mathematics

Special Issue Reprint

Modeling, Analysis and Optimization for Mathematical Finance, Economics and Risks

Edited by
Jing Yao, Xiang Hu and Jingchao Li

mdpi.com/journal/mathematics

MDPI

Modeling, Analysis and Optimization for Mathematical Finance, Economics and Risks

Modeling, Analysis and Optimization for Mathematical Finance, Economics and Risks

Editors

Jing Yao
Xiang Hu
Jingchao Li

MDPI

Basel • Beijing • Wuhan • Barcelona • Belgrade • Novi Sad • Cluj • Manchester

Editors

Jing Yao
Soochow University
Suzhou
China

Xiang Hu
Zhongnan University of Economics and Law
Wuhan
China

Jingchao Li
Shenzhen University
Shenzhen
China

Editorial Office
MDPI AG
Grosspeteranlage 5
4052 Basel, Switzerland

This is a reprint of articles from the Special Issue published online in the open access journal *Mathematics* (ISSN 2227-7390) (available at: https://www.mdpi.com/journal/mathematics/special_issues/M111VJGFG9).

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

Lastname, A.A.; Lastname, B.B. Article Title. *Journal Name* **Year**, *Volume Number*, Page Range.

ISBN 978-3-7258-1729-0 (Hbk)
ISBN 978-3-7258-1730-6 (PDF)
doi.org/10.3390/books978-3-7258-1730-6

Contents

About the Editors

Jing Yao

Jing Yao is currently a professor at the Center for Financial Engineering, Soochow University, China. He obtained his Ph.D. degree in Applied Economics from Vrije Universiteit Brussel in 2013. His current research interests focus on financial and actuarial risk management, stochastic controls and optimization, model uncertainty and applied probability and statistics. His research receives support (PI) from international research foundations such as the NSFC, FWO, Erasmus, Zimmerman Foundation, etc. He is also a research fellow in the Actuarial Research Center of Haifa University, a specially appointed professor of the Jiangsu Province, and a chair professor of the Bayu Scholar program.

Xiang Hu

Xiang Hu, Associate of the China Association of Actuaries (ACAA), is a professor of Insurance and Actuarial Science at the Department of Finane, Zhongnan University of Economics and Law (ZUEL). He received his Ph.D. degree from Nankai University in actuarial science. Prof. Hu's research focuses on risk management, actuarial mathematics, and insurance economics. His research has appeared in journals such as *The ASTIN Bulletin*, *Insurance: Mathematics and Economics*, *The North American Actuarial Journal*, *The Scandinavian Actuarial Journal*, *The North American Journal of Economics and Finance*, and *The Journal of Computational and Applied Mathematics*. He joined ZUEL in 2015 and has received support from the National Natural Science Foundation of China (NSFC) and several other projects.

Jingchao Li

Jingchao Li received her bachelor's degree, honors bachelor's degree, and Ph.D. from the University of Melbourne, Australia. Currently, she serves as an Associate Professor and a master's supervisor in the School of Mathematical Sciences at Shenzhen University. She is an actuary of the Institute of Actuaries Australia, a high-level overseas talent in Shenzhen, and a backup high-level talent in Shenzhen. Her primary research interests lie in actuarial science and risk theory related to insurance. She has led four scientific research projects, including those funded by the National Natural Science Foundation of China, as well as four actuarial and catastrophe-related teaching and research projects. Additionally, she has participated in four scientific research projects, serving as a key research member in a sub-project of the Ministry of Science and Technology's Key Research and Development Program. Her research has appeared in academic journals such as *Insurance: Mathematics and Economics*, *The Journal of Nonlinear and Variational Analysis*, *Methodology and Computing in Applied Probability*.

mathematics

MDPI

Article

Robust Optimal Investment Strategies with Exchange Rate Risk and Default Risk

Wei Wang, Qianyan Li, Quan Li and Song Xu *

School of Mathematics and Statistics, Ningbo University, No. 818 Fenghua Road, Ningbo 315211, China
* Correspondence: xusong@nbu.edu.cn; Tel.: +86-150-8883-2546

Abstract: The problem of robust optimal investment with exchange rate risk and default risk is studied. We assume that investors are ambiguity averse and they have access not only to the domestic market but also to the foreign market. The corresponding Hamilton–Jacobi–Bellman (HJB) equations are first obtained through the robust stochastic optimal control theory. Then, we discuss the optimal investment problems before and after default, and the value functions and optimal investment strategies are obtained. Finally, we find that the optimal investment strategies of pre-default are affected by the intensity of default and the credit spread, and the investors cannot hold defaultable bonds in the post-default case. Numerical results also show that the exchange rate risk, default risk and ambiguity aversion have a great effect on the optimal investment strategies.

Keywords: default risk; ambiguity aversion; HJB equation; optimal investment

MSC: 91B05; 91B70; 60H30

1. Introduction

Merton [1,2] first considered the problems of continuous time portfolio optimization. Since then, the optimal investment portfolio problems have been extensively investigated [3–6]. These works take the market risk into account when making a portfolio investment decision. However, a lot of empirical studies have shown that investors are not only risk averse but also ambiguity averse. In the field of behavioral economics, the investors' uncertainty about the probability distribution of random economic factors is called ambiguity aversion. Anderson et al. [7] first introduced the concept of ambiguity aversion and considered an optimal investment problem. By adopting a robust control method, the robust optimal investment strategies are obtained based on the so-called endogenous worst-case scenario. Different from Anderson et al. [7], Maenhout [8] presented a homothetic robustness framework to solve the optimal portfolio optimization problem and derived the closed-form solutions of robust optimal strategies when the investors have constant relative risk aversion utility. Many empirical results demonstrate that the financial market has the regime-switching property. Hence, Elliott and Siu [9] used a continuous time Markov chain to describe the state of economics and described the price process of the risky asset by a Markov-modulated Geometric Brownian. They derived the explicit solution of optimal investment strategies by using the stochastic differential game. Zheng et al. [10] consider an optimal portfolio optimization problem for an insurer with ambiguity aversion. The insurer's surplus process and the risky asset price process are assumed to follow the classical Cramer–Lundberg model and constant variance elasticity models, respectively. They obtained analytical solutions for robust optimal investment and reinsurance strategies by employing the Hamilton–Jacobi–Bellman (HJB) equations and the verification theorem. Most of the existing literature assumes that investors can only invest in bonds and stocks. However, there are many kinds of financial derivatives that can be invested in the market. Zeng et al. [11] further studied the robust optimal investment problem involving a European derivative with ambiguity aversion. Branger and Larsen [12] considered the jump

Citation: Wang, W.; Li, Q.; Li, Q.; Xu, S. Robust Optimal Investment Strategies with Exchange Rate Risk and Default Risk. *Mathematics* **2023**, *11*, 1550. https://doi.org/10.3390/math11061550

Academic Editors: Jing Yao, Xiang Hu and Jingchao Li

Received: 1 March 2023
Revised: 19 March 2023
Accepted: 20 March 2023
Published: 22 March 2023

phenomenon of the risky asset prices in the financial market and obtained the analytic solutions of optimal investment strategies in the complete and incomplete market. Moreover, they illustrated that ignoring ambiguity aversion about jump risk is more serious than ignoring ambiguity aversion about diffusion risk in the complete financial market, and the opposite holds true in the incomplete financial market.

In the financial market, the default risk has usually important effects on the optimal investment strategy for an investor. Since the 2008 financial crisis, the default risk has attracted increasing attention from investors and banking institutions and become one of the most important sources of financial risks. In fact, the default risk may exist almost in all financial products. Over the past decades, many kinds of corporate bonds with default risk appeared in the financial market. These corporate bonds are becoming more and more popular among investors because of their high yield. Therefore, the optimal portfolio problems with defaultable bonds are studied by many researchers. Bielecki and Jang [13] assumed that the financial market has three kinds of assets: a risk free bond, a defaultable bond and a stock. They considered a portfolio optimization decision problem by the HJB approach. Bo et al. [14] also considered that the financial market has a defaultable bond but studied an optimal investment and consumption problem over an infinite horizon time. Capponi and Figueroa-Lopez [15] used a continuous time Markov chain to describe the state of economic and obtained an optimal investment strategy with a defaultable bond in a Markov-modulated market. Deng et al. [16] supposed that an insurer is ambiguity averse and the stock price follows the constant elasticity of variance model. They employed the robust control method to study an optimal investment and reinsurance problem with a defaultable bond. One can refer to [17–20] for more results on the optimal investment with default risk.

In addition to the default risk, exchange rate risk also plays an important role in international portfolio investment and other businesses. The reason for arising exchange rate risk is that the value of the portfolio in local currency changes as the exchange rate changes. Kozo and Shujiro [21] presented the effect of exchange rate and its volatility on the investment decision. Guo et al. [22] supposed that there is an interest rate risk, exchange rate risk and inflation risk in the financial market. They studied the problem of the optimal reinsurance and investment when the risky assets follow the Geometry Brownian motions. Fei et al. [23] also assumed that the dynamics of exchange rate risk is described by the Geometry Brownian motion, but the risky assets price dynamics are modeled by a diffusion process which depend on time-varying underlying factors. They studied the problem of optimal investment of a multinational corporation and derived the explicit expression of the optimal investment strategies.

However, to our best knowledge, no discussion has yet been given on optimal investment selection problem with default risk, exchange rate risk and model uncertainty. This paper investigates a robust optimal investment selection problem of an investor with constant absolute risk aversion (CARA) utility, when the price processes of the defaultable bond and exchange rate follow the pure jump and Geometric Brownian models, respectively. The main contribution of our paper is given as follows.

(1) We consider a financial market with four assets, namely, a defaultable bond, a risk-free bond, a domestic stock and a foreign stock. An optimal portfolio selection problem with default risk and exchange risk is presented by maximizing the minimal expected utility of their terminal wealth over a class of probability measures.

(2) Using robust control and dynamic programming principle, we obtain the corresponding HJB equations in the pre-default case and post-default case. Solving these two HJB equations, we derive the explicit expression of the optimal investment strategies under the constant absolute risk aversion utility. We consider two cases, one with and one without the ambiguity aversion. Comparing the results of the two cases, we illustrate that the ambiguity aversion has a significant effect on the optimal investment strategies for an investor.

(3) For the international investors, they face not only market risk and default risk but also exchange rate risk. We suppose that the exchange rate is stochastic and study the joint effect of several kinds of risk and ambiguity aversion on the optimal portfolio choices. The numerical results show that default risk and exchange risk have different impacts on four different types of assets.

This article is organized as follows. In Section 2, the financial market model is set up. In Section 3, we provide an analytical solution to the optimal investment strategies under CARA utility. In Section 4, we present some numerical results of the optimal investment strategies and conclude the article in Section 5.

2. Financial Market Model

Suppose that the financial market has no transaction costs or taxes. All trades are continuously occuring on a finite time period $[0, T]$, where $T > 0$ is a finite constant. Let $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, P\right)$ be a filtered complete probability space, where P is a real-world probability measure and the filtration $(\mathcal{F}_t)_{t\in[0,T]}$ is the augmented natural filtration of a standard three-dimensional Brownian motion $\mathbf{W}(t) = \{W_1(t), W_2(t), W_3(t)\}$. Let τ be the first jump time of a Possion process with intensity h^p under the probability measure P, where h^p is a positive constant. We use τ to denote the time when the default risk occurs. For $t \in [0, T]$, denote the default process by $Z(t) := I_{\{\tau\leq t\}}$, where $I_{\{\cdot\}}$ is an indicator function.

Let $\mathcal{H} = (\mathcal{H}_t; t \in [0, T])$ be the augmented natural filtration of the default process $Z(t)$ and $\mathcal{G}_t = \mathcal{H}_t \vee \mathcal{F}_t$ for all $t \in [0, T]$; then, $\mathcal{G} = (\mathcal{G}_t; t \in [0, T])$ is the smallest filtration such that the time of default τ becomes a stopping time. In addition, we define $M^p(t)$ as follows:

$$M^p(t) = Z(t) - \int_0^t h^p(1 - Z(u))du, \tag{1}$$

where $h^p > 0$ is the arrival intensity of the default under the probability measure P. Then, $M^p(t)$ is a $(P, \mathcal{G})$ martingale. Note that $Z(t) = I_{\{\tau\leq t\}}$; then, $M^p(t)$ can be also represented by

$$M^p(t) = Z(t) - \int_0^{t\wedge\tau} h^p du.$$

where $M(t)$ is a compensate process. After the default time τ, the default process $Z(t)$ remains at the value 1. Hence, there is no need to compensate for the default process $Z(t)$ after the default time τ.

2.1. Model Assumptions

We assume that the financial market consists of four kinds of assets: a risk-free bond B and a zero-coupon defaultable bond p in the domestic financial market, a domestic stock S^{d} and a foreign stock S^{f}.

The risk-free bond $B := \{B(t)\}_{t\in[0,T]}$ whose price processes are given by:

$$\mathrm{d}B(t) = r_{\mathrm{d}}B(t)\mathrm{d}t, \tag{2}$$

where $r_{\mathrm{d}} > 0$ represents the risk-free interest rate of the domestic bank.

The stock price processes $S^{\mathrm{d}} := \left\{S^{\mathrm{d}}(t)\right\}_{t\in[0,T]}$ and $S^{\mathrm{f}} := \left\{S^{\mathrm{f}}(t)\right\}_{t\in[0,T]}$ are assumed to follow Geometric Brownian motions

$$\mathrm{d}S^{\mathrm{d}}(t) = S^{\mathrm{d}}(t)(\mu_{\mathrm{d}}\mathrm{d}t + \sigma_{\mathrm{d}}\mathrm{d}W_1(t)), \tag{3}$$

$$\mathrm{d}S^{\mathrm{f}}(t) = S^{\mathrm{f}}(t)(\mu_{\mathrm{f}}\mathrm{d}t + \sigma_{\mathrm{f}}\mathrm{d}W_2(t)), \tag{4}$$

where μ_{d} and μ_{f} denote the appreciate rate of the stocks S^{d} and S^{f}, respectively, σ_{d} and σ_{f} denote the volatility of the stocks S^{d} and S^{f}, respectively. These parameters are assumed to be positive constants.

Let $R := \{R(t)\}_{t\in[0,T]}$ be the exchange rate risk process. As in Amin and Jarrow [24], the dynamics of the exchange rate is assumed to satisfy the following SDE:

$$\mathrm{d}R(t) = R(t)\bigg((\widehat{\mu} - r_{\mathrm{f}})\mathrm{d}t + \rho_1\theta\mathrm{d}W_1(t) + \rho_2\theta\mathrm{d}W_2(t) + \sqrt{1-\rho_1^2-\rho_2^2}\theta\mathrm{d}W_3(t)\bigg), \tag{5}$$

where $\theta > 0$ is the volatility of exchange rate risk process, r_{f} is the risk-free interest rate of the foreign bank and $\frac{\widehat{\mu}-r_{\mathrm{d}}}{\theta}$ is the market price of the risk. Furthermore, we consider that the price dynamics of exchange rate is correlated with the price processes of domestic stock and foreign stock, $\rho_1 \in [0,1], \rho_2 \in [0,1]$ are the correlation coefficients, and $\rho_1^2 + \rho_2^2 \in [0,1]$. Let $\mu = \widehat{\mu} - r_{\mathrm{d}} + \mu_f$. If the exchange rate risk process and the foreign stock price process have the same volatility, $\frac{\mu-r_{\mathrm{f}}}{\theta}$ is the sum of market rice of two kinds of risk since $\frac{\mu-r_{\mathrm{f}}}{\theta} = \frac{\widehat{\mu}-r_{\mathrm{d}}}{\theta} + \frac{\mu_f-r_{\mathrm{f}}}{\theta}$.

Remark 1. *Melino and Turnbull [25] showed that using a diffusion process with stochastic volatility to describe the exchange rate is more appropriate from empirical results. Johnson and Schneeweis [26] and Huang and Hung [27] pointed out that a jump diffusion process is more suitable for capturing the exchange rate price dynamics. For convenience, we assume here that the exchange rate price dynamics is modeled by a Geometric Brownian motion.*

Let $S(t)$ be the price of the foreign stock in the domestic currency. Then, $S(t)$ and $S^{\mathrm{f}}(t)$ have this relationship, i.e., $S(t) = S^{\mathrm{f}}(t)R(t)$. By using the Itô formula, we have

$$\begin{aligned}\mathrm{d}S(t) &= S^{\mathrm{f}}(t)\mathrm{d}R(t) + R(t)\mathrm{d}S^{\mathrm{f}}(t) + \mathrm{d} < S^{\mathrm{f}}, R >_t \\ &= S(t)\Big[\bar{\mu}\mathrm{d}t + \rho_1\theta\mathrm{d}W_1(t) + (\sigma_{\mathrm{f}} + \rho_2\theta)\mathrm{d}W_2(t) + \sqrt{1-\rho_1^2-\rho_2^2}\theta\mathrm{d}W_3(t)\Big],\end{aligned} \tag{6}$$

where $\bar{\mu} = \mu - r_{\mathrm{f}} + r_{\mathrm{d}} + \rho_2\theta\sigma_{\mathrm{f}}$, $< S^{\mathrm{f}}, R >_t$ denotes the quadratic variation process of $S(t)$ and $R(t)$.

Now, we present the price dynamics of the default bond. The holder of a bond receives only a portion of its value when a default risk occurs. Let $\zeta \in (0,1)$ denote the constant loss rate; then, 1-ζ is the default recovery rate of a bond. Furthermore, we let T_1 denote the expiration date of the zero-coupon defaultable bond. In addition, we assume $T_1 \geq T$. This assumption means that the defaultable bonds have not matured before the end of the investment, and the defaultable bond can still be traded in the financial market. This assumption is also reasonable. Generally, the maturity period of the bond is relatively long. Let Q be a given risk-neutral probability measure which is equivalent to the statistical probability measure P. By the risk-neutral pricing formula, we have that the price dynamics of defaultable bond $p(t,T_1)$ at time t under Q is given by

$$p(t,T_1) = (1-Z(t))\mathbb{E}_Q\Big[(1-\zeta)e^{-r_d(\tau-t)}Y_{\tau-}|\mathcal{G}_t\Big] + Z(t)(1-\zeta)e^{r-\tau}Y_{\tau-}. \tag{7}$$

By using Itô's formula with jump, the price dynamic of a defaultable bond $p(.,T_1) = (p(t,T_1); t \geq 0)$ can be written as follows

$$\mathrm{d}p(t,T_1) = p(t,T_1)\Big[r_{\mathrm{d}}\mathrm{d}t + \delta(1-\frac{1}{\mu_p})(1-Z(t))\mathrm{d}t - (1-Z(t-))\zeta\mathrm{d}M^p(t)\Big], \tag{8}$$

where $\mu_p = \frac{h^Q}{h^P}$ denotes the default risk premium, $h^Q > 0$ is the arrival intensity of the default risk under the probability measure Q, and $\delta = h^Q\zeta$ denotes the credit spread under the probability measure Q. For more details of the Formulas (7) and (8), please see the Proposition 1 in Bo et al. [14].

2.2. The Optimal Investment Problem

Denote by $\pi_s^{\mathrm{d}}(t)$ the dollar amounts invested in the domestic stock S^{d} at time t. Similarly, $\pi_s(t)$ denotes the dollar amounts invested in the foreign stock S^{f} and $\pi_p(t)$ denotes the dollar amounts invested in the defaultable bond. Denote by $X^{\pi}(t)$ the wealth process of a portfolio $\pi = \{\pi_s^{\mathrm{d}}(t), \pi_s(t), \pi_p(t)\}_{t\in[0,T]}$ at time t. Furthermore, $X^{\pi}(t) - \pi_s^{\mathrm{d}}(t) - \pi_s(t) - \pi_p(t)$ represents the dollar amounts invested in the risk-free asset B. Using the self-financing condition, the dynamics of the wealth process associated with $(\pi_s^{\mathrm{d}}(\cdot), \pi_s(\cdot), \pi_p(\cdot))$ are given as

$$\begin{aligned}\mathrm{d}X^{\pi}(t) = &\frac{\pi_s^{\mathrm{d}}(t)}{S^{\mathrm{d}}(t)}\mathrm{d}S^{\mathrm{d}}(t) + \frac{\pi_s(t)}{S(t)}\mathrm{d}S(t) + \frac{\pi_p(t)}{p(t,T_1)}\mathrm{d}p(t,T_1)\\ &+ \frac{X^{\pi}(t) - \pi_s^{\mathrm{d}}(t) - \pi_s(t) - \pi_p(t)}{B(t)}\mathrm{d}B(t).\end{aligned} \tag{9}$$

Plugging (2), (3), (6) and (8) into (9), then

$$\begin{aligned}\mathrm{d}X^{\pi}(t) = &\Big[\pi_s^{\mathrm{d}}(t)(\mu_{\mathrm{d}} - r_{\mathrm{d}}) + \pi_s(t)(\mu - r_{\mathrm{f}} + \rho_2\theta\sigma_{\mathrm{f}}) + \pi_p(t)\delta(1 - \frac{1}{\mu_p})(1 - Z(t))\\ &+ X^{\pi}(t)r_{\mathrm{d}}\Big]\mathrm{d}t + (\pi_s^{\mathrm{d}}(t)\sigma_d + \pi_s(t)\rho_1\theta)\mathrm{d}W_1(t) + \pi_s(t)(\sigma_{\mathrm{f}} + \rho_2\theta)\mathrm{d}W_2(t)\\ &+ \pi_s(t)\sqrt{1 - \rho_1^2 - \rho_2^2}\theta\mathrm{d}W_3(t) - \pi_{\mathrm{p}}(t)(1 - Z(t-))\zeta\mathrm{d}M^P(t).\end{aligned} \tag{10}$$

By substituting (1) into (10), the price dynamics of the wealth process $X^{\pi}(t)$ is obtained by

$$\begin{aligned}\mathrm{d}X^{\pi}(t) = &[X^{\pi}(t)r_{\mathrm{d}} + \pi_s^{\mathrm{d}}(t)(\mu_{\mathrm{d}} - r_{\mathrm{d}}) + \pi_s(t)(\mu - r_{\mathrm{f}} + \rho_2\theta\sigma_{\mathrm{f}}) + \pi_{\mathrm{p}}(t)\delta(1 - Z(t))]\mathrm{d}t\\ &+ (\pi_s^{\mathrm{d}}(t)\sigma_{\mathrm{d}} + \pi_s(t)\rho_1\theta)\mathrm{d}W_1(t) + \pi_s(t)(\sigma_{\mathrm{f}} + \rho_2\theta)\mathrm{d}W_2(t)\\ &+ \pi_{\mathrm{s}}(t)\sqrt{1 - \rho_1^2 - \rho_2^2}\theta\mathrm{d}W_3(t) - \pi_{\mathrm{p}}(t)(1 - Z(t-))\zeta\mathrm{d}Z(t).\end{aligned} \tag{11}$$

Remark 2. *The wealth X^{π} here can be negative, which means that investors are allowed to sell short. If the short position can not be allowed, we can suppose $\pi_p(t) < \frac{1}{\zeta}$ for all $t > 0$.*

Definition 1. *A trading strategy $\pi := \{\pi_s^{\mathrm{d}}(t), \pi_s(t), \pi_p(t)\}_{t\in[0,T]}$ is said to be admissible if*
(i). $\pi_s^{\mathrm{d}}(t), \pi_s(t), \pi_p(t)$ are $\mathcal{G}_t$ progressively measurable;
(ii). $\mathrm{E}^P\Big[\exp\Big(\int_0^T(\mid \pi_s^d(t) \mid^2 + \mid \pi_s(t) \mid^2 + \mid \pi_p(t) \mid^2)\mathrm{d}t\Big)\Big] < \infty$, where $\mathrm{E}^P[\cdot]$ denotes the expectation under the probability measure P;
(iii). For any $t \in [0,T]$, the stochastic differential Equation (11) has a unique strong solution.

An investor hopes to maximize the expected utility of his terminal wealth by choosing an optimal investment strategy, and the objective functional is naturally defined by

$$\sup_{\pi(t)\in\Pi}\Big\{\mathrm{E}^P[U(X^{\pi}(T))|X^{\pi}(t) = x]\Big\}, \tag{12}$$

where Π represents the set of all admissible strategies, and $\mathrm{E}^P[\cdot|X^{\pi}(t) = x]$ denotes conditional expectation under the probability measure P.

In this article, we suppose that the investor is ambiguous adverse. That is, the investor cannot believe that the model under the probability measure P is completely correct. The investor has a certain ambiguity of model estimation under probability measure P. He needs to consider some alternative models defined by the other probability measures to replace the original model. These probability measures are defined by the world probability

measure P and the Radon–Nikodym derivatives. For $0 \le t \le T$, we suppose that there exists a Radon–Nikodym derivative process $\Lambda^{\phi}(t)$ defined by

$$\Lambda^{\phi}(t) = \exp\Big\{ -\int_0^t \phi_1(u)\mathrm{d}W_1(u) - \int_0^t \phi_2(u)\mathrm{d}W_2(u) - \int_0^t \phi_3(u)\mathrm{d}W_3(u) + \int_0^t \ln\phi_4(u-)\mathrm{d}Z(u) \\ -\frac{1}{2}\int_0^t \Big(\phi_1^2(u) + \phi_2^2(u) + \phi_3^2(u)\Big)\mathrm{d}u + h^p \int_0^t (1-\phi_4(u))(1-Z(u))\mathrm{d}u \Big\}, \tag{13}$$

where $\Phi(t) = \{\phi_i(t), i = 1,2,3,4\}$ should satisfy the following conditions:

(i). $\Phi(t)$ are progressively measurable processes;
(ii). For all $t \in [0,T]$, $\phi_4(t) > 0$;
(iii). $\mathrm{E}^P\Big\{\exp\Big[\int_0^T \Big(\frac{1}{2}\big(\phi_1^2(t) + \phi_2^2(t) + \phi_3^2(t)\big) + h^p(\phi_4(t)\ln\phi_4(t) - \phi_4(t) + 1)\Big)\mathrm{d}t\Big]\Big\} < \infty$.

Define a probability measure Q^{ϕ} equivalent to the probability measure P on $\mathcal{G}_T$ by

$$\frac{\mathrm{d}Q^{\phi}}{\mathrm{d}P}\Big|_{\mathcal{G}_T} = \Lambda^{\phi}(T).$$

By Girsanov's theorem, for $i = 1,2,3$, and $0 \le t \le T$, the processes

$$\mathrm{d}W_i^{Q^{\phi}}(t) = \mathrm{d}W_i(t) + \phi_i(t)\mathrm{d}t, \tag{14}$$

are three-dimensional standard Brownian motions under the probability measure Q^{ϕ}. Furthermore, Girsanov's theorem implies that

$$\widetilde{M}^{Q^{\phi}}(t) = Z(t) - \int_0^t \phi_4(u)h^p(1-Z(u))\mathrm{d}u \tag{15}$$

is a Q^{ϕ}-local martingale. For convenience, we still use h^Q to denote the arrival intensity of the default under the probability measure Q^{ϕ}.

Furthermore, the price dynamics of the domestic and foreign stocks under the probability measure Q^{ϕ} can be written as

$$\begin{aligned}
\mathrm{d}S^{\mathrm{d}}(t) &= S^{\mathrm{d}}(t)\Big((\mu_{\mathrm{d}} - \phi_1(t)\sigma_{\mathrm{d}})\mathrm{d}t + \sigma_{\mathrm{d}}\mathrm{d}W_1^{Q^{\phi}}(t)\Big), \\
\mathrm{d}S(t) &= S(t)\Big[\Big(\bar{\mu} - \rho_1\theta\phi_1(t) - (\sigma_{\mathrm{f}} + \rho_2\theta)\phi_2(t) - \sqrt{1-\rho_1^2-\rho_2^2}\theta\phi_3(t)\Big)\mathrm{d}t \\
&\quad + \rho_1\theta\mathrm{d}W_1^{Q^{\phi}}(t) + (\sigma_{\mathrm{f}} + \rho_2\theta)\mathrm{d}W_2^{Q^{\phi}}(t) + \sqrt{1-\rho_1^2-\rho_2^2}\theta\mathrm{d}W_3^{Q^{\phi}}(t)\Big].
\end{aligned} \tag{16}$$

Substituting (14) into (11), we have

$$\begin{aligned}
\mathrm{d}X^{\pi}(t) &= \Big[X^{\pi}(t)r_{\mathrm{d}} + \pi_s^{\mathrm{d}}(t)(\mu_{\mathrm{d}} - r_{\mathrm{d}}) + \pi_s(t)(\mu - r_{\mathrm{f}} + \rho_2\theta\sigma_{\mathrm{f}}) + \pi_p(t)\delta(1-Z(t)) \\
&\quad - (\pi_s^{\mathrm{d}}(t)\sigma_{\mathrm{d}} + \pi_s(t)\rho_1\theta)\phi_1(t) - \pi_s(t)(\sigma_{\mathrm{f}} + \rho_2\theta)\phi_2(t) - \pi_s(t)\sqrt{1-\rho_1^2-\rho_2^2}\theta\phi_3(t)\Big]\mathrm{d}t \\
&\quad + (\pi_s^{\mathrm{d}}(t)\sigma_{\mathrm{d}} + \pi_s(t)\rho_1\theta)\mathrm{d}W_1^{Q^{\phi}}(t) + \pi_s(t)(\sigma_{\mathrm{f}} + \rho_2\theta)\mathrm{d}W_2^{Q^{\phi}}(t) \\
&\quad + \pi_s(t)\sqrt{1-\rho_1^2-\rho_2^2}\theta\mathrm{d}W_3^{Q^{\phi}}(t) - \pi_p(t)(1-Z(t-))\zeta\mathrm{d}Z(t).
\end{aligned} \tag{17}$$

2.3. Robust Optimal Control

We adopt the method of robust optimal control presented in Anderson et al. [7]. The investor will choose the optimal investment decision to minimize the worst-case loss from a family of possible models, which is called the robust investment portfolio strategy. Suppose that the investor has a utility function $U(\cdot)$ over the terminal wealth W_T. He maximizes his expected utility of the terminal wealth by choosing an optimal investment strategy π in the risky and risk-free assets. Hence, the investor's indirect utility function is defined by:

$$J(t,x,z) = \sup_{\pi(t)\in\Pi}\left\{\inf_{\phi(t)\in\Phi} \mathrm{E}^{Q^{\phi}}\left[U(X^{\pi}(T)) + \int_t^T \Psi(u, X^{\pi}(u), \phi(u))\mathrm{d}u | X^{\pi}(t) = x, Z(t) = z\right]\right\}, \tag{18}$$

where the conditional expectation $\mathrm{E}^{Q^{\phi}}[\cdot]$ is given under the alternative probability measure Q^{ϕ} defined by ϕ. For an investor, he can determine the probability Q^{ϕ} by choosing ϕ. His purpose is considering the worst case, i.e, minimizing the expected utility. As in Branger and Larsen [12], $\Psi(u, X^{\pi}(u), \phi(u))$ is the penalty term of the model which is given by

$$\Psi(u,x,\phi(u)) = \sum_{i=1}^{3} \frac{\phi_i^2(u)}{2\varphi_i(u,x)} + \frac{h^p(\phi_4(u)\ln\phi_4(u) - \phi_4(u) + 1)(1 - Z(u))}{\varphi_4(u,x)}, \tag{19}$$

$$\varphi_i(u,x) = -\frac{\beta_i}{mJ(u,x,z)}, i = 1,2,3,4. \tag{20}$$

where $\beta_i \geq 0$ is the ambiguity aversion coefficient with respect to diffusion and jump risk, which can be understood as robustness preference or model uncertainty aversion, which measures investor confidence in the model. In order to avoid having a too complicated model and calculation, we suppose $\beta_i = \beta$ for $i = 1,2,3,4$. In general, we can also suppose that the β_i values are different. In this case, we can use the same method to obtain the result. The larger the β value, the greater the investor disconfidence in the model. m is a risk aversion coefficient that measures the investor's aversion to the market risk.

By using the stochastic dynamic programming principle, the robust Hamilton–Jacobi–Bellmann equation developed in Anderson et al. [7] can be obtained. Then, the value function $J(t,x,z)$ satisfies the following HJB equation:

$$\sup_{\pi(t)\in\Pi}\inf_{\phi(t)\in\Phi}\left(D^{\Phi,\pi}J(t,x,z) - \frac{mJ(t,x,z)}{\beta}\left[\sum_{i=1}^{3}\frac{\phi_i^2(t)}{2} + (\phi_4(t)\ln\phi_4(t) - \phi_4(t) + 1)h^p(1-z)\right]\right) = 0, \tag{21}$$

with boundary condition $J(T,x,z) = U(x)$. The differential operator $D^{\Phi,\pi}J(t,x,z)$ is given by

$$\begin{aligned} D^{\Phi,\pi}J(t,x,z) = {} & J_t + J_x\Big\{xr_{\mathrm{d}} + \pi_s^{\mathrm{d}}(t)(\mu_{\mathrm{d}} - r_{\mathrm{d}}) + \pi_s(t)(\mu - r_{\mathrm{f}} + \rho_2\theta\sigma_{\mathrm{f}}) + \pi_p(t)\delta(1-z) \\ & - \Big[\pi_s^{\mathrm{d}}(t)\sigma_{\mathrm{d}} + \pi_s(t)\rho_1\theta\Big]\phi_1(t) - \pi_s(t)(\sigma_{\mathrm{f}} + \rho_2\theta)\phi_2(t) - \pi_s(t)\sqrt{1-\rho_1^2-\rho_2^2}\theta\phi_3(t)\Big\} \\ & + \frac{1}{2}J_{xx}\Big\{[\pi_s^{\mathrm{d}}(t)\sigma_{\mathrm{d}} + \pi_s(t)\rho_1\theta]^2 + \pi_s(t)^2(\sigma_{\mathrm{f}} + \rho_2\theta)^2 + \pi_s(t)^2(1-\rho_1^2-\rho_2^2)\theta^2\Big\} \\ & + \Big[J(t,x-\pi_p(t)\zeta(1-z),1) - J(t,x,z)\Big]h^p(1-z)\phi_4(t), \end{aligned} \tag{22}$$

where J_t, J_x and J_{xx} are the partial derivatives of the value function $J(t,x,z)$ with respect to the corresponding variables t and x.

3. CARA Utility

For an investor, the defaultable bond may default during the investment time periods $[0,T]$. In this section, we split the problem (18) into the post-default case $(z = 1)$ and the pre-default case $(z = 0)$. Moreover, we assume that the investor has constant absolute risk aversion utility function defined by:

$$U(x) = -\frac{1}{m}\mathrm{e}^{-mx}, \tag{23}$$

where $m > 0$ is the risk aversion coefficient. The pre-default and post-default value functions are assumed to satisfy the following:

$$\begin{aligned} J^0(t,x) &= J(t,x,0), \\ J^1(t,x) &= J(t,x,1). \end{aligned}$$

Next, we discuss the optimal investment strategies with exchange rate risk in both cases before and after default; the main results are shown in Theorems 1 and 2.

3.1. The Post-Default Case

In this subsection, we consider the optimal investment problem in the post-default case. If the default risk occurs before time t, that is $z = 1$. Hence, we write HJB Equation (21) as follows:

$$\sup_{\pi(t)\in\Pi}\inf_{\phi(t)\in\Phi}\left(D^{\Phi,\pi}J^1(t,x) - mJ^1(t,x)\sum_{i=1}^{3}\frac{\phi_i^2(t)}{2\beta}\right) = 0. \tag{24}$$

Recall from (22) and $z = 1$, we have

$$\begin{aligned} &\sup_{\pi(t)\in\Pi}\inf_{\phi(t)\in\Phi}\Big\{J_t^1 + J_x^1\Big[xr_{\mathrm{d}} - \Big(\pi_s^{\mathrm{d}}(t)\sigma_{\mathrm{d}} + \pi_s(t)\rho_1\theta\Big)\phi_1(t) - \pi_s(t)(\sigma_{\mathrm{f}} + \rho_2\theta)\phi_2(t) \\ &+\pi_s^{\mathrm{d}}(t)(\mu_{\mathrm{d}} - r_{\mathrm{d}}) + \pi_s(t)(\mu - r_{\mathrm{f}} + \rho_2\theta\sigma_{\mathrm{f}}) - \pi_s(t)\sqrt{1-\rho_1^2-\rho_2^2}\theta\phi_3(t)\Big] - mJ^1(t,x)\sum_{i=1}^{3}\frac{\phi_i^2(t)}{2\beta} \\ &+\frac{1}{2}J_{xx}^1\Big[(\pi_s^{\mathrm{d}}(t)\sigma_{\mathrm{d}} + \pi_s(t)\rho_1\theta)^2 + \pi_s(t)^2(\sigma_{\mathrm{f}} + \rho_2\theta)^2 + \pi_s(t)^2(1-\rho_1^2-\rho_2^2)\theta^2\Big]\Big\} = 0. \end{aligned} \tag{25}$$

Theorem 1. *Under the exponential utility function, the post-default value function $J^1(t,x)$ is given by*

$$J^1(t,x) = -\frac{1}{m}\mathrm{e}^{-mx\mathrm{e}^{r_{\mathrm{d}}(T-t)}+A(t)}, \tag{26}$$

where

$$\begin{aligned} A(t) &= \frac{m(T-t)}{2\sigma_{\mathrm{d}}^2(m+\beta)}\Big\{(\mu_{\mathrm{d}} - r_{\mathrm{d}})^2 - M(\mu - r_{\mathrm{f}} + \rho_2\theta\sigma_{\mathrm{f}})\sigma_{\mathrm{d}} \\ &+ M(\mu_{\mathrm{d}} + r_{\mathrm{d}})\rho_1\theta - 2\sigma_{\mathrm{d}}^2 M\left[\frac{\rho_1\theta}{\sigma_{\mathrm{d}}} - (\mu - r_{\mathrm{f}} + \rho_2\theta\sigma_{\mathrm{f}})\right]\Big\}, \end{aligned} \tag{27}$$

and

$$M = \frac{(\mu - r_{\mathrm{f}} + \rho_2\theta\sigma_{\mathrm{f}})\sigma_{\mathrm{d}} - (\mu_{\mathrm{d}} + r_{\mathrm{d}})\rho_1\theta}{\sigma_{\mathrm{f}}^2 + 2\sigma_{\mathrm{f}}\rho_2\theta + (1-\rho_1^2)\theta^2}. \tag{28}$$

The optimal investment strategies are given as follows:

$$\begin{cases} \pi_s^{\mathrm{d}*}(t) &= \frac{(\mu_{\mathrm{d}} - r_{\mathrm{d}} - \rho_1\theta M)e^{-r_{\mathrm{d}}(T-t)}}{(m+\beta)\sigma_{\mathrm{d}}^2}, \\ \pi_s^*(t) &= \frac{Me^{-r_{\mathrm{d}}(T-t)}}{(m+\beta)\sigma_{\mathrm{d}}}, \\ \pi_p^*(t) &= 0. \end{cases} \tag{29}$$

Proof. We conjecture that the post-default value function $J^1(t,x)$ has the following form:

$$J^1(t,x) = -\frac{1}{m}\mathrm{e}^{-mx\mathrm{e}^{r_{\mathrm{d}}(T-t)}+A(t)}, \tag{30}$$

with boundary condition $A(T) = 0$.

The first-order partial derivative of $J^1(t,x)$ for t, the first and second-order partial derivatives of $J^1(t,x)$ for x are now calculated as follows:

$$\begin{cases} \frac{\partial J^1(t,x)}{\partial t} & = J^1(t,x)[r_{\mathrm{d}} mx\mathrm{e}^{r_{\mathrm{d}}(T-t)} + A_t], \\ \frac{\partial J^1(t,x)}{\partial x} & = J^1(t,x)[-m\mathrm{e}^{r_{\mathrm{d}}(T-t)}], \\ \frac{\partial^2 J^1(t,x)}{\partial x^2} & = J^1(t,x)[m^2\mathrm{e}^{2r_{\mathrm{d}}(T-t)}], \end{cases} \tag{31}$$

where A_t is the first-order partial derivative of $A(t)$ about t.

Substituting (30) and (31) into (25) and using the first-order optimal condition for $\phi(t)$, we have

$$\begin{cases} \phi_1^*(t) & = \mathrm{e}^{r_{\mathrm{d}}(T-t)}(\pi_s^{\mathrm{d}}(t)\sigma_{\mathrm{d}} + \pi_s(t)\rho_1\theta)\beta, \\ \phi_2^*(t) & = \mathrm{e}^{r_{\mathrm{d}}(T-t)}\pi_s(t)(\sigma_{\mathrm{f}} + \rho_2\theta)\beta, \\ \phi_3^*(t) & = \mathrm{e}^{r_{\mathrm{d}}(T-t)}\pi_s(t)\sqrt{1-\rho_1^2-\rho_2^2}\theta\beta. \end{cases} \tag{32}$$

Plugging (32) into (24), we obtain the following equation:

$$\begin{aligned} \sup_{\pi(t)\in\Pi}\Big\{ & A_t - m\mathrm{e}^{r_{\mathrm{d}}(T-t)}\Big[\pi_s^{\mathrm{d}}(t)(\mu_{\mathrm{d}} - r_{\mathrm{d}}) + \pi_s(t)(\mu - r_{\mathrm{f}} + \rho_2\theta\sigma_{\mathrm{f}})\Big] + \frac{1}{2}m(m+\beta)\mathrm{e}^{2r_{\mathrm{d}}(T-t)} \\ & \Big[(\pi_s^{\mathrm{d}}(t)\sigma_{\mathrm{d}} + \pi_s(t)\rho_1\theta)^2 + \pi_s(t)^2(\sigma_{\mathrm{f}} + \rho_2\theta)^2 + \pi_s(t)^2(1-\rho_1^2-\rho_2^2)\theta^2\Big]\Big\} = 0. \end{aligned} \tag{33}$$

According to the first-order optimal condition, $\pi^*(t) = \{\pi_s^{\mathrm{d}*}(t), \pi_s^*(t), \pi_p^*(t)\}$ are obtained as follows:

$$\begin{cases} \pi_s^{\mathrm{d}*}(t) & = \frac{(\mu_{\mathrm{d}} - r_{\mathrm{d}} - \rho_1\theta M)e^{-r_{\mathrm{d}}(T-t)}}{(m+\beta)\sigma_{\mathrm{d}}^2}, \\ \pi_s^*(t) & = \frac{Me^{-r_{\mathrm{d}}(T-t)}}{(m+\beta)\sigma_{\mathrm{d}}}, \\ \pi_p^*(t) & = 0. \end{cases} \tag{34}$$

Then, we substitute (34) into (32) and obtain

$$\begin{cases} \phi_1^*(t) & = \frac{(\mu_{\mathrm{d}} - r_{\mathrm{d}})\beta}{\sigma_{\mathrm{d}}(m+\beta)}, \\ \phi_2^*(t) & = \frac{M(\sigma_{\mathrm{f}} + \rho_2\theta)\beta}{(m+\beta)\sigma_{\mathrm{d}}}, \\ \phi_3^*(t) & = \frac{M\sqrt{1-\rho_1^2-\rho_2^2}\theta\beta}{(m+\beta)\sigma_d\mathrm{d}}. \end{cases} \tag{35}$$

Next, using (33), (34) and (35), we obtain the following equality

$$\begin{aligned} A_t - \frac{m}{2\sigma_{\mathrm{d}}^2(m+\beta)} & \times \Big[(\mu_{\mathrm{d}} - r_{\mathrm{d}})^2 - (\mu - r_{\mathrm{f}} + \rho_2\theta\sigma_{\mathrm{f}})\sigma_{\mathrm{d}} + (\mu_{\mathrm{d}} + r_{\mathrm{d}})\rho_1\theta\Big] \\ & + \frac{Mm}{m+\beta}\Big[\frac{\rho_1\theta}{\sigma_{\mathrm{d}}} - (\mu - r_{\mathrm{f}} + \rho_2\theta\sigma_{\mathrm{f}})\Big] = 0. \end{aligned} \tag{36}$$

Finally, we obtain (27) and complete the proof by solving the differential Equation (36). □

3.2. The Pre-Default Case

In this subsection, we present the robust optimal investment strategies in the pre-default case. We first provide a Lemma which is used to prove the Theorem 2.

Lemma 1. *Let the function $F(x) = h^p x + \frac{mh^p}{\beta} x\ln x - h^Q$; then, the equation $F(x) = 0$ has a unique positive solution.*

Proof. Taking the derivative for $F(x)$ with respect to the variable x, we get

$$\frac{\partial F(x)}{\partial x} = h^p + \frac{mh^p}{\beta}(\ln x + 1).$$

Hence, it is easy to find that $F(x)$ is a decreasing function when $x \in (0, e^{-(1+\frac{\beta}{m})})$, and it is an increasing function when $x \in [e^{-(1+\frac{\beta}{m})}, +\infty)$.

Noting that

$$\lim_{x\to 0^+} F(x) = -h^Q < 0; \lim_{x\to+\infty} F(x) = +\infty.$$

Then, $F(x)$ has a unique positive solution at $(0, +\infty)$ according to the zero-point theorem. □

Theorem 2. *Under the exponential utility function, the pre-default value function $J^0(t,x)$ is given by:*

$$J^0(t,x) = -\frac{1}{m}\mathrm{e}^{-mx\mathrm{e}^{r_\mathrm{d}(T-t)}+C(t)}, \tag{37}$$

$$C(t) = A(t)\mathrm{e}^{h^Q(T-t)} + \int_t^T H(s)\mathrm{e}^{h^Q(T-s)}\mathrm{d}s, \tag{38}$$

where $A(t)$ is defined by (27) and

$$H(t) = -h^Q \ln\frac{h^Q}{h^p\phi_4^*(t)} - \frac{mh^p}{\beta}(1-\phi_4^*(t)). \tag{39}$$

Furthermore, the optimal investment strategies are given by

$$\begin{cases} \pi_s^{\mathrm{d}*}(t) &= \frac{(\mu_\mathrm{d}-r_\mathrm{d}-\rho_1\theta M)\mathrm{e}^{-r_\mathrm{d}(T-t)}}{(m+\beta)\sigma_\mathrm{d}^2}, \\ \pi_s^*(t) &= \frac{M\mathrm{e}^{-r_\mathrm{d}(T-t)}}{(m+\beta)\sigma_\mathrm{d}}, \\ \pi_p^*(t) &= \frac{\ln\frac{h^Q}{\zeta h^p\phi_4^*(t)}+C(t)-A(t)}{m\zeta\mathrm{e}^{r_\mathrm{d}(T-t)}}. \end{cases} \tag{40}$$

Proof. We will use the same idea from the proof of Theorem 1. Let $z = 0$; then, the HJB Equation (21) is written as

$$\sup_{\pi(t)\in\Pi}\inf_{\phi(t)\in\Phi}\left(D^{\theta,\pi}J^0(t,x) - \frac{mJ^0(t,x)[\phi_1^2(t)+\phi_2^2(t)+\phi_3^2(t)]}{2\beta} - \frac{mJ^0(t,x)h^p[\phi_4(t)\ln\phi_4(t)-\phi_4(t)+1]}{\beta}\right) = 0. \tag{41}$$

where

$$\begin{aligned} D^{\theta,\pi}J^0(t,x) &= J_t^0 + J_x^0\Big[xr_\mathrm{d} + \pi_s^\mathrm{d}(t)(\mu_\mathrm{d}-r_\mathrm{d}) + \pi_s(t)(\mu - r_\mathrm{f} + \rho_2\theta\sigma_\mathrm{f}) + \pi_p(t)\delta \\ &\quad -(\pi_s^\mathrm{d}(t)\sigma_\mathrm{d} + \pi_s(t)\rho_1\theta)\phi_1(t) - \pi_s(t)(\sigma_\mathrm{f}+\rho_2\theta)\phi_2(t) - \pi_s(t)\sqrt{1-\rho_1^2-\rho_2^2}\theta\phi_3(t)\Big] \\ &\quad +\frac{1}{2}J_{xx}^0\Big[(\pi_s^\mathrm{d}(t)\sigma_\mathrm{d})^2 + 2\pi_s^\mathrm{d}(t)\pi_s(t)\sigma_\mathrm{d}\rho_1\theta + \pi_s(t)^2\Big(\sigma_\mathrm{f}^2 + 2\rho_2\theta\sigma_\mathrm{f} + \theta^2\Big)\Big] \\ &\quad +\Big[J^1(t,x-\pi_p(t)\zeta) - J^0(t,x)\Big]h^p\phi_4(t) \end{aligned}$$

Recalling the (30), we obtain

$$J^1(t,x-\pi_p(t)\zeta)-J^0(t,x)=J^0(t,x)\Big[\exp\{m\pi_p(t)\zeta \mathrm{e}^{r_\mathrm{d}(T-t)}-C(t)+A(t)\}-1\Big]. \quad (42)$$

Hence, (41) can be rewritten as

$$\begin{aligned}
&\sup_{\pi(t)\in\Pi}\inf_{\phi(t)\in\Phi}\Big\{J_t^0+J_x^0\Big[xr_\mathrm{d}+\pi_s^\mathrm{d}(t)(\mu_\mathrm{d}-r_\mathrm{d})+\pi_s(t)(\mu-r_\mathrm{f}+\rho_2\theta\sigma_\mathrm{f})+\pi_p(t)\delta\\
&-(\pi_s^\mathrm{d}(t)\sigma_\mathrm{d}+\pi_s(t)\rho_1\theta)\phi_1(t)-\pi_s(t)(\sigma_\mathrm{f}+\rho_2\theta)\phi_2(t)-\pi_s(t)\sqrt{1-\rho_1^2-\rho_2^2}\theta\phi_3(t)\Big]\\
&+\frac{1}{2}J_{xx}^0\Big[(\pi_s^\mathrm{d}(t)\sigma_\mathrm{d})^2+2\pi_s^\mathrm{d}(t)\pi_s(t)\sigma_\mathrm{d}\rho_1\theta+\pi_s(t)^2\Big(\sigma_\mathrm{f}^2+2\rho_2\theta\sigma_\mathrm{f}+\theta^2\Big)\Big] \quad (43)\\
&+J^0(t,x)\Big[\exp\{m\pi_p(t)\zeta \mathrm{e}^{r_\mathrm{d}(T-t)}-C(t)+A(t)\}-1\Big]h^p\phi_4(t)\\
&-\frac{mJ^0(t,x)[\phi_1^2(t)+\phi_2^2(t)+\phi_3^2(t)]}{2\beta}-\frac{mJ^0(t,x)h^p[\phi_4(t)\ln\phi_4(t)-\phi_4(t)+1]}{\beta}\Big\}=0.
\end{aligned}$$

We guess that the explicit solution of the HJB Equation (43) satisfies the following

$$J^0(t,x)=-\frac{1}{m}\mathrm{e}^{-mx\mathrm{e}^{r_\mathrm{d}(T-t)}+C(t)}, \quad (44)$$

with terminal condition $C(T)=0$. Taking the derivatives of $J^0(t,x)$ with respect to the variables t and x, we have

$$\begin{cases}
\frac{\partial J^0(t,x)}{\partial t} &= J^0(t,x)[r_\mathrm{d}mx\mathrm{e}^{r_\mathrm{d}(T-t)}+C_t],\\
\frac{\partial J^0(t,x)}{\partial x} &= J^0(t,x)[-m\mathrm{e}^{r_\mathrm{d}(T-t)}],\\
\frac{\partial J^0(t,x)}{\partial x^2} &= J^0(t,x)[m^2\mathrm{e}^{2r_\mathrm{d}(T-t)}],
\end{cases} \quad (45)$$

where B_t is the first derivative of $B(t)$ about t.

Substituting (45) into (43), we obtain the following results by the first-order optimal condition

$$\begin{cases}
\phi_1^*(t) &= \mathrm{e}^{r_\mathrm{d}(T-t)}\Big(\pi_s^\mathrm{d}(t)\sigma_\mathrm{d}+\pi_s(t)\rho_1\theta\Big)\beta,\\
\phi_2^*(t) &= \mathrm{e}^{r_\mathrm{d}(T-t)}\pi_s(t)(\sigma_\mathrm{f}+\rho_2\theta)\beta,\\
\phi_3^*(t) &= \mathrm{e}^{r_\mathrm{d}(T-t)}\pi_s(t)\sqrt{1-\rho_1^2-\rho_2^2}\theta\beta,
\end{cases} \quad (46)$$

and

$$\Big[\exp\{m\pi_p(t)\zeta \mathrm{e}^{r_\mathrm{d}(T-t)}-C(t)+A(t)\}-1\Big]h^p-\frac{mh^p}{\beta}\ln\phi_4^*(t)=0. \quad (47)$$

We apply the first-order optimal condition, (46) and (47) for (43), the optimal investment strategies are obtained as follows

$$\begin{cases}
\pi_s^{\mathrm{d}*}(t) &= \frac{(\mu_\mathrm{d}-r_\mathrm{d}-\rho_1\theta M)\mathrm{e}^{-r_\mathrm{d}(T-t)}}{(m+\beta)\sigma_\mathrm{d}^2},\\
\pi_s^*(t) &= \frac{M\mathrm{e}^{-r_\mathrm{d}(T-t)}}{(m+\beta)\sigma_\mathrm{d}},\\
\pi_p^*(t) &= \frac{\ln\frac{h^Q}{\zeta h^p\phi_4^*(t)}+C(t)-A(t)}{m\zeta\mathrm{e}^{r_\mathrm{d}(T-t)}}.
\end{cases} \quad (48)$$

Combining (46), (47) and (48), we conclude that

$$\begin{cases} \phi_1^*(t) &= \frac{(\mu_{\mathrm{d}}-r_{\mathrm{d}})\beta}{\sigma_{\mathrm{d}}(m+\beta)}, \\ \phi_2^*(t) &= \frac{M(\sigma_{\mathrm{f}}+\rho_2\theta)\beta}{(m+\beta)\sigma_{\mathrm{d}}}, \\ \phi_3^*(t) &= \frac{M\sqrt{1-\rho_1^2-\rho_2^2}\theta\beta}{(m+\beta)\sigma_{\mathrm{d}}}, \end{cases} \tag{49}$$

and

$$h^{\mathrm{Q}} - h^p\phi_4^*(t) - \frac{mh^p}{\beta}\phi_4^*(t)\ln\phi_4^*(t) = 0. \tag{50}$$

Furthermore, Lemma 1 shows that Equation (50) has a unique position root. Then, we substitute (45), (48) and (49) and (50) into (43), and a differential equation for $C(t)$ is obtained as follows:

$$\begin{aligned} &C_t - h^{\mathrm{Q}}C(t) - \frac{m}{2\sigma_{\mathrm{d}}^2(m+\beta)}\Big\{(\mu_{\mathrm{d}}-r_{\mathrm{d}})^2 - M[(\mu-r_{\mathrm{f}}+\rho_2\theta\sigma_{\mathrm{f}})\sigma_{\mathrm{d}} - (\mu_{\mathrm{d}}+r_{\mathrm{d}})\rho_1\theta]\Big\} \\ &+\frac{Mm}{m+\beta}\Big[\frac{\rho_1\theta}{\sigma_{\mathrm{d}}} - (\mu-r_{\mathrm{f}}+\rho_2\theta\sigma_{\mathrm{f}})\Big] + A(t)h^{\mathrm{Q}} - h^{\mathrm{Q}}\ln\frac{h^{\mathrm{Q}}}{h^p\phi_4^*(t)} - \frac{mh^p}{\beta}(1-\phi_4^*(t)) = 0, \end{aligned} \tag{51}$$

where $C(T) = 0$. By solving the differential Equation (51), we obtain (38). This completes the proof. □

We put the pre-default and post-default together and obtain the following result.

Theorem 3. *When the investors have ambiguity aversion, the optimal investment strategies with default risk and exchange rate risk are given as follows*

$$\pi_s^{\mathrm{d}*}(t) = \frac{(\mu_{\mathrm{d}}-r_{\mathrm{d}}-\rho_1\theta M)\mathrm{e}^{-r_{\mathrm{d}}(T-t)}}{(m+\beta)\sigma_{\mathrm{d}}^2}, \quad \pi_s^*(t) = \frac{M\mathrm{e}^{-r_{\mathrm{d}}(T-t)}}{(m+\beta)\sigma_{\mathrm{d}}},$$

and

$$\begin{cases} \pi_p^*(t) = 0, \quad 0 < \tau \le t, \\ \pi_p^*(t) = \frac{\ln\frac{h^{\mathrm{Q}}}{\zeta h^p\phi_4^*(t)} + \mathrm{C}(t) - A(t)}{m\zeta\mathrm{e}^{r_{\mathrm{d}}(T-t)}}, \quad \tau \in (t,T], \end{cases}$$

where $A(t)$, M, $C(t)$ and $\phi_4^*(t)$ are given by (27), (28), (38) and (50), respectively.

Remark 3. *If the investors are ambiguity neutral, i.e., $\beta = 0$, then the optimal investment strategies satisfies*

$$\pi_s^{\mathrm{d}*}(t) = \frac{(\mu_{\mathrm{d}}-r_{\mathrm{d}}-\rho_1\theta M)\mathrm{e}^{-r_{\mathrm{d}}(T-t)}}{m\sigma_{\mathrm{d}}^2}, \quad \pi_s^*(t) = \frac{M\mathrm{e}^{-r_{\mathrm{d}}(T-t)}}{m},$$

and

$$\begin{cases} \pi_p^*(t) = 0, \quad 0 < \tau \le t, \\ \pi_p^*(t) = \frac{\ln\frac{\delta}{\zeta h^p\phi_4^*(t)} + \mathrm{C}(t) - A(t)}{m\zeta\mathrm{e}^{r_{\mathrm{d}}(T-t)}}, \quad \tau \in (t,T], \end{cases}$$

where $A(t)$, M, $C(t)$ and $\phi_4^*(t)$ are given by (27), (28), (38) and (50), respectively.

4. Numerical Analysis

In this section, we provide some numerical results and analyze the impact of some financial model parameters on the robust optimal investment strategies. For convenience, we consider the value of the robust optimal investment strategies at time $t = 0$. In general, we can also suppose that t is a positive constant. In this case, we can obtain the numerical results of the optimal investment strategies by using the same method. To be more specific, we assume that unless stated otherwise, the model's parameters take the following values:

$$r_{\mathrm{d}} = 0.05, \theta = 0.1, \mu_{\mathrm{d}} = 0.08, \mu_{\mathrm{f}} = 0.1, \hat{\mu} = 0.05, \sigma_{\mathrm{d}} = 0.02, \sigma_{\mathrm{f}} = 0.25, \rho_1 = 0.2, \rho_2 = -0.2,$$

$$\mu_p = 0.4, T = 10, t = 0, \beta = 1, m = 2, r_f = 0.04, h^p = 0.4, \delta = 0.2, \zeta = 0.4.$$

Figure 1 depicts the plots of the robust optimal investment strategies against the volatilities σ_d and θ. From the left panel of Figure 1, we can find that the value of bond B held in the optimal investment strategy, denoted by $\pi_s^d(t)$, becomes much less as the volatility σ_d increases. The greater the volatility, the greater the risk of stock price. Hence, the investors will naturally reduce their investment in domestic stocks when σ_{d} increases. In addition, the right panel of Figure 1 describes the relationship between θ and $\pi_s^{\mathrm{d}}(t)$. We can observe that the optimal investment strategy $\pi_s^{\mathrm{d}}(t)$ is an increasing function with respect to the variable θ. When the volatility of the exchange rate risk increases, the foreign stock price risk will also increase; thus, the investors will reduce investment in foreign stocks and increase investment in domestic stocks. Finally, we also compare the optimal investment strategy in the ambiguity averse case with the corresponding strategy in the ambiguity neutral case. In Figure 1, $\beta = 1$ denotes ambiguity aversion, and $\beta = 0$ means ambiguity neutral. We can see that the value of the optimal investment strategy $\pi_s^{\mathrm{d}}(t)$ at $\beta = 0$ is greater than the value at $\beta = 1$. This verifies that the investors with ambiguity aversion will invest less in risky assets compared with those who are ambiguity neutral. The reason is that investors with ambiguity aversion are more risk averse, which brings less money to the risky assets.

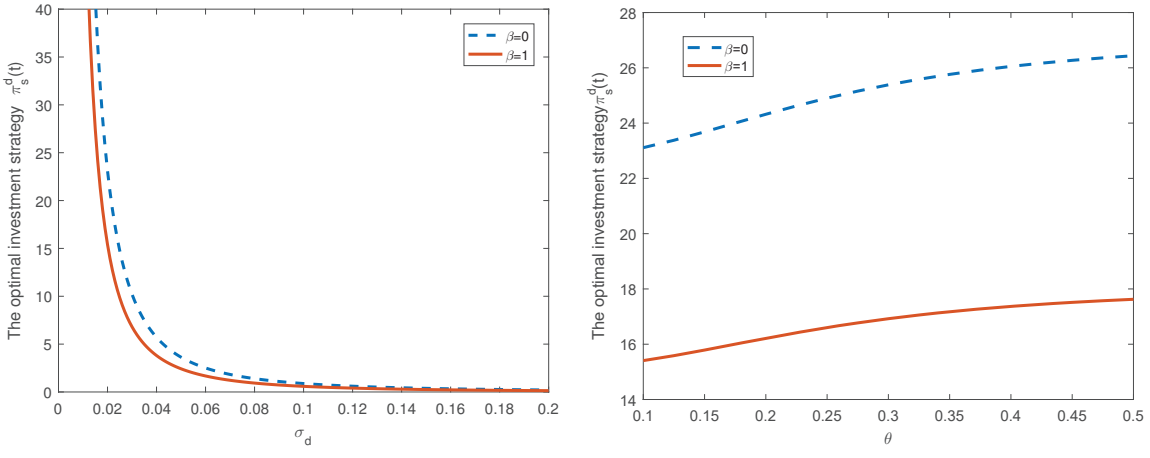

Figure 1. The impact of parameters σ_{d} and θ on $\pi_s^{\mathrm{d}}(t)$.

Figure 2 shows how the optimal investment strategy $\pi_s^{\mathrm{d}}(t)$ adjusts in response to the change of m and μ_d. In Figure 2, m is a risk aversion coefficient; it measures the level of investor's aversion to risk. As shown in the left panel of Figure 2, the optimal investment strategy $\pi_s^{\mathrm{d}}(t)$ decreases as the risk aversion coefficient m increases. This is because the investors will reduce their investment amount of the risky asset when the investors with a lower risk aversion parameter detest risk more. μ_d is the appreciate rate of the stock S^d. Since a high appreciate rate μ_d leads to a high yield, we think that there is a positive

relationship between μ_d and $\pi_s^d(t)$. Indeed, it can be seen from the right panel of Figure 3 that $\pi_s^d(t)$ increases as μ_d increases. Hence, this result is also consistent with our conjecture.

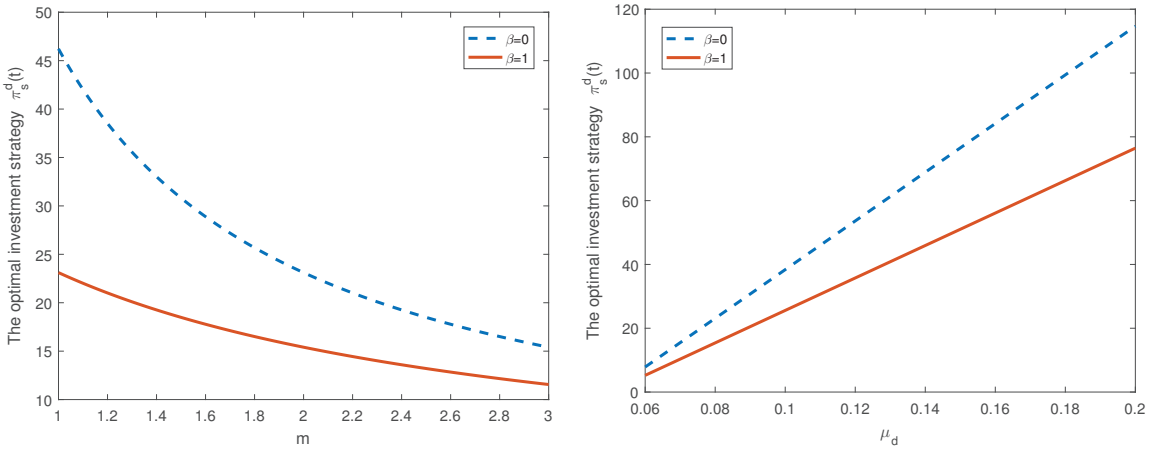

Figure 2. The impact of parameters m and μ_d on $\pi_s^\mathrm{d}(t)$.

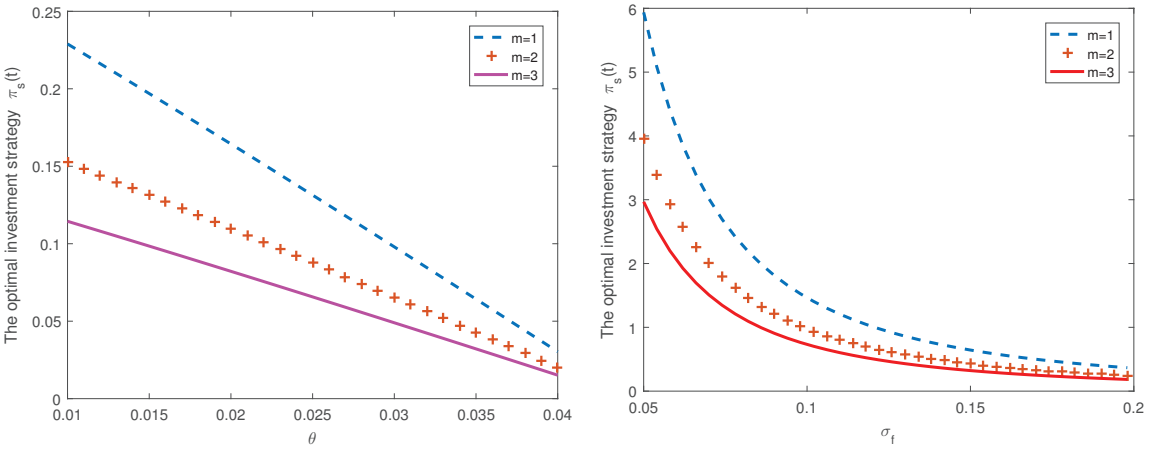

Figure 3. The impact of parameters m, θ and σ_f on $\pi_s(t)$.

Figure 3 presents the influence of different θ, m and σ_f values on the optimal investment strategy $\pi_s(t)$. We observe that the negative correlation between θ and the optimal investment strategy $\pi_s(t)$ from the left panel of Figure 3. The cause for such results is very obvious, the larger θ means the greater volatility of exchange rate risk. The increased volatility of exchange rate risk may lead to investment losses; then, the investors will reduce the investment amount in the foreign stocks. Compared with Figure 2, we conclude that the impact of m on the optimal investment strategy $\pi_s(t)$ is the same as the influence of the optimal investment strategy $\pi_s^\mathrm{d}(t)$. Moreover, we can also find that there is a negative relationship between σ_f and the optimal investment strategy $\pi_s(t)$. It implies that the volatility risk has a significant impact on the number of risky assets held in the optimal investment strategy $\pi_s(t)$.

From Figure 4, we find that the increase of both domestic and foreign risk-free interest rates will reduce the amount of investment in the risky assets. This is because the investors will buy more risk-free assets but less risky assets when the risk-free interest rate increases. Compared with Figure 3, we can find that the investors will reduce their investment in

foreign risky assets when the volatility of foreign risky assets increases, and the investors will choose to buy more foreign risky assets when the volatility of domestic risky assets increases. The reason is also obvious. When the volatility of domestic risky assets increases, the domestic risky assets will have a greater market price risk, and the investors will transfer part of their investment to foreign risky assets.

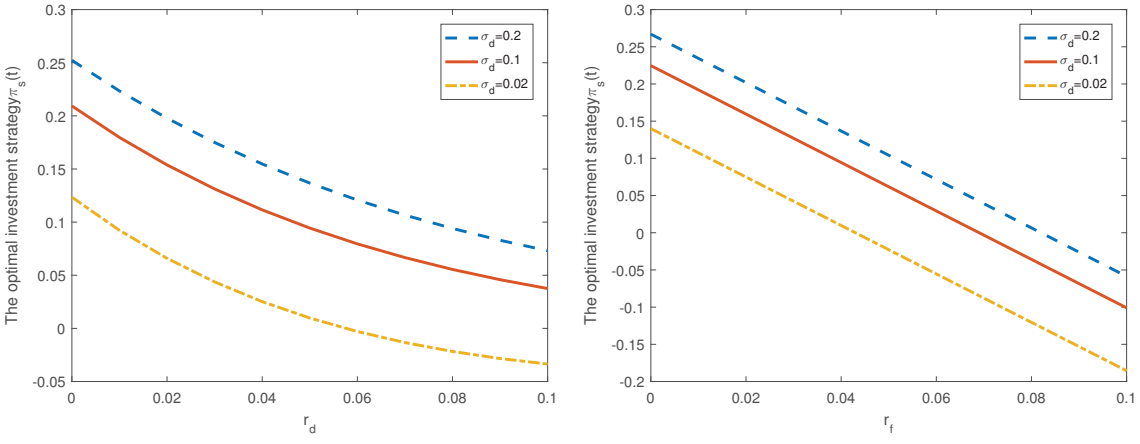

Figure 4. The impact of parameters r_d, r_f and σ_d on $\pi_s(t)$.

In Figure 5, we study the effect of model parameters δ, ζ, r_d and m on the optimal investment strategy $\pi_p(t)$ in the blue-default case. From the left panel of Figure 5, we find that the number of defaultable bonds held in the optimal investment strategy will decrease as the loss rate ζ increases. It is obvious that a higher loss rate will lead to more potential loss. In order to minimize the investment risk, it is reasonable to reduce the holding positions in defaultable bonds. Furthermore, we also observe that there is a positive relationship between $\pi_p(t)$ and δ. Our explanation is as follows: $\delta = \zeta h^Q$ is the default risk spread under the probability measure Q^ϕ, where h^Q is the default intensity. A higher default intensity will lead to a higher yield before the default occurs; thus, the investors prefer to purchase more defaultable bonds as the default intensity h^Q increases. Hence, $\pi_p(t)$ is an increase function with respect to δ when ζ is fixed. The right panel of Figure 5 plots the graph of the optimal trading strategy $\pi_p(t)$ invested in defaultable bonds as a function of the risk-free interest rate r_d. We observe that the optimal investment strategy is decreasing with respect to the risk-free interest rate r_d; it means that the investors should reduce the defaultable bond investment when the risk-free interest rate r_d is increasing. This behavior of the optimal investment strategy is consistent with economic intuition. Moreover, we also present the numerical results of the optimal trading strategy $\pi_p(t)$ by varying the risk aversion coefficient m. We can find that the optimal trading strategy $\pi_p(t)$ is a decreasing function with respect to the risk aversion coefficient m. This may be interpreted as follows: When m is increasing, i.e., the investors are more risk-averse, they will reduce the investment in defaultable bonds.

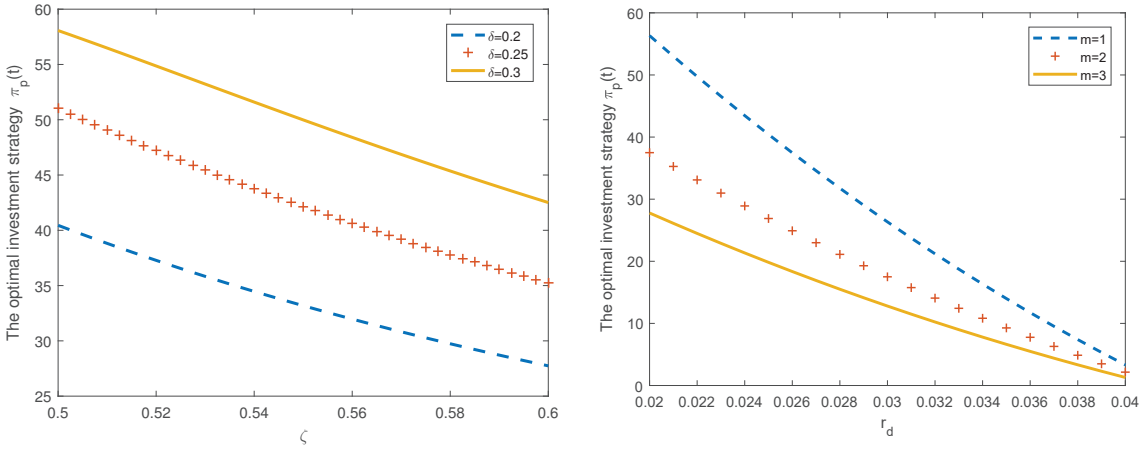

Figure 5. The impact of parameters δ, ζ, r_d and m on $\pi_p(t)$.

5. Conclusions

This article focuses on an optimal investment problem with exchange rate risk and default risk when the investors are ambiguity averse. The price dynamics of exchange rate and domestic and foreign stocks are modeled by the Geometric Brownian motions, and meanwhile, the defaultable price process follows a jump process. To obtain the explicit expression of optimal investment strategy, an optimal portfolio problem framework with ambiguity aversion is first set up by using the robust control method. Second, the optimal investment problem is transformed to the corresponding HJB equation by the dynamic principle. Due to the existence of default risk, the HJB equation is usually too complicated to solve. Hence, we divide the HJB equation into the pre-default case and the post-default case. Finally, we derive the analytical solutions of the optimal investment strategies and the value functions by solving two HJB equations with the first order optimal condition. We find that the model uncertainty has a significant effect on the optimal investment strategies, and the investors with ambiguity aversion prefer to invest less risky assets than that of the investors who are ambiguity neutral. Moreover, we illustrate that if the volatility of the exchange rate risk increases, the investors will reduce their investment in foreign risky assets and meanwhile increase investment in domestic risky assets. This implies that an international investor must not ignore the exchange rate risk. In particular, our results also show that the optimal investment strategies are affected by the intensity of the default risk spread in the pre-default case.

The method in our work can also be used to solve other optimal investment portfolio problems involving default risk and exchange risk and model uncertainty. This article assumes that the price dynamics of the stocks and exchange rate follows the geometric Brownian motion. However, the implied volatility curve looks like a "smile", not a constant. In the future research, we will incorporate the stochastic volatility into the stocks and exchange rate price dynamics to capture the volatility smile phenomena. Furthermore, this article shows that ambiguity aversion, exchange rate risk and default risk have a significant effect on the optimal investment strategies by some numerical results. It is challenging but necessary to analyze the impact of these factors on the optimal investment strategy through empirical analysis in future research.

Author Contributions: All authors have equally contributed to this paper. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the Humanity and Social Science Youth Foundation of the Ministry of Education of China (18YJC910012), the National Natural Science Foundation of China (12226510), and the Humanity and Social Science Foundation of Ningbo University (XPYB19002).

Data Availability Statement: Not applicable.

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

References

1. Merton, R.C. Lifetime portfolio selection under uncertainty: The continuous time case. *Rev. Econ. Stat.* **1969**, *51*, 247–257. [CrossRef]
2. Merton, R.C. Optimal consumption and portfolio rules in a continuous time model. *J. Econ. Theory* **1971**, *3*, 373–413. [CrossRef]
3. Ma, Y.; Shan, S.P.; Xu, W.D. Optimal investment and consumption in the market with jump risk and capital gains tax. *J. Ind. Manag. Optim.* **2018**, *15*, 1937–1953. [CrossRef]
4. Sun, Z.Y.; Guo, J.Y. Optimal mean variance investment and reinsurance problem for an insurer with stochastic volatility. *Math. Methods Oper. Res.* **2018**, *88*, 59–79. [CrossRef]
5. Bi, J.; Chen, K. Optimal investment-reinsurance problems with common shock dependent risks under two kinds of premium principles. *RAIRO Oper. Res.* **2019**, *53*, 179–206. [CrossRef]
6. Wang, W.Y.; Muravey, D.; Shen, Y.; Zeng, Y. Optimal investment and reinsurance strategies under 4/2 stochastic volatility model. *Scand. Actuar. J.* **2022**, 1–37. [CrossRef]
7. Anderson, E.L.; Hansen, L.P.; Sargent, T.J. A quartet of semigroups for model specification, robustness, prices of risk, and model detection. *J. Eur. Econ. Assoc.* **2003**, *1*, 68–123. [CrossRef]
8. Maenhout, P.J. Robust portfolio rules and asset pricing. *Rev. Financ. Stud.* **2004**, *17*, 951–983. [CrossRef]
9. Elliott, R.J.; Siu, T.K. Robust optimal portfolio choice under Markovian regime-switching model. *Methodol. Comput. Appl. Probab.* **2009**, *11*, 145–157. [CrossRef]
10. Zheng, X.X.; Zhou, J.M.; Sun, Z.Y. Robust optimal portfolio and proportional reinsurance for an insurer under a CEV model. *Insur. Math. Econ.* **2016**, *67*, 77–87. [CrossRef]
11. Zeng, Y.; Li, D.; Chen, Z.; Yang, Z. Ambiguity aversion and optimal derivative-based pension investment with stochastic income and volatility. *J. Econ. Dyn. Control* **2018**, *88*, 70–103. [CrossRef]
12. Branger, N.; Larsen, L.S. Robust portfolio choice with uncertainty about jump and diffusion risk. *J. Bank. Financ.* **2013**, *37*, 5036–5047. [CrossRef]
13. Bielecki, T.; Jang, I. Portfolio optimization with a defaultable security. *Asia-Pac. Financ. Mark.* **2006**, *13*, 113–127. [CrossRef]
14. Bo, L.J.; Li, X.D.; Wang, Y.J. Optimal investment and consumption with default risk: HARA utility. *Aisa-Pac. Financ. Mark.* **2013**, *20*, 261–281. [CrossRef]
15. Capponi, A.; Figueroa-López, J.E. Dynamic portfolio optimization with a defaultable security and regime switching. *Math. Financ.* **2014**, *24*, 207–249. [CrossRef]
16. Deng, Y.C.; Li, M.; Huang, Y.; Zhou, J.M. Robust optimal strategies for an insurer under generalized mean-variance premium principle with defaultable bond. *Commun. Stat.-Theory Methods* **2021**, *50*, 5126–5159. [CrossRef]
17. Liu, H. Dynamic portfolio choice under ambiguity and regime switching mean returns. *J. Econ. Dyn. Control* **2011**, *35*, 623–640. [CrossRef]
18. Escobar, M.; Ferrando, S.; Rubtsov, A. Robust portfolio choice with derivative trading under stochastic volatility. *J. Bank. Financ.* **2015**, *61*, 142–157. [CrossRef]
19. Wang, N.; Zhang, N.; Zhuo, J. Robust non zero sum investment and reinsurance game with default risk. *Insur. Math. Econ.* **2014**, *84*, 115–132. [CrossRef]
20. Schoberl, M.; Siuka, A. Robust optimal proportional reinsurance and investment strategy for an insurer with defaultable risks and jumps. *J. Comput. Appl. Math.* **2013**, *356*, 46–66.
21. Kozo, K.; Shujiro, U. Exchange rate, Exchange rate volatility and foreign direct investment. *World Econ.* **2004**, *27*, 1501–1536.
22. Guo, C.; Zhuo, X.Y.; Constantinescu, C.; Pamen, O.M. Optimal reinsurance-investment strategy under risks of interest rate, exchange rate and inflation. *Methodol. Comput. Appl. Probab.* **2018**, *20*, 1477–1502. [CrossRef]
23. Fei, C.; Fei, W.Y.; Rui, Y.; Yan, L.T. International investment with exchange rate risk. *Asia-Pac. J. Account. Econ.* **2021**, *28*, 225–241. [CrossRef]
24. Amin, K.I.; Jarrow, R.A. Pricing foreign currency options under stochastic interest rates. *J. Int. Money Financ.* **1991**, *10*, 310–329. [CrossRef]
25. Melino, A.; Turnbull, S.M. Pricing foreign currency options with stochastic volatility. *J. Econom.* **1990**, *45*, 239–265. [CrossRef]
26. Johnson, G.; Schneeweis, T. Jump diffusion processes in the foreign exchange markets and the release of macroeconomic news. *Comput. Econ.* **1994**, *7*, 309–329. [CrossRef]
27. Huang, S.C.; Hung, M.W. Pricing foreign equity options under Lévy process. *J. Futur. Mark.* **2005**, *25*, 917–944. [CrossRef]

mathematics

MDPI

Article

Two Generalizations of the Core Inverse in Rings with Some Applications

San-Zhang Xu [1,*], Julio Benítez [2], Ya-Qian Wang [1] and Dijana Mosić [3]

[1] Faculty of Mathematics and Physics, Huaiyin Institute of Technology, Huaian 223003, China
[2] Instituto de Matemática Multidisciplinar, Universitat Politècnica de Vaència, 46022 Valencia, Spain
[3] Faculty of Sciences and Mathematics, University of Niš, P.O. Box 224, 18000 Niš, Serbia
* Correspondence: xusanzhang5222@126.com or szxu@hyit.edu.cn

Abstract: In this paper, we introduce two new generalized core inverses, namely, the (p,q,m)-core inverse and the $\langle p,q,n\rangle$-core inverse; both extend the inverses of the $\langle \mathrm{i},\mathrm{m}\rangle$-core inverse, the (j,m)-core inverse, the core inverse, the core-EP inverse and the DMP-inverse.

Keywords: (p,q,m)-core inverse; $\langle p,q,n\rangle$-core inverse; $\langle \mathrm{i},\mathrm{m}\rangle$-core inverse; (j,m)-core inverse; core inverse; DMP-inverse; core-EP inverse

MSC: 15A09; 15A23; 16W10

1. Introduction

Throughout this paper, R denotes a unital ring with involution, i.e., a ring with unity 1, and a mapping $a \mapsto a^*$ satisfying $(a^*)^* = a$, $(ab)^* = b^*a^*$ and $(a+b)^* = a^* + b^*$, for all $a, b \in R$. Let $a, x \in R$, if $axa = a$, $xax = x$, $(ax)^* = ax$ and $(xa)^* = xa$, then x is called a *Moore–Penrose inverse* of a. If such an element x exists, then it is unique and denoted by $a^\dagger$. The set of all Moore–Penrose invertible elements will be denoted by $R^\dagger$.

An element $a \in R$ is said to be *Drazin invertible* if there exists $b \in R$ such that $ab = ba$, $bab = b$ and $a^m = a^{m+1}b$ for some integer m. The element b above is unique if it exists and denoted by a^D. The smallest positive integer m is called the Drazin index of a, denoted by $\mathrm{ind}(a)$. The set of all Drazin invertible elements in R will denoted by R^D. The DMP-inverse for a complex matrix was introduced by Malik and Thome [1]. Let $A \in \mathbb{C}^{n\times n}$ with $\mathrm{ind}(A) = m$, where $\mathbb{C}^{n\times n}$ denotes the set of all $n \times n$ matrices over the field of complex numbers. A matrix $X \in \mathbb{C}^{n\times n}$ is called a *DMP-inverse* of A if it satisfies $XAX = X$, $XA = A^DA$ and $A^mX = A^mA^\dagger$. It is unique (and denoted by $A^{d,\dagger}$). Malik and Thome gave several characterizations of the core inverse by using the decomposition of Hartwig and Spindelböck [2].

The notion of the core-EP inverse for a complex matrix was introduced by Manjunatha Prasad and Mohana [3]. A matrix $X \in \mathbb{C}^{n\times n}$ is a *core-EP inverse* of $A \in \mathbb{C}^{n\times n}$ if X is an outer inverse of A satisfying $\mathcal{R}(X) = \mathcal{R}(X^*) = \mathcal{R}(A^m)$, where m is the index of A and $\mathcal{R}(A)$ stands for the range (column space) of $A \in \mathbb{C}^{n\times n}$. It is unique and denoted by $A^{\textcircled{\dagger}}$. The core-EP inverse for a complex matrix can be investigated by the Core-EP decomposition of a complex matrix by Wang [4]. The notion of the core-EP inverse is extended from the complex matrix to an element in a ring with involution. We will also use the following notations: $aR = \{ax : x \in R\}$, $Ra = \{xa : x \in R\}$, $^\circ a = \{x \in R : xa = 0\}$ and $a^\circ = \{x \in R : ax = 0\}$. Let $a \in R$ with $\mathrm{ind}(a) = k$. An element $b \in R$ is called the core-EP inverse of a if it is an outer inverse of a and b is a $*$-EP element satisfies $bR = a^kR$.

The notion of the core inverse for a complex matrix was introduced by Baksalary and Trenkler [5]. In [6], Rakić et al. generalized the core inverse of a complex matrix to the case of an element in R. More precisely, let $a, x \in R$, if $axa = a$, $xR = aR$ and $Rx = Ra^*$, then x is called a *core inverse* of a. The core inverse can be investigated by three equations by Xu,

Citation: Xu, S.-Z.; Benítez, J.; Wang, Y.-Q.; Mosić, D. Two Generalizations of the Core Inverse in Rings with Some Applications. *Mathematics* **2023**, *11*, 1822. https://doi.org/10.3390/math11081822

Academic Editors: Jing Yao, Xiang Hu and Jingchao Li

Received: 24 February 2023
Revised: 5 April 2023
Accepted: 7 April 2023
Published: 12 April 2023

Chen and Zhang [7]. If such an element x exists, then it is unique and denoted by $a^{\textcircled{\#}}$. The set of all core invertible elements in R will be denoted by $R^{\textcircled{\#}}$.

In addition, $\mathbf{1}_n$ and $\mathbf{0}_n$ will denote the $n \times 1$ column vectors all of whose components are 1 and 0, respectively. The zero matrix of size $m \times n$ is denoted by $0_{m\times n}$ (abbr. 0). If $\mathcal{S}$ is a subspace of $\mathbb{C}^n$, then $P_{\mathcal{S}}$ stands for the *orthogonal projector* onto the subspace $\mathcal{S}$. A matrix $A \in \mathbb{C}^{n\times n}$ is *unitary* if $AA^* = I_n$, where I_n denotes the *identity matrix* of size n. Let $a \in R$, a is called idempotent if $a^2 = a$. The symbol $\mathbb{N}$ denotes the set of all positive integers.

2. Preliminaries

A related decomposition of the matrix decomposition of Hartwig and Spindelböck [2] was given in ([8], Theorem 2.1) by Benítez; in [9] a simpler proof of this decomposition can be found. Let us start this section with the concept of principal angles.

Definition 1 ([10])**.** *Let $\mathcal{S}_1$ and $\mathcal{S}_2$ be two nontrivial subspaces of $\mathbb{C}^n$. We define the principal angles $\theta_1, \ldots, \theta_r \in [0, \pi/2]$ between $\mathcal{S}_1$ and $\mathcal{S}_2$ by*

$$\cos\theta_i = \sigma_i(P_{\mathcal{S}_1} P_{\mathcal{S}_2}),$$

for $i = 1, \ldots, r$, where $r = \min\{\dim \mathcal{S}_1, \dim \mathcal{S}_2\}$. The real numbers $\sigma_i(P_{\mathcal{S}_1} P_{\mathcal{S}_2}) \geq 0$ are the singular values of $P_{\mathcal{S}_1} P_{\mathcal{S}_2}$.

The following theorem can be found in ([8], Theorem 2.1).

Theorem 1. *Let $A \in \mathbb{C}^{n\times n}$, $r = \mathrm{rk}(A)$, and let $\theta_1, \ldots, \theta_p$ be the principal angles between $\mathcal{R}(A)$ and $\mathcal{R}(A^*)$ belonging to $]0, \pi/2[$. Denote by x and y the multiplicities of the angles 0 and $\pi/2$ as a canonical angle between $\mathcal{R}(A)$ and $\mathcal{R}(A^*)$, respectively. There exists a unitary matrix $U \in \mathbb{C}^{n\times n}$ such that*

$$A = U \begin{bmatrix} MC & MS \\ 0 & 0 \end{bmatrix} U^*, \tag{1}$$

where $M \in \mathbb{C}^{r\times r}$ is nonsingular,

$$C = \mathrm{diag}(\mathbf{0}_y, \cos\theta_1, \ldots, \cos\theta_p, \mathbf{1}_x),$$

$$S = \begin{bmatrix} \mathrm{diag}(\mathbf{1}_y, \sin\theta_1, \ldots, \sin\theta_p) & 0_{p+y,n-(r+p+y)} \\ 0_{x,p+y} & 0_{x,n-(r+p+y)} \end{bmatrix},$$

and $r = y + p + x$. Furthermore, x and $y + n - r$ are the multiplicities of the singular values 1 and 0 in $P_{\mathcal{R}(A)} P_{\mathcal{R}(A^*)}$, respectively. We call (1) as the CS decomposition of A.

In this decomposition, one has $C^2 + SS^* = I_r$ and $C^* = C$. This decomposition can answer the question "how far is a matrix from being EP". Moreover, it can be applied to some partial matrix ordering, such as star ordering and sharp ordering.

3. (p, q, m)-Core Inverse

Let us start this section by introducing the notation of the (p, q, m)-core inverse.

Definition 2. *Let $a, p, q \in R$ and $m \in \mathbb{N}$. If $pa = ap$ and pa is idempotent, then $x \in R$ is called a* (p, q, m)-core inverse *of a, if it satisfies*

$$x = pax \ \text{ and } \ a^m x = q. \tag{2}$$

It will be proved that if x exists, then it is unique and denoted by $a^{\circledast}_{p,q,m}$.

Remark 1. *If $a \in R$ is (p,q,m)-core invertible, then we have $pa = ap$ and pa is idempotent. Since this property of the (p,q,m)-core inverse is used many times in the sequel, thus we emphasize it here.*

Theorem 2. *If equations in* (2) *have a solution, then it is unique.*

Proof. Let x_1 and x_2 be two candidates (p,q,m)-core inverse of a, that is $x_1 = pax_1$, $a^m x_1 = q$, $x_2 = pax_2$ and $a^m x_2 = q$. Thus by $pa = ap$ and pa is idempotent, we have

$$x_1 = p^m q = p^m a^m x_2 = pax_2 = x_2.$$

□

In the following lemma, we will show that $q = paq$ if a is (p,q,m)-core invertible.

Lemma 1. *Let $a, p, q \in R$ and $m, n \in \mathbb{N}$. If a is (p,q,m)-core invertible, then*

(1) $q = pa^{m+1}a^{\circledast}_{p,q,m}$;
(2) $q = paq$;
(3) $a^n a^{\circledast}_{p,q,m} = p^{m-n}q$, *where* $m \geq n$.

Proof. (1) and (2). If a is (p,q,m)-core invertible, then we have $a^{\circledast}_{p,q,m} = paa^{\circledast}_{p,q,m}$ and $a^m a^{\circledast}_{p,q,m} = q$. Having in mind that $ap = pa$ and the idempotency of pa, we obtain

$$q = a^m a^{\circledast}_{p,q,m} = a^m(paa^{\circledast}_{p,q,m}) = pa^{m+1}a^{\circledast}_{p,q,m}; \tag{3}$$
$$a^{\circledast}_{p,q,m} = paa^{\circledast}_{p,q,m} = p^m a^m a^{\circledast}_{p,q,m} = p^m q. \tag{4}$$

Thus, by (3) and (4), we have

$$q = pa^{m+1}a^{\circledast}_{p,q,m} = pa^{m+1}(p^m q) = p^{m+1}a^{m+1}q = paq. \tag{5}$$

(3). If $m \geq n$, then $a^n a^{\circledast}_{p,q,m} = a^n(paa^{\circledast}_{p,q,m}) = a^n p^m a^m a^{\circledast}_{p,q,m} = a^n p^m q = p^{m-n}p^n a^n q = p^{m-n}paq = p^{m-n}q$ by the definition of the (p,q,m)-core inverse and (2). □

Theorem 3. *If the solution of the equations in* (2) *exists, then the unique solution is $x = p^m q$.*

Proof. By Lemma 1, we have $q = paq$. Having in mind that $ap = pa$ and the idempotency of pa, we obtain

$$\begin{aligned} pax &= pap^m q = p^m(paq) = p^m q = x \\ a^m x &= a^m p^m q = paq = q. \end{aligned}$$

□

Remark 2. *If $a \in R^D$ and $a^i, a^j \in R^{\dagger}$, then the (p,q,m)-core inverse is the generalizations of the $\langle \mathrm{i}, \mathrm{m} \rangle$-core inverse and the (j,m)-core inverse* [11]*, respectively. More precisely, we have the following statements:*

(1) *If $p = a^D$ and $q = a^i(a^i)^{\dagger}$, then the (p,q,m)-core inverse coincides with the $\langle \mathrm{i}, \mathrm{m} \rangle$-core inverse;*
(2) *If $p = a^D$ and $q = a^m(a^j)^{\dagger}$, then the (p,q,m)-core inverse coincides with the (j,m)-core inverse.*

By Remarks 3.5, 4.7 and 4.8 in [11], we have the $\langle \mathrm{m}, \mathrm{j} \rangle$-core inverse for a complex matrix, which extends the notions of the core inverse defined by Baksalary and Trenkler [5] and the core-EP inverse defined by Manjunatha Prasad and Mohana [3], respectively. The (m, k)-core inverse for a complex matrix, which extends the notions of the core inverse and

the DMP-inverse defined by Malik and Thome [1], respectively. Therefore, we have the following remark by Remark 2. We can use generalized inverses to study the system of constrained matrix equations and Toeplitz matrix, etc. [12,13].

Remark 3. *If $a \in R^D$ and $a^j \in R^\dagger$, then the (p,q,m)-core inverse is a generalization of the core inverse, the DMP inverse and the core-EP inverse. More precisely, we have the following statements:*

(1) *If $p = a^\#$, $m = 1$ and $q = aa^\dagger$, then the (p,q,m)-core inverse coincides with the core inverse;*
(2) *If $p = a^D$, $m = \mathrm{ind}(a)$ and $q = a^m a^\dagger$, then the (p,q,m)-core inverse coincides with the DMP inverse;*
(3) *If $p = a^D$, $m = 1$, $j = \mathrm{ind}(a)$ and $q = a^j(a^j)^\dagger$, then the (p,q,m)-core inverse coincides with the core-EP inverse.*

Example 1. *The (p,q,m)-core inverse is different from the group inverse and the Moore–Penrose inverse. Let $A = \begin{bmatrix} 1 & i \\ 0 & 0 \end{bmatrix} \in \mathbb{C}^{2\times 2}$. Then $A^\# = A$ by $A^2 = A$, but A is not Moore–Penrose invertible by $AA^* = \begin{bmatrix} 1 & i \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ i & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$. Note that if A is Moore–Penrose invertible, then $A = AA^\dagger A = A(A^\dagger A)^* = AA^*(A^\dagger)^* = 0$, but $A \neq 0$. In fact, AA^* implies A is not $\{1,4\}$-invertible. If we let $p = a^\#$, $q = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, then $a^{\circledast}_{p,q,m} = \begin{bmatrix} 1+3i & 2+4i \\ 0 & 0 \end{bmatrix}$.*

Theorem 4. *Let $a, p, q \in R$ and $m \in \mathbb{N}$. If $pa = ap$ and pa is idempotent, then the following are equivalent:*

(1) *a is (p,q,m)-core invertible with $a^{\circledast}_{p,q,m} = x$;*
(2) *$x = pax$ and $q = pa^{m+1}x$;*
(3) *$x = pax$, $aq = a^{m+1}x$ and $q = paq$.*

Proof. (1) $\Rightarrow$ (2) and (1) $\Rightarrow$ (3) are trivial by Lemma 1 and the definition of the (p,q,m)-core inverse.

