}{(|x-y|)^{d-2}} = \frac{d\delta^2}{2}|B(0, 1)|, & \text{if } d \geq 3 \\ c_1 \int_{B(x,\delta)} \ln(\frac{1}{|x-y|})dy = \delta(\delta \ln(\frac{1}{\delta}) + \frac{\delta}{2})|B(0, 1)|, & \text{if } d = 2 \\ c_1 \int_{B(x,\delta)} dy = 2c_1\delta, & \text{if } d = 1. \end{cases} \qquad (4.21)
$$

No matter $d = 1$, $d = 2$ or $d \geq 3$, we have $\int_{B(x,\delta)} G_E^b(x, y)dy = o(\delta)$ as $\delta \to 0$. Finally, we consider the last term in (4.19). It can be bounded similarly to the second term. Since $\lim_{n\to\infty} x_n = x$, we have $B(x, \delta) \subset B(x_n, 2\delta)$ for sufficiently large n. Similar to the previous arguments, we can derive

$$
\limsup_{n\to\infty} \int_{B(x,\delta)} G_E^B(x_n, y)dy < \limsup_{n\to\infty} \int_{B(x_n,2\delta)} G_E^B(x_n, y)dy = o(\delta). \qquad (4.22)
$$

Therefore, combining (4.20), (4.21) and (4.22), we get

$$\lim_{\delta\to 0}\limsup_{n\to\infty}|\int_E G_E^b(x_n,y)dy - \int_E G_E^B(x,y)dy| = 0.$$

The continuity of $G_E^B(x,y)$ follows from the previous equation.

To prove the boundedness of $h_E(x)$, we have

$$h_E(x) = \int_E G_E^b(x,y)dy < \begin{cases} c_1\int_E \frac{1}{(|x-y|)^{d-2}} < c_1\int_{B(x,diam(E))}\frac{1}{(|x-y|)^{d-2}} < \infty, & \text{if } d\geq 3\\ c_1\int_E \ln(\frac{1}{|x-y|})\vee 1dy < c_1\int_{B(x,diam(E))}\ln(\frac{1}{|x-y|})\vee 1dy < \infty, & \text{if } d=2\\ c_1\int_E dy = c_1|E|\delta, \quad \text{if } d=1. \end{cases}$$

Then it is easy to see $h_E(x)$ is bounded.

For the last claim in this theorem, by the Chapman-Kolmogorov equation (4.12), we have

$$\begin{aligned}\int_E \mathbb{E}_y[f(X_t^{b,E})]G_E^b(x,y)dy =& \mathbb{E}_x\int_0^\infty \mathbb{E}_{X_x^{b,E}}[f(X_t^{b,E})]ds\\ =& \int_0^\infty \mathbb{E}_x[f(X_{t+s}^{b,E}]ds\\ =& \int_0^\infty f(y)p_E^b(t+s,x,y)dyds\\ \leq& \int_E f(y)G_E^b(x,y)dy, \qquad \text{for any Borel function } f\geq 0 \text{ and } x\in E.\end{aligned}$$

Integrating both sides of the last inequality with respect to x, we get

$$\int_E \mathbb{E}_x[f(X_t^{b,E})]h_E(y)dy \leq \int_E f(x)h_E(x)dx,$$

which proves the last claim of the theorem. □

It is easy to see that under the reference measure $\xi_E(dx)$, the transition density and the Green function of process $X^{b,E}$ are translated to

$$\overline{p}_E^b(t,x,y) := \frac{p_E^b(t,x,y)}{h_E(y)},$$

and

$$\overline{G}_E^b(x,y) := \int_0^\infty \overline{p}_E^b(t,x,y)dt = \frac{G_E^b(x,y)}{h_E(y)}.$$

The Green function $\overline{G}_E^b(t,x,y)$ also satisfies property **(A1) - (A5)** listed in [17, p2513] as well as [17, Theorem 5.4]. These facts, as stated in [17, Theorem 5.5], show that $X^{b,E}$ has a Hunt process as a dual process. We summarize these facts without proof in the following theorem, which was proved by Chen, Kim and Song in [17, Theorem 5.5].

Theorem 4.2.13 (Chen et al. (2012))**.** *The killed process $X^{b,E}$ has a strong dual process $\widehat{X}^{b,E}$ with respect to the reference measure ξ_E. In other words, the transition density $\overline{p}^b_E(t,y,x)$ also forms a semigroup $\{\widehat{P}^E_t, t \geq 0\}$ with respect to ξ_E so that*

$$\widehat{P}^E_t f(x) = \int_E f(y)\overline{p}^b_E(t,y,x)\xi_E(dy) = \int_E f(y)p^b_E(t,y,x)\frac{h_E(y)}{h_E(x)}dy,$$

and therefore

$$\int_E f(x)P^E_t g(x)\xi_E(dx) = \int_E g(x)\widehat{P}^E_t f(x)\xi(dx), \qquad \text{for any } f,g \in L^2(E,\xi_E(dx)).$$

Another goal of this subsection is to determine the Lévy system of the process $X^{b,E}$ and of the dual process $\widehat{X}^{b,E}$ under the Lebesgue measure. We have seen in Theorem 4.2.13, however, the duality of $X^{b,E}$ and $\widehat{X}^{b,E}$ is under the reference measure $\xi_E(dx)$. Let $(\overline{N}^E, \overline{H}^E)$ be the Lévy system of $X^{b,E}$ with respect to the reference measure $\xi_E(dx)$. It is easy to see that

$$\begin{cases} \overline{N}^E(x,dy) := \frac{J^b(x,y)}{h_E(y)}\xi_E(dy), & \text{for } (x,y) \in E \times E, \\ \overline{N}^E(x,\partial) := \int_{E^c} J^b(x,y)dy, & \text{for } x \in E, \end{cases}$$

and $\overline{H}^E_t := t$. By the duality of Lévy system ([36]), the Lévy system $(\widehat{N}^E, \widehat{H}^E)$ of $\widehat{X}^{b,E}$ should satisfy

$$\widehat{H}^E_t = t$$

and

$$\widehat{N}^E(y,dx)\xi_E(dy) = \overline{N}^E(x,dy)\xi_E(dx).$$

Note that $J^b(x,y) = \frac{b(x,x-y)}{|x-y|^{d+\beta}}$, hence from the previous equation we can derive that

$$\widehat{N}^E(x,dy) = \frac{J^b(y,x)h_E(y)}{h_E(x)}dy.$$

To be noted that when we introduce the reference measure $\xi_E(dx)$, E is arbitrary. The bounded open set of our interest is D, which is given. First, we can choose a sufficiently large open ball E centered at the origin so that $D \subset \frac{1}{2}E$. Under the reference measure $\xi_E(dx)$, the killed process $X^{b,E}$ has a strong dual process $\widehat{X}^{b,E}$. By Theorem 4.2.13, the transition density of the dual process $\widehat{X}^{b,E}$ under the Lebesgue measure is given by

$$\widehat{p}^{b,E}(t,x,y) := \frac{p_E^b(t,y,x)h_E(y)}{h_E(x)}, \tag{4.23}$$

which is also strictly positive and jointly continuous. Then, we consider further killing the processes $X^{b,E}$ and $\widehat{X}^{b,E}$ upon leaving the bounded open set D. We denote the killed processes by $X^{b,E,D}$ and $\widehat{X}^{b,E,D}$, respectively. It is easy to see that $X^{b,E,D}$ is the same as $X^{b,D}$, which is the process obtained by directly killing X^b upon leaving D. Moreover, by [8, Corollary III.3.16] and [61, Theorem 2, Remark 2], $X^{b,D}$ and $\widehat{X}^{b,E,D}$ are still Hunt processes and are dual of each other. Let

$$\overline{p}_D^{b,E}(t,x,y) := \frac{p_D^b(t,x,y)}{h_E(y)}, \qquad \widehat{p}_D^{b,E}(t,x,y) := \frac{p_D^b(t,y,x)h_E(y)}{h_E(x)}, \qquad \text{for any } x,y \in D. \tag{4.24}$$

$\overline{p}^{b,E}(t,x,y)$ and $\widehat{p}_D^{b,E}(t,x,y)$ are the transition densities of $X^{b,D}$ with respect to the reference measure $\xi_E(dx)$ and of $\widehat{X}^{b,E,D}$ with respect to the Lebesgue measure, respectively. Both of them are strictly positive and jointly continuous due to the strict positivity of $p_D^b(t,x,y)$ and the continuity of $h_E(x)$. Since

$$M = \sup_{x,y\in\frac{1}{2}E} \frac{h_E(x)}{h_E(y)} < \infty, \tag{4.25}$$

both $\overline{p}_D^{b,E}(t,x,y)$ and $\widehat{p}_D^{b,E}(t,x,y)$ are bounded for any fixed t. This is the reason why we consider killing process X^b upon leaving a larger set E first. Directly killing the process X^b upon leaving D can result in getting unbounded $\widehat{p}_D^b(t,x,y)$. Let

$$P_t^{b,E,D}f(x) := \int_D \overline{p}_D^{b,E}(t,x,y)f(y)\xi_E(dy)$$

and

$$\widehat{P}_t^{b,E,D}f(x) := \int_D \overline{p}_D^{b,E}(t,y,x)f(y)\xi_E(dy) = \int_D \widehat{p}_D^{b,E}(t,x,y)f(y)dy$$

be the semigroups of the processes $X^{b,D}$ and $\widehat{X}^{b,E,D}$, respectively. Denote the corresponding infinitesimal generators of them on $L^2(D,\xi_E)$ by $\mathcal{L}_D^{b,E}$ and $\widehat{\mathcal{L}}_D^{b,E}$, respectively. The following proposition is directly obtained due to the boundedness of the transition densities $\overline{p}_D^{b,E}(t,x,y)$ and $\widehat{p}_D^{b,E}(t,x,y)$.

Proposition 4.2.14. *$\{P_t^{b,E,D} : t \geq 0\}$ and $\{\widehat{P}_t^{b,E,D} : t \geq 0\}$ are ultracontractive with respect to ξ_E. i.e., for* *any $t > 0$, there exists positive constant c_t such that*

$$\overline{p}_D^{b,E}(t,x,y) \le c_t < \infty, \qquad \text{for any } (x,y) \in D \times D.$$

Moreover, $\{P_t^{b,E,D} : t \ge 0\}$ and $\{\widehat{P}_t^{b,E,D} : t \ge 0\}$ are both strongly continuous contraction semigroups in $L^2(D, \xi_E(dx))$.

Proof. By (4.4), (4.1) and Lemma 4.2.10, we know that for any fixed $t > 0$, there exists $c_t > 0$, such that $p_D^b(t,x,y) \le c_t$ for any $(x,y) \in \mathbb{R}^d \times \mathbb{R}^d$. Also, $h_E(y)$ is bounded below in D. Therefore, we have

$$\overline{p}_D^{b,E}(t,x,y) = \frac{p_D^b(t,x,y)}{h_E(y)} \le c_t.$$

This proves the ultracontractivity.