(2) $\Rightarrow$ (1). From $a^m x = a^m(pax) = pa^{m+1}x = q$ we have that x is the (p,q,m)-core inverse of a.

(3) $\Rightarrow$ (2). It is sufficient to prove $q = pa^{m+1}x$. We have $q = paq = pa^{m+1}x$. □

Remark 4. *Note that $x = pax$ iff $xR \subseteq paR$ iff ${}^\circ(pa) \subseteq {}^\circ x$. Moreover, $q = paq$ iff $qR \subseteq paR$ iff ${}^\circ(pa) \subseteq {}^\circ q$. Thus, we can obtain more conditions such that a is (p,q,m)-core invertible in Theorem 4.*

If $p = a^\#$, $m = 1$ and $q = aa^\dagger$, then the (p,q,m)-core inverse coincides with the core inverse, thus we have the following corollary by Theorem 4.

Corollary 1. *Let $a \in R$ with $a \in R^\# \cap R^\dagger$. Then the following are equivalent:*

(1) *a is core invertible with $a^{\circledast} = x$;*
(2) *$x = a^\# ax$ and $aa^\dagger = ax$;*
(3) *$x = a^\# ax$ and $a^2 a^\dagger = a^2 x$.*

Since the (p,q,m)-core inverse is a generalization of the core inverse, the core-EP inverse, the DMP-inverse, $\langle i,m\rangle$-core inverse and (j,m)-core inverse, we can obtain some analogous corollaries as Corollary 1.

Recall that for $e = e^2 \in R$, we can represent any $a \in R$ as a matrix

$$a = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}_{e\times e},$$

where $a_{11} = eae$, $a_{12} = ea(1-e)$, $a_{21} = (1-e)ae$ and $a_{22} = (1-e)a(1-e)$.

Now we present the result concerning the matrix form of (p,q,m)-core invertible element $a \in R$.

Theorem 5. *Let $a,p,q \in R$ and $m \in \mathbb{N}$. Then a is (p,q,m)-core invertible if and only if there exists $e \in R$ such that $e = e^2$,*

$$a = \begin{bmatrix} a_1 & 0 \\ 0 & a_2 \end{bmatrix}_{e\times e}, \quad p = \begin{bmatrix} p_1 & 0 \\ 0 & p_2 \end{bmatrix}_{e\times e} \quad \text{and} \quad q = \begin{bmatrix} q_1 & q_2 \\ 0 & 0 \end{bmatrix}_{e\times e},$$

where $p_1a_1 = a_1p_1 = (p_1a_1)^2$, $p_2a_2 = a_2p_2 = 0$, a_1 is (p_1,q_1,m)-core invertible and (p_1,q_2,m)-core invertible. The (p,q,m)-core inverse of a is given by

$$a^{\circledast}_{p,q,m} = \begin{bmatrix} (a_1)^{\circledast}_{p_1,q_1,m} & (a_1)^{\circledast}_{p_1,q_2,m} \\ 0 & 0 \end{bmatrix}_{e\times e} = \begin{bmatrix} p_1^m q_1 & p_1^m q_2 \\ 0 & 0 \end{bmatrix}_{e\times e}.$$

Proof. Suppose that a is (p,q,m)-core invertible and let $e = pa$. Then $e^2 = (pa)^2 = pa = e$, $ea(1-e) = apa(1-pa) = 0$ and $(1-e)ae = 0$. Hence,

$$a = \begin{bmatrix} a_1 & 0 \\ 0 & a_2 \end{bmatrix}_{e\times e},$$

where $a_1 = pa^2$ and $a_2 = (1-pa)a$. Similarly, we obtain, for $p_1 = p^2a$ and $p_2 = (1-pa)p$,

$$p = \begin{bmatrix} p_1 & 0 \\ 0 & p_2 \end{bmatrix}_{e\times e}.$$

The equalities $pa = ap$ and $(pa)^2 = pa$ give $p_1a_1 = a_1p_1 = (p_1a_1)^2$ and $p_2a_2 = a_2p_2 = (1-pa)ap(1-pa) = 0$. Set

$$a^{\circledast}_{p,q,m} = \begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix}_{e\times e} \quad \text{and} \quad q = \begin{bmatrix} q_1 & q_2 \\ q_3 & q_4 \end{bmatrix}_{e\times e}.$$

From $a^{\circledast}_{p,q,m} = paa^{\circledast}_{p,q,m} = \begin{bmatrix} p_1a_1 & 0 \\ 0 & 0 \end{bmatrix}_{e\times e} a^{\circledast}_{p,q,m}$, we obtain $x_1 = p_1a_1x_1$, $x_2 = p_1a_1x_2$ and $x_3 = x_4 = 0$. Since $q = paq$, then $q_3 = q_4 = 0$. Now, by

$$\begin{bmatrix} a_1^m x_1 & a_1^m x_2 \\ 0 & 0 \end{bmatrix}_{e\times e} = a^m a^{\circledast}_{p,q,m} = q = \begin{bmatrix} q_1 & q_2 \\ 0 & 0 \end{bmatrix}_{e\times e},$$

we conclude that $a_1^m x_1 = q_1$ and $a_1^m x_2 = q_2$. Hence, a_1 is (p_1,q_1,m)-core invertible and (p_1,q_2,m)-core invertible with $x_1 = (a_1)^{\circledast}_{p_1,q_1,m}$ and $x_2 = (a_1)^{\circledast}_{p_1,q_2,m}$.

Conversely, by the assumption $p_1a_1 = a_1p_1 = (p_1a_1)^2$ and $p_2a_2 = a_2p_2 = 0$, we check that $pa = ap = (pa)^2$. Since a_1 is (p_1,q_1,m)-core invertible and (p_1,q_2,m)-core invertible, if we let

$$x = \begin{bmatrix} (a_1)^{\circledast}_{p_1,q_1,m} & (a_1)^{\circledast}_{p_1,q_2,m} \\ 0 & 0 \end{bmatrix}_{e\times e},$$

we get $x = pax$ and $a^m x = q$. So, a is (p,q,m)-core invertible and $x = a^{\circledast}_{p,q,m}$. □

Under some conditions, we obtain that the (p,q,m)-core inverse of a and the (p,r,m)-core inverse of b commute.

Lemma 2. *Let $a,b,p,q,r \in R$ and $m \in \mathbb{N}$. If a is (p,q,m)-core invertible, b is (p,r,m)-core invertible and $qp^m r = rp^m q$ (or equivalently $qb^{\circledast}_{p,r,m} = ra^{\circledast}_{p,q,m}$), then $a^{\circledast}_{p,q,m}b^{\circledast}_{p,r,m} = b^{\circledast}_{p,r,m}a^{\circledast}_{p,q,m}$.*

Proof. Because $a^{\circledast}_{p,q,m} = p^m q$ and $b^{\circledast}_{p,r,m} = p^m r$ by Theorem 3, we get $a^{\circledast}_{p,q,m} b^{\circledast}_{p,r,m} = p^m q p^m r = p^m r p^m q = b^{\circledast}_{p,r,m} a^{\circledast}_{p,q,m}$. □

Now, we study when the product of one (p,q,m)-core invertible element and one (p,r,m)-core invertible element is (p^2,rq,m)-core invertible.

Theorem 6. *Let $a,b,p,q,r \in R$ and $m \in \mathbb{N}$ such that a is (p,q,m)-core invertible, b is (p,r,m)-core invertible, $ab = ba$ and $a^m r = r a^m$. We have the following statements:*

(1) *If $pap^m r = p^m rap$ (or equivalently $pab^{\circledast}_{p,r,m} = b^{\circledast}_{p,r,m} ap$), then ab is (p^2,rq,m)-core invertible and $(ab)^{\circledast}_{p^2,rq,m} = b^{\circledast}_{p,r,m} a^{\circledast}_{p,q,m}$;*

(2) *If $qb^{\circledast}_{p,r,m} = r a^{\circledast}_{p,q,m}$, then ab is (p^2,rq,m)-core invertible and $(ab)^{\circledast}_{p^2,rq,m} = b^{\circledast}_{p,r,m} a^{\circledast}_{p,q,m} = a^{\circledast}_{p,q,m} b^{\circledast}_{p,r,m}$.*

Proof. Since $pa = ap$ and $pb = bp$ are idempotents and $ab = ba$, notice that $p^2 ab = abp^2$ and $(p^2 ab)^2 = (pa)^2 (pb)^2 = p^2 ab$. The assumptions $ab = ba$ and $a^m r = r a^m$ imply $(ab)^m b^{\circledast}_{p,r,m} a^{\circledast}_{p,q,m} = a^m (b^m b^{\circledast}_{p,r,m}) a^{\circledast}_{p,q,m} = (a^m r) a^{\circledast}_{p,q,m} = r(a^m a^{\circledast}_{p,q,m}) = rq$.

(1). Since $pap^m r = p^m rap$, $ap = pa$ and $bp = pb$, we have $p^2 abb^{\circledast}_{p,r,m} a^{\circledast}_{p,q,m} = pa(pbb^{\circledast}_{p,r,m}) a^{\circledast}_{p,q,m} = (pab^{\circledast}_{p,r,m}) a^{\circledast}_{p,q,m} = b^{\circledast}_{p,r,m} (paa^{\circledast}_{p,q,m}) = b^{\circledast}_{p,r,m} a^{\circledast}_{p,q,m}$. Therefore, ab is (p^2,rq,m)-core invertible and $(ab)^{\circledast}_{p^2,rq,m} = b^{\circledast}_{p,r,m} a^{\circledast}_{p,q,m}$.

(2). From $qp^m r = rp^m q$ we can get that $b^{\circledast}_{p,r,m} a^{\circledast}_{p,q,m} = a^{\circledast}_{p,q,m} b^{\circledast}_{p,r,m}$ by Lemma 2. By $p^2 abb^{\circledast}_{p,r,m} a^{\circledast}_{p,q,m} = pa(b^{\circledast}_{p,r,m} a^{\circledast}_{p,q,m}) = (paa^{\circledast}_{p,q,m}) b^{\circledast}_{p,r,m} = a^{\circledast}_{p,q,m} b^{\circledast}_{p,r,m} = b^{\circledast}_{p,r,m} a^{\circledast}_{p,q,m}$, we deduce that $(ab)^{\circledast}_{p^2,rq,m} = b^{\circledast}_{p,r,m} a^{\circledast}_{p,q,m} = a^{\circledast}_{p,q,m} b^{\circledast}_{p,r,m}$. □

In the case that $ab = ba = 0$, the sum of (p,q,m)-core invertible element a and (p,r,m)-core invertible element b is $(p,q+r,m)$-core invertible.

Theorem 7. *Let $a,b,p,q,r \in R$ and $m \in \mathbb{N}$ such that a is (p,q,m)-core invertible, b is (p,r,m)-core invertible and $ab = ba = 0$. Then $a+b$ is $(p,q+r,m)$-core invertible and $(a+b)^{\circledast}_{p,q+r,m} = a^{\circledast}_{p,q,m} + b^{\circledast}_{p,r,m}$.*

Proof. First, observe that $p(a+b) = (a+b)p$ and $[p(a+b)]^2 = p^2(a^2+b^2) = pa + pb = p(a+b)$. Further,

$$a^m b^{\circledast}_{p,r,m} = a^m pbb^{\circledast}_{p,r,m} = pa^m bb^{\circledast}_{p,r,m} = 0$$

and $pab^{\circledast}_{p,r,m} = p^m (a^m b^{\circledast}_{p,r,m}) = 0$. Analogously, $b^m a^{\circledast}_{p,q,m} = 0 = pba^{\circledast}_{p,q,m}$. Thus,

$$p(a+b)(a^{\circledast}_{p,q,m} + b^{\circledast}_{p,r,m}) = (pa+pb)(a^{\circledast}_{p,q,m} + b^{\circledast}_{p,r,m}) = paa^{\circledast}_{p,q,m} + pbb^{\circledast}_{p,r,m} = a^{\circledast}_{p,q,m} + b^{\circledast}_{p,r,m}$$

and

$$(a+b)^m (a^{\circledast}_{p,q,m} + b^{\circledast}_{p,r,m}) = (a^m + b^m)(a^{\circledast}_{p,q,m} + b^{\circledast}_{p,r,m}) = a^m a^{\circledast}_{p,q,m} + b^m b^{\circledast}_{p,r,m} = q + r,$$

that is, $a+b$ is $(p,q+r,m)$-core invertible and $(a+b)^{\circledast}_{p,q+r,m} = a^{\circledast}_{p,q,m} + b^{\circledast}_{p,r,m}$. □

Lemma 3. *Let $a,p,q \in R$ and $m \in \mathbb{N}$ and a is (p,q,m)-core invertible. Then $aa^{\circledast}_{p,q,m} = a^{\circledast}_{p,q,m} a$ if and only if $p^{m-1} q = p^m qa$.*

Proof. By Lemma 1, we have $q = paq$. If $aa^{\circledast}_{p,q,m} = a^{\circledast}_{p,q,m} a$, then $p^m qa = a^{\circledast}_{p,q,m} a = aa^{\circledast}_{p,q,m} = ap^m q = p^{m-1}(paq) = p^{m-1} q$. For the opposite implication, we have $aa^{\circledast}_{p,q,m} = ap^m q = app^{m-1} q = app^m qa = p^m (paq) a = p^m qa = a^{\circledast}_{p,q,m} a$. □

Proposition 1. *Let $a,p,q \in R$ and $m \in \mathbb{N}$. If a is (p,q,m)-core invertible, then*

(1) *If $qa^m = a^m$, then $a^{\circledast}_{p,q,m}$ is an inner inverse of a^m and q is idempotent;*

(2) *If* $aq = qa$ *(or equivalently* $a^{m+1}a^{\circledast}_{p,q,m} = a^m a^{\circledast}_{p,q,m}a$*), then* $aa^{\circledast}_{p,q,m} = a^{\circledast}_{p,q,m}a$*;*
(3) *If* q *is idempotent, then* $a^{\circledast}_{p,q,m}$ *is an outer inverse of* a^m*;*
(4) *If* $q = q^*$*, then* $a^m a^{\circledast}_{p,q,m} = (a^m a^{\circledast}_{p,q,m})^*$*;*
(5) *If* $aq = qa$ *and* $q = q^*$*, then* $a^{\circledast}_{p,q,m}a^m = (a^{\circledast}_{p,q,m}a^m)^*$*.*

Proof. (1). Since $qa^m = a^m$ and $q = a^m a^{\circledast}_{p,q,m}$, we have that $a^m = qa^m = a^m a^{\circledast}_{p,q,m}a^m$ and $q = a^m a^{\circledast}_{p,q,m} = qa^m a^{\circledast}_{p,q,m} = q^2$.

(2). It is easy to check that $p^{m-1}q = p^m qa$ by $aq = qa$ and $q = paq$. Thus, we have $aa^{\circledast}_{p,q,m} = a^{\circledast}_{p,q,m}a$ by Lemma 3.

(3). The condition $q = q^2$ gives $a^{\circledast}_{p,q,m}a^m a^{\circledast}_{p,q,m} = a^{\circledast}_{p,q,m}a^m p^m q = a^{\circledast}_{p,q,m}paq = p^m q^2 = p^m q = a^{\circledast}_{p,q,m}$.

(4). By definition of the (p,q,m)-core inverse.

(5). It follows from (2) and (4). □

Applying Proposition 1, we obtain the next result.

Corollary 2. *Let* $a, p, q \in R$ *and* $m \in \mathbb{N}$*. If* a *is* (p,q,m)*-core invertible, then*

(1) *If* $qa^m = a^m$ *and* $aq = qa$*, then* $a^m \in R^{\#}$ *and* $(a^m)^{\#} = a^{\circledast}_{p,q,m}$*;*
(2) *If* $qa^m = a^m$*,* $q = q^*$ *and* $aq = qa$*, then* $a^m \in R^{\#} \cap R^{\dagger}$ *and* $(a^m)^{\dagger} = (a^m)^{\#} = a^{\circledast}_{p,q,m}$ *(that is,* a^m *is EP).*

4. $\langle p, q, n\rangle$-Core Inverse

Definition 3. *Let* $a, p, q \in R$ *and* $n \in \mathbb{N}$*. We say that* $x \in R$ *is a* $\langle p, q, n\rangle$-core inverse *of* a*, if it satisfies*

$$x = pa^n x \ \text{ and } \ a^n x = q. \tag{6}$$

It will be proved that if x exists, then it is unique and denoted by $a^{\circledcirc}_{p,q,n}$.

Theorem 8. *If equations in* (6) *have a solution, then it is unique and the unique solution is* $x = pq$*.*

Proof. Let x satisfy (6). Then $x = pa^n q = pq$". Observe that this implies the uniqueness of the equations (6): the unique element in R satisfying (6) is pq. □

If a is $\langle p, q, n\rangle$-core invertible, then we have $a^{\circledcirc}_{p,q,n} = pa^n a^{\circledcirc}_{p,q,n}$ and $a^n a^{\circledcirc}_{p,q,n} = q$ and

$$q = a^n a^{\circledcirc}_{p,q,n} = a^n(pa^n a^{\circledcirc}_{p,q,n}) = a^n pa^n a^{\circledcirc}_{p,q,n}$$

Thus, we obtain

$$q = a^n pa^n a^{\circledcirc}_{p,q,n} = a^n pa^n pq = (a^n p)^2 q.$$

By Theorem 8, we have $q = a^n x = a^n pq$; here, x is the $\langle p, q, n\rangle$-core inverse of a (see next Theorem 11).

Lemma 4. *Let* $a, p, q \in R$ *and* $n \in \mathbb{N}$*. If* a *is* $\langle p, q, n\rangle$*-core invertible, then* $q = a^n pa^n a^{\circledcirc}_{p,q,n} = (a^n p)^2 q$*.*

Remark 5. *If* $a \in R^D$ *and* $a^i, a^j \in R^{\dagger}$*, then the* $\langle p, q, n\rangle$*-core inverse is a generalization of the* $\langle \mathrm{i}, \mathrm{m}\rangle$*-core inverse and the* (j, m)*-core inverse* [11]*. More precisely, we have the following statements:*

(1) *If* $p = (a^D)^n$ *and* $q = a^i(a^i)^{\dagger}$*, then the* $\langle p, q, n\rangle$*-core inverse coincides with the* $\langle \mathrm{i}, \mathrm{m}\rangle$*-core inverse;*
(2) *If* $p = (a^D)^n$ *and* $q = a^m(a^j)^{\dagger}$*, then the* $\langle p, q, n\rangle$*-core inverse coincides with the* (j, m)*-core inverse.*

Theorem 9. *Let $a, p, q \in R$ and $n \in \mathbb{N}$. Then the following are equivalent:*

(1) *a is $\langle p, q, n\rangle$-core invertible with $a^{\circledcirc}_{p,q,n} = x$;*
(2) *$x = pa^n x$ and $q = a^n p a^n x$;*
(3) *$x = pa^n x$, $a^n pq = a^n x$ and $q = (a^n p)^2 q$.*

Proof. (1) $\Rightarrow$ (2) and (1) $\Rightarrow$ (3) are trivial by Lemma 4 and the definition of the $\langle p, q, n\rangle$-core inverse.

(2) $\Rightarrow$ (1). From $q = a^n p a^n x = a^n x$ we have that x is the $\langle p, q, n\rangle$-core inverse of a.

(3) $\Rightarrow$ (2). It is sufficient to prove $q = a^n p a^n x$. We have $q = (a^n p)^2 q = a^n p a^n pq = a^n p a^n x$. □

Under certain conditions, the product of a $\langle p, q, n\rangle$-core invertible element and a $\langle r, s, n\rangle$-core invertible element is $\langle pr, sq, n\rangle$-core invertible.

Theorem 10. *Let $a, b, p, q, r, s \in R$ and $n \in \mathbb{N}$ such that a is $\langle p, q, n\rangle$-core invertible, b is $\langle r, s, n\rangle$-core invertible, $ab = ba$, $a^n r = ra^n$, $a^n s = sa^n$ and $prs = rsp$. Then ab is $\langle pr, sq, n\rangle$-core invertible and $(ab)^{\circledcirc}_{pr,sq,n} = b^{\circledcirc}_{r,s,n} a^{\circledcirc}_{p,q,n}$.*

Proof. Notice that

$$(ab)^n b^{\circledcirc}_{r,s,n} a^{\circledcirc}_{p,q,n} = a^n (b^n b^{\circledcirc}_{r,s,n}) a^{\circledcirc}_{p,q,n} = (a^n s) a^{\circledcirc}_{p,q,n} = s(a^n a^{\circledcirc}_{p,q,n}) = sq$$

and

$$pr(ab)^n b^{\circledcirc}_{r,s,n} a^{\circledcirc}_{p,q,n} = (prs)q = (rs)(pq) = b^{\circledcirc}_{r,s,n} a^{\circledcirc}_{p,q,n}$$

imply ab is $\langle pr, sq, n\rangle$-core invertible and $(ab)^{\circledcirc}_{pr,sq,n} = b^{\circledcirc}_{r,s,n} a^{\circledcirc}_{p,q,n}$. □

We also study when the sum of a $\langle p, q, n\rangle$-core invertible element and a $\langle r, s, n\rangle$-core invertible element is $\langle p + r, q + s, n\rangle$-core invertible.

Theorem 11. *Let $a, b, p, q, r, s \in R$ and $n \in \mathbb{N}$ such that a is $\langle p, q, n\rangle$-core invertible, b is $\langle r, s, n\rangle$-core invertible, $ab = ba = 0$, $a^n rs = 0 = b^n pq$ and $ps + rq = 0$. Then $a + b$ is $\langle p + r, q + s, n\rangle$-core invertible and $(a + b)^{\circledcirc}_{p+r,q+s,n} = a^{\circledcirc}_{p,q,n} + b^{\circledcirc}_{r,s,n}$.*

Proof. Let x be the $\langle p, q, n\rangle$-core inverse of a and y be the $\langle r, s, n\rangle$-core inverse of b, then by Theorem 8, we have $(p + r)(a + b)^n (x + y) = (p + r)(q + s) = pq + ps + rq + rs = pq + rs = x + y$. □

It is easy to check the following propositions by Definition 3 and Theorem 8.

Proposition 2. *Let $a, p, q \in R$ and $n \in \mathbb{N}$ such that a is $\langle p, q, n\rangle$-core invertible. Then $aa^{\circledcirc}_{p,q,n} = a^{\circledcirc}_{p,q,n} a$ if and only if $apq = pqa$.*

Proposition 3. *Let $a, p, q \in R$ and $n \in \mathbb{N}$ such that a is $\langle p, q, n\rangle$-core invertible. Then*

(1) *If $qa^n = a^n$, then $a^{\circledcirc}_{p,q,n}$ is an inner inverse of a^n and q is idempotent;*
(2) *If $q = q^2$, then $a^{\circledcirc}_{p,q,n} a^n a^{\circledcirc}_{p,q,n} = a^{\circledcirc}_{p,q,n}$;*
(3) *If $q = q^*$, then $a^n a^{\circledcirc}_{p,q,n} = (a^n a^{\circledcirc}_{p,q,n})^*$;*
(4) *If $apq = pqa$ and $q = q^*$, then $a^{\circledcirc}_{p,q,n} a^n = (a^{\circledcirc}_{p,q,n} a^n)^*$.*

5. How to Compute the (P, Q, m)-Core Inverse and $\langle P, Q, n\rangle$-Core Inverse in $\mathbb{C}^{n\times n}$

5.1. How to Compute the (p, q, m)-Core Inverse in $\mathbb{C}^{n\times n}$

Let $A, P, Q \in \mathbb{C}^{n\times n}$ and $m \in \mathbb{N}$. We will assume in this subsection that A is (P, Q, m)-core invertible.If $A \in \mathbb{C}^{n\times n}$ is (P, Q, m)-core invertible, then we have $PA = AP$, PA

is idempotent, $X = PAX$ and $A^mX = Q$. Assume that A has the form (1). If we let $P = U\begin{bmatrix} P_1 & P_2 \\ P_3 & P_4 \end{bmatrix}U^*$, where $P_1 \in \mathbb{C}^{r\times r}$, then

$$PA = U\begin{bmatrix} P_1MC & P_1MS \\ P_3MC & P_3MS \end{bmatrix}U^*; \tag{7}$$

$$AP = U\begin{bmatrix} MCP_1 + MSP_3 & MCP_2 + MSP_4 \\ 0 & 0 \end{bmatrix}U^*. \tag{8}$$

From (7) and (8) and $PA = AP$ we obtain $P_3MC = 0$ and $P_3MS = 0$. Then we have $P_3MC^2 = 0$ and $P_3MSS^* = 0$, thus $P_3MC^2 + P_3MSS^* = P_3M(C^2 + SS^*) = P_3M$ by $C^2 + SS^* = I_r$. The nonsingularity of M implies that P_3 is zero matrix, which gives

$$PA = U\begin{bmatrix} P_1MC & P_1MS \\ 0 & 0 \end{bmatrix}U^* = U\begin{bmatrix} P_1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} MC & MS \\ 0 & 0 \end{bmatrix}U^* \tag{9}$$

Since PA is idempotent, $AP = PA$ and $(PA)^2 = U\begin{bmatrix} (P_1MC)^2 & P_1MCP_1MS \\ 0 & 0 \end{bmatrix}U^*$, hence

$$P_1MC = MCP_1 = (P_1MC)^2 \tag{10}$$

By Lemma 1, we have $Q = PAQ$. If we let $Q = U\begin{bmatrix} Q_1 & Q_2 \\ Q_3 & Q_4 \end{bmatrix}U^*$, then by (9) we have

$$\begin{aligned} PAQ =& U\begin{bmatrix} P_1MC & P_1MS \\ 0 & 0 \end{bmatrix}\begin{bmatrix} Q_1 & Q_2 \\ Q_3 & Q_4 \end{bmatrix}U^* \\ =& U\begin{bmatrix} P_1MCQ_1 + P_1MSQ_3 & P_1MCQ_2 + P_1MSQ_4 \\ 0 & 0 \end{bmatrix}U^*. \end{aligned} \tag{11}$$

From $Q = PAQ$ we have that Q_3 and Q_4 are zero matrices and

$$\begin{cases} Q_1 = P_1MCQ_1 \\ Q_2 = P_1MCQ_2 \end{cases} \tag{12}$$

By Theorem 3, we have $A^{\circledast}_{P,Q,m} = P^mQ$. Since $P_3 = 0$, $Q_3 = 0$ and $Q_4 = 0$, thus we have $P = U\begin{bmatrix} P_1 & P_2 \\ 0 & P_4 \end{bmatrix}U^*$ and $Q = U\begin{bmatrix} Q_1 & Q_2 \\ 0 & 0 \end{bmatrix}U^*$, thus $P^m = U\begin{bmatrix} P_1^m & \bigstar \\ 0 & P_4^m \end{bmatrix}U^*$; the entries that we are not interested in are marked with $\bigstar$. Therefore

$$\begin{aligned} A^{\circledast}_{P,Q,m} =& P^mQ = U\begin{bmatrix} P_1^m & \bigstar \\ 0 & P_4^m \end{bmatrix}\begin{bmatrix} Q_1 & Q_2 \\ 0 & 0 \end{bmatrix}U^* \\ =& U\begin{bmatrix} P_1^mQ_1 & P_1^mQ_2 \\ 0 & 0 \end{bmatrix}U^*. \end{aligned} \tag{13}$$

By $A^m = U\begin{bmatrix} (MC)^m & (MC)^{m-1}MS \\ 0 & 0 \end{bmatrix}U^*$ and $A^mA^{\circledast}_{P,Q,m} = Q$, we have

$$\begin{aligned} A^mA^{\circledast}_{P,Q,m} =& U\begin{bmatrix} (MC)^m & (MC)^{m-1}MS \\ 0 & 0 \end{bmatrix}\begin{bmatrix} P_1^mQ_1 & P_1^mQ_2 \\ 0 & 0 \end{bmatrix}U^* \\ =& U\begin{bmatrix} (MC)^mP_1^mQ_1 & (MC)^mP_1^mQ_2 \\ 0 & 0 \end{bmatrix}U^* \\ =& U\begin{bmatrix} Q_1 & Q_2 \\ 0 & 0 \end{bmatrix}U^*. \end{aligned} \tag{14}$$

Thus

$$\begin{cases} Q_1 = (MC)^m P_1^m Q_1 \\ Q_2 = (MC)^m P_1^m Q_2 \end{cases} \tag{15}$$

Therefore, by (10), (12), (15) and the definition of the (P, Q, m)-core inverse, we have

$$\begin{cases} (MC)^{\circledast}_{P_1,Q_1,m} = P_1^m Q_1 \\ (MC)^{\circledast}_{P_1,Q_2,m} = P_1^m Q_2 \end{cases} \tag{16}$$

From (13) and (16) we have

$$A^{\circledast}_{P,Q,m} = U \begin{bmatrix} (MC)^{\circledast}_{P_1,Q_1,m} & (MC)^{\circledast}_{P_1,Q_2,m} \\ 0 & 0 \end{bmatrix} U^*.$$

5.2. How to Compute the $\langle p, q, n\rangle$-Core Inverse in $\mathbb{C}^{n\times n}$

Let $A, P, Q \in \mathbb{C}^{n\times n}$ and $n \in \mathbb{N}$. We will assume in this subsection that A is $\langle P, Q, n\rangle$-core invertible. Here we suppose that $AP = PA$, thus we have $P = U \begin{bmatrix} P_1 & P_2 \\ 0 & P_4 \end{bmatrix} U^*$, where $P_1 \in \mathbb{C}^{r\times r}$. Moreover, we have

$$P_1 MC = MCP_1 \tag{17}$$

and

$$PA^n = U \begin{bmatrix} P_1(MC)^n & P_1(MC)^{n-1}MS \\ 0 & 0 \end{bmatrix} U^*. \tag{18}$$

By Lemma 4, we have $Q = (A^nP)^2 Q$. If we let $Q = U \begin{bmatrix} Q_1 & Q_2 \\ Q_3 & Q_4 \end{bmatrix} U^*$, then by (18) we have

$$\begin{aligned} (A^nP)^2 Q &= U \begin{bmatrix} P_1(MC)^n & P_1(MC)^{n-1}MS \\ 0 & 0 \end{bmatrix}^2 \begin{bmatrix} Q_1 & Q_2 \\ Q_3 & Q_4 \end{bmatrix} U^* \\ &= U \begin{bmatrix} \bigstar & \bigstar \\ 0 & 0 \end{bmatrix} U^*. \end{aligned} \tag{19}$$

where we marked with $\bigstar$ the entries that we are not interested in. Thus, from $Q = (A^nP)^2 Q$ we have Q_3 and Q_4 which are zero matrices. Therefore, we have $A^{\circledcirc}_{P,Q,n} = PQ = U \begin{bmatrix} P_1 & P_2 \\ 0 & P_4 \end{bmatrix} \begin{bmatrix} Q_1 & Q_2 \\ 0 & 0 \end{bmatrix} U^* = U \begin{bmatrix} P_1Q_1 & P_1Q_2 \\ 0 & 0 \end{bmatrix} U^*$. It is not difficult to see that we have

$$A^{\circledcirc}_{P,Q,n} = U \begin{bmatrix} (MC)^{\circledcirc}_{P_1,Q_1,n} & (MC)^{\circledcirc}_{P_1,Q_2,n} \\ 0 & 0 \end{bmatrix} U^*.$$

6. Conclusions with Some Applications

Two new generalized core inverse are introduced, namely, the (p, q, m)-core inverse and the $\langle p, q, n\rangle$-core inverse. These inverses extend the inverses of the $\langle$i, m$\rangle$-core inverse, the (j, m)-core inverse, the core inverse, the core-EP inverse and the DMP-inverse. The (p, q, m)-core inverse and the $\langle p, q, n\rangle$-core inverse can used in some areas such as statistics and matrix generalized inverses. There are a lot of research articles about matrix ordering and element partial ordering; by using the reverse order of the (p, q, m)-core inverse and the $\langle p, q, n\rangle$-core inverse, one can get some suitable applications in statistics, electrical networks, etc. We can obtain several partial ordering by using different generalized inverses, such as the minus ordering by using the inner inverse, the sharp ordering by using the group inverse and the core ordering by using the core inverse. The main results in this paper as follows:

If a is (p,q,m)-core invertible, then the (p,q,m)-core inverse of a is $p^m q$. Let $a,p,q \in R$ and $m \in \mathbb{N}$. Then a is (p,q,m)-core invertible if and only if there exists $e \in R$ such that $e = e^2$,

$$a = \begin{bmatrix} a_1 & 0 \\ 0 & a_2 \end{bmatrix}_{e\times e}, \quad p = \begin{bmatrix} p_1 & 0 \\ 0 & p_2 \end{bmatrix}_{e\times e} \quad and \quad q = \begin{bmatrix} q_1 & q_2 \\ 0 & 0 \end{bmatrix}_{e\times e},$$

where $p_1a_1 = a_1p_1 = (p_1a_1)^2$, $p_2a_2 = a_2p_2 = 0$, a_1 is (p_1,q_1,m)-core invertible and (p_1,q_2,m)-core invertible. The (p,q,m)-core inverse of a is given by

$$a^{\circledast}_{p,q,m} = \begin{bmatrix} (a_1)^{\circledast}_{p_1,q_1,m} & (a_1)^{\circledast}_{p_1,q_2,m} \\ 0 & 0 \end{bmatrix}_{e\times e} = \begin{bmatrix} p_1^m q_1 & p_1^m q_2 \\ 0 & 0 \end{bmatrix}_{e\times e}.$$

If $A \in \mathbb{C}^{n\times n}$ is (P,Q,m)-core invertible, then we have $PA = AP$, PA is idempotent, $X = PAX$, $A^mX = Q$ and

$$A^{\circledast}_{P,Q,m} = U\begin{bmatrix} (MC)^{\circledast}_{P_1,Q_1,m} & (MC)^{\circledast}_{P_1,Q_2,m} \\ 0 & 0 \end{bmatrix}U^*.$$

Author Contributions: Resources, Y.-Q.W.; Writing—original draft, S.-Z.X.; Writing—review & editing, J.B. and D.M. All authors have read and agreed to the published version of the manuscript.

Funding: The first author is supported by the National Natural Science Foundation of China (No. 12001223), the Qing Lan Project of Jiangsu Province and the Natural Science Foundation of Jiangsu Province of China (No. BK20220702). The fourth author is supported by the Ministry of Education, Science and Technological Development, Republic of Serbia (No. 451-03-47/2023-01/200124).

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

References

1. Malik, S.B.; Thome, N. On a new generalized inverse for matrices of an arbitrary index. *Appl. Math. Comput.* **2014**, *226*, 575–580. [CrossRef]
2. Hartwig, R.E.; Spindelböck, K. Matrices for which A^* and $A^\dagger$ commute. *Linear Multilinear Algebra* **1984**, *14*, 241–256. [CrossRef]
3. Prasad, K.M.; Mohana, K.S. Core-EP inverse. *Linear Multilinear Algebra* **2014**, *62*, 792–802. [CrossRef]
4. Wang, H.X. Core-EP decomposition and its applications. *Linear Algebra Appl.* **2016**, *508*, 289–300. [CrossRef]
5. Baksalary, O.M.; Trenkler, G. Core inverse of matrices. *Linear Multilinear Algebra* **2010**, *58*, 681–697. [CrossRef]
6. Rakić, D.S.; Dinčić, N.Č.; Djordjević, D.S. Group, Moore-Penrose, core and dual core inverse in rings with involution. *Linear Algebra Appl.* **2014**, *463*, 115–133. [CrossRef]
7. Xu, S.Z.; Chen, J.L.; Zhang, X.X. New characterizations for core inverses in rings with involution. *Front. Math. China* **2017**, *12*, 231–246. [CrossRef]
8. Benítez, J. A new decomposition for square matrices. *Electron. Linear Algebra* **2010**, *20*, 207–225. [CrossRef]
9. Benítez, J.; Liu, X.J. A short proof of a matrix decomposition with applications. *Linear Algebra Appl.* **2013**, *438*, 1398–1414. [CrossRef]
10. Wimmer, H.K. Canonical angles of unitary spaces and perturbations of direct complements. *Linear Algebra Appl.* **1999**, *287*, 373–379. [CrossRef]
11. Xu, S.Z.; Chen, J.L.; Benítez, J.; Wang, D.G. Generalized core inverses of matrices. *Miskolc Math. Notes* **2019**, *20*, 565–584. [CrossRef]
12. Li, P.Y.; Jiang, Z.L.; Zheng, Y.P. On determinants and inverses of some triband Toeplitz matrices with permuted columns. *J. Math. Comput. Sci.* **2020**, *20*, 196–206. [CrossRef]
13. Ren, B.Y.; Wang, Q.W.; Chen, X.Y. The η-anti-Hermitian solution to a system of constrained matrix equations over the generalized seggre quaternion algebra. *Symmetry* **2023**, *15*, 92. [CrossRef]

mathematics

MDPI

Article

Estimating the Gerber–Shiu Function in the Two-Sided Jumps Risk Model by Laguerre Series Expansion

Kang Hu [1], Ya Huang [2,*] and Yingchun Deng [3,*]

[1] Key Laboratory of Computing and Stochastic Mathematics (Ministry of Education), School of Mathematics and Statistics, Hunan Normal University, Changsha 410081, China
[2] School of Business, Hunan Normal University, Changsha 410081, China
[3] School of International Business, Hunan University of Information Technology, Changsha 410151, China
* Correspondence: huangya@hunnu.edu.cn (Y.H.); dengyc@hunnu.edu.cn (Y.D.)

Abstract: In this paper, we consider an insurance risk model with two-sided jumps, where downward and upward jumps typically represent claim amounts and random gains, respectively. We use the Laguerre series to expand the Gerber–Shiu function and estimate it based on observed information. Moreover, we show that the estimator is easily computed and has a fast convergence rate. Numerical examples are also provided to show the efficiency of our method when the sample size is finite.

Keywords: two-sided jumps; Gerber–Shiu function; Laguerre series; estimator

MSC: 91G05; 91G60

1. Introduction

In this paper, the surplus process of an insurance company is described by the classical risk model

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

where $U(0) = u \geq 0$ is the initial surplus and $c > 0$ is the constant premium rate per unit time. The claim number process $\{N(t)\}_{t\geq 0}$ is a homogeneous Poisson process with intensity $\lambda > 0$, and the claim sizes $\{Z_i\}_{i=1}^{\infty}$ form an independent and identically distributed sequence that may be positive or negative. For later use, the density of Z_i is denoted by $f(\cdot)$. We also assume that $\{N(t)\}_{t\geq 0}$ and $\{Z_i\}_{i=1}^{\infty}$ are independent. Furthermore, since the size of each jump Z_i can be positive or negative, we can think of it as jumping up or down, and the upward and downward jumps can be interpreted as company random gains and random losses, respectively. The size of each upward jump is defined as X_i and its density function is defined as $f_+(\cdot)$, the mean value is μ_+. Similarly, the size of each downward jump is defined as Y_i and the corresponding density function is $f_-(\cdot)$, the mean value is μ_-. Hence, we have

$$f(x) = pf_+(x)I_{\{x>0\}} + qf_-(-x)I_{\{x<0\}}, \tag{2}$$

where $p, q > 0, p + q = 1$, $I_{\{A\}}$ is an indicator function of the event A. To this end, we define $N^+(t) = \sum_{i=1}^{N(t)} I_{\{Z_i>0\}}$ to be the number of upward jumps until time t. Similarly, let

Citation: Hu, K.; Huang, Y.; Deng, Y. Estimating the Gerber–Shiu Function in the Two-Sided Jumps Risk Model by Laguerre Series Expansion. *Mathematics* **2023**, *11*, 1994. https://doi.org/10.3390/math11091994

Academic Editor: Anatoliy Swishchuk

Received: 5 March 2023
Revised: 15 April 2023
Accepted: 19 April 2023
Published: 23 April 2023

$N^-(t) = \sum_{i=1}^{N(t)} I_{\{Z_i<0\}}$ be the number of downward jumps until time t. Therefore, the surplus process (1) can be viewed as a risk model with stochastic premium income

$$U(t) = u + c - S(t) = u + c + \sum_{i=1}^{N^-(t)} Y_i - \sum_{i=1}^{N^+(t)} X_i,\ t \geq 0. \tag{3}$$

For a more detailed introduction of Equations (1) and (3), please refer to Cheung et al. [1]. Related works can be found in [2–5], among others.

Define the ruin time by $\tau = inf\{t \geq 0 : U(t) < 0\}$, and set $\tau = \infty$ if $U(t) \geq 0$ for all $t \geq 0$. In this paper, we are interested in the Gerber–Shiu expected discounted penalty function that is defined as

$$m(u) := E[e^{-\delta\tau}\omega(U(\tau-), |U(\tau)|)I_{\{\tau<\infty\}}],\ u \geq 0.$$

where $\delta \geq 0$ is the Laplace transform argument, and $\omega : [0, \infty) \times [0, \infty) \to [0, \infty)$ is a measurable penalty function of the $U(\tau-)$ and $|U(\tau)|$. This function was first introduced by Gerber and Shiu [6]. It has become an important and standard risk measure in ruin theory since various quantities of interests in ruin theory can be obtained for different values of the discount factor δ and different penalty functions ω. Interested readers are referred to [7–13], among others.

The above-mentioned papers assume that some probability characteristics of the surplus process are known, for example, the probability characteristics of the claim sizes and claim number process; however, these are usually unknown for an insurance company. In fact, we can only obtain some discrete data information about the surplus flow levels, claim numbers, and individual claim sizes (income numbers and individual income sizes). According to these data, more and more actuarial researchers use different methods to calculate statistical estimations of ruin probability and Gerber–Shiu function. Shimizu [14,15] used a regularized version of the inverse Laplace transform to estimate the Gerber–Shiu function in the Lévy risk model and the perturbed compound Poisson risk model, respectively; You and Cai [16] used a regularized version of the inverse Laplace transform to consider the nonparametric estimation of the survival probability for a spectrally negative Lévy risk model based on high-frequency data; Zhang and Yang [17,18] estimated the ruin probability based on high-frequency data and low-frequency data, respectively; Shimizu and Zhang [19] estimated the Gerber–Shiu function in a Lévy risk model based on high-frequency data by Fourier inversion transform. In addition, there are some effective estimation methods. Chau et al. [20,21] used the Fourier-cosine series expansion to estimate ruin probability and Gerber–Shiu function in the Lévy risk models; Yang et al. [22] applied two-dimensional Fourier cosine series expansion to estimate the discounted density function of the deficit at ruin; Xie and Zhang [23] applied the Fourier cosine series expansion to estimate the compound Poisson risk model under a constant barrier dividend strategy; Zhang [24] proposed a new estimator of the Gerber–Shiu function by Fourier sinc series expansion in the perturbed compound Poisson risk model; Chan [25,26] proposed a method based on the complex Fourier series expansion and used it in the actuarial field; Wang et al. [27] considered the pricing problem of variable annuities with guaranteed minimum death benefit by a complex Fourier series method under regime-switching jump diffusion models. For more detail on the statistical estimation of risk models, we refer the interested readers to [28–40].

The main goal of this paper is to use the Laguerre series expansion method to estimate the Gerber–Shiu function. The Laguerre series expansion method has been used by some authors for solving some statistical problems. For example, Comte and Genon-Catalot [41] used the appropriate Laguerre basis to take into account the estimation of the random strength of the mixed Poisson model; Zhang and Su [42,43] applied Laguerre series to approximate the Gerber–Shiu function in the class compound Poisson risk model and the Lévy risk model, respectively; Zhang and Yong [44] studied the valuation of equity-linked

annuity contracts with guaranteed minimum death benefits by Laguerre series expansion; Cheung and Zhang [45] proposed to use Laguerre series expansion as a function of the initial earnings level to approximate the ruin probability of the updated risk model; Albrecher et al. [46] considered the bivariate Laguerre expansions approach for joint ruin probabilities in a two-dimensional insurance risk process; Xie and Zhang [47] considered the finite-time dividend and ruin problems in a class of risk models under the constant dividend barrier strategy by Laguerre series expansion; Su et al. [48] considered the random deviation of premium income (or claim loss), so they studied the statistical estimation of Gerber–Shiu function in the compound Poisson risk model perturbed by diffusion. In the actual insurance business, the premium income of insurance companies, especially small companies, is sometimes random. Therefore, this paper considers the two-sided jumps risk model. For more on the Laguerre series expansion method, we refer the interested readers to [49–53].

The remainder of this paper is organized as follows: In Section 2, we first briefly introduce the Laguerre series expansion method, and then derive Laguerre series expansions of $m(u)$. In Section 3, we present how to construct estimators for the aforementioned quantities based on observed sample of the surplus process, and in Section 4, we study the consistency rate of our estimator. Finally, numerical examples are given in Section 5 to illustrate that the performance of the estimator behaves well when the sample size is finite.

2. Preliminaries

2.1. The Laguerre Series Expansion

In this subsection, we present some known results on the Laguerre series expansion method. Throughout, let $\mathbb{L}^1(\mathbb{R}_+)$ and $\mathbb{L}^2(\mathbb{R}_+)$ denote the classes of absolutely integrable functions and square integrable functions on the positive axis, respectively, and denote by $\mathbb{C}_+$ (respectively, $\mathbb{C}_{++}$) those complex numbers that have a non-negative (respectively positive) real part, that is

$$\mathbb{C}_+ := \{s \in \mathbb{C} : \mathrm{Re}(s) \geq 0\} \text{ and } \mathbb{C}_{++} := \{s \in \mathbb{C} : \mathrm{Re}(s) > 0\}.$$

For any complex number s, we denote its real part and imaginary part by $\mathrm{Re}(s)$ and $\mathrm{Im}(s)$, respectively. For two positive functions f_1, f_2 with a common domain $\mathcal{X} \in \mathbb{R}$, we use $f_1 \lesssim f_2$ to mean $f_1(x) \leq Cf_2(x)$ uniformly in $x \in \mathcal{X}$. Similarly, we use $f_1(x) \gtrsim f_2(x)$ to mean $f_1(x) \geq Cf_2(x)$ uniformly in $x \in \mathcal{X}$. For two sequences of functions $\{f_k\}$ and $\{g_k\}$, we use $f_k \lesssim$ (or $\gtrsim$)g_k to mean $f_k(x) \leq$ (or $\geq$)$Cg_k(x)$ uniformly in k and x. Denote the scalar product and $\mathbb{L}^2$-norm on $\mathbb{L}^2(\mathbb{R}_+)$ by

$$\langle f, g\rangle = \int_0^\infty f(x)g(x)dx,\ \|f\| = \sqrt{\int_0^\infty f(x)^2dx},\ \forall f, g \in \mathbb{L}^2(\mathbb{R}_+).$$

For convenience, let C be a generic positive constant that can take different values from line to line. For any $g \in \mathbb{L}^1$, we define its Laplace transform and Fourier transform by $\mathcal{L}g(s) = \int_0^\infty e^{-su}g(u)du$, $\mathrm{Re}(s) \geq 0$ and $\mathcal{F}g(s) = \int_0^\infty e^{isu}g(u)du$ $s \in \mathbb{R}$. Furthermore, let T_s be the Dickson–Hipp operator, such that

$$T_sf(y) = \int_y^\infty e^{-s(x-y)}f(x)dx = \int_0^\infty e^{-sx}f(x+y)dx,\ y \geq 0,$$

for any integrable real function f. The operator T_s was first introduced in Dickson and Hipp [54] and has many nice properties, which can be found in Li and Garrido [55]. The Laguerre functions are given by

$$\psi_k(x) = \sqrt{2}L_k(2x)e^{-x},\ x \geq 0,\ k = 0, 1, 2, \ldots, \tag{4}$$

where $\{L_k\}$ is a Laguerre polynomial defined as

$$L_k(x) = \sum_{j=0}^{k} (-1)^j \binom{k}{j} \frac{x^j}{j!},\ x \geq 0. \tag{5}$$

It is follows that the Laguerre functions are uniformly bounded, i.e.,

$$|\psi_k| \leq \sqrt{2},\ \forall\, k \geq 0 \text{ and } \forall\, x \in \mathbb{R}_+. \tag{6}$$

We also note that, for the Laguerre functions ψ_k and ψ_j, the following convolution formula holds:

$$\int_0^x \psi_k(x-y)\psi_j(y)dy = \frac{1}{\sqrt{2}}[\psi_{k+j}(x) - \psi_{k+j+1}(x)]. \tag{7}$$

For more details on the above results, refer to Abramowitz and Stegun [56].

Remark 1. *Suppose that the collection $\{\psi_k\}_{k\geq 0}$ is a complete orthonormal basis of $\mathbb{L}^2(\mathbb{R}_+)$ satisfying*

(1) $\|\psi_k\| = 1$;
(2) $\langle \psi_k, \psi_j \rangle = 0$ for $k \neq j$.

Using the orthonormal property of the Laguerre basis $\{\psi_k\}_{k\geq 0}$, for any $f \in \mathbb{L}^2(\mathbb{R}_+)$, we can develop it on the Laguerre basis

$$f(x) = \sum_{k=0}^{\infty} \langle f, \psi_k \rangle \psi_k(x).$$

In practical applications, we need to truncate the above infinite sum. Hence, for all $K \geq 0$, we have

$$f(x) \approx f_K(x) = \sum_{k=0}^{K} \langle f, \psi_k \rangle \psi_k(x).$$

To evaluate the convergence rate of the bias $\|f_K - f\|$, we introduce the Sobolev–Laguerre space (see Bongioanni and Torrea [57]) that is defined by

$$W(\mathbb{R}_+, r, B) = \left\{ f : \mathbb{R}_+ \to \mathbb{R}, f \in \mathbb{L}^2(\mathbb{R}_+),\ \sum_{k=0}^{\infty} k^r \langle f, \psi_k \rangle^2 \leq B < \infty \right\},$$

where $0 < r, B < \infty$. Suppose that r is a positive integer. If $f \in \mathbb{L}^2(\mathbb{R}_+)$, then the following properties are equivalent:

(1) $\sum_{k=0}^{\infty} k^r \langle f, \psi_k \rangle^2 < \infty$.
(2) For function f admits derivatives up to order $r-1$, with $f^{(r-1)}$ absolutely continuous and for $m = 0, 1, \ldots, r-1$, the functions

$$x^{\frac{(m+1)}{2}} (fe^x)^{(m+1)} e^{-x} = x^{\frac{(m+1)}{2}} \sum_{j=0}^{m+1} \binom{m+1}{j} f^{(j)} \tag{8}$$

belong to $\mathbb{L}^2(\mathbb{R}_+)$ (see Comte and Genon-Catalot [41]). If $f \in W(\mathbb{R}_+, r, B)$, using the orthonormal property of the Laguerre basis $\{\psi_k\}_{k\geq 0}$ (see Zhang and Su [43] and Zhang and Yong [44]), we have

$$\|f_K - f\| = \sum_{k=K+1}^{\infty} \langle f, \psi_k\rangle^2 \leq (K+1)^{-r} \sum_{k=K+1}^{\infty} k^r \langle f, \psi_k\rangle^2 \leq BK^{-r}.$$

2.2. The Laguerre Expansion of Gerber–Shiu Function

In this subsection, we show that the Gerber–Shiu function can be expressed by Laguerre functions. We focus on the $Erlang[n, \beta]$ distribution, for some $\beta > 0$ and a positive integer n, to model the premium sizes (see Labbé et al. [58]). No specific assumption is made on the claim's distribution. For $u > 0$, conditioning on the time of the first event (premium or claim), we obtain

$$m(u) = \int_0^{\infty} \lambda p e^{-(\delta+\lambda)t} \int_0^{u+ct} m(u+ct-y) f_+(y) dy dt + \int_0^{\infty} \lambda p e^{-(\delta+\lambda)t} \times \int_{u+ct}^{\infty} \omega(u+ct, y-u-ct) f_+(y) dy dt + \int_0^{\infty} \lambda q e^{-(\delta+\lambda)t} \int_0^{\infty} m(u+ct+y) f_-(y) dy dt,$$

hence

$$m(u) = \int_0^u m(u-y) f_\delta(y) dy + H_{\delta,w}(u), \tag{9}$$

where

$$f_\delta(y) = \frac{p\lambda}{c}\left[(-1)^n \sum_{i=1}^{n+1} \frac{(\beta-\rho_i)^n}{\prod_{j=1, j\neq i}^{n+1}(\rho_i-\rho_j)} T_{\rho_i} f_+(y)\right], y \geq 0,$$

$$H_{\delta,w}(u) = \frac{p\lambda}{c}\left[(-1)^n \sum_{i=1}^{n+1} \frac{(\beta-\rho_i)^n}{\prod_{j=1, j\neq i}^{n+1}(\rho_i-\rho_j)} T_{\rho_i} \eta(u)\right], u \geq 0,$$

$$\eta(u) = \int_u^{\infty} \omega(u, y-u) f_+(y) dy.$$

For any $\delta \geq 0$, in the following Lundberg's fundamental equation (in s)

$$\chi(s) := [\lambda + \delta - cs - p\lambda \mathcal{L} f(s)](\beta - s)^n - q\lambda\beta^n = 0,\ s \in \mathbb{C}_+, \tag{10}$$

ρ_i and ρ_j are the $n+1$ roots of the above equation.

Remark 2. *Assume, in addition, that $n = 1$ (i.e., the annuity income amounts follow the exponential distribution).*

$$\chi(s) := [\lambda + \delta - cs - p\lambda \mathcal{L} f_+(s)](\beta - s) - q\lambda\beta = 0,\ s \in \mathbb{C}_+. \tag{11}$$

The above equation has two positives roots, $\rho_1 \in (0, \beta)$ and $\rho_2 \in (\beta, \infty)$. It is clear from Equation (11) that the continuous function $\chi(s)$ is such that $\chi(0) = \delta\beta > 0, \chi(\beta) =$

$-q\lambda\beta < 0$ and $\lim_{s\to\infty}\chi(s) = \infty$. Thus, the existence of two distinct roots satisfying $0 \le \rho_1 < \beta < \rho_2 < \infty$ is established.

$$f_\delta(y) = \frac{p\lambda}{c}\left[\sum_{i=1}^{2}\frac{(\rho_i-\beta)}{\prod_{j=1,j\neq i}^{2}(\rho_i-\rho_j)}T_{\rho_i}f_+(y)\right], y \ge 0,$$

$$H_{\delta,w}(u) = \frac{p\lambda}{c}\left[\sum_{i=1}^{2}\frac{(\rho_i-\beta)}{\prod_{j=1,j\neq i}^{2}(\rho_i-\rho_j)}T_{\rho_i}\eta(u)\right], u \ge 0,$$

$$\eta(u) = \int_u^\infty \omega(u, y-u)f_+(y)dy.$$

In the following, we suppose that some conditions hold true in this paper, which has also been considered in Shimizu and Zhang [19].

Condition 1. *(Net profit condition.)*

$$ct - E[S(t)] = ct + q\lambda t\mu_- - p\lambda t\mu_+ > 0,\ t > 0.$$

The above condition guarantees that the expectation of the surplus process will always be positive at any time $t > 0$. From a practical point of view, we only consider the case of $c > p\lambda$ in this paper.

Condition 2. *For the penalty function w, it satisfies*

$$\int_0^\infty\int_0^\infty (1+x)\omega(x,y)f_+(x+y)dydx < \infty.$$

Condition 3. *For the penalty function w, there exist some integers α_1, α_2 such that*

$$w(x,y) \lesssim (1+x)^{\alpha_1}(1+y)^{\alpha_2}.$$

In order to use the Laguerre series expansion method to calculate Equation (9), we need to ensure that $m \in \mathbb{L}^2(\mathbb{R}^+)$. Using inequality $(x+y)^2 \le 2x^2 + 2y^2$, we obtain

$$\begin{aligned}\int_0^\infty m^2(u)du &= \int_0^\infty\left(\int_0^u m(u-y)f_\delta(y)dy + H_{\delta,w}(u)\right)^2 du \\ &\le 2\int_0^\infty\left(\int_0^u m(u-y)f_\delta(y)dy\right)^2 du + 2\int_0^\infty (H_{\delta,w}(u))^2 du.\end{aligned} \tag{12}$$

As can be seen from Equation (12), in order to determine $m \in \mathbb{L}^2(\mathbb{R}^+)$, we need some Lemmas.

Lemma 1. *For function f_δ, by $\delta > 0$, $\mu_- = \frac{1}{\beta}$ and Condition 1, we have $f_\delta \in \mathbb{L}^2(\mathbb{R}^+)$.*

Proof. Because

$$\begin{aligned}\int_0^\infty f_\delta(x)dx &= \frac{p\lambda}{c}\sum_{i=1}^{2}\frac{(\rho_i-\beta)}{\prod_{j=1,j\neq i}^{2}(\rho_i-\rho_j)}\int_0^\infty\int_x^\infty e^{-\rho_i(y-x)}f_+(y)dydx\\ &= \frac{p\lambda}{c}\sum_{i=1}^{2}\frac{(\rho_i-\beta)}{\prod_{j=1,j\neq i}^{2}(\rho_i-\rho_j)}\int_0^\infty\int_0^y e^{-\rho_i(y-x)}f_+(y)dxdy\\ &\leq \frac{p\lambda}{c}\sum_{i=1}^{2}\frac{(\rho_i-\beta)}{\prod_{j=1,j\neq i}^{2}(\rho_i-\rho_j)}\int_0^\infty yf_+(y)dy\\ &= \frac{p\lambda}{c}\cdot\mu_+ < \frac{\lambda}{c}\frac{q}{\beta}+1 < \frac{\lambda+c\beta}{c\beta}.\end{aligned}$$

Note that

$$f_\delta(x) \leq \frac{p\lambda}{c}\sum_{i=1}^{2}\frac{(\rho_i-\beta)}{\prod_{j=1,j\neq i}^{2}(\rho_i-\rho_j)}\int_0^\infty f_+(y)dy < \frac{\lambda}{c}.$$

Hence,

$$\int_0^\infty (f_\delta(x))^2dx \leq \frac{\lambda}{c}\int_0^\infty f_\delta(x)dx \leq \frac{\lambda^2+c\lambda\beta}{c^2\beta} < \infty. \tag{13}$$

This completes the proof. □

Lemma 2. *Under Condition 2, we have* $H_{\delta,w} \in \mathbb{L}^2(\mathbb{R}^+)$.

Proof.

$$\sup_{u\geq 0} H_{\delta,w}(u) \leq \frac{p\lambda}{c}\int_0^\infty \eta(u)du = \frac{p\lambda}{c}\int_0^\infty\int_0^\infty \omega(x,y)f(x+y)dydx < \infty$$

and

$$\begin{aligned}\int_0^\infty H_{\delta,w}(u)du =& \frac{p\lambda}{c}\sum_{i=1}^{2}\frac{(\rho_i-\beta)}{\prod_{j=1,j\neq i}^{2}(\rho_i-\rho_j)}\int_0^\infty\int_u^\infty\int_x^\infty e^{-\rho_i(y-x)}\omega(x,y-x)f(y)dydxdu\\ \leq& \frac{p\lambda}{c}\int_0^\infty\int_u^\infty\int_x^\infty \omega(x,y-x)f(y)dydxdu\\ =& \frac{p\lambda}{c}\int_0^\infty\int_0^\infty x\omega(x,y)f(x+y)dydx < \infty,\end{aligned}$$

which yield

$$\int_0^\infty (H_{\delta,w}(u))^2du \leq \sup_{u\geq 0} H_{\delta,w}(u)\times\left(\int_0^\infty H_{\delta,w}(u)du\right) < \infty. \tag{14}$$

This completes the proof. □

Lemma 3. *As for m, by Conditions 1 and 2, we obtain* $m \in \mathbb{L}^2(\mathbb{R}^+)$.

Proof. By Equation (9), we have

$$\int_0^\infty m(u)du = \frac{\int_0^\infty H_{\delta,w}(u)du}{1-\int_0^\infty f_\delta(y)dy} < \infty,$$

i.e., $m \in \mathbb{L}^1(\mathbb{R}^+)$. According to Theorem 1.4.5 in Stenger [59], we can obtain $\int_0^u m(u-x)f_\delta(x)dx \in \mathbb{L}^2(\mathbb{R}^+)$ due to $f_\delta \in \mathbb{L}^2(\mathbb{R}^+)$. Furthermore, according to Equations (12) and (14), we can obtain $m \in \mathbb{L}^2(\mathbb{R}^+)$. □

In the remainder of this paper, suppose that $m, f_\delta, H_{\delta,w} \in \mathbb{L}^2(\mathbb{R}^+)$. Then we can develop them on the Laguerre basis, i.e.,

$$m(u) = \sum_{k=0}^{\infty} P_k \psi_k(u),\ u \geq 0, \tag{15}$$

$$f_\delta(x) = \sum_{k=0}^{\infty} Q_k \psi_k(x),\ x \geq 0, \tag{16}$$

$$H_{\delta,w}(x) = \sum_{k=0}^{\infty} R_k \psi_k(x),\ x \geq 0, \tag{17}$$

where for $k = 0, 1, 2, \ldots$

$$P_k = \langle m, \psi_k \rangle,\ Q_k = \langle f_\delta, \psi_k \rangle,\ R_k = \langle H_{\delta,w}, \psi_k \rangle.$$

Plugging the Laguerre series expansion Equations (15)–(17) into the defective renewal Equation (9) and using the convolution Formula (7), we obtain

$$\begin{aligned}
\sum_{k=0}^{\infty} P_k \psi_k(u) &= \int_0^u \sum_{k=0}^{\infty} P_k \psi_k(u-x) \cdot \sum_{j=0}^{\infty} Q_j \psi_j(x) + \sum_{k=0}^{\infty} R_k \psi_k(u) \\
&= \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} P_k Q_j \int_0^u \psi_k(u-x)\psi_j(x) + \sum_{k=0}^{\infty} R_k \psi_k(u) \\
&= \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} \frac{1}{\sqrt{2}} P_k Q_j [\psi_{k+j}(u) - \psi_{k+j+1}(u)] + \sum_{k=0}^{\infty} R_k \psi_k(u).
\end{aligned} \tag{18}$$

Furthermore, by changing the order of summation we obtain

$$\sum_{k=0}^{\infty} \sum_{j=0}^{\infty} \frac{1}{\sqrt{2}} P_k Q_j \psi_{k+j}(u) = \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} \frac{1}{\sqrt{2}} P_j Q_{k-j} \psi_k(u)$$

and

$$\sum_{k=0}^{\infty} \sum_{j=0}^{\infty} \frac{1}{\sqrt{2}} P_k Q_j \psi_{k+j+1}(u) = \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} \frac{1}{\sqrt{2}} P_j Q_{k-j-1} \psi_k(u).$$

As a result, Equation (18) gives

$$\begin{aligned}
\sum_{k=0}^{\infty} P_k \psi_k(u) = &\left[\frac{1}{\sqrt{2}} P_0 Q_0 + R_0\right] \psi_0(u) \\
&+ \sum_{k=1}^{\infty} \left(\sum_{j=0}^{k-1} \frac{1}{\sqrt{2}} P_j (Q_{k-1} - Q_{k-j-1}) + \frac{1}{\sqrt{2}} P_k Q_0 + R_k \right) \psi_k(u).
\end{aligned} \tag{19}$$

After comparing the coefficients for each basis function $\psi_k(u)$ on both sides of Equation (19), we obtain an infinite triangular system of linear equations,

$$\begin{cases} P_0 = \frac{1}{\sqrt{2}} P_0 Q_0 + R_0, \\ P_k = \sum_{j=0}^{k-1} \frac{1}{\sqrt{2}} P_j (Q_{k-1} - Q_{k-j-1}) + \frac{1}{\sqrt{2}} P_k Q_0 + R_k, \quad k \geq 1. \end{cases} \tag{20}$$

Furthermore, let $\vec{\mathbf{p}} = (P_0, P_1, P_2, \ldots)^{\mathrm{T}}$, $\vec{\mathbf{r}} = (R_0, R_1, R_2, \ldots)^{\mathrm{T}}$ and

$$\mathbf{A} = \begin{pmatrix} 1 - \frac{1}{\sqrt{2}} Q_0 & 0 & 0 & \ldots \\ \frac{1}{\sqrt{2}}(Q_0 - Q_1) & 1 - \frac{1}{\sqrt{2}} Q_0 & 0 & \ldots \\ \frac{1}{\sqrt{2}}(Q_1 - Q_2) & \frac{1}{\sqrt{2}}(Q_0 - Q_1) & 1 - \frac{1}{\sqrt{2}} Q_0 & \ldots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} := (a_{ij})_{i,j \geq 1}.$$

Then we can write Equation (20) in the following matrix form

$$\mathbf{A}\vec{\mathbf{p}} = \vec{\mathbf{r}}. \tag{21}$$

Note that $\mathbf{A}$ is a lower triangular Toeplitz matrix, and for the non-zero elements in $\mathbf{A}$, we have

$$\begin{aligned} \left| \frac{1}{\sqrt{2}}(Q_k - Q_{k-1}) \right| &\leq \frac{1}{\sqrt{2}} \int_0^\infty f_\delta(x) |\psi_k(x)| dx + \frac{1}{\sqrt{2}} \int_0^\infty f_\delta(x) |\psi_{k-1}(x)| dx \\ &\leq 2 \int_0^\infty f_\delta(x) dx \leq 2\left(1 + \frac{q\lambda}{c\beta}\right), \ k \geq 1. \end{aligned}$$

Furthermore, we have

$$\begin{aligned} 1 - \frac{1}{\sqrt{2}} Q_0 &= 1 - \frac{1}{\sqrt{2}} \langle f_\delta, \psi_0 \rangle = 1 - \frac{1}{\sqrt{2}} \int_0^\infty f_\delta(x) \psi_0(x) dx \\ &= 1 - \int_0^\infty f_\delta(x) e^{-x} dx > 1 - \frac{p\lambda}{c} > 0, \end{aligned}$$

by Condition 1, then $\mathbf{A}$ is nonsingular and explicitly invertible.

Hence, for all $K \geq 0$, truncating the infinite dimension vectors and matrix in Equation (21) leads to

$$\mathbf{A}_K \vec{\mathbf{p}}_K = \vec{\mathbf{r}}_K, \tag{22}$$

where $\vec{\mathbf{p}}_K = (P_0, P_1, P_2, \ldots, P_K)^{\mathrm{T}}$, $\vec{\mathbf{r}}_K = (R_0, R_1, R_2, \ldots, R_K)^{\mathrm{T}}$, and $\mathbf{A}_K = (a_{ij})_{i,j=1}^{K+1}$. As a result, the matrix $\mathbf{A}_K$ is nonsingular and explicitly invertible. Then we have

$$\vec{\mathbf{p}}_K = \mathbf{A}_K^{-1} \vec{\mathbf{r}}_K. \tag{23}$$

After solving Equation (23), we can obtain $\vec{\mathbf{p}}_K$, and for a larger K we can approximate the Gerber–Shiu function as follows:

$$m(u) \approx m_K(u) := \sum_{k=0}^{K} P_k \psi_k(u), \ u \geq 0. \tag{24}$$

3. Estimation Procedure

In this section, we assume that both Poisson intensity and claim size density are unknown, but we can obtain discrete information about the surplus process and the aggregate claims. Assume that we can observe the surplus process over a long time interval

$[0, T]$. Let $\triangle > 0$ be a fixed inter-observation interval (or sampling interval). Without loss of generality, assume that $T/\triangle$ is an integer, and let $n = T/\triangle$.

(1) Data-set of surplus levels:

$$\{U_{j\triangle} : j = 0, 1, 2, \ldots, n\},$$

where $U_{j\triangle}$ is the observed surplus level at time $t = j\triangle$.

(2) Data-set of total claim numbers and claim sizes:

$$\{N_{j\triangle}, Z_1, Z_2, \ldots, Z_{N_{j\triangle}}\},\ j = 1, \ldots, n.$$

(3) Data-set of downward jump numbers and random loss sizes:

$$\{N^+_{j\triangle}, X_1, X_2, \ldots, X_{N^+_{j\triangle}}\},\ j = 1, \ldots, n.$$

(4) Data-set of upward jump numbers and random income sizes:

$$\{N^-_{j\triangle}, Y_1, Y_2, \ldots, Y_{N^-_{j\triangle}}\},\ j = 1, \ldots, n.$$

where $N_{j\triangle}$ is the total claim number up to time $t = j\triangle$ and $N_{j\triangle} = N^+_{j\triangle} + N^-_{j\triangle}$.

Next, we shall propose our estimator of the Gerber–Shiu function by Laguerre expansion based on Equation (24). To this end, we need to estimate the vector $\vec{\mathbf{p}}_K$, or equivalently, $\mathbf{A}_K$ and $\vec{\mathbf{r}}_K$. By the definitions of $\mathbf{A}_K$ and $\vec{\mathbf{r}}_K$, we only need to estimate the following quantities:

$$Q_k,\ R_k,\ k = 0, 1, 2, \ldots, K.$$

By the definitions of Q_k and R_k and changing the order of integrals, we can write Q_k and R_k as follows:

$$\begin{aligned} Q_k = \langle f_\delta, \psi_k \rangle &= \int_0^\infty f_\delta(x)\psi_k(x)dx = \frac{p\lambda}{c}\sum_{i=1}^{2}\frac{\rho_i - \beta}{\prod_{j=1,j\neq i}^{2}(\rho_i - \rho_j)}\int_0^\infty\int_x^\infty e^{-\rho_i(y-x)}f_+(y)dy\psi_k(x)dx \\ &= \frac{p\lambda}{c}\sum_{i=1}^{2}\frac{\rho_i - \beta}{\prod_{j=1,j\neq i}^{2}(\rho_i - \rho_j)}\int_0^\infty\int_0^y e^{-\rho_i(y-x)}\psi_k(x)dxf_+(y)dy \\ &= \frac{p\lambda}{c}\sum_{i=1}^{2}\frac{\rho_i - \beta}{\prod_{j=1,j\neq i}^{2}(\rho_i - \rho_j)}E\left[\int_0^X e^{-\rho_i(X-x)}\psi_k(x)dx\right] \end{aligned} \tag{25}$$

and

$$\begin{aligned} R_k = \langle H_{\delta,w}, \psi_k \rangle &= \int_0^\infty H_{\delta,w}(u)\psi_k(x)dx = \frac{p\lambda}{c}\sum_{i=1}^{2}\frac{\rho_i - \beta}{\prod_{j=1,j\neq i}^{2}(\rho_i - \rho_j)}\int_0^\infty\int_u^\infty e^{-\rho_i(y-u)}\eta(y)dy\psi_k(u)du \\ &= \frac{p\lambda}{c}\sum_{i=1}^{2}\frac{\rho_i - \beta}{\prod_{j=1,j\neq i}^{2}(\rho_i - \rho_j)}\int_0^\infty\int_u^\infty\int_y^\infty e^{-\rho_i(y-u)}\omega(y, x-y)f_+(x)dxdy\psi_k(u)du \\ &= \frac{p\lambda}{c}\sum_{i=1}^{2}\frac{\rho_i - \beta}{\prod_{j=1,j\neq i}^{2}(\rho_i - \rho_j)}\int_0^\infty\int_0^x\int_u^x e^{-\rho_i(y-u)}\omega(y, x-y)f_+(x)dy\psi_k(u)dudx \\ &= \frac{p\lambda}{c}\sum_{i=1}^{2}\frac{\rho_i - \beta}{\prod_{j=1,j\neq i}^{2}(\rho_i - \rho_j)}E\left[\int_0^X\int_u^X e^{-\rho_i(y-u)}\omega(y, X-y)dy\psi_k(u)du\right]. \end{aligned} \tag{26}$$

The above two formulae imply that we have to estimate the Poisson intensity λ, p, q, β, the root ρ_1, ρ_2, and the expectations appearing in Equations (25) and (26).

According to the property of Poisson distribution, we can estimate p and λ by

$$\widehat{p} = \frac{N_T^+}{N_T},\ \widehat{q} = 1 - \widehat{p},\ \widehat{\lambda} = \frac{N_T}{T}.$$

Since the premium size Y follows the $Erlang(1, \beta)$ distribution, we have $E[Y] = 1/\beta$, then we can estimate β by

$$\widehat{\beta} = \frac{1}{\frac{1}{N_T^-}\sum\limits_{j=1}^{N_T^-} Y_i},$$

which are all unbiased estimates. We estimate the root ρ_1, ρ_2 by $\widehat{\rho}_1, \widehat{\rho}_2$, which is a positive root of the following estimating equation:

$$\left[\widehat{\lambda} + \delta - cs - \widehat{p}\widehat{\lambda}\widehat{\mathcal{L}f}_+(s)\right](\widehat{\beta} - s) - \widehat{q}\widehat{\lambda}\widehat{\beta} = 0,\ s \in \mathbb{C}_+ \tag{27}$$

where $\widehat{\mathcal{L}f}_+(s) = \frac{1}{N_T^+}\sum\limits_{j=1}^{N_T^+} e^{-sX_j}$ is an estimate of the Laplace transform $\mathcal{L}f_+(s)$. It follows from Equation (25) that we have

$$\begin{aligned}
\widehat{Q}_k &= \frac{1}{cT}\sum_{i=1}^{2}\frac{\widehat{\rho}_i - \widehat{\beta}}{\prod\limits_{j=1,j\neq i}^{2}(\widehat{\rho}_i - \widehat{\rho}_j)}\sum_{j=1}^{N_T^+}\int_0^{X_j} e^{-\widehat{\rho}_i(X_j - x)}\psi_k(x)dx \\
&= \frac{\sqrt{2}}{cT}\sum_{i=1}^{2}\frac{\widehat{\rho}_i - \widehat{\beta}}{\prod\limits_{j=1,j\neq i}^{2}(\widehat{\rho}_i - \widehat{\rho}_j)}\sum_{j=1}^{N_T^+}\int_0^{X_j} e^{-\widehat{\rho}_i(X_j - x) - x}L_k(2x)dx \\
&= \frac{\sqrt{2}}{cT}\sum_{i=1}^{2}\frac{\widehat{\rho}_i - \widehat{\beta}}{\prod\limits_{j=1,j\neq i}^{2}(\widehat{\rho}_i - \widehat{\rho}_j)}\sum_{j=1}^{N_T^+}\sum_{m=0}^{k}(-2)^m\binom{k}{m}e^{-\widehat{\rho}_i X_j}\int_0^{X_j} e^{-(1-\widehat{\rho}_i)x}\cdot\frac{x^m}{m!}dx \\
&= \frac{\sqrt{2}}{cT}\sum_{i=1}^{2}\frac{\widehat{\rho}_i - \widehat{\beta}}{\prod\limits_{j=1,j\neq i}^{2}(\widehat{\rho}_i - \widehat{\rho}_j)}\sum_{j=1}^{N_T^+}\sum_{m=0}^{k}\frac{(-2)^m}{(1-\widehat{\rho}_i)^{m+1}}\binom{k}{m}e^{-\widehat{\rho}_i X_j}\left(1 - \sum_{l=0}^{m}e^{-(1-\widehat{\rho}_i)X_j}\frac{[(1-\rho_i)X_j]^l}{l!}\right).
\end{aligned} \tag{28}$$

Similarly, we can estimate R_k by

$$\widehat{R}_k = \frac{1}{cT}\sum_{i=1}^{2}\frac{\widehat{\rho}_i - \widehat{\beta}}{\prod\limits_{j=1,j\neq i}^{2}(\widehat{\rho}_i - \widehat{\rho}_j)}\sum_{j=1}^{N_T^+}\int_0^{X_j}\int_u^{X_j} e^{-\widehat{\rho}_i(x-u)}\omega(x, X_j - x)dx\psi_k(u)du. \tag{29}$$

Now, we define the estimates of $\mathbf{A}_K$ and $\vec{\mathbf{r}}_K$ by replacing Q_k and R_k with $\widehat{Q}_k$ and $\widehat{R}_k$ in their definitions, and denote them by $\widehat{\mathbf{A}}_K$ and $\vec{\widehat{\mathbf{r}}}_K$, respectively. Accordingly, the estimate of $\vec{\mathbf{p}}_K$, denoted by $\vec{\widehat{\mathbf{p}}}_K := (\widehat{P}_0, \widehat{P}_1, \ldots, \widehat{P}_K)^{\mathrm{T}}$, is defined to be the solution of the following linear system:

$$\widehat{\mathbf{A}}_K\vec{\widehat{\mathbf{p}}}_K = \vec{\widehat{\mathbf{r}}}_K. \tag{30}$$

Finally, replacing P_K by $\widehat{P}_K$ in Equation (23), we obtain the following estimate of the Gerber–Shiu function:

$$\widehat{m}_K(u) = \sum_{k=0}^{K} \widehat{P}_k \psi_k(u),\ u \geq 0. \tag{31}$$

Remark 3. *The estimator $\widehat{R}_k$ given in Equation (29) is expressed in a two-fold integral, which can be explicitly computed for most of the widely used penalty functions. Here are some examples.*

(1) $\delta = 0$ and $\omega = 1$. In this case, the Gerber–Shiu function becomes the ruin probability and we have $\widehat{\rho}_1 = 0$ and $\widehat{\rho}_2 \in (\widehat{\beta}, \infty)$. Then

$$\begin{aligned}\widehat{R}_k = {} & \frac{\sqrt{2}}{cT}\frac{\widehat{\beta}}{\widehat{\rho}_2}\sum_{j=1}^{N_T^+}\sum_{m=0}^{k}(-2)^m\binom{k}{m}\left[X_j\left(1-\sum_{l=0}^{m}e^{-X_j}\frac{X_j^l}{l!}\right)-(m+1)\left(1-\sum_{l=0}^{m+1}e^{-X_j}\frac{X_j^l}{l!}\right)\right] \\ & + \frac{\sqrt{2}}{cT}\frac{\widehat{\rho}_2-\widehat{\beta}}{\widehat{\rho}_2^2}\sum_{j=1}^{N_T^+}\sum_{m=0}^{k}(-2)^m\binom{k}{m}\left(1-\sum_{l=0}^{m}e^{-X_j}\frac{X_j^l}{l!}\right) \\ & - \frac{\sqrt{2}}{cT}\frac{\widehat{\rho}_2-\widehat{\beta}}{\widehat{\rho}_2^2}\sum_{j=1}^{N_T^+}\sum_{m=0}^{k}\frac{(-2)^m}{(1-\widehat{\rho}_2)^{m+1}}\binom{k}{m}e^{-\widehat{\rho}_2 X_j}\left(1-\sum_{l=0}^{m}e^{-(1-\widehat{\rho}_2)X_j}\frac{((1-\widehat{\rho}_2)X_j)^l}{l!}\right).\end{aligned}$$

(2) $\delta > 0$ and $\omega = 1$. In this case, the Gerber–Shiu function becomes the Laplace transform of ruin time and we have $\widehat{\rho}_1 \in (0, \widehat{\beta})$ and $\widehat{\rho}_2 \in (\widehat{\beta}, \infty)$. Then

$$\begin{aligned}\widehat{R}_k = {} & \frac{\sqrt{2}}{cT}\frac{\widehat{\rho}_1-\widehat{\beta}}{(\widehat{\rho}_1-\widehat{\rho}_2)\widehat{\rho}_1}\sum_{j=1}^{N_T^+}\sum_{m=0}^{k}(-2)^m\binom{k}{m}\left(1-\sum_{l=0}^{m}e^{-X_j}\frac{X_j^l}{l!}\right) \\ & - \frac{\sqrt{2}}{cT}\frac{\widehat{\rho}_1-\widehat{\beta}}{(\widehat{\rho}_1-\widehat{\rho}_2)\widehat{\rho}_1}\sum_{j=1}^{N_T^+}\sum_{m=0}^{k}\frac{(-2)^m}{(1-\widehat{\rho}_1)^{m+1}}\binom{k}{m}e^{-\widehat{\rho}_1 X_j}\left(1-\sum_{l=0}^{m}e^{-(1-\widehat{\rho}_1)X_j}\frac{((1-\widehat{\rho}_1)X_j)^l}{l!}\right) \\ & + \frac{\sqrt{2}}{cT}\frac{\widehat{\rho}_2-\widehat{\beta}}{(\widehat{\rho}_2-\widehat{\rho}_1)\widehat{\rho}_2}\sum_{j=1}^{N_T^+}\sum_{m=0}^{k}(-2)^m\binom{k}{m}\left(1-\sum_{l=0}^{m}e^{-X_j}\frac{X_j^l}{l!}\right) \\ & - \frac{\sqrt{2}}{cT}\frac{\widehat{\rho}_2-\widehat{\beta}}{(\widehat{\rho}_2-\widehat{\rho}_1)\widehat{\rho}_2}\sum_{j=1}^{N_T^+}\sum_{m=0}^{k}\frac{(-2)^m}{(1-\widehat{\rho}_2)^{m+1}}\binom{k}{m}e^{-\widehat{\rho}_2 X_j}\left(1-\sum_{l=0}^{m}e^{-(1-\widehat{\rho}_2)X_j}\frac{((1-\widehat{\rho}_2)X_j)^l}{l!}\right).\end{aligned}$$

(3) $\delta = 0$ and $\omega(x, y) = x + y$. In this case, the Gerber–Shiu function becomes the expected claim size causing ruin. Then

$$\begin{aligned}\widehat{R}_k = {} & \frac{\sqrt{2}}{cT}\frac{\widehat{\beta}}{\widehat{\rho}_2}\sum_{j=1}^{N_T^+}\sum_{m=0}^{k}(-2)^m\binom{k}{m}\left[X_j^2\left(1-\sum_{l=0}^{m}e^{-X_j}\frac{X_j^l}{l!}\right)-X_j(m+1)\left(1-\sum_{l=0}^{m+1}e^{-X_j}\frac{X_j^l}{l!}\right)\right] \\ & + \frac{\sqrt{2}}{cT}\frac{\widehat{\rho}_2-\widehat{\beta}}{\widehat{\rho}_2^2}\sum_{j=1}^{N_T^+}\sum_{m=0}^{k}X_j(-2)^m\binom{k}{m}\left(1-\sum_{l=0}^{m}e^{-X_j}\frac{X_j^l}{l!}\right) \\ & - \frac{\sqrt{2}}{cT}\frac{\widehat{\rho}_2-\widehat{\beta}}{\widehat{\rho}_2^2}\sum_{j=1}^{N_T^+}\sum_{m=0}^{k}\frac{X_j(-2)^m}{(1-\widehat{\rho}_2)^{m+1}}\binom{k}{m}e^{-\widehat{\rho}_2 X_j}\left(1-\sum_{l=0}^{m}e^{-(1-\widehat{\rho}_2)X_j}\frac{((1-\widehat{\rho}_2)X_j)^l}{l!}\right).\end{aligned}$$

(4) $\delta = 0$ and $\omega(x, y) = y$. In this case, the Gerber–Shiu function reduces to the expected deficit at ruin. Then

$$\begin{aligned}\widehat{R}_k = {} & \frac{\sqrt{2}}{2cT}\frac{\widehat{\beta}}{\widehat{\rho}_2}\sum_{j=1}^{N_T^+}\sum_{m=0}^{k}(-2)^m\binom{k}{m}\left[X_j^2\left(1-\sum_{l=0}^{m}e^{-X_j}\frac{X_j^l}{l!}\right)-2X_j(m+1)\left(1-\sum_{l=0}^{m+1}e^{-X_j}\frac{X_j^l}{l!}\right)\right]\\ & + \frac{\sqrt{2}}{2cT}\frac{\widehat{\beta}}{\widehat{\rho}_2}\sum_{j=1}^{N_T^+}\sum_{m=0}^{k}(-2)^m\binom{k}{m}(m+1)(m+2)\left(1-\sum_{l=0}^{m+2}e^{-X_j}\frac{X_j^l}{l!}\right)\\ & + \frac{\sqrt{2}}{cT}\frac{\widehat{\rho}_2-\widehat{\beta}}{\widehat{\rho}_2^2}\sum_{j=1}^{N_T^+}\sum_{m=0}^{k}(-2)^m\binom{k}{m}\left[X_j\left(1-\sum_{l=0}^{m}e^{-X_j}\frac{X_j^l}{l!}\right)-(m+1)\left(1-\sum_{l=0}^{m+1}e^{-X_j}\frac{X_j^l}{l!}\right)\right]\\ & - \frac{\sqrt{2}}{cT}\frac{\widehat{\rho}_2-\widehat{\beta}}{\widehat{\rho}_2^3}\sum_{j=1}^{N_T^+}\sum_{m=0}^{k}(-2)^m\binom{k}{m}\left(1-\sum_{l=0}^{m}e^{-X_j}\frac{X_j^l}{l!}\right)\\ & + \frac{\sqrt{2}}{cT}\frac{\widehat{\rho}_2-\widehat{\beta}}{\widehat{\rho}_2^3}\sum_{j=1}^{N_T^+}\sum_{m=0}^{k}\frac{(-2)^m}{(1-\widehat{\rho}_2)^{m+1}}\binom{k}{m}e^{-\widehat{\rho}_2X_j}\left(1-\sum_{l=0}^{m}e^{-(1-\widehat{\rho}_2)X_j}\frac{((1-\widehat{\rho}_2)X_j)^l}{l!}\right).\end{aligned}$$

4. Consistency Properties

In this section, we study the asymptotic properties of our estimator. We measure the performance of the estimator $\widehat{m}_K$ by the $\mathbb{L}^2$-norm distance $\|\widehat{m}_K - m\|$. By $\mathbb{L}^2$-norm inequality, we have

$$\|\widehat{m}_K - m\|^2 = \|\widehat{m}_K - m_K + m_K - m\|^2 \le 2\|\widehat{m}_K - m_K\|^2 + 2\|m_K - m\|^2, \tag{32}$$

where $\|m_K - m\|$ is the series truncation error and $\|\widehat{m}_K - m_K\|$ is the error due to statistical estimation. Now, if $m \in W(\mathbb{R}_+, r, B)$, we have

$$\|m_K - m\|^2 = \left\|\sum_{k=K+1}^{\infty} P_k \cdot \psi_k\right\|^2 = \sum_{k=K+1}^{\infty} P_k^2 = \sum_{k=K+1}^{\infty}\langle m, \psi_k\rangle^2 \le \frac{B}{(K+1)^r} = O(K^{-r}) \tag{33}$$

due to Remark 1. The polynomial convergence rate in Equation (33) can be improved when m has an exponential decay rate.

Next, it remains to study the convergence rate for $\|\widehat{m}_K - m_K\|$, and we obtain the result as follows:

Theorem 1. *Suppose $EX^2 < \infty$ and Conditions 1–3 hold. If $K = o(T^{\frac{1}{2}})$, then*

$$\|\widehat{m}_K - m\|^2 \le 2\|m_K - m\|^2 + O_p(K^2T^{-1}). \tag{34}$$

Further, if $m \in W(\mathbb{R}_+, r, B)$, then

$$\|\widehat{m}_K - m\|^2 = O(K^{-r}) + O_p(K^2T^{-1}). \tag{35}$$

In the following, we present some notations on matrix (and vector) norms. For a vector $\vec{\mathbf{b}} = (b_1, b_2, \ldots, b_n)^{\mathrm{T}}$, its 2-norm is defined by $\|\vec{\mathbf{b}}\|_2 = \sqrt{\sum_{i=1}^n |b_i|^2}$. For a matrix $\mathbf{B} = (b_{ij})_{i,j=1}^n$, its spectral norm is defined by $\|\mathbf{B}\|_2 = \sqrt{\lambda_{\max}(\mathbf{B}^{\mathrm{T}}\mathbf{B})}$, where $\lambda_{\max}(\mathbf{B}^{\mathrm{T}}\mathbf{B})$ is the largest eigenvalue of $\mathbf{B}^{\mathrm{T}}\mathbf{B}$. The Frobenius norm of $\mathbf{B}$ is defined by

$$\|\mathbf{B}\|_F = \sqrt{\mathrm{tr}(\mathbf{B}^{\mathrm{T}}\mathbf{B})} = \sqrt{\sum_{i=1}^n\sum_{j=1}^n |b_{i,j}|^2}.$$

It is known that

$$\|\mathbf{B}\vec{\mathbf{b}}\|_2 \le \|\mathbf{B}\|_2 \cdot \|\vec{\mathbf{b}}\|_2, \ \|\mathbf{B}\|_2 \le \|\mathbf{B}\|_F. \tag{36}$$

For two square matrices $\mathbf{B}_1$ and $\mathbf{B}_2$ with the same dimension, we have $\|\mathbf{B}_1\mathbf{B}_2\|_2 \leq \|\mathbf{B}_1\|_2 \cdot \|\mathbf{B}_2\|_2$.