The contraction property follows easily from the duality and Hölder's inequality: for any $f \in L^2(D, \xi_E)$,

$$\begin{aligned}
&\int_D |P_t^{b,E,D} f(x)|^2 \xi_E(dx) \\
=& \int_D \left| \int_D \overline{p}_D^{b,E}(t,x,y) f(y) \xi_E(dy) \right|^2 \xi_E(dx) \\
\le& \int_D \left(\int_D \overline{p}_D^{b,E}(t,x,y) \xi_E(dy) \right) \left(\int_D \overline{p}_D^{b,E}(t,x,y) f^2(y) \xi_E(dy) \right) \xi_E(dx) \\
=& \int_D f^2(y) \left(\int_D \overline{p}_D^{b,E}(t,x,y) \xi_E(dx) \right) \xi_E(dy) \\
=& \int_D f^2(y) \xi_E(dy).
\end{aligned}$$

For the contraction property of the dual semigroup $\{\widehat{P}_t^{b,E,D} : t \ge 0\}$, the proof follows the same logic. The proof of the strong continuity is very standard (See [42, Proposition 3.6]), so we omit the details. $\square$

Recalling the arguments in Chapter 1.1, a common eigenvalue $\lambda_0^{b,E,D}$ which lies on the top of both the spectrums of the generators $\mathcal{L}_D^{b,E}$ and $\widehat{\mathcal{L}}_D^{b,E}$ exists. Moreover, a normalized eigenfunction $\phi_D^{b,E}$ ($\psi_D^{b,E}$) of $\mathcal{L}_D^{b,E}$ ($\widehat{\mathcal{L}}_D^{b,E}$) associated with $\lambda_0^{b,E,D}$ can be chosen to be strictly positive and continuous on D. Since $\lambda^{b,E,D}$ is also the ground state eigenvalue of $\mathcal{L}^{b,D}$, it does not depend on E. We write it as $\lambda_0^{b,D}$ and for any $x \in D$ we have

$$e^{\lambda_0^{b,D} t} \phi_D^{b,E}(x) = \int_D p_D^b(t,x,z) \phi_D^{b,E}(z) dz, \qquad e^{\lambda_0^{b,D} t} \psi_D^{b,E}(x) = \int_D \widehat{p}_D^b(t,x,z) \psi_D^{b,E}(z) dz. \tag{4.26}$$

From now on, we denote the Green functions of the kernels $p_D^b(t,x,y)$ and $\widehat{p}_D^{b,E}(t,x,y)$ (with respect to the

Lebesgue measure) by $G^b_D(x,y)$ and $\widehat{G}^{b,E}_D(x,y)$. It is easy to see that

$$\widehat{G}^{b,E}_D(x,y) = \frac{G^b_D(y,x)h_E(y)}{h_E(x)}.$$

Integrating both sides of (4.26) with respect to t from 0 to ∞ , we have

$$-\frac{1}{\lambda_0^{b,D}}\phi^{b,E}_D(x) = \int_D G^b_D(t,x,z)\phi^{b,E}_D(z)dz, \qquad -\frac{1}{\lambda_0^{b,D}}\psi^{b,E}_D(x) = \int_D \widehat{G}^{b,E}_D(t,x,z)\psi^{b,E}_D(z)dz, \qquad \text{for any } x \in D. \tag{4.27}$$

4.2.4 Intrinsic ultracontractivity of the heat kernel

Recall the definition of intrinsic ultracontractivity in Definition 1.2.5. The semigroups $\{P^{b,E,D}_t : t \geq 0\}$ and $\{\widehat{P}^{b,E,D}_t : t \geq 0\}$ are intrinsically ultracontractive if we can prove

$$\overline{p}^{b,E}_D(t,x,y) \leq c_t \phi^{b,E}_D(x)\psi^{b,E}_D(y), \qquad \text{for any } x,y \in D \text{ and any } t > 0. \tag{4.28}$$

c_t is a positive constant which depends on t. In this subsection, we aim to prove (4.28). We follow a well-known scheme of establishing intrinsic ultracontractivity for processes with jumps (See, for example, [44, 43, 19, 17]). Essentially, the method relies on the Lévy systems of the processes and is related to condition (4.7). As we shall also see in the case of α-stable processes, the key fact that the jump intensities of the processes have positive lower bounds on compact sets guarantees the intrinsic ultracontractivity of their semigroups. Lemma 4.2.17 is of great importance in the scheme. Similar lemmas also exist in the proofs of [43, 17] ([43, Lemma 3.2] and [17, Theorem 8.2]). After we establish Lemma 4.2.17, other steps of the well-known scheme follow from it. For the completeness of the thesis, we give the detailed proof as follows.

Choose an arbitrary point $x_0 \in D$ and a sufficiently small $r_0 \in (0,\infty)$ such that $B(x_0,r_0) \subset \overline{B(x_0,r_0)} \subset D$. For any fixed $a \in (0,1)$, we put $B_0 := B(x_0, ar_0/2)$, $C_1 := \overline{B(x_0,ar_0)}$ and $B_2 := B(x_0,r_0)$. The constant a is to be determined by the following lemma, which is a prerequisite for establishing intrinsic ultracontractivity.

Lemma 4.2.15. *There exists a constant $a \in (0,1)$, such that for any $t \in [\frac{a^2r_0^2}{2}, a^2r_0^2]$, there exists a constant $c > 0$ (depending on a), such that*

$$p^b_{B_2}(t,x,y) > \frac{c}{r_0^d}, \qquad \textit{for any } x,y \in C_1,$$

where $p^b_{B_2}(t,x,y)$ is the transition density of the killed process of X^b upon leaving set B_2.

Proof. Similar to (4.9) and (4.10), the transition density $p^b_{B_2}(t,x,y)$ is given by

$$p^b_{B_2}(t,x,y) := p^b(t,x,y) - \mathbb{E}_x[p^b(t-\tau^b_{B_2}, X^b_{\tau^b_{B_2}}, y); \tau^b_{B_2} < t].$$

By the heat kernel estimates lower bounds in (4.4) and (4.1), we choose such a sufficiently small a that, for any $t \in [\frac{a^2 r_0^2}{2}, a^2 r_0^2]$ we have

$$\begin{aligned} p^b(t,x,y) &> c_1 t^{-d/2} \wedge (m_b t)^{-d/\beta} \wedge p_0(t, c_2 x, c_2 y) \\ &> c_1 \frac{1}{t^{d/2}} \wedge (\frac{1}{m_b^{d/\beta}} \frac{1}{t^{d/\beta}}) \wedge \frac{c_3}{t^{d/2}} e^{-\frac{4c_2 a^2 r_0^2}{4t}} \\ &\geq c_4 \frac{1}{(ar_0)^d}, \qquad \text{for any } x,y \in C_1. \end{aligned} \tag{4.29}$$

Similarly, using the heat kernel estimates upper bounds in (4.4) and (4.1), we have

$$\begin{aligned} p^b(t-\tau^b_{B_2}, X^b_{\tau^b_{B_2}}, y) &< \frac{c_5}{t^{d/2}} e^{-\frac{c_6^2 (1-a)^2 r_0^2}{4t}} + \frac{M_b t}{c_7 (1-a)^{d+\beta} r_0^{d+\beta}} \\ &< c_5 \frac{1}{a^d r_0^d} e^{-\frac{c_6^2(1-a)^2}{a^2}} + \frac{M_b}{c_7} \frac{a^2}{(1-a)^{d+\beta}} \frac{1}{r_0^{d+\beta-2}}. \end{aligned} \tag{4.30}$$

Therefore,

$$p^b_{B_2}(t,x,y) > \frac{1}{r_0^d} \frac{1}{a^d} \left(c_3 - c_4 e^{-\frac{c_5(1-a)^2}{a^2}} - \frac{M_b}{c_6} r_0^{2-\beta} \frac{a^{d+2}}{(1-a)^{d+\beta}} \right).$$

It is obvious that the term in the bracket will be strictly positive when $a \to 0$. Therefore, we can choose such an a (depending r_0 and the heat kernel estimates) that $p^b_{B_2}(t,x,y)$ is bounded below by c/r_0^d. Of course, the constant c also depends r_0 and the heat kernel estimates. The proof is completed.

□

The previous lemma shows the existence of a sufficiently small closed ball C_1 such that the killed transition density $p^b_{B_2}(t,x,y)$ has a positive lower bound for a positive length of time within C_1. Now we fix such an a which makes this condition hold (hence the sets B_0, C_1, B_2 are fixed). The following lemma is one of the key facts we need to prove the intrinsic ultracontractivity.

Lemma 4.2.16. *There exists a constant $c > 0$, such that*

$$\inf_{y \in C_1} \mathbb{E}_y[\tau^{b,E}_{B_2}] > c \quad \textit{and} \quad \inf_{y \in C_1} \mathbb{E}_y[\widehat{\tau}^{b,E}_{B_2}] > c.$$

Proof. First, we prove the inequality for the original process. From Lemma 4.2.15 we have

$$\begin{aligned}
\inf_{y\in C_1}\mathbb{E}_y[\tau^{b,E}_{B_2}] &= \inf_{y\in C_1}\int_{B_2}\int_0^\infty p^{b,E}_{B_2}(t,x,y)dtdy \\
&> \inf_{y\in C_1}\int_{C_1}\int_0^\infty p^{b}_{B_2}(t,x,y)dtdy \\
&> \inf_{y\in C_1}\int_{C_1}\int_{\frac{a^2r_0^2}{2}}^{a^2r_0^2} p^{b}_{B_2}(t,x,y)dtdy \\
&> \inf_{y\in C_1}\int_{C_1}\int_{\frac{a^2r_0^2}{2}}^{a^2r_0^2} \frac{c_1}{r_0^d}dtdy \\
&> c_2.
\end{aligned}$$

For the dual process, due to the boundless of the ratio $\sup_{x,y\in B_2}\frac{h_E(x)}{h_E(y)}$, we have

$$\begin{aligned}
\inf_{y\in C_1}\mathbb{E}_y[\widehat{\tau}^{b,E}_{B_2}] &= \inf_{y\in C_1}\int_{B_2}\int_0^\infty \widehat{p}^{b,E}_{B_2}(t,x,y)dtdy \\
&= \inf_{y\in C_1}\int_{B_2}\int_0^\infty p^{b}_{B_2}(t,x,y)\frac{h_E(y)}{h_E(x)}dtdy \\
&\geq \frac{1}{M}\inf_{y\in C_1}\int_{B_2}\int_0^\infty p^{b}_{B_2}(t,x,y)dtdy \\
&= \frac{1}{M}\inf_{y\in C_1}\mathbb{E}_y[\tau^{b,E}_{B_2}] \\
&\geq \frac{c_2}{M}. \qquad (4.31)
\end{aligned}$$

The proof is completed. □

To be noted, the establishment of the previous Lemma in some related works (For example, [43]) is more straightforward. In [43], it was obtained by the separation property of Feller processes (See the remarks between Lemma 3.4 and Lemma 3.5 in [43]). The previous Lemma, however, becomes not so obvious in our case since $X^{b,E}$ and $\widehat{X}^{b,E}$ are not necessarily Feller processes (they are Hunt processes). Therefore, they no longer have the separation property. This is why we made an effort to establish Lemma 4.2.15 first.

As we mentioned previously, the following lemma is directly related to condition (4.7). It reflects when a process' jump intensity can guarantee its semigroup's intrinsic ultracontractivity.

Lemma 4.2.17. *There exists a positive constant $c = c(d,\beta,D,b) > 0$ such that for any $x\in D\setminus C_1$,*

$$\mathbb{P}_x\left(X^{b,E}_{\tau^{b,E}_{D\setminus C_1}}\in C_1\right)\geq c\mathbb{E}_x[\tau^{b,E}_{D\setminus C_1}] \qquad \textit{and} \qquad \mathbb{P}_x\left(\widehat{X}^{b,E}_{\widehat{\tau}^{b,E}_{D\setminus C_1}}\in C_1\right)\geq c\mathbb{E}_x[\widehat{\tau}^{b,E}_{D\setminus C_1}].$$

Proof. For any $z\in B_0$ and $y\in D\setminus C_1$, we have $|w-y| < diam(D)$. Let K, L be two compact sets such that

$D \subset K$ and $\bigcup_{y \in D\setminus C_1}(y - B_0) \subset L$ respectively. By condition (4.7), we have

$$
\begin{aligned}
&\mathbb{P}_x\left(X^{b,E}_{\tau^{b,E}_{D\setminus C_1}} \in C_1\right) \geq \mathbb{P}_x\left(X^{b,E}_{\tau^{b,E}_{D\setminus C_1}} \in B_0\right) \\
=&\int_{D\setminus C_1} G^b_{D\setminus C_1}(x,y)\int_{B_0} J^b(y,z)dzdy \\
=&\int_{D\setminus C_1} G^b_{D\setminus C_1}(x,y)\int_{B_0} \frac{b(y,y-z)}{|y-z|^{d+\beta}}dzdy \\
\geq&\frac{\inf_{y\in K}\inf_{z\in L} b(y,z)|B_0|}{diam(D)^{d+\beta}}\mathbb{E}_x[\tau^{b,E}_{D\setminus C_1}].
\end{aligned}
$$

Let $c_1 = \frac{\inf_{y\in K}\inf_{z\in L} b(y,z)|B_0|}{diam(D)^{d+\beta}}$ and we have proved the first part. For the dual process, similarly we have

$$
\begin{aligned}
&\mathbb{P}_x\left(\widehat{X}^{b,E}_{\widehat{\tau}^{b,E}_{D\setminus C_1}} \in C_1\right) \geq \mathbb{P}_x\left(\widehat{X}^{b,E}_{\widehat{\tau}^{b,E}_{D\setminus C_1}} \in B_0\right) \\
=&\int_{D\setminus C_1} \widehat{G}^{b,E}_{D\setminus C_1}(x,y)\int_{B_0} \frac{J^b(z,y)h_E(z)}{h_E(y)}dzdy \\
=&\int_{D\setminus C_1} \widehat{G}^{b,E}_{D\setminus C_1}(x,y)\int_{B_0} \frac{b(z,z-y)h_E(z)}{|z-y|^{d+\beta}h_E(y)}dzdy \\
\geq&\frac{\inf_{y\in K}\inf_{z\in L} b(y,z)|B_0|}{\sup_{y,z\in D}\frac{h_E(y)}{h_E(z)}diam(D)^{d+\beta}}\mathbb{E}_x[\widehat{\tau}^{b,E}_{D\setminus C_1}].
\end{aligned}
$$

Let $c_2 = \frac{\inf_{y\in K}\inf_{z\in L} b(y,z)|B_0|}{\sup_{y,z\in D}\frac{h_E(y)}{h_E(z)}diam(D)^{d+\beta}}$ and the second part of the lemma is also proved. To find one constant for both inequalities, we set $c = c_1 \wedge c_2$ and the proof is completed. □

The two lower bounds in the previous two lemmas enable us to establish the following lemma, which is the last step before we prove the intrinsic ultracontractivity of $\{P^{b,E,D}_t : t \geq 0\}$ and $\{\widehat{P}^{b,E,D}_t : t \geq 0\}$. In fact, after we obtained the previous lemma, the remaining steps to establish intrinsic ultracontractivity directly follow the well-known scheme used in [44, 43, 19, 17]. The following lemma adopts the same method used in [43, Lemma 3.5] and [17, Theorem 8.2].

Lemma 4.2.18. *There exists a constant $c = c(d, \beta, D, b) > 0$ such that*

$$\int_{B_2} G^b_D(x,y)dy \geq c\int_{D\setminus B_2} G^b_D(x,y)dy, \qquad \text{for any } x \in D.$$

and

$$\int_{B_2} \widehat{G}^{b,E}_D(x,y)dy \geq c\int_{D\setminus B_2} \widehat{G}^{b,E}_D(x,y)dy, \qquad \text{for any } x \in D.$$

Proof. Since the proofs for the two inequalities are similar, we only show the inequality for the original process. For notational convenience, we assume that the usual shift operators for the Markov process $X^{b,E}$ exist and we denote them by $\{\theta_t, t > 0\}$. Define a series of stopping times S_n and T_n, $n = 1, 2, \ldots$ as

$$S_1 := 0, \quad T_n := S_n + \tau^{b,E}_{D\setminus C_1} \circ \theta_{S_n},$$

on event $S_n < \tau^{b,E}_D$ and

$$S_{n+1} := T_n + \tau^{b,E}_{B_2} \circ \theta_{T_n},$$

on event $T_n < \tau^{b,E}_D$. From the Green function estimates in Theorem 4.2.11, we have

$$\mathbb{E}_x[\tau^{b,E}_D] = \int_D G^{b,E}_D(x,y)dy < \infty,$$

which implies $\mathbb{P}_x(\tau^{b,E}_D < \infty) = 1$. Then by the quasi-left continuity of the Hunt process $X^{b,E,D}$, we have

$$\mathbb{P}_x\left(\lim_{n\to\infty} S_n = \lim_{n\to\infty} T_n = \tau^{b,E}_D\right) = 1. \tag{4.32}$$

Note that $X^{b,E,D}_t \in B_2$ for $T_n < t < S_{n+1}$, hence by (4.32) we have

$$\begin{aligned}
&\int_{B_2} G^{b,E}_D(x,y)dy = \mathbb{E}_x\left[\int_0^{\tau^{b,E}_D} \mathbb{1}_{B_2}(X^{b,E,D}_t)dt\right] \\
=&\mathbb{E}_x\left[\sum_{n=1}^{\infty}\left(\int_{S_n}^{T_n} \mathbb{1}_{B_2}(X^{b,E,D}_t)dt + \int_{T_n}^{\widehat{S}_{n+1}} \mathbb{1}_{B_2}(X^{b,E,D}_t)dt\right)\right] \\
\geq&\mathbb{E}_x\left[\sum_{n=1}^{\infty}\left(\int_{T_n}^{S_{n+1}} \mathbb{1}_{B_2}(X^{b,E,D}_t)dt\right)\right] \\
=&\mathbb{E}_x\left[\sum_{n=1}^{\infty}(S_{n+1} - T_n)\right].
\end{aligned}$$

By the strong Markov property of $X^{b,E,D}$, Lemma 4.2.16 and Lemma 4.2.17, we have

$$\mathbb{E}_x[S_{n+1} - T_n] = \mathbb{E}_x\left[\mathbb{E}_{X^{b,E,D}_{T_n}}[\tau^{b,E}_{B_2}]; T_n < \tau^{b,E}_D\right] \geq c_1\mathbb{P}_x(X^{b,E,D}_{T_n} \in C_1) \geq c_1c_2\mathbb{E}_x[T_n - S_n].$$

Since $X^{b,E,D}_t \notin D \setminus B_2$ for $t \in [T_n, S_{n+1})$, by the previous two inequalities, (4.32) and Fubini's theorem, we can derive

$$\begin{aligned}
\int_{B_2} G_D^{b,E}(x,y)dy &\geq c_1 c_2 \mathbb{E}_x\left[\sum_{n=1}^{\infty}(T_n - S_n)\right] \\
&= c_1 c_2 \sum_{n=1}^{\infty} \mathbb{E}_x\left[\int_{S_n}^{T_n} \mathbb{1}_D(X_t^{b,E,D})dt\right] \\
&\geq c_1 c_2 \mathbb{E}_x\left[\sum_{n=1}^{\infty}\int_{S_n}^{T_n} \mathbb{1}_{D\setminus B_2}(X_t^{b,E,D})dt + \sum_{n=1}^{\infty}\int_{T_n}^{S_{n+1}} \mathbb{1}_{D\setminus B_2}(X_t^{b,E,D})dt\right] \\
&= c_1 c_2 \mathbb{E}_x\left[\int_0^{\tau_D^{b,E}} \mathbb{1}_{D\setminus B_2}(X_t^{b,E,D})dt\right] \\
&= c_1 c_2 \int_{D\setminus B_2} G_D^b(x,y)dy.
\end{aligned}$$

Then the proof is completed.

□

The previous lemma is of great importance in establishing intrinsic ultracontractivity. It indicates that the integration of the Green function of the process $X^{b,E,D}(\widehat{X}^{b,E,D})$ in the outer region of D is bounded by that in the inner region of D. Recall that all the ground state eigenfunctions $\psi_D^{b,E}$ and $\phi_D^{b,E}$ are continuous and strictly positive in D. Therefore, they can achieve their minimums in the compact set B_2. This fact enables us to prove the following main theorem.