By the inequality $(x+y)^2 \leq 2x^2 + 2y^2$ and the first inequality in Equation (36), we obtain

$$\begin{aligned}
&\|\widehat{m}_K - m_K\|^2 \\
&= \|\sum_{k=0}^{K}(\widehat{P}_k - P_k)\psi_k\|^2 = \sum_{k=0}^{K}(\widehat{P}_k - P_k)^2 = \|\vec{\widehat{\mathbf{P}}}_K - \vec{\mathbf{P}}_K\|_2^2 \\
&= \|\widehat{\mathbf{A}}_K^{-1}\vec{\widehat{\mathbf{r}}}_K - \mathbf{A}_K^{-1}\vec{\mathbf{r}}_K\|_2^2 = \|(\widehat{\mathbf{A}}_K^{-1} - \mathbf{A}_K^{-1})\vec{\widehat{\mathbf{r}}}_K + \mathbf{A}_K^{-1}(\vec{\widehat{\mathbf{r}}}_K - \vec{\mathbf{r}}_K)\|_2^2 \\
&\leq 2\|(\widehat{\mathbf{A}}_K^{-1} - \mathbf{A}_K^{-1})\vec{\widehat{\mathbf{r}}}_K\|_2^2 + 2\|(\mathbf{A}_K^{-1}(\vec{\widehat{\mathbf{r}}}_K - \vec{\mathbf{r}}_K)\|_2^2 \\
&\leq 4\|(\widehat{\mathbf{A}}_K^{-1} - \mathbf{A}_K^{-1})(\vec{\widehat{\mathbf{r}}}_K - \vec{\mathbf{r}}_K))\|_2^2 + 4\|(\widehat{\mathbf{A}}_K^{-1} - \mathbf{A}_K^{-1})\vec{\widehat{\mathbf{r}}}_K\|_2^2 + 2\|(\mathbf{A}_K^{-1}(\vec{\widehat{\mathbf{r}}}_K - \vec{\mathbf{r}}_K)\|_2^2 \\
&\leq 4\|(\widehat{\mathbf{A}}_K^{-1} - \mathbf{A}_K^{-1})\|_2^2 \cdot \|\vec{\widehat{\mathbf{r}}}_K - \vec{\mathbf{r}}_K)\|_2^2 + 4\|(\widehat{\mathbf{A}}_K^{-1} - \mathbf{A}_K^{-1})\|_2^2 \cdot \|\vec{\widehat{\mathbf{r}}}_K\|_2^2 + 2\|\mathbf{A}_K^{-1}\|_2^2 \cdot \|\vec{\widehat{\mathbf{r}}}_K - \vec{\mathbf{r}}_K\|_2^2.
\end{aligned} \tag{37}$$

In order to prove Theorem 1, we can study the convergence rates for the three terms on the right-hand side of Equation (37). To obtain the convergence rates $\|\vec{\mathbf{r}}_K\|_2^2$ and $\|\vec{\widehat{\mathbf{r}}}_K - \vec{\mathbf{r}}_K\|_2^2$, we need the following Lemma:

Lemma 4. *Suppose that Condition 2 holds. Then*

$$\|\vec{\mathbf{r}}_K\|_2^2 < \|h\|^2 < \infty.$$

Moreover, if Conditions 1 and 3 hold and $EX^2 < \infty$, we have

$$\|\vec{\widehat{\mathbf{r}}}_K - \vec{\mathbf{r}}_K\|_2^2 < \|h\|^2 <= O_p(KT^{-1}). \tag{38}$$

Proof. First, under Condition 2 we have

$$\|\vec{\mathbf{r}}_K\|_2^2 = \sum_{k=0}^{K} R_k^2 < \sum_{k=0}^{\infty} R_k^2 = \|h\|_2^2 < \infty.$$

Next, we prove Equation (38). We only consider the case $\delta > 0$. Under Condition 1 and $EX^2 < \infty$,

$$\widehat{\rho_1} - \rho_1 = O_p(T^{-\frac{1}{2}}),\ \widehat{\rho_2} - \rho_2 = O_p(T^{-\frac{1}{2}}).$$

Because N_T is Poisson-distributed with intensity λT and is independent from X_j, we have

$$\begin{aligned}
R_k &= \frac{p\lambda}{c}\sum_{i=1}^{2}\frac{\rho_i - \beta}{\prod_{j=1,j\neq i}^{2}(\rho_i - \rho_j)} E\left[\int_0^X \int_u^X e^{-\rho_i(y-u)} w(y, X-y)dy\psi_k(u)du\right]. \\
&= \frac{p\lambda}{c}\sum_{i=1}^{2}\frac{\rho_i - \beta}{\prod_{j=1,j\neq i}^{2}(\rho_i - \rho_j)} E\left[\frac{N_T}{N_T^+}\cdot\frac{T}{N_T}\cdot\frac{1}{T}\cdot\sum_{j=1}^{N_T^+}\int_0^{X_j}\int_u^{X_j} e^{-\rho_i(x-u)} w(x, X_j - x)dx\psi_k(u)du\right] \\
&= \frac{1}{cT}\sum_{i=1}^{2}\frac{\rho_i - \beta}{\prod_{j=1,j\neq i}^{2}(\rho_i - \rho_j)} E\left[\sum_{j=1}^{N_T^+}\int_0^{X_j}\int_u^{X_j} e^{-\rho_i(x-u)} w(x, X_j - x)dx\psi_k(u)du\right].
\end{aligned} \tag{39}$$

Hence,

$$
\begin{aligned}
\widehat{R}_k - R_k =& \frac{1}{cT}\sum_{i=1}^{2}\frac{\widehat{\rho}_i-\widehat{\beta}}{\prod_{j=1,j\neq i}^{2}(\widehat{\rho}_i-\widehat{\rho}_j)}\sum_{j=1}^{N_T^+}\int_0^{X_j}\int_u^{X_j} e^{-\widehat{\rho}_i(x-u)}w(x,X_j-x)dx\psi_k(u)du \\
&- \frac{1}{cT}\sum_{i=1}^{2}\frac{\rho_i-\beta}{\prod_{j=1,j\neq i}^{2}(\rho_i-\rho_j)}E\left[\sum_{j=1}^{N_T^+}\int_0^{X_j}\int_u^{X_j} e^{-\rho_i(x-u)}w(x,X_j-x)dx\psi_k(u)du\right] \\
=& \frac{1}{cT}\sum_{i=1}^{2}\frac{\widehat{\rho}_i-\widehat{\beta}}{\prod_{j=1,j\neq i}^{2}(\widehat{\rho}_i-\widehat{\rho}_j)}\sum_{j=1}^{N_T^+}\int_0^{X_j}\int_u^{X_j} (e^{-\rho_i(x-u)}-e^{-\rho_i(x-u)})w(x,X_j-x)dx\psi_k(u)du \\
&+ \left(\frac{1}{cT}\sum_{i=1}^{2}\frac{\widehat{\rho}_i-\widehat{\beta}}{\prod_{j=1,j\neq i}^{2}(\widehat{\rho}_i-\widehat{\rho}_j)} - \frac{1}{cT}\sum_{i=1}^{2}\frac{\rho_i-\beta}{\prod_{j=1,j\neq i}^{2}(\rho_i-\rho_j)}\right) \\
&\times \sum_{j=1}^{N_T^+}\int_0^{X_j}\int_u^{X_j} e^{-\rho_i(x-u)}w(x,X_j-x)dx\psi_k(u)du \\
&+ \frac{1}{cT}\sum_{i=1}^{2}\frac{\rho_i-\beta}{\prod_{j=1,j\neq i}^{2}(\rho_i-\rho_j)}\left\{\sum_{j=1}^{N_T^+}\int_0^{X_j}\int_u^{X_j} e^{-\rho_i(x-u)}w(x,X_j-x)dx\psi_k(u)du\right. \\
&\left.- E\left[\sum_{j=1}^{N_T^+}\int_0^{X_j}\int_u^{X_j} e^{-\rho_i(x-u)}w(x,X_j-x)dx\psi_k(u)du\right]\right\} \\
:=& I_{k,1}+I_{k,2}+I_{k,3}.
\end{aligned}
\tag{40}
$$

Then, using inequality $(x+y)^2 \le 2x^2+2y^2$, we have

$$
\|\vec{\widehat{\mathbf{r}}}_K - \vec{\mathbf{r}}_K\|_2^2 = \sum_{k=0}^{K}(\widehat{R}_K - R_K)^2 \le 2\sum_{k=0}^{K} I_{k,1}^2 + 4\sum_{k=0}^{K} I_{k,2}^2 + 4\sum_{k=0}^{K} I_{k,3}^2. \tag{41}
$$

By the mean value theorem, it is easy to see that

$$
\begin{aligned}
\left|e^{-\widehat{\rho}_i(x-u)} - e^{-\rho_i(x-u)}\right| &= \left|(\widehat{\rho}_i-\rho_i)(x-u)e^{-\rho_i^*(x-u)}\right| \\
&\le |\widehat{\rho}_i-\rho_i|(x-u),\ i=1,2,
\end{aligned}
\tag{42}
$$

where ρ_1^*, ρ_2^* is a random number between $\widehat{\rho}_i$ and $\rho_i,\ i=1,2$. First, to estimate $\sum_{k=0}^{K} I_{k,1}^2$,

$$
\begin{aligned}
&\sum_{k=0}^{K} I_{k,1}^2 \\
=&\frac{2}{c^2T^2}\sum_{i=1}^{2}\left(\frac{\widehat{\rho}_i-\widehat{\beta}}{\prod_{j=1,j\neq i}^{2}(\widehat{\rho}_i-\widehat{\rho}_j)}\right)^2\sum_{k=0}^{K}\left[\sum_{j=1}^{N_T^+}\int_0^{X_j}\int_u^{X_j}(e^{-\widehat{\rho}_i(x-u)}-e^{-\rho_i(x-u)})w(x,X_j-x)dx\psi_k(u)du\right]^2 \\
=&\frac{2}{c^2T^2}\sum_{i=1}^{2}\left(\frac{\widehat{\rho}_i-\widehat{\beta}}{\prod_{j=1,j\neq i}^{2}(\widehat{\rho}_i-\widehat{\rho}_j)}\right)^2\sum_{k=0}^{K}\left(\int_0^{\infty}\left[\sum_{j=1}^{N_T^+}\mathbf{I}_{(u\le X_j)}\int_u^{X_j}(e^{-\widehat{\rho}_i(x-u)}-e^{-\rho_i(x-u)})w(x,X_j-x)dx\right]\psi_k(u)du\right)^2 \\
\le&\frac{2K}{c^2T^2}\sum_{i=1}^{2}\left(\frac{\widehat{\rho}_i-\widehat{\beta}}{\prod_{j=1,j\neq i}^{2}(\widehat{\rho}_i-\widehat{\rho}_j)}\right)^2\int_0^{\infty}\left[\sum_{j=1}^{N_T^+}\mathbf{I}_{(u\le X_j)}\int_u^{X_j}(e^{-\widehat{\rho}_i(x-u)}-e^{-\rho_i(x-u)})w(x,X_j-x)dx\right]^2du \\
\le&\frac{2K}{c^2T^2}\sum_{i=1}^{2}\left(\frac{\widehat{\rho}_i-\widehat{\beta}}{\prod_{j=1,j\neq i}^{2}(\widehat{\rho}_i-\widehat{\rho}_j)}\right)^2(\widehat{\rho}_i-\rho_i)^2N_T^+\int_0^{\infty}\sum_{j=1}^{N_T^+}\left[\mathbf{I}_{(u\le X_j)}\int_u^{X_j}(x-u)w(x,X_j-x)dx\right]^2du \\
=&\frac{2K\widehat{\lambda}\widehat{p}}{c^2}\sum_{i=1}^{2}\left(\frac{\widehat{\rho}_i-\widehat{\beta}}{\prod_{j=1,j\neq i}^{2}(\widehat{\rho}_i-\widehat{\rho}_j)}\right)^2(\widehat{\rho}_i-\rho_i)^2\frac{1}{T}\sum_{j=1}^{N_T^+}\int_0^{X_j}\left[\int_u^{X_j}(x-u)w(x,X_j-x)dx\right]^2du.
\end{aligned}
\tag{43}
$$

It follows from Condition 3 and Markov's inequality that

$$
\frac{1}{T}\sum_{j=1}^{N_T^+}\int_0^{X_j}\left[\int_u^{X_j}(x-u)w(x,X_j-x)dx\right]^2du=O_p(1),
$$

hence

$$
\sum_{k=0}^{K} I_{k,1}^2=O_p(KT^{-1}). \tag{44}
$$

As for $\sum_{k=0}^{K} I_{k,2}^2$, we can obtain

$$
\begin{aligned}
&\sum_{k=0}^{K} I_{k,2}^2 \\
&= \sum_{k=0}^{K}\left[\left(\frac{1}{cT}\sum_{i=1}^{2}\frac{\widehat{\rho}_i-\widehat{\beta}}{\prod_{j=1,j\neq i}^{2}(\widehat{\rho}_i-\widehat{\rho}_j)}-\frac{1}{cT}\sum_{i=1}^{2}\frac{\rho_i-\beta}{\prod_{j=1,j\neq i}^{2}(\rho_i-\rho_j)}\right)\sum_{j=1}^{N_T^+}\int_0^{X_j}\int_u^{X_j} e^{-\rho_i(x-u)}w(x,X_j-x)dx\psi_k(u)du\right]^2 \\
&\le 2\sum_{k=0}^{K}\sum_{i=1}^{2}\left(\frac{1}{cT}\frac{\widehat{\rho}_i-\widehat{\beta}}{\prod_{j=1,j\neq i}^{2}(\widehat{\rho}_i-\widehat{\rho}_j)}-\frac{1}{cT}\frac{\rho_i-\beta}{\prod_{j=1,j\neq i}^{2}(\rho_i-\rho_j)}\right)^2\left(\sum_{j=1}^{N_T^+}\int_0^{X_j}\int_u^{X_j} e^{-\rho_i(x-u)}w(x,X_j-x)dx\psi_k(u)du\right)^2 \\
&\le \frac{2}{c^2T^2}\sum_{i=1}^{2}\left(\frac{\widehat{\rho}_i-\widehat{\beta}}{\prod_{j=1,j\neq i}^{2}(\widehat{\rho}_i-\widehat{\rho}_j)}-\frac{\rho_i-\beta}{\prod_{j=1,j\neq i}^{2}(\rho_i-\rho_j)}\right)^2\sum_{k=0}^{K}\left[\int_0^{\infty}\left(\sum_{j=1}^{N_T^+}\mathbf{I}_{(u\le X_j)}\int_u^{X_j}w(x,X-x)dx\right)\psi_k(u)du\right]^2 \\
&\le \frac{2K}{c^2T^2}\sum_{i=1}^{2}\left(\frac{\widehat{\rho}_i-\widehat{\beta}}{\prod_{j=1,j\neq i}^{2}(\widehat{\rho}_i-\widehat{\rho}_j)}-\frac{\rho_i-\beta}{\prod_{j=1,j\neq i}^{2}(\rho_i-\rho_j)}\right)^2\int_0^{\infty}\left[\sum_{j=1}^{N_T^+}\mathbf{I}_{(u\le X_j)}\int_u^{X_j}w(x,X-x)dx\right]^2du \\
&\le \frac{2K\widehat{\lambda}\widehat{p}}{c^2}\sum_{i=1}^{2}\left(\frac{\widehat{\rho}_i-\widehat{\beta}}{\prod_{j=1,j\neq i}^{2}(\widehat{\rho}_i-\widehat{\rho}_j)}-\frac{\rho_i-\beta}{\prod_{j=1,j\neq i}^{2}(\rho_i-\rho_j)}\right)^2\frac{1}{T}\sum_{j=1}^{N_T^+}\int_0^{X_j}\left(\int_u^{X_j}w(x,X-x)dx\right)^2du.
\end{aligned}
\tag{45}
$$

According to $\widehat{\beta}-\beta=O_p(T^{-\frac{1}{2}})$, $\widehat{\rho_1}-\rho_1=O_p(T^{-\frac{1}{2}})$ and $\widehat{\rho_2}-\rho_2=O_p(T^{-\frac{1}{2}})$. Then

$$
\left(\frac{\widehat{\rho}_i-\widehat{\beta}}{\prod_{j=1,j\neq i}^{2}(\widehat{\rho}_i-\widehat{\rho}_j)}-\frac{\rho_i-\beta}{\prod_{j=1,j\neq i}^{2}(\rho_i-\rho_j)}\right)^2=O_p(T^{-1}). \tag{46}
$$

Hence,

$$
\sum_{k=0}^{K} I_{k,2}^2=O_p(KT^{-1}). \tag{47}
$$

For the summation $\sum_{k=0}^{K} I_{k,3}^2$, we have

$$
\begin{aligned}
&E\left[\sum_{k=0}^{K} I_{k,3}^2\right] = \sum_{k=0}^{K} E[I_{k,3}^2] \\
&= \sum_{k=0}^{K} \frac{2}{c^2T^2} \sum_{i=1}^{2} \left(\frac{\rho_i - \beta}{\prod_{j=1, j\neq i}^{2} (\rho_i - \rho_j)} \right)^2 Var\left\{ \sum_{j=1}^{N_T^+} \int_0^{X_j} \int_u^{X_j} e^{-\rho_i(x-u)} w(x, X_j - x) dx \psi_k(u) du \right\} \\
&\leq \frac{2p\lambda}{c^2T} \sum_{i=1}^{2} \left(\frac{\rho_i - \beta}{\prod_{j=1, j\neq i}^{2} (\rho_i - \rho_j)} \right)^2 \sum_{k=0}^{K} E\left[\int_0^{X} \int_u^{X} e^{-\rho_i(x-u)} w(x, X_j - x) dx \psi_k(u) du \right]^2 \\
&= \frac{2p\lambda}{c^2T} \sum_{i=1}^{2} \left(\frac{\rho_i - \beta}{\prod_{j=1, j\neq i}^{2} (\rho_i - \rho_j)} \right)^2 E\left\{ \sum_{k=0}^{K} \left[\int_0^{\infty} \left(\mathbf{I}_{(u\leq X)} \int_u^{X} e^{-\rho_i(x-u)} w(x, X - x) dx \right) \psi_k(u) du \right]^2 \right\} \\
&\leq \frac{2Kp\lambda}{c^2T} \sum_{i=1}^{2} \left(\frac{\rho_i - \beta}{\prod_{j=1, j\neq i}^{2} (\rho_i - \rho_j)} \right)^2 E\left[\int_0^{\infty} \left(\mathbf{I}_{(u\leq X)} \int_u^{X} e^{-\rho_i(x-u)} w(x, X - x) dx \right)^2 du \right] \\
&= \frac{2Kp\lambda}{c^2T} \sum_{i=1}^{2} \left(\frac{\rho_i - \beta}{\prod_{j=1, j\neq i}^{2} (\rho_i - \rho_j)} \right)^2 E\left[\int_0^{X} \left(\int_u^{X} e^{-\rho_i(x-u)} w(x, X - x) dx \right)^2 du \right] \\
&\leq \frac{2Kp\lambda}{c^2T} \sum_{i=1}^{2} \left(\frac{\rho_i - \beta}{\prod_{j=1, j\neq i}^{2} (\rho_i - \rho_j)} \right)^2 E\left[\int_0^{X} \left(\int_u^{X} w(x, X - x) dx \right)^2 du \right],
\end{aligned}
\tag{48}
$$

which, together with Condition 3 and Markov's inequality, yields

$$
\sum_{k=0}^{K} I_{k,3}^2 = O_p(KT^{-1}). \tag{49}
$$

Finally, we complete the proof. □

In order to obtain the convergence rates of $\|\widehat{\mathbf{A}}_K - \mathbf{A}_K\|_F^2$, $\|\mathbf{A}_K^{-1}\|_2$ and $\|\widehat{\mathbf{A}}_K^{-1} - \mathbf{A}_K^{-1}\|_2$, we have the following propositions:

Proposition 1. *Let Condition 1 hold and $EX^2 < \infty$. Then*

$$
\|\widehat{\mathbf{A}}_K - \mathbf{A}_K\|_F^2 = O_p(K^2T^{-1}). \tag{50}
$$

Proposition 2. *Suppose that Condition 1 holds. Then for all $K \geq 1$,*

$$
\|\mathbf{A}_K^{-1}\|_2 \leq \frac{2c}{c - \lambda(p\mu_+ - q\mu_-)}. \tag{51}
$$

Proposition 3. *Let Condition 1 hold and $EX^2 < \infty$. If $K = o(T^{\frac{1}{2}})$, then*

$$
\|\widehat{\mathbf{A}}_K^{-1} - \mathbf{A}_K^{-1}\|_2 = O_p(KT^{-\frac{1}{2}}). \tag{52}
$$

In the rest of the section, we give the proof of Propositions 1, 2 and 3 and Theorem 1.

Proof of Proposition 1. Using the definitions of $\widehat{\mathbf{A}}_K$ and $\mathbf{A}_K$, then

$$\begin{aligned}\|\widehat{\mathbf{A}}_K - \mathbf{A}_K\|_F^2 &= \sum_{k=0}^{K}\sum_{j=0}^{k}\left\{\frac{1}{\sqrt{2}}(Q_{k-j} - \widehat{Q}_{k-j}) - \frac{1}{\sqrt{2}}(Q_{k-j-1} - \widehat{Q}_{k-j-1})\right\}^2 \\ &= \frac{1}{2}\sum_{k=0}^{K}\sum_{j=0}^{k}\left\{(Q_{k-j} - \widehat{Q}_{k-j}) - (Q_{k-j-1} - \widehat{Q}_{k-j-1})\right\}^2 \\ &\leq \sum_{k=0}^{K}\sum_{j=0}^{k}\left\{(Q_{k-j} - \widehat{Q}_{k-j})^2 - (Q_{k-j-1} - \widehat{Q}_{k-j-1})^2\right\}, \end{aligned} \tag{53}$$

where we have put $Q_{-1} = \widehat{Q}_{-1} = 0$ for convenience.

Because N_T is Poisson-distributed with intensity λT and is independent from X_j, we have

$$\begin{aligned} Q_k &= \frac{p\lambda}{c}\sum_{i=1}^{2}\frac{\rho_i - \beta}{\prod_{j=1,j\neq i}^{2}(\rho_i - \rho_j)} E\left[\int_0^X e^{-\rho_i(X-x)}\psi_k(x)dx\right] \\ &= \frac{p\lambda}{c}\sum_{i=1}^{2}\frac{\rho_i - \beta}{\prod_{j=1,j\neq i}^{2}(\rho_i - \rho_j)} E\left\{\frac{N_T}{N_T^+}\cdot\frac{T}{N_T}\cdot\frac{1}{T}\cdot\sum_{j=1}^{N_T^+}\int_0^{X_j} e^{-\rho_i(X_j-x)}\psi_k(x)dx\right\} \\ &= \frac{1}{cT}\sum_{i=1}^{2}\frac{\rho_i - \beta}{\prod_{j=1,j\neq i}^{2}(\rho_i - \rho_j)} E\left\{\sum_{j=1}^{N_T^+}\int_0^{X_j} e^{-\rho_i(X_j-x)}\psi_k(x)dx\right\}. \end{aligned}$$

Then

$$\begin{aligned} \widehat{Q}_k - Q_k =& \frac{1}{cT}\sum_{i=1}^{2}\frac{\widehat{\rho}_i - \widehat{\beta}}{\prod_{j=1,j\neq i}^{2}(\widehat{\rho}_i - \widehat{\rho}_j)}\left(\sum_{j=1}^{N_T^+}\int_0^{X_j} e^{-\widehat{\rho}_i(X_j-x)}\psi_k(x)dx\right) \\ &- \frac{1}{cT}\sum_{i=1}^{2}\frac{\rho_i - \beta}{\prod_{j=1,j\neq i}^{2}(\rho_i - \rho_j)} E\left\{\sum_{j=1}^{N_T^+}\int_0^{X_j} e^{-\rho_i(X_j-x)}\psi_k(x)dx\right\} \\ =& \frac{1}{cT}\sum_{i=1}^{2}\frac{\widehat{\rho}_i - \widehat{\beta}}{\prod_{j=1,j\neq i}^{2}(\widehat{\rho}_i - \widehat{\rho}_j)}\left[\sum_{j=1}^{N_T^+}\int_0^{X_j}\left(e^{-\widehat{\rho}_i(X_j-x)} - e^{-\rho_i(X_j-x)}\psi_k(x)dx\right)\right] \\ &+ \frac{1}{cT}\sum_{i=1}^{2}\left[\frac{\widehat{\rho}_i - \widehat{\beta}}{\prod_{j=1,j\neq i}^{2}(\widehat{\rho}_i - \widehat{\rho}_j)} - \frac{\rho_i - \beta}{\prod_{j=1,j\neq i}^{2}(\rho_i - \rho_j)}\right]\left(\sum_{j=1}^{N_T^+}\int_0^{X_j} e^{-\widehat{\rho}_i(X_j-x)}\psi_k(x)dx\right) \\ &+ \frac{1}{cT}\sum_{i=1}^{2}\frac{\rho_i - \beta}{\prod_{j=1,j\neq i}^{2}(\rho_i - \rho_j)}\left\{\sum_{j=1}^{N_T^+}\int_0^{X_j} e^{-\widehat{\rho}_i(X_j-x)}\psi_k(x)dx - E\left[\sum_{j=1}^{N_T^+}\int_0^{X_j} e^{-\rho_i(X_j-x)}\psi_k(x)dx\right]\right\} \\ :=& II_{k,1} + II_{k,2} + II_{k,3}. \end{aligned} \tag{54}$$

Plugging the above result into Equation (48), we obtain

$$\begin{aligned}\|\widehat{\mathbf{A}}_K - \mathbf{A}_K\|_F^2 &\le \sum_{k=0}^{K}\sum_{j=0}^{k}\Big\{(II_{k-j,1}+II_{k-j,2}+II_{k-j,3})^2+(II_{k-j-1,1}+II_{k-j-1,2}+II_{k-j-1,3})^2\Big\}\\ &\le 4\sum_{k=0}^{K}\sum_{j=0}^{k}\Big[II_{k-j,1}+II_{k-j,2}+II_{k-j,3}+II_{k-j-1,1}+II_{k-j-1,2}+II_{k-j-1,3}\Big]\\ &\le 8K\sum_{k=0}^{K}\Big[II_{k,1}^2+II_{k,2}^2+II_{k,3}^2\Big].\end{aligned} \tag{55}$$

First, for $K\sum_{k=0}^{K} II_{k,1}^2$, we can obtain

$$\begin{aligned}K\sum_{k=0}^{K} II_{k,1}^2 &= \sum_{k=0}^{K}\frac{2K}{c^2T^2}\sum_{i=1}^{2}\left(\frac{\widehat{\rho}_i-\widehat{\beta}}{\prod_{j=1,j\neq i}^{2}(\widehat{\rho}_i-\widehat{\rho}_j)}\right)^2\left[\sum_{j=1}^{N_T^+}\int_0^{X_j}\left(e^{-\widehat{\rho}_i(X_j-x)}-e^{-\rho_i(X_j-x)}\right)\psi_k(x)dx\right]^2\\ &\le \frac{2K^2}{c^2T^2}\sum_{i=1}^{2}\left(\frac{\widehat{\rho}_i-\widehat{\beta}}{\prod_{j=1,j\neq i}^{2}(\widehat{\rho}_i-\widehat{\rho}_j)}\right)^2\int_0^{\infty}\left[\sum_{j=1}^{N_T^+}\mathbf{I}_{(x\le X_j)}\left(e^{-\widehat{\rho}_i(X_j-x)}-e^{-\rho_i(X_j-x)}\right)\right]^2\\ &\le \frac{2K^2}{c^2T^2}\sum_{i=1}^{2}\left(\frac{\widehat{\rho}_i-\widehat{\beta}}{\prod_{j=1,j\neq i}^{2}(\widehat{\rho}_i-\widehat{\rho}_j)}\right)^2(\widehat{\rho}_i-\rho_i)^2N_T^+\int_0^{\infty}\sum_{j=1}^{N_T^+}\mathbf{I}_{(x\le X_j)}\left|(X_j-x)e^{-\rho_i^*(X_j-x)}\right|^2dx\\ &\le \frac{2K^2\widehat{\lambda}\widehat{p}}{c^2}\sum_{i=1}^{2}\left(\frac{\widehat{\rho}_i-\widehat{\beta}}{\prod_{j=1,j\neq i}^{2}(\widehat{\rho}_i-\widehat{\rho}_j)}\right)^2\frac{(\widehat{\rho}_i-\rho_i)^2}{[\min\{\rho_i,\widehat{\rho}_i\}]^2}\cdot\frac{1}{T}\sum_{j=1}^{N_T^+}X_j.\end{aligned} \tag{56}$$

It follows from $E\left[\frac{1}{T}\sum_{j=1}^{N_T^+}X_j\right]=p\lambda\mu_x<\infty$ and Markov's inequality that $\frac{1}{T}\sum_{j=1}^{N_T^+}X_j=O_p(1)$. Then

$$K\sum_{k=0}^{K} II_{k,1}^2 = O_p(K^2T^{-1}). \tag{57}$$

Next, to compute $II_{k,2}$, we can obtain

$$
\begin{aligned}
K\sum_{k=0}^{K} II_{k,2}^2 &= 2K\sum_{k=0}^{K}\sum_{i=1}^{2}\left[\frac{1}{cT}\frac{\widehat{\rho}_i-\widehat{\beta}}{\prod_{j=1,j\neq i}^{2}(\widehat{\rho}_i-\widehat{\rho}_j)}-\frac{1}{cT}\frac{\rho_i-\beta}{\prod_{j=1,j\neq i}^{2}(\rho_i-\rho_j)}\right]^2\left(\sum_{j=1}^{N_T^+}\int_0^{X_j}e^{-\rho_i(X_j-x)}\psi_k(x)dx\right)^2\\
&= \frac{2K}{c^2T^2}\sum_{k=0}^{K}\sum_{i=1}^{2}\left[\frac{\widehat{\rho}_i-\widehat{\beta}}{\prod_{j=1,j\neq i}^{2}(\widehat{\rho}_i-\widehat{\rho}_j)}-\frac{\rho_i-\beta}{\prod_{j=1,j\neq i}^{2}(\rho_i-\rho_j)}\right]^2\left[\int_0^{\infty}\sum_{j=1}^{N_T^+}\mathbf{I}_{(x\le X_j)}e^{-\rho_i(X_j-x)}\psi_k(x)dx\right]^2\\
&\le \frac{2K^2}{c^2T^2}\sum_{i=1}^{2}\left[\frac{\widehat{\rho}_i-\widehat{\beta}}{\prod_{j=1,j\neq i}^{2}(\widehat{\rho}_i-\widehat{\rho}_j)}-\frac{\rho_i-\beta}{\prod_{j=1,j\neq i}^{2}(\rho_i-\rho_j)}\right]^2\int_0^{\infty}\left(\sum_{j=1}^{N_T^+}\mathbf{I}_{(x\le X_j)}\right)^2dx\\
&\le \frac{2K^2\widehat{\lambda}\widehat{p}}{c^2}\sum_{i=1}^{2}\left[\frac{\widehat{\rho}_i-\widehat{\beta}}{\prod_{j=1,j\neq i}^{2}(\widehat{\rho}_i-\widehat{\rho}_j)}-\frac{\rho_i-\beta}{\prod_{j=1,j\neq i}^{2}(\rho_i-\rho_j)}\right]^2\frac{1}{T}\sum_{j=1}^{N_T^+}X_j.
\end{aligned}
\tag{58}
$$

Then,

$$
K\sum_{k=0}^{K} II_{k,2}^2 = O_p(K^2T^{-1}),
\tag{59}
$$

due to $\widehat{\beta}-\beta = O_p(T^{-\frac{1}{2}})$, $\widehat{\rho}_1-\rho_1 = O_p(T^{-\frac{1}{2}})$ and $\widehat{\rho}_2-\rho_2 = O_p(T^{-\frac{1}{2}})$.

As for $II_{k,3}$, taking expectation, we have

$$
\begin{aligned}
E\left[K\sum_{k=0}^{K} II_{k,3}^2\right] &= K\sum_{k=0}^{K}E\left[II_{k,3}^2\right] = \sum_{k=0}^{K}\frac{2K}{c^2T^2}\sum_{i=1}^{2}\left(\frac{\rho_i-\beta}{\prod_{j=1,j\neq i}^{2}(\rho_i-\rho_j)}\right)^2\\
&\quad\times Var\left\{\sum_{j=1}^{N_T^+}\int_0^{X_j}e^{-\rho_i(X_j-x)}\psi_k(x)dx\right\}\\
&= \frac{2\lambda pK}{c^2T}\sum_{k=0}^{K}\sum_{i=1}^{2}\left(\frac{\rho_i-\beta}{\prod_{j=1,j\neq i}^{2}(\rho_i-\rho_j)}\right)^2E\left[\left(\int_0^{X}e^{-\rho_i(X-x)}\psi_k(x)dx\right)^2\right]\\
&\le \frac{2\lambda pK^2}{c^2T}\sum_{i=1}^{2}\left(\frac{\rho_i-\beta}{\prod_{j=1,j\neq i}^{2}(\rho_i-\rho_j)}\right)^2E\left[\int_0^{X}e^{-2\rho_i(X-x)}dx\right]\\
&\le \frac{2\lambda pK^2}{c^2T}\sum_{i=1}^{2}\left(\frac{\rho_i-\beta}{\prod_{j=1,j\neq i}^{2}(\rho_i-\rho_j)}\right)^2E(X).
\end{aligned}
\tag{60}
$$

Due to $\mu_X < \infty$ and Markov's inequality, hence

$$K \sum_{k=0}^{K} II_{k,3}^2 = O_p(K^2 T^{-1}). \tag{61}$$

Finally, substituting Equations (57), (59) and (61) into Equation (55) yields the convergence rate. □

Proof of Proposition 2. First, let $h = \frac{\lambda}{c} \sum_{i=1}^{2} \frac{(\rho_i - \beta)}{\prod_{j=1, j\neq i}^{2} (\rho_i - \rho_j)} \int_x^\infty e^{-\rho_i (y-x)} (p f_+(y) - q f_-(y)) dy$ and define a sequence $\{c_k\}_{k=0}^{\infty}$ by

$$c_0 = 1 - \frac{1}{\sqrt{2}} C_0, \quad c_k = \frac{1}{\sqrt{2}} (C_{k-1} - C_k), \quad k = 1, 2, \cdots$$

where

$$C_k = \langle h, \psi_k \rangle \quad \text{for} \quad k = 0, 1, 2 \cdots$$

such that $\mathbf{C}$ is an infinite lower triangular Toeplitz matrix generated by $\{c_k\}$ similar to $\mathbf{A}$

$$\mathbf{C} = \begin{pmatrix} 1 - \frac{1}{\sqrt{2}} C_0 & 0 & 0 & \dots \\ \frac{1}{\sqrt{2}} (C_0 - C_1) & 1 - \frac{1}{\sqrt{2}} C_0 & 0 & \dots \\ \frac{1}{\sqrt{2}} (C_1 - C_2) & \frac{1}{\sqrt{2}} (C_0 - C_1) & 1 - \frac{1}{\sqrt{2}} C_0 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

It is easy to see that

$$\|\mathbf{A}_K^{-1}\|_2 \leq \|\mathbf{C}_K^{-1}\|_2,$$

where $\mathbf{C}_K = (c_{ij})_{i,j=1}^{K+1}$.

By Lemma 4.3 in Zhang and Su [42], we know that $c_k, k \geq 0$ are Fourier coefficients of the function

$$c(e^{i\theta}) = \sum_{k=0}^{\infty} c_k e^{i\theta k} = 1 - \mathcal{L}h\left(\frac{1 + i\theta}{1 - i\theta}\right), \quad \theta \in \mathbb{R}.$$

Let $\zeta = \{z \in \mathbb{C} : |z| = 1\}$ denote the complex unite circle. We have

$$\begin{aligned} \inf_{z \in \zeta} |c(z)| &= \inf_{z \in \zeta} \left| 1 - \mathcal{L}h\left(\frac{1 + i\theta}{1 - i\theta}\right) \right| \geq 1 - \sup_{z \in \zeta} \left| \mathcal{L}h\left(\frac{1 + i\theta}{1 - i\theta}\right) \right| \\ &\geq 1 - \int_0^\infty h(x) dx = 1 - \frac{\lambda}{c} \sum_{i=1}^{2} \frac{(\rho_i - \beta)}{\prod_{j=1, j\neq i}^{2} (\rho_i - \rho_j)} \int_0^\infty \int_x^\infty e^{-\rho_i (y-x)} (p f_+(y) - q f_-(y)) dy dx \\ &\geq 1 - \frac{\lambda}{c} \int_0^\infty \int_x^\infty p f_+(y) - q f_-(y) dy dx = 1 - \frac{\lambda p \mu_+}{c} + \frac{\lambda q \mu_-}{c} > 0, \end{aligned}$$

by Condition 1. Then, by Lemma 3.8 in the work of Böttcher and Grudsky [60], we obtain

$$\|\mathbf{A}_K^{-1}\|_2 \leq \|\mathbf{C}_K^{-1}\|_2 \leq \frac{2}{1 - \frac{\lambda p \mu_+}{c} + \frac{\lambda q \mu_-}{c}} = \frac{2c}{c - \lambda(p\mu_+ - q\mu_-)}.$$

The proof is completed. □

Proof of Proposition 3. Note that $\widehat{\mathbf{A}}_K = \mathbf{A}_K + \widehat{\mathbf{A}}_K - \mathbf{A}_K$ and $\mathbf{A}_K$ is invertible. By Propositions 1 and 2,

$$\begin{aligned}\|\mathbf{A}_K^{-1} \cdot (\widehat{\mathbf{A}}_K - \mathbf{A}_K)\|_2 &\leq \|\mathbf{A}_K^{-1}\|_2 \cdot \|\widehat{\mathbf{A}}_K - \mathbf{A}_K\|_2 \leq \frac{2c}{c - \lambda(p\mu_+ - q\mu_-)} \cdot \|\widehat{\mathbf{A}}_K - \mathbf{A}_K\|_2 \\ &\leq \frac{2c}{c - \lambda(p\mu_+ - q\mu_-)} \cdot \|\widehat{\mathbf{A}}_K - \mathbf{A}_K\|_F = O_p(KT^{-\frac{1}{2}}) = o_p(1).\end{aligned}$$

Then, by the result of Theorem 2.5 of Stewart and Sun [61], we have

$$\begin{aligned}\|\widehat{\mathbf{A}}_K^{-1} - \mathbf{A}_K^{-1}\|_2 &\leq \frac{\|\widehat{\mathbf{A}}_K - \mathbf{A}_K\|_2 \cdot \|\mathbf{A}_K^{-1}\|_2^2}{1 - \|\mathbf{A}_K^{-1} \cdot (\widehat{\mathbf{A}}_K - \mathbf{A}_K)\|_2} \\ &\leq \left(\frac{2c}{c - \lambda(p\mu_+ - q\mu_-)}\right)^2 \frac{\|\widehat{\mathbf{A}}_K - \mathbf{A}_K\|_F}{1 - \|\mathbf{A}_K^{-1} \cdot (\widehat{\mathbf{A}}_K - \mathbf{A}_K)\|_2} \\ &= O_p(KT^{-\frac{1}{2}}). \end{aligned} \tag{62}$$

This completes the proof. □

Finally, by the three terms of (37), Lemma 4, and Propositions 1–3, the proof of Theorem 1 is as follows:

Proof of Theorem 1. By Lemma 4 and Propositions 1–3, we have

$$\|\widehat{\mathbf{A}}_K^{-1} - \mathbf{A}_K^{-1}\|_2^2 \cdot \|\vec{\widehat{\mathbf{r}}}_K - \vec{\mathbf{r}}_K\|_2^2 = O_p(K^3T^{-2}),$$

$$\|\widehat{\mathbf{A}}_K^{-1} - \mathbf{A}_K^{-1}\|_2^2 \cdot \|\vec{\mathbf{r}}_K\|_2^2 = O_p(K^2T^{-1}),$$

$$\|\mathbf{A}_K^{-1}\|_2^2 \cdot \|\vec{\widehat{\mathbf{r}}}_K - \vec{\mathbf{r}}_K\|_2^2 = O_p(KT^{-1}).$$

Then,

$$\|\widehat{m}_K - m_K\|^2 = O_p(K^3T^{-2}) + O_p(K^2T^{-1}) + O_p(KT^{-1}) = O_p(K^2T^{-1}) \tag{63}$$

under condition $K = o(T^{\frac{1}{2}})$. Furthermore, if $m \in W(\mathbb{R}_+, r, B)$, Equation (35) follows from Equation (34). □

Remark 4. *Suppose the conditions in Theorem 1. Then, by Equations (33) and (63), we have*

$$\|\widehat{m}_K - m\|^2 = O(K^{-r}) + O_p(K^2T^{-1}).$$

We can minimize the error bound $O(K^{-r}) + O_p(K^2T^{-1})$ to find the optimal truncation parameter, say m_{op}, has $O_p(T^{-\frac{1}{2}})$. (See Zhang and Su [43] and Su et al. [49].)

5. Numerical Illustration

In this section, we provide some numerical examples to show the performance of our estimator when the observed sample size is finite. Throughout this section, we set $c = 1.5$, $\lambda = 2$, $\beta = 1$, $p = 0.5$, and $q = 0.5$, and we consider the following three claim density functions at the same time:

(1) Exponential density function: $f_+(x) = e^{-x}$, $x > 0$.
(2) Erlang (2) density function: $f_+(x) = 4xe^{-2x}$, $x > 0$.
(3) Combination-of-exponentials density function: $f_+(x) = 3e^{-1.5x} - 3e^{-3x}$, $x > 0$.

As in Zhang [24], we estimate the following four classes of Gerber–Shiu functions:

(1) Ruin probability (RP): $w(x,y) \equiv 1$, $\delta = 0$.
(2) Laplace transform of ruin time (LT): $w(x,y) \equiv 1$, $\delta = 0.1$.
(3) Expected claim size causing ruin (ECS): $w(x,y) \equiv x+y$, $\delta = 0$.
(4) Expected deficit at ruin (ED): $w(x,y) \equiv y$, $\delta = 0$.

Note that the assumptions of the above three claim density functions all satisfy $\mu_1 = 1$, and through Equations (9) and (11), we can easily obtain the explicit formulae for the above Gerber–Shiu functions by Laplace inversion. For exponential claim density function, the explicit formulae for these Gerber–Shiu functions are given by

(1) $m(u) = 0.46482\mathrm{e}^{-0.53518u}, \quad u \geq 0.$
(2) $m(u) = 0.43217\mathrm{e}^{-0.56783u}, \quad u \geq 0.$
(3) $m(u) = 1.8593\mathrm{e}^{-0.53518u} - \mathrm{e}^{-u}, \ u \geq 0.$
(4) $m(u) = 0.46482\mathrm{e}^{-0.53518u}, \quad u \geq 0.$

For Erlang (2) claim size, the explicit formulae for these Gerber–Shiu functions are given by

(1) $m(u) = 0.53387\mathrm{e}^{-0.747u} - 0.06037\mathrm{e}^{-2.819u}, \quad u \geq 0.$
(2) $m(u) = 0.50866\mathrm{e}^{-0.764u} - 0.06266\mathrm{e}^{-2.816u}, \quad u \geq 0.$
(3) $m(u) = 1.2575\mathrm{e}^{-0.747u} + 0.41249\mathrm{e}^{-2.819u} - \mathrm{e}^{-2u}, \ u \geq 0.$
(4) $m(u) = 0.02046\mathrm{e}^{-2.819u} + 0.34454\mathrm{e}^{-0.747u}, \quad u \geq 0.$

For combination-of-exponential claim size, the explicit formulae for these Gerber- Shiu functions are given by

(1) $m(u) = 0.49493\mathrm{e}^{-0.707u} - 0.04003\mathrm{e}^{-30357u}, \quad u \geq 0.$
(2) $m(u) = 0.48606\mathrm{e}^{-0.746u} - 0.04136\mathrm{e}^{-3.352u}, \quad u \geq 0.$
(3) $m(u) = 1.3522\mathrm{e}^{-0.707u} - 0.6667\mathrm{e}^{-1.5u} - 0.3333\mathrm{e}^{-3u} + 0.33879\mathrm{e}^{-3.357u}, \ u \geq 0.$
(4) $m(u) = 0.35945\mathrm{e}^{-0.707u} + 0.01625\mathrm{e}^{-0.3.357u}, \quad u \geq 0.$

Here, we consider $T = 120,\ 180,\ 360$. For the cut-off parameter K, we use the result of Remark 4.1 in Su et al. [49] with $K = \lfloor 5T^{\frac{1}{10}} \rfloor$, where $\lfloor \cdot \rfloor$ means the integer part. Through simulation, we find that even if the truncation parameter K is very small, the satisfactory effect can be obtained. In the case of finite sample size, to test the performance of the estimator, we consider mean value, mean relative error, and integrated mean square error (IMSE) based on 300 experiments, which are computed by

$$\frac{1}{300}\sum_{j=1}^{300}\widehat{m}_{K,j}(u),\ \frac{1}{300}\sum_{j=1}^{300}\frac{\widehat{m}_{K,j}(u)}{m(u)} - 1,\ \frac{1}{300}\sum_{j=1}^{300}\int_0^{30}|\widehat{m}_{K,j}(u) - m(u)|^2 du,$$

where $\widehat{m}_{K,j}(u)$ is the estimate of Gerber–Shiu function in the j-th experiment. For IMSE, we computed the integral on the finite domain $[0, 30]$ instead of $[0, \infty]$, since when u is large, both the true value and the estimates are very close to zero.

For Figure 1, we consider the comparison between the 30 estimated curves and the value curves when $T = 180$ and the exponential claim size density. It is easily observed that the estimated curves are close to each other and close to the true value curve, which indicates that our estimation method has good stability. Next, in Figures 2 and 3, based on 300 experiments, we respectively show the mean value curves and true value curves of the exponential claim size density and the combination-of-exponentials claim size density at different observed intervals T. It is easy to see from the figure above that it is difficult to distinguish the true value curves from the mean value curves when T is larger.

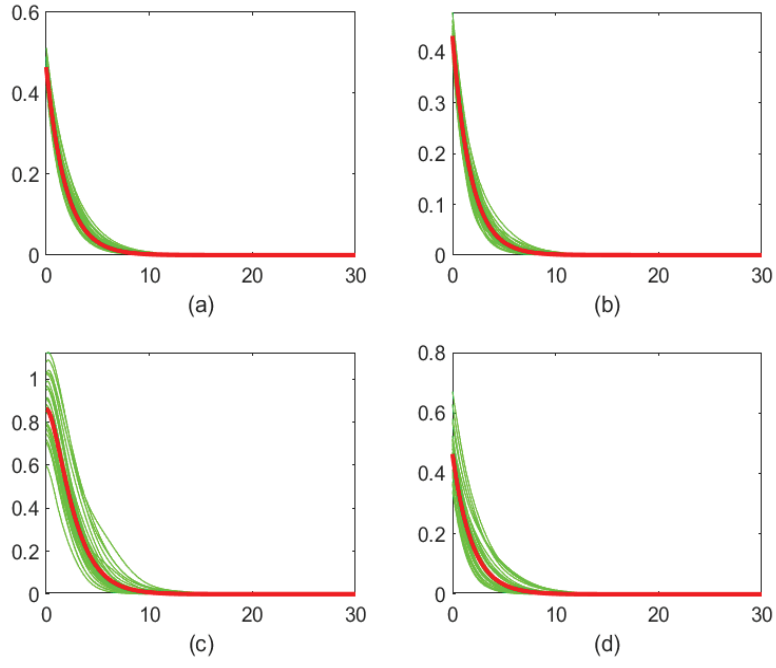

Figure 1. Estimation of the Gerber–Shiu function for exponential density function: red line (true values) and green lines (30 estimated curves) when $T = 180$. (**a**) Ruin probability; (**b**) Laplace transform of ruin time; (**c**) expected claim size causing ruin; (**d**) expected deficit at ruin.

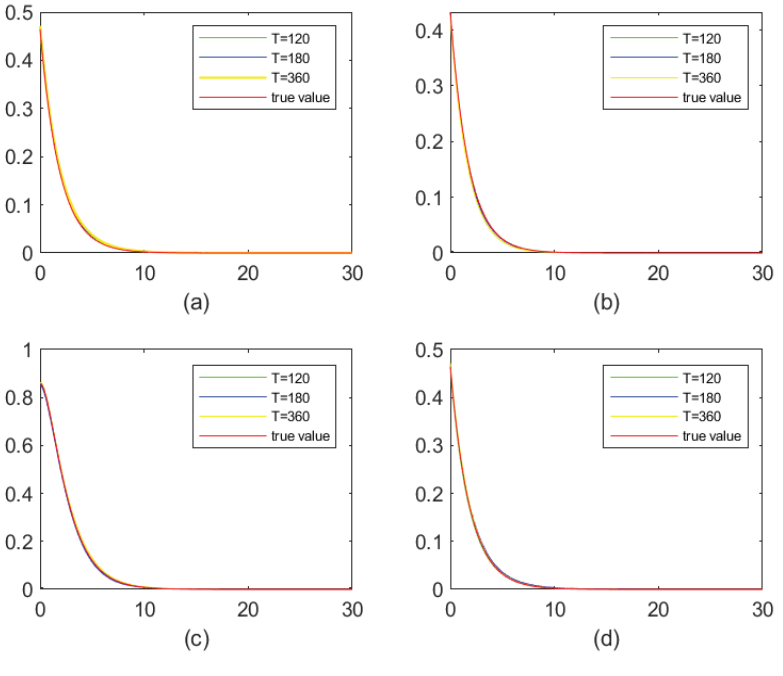

Figure 2. Estimation of the Gerber–Shiu function for exponential density function: mean curves. (**a**) Ruin probability; (**b**) Laplace transform of ruin time; (**c**) expected claim size causing ruin; (**d**) expected deficit at ruin.

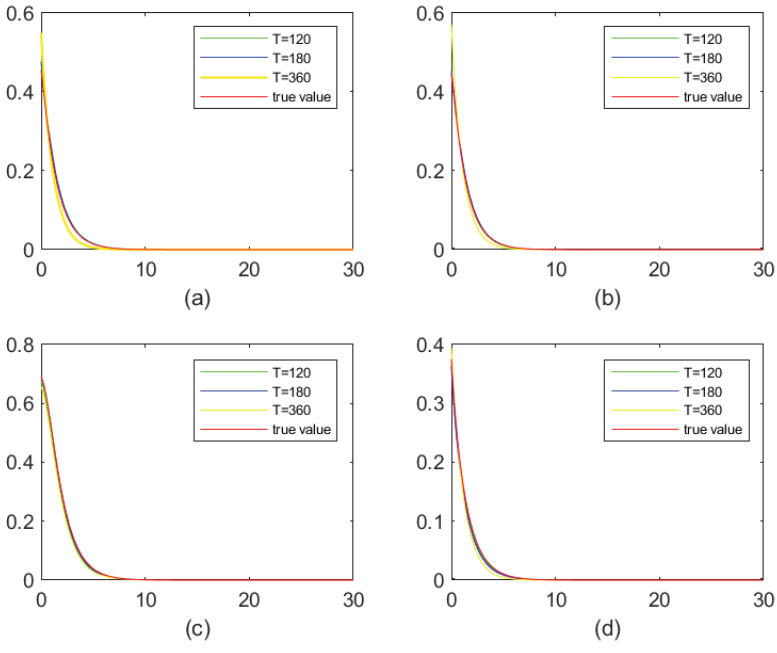

Figure 3. Estimation of the Gerber–Shiu function for combination-of-exponentials density function: mean curves. (**a**) Ruin probability; (**b**) Laplace transform of ruin time; (**c**) expected claim size causing ruin; (**d**) expected deficit at ruin.

We also provide the situation of the mean relative error curves at the Erlang (2) claim size density in Figure 4. It can be noted that (1) the mean relative error curves first increase and then decrease with the increase in u; (2) when T is larger, the average relative error curve is smaller. This fact can be explained as follows: (1) when the initial surplus u is small, the true value $m(u)$ as the denominator is large, which leads to a small mean relative error; (2) with the increase in u, the true value $m(u)$ decreases, so the mean relative error increases; (3) as u continues to increase, the estimated value $\widehat{m}_K(u)$ as the numerator decreases faster than the true value $m(u)$ as the denominator, which makes the subsequent mean relative error curve drop below zero level.

In addition, based on the above 300 repeated experiments, we give a series of IMSE values of Gerber–Shiu function estimation under three kinds of claim distribution assumptions in Table 1. All the numerical experiments in this paper were completed in MATLAB. Taking exponential density as an example, when $T = 120$, we completed 300 independent repeated experiments in 176.06 s. For each claim density function, the IMSE of the Gerber–Shiu function decreases as T increases. This conclusion also shows the stability of the estimation method in this paper. Finally, we compare the Laguerre series expansion method with FFT method used in Shimizu and Zhang [19]. The parameter setting of FFT is the same as in Shimizu and Zhang [19]. First, we present the IMSE values for both methods in Table 2, and we find that the Laguerre series expansion method can lead to smaller IMSEs compared with the FFT method. Moreover, we set $T = 120$ and display the mean relative error curves in Figure 5, and we find that the Laguerre series expansion method can yield smaller mean relative errors.

Table 1. IMSEs for the estimated Gerber–Shiu functions.

Claim Size	T	RP	LT	ECS	ED
Exponential	120	0.02167	0.00555	0.44172	0.01433
	180	0.01811	0.00304	0.42494	0.00682
	360	0.01649	0.00181	0.39796	0.00468
Erlang (2)	120	0.00196	0.00707	0.22613	0.32025
	180	0.00099	0.00153	0.16997	0.24374
	360	0.00097	0.00023	0.13564	0.16738
Combination-of-exponentials	120	0.00394	0.00192	0.02533	0.00501
	180	0.00271	0.00082	0.02261	0.00108
	360	0.00205	0.00055	0.00163	0.00053

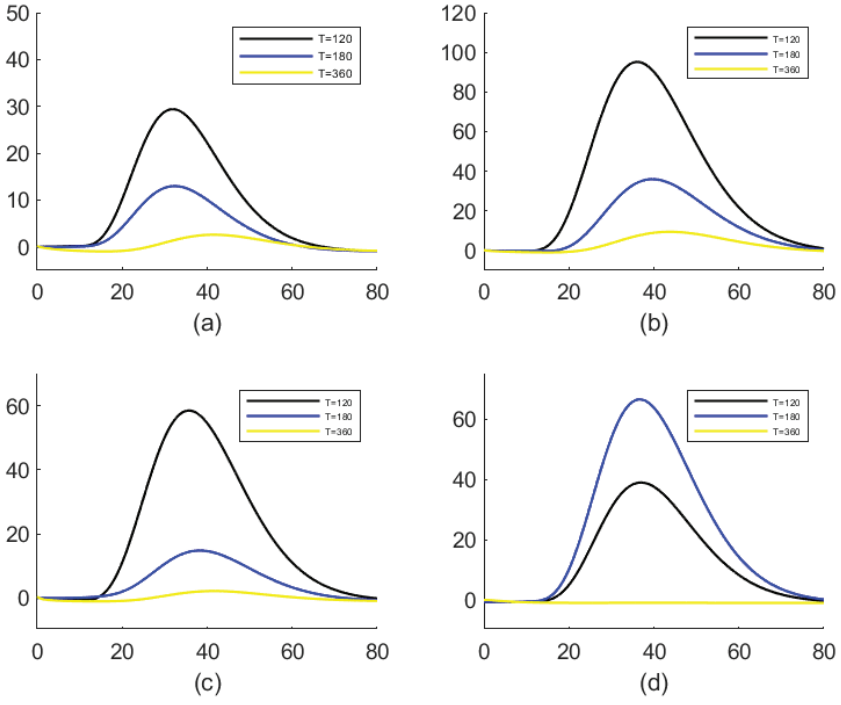

Figure 4. Estimation of the Gerber–Shiu function for Erlang (2) density function: mean relative error curves. (**a**) Ruin probability; (**b**) Laplace transform of ruin time; (**c**) expected claim size causing ruin; (**d**) expected deficit at ruin.

Table 2. IMSEs for the estimated Gerber–Shiu functions.

Claim Size	T	RP	LT	ECS	ED
Exponential	Laguerre	0.02167	0.00555	0.44172	0.01433
	FFT	0.02379	0.00613	0.47373	0.02041
Erlang (2)	Laguerre	0.00196	0.00707	0.22613	0.32025
	FFT	0.00214	0.00813	0.24715	0.41079
Combination-of-exponentials	Laguerre	0.00394	0.00192	0.02533	0.00501
	FFT	0.00424	0.00231	0.03141	0.00673

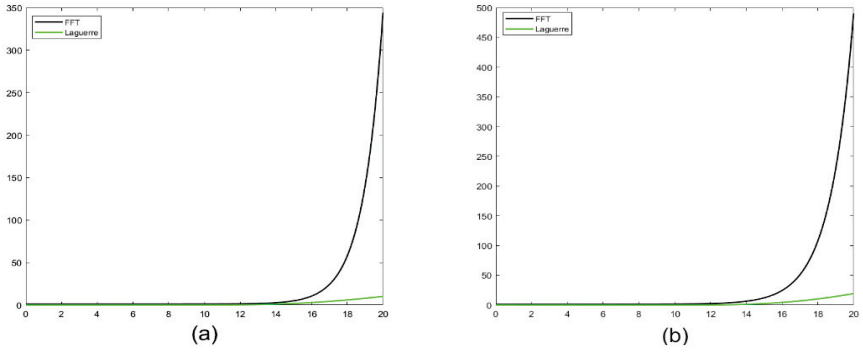

Figure 5. Comparing with FFT method for Erlang (2) density function: mean relative error curves. (**a**) Ruin probability; (**b**) Laplace transform of ruin time.

6. Conclusions

This paper introduces how to use the Laguerre series expansion method to estimate the Gerber–Shiu function of the two-sided jumps risk model and gives the nonparametric estimation of the corresponding ruin characteristic quantity. First, we prove that the Gerber–Shiu function of the two-sided jumps risk model can be expanded by Laguerre series, then Laguerre coefficient can be obtained by solving system of linear equations, and then the unknown coefficients can be estimated based on sample information on claim numbers and individual claim sizes. We derive the consistency property of this estimator when the sample size is large. Finally, when the sample size is limited, we demonstrate the high accuracy of the estimation method through numerical experiments. More importantly, it should be noted that our methods are not limited to be applied to the two-sided jumps risk model, but can be widely applied to other risk models in insurance. In addition, the following studies could be extended to other mathematical methods and models.

Author Contributions: Software, K.H.; Methodology, K.H. and Y.H.; Writing–original draft, K.H. and Y.H.; Writing–review and editing, Y.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Natural Science Foundation of China (Grant Nos. 71701068 and 11601147), and the Natural Science Foundation of Hunan Province (Grant No. 2021JJ30436), and the Scientific Research Fund of Hunan Provincial Education Department, China (Grant Nos. 20B381 and 20K084), and the Natural Science Foundation of Changsha City, China (Grant No. kq2014072).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data in this paper are randomly generated and are not available for readers.

Acknowledgments: The authors would like to thank the editors and reviewers for their help.

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

References

1. Cheung, E.C.K.; Liu, H.; Willmot, G.E. Joint moments of the total discounted gains and losses in the renewal risk model with two-sided jumps. *Appl. Math. Comput.* **2018**, *331*, 358–377. [CrossRef]
2. Boucheire, R.J.; Boxma, P.J.; Sigman, K. A note on negative customers, GI/G/1 workload, and risk processes. *Probab. Eng. Inf. Sci.* **1997**, *358*, 305–311. [CrossRef]
3. Zhang, Z.; Yang, H.; Li, S. The perturbed compound Poisson risk model with two-sided jumps. *J. Comput. Appl. Math.* **2010**, *233*, 1773–1784. [CrossRef]

4. Zou, W.; Gao, J.; Xie, J. On the expected discounted penalty function and optimal dividend strategy for a risk model with random incomes and interclaim-dependent claim sizes. *J. Comput. Appl. Math.* **2014**, *255*, 270–281. [CrossRef]
5. Palmowski, Z.; Vatamidou, E. Phase-type approximations perturbed by a heavy-tailed component for the Gerber-Shiu function of risk processes with two-sided jumps. *Stoch. Model.* **2020**, *26*, 337–363. [CrossRef]
6. Gerber, H.U.; Shiu, E.S.W. On the time value of ruin. *N. Am. Actuar. J.* **1998**, *2*, 48–478. [CrossRef]
7. Zhao, X.; Yin, C. The Gerber-Shiu expected discounted penalty function for Lévy insurance risk processes. *Acta Math. Appl. Sin.* **2010**, *26*, 575–586. [CrossRef]
8. Xie, J.; Zou, W. On the expected discounted penalty function for the compound Poisson risk model with delayed claims. *Comput. Appl. Math.* **2011**, *235*, 2392–2404. [CrossRef]
9. Preischl, M.; Thonhauser, S. Optimal reinsurance for Gerber-Shiu functions in the Cramér-Lundberg model. *Insur. Math. Econ.* **2019**, *87*, 82–91. [CrossRef]
10. Li, S.; Lu, Y.; Sendova, K.P. The expected discounted penalty function: From infinite time to finite time. *Scand. Actuar. J.* **2019**, *2019*, 336–354. [CrossRef]
11. Wang, W.; Zhang, Z. Computing the Gerber-Shiu function by frame duality projection. *Scand. Actuar. J.* **2019**, *4*, 291–307. [CrossRef]
12. Peng, X.; Su, W.; Zhang, Z. On a perturbed compound Poisson risk model under a periodic threshold-type dividend strategy. *J. Ind. Manag. Optim.* **2020**, *16*, 1967–1986. [CrossRef]
13. Wang, W.; Chen, P.; Li, S. Generalized expected discounted penalty function at general drawdown for Lévy risk processes. *Insur. Math. Econ.* **2020**, *91*, 12–25. [CrossRef]
14. Shimizu, Y. Estimation of the expected discounted penalty function for Lévy insurance risks. *Math. Methods Stat.* **2011**, *20*, 125–149. [CrossRef]
15. Shimizu, Y. Non-parametric estimation of the Gerber-Shiu function for the Winer-Poisson risk model. *Scand. Actuar. J.* **2012**, *1*, 56–69. [CrossRef]
16. You, H.; Cai, C. Nonparametric estimation for a spectrally negative Lévy process based on high frequency data. *J. Comput. Appl. Math.* **2019**, *345*, 196–205. [CrossRef]
17. Zhang, Z.; Yang, H. Nonparametric estimate of the ruin probability in a pure-jump Lévy risk model. *Insur. Math. Econ.* **2013**, *53*, 24–35. [CrossRef]
18. Zhang, Z.; Yang, H. Nonparametric estimation for the ruin probability in a Lévy risk model under low-frequency observation. *Insur. Math. Econ.* **2014**, *59*, 168–177. [CrossRef]
19. Shimizu, Y.; Zhang, Z. Estimating Gerber-Shiu functions from discretely observed Lévy driven surplus. *Insur. Math. Econ.* **2017**, *74*, 84–98. [CrossRef]
20. Chau, K.W.; Yam, S.C.P.; Yang, H. Fourier-cosine method for Gerber-Shiu functions. *Insur. Math. Econ.* **2015**, *61*, 170–180. [CrossRef]
21. Chau, K.W.; Yam, S.C.P.; Yang, H. Fourier-cosine method for ruin probabilities. *J. Comput. Appl. Math.* **2015**, *281*, 94–106. [CrossRef]
22. Yang, Y.; Su, W.; Zhang, Z. Estimating the discounted density of the deficit at ruin by Fourier-cosine series expansion. *Stat. Probab. Lett.* **2019**, *146*, 147–155. [CrossRef]
23. Xie, J.; Zhang, Z. Statistical estimation for some dividend problems under the compound Poisson risk model. *Insur. Math. Econ.* **2020**, *95*, 101–115. [CrossRef]
24. Zhang, Z. Estimating the Gerber-Shiu function by Fourier-Sinc series expansion. *Scand. Actuar. J.* **2017**, *10*, 898–979. [CrossRef]
25. Chan, T.L. Efficient computation of European option prices and their sensitivities with the complex Fourier series method. *N. Am. Econ. Financ.* **2019**, *50*, 1–23. [CrossRef]
26. Chan, T.L. Hedging and pricing early-exercise options with complex Fourier series expansion. *N. Am. Econ. Financ.* **2019**, *22*, 1–36.
27. Wang, Y.; Zhang, Z.; Yu, W. Pricing equity-linked death benefits by complex Fourier series expansion in a regime-switching jump diffusion model. *Appl. Math. Comput.* **2021**, *399*, 126301. [CrossRef]
28. You, H.; Yin, C. Threshold estimation for a spectrally negative Lévy process. *Math. Probl. Eng.* **2020**, *2020*, 3561089. [CrossRef]
29. You, H.; Guo, J.; Jiang, J. Interval estimation of the ruin probability in the classical compound Poisson risk model. *Comput. Stat. Data Anal.* **2020**, *144*, 106890. [CrossRef]
30. Wang, Y.; Yu, W.; Huang, Y. Estimating the Gerber-Shiu function in a compound Poisson risk model with stochastic premium income. *Discret. Dyn. Nat. Soc.* **2019**, *2019*, 5071268. [CrossRef]
31. Wang, Y.; Yu, W.; Huang, Y.; Yu, X.; Fan, H. Estimating the expected discounted penalty function in a compound poisson insurance risk model with mixed premium income. *Mathematics* **2019**, *7*, 305. [CrossRef]
32. Li, J.; Yu, W.; Liu, C. Nonparametric estimation of ruin probability by complex Fourier series expansion in the compound Poisson model. *Commun. Stat. Theory Methods* **2020**, *51*, 5048–5063. [CrossRef]
33. Xie, J.; Zhang, Z. Finite-time dividend problems in a Lévy risk model under periodic observation. *Appl. Math. Comput.* **2021**, *398*, 125981.
34. Huang, Y.; Li, J.; Liu, H.; Yu, W. Estimating ruin probability in an insurance risk model with stochastic premium income based on the CFS method. *Mathematics* **2021**, *9*, 982. [CrossRef]

35. Su, W.; Wang, Y. Estimating the Gerber-Shiu function in Lévy insurance risk model by Fourier-cosine series expansion. *Mathematics* **2021**, *9*, 1402. [CrossRef]
36. Wang, W.; Xie, J.; Zhang, Z. Estimating the time value of ruin in a Lévy risk model under low-frequency observation. *Insur. Math. Econ.* **2022**, *104*, 133–157. [CrossRef]
37. Yang, Y.; Xie, J.; Zhang, Z. Nonparametric estimation of some dividend problems in the perturbed compound Poisson model. *Probab. Eng. Inf. Sci.* **2022**, *37*, 418–441. [CrossRef]
38. Ai, M.; Zhang, Z.; Zhong, W. Valuation of a DB underpin hybrid pension under a regime-switching Lévy model. *J. Comput. Appl. Math.* **2023**, *419*, 114736. [CrossRef]
39. He, Y.; Kawai, R.; Shimizu, Y.; Yamazaki, K. The Gerber-Shiu discounted penalty function: A review from practical perspectives. *Insur. Math. Econ.* **2023**, *109*, 1–28. [CrossRef]
40. Zhong, W.; Cui, Z.; Zhang, Z. Efficient valuation of guaranteed minimum maturity benefits in regime switching jump diffusion models with surrender risk. *J. Comput. Appl. Math.* **2023**, *422*, 114914. [CrossRef]
41. Comte, F.; Genon-Catalot, V. Adaptive Laguerre density estimation for mixed Poisson models. *Electron. J. Stat.* **2015**, *9*, 1113–1149. [CrossRef]
42. Zhang, Z.; Su, W. A new efficient method for estimating the Gerber-Shiu function in the classical risk model. *Scand. Actuar. J.* **2018**, *5*, 426–449. [CrossRef]
43. Zhang, Z.; Su, W. Estimating the Gerber-Shiu function in a Lévy risk model by Laguerre series expansion. *J. Comput. Appl. Math.* **2019**, *346*, 133–149. [CrossRef]
44. Zhang, Z.; Yong, Y. Valuing guaranteed equity-linked contracts by Laguerre series expansion. *J. Comput. Appl. Math.* **2019**, *357*, 329–348. [CrossRef]
45. Cheung, E.C.K.; Zhang, Z. Simple approximation for the ruin probability in renewal risk model under interest force via Laguerre series expansion. *Scand. Actuar. J.* **2021**, *9*, 804–831. [CrossRef]
46. Albrecher, H.; Cheung, E.C.K.; Liu, H.; Woo, J.K. A bivariate Laguerre expansions approach for joint ruin probabilities in a two-dimensional insurance risk process. *Insur. Math. Econ.* **2022**, *103*, 96–118. [CrossRef]
47. Xie, J.; Zhang, Z. Infinite series expansion of some finite-time dividend and ruin related functions. *Commun. Stat. Theory Methods* **2022**. [CrossRef]
48. Su, W.; Yong, Y.; Zhang, Z. Estimating the Gerber-Shiu function in the perturbed compound Poisson model by Laguerre series expansion. *J. Math. Anal. Appl.* **2019**, *469*, 705–729. [CrossRef]
49. Su, W.; Shi, B.; Wang, Y. Estimating the Gerber-Shiu function under a risk model with stochastic income by Laguerre series expansion. *Commun. Stat. Theory Methods* **2019**, *49*, 5686–5708. [CrossRef]
50. Huang, Y.; Yu, W.; Pan, Y.; Cui, C. Estimating the Gerber-Shiu expected discounted penalty function for Lévy risk model. *Discret. Dyn. Nat. Soc.* **2019**, *2019*, 3607201. [CrossRef]
51. Su, W.; Yu, W. Asymptotically normal estimators of the Gerber-Shiu function in classical insurance risk model. *Mathematics* **2020**, *8*, 1638. [CrossRef]
52. Xie, J.; Yu, W.; Zhang, Z.; Cui, Z. Gerber-Shiu analysis in the compound Poisson model with constant inter-observation times. *Probab. Eng. Inf. Sci.* **2022**, *37*, 324–356. [CrossRef]
53. Cheung, E.C.K.; Lau, H.; Willmot, G.E. ; Woo, J.K. Finite-time ruin probabilities using bivariate Laguerre series. *Scand. Actuar. J.* **2023**, *2*, 153–190. [CrossRef]
54. Dickson, D.C.M.; Hipp, C. On the time to ruin for Erlang(2) risk processes. *Insur. Math. Econ.* **2001**, *29*, 333–344. [CrossRef]
55. Li, S.; Garrido, J. On ruin for the Erlang(n) risk process. *Insur. Math. Econ.* **2004**, *34*, 391–408. [CrossRef]
56. Abramowitz, M.; Stegun, I.A. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*; National Bureau of Standards Applied Mathematics Series; Courier Corporation: North Chelmsford, MA, USA, 1964.
57. Bongioanni, B.; Torrea, J.L. What is a Sobolev space for the Laguerre function system? *Stud. Math.* **2009**, *192*, 147–172. [CrossRef]
58. Labbé, C.; Sendov, H.S.; Sendova, K.P. The Gerber-Shiu function and the generalized Cramér-Lundberg model. *Appl. Math. Comput.* **2011**, *218*, 3035–3056.
59. Stenger, F. *Numerical Methods Based on Sinc and Analytic Functions*; Springer: New York, NY, USA, 1993.
60. Böttcher, A.B.; Grudsky, S.M. *Toeplitz Matrices, Asymptotic Linear Algebra, and Functional Analysis*; Birkhauser: Basel, Switzerland, 2000.
61. Stewart, G.W.; Sun, J.G. *Matrix Perturbation Theory*; Academic Press: Beijing, China, 1990.

mathematics

MDPI

Article

Perturbed Skew Diffusion Processes

Yingxu Tian and Haoyan Zhang *

College of Science, Civil Aviation University of China, Tianjin 300300, China; yxtian@cauc.edu.cn
* Correspondence: zhang-hy@cauc.edu.cn

Abstract: This work investigates whether there uniquely exists a solution to the perturbed skew diffusion process. We construct the solution by iteration and divide the whole time interval into parts on which we disperse the perturbed skew diffusion process into two tractable portions, one for perturbed diffusion process, the other for skew diffusion process. After this disposition, we only focus on the process in each time interval. Noticing the continuity on every time interval boundaries generalized by a sequence of stopping times, we acquire the main result of this paper as well as a time change for the perturbed skew process.

Keywords: skew diffusion process; perturbed diffusion process; perturbed skew diffusion process; local time; change in time

MSC: 60J55; 60J60

1. Introduction

This work documents the properties of the solution to the perturbed skew diffusion process, which writes as

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t + \alpha d \max_{0\le s\le t} X_s + (2p-1)d\hat{L}_t^X(0), \tag{1}$$

where W denotes a standard Brownian motion on a filtered complete probability space $(\Omega, \mathcal{F}, \mathrm{P})$ with respect to a filtration $\{\mathcal{F}_t, t \ge 0\}$, μ, σ are supposed to be globally Lipschitz, σ satisfies additionally the Engelbert–Schimdt condition (Lemmas 2 and 3), $\alpha < 1$, $p \in]0,1[$, and the symmetric local time $\hat{L}^X(0)$ is an increasing process starting from $0 \in \mathbf{R}$ such that

$$\int_0^t 1_{\{X_s=0\}} d\hat{L}_s^X(0) = \hat{L}_t^X(0).$$

Brownian motion with skew point was first practiced by Itô and McKean [1] to describe certain stochastic dynamics related to Feller's classification. Later Walsh [2] researched the skew Brownian motion with a discontinuous local time. Afterwards, the classical stochastic differential equation (SDE) expression was established by Harrison and Shepp [3]. Then, Le Gall [4] solved the process as a solution of generalized SDE with local time. Recently, multiskewed Brownian motion was studied by Ramirez [5]. Before arriving at skew point, skew Brownian motion just behaves as a standard Brownian motion. Once hitting skew point, skew Brownian motion moves up and down with different probability, being p and $1-p$, respectively. Note that skew Brownian motion reduces to standard Brownian motion (reflected Brownian motion) when $p = 0.5$ ($p = 0$ or 1). Naturally, skew Brownian motion is more flexible than Brownian motion. The reader may consult more details on recent theoretical development and applications of skew Brownian motion in Lejay [6].

Based on this advantage, the skew diffusion process as a generalization of typical diffusion processes has diverse applications, ranging from mathematical finance in Decamps et al. [7] and Monte Carlo simulation schemes in Lejay and Martinez [8] to heterogeneous

Citation: Tian, Y.; Zhang, H. Perturbed Skew Diffusion Processes. *Mathematics* **2023**, *11*, 2417. https://doi.org/10.3390/math11112417

Academic Editors: Jing Yao, Xiang Hu, Jingchao Li

Received: 22 April 2023
Revised: 17 May 2023
Accepted: 22 May 2023
Published: 23 May 2023

media in Freidlin and Sheu [9]. In this paper, the skew diffusion process satisfies the following SDE

$$dZ_t = \mu(Z_t)dt + \sigma(Z_t)dW_t + (2p-1)d\hat{L}_t^Z(0). \tag{2}$$

Just as the skew diffusion process solving the SDE with the term of symmetric local time, the perturbed diffusion process arises with the term of maximum

$$dU_t = \mu(U_t)dt + \sigma(U_t)dW_t + \alpha d \max_{0\le s\le t} U_s. \tag{3}$$

This version of perturbed process has attracted a crowd of scholars who have devoted themselves to creating a rich literature (see e.g., Carmona, Petit, and Yor [10,11], Chaumont and Doney [12,13], Le Gall and Yor [14,15] and Perman and Werner [16]). Lately, the existence and pathwise uniqueness of the solution of the perturbed reflected process and the doubly perturbed jump-diffusion processes have been practiced by Doney and Zhang [17] and Hu and Ren [18], respectively.

In addition, classical models without skew point have recently failed to capture the actual dynamics caused by more and more world events. It is easy in Figure 1 to observe that from 2005 to 2017, the federal funds rate of America expresses a novel trend. The rate approaches zero from 2008, then stays near zero until 2015. Evidently, a special level governs such a trend and naturally should be taken into consideration in our setting. Hence, we introduce skew diffusion process in our setting to show the mean-reverting and bounded situation.

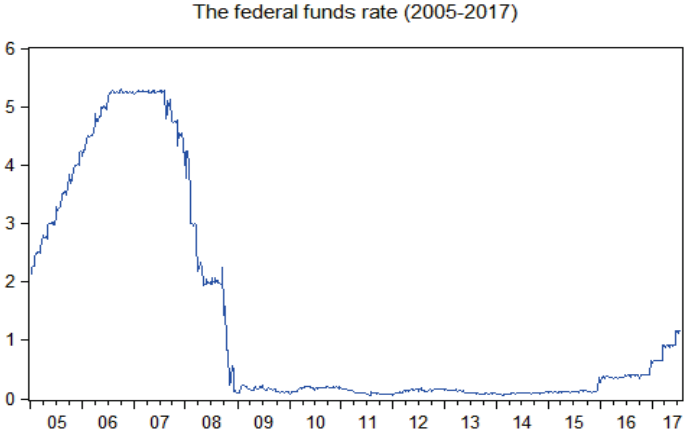

Figure 1. The federal funds rate from 2005 to 2017.

To our knowledge, because there are no previous works concerning the perturbed skew diffusion process, we must handle with proving the properties of the solution defined in Equation (1). However, it seems not easy to prove the result when both perturbed item and skew item exist. To overcome this obstacle, we divide the whole time interval into many parts, hence we are able to focus on the perturbed skew diffusion process in these interval parts, instead of the whole interval. With this division, we disperse the perturbed skew process into two tractable elements, perturbed diffusion process and skew diffusion process, respectively. Then we give a clear proof by iteration on these time intervals. In the meantime, we check the continuity in each time interval, which are generated by a sequence of stopping times. Hence, the existence and uniqueness of solution to the perturbed skew diffusion process is proved.

The remainder of our work is arranged as follows. In Section 2, we provide the iteration lemma as well as prove existence and uniqueness of solution to the perturbed diffusion process by means of this lemma. Section 3 puts forward the relevant analysis

about the solution to skew diffusion with the help of signed measure. Section 4 deduces the basic property of solution to the perturbed skew diffusion process and performs one time change version for this process.

2. Perturbed Diffusion Process

This section answers the question of whether the solution to the perturbed diffusion process Equation (3) uniquely exists.

Let $\{W_t, t \geq 0\}$ be a standard Brownian motion with respect to the filtration $\{\mathcal{F}_t, t \geq 0\}$ on a probability space $(\Omega, \mathcal{F}, \mathrm{P})$. We consider the following SDE

$$U_t = U_0 + \int_0^t \mu(U_s)ds + \int_0^t \sigma(U_s)dW_s + \alpha \max_{0\leq s\leq t} U_s, \tag{4}$$

with the assumption that the coefficients of perturbed diffusion process satisfy the Lipschitz continuous condition, i.e., for an existing constant b, the following inequalities hold:

$$|\sigma(u) - \sigma(v)| \leq b|u - v|$$

and

$$|\mu(u) - \mu(v)| \leq b|u - v|.$$

Before obtaining the result of the solution to the perturbed diffusion process, we show a useful lemma.

Lemma 1. *Suppose that there are two continuous functions U_t and H_t. If*

$$U_t = H_t + \alpha \max_{0\leq s\leq t} U_s,$$

then

$$U_t = H_t + \frac{\alpha}{1-\alpha} \max_{0\leq s\leq t} H_s.$$

Proof. We write the equation into iteration for two steps,

$$\begin{aligned} U_t &= H_t + \alpha \max_{0\leq s\leq t} U_s \\ &= H_t + \alpha \max_{0\leq s\leq t} [H_s + \alpha \max_{0\leq s_1\leq s} U_{s_1}] \\ &= H_t + \alpha \max_{0\leq s\leq t} [H_s + \alpha \max_{0\leq s_1\leq s} [H_{s_1} + \alpha \max_{0\leq s_2\leq s_1} U_{s_2}]]. \end{aligned}$$

In order to establish the equality, we need two procedures, as follows.

First, we check the upper bound. It is obvious to see the maximum of the sum is less than or equal to the sum of the maximum, that is

$$\begin{aligned} U_t &= H_t + \alpha \max_{0\leq s\leq t} [H_s + \alpha \max_{0\leq s_1\leq s} [H_{s_1} + \cdots]] \\ &\leq H_t + \alpha \max_{0\leq s\leq t} [H_s + \alpha \max_{0\leq s_1\leq t} [H_{s_1} + \cdots]] \\ &= H_t + \alpha \max_{0\leq s\leq t} [H_s + \alpha \max_{0\leq s\leq t} [H_s + \cdots]] \\ &= H_t + \alpha \max_{0\leq s\leq t} H_s + \alpha^2 \max_{0\leq s\leq t} H_s + \cdots \\ &= H(t) + \frac{\alpha}{1-\alpha} \max_{0\leq s\leq t} H_s. \end{aligned}$$

Second, we check the lower bound, noticing the fact that for any two functions $P(t), Q(t)$,

$$\max_{0\leq s\leq t} \{P(s) + \max_{0\leq s_1\leq s} Q(s_1)\} \geq \max_{0\leq s\leq t} \{P(s) + Q(s)\},$$

Thus, we obtain

$$\begin{aligned} U_t &= H_t + \alpha \max_{0\le s\le t}[H_s + \alpha \max_{0\le s_1\le s}[H_{s_1} + \alpha \max_{0\le s_2\le s_1}[H_{s_2} + \cdots]]] \\ &\ge H_t + \alpha \max_{0\le s\le t}[H_s + \alpha[H_s + \alpha \max_{0\le s_2\le s}[H_{s_2} + \cdots]]] \\ &\ge H_t + \alpha \max_{0\le s\le t}[H_s + \alpha[H_s + \alpha[H_s + \cdots]]] \\ &= H(t) + \frac{\alpha}{1-\alpha}\max_{0\le s\le t} H_s. \end{aligned}$$

This proves the lemma. □

The following theorem states the existence and uniqueness of the solution to the perturbed diffusion process.

Theorem 1. *Let U_0 be a random variable which is independent of W and $E[|U_0|^2] < \infty$. Then, for any fixed $T > 0$, there uniquely exists a continuous solution U_t (adapted with respect to $\mathcal{F}_t$), $t \ge 0$ to (4) satisfying $E[\max_{0\le s\le T}|U_s|^2] < \infty$.*

Proof. Set

$$U_t^0 = \frac{U_0}{1-\alpha}, \quad 0 \le t < \infty.$$

Then, we denote by

$$U_t^{n+1} = U_0 + \int_0^t \mu(U_s^n)ds + \int_0^t \sigma(U_s^n)dW_s + \alpha \max_{0\le s\le t} U_s^{n+1} \tag{5}$$

the unique continuous adapted solution to perturbed diffusion process with $n \ge 0$. By Lemma 1, set $H_t = U_0 + \int_0^t \mu(U_s^n)ds + \int_0^t \sigma(U_s^n)dW_s$ in Equation (5), we have

$$\begin{aligned} U_t^{n+1} = \frac{U_0}{1-\alpha} &+ \int_0^t \mu(U_s^n)ds + \int_0^t \sigma(U_s^n)dW_s \\ &+ \frac{\alpha}{1-\alpha}\max_{0\le s\le t}[\int_0^s \mu(U_\eta^n)d\eta + \int_0^s \sigma(U_\eta^n)dW_\eta]. \end{aligned} \tag{6}$$

With this iteration expression, as well as the Theorem 2.1 in Doney and Zhang [17], we complete this proof. □

Remark 1. *In fact, Theorem 1 is borrowed from Doney and Zhang [17]. Because Equation (6) in Doney and Zhang [17] is straightly provided, we present more details in Lemma 1.*

3. Skew Diffusion Process

This section answers the question of whether the solution to the skew diffusion process Equation (2) uniquely exists.

When $p = 1$ or 0 in Equation (2), the skew diffusion process degenerates into the reflected diffusion process, and the existence and uniqueness of the solution to this reflected diffusion has been studied by Lions and Sznitman [19]. Therefore, we only study the case $0 < p < 1$.

To begin with, consider the stochastic equation with generalized drift as follows

$$X_t = X_0 + \int_R \hat{L}_t^X(y)\nu(dy) + \int_0^t b(X_s)dB_s, \tag{7}$$

in which B denotes a Brownian motion, $\hat{L}^X$ is the symmetric local time of the unknown process X, and $\nu(dy)$ signifies a signed measure. To help us prove the result, we need the description in Engelbert and Schmidt [20]. Set $N_f = \{y \in \mathbf{R} : f(y) = 0\}$ and $E_f = \{y \in \mathbf{R} :$

$\int_G f^{-2}(z)dz = +\infty\}$, where G is any open set including y. Here, we introduce two useful lemmas without proving them.

Lemma 2 (Theorem 4.35 in Engelbert and Schmidt [20])**.** *There exist three equivalent conclusions as follows:*
(a) There exists a fundamental solution X to Equation (7).
(b) There exists a solution X to Equation (7).
(c) $E_b \subseteq N_b$.

Lemma 3 (Theorem 4.37 in Engelbert and Schmidt [20])**.** *(a) The fundamental solution to Equation (7) is unique.*
(b) The solution to Equation (7) is unique if and only if the following condition is satisfied:
If $E_b \subseteq N_b$ then $E_b = N_b$.

Remark 2. *On page 153 in Engelbert and Schmidt [20], it states that "Firstly, we establish the existence of a unique solution to Equation (7) which spends minimal time at the zeros of the diffusion coefficient b. We call it fundamental solution. This solution is a strong MARKOV continuous semimartingale up to the explosion time."*

What follows next is the theorem to obtain the existence and uniqueness of the solution to the skew diffusion process.

Theorem 2. *Assume $0 < p < 1$ if the coefficient $\sigma \neq 0$ in Equation (2) satisfies the bounded condition, i.e., there exists a constant $M < +\infty$ such that for all open sets G containing x, $\int_G 1/\sigma^2(x)dx \leq M$ and both coefficients μ, σ satisfy the Lipschitz condition. Then, there exists a unique solution to Equation (2).*

Proof. Define a signed measure by

$$\nu(A) \triangleq \int_A \frac{\mu(x)}{\sigma^2(x)}dx + (2p-1)\delta_0(A),$$

where $\delta_0(\cdot)$ is the Delta function. By the occupation time formula in Equation (A3) (see Appendix A), Equation (7) becomes

$$\begin{aligned}
&Z_0 + \int_R \hat{L}_t^Z(y)\nu(dy) + \int_0^t \sigma(Z_s)dW_s \\
=&Z_0 + \int_R \hat{L}_t^Z(y)\frac{\mu(y)}{\sigma^2(y)}dy + \int_R \hat{L}_t^Z(y)(2p-1)\delta_0(dy) + \int_0^t \sigma(Z_s)dW_s \\
=&Z_0 + \int_0^t \frac{\mu(Z_s)}{\sigma^2(Z_s)}d\langle Z\rangle_s + (2p-1)\hat{L}_t^Z(0) + \int_0^t \sigma(Z_s)dW_s \\
=&Z_0 + \int_0^t \mu(Z_s)dZ_s + (2p-1)\hat{L}_t^Z(0) + \int_0^t \sigma(Z_s)dW_s \\
=&Z_t.
\end{aligned}$$

Noticing that let $M \in [1, +\infty)$, we have a finite signed measure ν defined on $\mathcal{B}([-M, M])$. As the fundamental assumption in Engelbert and Schmidt [20], $|\nu(\{x\})| \in [0, 1)$ holds. On the other hand, because the coefficient σ satisfies the bounded condition, we know that the set $E_\sigma = \{x \in \mathbb{R} : \int_G \frac{1}{\sigma^2(x)}dx = \infty\} = \emptyset$. Note that $N_\sigma = \{x \in \mathbf{R} : \sigma(x) = 0\} = \emptyset$, then by Lemmas 2 and 3, we prove the existence and uniqueness of solution to Equation (2), and the solution (fundamental solution) is also a strong Markov process. □

4. Perturbed Skew Diffusion Process

With the last two sections devoted to obtaining the results of solutions to the perturbed diffusion process and skew diffusion process, respectively, this section answers the question of whether the solution to the perturbed skew diffusion process in Equation (8) uniquely exists.

$$X_t = x + \int_0^t \mu(X_s)ds + \int_0^t \sigma(X_s)dW_s + (2p-1)\hat{L}_t^X(0) + \alpha \max_{0\le s\le t} X_s, \tag{8}$$

where the coefficients μ, σ, p, α, and $W, \hat{L}^X(0)$ are the same in (1). Because there exists a big difference under case $x = 0$ and $x \neq 0$, we study them separately. What follows is the main result.

Theorem 3. *Let $x(\neq 0)$ be a random variable which is independent of W and $E(|x|^2) < \infty$. There exists a unique continuous solution to Equation (8).*

Proof. Iterative technique is adopted to prove the solution property parallel to Le Gall and Yor [15] and Doney and Zhang [17].

Set U^0 to be the unique solution to the following equation:

$$U_t^0 = x + \int_0^t \mu(U_s^0)ds + \int_0^t \sigma(U_s^0)dW_s^0 + \alpha \max_{0\le s\le t} U_s^0.$$

Obviously, such a unique solution to the perturbed diffusion process exists from Section 2. Then, put $T_1 = \inf\{t \ge 0, U_t^0 = 0\}$; we know $T_1 > 0$, as $x \neq 0$. Define $X_t = U_t^0$ and $\hat{L}_t^X(0) = \hat{L}_t^U(0) = 0$ for $t \in [0, T_1]$ with $W_t^1 = W_{t+T_1} - W_{T_1}$ for $t \in [0, +\infty)$, as it is known to all that W_t^1 denotes a standard Brownian motion independent of $\mathcal{F}_{T_1}$.

Next let us focus on the skew diffusion process

$$\begin{cases} V_t^1 = \int_0^t \mu(V_s^1)ds + \int_0^t \sigma(V_s^1)dW_s^1 + (2p-1)\hat{L}_t^{V,1}(0), \\ V_0^1 = 0, \\ \hat{L}_0^{V,1}(0) = 0, \int_0^t 1_{\{V_s^1=0\}} d\hat{L}_s^{V,1}(0) = \hat{L}_t^{V,1}(0). \end{cases} \tag{9}$$

It is known that such a unique solution exists from Section 3.

In general, assume that X has been defined in the time interval $t \in [0, T_{2n-1}]$. We establish X for $T_{2n-1} \le t \le T_{2n+1}$, $n \ge 1$, as follows.

First, suppose V^{2n-1} to be the solution to the skew diffusion process:

$$\begin{cases} V_t^{2n-1} = \int_0^t \mu(V_s^{2n-1})ds + \int_0^t \sigma(V_s^{2n-1})dW_s^{2n-1} + (2p-1)\hat{L}_t^{V,2n-1}(0), \\ V_0^{2n-1} = 0, \\ \hat{L}_0^{V,2n-1}(0) = 0, \int_0^t 1_{\{V_s^{2n-1}=0\}} d\hat{L}_s^{V,2n-1}(0) = \hat{L}_t^{V,2n-1}(0), \end{cases}$$

where $W_t^{2n-1} = W_{t+T_{2n-1}} - W_{T_{2n-1}}$. It should be noted that $\hat{L}^{V,2n-1}(0)$ stands for the symmetric local time of V^{2n-1} at 0 for all $n \ge 1$. Put $T_{2n} = \inf\{t > T_{2n-1}; V_{t-T_{2n-1}}^{2n-1} = \max_{0\le s\le T_{2n-1}} X_s\}$ and for $T_{2n-1} \le t \le T_{2n}$, define

$$\begin{cases} X_t = V_{t-T_{2n-1}}^{2n-1}, \\ \hat{L}_t^X(0) = \hat{L}_{T_{2n-1}}^X(0) + \hat{L}_{t-T_{2n-1}}^{V,2n-1}(0). \end{cases} \tag{10}$$

Second, suppose U^{2n} be the solution to the perturbed diffusion process:

$$U_t^{2n} = (1-\alpha)X_{T_{2n}} + \int_0^t \mu(U_s^{2n})ds + \int_0^t \sigma(U_s^{2n})dW_s^{2n} + \alpha \max_{0\le s\le t} U_s^{2n},$$

with $W_t^{2n} = W_{t+T_{2n}} - W_{T_{2n}}$. It should be pointed out that $X_{T_{2n}}$ in the above perturbed diffusion process is known and is exactly equal to $\max\limits_{0\le s\le T_{2n-1}} X_s$. Put $T_{2n+1} = \inf\{t > T_{2n}; U_{t-T_{2n}}^{2n} = 0\}$ and for $T_{2n} \le t \le T_{2n+1}$, define

$$\begin{cases} X_t = U_{t-T_{2n}}^{2n}, \\ \hat{L}_t^X(0) = \hat{L}_{T_{2n}}^{U,2n}(0). \end{cases} \tag{11}$$

By this construction, we obtain a line of nondecreasing stopping times T_n, $n \ge 0$. We put $0 = T_0$ and $T = \lim\limits_{n\to\infty} T_n$. Then, X is a continuous process defined on the time interval $[0, T]$ and T is also a stopping time. At this rate, our aim is to show that X defined by the above construction satisfies Equation (8) for $T_{2n} \le t \le T_{2n+1}$, $n = 0, 1, 2, ...,$. First, we show the continuity for X on the time interval boundary. To see this, we show the following equalities.

If X satisfies Equation (10), we obtain

$$\begin{cases} X_{T_{2n-1}} = V_{T_{2n-1}-T_{2n-1}}^{2n-1} = V_0^{2n-1} = 0, \\ X_{T_{2n}} = V_{T_{2n}-T_{2n-1}}^{2n-1} = \max\limits_{0\le s\le T_{2n-1}} X_s. \end{cases}$$

If X satisfies Equation (11), we have

$$\begin{cases} X_{T_{2n+1}} = U_{T_{2n+1}-T_{2n}}^{2n} = 0, \\ X_{T_{2n}} = U_{T_{2n}-T_{2n}}^{2n} = U_0^{2n} = X_{T_{2n}} = \max\limits_{0\le s\le T_{2n-1}} X_s. \end{cases}$$

Noticing the value of X at the time interval boundaries ($T_n, n \ge 0$), the continuity property follows.

Now, we show that X is the unique solution to Equation (8). For $n = 0$, e.g., $0 \le t \le T_1$, we know that X is defined by

$$X_t = U_t^0.$$

Recall the definition that $\hat{L}_t^X(0) = \hat{L}_t^U(0) = 0$ for $0 \le t \le T_1$, hence

$$\begin{aligned} X_t &= (1-\alpha)X_0 + \int_0^t \mu(U_s^0)ds + \int_0^t \sigma(U_s^0)dW_s^0 + \alpha \max_{0\le s\le t} U_s^0 \\ &= (1-\alpha)U_0^0 + \int_0^t \mu(U_s^0)ds + \int_0^t \sigma(U_s^0)dW_s^0 + \alpha \max_{0\le s\le t} U_s^0 + (2p-1)\hat{L}_t^X(0) \\ &= x + \int_0^t \mu(U_s^0)ds + \int_0^t \sigma(U_s^0)dW_s + \alpha \max_{0\le s\le t} U_s^0 + (2p-1)\hat{L}_t^X(0). \end{aligned} \tag{12}$$

We conclude that X is the solution to the perturbed skew diffusion process in $0 \le t \le T_1$. Then, for $n = 1$, i.e., $T_1 \le t \le T_2$, X is defined by

$$X_t = V_{t-T_1}^1.$$

It is easy to derive

$$\begin{aligned} X_t &= \int_0^{t-T_1} \mu(V_s^1)ds + \int_0^{t-T_1} \sigma(V_s^1)dW_s^1 + (2p-1)\hat{L}_{t-T_1}^{V,1}(0) \\ &= X_{T_1} + \int_0^{t-T_1} \mu(V_s^1)ds + \int_0^{t-T_1} \sigma(V_s^1)dW_s^1 + (2p-1)\hat{L}_{t-T_1}^{V,1}(0) \\ &= x + \int_0^t \mu(X_s)ds + \int_0^t \sigma(X_s)dW_s + (2p-1)\hat{L}_t^X(0) + \alpha \max_{0\le s\le T_1} X_s \\ &= x + \int_0^t \mu(X_s)ds + \int_0^t \sigma(X_s)dW_s + (2p-1)\hat{L}_t^X(0) + \alpha \max_{0\le s\le t} X_s, \end{aligned}$$

where we use the following facts: $\max_{0\le s\le T_1} X_s = \max_{0\le s\le t} X_s$ for $T_1 \le t \le T_2$, $X_{T_1} = 0$, substitute T_1 for t in Equation (12), and the definition by Equation (10). Furthermore, if $T_1 \le t \le T_2$, we also derive

$$\int_0^t 1_{\{X_s=0\}} d\hat{L}_s^X(0) = \int_{T_1}^t 1_{\{X_s=0\}} d\hat{L}_{s-T_1}^{V,1}(0) = \int_0^{t-T_1} 1_{\{V_u^1=0\}} d\hat{L}_u^{V,1}(0) = \hat{L}_{t-T_1}^{V,1}(0) = \hat{L}_t^X(0).$$

Thus, we obtain the solution to the perturbed skew diffusion process for $0 \le t \le T_2$. In general, assume that we establish the solution X to Equation (8) at the time interval $[0, T_{2n}]$. When $T_{2n} \le t \le T_{2n+1}$, from the definition, X takes the form of

$$X_t = U_{t-T_{2n}}^{2n},$$

which leads to

$$\begin{aligned}
X_t &= (1-\alpha)X_{T_{2n}} + \int_0^{t-T_{2n}} \mu(U_s^{2n})ds + \int_0^{t-T_{2n}} \sigma(U_s^{2n})dW_s^{2n} + \alpha \max_{0\le s\le t-T_{2n}} U_s \\
&= (1-\alpha)(x + \int_0^{T_{2n}} \mu(X_s)ds + \int_0^{T_{2n}} \sigma(X_s)dW_s + (2p-1)\hat{L}_{T_{2n}}^X(0) + \alpha \max_{0\le s\le T_{2n}} X_s) \\
&\quad + \int_0^{t-T_{2n}} \mu(U_s^{2n})ds + \int_0^{t-T_{2n}} \sigma(U_s^{2n})dW_s^{2n} + \alpha \max_{0\le s\le t-T_{2n}} U_s \\
&= x + \int_0^{T_{2n}} \mu(X_s)ds + \int_0^{T_{2n}} \sigma(X_s)dW_s + (2p-1)\hat{L}_{T_{2n}}^X(0) + \alpha \max_{0\le s\le T_{2n}} X_s - \alpha X_{T_{2n}} \\
&\quad + \int_0^{t-T_{2n}} \mu(U_s^{2n})ds + \int_0^{t-T_{2n}} \sigma(U_s^{2n})dW_s^{2n} + \alpha \max_{0\le s\le t-T_{2n}} U_s \\
&= x + \int_0^t \mu(X_s)ds + \int_0^t \sigma(X_s)dW_s + (2p-1)\hat{L}_t^X(0) + \alpha \max_{0\le s\le t} X_s,
\end{aligned}$$

Note that $X_{T_{2n}} = \max_{0\le s\le t-T_{2n-1}} X_s$, and $\max_{0\le s\le T_{2n}} X_s = \max_{0\le s\le t-T_{2n-1}} X_s$, the expression $\max_{T_{2n}\le s\le t} X_s = \max_{0\le s\le t} X_s$ holds.

Then, X satisfies the equation Equation (8) for $T_{2n} \le t \le T_{2n+1}$. In addition, from the definition of stopping times T_n, we see $X_t \ne 0$, $t \in [T_{2n}, T_{2n+1})$, implying

$$\hat{L}_t^X(0) = \int_0^t 1_{\{X_s=0\}} d\hat{L}_s^X(0) = \int_0^{T_{2n}} 1_{\{X_s=0\}} d\hat{L}_s^X(0) = \hat{L}_{T_{2n}}^X(0).$$

In a similar way, we can show the solution X for $T_{2n+1} \le t \le T_{2n+2}$ as well.

Lastly, we prove $T = \infty$, *a.s.* With the definition of T_{2n+1}, we have

$$X_{T_{2n+1}} = 0.$$

We can also write $X_{T_{2n+1}}$ by

$$\begin{aligned}
X_{T_{2n+1}} &= X_{T_{2n+1}} + X_{T_{2n}} - X_{T_{2n}} \\
&= \max_{0\le s\le T_{2n}} X_s + \int_{T_{2n}}^{T_{2n+1}} \mu(X_s)ds + \int_{T_{2n}}^{T_{2n+1}} \sigma(X_s)dW_s \\
&\quad + \alpha(\max_{0\le s\le T_{2n+1}} X_s - \max_{0\le s\le T_{2n}} X_s) + (2p-1)(\hat{L}_{T_{2n+1}}(0) - \hat{L}_{T_{2n}}(0)).
\end{aligned}$$

Suppose $T < \infty$ *a.s.* with positive probability and let $n \to \infty$; we have $\max_{0\le s\le T} X_s = X_{T_{2n+1}} = 0$, which contradicts the definition of $X_0 = \frac{x}{1-\alpha} \ne 0$.