Theorem 4.2.19. *The semigroups $\{P_t^{b,E,D} : t \geq 0\}$ and $\{\widehat{P}_t^{b,E,D} : t \geq 0\}$ are intrinsically ultracontractive (with respect to the reference measure $\xi_E(dx)$). It means that for any $t \geq 0$, there exists a constant $c_t > 0$ such that*

$$c_t^{-1}\phi_D^{b,E}(x)\psi_D^{b,E}(y) \leq \overline{p}_D^{b,E}(t,x,y) \leq c_t\phi_D^{b,E}(x)\psi_D^{b,E}(y), \qquad \text{for } x,y \in D.$$

Proof. By Lemma 4.2.18 and (4.27), there exists a constant $c_2 > 0$ such that

$$\begin{aligned}
\mathbb{E}_x[\tau_D^{b,E}] &= \int_{B_2} G_D^{b,E}(x,z)dz + \int_{D\setminus B_2} G_D^{b,E}(x,z)dz \\
&\leq c_1 \int_{B_2} G_D^{b,E}(x,z)dz \\
&\leq c_2 \int_{B_2} G_D^{b,E}(x,z)\psi_D^{b,E}(z)dz \\
&\leq c_2 \int_D G_D^b(z,y)\psi_D^{b,E}(z)dz \\
&= -\frac{c_2}{\lambda_0^{b,D}}\psi_D^{b,E}(y),
\end{aligned} \tag{4.33}$$

where the third inequality comes from the boundedness of $\psi_D^{b,E}(z)$ in D. Similar argument applies to the case of the dual process, from which we can derive

$$\mathbb{E}_x[\widehat{\tau}_D^{b,E}] \leq -\frac{c_2}{\lambda_0^{b,D}}\phi_D^{b,E}(y). \tag{4.34}$$

Using the semigroup property (4.12), the ultracontractivity of $X^{b,E,D}$, the duality of $X^{b,E,D}$ and $\widehat{X}^{b,E,D}$, and Markov's inequality, we have

$$\begin{aligned}
&\overline{p}_D^{b,E}(t,x,y)\\
=&\int_D\int_D \overline{p}_D^{b,E}(t/3,x,z)\overline{p}_D^{b,E}(t/3,z,w)\overline{p}_D^{b,E}(t/3,w,y)\xi_E(dw)\xi_E(dz)\\
\leq& c_t\int_D \overline{p}_D^{b,E}(t/3,x,z)\xi_E(dz)\int_D \overline{p}_D^{b,E}(t/3,w,y)\xi_E(dw)\\
=&c_t\mathbb{P}_x(\tau_D^{b,E}>t/3)\mathbb{P}_y(\widehat{\tau}_D^{b,E}>t/3)\leq c_t 9/t^2\mathbb{E}_x[\tau_D^b]\mathbb{E}_y[\widehat{\tau}_D^{b,E}].
\end{aligned}$$

Together with (4.33) and (4.34), it implies that

$$\overline{p}_D^{b,E}(t,x,y)\leq c_t(\frac{c_2}{\lambda_0^{b,D}})^2\psi_D^{b,E}(x)\phi_D^{b,E}(y).$$

The intrinsic ultracontractivities of $\{P_t^{b,E,D}: t\geq 0\}$ and $\{\widehat{P}_t^{b,E,D}: t\geq 0\}$ follow from the previous inequality.

□

4.3 Intrinsic ultracontracitivity of $\Delta^\alpha+\mathcal{S}^b$ in bounded open sets

In this section, we focus on the intrinsic ultracontractivity of $\Delta^\alpha+\mathcal{S}^b$ $(0<\alpha<2)$ in a bounded open set. As we mentioned at the beginning of this chapter, the way to establish it is the same as the case of $\Delta+\mathcal{S}^b$. The differences lie in the different heat kernel estimates and the different jump mechanisms of the two processes. Therefore, some of the intermediate steps in the previous proof need to be reproved in regard to different heat kernel estimates. Moreover, one key difference between the operator Δ and Δ^α is that: the process corresponding to the operator Δ (the standard Brownian motion) is a continuous sample path process; in contrast, the process corresponding to Δ^α (the α-stable process) is a pure jump process. The perturbation operator $\mathcal{S}^b$ is essentially adding a jump perturbation to the original process. Hence, the jump mechanism of the operator $\Delta+\mathcal{S}^b$ is determined by $\mathcal{S}^b$; while the jump mechanism of the operator $\Delta^\alpha+\mathcal{S}^b$

is determined by "the sum of the jump of Δ^α and the jump of $\mathcal{S}^b$". Since our proof depends on the Lévy system of the process, the condition on function $b(x,z)$ should be different. Instead of Condition (4.7), we will impose Condition (4.42) in the case of $\Delta^\alpha + \mathcal{S}^b$.

4.3.1 Preliminaries

Similar to the last section, we begin with reviewing some basic properties of the process X^b corresponding to the operator $\mathcal{L}^b := \Delta^\alpha + \mathcal{S}^b$ on $\mathbb{R}^d$. This class of operators were first introduced in [21]. The authors considered the fundamental solutions of the operators and obtained sharp two-sided estimates of the fundamental solution. Previously, [18] considered the heat kernel estimates for the operator $\Delta^\alpha + a\Delta^\beta$ with $0 < \beta < \alpha < 2$ and $a > 0$, which corresponds to the independent sum of an α-stable process and a scaled lower order β-stable process. Therefore, $\mathcal{L}^b$ could be seen as a generalization of the operators for an α-stable process perturbed by an independent scaled lower order β-stable process ($\mathcal{L}^b = \Delta^{\alpha/2} + a\Delta^{\beta/2}$ when $b(x,z) \equiv a$).

Let Y_t be a symmetric α-stable process in $\mathbb{R}^d$ and $\overline{Y}_t$ be the finite range symmetric α-stable process corresponding to Y_t: $\overline{Y}_t$ only has the jumps of Y_t whose size is less than 1. The infinitesimal generator of the process $\overline{Y}_t$ is the truncated fractional Laplacian:

$$\overline{\Delta}^{\alpha/2} f(x) := \int_{|z|<1} (f(x+z) - f(x) - \langle \nabla f(x), z \rangle) \frac{\mathcal{A}(d,-\alpha)}{|z|^{d+\alpha}} dz. \tag{4.35}$$

It was shown in [21] that if $b(x,z)$ satisfies

$$b(x,z) = b(x,-z), \qquad \text{for every } x \in \mathbb{R}^d, \text{ a.e } z \in \mathbb{R}^d, \tag{4.36}$$

and

$$b(x,z) \geq -\mathcal{A}(d,-\alpha)|z|^{\beta-\alpha}, \qquad \text{for every } x \in \mathbb{R}^d, \text{ a.e } z \in \mathbb{R}^d, \tag{4.37}$$

$\mathcal{L}^b$ uniquely determines a conservative Feller process X^b in $\mathbb{R}^d$ which has a strictly positive and continuous transition density $p^b(t,x,y)$. Sharp two-sided estimates of $p^b(t,x,y)$ were obtained in [21]. Later, [22] considered the killed process of X^b upon leaving a $C^{1,1}$ open set (see [22] for the definition) in $\mathbb{R}^d$ and obtained sharp two-sided estimates for the heat kernel and the Green function. By imposing additional assumptions on the region, the authors obtained the two-sided heat kernel estimates in terms of the distance functions. We are interested in the properties of the semigroup associated with the killed process of X^b in more general regions. Therefore, in this section, we consider the intrinsic ultracontractivity of the killed process of X^b upon leaving a bounded open set $D \subset \mathbb{R}^d$.

We first recall some basic definitions and results in [22, 21, 18, 15]. For any $a \geq 0$, let $p_a(t,x,y)$ be the fundamental solution of operator $\Delta^{\alpha/2} + a\Delta^{\beta/2}$ under the Lebesgue measure in $\mathbb{R}^d$. It was proved in [18] that for any $(t,x,y) \in (0,\infty) \times \mathbb{R}^d \times \mathbb{R}^d$,

$$p_a(t,x,y) \asymp (t^{-d/\alpha} \wedge (at)^{-d/\beta}) \wedge (\frac{t}{|x-y|^{d+\alpha}} + \frac{at}{|x-y|^{d+\beta}}). \tag{4.38}$$

For the truncated α-stable process $\overline{Y}_t$ in (4.35), we denote its transition density by $\overline{p}_0(t,x,y)$. It was shown in [15] that for $t \in (0,1)$ and $|x-y| \leq 1$,

$$\overline{p}_0(t,x,y) \asymp t^{-\frac{d}{\alpha}} \wedge \frac{t}{|x-y|^{d+\alpha}}, \tag{4.39}$$

and for $t \in (0,1)$ and $|x-y| > 1$,

$$c_1 \left(\frac{t}{|x-y|}\right)^{c_2|x-y|} \leq \overline{p}_0(t,x,y) \leq c_3 \left(\frac{t}{|x-y|}\right)^{c_4|x-y|}. \tag{4.40}$$

The constants $c_i = c_i(d,\alpha)$, $i = 1,\dots,4$, are strictly positive. To keep the same notation as [21], for each bounded function $b(x,z)$ and $\lambda > 0$ we define $b^+(x,z) = \max\{b(x,z), 0\}$ and

$$m_{b,\lambda} = \inf_x \text{ess inf}_{|z|>\lambda} b(x,z) \quad \text{and} \quad M_{b,\lambda} = \text{ess sup}_{x,|z|>\lambda} b(x,z).$$

The next theorem summarizes some of the key facts about the heat kernel $p^b(t,x,y)$. Readers are referred to [21, Theorem 1.1, 1.2, 1.3] for details.

Theorem 4.3.1 (Chen and Wang (2013))**.** *For every $A, \lambda > 0$, there are positive constants $C_k(d,\alpha,\beta,A) > 0, k = 1,2,3$ such that for any bounded function $b(x,z)$ satisfying* (4.36) *and* (4.37) *with $||b||_\infty \leq A$, we have*

$$C_1\overline{p}_0(t, C_2x, C_2y) \leq p^b(t,x,y) \leq C_3 p_{M_{b^+,\lambda}}(t,x,y), \qquad \textit{for } t \in (0,1) \textit{ and } x,y \in \mathbb{R}^d. \tag{4.41}$$

The heat kernel $p^b(t,x,y)$ uniquely determines a Feller process $X^b = \{X^b_t, t \geq 0, \mathbb{P}_x, x \in \mathbb{R}^d\}$ on the canonical Skorokhod space $\mathbb{D}([0,\infty), \mathbb{R}^d)$. The Feller process is conservative and has a Lévy system $(J^b(x,y)dy, t)$, where

$$J^b(x,y) = \frac{\mathcal{A}(d,-\alpha)}{|y-x|^{d+\alpha}}(1 + \frac{b(x,y-x)}{\mathcal{A}(d,-\alpha)}|y-x|^{\alpha-\beta}).$$

From the Lévy system of the perturbed process X^b, we observe that the function $b(x,z)$ here can be negative which intuitively means the jumps of the original α-stable are canceled. Similar to the case of the Laplacian under perturbation, we need to guarantee that the killed process $X^{b,D}$ can traverse every connected component of D. Therefore, the following assumption is an analog of condition (4.7): for any compact subset $K, L \in \mathbb{R}^d$,

$$\inf_{x\in K}\inf_{z\in L}\left(1+\frac{b(x,z)}{\mathcal{A}(d,-\alpha)}|z|^{\alpha-\beta}\right)>0. \tag{4.42}$$

We shall prove that condition (4.42) is strong enough to guarantee the intrinsic ultracontractivity of the killed process $X^{b,D}$. In the rest of this section, we follow the same scheme we used in the case of the perturbed Laplacian. Most of the intermediate steps are the same as the last subsection, so only the different steps needed in the proof are shown in this section.

4.3.2 Properties of the killed process

In this section, we use X^b to present the process corresponding to the operator $\Delta^\alpha + \mathcal{S}^b$ in $\mathbb{R}^d$ in Theorem 4.3.1. For any bounded open subset $D \subset \mathbb{R}^d$, define the killed process $X^{b,D}$ of X^b as (4.8), then its transition density $p^b_D(t,x,y)$ is given by (4.9) and (4.10). Use $\{P^{b,D}_t : t \geq 0\}$ to denote the semigroup of $X^{b,D}$ and $\mathcal{L}^b|_D$ to denote the infinitesimal generator of $\{P^{b,D}_t : t \geq 0\}$. Then with a few minor modifications, we can use the same method to establish Lemma 4.2.3 (It was also proved in [22, Lemma 3.1], so we omit the proof). As a result, we can obtain Theorem 4.2.4 which guarantees the continuity of the transition density $p^b_D(t,x,y)$.

Next, we consider the strict positivity of $p^b_D(t,x,y)$. In fact, [22] already obtained a stronger result with regard to this (See [22, Lemma 3.3, Proposition 3.4, Lemma 3.5]), which gives a local lower bound of $p^b_D(t,x,y)$ in a delicately chosen time interval. Here, we provide a simpler proof to account for the strict positivity of $p^b_D(t,x,y)$. The idea of proof aligns with what we used in Subsection 4.2.2. The following lemma is an analog of Lemma 4.2.5.

Lemma 4.3.2. *For any $A > 0$ with $||b||_\infty < A$. For any sufficiently small $r_0 > 0$, there exists*

$$t_0 = t_0(r_0, A, d, \alpha, \beta) > 0 \qquad \text{and} \qquad r_1 = r_1(r_0, A, d, \alpha, \beta) > 0$$

such that for any $x, y \in D$ and $t < t_0$, satisfying $r_0 < \rho(x,\partial D) \wedge \rho(y,\partial D)$ and $\rho(x,y) < r_1$, we have $p^b_D(t,x,y) > 0$.

Proof. For any $x, y \in D$ satisfying $\rho(x,\partial D) > r_0$ and $\rho(y,\partial D) > r_0$, we recall the definition of $p^b_D(t,x,y)$ in (4.9) and (4.10),

$$p^b_D(t,x,y) = p^b(t,x,y) - \mathbb{E}_x[p^b(t-\tau^b_D, X^b_{\tau^b_D}, y); \tau^b_D < t]. \tag{4.43}$$

For $0 < t < 1$ and $|x-y| < 1$, using the estimates in (4.41), we have

$$p^b(t,x,y) \geq c_1 \overline{p}_0(t, c_2 x, c_2 y) \geq c_1 t^{-\frac{d}{\alpha}} \wedge \frac{t}{|x-y|^{d+\alpha}},$$

and

$$\begin{aligned}
& p^b(t-\tau_D^b, X^b_{\tau_D^b}, y) \\
\leq & c_3 (t-\tau_D^b)^{-d/2} \wedge \left(A(t-\tau_D^b)\right)^{-d/\beta} \wedge \left(\frac{t-\tau_D^b}{|X^b_{\tau_D^b} - y|^{d+\alpha}} + \frac{A(t-\tau_D^b)}{|X^b_{\tau_D^b} - y|^{d+\beta}} \right) \\
\leq & c_3 (t-\tau_D^b)^{-d/2} \wedge \left(A(t-\tau_D^b)\right)^{-d/\beta} \wedge \left(\frac{t}{r_0^{d+\alpha}} + \frac{At}{r_0^{d+\beta}} \right).
\end{aligned} \tag{4.44}$$

It is easy to see that

$$\lim_{t\to 0} \left(\frac{t}{r_0^{d+\alpha}} + \frac{At}{r_0^{d+\beta}} \right) = 0,$$

and

$$\lim_{t\to 0} t^{-\frac{d}{\alpha}} = \infty \quad \text{and} \quad \lim_{t\to 0} (At)^{-\frac{d}{\beta}} = \infty.$$

Therefore, there exists a sufficiently small $t_0 = t_0(r_0, A, d, \alpha, \beta) > 0$, such that for any $t < t_0$, we have

$$\frac{t}{r_0^{d+\alpha}} + \frac{At}{r_0^{d+\beta}} < \min\{t^{-\frac{d}{\alpha}}, (At)^{-\frac{d}{\beta}}, \frac{c_1}{c_3} t^{-\frac{d}{\alpha}}\}. \tag{4.45}$$

In this case, we have

$$\begin{aligned}
& k_D^b(t,x,y) \\
= & \mathbb{E}_x[p^b(t-\tau_D^b, X^b_{\tau_D^b}, y); \tau_D^b < t] \\
\leq & \mathbb{E}_x[c_3 \left(\frac{t}{r_0^{d+\alpha}} + \frac{At}{r_0^{d+\beta}} \right); \tau_D^b < t] \\
\leq & c_3 \left(\frac{t}{r_0^{d+\alpha}} + \frac{At}{r_0^{d+\beta}} \right).
\end{aligned}$$

Then it is sufficient to show that

$$c_3 \left(\frac{t}{r_0^{d+\alpha}} + \frac{At}{r_0^{d+\beta}} \right) < c_1 \frac{t}{|x-y|^{d+\alpha}}.$$

It is obvious that the right-hand side will go to infinity when $|x-y| \to 0$. Hence, there exists a sufficiently small $r_1 = r_1(r_0, A, d, \alpha, \beta) > 0$, such that for any $x, y \in \mathbb{R}^d$ with $|x-y| < r_1$, we have

$$\begin{aligned} p_D^b(t,x,y) &= p^b(t,x,y) - k_D^b(t,x,y) \\ &> c_1 t^{-\frac{d}{\alpha}} \wedge \frac{t}{|x-y|^{d+\alpha}} - c_3\left(\frac{t}{r_0^{d+\alpha}} + \frac{At}{r_0^{d+\beta}}\right) \\ &> 0. \end{aligned}$$

The proof is completed. $\square$

Theorem 4.2.6 and 4.2.9 in the last section can also be proved in a similar manner. We prove Lemma 4.2.10 and Theorem 4.2.11 for the case of $\Delta^\alpha + \mathcal{S}^b$.

Lemma 4.3.3. *There exist constants $C_1(d, A, \alpha, \beta, diam(D)) > 0$ and $C_2(d, A, \alpha, \beta, diam(D)) > 0$ such that*

$$p_D^b(t,x,y) \le C_1 e^{-C_2 t}, \qquad \textit{for any } (t,x,y) \in (1,\infty) \times D \times D.$$

Proof. Let $L := diam(D) \vee 1$. It follows from the lower bound in (4.41) and (4.40) that, for any $x \in D$ we have

$$\begin{aligned} \mathbb{P}_x(\tau_D^b \le 1) &\ge \mathbb{P}_x(X_1^b \notin D) \\ &= \int_{D^c} p^b(1,x,y)dy \\ &\ge c_1 \int_{D^c} \left(\frac{1}{|x-y|}\right)^{c_2|x-y|} dy. \end{aligned}$$

For any $x \in D$, $D \subset B(x, L)$, we have

$$\begin{aligned} \mathbb{P}_x(\tau_D^b \le 1) &\ge c_1 \int_{|z|>L} \left(\frac{1}{|z|}\right)^{c_2|z|} dz \\ &= c_1 \int_L^\infty \frac{1}{r^{c_2 r}} dr > 0. \end{aligned} \tag{4.46}$$

Therefore,

$$\sup_{x\in D} \int_D p_D^b(1,x,y)dy = \sup_{x \in D} \mathbb{P}_x(\tau_D^b > 1) < 1.$$

The rest of the proof is exactly the same as Lemma 4.2.10, so we omit it and the proof is completed. $\square$

Theorem 4.3.4. *The Green function $G_D^b(x,y) := \int_0^\infty p_D^b(t,x,y)dt$ is finite for any $x, y \in D$ and it is continuous off the diagonal. Moreover, there exists a constant $C_3 = C_3(d, A, \alpha, \beta, diam(D)) > 0$, such that*

- *If $d \geq 2$, we have $G_D^b(x,y) \leq C_3|x-y|^{\alpha-d}$*
- *If $d = 1$ and $1 < \alpha < 2$, we have $G_D^b(x,y) \leq C_3$*
- *If $d = 1$ and $\alpha = 1$, we have $G_D^b(x,y) \leq C_3 \ln(\frac{1}{|x-y|}) \vee 1$*
- *If $d = 1$ and $0 < \alpha < 1$, we have $G_D^b(x,y) \leq C_3|x-y|^{\alpha-1}$*

Proof. For the case $d \geq 2$, it was proved in [22, Lemma 4.2]. We only give the proof for $d = 1$ where we use a similar argument to the one in Theorem 4.2.11. Using the upper bound in (4.41) and the domain monotonicity, we have

$$\begin{aligned}
p_D^b(t,x,y) \leq p^b(t,x,y) &\leq c_1(t^{-1/\alpha} \wedge (At)^{-1/\beta}) \wedge (\frac{t}{|x-y|^{1+\alpha}} + \frac{At}{|x-y|^{1+\beta}}) \\
&\leq c_3 t^{-\frac{1}{\alpha}} \wedge (At)^{-\frac{1}{\beta}} \wedge \left(\frac{t}{|x-y|^{1+\alpha}}(1 + A|x-y|^{\alpha-\beta})\right) \\
&\leq c_3 t^{-\frac{1}{\alpha}} \wedge (At)^{-\frac{1}{\beta}} \wedge \left(\frac{t}{|x-y|^{1+\alpha}}(1 + A \times diam(D)^{\alpha-\beta})\right) \\
&\leq c_3 t^{-\frac{1}{\alpha}} \wedge \frac{t}{|x-y|^{1+\alpha}}, \qquad \text{for } 0 < t < 1.
\end{aligned}$$

Together with Lemma 4.3.3, for $1 < \alpha < 2$, we have

$$\begin{aligned}
G_D^b(x,y) &= \int_0^\infty p_D^b(t,x,y)dt \\
&\leq c_3 \int_0^1 t^{-\frac{1}{\alpha}} \wedge \frac{t}{|x-y|^{1+\alpha}} dt + c_1 \int_1^\infty e^{-c_2 t} dt \\
&= c_3 \int_0^{|x-y|^\alpha \wedge 1} \frac{t}{|x-y|^{1+\alpha}} dt + c_3 \int_{|x-y|^\alpha \wedge 1}^1 t^{-\frac{1}{\alpha}} dt + \frac{c_1}{c_2} \\
&= \frac{c_1}{c_2} + \frac{c_3}{1-\alpha} + \frac{c_3}{2}|x-y|^{\alpha-1} \wedge \frac{1}{|x-y|^{1+\alpha}} - \frac{c_3}{1-\alpha}|x-y|^{\alpha-1} \wedge 1 \\
&\leq c_1 + c_3|x-y|^{\alpha-1} \\
&\leq c_1.
\end{aligned}$$