On the other hand, the construction of the processes in every time intervals gives the uniqueness of the solution, and such a solution is unique in the whole time interval. We complete the proof. □

Theorem 4. *Suppose $x = 0$, if $0 \le \alpha < \frac{1}{2}$ and the coefficient σ satisfies the bounded condition in the previous section (Theorem 2, Section 3), then there exists a unique continuous solution to Equation (8).*

Proof. We want to use the iteration scheme. Set $X_t^0 = 0$, and $\{X_t^{n+1}\}_{n\ge 0}$ satisfies

$$X_t^{n+1} = \int_0^t \mu(X_s^n)ds + \int_0^t \sigma(X_s^n)dW_s + (2p-1)\hat{L}_t^{X,n+1}(0) + \alpha \sup_{0\le s\le t} X_s^{n+1}. \tag{13}$$

By the reflection principle,

$$\hat{L}_t^{X,n+1}(0) = -\inf_{s\le t}\Big\{\big(\int_0^s \mu(X_u^n)du + \int_0^s \sigma(X_u^n)dW_u + \alpha \sup_{0\le u\le s} X_u^{n+1}\big) \wedge 0\Big\}. \tag{14}$$

Now, Equations (13) and (14) lead to

$$\begin{aligned}|X_t^{n+1} - X_t^n| \le &\Big|\int_0^t (\mu(X_s^n) - \mu(X_s^{n-1}))ds\Big| + \Big|\int_0^t (\sigma(X_s^n) - \sigma(X_s^{n-1}))dW_s\Big| \\ &+ \alpha\Big|\sup_{0\le s\le t} X_s^{n+1} - \sup_{0\le s\le t} X_s^n\Big| + (2p-1)|\hat{L}_t^{X,n+1}(0) - \hat{L}_t^{X,n}(0)|.\end{aligned}$$

As a result,

$$\begin{aligned}\sup_{s\le t}|X_s^{n+1} - X_s^n| \le &\frac{2p}{1-2\alpha p}\sup_{s\le t}\Big|\int_0^s (\mu(X_u^n) - \mu(X_u^{n-1}))du\Big| \\ &+ \frac{2p}{1-2\alpha p}\sup_{s\le t}\Big|\int_0^s \sigma(X_u^n) - \sigma(X_u^{n-1})dW_u\Big|.\end{aligned}$$

By Burkholder's inequality,

$$\begin{aligned}E\big[\sup_{s\le t}|X_s^{n+1} - X_s^n|^2\big] &\le C_{\alpha,p}E\big[\int_0^t (\sigma(X_s^n) - \sigma(X_s^{n-1}))^2 ds\big] \\ &\le C_{\alpha,p}E\big[\int_0^t (X_s^n - X_s^{n-1})^2 ds\big].\end{aligned}$$

Thus, we deduce that for any fixed $T > 0$,

$$E\big[\sup_{s\le T}|X_s^{n+1} - X_s^n|^2\big] \le \frac{(C_{\alpha,p,T})^n}{n!},$$

which yields

$$P\big[\max_{0\le s\le T}|X_s^{n+1} - X_s^n| > \frac{1}{2^n}\big] \le \frac{(4C_{\alpha,p,T})^n}{n!}.$$

With the lemma of Borel–Cantelli, we can deduce that $X^n \to X$ on $[0,T]$ *a.s.* We can also obtain the convergence property of $\int_0^t \mu(X_s^n)ds + \int \sigma(X_s^n)dW_s$ *a.s.* Accordingly, in Equation (13), it is seen that $\hat{L}^{X,n}(0)$ performs the convergence property to some nondecreasing process $\hat{L}^X(0)$. Lastly, we have the fact that

$$\hat{L}_t^X(0) = \lim_{n\to\infty}\hat{L}_t^{X,n}(0) = \lim_{n\to\infty}\int_0^t 1_{\{X_s^n=0\}}d\hat{L}_s^{X,n}(0) = \int_0^t 1_{\{X_s=0\}}d\hat{L}_s^X(0).$$

Indeed, for any $f \in C_0(0,+\infty)$, $\int_0^t f(X_s)d\hat{L}_s^X(0) = \lim_{n\to\infty}\int_0^t f(X_s^n)d\hat{L}_s^{X,n}(0)$ holds. Then, we turn to proving the uniqueness of the solution. Using a similar way as the one above, let X^1 and X^2 be two solutions to Equation (13); we find

$$E[\sup_{s\le t}|X_s^1 - X_s^2|^2] \le C_{\alpha,p}E[\int_0^t (X_u^1 - X_u^2)^2 du].$$

By Gronwall's inequality, it is obvious that $X_t^1 = X_t^2$. Thus, we complete the proof. □

To further explore the time change in the perturbed skew process, we provide the next corollary.

Corollary 1. *Set $T_0 = 0$ and $X_0 = x$. Define $T_i(t) = \inf\{t > T_{i-1}, X_t > X_{T_{i-1}}\}$, $i \ge 1$ with $T_\infty(t) = \infty$. Define a new process $Y_t = X_{T_i(t)}$, then Y is a continuous skew diffusion process. Furthermore, $\{T_i(t), i \ge 0\}$ form a sequence of stopping times.*

Proof. It is obvious to see that $\{T_i(t), i \ge 0\}$ are stopping times depending on t, and for each time interval $[T_{i-1}(t), T_i(t)]$ it follows that

$$\max_{T_{i-1}(t)\le s\le T_i(t)} X_s = \max_{0\le s\le T_i(t)} X_s = X_{T_i(t)},$$

where the last equality comes from the apparent relationships

$$0 = T_0(t) < T_1(t) < \cdots < T_\infty(t) = \infty,$$

and

$$x = X_0 < X_{T_1(t)} < \cdots < X_{T_\infty(t)} = \infty.$$

Then, rewrite Equation (8) by

$$(1-\alpha)X_{T_i(t)} = x + \int_0^{T_i(t)} \mu(X_s)ds + \int_0^{T_i(t)} \sigma(x_s)dW_s + (2p-1)\hat{L}_{T_i(t)}^X(0).$$

We conclude that $Y_t = (1-\alpha)X_{T_i(t)}$ is a continuous-time skew diffusion after this time change, and we prove this corollary. □

5. Conclusions and Summary

In this work, we consider a novel dynamic called the perturbed skew diffusion process. Such a process contains perturbed and skew phenomena. Perturbed phenomenon means that the model will reflect the maximum of the model in the past time, while the skew phenomenon reflects different probabilities of the upward and downward movement by p and $1-p$, respectively. We first prove the existence and uniqueness of the solution to the perturbed skew diffusion process. The idea is to disperse the perturbed skew diffusion process into the perturbed diffusion and skew diffusion processes, respectively. To learn more construction about this model, we study the relation between the perturbed skew diffusion process and skew diffusion process. In the future, perturbed skew diffusion may be applied to lookback options. Lookback options are options where the return depends not only on the strike price of the underlying asset but also on the highest or lowest price of the underlying asset over the life of the option. As with many exotic options, the payoff structure of a lookback option is related to the maximum or minimum value reached by the underlying asset price during the term of the contract. Lookback options are also path-dependent options, which enable the holder to execute an option at the most beneficial price of the underlying dynamic during the term of an option. A benefit from this exotic option is that the investor can "look back" or have retrospect on the underlying setting of such an option after getting into a long or short position, and then they can seek to

maximize the value of the options. In summary, perturbed phenomena may help investors know the past maximum dynamics, hence maximizing their benefits.

Author Contributions: Methodology, Y.T.; Writing—original draft, Y.T.; Writing—review & editing, H.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the National Natural Science Foundation of China (No. 12101602), the Counterpart Fund of National Natural Science Foundation of China (No. 3122022PT17).

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank the editors and the referees for valuable comments and suggestions which helped to improve the paper significantly.

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

Appendix A

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space. Denote by a continuous semimartingale $\{X_t, \mathcal{F}_t; 0 \leq t \leq \infty\}$ an adapted process which can uniquely be expressed by

$$X_t = X_0 + M_t + V_t.$$

M_t writes on a continuous martingale ($M_0 = 0$), and V_t is an adapted bounded variation process with continuous sample trajectory ($V_0 = 0$).

We borrow a similar expression of local times with respect to X in Protter [21]. Denote by $\text{sign}(y)$ the sign function as follows

$$\text{sign}(y) = \begin{cases} 1, & y > 0, \\ -1, & y \leq 0, \end{cases}$$

and provide the local time with respect to X defined by

$$\frac{1}{2}L_t^X(\beta) = (X_t - \beta)^+ - (X_0 - \beta)^+ - \int_0^t 1_{(\beta,+\infty)}(X_s)dX_s,$$

$$\frac{1}{2}L_t^X(\beta) = (X_t - \beta)^- - (X_0 - \beta)^- + \int_0^t 1_{(-\infty,\beta]}(X_s)dX_s,$$

$$L_t^X(\beta) = |X_t - \beta| - |X_0 - \beta| - \int_0^t \text{sign}(X_s - \beta)dX_s.$$

Denote by $L_t^X(\beta)$ (resp. $L_t^X(\beta-)$) the local time (resp. left local time) for X_t, where $L^X(\beta-) = \lim_{c\to\beta, c<\beta} L^X(c)$. Then, the symmetric local time for X at the point β takes the following form

$$\hat{L}_t^X(\beta) = \frac{L_t^X(\beta) + L_t^X(\beta-)}{2}, \tag{A1}$$

$\mathbf{P}^\beta$-a.s. for every $\beta \in \mathbf{R}$. It is a continuous increasing process in t and is constant on any interval on which $X_t \neq \beta$.

Suppose that f is a function whose left and right derivatives are denoted by f'_+ and f'_- and the second derivative measure μ (μ can be viewed as a general second differential measure of f ($\mu = f''$ in much of the literature)). Then, from the Meyer–Tanaka formula (see also Salins and Spiliopoulos [22]), we have:

$$f(X_t) = f(X_0) + \frac{1}{2}\int_0^t [f'_+(X_s) + f'_-(X_s)]dX_s + \frac{1}{2}\int_{\mathbb{R}} \hat{L}_t^X(y)\mu(dy). \tag{A2}$$

Here, μ has the properties $\mu((a,b]) = f'_+(b) - f'_+(a)$ and $\mu(\{x\}) = \lim_{a \uparrow x} \mu((a,x])$.

We have almost surely, for every $t \geq 0$ and every non-negative measurable function φ on $\mathbb{R}$,

$$\int_0^t \varphi(X_s) d\langle X, X\rangle_s = \int_{\mathbb{R}} \varphi(a) L_t^X(a) da. \tag{A3}$$

For more details on symmetric local time, we refer the reader to Karatzas and Shreve [23] and Revuz and Yor [24]).

References

1. Itô, K.; McKean, H.P. *Diffusion Processes and Their Sample Paths*; Springer: Berlin/Heidelberg, Germany, 1965.
2. Walsh, J.B. A diffusion with a discontinuous local time. *Asterisque* **1978**, *52–53*, 37–45.
3. Harrison, J.M.; Shepp, L.A. On skew Brownian motion. *Ann. Probab.* **1981**, *9*, 309–313. [CrossRef]
4. Le Gall, J.F. One-dimensional stochastic differential equations involving the local times of the unknown process. In *Stochastic Analysis and Applications, Proceedings of the International Conference Held in Swansea, 11–15 April 1983*; Lecture Notes in Math.; Springer: Berlin/Heidelberg, Germany, 1984; Volume 1095, pp. 51–82.
5. Ramirez, J.M. Multi-skewed Brownian motion and diffusion. *Proc. Am. Math. Soc.* **2010**, *139*, 3739–3752. [CrossRef]
6. Lejay, A. On the construction of the skew Brownian motion. *Probab. Surv.* **2006**, *3*, 413–466. [CrossRef]
7. Decamps, M.; Goovaerts, M.; Schoutens, W. Asymmetric skew Bessel processes and their applications to finance. *J. Comput. Appl. Math.* **2006**, *186*, 130–147. [CrossRef]
8. Lejay, A.; Martinez, M. A scheme for simulating one-dimensional diffusion processes with discontinuous coefficients. *Ann. Appl. Probab.* **2006**, *16*, 107–139. [CrossRef]
9. Freidlin, M.; Sheu, S.J. Diffusion processes on graphs: Stochastic differential equations, large deviation priciple. *Probab. Theory Relat.* **2000**, *116*, 181–220. [CrossRef]
10. Carmona, P.; Petit, F.; Yor, M. Some extentions of the arcsine law as partial consequences of the scaling property of Brownian motion. *Probab. Theory Relat.* **1994**, *100*, 1–29. [CrossRef]
11. Carmona, P.; Petit, F.; Yor, M. Beta variables as times spent in $[0,\infty)$ by certain perturbed Brownian motions. *J. Lond. Math. Soc.* **1998**, *58*, 239–256. [CrossRef]
12. Chaumont, L.; Doney, R.A. Pathwise uniqueness for perturbed versions of Brownian motion and reflected Brownian motion. *Probab. Theory Relat.* **1999**, *113*, 519–534. [CrossRef]
13. Chaumont, L.; Doney, R.A. Some calculations for doubly perturbed Brownian motion. *Stoch. Proc. Appl.* **2000**, *85*, 61–74. [CrossRef]
14. Gall, J.F.L.; Yor, M. Excursions browniennes et carrés de processus de Bessel. *C. R. Acad. Sci. Paris Série I* **1986**, *303*, 73–76
15. Gall, J.F.L.; Yor, M. Enlacements du mouvement brownien autour des courbes de l'espace. *Trans. Am. Math. Soc.* **1990**, *317*, 687–722.
16. Perman, M.; Werner, W. Perturbed Brownian motions. *Probab. Theory Relat.* **1997**, *108*, 357–383. [CrossRef]
17. Doney, R.A.; Zhang, T. Perturbed Skorohod equations and perturbed reflected diffusion. *Ann. I'HPoincaré-PR* **2005**, *41*, 107–121. [CrossRef]
18. Hu, L.; Ren, Y. Doubly perturbed neutral stochastic functional equations. *J. Comput. Appl. Math.* **2009**, *231*, 319–326. [CrossRef]
19. Lions, P.L.; Sznitman, A.S. Stochastic differential equations with reflecting boundary conditions. *Comm. Pur. Math. Soc.* **1984**, *37*, 511–537. [CrossRef]
20. Engelbert, H.J.; Schmidt, W. Strong Markov continuous local martingales and solutions of one-dimensional stochastic differential equations (Part III). *Math. Nachr.* **1991**, *151*, 149–197. [CrossRef]
21. Protter, P. *Stochastic Integration and Differential Equations*; Springer: Berlin/Heidelberg, Germany, 2004.
22. Salins, M.; Spiliopoulos, K. Markov processes with spatial delay: Path space characterization, occupation time and properties. *Stoch. Dyn.* **2017**, *17*, 1750042.
23. Karatzas, I.; Shreve, S.E. *Brownian Motion and Stochastic Calculus*; Springer: Berlin/Heidelberg, Germany, 1991. [CrossRef]
24. Revuz, D.; Yor, M. *Continuous Martingales and Brownian Motion*; Springer: Berlin/Heidelberg, Germany, 1999.

mathematics

MDPI

Article

Finite-Time Ruin Probabilities of Bidimensional Risk Models with Correlated Brownian Motions

Dan Zhu [1], Ming Zhou [2] and Chuancun Yin [1,*]

[1] School of Statistics and Data Science, Qufu Normal University, Qufu 273165, China; zhudanspring@qfnu.edu.cn
[2] Center for Applied Statistics, School of Statistics, Renmin University of China, Beijing 100872, China; mingzhou@ruc.edu.cn
* Correspondence: ccyin@qfnu.edu.cn

Abstract: The present work concerns the finite-time ruin probabilities for several bidimensional risk models with constant interest force and correlated Brownian motions. Under the condition that the two Brownian motions $\{B_1(t), t \geq 0\}$ and $\{B_2(t), t \geq 0\}$ are correlated, we establish new results for the finite-time ruin probabilities. Our research enriches the development of the ruin theory with heavy tails in unidimensional risk models and the dependence theory of stochastic processes.

Keywords: bidimensional perturbed risk model; correlated brownian motions; finite-time ruin probability; heavy-tailed risk model; interest force

MSC: 60H05

Citation: Zhu, D.; Zhou, M.; Yin, C. Finite-Time Ruin Probabilities of Bidimensional Risk Models with Correlated Brownian Motions. *Mathematics* **2023**, *11*, 2767. https://doi.org/10.3390/math11122767

Academic Editors: Jing Yao, Xiang Hu and Jingchao Li

Received: 8 May 2023
Revised: 15 June 2023
Accepted: 16 June 2023
Published: 19 June 2023

1. Introduction

In traditional studies, many researchers have investigated ruin probability problems of insurers under unidimensional models. For example, ref. [1] studied ruin probability problems with constant interest force. Other studies about these problems can be found in [2–5]. An assumption behind these models is that the insured businesses homogeneous and can be described by a unidimensional model; however, this assumption is too strong. Thus, bidimensional or multidimensional insurance risk models have received growing interest in recent years, such as [6–8]. Various assumptions have been considered regarding the claim arrival process and the distribution of claim amounts; see, e.g., [9–12]. Ref. [13] considered finite-time ruin probabilities for nonstandard bidimensional renewal risk models with constant interest forces and diffusion generated by Brownian motions; they assumed that the two Brownian motions $\{B_1(t), t \geq 0\}$ and $\{B_2(t), t \geq 0\}$ are mutually independent. Similar results were obtained by [14], although they considered dependent subexponential claims. More papers can be found in [15,16], and the references therein. In this paper, we consider uniform asymptotics for the finite-time ruin probabilities for several bidimensional risks models with constant interest force and correlated Brownian motions, meaning that the businesses of the insurer have a relationship with each other. We introduce risk models and different types of ruin times with corresponding ruin probabilities as follows.

The bidimensional risk model $\vec{U}(t) = (U_1(t), U_2(t))^\tau$ is the surplus vector of an insurance company at time $t \geq 0$; in this paper, we state this formally as

$$U_i(t) = u_i e^{rt} + \int_0^t e^{r(t-s)} dC_i(s) - \int_0^t e^{r(t-s)} dS_i(s) + \sigma_i \int_0^t e^{r(t-s)} dB_i(s),\ t \geq 0, \qquad (1)$$

where $\vec{u} = (u_1, u_2)^\tau$ stands for the initial surplus vector and $\vec{C}(t) = (C_1(t), C_2(t))^\tau$ for the total premiums received up to time t; here, $\{C_1(t), t \geq 0\}$, $\{C_2(t), t \geq 0\}$ are mutually independent. Moreover, $r \geq 0$ stands for the interest rate and $(S_1(t), S_2(t)) =$

$(\sum_{i=1}^{N_1(t)} X_{1i}, \sum_{i=1}^{N_2(t)} X_{2i})$ for the total amount of claims vector up to time t. Here, $\vec{X}_i = (X_{1i}, X_{2i})^\tau, i = 1, 2, \cdots$ denote pairs of claims with arrival times that constitute a counting process vector $\{\vec{N}(t), t \geq 0\}$, where $\vec{N}(t) = (N_1(t), N_2(t))$, while $\{N_1(t), t \geq 0\}$, $\{N_2(t), t \geq 0\}$ are mutually independent. The process $\{N_i(t), t \geq 0\}$ is a Poisson process with intensity $\lambda_i > 0$, and $\{\vec{X}_i, i = 1, 2, \cdots\}$ is a sequence of independent copies of the random pair $\vec{X} = (X_1, X_2)^\tau$ with the joint distribution function $F(x_1, x_2)$ and the marginal distribution functions $F_1(x_1)$ and $F_2(x_2)$. For all vectors, the $\vec{X}_i$s and $\vec{C}$ consist of only non-negative components $\vec{C}(0) = (0, 0)^\tau$. Moreover, each $C_i(t)$ is a non-decreasing and right-continuous stochastic process. The vector $\vec{B}(t) = (B_1(t), B_2(t))^\tau$ denotes a standard bidimensional Brownian motion with a constant correlation coefficient $\rho \in [-1, 1]$, while $\sigma_1 \geq 0$ and $\sigma_2 \geq 0$ are constants. For simplicity, we assume that $\{\vec{X}_i, i = 1, 2, \cdots\}$, $\{\vec{N}(t), t \geq 0\}$ and $\{\vec{C}(t), t \geq 0\}$ are independent and that both of them are independent of $\{\vec{B}(t), t \geq 0\}$. To avoid the certainty of ruin in each class, we assume that the following safety loading conditions hold when $r = 0$:

$$EC_i(t) - \lambda_i EX_{i1} > 0, \ i = 1, 2.$$

In this paper, we consider the following four types of ruin probabilities. For a finite horizon $T > 0$, we define

$$\psi_{\max}(\vec{u}, T) = P(T_{\max} \leq T | \vec{U}(0) = \vec{u}), \tag{2}$$

where

$$T_{\max} = \inf\{t > 0 | \max\{U_1(t), U_2(t)\} < 0\};$$

$$\psi_{\min}(\vec{u}, T) = P(T_{\min} \leq T | \vec{U}(0) = \vec{u}), \tag{3}$$

where

$$T_{\min} = \inf\{t > 0 | \min\{U_1(t), U_2(t)\} < 0\};$$

and

$$\psi_{\text{sum}}(\vec{u}, T) = P(T_{\text{sum}} \leq T | \vec{U}(0) = \vec{u}), \tag{4}$$

where

$$T_{\text{sum}} = \inf\{t > 0 | U_1(t) + U_2(t) < 0\};$$

$$\psi_{\text{and}}(\vec{u}, T) = P(T_{\text{and}} \leq T | \vec{U}(0) = \vec{u}), \tag{5}$$

where $T_{\text{and}} = \max\{T_1, T_2\}$ and

$$T_i = \inf\{t > 0 | U_i(t) < 0 \text{ for some } 0 \leq t \leq T), \ i = 1, 2,$$

with $\inf \emptyset = \infty$ by convention.

We remark that the probability in (2) denotes the probability of ruin occurring when both $U_1(t)$ and $U_2(t)$ are below zero at the same time within finite time $T > 0$, the probability in (3) denotes the probability of ruin occurring when at least one of $\{U_i(t), i = 1, 2\}$ is below zero within finite time $T > 0$, the probability in (4) denotes the probability of ruin occurring when the total of $U_1(t)$ and $U_2(t)$ is negativ within finite time $T > 0$, and the probability in (5) denotes the probability of ruin occurring when both $U_1(t)$ and $U_2(t)$ are below zero, not necessarily simultaneously, within a finite time $T > 0$. T_{and} represents a more critical time than $T_{\max}$, and the ruin probability defined by T_{sum} is reduced to that in the unidimensional model. The following relation between the four ruin probabilities defined above holds:

$$\psi_{\max}(\vec{u}, T) \leq \psi_{\text{and}}(\vec{u}, T) \leq \psi_{\min}(\vec{u}, T), \ \psi_{\text{sum}}(\vec{u}, T) \leq \psi_{\min}(\vec{u}, T),$$

and

$$\psi_{\min}(\vec{u},T)+\psi_{\text{and}}(\vec{u},T)=P(T_1\le T|U_1(0)=u_1)+P(T_2\le T|U_2(0)=u_2). \tag{6}$$

The rest of this paper is organized as follows. In Section 2 we review the related results after briefly introducing preliminaries about heavy-tailed distributions, in Section 3 we provide several important definitions and lemmas, and the main results and the proof procedure are presented in Section 4.

2. Review of Related Results

Unless otherwise stated herein, all limit relations are for $(u_1,u_2)\to(\infty,\infty)$. We denote $a\lesssim b$ and $a\gtrsim b$ if $\limsup a/b\le 1$ and $\limsup a/b\ge 1$, respectively, and $a\sim b$ if both, where, $a(\cdot,\cdot)$ and $b(\cdot,\cdot)$ are two positive functions. Let $F_1*\cdots*F_n$ be the convolution of the distributions $F_1,\cdots,F_n$ and let F^{*n} denote the n-fold convolution of a distribution F.

In this section, we review definitions and properties that are relevant to the results of this paper, considering only the case of the distribution of heavy-tail claims. An r.v. X or its d.f. $F(x)=1-\overline{F}(x)$ satisfying $\overline{F}(x)>0$ for all $x\in(-\infty,\infty)$ is called heavy-tailed to the right, or simply heavy-tailed, if $E[e^{\gamma X}]=\infty$ for all $\gamma>0$. In the following, we recall several important classes of heavy-tailed distributions.

F is a long tailed distribution, written as $F\in\mathcal{L}$, if $\lim\limits_{x\to\infty}\frac{\overline{F}(x-t)}{\overline{F}(x)}=1$ holds for some $t>0$. Note that the convergence is uniform over t in compact intervals. If $\lim\limits_{x\to\infty}\frac{\overline{F^{*n}}(x)}{\overline{F}(x)}=n$ holds $(n=2,3,\cdots)$, then F is a subexponential distribution on $(0,\infty)$, written as $F\in\mathcal{S}$. For some $0<t<1$, if $\limsup\limits_{x\to\infty}\frac{\overline{F}(tx)}{\overline{F}(x)}<\infty$ holds, F is said to have a dominatedly varying tailed distribution, written as $F\in\mathcal{D}$. We call F a consistently varying tailed distribution, written as $F\in\mathcal{C}$, if

$$\lim_{t\downarrow 1}\liminf_{x\to\infty}\frac{\overline{F}(tx)}{\overline{F}(x)}=1,\ \text{ or equivalently if }\ \lim_{t\uparrow 1}\limsup_{x\to\infty}\frac{\overline{F}(tx)}{\overline{F}(x)}=1$$

holds. A distribution F is extended regularly-varying tailed, written as $F\in\mathcal{ERV}(-\alpha,-\beta)$ for some $0\le\alpha\le\beta<\infty$, if $s^{-\beta}\le\liminf\limits_{x\to\infty}\frac{\overline{F}(sx)}{\overline{F}(x)}\le\limsup\limits_{x\to\infty}\frac{\overline{F}(sx)}{\overline{F}(x)}\le s^{-\alpha}$ holds for $s\ge 1$.

It is obvious that the following formula holds:

$$\mathcal{ERV}(-\alpha,-\beta)\subset\mathcal{C}\subset\mathcal{D}\cap\mathcal{L}\subset\mathcal{S}\subset\mathcal{L}.$$

There are many other references to heavy-tailed distributions; readers may refer to [17–22] among others.

The asymptotic behavior of the finite-time ruin probability of bidimensional or multidimensional risk models has previously been investigated by [23]. They proved that under the conditions $F_1,F_2\in\mathcal{S}$, $N_1(t)=N_2(t)$, and $\sigma_1=\sigma_2=0$, it is the case that $r>0$ and the claim vector $\vec{X}$ consist of independent components

$$\psi_{\max}(\vec{u};T)\sim\frac{\lambda(\lambda+\frac{1}{T})}{r^2}\int_{u_1}^{u_1e^{rT}}\frac{\overline{F_1}(y)}{y}dy\int_{u_2}^{u_2e^{rT}}\frac{\overline{F_2}(y)}{y}dy,\ \text{ as }\ (u_1,u_2)\to(\infty,\infty).$$

Under the conditions $F_1,F_2\in\mathcal{S}$, $r=0$, and $N_1(t)=N_2(t)$, it is the case that $C_i(\cdot)$ are deterministic linear functions, and both the claim vector $\vec{X}$ and the bidimensional Brownian motion $\vec{B}$ consist of independent components. Li et al. [12] found that for each fixed time $T>0$,

$$\psi_{\max}(\vec{u};T)\sim\lambda T(1+\lambda T)\overline{F_1}(u_1)\overline{F_2}(u_2),\ \text{ as }\ (u_1,u_2)\to(\infty,\infty).$$

Chen et al. [11] investigated the uniform asymptotics of $\psi_{\text{and}}(\vec{u},T)$ and $\psi_{\min}(\vec{u},T)$ for an ordinary renewal risk model with the claim amounts belonging to the consistently varying tailed distributions class for large T. Zhang and Wang [24] considered model (1) with

$r = 0$ and assumed that all sources of randomness, $\{X_{1k}, k = 1, 2, \cdots\}$, $\{X_{2k}, k = 1, 2, \cdots\}$, $\{N_1(t) = N_2(t), t \geq 0\}$, $\{B_1(t), t \geq 0\}$ and $\{B_2(t), t \geq 0\}$ are mutually independent. They obtained that if $F_1, F_2 \in \mathcal{EVR}(-\alpha, -\beta)$ for some $0 < \alpha \leq \beta < \infty$, then, for each fixed time $T \geq 0$,

$$\psi_{\max}(\vec{u}; T) \sim \lambda T(1 + \lambda T)\overline{F_1}(u_1)\overline{F_2}(u_2), \text{ as } (u_1, u_2) \to (\infty, \infty).$$

The analogous result for multidimensional risk models can be found in Asmussen and Albrecher [17].

3. Some Lemmas

Before providing the main results, we first provide several lemmas.

Lemma 1. *If $F \in \mathcal{S}$, then for each $\varepsilon > 0$ there exists some constant $C_\varepsilon > 0$ such that the inequality*

$$\overline{F^{*n}}(x) \leq C_\varepsilon(1 + \varepsilon)^n \overline{F}(x)$$

holds for all $n = 1, 2, \cdots$ and $x \geq 0$.

Proof. See Lemma 1.3.5 of Embrechts et al. [25]. □

Lemma 2. *Let G_1 and G_2 be two distribution functions. If $G_1 \in \mathcal{S}$ and $\overline{G_2}(x) = o(\overline{G_1}(x))$, then we have $\overline{G_1 * G_2}(x) \sim \overline{G_1}(x)$ as $x \to \infty$.*

Proof. See Proposition 1 of Embrechts et al. [25]. □

Lemma 3. *Consider a unidimensional risk model*

$$U_i(t) = u_i + C_i(t) - S_i(t) + \sigma_i B_i(t), \ t \geq 0, i = 1, 2. \tag{7}$$

If $F_i \in \mathcal{S}$, then the ruin probability with finite-horizon T satisfies

$$\psi_i(u_i; T) = P(U_i(t) < 0 \text{ for some } t \leq T | U_i(0) = u_i) \sim \lambda T \overline{F_i}(u_i), u_i \to \infty.$$

Proof. Clearly, on the one hand,

$$\begin{aligned}
\psi_i(u_i; T) &\geq P(S_i(T) \geq u_i + C_i(T) + \sigma_i \sup_{0 \leq t \leq T} B_i(t)) \\
&= \int_0^\infty P(S_i(T) \geq u_i + C_i(T) + \sigma_i z) dP(\sup_{0 \leq t \leq T} B_i(t)) \leq z) \\
&= P(S_i(T) \geq u_i) \int_0^\infty \int_0^\infty \frac{P(S_i(T) \geq u_i + l_i + \sigma_i z)}{P(S_i(T) \geq u_i)} dP(\sup_{0 \leq t \leq T} B_i(t)) \leq z) \\
&\quad \times dP(C_i(T) \leq l_i) \\
&\sim P(S_i(T) \geq u_i),
\end{aligned} \tag{8}$$

where we have used the fact that $P(S_i(T) \geq u_i + l_i + \sigma_i z) \leq P(S_i(T) \geq u_i)$ and the dominated convergence theorem.

On the other hand,

$$\begin{aligned}
\psi_i(u_i; T) &\leq P(S_i(T) + \sigma_i \sup_{0 \leq t \leq T} (-B_i(t)) \geq u_i) \\
&\sim P(S_i(T) \geq u_i),
\end{aligned} \tag{9}$$

where we have used Lemma 2 and the fact that

$$P(\sigma_i \sup_{0 \leq t \leq T} (-B_i(t)) \geq u_i) = o(P(S_i(T) \geq u_i)).$$

Per Lemma 1 and dominated convergence theorem, we have

$$P(S_i(T) \geq u_i) \sim \overline{F}_i(u_i) \sum_{n=1}^{\infty} nP(N(T) = n) = \lambda T \overline{F}_i(u_i), \text{ as } u_i \to \infty.$$

The result follows from (8) and (9). □

Lemma 4. *Consider a unidimensional risk model*

$$U_i(t) = u_i e^{rt} + \int_0^t e^{r(t-s)} C_i(ds) - \int_0^t e^{r(t-s)} dS_i(s) + \sigma_i \int_0^t e^{r(t-s)} dB_i(s),\ t \geq 0, i = 1, 2.$$

If $F_i \in \mathcal{S}$, then the ruin probability with finite-horizon T satisfies

$$\psi_i(u_i; T) = P(U_i(t) < 0 \text{ for some } t \leq T | U_i(0) = u_i) \sim \frac{\lambda}{r} \int_{u_i}^{u_i e^{rT}} \frac{\overline{F}_i(y)}{y} dy,\ u_i \to \infty.$$

Proof. By simply modifying the proof of Lemma 3, we have

$$\psi_i(u_i; T) \sim P\left(\sum_{j=1}^{N(T)} X_{ij} e^{-r\tau_j} \geq u_i \right) \sim \lambda \int_0^T P(X_{i1} e^{-rz} > u_i) dz,\ u_i \to \infty,$$

where in the last step we use (28) from [26]. Here, τ_j are the arrival times of the Poisson process $N(t)$. In fact,

$$z = \frac{1}{r} \log \frac{y}{u_i},$$

and we have that

$$dz = d\left(\frac{1}{r} \log \frac{y}{u_i} \right) = \frac{1}{r} \cdot \frac{u_i}{y} \cdot \frac{1}{u_i} dy = \frac{1}{ry} dy.$$

Then,

$$\lambda \int_0^T P(X_{i1} > u_i e^{rz}) dz = \frac{\lambda}{r} \int_{u_i}^{u_i e^{rT}} \frac{\overline{F}_i(y)}{y} dy.$$

Upon a trivial substitution, the required result is implied. □

Definition 1.

(i) Two processes $\{X_1(t); t \geq 0\}$ and $\{X_2(t); t \geq 0\}$ are said to be positively associated if

$$\text{Cov}(f(X_1(t_1), X_2(t_2)), g(X_1(t_1), X_2(t_2)) | X_1(0) = x_1, X_2(0) = x_2) \geq 0$$

for all non-decreasing real valued functions f and g such that covariance exists, all $t_1, t_2 \geq 0$, and all $x_1, x_2 \in \mathbb{R}$.

(ii) Two processes $\{X_1(t); t \geq 0\}$ and $\{X_2(t); t \geq 0\}$ are said to be negatively associated if

$$\text{Cov}(f(X_1(t_1)), g(X_2(t_2)) | X_1(0) = x_1, X_2(0) = x_2) \leq 0,$$

for all non-decreasing real valued functions f and g such that covariance exists, all $t_1, t_2 \geq 0$, and all $x_1, x_2 \in \mathbb{R}$.

Definition 2. *Two processes $\{X_1(t); t \geq 0\}$ and $\{X_2(t); t \geq 0\}$ are said to be positively (negatively) quadrant-dependent if*

$$\begin{aligned} &P(X_1(t_1) > y_1, X_2(t_2) > y_2 | X_1(0) = x_1, X_2(0) = x_2) \\ &\geq (\leq) P(X_1(t_1) > y_1 | X_1(0) = x_1) P(X_2(t_2) > y_2 | X_2(0) = x_2) \end{aligned} \tag{10}$$

for all $t_1, t_2 \geq 0$ and for all $y_1, y_2, x_1, x_2 \in \mathbb{R}$.

It is well known (cf. Ebrahimi [27]) that $(X_1(t), X_2(t))$ *being positively (negatively) associated implies that* $X_1(t)$ *and* $X_2(t)$ *are positively (negatively) quadrant-dependent.*

Let $\vec{B}(t) = (B_1(t), B_2(t))^\tau$ *be a standard bidimensional Brownian motion with constant correlation coefficient* $\rho \in (-1, 1)$. *For notional convenience, for* $t \geq 0$ *we write* $\underline{B}_i(t) = \inf_{0\leq s\leq t} B_i(s)$, $\overline{B}_i(t) = \sup_{0\leq s\leq t} B_i(s)$, $i = 1, 2$. *It is well known that* $P(\underline{B}_i(t) < -x) = P(\overline{B}_i(t) > x) = 2P(B_i(t) > x)$ *for* $x > 0$. *The following lemma is essential to proving our main results. Moreover, it is of independent interest.*

Lemma 5. *For any* $x_1 > 0, x_2 > 0$, *if* $\rho \in [0, 1)$, *then*

$$P(\overline{B}_1(t) > x_1, \overline{B}_2(t) > x_2) \geq P(\overline{B}_1(t) > x_1)P(\overline{B}_2(t) > x_2), \tag{11}$$

and

$$P(\underline{B}_1(t) < -x_1, \underline{B}_2(t) < -x_2) \geq P(\underline{B}_1(t) < -x_1)P(\underline{B}_2(t) < -x_2); \tag{12}$$

If $\rho \in (-1, 0]$, *then*

$$P(\overline{B}_1(t) > x_1, \overline{B}_2(t) > x_2) \leq P(\overline{B}_1(t) > x_1)P(\overline{B}_2(t) > x_2), \tag{13}$$

and

$$P(\underline{B}_1(t) < -x_1, \underline{B}_2(t) < -x_2) \leq P(\underline{B}_1(t) < -x_1)P(\underline{B}_2(t) < -x_2). \tag{14}$$

Proof. For any $t_1, t_2 \geq 0$, we have $\mathrm{Cov}(B_1(t_1), B_2(t_2)) = \rho \min\{t_1, t_2\}$. It follows from the Theorem in Pitt [28] that $\rho \geq 0$ is necessary and sufficient for $(B_1(t), B_2(t))^\tau$ to be positively associated, as $(B_1(t_1), B_2(t_2))^\tau$ is bivariate normal, which implies that $(B_1(t), B_2(t))^\tau$ is positively quadrant-dependent. Thus, (11) holds. To prove (12), we use (11) and the facts that $-\sup_{0\leq s\leq t} B_i(s) = \inf_{0\leq s\leq t}(-B_i(s))$ and $(-B_1(t), -B_2(t))^\tau$ is a standard bidimensional Brownian motion with correlation coefficient ρ. Inequalities (13) and (14) can be proved similarly. This completes the proof.

For $r \geq 0$, consider a bidimensional Gaussian process $(\int_0^t e^{-rs} dB_1(s), \int_0^t e^{-rs} dB_2(s))^\tau$, where $\vec{B}(t) = (B_1(t), B_2(t))^\tau$ is a standard bidimensional Brownian motion with constant correlation coefficient $\rho \in (-1, 1)$. For $t \geq 0$, we can write

$$\underline{\Delta}_i(t) = \inf_{0\leq s\leq t} \int_0^s e^{-rl} dB_1(l),\ \overline{\Delta}_i(t) = \sup_{0\leq s\leq t} \int_0^s e^{-rl} dB_2(l),\ i = 1, 2.$$

The following lemma is an extension of Lemma 5. □

Lemma 6. *For any* $x_1 > 0, x_2 > 0$, *if* $\rho \in [0, 1)$, *then*

$$P\big(\overline{\Delta}_1(t) > x_1, \overline{\Delta}_2(t) > x_2\big) \geq P\big(\overline{\Delta}_1(t) > x_1\big)P\big(\overline{\Delta}_2(t) > x_2\big),$$

and

$$P(\underline{\Delta}_1(t) < -x_1, \underline{\Delta}_2(t) < -x_2) \geq P(\underline{\Delta}_1(t) < -x_1)P(\underline{\Delta}_2(t) < -x_2);$$

If $\rho \in (-1, 0]$, *then*

$$P\big(\overline{\Delta}_1(t) > x_1, \overline{\Delta}_2(t) > x_2\big) \leq P\big(\overline{\Delta}_1(t) > x_1\big)P\big(\overline{\Delta}_2(t) > x_2\big),$$

and

$$P(\underline{\Delta}_1(t) < -x_1, \underline{\Delta}_2(t) < -x_2) \leq P(\underline{\Delta}_1(t) < -x_1)P(\underline{\Delta}_2(t) < -x_2).$$

Remark 1. *Several distributions of interest are available in closed form (see, e.g., He, Keirstead, and Rebholz [29]). These include the joint distributions of* $(\underline{X}_1(t), \underline{X}_2(t))$, $(\overline{X}_1(t), \overline{X}_2(t))$, $(\underline{X}_1(t), \overline{X}_1(t))$, *and so on. However, those closed-form results cannot apply our proofs to the main results. The results of Lemmas 5 and 6 cannot be obtained from the results of Shao and Wang [30].*

Lemma 7. *Let $\{N(t), t \geq 0\}$ be a Poisson process with arrival times $\tau_k, k = 1, 2, \cdots$. Considering $N(T) = n$ for arbitrarily fixed $T > 0$ and $n = 1, 2, \cdots$, the random vector $(\tau_1, \cdots, \tau_n)$ is equal in distribution to the random vector $(TU_{(1,n)}, \cdots, TU_{(n,n)})$, where $U_{(1,n)}, \cdots, U_{(n,n)}$ denote the order statistics of n i.i.d. (0, 1) uniformly distributed random variables $U_1, \cdots, U_n$.*

Proof. See Theorem 2.3.1 of Ross [26]. □

Lemma 8. *Let X and Y be two independent and non-negative random variables. If X is subexponentially distributed while Y is bounded and non-degenerate at 0, then the product XY is subexponentially distributed.*

Proof. See Corollary 2.3 of Cline and Samorodnitsky [19]. □

The following result is due to Tang [1].

Lemma 9. *Let X and Y be two independent random variables with distributions F_X and F_Y. Moreover, let Y be non-negative and non-degenerate at 0. Then,*

$$F_{X-Y} \in \mathcal{L} \Leftrightarrow F_X \in \mathcal{L} \Leftrightarrow \overline{F}_{X-Y}(x) \sim \overline{F}_X(x).$$

4. Main Results and Proofs

In this paper, we establish new results for the finite-time ruin probabilities. Unlike the above-motioned articles, we assume that the two Brownian motions $\{B_1(t), t \geq 0\}$ and $\{B_2(t), t \geq 0\}$ are correlated with a constant correlation coefficient $\rho \in (-1, 1)$. The following are the main results of this paper.

Theorem 1. *Consider the insurance risk model introduced in Section 1. Assume that $N_1(t) = N_2(t) = N(t)$, $\rho \in (-1, 0]$, $r = 0$ and that $\{X_{1k}, k = 1, 2, \cdots\}$, $\{X_{2k}, k = 1, 2, \cdots\}$, $\{C_1(t), t \geq 0\}$, $\{C_2(t), t \geq 0\}$, $\{N(t), t \geq 0\}$, $\{(B_1(t), B_2(t)), t \geq 0\}$ are mutually independent.*

(a) If $F_1, F_2 \in \mathcal{S}$, then, for each fixed time $T \geq 0$,

$$\psi_{\max}(\vec{u}; T) \sim \lambda T(1 + \lambda T)\overline{F_1}(u_1)\overline{F_2}(u_2), \text{ as } (u_1, u_2) \to (\infty, \infty), \tag{15}$$

$$\psi_{\min}(\vec{u}; T) \sim \lambda T\big(\overline{F_1}(u_1) + \overline{F_2}(u_2)\big), \text{ as } (u_1, u_2) \to (\infty, \infty). \tag{16}$$

(b) If $F_1 * F_2 \in \mathcal{S}$, then, for each fixed time $T \geq 0$,

$$\psi_{\text{sum}}(\vec{u}; T) \sim \lambda T\big(\overline{F_1}(u_1 + u_2) + \overline{F_2}(u_1 + u_2)\big), \text{ as } u_1 + u_2 \to \infty. \tag{17}$$

Proof. First, we establish the asymptotic upper bound for $\psi_{\max}(\vec{u}; T)$. Clearly,

$$\begin{aligned}
\psi_{\max}(\vec{u}; T) &\leq P\left(\sum_{i=1}^{N(T)} \vec{X}_i - \begin{pmatrix} \sigma_1 \underline{B}_1(T) \\ \sigma_2 \underline{B}_2(T) \end{pmatrix} > \vec{u}\right) \\
&= \sum_{n=0}^{\infty} P(N(T) = n) P\left(\sum_{i=1}^{n} \vec{X}_i - \begin{pmatrix} \sigma_1 \underline{B}_1(T) \\ \sigma_2 \underline{B}_2(T) \end{pmatrix} > \vec{u}\right) \\
&= \sum_{n=0}^{\infty} P(N(T) = n) \int_0^{\infty} \int_0^{\infty} P(\sum_{i=1}^{n} \vec{X}_i \in d\vec{z}) \\
&\quad \times P\left(\vec{z} - \begin{pmatrix} \sigma_1 \underline{B}_1(T) \\ \sigma_2 \underline{B}_2(T) \end{pmatrix} > \vec{u}\right).
\end{aligned} \tag{18}$$

Because $\rho \in (-1, 0]$, by using (14) we have

$$\begin{aligned} & P\left(\vec{z} - \begin{pmatrix} \sigma_1 \underline{B}_1(T) \\ \sigma_2 \underline{B}_2(T) \end{pmatrix} > \vec{u}\right) \\ & \leq P(z_1 - \sigma_1 \underline{B}_1(T) > u_1) P(z_2 - \sigma_2 \underline{B}_2(T) > u_2). \end{aligned} \tag{19}$$

Using the independence of $\{X_{1k}, k = 1, 2, \cdots\}$ and $\{X_{2k}, k = 1, 2, \cdots\}$, we have

$$P\left(\sum_{i=1}^{n} \vec{X}_i \in \vec{z}\right) = P\left(\sum_{i=1}^{n} X_{1i} \in dz_1\right) P\left(\sum_{i=1}^{n} X_{2i} \in dz_2\right). \tag{20}$$

Substituting (19) and (20) into (18) and using the dominated convergence theorem, we obtain

$$\begin{aligned} \psi_{\max}(\vec{u}; T) &\leq \sum_{n=0}^{\infty} P(N(T) = n) P\left(\sum_{i=1}^{n} X_{1i} - \sigma_1 \underline{B}_1(T) > u_1\right) P\left(\sum_{i=1}^{n} X_{2i} - \sigma_2 \underline{B}_2(T) > u_1\right) \\ &\sim \sum_{n=0}^{\infty} P(N(T) = n) n^2 \overline{F_1}(u_1) \overline{F_2}(u_2) \\ &= \lambda T(1 + \lambda T) \overline{F_1}(u_1) \overline{F_2}(u_2), \text{ as } (u_1, u_2) \to (\infty, \infty), \end{aligned} \tag{21}$$

where in the second step we have used Lemma 2 and the fact that

$$P\left(\sigma_j \sup_{0 \leq t \leq T} (-B_j(t)) \geq u_j\right) = o\left(P(\sum_{i=1}^{n} X_{ji} \geq u_j)\right), \ j = 1, 2.$$

Next, we establish the asymptotic lower bound for $\psi_{\max}(\vec{u}; T)$. Clearly,

$$\begin{aligned} \psi_{\max}(\vec{u}; T) &\geq P\left(\sum_{i=1}^{N(T)} \vec{X}_i - \vec{C}(T) - \begin{pmatrix} \sigma_1 \overline{B}_1(T) \\ \sigma_2 \overline{B}_2(T) \end{pmatrix} > \vec{u}\right) \\ &= \sum_{n=0}^{\infty} P(N(T) = n) P\left(\sum_{i=1}^{n} \vec{X}_i - \begin{pmatrix} \sigma_1 \overline{B}_1(T) \\ \sigma_2 \overline{B}_2(T) \end{pmatrix} - \vec{C}(T) > \vec{u}\right) \\ &\equiv \sum_{n=0}^{\infty} P(N(T) = n) I_1, \end{aligned} \tag{22}$$

where I_1 can be written as

$$I_1 = \int_0^{\infty} \int_0^{\infty} P(\overline{B_1}(T) \in dy_1, \overline{B_2}(T) \in dy_2) J_1 J_2. \tag{23}$$

Here,

$$J_1 = P\left(\sum_{i=1}^{n} X_{1i} - C_1(T) - \sigma_1 y_1 > u_1\right),$$

and

$$J_2 = P\left(\sum_{i=1}^{n} X_{2i} - C_2(T) - \sigma_2 y_2 > u_2\right).$$

For large constants $a > 0$ and $b > 0$, we can further write I_1 as

$$\begin{aligned} I_1 &= \left(\int_0^a \int_0^b + \int_0^a \int_b^{\infty} + \int_a^{\infty} \int_0^b + \int_a^{\infty} \int_b^{\infty}\right) P(\overline{B_1}(T) \in dy_1, \overline{B_2}(T) \in dy_2) J_1 J_2 \\ &\equiv k_1 + k_2 + k_3 + k_4. \end{aligned} \tag{24}$$

First, we consider k_1. Then, per Lemma 9, it holds uniformly for all $y_1 \in [0, a]$ that

$$J_1 \sim n\overline{F}_1(u_1), \text{ as } u_1 \to \infty \tag{25}$$

and it holds uniformly for all $y_2 \in [0, b]$ that

$$J_2 \sim n\overline{F}_2(u_2), \text{ as } u_2 \to \infty. \tag{26}$$

Using Lemma 1 and the dominated convergence theorem, we obtain

$$k_1 \sim n^2\overline{F}_1(u_1)\overline{F}_2(u_2)\int_0^a \int_0^b P(\overline{B_1}(T) \in dy_1, \overline{B_2}(T) \in dy_2), \text{ as } (u_1, u_2) \to (\infty, \infty).$$

Thus,

$$\lim_{(a,b)\to(\infty,\infty)} \lim_{(u_1,u_2)\to(\infty,\infty)} \frac{k_1}{n^2\overline{F}_1(u_1)\overline{F}_2(u_2)} = 1. \tag{27}$$

Now, we consider k_2. Using (25), Lemma 1, and the dominated convergence theorem,

$$\begin{aligned} k_2 &\sim n\overline{F}_1(u_1)\int_0^a \int_b^\infty P(\overline{B_1}(T) \in dy_1, \overline{B_2}(T) \in dy_2) J_2 \\ &\leq n\overline{F}_1(u_1) P\left(\sum_{i=1}^n X_{2i} - C_2(T) - \sigma_2 b > u_2\right) \int_0^a \int_b^\infty P(\overline{B_1}(T) \in dy_1, \overline{B_2}(T) \in dy_2) \\ &\sim n^2\overline{F}_1(u_1)\overline{F}_2(u_2)\int_0^a \int_b^\infty P(\overline{B_1}(T) \in dy_1, \overline{B_2}(T) \in dy_2), \text{ as } (u_1, u_2) \to (\infty, \infty). \end{aligned}$$

Thus,

$$\lim_{(a,b)\to(\infty,\infty)} \lim_{(u_1,u_2)\to(\infty,\infty)} \frac{k_2}{n^2\overline{F}_1(u_1)\overline{F}_2(u_2)} = 0. \tag{28}$$

Likewise,

$$\lim_{(a,b)\to(\infty,\infty)} \lim_{(u_1,u_2)\to(\infty,\infty)} \frac{k_3}{n^2\overline{F}_1(u_1)\overline{F}_2(u_2)} = 0. \tag{29}$$

Finally, we deal with k_4:

$$\begin{aligned} k_4 &\leq P\left(\sum_{i=1}^n X_{1i} - C_1(T) - \sigma_1 a > u_1\right) P\left(\sum_{i=1}^n X_{2i} - C_2(T) - \sigma_2 b > u_2\right) \\ &\quad \times \int_a^\infty \int_b^\infty P(\overline{B_1}(T) \in dy_1, \overline{B_2}(T) \in dy_2) \\ &\sim n^2\overline{F}_1(u_1)\overline{F}_2(u_2)\int_a^\infty \int_b^\infty P(\overline{B_1}(T) \in dy_1, \overline{B_2}(T) \in dy_2), \text{ as } (u_1, u_2) \to (\infty, \infty), \end{aligned}$$

from which we obtain

$$\lim_{(a,b)\to(\infty,\infty)} \lim_{(u_1,u_2)\to(\infty,\infty)} \frac{k_4}{n^2\overline{F}_1(u_1)\overline{F}_2(u_2)} = 0. \tag{30}$$

From (23) and (27)–(30), we obtain

$$\lim_{(u_1,u_2)\to(\infty,\infty)} \frac{I_1}{n^2\overline{F}_1(u_1)\overline{F}_2(u_2)} = 1. \tag{31}$$

Now, it follows from (22), (31), and the dominated convergence theorem that

$$\lim_{(u_1,u_2)\to(\infty,\infty)} \frac{\psi_{\max}(\vec{u}; T)}{\lambda T(1 + \lambda T)\overline{F_1}(u_1)\overline{F_2}(u_2)} \geq 1,$$

from which, along with (21), we obtain (15).

Note that

$$\psi_{\text{and}}(\vec{u};T) \le P\left(\sum_{i=1}^{N(T)} X_{1i} - \sigma_1 \underline{B}_1(T) > u_1, \sum_{i=1}^{N(T)} X_{2i} - \sigma_2 \underline{B}_2(T) > u_2\right),$$

from which, along with (18) and (21), we have

$$\lim_{(u_1,u_2)\to(\infty,\infty)} \frac{\psi_{\text{and}}(\vec{u};T)}{\overline{F_1}(u_1)+\overline{F_2}(u_2)} \le \lim_{(u_1,u_2)\to(\infty,\infty)} \frac{\lambda T(1+\lambda T)\overline{F_1}(u_1)\overline{F_2}(u_2)}{\overline{F_1}(u_1)+\overline{F_2}(u_2)} = 0.$$

Thus, it is the case that $\psi_{and}(\vec{u};T) \sim 0$, as $(u_1,u_2) \to (\infty,\infty)$. From (6), we have

$$\psi_{min}(\vec{u};T) \sim P(T_1 \le T \mid U_1(0) = u_1) + P(T_2 \le T \mid U_2(0) = u_1) = \psi_1(u_1;T) + \psi_2(u_2;T).$$

From Lemma 3, we can obtain (16).

Next, we prove relation (17). Using Theorem 7.2 in Ikeda and Watanabe [31] (and see Yin and Wen [32]), for all $t \ge 0$ we have

$$\sqrt{\sigma_1^2+\sigma_2^2+2\rho\sigma_1\sigma_2}W(t) \stackrel{d}{=} \sigma_1 B_1(t) + \sigma_2 B_2(t),$$

where '$\stackrel{d}{=}$' denotes equality in distribution, W is a standard Brownian motion independent of $\{X_{1k}, k = 1,2,\cdots\}$, $\{X_{2k}, k = 1,2,\cdots\}$, $\{C_1(t), t \ge 0\}$, $\{C_2(t), t \ge 0\}$, and $\{N(t), t \ge 0\}$. Thus, for all $t \ge 0$, $U_1(t) + U_2(t)$ can be written as

$$U_1(t) + U_2(t) \stackrel{d}{=} u_1 + u_2 + C_1(t) + C_2(t) - \sum_{i=1}^{N(t)} (X_{1i} + X_{2i}) + \sqrt{\sigma_1^2+\sigma_2^2+2\rho\sigma_1\sigma_2}W(t).$$

Applying Lemma 3 to this model, we find that if $F_1 * F_2 \in \mathcal{S}$, then

$$\psi_{\text{sum}}(\vec{u};T) \sim \lambda T\overline{F_1 * F_2}(u_1+u_2) \sim \lambda T(\overline{F_1}(u_1+u_2) + \overline{F_2}(u_1+u_2)),\ u_1+u_2 \to \infty,$$

where, in the last step, we have relied on the statement in [33] (and see Geluk and Tang [34]) that

$$F_1 * F_2 \in \mathcal{S} \text{ if and only if } P(X_1+X_2 > x) \sim \overline{F_1}(x) + \overline{F_2}(x).$$

This ends the proof of Theorem 1. □

Remark 2. *Letting $\{C_i(t) = c_i t, i = 1,2$ and $\rho = 0$ in Theorem 1, we obtain Theorem 1 in [12].*

Theorem 2. *Consider the insurance risk model introduced in Section 1. Assume that $N_1(t) = N_2(t) = N(t)$, $\rho \in (-1,0]$, $r > 0$ and that $\{X_{1k}, k = 1,2,\cdots\}$, $\{X_{2k}, k = 1,2,\cdots\}$, $\{C_1(t), t \ge 0\}$, $\{C_2(t), t \ge 0\}$, $\{N(t), t \ge 0\}$, $\{(B_1(t), B_2(t)), t \ge 0\}$ are mutually independent.*

(a) If $F_1, F_2 \in \mathcal{S}$, then for each fixed time $T \ge 0$,

$$\psi_{\max}(\vec{u};T) \sim \frac{\lambda(\lambda+\frac{1}{T})}{r^2}\int_{u_1}^{u_1e^{rT}} \frac{\overline{F_1}(y)}{y}dy \int_{u_2}^{u_2e^{rT}} \frac{\overline{F_2}(y)}{y}dy,\ \text{as}\ (u_1,u_2) \to (\infty,\infty), \quad (32)$$

$$\psi_{\min}(\vec{u};T) \sim \frac{\lambda}{r}\left(\int_{u_1}^{u_1e^{rT}} \frac{\overline{F_1}(y)}{y}dy + \int_{u_2}^{u_2e^{rT}} \frac{\overline{F_2}(y)}{y}dy\right),\ \text{as}\ (u_1,u_2) \to (\infty,\infty). \quad (33)$$

(b) If $F_1 * F_2 \in \mathcal{S}$, then for each fixed time $T \ge 0$,

$$\psi_{\text{sum}}(\vec{u};T) \sim \lambda T\int_0^1 \overline{F_1 * F_2}(e^{rTz}(u_1+u_2))dz,\ \text{as}\ u_1+u_2 \to \infty. \quad (34)$$

In particular, if there are two positive constants l_1 and l_2 such that $\overline{F}_i(x) \sim l_i\overline{F}(x)$, $i = 1, 2$, then

$$\psi_{\text{sum}}(\vec{u}; T) \sim \lambda T\left(\int_0^1 \overline{F_1}(e^{rTz}(u_1+u_2)) + \int_0^1 \overline{F_2}(e^{rTz}(u_1+u_2))\right), \text{ as } u_1+u_2 \to \infty. \quad (35)$$

Proof. We can write $\psi_{\max}(\vec{u}; T)$ as

$$\psi_{\max}(\vec{u}; T) = P(e^{-rt}U_i(t) < 0, i = 1, 2 \text{ for some } 0 < t \le T | \vec{U}(0) = \vec{u}).$$

For $t \in [0, T]$ and each $i = 1$ or 2, we have

$$\begin{aligned} u_i - \int_0^t e^{-rs}dS_i(s) + \sigma_i\int_0^t e^{-rs}dB_i(s) \quad &\le \quad e^{-rt}U_i(t) \le u_i + \int_0^T e^{-rs}dC_i(s) \\ &\quad - \int_0^t e^{-rs}dS_i(s) + \sigma_i\int_0^t e^{-rs}dB_i(s). \end{aligned}$$

It follows that $\psi_{\max}(\vec{u}; T)$ satisfies

$$\begin{aligned} \psi_{\max}(\vec{u}; T) \quad &\le \quad P\left(\sum_{i=1}^{N(T)} \vec{X}_i e^{-r\tau_i} - \begin{pmatrix}\sigma_1\underline{\Delta}_1(T)\\ \sigma_2\underline{\Delta}_2(T)\end{pmatrix} > \vec{u}\right)) \\ &\le \quad \sum_{n=0}^{\infty} P(N(T) = n)P\left(\sum_{i=1}^{n} \vec{X}_i e^{-r\tau_i} - \begin{pmatrix}\sigma_1\underline{\Delta}_1(T)\\ \sigma_2\underline{\Delta}_2(T)\end{pmatrix} > \vec{u} \middle| N(t) = n\right) \\ &\le \quad \sum_{n=0}^{\infty} P(N(T) = n)\int_0^{\infty}\int_0^{\infty} P\left(\sum_{i=1}^{n} \vec{X}_i e^{-rTU_i} \in d\vec{z}\right) \\ &\quad \times P\left(\vec{z} - \begin{pmatrix}\sigma_1\underline{\Delta}_1(T)\\ \sigma_2\underline{\Delta}_2(T)\end{pmatrix} > \vec{u}\right). \end{aligned} \quad (36)$$

where we have used Lemma 7 in the last steps. Because $\rho \in (-1, 0]$, using Lemma 6, we have

$$\begin{aligned} &P\left(\vec{z} - \begin{pmatrix}\sigma_1\underline{\Delta}_1(T)\\ \sigma_2\underline{\Delta}_2(T)\end{pmatrix} > \vec{u}\right) \\ &\le P(z_1 - \sigma_1\underline{\Delta}_1(T) > u_1)P(z_2 - \sigma_2\underline{\Delta}_2(T) > u_2). \end{aligned} \quad (37)$$

Using the independence of $\{X_{1k}, k = 1, 2, \cdots\}$ and $\{X_{2k}, k = 1, 2, \cdots\}$, we have

$$\begin{aligned} P\left(\sum_{i=1}^{n} \vec{X}_i e^{-rTU_i} \in d\vec{z}\right) \quad &= \quad \int_0^1 \cdots \int_0^1 P\left(\sum_{i=1}^{n} X_{1i}e^{-rTv_i} \in dz_1\right)P\left(\sum_{i=1}^{n} X_{2i}e^{-rTv_i} \in dz_2\right) \\ &\quad \times \prod_{j=1}^{n} P(U_j \in dv_j). \end{aligned} \quad (38)$$

Substituting (37) and (38) into (36) and using

$$P\left(\sum_{i=1}^{n} X_{1i}e^{-rTv_i} - \sigma_1\underline{\Delta}_1(T) > u_1\right) \sim P\left(\sum_{i=1}^{n} X_{1i}e^{-rTv_i} > u_1\right), u_1 \to \infty,$$

and

$$P\left(\sum_{i=1}^{n} X_{2i}e^{-rTv_i} - \sigma_2\underline{\Delta}_2(T) > u_2\right) \sim P\left(\sum_{i=1}^{n} X_{2i}e^{-rTv_i} > u_2\right), u_2 \to \infty,$$

uniformly for $(v_1, \cdots, v_n) \in [0,1]^n$, we obtain

$$\begin{aligned}\psi_{\max}(\vec{u};T) &\lesssim \sum_{n=0}^{\infty} P(N(T)=n) P\left(\sum_{i=1}^{n} X_{1i} e^{-rTU_i} > u_1, \sum_{i=1}^{n} X_{2i} e^{-rTU_i} > u_2\right) \\ &\equiv \sum_{n=0}^{\infty} P(N(T)=n) k_5. \end{aligned} \tag{39}$$

We apply Proposition 5.1 of Tang and Tsitsiashvili [22], which says that for i.i.d. subexponential random variables $\{X_k\}$ and for arbitrarily a and b where $0 < a \le b < \infty$, the relation

$$P\left(\sum_{i=1}^{n} c_i X_i > x\right) \sim \sum_{i=1}^{n} P(c_i X_i > x)$$

holds uniformly for $(c_1, \cdots, c_n) \in [a,b] \times \cdots \times [a,b]$. Hence, by conditioning on $(U_1, \cdots, U_n)$, we find that where

$$k_5 \sim n^2 P\Big(X_{11} e^{-rTU_1} > u_1\Big) P\Big(X_{21} e^{-rTU_1} > u_2\Big), \tag{40}$$

by substituting (40) into (39) and using the dominated convergence theorem, we obtain

$$\limsup_{(u_i,u_2)\to(\infty,\infty)} \frac{\psi_{\max}(\vec{u};T)}{\frac{\lambda(\lambda+\frac{1}{\gamma})}{r^2} \int_{u_1}^{u_1 e^{rT}} \frac{\overline{F_1}(y)}{y} dy \int_{u_2}^{u_2 e^{rT}} \frac{\overline{F_2}(y)}{y} dy} \le 1. \tag{41}$$

Next, we establish the asymptotic lower bound for $\psi_{\max}(\vec{u};T)$. Clearly,

$$\begin{aligned}\psi_{\max}(\vec{u};T) &\ge P\left(\sum_{i=1}^{N(T)} \vec{X}_i e^{-r\tau_i} - \int_0^T e^{-rs} d\vec{C}(s) - \begin{pmatrix}\sigma_1 \overline{\Delta}_1(T) \\ \sigma_2 \overline{\Delta}_2(T)\end{pmatrix} > \vec{u}\right) \\ &= \sum_{n=0}^{\infty} P(N(T)=n) P\left(\sum_{i=1}^{n} \vec{X}_i e^{-rTU_i} - \begin{pmatrix}\sigma_1 \overline{\Delta}_1(T) \\ \sigma_2 \overline{\Delta}_2(T)\end{pmatrix} - \int_0^T e^{-rs} d\vec{C}(s) > \vec{u}\right) \\ &\equiv \sum_{n=0}^{\infty} P(N(T)=n) I_2, \end{aligned} \tag{42}$$

where, for some positive constants c and d,

$$I_2 = \left(\int_0^c \int_0^d + \int_0^c \int_d^{\infty} + \int_c^{\infty} \int_0^d + \int_c^{\infty} \int_d^{\infty}\right) P(\overline{\Delta_1}(T) \in dy_1, \overline{\Delta_2}(T) \in dy_2) J_3 J_4.$$

Here,

$$J_3 = P\left(\sum_{i=1}^{n} X_{1i} e^{-rTU_i} - \int_0^T e^{-rs} dC_1(s) - \sigma_1 y_1 > u_1\right),$$

and

$$J_4 = P\left(\sum_{i=1}^{n} X_{2i} e^{-rTU_i} - \int_0^T e^{-rs} dC_2(s) - \sigma_2 y_2 > u_2\right).$$

Per Lemma 8, we know that $\sum_{i=1}^{n} X_{ji} e^{-rTU_i} \in \mathcal{S}, j = 1,2$, as all $X_{ji} \in \mathcal{S}$. Then, invoking Lemma 9, we obtain

$$J_3 \sim nP(X_{11} e^{-rTU_1} > u_1), \text{ as } u_1 \to \infty, \; J_4 \sim nP(X_{21} e^{-rTU_1} > u_2) \text{ as } u_2 \to \infty$$

uniformly for all $y_1 \in [0,c]$ and $y_2 \in [0,d]$, respectively. Now, using the same argument by which we reached (31), we have

$$\lim_{(u_1,u_2)\to(\infty,\infty)} \frac{I_2}{n^2 P(X_{11} e^{-rTU_1} > u_1) P(X_{21} e^{-rTU_1} > u_2)} = 1. \tag{43}$$

Now, it follows from (42), (43), Lemma 1, and the dominated convergence theorem that

$$\lim_{(u_i,u_2)\to(\infty,\infty)} \frac{\psi_{\max}(\vec{u};T)}{\lambda T(1+\lambda T)P(X_{11}e^{-rTU_1}>u_1)P(X_{21}e^{-rTU_1}>u_2)} \geq 1,$$

or, equivalently,

$$\lim_{(u_i,u_2)\to(\infty,\infty)} \frac{\psi_{\max}(\vec{u};T)}{\frac{\lambda(\lambda+\frac{1}{T})}{r^2}\int_{u_1}^{u_1e^{rT}}\frac{\overline{F_1}(y)}{y}dy\int_{u_2}^{u_2e^{rT}}\frac{\overline{F_2}(y)}{y}dy} \geq 1,$$

from which, along with (41), we obtain (32).

The relation (33) follows from (6) and Lemma 4 because, as above,

$$\begin{aligned}\lim_{(u_1,u_2)\to(\infty,\infty)} & \frac{\psi_{\text{and}}(\vec{u};T)}{\int_{u_1}^{u_1e^{rT}}\frac{\overline{F_1}(y)}{y}dy+\int_{u_2}^{u_2e^{rT}}\frac{\overline{F_2}(y)}{y}dy} \\ &\leq \lim_{(u_1,u_2)\to(\infty,\infty)} \frac{\frac{\lambda(\lambda+\frac{1}{T})}{r^2}\int_{u_1}^{u_1e^{rT}}\frac{\overline{F_1}(y)}{y}dy\int_{u_2}^{u_2e^{rT}}\frac{\overline{F_2}(y)}{y}dy}{\int_{u_1}^{u_1e^{rT}}\frac{\overline{F_1}(y)}{y}dy+\int_{u_2}^{u_2e^{rT}}\frac{\overline{F_2}(y)}{y}dy}=0.\end{aligned}$$

From (6), we have

$$\psi_{min}(\vec{u};T)\sim\psi_1(u_1;T)+\psi_2(u_2;T),\ \text{as}\ (u_1,u_2)\to(\infty,\infty).$$

From Lemma 4, we have

$$\psi_i(u_i;T)\sim\frac{\lambda}{r}\int_{u_i}^{u_ie^{rT}}\frac{\overline{F_i}(y)}{y}dy,\ u_i\to\infty,\ i=1,2.$$

Then,

$$\psi_{min}(\vec{u};T)\sim\frac{\lambda}{r}\left(\int_{u_1}^{u_1e^{rT}}\frac{\overline{F_1}(y)}{y}dy+\int_{u_2}^{u_2e^{rT}}\frac{\overline{F_2}(y)}{y}dy\right),\ \text{as}\ (u_1,u_2)\to(\infty,\infty).$$

Thus, we have completed the proof of (33).

Next, we prove relation (34). Similarly, for all $t\geq 0$, we have

$$\begin{aligned}U_1(t)+U_2(t) \ &\stackrel{d}{=}\ (u_1+u_2)e^{rt}+\int_0^t e^{r(t-s)}d(C_1(s)+C_2(s)) \\ &\quad-\int_0^t e^{r(t-s)}d\sum_{i=1}^{N(s)}(X_{1i}+X_{2i}) \\ &\quad+\sqrt{\sigma_1^2+\sigma_2^2+2\rho\sigma_1\sigma_2}\int_0^t e^{r(t-s)}dW(s), \qquad (44)\end{aligned}$$

where $\{W(t),t\geq 0\}$ is a standard Brownian motion independent of $\{X_{1k},k=1,2,\cdots\}$, $\{X_{2k},k=1,2,\cdots\}$, $\{C_1(t),t\geq 0\}$, $\{C_2(t),t\geq 0\}$, and $\{N(t),t\geq 0\}$.

From Lemma 4, we have

$$\psi_{sum}(\vec{u};T)\sim\frac{\lambda}{r}\int_{u_1+u_2}^{(u_1+u_2)e^{rT}}\frac{\overline{F_1*F_2}(y)}{y}dy,\ u_1+u_2\to\infty.$$

Let $y = (u_1 + u_2)e^{rTz}$; then, $dy = rT(u_1 + u_2)e^{rTz}dz$. Therefore,

$$\begin{aligned}\psi_{sum}(\vec{u};T) &\sim \frac{\lambda}{r}\int_0^1 \frac{\overline{F_1 * F_2}((u_1+u_2)e^{rTz})}{(u_1+u_2)e^{rTz}} rT(u_1+u_2)e^{rTz}dz \\ &= T\lambda \int_0^1 \overline{F_1 * F_2}((u_1+u_2)e^{rTz})dz, \text{as } (u_1,u_2)\to(\infty,\infty).\end{aligned}$$

This completes the proof of (34). The result (35) follows from (34) and Lemma 3.1 in [5]. This ends the proof of Theorem 2. □

Remark 3. *When letting* $\{C_i(t) = c_i t, i = 1,2, \rho = 0, \sigma_1 = 0, \sigma_2 = 0$ *in Theorem 2, we obtain the result in Liu et al. [23].*

Theorem 3. *Consider the insurance risk model introduced in Section 1. Assume that* $\rho \in (-1,0]$, $r = 0$ *and* $\{X_{1k}, k = 1,2,\cdots\}$, $\{X_{2k}, k = 1,2,\cdots\}$, $\{C_1(t), t \geq 0\}$, $\{C_2(t), t \geq 0\}$, $\{N_i(t), t \geq 0\}$, *and* $i = 1,2$, $\{(B_1(t), B_2(t)), t \geq 0\}$ *are mutually independent.*
(a) If $F_1, F_2 \in \mathcal{S}$, *then for each fixed time* $T \geq 0$,

$$\psi_{\max}(\vec{u};T) \sim \lambda_1\lambda_2 T^2 \overline{F_1}(u_1)\overline{F_2}(u_2), \text{ as } (u_1,u_2)\to(\infty,\infty), \tag{45}$$

$$\psi_{\min}(\vec{u};T) \sim T\big(\lambda_1\overline{F_1}(u_1) + \lambda_2\overline{F_2}(u_2)\big), \text{ as } (u_1,u_2)\to(\infty,\infty). \tag{46}$$

(b) If $F_{\xi X_{11}+(1-\xi)X_{21}} \in \mathcal{S}$, *where* ξ *is a random variable independent of* $\{X_{1k}, k = 1,2,\cdots\}$ *and* $\{X_{2k}, k = 1,2,\cdots\}$ *and* $P(\xi = 1) = 1 - P(\xi = 0) = \frac{\lambda_1}{\lambda_1+\lambda_2}$; *then, for each fixed time* $T \geq 0$,

$$\psi_{\text{sum}}(\vec{u};T) \sim T\big(\lambda_1\overline{F_1}(u_1+u_2) + \lambda_2\overline{F_2}(u_1+u_2)\big), \text{ as } u_1+u_2\to\infty. \tag{47}$$

Proof. As the proof is similar to that of Theorem 1, we only provide the main steps. First, we establish the asymptotic upper bound for $\psi_{\max}(\vec{u};T)$. Clearly,

$$\begin{aligned}\psi_{\max}(\vec{u};T) &\leq P\left(\begin{pmatrix}\sum_{i=1}^{N_1(T)} X_{1i}\\ \sum_{i=1}^{N_2(T)} X_{2i}\end{pmatrix} - \begin{pmatrix}\sigma_1\underline{B}_1(T)\\ \sigma_2\underline{B}_2(T)\end{pmatrix} > \begin{pmatrix}u_1\\u_2\end{pmatrix}\right) \\ &= \int_0^\infty\int_0^\infty P\left(\sum_{i=1}^{N_1(T)} X_{1i} \in dz_1\right) P\left(\sum_{i=1}^{N_2(T)} X_{2i} \in dz_2\right) \\ &\quad \times P\left(\begin{pmatrix}z_1\\z_2\end{pmatrix} - \begin{pmatrix}\sigma_1\underline{B}_1(T)\\ \sigma_2\underline{B}_2(T)\end{pmatrix} > \begin{pmatrix}u_1\\u_2\end{pmatrix}\right).\end{aligned} \tag{48}$$

Because $\rho \in (-1,0]$, using (14), we have

$$\begin{aligned}P &\left(\begin{pmatrix}z_1\\z_2\end{pmatrix} - \begin{pmatrix}\sigma_1\underline{B}_1(T)\\ \sigma_2\underline{B}_2(T)\end{pmatrix} > \begin{pmatrix}u_1\\u_2\end{pmatrix}\right) \\ &\leq P(z_1 - \sigma_1\underline{B}_1(T) > u_1)P(z_2 - \sigma_2\underline{B}_2(T) > u_2).\end{aligned} \tag{49}$$

Substituting (49) into (48), we obtain

$$\begin{aligned}\psi_{\max}(\vec{u};T) &\leq P\left(\sum_{i=1}^{N_1(T)} X_{1i} - \sigma_1\underline{B}_1(T) > u_1\right) P\left(\sum_{i=1}^{N_2(T)} X_{2i} - \sigma_2\underline{B}_2(T) > u_1\right) \\ &\sim \lambda_1\lambda_2 T^2 \overline{F_1}(u_1)\overline{F_2}(u_2), \text{ as } (u_1,u_2)\to(\infty,\infty),\end{aligned} \tag{50}$$

where in the last step we have used Lemma 3.

Next, we establish the asymptotic lower bound for $\psi_{\max}(\vec{u};T)$. Clearly,

$$\begin{aligned}\psi_{\max}(\vec{u};T) &\geq P\left(\begin{pmatrix}\sum_{i=1}^{N_1(T)} X_{1i}\\ \sum_{i=1}^{N_2(T)} X_{2i}\end{pmatrix} - \begin{pmatrix}C_1(T)\\ C_2(T)\end{pmatrix} - \begin{pmatrix}\sigma_1\overline{B}_1(T)\\ \sigma_2\overline{B}_2(T)\end{pmatrix} > \begin{pmatrix}u_1\\ u_2\end{pmatrix}\right)\\ &= \sum_{n=0}^{\infty} P(N_1(T)=n)\sum_{m=0}^{\infty} P(N_1(T)=m)I_3, \end{aligned} \tag{51}$$

where

$$I_3 = P\left(\begin{pmatrix}\sum_{i=1}^{n} X_{1i}\\ \sum_{i=1}^{m} X_{2i}\end{pmatrix} - \begin{pmatrix}C_1(T)\\ C_2(T)\end{pmatrix} - \begin{pmatrix}\sigma_1\overline{B}_1(T)\\ \sigma_2\overline{B}_2(T)\end{pmatrix} > \begin{pmatrix}u_1\\ u_2\end{pmatrix}\right).$$

Using the same arguments as those used to prove (31), we obtain

$$\lim_{(u_1,u_2)\to(\infty,\infty)} \frac{I_3}{nm\overline{F}_1(u_1)\overline{F}_2(u_2)} = 1,$$

from which, together with (51), we have

$$\lim_{(u_1,u_2)\to(\infty,\infty)} \frac{\psi_{\max}(\vec{u};T)}{\lambda_1\lambda_2 T^2\overline{F_1}(u_1)\overline{F_2}(u_2)} \geq 1.$$

The proof of (46) is straightforward, and is omitted here. Next, we prove (47). Using the properties of two independent compound Poisson processes and two independent Brownian motions, for all $t \geq 0$ we have

$$\begin{aligned}U_1(t)+U_2(t) \overset{d}{=}\ & u_1+u_2+C_1(t)+C_2(t)-\sum_{i=1}^{N_0(t)}(\xi X_{1i}+(1-\xi)X_{2i})\\ &+\sqrt{\sigma_1^2+\sigma_2^2+2\rho\sigma_1\sigma_2}W(t),\end{aligned}$$

where $\{W(t), t \geq 0\}$ is a standard Brownian motion, $\{N_0(t), t \geq 0\}$ is a Poisson process with intensity $\lambda_1+\lambda_2$, and ξ is a Bernoulli random variable with $P(\xi=1) = 1-P(\xi=0) = \frac{\lambda_1}{\lambda_1+\lambda_2}$. Moreover, ξ, $\{W(t), t \geq 0\}$, $\{N_0(t), t \geq 0\}$, $\{X_{1k}, k = 1,2,\cdots\}$, $\{X_{2k}, k = 1,2,\cdots\}$, $\{C_1(t), t \geq 0\}$, $\{C_2(t), t \geq 0\}$, and $\{N(t), t \geq 0\}$ are independent. Applying Lemma 3 to this model, we obtain

$$\psi_{\text{sum}}(\vec{u};T) \sim (\lambda_1+\lambda_2)T\overline{F}_{\xi X_{11}+(1-\xi)X_{21}}(u_1+u_2),\ u_1+u_2\to\infty,$$

and result (47) follows (c.f. Kaas et al. [35].)

$$P(\xi X_{11}+(1-\xi)X_{21} > u_1+u_2) = \frac{\lambda_1}{\lambda_1+\lambda_2}\overline{F}_1(u_1+u_2)+\frac{\lambda_2}{\lambda_1+\lambda_2}\overline{F}_2(u_1+u_2).$$

This ends the proof of Theorem 3. □

Theorem 4. *Consider the insurance risk model introduced in Section 1. Assume that* $\rho \in (-1,0]$, $r > 0$, *and that* $\{X_{1k}, k = 1,2,\cdots\}$, $\{X_{2k}, k = 1,2,\cdots\}$, $\{C_1(t), t \geq 0\}$, $\{C_2(t), t \geq 0\}$, $\{N_i(t), t \geq 0\}$, $i = 1,2$, *and* $\{(B_1(t), B_2(t)), t \geq 0\}$ *are mutually independent.*
(a) If $F_1, F_2 \in \mathcal{S}$, *then for each fixed time* $T \geq 0$,

$$\psi_{\max}(\vec{u};T) \sim \frac{\lambda_1\lambda_2}{r^2}\int_{u_1}^{u_1e^{rT}}\frac{\overline{F_1}(y)}{y}dy\int_{u_2}^{u_2e^{rT}}\frac{\overline{F_2}(y)}{y}dy,\ \text{as}\ (u_1,u_2)\to(\infty,\infty), \tag{52}$$

$$\psi_{\min}(\vec{u};T) \sim \frac{1}{r}\left(\lambda_1 \int_{u_1}^{u_1 e^{rT}} \frac{\overline{F_1}(y)}{y}dy + \lambda_2 \int_{u_2}^{u_2 e^{rT}} \frac{\overline{F_2}(y)}{y}dy\right), \text{ as } (u_1,u_2) \to (\infty,\infty). \quad (53)$$

(b) If $F_{\xi X_{11}+(1-\xi)X_{21}} \in \mathcal{S}$, where ξ is defined as in Theorem 3, then for each fixed time $T \geq 0$,

$$\psi_{\text{sum}}(\vec{u};T) \sim \frac{1}{r}\left(\lambda_1 \int_{u_1+u_2}^{(u_1+u_2)e^{rT}} \frac{\overline{F_1}(y)}{y}dy + \lambda_2 \int_{u_1+u_2}^{(u_1+u_2)e^{rT}} \frac{\overline{F_2}(y)}{y}dy\right), \text{ as } u_1+u_2 \to \infty. \quad (54)$$

Proof. As in the proof of Theorem 2, we have

$$\begin{aligned}
\psi_{\max}(\vec{u};T) &\leq P\left(\begin{pmatrix}\sum_{i=1}^{N_1(T)} X_{1i}e^{-r\tau_i}\\ \sum_{i=1}^{N_2(T)} X_{2i}e^{-r\tau_i}\end{pmatrix} - \begin{pmatrix}\sigma_1\underline{\Delta}_1(T)\\ \sigma_2\underline{\Delta}_2(T)\end{pmatrix} > \vec{u}\right)\\
&\leq \sum_{n=0}^{\infty} P(N_1(T)=n)\sum_{m=0}^{\infty} P(N_2(T)=m)\\
&\quad \times P\left(\begin{pmatrix}\sum_{i=1}^{n} X_{1i}e^{-r\tau_i}\\ \sum_{i=1}^{m} X_{2i}e^{-r\tau_i}\end{pmatrix} - \begin{pmatrix}\sigma_1\underline{\Delta}_1(T)\\ \sigma_2\underline{\Delta}_2(T)\end{pmatrix} > \vec{u}\right)\\
&\lesssim \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} nmP(N(T)=n)P(N_2(T)=m)P\Big(X_{11}e^{-rTU_1} > u_1\Big)P\Big(X_{21}e^{-rTU_1} > u_2\Big)\\
&= \lambda_1\lambda_2 T^2 P\Big(X_{11}e^{-rTU_1} > u_1\Big)P\Big(X_{21}e^{-rTU_1} > u_2\Big).
\end{aligned}$$

It follows that

$$\limsup_{(u_i,u_2)\to(\infty,\infty)} \frac{\psi_{\max}(\vec{u};T)}{\frac{\lambda_1\lambda_2}{r^2}\int_{u_1}^{u_1e^{rT}}\frac{\overline{F_1}(y)}{y}dy\int_{u_2}^{u_2e^{rT}}\frac{\overline{F_2}(y)}{y}dy} \leq 1.$$

The asymptotic lower bound for $\psi_{\max}(\vec{u};T)$ can be established similarly. The relation (53) follows from (6), Lemma 4, and the fact that

$$\begin{aligned}
&\lim_{(u_1,u_2)\to(\infty,\infty)} \frac{\psi_{\text{and}}(\vec{u};T)}{\lambda_1\int_{u_1}^{u_1e^{rT}}\frac{\overline{F_1}(y)}{y}dy + \lambda_2\int_{u_2}^{u_2e^{rT}}\frac{\overline{F_2}(y)}{y}dy}\\
&\leq \lim_{(u_1,u_2)\to(\infty,\infty)} \frac{\frac{\lambda_1\lambda_2}{r^2}\int_{u_1}^{u_1e^{rT}}\frac{\overline{F_1}(y)}{y}dy\int_{u_2}^{u_2e^{rT}}\frac{\overline{F_2}(y)}{y}dy}{\lambda_1\int_{u_1}^{u_1e^{rT}}\frac{\overline{F_1}(y)}{y}dy + \lambda_2\int_{u_2}^{u_2e^{rT}}\frac{\overline{F_2}(y)}{y}dy} = 0.
\end{aligned}$$

Finally, we prove (54). Using the same arguments as above, we have

$$\begin{aligned}
U_1(t)+U_2(t) &\stackrel{d}{=} (u_1+u_2)e^{rt} + \int_0^t e^{r(t-s)}d(C_1(s)+C_2(s))\\
&\quad - \int_0^t e^{r(t-s)}d\sum_{i=1}^{N_0(t)}(\xi X_{1i} + (1-\xi)X_{2i})\\
&\quad + \sqrt{\sigma_1^2+\sigma_2^2+2\rho\sigma_1\sigma_2}\int_0^t e^{r(t-s)}dW(s),\ t \geq 0, \quad (55)
\end{aligned}$$

where ξ, $\{W(t), t \geq 0\}$, $\{N_0(t), t \geq 0\}$ are the same as in the proof of Theorem 3. It follows from Lemma 4 that

$$\psi_{\text{sum}}(\vec{u};T) \sim \frac{\lambda_1+\lambda_2}{r}\int_{u_1+u_2}^{(u_1+u_2)e^{rT}} \frac{\overline{F}_{\xi X_{11}+(1-\xi)X_{21}}(y)}{y}dy,\ u_1+u_2 \to \infty,$$

and the result (54) follows, as

$$\overline{F}_{\xi X_{11}+(1-\xi)X_{21}}(y) = \frac{\lambda_1}{\lambda_1+\lambda_2}\overline{F}_1(y) + \frac{\lambda_2}{\lambda_1+\lambda_2}\overline{F}_2(y).$$

This completes the proof of Theorem 4. □

5. Conclusions

In this paper, we have investigated a bidimensional risk model that describes the surplus process of an insurer. We provide new results for the different types of finite-time ruin probabilities under the circumstance of that the Brownian motions are correlated with a constant correlation coefficient. We remark that the extension to multidimensional models is more complicated. However, multidimensional models can better describe different insurance businesses. In addition, we might consider the relationship between different businesses in the future research, which could be an even more interesting problem.

Author Contributions: Methodology, M.Z.; Writing—original draft, D.Z.; Writing—review & editing, C.Y. All authors have equally contributed to this paper. All authors have read and agreed to the published version of the manuscript.

Funding: The research was supported by the National Natural Science Foundation of China (No. 12071251); the Youth Innovation Team of Shandong Universities (Grant No. 2022KJ174).

Data Availability Statement: Not applicable.

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

References

1. Tang, Q.H. The ruin probability of a discrete time risk model under constant interest rate with heavy tails. *Scand. Actuar. J.* **2004**, *3*, 229–240. [CrossRef]
2. Tang, Q.H. The finite-time ruin probability of the compound Poisson model with constant interest force. *J. Appl. Probab.* **2005**, *42*, 608–619. [CrossRef]
3. Wang, D. Finite-time ruin probabiliy with heavy-tailed claims and constant interest rate. *Stoch. Model.* **2008**, *24*, 41–57. [CrossRef]
4. Wang, Y.; Cui, Z.; Wang, K.; Ma, X. Uniform asymptotics of the finite-time ruin probability for all times. *J. Math. Anal. Appl.* **2012**, *390*, 208–223. [CrossRef]
5. Hao, X.; Tang, Q.H. A uniform asymptotic estimate for discounted aggregate claims with subexponential tails. *Insur. Math. Econom.* **2008**, *43*, 116–120. [CrossRef]
6. Chen, Y.; Yang, Y.; Jiang, T. Uniform asymptotics for finite-time ruin probability of a bidimensional risk model. *J. Math. Anal. Appl.* **2019**, *469*, 525–536. [CrossRef]
7. Cheng, D. Uniform asymptotics for the finite-time ruin probability of a generalized bidimensional risk model with Brownian perturbations. *Stoch. Int. J. Probab. Stoch. Process.* **2019**, *93*, 1–16. [CrossRef]
8. Li, J.Z. The infinite-time ruin probability for a bidimensional renewal risk model with constant force of interest and dependent claims. *Commun. Stat. Theory Methods* **2017**, *46*, 1959–1971. [CrossRef]
9. Chan, W.; Yang, H.L.; Zhang, L. Some results on ruin probabilities in a two-dimensional risk model. *Insur. Math. Econom.* **2003**, *32*, 345–358. [CrossRef]
10. Avram, F.; Palmowski, Z.; Pistorius, M. A two-dismensional ruin problem on the positive quadrant. *Insur. Math. Econom.* **2008**, *42*, 227–234. [CrossRef]
11. Chen, Y.; Yuen, K.C.; Ng, K.W. Asymptotics for the ruin probabilities of a two-dimensional renewal risk model with heavy-tailed claim. *Appl. Stoch. Model. Bus. Ind.* **2011**, *27*, 290–300. [CrossRef]
12. Li, J.; Liu, Z.; Tang, Q.H. On the ruin probabilities of a bidimensional perturbed risk model. *Insur. Math. Econom.* **2007**, *41*, 85–195. [CrossRef]
13. Chen, Y.; Wang, L.; Wang, Y.B. Uniform asymptotics for the finite-time ruin probabilities of two kinds of nonstandard bidimensional risk models. *J. Math. Anal. Appl.* **2013**, *401*, 124–129. [CrossRef]
14. Yang, H.; Li, J. Asymptotic finite-time ruin probability for a bidimensional renewal risk model with constant interest force and dependent subexponential claims. *Insur. Math. Econom.* **2014**, *58*, 185–192. [CrossRef]
15. Lu, D.; Yuan, M. Asymptotic finite-time ruin probabilities for a bidimensional delay-claim risk model with subexponential claims. *Methodol. Comput. Appl. Probab.* **2022**, *24*, 2265–2286. [CrossRef]
16. Wang, S.; Qian, H.; Sun, H.; Geng, B. Uniform asymptotics for ruin probabilities of a non standard bidimensional perturbed risk model with subexponential claims. *Commun. Stat. Theory Methods* **2022**, *51*, 7871–7884. [CrossRef]
17. Asmussen, S.; Albrecher, H. *Ruin Probabilities*; World Scientific: Singapore, 2010.

18. Bingham, N.H.; Goldie, C.M.; Teugels, J.L. *Regular Variation*; Cambridge University Press: Cambridge, UK, 1987.
19. Cline, D.; Samorodnitsky, G. Subexponentiality of the product of independent random variables. *Stoch. Proc. Appl.* **1994**, *49*, 75–98. [CrossRef]
20. Ke, Y.; Minsker, S.; Ren, Z.; Sun, Q.; Zhou, W.X. User-friendly covariance estimation for heavy-tailed distributions. *Stat. Sci.* **2019**, *34*, 454–471. [CrossRef]
21. Konstantin, T. Sample covariance matrices of heavy-tailed distributions. *Int. Math. Res. Not.* **2018**, *2018*, 6254–6289.
22. Tang, Q.H.; Tsitsiashvili, G. Precise estimates for the ruin probability in finite horizon in a discrete-time model with heavy-tailed insurance and financial risks. *Stoch. Proc. Appl.* **2003**, *108*, 299–325. [CrossRef]
23. Liu, L.; Wang, B.; Long, G. The finite-time ruin probability of a bidimensional risk model with heavy-tailed clsims. *Math. Theory Appl.* **2007**, *27*, 67–71. (In Chinese)
24. Zhang, Y.; Wang, W. Ruin probabilities of a bidimensional risk model with investment. *Stat. Probab. Lett.* **2012**, *82*, 130–138. [CrossRef]
25. Embrechts, P.; Klüppelberg, C.; Mikosch, T. *Modelling Extremal Events for Insurance and Finance*; Springer: Berlin, Germany, 1997.
26. Ross, R. *Stochastic Processes*; Wiley: New York, NY, USA, 1983.
27. Ebrahimi, N. On the dependence structure of certain multi-dimensional Ito processes and corresponding hitting times. *J. Multivar. Anal.* **2002**, *81*, 128–137. [CrossRef]
28. Pitt, L. Positively correlated normal variables are associated. *Ann. Probab.* **1982**, *10*, 496–499. [CrossRef]
29. He, H.; Keirstead, W.P.; Rebholz, J. Double lookbacks. *Math. Financ.* **1998**, *8*, 201–228. [CrossRef]
30. Shao, J.; Wang, X. Estimates of the exit probability for two correlated Brownian motions. *Adv. Appl. Probab.* **2013**, *45*, 37–50. [CrossRef]
31. Ikeda, N.; Watanabe, S. *Stochastic Differential Equations and Diffusion Processes*; North-Holland Publishing Company: Amsterdam, The Netherlands, 1981.
32. Yin, C.C.; Wen, Y. An extension of Paulsen-Gjessing's risk model with stochastic return on investments. *Insur. Math. Econom.* **2013**, *52*, 469–476. [CrossRef]
33. Embrechts, P.; Goldie, C.M. On closure and factorization properties of subexponential and related distributions. *J. Aust. Math. Soc. Ser. A* **1980**, *29*, 243–256. [CrossRef]
34. Geluk, J.; Tang, Q.H. Asymptotic tail probabilities of sums of dependent subexponential random variables. *J. Theor. Probab.* **2009**, *22*, 871–882. [CrossRef]
35. Kaas, R.; Goovaerts, M.J.; Dhaene, J.; Denuit, M. *Modern Actuarial Risk Theory*; Kluwer Academic Publishers: Boston, MA, USA, 2008.

mathematics

MDPI

Article

Optimal Defined Contribution Pension Management with Jump Diffusions and Common Shock Dependence

Wujun Lv [1], Linlin Tian [1,*] and Xiaoyi Zhang [2]

[1] College of Science, Donghua University, Shanghai 201620, China; lvwujun@dhu.edu.cn
[2] School of Economics and Management, Hebei University of Technology, Tianjin 300401, China; 2019027@hebut.edu.cn
* Correspondence: linlin.tian@dhu.edu.cn

Abstract: This work deals with an optimal asset allocation problem for a defined contribution (DC) pension plan during its accumulation phase. The contribution rate is assumed to be proportional to the individual's salary. The salary follows a Heston stochastic volatility model with jumps, and there exists common shock dependence between the salary and the volatility. Since the time horizon of pension management is quite long, the influence of inflation is considered in the given context. The aim of the pension plan described in this paper is to reduce fluctuations in terminal wealth by investing in the bond and the stock. Through the dynamic programming principle, the Hamilton–Jacobi–Bellman equation is shown. The explicit expression of the investment decision is derived by solving the Hamilton–Jacobi–Bellman equation. In the last part, a numerical analysis is shown to illustrate the impacts of different parameters on the optimal investment policy.

Keywords: DC pension plan; stochastic volatility; Poisson process; common shock dependence; inflation; Hamilton–Jacobi–Bellman equation

MSC: 93E20; 91G30

Citation: Lv, W.; Tian, L.; Zhang, X. Optimal Defined Contribution Pension Management with Jump Diffusions and Common Shock Dependence. *Mathematics* **2023**, *11*, 2954. https://doi.org/10.3390/math11132954

Academic Editors: Jing Yao, Xiang Hu and Jingchao Li

Received: 31 May 2023
Revised: 29 June 2023
Accepted: 29 June 2023
Published: 2 July 2023

1. Introduction

A pension fund is an important financial instrument for individuals to reallocate incomes and sustain consumption after retirement. Generally, according to the determination of benefits, there are two typical types of pension plans: defined benefit (DB) and defined contribution (DC) pension plans. In DB plans, benefits are fixed in advance, while, in the DC case, contributions are fixed by the trustee. There are two phases in a pension scheme: the accumulation phase, which is the period from the entry time to the retirement time, and the decumulation phase, which is the period from the retirement time to death.

Next, we review the literature relevant to our paper. In the accumulation phase of a DC pension scheme, the contributor contributes part of his/her salary to the fund. Since the salary is related to the profitability of the company, the works of Bodie et al. [1] and Dybvig and Liu [2] assume that the salary process is spanned by the stock price. In addition, Guan and Liang [3] and Li and Wang [4] describe the salary process using a Heston stochastic volatility model, i.e., the salary is correlated with the volatility of the stock. Furthermore, Zeng et al. [5] assume that the salary process is related to stochastic volatility. Based on [3–5], we add an independent random process to the stochastic salary process to be closer to the reality.