Similarly, for $d = 1$ and $\alpha = 1$, we have

$$\begin{aligned} G_D^b(x,y) &= \int_0^\infty p_D^b(t,x,y)dt \\ &< c_3 \int_0^1 t^{-1} \wedge \frac{t}{|x-y|^2} dt + c_1 \int_1^\infty e^{-c_2 t} dt \\ &= \frac{c_1}{c_2} + \frac{c_3}{2} 1 \wedge \frac{1}{|x-y|^2} - c_3 \ln(|x-y| \wedge 1) \\ &\le c_1 + c_3 \ln(\frac{1}{|x-y|} \vee 1) \\ &\le c_3 \ln(\frac{1}{|x-y|}) \vee 1. \end{aligned} \tag{4.47}$$

For $d = 1$ and $0 < \alpha < 1$, we have

$$\begin{aligned} G_D^b(x,y) &= \int_0^\infty p_D^b(t,x,y)dt \\ &< c_3 \int_0^1 t^{-\frac{d}{2}} \wedge \frac{t}{|x-y|^{d+2}} dt + c_1 \int_1^\infty e^{-c_2 t} dt \\ &= \frac{c_1}{c_2} + \frac{c_3}{2} \frac{1}{|x-y|^{1-\alpha}} \wedge \frac{1}{|x-y|^{1+\alpha}} + c_3(\frac{1}{\alpha} - 1) \frac{1}{|x-y|^{1-\alpha}} \vee 1 \\ &\le c_1 + c_3 \frac{1}{|x-y|^{1-\alpha}} \\ &\le c_3 \frac{1}{|x-y|^{1-\alpha}}. \end{aligned}$$

All the above inequalities indicate the existence of the Green functions. The continuity of $G_D^b(x,y)$ follows from the dominated convergence theorem. □

4.3.3 Intrinsic ultracontractivity of the heat kernel

We can construct the reference measure and find the dual process of $X^{b,D}$ in the same manner as in Subsection 4.2.3. It is easy to see that all the theorems and propositions in Subsection 4.2.3 also hold for the case of $\Delta^\alpha + \mathcal{S}^b$. As a result, we can obtain a pair of dual processes $X^{b,D}$ and $\widehat{X}^{b,E,D}$, together with their dual semigroups $\{P_t^{b,E,D} : t \ge 0\}$ and $\{\widehat{P}_t^{b,E,D} : t \ge 0\}$ with respect to the reference measure $\xi_E(dx)$ defined in Subsection 4.2.3.

The difference lies in the Lévy systems of the killed processes $X^{b,E}$ and $\widehat{X}^{b,E}$. By Theorem 4.3.1, the Feller process X^b is conservative and has a Lévy system $(N, H) = (J^b(x,y)dy, t)$ with

$$J^b(x,y) = \frac{\mathcal{A}(d,-\alpha)}{|x-y|^{d+\alpha}} \left(1 + \frac{b(x,x-y)}{\mathcal{A}(d,-\alpha)} |x-y|^{\alpha-\beta}\right).$$

$J^b(x,y)$ describes the jumps of the process X^b. The Lévy system $(\overline{N}^E, \overline{H}^E)$ of the killed process $X^{b,E}$ with respect to the reference measure $\xi_E(dx)$ is given by

$$\begin{cases} \overline{N}^E(x,dy) := \frac{J^b(x,y)}{h_E(y)}\xi_E(dy), & \text{for } (x,y) \in E \times E \\ \overline{H}^E_t := t, \end{cases}$$

and the Lévy system $(\widehat{N}^E, \widehat{H}^E)$ of the dual process $\widehat{X}^{b,E}$ with respect to the Lebesgue measure is given by

$$\begin{cases} \widehat{N}^E(x,dy) = \frac{J^b(y,x)h_E(y)}{h_E(x)}dy, & \text{for } (x,y) \in E \times E \\ \widehat{H}^E_t = t. \end{cases}$$

To prove the intrinsic ultracontractivity of the semigroups $\{P^{b,E,D}_t : t \geq 0\}$ and $\{\widehat{P}^{b,E,D}_t : t \geq 0\}$, the techniques in Subsection 4.2.4 can be employed. It is not hard to see that we only need to re-establish Lemma 4.2.17. The rest of the proof is the same as Subsection 4.2.4. So after we prove the following lemma we will jump to the conclusion.

Similar to Subsection 4.2.4, we choose an arbitrary point $x_0 \in D$ and a sufficiently small $r_0 \in (0,\infty)$ such that $B(x_0,r_0) \subset \overline{B(x_0,r_0)} \subset D$. We put $B_0 := B(x_0,r_0/2)$, $C_1 := \overline{B(x_0,r_0)}$ and $B_2 := B(x_0,r_0)$.

Lemma 4.3.5. *There exists a constant $c = c(d,\alpha,\beta,D,b) > 0$ such that for every $x \in \mathbb{R}^d \setminus C_1$,*

$$\mathbb{P}_x\left(X^{b,E}_{\tau^{b,E}_{D\setminus C_1}} \in C_1\right) \geq c\mathbb{E}_x[\tau^{b,E}_{D\setminus C_1}] \qquad \textit{and} \qquad \mathbb{P}_x\left(\widehat{X}^{b,E}_{\widehat{\tau}^{b,E}_{D\setminus C_1}} \in C_1\right) \geq c\mathbb{E}_x[\widehat{\tau}^{b,E}_{D\setminus C_1}]$$

Proof. For any $z \in B_0$ and $y \in D \setminus C_1$, we have $|z-y| < diam(D)$. Let K, L be two compact sets such that $D \subset K$ and $\bigcup_{y\in D\setminus C_1}(y - B_0) \subset L$. By condition (4.42), we have

$$\begin{aligned}
&\mathbb{P}_x\left(X^{b,E}_{\tau^{b,E}_{D\setminus C_1}} \in C_1\right) \geq \mathbb{P}_x\left(X^{b,E}_{\tau^{b,E}_{D\setminus C_1}} \in B_0\right) \\
=& \int_{D\setminus C_1} G^b_{D\setminus C_1}(x,y)\int_{B_0} J^b(y,z)dzdy \\
=& \int_{D\setminus C_1} G^b_{D\setminus C_1}(x,y)\int_{B_0} \frac{\mathcal{A}(d,-\alpha)}{|y-z|^{d+\alpha}}(1+\frac{b(y,y-z)}{\mathcal{A}(d,-\alpha)}|y-z|^{\alpha-\beta})dzdy \\
\geq& \inf_{y\in K}\inf_{z\in L}(1+\frac{b(y,z)}{\mathcal{A}(d,-\alpha)}|z|^{\alpha-\beta})\int_{D\setminus C_1} G^b_{D\setminus C_1}(x,y)\int_{B_0}\frac{\mathcal{A}(d,-\alpha)}{|z-y|^{d+\alpha}}dzdy \\
\geq& \frac{\mathcal{A}(d,-\alpha)|B_0|\inf_{y\in K}\inf_{z\in L}(1+\frac{b(x,z)}{\mathcal{A}(d,-\alpha)}|z|^{\alpha-\beta})}{diam(D)^{d+\alpha}}\int_{D\setminus C_1} G^b_{D\setminus C_1}(x,y)dy \\
\geq& \frac{\mathcal{A}(d,-\alpha)|B_0|\inf_{y\in K}\inf_{z\in L}(1+\frac{b(x,z)}{\mathcal{A}(d,-\alpha)}|z|^{\alpha-\beta})}{diam(D)^{d+\alpha}}\mathbb{E}_x[\tau^{b,E}_{D\setminus C_1}].
\end{aligned}$$

Let $c_1 = \frac{\mathcal{A}(d,-\alpha)|B_0| \inf_{y\in K} \inf_{z\in L}(1+\frac{b(x,z)}{\mathcal{A}(d,-\alpha)}|z|^{\alpha-\beta})}{diam(D)^{d+\alpha}}$ and it depends on d, α, β, D and b. We have proved the first part. For the dual process, we have

$$
\begin{aligned}
&\mathbb{P}_x\left(\widehat{X}^{b,E}_{\widehat{\tau}^{b,E}_{D\setminus C_1}} \in C_1\right) \geq \mathbb{P}_x\left(\widehat{X}^{b,E}_{\widehat{\tau}^{b,E}_{D\setminus C_1}} \in B_0\right)\\
&= \int_{D\setminus C_1} \widehat{G}^{b,E}_{D\setminus C_1}(x,y)\int_{B_0} \frac{J^b(z,y)h_E(z)}{h_E(y)}dzdy\\
&= \int_{D\setminus C_1} \widehat{G}^{b,E}_{D\setminus C_1}(x,y)\int_{B_0} \frac{h_E(z)}{h_E(y)}\frac{\mathcal{A}(d,-\alpha)}{|y-z|^{d+\alpha}}(1+\frac{b(z,y-z)}{\mathcal{A}(d,-\alpha)}|y-z|^{\alpha-\beta})dzdy\\
&\geq \frac{\mathcal{A}(d,-\alpha)|B_0| \inf_{z\in K}\inf_{y\in L}(1+\frac{b(z,y)}{\mathcal{A}(d,-\alpha)}|y|^{\alpha-\beta})}{\sup_{y,z\in D}\frac{h_E(y)}{h_E(z)} diam(D)^{d+\beta}}\mathbb{E}_x[\widehat{\tau}^{b,E}_{D\setminus C_1}].
\end{aligned}
$$

Let $c_2 = \frac{\mathcal{A}(d,-\alpha)|B_0| \inf_{z\in K}\inf_{y\in L}(1+\frac{b(z,y)}{\mathcal{A}(d,-\alpha)}|y|^{\alpha-\beta})}{\sup_{y,z\in D}\frac{h_E(y)}{h_E(z)} diam(D)^{d+\beta}} > 0$ then the second part is proved. To get an uniform constant, we can set $c = c_1 \wedge c_2 > 0$ and the proof is completed. □

The following theorem is our final conclusion.

Theorem 4.3.6. *The semigroups* $\{P^{b,E,D}_t : t \geq 0\}$ *and* $\{\widehat{P}^{b,E,D}_t : t \geq 0\}$ *are intrinsically ultracontractive (with respect to the reference measure* $\xi_E(dx)$*). For any* $t \geq 0$*, there exists a constant* $c_t > 0$ *such that*

$$
c_t^{-1}\phi^{b,E}_D(x)\psi^{b,E}_D(y) \leq \overline{p}^b_D(t,x,y) \leq c_t\phi^{b,E}_D(x)\psi^{b,E}_D(y), \qquad \textit{for any } x, y \in D.
$$

Proof. The proof is the same as Theorem 4.2.19. □

The results of Theorem 4.2.19 and Theorem 4.3.6 conclude this chapter. From the previous two sections, it is easy to see that our proofs rely on the jump property of the original process X^b. Even though for $0 < \alpha < 2$ and $\alpha = 2$ we have different analyses, the principle of our method is to control the jump behavior of the perturbed process X^b. Condition 4.7 and Condition 4.42 impose a positive lower bound on the intensity of the jump of the process X^b. This fact helps us establish Lemma 4.2.18 and finally prove the intrinsic ultracontractivity. So our proofs here could suggest that the semigroups of a larger class of jump processes can have intrinsic ultracontractivity if their Lévy systems satisfy similar conditions. Of course, one prerequisite is to find the dual process of the original process under a certain reference measure. In this work, we achieved this by using the heat kernel estimates of the original process. But this would limit ourselves to the operators whose heat kernel estimates are already known. For future work, we would like to consider establishing the intrinsic ultracontractivity from a more abstract view.