It is appropriate to insert a jump process in the stochastic salary due to a promotion and job-hopping. Moreover, it is realistic to introduce jumps in volatility, which represent some unexpected events, such as an economic crisis or policy adjustments by the government. In our model, the contribution rate of the pension scheme is proportional to the salary of the individual, the dynamics of which follows a Heston stochastic volatility model with jumps. In addition, salary and variance are correlated by means of a common shock. In reality, a common component may depict an event that has an impact on both salary and volatility.

Common shock models are widely used in the area of actuarial science. For instance, in Liang et al. [6], the insurance risk model is modulated by a compound Poisson process, and the two-jump-number processes are correlated through a common shock. Liang et al. [7] assume that jumps in both the risky asset and insurance risk process are correlated through common shock dependence.

Since the period of a pension scheme is usually long, the inflation risk should be considered during the optimization phase. There are various studies focused on the stochastic optimization problem for DC pension plans under the inflation risk. For example, ref. [8] explores the optimal asset allocation problem with downside protection and stochastic inflation risk. Yao et al. [9] solve an optimal portfolio decision problem under the inflation risk and mean-variance criterion. Other relevant works on optimal control under inflation risk can be found in [10,11]. The inflation risk is also involved in our model. Instead of a pure diffusion process, we introduce a jump diffusion process to model the nominal price level of a representative bundle of commodity goods in the market. In other words, the dynamics of the price index given by Zhang et al. [12] and Zhang and Ewald [13] are extended in our model, and a Poisson jump is included in the evolution of the index price.

Stochastic processes are commonly used to model the uncertainty in the financial market. For further study of stochastic processes, we refer interested readers to [14–20]. In our paper, we assume that the pension manager is allowed to invest in two types of assets: the bond and the stock. The dynamics of the bond price follows a geometric Brownian process. The stock price is driven by a drifted Brownian motion and a Poisson jump. The similar asset model is widely used in various asset allocation problems. For example, Merton [21] considers Poisson jumps in an optimal dynamic portfolio decision problem. In a DC pension funding framework, Sun et al. [22] deal with the pre-commitment and equilibrium investment strategies by incorporating jumps into the risky asset process. More relevant works on jump diffusion asset allocation problems in pension management can be found in [23–28].

The aim of pension management is to find the optimal investment and minimize the expected distance between the terminal wealth and two given targets. To find the optimal policy, the dynamic programming principle is used to derive the Hamilton–Jacobi–Bellman (HJB) equation. From the classical optimal control theory, once a continuously differentiable solution of the HJB equation is explicitly solved, the optimal value function and the optimal policy can be derived. In our paper, by solving the HJB equation, we show the explicit form of the optimal investment policy and the optimal value function. The dynamic programming principle and HJB equation are applicable to various optimization problems. However, the drawback is that if there is no explicit solution for the HJB equation, then the dynamic programming principle fails to solve the problem. In our paper, there is a continuously differentiable solution for the HJB equation, since there is only one specific boundary condition in the HJB equation. If there are more than two boundary conditions for the optimization problem, it will be more difficult to find the explicit expression for the value function as well as the optimal policy. When there is no explicit solution for the HJB equation, it is more appropriate to adopt the maximum principle, Martingale approach or viscosity solution.

The rest of the paper is structured as follows. Section 2 describes the financial market with the jump diffusion price index, as well as two tradable assets that are of interest for pension management. This section also gives the pension model. Section 3 deals with a stochastic optimal control problem in order to minimize the fluctuation in the final real wealth over a finite horizon. The closed form of the investment strategy is given by solving the HJB equation. Finally, Section 4 gives the sensitivity analysis and Section 5 establishes the conclusions.

2. Model Assumptions and Notations

Consider a probability space $(\Omega, \mathscr{F}, \mathbb{P})$, with $\mathbb{P}$ as the real-world probability measure on Ω and $\mathscr{F} = \mathscr{F}^W \vee \mathscr{F}^N$. The filtration $\mathscr{F}^W = \{\mathscr{F}_t^W\}_{t\geq 0}$ is generated by a

five-dimensional standard Brownian motion $(W_r, W_\Pi, W_S, W_L, W_V)$, i.e., $\mathscr{F}_t^W = \sigma\{(W_r(s), W_\Pi(s), W_S(s), W_L(s), W_V(s)); 0 \le s \le t\}, t \ge 0$, which represents the risk sources of the interest rate, inflation, stock price, salary and its volatility, respectively. W_r and W_Π are correlated, which is captured by the coefficient $\rho_{\Pi r} \in (-1, 1)$. W_L and W_V are also correlated, which is captured by the coefficient $\rho_{LV} \in (-1, 1)$. Let the filtration $\mathscr{F}^N = \{\mathscr{F}_t^N\}_{t\ge 0}$ be generated by a five-dimensional Poisson process $(N_\Pi, N_S, N_L, N_V, N_c)$ with intensity $(\lambda_\Pi, \lambda_S, \lambda_L, \lambda_V, \lambda_c)$, where $\lambda_\Pi, \lambda_S, \lambda_L, \lambda_V, \lambda_c \in \mathbb{R}^+$, i.e., $\mathscr{F}_t^N = \sigma\{N_\Pi(s), N_S(s), N_L(s), N_V(s), N_c(s); 0 \le s \le t\}, t \ge 0$, which represent the jumps in inflation, stock price, salary, volatility and common shocks between two jumps, respectively. Suppose that Poisson processes are mutually independent. Moreover, Brownian motions are independent of Poisson processes on $(\Omega, \mathscr{F}, \mathbb{P})$.

2.1. The Financial Market

Following the work of Eisenberg [29], we assume that the discount factor is a geometric Brownian motion:

$$\exp\left[r + mt + \zeta W_r(t)\right],$$

where $r, m \in \mathbb{R}^+$, and $\zeta \in \mathbb{R}$.

The financial market consists of two underlying instruments that are traded continuously over time and perfectly divisible. Suppose that there are no transaction costs or taxes in the given context. The bond $S_0(t)$ evolves according to the the following dynamics:

$$\frac{dS_0(t)}{S_0(t)} = (m + \frac{\zeta^2}{2})dt + \zeta dW_r(t), \tag{1}$$

with initial price $S_0(0) = e^r$.

Besides the cash account, the trustee also has the opportunity to invest the fund into a stock with the dynamics

$$\frac{dS(t)}{S(t-)} = \mu_S(t)dt + \sigma_{SS}(t)dW_S(t) + \eta_S(t)dN_S(t), \tag{2}$$

where $\mu_S(t)$ is the appreciation rate for the stock. $\sigma_{SS}(t)$ is the volatility associated with the diffusion component of the stock price. $\eta_S(t)$ denotes the magnitude of a jump. We state that $\eta_S(t) > -1$ to prevent the process from jumping to a value below zero. W_S describes the fluctuation, and N_S describes the jump in the stock price. For simplicity, it is assumed that W_S and N_S are independent stochastic processes.

2.2. The Pension Model

This paper considers the accumulation phase of a DC-type pension plan. Assume that the entry time of a pensioner is the initial time 0, and his/her retirement time is the terminal time in our model. We denote the pensioner's death time as τ, which is a positive random variable defined on the probability space $(\Omega, \mathscr{F}, \mathbb{P})$. The mortality rate $\lambda(t)$ is defined as

$$\lambda(t) = \lim_{\triangle t \to 0} \frac{\mathbb{P}(t < \tau < t + \triangle t \mid \tau > t)}{\triangle t}.$$

In a general pension plan, the pensioner pays contributions before the retirement time T, where $T \in R^+$. The level of the contribution rate is usually defined as a proportion $\xi (0 \le \xi \le 1)$ of the pensioner's salary. In previous works, such as refs. [3–5], it is assumed that the stochastic salary is driven by the Heston stochastic volatility model, i.e., the salary process has stochastic volatility and the salary return variance is governed by a mean-reverting process. To be more realistic, we add a Brownian motion W_L in the salary process to describe the fluctuation in the salary itself. In addition, we also assume that there are

possible common Poisson jumps between the salary process and the stochastic variance. The dynamics of the investor's salary $L(t)$ follow

$$\begin{cases} \dfrac{dL(t)}{L(t-)} = \mu_L(t)dt + \sigma_{LS}(t)dW_S(t) + \sqrt{V(t)}dW_L(t) + \eta_{LL}(t)dN_L(t) + \eta_{Lc}(t)dN_c(t), \\ L(0) = L_0, \end{cases} \tag{3}$$

while the stochastic volatility $V(t)$ is governed by

$$dV(t) = \kappa(\delta - V(t))dt + \sigma_V\sqrt{V(t)}dW_V(t) + \eta_{VV}(t)dN_V(t) + \eta_{Vc}(t)dN_c(t). \tag{4}$$

Regarding the parameters in (3), $\mu_L(t)$ denotes the instantaneous expected rate of the salary, and $\sigma_{LS}(t)$ is the instantaneous volatility scale factor measuring how the risk source of the stock price affects the salary. $\eta_{LL}(t)(> -1)$ and $\eta_{Lc}(t)(> -1)$ denote the magnitude of the jumps associated with Poisson processes $N_L(t)$ and $N_c(t)$, respectively. N_L describes the jump in the salary itself, and N_c describes the possible common jumps between the salary (given by Equation (3)) and the stochastic volatility $V(t)$ (given by Equation (4)). As introduced earlier, we state that the Brownian motion W_L in Equation (3) and the Brownian motion W_V in Equation (4) are correlated with the coefficient $\rho_{LV} \in (-1, 1)$.

Regarding the parameters in (4), κ denotes the mean-reversion rate, δ denotes the long-run mean, and σ_V is the volatility coefficient. The assumption $2\kappa\delta > \sigma_V^2$ is proposed to guarantee the volatility process $V(t) > 0$. The Brownian motion W_V describes the fluctuation in the volatility, and the Poisson process N_V describes the jump in the volatility. We state that $\eta_{VV} > -1$ and $\eta_{Vc} > -1$ to prevent the process $V(t)$ from jumping to a value below zero. It should be noted that all Poisson processes are mutually independent. Moreover, Brownian motions are independent of Poisson processes.

3. The Optimal Portfolio

The aim of the stochastic control problem is to find the optimal investment decision. The pension trustee continuously decides on the weights invested into the cash account and the stock. We denote the nominal wealth at time t as $X(t)$. Under the investment policy chosen, it is easy to obtain the following stochastic differential equation, which describes the evolution of the wealth:

$$dX(t) = X(t)(1 - \pi(t))\frac{dS_0(t)}{S_0(t)} + X(t-)\pi(t)\frac{dS(t)}{S(t-)} + \xi(t)L(t)dt, \tag{5}$$

with $X(0) = X_0 > 0$. $\pi(t)$ denotes the weight invested into the stock at time t. The remainder, $1 - \pi(t)$, is the proportion invested into the cash account. Borrowing and short-selling are permitted in the given context. A negative value of $\pi(t)$ means that the pension trustee takes a short position in the stock, while a negative value of $1 - \pi(t)$ reflects that the trustee borrows money from the bank to purchase the risky asset.

By substituting Equations (1) and (2) into Equation (5), we obtain that

$$\begin{aligned} dX(t) \;=\; & X(t)\Big[(m + \frac{\zeta^2}{2}) + \pi(t)(\mu_S(t) - (m + \frac{\zeta^2}{2}))\Big]dt + \xi(t)L(t)dt \\ & + X(t)(1 - \pi(t))\zeta(t)dW_r(t) + X(t)\pi(t)\sigma_{SS}(t)dW_S(t) + X(t-)\pi(t-)\eta_S(t)dN_S(t). \end{aligned}$$

As mentioned in Section 1, the time horizon for the accumulation phase of a pension fund (in our model, from time 0 to T) is usually long; hence, the influence of inflation is considered in the given context.

The price index at time t is denoted by $\Pi(t)$, which refers to the purchase power per unit of money. The dynamics are driven by a jump diffusion process of the following type:

$$\frac{d\Pi(t)}{\Pi(t-)} = \mu_\Pi(t)dt + \sigma_\Pi(t)dW_\Pi(t) + \eta_\Pi(t)dN_\Pi(t), \tag{6}$$

with initial value $\Pi(0) = \Pi_0 > 0$. $\mu_\Pi(t)$ is the instantaneous expected inflation rate. $\sigma_\Pi(t)$ is the instantaneous volatility associated with the diffusion component, and $\eta_\Pi(t)$ denotes the magnitude of a jump with the condition $\eta_\Pi > -1$ to ensure that the price index remains strictly positive. In Equation (6), W_Π and N_Π are independent stochastic processes.

Next, we define the corresponding real salary process as the following:

Definition 1. *The real salary process is defined by*

$$\overline{L}(t) = \frac{L}{\Pi}(t).$$

Applying the quotient rule of Itô's formula, $\overline{L}$ is given by

$$\begin{aligned} d\overline{L}(t) &= d\Big[\frac{L}{\Pi}\Big](t) \\ &= \overline{L}(t)(\mu_L(t) - \mu_\Pi(t) + \sigma_\Pi^2(t))dt + \overline{L}(t)\sigma_{LS}(t)dW_S(t) + \overline{L}(t)\sqrt{V(t)}dW_L(t) \\ &\quad -\overline{L}(t)\sigma_\Pi(t)dW_\Pi(t) + \overline{L}(t-)\eta_{LL}(t)dN_L(t) + \overline{L}(t-)\eta_{Lc}(t)dN_c(t) \\ &\quad +\overline{L}(t-)(\eta_\Pi^2(t) - \eta_\Pi(t))dN_\Pi(t). \end{aligned} \tag{7}$$

with initial value $\overline{L}(0) = L_0/\Pi_0 \triangleq \overline{L}_0$.

Then, the real wealth process with the consideration of inflation follows

$$\begin{aligned} d\overline{X}(t) &= d\Big[\frac{X}{\Pi}\Big](t) \\ &= \overline{X}(t)\Big[(m + \frac{\zeta^2}{2}) + \pi(t)(\mu_S(t) - (m + \frac{\zeta^2}{2}) + \zeta\sigma_\Pi(t)\rho_{\Pi r}(t)) - \mu_\Pi(t) + \sigma_\Pi^2(t) \\ &\quad -\zeta\sigma_\Pi(t)\rho_{\Pi r}(t)\Big]dt + \xi(t)\overline{L}(t)dt + \overline{X}(t)(1 - \pi(t))\zeta dW_r(t) + \overline{X}(t)\pi(t)\sigma_{SS}(t)dW_S(t) \\ &\quad -\overline{X}(t)\sigma_\Pi(t)dW_\Pi(t) + \overline{X}(t-)\pi(t)\eta_S(t)dN_S(t) + \overline{X}(t-)(\eta_\Pi^2(t) - \eta_\Pi(t))dN_\Pi(t), \end{aligned} \tag{8}$$

with initial condition $\overline{X}(0) = X_0/\Pi_0 \triangleq \overline{X}_0$.

Next, we restrict the strategies in order to fulfil some technical conditions. We call a strategy $\pi(\cdot)$ an admissible control process if it is $\mathscr{F}_t$-measurable, Markovian and stationary and satisfies the condition

$$\mathbb{E}\Big\{\int_0^\infty \pi^2(t)dt\Big\} < \infty. \tag{9}$$

Denote $\mathscr{A}_{\overline{X}_0,\overline{L}_0}$ the set of all admissible controls, i.e., it is the set of all measurable processes $\{\pi(t)\}_{t\geq 0}$, which satisfies Equation (9). Next, we try to find the optimal investment strategy for the DC pension plan manager under $\mathscr{A}_{\overline{X}_0,\overline{L}_0}$.

Assume that the pension trustee has a preference to minimize the expected value of the fluctuations in the terminal wealth until time $\tau \wedge T$, where T is the terminal time of the control problem. The objective is to minimize

$$J(t,\overline{X},\overline{L},V) = \mathbb{E}_t\Big[\big[\alpha_1 + \beta_1(\overline{X}(T) - X_1^*)\big]^2 \cdot \mathbb{1}_{\{\tau>T\}} + \big[\alpha_2 + \beta_2(\overline{X}(\tau) - X_2^*)\big]^2 \cdot \mathbb{1}_{\{\tau\leq T\}}\Big|\tau > t\Big], \tag{10}$$

with $\mathbb{E}_t$ as the conditional expectation given the filtration $\{\mathscr{F}_t\}_{t\geq 0}$. X_1^* and X_2^* are two positive constants representing the target funds of the plan at time T and τ, respectively. The deviation between the actual fund and the target fund is called the discontinuity risk; see Wang et al. [30].

In quadratic loss functions (10), any deviations between $\overline{X}$ and X_1^* (or X_2^*) are penalized. To be more specific, we assumed that $\alpha_1, \alpha_2 > 0$ and $\beta_1, \beta_2 < 0$ in Equation (10) to characterize that under-funding is more penalized than over-funding. A similar setting can be seen in Devolder, Janssen and Manca [31] and Zhang and Guo [32].

For a better understanding, we use a flowchart to describe the whole research process in Figure 1.

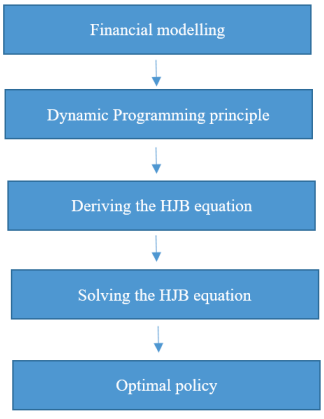

Figure 1. Research process.

According to

$$\mathbb{E}_t\left[\left[\alpha_1+\beta_1(\overline{X}(T)-X_1^*)\right]^2\cdot\mathbb{1}_{\{\tau>T\}}\middle|\tau>t\right]=\mathbb{E}_t\left[\left[\alpha_1+\beta_1(\overline{X}(T)-X_1^*)\right]^2 e^{-\int_t^T\lambda(u)du}\right],$$

and

$$\mathbb{E}_t\left[\left[\alpha_2+\beta_2(\overline{X}(\tau)-X_2^*)\right]^2\cdot\mathbb{1}_{\{\tau\le T\}}\middle|\tau>t\right]=\mathbb{E}_t\left[\int_t^T\left[\alpha_2+\beta_2(\overline{X}(s)-X_2^*)\right]^2\lambda(s)e^{-\int_t^s\lambda(u)du}ds\right],$$

the objective function with an uncertain lifetime can be converted into the following deterministic horizontal function:

$$\begin{aligned}&J(t,\overline{X},\overline{L},V)\\&=\mathbb{E}_t\left[\int_t^T\left[\alpha_2+\beta_2(\overline{X}(s)-X_2^*)\right]^2\lambda(s)e^{-\int_t^s\lambda(u)du}ds+\left[\alpha_1+\beta_1(\overline{X}(T)-X_1^*)\right]^2e^{-\int_t^T\lambda(u)du}\right].\end{aligned}$$

The dynamic programming approach is used to solve the stochastic optimization problem. Define the value function as

$$\varphi(t,\overline{X},\overline{L},V)=\min_{\{\pi\}}\{J(t,(\overline{X},\overline{L},V);\pi):\text{subject to }(8),(7),(4)\}.$$

In stochastic optimal control theory, the HJB equation accomplishes the connection between the value function and the optimal control; see, for instance, the books [33–37] and the papers [38–40]. The HJB equation is

$$\min_{\{\pi\}}\Psi(\pi)=0, \tag{11}$$

where

$$
\begin{aligned}
&\Psi(\pi)\\
&= \varphi_t + \lambda\left[\alpha_2 + \beta_2(\overline{X} - X_2^*)\right]^2 - \lambda\varphi + \varphi_{\overline{X}}\overline{X}\Big[(m + \frac{\zeta^2}{2}) + \pi(\mu_S - (m + \frac{\zeta^2}{2}) + \zeta\sigma_{\Pi}\rho_{\Pi r})\\
&\quad -\mu_{\Pi} + \sigma_{\Pi}^2 - \zeta\sigma_{\Pi}\rho_{\Pi r}\Big] + \varphi_{\overline{X}}\xi\overline{L} + \frac{1}{2}\varphi_{\overline{X}\,\overline{X}}\overline{X}^2\Big[(1-\pi)^2\zeta^2 + \pi^2\sigma_{SS}^2 + \sigma_{\Pi}^2 - 2(1-\pi)\\
&\quad \cdot\zeta\sigma_{\Pi}\rho_{\Pi r}\Big] + \varphi_{\overline{L}}\overline{L}(\mu_L - \mu_{\Pi} + \sigma_{\Pi}^2) + \frac{1}{2}\varphi_{\overline{L}\,\overline{L}}\overline{L}^2(\sigma_{LS}^2 + V + \sigma_{\Pi}^2) + \varphi_V\kappa(\delta - V) + \frac{1}{2}\varphi_{VV}\sigma_V^2 V\\
&\quad +\varphi_{\overline{X}\,\overline{L}}\overline{X}\,\overline{L}\Big[\pi\sigma_{SS}\sigma_{LS} + \sigma_{\Pi}^2 - (1-\pi)\zeta\sigma_{\Pi}\rho_{\Pi r}\Big] + \varphi_{\overline{L}V}\overline{L}V\sigma_V\rho_{LV} + \lambda_S\Big[\varphi(t, \overline{X}(1+\pi\eta_S), \overline{L}, V)\\
&\quad -\varphi(t, \overline{X}, \overline{L}, V)\Big] + \lambda_L\Big[\varphi(t, \overline{X}, \overline{L}(1+\eta_{LL}), V) - \varphi(t, \overline{X}, \overline{L}, V)\Big] + \lambda_V\Big[\varphi(t, \overline{X}, \overline{L}, V + \eta_{VV})\\
&\quad -\varphi(t, \overline{X}, \overline{L}, V)\Big] + \lambda_C\Big[\varphi(t, \overline{X}, \overline{L}(1+\eta_{Lc}), V + \eta_{Vc}) - \varphi(t, \overline{X}, \overline{L}, V)\Big]\\
&\quad +\lambda_{\Pi}\Big[\varphi(t, \overline{X}(1 + (\eta_{\Pi}^2 - \eta_{\Pi})), \overline{L}(1 + (\eta_{\Pi}^2 - \eta_{\Pi})), V) - \varphi(t, \overline{X}, \overline{L}, V)\Big],
\end{aligned}
\tag{12}
$$

with terminal condition $\varphi(T, \overline{X}, \overline{L}, V) = \left[\alpha_1 + \beta_1(\overline{X}(T) - X_1^*)\right]^2$. φ_t, $\varphi_{\overline{X}}$, $\varphi_{\overline{L}}$, φ_V, $\varphi_{\overline{X}\,\overline{X}}$, $\varphi_{\overline{L}\,\overline{L}}$, φ_{VV}, $\varphi_{\overline{X}\,\overline{L}}$ and $\varphi_{\overline{L}V}$ denote the first- and second-order partial derivatives of the value function φ with respect to t, $\overline{X}$, $\overline{L}$ and V, respectively.

If there exists a twice continuously differentiable solution of Equation (12), strictly convex, then the minimizer of the investment strategy is obtained by the optimal functional π^*, which satisfies the following necessary conditions:

$$\Psi(\pi^*) = 0, \tag{13}$$

$$\frac{d\Psi}{d\pi}(\pi^*) = 0. \tag{14}$$

We shall frequently use the following notations. Define

$$\varpi_1 = \mu_S - (m + \frac{\zeta^2}{2}) + \zeta\sigma_{\Pi}\rho_{\Pi r} + \lambda_S\eta_S, \tag{15}$$

$$\varpi_2 = \zeta\sigma_{\Pi}\rho_{\Pi r} + \sigma_{SS}\sigma_{LS}, \tag{16}$$

$$\varpi_3 = \zeta\sigma_{\Pi}\rho_{\Pi r} - \zeta^2, \tag{17}$$

$$\varpi_4 = \zeta^2 + \sigma_{SS}^2 + \lambda_S\eta_S^2. \tag{18}$$

By using the first-order condition and solving the HJB equation, the explicit form of the optimal investment decision is given by the following theorem.

Theorem 1. *(Main result) The optimal investment strategy on the stock is given by*

$$\pi^*(t) = -\frac{2\varphi_1(t)\overline{X} + \varphi_2(t) + \varphi_5(t)\overline{L}}{2\varphi_1(t)\overline{X}} \cdot \frac{\varpi_1}{\varpi_4} - \frac{\varphi_5(t)\overline{L}}{2\varphi_1(t)\overline{X}} \cdot \frac{\varpi_2}{\varpi_4} - \frac{\varpi_3}{\varpi_4}.$$

The value function is

$$\varphi(t, \overline{X}, \overline{L}, V) = \varphi_1(t)\overline{X}^2 + \varphi_2(t)\overline{X} + \varphi_3(t, V)\overline{L}^2 + \varphi_4(t)\overline{L} + \varphi_5(t)\overline{X}\,\overline{L} + \varphi_6(t).$$

In the above equations,

$$\varphi_1(t) = \lambda\beta_2^2\int_t^T e^{\int_t^s a_1(u)du}ds + \beta_1^2 e^{\int_t^T a_1(s)ds}, \tag{19}$$

$$\varphi_2(t) = 2\lambda(\alpha_2\beta_2 - \beta_2^2 X_2^*)\int_t^T e^{\int_t^s a_2(u)du}ds + 2\beta_1(\alpha_1 - \beta_1 X_1^*)e^{\int_t^T a_2(s)ds}, \tag{20}$$

$$\varphi_3(t,V) = \int_t^T \widetilde{\varphi}_{31}(t;\tau)e^{\widetilde{\varphi}_{32}(t;\tau)V}d\tau, \tag{21}$$

$$\varphi_4(t) = \int_t^T e^{\int_0^s [a_4(u)+\lambda_c\eta_{Lc}]du}\Big[\xi\varphi_2(s) - \frac{\varphi_2(s)\varphi_5(s)}{2\varphi_1(s)}\cdot\frac{\varpi_1(\varpi_1+\varpi_2)}{\varpi_4}\Big]ds\cdot e^{-\int_0^t [a_4(s)+\lambda_c\eta_{Lc}]ds}, \tag{22}$$

$$\varphi_5(t) = 2\int_t^T e^{\int_0^s [a_5(u)+\lambda_c\eta_{Lc}]du}\xi\varphi_1(s)ds\cdot e^{-\int_0^t [a_5(s)+\lambda_c\eta_{Lc}]ds}, \tag{23}$$

$$\varphi_6(t) = \int_t^T e^{\int_t^s a_1(u)du}\Big[\lambda(\alpha_2 - \beta_2 X_2^*)^2 - \frac{\varphi_2^2}{4\varphi_1}\cdot\frac{\varpi_1^2}{\varpi_4}\Big]ds + (\alpha_1 - \beta_1 X_1^*)^2 e^{\int_t^T a_1(s)ds}, \tag{24}$$

where a_1, a_2, $\widetilde{\varphi}_{31}$, $\widetilde{\varphi}_{32}$, a_4 and a_5 are given by Equation (A6), Equation (A7), Equation (A28), Equation (A27), Equation (A8) and Equation (A9), respectively.

Proof. See Appendix A. □

4. Sensitivity Analysis

In order to investigate the influence of the parameters on the optimal investment decision, we provide a sensitivity analysis in this section. Unless otherwise stated, the employed parameters of the model are based on the following annualized benchmark values presented in Table 1. In what follows, we mainly explore the impacts of the volatility σ_{LS}, the jump magnitude η_{LL} and the jump intensity λ_L, λ_c on optimal dividend policy π^*.

Table 1. Model parameters.

T	X_0	L_0	X_1^*	X_2^*	α_1	α_2	β_1	β_2	λ	ζ	**m**	ξ	μ_Π	μ_S
35	1	1	100	100	0.1	0.1	−0.01	−0.01	0.01	0.05	0.01	0.02	0.2	0.1
μ_L	σ_Π	σ_{SS}	σ_{LS}	σ_V	η_Π	η_S	η_{LL}	η_{VV}	η_{Lc}	η_{Vc}	λ_Π	λ_S	λ_L	λ_c
0.1	0.5	0.5	0.5	0.5	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3

Figure 2 gives a possible path simulation of the optimal investment decision π^*. Denoting the optimal wealth process by X^*, a possible path simulation is given in Figure 3. As we can see, the path simulation of the wealth of the pension fund is increasing as time passes, which highly coincides with reality.

Figure 4 analyzes the relationship between the optimal investment decision π^* and the stock volatility σ_{LS}. First, we can see that the optimal investment decision π^* is always negative under the given parameters in Table 1, which means that the pension manager prefers to be a short-seller and invests more money in less risky bonds. From Figure 4, we can also see that when σ_{LS} increases, the optimal investment policy π^* increases. This can be explained by the fact that as σ_{LS} increases, the risk and return of the pension fund simultaneously increase. To achieve the given target value X_1^* and X_2^*, the manager prefers to take more risks as well as achieve more profits.

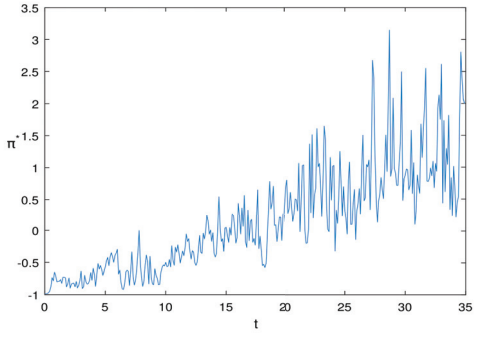

Figure 2. A path of π^*.

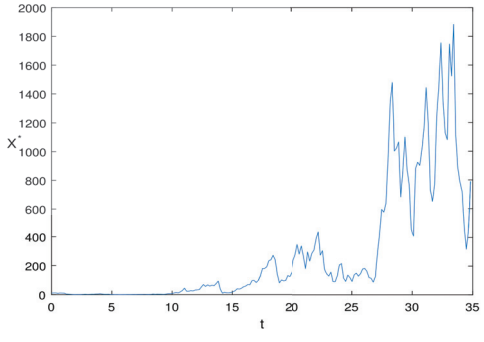

Figure 3. A path of X^*.

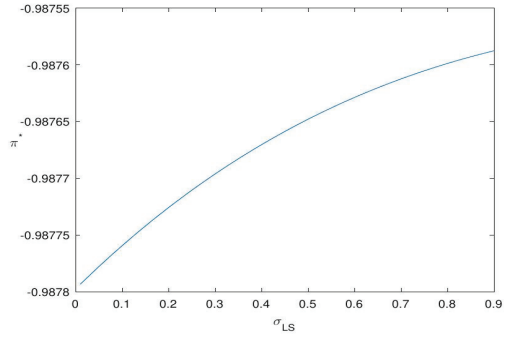

Figure 4. Impact of σ_{LS} on π^*.

In Figure 5, as η_{LL} increases, the optimal investment policy π^* also increases. This phenomenon can be explained as follows. The parameter η_{LL} measures the jump magnitude of the salary. When η_{LL} is positive and increasing, which means that there is a higher jump in the pensioner's salary surplus, or, in other words, the pensioner will input more money into the pension fund and eventually the manager of the pension fund will increase the investment amount for the risky stocks to achieve higher profits.

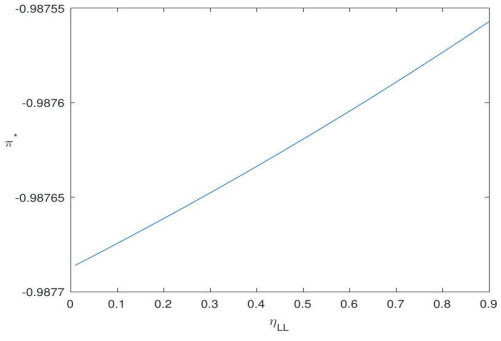

Figure 5. Impact of η_{LL} on π^*.

Figures 6 and 7 show the impact of the Poisson intensity λ_L, λ_c on optimal policy π^*, respectively. As we can see, as λ_L, λ_c increase, the optimal investment amount π^* increases. This can be explained as follows. λ_L, λ_c represent the intensity of a positive jump in salary surplus. If λ_L (or λ_c) increases, then the pensioner will be more positive about the future and will make more contributions to the pension fund. Eventually, the pension manager is able to invest more money in the risky asset.

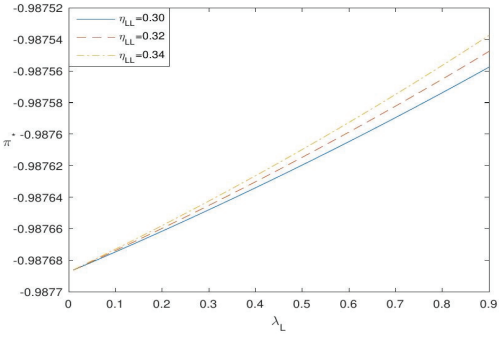

Figure 6. Impact of λ_L on π^*.

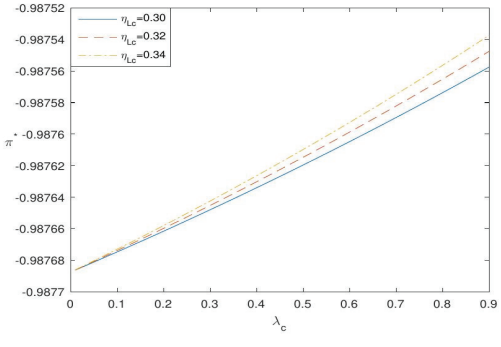

Figure 7. Impact of λ_c on π^*.

Figure 6 depicts the impact of λ_L on the optimal policy under different parameters $\eta_{LL} = 0.30, 0.32, 0.34$. Figure 7 depicts the impact of λ_L on the optimal policy under different parameters $\eta_{Lc} = 0.30, 0.32, 0.34$. Both figures show the same conclusion, i.e., that the higher the salary jump magnitude, the more money should be invested in the stock market. The impact of η_{Lc} on π^* is similar to that of η_{LL} since η_{Lc} and η_{LL} both measure the jump magnitude of the salary.

Figure 8 depicts the impact of the jump magnitude of the salary on the optimal policy π^* with different μ_L. As we can see, as μ_L increases, the investment policy π^* also increases. This can be explained as follows: if the average wage of society increases, then the investment enthusiasm will increase. Figure 9 depicts the impact of η_{Lc} with different σ_{LS}. The increasing σ_{LS} leads to an increase in investment policy π^*. This shows that if the volatility of the salary is high, then the optimal choice is to increase the investment in the risky asset to reach the desired target as soon as possible, to avoid possible losses. From the above analysis, from the perspective of the government, an increase in salary will increase the investment in risky assets. On the other hand, to encourage the pension manager to invest more money in a risk-less bond, the government should be reduce the intensity of wage growth.

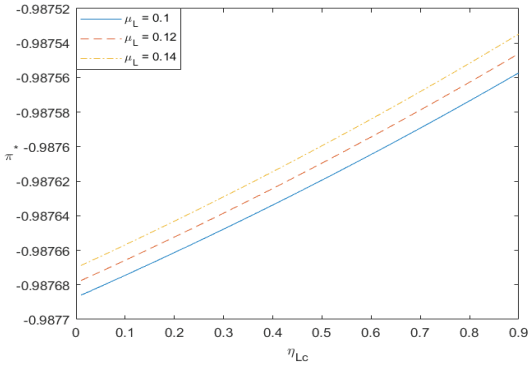

Figure 8. Impact of η_{Lc} with different μ_L.

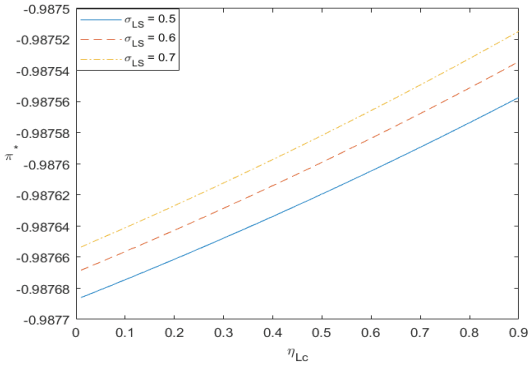

Figure 9. Impact of η_{Lc} with different σ_{LS}.

5. Conclusions

This paper analyzes the optimal investment strategy for a DC-type pension scheme during its accumulation phase, where the price of the risky asset follows a jump diffusion process. The price index as well as the common shock between the salary and the variance are involved. The aim of pension management is to minimize the fluctuations in terminal

wealth, and the dynamic programming technique is used to derive the HJB equation. By solving an explicit continuously differentiable solution for the HJB equation, we give a classical expression for the optimal value function as well as the optimal policy. Sensitivity analysis shows the impact of different parameters on the investment policy, from which we conclude that several essential factors (the volatility of the salary process, the magnitude of salary jumps, the intensity of salary jumps, the jump intensity of the volatility of the salary, the expected rate of salary) control the investment in the risky and risk-less assets. We suggest that the government should regulate income levels, wage increases and financial market volatility to stabilize the pension market. If it is necessary to stimulate pension managers to buy risk-less assets such as treasury bonds, then the optimal policy is to reduce the magnitude of salary jumps and the intensity of salary jumps.

It is also important to study the optimization problem in a defined benefit pension plan during the accumulation phase or decumulation phase. We will use the CIR model to describe the interest rate and mortality rate in a further study.

Author Contributions: Methodology, W.L. and L.T.; computation, X.Z. All authors have read and agreed to the published version of the manuscript.

Funding: The work was sponsored by the Fundamental Research Funds for the Central Universities, No. 2232021D-29, No. 2232021D-31, No. 2232023D-22, No. 2232022G-13 and No. 2232023G-13, and the High School Natural Science Foundation of Hebei Province, No. QN2021215.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author.

Acknowledgments: The authors thank the editor and the referees for their valuable comments and suggestions, which improved greatly the quality of this paper.

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

Appendix A

The proof of Theorem 1.

From the HJB Equation (11), we conjecture that the solution of (11) takes a quadratic homogeneous form with $\varphi \in C^{1,2}$ and $\varphi_{\overline{X}\,\overline{X}} > 0$, as the following:

$$\varphi(t,\overline{X},\overline{L},V) = \varphi_1(t,V)\overline{X}^2 + \varphi_2(t,V)\overline{X} + \varphi_3(t,V)\overline{L}^2 + \varphi_4(t,V)\overline{L} + \varphi_5(t,V)\overline{X}\,\overline{L} + \varphi_6(t,V), \tag{A1}$$

where $\varphi_1(\cdot,\cdot)$, $\varphi_2(\cdot,\cdot)$, $\varphi_3(\cdot,\cdot)$, $\varphi_4(\cdot,\cdot)$, $\varphi_5(\cdot,\cdot)$, and $\varphi_6(\cdot,\cdot)$ are six suitable functions with terminal conditions $\varphi_1(T,V) = \beta_1^2$, $\varphi_2(T,V) = 2\beta_1(\alpha_1 - \beta_1 X_1^*)$, $\varphi_3(T,V) = \varphi_4(T,V) = \varphi_5(T,V) = 0$ and $\varphi_6(T,V) = (\alpha_1 - \beta_1 X_1^*)^2$.

Differentiating Equation (A1) with respect to t, $\overline{X}$, $\overline{L}$, V, we have

$$\begin{aligned}
&\varphi_t = \varphi_{1t}\overline{X}^2 + \varphi_{2t}\overline{X} + \varphi_{3t}\overline{L}^2 + \varphi_{4t}\overline{L} + \varphi_{5t}\overline{X}\,\overline{L} + \varphi_{6t}, \qquad \varphi_{\overline{X}} = 2\varphi_1\overline{X} + \varphi_2 + \varphi_5\,\overline{L},\\
&\varphi_{\overline{X}\,\overline{X}} = 2\varphi_1, \qquad \varphi_{\overline{L}} = 2\varphi_3\overline{L} + \varphi_4 + \varphi_5\,\overline{X}, \qquad \varphi_{\overline{L}\,\overline{L}} = 2\varphi_3, \qquad \varphi_{\overline{X}\,\overline{L}} = \varphi_5,\\
&\varphi_V = \varphi_{1V}\overline{X}^2 + \varphi_{2V}\overline{X} + \varphi_{3V}\overline{L}^2 + \varphi_{4V}\overline{L} + \varphi_{5V}\overline{X}\,\overline{L} + \varphi_{6V},\\
&\varphi_{VV} = \varphi_{1VV}\overline{X}^2 + \varphi_{2VV}\overline{X} + \varphi_{3VV}\overline{L}^2 + \varphi_{4VV}\overline{L} + \varphi_{5VV}\overline{X}\,\overline{L} + \varphi_{6VV},
\end{aligned} \tag{A2}$$

where φ_{1t}, φ_{1V} and φ_{1VV} denote the first- and second-order derivatives of φ_1 with respect to t and V, respectively. The derivatives of φ_2, φ_3, φ_4, φ_5 and φ_6 are defined in the same way.

Substituting Equations (A1) and (A2) into Equations (11) and (12), and rearranging the terms by the order of π, we obtain that

$$
\begin{aligned}
&\min_{\{\pi\}} \varphi_1\overline{X}^2\varpi_4\pi^2 + (2\varphi_1\overline{X} + \varphi_2 + \varphi_5\,\overline{L})\overline{X}\varpi_1\pi + \varphi_5\overline{X}\,\overline{L}\varpi_2\pi + 2\varphi_1\overline{X}^2\varpi_3\pi + \varphi_1\overline{X}^2(\zeta^2\\
&-2\zeta\sigma_{\Pi}\rho_{\Pi r}) - \varphi_5\overline{X}\,\overline{L}\zeta\sigma_{\Pi}\rho_{\Pi r} + \varphi_{1t}\overline{X}^2 + \varphi_{2t}\overline{X} + \varphi_{3t}\overline{L}^2 + \varphi_{4t}\overline{L} + \varphi_{5t}\overline{X}\,\overline{L} + \varphi_{6t} + \lambda\beta_2^2\overline{X}^2\\
&+2\lambda(\alpha_2\beta_2 - \beta_2^2X_2^*)\overline{X} + \lambda(\alpha_2 - \beta_2X_2^*)^2 - \lambda(\varphi_1\overline{X}^2 + \varphi_2\overline{X} + \varphi_3\overline{L}^2 + \varphi_4\overline{L} + \varphi_5\overline{X}\,\overline{L} + \varphi_6)\\
&+(2\varphi_1\overline{X} + \varphi_2 + \varphi_5\,\overline{L})\overline{X}\Big[(m + \frac{\zeta^2}{2}) - \mu_{\Pi} + \sigma_{\Pi}^2 - \zeta\sigma_{\Pi}\rho_{\Pi r}\Big] + (2\varphi_1\overline{X} + \varphi_2 + \varphi_5\,\overline{L})\overline{L}\xi\\
&+\varphi_1\overline{X}^2\sigma_{\Pi}^2 + (2\varphi_3\overline{L} + \varphi_4 + \varphi_5\overline{X})\overline{L}(\mu_L - \mu_{\Pi} + \sigma_{\Pi}^2) + \varphi_3\overline{L}^2(\sigma_{LS}^2 + V + \sigma_{\Pi}^2) + (\varphi_{1V}\overline{X}^2\\
&+\varphi_{2V}\overline{X} + \varphi_{3V}\overline{L}^2 + \varphi_{4V}\overline{L} + \varphi_{5V}\overline{X}\,\overline{L} + \varphi_{6V})\kappa(\delta - V) + \frac{1}{2}(\varphi_{1VV}\overline{X}^2 + \varphi_{2VV}\overline{X} + \varphi_{3VV}\overline{L}^2\\
&+\varphi_{4VV}\overline{L} + \varphi_{5VV}\overline{X}\,\overline{L} + \varphi_{6VV})\sigma_V^2V + \varphi_5\overline{X}\,\overline{L}\sigma_{\Pi}^2 + (2\varphi_{3V}\overline{L} + \varphi_{4V} + \varphi_{5V}\,\overline{X})\overline{L}V\sigma_V\rho_{LV}\\
&+\varphi_3\overline{L}^2\lambda_L\eta_L^2 + (2\varphi_3\overline{L}^2 + \varphi_4\overline{L} + \varphi_5\,\overline{X}\,\overline{L})\lambda_L\eta_L + \lambda_V\Big[(\varphi_1(t, V + \eta_{VV}) - \varphi_1(t, V))\overline{X}^2\\
&+(\varphi_2(t, V + \eta_{VV}) - \varphi_2(t, V))\overline{X} + (\varphi_3(t, V + \eta_{VV}) - \varphi_3(t, V))\overline{L}^2 + (\varphi_4(t, V + \eta_{VV})\\
&-\varphi_4(t, V))\overline{L} + (\varphi_5(t, V + \eta_{VV}) - \varphi_5(t, V))\overline{X}\,\overline{L} + (\varphi_6(t, V + \eta_{VV}) - \varphi_6(t, V))\Big]\\
&+\lambda_c\Big[(\varphi_1(t, V + \eta_{Vc}) - \varphi_1(t, V))\overline{X}^2 + (\varphi_2(t, V + \eta_{Vc}) - \varphi_2(t, V))\overline{X} + (\varphi_3(t, V + \eta_{Vc})\\
&-\varphi_3(t, V))\overline{L}^2 + \varphi_3(t, V + \eta_{Vc})(\eta_{Lc}^2 + 2\eta_{Lc})\overline{L}^2 + (\varphi_4(t, V + \eta_{Vc}) - \varphi_4(t, V))\overline{L}\\
&+\varphi_4(t, V + \eta_{Vc})\eta_{Lc}\overline{L} + (\varphi_5(t, V + \eta_{Vc}) - \varphi_5(t, V))\overline{X}\,\overline{L} + \varphi_5(t, V + \eta_{Vc})\eta_{Lc}\overline{X}\,\overline{L}\\
&+(\varphi_6(t, V + \eta_{Vc}) - \varphi_6(t, V))\Big] + \lambda_{\Pi}\Big[\varphi_1(t, V)((\eta_{\Pi}^2 - \eta_{\Pi})^2 + 2(\eta_{\Pi}^2 - \eta_{\Pi}))\overline{X}^2\\
&+\varphi_2(t, V)(\eta_{\Pi}^2 - \eta_{\Pi})\overline{X} + \varphi_3(t, V)((\eta_{\Pi}^2 - \eta_{\Pi})^2 + 2(\eta_{\Pi}^2 - \eta_{\Pi}))\overline{L}^2 + \varphi_4(t, V)(\eta_{\Pi}^2 - \eta_{\Pi})\overline{L}\\
&+\varphi_5(t, V)((\eta_{\Pi}^2 - \eta_{\Pi})^2 + 2(\eta_{\Pi}^2 - \eta_{\Pi}))\overline{X}\,\overline{L}\Big] = 0.
\end{aligned}
\tag{A3}
$$

where ϖ_1, ϖ_2, ϖ_3 and ϖ_4 are given by Equation (15), Equation (16), Equation (17) and Equation (18), respectively. By Equations (13) and (14), we have

$$
\pi^*(t, V) = -\frac{2\varphi_1(t, V)\overline{X} + \varphi_2(t, V) + \varphi_5(t, V)\overline{L}}{2\varphi_1(t, V)\overline{X}} \cdot \frac{\varpi_1}{\varpi_4} - \frac{\varphi_5(t, V)\overline{L}}{2\varphi_1(t, V)\overline{X}} \cdot \frac{\varpi_2}{\varpi_4} - \frac{\varpi_3}{\varpi_4}, \tag{A4}
$$

where π^* denotes the optimal investment decision regarding the risky asset. Substituting π^* into Equation (A3), and rearranging the terms by the order of $\overline{X}^2$, $\overline{L}^2$ and $\overline{X}\,\overline{L}$, we obtain the following bivariate polynomial function of $\overline{X}$ and $\overline{L}$:

$$
\begin{aligned}
&\Big[\varphi_{1t} + a_1(t)\varphi_1 + \kappa(\delta - V)\varphi_{1V} + \frac{1}{2}\sigma_V^2 V\varphi_{1VV} + \lambda_V(\varphi_1(t, V + \eta_{VV}) - \varphi_1(t, V)) \\
&+\lambda_c(\varphi_1(t, V + \eta_{Vc}) - \varphi_1(t, V)) + \lambda\beta_2^2\Big]\overline{X}^2 \\
&+\Big[\varphi_{2t} + a_2(t)\varphi_2 + \kappa(\delta - V)\varphi_{2V} + \frac{1}{2}\sigma_V^2 V\varphi_{2VV} + \lambda_V(\varphi_2(t, V + \eta_{VV}) - \varphi_2(t, V)) \\
&+\lambda_c(\varphi_2(t, V + \eta_{Vc}) - \varphi_2(t, V)) + 2\lambda(\alpha_2\beta_2 - \beta_2^2 X_2^*)\Big]\overline{X} \\
&+\Big[\varphi_{3t} + (a_3(t) + V)\varphi_3 + (\kappa(\delta - V) + 2\sigma_V\rho_{LV}V)\varphi_{3V} + \frac{1}{2}\sigma_V^2 V\varphi_{3VV} + \lambda_V(\varphi_3(t, V + \eta_{VV}) \\
&-\varphi_3(t, V)) + \lambda_c(\varphi_3(t, V + \eta_{Vc}) - \varphi_3(t, V)) + \lambda_c\varphi_3(t, V + \eta_{Vc})(\eta_{Lc}^2 + 2\eta_{Lc}) - \frac{\varphi_5^2}{4\varphi_1} \\
&\cdot\frac{(\omega_1 + \omega_2)^2}{\omega_4} + \xi\varphi_5\Big]\overline{L}^2 \\
&+\Big[\varphi_{4t} + a_4(t)\varphi_4 + (\kappa(\delta - V) + \sigma_V\rho_{LV}V)\varphi_{4V} + \frac{1}{2}\sigma_V^2 V\varphi_{4VV} + \lambda_V(\varphi_4(t, V + \eta_{VV}) - \varphi_4(t, V)) \\
&+\lambda_c(\varphi_4(t, V + \eta_{Vc}) - \varphi_4(t, V)) + \lambda_c\varphi_4(t, V + \eta_{Vc})\eta_{Lc} - \frac{\varphi_2\varphi_5}{2\varphi_1}\cdot\frac{\omega_1(\omega_1 + \omega_2)}{\omega_4} + \xi\varphi_2\Big]\overline{L} \\
&+\Big[\varphi_{5t} + a_5(t)\varphi_5 + \lambda_c\eta_{Lc}\varphi_5(t, V + \eta_{Vc}) + (\kappa(\delta - V) + \sigma_V\rho_{LV}V)\varphi_{5V} + \frac{1}{2}\sigma_V^2 V\varphi_{5VV} \\
&+\lambda_V(\varphi_5(t, V + \eta_{VV}) - \varphi_5(t, V)) + \lambda_c(\varphi_5(t, V + \eta_{Vc}) - \varphi_5(t, V)) + 2\xi\varphi_1\Big]\overline{XL} \\
&+\Big[\varphi_{6t} - \lambda\varphi_6 + \kappa(\delta - V)\varphi_{6V} + \frac{1}{2}\sigma_V^2 V\varphi_{6VV} + \lambda_V(\varphi_6(t, V + \eta_{VV}) - \varphi_6(t, V)) \\
&+\lambda_c(\varphi_6(t, V + \eta_{Vc}) - \varphi_6(t, V)) - \frac{\varphi_2^2}{4\varphi_1}\cdot\frac{\omega_1^2}{\omega_4} + \lambda(\alpha_2 - \beta_2 X_2^*)^2\Big] = 0,
\end{aligned}
\tag{A5}
$$

where

$$
\begin{aligned}
a_1(t) &= \zeta^2 - 4\zeta\sigma_\Pi\rho_{\Pi r} - \lambda + 2((m + \frac{\zeta^2}{2}) - \mu_\Pi + \sigma_\Pi^2) + \sigma_\Pi^2 + \lambda_\Pi((\eta_\Pi^2 - \eta_\Pi)^2 \\
&\quad +2(\eta_\Pi^2 - \eta_\Pi)) - \frac{(\omega_1 + \omega_3)^2}{\omega_4},
\end{aligned}
\tag{A6}
$$

$$
a_2(t) = (m + \frac{\zeta^2}{2}) - \mu_\Pi + \sigma_\Pi^2 - \zeta\sigma_\Pi\rho_{\Pi r} - \lambda + \lambda_\Pi(\eta_\Pi^2 - \eta_\Pi) - \frac{\omega_1(\omega_1 + \omega_3)}{\omega_4}, \tag{A7}
$$

$$
a_3(t) = 2(\mu_L - \mu_\Pi + \sigma_\Pi^2) + \sigma_{LS}^2 + \sigma_\Pi^2 - \lambda + \lambda_L\eta_L^2 + 2\lambda_L\eta_L + \lambda_\Pi((\eta_\Pi^2 - \eta_\Pi)^2 + 2(\eta_\Pi^2 - \eta_\Pi)),
$$

$$
a_4(t) = \mu_L - \mu_\Pi + \sigma_\Pi^2 - \lambda + \lambda_L\eta_L + \lambda_\Pi(\eta_\Pi^2 - \eta_\Pi), \tag{A8}
$$

$$
\begin{aligned}
a_5(t) &= (m + \frac{\zeta^2}{2}) - 2\mu_\Pi + 3\sigma_\Pi^2 - 2\zeta\sigma_\Pi\rho_{\Pi r} + \mu_L + \lambda_L\eta_L + \lambda_\Pi((\eta_\Pi^2 - \eta_\Pi)^2 + 2(\eta_\Pi^2 - \eta_\Pi)) \\
&\quad -\lambda - \frac{\omega_1^2 + \omega_1\omega_2 + \omega_1\omega_3 + \omega_2\omega_3}{\omega_4},
\end{aligned}
\tag{A9}
$$

Since Equation (A5) holds for every $\overline{X}$ and $\overline{L}$, the following six PDEs hold with the boundary conditions:

$$\begin{cases} \varphi_{1t} + a_1(t)\varphi_1 + \kappa(\delta - V)\varphi_{1V} + \frac{1}{2}\sigma_V^2 V\varphi_{1VV} + \lambda_V(\varphi_1(t, V + \eta_{VV}) - \varphi_1(t, V)) \\ + \lambda_c(\varphi_1(t, V + \eta_{Vc}) - \varphi_1(t, V)) + \lambda\beta_2^2 = 0, \\ \varphi_1(T, V) = \beta_1^2, \end{cases} \tag{A10}$$

$$\begin{cases} \varphi_{2t} + a_2(t)\varphi_2 + \kappa(\delta - V)\varphi_{2V} + \frac{1}{2}\sigma_V^2 V\varphi_{2VV} + \lambda_V(\varphi_2(t, V + \eta_{VV}) - \varphi_2(t, V)) \\ + \lambda_c(\varphi_2(t, V + \eta_{Vc}) - \varphi_2(t, V)) + 2\lambda(\alpha_2\beta_2 - \beta_2^2 X_2^*) = 0, \\ \varphi_2(T, V) = 2\beta_1(\alpha_1 - \beta_1 X_1^*), \end{cases}$$

$$\begin{cases} \varphi_{3t} + (a_3(t) + V)\varphi_3 + (\kappa(\delta - V) + 2\sigma_V\rho_{LV}V)\varphi_{3V} + \frac{1}{2}\sigma_V^2 V\varphi_{3VV} + \lambda_V(\varphi_3(t, V + \eta_{VV}) \\ - \varphi_3(t, V)) + \lambda_c(\varphi_3(t, V + \eta_{Vc}) - \varphi_3(t, V)) + \lambda_c\varphi_3(t, V + \eta_{Vc})(\eta_{Lc}^2 + 2\eta_{Lc}) - \frac{\varphi_5^2}{4\varphi_1} \\ \cdot \frac{(\varpi_1 + \varpi_2)^2}{\varpi_4} + \xi\varphi_5 = 0, \\ \varphi_3(T, V) = 0, \end{cases} \tag{A11}$$

$$\begin{cases} \varphi_{4t} + a_4(t)\varphi_4 + (\kappa(\delta - V) + \sigma_V\rho_{LV}V)\varphi_{4V} + \frac{1}{2}\sigma_V^2 V\varphi_{4VV} + \lambda_V(\varphi_4(t, V + \eta_{VV}) - \varphi_4(t, V)) \\ + \lambda_c(\varphi_4(t, V + \eta_{Vc}) - \varphi_4(t, V)) + \lambda_c\varphi_4(t, V + \eta_{Vc})\eta_{Lc} - \frac{\varphi_2\varphi_5}{2\varphi_1} \cdot \frac{\varpi_1(\varpi_1 + \varpi_2)}{\varpi_4} + \xi\varphi_2 = 0, \\ \varphi_4(T, V) = 0, \end{cases} \tag{A12}$$

$$\begin{cases} \varphi_{5t} + a_5(t)\varphi_5 + \lambda_c\eta_{Lc}\varphi_5(t, V + \eta_{Vc}) + (\kappa(\delta - V) + \sigma_V\rho_{LV}V)\varphi_{5V} + \frac{1}{2}\sigma_V^2 V\varphi_{5VV} \\ + \lambda_V(\varphi_5(t, V + \eta_{VV}) - \varphi_5(t, V)) + \lambda_c(\varphi_5(t, V + \eta_{Vc}) - \varphi_5(t, V)) + 2\xi\varphi_1 = 0, \\ \varphi_5(T, V) = 0, \end{cases} \tag{A13}$$

$$\begin{cases} \varphi_{6t} - \lambda\varphi_6 + \kappa(\delta - V)\varphi_{6V} + \frac{1}{2}\sigma_V^2 V\varphi_{6VV} + \lambda_V(\varphi_6(t, V + \eta_{VV}) - \varphi_6(t, V)) \\ + \lambda_c(\varphi_6(t, V + \eta_{Vc}) - \varphi_6(t, V)) - \frac{\varphi_2^2}{4\varphi_1} \cdot \frac{\varpi_1^2}{\varpi_4} + \lambda(\alpha_2 - \beta_2 X_2^*)^2 = 0, \\ \varphi_6(T, V) = (\alpha_1 - \beta_1 X_1^*)^2. \end{cases} \tag{A14}$$

Next, we solve the above equations, from Equation (A10) to Equation (A14), one by one. First, we solve Equation (A10). Assume that $\widetilde{\varphi}_1(t, V)$ is the solution of the following system:

$$\begin{cases} \widetilde{\varphi}_{1t} + a_1(t)\widetilde{\varphi}_1 + \kappa(\delta - V)\widetilde{\varphi}_{1V} + \frac{1}{2}\sigma_V^2 V\widetilde{\varphi}_{1VV} + \lambda_V(\widetilde{\varphi}_1(t, V + \eta_{VV}) - \widetilde{\varphi}_1(t, V)) \\ + \lambda_c(\widetilde{\varphi}_1(t, V + \eta_{Vc}) - \widetilde{\varphi}_1(t, V)) = 0, \\ \widetilde{\varphi}_1(T, V) = \beta_1^2, \end{cases} \tag{A15}$$

which has the following form

$$\widetilde{\varphi}_1(t, V) = e^{\widetilde{\varphi}_{11}(t) + \widetilde{\varphi}_{12}(t)V}, \tag{A16}$$

with terminal condition $\widetilde{\varphi}_1(T,V)=\beta_1^2$. Thus,

$$\begin{array}{lll}\widetilde{\varphi}_{1t}=(\widetilde{\varphi}'_{11}+\widetilde{\varphi}'_{12}V)\widetilde{\varphi}_1, & \widetilde{\varphi}_{1V}=\widetilde{\varphi}_{12}\widetilde{\varphi}_1, & \widetilde{\varphi}_{1VV}=\widetilde{\varphi}_{12}^2\widetilde{\varphi}_1,\\ \widetilde{\varphi}_1(t,V+\eta_{VV})-\widetilde{\varphi}_1(t,V)=\left[e^{\widetilde{\varphi}_{12}(t)\eta_{VV}}-1\right]\widetilde{\varphi}_1, & & \\ \widetilde{\varphi}_1(t,V+\eta_{Vc})-\widetilde{\varphi}_1(t,V)=\left[e^{\widetilde{\varphi}_{12}(t)\eta_{Vc}}-1\right]\widetilde{\varphi}_1, & & \end{array} \tag{A17}$$

Substituting Equations (A16) and (A17) into Equation (A15), we obtain

$$\widetilde{\varphi}'_{11}+\widetilde{\varphi}'_{12}V+a_1(t)+\kappa(\delta-V)\widetilde{\varphi}_{12}+\frac{1}{2}\sigma_V^2V\widetilde{\varphi}_{12}^2+\lambda_V\left[e^{\widetilde{\varphi}_{12}\eta_{VV}}-1\right]+\lambda_c\left[e^{\widetilde{\varphi}_{12}\eta_{Vc}}-1\right]=0. \tag{A18}$$

Since Equation (A18) holds for every V, the following two equation systems hold:

$$\begin{cases}\widetilde{\varphi}'_{12}-\kappa\widetilde{\varphi}_{12}+\dfrac{1}{2}\sigma_V^2\widetilde{\varphi}_{12}^2=0,\\ \widetilde{\varphi}_{12}(T)=0,\end{cases}$$

$$\begin{cases}\widetilde{\varphi}'_{11}+a_1(t)+\kappa\delta\widetilde{\varphi}_{12}+\lambda_V\left[e^{\widetilde{\varphi}_{12}\eta_{VV}}-1\right]+\lambda_c\left[e^{\widetilde{\varphi}_{12}\eta_{Vc}}-1\right]=0,\\ \widetilde{\varphi}_{11}(T)=\ln\beta_1^2.\end{cases}$$

Solving the above two systems, we have $\widetilde{\varphi}_{11}(t)=\ln\beta_1^2+\int_t^T a_1(s)ds$ and $\widetilde{\varphi}_{12}(t)=0$; thus, $\widetilde{\varphi}_1(t,V)$ is independent of the variable V, which can be written as

$$\widetilde{\varphi}_1(t)=\beta_1^2e^{\int_t^T a_1(s)ds},$$

and the system (A15) can be rewritten as

$$\begin{cases}\widetilde{\varphi}_{1t}+a_1(t)\widetilde{\varphi}_1=0,\\ \widetilde{\varphi}_1(T)=\beta_1^2.\end{cases}$$

Now, we solve Equation (A10). Let T be a variable in $\widetilde{\varphi}_1$, i.e., $\widetilde{\varphi}_1(t)=\widetilde{\varphi}_1(t,T)=e^{\widetilde{\varphi}_{11}(t,T)}$, where $\widetilde{\varphi}_{11}(t,s)=\ln\beta_1^2+\int_t^s a_1(u)du$. We conjecture that

$$\varphi_1(t)=\widetilde{\varphi}_1(t,T)+\Big[\int_t^T\widetilde{\varphi}_1(t,s)\lambda\beta_2^2ds\Big]\beta_1^{-2}, \tag{A19}$$

thus

$$\varphi_{1t}=\widetilde{\varphi}_{1t}+\Big[\int_t^T\widetilde{\varphi}_{1t}(t,s)\lambda\beta_2^2ds\Big]\beta_1^{-2}-\lambda\beta_2^2. \tag{A20}$$

Substituting Equations (A19) and (A20) into the left-hand side of Equation (A10), we have

$$\begin{aligned}&\widetilde{\varphi}_{1t}+\Big[\int_t^T\widetilde{\varphi}_{1t}(t,s)\lambda\beta_2^2ds\Big]\beta_1^{-2}-\lambda\beta_2^2+a_1(t)\left[\widetilde{\varphi}_1(t,T)+\Big[\int_t^T\widetilde{\varphi}_1(t,s)\lambda\beta_2^2ds\Big]\beta_1^{-2}\right]+\lambda\beta_2^2\\ &=\widetilde{\varphi}_{1t}+a_1(t)\widetilde{\varphi}_1(t,T)+\Big[\int_t^T\left[\widetilde{\varphi}_{1t}(t,s)+a_1(t)\widetilde{\varphi}_1(t,s)\right]\lambda\beta_2^2ds\Big]\beta_1^{-2}\\ &=0.\end{aligned}$$

Thus, $\varphi_1(t)$ given by Equation (A19) is the solution of system (A10), which is finally given by Equation (19). Similarly, $\varphi_2(t)$ is given by Equation (20) and $\varphi_6(t)$ is given by Equation (24).

Next, we solve Equation (A13). Since the coefficient of φ_5 and the constant term $2\xi\varphi_1$ are both independent of V, we suppose that $\varphi_5(t,V)$ is independent of V and rewrite it as $\varphi_5(t)$. Thus, Equation (A13) can be rewritten as

$$\varphi_{5t}+(a_5(t)+\lambda_c\eta_{Lc})\varphi_5+2\xi\varphi_1=0,$$

with terminal value $\varphi_5(T) = 0$, and $\varphi_5(t)$ is given by Equation (23).

Next, we solve Equation (A12). Since φ_1, φ_2 and φ_5 are all independent of V, and the coefficient a_4 is only dependent on time t, we assume that φ_4 is also independent of V, which satisfies the following system:

$$\begin{cases} \varphi_{4t} + (a_4(t) + \lambda_c \eta_{Lc})\varphi_4 - \dfrac{\varphi_2 \varphi_5}{2\varphi_1} \cdot \dfrac{\varpi_1(\varpi_1 + \varpi_2)}{\varpi_4} + \xi \varphi_2 = 0, \\ \varphi_4(T) = 0, \end{cases}$$

thus, φ_4 is given by Equation (22).

Since φ_1 and φ_5 are both independent of V, set

$$f_3(t) = -\frac{\varphi_5^2}{4\varphi_1} \cdot \frac{(\varpi_1 + \varpi_2)^2}{\varpi_4} + \xi \varphi_5.$$

Let $\widetilde{\varphi}_3 = \widetilde{\varphi}_3(t, V; \tau)$ be the solution of

$$\begin{cases} \widetilde{\varphi}_{3t}(t, V; \tau) + (a_3(t) + V)\widetilde{\varphi}_3(t, V; \tau) + (\kappa(\delta - V) + 2\sigma_V \rho_{LV} V)\widetilde{\varphi}_{3V}(t, V; \tau) \\ + \dfrac{1}{2}\sigma_V^2 V \widetilde{\varphi}_{3VV}(t, V; \tau) + \lambda_V(\widetilde{\varphi}_3(t, V + \eta_{VV}; \tau) - \widetilde{\varphi}_3(t, V; \tau)) + \lambda_c(\widetilde{\varphi}_3(t, V + \eta_{Vc}; \tau) \\ - \widetilde{\varphi}_3(t, V; \tau)) + \lambda_c \widetilde{\varphi}_3(t, V + \eta_{Vc}; \tau)(\eta_{Lc}^2 + 2\eta_{Lc}) = 0, \\ \widetilde{\varphi}_3(\tau, V; \tau) = f_3(\tau), \end{cases} \tag{A21}$$

and we have the following proposition.

Proposition A1. *The solution of Equation (A11) can be expressed as*

$$\varphi_3(t, V) = \int_t^T \widetilde{\varphi}_3(t, V; \tau) d\tau. \tag{A22}$$

Proof. First, we have $\varphi_3(T, V) = \int_T^T \widetilde{\varphi}_3 d\tau = 0$. Set $\tau = t$ in the second equation of Equation (A21); thus, we have $\widetilde{\varphi}_3(t, V; t) = f_3(t)$. Differentiating Equation (A22) with respect to t and V, respectively, we have

$$\begin{aligned} \varphi_{3t} &= \int_t^T \widetilde{\varphi}_{3t}(t, V; \tau) d\tau - \widetilde{\varphi}_3(t, V; t) = \int_t^T \widetilde{\varphi}_{3t}(t, V; \tau) d\tau - f_3(t), \\ \varphi_{3V} &= \int_t^T \widetilde{\varphi}_{3V}(t, V; \tau) d\tau, \qquad \varphi_{3VV} = \int_t^T \widetilde{\varphi}_{3VV}(t, V; \tau) d\tau. \end{aligned}$$

Substituting φ_{3t}, φ_{3V} and φ_{3VV} into Equation (A11)

$$\begin{aligned} &\int_t^T \widetilde{\varphi}_{3t}(t, V; \tau) d\tau - f_3(t) + (a_3(t) + V) \int_t^T \widetilde{\varphi}_3(t, V; \tau) d\tau + (\kappa(\delta - V) + 2\sigma_V \rho_{LV} V) \\ &\cdot \int_t^T \widetilde{\varphi}_{3V}(t, V; \tau) d\tau + \frac{1}{2}\sigma_V^2 V \int_t^T \widetilde{\varphi}_{3VV}(t, V; \tau) d\tau + \lambda_V \int_t^T \widetilde{\varphi}_3(t, V + \eta_{VV}; \tau) - \widetilde{\varphi}_3(t, V; \tau) d\tau \\ &+ \lambda_c \int_t^T \widetilde{\varphi}_3(t, V + \eta_{Vc}; \tau) - \widetilde{\varphi}_3(t, V; \tau) d\tau + \lambda_c \int_t^T \widetilde{\varphi}_3(t, V + \eta_{Vc}; \tau) d\tau(\eta_{Lc}^2 + 2\eta_{Lc}) + f_3(t) \\ &= \int_t^T 0 d\tau = 0. \end{aligned}$$

□

Now, we start solving Equation (A21). Suppose $\widetilde{\varphi}_3 = \widetilde{\varphi}_3(t, V; \tau) = \widetilde{\varphi}_{31}(t; \tau) e^{\widetilde{\varphi}_{32}(t;\tau)V}$, with terminal value $\widetilde{\varphi}_3(\tau, V; \tau) = \widetilde{\varphi}_{31}(\tau) e^{\widetilde{\varphi}_{32}(\tau)V} = f_3(\tau)$. Thus,

$$\begin{aligned}&\widetilde{\varphi}_{3t}=\left[\frac{\widetilde{\varphi}_{31}^{'}}{\widetilde{\varphi}_{31}}+\widetilde{\varphi}_{32}^{'}V\right]\widetilde{\varphi}_3,\qquad \widetilde{\varphi}_{3V}=\widetilde{\varphi}_{32}\widetilde{\varphi}_3,\qquad \widetilde{\varphi}_{3VV}=\widetilde{\varphi}_{32}^2\widetilde{\varphi}_3,\\&\widetilde{\varphi}_3(t,V+\eta_{VV};\tau)-\widetilde{\varphi}_3=(e^{\widetilde{\varphi}_{32}\eta_{VV}}-1)\widetilde{\varphi}_3,\qquad \widetilde{\varphi}_3(t,V+\eta_{Vc};\tau)=e^{\widetilde{\varphi}_{32}\eta_{Vc}}\widetilde{\varphi}_3.\end{aligned}\tag{A23}$$

Substituting Equation (A23) into Equation (A21) with the consideration of the terminal value, we obtain the following two systems:

$$\begin{cases}\widetilde{\varphi}_{32}^{'}+1+(2\sigma_V\rho_{LV}-\kappa)\widetilde{\varphi}_{32}+\frac{1}{2}\sigma_V^2\widetilde{\varphi}_{32}^2=0,\\ \widetilde{\varphi}_{32}(\tau)=0,\end{cases}\tag{A24}$$

$$\begin{cases}\frac{\widetilde{\varphi}_{31}^{'}}{\widetilde{\varphi}_{31}}+a_3+\kappa\delta\widetilde{\varphi}_{32}+\lambda_V(e^{\widetilde{\varphi}_{32}\eta_{VV}}-1)+\lambda_c(e^{\widetilde{\varphi}_{32}\eta_{Vc}}-1)+\lambda_c e^{\widetilde{\varphi}_{32}\eta_{Vc}}(\eta_{Lc}^2+2\eta_{Lc})=0,\\ \widetilde{\varphi}_{31}(\tau)=f_3(\tau).\end{cases}\tag{A25}$$

We solve system (A24) first. Rewrite the first equation as

$$\widetilde{\varphi}_{32}^{'}=-\frac{1}{2}\sigma_V^2\widetilde{\varphi}_{32}^2-(2\sigma_V\rho_{LV}-\kappa)\widetilde{\varphi}_{32}-1.$$

Let $\triangle_3=(2\sigma_V\rho_{LV}-\kappa)^2-2\sigma_V^2$ be the discriminant of the following quadratic equation

$$-\frac{1}{2}\sigma_V^2\widetilde{\varphi}_{32}^2-(2\sigma_V\rho_{LV}-\kappa)\widetilde{\varphi}_{32}-1=0.\tag{A26}$$

If $\triangle_3>0$, then the two real roots $h_{1,2}$ of Equation (A26) can be expressed as

$$h_{1,2}=\frac{(2\sigma_V\rho_{LV}-\kappa)-\sqrt{\triangle_3}}{\sigma_V^2}.$$

Thus,

$$\widetilde{\varphi}_{32}(t)=\frac{h_1h_2e^{-\sqrt{\triangle_3}(\tau-t)}-h_1h_2}{h_1e^{-\sqrt{\triangle_3}(\tau-t)}-h_2}.$$

If $\triangle_3=0$, then we have

$$\widetilde{\varphi}_{32}(t)=\frac{2\sigma_V\rho_{LV}-\kappa}{\sigma_V^2+\frac{1}{2}\sigma_V^2(\tau-t)(2\sigma_V\rho_{LV}-\kappa)}-\frac{2\sigma_V\rho_{LV}-\kappa}{\sigma_V^2}.\tag{A27}$$

If $\triangle_3<0$, then

$$\widetilde{\varphi}_{32}(t)=\sqrt{-\frac{\triangle_3}{\sigma_V^4}}\tan\left[\arctan\left[\frac{2\sigma_V\rho_{LV}-\kappa}{\sqrt{-\triangle_3}}\right]+\frac{1}{2}\sqrt{-\triangle_3}(\tau-t)\right]-\frac{2\sigma_V\rho_{LV}-\kappa}{\sigma_V^2}.$$

The solution of system (A25) is

$$\widetilde{\varphi}_{31}(t)=e^{\int_t^\tau f_{31}(s)ds}\cdot f_3(\tau),$$

where

$$f_{31}(t)=a_3+\kappa\delta\widetilde{\varphi}_{32}+\lambda_V(e^{\widetilde{\varphi}_{32}\eta_{VV}}-1)+\lambda_c(e^{\widetilde{\varphi}_{32}\eta_{Vc}}-1)+\lambda_c e^{\widetilde{\varphi}_{32}\eta_{Vc}}(\eta_{Lc}^2+2\eta_{Lc}),\tag{A28}$$

thus, $\varphi_3(t,V)$ is given by Equation (21).

It is obvious that $2\varphi_1(t)>0$. Inserting φ_1, φ_2 and φ_5 into Equation (A4), the optimal investment strategy is given by Theorem 1.

References

1. Bodie, Z.; Detemple, J.B.; Otruba, S.; Walter, S. Optimal consumption-portfolio choices and retirement planning. *J. Econ. Dyn. Control* **2004**, *28*, 1115–1148. [CrossRef]
2. Dybvig, P.H.; Liu, H. Lifetime consumption and investment: Retirement and constrained borrowing. *J. Econ. Theory* **2010**, *145*, 885–907. [CrossRef]
3. Guan, G.; Liang, Z. Optimal management of DC pension plan in a stochastic interest rate and stochastic volatility framework. *Insur. Math. Econ.* **2014**, *57*, 58–66. [CrossRef]
4. Wang, P.; Lu, Z. Robust optimal investment strategy for an AAM of DC pension plans with stochastic rate and stochastic volatility. *Insur. Math. Econ.* **2018**, *80*, 67–83. [CrossRef]
5. Zeng, Y.; Li, D.; Chen, Z.; Yang, Z. Ambiguity aversion and optimal derivative-based pension investment with stochastic income and volatility. *J. Econ. Dyn. Control* **2018**, *88*, 70–103. [CrossRef]
6. Liang, Z.; Bi, J.; Yuen, K.C.; Zhang, C. Optimal mean-variance reinsurance and investment in a jump-diffusion financial market with common shock dependence. *Math. Methods Oper. Res.* **2016**, *84*, 155–181. [CrossRef]
7. Liang, Z.; Yuen, K.C.; Zhang, C. Optimal reinsurance and investment in a jump-diffusion financial market with common shock dependence. *J. Appl. Math. Comput.* **2018**, *56*, 637–664. [CrossRef]
8. Han, N.; Hung, M. Optimal asset allocation for DC pension plans under inflation. *Insur. Math. Econ.* **2012**, *51*, 172–181. [CrossRef]
9. Yao, H.; Yang, Z.; Chen, P. Markowitz's mean-variance defined contribution pension fund management under inflation: A continuous-time model. *Insur. Math. Econ.* **2013**, *53*, 851–863. [CrossRef]
10. Chen, Z.; Li, Z.; Yan, Z.; Sun, J. Asset allocation under loss aversion and minimum performance constraint in a DC pension plan with inflation risk. *Insur. Math. Econ.* **2017**, *75*, 137–150. [CrossRef]
11. Tang, M.; Chen, S.; Lai, G.C.; Wu, T. Asset allocation for a DC pension fund under stochastic interest rates and inflation-protected guarantee. *Insur. Math. Econ.* **2018**, *78*, 87–104. [CrossRef]
12. Zhang, A.; Ralf, K.; Ewald, C. Optimal management and inflation protection for defined contribution pension plans. *Bl. DGVFM* **2007**, *28*, 239–258. [CrossRef]
13. Zhang, A.; Ewald, C. Optimal investment for a pension fund under inflation risk. *Math. Methods Oper. Res.* **2010**, *71*, 353–369. [CrossRef]
14. Ma, Y.K.; Johnson, M.; Vijayakumar, V.; Radhika, T.; Shukla, A.; Nisar, K.S. A note on approximate controllability of second-order impulsive stochastic Volterra-Fredholm integrodifferential system with infinite delay. *J. King Saud Univ.-Sci.* **2023**, *35*, 1018–3647. [CrossRef]
15. Hakkar, N.; Dhayal, R.; Debbouche, A.; Torres, D.F.M. Approximate Controllability of Delayed Fractional Stochastic Differential Systems with Mixed Noise and Impulsive Effects. *Fractal Fract.* **2023**, *7*, 104. [CrossRef]
16. Dhayal, R.; Malik, M.; Abbas, S.; Debbouche, A. Optimal controls for second-order stochastic differential equations driven by mixed-fractional Brownian motion with impulses. *Math. Meth. Appl. Sci.* **2020**, *43*, 4107–4124. [CrossRef]
17. Johnson, M.; Vijayakumar, V. An analysis on the optimal control for fractional stochastic delay integrodifferential systems of order $1 < \gamma < 2$. *Fractal Fract.* **2023**, *7*, 284.
18. Halder, S.; Sharma, H.K.; Biswas, A.; Prentkovskis, O.; Majumder, S.; Skačkauskas, P. On Enhanced Intelligent Water Drops Algorithm for Travelling Salesman Problem under Uncertain Paradigm. *Transp. Telecommun.* **2023**, *24*, 228–255. [CrossRef]
19. Mwanakatwe, P.K.; Wang, X.; Su, Y. Optimal investment and risk control strategies for an insurance fund in stochastic framework. *J. Math. Financ.* **2019**, *9*, 254–265. [CrossRef]
20. Swishchuk, A.; Zagst, R.; Zeller, G. Hawkes processes in insurance: Risk model, application to empirical data and optimal investment. *Insur. Math. Econ.* **2021**, *101*, 107–124. [CrossRef]
21. Merton, R.C. Optimal consumption and portfolio rules in a continuous-time model. *J. Econ. Theory* **1971**, *3*, 373–413. [CrossRef]
22. Sun, J.; Li, Z.; Zeng, Y. Precommitment and equilibrium investment strategies for defined contribution pension plans under a jump-diffusion model. *Insur. Math. Econ.* **2016**, *67*, 158–172. [CrossRef]
23. Delong, L.; Gerrard, R.; Haberman, S. Mean-variance optimization problems for an accumulation phase in a defined benefit plan. *Insur. Math. Econ.* **2008**, *42*, 107–118. [CrossRef]
24. Liang, X.; Bai, L.; Guo, J. Optimal time-consistent portfolio and contribution selection for defined benefit pension schemes under mean-variance criterion. *ANZIAM J.* **2014**, *56*, 66–90.
25. Ngwira, B.; Gerrard, R. Stochastic pension fund control in the presence of Poisson jumps. *Insur. Math. Econ.* **2007**, *40*, 283–292. [CrossRef]
26. Mudzimbabwe, W. A simple numerical solution for an optimal investment strategy for a DC pension plan in a jump diffusion model. *J. Comput. Appl. Math.***2019**, *360*, 55–61. [CrossRef]
27. Zhang, X. Optimal DC pension management under inflation risk with jump diffusion price index and cost of living process. *Methodol. Comput. Appl. Probab.* **2022**, *24*, 1253–1270. [CrossRef]
28. Guambe, C.; Kufakunesu, R.; Van Zyl, G.; Beyers, C. Time consistent mean-variance asset allocation for a DC plan with regime switching under a jump-diffusion model. *Jpn. J. Ind. Appl. Math.* **2022**, *39*, 119–143. [CrossRef]
29. Eisenberg, J. Optimal dividends under a stochastic interest rate. *Insur. Math. Econ.* **2015**, *65*, 259–266. [CrossRef]
30. Wang, S.; Lu, Y.; Sanders, B. Optimal investment strategies and intergenerational risk sharing for target benefit pension plans. *Insur. Math. Econ.* **2018**, *80*, 1–14. [CrossRef]

31. Devolder, P.; Janssen, J.; Manca, R. *Stochastic Methods for Pension Funds*; Wiley: New York, NY, USA, 2012.
32. Zhang, X.; Guo, J. Optimal defined contribution pension management when risky asset and salary follow jump diffusion processes. *East Asian J. Appl. Math.* **2020**, *10*, 22–39.
33. Fleming, W.H.; Soner, H.M. *Controlled Markov Processes and Viscosity Solutions*; Springer: New York, NY, USA, 1993.
34. Yong, X.; Zhou, X.Y. *Stochastic Controls: Hamiltonian Systems and HJB Equations*; Springer: New York, NY, USA, 1999.
35. Øksendal B.K.; Sulem A. *Applied Stochastic Control of Jump Diffusions*; Springer: Berlin, Germany, 2007.
36. Schmidli, H. *Stochastic Control in Insurance*; Springer: London, UK, 2007.
37. Azcue P.; Muler, N. *Stochastic Optimization in Insurance: A Dynamic Programming Approach*; Springer: Berlin/Heidelberg, Germany, 2014.
38. Browne, S. Optimal investment policies for a firm with a random risk process: Exponential utility and minimizing the probability of ruin. *Math. Oper. Res.* **1995**, *20*, 937–958. [CrossRef]
39. Hipp, C.; Plum, M. Optimal investment for insurers. *Insur. Math. Econ.* **2000** *27*, 215–228. [CrossRef]
40. Hipp, C.; Taksar, M. Optimal non-proportional reinsurance control. *Insur. Math. Econ.* **2010**, *47*, 246–254. [CrossRef]

mathematics

MDPI

Article

Tariff Analysis in Automobile Insurance: Is It Time to Switch from Generalized Linear Models to Generalized Additive Models?

Zuleyka Díaz Martínez [1,†], José Fernández Menéndez [2,†] and Luis Javier García Villalba [3,*,†]

[1] Group of Analysis, Security and Systems (GASS), Department of Financial and Actuarial Economics & Statistics, Faculty of Economics and Business Administration, Universidad Complutense de Madrid (UCM), Campus Somosaguas, 28223 Madrid, Spain; zuleyka@ccee.ucm.es
[2] Department of Business Administration, Faculty of Economics and Business Administration, Universidad Complutense de Madrid (UCM), Campus Somosaguas, 28223 Madrid, Spain; jfernan@ccee.ucm.es
[3] Group of Analysis, Security and Systems (GASS), Department of Software Engineering and Artificial Intelligence (DISIA), Faculty of Computer Science and Engineering, Office 431, Universidad Complutense de Madrid (UCM), Calle Profesor José García Santesmases, 9, Ciudad Universitaria, 28040 Madrid, Spain
* Correspondence: javiergv@fdi.ucm.es; Tel.: +34-91-394-7638
† These authors contributed equally to this work.

Abstract: Generalized Linear Models (GLMs) are the standard tool used for pricing in the field of automobile insurance. Generalized Additive Models (GAMs) are more complex and computationally intensive but allow taking into account nonlinear effects without the need to discretize the explanatory variables. In addition, they fit perfectly into the mental framework shared by actuaries and are easier to use and interpret than machine learning models, such as trees or neural networks. This work compares both the GLM and GAM approaches, using a wide sample of policies to assess their differences in terms of quality of predictions, complexity of use, and time of execution. The results show that GAMs are a powerful alternative to GLMs, particularly when "big data" implementations of GAMs are used.

Keywords: automobile insurance; generalized additive models; splines; tariff analysis

MSC: 97M30

Citation: Díaz Martínez, Z.; Fernández Menéndez, J.; García Villalba, L.J. Tariff Analysis in Automobile Insurance: Is It Time to Switch from Generalized Linear Models to Generalized Additive Models? *Mathematics* **2023**, *11*, 3906. https://doi.org/10.3390/math11183906

Academic Editors: Jing Yao, Xiang Hu and Jingchao Li

Received: 16 August 2023
Revised: 8 September 2023
Accepted: 11 September 2023
Published: 14 September 2023

1. Introduction: The Classical Tariff Analysis in the Actuarial Field

Since their original development by J. A. Nelder and R. W. M. Wedderburn [1], Generalized Linear Models (GLMs) have become a key methodology in applied statistics. GLMs burst onto the insurance scene in the early 1990s in works such as [2,3] and became part of the set of tools commonly used by actuaries, with a number of monographs devoted to showing their application, as well as those of their extensions, to the field of insurance [4–6].

The GLM provides a common framework that includes a broad range of regression models (ordinary least squares, logistic, Poisson, etc.) previously lacking a unified treatment. The essential requirement for a particular model that can be treated as a GLM is that the probability distribution of the variable analyzed belongs to the class of the exponential family [7,8]. In a GLM, the expectation of the dependent variable, $\mu = E(y)$, is modeled as a function of the linear predictor $\eta = \sum \beta_i x_i$, which includes the explanatory variables x_i. Therefore, we have that $\mu = l^{-1}(\eta)$ or, conversely, that $\eta = l(\mu)$. The function $l()$ is called a link function and can be selected with a certain degree of flexibility [9]. The variance $var(y)$ is customarily expressed in the GLM setting as a function of the mean in the following manner: $var(y) = \frac{\varphi}{A} V(\mu)$. In this expression, $V(\mu)$ is the so-called variance function, φ is the dispersion parameter or dispersion for short, and A is the known weight of each observation.

For instance, in the case of a Poisson regression model, the dependent variable, y, is the number of occurrences of an event in a time interval. By defining $\mu = E(y) = \lambda$ and taking, as is usually the case, a *log* link, the model becomes $\log(\lambda) = \sum \beta_i x_i$. Also, if the weights are $A = 1$, and because of the equality of mean and variance in a Poisson-distributed random variable, we obtain $var(y) = \lambda = \mu = \varphi V(\mu)$, which implies that the dispersion parameter is $\varphi = 1$ and the variance function is $V(\mu) = \mu$.

In any case, the explanatory variables enter the model through the linear predictor $\eta = \sum \beta_i x_i$, and therefore their effects are linear, making GLMs "linear" models. A classical ordinary least-squares model (OLS) is a GLM with $V(\mu) = 1$ as the variance function and the identity as the link function. This means that the parameter modeled is the mean of the dependent variable, μ, as a linear function of the covariates $x_i : \mu = \sum \beta_i x_i$. If x_i are continuous, their eventual nonlinear effects can be included in an OLS model, adding to the linear predictor of the corresponding quadratic, cubic, etc., terms, so that the linear predictor has the form $\eta = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \ldots$, or, likewise, in the case of a GLM, $l(\eta) = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \ldots$.

Another, simpler way to introduce nonlinear effects into a GLM is by discretizing the continuous variables x_i.

The discretization implies an information loss regarding the original variable and requires the determination of the number of categories and the intervals that define the bins, which can be complex and subject to some degree of arbitrariness. Nevertheless, the discretization of an explanatory variable can help in applying and interpreting the model and provides a straightforward procedure to take into account an eventual nonlinear effect of the variable. Also, the discretization of the variable increases the number of parameters to be estimated. For a categorical variable with k levels, $k - 1$ parameters should be estimated. This provides enough flexibility for the model to be capable of including nonlinearities.

In the classical tariff analysis in the automobile insurance line of business, the customary practice consists of discretizing all the continuous explanatory variables used to estimate the risk associated with a policy (like the driver's age, horsepower of the vehicle, etc.) and generating, through the Cartesian product of the levels of these discretized variables, a set of "cells" or "policy groups". The cell a policy belongs to determines the premium that is charged. This simplifies the tariff application (the calculus of the premium of a new policy merely consists of identifying the cell that corresponds to it). The pure premium of a cell is usually estimated as the product of the claim frequency and the claim severity associated with the cell. Both the claim frequency and claim severity share the same structure, and they are quotients between a risk measure and a risk exposure [10]. The claim frequency is the ratio between the total number of claims in a cell and the total exposure of the cell, which is the total time that the policies in the cell have been in force. The claim severity is the mean claim size, i.e., the quotient between the total size of the claims in a cell and the number of claims originated by the policies in the cell.

The claim frequency is usually modeled as a rate model, i.e., a Poisson regression model where the number of claims in each cell is the dependent variable and the exposure (in this case, the total time the policies are in force) is taken as an offset [2,10], i.e., an explanatory variable whose coefficient is not estimated but taken as 1 [7].

With regard to the claim severity, as this is a continuous and positive variable, it is customarily modeled as the dependent variable in a Gamma regression in which the exposure (number of claims) is introduced as a weight for each observation, i.e., for each cell in this case [2,10]. Usually, both for the claim frequency and the claim severity, the link function of the regression model is the logarithm. This means that the risk and the associated premium corresponding to a cell can be expressed as the product of a factor specific to the cell multiplied by the risk of the cell taken as a baseline or reference. In the motor insurance field, this multiplicative effects model is generally considered more suitable for calculating premiums than an additive effects model [2,10].

These kinds of models are widely used for the claim frequency and the claim severity and can be considered standard in the automobile insurance field. A drawback is the need to determine how to discretize the explanatory variables, which can be somewhat arbitrary and may also require a huge amount of trial and error to find a discretization that groups the policies in sufficiently homogenous cells, but at the same time keeps low enough the number of cells so that the number of parameters in the model does not become unmanageable. Another problem is that the within-cell variability neglected by the model can be appreciable unless the cell size becomes very small and the number of cells becomes very high.

There exists, however, the possibility of working with the original non-discretized continuous variables. The main problem with this approach is how to treat nonlinear effects, for example, that of the driver's age, which does not show a linear relationship with the claim risk.

The most traditional way is to include polynomial terms (quadratic, cubic, etc.) as regressors to deal with the nonlinearities of the data, but this is far from being satisfactory because polynomials, in a regression context, tends to cause problems [11–14]. These problems are a consequence of the fact that polynomial regression is an ill-conditioned problem because its matrix model includes a Vandermonde matrix, and it is well-known that Vandermonde matrices have a very high condition number and pose serious problems of numerical instability [15–18]. Hence, polynomial regression should be used with caution, and the order of the polynomials should be kept as low as possible.

However, there are more elaborate alternatives that can solve these problems, which are based on the use of splines with Generalized Additive Models (GAM). The GAM [19] is a generalization of the GLM that introduces the explanatory variables in the linear predictor, not directly but through a spline of the variable, i.e., through a continuous piecewise polynomial of the variable that allows for the modeling of nonlinearities, thereby eliminating the problems associated with the simpler and more conventional polynomial terms. These models have been suggested as a reasonable alternative to the traditional pricing models based on discretized risk factors [10,20]. Instead of laboriously and somewhat arbitrarily calculating the values according to which each variable will be divided into a series of levels, GAMs allow this to be done automatically and maintain the traditional structure of regression models, which is easy to understand and interpret. For this reason, the aim of this work is to compare a classical model of tariff analysis in the automobile insurance field estimated using a GLM and discretized explanatory variables to a similar model estimated using a GAM, with its explanatory variables retaining their original, continuous, and non-discretized form. Our goal is to assess the improvement achieved by the greater amount of information supplied by the continuous variables and evaluate if this improvement exceeds the greater complexity, both conceptual and computational, of the GAM model.

Another kind of method widely used for tariff analysis in automobile insurance is de Bonus–Malus Systems (BMSs). A BMS includes a posteriori information about the track record of each policy or each policyholder in the portfolio. With this information, the policies are classified into different risk levels. Some a priori rating variables, like age, cubic capacity of the car, etc., can also be included in the model. Examples of this approach can be found in [21,22]. Nevertheless, we do not consider the Bonus–Malus Systems here.

The main finding of this work is that a Generalized Additive Model is a tool for calculating premiums in the field of automobile insurance that can take into account the nonlinear effect of some of the pricing variables in an automatic way and without the need to identify the intervals in which to divide the values of the rating factors, similar to a traditional GLM.

The rest of the work is organized as follows. Section 2 briefly reviews the scientific literature on GAM usage in several fields. Section 3 provides an overview of the theoretical foundations of Generalized Additive Models. Section 4 compares the results obtained using GAMs and GLMs for motor insurance ratings with a large sample from a Spanish insurance company. Finally, Section 5 presents the main results obtained in this work.

2. Literature Review

One possibility for addressing nonlinearities in data is to use modern methods from the field of machine learning and neural networks. For example, an early attempt by Mulquiney [23] to compare GLMs to MARS (Multivariate Adaptive Regression Splines) [24], MARTs (Multiple Additive Regression Trees) [25], and neural networks achieved mixed results but ultimately favored GLMs. Furthermore, a comparison between XGBoost and GLMs (logistic regression) for predicting motor insurance claims in [26] also favored GLMs. In contrast, Henckaerts and colleagues [27] found that carefully crafted Boosted Trees outperformed classical GLMs for insurance pricing. Ticconi [28] compared GLMs to neural networks and Support Vector Machines (SVMs) for credit insurance analysis, with the results demonstrating the superiority of SVMs. Baillargeon and colleagues [29] used a Hierarchical Attention Network (a type of neural network for document classification) to analyze risk factors from a textual dataset describing car accidents, using a GLM as a baseline for comparison. Delong and Wüthrich [30] trained a neural network to develop models for the process of claim payments and claims incurred for property or bodily injury. They used GLMs, Generalized Additive Models, and Regression Trees as starting models to feed into the neural network. A good overview of other applications of machine learning tools to actuarial science can be found in [31].

As GAMs are particularly suitable for modeling nonlinear relationships between variables, they have been widely and routinely used in several research fields since their inception. For example, in the field of biology, they have been used to analyze the spatial distribution of fishes and vegetal species [32–40]. Comparisons of the relative performance of models estimated using GLMs or GAMs can be found in [41–45]. GAMs have also been frequently combined or compared to GLMs in environmental sciences [46–49], climatology [50], meteorology [51], neuroscience [52,53], and genetics [54]. In the actuarial literature, some works have applied the GAM methodology to various problems in the insurance field [55–60].

3. Generalized Additive Models

A Generalized Additive Model (GAM) is very similar to a GLM but with the key difference that the explanatory variables are introduced in the model through an "additive" predictor of the form $\eta = \beta_0 + f_1(x_1) + \ldots + f_p(x_p)$. This predictor replaces the linear predictor of the form $\eta = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p$, typical of a GLM. The result is that some smooth functions $f_k(x_k)$ of the independent variables replace the simpler linear terms of the form $\beta_k x_k$, typical of a GLM. The functions $f_k()$ can be selected from different function spaces, provided that they are flexible enough to approximate any functional shape; however, usually, they are some type of spline or kernel smoother [13,14,19]. Also, the fact that each explanatory variable enters the model through a specific additive term in the predictor, enables individualizing and clearly analyzing its effect on the response, independently of the other variables. Efficient algorithms for estimating these models can be found in [19].

In each model, the specific peculiarities of the functions $f_k()$ are evaluated as part of the fitting process. The model loses the "parametric" character of the GLM and becomes what is usually termed a "semiparametric" model [61]. These semiparametric models are a good compromise between a fully parametric model (simpler to fit and interpret but less flexible in incorporating nonlinear effects) and a totally nonparametric one, where the functions $f_k()$ are a priori entirely arbitrary. A nonparametric model has enormous flexibility but is very difficult to fit and interpret [62]. In the semiparametric case, the functions $f_k()$ are usually constructed, as mentioned previously, using some kind of kernel smoother or, more frequently, splines.

A kernel smoother enables the implementation of "local" regression: the estimated value in a point x_0 is calculated using the observed values at points close to it. To do this, a weight is assigned to every point x. The weight is determined by a kernel $K(x_0, x)$,

and the closer x is to x_0, the greater the weight value. Different kinds of kernels are used in practice. Some good revisions of this kind of technique can be found in [63,64].

A spline is a linear combination of polynomials, each defined over a specific bounded interval and null outside of this interval. In this way, it is possible to address one of the key problems of polynomials when used in interpolation and estimation: their non-local character, i.e., the fact that they are defined over the entire real line, so small changes in their values in one region can result in big changes in other, very distant regions. The most frequently used splines are cubic ones, composed of third-order polynomials. Splines are widely used in statistics and numerical analysis, and there is a huge amount of literature devoted to them. A classical reference on the subject is [65].

The boundaries of the intervals in which the splines are defined are a set of points termed knots. Given a set of knots $\xi_1 \leq \xi_2 \ldots \leq \xi_n$ belonging to an interval $[a, b]$, a cubic spline is a function based on a set of cubic polynomials, each one defined in one of the intervals $(a, \xi_1), (\xi_1, \xi_2), \ldots, (\xi_n, b)$ and null in the rest of them. The spline is defined so that it is continuous, and with first- and second-order derivatives, it is also continuous at the knots [13]. With these conditions, it can be easily seen that given n knots, a cubic spline is defined by $n + 4$ parameters. A natural cubic spline is a cubic spline that satisfies the four additional conditions that its second and third derivatives at the points a and b are zero. A natural cubic spline is defined by n parameters. In fact, given n knots in an interval $[a, b]$, the set of splines in this interval and with these knots forms a vector space of dimension $n + 4$ in the case of cubic splines, or n in the case of natural cubic ones.

There are many different ways to express the basis of these vector spaces. One of them is the basis of truncated powers of the space of cubic splines [14]. This basis includes the four monomial terms 1, x, x^2, and x^3 plus a set of n (one for each knot) third powers of the positive part functions of the form $\left(X - \xi_j\right)_+^3$. The positive part function $\left(X - \xi_j\right)_+$ is defined as:

$$\begin{cases} 0, & \text{if } X - \xi_j < 0 \\ X - \xi_j, & \text{if } X - \xi_j > 0 \end{cases} \tag{1}$$

Independently of the basis selected, if we denote $s_j()$ as the functions of the basis of the cubic splines defined over an interval (and with a specified set of knots), we find that every smooth function defined over such an interval can be approximated with a linear combination of s_j. In this way, the functions $f_k()$ of the predictor of a GAM can be expressed as linear combinations of the functions of the basis $f_k(x_k) \approx \sum_j \beta_j s_j(x_k)$, and the estimation of $f_k()$ reduces to the estimation of the coefficients β_j of these linear combinations. If the GAM reduces to an OLS model (because the dependent variable is normally distributed, and the link function is the identity), β_j can be estimated using least squares by minimizing the quantity $\| Y - \sum \beta_j s_j(x) \|^2$ in the usual way. This is completely analogous to the minimization of $\| Y - \sum \beta_j x^j \|^2$ in the case of a polynomial regression, where the monomials x^j play the role of the $s_k(x)$ basis functions.

This approach generalizes straightforwardly when the GAM corresponds to a Generalized Linear Model (i.e., when the variable is not normal or the link function is not the identity). Thus, the estimation of a GAM would be similar to that of a GLM, with the only change being the replacement of x^j with the splines $s_k(x)$. These types of splines, with the knots fixed in advance, are often called regression splines [14]. When used, the complexity or roughness of the model is controlled by adding or removing knots. The use of regression splines is very simple because it is merely a slight generalization of a conventional regression model. However, they pose the problem that one has to select the number and location of the knots, and the results can be quite sensitive to this choice. Moreover, the models constructed by adding and removing knots are not nested, which makes it difficult to select the most appropriate model.

It is possible to eliminate the problem of selecting knots through regularization, i.e., by adding a quadratic additional term that penalizes the curvature of the estimated functions

f_k to the quantity to be minimized, $\| Y - \sum \beta_j s_j(x) \|^2$. Usually, this term is of the form $\int (f''(x))^2 dx$. In this case, a penalized regression spline is obtained, which controls the complexity or roughness of the model through the weight assigned to the quadratic penalization term: the more the weight, the lesser the roughness. In the case of penalized regression splines, it is still necessary to choose the knots, but this selection has little impact on the final results, provided that the number of knots selected is high enough to lead to a good fitting of the model.

An additional generalization is provided by smoothing splines. In this case, the set of knots is maximal: every available value of the independent variable for which there is an observation becomes a knot. Such a model has an exceedingly high number of parameters and perfectly interpolates all the observations. This model lacks any interest, so it is necessary to reduce its complexity through regularization. This is done by selecting a function f from a determined function space $\mathcal{H}$ that minimizes the sum of a term penalizing curvature, along with another term that, for example, in the case of an OLS model, represents the sum of squared errors:

$$\min_{f \in \mathcal{H}} \left\{ \sum (y_i - f(x_i))^2 + \lambda \int (f''(x))^2 dx \right\} \quad (2)$$

The value of the smoothing parameter λ is chosen a priori and controls the amount of penalization of the curvature of f, so that the greater its value, the smoother the estimated model. If the function space $\mathcal{H}$ consists of differentiable functions with an absolutely continuous first derivative in an interval that contains the observations, it can be proven that the solution to the minimization problem above is a natural cubic spline with knots at the observations [13]. The function $f()$, the solution to the minimization problem, can thus be represented as a linear combination of the basis elements of the vector space of natural cubic splines, $f = \sum \beta_i s_i$. Therefore, the minimization problem becomes the following: $\min_\beta (y - S\beta)'(y - S\beta) + \lambda \beta' \Omega \beta$.

Here, we have that S is a matrix whose elements are $S_{ij} = s_j(x_i)$; Ω is a matrix of the elements $\Omega_{ij} = \int s_i'' s_j'' dx$; β is a vector whose elements are the parameters β_i that we want to estimate; and y is the vector with the observed values of the dependent variable. Formulated in this way, the problem is easily solved, and its solution is $\hat{\beta} = (S'S + \lambda\Omega)^{-1} S'y$ [13,14].

The conclusion is that the smoothing spline that best fits the data is the natural cubic spline with knots at the observations, which can be expressed as $\hat{f} = \sum \hat{\beta}_i s_i$. The values predicted by the model are $\hat{y} = S\hat{\beta} = S(S'S + \lambda\Omega)^{-1} S'y = Hy$. They are obtained as a linear transformation, Hy, of the observations. The matrix $H = S(S'S + \lambda\Omega)^{-1} S'$ plays a similar role to that of the hat matrix in an OLS regression. In OLS, the trace of the hat matrix provides the dimension of the linear subspace over which it projects the vector y of the observations. This dimension indicates the number of parameters, or degrees of freedom, of the model. Similarly, in the case of smoothing splines, the trace $tr(H)$ also indicates the "effective" degrees of freedom of the model. In general, they are not integer numbers, depend on the value of the smoothing parameter λ, and provide a measure of the complexity of the model. A high value of λ strongly penalizes the curvature of the spline f, which, therefore, tends to approach a straight line. Conversely, a low value of λ implies that the effective degrees of freedom of the model are high, and f becomes rougher, more complex, and "wrinkled". In fact, the complexity of the model can be controlled by selecting both the value of λ and the number of effective degrees of freedom.

The smoothing splines tend to be computationally expensive, but other than this, they have a good number of desirable properties and also easily generalize to higher dimensions. For example, in two dimensions, the term that penalizes the curvature of the spline becomes $\int \int \left[\left(\frac{\partial^2 f}{\partial x_1^2} \right)^2 + 2 \left(\frac{\partial^2 f}{\partial x_1 \partial x_2} \right) + \left(\frac{\partial^2 f}{\partial x_2^2} \right)^2 \right] dx_1 dx_2$ (instead of $\int (f''(x))^2 dx$). These smoothing splines in two or more dimensions are usually termed thin-plate splines [66].

4. Comparing the Two Approaches

As mentioned, the aim of this work is to compare the traditional approach to tariff analysis in the automobile insurance field, i.e., the one based on GLMs with discretized explanatory variables—the rating factors—with the one based on GAMs, which allows for the introduction of nonlinearities in a very flexible way and without discretizing the rating factors. To do this, we used data from the automobile policy portfolio of a Spanish insurance firm for the year 2005. The data were slightly cleaned, discarding some extremely atypical or clearly erroneous values; otherwise, they were used as found in the insurer's database. From the full portfolio, five random training samples, each containing 500,000 policies, were extracted. With each of these samples, the GLMs and GAMs were estimated and compared. Also, five additional random samples of 200,000 policies were extracted and used as test samples. Each one of these test samples was paired with one of the training samples to check the predictive accuracy of the estimated models. Initially, two kinds of models were estimated: one for the number of claims using Poisson regressions, and one for the claim size using Gamma regressions. The rating factors used as the independent variables in the models are described in Table 1.

Table 1. Rating Factors.

Name	Variable	Characteristics
TYPE_VE	Type or category of vehicle	Categorical variable with 6 levels.
USAGE	Usage of the vehicle	Categorical variable with 20 levels.
NATURE	Nature of the vehicle	Categorical variable with 2 levels.
PLACES	Number of seats	Count variable; treated as categorical with 8 levels.
AMBIT	Circulation area of the vehicle	Categorical variable with 8 levels.
VEH_AGE	Age of the vehicle	Continuous variable. Discretized with 16 levels.
DRI_AGE	Age of the driver	Continuous variable. Discretized with 11 levels.
LIC_YEARS	Number of years of the driving license	Continuous variable. Discretized with 9 levels.
WEIGHT_POW	Power-to-weight ratio	Continuous variable. Discretized with 3 levels.
GENDER	Gender of the driver	Categorical variable with 2 levels (male/female).
ZONA	The different Spanish regions and some big cities	Categorical variable with 65 levels.
DIESEL	Does the vehicle have a diesel engine?	Categorical variable with 2 levels (diesel/gasoline).

For the Generalized Linear Models, the continuous rating factors were discretized in the same way as the insurance firm did. The levels of the rating factors defined a set of cells, and within these cells, the values of the dependent variable (number of claims and claim size) were aggregated. In this manner, all the explanatory variables used in the GLMs were categorical. Additionally, a set of "naïve" GLMs was also estimated. In these models, the rating factors were discretized, but not according to the levels used by the insurance firm (levels that reflect its ample experience in policy pricing and require a long process of fine-tuning to find the most appropriate discretization of the variables), but using a set of categories obtained simply from the quantiles of each variable. Nevertheless, the number of levels chosen for the discretization was the same as that used by the insurer, so the discretization was not fully "naïve", as it included some of the insurer's ratings.

For the GAMs, the continuous explanatory variables were introduced into the models using splines.

For each of the five training samples of 500,000 policies, three models (a GLM, a naive GLM, and a GAM) were estimated. These models were used to predict the values of the dependent variables for the test sets, calculating the sum of their respective absolute prediction errors as a measure of predictive power. The absolute prediction errors obtained for the test samples of 200,000 observations are shown in Table 2.

Table 2. Absolute prediction errors.

Poisson									
Sample	**Naïve GLM**	**GLM**	**GAM**					**BAM**	
			$k = 10$	$k = 15$	$k = 20$	$k = 30$	$k = 50$	$k = 30$	
1	15,785.96	15,774.31	15,772.9	15,770.87	15,771.02	15,770.24	15,770.84	15,772.34	
2	16,065.91	16,050.05	16,050.54	16,048.58	16,048.20	16,047.72	16,046.31	16,048.67	
3	15,792.93	15,779.81	15,780.58	15,778.65	15,778.43	15,777.71	15,777.37	15,778.17	
4	15,765.74	15,750.54	15,750.68	15,748.98	15,749.11	15,748.74	15,748.12	15,749.06	
5	15,883.12	15,872.55	15,868.71	15,867.85	15,867.50	15,867.49	15,868.31	15,868.63	
Gamma									
Sample	**Naïve GLM**	**GLM**	**GAM**						**BAM**
			$k = 4$	$k = 10$	$k = 15$	$k = 20$	$k = 30$	$k = 50$	$k = 4$
1	3,085,250	3,069,263	3,056,845	3,076,293	3,077,411	3,091,096	3,106,039	3,118,692 *	3,056,219
2	3,053,085	3,045,601	3,037,337	3,043,671	3,052,295	3,062,141	3,072,514	3,083,956 *	3,038,045
3	3,175,236	3,167,669	3,155,742	3,166,577	3,177,928	3,182,010	3,189,077	3,206,048 *	3,153,131
4	2,887,690	2,884,199	2,880,680	2,882,199	2,885,353	2,890,385	2,897,359	2,911,412 *	2,880,103
5	3,225,151	3,216,130	3,203,817	3,209,767	3,215,150	3,219,094	3,234,659	3,239,390 *	3,203,398

(*) A spline with $k = 30$ has been used for VEH_AGE.

The GLMs were estimated using the standard R function, glm. For the GAMs, the gam function of the R mgcv package [67] was used with thin-plate regression splines and different values of the parameter k. This value indicated the maximum allowable dimension of the spline space and hence its degrees of freedom. Also, Table 2 shows the sums of the absolute prediction errors for these different values of k. For the Poisson models, in general, the smallest prediction error was reached for $k = 50$. Note that the huge differences between the values for the Poisson and Gamma models are due to the nature of the dependent variables (number of claims in the Poisson case, claim size in the Gamma case).

The mgcv package includes a "big data" function (bam) used to estimate the GAMs. This is very similar to the gam function but is designed to work with very big datasets and use multiple CPU cores in parallel, resulting in very efficient memory usage and shorter execution times. With the bam function, it is not advisable to use thin-plate splines because of their high computational complexity [68], so other types of splines, like the cubic regression ones that we used here, are preferable. Although these splines may lead to slightly worse results, they are much quicker to evaluate. We can see in Table 2 that, in fact, for $k = 4$, the prediction errors of the GAMs evaluated with the bam function are very close to those of the gam function, but the execution times are, as discussed below, a lot shorter and comparable to those of the GLM models.

In Table 2, the values corresponding to the Gamma regressions with $k = 50$ are marked with an asterisk because, in this case, the continuous variable VEH_AGE does not have enough different values to use a spline with $k = 50$ (the covariate has fewer unique values than the specified maximum degrees of freedom). Hence, we used a spline with $k = 30$ for this variable, and splines with $k = 50$ for the rest of the continuous variables (DRI_AGE, LIC_YEARS, and WEIGHT_POW).

To check if the improvement in the absolute prediction error obtained when using a GAM, instead of a GLM or naïve GLM, was statistically significant, we followed [69], which recommended conducting a Kolmogorov–Smirnov test. Although this test was used in [69] in a time-series prediction context, nothing prevents its usage in the more general case of comparing the prediction errors of two different models. Table 3 summarizes the p-values

of the Kolmogorov–Smirnov tests for the null hypothesis that the absolute prediction errors obtained with the GAM are not stochastically lesser than the errors obtained with the corresponding GLM or naïve GLM. Therefore, rejecting the null hypothesis means that the cumulative distribution function (cdf) of the absolute error (AE) of the prediction of the GAM lies above and to the left of the corresponding cdf of the AE of the GLM or naïve GLM. This would mean that the bulk of the distribution of the GAM prediction errors concentrates on values significantly smaller than those of the GLM or naïve GLM prediction errors.

Table 3 shows the *p*-values obtained by comparing the prediction errors of the GAMs to those of the corresponding GLMs and checking the hypothesis that the errors of the GAMs were not (stochastically) less than those of the GLMs. The lower the *p*-values, the greater the confidence that the GAMs outperformed the GLMs in terms of the absolute prediction error. In general, the absolute prediction errors of the GAMs were significantly better than those of the GLMs (at a 95% or 90% confidence level), except in the case of the Poisson regression models, where the absolute errors of the GAMs were not significantly better than those of the GLMs (although they were significantly better than those of naïve GLMs).

Table 3. *p*-values of the Kolmogorov–Smirnov tests.

	Poisson		**Gamma**	
Sample	**GAM vs. Naïve GLM**	**GAM vs. GLM**	**GAM vs. Naïve GLM**	**GAM vs. GLM**
1	0.0256	0.0966	0.0092	0.0033
2	0.0561	0.1624	0.0351	0.0651
3	0.0246	0.2012	0.0971	0.0499
4	0.0804	0.3357	0.1945	0.0202
5	0.0335	0.1716	0.0131	0.09861

A problem that was found in the Poisson GAMs was that of overdispersion, but this can be treated in a similar manner to that in the case of the more conventional GLMs. Among these solutions, the mgcv package enables estimating, for example, a negative binomial, a quasi-likelihood model, etc.

A relatively simple way of estimating the dispersion parameter and checking if it is close to 1, as it should be, is by using the quotient between the sum of the squared Pearson residuals and the degrees of freedom of the model [70]. In our case, for all the samples and models estimated, the values of this quotient ranged between 1.17 and 1.21, and therefore the overdispersion did not cause a serious problem in any case.

As for the Gamma regressions, it appears that there is a certain overfitting, both for the default value of k ($k = 10$) and the rest of the values used ($k = 15$, 20, 30, and 50). This can be seen in Figure 1, that depicts the values taken by the splines of the continuous rating factors for the Gamma regressions with a value of $k = 30$ (the red lines, with a 95% confidence band in grey around them). In this figure, there is a series of oscillations in the splines that do not reflect any foreseeable effect of the explanatory variables, but are a mere consequence of a too-high value for the degrees of freedom allowed in the splines.

As the absolute prediction errors shown in Table 2 grew with the value of k, different values of this parameter, smaller than the default value $k = 10$, were tested. The value of k with the lesser value of the absolute prediction error was $k = 4$. Figure 2 shows the splines for the continuous independent variables of the Gamma regression models with $k = 4$. One can see that the apparent oscillations visible in the models with higher values of k, like the ones displayed in Figure 1, have disappeared.

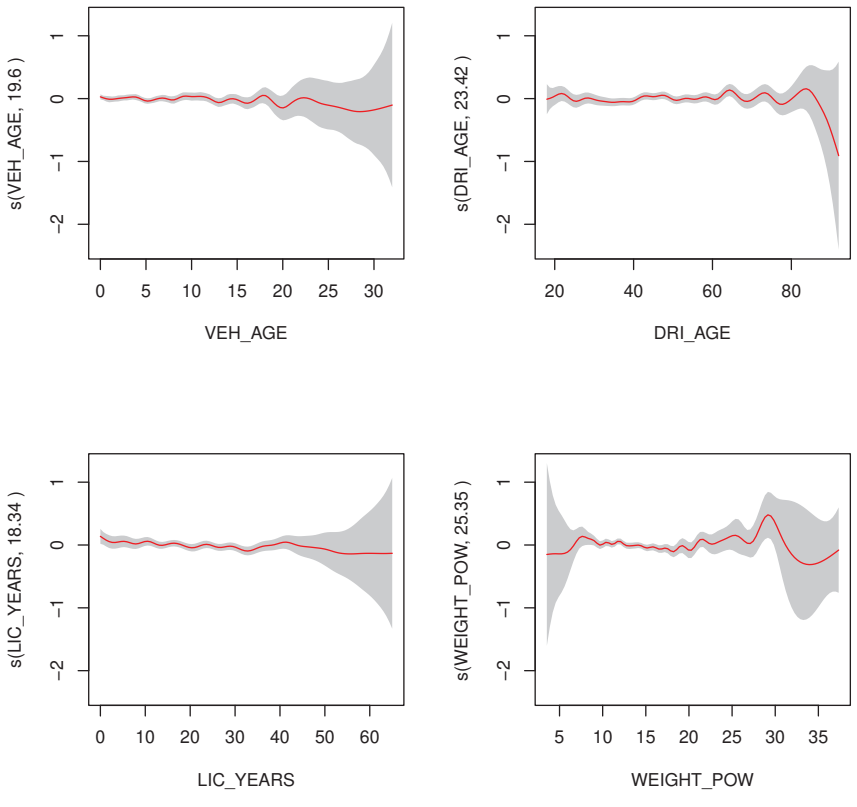

Figure 1. Overfitting in Gamma regression with $k = 30$.

In general, the results were very similar for the GAMs and the more classical GLMs with the "sophisticated" discretization of the continuous variables implemented by the insurance firm. Both kinds of models yielded results that outperformed those obtained with the GLMs with "naive" discretization, so they can be considered alternative solutions to the problem of coping with nonlinear effects in the automobile insurance field. The GLMs with discretized rating factors require fine-tuning, which can be painstaking and tedious, to determine how to discretize the continuous explanatory variables, i.e., to determine the number of levels and the intervals that define them. As for the GAMs, they are more complex models, but the fine-tuning process is more "automated" and quick, and in practice, it reduces to check that the complexity of the model, as measured by the degrees of freedom of the splines used, is high enough to properly reflect the nonlinear effect of every independent variable.

A problem with the GAMs was that their higher complexity entailed greater requirements in terms of RAM and significantly longer CPU execution times. Table 4 shows the execution times (in seconds) for a set of Poisson and Gamma regressions, with different sample sizes, both for GLMs and GAMs.

The aforementioned bam function of the mgcv package not only makes more efficient use of the RAM and is faster than the gam function but is also designed to be executed in parallel using multiple threads or multiple cores through the parallel package [71]. In our case, it was executed using three cores in a single computer. It can be seen in Table 3 that the execution times of the bam function are comparable to those of the GLMs and one order of magnitude lower than those of the gam function. Moreover, their predictive performance, as measured with the absolute prediction error, is only slightly worse than that of the GAMs with thin-plate splines. This clearly shows that the bam function provides a neat and quick

alternative for estimating GAMs since it is as fast as the more conventional glm function for GLMs with discretized variables. It should be noted that the sample sizes in Table 4 are far smaller for the Gamma than for the Poisson models because the former were estimated using the observations with at least one claim, and thus the vast majority of policies, that is, the ones with zero claims, were discarded. In practice, only one out of twenty-five policies has one or more claims per year. We chose sample sizes according to this proportion.

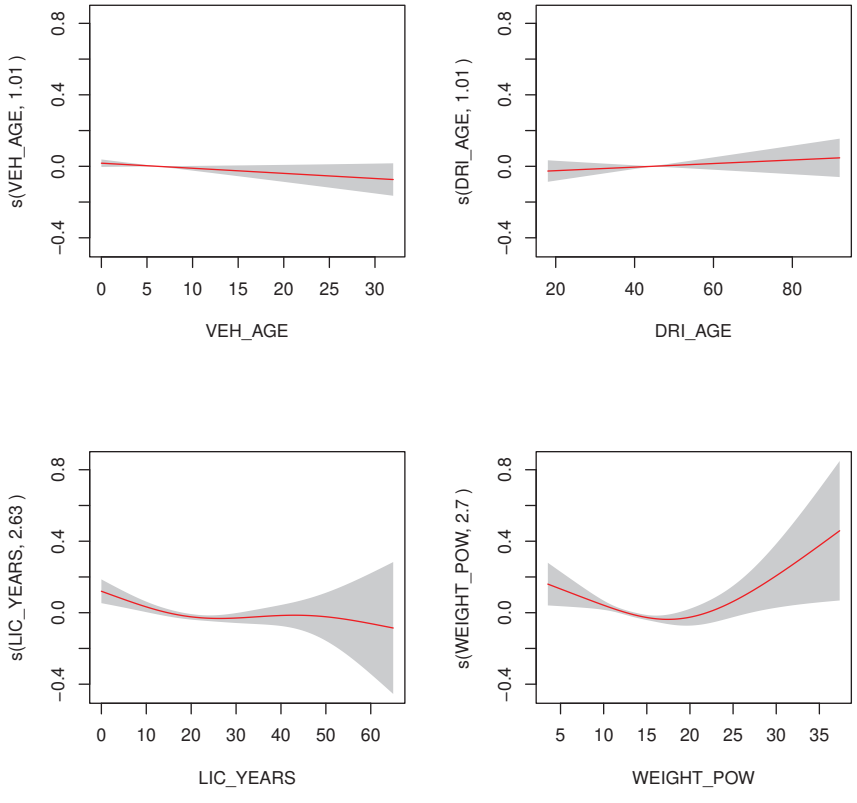

Figure 2. Gamma regression with $k = 4$.

Table 4. Execution times (in seconds).

Poisson				Gamma			
Sample Size	**GLM**	**GAM**	**BAM**	**Sample Size**	**GLM**	**GAM**	**BAM**
125,000	136.19	1591.18	201.77	5000	3.59	183.75	36.12
250,000	259.34	2326.81	348.30	10,000	6.39	289.92	36.08
375,000	370.55	3505.14	457.88	15,000	9.61	440.50	37.47
500,000	506.30	4793.21	616.93	20,000	12.48	763.18	40.33
625,000	640.68	6128.69	780.73	25,000	13.10	1014.32	54.19
750,000	757.25	8707.59	865.03	30,000	17.87	1570.21	56.51

Figure 3 shows a graphical image of the runtimes. Although these times increased exponentially with the sample size for the GAMs, with the bam function, they were almost identical to the times for estimating a GLM using the glm function.

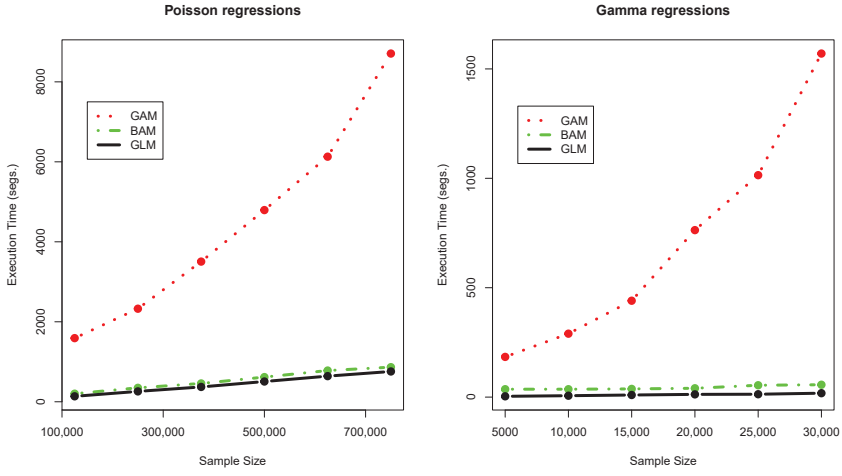

Figure 3. Execution times.

5. Conclusions

GLMs are the basic pricing tool used in the field of automobile insurance. Usually, they are used by taking the continuous rating factors and discretizing them. This implies the definition of a series of "cells", so that each policy belongs to one of them, and its premium is the one corresponding to that cell (all the policies in the same cell have the same premium). This facilitates the computation of the premium of a new policy because it reduces, in practice, to determine the cell the new policy belongs to. It also enables taking into account in a very simple manner the eventual nonlinear effects of the rating factors. As a drawback, the discretization of the rating factors implies a certain amount of information loss and also entails a certain amount of arbitrariness in the choice of the number of levels and intervals that define the discretized variables. A careful analysis and choice of these levels are, therefore, absolutely necessary to obtain an acceptable tariff model.

As a more sophisticated alternative to discretization, it is possible to replace the GLMs with their generalization, the GAMs. The latter are more computationally intensive and with them, it is no longer possible to compute the premium by merely looking at the cell the policy belongs to (since there are no cells at all), but they enable taking into account the nonlinear effects of the rating factors in a very straightforward fashion and without discretizing them. Consequently, there is no information loss, as happens in the case of GLMs with discretized explanatory variables, and it is possible to model the relationship between the rating factors and the key ratios (claim frequency, claim severity, etc.) the insurer is interested in with greater detail. The downsides of the GAMs are the greater complexity in the models, the greater computational load, and, therefore, the longer CPU time required for their estimation. Furthermore, it is not possible to determine the premium for a new policy simply by using a table that shows the set of premium values according to the levels of the rating factors. In contrast, in the case of GAMs, the calculation of the premium for a new policy requires determining the value predicted by the model for the specific values of the rating factors of that policy.

5.1. Discussion and Future Lines of Work

In this work, we compared both approaches—traditional GLMs with discretized variables and more sophisticated GAMs without discretization—with the aim of determining whether the use of GAMs, which are more complex and computationally intensive, is advantageous enough in comparison to GLMs to justify their utilization. To do this, we used a huge sample of car policies from a Spanish insurer and estimated and compared

a series of GLMs and GAMs for the two key ratios, claim frequency and claim severity, commonly used for tariff analysis in the automobile insurance field. As is conventional, Poisson regressions were used to model the claim frequency, and Gamma regressions were used to model the claim severity.

Regarding the quality of the predictions of both types of models, the most relevant issue is that the prediction errors for the GAMs were, in general, slightly better (i.e., minor) than those for the GLMs (see Table 2). Also, in the case of the Poisson regressions, the reduction in the prediction error when passing from GLMs to GAMs was notably smaller (approximately 25% on average) than the reduction when passing from naive GLMs to GLMs (we used this last reduction as a reference). In the case of the Gamma regressions, the reduction in the prediction error when passing from GLMs to GAMs was slightly greater (16.6% on average) than the reduction when passing from naive GLMs to GLMs, although we needed to manually adjust the degrees of freedom of the models (through the value of k) to avoid overfitting.

Table 5 shows the improvements in the prediction error when moving from GLMs to GAMs, relative to the improvement in the error when moving from naïve GLMs to GLMs (i.e., the values recorded in the table are of the form $\frac{error\ GAM - error\ GLM}{error\ GLM - error\ naïve}$). We can see in this table that the improvement in the predictions' quality (measured by the absolute prediction error) is not very appealing when considered in absolute terms, but compared to the improvement experienced when moving from naive GLMs to GLMs, one can see that the error reduction is very noticeable, especially in the case of the Gamma regressions.

Table 5. Improvement in the prediction error.

	Sample					**Mean**	**Standard Deviation**
	1	**2**	**3**	**4**	**5**		
Poisson	0.3494	0.1469	0.1601	0.1184	0.4787	0.2507	0.1567
Gamma	0.7768	1.1042	1.5762	1.0080	1.3649	1.1660	0.3115

With regard to execution times, the greater complexity of the GAMs implies longer execution times compared to the GLMs. According to Table 3, in the case of the Poisson regressions, these times were, on average, approximately ten times higher for the GAMs than for the GLMs and about 61 times higher in the case of the Gamma regressions. These differences in the execution times are large enough to pose a serious problem when working with big samples and tilt the balance in favor of the less demanding GLMs. Nevertheless, these differences almost vanish if the bam function of the same R package (mgcv) is used instead of the gam (especially for the Poisson regressions). The bam function was designed to deal with big datasets, and its results were very similar to those obtained with the gam function but with a dramatic reduction in execution times. Also, this function can be executed in parallel in a cluster with multiple CPUs, speeding up the calculations and providing additional reductions in execution times. Table 3 shows that the execution times of the bam function were very similar and of the same order of magnitude as those of the gam function in the case of the Poisson regressions, and only about five times greater (instead of 61 times if the bam function was not used) for the Gamma regressions. Thus, GAMs estimated using the bam function are a competitive alternative to GLMs.

The rise of modern machine learning and deep learning methods offers a novel and powerful alternative to traditional, regression-based methods for rate analysis in insurance. Many of these new tools, such as those based on trees (random forest, for example) and those based on neural networks (deep learning), lack the interpretability of classical methods and are also very computationally intensive, but their success in areas such as image analysis and natural language processing clearly shows that they will be increasingly important in the future in the actuarial field. A natural way of extending this work is to compare GAMs and their ability to deal with nonlinearities to methods such as deep learning. Some work has been carried out in this direction, for example in [72,73].

5.2. Managerial Implications

As for the managerial implications, we can conclude that GAMs should be considered a powerful and realistic choice for tariff analysis in the automobile insurance field. They improve the quality of the models by reducing the prediction errors but at the expense of a higher complexity of the models and longer execution times. When using GAMs, it is no longer necessary to discretize the continuous risk factors, which saves a significant amount of work in determining the levels of the discretized variables and fine-tuning the models. Nevertheless, in exchange for this capability of working directly with the continuous variables, we lose the possibility of assigning a premium to a new policy by simply searching in a table that collects the premiums corresponding to each level of the rating factors.

Moreover, the use of GAMs is not completely automatic (in the sense that one could trust the results obtained using the default values of the parameters), but sometimes it is necessary for some manual fine-tuning of the models, as seen in the case of the Gamma regressions, where there was certain overfitting that forced us to manually choose the models' degrees of freedom.

Broadly speaking, the nonlinearities observed in the rating factors' effects on the key ratios (claim frequency and claim severity) used in car insurance do not seem very pronounced, as can be seen, for example, in Figure 2. This means that GAMs will not be as useful for modeling these nonlinearities as they are in certain fields of natural sciences where they are commonly used, as previously mentioned, and where the nonlinear effects of the explanatory variables are noticeably more pronounced. Nonetheless, nonlinearities exist, and GAMs prove to be a very useful tool and more sophisticated than traditional GLMs with discretized variables, which, nowadays, is standard practice in automobile insurance pricing.

Author Contributions: All authors contributed equally to this work. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the THEIA (Techniques for Integrity and Authentication of Multimedia Files of Mobile Devices) UCM project (FEI-EU-19-04). This work was also funded by the THEIA I (Techniques for Integrity, Authentication, and Scene Recognition in Multimedia Files of Mobile Devices—Part I) UCM project (FEI-EU-21-01). This work was also funded by the Ministerio de Ciencia e Innovación under project number PID2021-125133NB-I00.

Data Availability Statement: Data has been obtained under a confidentiality agreement and cannot be publicly disclosed.

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

References

1. Nelder, J.A.; Wedderburn, W.M. Generalized Linear Models. *J. R. Stat. Soc. Ser. A* **1972**, *135*, 370–384. [CrossRef]
2. Brockman, M.J.; Wright, T.S. Statistical motor rating: Making effective use of your data. *J. Inst. Actuar.* **1992**, *119*, 457–543. [CrossRef]
3. Haberman, S.; Renshaw, A.E. Generalized Linear Models and Actuarial Science. *J. R. Stat. Soc. Ser. D* **1996**, *45*, 407–436. [CrossRef]
4. Denuit, M.; Hainaut, D.; Trufin, J. *Effective Statistical Learning Methods for Actuaries I. GLMs and Extensions*; Springer: Cham, Switzerland, 2019.
5. Goldburd, M.; Khare, A.; Tevet, D.; Guller, D. *Generalized Linear Models for Insurance Rating*, 2nd ed.; Casualty Actuarial Society: Arlington, VA, USA, 2020.
6. Wüthrich, M.V.; Merz, M. *Statistical Foundations of Actuarial Learning and Its Applications*; Springer International Publishing: Berlin/Heidelberg, Germany, 2023. [CrossRef]
7. Agresti, A. *Foundations of Linear and Generalized Linear Models*; John Wiley & Sons: Hoboken, NJ, USA, 2015.
8. Dobson, A.J. *An Introduction to Generalized Linear Models*, 2nd ed.; Chapman & Hall/CRC: Boca Raton, FL, USA, 2002.
9. McCullagh, P.; Nelder, J.A. *Generalized Linear Models*, 2nd ed.; Chapman & Hall/CRC: Boca Raton, FL, USA, 1989.
10. Ohlsson, E.; Johansson, B. *Non-Life Insurance Pricing with Generalized Linear Models*; Springer: Heidelberg, Germany, 2010.
11. Fox, J. *Applied Regression Analysis and Generalized Linear Models*, 3rd ed.; Sage Publications: Thousand Oaks, CA, USA, 2008.
12. Gentle, J.E. *Computational Statistics*; Springer: New York, NY, USA, 2009.
13. Green, P.J.; Silverman, B.W. *Nonparametric Regression and Generalized Linear Models*; Chapman & Hall: London, UK, 1994.

14. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning*, 2nd ed.; Springer: New York, NY, USA, 2009.
15. Demmel, J.W. *Applied Numerical Linear Algebra*; SIAM: Philadelphia, PA, USA, 1997.
16. Gentle, J.E. *Matrix Algebra. Theory, Computations, and Applications in Statistics*; Springer: New York, NY, USA, 2007.
17. Seber, G.A.F.; Lee, A.J. *Linear Regression Analysis*, 2nd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2003.
18. Seber, G.A.F. *A Matrix Handbook for Statisticians*; John Wiley & Sons: Hoboken, NJ, USA, 2008.
19. Hastie, T.; Tibshirani, R. *Generalized Additive Models*; Chapman & Hall/CRC: London, UK, 1990.
20. de Jong, P.; Heller, G.Z. *Generalized Linear Models for Insurance Data*; Cambridge University Press: New York, NY, USA, 2008.
21. Mahmoudvand, R.; Hassani, H. Generalized Bonus-Malus Systems with a Frequency and a Severity Component on an Individual Basis in Automobile Insurance. *ASTIN Bull. J. IAA* **2009**, *39*, 307–315. [CrossRef]
22. Si, J.; He, H.; Zhang, J.; Cao, X. Automobile insurance claim occurrence prediction model based on ensemble learning. *Appl. Stoch. Model. Bus. Ind.* **2022**, *38*, 1099–1112. [CrossRef]
23. Mulquiney, P. *Application of Soft-Computing Techniques in Accident Compensation*; Institute of Actuaries of Australia's (IAAust) Accident Compensation Seminar, 2004; Institute of Actuaries of Australia: Sydney, Australia, 2004.
24. Friedman, J.H. Multivariate adaptive regression splines. *Ann. Stat.* **1991**, *19*, 1–67. [CrossRef]
25. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29* , 1189–1232. [CrossRef]
26. Pesantez-Narvaez, J.; Guillen, M.; Alcañiz, M. Predicting Motor Insurance Claims Using Telematics Data—XGBoost versus Logistic Regression. *Risks* **2019**, *7*, 70. [CrossRef]
27. Henckaerts, R.; Côté, M.P.; Antonio, K.; Verbelen, R. Boosting insights in insurance tariff plans with tree-based machine learning methods. *N. Am. Actuar. J.* **2021**, *25*, 255–285. [CrossRef]
28. Ticconi, D. *Individual Claims Reserving in Credit Insurance Using GLM and Machine Learning*; Dipartimento di Scienze Statistiche, Sapienza Università di Roma: Rome, Italy, 2018.
29. Baillargeon, J.T.; Lamontagne, L.; Marceau, E. Mining actuarial risk predictors in accident descriptions using recurrent neural networks. *Risks* **2020**, *9*, 7. [CrossRef]
30. Delong, Ł.; Wüthrich, M.V. Neural Networks for the Joint Development of Individual Payments and Claim Incurred. *Risks* **2020**, *8*, 33. [CrossRef]
31. Blier-Wong, C.; Cossette, H.; Lamontagne, L.; Marceau, E. Machine learning in P&C insurance: A review for pricing and reserving. *Risks* **2020**, *9*, 4. [CrossRef]
32. Bailey, D.; Collins, M.; Gordon, J.; Zuur, A.; Priede, I. Long-term changes in deep-water fish populations in the northeast Atlantic: A deeper reaching effect of fisheries? *Proc. R. Soc. Lond. B Biol. Sci.* **2009**, *276*, 1965–1969. [CrossRef] [PubMed]
33. Drexler, M.; Ainsworth, C.H. Generalized Additive Models Used to Predict Species Abundance in the Gulf of Mexico: An Ecosystem Modeling Tool. *PLoS ONE* **2013**, *8*, e64458 . [CrossRef] [PubMed]
34. Grüss, A.; Drexler, M.; Ainsworth, C.H. Using delta generalized additive models to produce distribution maps for spatially explicit ecosystem models. *Fish. Res.* **2014**, *159*, 11–24. [CrossRef]
35. Heger, A.; Ieno, E.; King, N.; Morris, K.; Bagley, P.; Priede, I. Deep-sea pelagic bioluminescence over the Mid-Atlantic Ridge. *Deep Sea Res. Part II Top. Stud. Oceanogr.* **2008**, *55*, 126–136. [CrossRef]
36. Mondal, S.; Vayghan, A.H.; Lee, M.A.; Wang, Y.C.; Semedi, B. Habitat Suitability Modeling for the Feeding Ground of Immature Albacore in the Southern Indian Ocean Using Satellite-Derived Sea Surface Temperature and Chlorophyll Data. *Remote Sens.* **2021**, *13*, 2669. [CrossRef]
37. Murase, H.; Nagashima, H.; Yonezaki, S.; Matsukura, R.; Kitakado, T. Application of a generalized additive model (GAM) to reveal relationships between environmental factors and distributions of pelagic fish and krill: A case study in Sendai Bay, Japan. *ICES J. Mar. Sci. J. Du Cons.* **2009**, *66*, 1417–1424. [CrossRef]
38. Potts, S.E.; Rose, K.A. Evaluation of GLM and GAM for estimating population indices from fishery independent surveys. *Fish. Res.* **2018**, *208*, 167–178. [CrossRef]
39. Sagarese, S.R.; Frisk, M.G.; Cerrato, R.M.; Sosebee, K.A.; Musick, J.A.; Rago, P.J. Application of generalized additive models to examine ontogenetic and seasonal distributions of spiny dogfish (Squalus acanthias) in the Northeast (US) shelf large marine ecosystem. *Can. J. Fish. Aquat. Sci.* **2014**, *71*, 847–877. [CrossRef]
40. Valavi, R.; Guillera-Arroita, G.; Lahoz-Monfort, J.J.; Elith, J. Predictive performance of presence-only species distribution models: A benchmark study with reproducible code. *Ecol. Monogr.* **2022**, *92* . [CrossRef]
41. Hua, C.; Zhu, Q.; Shi, Y.; Liu, Y. Comparative analysis of CPUE standardization of Chinese Pacific saury (Cololabis saira) fishery based on GLM and GAM. *Acta Oceanol. Sin.* **2019**, *38*, 100–110. [CrossRef]
42. Thuiller, W.; Araújo, M.B.; Lavorel, S. Generalized models vs. classification tree analysis: Predicting spatial distributions of plant species at different scales. *J. Veg. Sci.* **2003**, *14*, 669–680. [CrossRef]
43. Momberg, M.; Ryan, P.G.; Hedding, D.W.; Schoombie, J.; Goddard, K.A.; Craig, K.J.; Le Roux, P.C. Factors determining nest-site selection of surface-nesting seabirds: A case study on the world's largest pelagic bird, the Wandering Albatross (Diomedea exulans). *IBIS* **2023**, *165*, 190–203. [CrossRef]
44. Yu, H.; Jiao, Y.; Carstensen, L.W. Performance comparison between spatial interpolation and GLM/GAM in estimating relative abundance indices through a simulation study. *Fish. Res.* **2013**, *147*, 186–195. [CrossRef]
45. Rocca, F.D.; Milanesi, P. The Spread of the Japanese Beetle in a European Human-Dominated Landscape: High Anthropization Favors Colonization of Popillia japonica. *Diversity* **2022**, *14*, 658. [CrossRef]

46. Gujral, H.; Sinha, A. Association between exposure to airborne pollutants and COVID-19 in Los Angeles, United States with ensemble-based dynamic emission model. *Environ. Res.* **2021**, *194*, 110704. [CrossRef]
47. Lee, W.; Lim, Y.H.; Ha, E.; Kim, Y.; Lee, W.K. Forecasting of non-accidental, cardiovascular, and respiratory mortality with environmental exposures adopting machine learning approaches. *Environ. Sci. Pollut. Res.* **2022**, *29*, 88318–88329. [CrossRef]
48. Li, L.; Blomberg, A.J.; Stern, R.A.; Kang, C.M.; Papatheodorou, S.; Wei, Y.; Liu, M.; Peralta, A.A.; Vieira, C.L.; Koutrakis, P. Predicting Monthly Community-Level Domestic Radon Concentrations in the Greater Boston Area with an Ensemble Learning Model. *Environ. Sci. Technol.* **2021**, *55*, 7157–7166. [CrossRef]
49. Tan, Y.; Zeng, Z.; Liang, H.; Weng, X.; Yao, H.; Fu, Y.; Li, Y.; Chen, J.; Wei, X.; Jing, C. Association between Perfluoroalkyl and Polyfluoroalkyl Substances and Women's Infertility, NHANES 2013–2016. *Int. J. Environ. Res. Public Health* **2022**, *19*, 15348. [CrossRef]
50. Pourghasemi, H.R.; Rossi, M. Landslide susceptibility modeling in a landslide prone area in Mazandarn Province, north of Iran: A comparison between GLM, GAM, MARS, and M-AHP methods. *Theor. Appl. Climatol.* **2016**, *130*, 609–633. [CrossRef]
51. Osah, S.; Acheampong, A.A.; Fosu, C.; Dadzie, I. Regression models for predicting daily IGS zenith tropospheric delays in West Africa: Implication for GNSS meteorology and positioning applications. *Meteorol. Appl.* **2021**, *28*, e2030. [CrossRef]
52. Egger, S.T.; Bobes, J.; Seifritz, E.; Vetter, S.; Schuepbach, D. Functional transcranial Doppler: Selection of methods for statistical analysis and representation of changes in flow velocity. *Health Sci. Rep.* **2021**, *4*, e400. [CrossRef] [PubMed]
53. Thompson, P.A.; Watkins, K.E.; Woodhead, Z.V.J.; Bishop, D.V.M. Generalized models for quantifying laterality using functional transcranial Doppler ultrasound. *Hum. Brain Mapp.* **2023**, *44*, 35–48. [CrossRef]
54. Cui, E.H.; Song, D.; Wong, W.K.; Li, J.J. Single-cell generalized trend model (scGTM): A flexible and interpretable model of gene expression trend along cell pseudotime. *Bioinformatics* **2022**, *38*, 3927–3934. [CrossRef]
55. Antonio, K.; Beirlant, J. Issues in claims reserving and credibility: A semiparametric approach with mixed models. *J. Risk Insur.* **2008**, *75*, 643–676. [CrossRef]
56. Breuer, A.; Staudt, Y. Equalization Reserves for Reinsurance and Non-Life Undertakings in Switzerland. *Risks* **2022**, *10*, 55. [CrossRef]
57. Denuit, M.; Lang, S. Non-life rate-making with Bayesian GAMs. *Insur. Math. Econ.* **2004**, *35*, 627–647. [CrossRef]
58. England, P.D.; Verrall, R.J. Stochastic claims reserving in general insurance. *Br. Actuar. J.* **2002**, *8*, 443–518. [CrossRef]
59. Staudt, Y.; Wagner, J. Assessing the Performance of Random Forests for Modeling Claim Severity in Collision Car Insurance. *Risks* **2021**, *9*, 53. [CrossRef]
60. Verschuren, R.M. Predictive claim scores for dinamic multi-product risk clasiffication in insurance. *ASTIN Bull. J. IAA* **2021**, *51*, 1–25. [CrossRef]
61. Wang, Y. *Smoothing Splines. Methods and Applications*; Chapman & Hall/CRC: Boca Raton, FL, USA, 2011.
62. Faraway, J.J. *Extending the Linear Model with R*; Chapman & Hall/CRC: Boca Raton, FL, USA, 2006.
63. Bowman, A.W.; Azzalini, A. *Applied Smoothing Techniques for Data Analysis*; Oxford University Press: New York, NY, USA, 1997.
64. Loader, C. *Local Regression and Likelihood*; Springer: New York, NY, USA, 1999.
65. Wahba, G. *Spline Models for Observational Data*; SIAM: Philadelphia, PA, USA, 1990.
66. Wood, S.N. Thin Plate Regression Splines. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **2003**, *65*, 95–114. [CrossRef]
67. Wood, S. mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation; R Package Version 1.9.0. 2023. Available online: https://cran.r-project.org/web/packages/mgcv/index.html (accessed on 13 July 2023).
68. Wood, S.N. *Generalized Additive Models: An Introduction with R*, 2nd ed.; CRC Press: Boca Raton, FL, USA, 2017.
69. Hassani, H.; Silva, E.S. A Kolmogorov-Smirnov Based Test for Comparing the Predictive Accuracy of Two Sets of Forecasts. *Econometrics* **2015**, *3*, 590–609. [CrossRef]
70. Venables, W.N.; Ripley, B.D. *Modern Applied Statistics with S*, 4th ed.; Springer: New York, NY, USA, 2003.
71. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2023.
72. Denuit, M.; Hainaut, D.; Trufin, J. *Effective Statistical Learning Methods for Actuaries I. Neural Networks and Extensions*; Springer: Cham, Switzerland, 2019.
73. Denuit, M.; Hainaut, D.; Trufin, J. *Effective Statistical Learning Methods for Actuaries I. Tree-Based Methods and Extensions*; Springer: Cham, Switzerland, 2020.

mathematics

MDPI

Article

From Transience to Recurrence for Cox–Ingersoll–Ross Model When $b < 0$

Mingli Zhang [1] and Gaofeng Zong [2,*]

[1] School of Science, Shandong Jianzhu University, Jinan 250101, China; mingli@sdjzu.edu.cn
[2] School of Statistics and Mathematics, Shandong University of Finance and Economics, Jinan 250014, China
* Correspondence: zonggf@sdufe.edu.cn

Abstract: We consider the Cox–Ingersoll–Ross (CIR) model in time-dependent domains, that is, the CIR process in time-dependent domains reflected at the time-dependent boundary. This is a very meaningful question, as the CIR model is commonly used to describe interest rate models, and interest rates are often artificially set within a time-dependent domain by policy makers. We consider the most fundamental question of recurrence versus transience for normally reflected CIR process with time-dependent domains, and we examine some precise conditions for recurrence versus transience in terms of the growth rates of the boundary. The drift terms and the diffusion terms of the CIR processes in time-dependent domains are carefully provided. In the transience case, we also investigate the last passage time, while in the case of recurrence, we also consider the positive recurrence of the CIR processes in time-dependent domains.

Keywords: transience; recurrence; CIR model; time-dependent region; reflection

MSC: 60J60; 60K35; 60J80; 60J10

Citation: Zhang, M.; Zong, G. From Transience to Recurrence for Cox–Ingersoll–Ross Model When $b < 0$. *Mathematics* **2023**, *11*, 4485. https://doi.org/10.3390/math11214485

Academic Editors: Jing Yao, Xiang Hu and Jingchao Li

Received: 12 September 2023
Revised: 26 October 2023
Accepted: 28 October 2023
Published: 30 October 2023

1. Introduction

In mathematical finance, especially in the field of interest rate theory, the Cox–Ingersoll–Ross (CIR for short) model explains the evolution of interest rates. The CIR model is a type of one-factor model (short-rate model), as it describes interest rate movements as driven by only one source of market risk. The model was introduced by [1] as an extension of the Vasicek's interest rate model, and it has the following stochastic differential equation (SDE for short):

$$dX(t) = (a - bX(t))dt + \sigma\sqrt{X(t)}dW(t), \tag{1}$$

where $W(t)$ is a Wiener process (modeling the random market risk factor) and a, b, and σ are positive constants. The parameter a is the mean level or long-term interest rate constant, the parameter b is the speed of the mean reversion and corresponds to the speed of adjustment to the mean a, and σ regulates the volatility. The drift factor, $(a - bX(t))$, is the same as in the Vasicek model; see [2]. It ensures the mean reversion of the interest rate towards the long-run value a, with the speed of adjustment governed by the strictly positive parameter b. The stochastic volatility term $\sigma\sqrt{X(t)}dW(t)$ has a standard deviation that is proportional to the square root of the current rate. This implies that as the interest rate increases, its standard deviation increases, and as it falls and approaches zero, the stochastic volatility term also approaches 0.

In the following section, we mainly study the Equation (1) as the CIR model or CIR process. The same process is used in the Heston model, see [3], to model stochastic volatility. The SDE (1) has no explicit solution in general, even though its mean and variance can be calculated explicitly, and the probability transition density can be easily determined by using the time–space transformation. This CIR process $X(t)$ can be defined as a sum of

squared Ornstein–Uhlenbeck process or be constructed using a BESQ process of dimension $d = \frac{4a}{\sigma^2}$; see [4]. Refs [5,6] proved that the CIR process is an affine process and the semigroup of every stochastic continuous affine process is a Feller semigroup; hence, the CIR process is a regular Feller process on the interval $(0, +\infty)$. The CIR process $X(t)$ is an ergodic process, and it possesses a stationary distribution. Furthermore, the CIR process is positive recurrent and nonexplosive on the interval $(0, +\infty)$.

Some diffusion processes in time-dependent domains have always been the focus of scholars' attention in the field of probability, and some various sample path properties involving diffusion processes in the time domain are constantly being discovered. The time-dependent domain problem that this article focuses on is actually a domain problem with deterministic moving boundaries, also known as noncylindrical domains. This type of time-dependent domain problem originates from both random environment problems and classic PDE problems with various boundary conditions; see [7]. In [8], the authors provided motivation for studying this issue of diffusion processes in time-dependent domains, through the theoretical explanation of a partial differential equation, and they focused on the heat equation in the time-dependent domain with Neumann rather than Dirichlet boundary conditions, that is, Brownian motion reflected on rather than killed at the boundary of a time-dependent domain. In [9], the most fundamental question of recurrence versus transience for normally reflected Brownian motion with time-dependent domains has been carefully studied, and the authors provided some sharp criterions for the recurrence versus transience of normally reflected Brownian motion in terms of the growth rate of the boundary. In [10] the author provided precise conditions for the recurrence versus transience of one-dimensional Brownian motion with a locally bounded drift, which belongs to the time-dependent domain with a normal reflection at the time-dependent boundary, and the precise conditions provided by the author naturally depend on the growth rates of the boundary and the drift terms of the diffusion processes.

Considering that the CIR model in time-dependent domains has important practical significance and value in the financial field, due to the fact that the evolution of interest rates is often limited to a regional scope, it often changes with the policies of interest rate makers or government management departments. In addition, this CIR model in time-dependent domains has theoretical significance in the field of mathematics and also promotes the research of the properties of transience versus recurrence for stochastic processes. Table 1 below gives the related progress in this field of transience versus recurrence for stochastic processes in time-dependent domains through the aid of the expression of the generator corresponding to the one-dimensional diffusion process. For more topics on the aspect of transience versus recurrence for stochastic processes, please refer to [11–14]. It should be pointed out that, in addition to transience versus recurrence for the conservative random walk, scaling limits for the conservative random walk have also been studied in the work of [11]. However, we did not address scaling limits for stochastic processes in this article.

Table 1. $L = \frac{1}{2}\sigma^2(x)\frac{d^2}{dx^2} + b(x)\frac{d}{dx}$.

$\sigma^2(x)$	$b(x)$	Ref.
1	0	[8,9]
1	bx^γ	[10]
$\sigma^2 x$	$a - bx$	This paper

Here, we need to emphasize that in [8–10], they not only deal with one-dimensional situations, but also with multidimensional situations. For more detailed conclusions, please refer to the literature above for interested readers. In this paper, we only deal with the one-dimensional situations for technical reasons, but we deal with situations where $\sigma^2(x)$ is not a constant. At present, in this paper, we only deal with the case where $\sigma^2(x) = \sigma^2 x$ is linear, and of course, we can also consider the nonlinear case (which is not the CIR model). This problem will also be considered in a future work.

When $a \leq 0$, the CIR process hits zero repeatedly but after each hit becomes positive again; this behavior of hitting zero will also occur even if $a < \frac{1}{2}\sigma^2$. Therefore, we do not intend to handle this simple situation; we will only consider $a > 0$. At this point, the CIR process will have an upward positive constant slope a, and the evolution of the CIR process will still have a mean reversion property when $b > 0$. However, when we started considering $b < 0$, we saw that the CIR process will have a completely positive slope, which will encourage the CIR process to continuously move upwards and hit our constantly changing time-dependent upper boundary. If there are no time-dependent boundary restrictions, this will cause the CIR process to explode, thus possessing the property of transience. How is it possible to conduct the CIR process so as not to explode? In other words, how is it possible to transfer from transience to recurrence for the CIR process when $b < 0$? A natural idea is to add a boundary to the explosion diffusion process, just like the boundary of a time-dependent domain we mentioned above, and when this diffusion process hits the boundary, it will reflect back to our time-dependent domain. This is the main research topic of this paper, which is a fundamental problem in the field of probability, that is, recurrence versus transience, for this normally reflected CIR process with time-dependent domains.

In addition, in the transience case, we also investigate the last passage time, which plays an important and increasing role in financial modelling. The theory of the last passage time is a very important topic in the field of mathematical finance. In this paper, we only provide the probability distribution of the last passage time through the scale function, without exploring its application in the financial modelling field. See [15], as well as [16], for the applications the last passage time to hazard processes and models of default risk.

Let us briefly explain the analytical method we used to prove recurrence versus transience for this normally reflected CIR process with time-dependent domains. The first major tool is the well-known Feynman–Kac formula of diffusion process, which provides the stochastic representation for the solution to the boundary value problem. It is worth noting here that the common Feynman–Kac formula is a boundary value problem with a Dirichlet condition or Cauchy condition; however, we still need the Feynman–Kac formula for the boundary value problem with a Neumann boundary condition here, as we need to handle the normally reflected CIR process with time-dependent domains. The second tool we use is the criticality theory of second-order elliptic operators; in particular, the maximum principle or comparison theorem is frequently used in our proofs. It is worth mentioning that some comparison theorems are not clearly found in the literature, and we provide detailed proofs of them in the Appendix. Regarding the criticality theory, we refer the reader to [17] for more details. Due to the need to obtain precise conditions for coefficients in the CIR process, the selection of certain parameters is also crucial in our proof process.

This paper is structured as follows. In Section 2, we give some basic notations used throughout this paper and provide some auxiliary results about the moment generating function of the first hitting time. In Section 3, we prove the results of two recurrent properties, recurrence and positive recurrence, and provide the precise conditions that the coefficients of the CIR process should meet for recurrence and positive recurrence in terms of the growth rates of the boundary, the drift terms, and the diffusion terms of the CIR processes in time-dependent domains. In Section 4, we prove the result of the transient for the CIR process in time-dependent domains and also provide the precise conditions that the coefficients of the CIR process should meet. Section 5 concludes, and in Appendix A, we provide some comparison theorems of second-order ordinary differential equations with nonconstant coefficients. in Appendix B, the exact solution of a second-order ordinary differential equation with nonconstant coefficients is given by transforming it into one-dimensional Riccati equation.

2. Auxiliary Results

We will first introduce some notations, which we will frequently use in the following sections. Let $X(t)$ denote a canonical, continuous, real-valued path, and let $T_\alpha = \inf\{t \geq 0 : X(t) = \alpha\}$. We introduce some generators for some diffusion processes:

$$\begin{aligned} L_{bx^\gamma} &= \frac{1}{2}\frac{d^2}{dx^2} + bx^\gamma \frac{d}{dx}; \\ L_D &= \frac{1}{2}\frac{d^2}{dx^2} + D\frac{d}{dx}; \\ L_{CIR} &= \frac{1}{2}\sigma^2 x\frac{d^2}{dx^2} + (a - bx)\frac{d}{dx}. \end{aligned}$$

Let $P_x^{*;Ref\leftarrow:\beta}$ and $E_x^{*;Ref\leftarrow:\beta}$ denote probabilities and expectations for diffusion process corresponding to the generator L_* on $[1, \beta]$, starting from $x \in [1, \beta]$, with a reflection at β and stopped at 1. Let $P_x^{*;Ref\rightarrow:\alpha}$ and $E_x^{*;Ref\rightarrow:\alpha}$ denote the probabilities and expectations for diffusion process corresponding to the generator L_* on $[\alpha, \infty)$, starting from $x \in [\alpha, \infty)$, with a reflection at α.

2.1. Moment-Generating Functions

Next, we will provide some auxiliary results about the moment-generating function of the first hitting time using some diffusion process without proofs. Actually, these conclusions can be easily obtained from the well-known Feynman–Kac formula and the criticality theory of second-order elliptic operators after simple calculations.

(A) It follows from the Feynman–Kac formula that

$$u(x) = E_x^{D;Ref\rightarrow:1}[e^{\frac{D^2}{2}T_\beta}]$$

solves the boundary value problem

$$\begin{cases} (L_D + \frac{D^2}{2})u = 0, & \text{in } (1, \beta), \\ u'(1) = 0, \\ u(\beta) = 1. \end{cases}$$

The solution of this linear equation is given by the function

$$u(x) = \frac{1}{1 + D(\beta - 1)}(1 + D(x - 1))e^{D(\beta-1)}.$$

According to the criticality theory of second-order elliptic operators, for instance, see [17], it follows that the principal eigenvalue λ_1 for $-L_D$ satisfies

$$\lambda_1(-L_D) \geq \frac{D^2}{2}.$$

(B) It follows from the Feynman–Kac formula that

$$u(x) = E_x^{D;Ref\leftarrow:\beta}[e^{-\lambda T_\alpha}]$$

solves the boundary value problem, $\lambda > 0$:

$$\begin{cases} (L_D - \lambda)u = 0, & \text{in } (\alpha, \beta), \\ u(\alpha) = 1, \\ u'(\beta) = 0 \end{cases}$$

The solution of this linear equation is given by the function

$$u(x) = \frac{r_1 e^{r_2 x - r_1 \beta} + r_2 e^{-r_1 x + r_2 \beta)}}{r_1 e^{r_2 \alpha - r_1 \beta} + r_2 e^{-r_1 \alpha + r_2 \beta)}},$$

where $r_1 = D + \sqrt{D^2 + 2\lambda}, r_2 = -D + \sqrt{D^2 + 2\lambda}$.

(C) It follows from the Feynman–Kac formula that

$$u(x) = E_x^{D;Ref\to:1}[e^{DT_\beta}]$$

solves the boundary value problem

$$\begin{cases} (L_D + D)u = 0, & \text{in } (1, \beta), \\ u'(1) = 0, \\ u(\beta) = 1. \end{cases}$$

The solution of this linear equation is given by the function

$$u(x) = \frac{r_1 e^{-(r_2 x + r_1)} - r_2 e^{-(r_1 x + r_2)}}{r_1 e^{-(r_2 \beta + r_1)} - r_2 e^{-(r_1 \beta + r_2)}},$$

where $r_1 = D - \sqrt{D^2 - 2D}, r_2 = D + \sqrt{D^2 - 2D}$.

(D) It follows from the Feynman–Kac formula that

$$u(x) = E_x^{D;Ref\to:1}[e^{\lambda T_\beta}]$$

solves the boundary value problem

$$\begin{cases} (L_D + \lambda)u = 0, & \text{in } (1, \beta), \\ u'(1) = 0, \\ u(\beta) = 1. \end{cases}$$

The solution of this linear equation is given by the function

$$u(x) = \frac{r_1 e^{-r_2 x - r_1} - r_2 e^{-r_1 x - r_2}}{r_1 e^{-r_2 \beta - r_1} - r_2 e^{-r_1 \beta - r_2}},$$

where $r_1 = D - \sqrt{D^2 - 2\lambda}, r_2 = D + \sqrt{D^2 - 2\lambda}$.

2.2. Moment-Generating Function for CIR Model

Consider the following CIR model:

$$dX_t = (a - bX_t)dt + \sigma\sqrt{X_t}dW_t,$$

with its operator

$$L_{CIR} = \frac{1}{2}\sigma^2 x \frac{d^2}{dx^2} + (a - bx)\frac{d}{dx}.$$

Lemma 1. *(i) The function*

$$u_\lambda(x) := E_x^{CIR;Ref\leftarrow:\beta} e^{\lambda T_\alpha}$$

satisfies the following boundary value problem

$$\begin{cases} (L_{CIR} + \lambda)u = 0, & \text{in } (\alpha, \beta), \\ u(\alpha) = 1, \\ u'(\beta) = 0. \end{cases}$$

(ii) For $\alpha \in [1, \beta]$ and $\lambda \leq \hat{\lambda}(\alpha, \beta)$,

$$E_x^{CIR;Ref\leftarrow:\beta} e^{\lambda T_\alpha} \leq 2,$$

where $x \in [\alpha, \beta]$, and

$$\hat{\lambda}(\alpha, \beta) = -\frac{b}{\sigma^2} e^{-\frac{2(1+a-b\alpha)}{\alpha\sigma^2}(\beta-\alpha)} = -\frac{b}{\sigma^2} e^{(\frac{2b}{\sigma^2} - \frac{2(1+a)}{\alpha\sigma^2})(\beta-\alpha)}.$$

Proof. It is easy to obtain (i) from the well-known Feynman–Kac formula, and we will only prove that (ii) holds. Consider the function

$$u(x) = 2 - e^{-r(x-\alpha)}, \ \ \alpha \leq x \leq \beta,$$

where $r > 0$. Choose $-\frac{1}{2}\sigma^2 r < b < 0$. Then,

$$\begin{aligned} e^{r(x-\alpha)}(L_{CIR} + \lambda)u &= -r(b + \frac{1}{2}\sigma^2 r)x + ar - \lambda + 2\lambda e^{r(x-\alpha)} \\ &\leq -r(b + \frac{1}{2}\sigma^2 r)\alpha + ar - \lambda + 2\lambda e^{r(\beta-\alpha)} \\ &= -\frac{1}{2}\alpha\sigma^2 r^2 + (a - b\alpha)r + \lambda(2e^{r(\beta-\alpha)} - 1). \end{aligned}$$

Next, we solve the following inequality,

$$\begin{aligned} & -\frac{1}{2}\alpha\sigma^2 r^2 + (a - b\alpha)r + \lambda(2e^{r(\beta-\alpha)} - 1) \leq 0 \\ \text{and} \qquad & \lambda(2e^{r(\beta-\alpha)} - 1) \leq \frac{1}{2}\alpha\sigma^2 r^2 - (a - b\alpha)r = (\frac{1}{2}\alpha\sigma^2 r - (a - b\alpha))r, \end{aligned}$$

so that we obtain

$$\lambda \leq \frac{(\frac{1}{2}\alpha\sigma^2 r - (a - b\alpha))r}{2e^{r(\beta-\alpha)} - 1}.$$

Hence, we have

$$(L_{CIR} + \lambda)u \leq 0, \ \text{ in } (\alpha, \beta),$$

if

$$0 \leq \lambda \leq \frac{(\frac{1}{2}\alpha\sigma^2 r - (a - b\alpha))r}{2e^{r(\beta-\alpha)} - 1}.$$

Let

$$\frac{1}{2}\alpha\sigma^2 r - (a - b\alpha) = 1,$$

and we can choose

$$r = \frac{2(1 + a - b\alpha)}{\alpha\sigma^2} = -\frac{2b}{\sigma^2} + \frac{2(1+a)}{\alpha\sigma^2}.$$

Obviously, r satisfies $r > -\frac{2b}{\sigma^2}$. After tedious calculations,

$$\begin{aligned} & \frac{(\frac{1}{2}\alpha\sigma^2 r - (a - b\alpha))r}{2e^{r(\beta-\alpha)} - 1} \\ = \ & \frac{r}{2e^{r(\beta-\alpha)} - 1} = \frac{\frac{2(1+a-b\alpha)}{\alpha\sigma^2}}{2e^{r(\beta-\alpha)} - 1} \\ \geq \ & \frac{\frac{2(1+a-b\alpha)}{\alpha\sigma^2}}{2e^{r(\beta-\alpha)}} \geq \frac{\frac{-b\alpha}{\alpha\sigma^2}}{e^{r(\beta-\alpha)}} = \frac{-b}{\sigma^2} e^{-r(\beta-\alpha)}, \end{aligned}$$

and choosing

$$r = \frac{2(1+a-b\alpha)}{\alpha\sigma^2},$$

we can obtain

$$\frac{(\frac{1}{2}\alpha\sigma^2 r - (a-b\alpha))r}{2e^{r(\beta-\alpha)}-1} \geq -\frac{b}{\sigma^2}e^{-\frac{2(1+a-b\alpha)}{\alpha\sigma^2}(\beta-\alpha)} := \hat{\lambda}(\alpha,\beta),$$

Here, we provide the definition of $\hat{\lambda}(\alpha,\beta)$, which we will frequently use below.

We have thus shown that there exists a function $u > 0$ on $[\alpha,\beta]$ satisfying

$$\begin{cases} (L_{CIR}+\hat{\lambda})u \leq 0, & \text{in } (\alpha,\beta), \\ u(\alpha) = 1, \\ u'(\beta) \geq 0. \end{cases}$$

Let $\lambda_1(-L)$ be the principal eigenvalue of the second-order elliptic operator for $-L$; according to the criticality theory of the second-order elliptic operators, it follows that the principal eigenvalue $\lambda_1(-L_{CIR})$ satisfies

$$\lambda_1(-L_{CIR}) \geq \hat{\lambda},$$

where the second-order elliptic operator L_{CIR} satisfies

$$\begin{cases} L_{CIR}u = 0, & \text{in } (\alpha,\beta), \\ u(\alpha) = 1, \\ u'(\beta) = 0. \end{cases}$$

According to the Feynman–Kac formula, if $\lambda \leq \lambda_1(-L_{CIR})$, then the function

$$u_\lambda(x) := E_x^{CIR;Ref\leftarrow:\beta}e^{\lambda T_\alpha}$$

satisfies the following boundary value problem

$$\begin{cases} (L_{CIR}+\lambda)u = 0, & \text{in } (\alpha,\beta), \\ u(\alpha) = 1, \\ u'(\beta) = 0. \end{cases}$$

According to the generalized maximum principal, it follows from $\lambda \leq \lambda_1(-L_{CIR})$ that

$$u_\lambda \leq u,$$

where, u satisfies

$$\begin{cases} (L_{CIR}+\lambda)u \leq 0, & \text{in } (\alpha,\beta), \\ u(\alpha) = 1, \\ u'(\beta) \geq 0, \end{cases} \tag{2}$$

Obviously, $u(x) = 2 - e^{-r(x-\alpha)}$ satisfies (2). Hence, in particular, we have

$$E_x^{CIR;Ref\leftarrow:\beta}e^{\lambda T_\alpha} = u_\lambda(x) \leq u(x) \leq 2.$$

This completes the proof of this lemma. □

3. Recurrence of the CIR Model When $b < 0$

Transience recurrence dichotomous issues are central to the study of stochastic processes and help describe the stochastic process's overall structure. There are many equivalent definitions of transience versus recurrence dichotomy in many of the literature; here, we can refer to [18,19].

Definition 1. *The stochastic process $X(t)$ is recurrent if $X(t)$ belongs to $\mathcal{O}$ at arbitrarily large times t, with a probability of one, and is transient if $X(t)$ belongs to $\mathcal{O}$ at arbitrarily large times t, with a probability of zero, for any set $\mathcal{O}$.*

In this section, we prove the results of two recurrent properties, that is, recurrence and positive recurrence. The definition of positive recurrence for a stochastic process $X(t)$ is given in the following subsection.

3.1. Recurrence

Theorem 1. *Consider the CIR model corresponding to the generator*

$$\frac{1}{2}\sigma^2 x\frac{d^2}{dx^2} + (a - bx)\frac{d}{dx}$$

in the time-dependent region $[1, f(t)]$ with reflection at both the fixed endpoint and the time-dependent one. Let $\sigma > 0, b < 0$, and $a > 0$ satisfy that there is an $a_0 < \frac{1}{2}$ and an $a \leq a_0$, where a_0 solves $2a = e^{\frac{2b}{\sigma^2} - \frac{2(1+a)}{\sigma^2}}$. Assume that $f(t) \leq \ln t$ for sufficiently large t. If

$$b > -\frac{\sigma^2}{2},$$

or if

$$b = -\frac{\sigma^2}{2} \text{ and } a < \frac{\sigma^2}{2},$$

then the CIR model is recurrent.

Proof. Let $j_0 \geq 3$ and $t_j = e^j$. Then, we have $f(t_j) > 2$ for $j \geq j_0$. For $j \geq j_0$, let A_{j+1} denote the event that the CIR process hits 1 at some time $t \in [t_j, t_{j+1}]$. The conditional version of the Borel–Cantelli lemma shows that if

$$\sum_{j=j_0}^{\infty} P_1(A_{j+1}|\mathcal{F}_{t_j}) = \infty, \quad \text{a.s.,} \tag{3}$$

then $P_1(A_j, \text{i.o.}) = 1$, and thus the CIR process is recurrent. Thus, to show recurrence of the CIR process, it suffices to show (3).

Since up to time t_j, the largest that the CIR process can be is $f(t_j)$, and since up to time t_{j+1}, the time-dependent region is contained in $[1, f(t_{j+1})]$, it follows by comparison that

$$P_1(A_{j+1}|\mathcal{F}_{t_j}) \geq P_{f(t_j)}^{CIR;Ref\leftarrow:f(t_{j+1})}(T_1 \leq t_{j+1} - t_j), \quad \text{a.s.} \tag{4}$$

Now, we estimate $P_{f(t_j)}^{CIR;Ref\leftarrow:f(t_{j+1})}(T_1 \leq t_{j+1} - t_j)$. Let

$$\begin{aligned}
\sigma_0^{(j)} &= 0, \\
\tau_i^{(j)} &= \inf\{t \geq \sigma_{i-1}^j | X(t) = f(t_{j+1})\}, \\
\sigma_i^{(j)} &= \inf\{t > \tau_i^{(j)} | X(t) = f(t_j)\}, \quad j \geq j_0, \ i = 1, 2, \cdots.
\end{aligned}$$

For any $l_j \in \mathbb{N}$,

$$\{T_1 < \sigma_{l_j}^{(j)}\} - \{\sigma_{l_j}^{(j)}\} > t_{j+1} - t_j\} \subset \{T_1 < t_{j+1} - t_j\}.$$

It follows from the strong Markov property that

$$P^{CIR;Ref\leftarrow:f(t_{j+1})}_{f(t_j)}(T_1 < \sigma^{(j)}_{l_j}) = 1 - \left(P^{CIR;Ref\leftarrow:f(t_{j+1})}_{f(t_j)}(T_{f(t_{j+1})} < T_1)\right)^{l_j}.$$

Thus, we have

$$\begin{aligned} & P^{CIR;Ref\leftarrow:f(t_{j+1})}_{f(t_j)}(T_1 \le t_{j+1} - t_j) \\ \ge\ & 1 - \left(P^{CIR;Ref\leftarrow:f(t_{j+1})}_{f(t_j)}(T_{f(t_{j+1})} < T_1)\right)^{l_j} - P^{CIR;Ref\leftarrow:f(t_{j+1})}_{f(t_j)}(\sigma^{(j)}_{l_j} > t_{j+1} - t_j). \end{aligned}$$

We then will obtain $P_1(A_j \text{ i.o.}) = 1$, and thus recurrence, if we can select $\{l_j\}_{j=1}^{\infty}$ such that

$$\sum_{j=j_0}^{\infty} \left(1 - \left(P^{CIR;Ref\leftarrow:f(t_{j+1})}_{f(t_j)}(T_{f(t_{j+1})} < T_1)\right)^{l_j}\right) = \infty, \tag{5}$$

and

$$\sum_{j=j_0}^{\infty} P^{CIR;Ref\leftarrow:f(t_{j+1})}_{f(t_j)}(\sigma^{(j)}_{l_j} > t_{j+1} - t_j) < \infty. \tag{6}$$

Define

$$\phi(x) := \int_x^{\infty} t^{-\frac{2a}{\sigma^2}} e^{\frac{2b}{\sigma^2}(t-1)} dt.$$

Obviously,

$$L^{CIR}\phi(x) = 0.$$

According to the standard probabilistic potential theory, it follows that

$$\begin{aligned} & P^{CIR;Ref\leftarrow:f(t_{j+1})}_{f(t_j)}(T_{f(t_{j+1})} < T_1) \\ =\ & \frac{\phi(1) - \phi(f(t_j))}{\phi(1) - \phi(f(t_{j+1}))} \\ =\ & 1 - \frac{\phi(f(t_j)) - \phi(f(t_{j+1}))}{\phi(1) - \phi(f(t_{j+1}))}. \end{aligned}$$

We compute using L'Hôpital's rule that

$$\lim_{x\to\infty} \frac{\phi(x)}{-\phi'(x)} = -\frac{\sigma^2}{2b},$$

and, we get, as $x \to \infty$,

$$\phi(x) \sim \frac{\sigma^2}{2b}\phi'(x) = -\frac{\sigma^2}{2b} x^{-\frac{2a}{\sigma^2}} e^{\frac{2b}{\sigma^2}(x-1)},$$

where $\sim$ indicates asymptotic equality in the sense that the ratio of the two sides goes to 1 as $x \to \infty$. Using the fact that

$$(1-t)^l \le e^{-lt} \le 1 - lt + \frac{1}{2}(lt)^2 \le 1 - \frac{1}{2}lt,$$

if $l, t \ge 0$ and $lt \le 1$, we have

$$1 - \left(P^{CIR;Ref\leftarrow:f(t_{j+1})}_{f(t_j)}(T_{f(t_{j+1})} < T_1)\right)^{l_j} \ge \frac{1}{2} l_j \frac{\phi(f(t_j)) - \phi(f(t_{j+1}))}{\phi(1) - \phi(f(t_{j+1}))}, \tag{7}$$

for sufficiently large j, if $\lim_{j\to\infty} l_j\phi(f(t_j)) = 0$. Obviously, we can choose a $C_0 \in (0,1)$ such that $\phi(f(t_{j+1})) \le C_0\phi(f(t_j))$ for all large j. Thus, for all sufficiently large j, we have

$$\frac{\phi(f(t_j)) - \phi(f(t_{j+1}))}{\phi(1) - \phi(f(t_{j+1}))} \ge C_1\phi(f(t_j)) \ge C_2 j^{-\frac{2a}{\sigma^2}} e^{\frac{2b}{\sigma^2}(j-1)},$$

for some constants $C_1, C_2 > 0$. Now, we choose $l_j \in \mathbb{N}$ according to

$$l_j := \Big[\frac{1}{\log j} j^{\frac{2a}{\sigma^2}-1} e^{-\frac{2b}{\sigma^2}(j-1)}\Big].$$

Hence, we obtain

$$\begin{aligned}
& \sum_{j=j_0}^{\infty} \left(1 - \left(P_{f(t_j)}^{CIR;Ref\leftarrow:f(t_{j+1})}(T_{f(t_{j+1})} < T_1)\right)^{l_j}\right) \\
\ge & \sum_{j=j_0}^{\infty} C l_j j^{-\frac{2a}{\sigma^2}} e^{\frac{2b}{\sigma^2}(j-1)} \\
\ge & \sum_{j=j_0}^{\infty} C \frac{1}{j\log j} \\
= & \infty,
\end{aligned}$$

for the constant $C > 0$.

With l_j chosen as above, we now analyze the second term

$$P_{f(t_j)}^{CIR;Ref\leftarrow:f(t_{j+1})}(\sigma_{l_j}^{(j)} > t_{j+1} - t_j).$$

It follows from the strong Markov property that

$$\sigma_{l_j}^{(j)} = \sum_{i=1}^{l_j} X_i + \sum_{i=1}^{l_j} Y_i,$$

where $\{X_i\}_{i=1}^{\infty}$ is an independent and identically distributed sequence distributed according to $T_{f(t_{j+1})}$ under $P_{f(t_j)}^{CIR;Ref\to:1}$, $\{Y_i\}_{i=1}^{\infty}$ is an independent and identically distributed sequence distributed according to $T_{f(t_j)}$ under $P_{f(t_{j+1})}^{CIR;Ref\leftarrow:f(t_{j+1})}$, and the two sequences are independent of one another.

For any $\lambda > 0$, according to Markov's inequality,

$$\begin{aligned}
& P_{f(t_j)}^{CIR;Ref\leftarrow:f(t_{j+1})}(\sigma_{l_j}^{(j)} > t) \\
\le & e^{-\lambda t} E_{f(t_j)}^{CIR;Ref\leftarrow:f(t_{j+1})}[e^{\lambda\sigma_{l_j}^{(j)}}] \\
= & e^{-\lambda t} E_{f(t_j)}^{CIR;Ref\leftarrow:f(t_{j+1})}[e^{\lambda\sum_{i=1}^{l_j} X_i} e^{\lambda\sum_{i=1}^{l_j} Y_i}] \\
= & e^{-\lambda t}\left(E_{f(t_j)}^{CIR;Ref\to:1}[e^{\lambda T_{f(t_{j+1})}}]\right)^{l_j}\left(E_{f(t_{j+1})}^{CIR;Ref\leftarrow:f(t_{j+1})}[e^{\lambda T_{f(t_j)}}]\right)^{l_j}. \qquad (8)
\end{aligned}$$

According to Lemma 1,

$$E_{f(t_{j+1})}^{CIR;Ref\leftarrow:f(t_{j+1})}[e^{\lambda T_{f(t_j)}}] \le 2, \qquad (9)$$

for $\lambda \le \hat{\lambda}(f(t_j), f(t_{j+1}))$, where $\hat{\lambda}(\cdot,\cdot)$ is as in Lemma 1. Using the fact that $f(t_j) = \ln(e^j) = j$, it is easy to check that there exists a $\hat{\lambda}_0 > 0$ such that $\hat{\lambda}(f(t_j), f(t_{j+1})) \ge \hat{\lambda}_0$ for all $j \ge 1$. In fact, by the definition of $\hat{\lambda}(\cdot,\cdot)$ in Lemma 1,

$$\hat{\lambda}(f(t_j), f(t_{j+1})) = -\frac{b}{\sigma^2} e^{\frac{2b}{\sigma^2} - \frac{2(1+a)}{\sigma^2 j}} \to -\frac{b}{\sigma^2} e^{\frac{2b}{\sigma^2}}$$

as $j \to \infty$ if $1 + a > 0$. Hence, we obtain

$$\hat{\lambda}_0 = -\frac{b}{\sigma^2} e^{\frac{2b}{\sigma^2} - \frac{2(1+a)}{\sigma^2}} > 0.$$

By choosing $\lambda_j = -\frac{2ab}{\sigma^2 j}$, there exists a j_0, and we have $\lambda_j \le \hat{\lambda}_0$ for all $j \ge j_0$. By choosing $\lambda = \lambda_1 = -\frac{2ab}{\sigma^2}$, we have $\lambda \le \hat{\lambda}_0$ if a satisfies the following inequality:

$$a \le \frac{1}{2} e^{\frac{2b}{\sigma^2} - \frac{2(1+a)}{\sigma^2}},$$

that is, there is a $a_0 < \frac{1}{2}$, $a \le a_0$, where a_0 solves $2a = e^{\frac{2b}{\sigma^2} - \frac{2(1+a)}{\sigma^2}}$.

Using Lemma A6, we substitute $x = f(t_j) = j$ and $\beta = f(t_{j+1}) = (j+1)$ in the expression on the right-hand side of (A11); the resulting expression is bounded in j. In fact, it follows from (A11) that

$$\begin{aligned}
u(j) &= e^{-\frac{2b}{\sigma^2 j}} \exp\left(-\int_j^{j+1} \frac{\exp(-\frac{2b(y-1)}{\sigma^2 j}) y^{-\frac{2a}{\sigma^2}}}{\int_1^y \exp(-\frac{2b(t-1)}{\sigma^2 j}) t^{-\frac{2a}{\sigma^2}} dt - \frac{\sigma^2 j}{2b}} dy \right) \\
&\le e^{-\frac{2b}{\sigma^2 j}} \exp\left(-\int_j^{j+1} \frac{\exp(-\frac{2b(j-1)}{\sigma^2 j})(j+1)^{-\frac{2a}{\sigma^2}}}{\int_1^{j+1} \exp(-\frac{2b(t-1)}{\sigma^2 j}) t^{-\frac{2a}{\sigma^2}} dt - \frac{\sigma^2 j}{2b}} dy \right) \\
&\le e^{-\frac{2b}{\sigma^2 j}} \exp\left(-\frac{\exp(-\frac{2b(j-1)}{\sigma^2 j})(j+1)^{-\frac{2a}{\sigma^2}}}{\int_1^{j+1} \exp(-\frac{2bj}{\sigma^2 j}) dt - \frac{\sigma^2 j}{2b}} \right) \\
&= e^{-\frac{2b}{\sigma^2 j}} \exp\left(-\frac{\exp(-\frac{2b(j-1)}{\sigma^2 j})(j+1)^{-\frac{2a}{\sigma^2}}}{\int_1^{j+1} \exp(-\frac{2b}{\sigma^2}) dt - \frac{\sigma^2 j}{2b}} \right) \\
&\le e^{-\frac{2b}{\sigma^2 j}} \exp\left(-\frac{\exp(-\frac{2b(j-1)}{\sigma^2 j})(j+1)^{-\frac{2a}{\sigma^2}}}{(j+1)\exp(-\frac{2b}{\sigma^2}) - \frac{\sigma^2}{2b}(j+1)} \right) \\
&= e^{-\frac{2b}{\sigma^2 j}} \exp\left(-\frac{1}{(j+1)^{1+\frac{2a}{\sigma^2}} \exp(-\frac{2b}{\sigma^2 j}) - \frac{\sigma^2}{2b} \exp(\frac{2b(j-1)}{\sigma^2 j})(j+1)^{1+\frac{2a}{\sigma^2}}} \right).
\end{aligned}$$

Obviously, notice that when $b < 0$,

$$\lim_{j\to\infty} \frac{1}{(j+1)^{1+\frac{2a}{\sigma^2}} \exp(-\frac{2b}{\sigma^2 j}) - \frac{\sigma^2}{2b} \exp(\frac{2b(j-1)}{\sigma^2 j})(j+1)^{1+\frac{2a}{\sigma^2}}} = 0.$$

Hence, we have

$$u(j) \le e^{-\frac{2b}{\sigma^2 j}} \le e^{-\frac{2b}{\sigma^2}},$$

for sufficiently large $j \ge 1$ when $b < 0$.

By letting $M := e^{-\frac{2b}{\sigma^2}} > 1$ be an upper bound, it follows that

$$E^{CIR;Ref\to:1}_{f(t_j)} e^{\lambda T_{f(t_{j+1})}} = E^{CIR;Ref\to:1}_{f(t_j)} e^{-\frac{2ab}{\sigma^2} T_{f(t_{j+1})}} \leq M. \tag{10}$$

By noting that $t_{j+1} - t_j = e^{j+1} - e^j \geq e^j$, it follows from (8) that

$$P^{CIR;Ref\leftarrow:f(t_{j+1})}_{f(t_j)}(\sigma^{(j)}_{l_j} > t_{j+1} - t_j) \leq e^{\frac{2ab}{\sigma^2} e^j} (2M)^{l_j}, \tag{11}$$

for sufficiently large j. Recalling the expression of l_j, we can have

$$\begin{aligned} & P^{CIR;Ref\leftarrow:f(t_{j+1})}_{f(t_j)}(\sigma^{(j)}_{l_j} > t_{j+1} - t_j) \\ \leq\ & e^{\frac{2ab}{\sigma^2} e^j} (2M)^{\frac{1}{\log j} j^{\frac{2a}{\sigma^2}-1} e^{-\frac{2b}{\sigma^2}(j-1)}} \\ =\ & e^{\frac{2ab}{\sigma^2} e^j} e^{\frac{1}{\log j} j^{\frac{2a}{\sigma^2}-1} e^{-\frac{2b}{\sigma^2}(j-1)} \log 2M}, \end{aligned} \tag{12}$$

for sufficiently large j. It follows that the right-hand side of (12) is summable in j if $1 > -\frac{2b}{\sigma^2}$, that is,

$$b > -\frac{\sigma^2}{2},$$

or if

$$b = -\frac{\sigma^2}{2} \text{ and } a \leq \frac{\sigma^2}{2}.$$

Thus, (6) holds for this range of a, b and σ. This completes the proof of this theorem. □

Remark 1. *In the time-independent region case, it is known that the drift $a - bX_t$ ensures a mean reversion of the CIR model towards the long-term value $\frac{a}{b}$. In the time-dependent region case, however, the CIR model can reflect at the fixed endpoint* 1*. Obviously, $a \leq \frac{1}{2} e^{\frac{2b}{\sigma^2}} < \frac{1}{2}$. This guarantees that the CIR model can down-cross the boundary* 1*; hence, the CIR model can reflect at the fixed point* 1 *infinitely often.*

3.2. Positive Recurrence

Now that we have the recurrence of the CIR model, it is natural to consider the positive recurrence in the following sense. The following definition of positive recurrence for a stochastic process can be found in [18].

Definition 2. *We say that a one-dimensional process is a positive recurrence if, starting from $x > 1$, the expected value of the first hitting time of* 1 *is finite, that is,*

$$E_x T_1 < \infty.$$

Theorem 2. *Consider the CIR model corresponding to the generator*

$$\frac{1}{2}\sigma^2 x \frac{d^2}{dx^2} + (a - bx)\frac{d}{dx}$$

in the time-dependent region $[1, f(t)]$, with reflection at both the fixed endpoint 1 *and the time-dependent endpoint $f(t)$ at time t. Let $\sigma > 0, b < 0$, and $a > 0$ satisfy that there is an $a_0 < \frac{1}{2}$ and an $a \leq a_0$, where a_0 solves $2a = e^{\frac{2b}{\sigma^2} - \frac{2(1+a)}{\sigma^2}}$. Assume that $f(t) \leq \ln t$, for sufficiently large t. If*

$$b > -\frac{\sigma^2}{2},$$

then the CIR model is positive recurrent.

Proof. Let P_2 and E_2 denote probabilities and expectations for the process starting from $x = 2$ at time 0. Let $t_j = e^j$ as in the proof of Theorem 1. We have

$$E_2 T_1 \leq t_1 + \sum_{j=1}^{\infty} t_{j+1} P_2(T_1 \geq t_j) = e + \sum_{j=1}^{\infty} e^{j+1} P_2(T_1 \geq t_j).$$

Let A_{j+1} denote the event that the process hits 1 at some time $t \in [t_j, t_{j+1}]$. We have, for $j \geq j_0 + 1$,

$$\begin{aligned} P_2(T_1 \geq t_j) &\leq P_2(\cap_{i=j_0}^{j-1} A_{i+1}^c) \\ &\leq \prod_{i=j_0}^{j-1} \left(1 - P_{f(t_j)}^{CIR;Ref\leftarrow:f(t_{j+1})}(T_1 \leq t_{i+1} - t_i)\right). \end{aligned} \tag{13}$$

If we show that

$$\lim_{j\to\infty} P_{f(t_j)}^{CIR;Ref\leftarrow:f(t_{j+1})}(T_1 \leq t_{j+1} - t_j) = 1,$$

then it will certainly follow that

$$E_2 T_1 < \infty,$$

thereby proving positive recurrence. In order to prove this, it suffices from

$$\begin{aligned} & P_{f(t_j)}^{CIR;Ref\leftarrow:f(t_{j+1})}(T_1 \leq t_{j+1} - t_j) \\ \geq\ & 1 - \left(P_{f(t_j)}^{CIR;Ref\leftarrow:f(t_{j+1})}(T_{f(t_{j+1})} < T_1)\right)^{l_j} - P_{f(t_j)}^{CIR;Ref\leftarrow:f(t_{j+1})}(\sigma_{l_j}^{(j)} > t_{j+1} - t_j). \end{aligned}$$

to prove that for some choice of positive integers $\{l_j\}_{l=j_0}^{\infty}$,

$$\lim_{j\to\infty} \left(P_{f(t_j)}^{CIR;Ref\leftarrow:f(t_{j+1})}(T_{f(t_{j+1})} < T_1)\right)^{l_j} = 0, \tag{14}$$

and

$$\lim_{j\to\infty} P_{f(t_j)}^{CIR;Ref\leftarrow:f(t_{j+1})}(\sigma_{l_j}^{(j)} > t_{j+1} - t_j) = 0. \tag{15}$$

According to the standard probabilistic potential theory, we have

$$\begin{aligned} & P_{f(t_j)}^{CIR;Ref\leftarrow:f(t_{j+1})}(T_{f(t_{j+1})} < T_1) \\ =\ & \frac{\phi(1) - \phi(f(t_j))}{\phi(1) - \phi(f(t_{j+1}))} \\ =\ & 1 - \frac{\phi(f(t_j)) - \phi(f(t_{j+1}))}{\phi(1) - \phi(f(t_{j+1}))}, \end{aligned} \tag{16}$$

Here, we have $\phi(x)$ as in the proof of Theorem 1, that is,

$$\phi(x) := \int_x^{\infty} t^{-\frac{2a}{\sigma^2}} e^{\frac{2b}{\sigma^2}(t-1)} dt.$$

Thus, for all sufficiently large j, by combining (16) with

$$\frac{\phi(f(t_j)) - \phi(f(t_{j+1}))}{\phi(1) - \phi(f(t_{j+1}))} \geq C_1 \phi(f(t_j)) \geq C_2 j^{-\frac{2a}{\sigma^2}} e^{\frac{2b}{\sigma^2}(j-1)},$$

we obtain

$$\begin{aligned} & P_{f(t_j)}^{CIR;Ref\leftarrow:f(t_{j+1})}(T_{f(t_{j+1})} < T_1) \\ = \;& 1 - \frac{\phi(f(t_j)) - \phi(f(t_{j+1}))}{\phi(1) - \phi(f(t_{j+1}))} \\ \leq \;& 1 - C_2 j^{-\frac{2a}{\sigma^2}} e^{\frac{2b}{\sigma^2}(j-1)}, \end{aligned}$$

Hence,

$$\begin{aligned} \left(P_{f(t_j)}^{CIR;Ref\leftarrow:f(t_{j+1})}(T_{f(t_{j+1})} < T_1)\right)^{l_j} &\leq \left(1 - C_2 j^{-\frac{2a}{\sigma^2}} e^{\frac{2b}{\sigma^2}(j-1)}\right)^{l_j} \\ &= \left(1 - \frac{C_2}{j^{\frac{2a}{\sigma^2}} e^{-\frac{2b}{\sigma^2}(j-1)}}\right)^{l_j}. \end{aligned}$$

We choose

$$l_j := \left[j^{\frac{2a}{\sigma^2}} \log j e^{-\frac{2b}{\sigma^2}(j-1)} \right].$$

It follows from the fact that

$$\lim_{y\to\infty} (1 - \frac{1}{y})^{yg(y)} = 0, \quad \text{if } \lim_{y\to\infty} g(y) = \infty,$$

that (14) holds. With this choice of l_j, we have, by (11),

$$\begin{aligned} & P_{f(t_j)}^{CIR;Ref\leftarrow:f(t_{j+1})}(\sigma_{l_j}^{(j)} > t_{j+1} - t_j) \\ \leq \;& e^{\frac{2ab}{\sigma^2} e^j} (2M)^{l_j} \\ = \;& e^{\frac{2ab}{\sigma^2} e^j} e^{j^{\frac{2a}{\sigma^2}} \log j e^{-\frac{2b}{\sigma^2}(j-1)} \log(2M)}. \end{aligned} \tag{17}$$

Thus, if

$$a > 0 \text{ and } -\frac{2b}{\sigma^2} < 1, \ (i.e., b > -\frac{\sigma^2}{2}),$$

it follows from (17) that

$$\lim_{j\to\infty} P_{f(t_j)}^{CIR;Ref\leftarrow:f(t_{j+1})}(\sigma_{l_j}^{(j)} > t_{j+1} - t_j) = 0.$$

This completes the proof of the theorem. □

4. Transience of the CIR Model When $b < 0$

Theorem 3. *Consider the CIR model corresponding to the generator*

$$\frac{1}{2}\sigma^2 x \frac{d^2}{dx^2} + (a - bx)\frac{d}{dx}$$

in the time-dependent region $[1, f(t)]$*, with reflection at both the fixed endpoint and the time-dependent one. Let* $\sigma > 0, b < 0$*, and* $a > 0$ *satisfy that there is an* $a_0 < \frac{1}{2}$ *and an* $a \leq a_0$*, where* a_0 *solves* $2a = e^{\frac{2b}{\sigma^2} - \frac{2(1+a)}{\sigma^2}}$*. Assume that* $f(t) \geq \ln t$*, for sufficiently large* t*. If*

$$b < -\sigma^2,$$

then the CIR model is transient.