Chapter 5

Future Research

One possible direction for future research lies in the part of the GMWB rider with the annual step-up. As we have seen in Chapter 3 Section 3.3, the dynamics of the system of the account value process U_t and the guarantee base process G_t is given by

$$\begin{aligned}\frac{dU_t}{U_{t-}} &= \left((\mu - m) - (w + m_r)\frac{G_{t-}}{U_{t-}}\right) dt + dY_t^+ - dY_t^-,\\ \frac{dG_t}{G_{t-}} &= \left((\mu - m) - (w + m_r)\frac{G_{t-}}{U_{t-}}\right)_+ dt + dY_t^+.\end{aligned} \tag{5.1}$$

In Chapter 3, we considered a simplified case by setting $\left((\mu - m) - (w + m_r)\frac{G_{t-}}{U_{t-}}\right)_+ = 0$. This simplification is due to the usual market conditions. It is worth going beyond this assumption and trying to completely describe the evolution of the model. If $w + m_r < \mu - m$, we let

$$\tau_1 := \inf\{t > 0 : \frac{U_t}{G_t} < \frac{w + m_r}{\mu - m}\},$$

which is the hitting time of 0 for the drift term in the second equation in (5.1). Note that when $t \le \tau_1$, the drift terms of the two equations in (5.1) are the same. This leads us to consider the dynamics of the process $\{\frac{U_t}{G_t} : t \le \tau_1\}$. Using Itô's rule for semimatingales, we can obtain that

$$d(\frac{U_t}{G_t}) = -\frac{U_{t-}}{G_{t-}} dY_t^-, \qquad t \le \tau_1.$$

The solution to the previous equation is $\frac{U_t}{G_t} = e^{-X_t^-}$. This implies $\frac{U_t}{G_t}$ is an exponential compound Poisson subordinator. The hitting time τ_1 is in fact the hitting time for a compound Poisson subordinator. Then the dynamics of System 5.1 can be described as

$$
\begin{aligned}
\frac{dU_t}{U_{t-}} &= \left((\mu - m) - (w + m_r)\frac{G_{t-}}{U_{t-}}\right) dt + dY_t^+ - dY_t^- \\
d(\frac{U_t}{G_t}) &= -\frac{U_{t-}}{G_{t-}} dY_t^-,
\end{aligned} \tag{5.2}
$$

for $t \leq \tau_1$, and

$$
\begin{aligned}
\frac{dU_t}{U_{t-}} &= \left((\mu - m) - (w + m_r)\frac{G_{t-}}{U_{t-}}\right) dt + dY_t^+ - dY_t^- \\
\frac{dG_t}{G_{t-}} &= dY_t^+,
\end{aligned}
$$

for $t > \tau_1$. This two-period model could be considered for pricing the GMWB with the annual high step-up. The difficulties are in the linking of these two periods. Since we are considering jump processes, after the process $\frac{U_t}{G_t}$ crosses the level $\frac{w+m_r}{\mu-m}$, there is an overshoot. If we would like to calculate the quantities related to the pricing problem, we need to deal with the overshoot distribution. This is related to the renewal theory and the potential measure of the compound Poisson subordinator. Then it becomes more challenging to derive analytical result and develop efficient pricing algorithm. This part is our ongoing research.

Appendix A

Certain Special Functions

A.1 The Gamma function

The Gamma function $\Gamma(z)$ is given by the Euler integral

$$\Gamma(z) := \int_0^\infty x^{z-1} e^{-x} dx,$$

for any complex number $z \neq 0, -1, -2, \ldots$. Obviously $\Gamma(1) = 1$. Its most well-known properties are the following recurrence relation

$$\Gamma(z+1) = z\Gamma(z), \qquad z \neq 0, -1, -2, \ldots,$$

and reflection formula

$$\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin(\pi z)}, \qquad z \neq 0, \pm 1, \ldots.$$

Therefore, from the above recurrence relation , it is to see that for any $n \in \mathbb{Z}^+$, we have $\Gamma(n) = (n-1)!$. From the reflection formula, we can derive that $\Gamma(\frac{1}{2}) = \sqrt{\pi}$. It can be shown that $\Gamma(z)$ is a meromorphic function on $\mathbb{C}$ with simple poles at non-positive integers. The residue of $\Gamma(z)$ at $z = -n$ is $\frac{(-1)^n}{n!}$, where $n \in \mathbb{Z}^+$ (See [54, p136]). The Gamma function is one of the most common types of special functions and it is closely related to other special functions, such as the Beta function and other hypergeometric functions. It is widely used in many applications.

A.2 Whittaker functions

The Whittaker functions $M_{\kappa,\mu}(z)$ and $W_{\kappa,\mu}(z)$ are defined to be the fundamental solutions of the Whittaker's ODE

$$\frac{d^2W}{dz^2} + \left(-\frac{1}{4} + \frac{\kappa}{z} + \frac{\frac{1}{4} - \mu^2}{z^2}\right) W = 0,$$

for $\kappa, \mu \in \mathbb{C}$. $M_{\kappa,\mu}(z)$ is the one that does not exist when $2\mu = -1, -2, -3, \ldots$. We introduce the Pochhammer's symbol: for any $x \in \mathbb{R}$ and $n \in \mathbb{N}$, we define

$$\begin{aligned}(x)_0 &:= 1,\\ (x)_n &:= x(x+1)(x+2)\cdots(x+n-1).\end{aligned} \tag{A.1}$$

Therefore, from [54, p334, 13.14.6], we have the following series representation of $M_{\kappa,\mu}(z)$:

$$M_{\kappa,\mu}(z) = e^{-\frac{1}{2}}z^{\frac{1}{2}+\mu}\sum_{n=0}^{\infty}\frac{(\frac{1}{2}+\mu-\kappa)_n}{(1+2\mu)_n n!}z^n,$$

for any $z \in \mathbb{C}$ and $2\mu \neq -1, -2, -3, \ldots$. When $2\mu \notin \mathbb{Z}$, the other fundamental solution $W_{\kappa,\mu}(z)$ can be represented as ([54, p335, 13.14.33])

$$W_{\kappa,\mu}(z) = \frac{\Gamma(-2\mu)}{\Gamma(\frac{1}{2}-\mu-\kappa)}M_{\kappa,\mu}(z) + \frac{\Gamma(2\mu)}{\Gamma(\frac{1}{2}+\mu-\kappa)}M_{\kappa,-\mu}(z).$$

For more properties of Whittaker functions and their relationships with other special functions, readers are referred to [54]. Symbolic and numerical computations of Whittaker functions are handled very well in computation software, such as Mathematica, Maple and MATLAB. There are packages which use specialized algorithms to handle their evaluations.

A.3 The Meijer G-function

First, we introduce the generalized hypergeometric series and the generalized hypergeometric functions. Let $a_1, \ldots, a_p$ and $b_1, \ldots, b_q$ be $p+q$ complex (or real) parameters and none of them is a nonpositive integer. Recalling the Pochhammer's symbol we introduced in (A.1), we define formally

$${}_pF_q\begin{pmatrix}a_1,\ldots,a_p\\ & ;z\\ b_1,\ldots,b_q\end{pmatrix} := \sum_{n=0}^{\infty}\frac{(a_1)_n\cdots(a_p)_n}{(b_1)_n\cdots(b_q)_n}\frac{z^n}{n!}, \tag{A.2}$$

for $z \in \mathbb{C}$. The right-hand side of (A.2) is called the (formal) generalized hypergeometric series. The left-hand side of (A.2) is called the generalized hypergeometric function providing that the series on the right-hand side is convergent. If $p \leq q$, the generalized hypergeometric series converges for all $z \in \mathbb{C}$ and the ${}_pF_q$ is an entire function on $\mathbb{C}$. For other convergent conditions of the generalized hypergeometric series (A.2), readers can refer to [54, p404].

To introduce the Meijer G-function ([54, p415]), we assume m, n to be two integers satisfying $0 \leq m \leq q$ and $0 \leq n \leq p$. For any k, j with $1 \leq k \leq n$ and $1 \leq j \leq m$, we also assume $a_k - b_j$ is not a positive integer. Moreover, the following additional conditions hold:

- (B1) $p \leq q$;
- (B2) for any $1 \leq j_1, j_2 \leq m$, $b_{j_1} - b_{j_2}$ is not an integer.

The Meijer G-function $G^{m,n}_{p,q}$ can be defined as

$$G^{m,n}_{p,q}\left(z; \begin{matrix} a_1, \ldots, a_p \\ b_1, \ldots, b_q \end{matrix}\right)$$
$$:= \sum_{k=1}^{m} A^{m,n}_{p,q,k}(z) {}_pF_q \left(\begin{matrix} 1 + b_k - a_1, & \ldots & , 1 + b_k - a_p \\ 1 + b_k - b_1, \ldots & 1 + b_k - b_{k-1}, 1 + b_k - b_{k+1}, & \ldots, 1 + b_k - b_q \end{matrix} ; (-1)^{p-m-n} z \right),$$

with

$$A^{m,n}_{p,q,k}(z) := \frac{\Pi^m_{l=1, l \neq k} \Gamma(b_l - b_k) \Pi^n_{l=1} \Gamma(1 + b_k - a_l) z^{b_k}}{\Pi^{q-1}_{l=m} \Gamma(1 + b_k - b_{l+1}) \Pi^{p-1}_{l=n} \Gamma(a_{l+1} - b_k)}.$$

More generally, if additional assumptions (B1) and (B2) do not hold, we can still define the Meijer G-function through the Mellin-Barnes integral representation ([54, p415]):

$$G^{m,n}_{p,q}\left(z; \begin{matrix} a_1, \ldots, a_p \\ b_1, \ldots, b_q \end{matrix}\right) := \frac{1}{2\pi i} \int_L \frac{\Pi^m_{l=1} \Gamma(b_l - s) \Pi^n_{l=1} \Gamma(1 - a_l + s)}{\Pi^{q-1}_{l=m} \Gamma(1 - b_{l+1} + s) \Pi^{p-1}_{l=n} \Gamma(a_{l+1} - s)} z^s ds,$$

where the path L should separate the poles of the factors $\Gamma(b_l - s)$ from the poles of the factors $\Gamma(1 - a_l + s)$. For more properties about the Meijer G-function and its relationships with the generalized hypergeometric functions, readers can refer to [54, Chapter 15]. Symbolic and numerical evaluations of the Meijer G-function are available in many computation software, such as Mathematica, Maple and MATLAB.

References

[1] Abate, J., & Whitt, W. (1992). The Fourier-series method for inverting transforms of probability distributions. *Queueing systems*, 10(1), 5-87.