Proof. Let $j_1 = e^2 + 1$; then, $f(j) = \ln j > 2$ for all $j \geq j_1$. Let B_1 be the event that the CIR process hits 1 sometimes between the first time it hits $f(j)$ and the first time it hits $f(t_{j+1})$:

$$B_j := \{X(t) = 1, \text{ for some } t \in (T_{f(t_j)}, T_{f(t_{j+1})})\}.$$

If we show that

$$\sum_{j=j_1}^{\infty} P_1(B_j) < \infty, \tag{18}$$

then, according to the Borel–Cantelli lemma, it will follow that $P_1(B_j, \text{ i.o.}) = 0$, and consequently the CIR process is transient.

To consider whether or not the event B_j occurs, we first wait until time $T_{f(t_j)}$. Hence, we have $T_{f(j)} \geq j$, since $f(j)$ is not accessible to the process before time j. Since we may have $T_{f(j)} < j + 1$, the point $f(j + 1)$ may not be accessible to the process at time $T_{f(j)}$. However, when we wait for one unit of time, then after that, the point $f(j + 1)$ certainly will be accessible because of $T_{f(j)} + 1 \geq j + 1$.

Let $M_j < f(j) - 1$. So, the process never got to the level $f(j) - M_j$ in that one unit of time; then, the probability of B_j occurring is no more than $P^{CIR;Ref\leftarrow:f(j+1)}_{f(j)-M_j}(T_1 < T_{f(j+1)})$. By comparison with the process that is reflected at the fixed point $f(j)$, the probability that the process will get to the level $f(j) - M_j$ in that one unit of time is bounded from above by $P^{CIR;Ref\leftarrow:f(j)}_{f(j)}(T_{f(j)-M_j} \leq 1)$. From these considerations above, we have

$$P_1(B_j) \leq P^{CIR;Ref\leftarrow:f(j+1)}_{f(j)-M_j}(T_1 < T_{f(j+1)}) + P^{CIR;Ref\leftarrow:f(j)}_{f(j)}(T_{f(j)-M_j} \leq 1). \tag{19}$$

It follows by standard probabilistic potential theory that

$$P^{CIR;Ref\leftarrow:f(j+1)}_{f(j)-M_j}(T_1 < T_{f(j+1)}) = \frac{\phi(f(j) - M_j) - \phi(f(j+1))}{\phi(1) - \phi(f(j+1))}. \tag{20}$$

We choose $M_j = \frac{1}{2}f(j)$ because of $M_j < f(j) - 1$. Recall that $f(j) \geq \log j$. Then, we have

$$\begin{aligned} \phi(f(j) - M_j) &= \phi(\frac{1}{2}f(j)) = \phi(\frac{1}{2}\log j) \\ &\sim -\frac{\sigma^2}{2b}(\frac{1}{2}\log j)^{-\frac{2a}{\sigma^2}} e^{\frac{2b}{\sigma^2}(\frac{1}{2}\log j)} \\ &= -\frac{\sigma^2}{2b}(\frac{1}{2}\log j)^{-\frac{2a}{\sigma^2}} j^{\frac{b}{\sigma^2}}. \end{aligned}$$

By the assumption that $b < -\sigma^2$, we have

$$-\frac{b}{\sigma^2} > 1.$$

Hence, it follows from (20) that

$$\sum_{j=j_1}^{\infty} P^{CIR;Ref\leftarrow:f(j+1)}_{f(j)-M_j}(T_1 < T_{f(j+1)}) < \infty. \tag{21}$$

We now estimate

$$P^{CIR;Ref\leftarrow:f(j)}_{f(j)}(T_{f(j)-M_j} \leq 1),$$

where $M_j = \frac{1}{2}f(j)$. According to Markov's inequality, we have, for $\lambda > 0$,

$$P^{CIR;Ref\leftarrow:f(j)}_{f(j)}(T_{f(j)-M_j} \leq 1) \leq e^{\lambda} E^{CIR;Ref\leftarrow:f(j)}_{f(j)}[e^{-\lambda T_{f(j)-M_j}}]. \tag{22}$$

By comparison, we have

$$E_{f(j)}^{CIR;Ref\leftarrow:f(j)}[e^{-\lambda T_{f(j)-M_j}}] \leq E_{\beta}^{\frac{a-b\alpha}{\sigma^2\beta};Ref\leftarrow:\beta} \exp\Big(-\frac{\lambda}{\sigma^2\beta}T_{\alpha}\Big). \tag{23}$$

According to Lemma A5 with $\alpha = f(j) - M_j = \frac{1}{2}f(j)$ and $\beta = f(j)$, we have

$$\begin{aligned}
& P_{f(j)}^{CIR;Ref\leftarrow:f(j)}(T_{f(j)-M_j} \leq 1) \\
\leq\ & e^{\lambda}\frac{(r_1+r_2)e^{-2\frac{a-b\beta}{\sigma^2\beta}(\beta-\alpha)}}{r_1e^{-r_1(\beta-\alpha)}+r_2e^{r_2(\beta-\alpha)}} \\
=\ & e^{\lambda}\frac{(r_1+r_2)e^{-2\frac{a-bf(j)}{\sigma^2f(j)}M_j}}{r_1e^{-r_1M_j}+r_2e^{r_2M_j}} \\
\leq\ & e^{\lambda}\frac{(r_1+r_2)e^{-2\frac{a-bf(j)}{\sigma^2f(j)}M_j}}{r_2e^{r_2M_j}} \\
=\ & e^{\lambda}(1+\frac{r_1}{r_2})e^{-r_2M_j-2\frac{a-bf(j)}{\sigma^2f(j)}M_j}.
\end{aligned}$$

Due to the complexity of $\frac{r_1}{r_2}$, we handle it separately. By substituting the specific expressions of r_1 and r_2 from Lemma A5 into $\frac{r_1}{r_2}$, after some calculations, we obtain

$$\begin{aligned}
\frac{r_1}{r_2} &= \frac{\sqrt{\Big(\frac{a-b\beta}{\sigma^2\beta}\Big)^2+\frac{2\lambda}{\sigma^2\beta}}+\frac{a-b\beta}{\sigma^2\beta}}{\sqrt{\Big(\frac{a-b\beta}{\sigma^2\beta}\Big)^2+\frac{2\lambda}{\sigma^2\beta}}-\frac{a-b\beta}{\sigma^2\beta}} \\
&= \frac{\sqrt{\Big(\frac{a-bf(j)}{\sigma^2f(j)}\Big)^2+\frac{2\lambda}{\sigma^2f(j)}}+\frac{a-bf(j)}{\sigma^2f(j)}}{\sqrt{\Big(\frac{a-bf(j)}{\sigma^2f(j)}\Big)^2+\frac{2\lambda}{\sigma^2f(j)}}-\frac{a-bf(j)}{\sigma^2f(j)}} \\
&= \frac{\Big(\sqrt{\Big(\frac{a-bf(j)}{\sigma^2f(j)}\Big)^2+\frac{2\lambda}{\sigma^2f(j)}}+\frac{a-bf(j)}{\sigma^2f(j)}\Big)^2}{\Big(\frac{a-bf(j)}{\sigma^2f(j)}\Big)^2+\frac{2\lambda}{\sigma^2f(j)}-(\frac{a-bf(j)}{\sigma^2f(j)})^2} \\
&= \frac{\sigma^2f(j)}{2\lambda}\Big(\sqrt{\Big(\frac{a-bf(j)}{\sigma^2f(j)}\Big)^2+\frac{2\lambda}{\sigma^2f(j)}}+\frac{a-bf(j)}{\sigma^2f(j)}\Big)^2 \\
&= \frac{\sigma^2}{2\lambda}\Big(\sqrt{\Big(\frac{a}{\sigma^2\sqrt{f(j)}}-\frac{b\sqrt{f(j)}}{\sigma^2}\Big)^2+\frac{2\lambda}{\sigma^2}}+\frac{a}{\sigma^2\sqrt{f(j)}}-\frac{b\sqrt{f(j)}}{\sigma^2}\Big)^2 \\
&\leq \frac{\sigma^2}{2\lambda}\Big[4(\frac{a}{\sigma^2\sqrt{f(j)}}-\frac{b\sqrt{f(j)}}{\sigma^2})^2+2(\frac{2\lambda}{\sigma^2})\Big] \\
&\leq \frac{\sigma^2}{2\lambda}\Big[8(\frac{a}{\sigma^2\sqrt{f(j)}})^2+8(\frac{b\sqrt{f(j)}}{\sigma^2})^2+2(\frac{2\lambda}{\sigma^2})\Big] \\
&= \frac{\sigma^2}{2\lambda}\Big[\frac{8a^2}{\sigma^4f(j)}+\frac{8b^2f(j)}{4\sigma^4}+\frac{4\lambda}{\sigma^2}\Big] \\
&= \frac{1}{\lambda\sigma^2}\Big[\frac{4a^2}{f(j)}+b^2f(j)+2\lambda^2\sigma^2\Big].
\end{aligned}$$

Hence, we have

$$\begin{aligned}
& P_{f(j)}^{CIR;Ref\leftarrow:f(j)}(T_{f(j)-M_j} \le 1) \\
= \; & e^{\lambda}(1+\frac{r_1}{r_2})\exp\Big(-\Big(\sqrt{\Big(\frac{a}{\sigma^2 f(j)}-\frac{b}{\sigma^2}\Big)^2+\frac{2\lambda}{\sigma^2 f(j)}} \\
& \qquad -\Big(\frac{a}{\sigma^2 f(j)}-\frac{b}{\sigma^2}\Big)+2\frac{a-bf(j)}{\sigma^2 f(j)}\Big)M_j\Big) \\
= \; & e^{\lambda}(1+\frac{r_1}{r_2})\exp\Big(-\Big(\sqrt{\Big(\frac{a}{\sigma^2 f(j)}-\frac{b}{\sigma^2}\Big)^2+\frac{2\lambda}{\sigma^2 f(j)}} \\
& \qquad +\Big(\frac{a}{\sigma^2 f(j)}-\frac{b}{\sigma^2}\Big)\Big)M_j\Big) \\
\sim \; & e^{\lambda}(1+\frac{1}{\lambda\sigma^2}\Big[b^2 f(j)+2\lambda^2\Big])\exp\Big(\frac{2b}{\sigma^2}M_j\Big) \\
= \; & e^{\lambda}(1+\frac{1}{\lambda\sigma^2}\Big[b^2 f(j)+2\lambda^2\Big])\exp\Big(\frac{b}{\sigma^2}f(j)\Big) \\
= \; & e^{\lambda}(1+\frac{1}{\lambda\sigma^2}\Big[b^2\ln j+2\lambda^2\Big])j^{\frac{b}{\sigma^2}}.
\end{aligned}$$

By the assumption $b < -\sigma^2$, we have

$$-\frac{b}{\sigma^2} > 1.$$

Then, we have

$$\sum_{j=j_0}^{\infty} P_{f(j)}^{CIR;Ref\leftarrow:f(j)}(T_{f(j)-M_j} \le 1) < \infty. \tag{24}$$

Now, (19), (21), and (24) give us (18), and this completes the proof of the theorem. □

Remark 2. *By comparing Theorems 1–3, we clearly find that there is a gap for b, that is,* $-\sigma^2 \le b < -\frac{\sigma^2}{2}$. *We expect that the CIR process is also recurrent in this gap of b. However, we cannot confirm this assertion because the estimates we use here cannot guarantee it.*

Last Passage Time

In the transient case, it is natural to consider the last passage time, which is a random time but not a stopping time. In recent years, last passage time has also played an increasing role in financial modeling, such as in models of default risk, models of insider trading, and the prices of European put and call options.

For the case of diffusion in the form of the CIR model, a differentiable increasing scale function is

$$s(x) = \int_c^x \exp\Big(-2\int_c^u \frac{b(v)}{\sigma^2(v)}dv\Big)du$$

for some choice of $c \in (0,\infty)$. Here in the CIR model, the drift coefficient $b(x) = a - bx$ and the diffusion coefficient $\sigma(x) = \sigma\sqrt{x}$; hence, we obtain the scale function

$$s(x) = C\int_c^x u^{-\frac{2a}{\sigma^2}} e^{\frac{2b}{\sigma^2}u}du, \tag{25}$$

as well as the constants $b < 0$ and $C = c^{\frac{2a}{\sigma^2}} e^{-\frac{2b}{\sigma^2}c}$.

Theorem 4. *Let X be a transient CIR process in Theorem 3 such that* $X_t \to +\infty$ *when* $t \to \infty$, *and the last time that X hits y is defined as*

$$\Gamma_y := \sup\{t : X_t = y\}.$$

Then,

$$P_x(\Gamma_y > t|\mathcal{F}_t) = \frac{s(X_t)}{s(y)} \wedge 1,$$

and the scale function is given in (25).

Proof. The following proof is classic and can be found in many classic textbooks of stochastic process, such as in [20,21], as well as [22]. Observe that $P_x(\Gamma_y > t|\mathcal{F}_t) = P_x(\inf_{s\geq t} X_s < y|\mathcal{F}_t)$; it follows from the Markov property of X that

$$P_x\Big(\inf_{s\geq t} X_s < y|\mathcal{F}_t\Big) = P_{X_t}\Big(\inf_{s\geq 0} X_s < y\Big) = P_{X_t}\Big(\sup_{s\geq 0}(-s(X_s)) > -s(y)\Big).$$

In the following fact, let M be a positive continuous local martingale such that $M_0 = x$, $M_t \geq 0$, and $\lim_{t\to\infty} M_t = 0$; then,

$$\sup_{t\geq 0} M_t \stackrel{\mathrm{d}}{=} \frac{x}{U},$$

where U is a random variable with a uniform law on $[0, 1]$. Hence,

$$P_x(\Gamma_y > t|\mathcal{F}_t) = P_{X_t}\Big(\sup_{s\geq 0}(-s(X_s)) > -s(y)\Big) = \frac{s(X_t)}{s(y)} \wedge 1.$$

This completes the proof of this theorem. □

5. Conclusions

This paper studies the transience/recurrence for CIR process when $b < 0$. By adding boundaries to a time-dependent domain, we obtained a CIR process when $b < 0$ with the transient property that became a CIR with a property of recurrence; however, the boundaries continue to grow over time. We have specified the conditions that the coefficients of the CIR process must meet when it is recurrent, positive recurrent and transient.

Author Contributions: Writing—original draft, M.Z. and G.Z.; Writing—review & editing, M.Z. and G.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the National Science Foundation of China (Grant No.11501325).

Data Availability Statement: All data used to support the findings of this study are included within the article.

Acknowledgments: The authors thank the editor and the referees for their valuable comments and suggestions, which improved greatly the quality of this paper.

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

Appendix A

We assume that $u'(x) \leq 0$ and $v'(x) \leq 0$ for all $x \in [1, \beta]$.

Lemma A1. *For $\lambda > 0$,*

$$\begin{cases} \frac{1}{2}u'' + bx^{\gamma}u' + \lambda u = 0, & in\ (1, \beta) \\ u'(1) = 0, \\ u(\beta) = 1, \end{cases} \tag{A1}$$

and

$$\begin{cases} \frac{1}{2}v'' + Dv' + \lambda v = 0, & in\ (1, \beta) \\ v'(1) = 0, \\ v(\beta) = 1, \end{cases} \tag{A2}$$

Then, $D \leq \min_{x\in[1,\beta]} bx^{\gamma}$ implies that

$$u(x) \leq v(x), \text{ for all } x \in [1,\beta].$$

Proof. Assume that the conclusion is false. According to $u, v \in C^2$, we assume $u(x) > v(x)$ for all $x \in [1,\beta)$ WLOG.

So, $u(1) > v(1), u'(1) = v'(1) = 0$ implies $u''(1) < v''(1)$ according to (A3) and (A4). Then, it follows that $u'(x) < v'(x)$ in $(1,c)$, with $1 < c \leq \beta$ and $u'(c) = v'(c)$. Then, it follows from

$$\frac{u'(c) - u'(c-h)}{h} > \frac{v'(c) - v'(c-h)}{h}$$

that $u''(c) \geq v''(c)$.

$u(c) > v(c), u'(c) = v'(c)$ and $u''(c) \geq v''(c)$, which are contradictory to (A3) and (A4). □

Lemma A2. *For $\lambda > 0$,*

$$\begin{cases} \frac{1}{2}u'' + bx^{\gamma}u' + \lambda_u u = 0, & in\ (1,\beta) \\ u'(1) = 0, \\ u(\beta) = 1, \end{cases} \tag{A3}$$

and

$$\begin{cases} \frac{1}{2}v'' + Dv' + \lambda_v v = 0, & in\ (1,\beta) \\ v'(1) = 0, \\ v(\beta) = 1, \end{cases} \tag{A4}$$

Then, $D \leq \min_{x\in[1,\beta]} bx^{\gamma}$ and $\lambda_u \geq \lambda_v$ imply that

$$u(x) \leq v(x), \text{ for all } x \in [1,\beta].$$

Proof. This proof is similar to the one in Lemma A1, so we omit it. □

Lemma A3. *Let u be the solution to the ODE*

$$\begin{cases} \frac{1}{2}u'' + \frac{a-bx}{\sigma^2 x}u' - \frac{\lambda}{\sigma^2 x}u = 0, & in\ (\alpha,\beta) \\ v(\alpha) = 1, \\ v'(\beta) = 0, \end{cases} \tag{A5}$$

and let v be the solution to the ODE

$$\begin{cases} \frac{1}{2}v'' + \frac{a-b\beta}{\sigma^2 \beta}v' - \frac{\lambda}{\sigma^2 \beta}v = 0, & in\ (\alpha,\beta) \\ v(\alpha) = 1, \\ v'(\beta) = 0. \end{cases} \tag{A6}$$

Then,

$$u(x) \leq v(x), \forall x \in [\alpha,\beta].$$

Proof. Assume that the conclusion is false. According to $u, v \in C^2$, we assume $u(x) > v(x)$ for all $x \in (\alpha,\beta]$ WLOG.

So, $u(\beta) > v(\beta), u'(\beta) = v'(\beta) = 0$ implies that $u''(\beta) < v''(\beta)$ according to (A5) and (A6). Then, it follows that $u'(x) < v'(x)$ in (c,β), with $\alpha \leq c < \beta$ and $u'(c) = v'(c)$. Then, it follows from

$$\frac{u'(c) - u'(c+h)}{h} > \frac{v'(c) - v'(c+h)}{h}$$

that $u''(c) \geq v''(c)$.

$u(c) > v(c) > 0$ implies that $-\frac{\lambda}{\sigma^2\beta}v(c) < -\frac{\lambda}{\sigma^2 x}u(x)|_{x=c}$, and $u'(c) = v'(c) < 0$ implies that $\frac{a-b\beta}{\sigma^2\beta}v'(c) \leq \frac{a-bx}{\sigma^2 x}u'(x)|_{x=c}$, together with $u''(c) \geq v''(c)$, which are contradictory to (A5) and (A6). □

Let $T_\alpha = \inf\{t \geq 0 : X(t) = \alpha\}$. Let

$$L^{CIR} = \frac{1}{2}\sigma^2 x\frac{d^2}{dx^2} + (a - bx)\frac{d}{dx},$$

and

$$L^{\Gamma} = \frac{1}{2}\frac{d^2}{dx^2} + \Gamma\frac{d}{dx}.$$

Let $P_x^{CIR;Ref\leftarrow:\beta}$ and $E_x^{CIR;Ref\leftarrow:\beta}$ denote the probabilities and expectations, respectively, for the CIR model corresponding to L^{CIR} on $[1, \beta]$, starting from $x \in [1, \beta]$, with reflection at β and stopped at 1; let $P_x^{CIR;Ref\rightarrow:\alpha}$ and $E_x^{CIR;Ref\rightarrow:\alpha}$ denote the probabilities and expectations, respectively, for the CIR model corresponding to L^{CIR} on $[\alpha, \infty)$, starting from $x \in [\alpha, \infty]$, with reflection at α. We will sometimes work for L^{Γ} with only a constant drift, which we will denote by Γ, in which case Γ will replace the CIR in all of the above notions.

The following lemma comes from the Proposition 2.3 in [10]; for the convenience of readers, we will now provide a proof of this lemma.

Lemma A4. *For $\lambda > 0$ and $1 < \alpha < \beta$,*

$$E_\beta^{\Gamma;Ref\leftarrow:\beta}\exp(-\lambda T_\alpha) = \frac{(r_1 + r_2)e^{-2\Gamma(\beta-\alpha)}}{r_1 e^{-r_1(\beta-\alpha)} + r_2 e^{r_2(\beta-\alpha)}},$$

where $r_1 = \sqrt{\Gamma^2 + 2\lambda} + \Gamma$ and $r_2 = \sqrt{\Gamma^2 + 2\lambda} - \Gamma$.

Proof. According to the Feynman–Kac formula, for any $x \in [\alpha, \beta]$,

$$w(x) = E_x^{\Gamma;Ref\leftarrow:\beta}\exp(-\lambda T_\alpha),$$

solves the boundary value problem $(L^{\Gamma} - \lambda)w = 0$ in (α, β), with the Dirichlet boundary condition at α and the Neumann boundary condition at β, that is,

$$\begin{cases} (L_\Gamma - \lambda)w = 0, & \text{in } (\alpha, \beta) \\ w(\alpha) = 1, \\ w'(\beta) = 0, \end{cases} \tag{A7}$$

The solution of this linear equation is given by

$$w(x) = \frac{r_1 e^{-r_1(\beta-\alpha)}e^{r_2(x-\alpha)} + r_2 e^{r_2(\beta-\alpha)}e^{-r_1(x-\alpha)}}{r_1 e^{-r_1(\beta-\alpha)} + r_2 e^{r_2(\beta-\alpha)}},$$

where $r_1 = \sqrt{\Gamma^2 + 2\lambda} + \Gamma$ and $r_2 = \sqrt{\Gamma^2 + 2\lambda} - \Gamma$.

Substituting $x = \beta$ completes the proof. □

Lemma A5. *For $\lambda > 0$ and $1 < \alpha < \beta$,*

$$E_\beta^{\frac{a-b\beta}{\sigma^2\beta};Ref\leftarrow:\beta}\exp\Big(-\frac{\lambda}{\sigma^2\beta}T_\alpha\Big) = \frac{(r_1 + r_2)e^{-2\frac{a-b\beta}{\sigma^2\beta}(\beta-\alpha)}}{r_1 e^{-r_1(\beta-\alpha)} + r_2 e^{r_2(\beta-\alpha)}},$$

where

$$r_1 = \sqrt{\Big(\frac{a-b\beta}{\sigma^2\beta}\Big)^2 + \frac{2\lambda}{\sigma^2\beta}} + \frac{a-b\beta}{\sigma^2\beta}$$

and

$$r_2 = \sqrt{\left(\frac{a-b\beta}{\sigma^2\beta}\right)^2 + \frac{2\lambda}{\sigma^2\beta}} - \frac{a-b\beta}{\sigma^2\beta}.$$

Proof. This lemma can be directly obtained from Lemma A4 by simply replacing λ with $\frac{\lambda}{\sigma^2\beta}$ and replacing Γ with $\frac{a-b\beta}{\sigma^2\beta}$. □

Appendix B

Lemma A6. *For $x \in [1, \beta]$ and $\lambda = -\frac{2ab}{\sigma^2} > 0$,*

$$u(x) = e^{-\frac{2b}{\sigma^2}(\beta - x)} \exp\left(- \int_x^\beta \frac{\exp(-\frac{2b(y-1)}{\sigma^2}) y^{-\frac{2a}{\sigma^2}}}{\int_1^y \exp(-\frac{2b(t-1)}{\sigma^2}) t^{-\frac{2a}{\sigma^2}} dt - \frac{\sigma^2}{2b}} dy \right) \tag{A8}$$

solves the following equation

$$\begin{cases} (L_{CIR} + \lambda)u = 0, \quad in\ (1, \beta) \\ u(\beta) = 1, \\ u'(1) = 0, \end{cases} \tag{A9}$$

that is,

$$\frac{1}{2}\sigma^2 x u'' + (a - bx)u' + \lambda u = 0$$

with the Dirichlet boundary condition at β and $u(\beta) = 1$ and the Neumann boundary condition at 1 and $u'(1) = 0$.

Proof. For the eigenvalue $\lambda = -\frac{2ab}{\sigma^2} > 0$ (due to $b < 0$), obviously, $u(x) = e^{\frac{2b}{\sigma^2}x}$ is a eigenfunction of Equation (A9).

Using the transformation

$$r(x) = \frac{u'(x)}{u(x)}, \text{ i.e., } u(x) = \exp\Big(\int_1^x r(t)dt \Big),$$

the linear differential equation of the second order

$$\frac{1}{2}\sigma^2 x u'' + (a - bx)u' + \lambda u = 0,$$

i.e.,

$$u'' + \frac{a - bx}{\frac{1}{2}\sigma^2 x} u' + \frac{\lambda}{\frac{1}{2}\sigma^2 x} u = 0$$

can be transformed into the Riccati differential equation

$$r' + r^2 + \frac{a - bx}{\frac{1}{2}\sigma^2 x} r + \frac{\lambda}{\frac{1}{2}\sigma^2 x} = 0,$$

where $\lambda = -\frac{2ab}{\sigma^2} > 0$. Obviously, $r_0 = \frac{2b}{\sigma^2}$ is a solution of the Riccati equation. If a solution r_0 of the Riccati equation is known, then all of the other solutions can be obtained in the form

$$r(x) = r_0 + \frac{1}{z(x)},$$

where $z(x)$ ia an arbitrary solution of the following linear equation

$$z' - \Big[\frac{a - bx}{\frac{1}{2}\sigma^2 x} + 2r_0 \Big] z = 1.$$

Since $u'(1) = 0$, we have $r(1) = \frac{u'(1)}{u(1)} = 0$; hence,

$$z(1) = -\frac{1}{r_0} = -\frac{\sigma^2}{2b}.$$

Next, we will solve the following Bernoulli's equation

$$z' - \Big[\frac{2(a-bx)}{\sigma^2 x} + \frac{4b}{\sigma^2}\Big]z = 1,$$

with the Dirichlet boundary condition at 1, i.e., $z(1) = -\frac{\sigma^2}{2b}$. The general solution to the homogeneous equation is

$$z_0(x) = C\exp\Big(\frac{2b(x-1)}{\sigma^2}\Big)x^{\frac{2a}{\sigma^2}},$$

from which a particular solution z_1 of the nonhomogeneous equation can be obtained

$$z_1(x) = \exp(\frac{2b(x-1)}{\sigma^2})x^{\frac{2a}{\sigma^2}}\int_1^x \exp(-\frac{2b(y-1)}{\sigma^2})y^{-\frac{2a}{\sigma^2}}dy.$$

Thus, $z(1) = -\frac{\sigma^2}{2b}$, $C = -\frac{\sigma^2}{2b}$ can be immediately obtained. The general solution of the Bernoulli's equation is

$$z(x) = z_0(x) + z_1(x)$$

with $C = -\frac{\sigma^2}{2b}$.

Hence, the general solution of the original Riccati equation is now obtained in the form

$$\begin{aligned} r(x) &= r_0 + \frac{1}{z(x)} = r_0 + \frac{1}{z_0(x) + z_1(x)} \\ &= \frac{2b}{\sigma^2} + \frac{\exp(-\frac{2b(x-1)}{\sigma^2})x^{-\frac{2a}{\sigma^2}}}{\int_1^x \exp(-\frac{2b(t-1)}{\sigma^2})t^{-\frac{2a}{\sigma^2}}dt - \frac{\sigma^2}{2b}}. \end{aligned} \tag{A10}$$

So, we obtain the solution of the linear differential equation of the second order (A9):

$$\begin{aligned} u(x) &= C\exp\Big(\int_1^x r(y)dy\Big) \\ &= Ce^{\frac{2b}{\sigma^2}(x-1)}\exp\left(\int_1^x \frac{\exp(-\frac{2b(y-1)}{\sigma^2})y^{-\frac{2a}{\sigma^2}}}{\int_1^y \exp(-\frac{2b(t-1)}{\sigma^2})t^{-\frac{2a}{\sigma^2}}dt - \frac{\sigma^2}{2b}}dy\right). \end{aligned}$$

Since $u(\beta) = 1$,

$$C = e^{-\frac{2b}{\sigma^2}(\beta-1)}\exp\left(-\int_1^\beta \frac{\exp(-\frac{2b(y-1)}{\sigma^2})y^{-\frac{2a}{\sigma^2}}}{\int_1^y \exp(-\frac{2b(t-1)}{\sigma^2})t^{-\frac{2a}{\sigma^2}}dt - \frac{\sigma^2}{2b}}dy\right)$$

can be obtained. Therefore, we obtain that

$$u(x) = e^{-\frac{2b}{\sigma^2}(\beta-x)}\exp\left(-\int_x^\beta \frac{\exp(-\frac{2b(y-1)}{\sigma^2})y^{-\frac{2a}{\sigma^2}}}{\int_1^y \exp(-\frac{2b(t-1)}{\sigma^2})t^{-\frac{2a}{\sigma^2}}dt - \frac{\sigma^2}{2b}}dy\right),$$

solves the Equation (A9). This completes the proof of the lemma. □

Lemma A7. *For $x \in [1, \beta]$ and $\lambda_j = -\frac{2ab}{\sigma^2 j} > 0$, then for all $j \geq 1$,*

$$u(x) = e^{-\frac{2b}{\sigma^2 j}(\beta - x)} \exp\left(- \int_x^\beta \frac{\exp(-\frac{2b(y-1)}{\sigma^2 j}) y^{-\frac{2a}{\sigma^2}}}{\int_1^y \exp(-\frac{2b(t-1)}{\sigma^2 j}) t^{-\frac{2a}{\sigma^2}} dt - \frac{\sigma^2 j}{2b}} dy \right) \tag{A11}$$

solves the following equation

$$\begin{cases} (L_{CIR;j} + \lambda_j)u = 0, \quad in\ (1, \beta) \\ u(\beta) = 1, \\ u'(1) = 0, \end{cases} \tag{A12}$$

that is

$$\frac{1}{2}\sigma^2 x u'' + (a - \frac{b}{j}x)u' + \lambda_j u = 0$$

with the Dirichlet boundary condition at β and $u(\beta) = 1$ and the Neumann boundary condition at 1 and $u'(1) = 0$.

Proof. This lemma can be directly obtained from Lemma A6. □

References

1. Cox, J.C.; Ingersoll, J.E., Jr.; Ross, S.A. An intertemporal general equilibrium model of asset prices. *Econom. J. Econom. Soc.* **1985**, *53*, 363–384. [CrossRef]
2. Vasicek, O. An equilibrium characterization of the term structure. *J. Financ. Econ.* **1977**, *5*, 177–188. [CrossRef]
3. Heston, S.L. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Rev. Financ. Stud.* **1993**, *6*, 327–343. [CrossRef]
4. Pitman, J.; Yor, M. A decomposition of Bessel bridges. *Z. Flzr Wahrscheinlichkeitstheorie Verwandte Geb.* **1982**, *59*, 425–457. [CrossRef]
5. Duffie, D.; Filipović, D.; Schachermayer, W. Affine processes and applications in finance. *Ann. Appl. Probab.* **2003**, *13*, 984–1053. [CrossRef]
6. Keller-Ressel, M.; Schachermayer, W.; Teichmann, J. Affine processes are regular. *Probab. Theory Relat. Fields* **2011**, *151*, 591–611. [CrossRef]
7. Burdzy, K.; Chen, Z.Q.; Sylvester, J. The heat equation and reflected Brownian motion in time-dependent domains: II. Singularities of solutions. *J. Funct. Anal.* **2003**, *204*, 1–34. [CrossRef]
8. Burdzy, K.; Chen, Z.Q.; Sylvester, J. The heat equation and reflected Brownian motion in time-dependent domains. *Ann. Probab.* **2004**, *32*, 775–804. [CrossRef]
9. Dembo, A.; Huang, R.; Sidoravicius, V. Walking within growing domains: Recurrence versus transience. *Elctronic J. Probab.* **2014**, *19*, 1–20. [CrossRef]
10. Pinsky, R.G. Transience/recurrence and growth rates for diffusion processes in time-dependent regions. *Electron. J. Probab.* **2016**, *21*, 46. [CrossRef]
11. Engländer, J.; Volkov, S. Conservative random walk. *Electron. J. Probab.* **2020**, *27*, 1–29. [CrossRef]
12. Hoffman, C.; Johnson, T.; Junge, M. From transience to recurrence with poisson tree frogs. *Mathematics* **2016**, *26*, 1620–1635. [CrossRef]
13. Pinsky, R.G. Recurrence, transience and bounded harmonic functions for diffusions in the plane. *Ann. Probab.* **1987**, *15*, 954–984. [CrossRef]
14. Pinsky, R.; Scheutzow, M. Some remarks and examples concerning the transience and recurrence of random diffusions. *Ann. Lihp Probab. Stat.* **1992**, *28*, 519–536.
15. Coculescu, D.; Nikeghbali, Hazard Process. Martingales Hazard Process. *Math. Financ.* **2012**, *22*, 519–537. [CrossRef]
16. Elliott, R.J.; Jeanblanc, M.; Yor, M. On models of default risk. *Math. Finance* **2020**, *10*, 179–196. [CrossRef]
17. Pinsky, R.G. P*ositive Harmonic Functions and Diffusion*; Cambridge University Press: Cambridge, UK, 1995; Volume 45.
18. Durrett, R. *Probability: Theory and Examples*; Cambridge University Press: Cambridge, UK, 2019; Volume 49.
19. Ross, S.M. *Introduction to Probability Models*; Academic Press: Cambridge, MA, USA, 2014.
20. Rogers, L.C.; Williams, D. *Diffusions, Markov Processes, and Martingales: Volume 1, Foundations*; Cambridge University Press: Cambridge, UK, 2000; Volume 1.
21. Rogers, L.C.; Williams, D. *Diffusions, Markov Processes and Martingales: Volume 2, Itô Calculus*; Cambridge University Press: Cambridge, UK, 2000; Volume 2.
22. Jeanblanc, M.; Yor, M.; Chesney, M. *Mathematical Methods for Financial Markets*; Springer: Berlin/Heidelberg, Germany, 2009.

mathematics

MDPI

Article

Robust and Sparse Portfolio: Optimization Models and Algorithms

Hongxin Zhao [1,*], Yilun Jiang [2] and Yizhou Yang [3]

[1] School of Mathematics and Statistics, Beijing Jiaotong University, Beijing 100044, China
[2] Department of Economic Management, Shijiazhuang Institute of Railway Technology, Shijiazhuang 050000, China; yljiangstat@163.com
[3] Personnel Department, Shijiazhuang University, Shijiazhuang 050035, China; joeyyangnz@gmail.com
* Correspondence: hxzhao@bjtu.edu.cn

Abstract: The robust and sparse portfolio selection problem is one of the most-popular and -frequently studied problems in the optimization and financial literature. By considering the uncertainty of the parameters, the goal is to construct a sparse portfolio with low volatility and decent returns, subject to other investment constraints. In this paper, we propose a new portfolio selection model, which considers the perturbation in the asset return matrix and the parameter uncertainty in the expected asset return. We define three types of stationary points of the penalty problem: the Karush–Kuhn–Tucker point, the strong Karush–Kuhn–Tucker point, and the partial minimizer. We analyze the relationship between these stationary points and the local/global minimizer of the penalty model under mild conditions. We design a penalty alternating-direction method to obtain the solutions. Compared with several existing portfolio models on seven real-world datasets, extensive numerical experiments demonstrate the robustness and effectiveness of our model in generating lower volatility.

Keywords: portfolio optimization; robustness; sparsity; uncertainty set; penalty-alternating-direction method

MSC: 91G10; 90C90; 90C30

Citation: Zhao, H.; Jiang, Y.; Yang, Y. Robust and Sparse Portfolio: Optimization Models and Algorithms. *Mathematics* **2023**, *11*, 4925. https://doi.org/10.3390/math11244925

Academic Editors: Jing Yao, Xiang Hu and Jingchao Li

Received: 8 November 2023
Revised: 7 December 2023
Accepted: 8 December 2023
Published: 11 December 2023

1. Introduction

In 1952, Harry M. Markowitz [1] published the classic "Portfolio Selection" in *The Journal of Finance*, which ushered in a new era of financial mathematical analysis. Markowitz pointed out that investors who care about return and risk should hold portfolios located at the efficient boundary of mean-variance, which is the famous mean-variance portfolio (MVP) selection model. Since then, many portfolio selection strategies have been proposed by referring to the MVP and its variants. However, MVPs exhibit instability due to estimation errors in the input parameters [2], especially in large-scale conditions. The instability means that the solution obtained under sample fluctuation may be optimal for a given sample, but it is not optimal from the perspective of risk. For more comments on this model, we refer to [3–6] and the references therein.

This paper focuses attention on sample fluctuations and parameter uncertainty in the portfolio selection problem. We now review some relevant methods for the parameter uncertainty. Among various approaches, the attractive one is the robust portfolio (RP), which corresponds to a robust optimization, since it does not use any information about the probability distribution of the uncertain parameters. RP we considered is a conservative approach that minimizes the loss function within an uncertainty set and then solves the problem under the worst-case scenario. In the last two decades, robust portfolio selection problems have gained the increasing interest of researches. These researches constructed well-known optimal portfolios from the perspective of robust optimization [7–10]. In this way, Goldfarb and Iyengar [11] formulated and solved RP problems. They introduced the

uncertainty structures for the input parameters, then they showed that the RP problems corresponding to the second-order cone programs and these uncertainty structures correspond to confidence regions employed to estimate the market parameters. Given the uncertainty in the mean and covariance matrix of the asset return, Lobo and Boyd [12] computed the maximum risk of a portfolio in a numerically efficient way. They proved that this is a semi-definite programming problem and is readily solved by interior-point methods for convex optimization. Min et al. [13] proposed the hybrid RP models under ellipsoidal uncertainty sets, and they considered both the best-case and the worst-case counterparts. Won and Kim [14] considered RP problems involving a trade-off between the worst-case utility and the worst-case regret, or the largest difference between the best utility achievable under the model and that achieved by a given portfolio. They showed that the entire optimal trade-off curve can be found via solving a series of semi-definite programs under the ellipsoidal uncertainty model. Some research works [15,16] concentrated on the application of robust optimization on basic mean-variance, mean value-at-risk (mean-VaR), and mean conditional-value-at-risk (mean-CVaR) problems, but did not consider variants of the problem like robust index tracking, robust and sparse portfolio selection problems, and so on. More relevant works can be found in [17–20] and the references therein.

RPs have a wide range of applications, among these, one essential step is the construction of uncertainty sets. Two types of uncertainty sets are widely used, namely the box uncertainty set and the ellipsoidal uncertainty set. Tütüncü and Koenig [21] used symmetric box uncertainty sets defined as $U_\mu = \{\mu \in \mathbb{R}^n | \mu^L \leq \mu \leq \mu^U\}$ and $U_\Sigma = \{\Sigma \in \mathbb{R}^{n\times n} | \Sigma^L \leq \Sigma \leq \Sigma^U, \Sigma \succeq 0\}$, where $\mu^L \in \mathbb{R}^n$ and $\mu^U \in \mathbb{R}^n$ are the lower and upper bounds of mean vector μ, $\Sigma^L \in \mathbb{R}^{n\times n}$ and $\Sigma^U \in \mathbb{R}^{n\times n}$ are the lower and the upper bounds of the covariance matrix Σ, respectively, and Σ is positive semi-definite. Khodamoradi et al. [22] used box uncertainty sets for a cardinal-constrained mean-variance portfolio problem which allows short selling. Swain and Ojha [10] analyzed the robust version of the mean-variance portfolio problem and mean-semi-variance portfolio problem with box uncertainty sets. Alternatively, Fabozzi et al. [23] defined an ellipsoidal uncertainty set for the expected asset return as $U_\mu = \{\mu | (\mu - \overline{\mu})\Sigma^{-1}(\mu - \overline{\mu})^\top \leq \epsilon^2\}$, where $\overline{\mu}$ is the nominal asset return and ϵ^2 is a small scalar, which controls the size of the uncertainty set. However, they did not consider the uncertainty of the covariance matrix, thus the solution was robust only against perturbations in the asset return vector. Pınar [24] developed a multi-period robust mean-variance portfolio problem with an ellipsoidal uncertainty set while allowing short selling. As we all know, the estimation error is more sensitive to the mean vector than the covariance matrix. On the other hand, dealing with the uncertainty in the covariance matrix is more complicated than dealing with the uncertainty set of the mean vector. Thus, in this paper, we consider two types of uncertainty sets for the mean vector.

Financial data have some remarkable features, such as multicollinearity and a heavy tail. Therefore, the perturbations of these data should not be underestimated. By referring to Brodie et al. [2], who transferred the MVP into a Lasso-type portfolio, we consider the perturbations in the asset return matrix and design its uncertainty set. In addition, from the perspective of transaction costs and administrative expenses, more assets are not always better. Therefore, it is also necessary to consider sparsity when constructing a portfolio [25–27]. After these discussions, a natural question follows: How do we find better RPs that not only reduce the undesired impact of parameter uncertainty, but also improve sparsity and reduce cost?

Following the above considerations, this paper proposes a sparsity constrained robust portfolio optimization model with parameter uncertainty and data perturbation. Specifically, we consider the perturbation in the asset return matrix and the parameter uncertainty in the expected asset return. By using the equivalence of robustness and regularization, the Lasso-type objective function can be converted into the sum of a square root and the ℓ_1 norm. We consider two kinds of uncertainty sets: the box uncertainty set and the ellipsoidal uncertainty set. For its penalty model, we define three types of stationary points:

the Karush–Kuhn–Tucker (KKT) point, the strong KKT point, and the partial minimizer. Under mild constraint qualification (CQ), we prove that any local minimizer of the penalty model is a KKT point. Moreover, the global minimizer of the penalty model is proven to be a partial minimizer and, then, a stronger KKT point under Slater's CQ. Finally, a penalty alternating direction method is proposed to obtain a portfolio, and its convergence is established. We confirm the effectiveness of our approach by comparing with nine widely studied portfolio models on seven real-world data sets. The numerical results show that the portfolios we proposed have less volatility, that is less risk. Moreover, our portfolio strategies can yield higher Sharpe ratios when the appropriate parameters are selected.

This paper is organized as follows. Some notations and preliminaries used in this paper are given in the next section. The model of robust and sparse portfolios and the analysis of their optimization theory are stated in Section 3. Two types of uncertainty sets of mean vectors are presented in Section 4. The optimization algorithm named the penalty alternating direction method is established in Section 5. Extensive numerical experiments are conducted in Section 6. Conclusions are drawn in Section 7.

2. Notations and Preliminary

We use $\mathbb{R}$ and $\mathbb{R}^n$ and $\mathbb{R}^{m\times n}$ to denote the set of real numbers and the n-dimensional and $m \times n$-dimensional Euclidean space. We use boldfaced small letters to denote vectors, e.g., $\mathbf{w} \in \mathbb{R}^n$ is a column vector with n elements w_i, $i = 1, \ldots, n$. The transpose of $\mathbf{w}$ is denoted as $\mathbf{w}^T$, which is a row vector. In particular, $\mathbf{1}_n$ is the vector of all ones of size n. For a vector $\mathbf{a} \in \mathbb{R}^n$, we define its absolute value vector by $|\mathbf{a}| := (|a_1|, \cdots, |a_n|)$. We use capital letters to denote matrices, e.g., $A \in \mathbb{R}^{m\times n}$ and a_{ij} denote the (i, j)-th entry of A. Given an index $\Gamma \subset \{1, \ldots, n\}$, $\mathbf{a}_\Gamma$ denotes the sub-vector of $\mathbf{a}$. We write the Euclidean norm of $\mathbf{w}$ by $\|\mathbf{w}\|_2$, the ℓ_1 norm by $\|\mathbf{w}\|_1$, and the infinity norm by $\|\mathbf{w}\|_\infty$. For two vectors $\mathbf{a} \in \mathbb{R}^n$ and $\mathbf{b} \in \mathbb{R}^n$, $\langle \mathbf{a}, \mathbf{b} \rangle$ denotes the standard inner product.

We now provide some existing results of optimization that are crucial for the theory of this paper. For the convenience of expression, we define the following convex programming:

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^n} \quad & f(\mathbf{x}), \\ \text{s.t.} \quad & g_i(\mathbf{x}) \geq 0,\ i = 1, \ldots, m,\ \mathbf{x} \in \Omega, \end{aligned} \tag{1}$$

where Ω is a nonempty convex set, f is a convex function, and the $g_i(\mathbf{x})$s are concave functions. For problem (1), Slater's CQ builds a bridge between its solution and the KKT point (the point satisfying the conditions in Theorem 1).

Definition 1 ([28], Definition 4.17)**.** *Slater's CQ holds in problem (1) if there exists* $\mathbf{u} \in \Omega$ *such that* $g_i(u) > 0$ *for all* $i = 1, \ldots, m$.

Theorem 1 ([28], Theorem 4.18)**.** *Suppose that Slater's CQ holds in problem (1). Then,* $\mathbf{x}^*$ *is an optimal solution to problem (1) if and only if there exist non-negative Lagrange multipliers* $(\lambda_1, \ldots, \lambda_m) \in \mathbb{R}^m$ *such that*

$$0 \in \partial f(\mathbf{x}^*) - \sum_{i=1}^{m} \lambda_i \partial g_i(\mathbf{x}^*) + N(\mathbf{x}^*; \Omega)$$

and $\lambda_i \partial g_i(\mathbf{x}^*) = 0$ *for all* $i = 1, \ldots, m$*, where* $\partial f(\mathbf{x}^*)$ *denotes the classical sub-differential set ([28], Definition 2.30) of* f *at* $\mathbf{x}^*$ *and* $N(\mathbf{x}^*; \#)$ *denotes the classical normal cone ([28], Definition 2.9) of* $\#$ *at* $\mathbf{x}^*$.

We also introduce some crucial terminologies and results for sparsity nonlinear programming:

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^n} \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & g_i(\mathbf{x}) \geq 0,\ i = 1,\ldots,m, \\ & h_j(\mathbf{x}) = 0,\ j = 1,\ldots,l, \\ & \|\mathbf{x}\|_0 \leq s, \end{aligned} \tag{2}$$

where f is a convex function and g and h are continuously differentiable. A restricted linear independence constraint qualification (R-LICQ) used for sparsity nonlinear programming (2) was defined by [29] as follows.

Definition 2 ([29], Definition 2.4)**.** *We say that the R-LICQ holds at* $\mathbf{x}^*$*, where* $\mathbf{x}^*$ *is feasible for the problem (2):*

- *When* $\|\mathbf{x}^*\|_0 = s$, $\nabla g_i(\mathbf{x}^*)$, $i \in I(\mathbf{x}^*)$, $\nabla h_j(\mathbf{x}^*)$, $j = 1,\ldots,l$, *are linearly independent.*
- *When* $\|\mathbf{x}^*\|_0 < s$, $\nabla_{\Gamma^*} g_i(\mathbf{x}^*)$, $i \in I(\mathbf{x}^*)$, $\nabla_{\Gamma^*} h_j(\mathbf{x}^*)$, $j = 1,\ldots,l$, *are linearly independent.*

Based on the R-LICQ, the following decomposition result holds.

Theorem 2 ([29], Proposition 2.5)**.** *Let* $\mathbf{x}^*$ *be a feasible point of problem (2) and the R-LICQ hold at* $\mathbf{x}^*$*. Then,*

$$\hat{N}(\mathbf{x}^*; S \cap Q) = \hat{N}(\mathbf{x}^*; S) + \hat{N}(\mathbf{x}^*; Q),$$

where $S := \{\boldsymbol{x} : \|\boldsymbol{x}\|_0 \leq s\}$, $Q := \{\boldsymbol{x} : g_i(\mathbf{x}) \geq 0,\ i = 1,\ldots,m,\ h_j(\boldsymbol{x}) = 0,\ j = 1,\ldots,l\}$, *and* $\hat{N}(\mathbf{x}^*; \#)$ *denotes the Frechét normal cone ([30], Definition 6.3) of* $\#$ *at* $\mathbf{x}^*$*, which degenerates into the classical norm cone described in Theorem 1 if* $\#$ *is a convex set.*

For the partial problem (10) of the portfolio model (6) in Section 3.1, the R-LICQ holds automatically at $\mathbf{x}^*$, where $m = 0$, $l = 1$, and $h(\boldsymbol{x}) := \mathbf{1}^T\mathbf{x} - 1$. Next, we establish the relationship between the local minimizer of problem (2) and its KKT point (the point satisfying the KKT system in Theorem 3).

Theorem 3. *Suppose that* $\mathbf{x}^*$ *is a local minimizer of problem (2) and the R-LICQ holds at* $\mathbf{x}^*$*. Then, there exist non-negative Lagrange multipliers* $(\lambda_1^*,\ldots,\lambda_m^*) \in \mathbb{R}_+^m$ *and* $(\mu_1^*,\ldots,\mu_m^*) \in \mathbb{R}^l$ *such that*

$$\begin{cases} 0 \in \partial f(\mathbf{x}^*) - \sum_{i=1}^m \lambda_i \partial g_i(\mathbf{x}^*) + \sum_{j=1}^l \lambda_i \partial h_j(\mathbf{x}^*) + \hat{N}(\mathbf{x}^*; S), \\ g_i(\mathbf{x}) \geq 0,\ \lambda_i g_i(\boldsymbol{x}) = 0,\ i = 1,\ldots,m, \\ h_j(\mathbf{x}) = 0,\ j = 1,\ldots,l, \\ \|\mathbf{x}\|_0 \leq s. \end{cases} \tag{3}$$

Proof. It follows from Theorem 6.12 of [30] that

$$0 \in \partial f(\mathbf{x}^*) + \hat{N}(\mathbf{x}^*; S \cap Q).$$

Combing Theorem 2 with the proof of Theorem 3.2 of [29], this result holds. □

This result is different from Theorem [29]. We allow the objective function of problem (2) to be non-differentiable. The analysis process of this result is completely consistent with that of Theorem [29].

3. Model and Optimization Theory

In this section, we first propose a robust and sparse portfolio model (4) with an uncertainty set constraint and a sparsity constraint. For the convenience of the numerical

calculation, we consider its ℓ_1 norm penalization variant (6). We define three types of stationary points of the penalization variant: the KKT point, the strong KKT point, and the partial minimizer. The relationships of these stationary points and the local/global minimizer of the penalization problem (6) are established in Section 3.2.

3.1. Robust and Sparse Portfolio Model

Consider n risky assets, denoting the asset return at period t by $\mathbf{r}_t = (r_1, ..., r_n)^\top \in \mathbb{R}^n$. The expected return vector of different assets is denoted by $E(\mathbf{r}_t) = \mu$, and the covariance matrix is denoted by $E[(\mathbf{r}_t - \mu)(\mathbf{r}_t - \mu)^\top] = V$. In the traditional Markowitz portfolio selection problem, the portfolio construction is based on the trade-off between risk and return. For a given level of acceptable portfolio return $\rho = \mathbf{w}^\top \mu$, the mean-variance optimization can be formulated as

$$\min_{\mathbf{w}\in\mathbb{R}^n} \frac{1}{2}\mathbf{w}^\top V\mathbf{w}, \quad \text{s.t.} \quad \mathbf{w}^\top\mu = \rho, \quad \mathbf{w}^\top\mathbf{1}_n = 1,$$

and its aim is to find a portfolio that has minimal risk for a given expected return. A significant model that has been developed from the Markowitz model is the Lasso-type portfolio proposed by Brodie et al. [2], which is given as:

$$\min_{\mathbf{w}\in\mathbb{R}^n} \frac{1}{T}\|\rho\mathbf{1}_T - R\mathbf{w}\|_2^2 + \alpha\|\mathbf{w}\|_1, \quad \text{s.t.} \quad \mathbf{w}^\top\overline{\mu} = \rho, \quad \mathbf{w}^\top\mathbf{1}_n = 1,$$

where $\overline{\mu} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{r}_t$, α is the penalty parameter, and $R \in \mathbb{R}^{T\times n}$ is the asset return matrix. Brodie et al. [2] confirmed that the ℓ_1 norm can produce a sparse portfolio, and this method can stabilize the problem. In this paper, we start with the square root Lasso-type portfolio, while adding more consideration about the perturbation in asset return matrix R and the parameter uncertainty in μ. We propose the following robust and sparse portfolio selection model:

$$\begin{aligned} \min_{\mathbf{w}\in\mathbb{R}^n}\max_{\Delta\in U_0} \quad & \|\rho\mathbf{1}_T - (R+\Delta)\mathbf{w}\|_2 \\ \text{s.t.} \quad & \min_{\mu\in U} \mathbf{w}^T\mu \geq \rho, \\ & \mathbf{w}^T\mathbf{1}_n = 1,\ \mathbf{w} \geq 0, \\ & \|\mathbf{w}\|_0 \leq s, \end{aligned} \tag{4}$$

where Δ is the data perturbation matrix and $U_0 = \{\Delta \in \mathbb{R}^{T\times n} : \|\Delta_i\|_2 \leq \alpha, \forall i \in \{1, ..., n\}\}$. The uncertainty set of the asset return is denoted by U, and we will discuss two selections of U in the last section.

In [31] (Chapter 2), they showed the equivalence of robustness and regularization. Specifically, they precisely characterized the conditions on the model of uncertainty and loss function under which robustness is equivalent to regularization for linear regression.

Definition 3. *Let $g : \mathbb{R}^T \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}$ be the norm, then the induced norm $\|\cdot\|_{(h,g)}$ is defined as*

$$\|\Delta\|_{(h,g)} = \max_{\mathbf{w}\in\mathbb{R}^n} \frac{g(\Delta\mathbf{w})}{h(\mathbf{w})}.$$

Theorem 4 ([31], Chapter 2)**.** *If $r, q \in [1, \infty]$, then*

$$\min_{\mathbf{w}}\max_{\Delta\in U_{(\ell_q,\ell_r)}} \|\mathbf{y} - (R+\Delta)\mathbf{w}\|_r = \min_{\mathbf{w}} \|\mathbf{y} - R\mathbf{w}\|_r + \alpha\|\mathbf{w}\|_q,$$

where $U_{(\ell_q,\ell_r)} = \{\Delta : \|\Delta\|_{(\ell_q,\ell_r)} \leq \alpha\}$. Moreover, if $U_0 = \{\Delta : \|\Delta_i\|_2 \leq \alpha, \forall i \in \{1, ..., n\}\}$, then $U_{(\ell_1,\ell_2)} = U_0$, and this implies

$$\min_{\mathbf{w}} \max_{\|\Delta_i\|_2 \le \alpha} \|\mathbf{y} - (R+\Delta)\mathbf{w}\|_2 = \min_{\mathbf{w}} \|\mathbf{y} - R\mathbf{w}\|_2 + \alpha\|\mathbf{w}\|_1.$$

From the relationship of the robustness and the regularization, problem (4) can be rewritten as

$$\begin{aligned} \min_{\mathbf{w}\in\mathbb{R}^n} \quad & \|\rho\mathbf{1}_T - R\mathbf{w}\|_2 + \alpha\|\mathbf{w}\|_1 \\ \text{s.t.} \quad & \min_{\mu\in U} \mathbf{w}^T\mu \ge \rho, \\ & \mathbf{w}^T\mathbf{1}_n = 1,\ \mathbf{w} \ge 0, \\ & \|\mathbf{w}\|_0 \le s. \end{aligned} \tag{5}$$

Under this transformation, the problem (5) actually enjoys robustness. We plan to use an alternating penalty method to solve problem (5). To ensure the implementation of the alternating penalty method, we add a copy constraint $\mathbf{w} = \mathbf{v}$ to the problem (5) and, then, move it to the objective function by means of the ℓ_1 norm penalty, then the penalization formulation is

$$\begin{aligned} \min_{\mathbf{w},\mathbf{v}\in\mathbb{R}^n} \quad & f(\mathbf{w},\mathbf{v}) := \|\rho\mathbf{1}_T - R\mathbf{w}\|_2 + \alpha\|\mathbf{w}\|_1 + \beta\|\mathbf{w}-\mathbf{v}\|_1 \\ \text{s.t.} \quad & \mathbf{w} \in \Omega_1 := \{\mathbf{w} | \min_{\mu\in U} \mathbf{w}^T\mu \ge \rho,\ \mathbf{w} \ge 0\} \\ & \mathbf{v} \in \Omega_2 := \{\mathbf{v} | \mathbf{v}^T\mathbf{1}_n = 1, \|\mathbf{v}\|_0 \le s\}. \end{aligned} \tag{6}$$

We conduct its optimality analysis in the next subsection.

3.2. Optimization Theory

We now analyze the optimality of the penalization problem (6). Obviously, the objective function of the problem (6) is a lower semi-continuous and coercive function. Theorem 5 in the next subsection provides the existence of optimal solutions. The selection of the uncertainty set U is discussed in Section 4.

This subsection provides a few theoretical results of the problem (6) including the existence of the solution and three classes of the first-order necessary optimal condition.

Theorem 5. *For any given $\alpha \in \mathbb{R}_+$ and $\beta \in \mathbb{R}_+$, the optimal solutions of the problem (6) can be attained.*

Proof. It is clear that f is a proper, closed, and coercive function and $\Omega_1 \times \Omega_2$ is a nonempty closed set satisfying $\Omega_1 \times \Omega_2 \cap \mathrm{dom}(f) \neq \emptyset$. It follows from Theorem 2.14 of [32] that this theorem holds. □

We now define a class of KKT points of the problem (6). For the convenience of expression and the generality of optimality, we write $\min_{\mu\in U} \mathbf{w}^T\mu \ge \rho$ as $g(\mathbf{w}) \ge 0$ and suppose that g is a concave function and is not necessarily differentiable. Indeed, the quadratic uncertainty set and the absolute uncertainty set introduced in Section 4 satisfy these terminologies.

Definition 4. *The point $(\mathbf{w}^*, \mathbf{v}^*) \in \Omega_1 \times \Omega_2$ is called a KKT point of the problem (6), if there exist Lagrange multipliers $\lambda_1^* \in \mathbb{R}_+$ and $\lambda_2^* \in \mathbb{R}$ such that the following system holds:*

$$\begin{cases} 0 \in \partial_{\mathbf{w}} f(\mathbf{w}^*, \mathbf{v}^*) - \lambda_1^* \partial_{\mathbf{w}} g(\mathbf{w}^*) + N(\mathbf{w}^*; \mathbb{R}_+), \\ 0 \in \partial_{\mathbf{v}} f(\mathbf{w}^*, \mathbf{v}^*) - \lambda_2^* \mathbf{1} + \hat{N}(\mathbf{v}^*; S), \\ g(\mathbf{w}^*) \ge 0,\ \lambda^* g(\mathbf{w}^*) = 0, \\ \mathbf{1}^T\mathbf{v}^* = 1,\ \|\mathbf{v}^*\|_0 \le s. \end{cases} \tag{7}$$

Although the functions corresponding to the quadratic uncertainty set and the absolute uncertainty set introduced in Section 4 are all concave and may both be non-differentiable and Slater's CQ automatically holds for both functions, we still considered Slater's CQ as a condition of Theorem 6 for the sake of generality. Moreover, it is stated in Section 2 that the R-LICQ of Ω_2 holds at every point. Then, only under the condition that Slater's CQ holds, the relationship between the local minimizer of the problem (6) and the KKT point of the problem (6) can be established.

Theorem 6. *Let $(\mathbf{w}^*, \mathbf{v}^*) \in \Omega_1 \times \Omega_2$ be a local minimizer of the problem (6). If Slater's CQ holds on Ω_1, then it is a KKT point of the problem (6).*

Proof. On the one hand, since $(\mathbf{w}^*, \mathbf{v}^*) \in \Omega_1 \times \Omega_2$ is a local minimizer of the problem (6), $\mathbf{w}^*$ is a local minimizer of the following optimization:

$$\begin{aligned} \min_{\mathbf{w}\in\mathbb{R}^n} \quad & f(\mathbf{w}, \mathbf{v}^*) = \|\rho\mathbf{1}_T - R\mathbf{w}\|_2 + \alpha\|\mathbf{w}\|_1 + \beta\|\mathbf{w} - \mathbf{v}^*\|_1 \\ \text{s.t.} \quad & \mathbf{w} \in \Omega_1. \end{aligned} \tag{8}$$

Notice that $f(\mathbf{w}, \mathbf{v}^*)$ is a convex function about $\mathbf{w}$ and Ω_1 is a convex set. Then, problem (8) is a convex optimization. Since Slater's CQ holds on Ω_1, it follows from Theorem 1 that there exists a Lagrange multiplier $\lambda_1^* \in \mathbb{R}_+$ such that

$$\begin{cases} 0 \in \partial_{\mathbf{w}} f(\mathbf{w}^*, \mathbf{v}^*) + \lambda_1^* \partial_{\mathbf{w}} g(\mathbf{w}^*) + N(\mathbf{w}^*; \mathbb{R}_+), \\ g(\mathbf{w}^*) \geq 0,\ \lambda_1^* g(\mathbf{w}^*) = 0. \end{cases} \tag{9}$$

On the other hand, $\mathbf{v}^*$ is a local minimizer of the following optimization:

$$\begin{aligned} \min_{\mathbf{v}\in\mathbb{R}^n} \quad & f(\mathbf{w}^*, \mathbf{v}) = \|\rho\mathbf{1}_T - R\mathbf{w}^*\|_2 + \alpha\|\mathbf{w}^*\|_1 + \beta\|\mathbf{w}^* - \mathbf{v}\|_1 \\ \text{s.t.} \quad & \mathbf{v} \in \Omega_2. \end{aligned} \tag{10}$$

Since the R-LICQ of Ω_2 holds at every point, it follows from Theorem 3 that there exists a Lagrange multiplier $\lambda_2^* \in \mathbb{R}$ such that

$$\begin{cases} 0 \in \partial_{\mathbf{v}} f(\mathbf{w}^*, \mathbf{v}^*) + \lambda_2^* \mathbf{1} + \hat{N}(\mathbf{v}^*; S), \\ \mathbf{1}^T\mathbf{v}^* = 1,\ \|\mathbf{v}^*\|_0 \leq s. \end{cases} \tag{11}$$

Combing the system (9) and (11), this theorem holds. □

Again, problem (10) can be simply written as

$$\begin{aligned} \min_{\mathbf{v}\in\mathbb{R}^n} \quad & \|\mathbf{w}^* - \mathbf{v}\|_1 \\ \text{s.t.} \quad & \mathbf{v} \in \Omega_2, \end{aligned}$$

and it has a closed-form solution; see [33], i.e.,

$$v_i^* = \begin{cases} \frac{w_i^*}{(\mathbf{w}_s^*)^T\mathbf{1}_s}, & \text{if } i \in I_s^* \\ 0, & \text{otherwise,} \end{cases} \tag{12}$$

where $I_s^* := \{i \mid w_{\langle 1\rangle}^* \geq \ldots \geq w_{\langle s\rangle}^*\}$ and $w_{\langle i\rangle}^*$ denotes the i-th largest absolute value among the n elements of $\mathbf{w}^*$. Thus, we can define a class of strong KKT points of the problem (6) as follows.

Definition 5. *The point $(\mathbf{w}^*, \mathbf{v}^*) \in \Omega_1 \times \Omega_2$ is called a strong KKT point of the problem (6), if there exists a Lagrange multiplier $\lambda^* \in \mathbb{R}_+$ such that the following system holds:*

$$\begin{cases} 0 \in \partial_{\mathbf{w}} f(\mathbf{w}^*, \mathbf{v}^*) + \partial_{\mathbf{w}} g(\mathbf{w}^*) + N(\mathbf{w}^*; \mathbb{R}_+), \\ g(\mathbf{w}^*) \geq 0,\ \lambda^* g(\mathbf{w}^*) = 0, \\ v_i^* = \begin{cases} \frac{w_i^*}{(\mathbf{w}_s^*)^T \mathbf{1}_s}, & \text{if } i \in I_s^* \\ 0, & \text{otherwise.} \end{cases} \end{cases} \tag{13}$$

It is easy to prove that, if $(\mathbf{w}^*, \mathbf{v}^*)$ is a strong KKT point of the problem (6), then it is a KKT point of the problem (6). The following result provides the relationship between the global minimizer of the problem (6) and the strong KKT point of the problem (6).

Theorem 7. *Let $(\mathbf{w}^*, \mathbf{v}^*) \in \Omega_1 \times \Omega_2$ be a global minimizer of the problem (6). If Slater's CQ holds on Ω_1 at w^*, then it is a strong KKT point of the problem (6).*

Proof. The part of $\mathbf{w}^*$ in (13) follows from (7). We only need to discuss the part of $\mathbf{v}^*$ in (13). Since $\mathbf{v}^*$ is the global minimizer of (10), it follows from (12) that the part of $\mathbf{v}^*$ in (13) holds. □

Note that the local minimizer of the problem (6) cannot be guaranteed to be a strong KKT point.

Finally, we introduce the third stationary point of the problem (6), which is called the partial minimizer.

Definition 6. *The point $(\mathbf{w}^*, \mathbf{v}^*) \in \Omega_1 \times \Omega_2$ is called a partial minimizer of the problem (6), if it satisfies*

$$f(\mathbf{w}^*, \mathbf{v}^*) \leq f(\mathbf{w}, \mathbf{v}^*),\ \forall\ \mathbf{w} \in \Omega_1,\ f(\mathbf{w}^*, \mathbf{v}^*) \leq f(\mathbf{w}^*, \mathbf{v}),\ \forall\ \mathbf{v} \in \Omega_2.$$

Clearly, any global minimizer of the problem (6) is a partial minimizer. Moreover, on the one hand, the partial problem (8) is a convex optimization, and Slater's CQ ensures that its KKT point and global minimizer are consistent. On the other hand, the partial problem (10) has a closed-form solution. Thus, the equivalence relationship between the KKT point of the problem (6) and the partial minimizer of the problem (6) can be established under Slater's CQ.

Theorem 8. *Let $(\mathbf{w}^*, \mathbf{v}^*) \in \Omega_1 \times \Omega_2$ be a feasible point of the problem (6). Suppose that Slater's CQ holds on Ω_1. Then, $(\mathbf{w}^*, \mathbf{v}^*)$ is a partial minimizer of the problem (6) if and only if $(\mathbf{w}^*, \mathbf{v}^*)$ is a strong KKT point of the problem (6).*

Proof. Suppose that $(\mathbf{w}^*, \mathbf{v}^*)$ is a strong KKT point of the problem (6), then

$$0 \in \partial_{\mathbf{w}} f(\mathbf{w}^*, \mathbf{v}^*) + \partial_{\mathbf{w}} g(\mathbf{w}^*) + N(\mathbf{w}^*; \mathbb{R}_+),\ g(\mathbf{w}^*) \geq 0, \text{ and } \lambda_1^* g(\mathbf{w}^*) = 0.$$

Since Slater's CQ holds at $\mathbf{w}^*$, $\mathbf{w}^*$ is a global minimizer of the problem (8). Then, we have that $f(\mathbf{w}^*, \mathbf{v}^*) \leq f(\mathbf{w}, \mathbf{v}^*)\ \forall\ \mathbf{w} \in \Omega_1$. Moreover, it follows from the definition of the strong KKT point of the problem (6) that $\mathbf{v}^*$ is a global minimizer of the problem (10). Then, we have that $f(\mathbf{w}^*, \mathbf{v}^*) \leq f(\mathbf{w}^*, \mathbf{v})$, $\forall \mathbf{v} \in \Omega_2$. Thus, $(\mathbf{w}^*, \mathbf{v}^*)$ is a partial minimum of the problem (6). The opposite conclusion clearly holds. □

4. The Uncertainty Set U

In Subsection 3.2, we rewrite the uncertainty set constraint as $g(\boldsymbol{w}) \geq 0$, where g is a generalized concave function and is not necessarily differentiable. This section introduces two mainstream formulations for the uncertainty set in asset mean return vector μ (see [34]), which corresponds to the quadratic uncertainty set and the absolute uncertainty set, respectively.

4.1. The Quadratic Uncertainty Set

The first one is the quadratic formulation, $U = \{\mu | (\mu - \overline{\mu})^T \Omega (\mu - \overline{\mu}) \leq \kappa^2\}$, where $\overline{\mu}$ is the nominal expected return and κ is the error. Assume that asset returns are independent and identically distributed and $\mu - \overline{\mu}$ follows a normal distribution with mean value $\mathbf{0}$ and covariance matrix Ω, where Ω is the covariance matrix of errors in the expected asset return. In Yin et al. [35], they discussed the choice of uncertainty matrix Ω in the quadratic uncertainty set and proposed the selection criteria. In the quadratic uncertainty case, $\min_{\mu \in U} \mathbf{w}^T \mu$ in problem (5) is equivalent to the following problem:

$$\max_{\mu \in U} \ \mathbf{w}^T \overline{\mu} - \mathbf{w}^T \mu.$$

Solving the above problem, we obtain

$$\mu = \overline{\mu} - \sqrt{\frac{\kappa^2}{\mathbf{w}^T \Omega \mathbf{w}}}\, \Omega \mathbf{w}.$$

Then, the problem (5) is rewritten as

$$\begin{aligned} \min_{\mathbf{w},\mathbf{v} \in \mathbb{R}^n} \quad & \|\rho \mathbf{1}_T - R\mathbf{w}\|_2 + \alpha \|\mathbf{w}\|_1 \\ \text{s.t.} \quad & \overline{\mu}^T \mathbf{w} - \kappa \sqrt{\mathbf{w}^T \Omega \mathbf{w}} \geq \rho, \\ & \mathbf{w}^T \mathbf{1}_n = 1, \ \mathbf{w} \geq 0, \\ & \|\mathbf{w}\|_0 \leq s. \end{aligned} \tag{14}$$

Here, $g(\mathbf{w}) = \overline{\mu}^T \mathbf{w} - \kappa \sqrt{\mathbf{w}^T \Omega \mathbf{w}} - \rho$. The penalization form of problem (14) can be rewritten as

$$\begin{aligned} \min_{\mathbf{w},\mathbf{v} \in \mathbb{R}^n} \quad & \|\rho \mathbf{1}_T - R\mathbf{w}\|_2 + \alpha \|\mathbf{w}\|_1 + \beta \|\mathbf{w} - \mathbf{v}\|_1 \\ \text{s.t.} \quad & \mathbf{w} \in \Omega_1 = \{\mathbf{w} \mid \kappa \|\sqrt{\Omega}\mathbf{w}\|_2 \leq \overline{\mu}^T \mathbf{w} - \rho, \ \mathbf{w} \geq 0\} \\ & \mathbf{v} \in \Omega_2 = \{\mathbf{v} \mid \mathbf{v}^T \mathbf{1}_n = 1, \ \|\mathbf{v}\|_0 \leq s\}. \end{aligned}$$

According to the proposition of Yin et al. [35], we choose $\Omega \propto diag(V)$. By using this uncertainty matrix, it is expected to reduce the sensitivity to the inputs, as well as keep the original volatility unchanged.

4.2. The Absolute Uncertainty Set

Fabozzi et al. [23] used the absolute uncertainty set in mean returns that ask that the sum of absolute spreads between estimated and possible mean returns should not be too large. The absolute formulation is $U = \{\mu | \sum_i |\mu_i - \overline{\mu}_i| \leq \frac{\kappa \overline{\sigma}}{\sqrt{T}}\}$. In this case,

$$\mu^T \mathbf{w} - \overline{\mu}^T \mathbf{w} \geq -\sum_i |\mu_i - \overline{\mu}_i| \max(|\mathbf{w}_i|) \geq -\frac{\kappa \overline{\sigma}}{\sqrt{T}} \max(|\mathbf{w}_i|);$$

thus,

$$\mu^T \mathbf{w} \geq \overline{\mu}^T \mathbf{w} - \frac{\kappa \overline{\sigma}}{\sqrt{T}} \max(|\mathbf{w}_i|).$$

Then, the problem (5) is equivalent to

$$\begin{aligned} \min_{\mathbf{w},\mathbf{v}\in\mathbb{R}^n} \quad & \|\rho\mathbf{1}_T - R\mathbf{w}\|_2 + \alpha\|\mathbf{w}\|_1 \\ \text{s.t.} \quad & \overline{\mu}^T\mathbf{w} - \frac{\kappa\overline{\sigma}}{\sqrt{T}}\max(|\mathbf{w}_i|) \geq \rho, \\ & \mathbf{w}^T\mathbf{1}_n = 1,\ \mathbf{w} \geq 0, \\ & \|\mathbf{w}\|_0 \leq s. \end{aligned} \tag{15}$$

Here, $g(\mathbf{w}) = \overline{\mu}^T\mathbf{w} - \kappa\overline{\sigma}\max(|\mathbf{w}_i|)/\sqrt{T} - \rho$. Similarly, the penalization from of problem (15) can be written as

$$\begin{aligned} \min_{\mathbf{w},\mathbf{v}\in\mathbb{R}^n} \quad & \|\rho\mathbf{1}_T - R\mathbf{w}\|_2 + \alpha\|\mathbf{w}\|_1 + \beta\|\mathbf{w}-\mathbf{v}\|_1 \\ \text{s.t.} \quad & \mathbf{w} \in \Omega_1 = \{\mathbf{w} \mid |w_i| \leq \frac{\sqrt{T}}{\kappa\overline{\sigma}}(\mathbf{w}^T\overline{\mu} - \rho),\ i = 1, ..., n,\ \mathbf{w} \geq 0\} \\ & \mathbf{v} \in \Omega_2 = \{\mathbf{v} \mid \mathbf{v}^T\mathbf{1}_n = 1,\ \|\mathbf{v}\|_0 \leq s\}. \end{aligned}$$

5. Optimization

This section introduces a penalty alternating direction method (PADM) to solve problem (5).

5.1. Alternating Direction Methods

We first discuss the optimization of the problem (6). Due to the complexity of this problem, alternating direction methods (ADMs) can be used to solve this problem. The framework of ADMs is described as follows:

Next, we state the general convergence result of Algorithm 1, and one can refer to Geissler et al. [36] for a proof (Theorem 8) and for further details about this method.

Algorithm 1 ADM: Alternating Direction Method.

1: Set the problem parameters: $\alpha, \kappa, \rho, T > 0$, asset return matrix $R \in \mathbb{R}^{T\times n}$, and nominal expected return vector $\overline{\mu} \in \mathbb{R}^n$. Initialize $\varepsilon > 0$, $(\mathbf{w}^0, \mathbf{v}^0)$ and penalty parameter $\beta > 0$. Set the iteration index $k := 0, 1, ...$

2: Compute

$$\mathbf{w}^{k+1} \in \arg\min_{\mathbf{w}} \{f(\mathbf{w}, \mathbf{v}^k) : \mathbf{w} \in \Omega_1\}, \tag{16}$$

and

$$\mathbf{v}^{k+1} \in \arg\min_{\mathbf{v}} \{f(\mathbf{w}^{k+1}, \mathbf{v}) : \mathbf{v} \in \Omega_2\}. \tag{17}$$

3: If $\frac{\|\mathbf{w}^{k+1}-\mathbf{w}^k\|_2+\|\mathbf{v}^{k+1}-\mathbf{v}^k\|_2}{\|\mathbf{w}^k\|_2+\|\mathbf{v}^k\|_2} \leq \varepsilon$, then stop with $(\mathbf{w}^k, \mathbf{v}^k)$ being an output point of (6).

Theorem 9. *Let $\{(\mathbf{w}^k, \mathbf{v}^k)\}$ be a sequence generated by Algorithm 1. Then, the following holds:*

(a) *$\{(\mathbf{w}^k, \mathbf{v}^k)\}$ is bounded.*

(b) *Any limiting point $\{(\mathbf{w}^*, \mathbf{v}^*)\}$ of $\{(\mathbf{w}^k, \mathbf{v}^k)\}$ is a partial minimizer of the problem (6).*

(c) *If Slater's CQ holds on Ω_1, the limiting point of $\{(\mathbf{w}^k, \mathbf{v}^k)\}$ is also a strong KKT of the problem (6).*

Proof. (a) It follows from Algorithm 1 that

$$f(\mathbf{w}^{k+1}, \mathbf{v}^{k+1}) \leq f(\mathbf{w}^{k+1}, \mathbf{v}^k) \leq f(\mathbf{w}^k, \mathbf{v}^k).$$

Since f is a coercive function, then the level set of f is bounded. Thus, $\{(\mathbf{w}^k, \mathbf{v}^k)\}$ is bounded.

(b) Clearly, $\{f(\mathbf{w}^k, \mathbf{v}^k)\}$ is a decreasing sequence and $f(\mathbf{w}^k, \mathbf{v}^k) \geq 0$, then there exists a value f^* such that $\lim_{k\to\infty} f(\mathbf{w}^k, \mathbf{v}^k) = f^*$. Suppose that $\{(\mathbf{w}^*, \mathbf{v}^*)\}$ is a limiting point of $\{(\mathbf{w}^k, \mathbf{v}^k)\}$. Then, there exists a sequence $\{k_j\}$ such that $\lim_{j\to\infty} k_j = \infty$, $\lim_{j\to\infty}(\mathbf{w}^{k_j}, \mathbf{v}^{k_j}) = (\mathbf{w}^*, \mathbf{v}^*)$ and $\lim_{j\to\infty} f(\mathbf{w}^{k_j}, \mathbf{v}^{k_j}) = f(\mathbf{w}^*, \mathbf{v}^*) = f^*$, where the last equality holds since $\lim_{k\to\infty} f(\mathbf{w}^k, \mathbf{v}^k) = f^*$. Without loss of generality, let $\lim_{k\to\infty}(\mathbf{w}^k, \mathbf{v}^k) = (\mathbf{w}^*, \mathbf{v}^*)$. It follows from Algorithm 1 that

$$\mathbf{w}^{k+1} \in \arg\min_{\mathbf{w}\in\Omega_1} f(\mathbf{w}, \mathbf{v}^k).$$

Since f is continuous with respect to $\mathbf{w}$, then taking $k \to \infty$, we have that

$$\mathbf{w}^* \in \arg\min_{\mathbf{w}\in\Omega_1} f(\mathbf{w}, \mathbf{v}^*).$$

Similarly,

$$\mathbf{v}^* \in \arg\min_{\mathbf{v}\in\Omega_2} f(\mathbf{w}^*, \mathbf{v}).$$

Thus, the limiting point $\{(\mathbf{w}^*, \mathbf{v}^*)\}$ of $\{(\mathbf{w}^k, \mathbf{v}^k)\}$ is a partial minimizer of the problem (6).

(c) Under Slater's CQ, the partial minimizer of the problem (6) is a strong KKT point of the problem (6) and the opposite also holds. Thus, this result holds. □

5.2. The Optimization for the Partial Problem (8)

We now discuss the optimization of the partial problem (8) at the k-th iteration of the ADM. Some non-exact penalty methods and smoothing methods can be used to solve this problem. Here, we obtain $\mathbf{w}^{k+1}$ by solving the following optimization:

$$\begin{aligned} \min_{\mathbf{w}\in\mathbb{R}^n} \quad & f(\mathbf{w}, \mathbf{v}^k) = \|\rho\mathbf{1}_T - R\mathbf{w}\|_2 + \alpha\|\mathbf{w}\|_1 + \beta\|\mathbf{w} - \mathbf{v}^k\|_1 + \gamma|g(\mathbf{w})_-| \\ \text{s.t.} \quad & \mathbf{w} \geq 0, \end{aligned} \tag{18}$$

where $\gamma > 0$ is a penalty parameter. Let

$$\psi_\mu(t) = \begin{cases} |t|, & |t| \geq \mu, \\ \frac{t^2}{2\mu} + \frac{\mu}{2}, & |t| < \mu, \end{cases} \quad \phi_\mu(t) = \frac{1}{2}(t + \sqrt{t^2 + \mu}),$$

where $\mu > 0$ is a smoothing parameter. Then, a class of the smoothing optimization of problem (18) can be given as follows:

$$\begin{aligned} \min_{\mathbf{w}\in\mathbb{R}^n} \quad & f_\mu(\mathbf{w}, \mathbf{v}^k) = \sqrt{\|\rho\mathbf{1}_T - R\mathbf{w}\|_2^2 + \mu} + \\ & \alpha\sum_{i=1}^{n}\psi_\mu(w_i) + \beta\sum_{i=1}^{n}\psi_\mu(w_i - v_i^k) + \gamma\phi_\mu(-g(\mathbf{w})) \\ \text{s.t.} \quad & \mathbf{w} \geq 0. \end{aligned} \tag{19}$$

The projection gradient method (PGM) can be used to solve the problem (19), and its iteration formula is

$$\mathbf{w}^{j+1,k} = P_{\mathbb{R}_+}(\mathbf{w}^{j,k} + \eta\nabla f_\mu(\mathbf{w}^{j,k}, \mathbf{v}^k)),$$

where $\eta > 0$ denotes the step length at the j-th iteration of the PGM at the k-th iteration of the ADM and $P_\#(t)$ denotes the projection point of t onto #. The framework of the above method is called the penalty projection gradient method (PPGM) and can be described in Algorithm 2.

Algorithm 2 PPGM: Penalty Projection Gradient Method.

1: Set the problem parameters: $\alpha, \kappa, \rho, T, \beta, \gamma_{\max} > 0$, asset return matrix $R \in \mathbb{R}^{T\times n}$, and nominal expected return vector $\overline{\mu} \in \mathbb{R}^n$. Initialize penalty parameters $\gamma^0 > 0$, $\tau > 1$. Set the iteration index $j = 0, 1, \cdots$.
2: Computing $\mathbf{w}^{j+1,k} = P_{\mathbb{R}_+}(\mathbf{w}^{j-1,k} + \eta\nabla f_\mu(\mathbf{w}, \mathbf{v}^k))$.
3: If $\mathbf{w}^{j+1,k}$ satisfies $\frac{\|\mathbf{w}^{j+1,k}-\mathbf{w}^{j,k}\|_2}{\|\mathbf{w}^{j,k}\|_2} \leq \varepsilon$ and $g(\mathbf{w}^{j,k}) \geq -\varepsilon$, then stop and $\mathbf{w}^k = \mathbf{w}^{j,k}$.
4: If $\frac{\|\mathbf{w}^{j+1,k}-\mathbf{w}^{j,k}\|_2}{\|\mathbf{w}^{j,k}\|_2} \leq \varepsilon$ and $g(w^{j,k}) < -\varepsilon$, then choose new penalty parameter $\gamma = \min\{\tau\gamma, \gamma_{\max}\}$. Otherwise, return to Step 2.

5.3. Penalty Alternating Direction Method

At the end of this section, we describe the PADM for the general problem (5). At iteration l, set the value of penalty parameter β^l and obtain $(\mathbf{w}^l, \mathbf{v}^l)$ by the ADM with β^l. If the inequality $\|\mathbf{w}^l - \mathbf{v}^l\|_1 \leq tol$ holds, where tol is a small positive constant, we stop with a feasible solution of problem (6). Otherwise, the penalty parameter β^l is updated to β^{l+1}. In this way, the PADM generates a sequence of the partial minimizer of problem (6) with β^l. The framework of the PADM is formally stated in Algorithm 3.

Algorithm 3 PADM: Penalty Alternating Direction Method.

1: Set the problem parameters: $\alpha, \kappa, \rho, T, \beta_{\max} > 0$, asset return matrix $R \in \mathbb{R}^{T\times n}$, and nominal expected return vector $\overline{\mu} \in \mathbb{R}^n$. Initialize penalty parameters $\beta^0 > 0$, $\tau > 1$. Set the iteration index $l = 0, 1, \cdots$.
2: Obtain $(\mathbf{w}^l, \mathbf{v}^l)$ by the ADM with β^l.
3: If $(\mathbf{w}^l, \mathbf{v}^l)$ satisfies $\|\mathbf{w}^l - \mathbf{v}^l\|_1 \leq tol$, then stop with $(\mathbf{w}^l, \mathbf{v}^l)$. Otherwise, choose new penalty parameter $\beta^{l+1} = \min\{\tau\beta^l, \beta_{\max}\}$, and return to Step 2.

6. Numerical Results

This section shows extensive numerical experiments. In Section 6.1, we first present six real data sets, explain some existing models to be compared with and describe the performance measures to be used. In Section 6.2, we demonstrate that our methods lead to robust and sparse portfolios. In Section 6.3, we compare nine popular portfolios in terms of *out-of-sample* (OOS) performance measures. Finally, in Section 6.4, we show the cumulative return of different portfolio strategies. All of our computations are conducted in the Matlab R2019a environment, on a PC with an Intel(R) Core(TM) i5-7200U CPU (2.50 GHz, 4 CPUs) and 4G RAM processors.

6.1. Models of Comparison, Data, and Performance Measures

(a) Eleven portfolio models compared. We compare the OOS performance of 11 portfolio models across six real data sets of weekly and monthly returns. Those models are well studied, and we divide them into four groups, which are summarized in Table 1. The first group is the robust and sparse portfolio strategies developed in this paper. The second group includes some well-studied portfolio strategies. The third group includes three benchmark portfolio strategies. The last group consists of two portfolios that use the shrinkage technique to estimate the covariance matrix.

Table 1. List of portfolio strategies considered.

Group	Model	*Abbr.*	*Refer.*
(1)	Robust and sparse portfolios with		
	quadratic uncertainty set	RSQ	this paper
	absolute uncertainty set	RSA	this paper
(2)	Some well-studied portfolio strategies with		
	ℓ_1 regularization	L1	Brodie et al. [2]
	$\ell_{1,2}$ regularization	L12	Zhao et al. [37]
	Elastic Net regularization	EN	Yen and Yen [38]
	upper and lower bound	Box	Behr et al. [39]
(3)	Benchmarks' portfolio strategies with		
	short-sales constrained	SC	Jagannathan and Ma [5]
	short-sales unconstrained	SU	Jagannathan and Ma [5]
	equally weighted (1/N) portfolio	EW	DeMiguel et al. [40]
(4)	Shrinkage of covariance		
	sample covariance and identity matrix	SCID	Olivier and Wolf [41]
	sample covariance and 1-factor matrix	SC1F	Olivier and Wolf [3]

(b) Seven data sets tested. Table 2 lists some real-world data sets: DJIA [42], NASDAQ [43], S&P [44,45], Russell2000 [46], Russell3000 [47], and FF100 [38]. All the data are obtained from Yahoo finance (https://finance.yahoo.com/, accessed on 10 January 2023) and Ken French's website (https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html, accessed on 10 January 2023). In all cases, we remove those assets that have missing values.

Table 2. Information of the seven real data sets.

#	Data Sets	Stocks	Time Period	Source	Frequency
1	DJIA	29	01/10/2017–30/10/2022	Yahoo finance	Weekly
2	NASDAQ	95	01/10/2017–30/10/2022	Yahoo finance	Weekly
3	SP500	336	01/10/2017–30/10/2022	Yahoo finance	Weekly
4	Russell2000	1340	01/10/2017–30/10/2022	Yahoo finance	Weekly
5	Russell3000	2166	01/10/2017–30/10/2022	Yahoo finance	Weekly
6	SP100	71	01/10/2017–30/10/2022	Yahoo finance	Weekly
7	FF100	100	11/1999–06/2022	K.French	Monthly

(c) Measuring the OOS performance and its setup. We largely follow the "rolling-window" procedures in [2,37] to conduct our comparison. Let T be the length of a data set and τ be the window length (e.g., $\tau = 120$) used to construct the optimal portfolio by a model. In each period $(t+1)$, $t = \tau, ..., T-1$, we compute different portfolios over the previous τ periods. We then compute the OOS return in the $(t+1)$-th period based on the obtained portfolio. We repeat this procedure until we reach the end of the data set. In this way, we will obtain a series of $(T-\tau)$ portfolio vectors for each model listed in Table 1. To make it precise, let $\mathbf{w}_t^s$ be the optimal portfolio obtained by the portfolio strategy s over the date from $t-\tau+1, \ldots, t$. The OOS return in the $t+1$ period is computed as $\mathbf{r}_{t+1}^s = \mathbf{w}_t^{s\intercal}\mathbf{r}_{t+1}$, where $\mathbf{r}_{t+1}$ is the return in the $(t+1)$-th period. Thus, we obtain a time series of $(T-\tau-1)$ periods OOS returns for all strategies. Note that we use the traditional "rolling-window" procedures for the numerical analysis, and some new methods could provide new ideas for the analysis of portfolio selection problems, see [48].

The OOS performance of each portfolio strategy is assessed by using four quantities: (i) the OOS portfolio variance ($\widehat{\sigma}^2$), (ii) the OOS portfolio Sharpe ratio ($\widehat{SR}$), (iii) portfolio turnover ($TURN$), and (iv) the average short positions (ASP). The specific definitions can be found in DeMiguel et al. [6], Yen and Yen [38], and Zhao et al. [37]. We evaluate the cumulative return (CR). The CR of a portfolio scores the total payoffs that are yielded by

the investment strategy across the investment periods without considering any risk or cost, see Shen et al. [49]. We also consider some quantities studied in [38] on the profiles of the portfolio weights: PAP represents the proportion of active positions and PZP is the proportion of zero positions, respectively, defined as $PAP_t = \frac{|S_t^1|}{N}$, $PZP_t = \frac{|S_t^0|}{N}$, where $S_t^1 = \{i : w_{i,t} \neq 0\}$ and $S_t^0 = \{i : w_{i,t} = 0\}$.

6.2. Robust and Sparse Portfolio

This section shows the weight of robust and sparse portfolios. We use the DJIA data set and the sparse levels $s_1 = \lceil 30\%n \rceil$ and $s_2 = \lceil 50\%n \rceil$. The parameter $\alpha = \beta = 10\lambda$, and the value of λ varies from 10^{-2} to 10^1.

Figure 1 shows the portfolio weights, PAP, and PZP. The two plots in the top panel correspond to a robust portfolio under the quadratic uncertainty set, and the sparsity is s_1. The two plots in the bottom panel correspond to a robust portfolio under the absolute uncertainty set, and the sparsity is s_2. With the increase of penalty parameter λ, the portfolio weights tended to be sparse. The PAP and PZP indicate that we can obtain sparse portfolios that satisfy the specified sparsity.

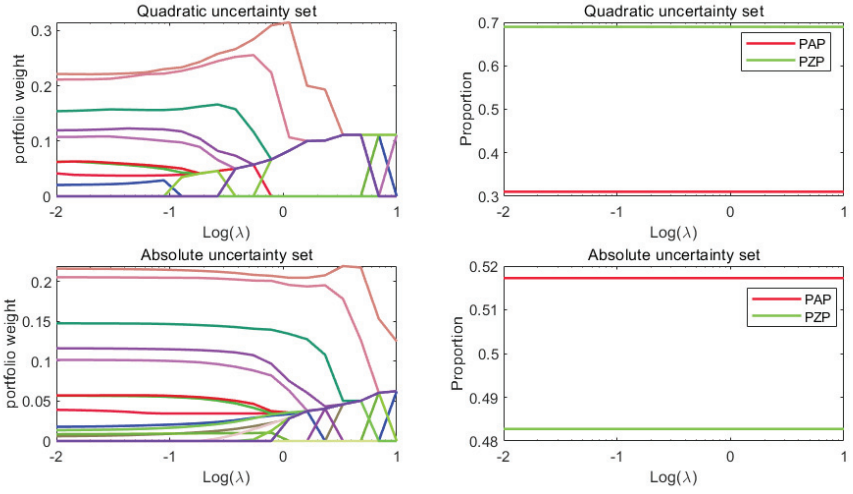

Figure 1. Portfolio weights.

Figure 2 shows the sparse portfolio. We use four different data sets. The sparsity level on DJIA is $s_1 = \lceil 30\%n \rceil$, on NASDAQ and FF100 is $s_2 = \lceil 10\%n \rceil$, and on Russell2000 is $s_3 = \lceil 1\%n \rceil$. We solve the robust portfolio under the quadratic uncertainty set to show the results. We obtain the portfolio with the specified sparsity and the distribution of different asset weight values.

6.3. Out-of-Sample Performance

The Sharpe ratio considers return and risk at the same time; it is a comprehensive measurement for us to observe the performance of a portfolio. Thus, we first test the Sharpe ratio of different portfolio strategies. We use the SP100 data set. The parameter $\alpha = 2\beta$ and the value of β varies from 10 to $10^{1.5}$. The sparsity level $s_1 = \lceil 15\%n \rceil$ and $s_2 = \lceil 5\%n \rceil$.

By comparing with two benchmark portfolios, Figure 3 shows that the RSQ and RSA can produce a higher Sharpe ratio when choosing a suitable penalty parameter.

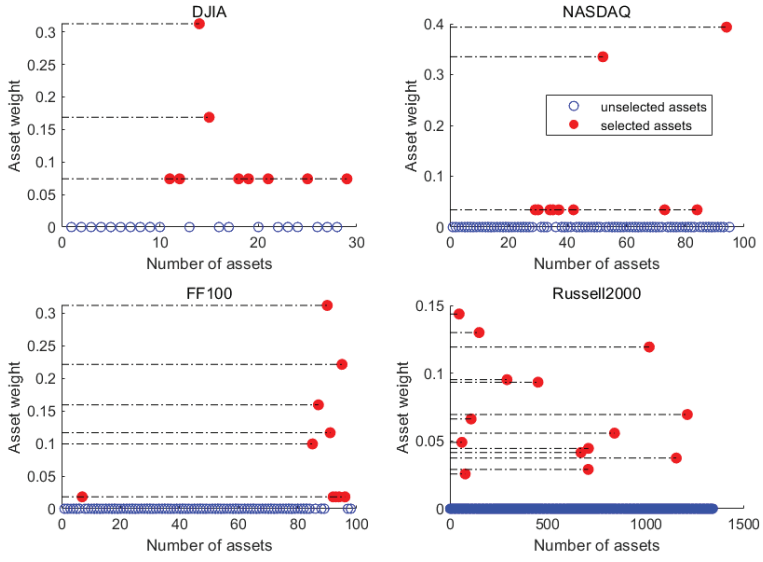

Figure 2. Sparse solutions.