[2] Abate, J., & Whitt, W. (2006). A unified framework for numerically inverting Laplace transforms. *INFORMS Journal on Computing*, 18(4), 408-421.

[3] Bacinello, A. R., Millossovich, P., Olivieri, A., & Pitacco, E. (2011). Variable annuities: A unifying valuation approach. *Insurance: Mathematics and Economics*, 49(3), 285-297.

[4] Bartholomew, D. J. (1969). Sufficient conditions for a mixture of exponentials to be a probability density function. *The Annals of Mathematical Statistics*, 40(6), 2183-2188.

[5] Bauer, D., Kling, A., & Russ, J. (2008). A universal pricing framework for guaranteed minimum benefits in variable annuities. *Astin Bulletin*, 38(02), 621-651.

[6] Bell, F. C., Wade, A. H., & Goss, S. C. (1992). Life tables for the United States social security area. *Actuarial study*, 116, 1900-2100.

[7] Bertoin, J. (1998). Lévy processes (Vol. 121). Cambridge university press.

[8] Blumenthal, R. M., & Getoor, R. K. (1968). Markov processes and potential theory.

[9] Botta, R. F., & Harris, C. M. (1986). Approximation with generalized hyperexponential distributions: weak convergence results. *Queueing Systems*, 1(2), 169-190.

[10] Boyle, P., & Schwartz, E. (1977). Equilibrium prices of equity linked insurance policies with an asset value guarantee. *Journal of Risk and Insurance*, 44, 639-660.

[11] Brennan, M. J., & Schwartz, E. S. (1976). The pricing of equity-linked life insurance policies with an asset value guarantee. *Journal of Financial Economics*, 3(3), 195-213.

[12] Cai, N., & Kou, S. G. (2010). Pricing Asian options under a general jump diffusion model. *Operations Research*.

[13] Cai, N., & Kou, S. G. (2011). Option pricing under a mixed-exponential jump diffusion model. *Management Science*, 57(11), 2067-2081.

[14] Carmona, P., Petit, F., & Yor, M. (1997). On the distribution and asymptotic results for exponential functionals of Lévy processes. *Exponential functionals and principal values related to Brownian motion*, 73-121.

[15] Chen, Z. Q., Kim, P., & Kumagai, T. (2008). Weighted Poincar inequality and heat kernel estimates for finite range jump processes. *Mathematische Annalen*, 342(4), 833-883.

[16] Chen, Z. Q., Kim, P., & Song, R. (2010). Two-sided heat kernel estimates for censored stable-like processes. *Probability theory and related fields*, 146(3), 361-399.

[17] Chen, Z. Q., Kim, P., & Song, R. (2012). Dirichlet heat kernel estimates for fractional Laplacian with gradient perturbation. *The Annals of Probability*, 40(6), 2483-2538.

[18] Chen, Z. Q., & Kumagai, T. (2008). Heat kernel estimates for jump processes of mixed types on metric measure spaces. *Probability Theory and Related Fields*, 140(1-2), 277-317.

[19] Chen, Z. Q., & Kumagai, T. (2010). A priori Hlder estimate, parabolic Harnack principle and heat kernel estimates for diffusions with jumps. *Revista Matematica Iberoamericana*, 26(2), 551-589.

[20] Chen, Z., Vetzal, K., & Forsyth, P. A. (2008). The effect of modelling parameters on the value of GMWB guarantees. *Insurance: Mathematics and Economics*, 43(1), 165-173.

[21] Chen, Z. Q., & Wang, J. M. (2013). Perturbation by non-local operators. *arXiv preprint* arXiv:1312.7594.

[22] Chen, Z. Q., & Yang, T. (2015). Dirichlet heat kernel estimates for fractional Laplacian under non-local perturbation. *arXiv preprint* arXiv:1503.05302.

[23] Chung, K. L., & Walsh, J. B. (2006). Markov processes, Brownian motion, and time symmetry (Vol. 249). Springer Science & Business Media.

[24] Chung, K. L., & Zhao, Z. (1995). From Brownian motion to Schrödinger's equation.

[25] Davies, E. B. (1980). One-parameter semigroups (Vol. 15). Academic Pr.

[26] Davies, E. B., & Simon, B. (1984). Ultracontractivity and the heat kernel for Schrödinger operators and Dirichlet Laplacians. *Journal of Functional Analysis*, 59(2), 335-395.

[27] Coleman, T. F., Kim, Y., Li, Y., & Patron, M. (2007). Robustly hedging variable annuities with guarantees under jump and volatility risks. *Journal of Risk and Insurance*, 74(2), 347-376.

[28] Dai, M., Kuen Kwok, Y., & Zong, J. (2008). Guaranteed minimum withdrawal benefit in variable annuities. Mathematical Finance, 18(4), 595-611.

[29] Feng, R., & Huang, H. (2016). Statutory financial reporting for variable annuity guaranteed death benefits: Market practice, mathematical modeling and computation. *Insurance: Mathematics and Economics*, 67, 54-64.

[30] Feng, R., Jing, X., & Dhaene, J. (2017). Comonotonic approximations of risk measures for variable annuity guaranteed benefits with dynamic policyholder behavior. *Journal of Computational and Applied Mathematics*, 311, 272-292.

[31] Feng, R., & Jing, X. (2017). Analytical valuation and hedging of variable annuity guaranteed lifetime withdrawal benefits. *Insurance: Mathematics and Economics*, 72, 36-48.

[32] Feng, R., & Vecer, J. (2017). Risk based capital for guaranteed minimum withdrawal benefit. *Quantitative Finance*, 17(3), 471-478.

[33] Feng, R., & Volkmer, H. W. (2012). Analytical calculation of risk measures for variable annuity guaranteed benefits. *Insurance: Mathematics and Economics*, 51(3), 636-648.

[34] Feng, R., & Volkmer, H. W. (2014). Spectral methods for the calculation of risk measures for variable annuity guaranteed benefits. *Astin Bulletin*, 44(03), 653-681.

[35] Feng, R., & Volkmer, H. W. (2016). An identity of hitting times and its application to the valuation of guaranteed minimum withdrawal benefit. *Mathematics and Financial Economics*, 10(2), 127-149.

[36] Getoor, R. K. (1971). Duality of Lévy systems. Zeitschrift für *Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 19(4), 257-270.

[37] Grosen, A., & Jørgensen, P. L. (2002). Life insurance liabilities at market value: an analysis of insolvency risk, bonus policy, and regulatory intervention rules in a barrier option framework. *Journal of risk and insurance*, 69(1), 63-91.

[38] Hull, J. C. (2006). Options, futures, and other derivatives. Pearson Education India.

[39] Hardy, M. (2003). Investment guarantees: modeling and risk management for equity-linked life insurance (Vol. 215). John Wiley & Sons.

[40] Huang, H., Milevsky, M. A., & Salisbury, T. S. (2014). Optimal initiation of a GLWB in a variable annuity: No Arbitrage approach. *Insurance: Mathematics and Economics*, 56, 102-111.

[41] Karatzas, I., & Shreve, S. (2012). Brownian motion and stochastic calculus (Vol. 113). Springer Science & Business Media.

[42] Kim, P., & Song, R. (2008). Intrinsic ultracontractivity of non-symmetric diffusion semigroups in bounded domains. *Tohoku Mathematical Journal*, Second Series, 60(4), 527-547.

[43] Kim, P., & Song, R. (2009). Intrinsic ultracontractivity for non-symmetric Lévy processes. *In Forum Mathematicum* (Vol. 21, No. 1, pp. 43-66).

[44] Kulczycki, T. (1998). Intrinsic ultracontractivity for symmetric stable processes. *Bulletin of the Polish Academy of Sciences. Mathematics*, 46(3), 325-334.

[45] Kuznetsov, A. (2012). On the distribution of exponential functionals for Lvy processes with jumps of rational transform. *Stochastic Processes and their Applications*, 122(2), 654-663.

[46] Kyprianou, A. E. (2014). Fluctuations of Lévy processes with applications. Introductory Lectures (second ed.) Universitext. Springer, Heidelberg.

[47] Liao, M. (1984). Riesz representation and duality of Markov processes. Ph.D. dissertation, Dept. Mathematics, Stanford Univ.

[48] Liao, M. (1985). Riesz representation and duality of Markov processes. *Séminaire de probabilités de Strasbourg*, 19, 366-396.

[49] Linetsky, V. (2004). The spectral decomposition of the option value. *International Journal of Theoretical and Applied Finance*, 7(03), 337-384.

[50] Linetsky, V. (2004). Spectral expansions for Asian (average price) options. *Operations Research*, 52(6), 856-867.

[51] Milevsky, M. A., & Salisbury, T. S. (2006). Financial valuation of guaranteed minimum withdrawal benefits. *Insurance: Mathematics and Economics*, 38(1), 21-38.

[52] Milevsky, M. A., & Posner, S. E. (2001). The titanic option: valuation of the guaranteed minimum death benefit in variable annuities and mutual funds. *Journal of Risk and Insurance*, 93-128.

[53] Milevsky, M. A., & Salisbury, T. S. (2002). The real option to lapse and the valuation of death-protected investments.

[54] Olver, F. W., Lozier, D. W., Boisvert, R. F., & Clark, C. W. (2010). NIST Handbook of Mathematical Functions.

[55] Peng, J., Leung, K. S., & Kwok, Y. K. (2012). Pricing guaranteed minimum withdrawal benefits under stochastic interest rates. *Quantitative Finance*, 12(6), 933-941.

[56] Piscopo, G., & Haberman, S. (2011). The valuation of guaranteed lifelong withdrawal benefit options in variable annuity contracts and the impact of mortality risk. *North American Actuarial Journal*, 15(1), 59-76.

[57] Reed, M., & Simon, B. (1980). Methods of Modern Mathematical Physics (Academic, New York, 1975). Vol. II, 166-168.

[58] Revuz, D., & Yor, M. (2013). Continuous martingales and Brownian motion (Vol. 293). Springer Science & Business Media.

[59] Schaefer, H. H., & Schaefer, H. H. (1974). Banach lattices (pp. 46-153). Springer Berlin Heidelberg.

[60] Sheldon, T. J., & Smith, A. D. (2004). Market consistent valuation of life assurance business. *British Actuarial Journal*, 10(03), 543-605.

[61] Shur, M. G. E. (1978). On dual Markov processes. *Theory of Probability & Its Applications*, 22(2), 257-270.

[62] Song, R., & Vondracek, Z. (2007). Parabolic Harnack inequality for the mixture of Brownian motion and stable process. *Tohoku Mathematical Journal*, Second Series, 59(1), 1-19.

[63] Ulm, E. R. (2008). Analytic solution for return of premium and rollup guaranteed minimum death benefit options under some simple mortality laws. *Astin Bulletin*, 38(02), 543-563.

[64] Wang, J. M. (2015). Laplacian perturbed by non-local operators. *Mathematische Zeitschrift*, 279(1-2), 521-556.