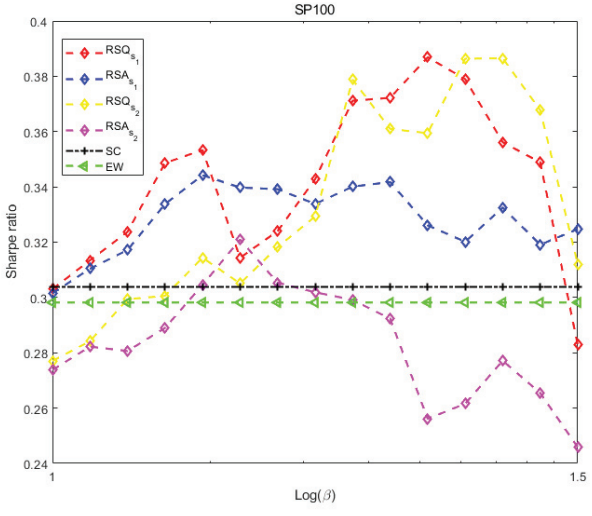

Figure 3. The Sharpe ratio.

Table 3 reports the OOS performance by using four quantities defined in Section 6.1. We set $\alpha = \beta = 10$ and the sparsity level $s_1 == \lceil 30\% n \rceil$ (on the DJIA, NASDAQ, SP500, and FF100 data sets) and $s_2 == \lceil 30\% n \rceil$ (on the Russell2000 and Russell3000 data sets). We can observe that the RSA and RSQ portfolios achieve the smallest variances across all portfolio strategies, i.e., on average with $10.84(\%)^2$ and $11.09(\%)^2$, respectively. This means they are less volatile, i.e., less risky. SU, SC1F, and SCID have the highest variance on average, $995.73(\%)^2$, $442.81(\%)^2$ and $404.20(\%)^2$ in this setting. The variance of the remaining portfolio strategies is $11.94(\%)^2$ (L1), $11.98(\%)^2$ (EN), $16(\%)^2$ (L12), $27.78(\%)^2$ (SC), $28.71(\%)^2$ (EW), and $71.11(\%)^2$ (Box), respectively. In addition, we observe that the Sharpe ratios of the various portfolios on average are 12.34% (SC), 11.82% (EW), 11.77% (RSA), 11.70% (RSQ), 11.61% (L12), 11.21% (EN), 11.18% (L1), 10.58% (SCID), 10.09% (SC1F),

9.23% (Box), and 7.92% (SU). We see that the RSA and RSQ portfolios do not result in a significantly different OOS Sharpe ratio when compared with SC and EW; however, they are higher than the rest of the portfolio strategies.

Table 3. Portfolio out-of-sample variance ($\widehat{\sigma}^2$) ($(\%)^2$), Sharpe ratio ($\widehat{SR}$), turnover ($TURN$), and the average short positions (ASP).

		DJIA	NASDAQ	SP500	Russell2000	Russell3000	FF100
		n = 29	**n = 95**	**n = 336**	**n = 1340**	**n = 2166**	**n = 100**
RSA	var	5.3741	5.9065	11.5604	12.1813	12.7837	17.2215
	SR	0.0703	0.1807	0.1008	0.0063	0.0267	0.3215
	TURN	0.1194	0.1721	0.1430	0.1033	0.1021	0.1698
	ASP	−1.11e-18	2.22e-18	0	−3.08e-18	−4.01e-18	−6.28e-18
RSQ	var	5.3949	5.8633	12.5936	12.7041	12.7702	17.2097
	SR	0.0737	0.1889	0.0863	0.0059	0.0258	0.3216
	TURN	0.1063	0.1711	0.1334	0.1112	0.1003	0.1698
	ASP	−6.66e-18	−3.33e-18	−1.43e-17	−9.25e-18	−9.25e-18	0
L12	var	9.5931	7.1442	32.1907	14.8927	13.7879	18.3918
	SR	0.0860	0.1768	0.1000	0.0072	0.0295	0.2971
	TURN	0.0220	0.0286	0.0369	0.0675	0.0605	0.0296
	ASP	−0.0216	−0.0209	−0.0240	−0.0057	−0.0050	−0.0276
L1	var	8.0373	6.7078	13.1228	14.2679	13.0166	16.4937
	SR	0.0864	0.1831	0.0414	0.0042	0.0279	0.3279
	TURN	0.1063	0.0677	0.0369	0.1250	0.1136	0.0629
	ASP	−0.0036	−0.0012	−0.0124	−0.0039	0.0020	−0.0195
EN	var	8.5871	6.5362	12.9449	14.2074	13.0817	16.5823
	SR	0.0911	0.1790	0.0418	0.0029	0.0402	0.3181
	TURN	0.0256	0.0472	0.0408	0.1208	0.1181	0.0510
	ASP	−0.0257	−0.0198	−0.0272	0.0015	0.0011	−0.0257
BOX	var	9.9181	8.7749	3.59e+02	7.3636	7.0004	34.5878
	SR	0.0250	0.0768	−0.1204	−0.0495	−0.0023	0.6244
	TURN	0.8322	1.7146	3.2709	0.2820	0.2973	5.1895
	ASP	1.1172	2.7850	6.5137	0.5744	0.5886	8.4327
SC	var	10.0132	8.0089	86.1291	24.3110	16.9741	21.2942
	SR	0.0871	0.1891	0.1202	0.0233	0.0399	0.2812
	TURN	0.0388	0.0313	0.0411	0.0512	0.0481	0.0247
	ASP	1.38e-16	1.23e-16	−1.52e-16	3.12e-16	−4.19e-16	0
SU	var	9.9181	15.0721	5.91e+03	12.1721	9.9999	17.2369
	SR	0.0250	0.1800	−0.1270	−0.0482	−0.0167	0.4624
	TURN	0.8322	2.7947	3.9150	0.3919	0.3722	5.2429
	ASP	1.1172	2.5580	5.8870	0.5542	0.5494	6.3456
EW	var	11.1339	8.1780	89.8252	21.3370	19.5742	22.2346
	SR	0.0701	0.1790	0.1151	0.0194	0.0337	0.2922
	TURN	0.0208	0.0253	0.0400	0.0429	0.0381	0.0252
	ASP	1.13e-16	1.13e-16	−1.12e-16	4.52e-16	−3.39e-16	0
SCID	var	6.9600	6.9782	2.38e+03	7.3676	7.0304	16.8727
	SR	0.0294	0.1097	−0.1259	−0.0497	−0.0025	0.6740
	TURN	0.4115	0.8589	0.9940	0.2803	0.2884	1.5393
	ASP	0.6362	1.6735	2.8275	0.5727	0.5871	3.5605
SC1F	var	6.2670	6.5347	2.6140e+03	7.3790	7.0367	15.6203
	SR	0.0380	0.0902	−0.1253	−0.0510	−0.0034	0.6570
	TURN	0.3032	1.1648	0.8972	0.2964	0.2962	1.5858
	ASP	0.4690	1.3556	2.1771	0.5722	0.5863	2.8086

As for the portfolio turnover, unsurprisingly, the EW portfolio strategy exhibit the lowest turnover of all portfolio strategies, amounting to 3.20%. The RSA and RSQ portfolio strategies have moderate levels of turnover on average, 13.49% and 13.20%. The highest average turnover is generated by the SU portfolio and, then, by the Box portfolio, amounting on average to 225.81% and 193.10%, meaning that they are very costly. The turnover of the remaining portfolio strategies range between 11.85% (L12), 25.76% (L2), 16.66% (L1), and 13.98% (EN), respectively. The high turnover of SU and Box was reflect in the enormous average short positions of over 283.52% and 333.52% on average across the six data sets. The second two highest average short positions are by SCID and SC1F, respectively, amounting to 164.29% and 132.81%. The average short positions of the SC and EW portfolios are on average approximately 0% across the six data sets. The average short positions of the RSQ and RSA portfolio strategies also tend to zero. Therefore, considering the moderate turnover and the average short positions, the proposed RSQ and RSA strategies represent a practically implementable method that outperform the portfolio strategies listed in Table 1.

6.4. Cumulative Return

In this subsection, we show the CR of several portfolio strategies. We use the FF100 data set. The sparsity level $s = \lceil 10\%n \rceil$. The parameter $\alpha = \beta = 10$. According to the OOS performance, we choose RSQ, RSA, L12, L1, EN, EW, and SC to compare the CR.

Figure 4 shows the curves of the CR over the corresponding investment periods for the different portfolio strategies. Apparently, RSQ and RSA outperform the others with visible margins. However, RSA and RSQ do not produce significant differences. This result suggest that, compared with the other portfolios, the sparse portfolios RSA and RSQ grow more steadily together with a reduced volatility across most of the investment periods.

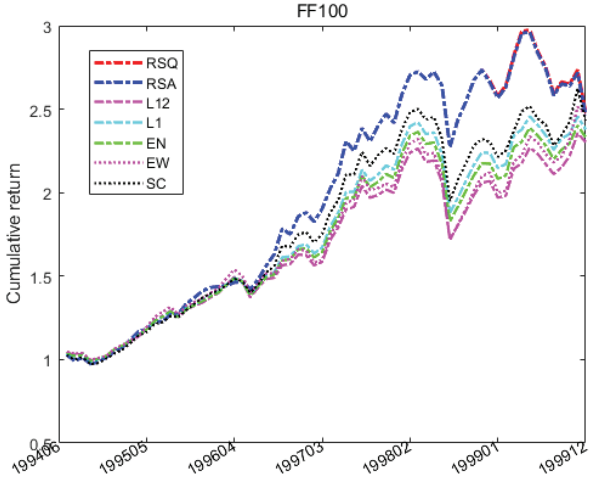

Figure 4. The cumulative return.

7. Conclusions

Portfolio selection has been a fertile area for robust optimization techniques. We proposed a robust and sparse portfolio selection optimization model by considering the perturbation in the asset return matrix and the parameter uncertainty in the expected asset return. We used the equivalence of robustness and regularization to deal with the perturbation in the asset return matrix. To deal with the uncertainty in the expected asset return, we considered two kinds of uncertainty sets and solved the worst-case scenario. We defined three types of stationary points of the penalty problem and then analyzed the relationship between these stationary points and local/global minimizers. Then, we

designed the penalty alternating direction method to solve each problem. Although there is no theoretical guarantee for the equivalence between problems (5) and (6), as well as problems (8) and (18), we confirmed the effectiveness of our approach by comparing with nine widely studied portfolio models on seven real-world data sets. Extensive numerical experiments confirm that the portfolios we proposed have lower volatility, that is less risk. Moreover, our portfolio strategies can yield higher Sharpe ratios when the appropriate parameters are selected.

We note that the robust optimization (RO) mainly consider the uncertainty sets of parameters and thus it do not consider any distribution information of the data. This characteristic makes RO attractive, but at the same time, this method loses the comprehensive characterization of the data. Recently, distributed robust optimization (DRO) has attracted widespread attention and research. Although DRO takes into account the distribution information of the data, the cost paid is that it is difficult to solve. We will consider how to apply DRO to sparse portfolio problems, while considering the distribution information of financial data and improving the sparsity. The most direct extension is the distributed robust portfolio optimization with the ℓ_0 norm constraint, which is a worthwhile and challenging issue.

Author Contributions: Conceptualization, H.Z.; methodology, H.Z. and Y.J.; software, H.Z., Y.J., and Y.Y.; formal analysis, H.Z. and Y.J.; writing—original draft preparation, H.Z.; writing—review and editing, H.Z., Y.J., and Y.Y.; visualization, H.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data source has been presented in Section 6 of the paper.

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

References

1. Markowitz, H. Portfolio selection. *J. Financ.* **1952**, *7*, 142–149.
2. Brodie, J.; Daubechies, I.; De Mol, C.; Giannone, D.; Loris, I. Sparse and stable markowitz portfolios. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 12267–12272. [CrossRef] [PubMed]
3. Ledoit, O.; Wolf, M. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *J. Empir. Financ.* **2003**, *10*, 603–621. [CrossRef]
4. Ledoit, O.; Wolf, M. Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets goldilocks. *Rev. Financ. Stud.* **2017**, *30*, 4349–4388. [CrossRef]
5. Jagannathan, R.; Ma, T. Risk reduction in large portfolios: Why imposing the wrong constraints helps. *J. Financ.* **2003**, *58*, 1651–1683. [CrossRef]
6. DeMiguel, V.; Garlappi, L.; Nogales, F.J.; Uppal, R. A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Manag. Sci.* **2009**, *55*, 798C812. [CrossRef]
7. Ben-Tal, A.; Nemirovski, A.; Roos, C. Robust solutions of uncertain quadratic and conic-quadratic problems. *SIAM J. Optim.* **2002**, *13*, 535–560. [CrossRef]
8. El Ghaoui, L.; Oustry, F.; Lebret, H. Robust solutions to uncertain semidefinite programs. *SIAM J. Optim.* **1998**, *9*, 33–52. [CrossRef]
9. Lee, Y.; Kim, M.J.; Kim, J.H.; Jang, J.R.; Chang, K.W. Sparse and robust portfolio selection via semi-definite relaxation. *J. Oper. Res.* **2020**, *71*, 687–699. [CrossRef]
10. Swain, P.; Ojha, A.K. Robust approach for uncertain portfolio allocation problems under box uncertainty. In *Recent Trends in Applied Mathematics: Select Proceedings of AMSE*; Springer: Singapore, 2019; pp. 347–356.
11. Goldfarb, D.; Iyengar, G. Robust portfolio selection problems. *Math. Oper. Res.* **2003**, *28*, 1–38. [CrossRef]
12. Lobo, M.S.; Boyd, S. The Worst-Case Risk of a Portfolio. Unpublished Manuscript. Available online: Available online: http://faculty.fuqua.duke.edu/(2000)%7Emlobo/bio/researchfiles/rsk-bnd.pdf (accessed on 13 July 2022).
13. Min, L.; Dong, J.; Liu, J.; Gong, X. Robust mean-risk portfolio optimization using machine learning-based trade-off parameter. *Appl. Soft Comput.* **2021**, *113*, 107948. [CrossRef]
14. Won, J.H.; Kim, S.J. Robust trade-off portfolio selection. *Optim. Eng.* **2020**, *21*, 867-904. [CrossRef]
15. Fabozzi, F.J.; Huang, D.; Zhou, G. Robust portfolios: Contributions from operations research and finance. *Ann. Oper. Res.* **2010**, *176*, 191–220. [CrossRef]
16. Scutella, M.G.; Recchia, R. Robust portfolio asset allocation and risk measures. *Ann. Oper. Res.* **2013**, *204*, 145–169. [CrossRef]
17. Xidonas, P.; Steuer, R.; Hassapis, C. Robust portfolio optimization: A categorized bibliographic review. *Ann. Oper. Res.* **2020**, *292*, 533–552. [CrossRef]

18. Ghahtarani, A.; Saif, A.; Ghasemi, A. Robust portfolio selection problems: A comprehensive review. *Oper. Res.* **2022**, *22*, 3203–3264. [CrossRef]
19. Leyffer, S.; Menickelly, M.; Munson, T.; Vanaret, C.; Wild, S.M. A survey of nonlinear robust optimization. *INFOR: Inf. Syst. Oper. Res.* **2020**, *58*, 342–373. [CrossRef]
20. Zhao, Z.; Xu, F.; Du, D.; Meihua, W. Robust portfolio rebalancing with cardinality and diversification constraints. *Quant. Financ.* **2021**, *21*, 1707–1721. [CrossRef]
21. Tütüncü, R.H.; Koenig, M. Robust asset allocation. *Ann. Oper. Res.* **2004**, *132*, 157–187. [CrossRef]
22. Khodamoradi, T.; Salahi, M.; Najafi, A.R. Robust CCMV model with short selling and risk-neutral interest rate. *Phys. A Stat. Mech. Its Appl.* **2020**, *547*, 124429. [CrossRef]
23. Fabozzi, F.J.; Kolm, P.N.; Pachamanova, D.A.; Focardi, S.M. Robust portfolio optimization. *J. Portf. Manag.* **2007**, *33*, 40. [CrossRef]
24. Pınar, M.Ç. On robust mean-variance portfolios. *Optimization* **2016**, *65*, 1039–1048. [CrossRef]
25. Busse, J.A.; Chordia, T.; Jiang, L.; Tang, Y. Transaction costs, portfolio characteristics, and mutual fund performance. *Manag. Sci.* **2021**, *67*, 1227–1248. [CrossRef]
26. Hautsch, N.; Voigt, S. Large-scale portfolio allocation under transaction costs and model uncertainty. *J. Econom.* **2019**, *212*, 221–240. [CrossRef]
27. Yu, J.R.; Chiou, W.J.P.; Lee, W.Y.; ; Lin, S.J. Portfolio models with return forecasting and transaction costs. *Int. Rev. Econ. Financ.* **2020**, *66*, 118–130. [CrossRef]
28. Mordukhovich, B.S.; Nguyen, M.M. *An Easy Path to Convex Analysis and Applications*; Morgan and Claypool Publishers Series: San Rafael, CA, USA, 2014.
29. Pan, L.; Xiu, N.; Fan, J. Optimality conditions for sparsity nonlinear programming. *Sci. China Math.* **2017**, *60*, 759–776. [CrossRef]
30. Rockafellar, R.T.; Wets, R.J. *Variational Analysis*; Springer: Berlin/Heidelberg, Germany, 1998.
31. Bertsimas, D.; Dunn, J. *Machine Learning under a Modern Optimization Lens*; Dynamic Ideas LLC: Charlestown, MA, USA, 2019.
32. Beck, A. *First-Order Methods in Optimization*; Society for Industrial and Applied Mathematics and Mathematical Optimization Society: Philadelphia, PA, USA, 2017.
33. Costa, C.M.; Kreber, D.; Schmidta, M. An alternating method for cardinality-constrained optimization: A computational study for the best subset selection and sparse portfolio problems. *Informs J. Comput.* **2022**, *34*, 2968–2988. [CrossRef]
34. Heckel, T.; de Carvahlo, R.; Lu, X.; Perchet, R. Insights into robust optimization: Decomposing into mean-variance and risk-based portfolios. *J. Invest. Strateg.* **2016**, *6*, 1–24. [CrossRef]
35. Yin, C.; Perchet, R.; Soupé, F. A practical guide to robust portfolio optimization. *Quant. Financ.* **2021**, *21*, 911–928. [CrossRef]
36. Geissler, B.; Morsi, A.; Schewe, L.; Schmidt, M. Penalty alternating direction methods for mixed-integer optimization: A new view on feasibility pumps. *SIAM J. Optim.* **2017**, *27*, 1611–1636. [CrossRef]
37. Zhao, H.; Kong, L.; Qi, H.D. Optimal portfolio selections via ℓ_{12}-norm regularization. *Comput. Optim. Appl.* **2021**, *80*, 853–881. [CrossRef]
38. Yen, Y.M.; Yen, T.J. Solving norm constrained portfolio optimization via coordinate-wise descent algorithms. *Comput. Stat. Data Anal.* **2014**, *76*, 737–759. [CrossRef]
39. Behr, P.; Guettler, A.; Miebs, F. On portfolio optimization: Imposing the right constraints. *J. Bank. Financ.* **2013**, *37*, 1232–1242. [CrossRef]
40. DeMiguel, V.; Garlappi, L.; Uppal, R. Optimal versus naive diversification: How in-efficient is the $1/n$ portfolio strategy? *Rev. Financ. Stud.* **2009**, *22*, 1915–1953. [CrossRef]
41. Ledoit, O.; Wolf, M. A well-conditioned estimator for large-dimensional covariance matrices. *J. Multivar. Anal.* **2004**, *88*, 365–411. [CrossRef]
42. Lai, Z.R.; Yang, P.Y.; Fang, L.; Wu, X. Short-term sparse portfolio optimization based on alternating direction method of multipliers. *J. Mach. Learn. Res.* **2018**, *19*, 2547–2574.
43. Chou, R.K.; Chung, H. Decimalization, trading costs, and information transmission between etfs and index futures. *J. Futur. Mark.* **2006**, *26*, 131–151. [CrossRef]
44. Kan, R.; Wang, X.; Zhou, G. Optimal portfolio choice with estimation risk: No risk-free asset case. *Manag. Sci.* **2022**, *68*, 2047–2068. [CrossRef]
45. Mutunge, P.; Haugl, ; D. Minimizing the tracking error of cardinality constrained portfolios. *Comput. Oper. Res.* **2018**, *90*, 33–41. [CrossRef]
46. Fan, J.; Zhang, J.; Yu, K. Vast portfolio selection with gross-exposure constraints. *J. Am. Stat. Assoc.* **2012**, *107*, 592–606. [CrossRef]
47. Teng, Y.; Yang, L.; Yu, B.; Song, X. A penalty PALM method for sparse portfolio selection problems. *Optim. Methods Softw.* **2017**, *32*, 126–147. [CrossRef]

48. Wang, Y.; Gao, S.; Yu, Y.; Cai, Z.; Wang, Z. A gravitational search algorithm with hierarchy and distributed framework. *Knowl.-Based Syst.* **2021**, *218*, 106877. [CrossRef]
49. Shen, W.; Wang, J.; Ma, S. Doubly regularized portfolio with risk minimization. In Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, Québec City, QC, Canada, 27–31 July 2014.

mathematics

MDPI

Article

Optimal Debt Ratio and Dividend Payment Policies for Insurers with Ambiguity

Dan Zhu [1], Cuixia Chen [2,*] and Bing Liu [3]

[1] School of Statistics and Data Science, Qufu Normal University, Qufu 273165, China; zhudanspring@qfnu.edu.cn
[2] School of Insurance and Public Finance, Hebei Finance University, Baoding 071051, China
[3] School of Finance, Nanjing University of Finance and Economics, Nanjing 210023, China; liubingsdly@nufe.edu.cn
* Correspondence: chencuixia@hbfu.edu.cn

Abstract: This study considers the optimal debt ratio and dividend payment policies for an insurer concerned about model misspecification. We assume that the insurer can invest all of its asset to the financial market and the ambiguity may exist in the risky asset. Taking into account the ambiguous situation, the insurer aims to maximize the expected utility of a discounted dividend payment until it ruins. Under some assumption, we prove that there exists classical solutions of the optimal debt ratio, dividend payment policies, and value functions that show that the existence of ambiguity can affect the optimal debt ratio and dividend policies significantly.

Keywords: dividend payment; model ambiguity; optimal debt ratio

MSC: 60H30

1. Introduction

Both dividends and debt are crucial factors in an insurance company. Dividends are a critical component of shareholder returns. Their demonstration shows the company's financial strength and ability to generate profits. Dividends also serve as a signal to investors about the company's future prospects. If the dividend payout ratio is consistent and reliable, it indicates a stable and profitable business model. On the other hand, insurance company debt is also crucial. It is a crucial element in risk management and a key factor in ensuring solvency. The debt-to-asset ratio indicates how well the company manages its balance sheet and risks. A debt ratio that is too high may indicate a leveraged balance sheet and potentially increase the default risk. Conversely, a debt ratio that is too low may indicate underutilized capital and potentially missed opportunities for growth. Therefore, both dividends and debt are important factors to consider when analyzing insurance company performance and financial health.

Due to the nature of their insurance product, insurers sometimes collect substantial sums of cash, cash equivalents, and pursue capital gains in order to cover future claims and prevent bankruptcy. The appropriate debt level and prospective insurance liabilities is of great importance for an insurer. In actuarial science, the appropriate debt level and prospective insurance liabilities of an insurance company should be discussed in detail. Many researchers have investigated the optimal debt policy of an insurer in the last decade. For example, Jin et al. (2015) [1] studied the optimal debt ratio problem considering reinsurance, where they used the subsolution–supersolution method to deal with the existence of solutions of the optimal debt ratio policy. Zhao et al. (2018) [2] considered optimal debt ratio policies for an insurer with a regime-switching model. Qian et al. (2018) [3] investigated the optimal liability ratio under catastrophic risk. In

Citation: Zhu, D.; Chen, C.; Liu, B. Optimal Debt Ratio and Dividend Payment Policies for Insurers with Ambiguity. *Mathematics* **2024**, *12*, 40. https://doi.org/10.3390/math12010040

Academic Editors: Jing Yao, Xiang Hu and Jingchao Li

Received: 14 November 2023
Revised: 11 December 2023
Accepted: 13 December 2023
Published: 22 December 2023

continuous-time setting, Li et al. (2023) [4] researched a state-dependent optimal asset–liability management problem. The optimal debt ratio problem can be seen in other words such as Zhu and Yin (2018) [5], Zhang et al. (2020) [6], Meng and Bi (2020) [7], Abid and Abid (2023) [8], and the references therein.

The seminal work of De Finetti (1957) [9] has led to the classical problem of optimal dividend payment in insurance mathematics. Paulsen and Gjessing (1997) [10] analyzed a risk process with stochastic return on investments and obtained the optimal dividend barrier policy. Cai et al. (2006) [11] investigated the Ornstein–Uhlenbeck-type model with credit and debit interest for the optimal dividend problem. Cheung and Wong (2017) [12] studied the dividend payment in the dual risk model considering delays. Xie and Zhang (2021) [13] researched the finite-time dividend problems in a Levy risk model under periodic observation. Chakraborty et al. (2023) [14] considered a diffusive model for optimally distributing dividends under the situation of Knightian model ambiguity. For more studies on the optimal dividend problem for an insurer, we refer the reader to Avanzi (2009) [15], Yao et al. (2011) [16], Yin and Wen (2013) [17], Bi and Meng (2016) [18], Marciniak and Palmowski (2016) [19], Feng et al. (2021) [20], and so on.

Actuarial research has recently revealed a pattern of diverse development. It is clear that there are growing connections between risk theory and mathematics and some optimum control issues are also becoming more significant and fascinating. As a result, scholars have studied the optimal debt ratio combined with optimal dividend problems in great detail. For example, see Meng et al. (2016) [21], Jin et al. (2022) [22], etc. In the above studies, the scholars did not consider the existence of ambiguity and assumed that the models used are exactly true. In reality, the insurance business uses a wealth of data and a variety of technology to predict actuarial models. It is clear that the insurer is unsure whether a model is the correct model or whether there is a misspecification error. Thus, the aim of this study is to analyze the optimal debt ratio and dividend payment policies for an insurer that is concerned about model misspecification. We assume that the insurer has the ability to invest all of its assets in the financial market and that there may be ambiguity in the risky asset. The insurer's goal is to maximize the expected utility of a discounted dividend payment until it ruins, taking into account the ambiguous situation. Based on some assumptions, we prove that there exists classical solutions of the optimal debt ratio, dividend payment policies, and value functions.

2. Model Formulation

We firstly describe an insurer's surplus process as follows

$$X(t) = K(t) - L(t), \tag{1}$$

where $K(t)$ and $L(t)$ denote asset value and liabilities at time t, respectively. Denote π_t as a debt ratio, i.e.,

$$\pi_t = \frac{L(t)}{X(t)}. \tag{2}$$

Then, we have

$$1 + \pi_t = \frac{K(t)}{X(t)}. \tag{3}$$

Intuitively, if the insurer holds a liability, it will earn premium. Denote α as the premium rate, which means that the insurer can earn α dollars when it has provided a dollar insurance protection. Thus, the increase in the asset value of insurance sales during $[t, t + \mathrm{d}t]$ can be determined by $\alpha L(t)\mathrm{d}t$. Consequently, the insurer aims to know how much of the debt ratio is suitable. For the sake of simplicity, we assume that there is only one risky asset in the financial market. Thus, the price of the risky asset $M(t)$ satisfies that

$$\frac{\mathrm{d}M(t)}{M(t)} = \mu \mathrm{d}t + \sigma \mathrm{d}B_1(t), \tag{4}$$

where $\mu > 0$ and $\sigma > 0$ are real numbers. We assume that the insurer invests all of its asset value $K(t)$ into the financial market. Without considering claims and dividend payment, the surplus process of the insurer can be written as

$$\mathrm{d}X_1(t) = \mathrm{d}K(t) = [\alpha L(t) + K(t)][\mu \mathrm{d}t + \sigma \mathrm{d}B_1(t)]. \tag{5}$$

Then, we assume the accumulated claims up to time t are proportional to the insurer's liabilities $L(t)$, denoted as $S(t)$,

$$S(t) = \int_0^t c(s)L(s)\mathrm{d}t. \tag{6}$$

where $c(t)$ is served as a claim rate which can be described by a diffusion process as

$$\mathrm{d}c(t) = h(c(t))\mathrm{d}t + v\mathrm{d}B_2(t), \tag{7}$$

with $c(0) = c$, where $h(c(t))$: $R \to R$ is an expected claim rate, and $v > 0$ is the volatility of the claim rate. $Cov(B_1(t), B_2(t)) = \rho t$, $-1 < \rho < 1$ represents the correlation between the future claims and the risky asset. Using Gaussian linear regression, we can obtain that

$$\mathrm{d}B_1(t) = \rho \mathrm{d}B_2(t) + \sqrt{1-\rho^2}\mathrm{d}B_3(t), \tag{8}$$

where $B_3(t)$ is a standard Brownian motion, and $B_2(t)$ and $B_3(t)$ are independent.

What is more, we also consider the dividend payment in this paper, and we denote $D(t)$ as accumulated dividend payments up to time t,

$$\mathrm{d}D(t) = z_t X(t)\mathrm{d}t, \ \ with \ \ D(0^-) = 0. \tag{9}$$

where $0 < z_t < M$ is $\mathcal{F}_t$-adapted, and M is a positive constant. Thus, the wealth process of the insurer considering claims and dividend payments can be given in the following

$$\mathrm{d}X(t) = \mathrm{d}X_1(t) - \mathrm{d}S(t) - \mathrm{d}D(t), \ \ X(0) = x. \tag{10}$$

Substituting (5)–(8) into (10), we have

$$\frac{\mathrm{d}X(t)}{X(t)} = (\alpha\pi_t + \pi_t + 1)\mu\mathrm{d}t - c(t)\pi_t\mathrm{d}t - z_t\mathrm{d}t + (\alpha\pi_t + \pi_t + 1)\sigma\rho\mathrm{d}B_2(t) + (\alpha\pi_t + \pi_t + 1)\sigma\sqrt{1-\rho^2}\mathrm{d}B_3(t). \tag{11}$$

The value function is usually set to maximize the expected utility of a discounted dividend payment until it ruins (for example, see Jin et al. (2015) [1]).

$$V_1(x,c) = \sup_{\pi,z} E^P\left[\int_0^\tau U(z_s X(s))\mathrm{d}s \middle| X(0) = x, c(0) = c\right], \tag{12}$$

where τ is the time of ruin and $\tau = \inf\{t \geq 0 : X(t) < 0\}$, where E^P denotes the expectation operator under probability measure P, U is a utility function satisfying $U' > 0$, $U'' < 0$, π, and z are some admissible policies that will be described later. The insurer's understanding of the true probability measure (P) used in computing the equation above is the basic assumption behind this model. The assumption being too strong has been argued by some researchers. Insurers should be permitted to consider optimal policies for other measures of probability. Otherwise, if an insurer ignores the ambiguity of the probability measure and trusts P completely, the insurer may make some mistakes in some decision problems. It is our assumption that the insurer's ambiguity about the financial market model is only due to its limited information about the financial market. The purpose of this paper is to examine the optimal debt ratio and dividend strategy policy with ambiguity in the model against the financial market only. The model ambiguity in our optimal control problem is presented in the following.

We are aware that the probability measure P mentioned above is created using the insurer's limited information. The insurance company computes P by utilizing a vast amount of data and various technologies. This P is referred to as the reference model or reference probability. It is clear that the insurer is unsure whether P is the correct model or whether there was a misspecification error. Naturally, the insurer would take other models into account. We call the aforementioned phenomenon ambiguity. Additionally, we presumptively believe that the model ambiguity is limited to the financial market. The alternative models that the insurer considers should then be identical to the reference model and cannot affect the (7) that corresponds to the claims' arrival rate, so we define the alternative models by a class of probability measures that are equivalent to P:

$$\mathcal{Q} := \{Q | Q \sim P, \ \ Q \text{ will not change } (7)\}, \sim \text{means equivalent} \tag{13}$$

In a probability space, two measures, P and Q, are equivalent, denoted by $Q \sim P$, if they have same null sets, i.e., $Q(A) = 0$, if and only if $P(A) = 0$)

By Girsanov's theorem (Klebaner (2008) [23]), Q satisfies the following conditions

$$\frac{\mathrm{d}Q}{\mathrm{d}P}(B_3[0,T]) = \Lambda(T), \tag{14}$$

where

$$\Lambda(t) = \exp\left\{ \int_0^t m(s)\mathrm{d}B_3(s) - \frac{1}{2}\int_0^t m(s)^2 \right\}$$

is a P-martingale with filtration $\{\mathcal{F}_t\}_{t\in[0,T]}$, and $m(t)$ is a regular adapted process satisfying Novikov's condition, i.e.,

$$E^P\left[\exp\left(\frac{1}{2}\int_0^T [m(t)]^2 \mathrm{d}s\right)\right] < \infty.$$

Then, we have

$$\mathrm{d}B_3(t) = m(t)\mathrm{d}t + \mathrm{d}B_3^Q(t), \tag{15}$$

where $B_3^Q(t)$ is the standard Brownian motion corresponding to the probability measure Q.

We use relative entropy to calculate the difference between each alternative model and the reference model in order to take the alternative model Q into consideration. Relative entropy is a tried-and-true method for calculating the difference between Q and P. Relative entropy has been employed to measure it in numerous research; for examples, see Uppal and Wang (2003) [24], Maenhout (2004) [25], and Yi et al. (2013) [26]. The following is the relative entropy between Q and P.

$$\begin{aligned} H(Q \parallel P) = E^Q\left[\ln\frac{\mathrm{d}Q}{\mathrm{d}P}\right] &= E^Q\left\{\int_0^T m(s)\mathrm{d}B_3(s) - \frac{1}{2}\int_0^T m(s)^2\mathrm{d}s\right\} \\ &= E^Q\left\{\int_0^T m(s)\mathrm{d}B_3^Q(s) + \frac{1}{2}\int_0^T m(s)^2\mathrm{d}s\right\}. \end{aligned}$$

Given that standard Brownian motion is defined by $B_3^Q(t)$ under the probability measure Q, we have

$$H(Q \parallel P) = E^Q\left\{\int_0^T Z(s)\mathrm{d}s\right\} = \int_0^T \frac{1}{2}m(s)^2\mathrm{d}s, \tag{16}$$

where $Z(t) = \frac{1}{2}m(t)^2$. Therefore, the $H(Q \parallel P)$ is measured by $Z(t)$. A penalty is charged if the insurer chooses to use the alternative model Q instead of the reference model P. Naturally, the penalty is increased by the size of the $H(Q \parallel P)$. In a manner similar to Uppal and Wang (2003) [24], the following is how we design a robust control problem:

$$V(x,c) = \sup_{\pi,z\in\mathbb{V}} \inf_{Q\in\mathcal{Q}} E^{Q}_{c,x}\left[\int_0^{\tau} [\xi\phi(V(x,c))Z(s) + U(z_s X(s))\mathrm{d}s]\right], \tag{17}$$

where $E^{Q}_{c,x}[\cdot] = E^{Q}[\cdot|c(0) = c, X(0) = x]$. The standardizing function $\phi(V(t,x)) > 0$ converts the penalty to the same order of magnitude as $V(x,c)$, where this specific version of $\phi(\cdot)$ is typically chosen for analytical simplicity. The size of the ξ indicates how confident the insurer is in the reference model P, where the larger the ξ is, the more confident the insurer is in P, which we assume as $0 < \xi < \infty$ in this paper. The inf term shows the insurer's aversion to ambiguity. In other words, the insurer is conservative and will take into account the worst outcome with ambiguity; it will be further explained that the $\mathbb{V}$ is the set of admissible policies.

3. Main Results

For the purpose of solving (17), the wealth process should be derived under Q. Substituting (15) into (11), we can obtain that

$$\begin{aligned}\frac{\mathrm{d}X(t)}{X(t)} &= (\alpha\pi_t + \pi_t + 1)\mu\mathrm{d}t - c(t)\pi_t\mathrm{d}t - z_t\mathrm{d}t + (\alpha\pi_t + \pi_t + 1)\sigma m(t)\sqrt{1-\rho^2}\mathrm{d}t \\ &\quad +(\alpha\pi_t + \pi_t + 1)\sigma\rho\mathrm{d}B_2(t) + (\alpha\pi_t + \pi_t + 1)\sigma\sqrt{1-\rho^2}\mathrm{d}B^{Q}_3(t).\end{aligned} \tag{18}$$

The authors of (18) show that the alternative models only change the drift coefficient, which exactly corresponds to the use of Girsanov's theorem. Let the generator of (18) be

$$\begin{aligned}\mathcal{A}f(x,c) &= \left[(\alpha\pi_t + \pi_t + 1)\mu\mathrm{d}t - c\pi_t - z_t + (\alpha\pi_t + \pi_t + 1)\sigma m(t)\sqrt{1-\rho^2}\right]x\frac{\partial}{\partial x}f(x,c) \\ &\quad +\frac{1}{2}(\alpha\pi_t + \pi_t + 1)^2\sigma^2x^2\frac{\partial^2}{\partial x^2}f(x,c) + h(c)\frac{\partial}{\partial c}f(x,c) + \frac{1}{2}v^2\frac{\partial^2}{\partial c^2}f(x,c) \\ &\quad +(\alpha\pi_t + \pi_t + 1)\sigma v\rho x\frac{\partial^2}{\partial c\partial x}f(x,c).\end{aligned} \tag{19}$$

We also provide a definition of the set that includes all admissible policies.

Definition 1. *$\mathbb{V} = \{\pi_t, z_t\}$ is admissible, if*

(i) The process $z = \{z_t, t \geq 0\}$ is a predictable and satisfy that $0 \leq z_t \leq M$;

(ii) The process $\pi = \{\pi_t, t \geq 0\}$ is a predictable and satisfy that

$$E^{Q}\int_0^{T} \pi_s^2 ds < \infty, \quad 0 < T < \infty, \ Q \in \mathcal{Q};$$

(iii) The stochastic differential Equation (18) determines a unique strong solution.

Additionally, we state that a pair of policies (π, z) is admissible if $(\pi, z) \in \mathbb{V}$.

It is obvious that $z_t = 0$ for $t \geq \tau$. So, we rewrite function (17) as

$$V(x,c) = \sup_{\pi,z\in\mathbb{V}} \inf_{Q\in\mathcal{Q}} E^{Q}_{c,x}\left[\int_0^{\infty} [\xi\phi(V(x,c))Z(s) + U(z_s X(s))\mathrm{d}s]\right], \tag{20}$$

The Hamilton–Jacobi–Bellman(HJB) (See Fleming and Soner (2006) [27]) equation, which is satisfied by the value function (20), can be given as follows.

$$\max_{\pi,z}\inf_{m}\left\{\mathcal{A}V(x,c) - rV(x,c) + U(z_t x) + \frac{1}{2}\xi\phi(V(x,c))m^2\right\} = 0 \tag{21}$$

Let $G_t := \alpha\pi_t + \pi_t + 1$, then $\pi_t = \frac{G_t - 1}{\alpha + 1}$. Additionally, as previously noted, we select a suitable form of $\phi(\cdot)$ (the form has been employed in Uppal and Wang (2003) [24] and other instances),

$$\phi(V(x,c)) = V(x,c).$$

Hence, the Equation (21) can be simplified as

$$\begin{aligned}\max_{G,z}\inf_{m} \quad & \left\{\left[G_t\mu \mathrm{d}t - \frac{c(G_t-1)}{\alpha+1} - z_t + G_t\sigma m(t)\sqrt{1-\rho^2}\right]x\frac{\partial}{\partial x}V(x,c)\right. \\ & + \frac{1}{2}G_t^2\sigma^2x^2\frac{\partial^2}{\partial x^2}V(x,c) + h(c)\frac{\partial}{\partial c}V(x,c) + \frac{1}{2}v^2\frac{\partial^2}{\partial c^2}V(x,c) \\ & \left. + G_t\sigma v\rho x\frac{\partial^2}{\partial c\partial x}V(x,c) - rV(x,c) + U(z_tx) + \frac{1}{2}\xi V(x,c)m^2\right\} = 0 \end{aligned} \tag{22}$$

Since $\frac{1}{2}\xi V(x,c) > 0$, the function m^* minimizes (22) according to the first-order condition, which takes the following form.

$$m^* = \frac{-G_t\sigma\sqrt{1-\rho^2}x\frac{\partial}{\partial x}V(x,c)}{\xi V(x,c)}. \tag{23}$$

Substituting (23) into (22) yields

$$\begin{aligned}&\max_{G}\left\{\frac{1}{2}G_t^2\sigma^2x^2\frac{\partial^2}{\partial x^2}V(x,c) - \frac{G_t^2\sigma^2(1-\rho^2)x^2[\frac{\partial}{\partial x}V(x,c)]^2}{2\xi V(x,c)} + [G_t\mu - \frac{cG_t}{\alpha+1}]x\frac{\partial}{\partial x}V(x,c)\right. \\ &\left. + G_t\sigma v\rho x\frac{\partial^2}{\partial c\partial x}V(x,c)\right\} + \max_{z}\left\{-z_tx\frac{\partial}{\partial x}V(x,c) + U(z_tx)\right\} \\ &+ \frac{1}{2}v^2\frac{\partial^2}{\partial c^2}V(x,c) + h(c)\frac{\partial}{\partial c}V(x,c) - rV(x,c) + \frac{cx}{\alpha+1}\frac{\partial}{\partial x}V(x,c) = 0 \end{aligned} \tag{24}$$

Assume that the utility function has the following form

$$U(x) = \frac{x^\gamma}{\gamma}, \tag{25}$$

where $0 < \gamma < 1$. We speculate that the value function has the following form given the utility function.

$$V(x,c) = Y(c)\frac{x^\gamma}{\gamma}, \tag{26}$$

where $Y(c)$ is a function of c. To determine $Y(c)$, we derive the following functions

$$\begin{cases} \frac{\partial}{\partial x}V(x,c) = Y(c)x^{\gamma-1}, \\ \frac{\partial^2}{\partial x^2}V(x,c) = Y(c)(\gamma-1)x^{\gamma-2}, \\ \frac{\partial}{\partial c}V(x,c) = Y'(c)\frac{x^\gamma}{\gamma}, \\ \frac{\partial^2}{\partial c^2}V(x,c) = Y''(c)\frac{x^\gamma}{\gamma}, \\ \frac{\partial^2}{\partial x\partial c}V(x,c) = Y'(c)x^{\gamma-1}. \end{cases} \tag{27}$$

Substituting (27) into (24) and simplifying it, we can obtain that

$$\max_{G}\left\{\frac{1}{2}G_t^2[\sigma^2(\gamma-1)Y(c)-\frac{\gamma}{\xi}\sigma^2(1-\rho^2)Y(c)]+G_t\left[(\mu-\frac{c}{\alpha+1})Y(c)+\sigma v\rho Y'(c)\right]\right\}$$
$$+\max_{z}\left\{-z_tY(c)+\frac{z_t^{\gamma}}{\gamma}\right\}+\frac{1}{2}v^2Y''(c)\frac{1}{\gamma}+h(c)Y'(c)\frac{1}{\gamma}-rY(c)\frac{1}{\gamma}+\frac{c}{\alpha+1}Y(c)=0. \tag{28}$$

According to the first-order conditions, we can obtain that

$$\begin{cases} G_t*=\dfrac{(\mu-\frac{c}{\alpha+1})Y(c)+\sigma\nu\rho Y'(c)}{\sigma^2Y(c)[(1-\gamma)+\frac{(1-\rho^2)\gamma}{\xi}]}, \\ z_t^*=Y(c)^{\frac{1}{\gamma-1}}. \end{cases} \tag{29}$$

Substituting (29) into (28), we have

$$\frac{1}{2}\frac{[(\mu-\frac{c}{\alpha+1})Y(c)+\sigma\nu\rho Y'(c)]^2}{\sigma^2Y(c)[(1-\gamma)+\frac{(1-\rho^2)\gamma}{\xi}]}+(\frac{1}{\gamma}-1)Y(c)^{\frac{\gamma}{\gamma-1}}$$
$$+\frac{1}{2}v^2Y''(c)\frac{1}{\gamma}+h(c)Y'(c)\frac{1}{\gamma}-rY(c)\frac{1}{\gamma}+\frac{c}{\alpha+1}Y(c)=0. \tag{30}$$

Multiplying both sides by γ in the above equation and simplifying it, (30) can be represented as

$$\frac{1}{2}\nu^2Y''(c)+\left[h(c)+\frac{\nu\rho\gamma(\mu-\frac{c}{\alpha+1})}{\sigma[(1-\gamma)+\frac{\gamma(1-\rho^2)}{\xi}]}\right]Y'(c)+\left[\frac{1}{2}\frac{\gamma(\mu-\frac{c}{\alpha+1})^2}{\sigma^2[(1-\gamma)+\frac{(1-\rho^2)\gamma}{\xi}]}+\frac{c\gamma}{\alpha+1}-r\right]Y(c)$$
$$+\frac{1}{2}\frac{\nu^2\rho^2\gamma}{(1-\gamma)+\frac{(1-\rho^2)\gamma}{\xi}}\frac{Y'(c)^2}{Y(c)}+(1-\gamma)Y(c)^{\frac{\gamma}{\gamma-1}}=0. \tag{31}$$

For the sake of simplicity, denote $\Lambda(c)=\ln Y(c)$. Additionally, let

$$\begin{cases} H(c)=h(c)+\dfrac{\nu\rho\gamma(\mu-\frac{c}{\alpha+1})}{\sigma[(1-\gamma)+\frac{\gamma(1-\rho^2)}{\xi}]}, \\ K(c)=\dfrac{1}{2}\dfrac{\gamma(\mu-\frac{c}{\alpha+1})^2}{\sigma^2[(1-\gamma)+\frac{(1-\rho^2)\gamma}{\xi}]}+\dfrac{c\gamma}{\alpha+1}, \\ L=\dfrac{1}{2}\dfrac{\nu^2\rho^2\gamma}{(1-\gamma)+\frac{(1-\rho^2)\gamma}{\xi}}. \end{cases} \tag{32}$$

Then (31) can be represented as

$$\frac{1}{2}\nu^2\Lambda''(c)+H(c)\Lambda'(c)+[L+\frac{\nu^2}{2}]\Lambda'(c)^2+N(c)=0, \tag{33}$$

where $N(c)=K(c)-r+(1-\gamma)e^{\frac{\Lambda(c)}{\gamma-1}}$. Next, we will verify the existence of the classical solution of $\Lambda(c)$. Naturally, from $Y(c)=e^{\Lambda(c)}$, we can also obtain $Y(c)$. In order to obtain the classical solution of (33), we apply the method that Jin et al. (2015) [1] used named subsolution and supersolution. The definition of subsolution and supersolution can be presented in the following.

Definition 2. *A solution $\Lambda_1(c)$ is said to be a subsolution of (33) if $\forall c \in R$, $\Lambda_1(c) \in C^2(R)$ and $\Lambda_1(c)$ satisfies*

$$\frac{1}{2}\nu^2\Lambda_1''(c) + H(c)\Lambda_1'(c) + [L(c) + \frac{\nu^2}{2}]\Lambda_1'(c)^2 + N(c) \geq 0 \tag{34}$$

A solution $\Lambda_2(c)$ is said to be a supersolution of (33) if $\forall\, c \in R$, $\Lambda_2(c) \in C^2(R)$ and $\Lambda_2(c)$ satisfies

$$\frac{1}{2}\nu^2\Lambda_2''(c) + H(c)\Lambda_2'(c) + [L(c) + \frac{\nu^2}{2}]\Lambda_2'(c)^2 + N(c) \leq 0. \tag{35}$$

Furthermore, if $\forall c \in R\ \Lambda_1(c) \leq \Lambda_2(c)$, we say $\Lambda_1(c)$ and $\Lambda_2(c)$ are an ordered pair of subsolution and supersolution, respectively.

In order to obtain the existence of the classical solution of (33), we give the following lemmas.

Lemma 1. *Suppose that*

$$r > \gamma[\mu - \frac{\sigma^2(1-\gamma) + \frac{(1-\rho^2)\gamma}{\xi}}{2}], \tag{36}$$

then

$$\underline{\Lambda} = (\gamma - 1)\ln\left\{\frac{2}{\gamma - 1}[r - \gamma(\mu - \frac{\sigma^2(1-\gamma) + \frac{(1-\rho^2)\gamma}{\xi}}{2})]\right\} \tag{37}$$

is a subsolution of (33).

Proof. Since

$$\begin{aligned} K(c) =& \frac{1}{2}\frac{\gamma(\mu - \frac{c}{\alpha+1})^2}{\sigma^2[(1-\gamma) + \frac{(1-\rho^2)\gamma}{\xi}]} + \frac{c\gamma}{\alpha+1} \\ =& \frac{\gamma\left(\mu - \frac{c}{\alpha+1} - \sigma^2[(1-\gamma) + \frac{(1-\rho^2)\gamma}{\xi}]\right)^2}{2\sigma^2[(1-\gamma) + \frac{(1-\rho^2)\gamma}{\xi}]} + \gamma\mu - \frac{\gamma\sigma^2[(1-\gamma) + \frac{(1-\rho^2)\gamma}{\xi}]}{2} \\ \geq& \gamma\mu - \frac{\gamma\sigma^2[(1-\gamma) + \frac{(1-\rho^2)\gamma}{\xi}]}{2}, \end{aligned} \tag{38}$$

hence

$$\begin{aligned} N(c) =& K(c) - r + (1-\gamma)e^{\frac{\Lambda(c)}{\gamma-1}} \geq \gamma\mu - \frac{\gamma\sigma^2[(1-\gamma) + \frac{(1-\rho^2)\gamma}{\xi}]}{2} - r + 2[r - \gamma(\mu - \frac{\sigma^2(1-\gamma) + \frac{(1-\rho^2)\gamma}{\xi}}{2})] \\ =& r - [\gamma\mu - \frac{\gamma\sigma^2[(1-\gamma) + \frac{(1-\rho^2)\gamma}{\xi}]}{2}] > 0. \end{aligned} \tag{39}$$

Combining with $\frac{1}{2}\nu^2\underline{\Lambda}'' + H(c)\underline{\Lambda}' + [L + \frac{\nu^2}{2}]\underline{\Lambda}'^2 = 0$, we complete the proof. □

Let

$$\begin{aligned} l_1 &= 2\nu^2(\frac{\gamma\rho^2}{1 - \gamma + \frac{\gamma(1-\rho^2)}{\xi}} + 1), \\ l_2 &= -2\frac{\nu\rho\gamma\frac{1}{\alpha+1}}{\sigma[(1-\gamma) + \frac{\gamma(1-\rho^2)}{\xi}]} + 2\kappa, \\ l_3 &= \frac{\gamma}{2\sigma^2(1 - \gamma + \frac{(1-\rho^2)\gamma}{\xi})(\alpha+1)^2}, \end{aligned}$$

where $\kappa > \frac{\nu\rho\gamma}{\sigma[(1-\gamma)+\frac{\gamma(1-\rho^2)}{\xi}](\alpha+1)} - \frac{\nu\sqrt{\gamma[\gamma\rho^2+1-\gamma+\frac{\gamma(1-\rho^2)}{\xi}]}}{\sigma[1-\gamma+\frac{(1-\rho^2)\gamma}{\xi}](\alpha+1)}$. Then the equation $l_1k^2 + l_2k + l_3 = 0$ has two positive real roots denoted by k_1 and k_2.

Obviously, $l_1, l_3 > 0$, $l_2 < 0$ and $l_2^2 - 4l_1l_3 > 0$. So the equation has two positive roots.

Lemma 2. *Let* $k_0 = \frac{k_1+k_2}{2}$, *then* $k_0 > 0$. *Additionally, assume that* $h'(c) < \kappa$ *and*

$$r > K_1(k_0), \tag{40}$$

where $K_1(k_0)$ *will be given later. Then*

$$\tilde{\Lambda}(c) = k_0c^2 + K_0 \tag{41}$$

is a supersolution of (33), *where* K_0 *is a constant which is large enough such that* $\tilde{\Lambda}(c) > \Lambda_1$ *and satisfies that* $K_0 > (\gamma - 1)\ln\left(\frac{r-K_1(k_0)}{1-\gamma}\right)$.

Proof. From (33) and (41), we have

$$\begin{aligned}
&\frac{1}{2}\nu^2\tilde{\Lambda}''(c) + H(c)\tilde{\Lambda}'(c) + [L + \frac{\nu^2}{2}]\tilde{\Lambda}'(c)^2 \\
= \quad & k_0\nu^2 + 2H(c)k_0c + [L + \frac{\nu^2}{2}](2k_0c)^2 \\
= \quad & k_0\nu^2 + [2h(c) + 2\frac{\nu\rho\gamma(\mu - \frac{c}{\alpha+1})}{\sigma[(1-\gamma) + \frac{\gamma(1-\rho^2)}{\xi}]}]k_0c + [\frac{\nu^2\rho^2\gamma}{(1-\gamma) + \frac{(1-\rho^2)\gamma}{\xi}} + \nu^2]2k_0^2c^2 \\
= \quad & 2k_0^2c^2\nu^2[\frac{\rho^2\gamma}{(1-\gamma) + \frac{(1-\rho^2)\gamma}{\xi}} + 1] + k_0\nu^2 + 2k_0c[\frac{\nu\rho\gamma(\mu - \frac{c}{\alpha+1})}{\sigma[(1-\gamma) + \frac{\gamma(1-\rho^2)}{\xi}]}] + 2h(c)k_0c
\end{aligned} \tag{42}$$

We know that $\exists\ \hat{c}$ s.t.

$$h(c) = h(0) + ch'(\hat{c}) < h(0) + c\kappa. \tag{43}$$

Hence,

$$\begin{aligned}
&\frac{1}{2}\nu^2\tilde{\Lambda}''(c) + H(c)\tilde{\Lambda}'(c) + [L + \frac{\nu^2}{2}]\tilde{\Lambda}'(c)^2 \\
< \quad & 2k_0^2c^2\nu^2[\frac{\rho^2\gamma}{(1-\gamma) + \frac{(1-\rho^2)\gamma}{\xi}} + 1] + k_0\nu^2 + 2k_0c[\frac{\nu\rho\gamma(\mu - \frac{c}{\alpha+1})}{\sigma[(1-\gamma) + \frac{\gamma(1-\rho^2)}{\xi}]}] + 2(h(0) + c\kappa)k_0c. \\
= \quad & c^2(k_0^2l_1 + k_0l_2) + 2k_0c[\frac{\nu\rho\gamma\mu}{\sigma[(1-\gamma) + \frac{\gamma(1-\rho^2)}{\xi}]} + h(0)] + k_0\nu^2.
\end{aligned} \tag{44}$$

So

$$
\begin{aligned}
& \frac{1}{2}\nu^2\tilde{\Lambda}''(c) + H(c)\tilde{\Lambda}'(c) + [L + \frac{\nu^2}{2}]\tilde{\Lambda}'(c)^2 + N(c) \\
< \quad & c^2(k_0^2 l_1 + k_0 l_2) + 2k_0 c[\frac{\nu\rho\gamma\mu}{\sigma[(1-\gamma) + \frac{\gamma(1-\rho^2)}{\xi}]} + h(0)] + k_0\nu^2 \\
& + \frac{1}{2}\frac{\gamma(\mu - \frac{c}{\alpha+1})^2}{\sigma^2[(1-\gamma) + \frac{(1-\rho^2)\gamma}{\xi}]} + \frac{c\gamma}{\alpha+1} - r + (1-\gamma)e^{\frac{\tilde{\Lambda}(c)}{\gamma-1}} \\
= \quad & c^2(k_0^2 l_1 + k_0 l_2) + 2k_0 c[\frac{\nu\rho\gamma\mu}{\sigma[(1-\gamma) + \frac{\gamma(1-\rho^2)}{\xi}]} + h(0)] + k_0\nu^2 \\
& + \frac{1}{2}\frac{\gamma\left\{\frac{c^2}{(\alpha+1)^2} + \left\{\mu - \sigma^2[1-\gamma+\frac{(1-\rho^2)\gamma}{\xi}]\right\}^2 - \frac{2c}{\alpha+1}\left\{\mu - \sigma^2[1-\gamma+\frac{(1-\rho^2)\gamma}{\xi}]\right\}\right\}}{\sigma^2[(1-\gamma) + \frac{(1-\rho^2)\gamma}{\xi}]} \\
& + \gamma(\mu - \frac{c}{\alpha+1}) - \frac{\gamma\sigma^2[1-\gamma+\frac{(1-\rho^2)\gamma}{\xi}]}{2} + \frac{c\gamma}{\alpha+1} - r + (1-\gamma)e^{\frac{\tilde{\Lambda}(c)}{\gamma-1}} \\
= \quad & c^2\varphi_1^2 + c\varphi_2 + \varphi_3 - r + (1-\gamma)e^{\frac{\tilde{\Lambda}(c)}{\gamma-1}}, \qquad (45)
\end{aligned}
$$

where

$$
\begin{aligned}
\varphi_1 &= k_0^2 l_1 + k_0 l_2 + l_3, \\
\varphi_2 &= 2k_0[\frac{\nu\rho\gamma\mu}{\sigma[(1-\gamma) + \frac{\gamma(1-\rho^2)}{\xi}]} + h(0)] - \frac{\gamma[\mu - \sigma^2[(1-\gamma) + \frac{(1-\rho^2)\gamma}{\xi}]]}{\sigma^2[(1-\gamma) + \frac{(1-\rho^2)\gamma}{\xi}](\alpha+1)}, \\
\varphi_3 &= k_0\nu^2 + \frac{1}{2}\frac{\gamma\{\mu - \sigma^2[1-\gamma+\frac{(1-\rho^2)\gamma}{\xi}]\}^2}{\sigma^2[(1-\gamma) + \frac{(1-\rho^2)\gamma}{\xi}]} + \gamma\mu - \frac{\gamma\sigma^2[1-\gamma+\frac{(1-\rho^2)\gamma}{\xi}]}{2}.
\end{aligned}
$$

Let $K_1(k_0) = \varphi_3 - \frac{\varphi_2^2}{4\varphi_1}$, we obtain that

$$
\begin{aligned}
& \frac{1}{2}\nu^2\tilde{\Lambda}''(c) + H(c)\tilde{\Lambda}'(c) + [L + \frac{\nu^2}{2}]\tilde{\Lambda}'(c)^2 + N(c) \\
< \quad & K(k_0) - r + (1-\gamma)e^{\frac{k_0 c^2 + K_0}{\gamma-1}} \\
< \quad & (r - K(k_0))[e^{\frac{k_0 c^2}{\gamma-1}} - 1] < 0 \qquad (46)
\end{aligned}
$$

□

By Lemmas 1 and 2 , we have Theorem 1.

Theorem 1. *There exists a classical solution of* (33) *denoted by* $\hat{\Lambda}(c)$ *such that*

$$\underline{\Lambda} \leq \hat{\Lambda}(c) \leq \tilde{\Lambda}(c). \qquad (47)$$

Proof. An ordered pair of subsolution and supersolution of (42) are obtained by Lemmas 1 and 2. Then the existence of a classical solution can be proved by Theorem 3.4 in Jin et al. (2015) [1]. □

Then, the value function, optimal debt ratio strategy and optimal dividend strategy are given as follows.

Theorem 2. *Suppose that a function* $\hat{\Lambda}(c)$ *solves* (33)*, then there exists* $\hat{Y}(c)$ *that solves* (31)*. Additionally, assume that* (36) *and* (40) *hold. Then,*

(i)

$$\hat{V}(x,c) = \hat{Y}(c)\frac{x^{\gamma}}{\gamma}, \tag{48}$$

is the value function of (20);

(ii) The optimal debt ratio and optimal dividend policies are given by

$$\begin{cases} \pi_t^* = \dfrac{G_t^* - 1}{\alpha + 1}, \\ z_t^* = \hat{Y}(c)^{\frac{1}{\gamma - 1}}, \end{cases} \tag{49}$$

where $G_t^* = \frac{(\mu - \frac{c}{\alpha+1})\hat{Y}(c) + \sigma v \rho \hat{Y}'(c)}{\sigma^2 \hat{Y}(c)[(1-\gamma) + \frac{(1-\rho^2)\gamma}{\xi}]}$.

Remark 1. *We can see that in Theorem 2, the optimal policies and the value function can be affected by the ambiguity parameter* ξ*, which means that the existence of the ambiguity can affect the optimal debt ratio and dividend policies and that insurers cannot ignore the existence of ambiguity when making their decisions.*

4. Conclusions

In the modern field of actuarial science, optimal debt ratio decisions and dividend problems are extremely important. Most of the existing works only deal with this interesting topic under the assumption of an accurate model. We investigate the optimal debt ratio and dividend payment policies for an insurer concerned about model misspecification and prove that there exists classical solutions of the optimal debt ratio, dividend payment policies, and value functions.

Author Contributions: Methodology, B.L.; Writing—original draft, D.Z.; Writing—review & editing, C.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Natural Science Foundation of China (12001267, 12301605), Youth Innovation Team of Shandong Universities (Grant No. 2022KJ174).

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

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

References

1. Jin, Z.; Yang, H.L.; Yin, G. Optimal debt ratio and dividend payment strategies with reinsurance. *Insur. Math. Econ.* **2015**, *64*, 351–363. [CrossRef]
2. Zhao, Q.; Tin, Z.; Wei, J.Q. Optimal investment and dividend payment strategies with debt management and reinsurance. *J. Ind. Manag. Optim.* **2018**, *14*, 1323–1348. [CrossRef]
3. Qian, L.Y.; Chen, L.; Tin, Z.; Wang, R.M. Optimal liability ratio and dividend payment strategies under catastrophic risk. *J. Ind. Manag. Optim.* **2018**, *14*, 1443–1461. [CrossRef]
4. Li, X.; Zhao, D.X.; Chen, X.W. Asset-liability management with state-dependent utility in the regime-switching market. *Stoch. Model.* **2023**, *39*, 566–591. [CrossRef]
5. Zhu, D.; Yin, C.C. Stochastic optimal control of investment and dividend payment model under debt control with time inconsistency. *Math. Probl. Eng.* **2018**, *2018*, 7928953. [CrossRef]
6. Zhang, J.N.; Chen, P.; Jin, Z.; Li, S.M. Open-loop equilibrium strategy for mean-variance asset-liability management portfolio selection problem with debt ratio. *J. Comput. Appl. Math.* **2020**, *380*, 11295. [CrossRef]
7. Meng, Q.B.; Bi, J.N. On the dividends of the risk model with Markovian barrier. *Commun. Stat.-Theory Methods* **2020**, *49*, 1272–1280. [CrossRef]
8. Abid, A.; Abid, F. A methodology to estimate the optimal debt ratio when asset returns and default probability follow stochastic processes. *J. Ind. Manag. Optim.* **2023**, *19*, 7735–7752. [CrossRef]

9. De Finetti, B. Su un'impostazione alternativa della teoria collettiva del rischio. *Trans. XVth Lnternational Congr. Actuar.* **1957**, *2*, 433–443.
10. Paulsen, J.; Gjessing, H.K. Optimal choice of dividend barriers for a risk process with stochastic return on investments. *Insur. Math. Econ.* **1997**, *20*, C215–C223. [CrossRef]
11. Cai, J.; Gerber, H.U.; Yang, H.L. Optimal dividends in an Ornstein-Uhlenbeck type model with credit and debit interest. *N. Am. Actuar. J.* **2006**, *10*, 94–108. [CrossRef]
12. Cheung, E.C.K.; Wong, J.T.Y. On the dual risk model with Parisian implementation delays in dividend payments. *Eur. J. Oper. Res.* **2017**, *257*, 159–173. [CrossRef]
13. Xie, J.Y.; Zhang, Z.M. Finite-time dividend problems in a Levy risk model under periodic observation. *Appl. Math. Comput.* **2021**, *398*, 125981. [CrossRef]
14. Chakraborty, P.; Cohen, A.; Young, V.R. Optimal Dividends Under Model Uncertainty. *SIAM J. Financ. Math.* **2023**, *14*, 497–524. [CrossRef]
15. Avanzi, B. Strategies for dividend distribution: A review. *N. Am. Actuar. J.* **2009**, *13*, 217–251. [CrossRef]
16. Yao, D.J.; Yang, H.L.; Wang, R.M. Optimal dividend and capital injection problem in the dual model with proportional and fixed transaction costs. *Eur. J. Oper. Res.* **2011**, *211*, 568–576. [CrossRef]
17. Yin, C.C.; Wen, Y.Z. Optimal dividend problem with a terminal value for spectrally positive Levy processes. *Insur. Math. Econ.* **2013**, *53*, 769–773. [CrossRef]
18. Bi, J.N.; Meng, Q.B. Optimal investment with transaction costs and dividends for an insurer. *RAIRO-Oper. Res.* **2016**, *50*, 845–855. [CrossRef]
19. Marciniak, E.; Palmowski, Z. On the optimal dividend problem for insurance risk models with surplus-dependent premiums. *J. Optim. Theory Appl.* **2016**, *168*, 723–742. [CrossRef]
20. Feng, Y.; Zhu, J.X.; Siu, T.K. Optimal risk exposure and dividend payout policies under model uncertainty. *Insur. Math. Econ.* **2021**, *100*, 1–29. [CrossRef]
21. Meng, H.; Siu, T.K.; Yang, H.L. Optimal insurance risk control with multiple reinsurers. *J. Comput. Appl. Math.* **2016**, *306*, 40–52. [CrossRef]
22. Jin, Z.; Xu, Z.Q.; Zou, B. A perturbation approach to optimal investment, liability ratio, and dividend strategies. *Scand. Actuar. J.* **2022**, *2022*, 165–188. [CrossRef]
23. Klebaner, F.C. *Introduction to Stochastic Calculus with Applications*, 2nd ed.; Imperial College Press: London, UK, 2008.
24. Uppal, R.; Wang, T. Model misspecification and underdiversification. *J. Financ.* **2003**, *58*, 2465–2486. [CrossRef]
25. Maenhout, P.J. Robust portfolio rules and asset pricing. *Rev. Financ. Stud.* **2004**, *17*, 951–983. [CrossRef]
26. Yi, B.; Li, Z.; Viens, F.G.; Zeng, Y. Robust optimal control for an insurer with reinsurance and investment under heston's stochastic volatility model. *Insur. Math. Econ.* **2013**, *53*, 601–614. [CrossRef]
27. Fleming, W.H.; Soner, M. *Controlled Markov Processes and Viscosity Solutions*, 2nd ed.; Springer: New York, NY, USA, 2006.

mathematics

MDPI

Article

Identifying Hidden Factors Associated with Household Emergency Fund Holdings: A Machine Learning Application

Wookjae Heo [1], Eunchan Kim [2,*], Eun Jin Kwak [3] and John E. Grable [4]

[1] Division of Consumer Science, White Lodging-J.W. Marriot Jr. School of Hospitality & Tourism Management, Purdue University, West Lafayette, IN 47907, USA; heo28@purdue.edu
[2] College of Business Administration, Seoul National University, Seoul 08826, Republic of Korea
[3] Department of Accounting and Finance, University of Wisconsin-Green Bay, Green Bay, WI 54311, USA; kwake@uwgb.edu
[4] Department of Financial Planning, Housing, and Consumer Economics, University of Georgia, Athens, GA 30602, USA; grable@uga.edu
* Correspondence: eunchan@snu.ac.kr

Abstract: This paper describes the results from a study designed to illustrate the use of machine learning analytical techniques from a household consumer perspective. The outcome of interest in this study is a household's degree of financial preparedness as indicated by the presence of an emergency fund. In this study, six machine learning algorithms were evaluated and then compared to predictions made using a conventional regression technique. The selected ML algorithms showed better prediction performance. Among the six ML algorithms, Gradient Boosting, *k*NN, and SVM were found to provide the most robust degree of prediction and classification. This paper contributes to the methodological literature in consumer studies as it relates to household financial behavior by showing that when prediction is the main purpose of a study, machine learning techniques provide detailed yet nuanced insights into behavior beyond traditional analytic methods.

Keywords: financial preparedness; emergency fund; machine learning; consumer studies

MSC: 68T07; 68T09

Citation: Heo, W.; Kim, E.; Kwak, E.J.; Grable, J.E. Identifying Hidden Factors Associated with Household Emergency Fund Holdings: A Machine Learning Application. *Mathematics* **2024**, *12*, 182. https://doi.org/10.3390/math12020182

Academic Editors: Jing Yao, Xiang Hu and Jingchao Li

Received: 24 November 2023
Revised: 29 December 2023
Accepted: 3 January 2024
Published: 5 January 2024

1. Introduction

As is the case with nearly all fields of study that fall under the area of the social sciences, much of the body of knowledge in the field of consumer studies is based on statistical results from conventional data methodological approaches, with regression procedures dominating the way researchers attempt to describe variable relationships and explain phenomena. Traditional regression techniques are designed to identify the marginal effects of specified and pre-selected factors based on theory and the existing literature. Conventional analytical techniques have been refined over the past half-century to increase explanatory power; however, even with advancements, conventional approaches remain limited in their explanatory power. Factors that might be possibly related to an outcome of interest, but have not been reported in the literature or thought to be theoretically relevant, are generally excluded from subsequent analyses. This means that the amount of explained variance across a wide number and variety of consumer studies outcomes is inevitably limited.

Big data analytical techniques, which tend to be atheoretical, have increasingly gained traction across the social sciences to acquire a deeper understanding of human attitudes and behaviors. Machine learning (ML)—a type of artificial intelligence application—is both a field of study and an umbrella term that describes algorithms that are built in such a way that hidden layers of information can be identified through a learning process based on training data and computational proofs. ML approaches are intended to supplement

the role of researchers by showing that variables that might have once been discarded in previous studies or not included at all in an empirical analysis can add insight into describing and explaining outcomes.

The purpose of this study is to illustrate the use of ML from a consumer studies perspective to improve data descriptions when compared to a conventional regression approach. The outcome of interest in this study is a household's degree of financial preparedness as indicated by the presence of an emergency fund (i.e., a measure based on household liquidity). As will be discussed later in this paper, numerous researchers have examined factors associated with holding an emergency fund, explaining the components of emergency savings, and predicting which households are most likely to meet liquidity ratio guidelines. A unique feature of much of the existing literature is that regardless of the research purpose, analysts tend to use similar variables when describing and predicting household emergency funds. These variables have come to represent the basis of many consumer-focused financial recommendations. A cursory review of this literature suggests, however, that other variables or relationships among variables is needed to gain a more comprehensive understanding of consumer financial preparedness to improve prediction rates.

When asked, financial service professionals, financial counselors, and financial educators tend to agree that managing household emergency funds involves the ongoing management of interacting variables. This is one reason why ecological systemic theory is prominently mentioned as a key explanatory model when emergency fund analyses are conducted at the household level [1,2]. As previously mentioned, much of the existing research has primarily sought to understand emergency funds within the confines of economic or financial theories using a delimited number of factors such as financial status or sociodemographic variables (e.g., [3,4]). While such studies have contributed positively to the literature by reinforcing existing theories and research findings, they may overlook the potential relevance of variables highly pertinent to how households manage emergency funds in practice. Methodologically, this signifies the need for an approach centered on pattern recognition and classification, as opposed to the identification of linear relationships upon which conventional studies have been based (e.g., [3–5]). Consequently, the combination of ecological systemic theory, pattern recognition, and classification underscores the necessity to consider complex system science models [6,7]. Furthermore, in the context of the social sciences and economics, where complex system science models are gaining acceptance, there is a need for research in personal finance utilizing ML techniques [6,8].

This study adds to the existing literature in several important ways. First, it employs ML in the context of a consumer studies topic. While some prior attempts within the field have been made (e.g., [9–15]), these efforts have been limited in their ability to compare various ML methods comprehensively. Another limitation is that some prior studies have relied on macro, rather than micro or household, data, which produce outcomes that are disconnected from a household's actual financial management activities. Consequently, this study is one of the few initial attempts to explain emergency fund management by integrating various ML techniques at the household level.

Second, previous studies have been limited to the assessment of a few central variables, including financial factors and sociodemographic factors, when studying emergency funds (e.g., [3,4]); this study is more expansive. Specifically, the analyses conducted in this study relied on a diverse set of variables that align with the research objectives. For instance, in addition to financial and sociodemographic factors, this study introduces a broad array of variables, including financial education, psychological factors, COVID-19-related factors, distance to financial service providers, and types of loans. This approach aligns well with the strengths of ML, which are designed to enhance predictive capabilities by combining numerous variables when classifying and describing relationships [16]. This study carries the potential to discover meaningful variables that have been previously unnoticed in existing research by supplementing ML predictions with additional variables potentially related to the management of emergency funds at the household level.

Third, as mentioned earlier, previous studies have typically assumed that variable relationships are linear, even when this assumption may not be practically relevant. Rather than rely on a linear assumption, this study is premised on pattern recognition and classification, distinct from models based on linear assumptions. Specifically, this study utilizes six ML algorithms as complex systems science models. While the six ML methods in this study have been widely used in empirical studies, their application in comparison to traditional linear assumption-based analytical methods is limited, particularly in relation to personal finance and consumer studies topics.

In summary, this paper contributes to the methodological literature in consumer studies by showing that when prediction is the main purpose of analysis (i.e., for use when making policy, creating education interventions, and advice giving), conventional analytical techniques may not always be the best solution. ML incorporating a larger set of variables that accounts for interactions between and among factors can offer a more robust and powerful way to increase predictive validity. In this regard, the research questions associated with this study are (a) What is the optimal ML algorithm to predict the presence of an emergency fund? (b) How do ML predictions perform when compared to a conventional logistic regression analysis? and (c) What are the most important factors associated with holding an emergency fund when viewed with an ML algorithm lens?

This study consists of sub-sections to arrive at the answer to these questions and deliver contributory points. Section 2 includes a background discussion about emergency funds and the methodological background of ML. Section 3 introduces the empirical model based on the background and methodological review. Section 4 describes the data and measurements utilized in the ML and logistic models. Section 5 illustrates the findings from each ML and the logistic model. Section 6 discusses the results. This paper concludes by describing this study's implications in Section 7.

2. Background

2.1. Household Emergency Funds

The ability of households to pay for unexpected emergencies and situations associated with unanticipated unemployment is a topic of interest to those who study and research consumer issues [17]. Household financial ratio analysis originates in consumer studies research that began in earnest in the last two decades of the 20th century. Johnson and Widdows [18] are generally given credit for being the first to adapt traditional business valuation ratios for use with households [19]. The liquidity ratio, also known as the emergency fund ratio, appears prominently in the early literature as a marker of household financial preparedness. Prather and Hanna [20] were among the first to publish standards and norms associated with the liquidity ratio, which is defined as the number of months a household can viably meet expenses in an emergency. The most commonly applied liquidity ratio formula is: Liquid Assets/(Minimum Monthly Fixed + Monthly Variable Expenses). The ratio indicates the number of months a household can weather an emergency. According to Lytton et al. [19], a household's goal should be to maintain an emergency fund equal to three months of living expenses (see also [21]). Based on this guideline, it has been estimated that less than one-third of U.S. households can adequately meet a financial emergency [22].

Gaining a unified understanding of the factors associated with holding an emergency fund that meets the liquidity ratio guideline can be complicated. Hanna et al. [23] noted that savings can be influenced directly by a household's stage in the lifecycle, which implies that the role of certain variables in describing savings patterns may differ across the lifecycle. Lifecycle theory suggests that households that expect higher income uncertainty should allocate more assets to precautionary saving [24]. Beyond anticipatory behavior, the literature also indicates that a number of personal and household characteristics are associated with an adequately funded emergency account. Bi and Montalto [22] reviewed the literature and they found age, education, income, race/ethnicity, spending behavior, risk tolerance, a willingness to borrow, holding negative economic expectations, motivation, diversification

of household income, the presence of other savings (e.g., retirement accounts), home equity, and available lines of credit provide needed information when attempting to describe who does or does not hold an emergency fund. In their study, Bi and Montalto concluded that the ability to save was more important than documenting a need to save when explaining emergency fund holdings. Others have identified factors such as financial confidence and financial knowledge as important when explaining emergency fund saving behavior.

2.2. An Introduction to Machine Learning

As the previous discussion highlights, the literature describing the characteristics associated with household emergency fund holdings has a long and robust history. Almost all previous studies that have been undertaken to describe the characteristics associated with holding emergency funds have been conducted using conventional linear-based modeling techniques. What has emerged from this literature is a common set of factors that are thought to be associated with the decision to build and maintain emergency fund assets (see [22]). An important caveat when evaluating the existing literature is the general lack of a description of the effect sizes of significant variable associations and very little discussion regarding the degree of model-explained variance. A careful examination of existing studies shows that while all the models described in the literature are statistically significant, the amount of explained variance rarely exceeds 40%. This means that other variables (or variable relationships) that have yet to be identified or used in models contribute significant explanatory power. What these variables are or how these variables interact is yet unknown.

Researchers are increasingly using ML techniques because it is now known that artificial intelligence algorithms can provide a deeper insight into the mechanisms underlying human attitudes and behaviors. ML algorithms can be used to identify what are sometimes referred to as hidden layers within data. Within these hidden layers are functions that may not be linearly related to the outcome of interest but are, nonetheless, important when viewed holistically in combination with other variables in a network [6]. A now ubiquitous example illustrates how hidden layers and networks perform in practice. In this example, assume a researcher wants to understand how people identify faces when viewed as an image. When the researcher shows study participants extracts of a subject's face (e.g., one eye, a tooth, nose), the researcher finds that these independent factors fail to reach statistical significance and thus do not provide enough information to describe a face accurately. In this example, the researcher wrongly concludes that people fail to use some visual cues when creating descriptions. What a person actually does is compile, through hidden layers of analyses, all relevant snippets of information to derive an identification. A single viewpoint cannot provide enough information to build a valid description, nor can eliminating some pieces of information improve validity. Similarly, researchers relying solely on conventional linear statistical techniques may inadvertently dismiss variables as irrelevant or unimportant when describing or predicting a social science outcome. Some researchers may dismiss potential explanatory variables altogether. Like limited pictorial extracts used when describing a face, traditional analytical techniques rarely provide more than a rough outline of an outcome or phenomenon.

This is where ML adds explanatory power beyond what can be obtained from most conventional data analysis methodologies. Kudyba and Kwaitinetz [25] and Thompson [26] described ML as improving classification by identifying patterns within large datasets. ML is generally used when a project aims to improve predictions. As with any statistical approach, the reliability of ML protocols depends on the data source and how variables are coded [27]. Numerous ML algorithms and models have been proposed and tested over the past two decades. Examples of early ML approaches include Naïve Bayes, Linear Discriminant Analysis, logistic regression, *k*-Nearest Neighbors, decision trees, Supportive Vector Machine, adaptive boosting, and Gradient Boosting methodologies. It is important to note that ML approaches do not always outperform conventional approaches. When an outcome is measured continuously, linear, polynomial, lasso, and ridge regressions sometimes provide a more robust level of prediction compared to more complex ML

techniques. According to Abiodun et al. [28], however, the sophistication of ML approaches has increased exponentially over the past decade, resulting in increasingly higher levels of reliability and robust prediction levels.

In this study, six ML algorithms are introduced and tested using the Orange package with Python [29] and then compared to predictions made using a conventional regression technique. The algorithms evaluated in this study included (a) *k*-Nearest Neighbor (*k*NN), (b) Gradient Boosting, (c) Naïve Bayes, (d) Support Vector Machine (SVM), (e) Stochastic Gradient Descent (SGD), and (f) Neural Networks (NN) (for more information about these techniques, see [28,30–32]). By comparing these six ML techniques, this study adds to the consumer studies methodology literature by illustrating how hidden connections can bring new and interesting variable associations that describe and predict consumer attitudes and behaviors to light.

2.3. Methodological Background: Machine Learning (ML) Algorithms and Their Applications in Financial and Consumer Research

As noted above, six ML algorithms were tested in this study. More than one algorithm was chosen because the literature shows that each offers unique advantages and disadvantages. A particular ML algorithm may perform well when the outcome is financial distress or bankruptcy but perform less well when applied to a credit scoring situation. The following discussion reviews the six ML algorithms tested in this study.

2.3.1. *k*-Nearest Neighbor (*k*NN)

As the name implies, *k*NN utilizes instance-based learning as a classification tool [33,34]. Instance-based learning means that the algorithm utilizes the vector space (i.e., space between objects) model, which makes *k*NN different from other classification algorithms. Because it relies on the vector space model, *k*NN can be utilized with cross-sectional data [35]. Various approaches can be used when assessing vector space [36]. When the outcome variable is categorical, Hamming distance can be utilized as shown in Equation (1):

$$Hamming\ distance = \sum_{i=1}^{I} Int(x_i \neq y_i) \quad (1)$$

where i indicates each observation; I is a set of observations i; x_i and y_i are the predictor and the outcome value with ith observation. When the outcome variable is a continuous variable, Euclidean distance, using the root of squared differences among observed samples, can be applied [37], or the Manhattan distance, using the absolute value of differences, can also be utilized as shown in Equations (2) and (3).

$$Euclidean\ distance = \sqrt{\sum_{i=1}^{I}(x_i - y_i)^2} \quad (2)$$

$$Manhattan\ distance = \sum_{i=1}^{I}|x_i - y_i| \quad (3)$$

The combination of predictors and the outcome can be shown as (x_i,y_i) where i means the ith observation from the data (i = 1, 2, 3, … I). By using ascending order of distance, the observations can be allocated on a matrix as $d(x_1,\ y_1) \leq \cdots \leq d(x_i,\ y_i)$, where d is the distance from Equations (1), (2), or (3). When the outcome variable is categorical, the most frequent occurrence indicates the highest probability of belonging to the category shown in Equation (4). By using the probability, the expected category of the outcome is the maximum value from Function (4), as indicated in Equation (5):

$$\widehat{p_k} = \frac{\sum_{i=1}^{I}(y_{i=}k)}{\underset{\sim}{i}} \quad (4)$$

$$\hat{y} = argmax\widehat{p_k} \tag{5}$$

where a predictor is a categorical variable from 1 to K, k means the kth category; $\widehat{p_k}$ is the probability to be founded; and i is observed as the optimal observation (ith). In the case that the outcome variable is a continuous variable, a certain number of observations are selected ($n = i$) from $d(x_1,\ y_1) \leq \cdots \leq d(x_I,\ y_I)$. The selected observations are utilized to calculate the inverse distance weighted average, which produces the predicted value of an outcome from Equation (6):

$$\hat{y} = \frac{\sum_{i=1}^{I} \frac{1}{d(x_i,x)} y_i}{i} \tag{6}$$

As a classification algorithm, kNN is widely used for forecasting underweighted regression conditions. When kNN is combined with fuzzy vectoring, Östermark [38] suggested that kNN can be a useful tool for detecting data outliers, specifically when forecasting using finance and economic datasets. Because of the usability of kNN when making forecasts, this classification method has been adopted in various financial studies [39]. For instance, Meng et al. [33] adopted kNN to predict internet financial risk. They found an optimal number of categories for internet financial institutions. Phongmekin and Jarumaneeroj [40] utilized various algorithms (i.e., logistic regression, decision trees, Linear Discriminant Analysis, and kNN) to forecast stock exchange returns in Thailand. They found that kNN offers the best performance when predicting stock returns.

2.3.2. Gradient Boosting

Gradient Boosting was introduced by Breiman [41], which was then merged with a regression algorithm developed by Friedman [42]. Gradient boosting is an ensemble modeling technique that combines classification and regression methods [42,43]. As the term 'boosting' implies, weak patterns from a dataset can be strengthened through a learning process when the goal is to find the highest probability of prediction [38]. 'Gradient' means an error from each strengthened stage gradually decreases until the lowest error level is reached [44]. The basic learning process begins by measuring the error (i.e., residuals) between a predicted value and an observed value [45], as shown in Equation (7), which is called a loss function:

$$l(y_i, f(x_i)) = \frac{1}{2}(y_i - f(x_i))^2 \tag{7}$$

where i is the ith observation. The negative gradient format of Equation (7) produces residuals like those in Equation (8), which is a derivative of $l(y_i,\ f(x_i))$:

$$-\frac{\delta(y_i, f(x_i))}{\delta f(x_i)} = y_i - f(x_i) \tag{8}$$

As shown in Equation (8), the negative gradient produces a function similar to that of a regression residual (i.e., the difference between the predicted outcome and the actual outcome), which is how the name Gradient Boosting originated. Until the residuals are minimized, Gradient Boosting is iterated to make weak learners be combined, as shown in Equation (9):

$$\hat{y} = f(x) = \sum_{k=1}^{K} L_k + e \tag{9}$$

where k indicates each predictor; K is the optimal number to minimize the residual; and L_k is each different weak learner. Usually, the weak learner is a tree model developed using a predictor.

In practice, there are multiple types of Gradient Boosting, including categorical Gradient Boosting, scikit-learn Gradient Boosting, Extreme Gradient Boosting, and Extreme Gradient Boosting with random forest. Categorical Gradient Boosting utilizes features as categories [46]. Scikit-learn gradient boosting is a type of Gradient Boosting algorithm offered in Python (https://scikit-learn.org/stable/ accessed on 1 November 2023), whereas

Extreme Gradient Boosting is the most recent version of Gradient Boosting [9,47]. Each method was evaluated in this study.

The use of Gradient Boosting fits well with the research of interest in this study. Gradient Boosting is an ensemble model, which makes it particularly useful when conducting finance and business analyses [10,15]. Consider the work of Zhang and Haghni [15]. They utilized Gradient Boosting to improve travel time prediction in the transportation business. Specifically, they compared autoregressive integrated moving averages, random forest, and Gradient Boosting and concluded that Gradient Boosting showed better performance prediction. Guelman [10] investigated loss costs from Canadian insurers by comparing Gradient Boosting and a generalized linear model. Gradient Boosting was found to offer better performance in terms of prediction. Gradient Boosting has also been utilized in credit analyses. For instance, Chang et al. [44] compared various ML algorithms (i.e., group method of data handling, logistic, SVM, and Extreme Gradient Boosting). They observed Extreme Gradient Boosting to have outstanding performance when predicting credit risk. The approach has also been used to predict financial distress. Liu et al. [45] compared logistic, random forest, NN, SVM, and Gradient Boosting and noted that Gradient Boosting outperformed financial distress predictions. Carmona et al. [9] found the most impactful factors associated with bank failures using Gradient Boosting. Specifically, they compared bank failure prediction performance across logistic, random forest, and Extreme Gradient Boosting. They noted that Gradient Boosting provided the most meaningful insight when understanding bank failures.

2.3.3. Naïve Bayes

As the name implies, Naïve Bayes relies on Bayes' theorem; sometimes researchers refer to the approach as Bayes or independent Bayes [48]. In practical applications, Naïve Bayes is useful for clustering and classification purposes [49]. All variables or features in a prediction model are assumed to be independent [50]. Naïve Bayes utilizes conditional probability modeling by combining various predictors ($X_k \ni x_1, x_2, \ldots, x_k$) with a set of probabilities ($p(C_m|X_k)$), where k is the number of predictors and m means the number of probabilities found. Because Naïve Bayes assumes the independence of all predictors, the maximized probability of having a certain value (or category) can be found using Equations (10) and (11):

$$p(C_m|X_i) = \frac{1}{Z} p(C_m) \prod_{k=1}^{K} p(x_k|C_m) \tag{10}$$

$$\hat{y} = argmax_{m \in \{1,\ldots,M\}} p(C_m) \prod_{k=1}^{K} p(x_k|C_m) \tag{11}$$

Some researchers have criticized the approach because the independent assumption is unnatural and unrealistic [51]. This is the reason that the approach is termed naïve. However, because of the assumption of independence, Naïve Bayes offers a mathematical transformation advantage, making the dataset analysis more predictable [51].

Naïve Bayes has been utilized in various financial studies as a classification algorithm. Jadhav et al. [12] compared the efficacy of SVM, *k*NN, and Naïve Bayes as algorithms to predict credit ratings. After comparing the algorithms, they concluded that Naïve Bayes performed best. Deng [52] utilized Naïve Bayes to detect fraudulent financial statements in auditing. Deng noted that Naïve Bayes can provide unique insights. Similarly, Viaene et al. [14] utilized Naïve Bayes to detect financial fraud (i.e., consumers' faulty insurance claims). They concluded that the approach can improve prediction rates. Naïve Bayes has also been utilized in text classifications, such as when conducting a financial news analysis. Shihavuddin et al. [53] collected news articles about the Financial Times Stock Exchange 100 (FTSE100). Using Naïve Bayes, they concluded that not only does Naïve Bayes improve classification, but the approach can also be used to predict stock prices.

2.3.4. Support Vector Machine (SVM)

SVM classification is based on the concept of a hyperplane, which combines two separate classes [30]. The easiest way to understand classification by SVM is that a hyperplane is drawn among total samples. By drawing the hyperplane, two separate groups can be identified (e.g., upper and lower hyperplanes) as shown in Equations (12) and (13):

$$y = 1,\ when\ [B\sum x_k + a] > 0 \tag{12}$$

$$y = -1,\ when\ [B\sum x_k + a] < 0 \tag{13}$$

where k means each predictor and a is the constant in each hyperplane. Because of the complexities built into most datasets, the hyperplane is generally not well specified. Therefore, SVM sets the hyperplane by considering the maximum margin, the nearest vector from the potential hyperplane [54]. To draw a hyperplane when the maximum margin is found (Max M), SVM secures the optimal prediction performance. The function is shown in Equation (14), where B and a are assumed to be 1.00:

$$Max\ M,\ where\ y_k\left(B\sum x_k + a\right) \geq M \tag{14}$$

In addition to a hyperplane and maximum margin in SVM, kernel functioning is often used to help classify samples when the dataset and vectors are highly dimensional [54]. Because one straight hyperplane cannot easily be optimally identified when the dataset is highly dimensional, different types of hyperplanes can be utilized, including linear (i.e., straight), polynomial, radial basis function (RBF), and sigmoid. These types function in the hyperplane, called a kernel [30]. In the current study, four types of kernels were utilized.

SVM has been utilized widely in credit risk studies [55]. For example, the approach has been employed to predict credit scores [56,57]. Baesens et al. [58] compared various algorithms (i.e., SVM, logistic, discriminant analysis, *k*NN, Neural Networks (NN), and decision trees) to predict credit scores. They found that SVM and NN showed the best prediction performance compared to the other algorithms. Yang [59] introduced an adaptive credit-scoring system using a kernel-based SVM. Yang noted that the non-linear feature of datasets can be managed through kernel transformation. Kim and Ahn [60] utilized various ML algorithms (i.e., multiple discriminant analysis, multinomial logistic analysis, case-based reasoning, and an artificial neural network) to examine corporate credit rates. They found that SVM outperformed in detecting multiclass classification of corporate credit ratings. Similar findings have been reported by Chaudhuri and De [61], Chen and Hsiao [62], and Hsieh et al. [63] when making bankruptcy and financial distress predictions.

2.3.5. Stochastic Gradient Descent (SGD)

SGD emerged as an extension of previous theories, including the theory of adaptive pattern classifiers [64,65]. SGD is primarily used to help with data classifications. SGD begins by minimizing the errors (i.e., residuals) between predicted and observed values [66]. Specifically, SGD employs multiple iterations to minimize the errors in each gradient step [67] using Equation (15):

$$\theta = \Theta - \eta\nabla_{\theta} J(\theta) \tag{15}$$

where θ is the parameters of all networks from predictors; $J(\theta)$ is the loss function by using θ; and η is the size of the learning rate. By repeating Equation (15), the parameters to minimize the value of the loss function can be estimated. SGD is popular because it is mathematically tractable and scalable [67]. Researchers like SGD because it helps solve optimization issues through stochastic approximation [68]. Because SGD relies on minimizing errors, regularization needs to be considered. Ridge and lasso are popular regularizations [69]. Elastic regularization can also be utilized [70]. The SGD approach can be employed when pre-selection or the transformation of explanatory variables is required and in situations where predictive machine learning scenarios are needed. The technique

showcases robustness against outliers, as the steepest gradient algorithm emphasizes the correct classification of data points closely aligned with their true labels. As such, SGD extends beyond a mere method for optimizing objective functions with appropriate smoothness properties. SGD applies to a diverse set of machine learning prediction methods (e.g., [71,72]).

Similar to the other ML algorithms, SGD has been used in various consumer and finance studies. Deepa et al. [69] utilized SGD to predict the early onset of diabetes. Compared to logistic models, SGD showed a better prediction outcome. Using SGD algorithms, they noted that SGD can be used to enhance prediction rates.

2.3.6. Neural Networks (NN)

NN is unquestionably the most mature of all algorithms within the ML area. NN offers flexibility when attempting to make classifications and when the goal of a project is to engage in future pattern recognition [25,26]. The uniqueness of NN is the approach's use of neurons as hidden layers. Neurons resemble the human brain architecture [73]. Because of the unique architecture, all inputs (i.e., features or variables) are assumed to be connected to all neurons. All neurons are also assumed to be connected to all expected outcomes [6]. The basic function of NN is shown in Equation (16):

$$y = a(\sum_{k=1}^{K} w_k x_k + e) \tag{16}$$

where k denotes the predictors; w_k is each predictor's weight; and a is e bias like the error terms. Because of the complex connectivity through neurons between inputs and outcomes, NN can be expected to improve the prediction rate of outcomes. For instance, if five variables are used as inputs to predict two outcomes, employing four neurons, then there are 20 connections between the five variables and four neurons and an additional eight connections between the four neurons and two outcomes. This interconnectedness means 160 possible pathways from the five variables to the two outcomes through the four neurons. As this example illustrates, neurons make all connectivity from inputs to outcomes so that the prediction of outcomes can be improved.

The first step when conducting an NN analysis is to define the optimal number of neurons. Because NN can employ any possible number of neurons, the number of neurons should be tested first to find the best performing model [74]. In this study, the number of neurons was first tested, and then the optimal number of neurons was employed in the final model.

As noted above, NN is a very popular ML technique. NN has been utilized to predict credit scores and other consumer behaviors. Baesens et al. [58] compared various algorithms, including SVM, logistic, discriminant analysis, *k*NN, NN, and decision trees, to conclude that SVM and NN show the best prediction performance compared to the other algorithms. Some researchers have utilized NN to detect financial fraud (e.g., fraud reporting, fraudulent use of credit cards, fraudulent financial statements, fraud claims) (e.g., [74–76]), whereas others have utilized NN for the prediction of bankruptcy and financial distress [57,77,78]. Heo et al. [11] applied NN to predict the savings-to-income and debt-to-asset ratios among U.S. households. They compared the prediction accuracy between NN and conventional regression models and found that NN provides a deeper and more meaningful insight into the savings-to-income ratio and the debt-to-asset ratio.

2.3.7. Comparison Analysis

As alluded to in the preceding discussion, it is common for researchers to check whether ML algorithms enhance predictions by comparing outcomes to the results generated from a conventional analytic tool. When the outcome variable is binary, a logistic

regression model [79] is most often the comparison. A logistic regression model can be estimated from Equation (17):

$$\ln\left(\frac{p(x)}{1-p(x)}\right) = a + \sum_{k=1}^{K} x_k \quad (17)$$

where k denotes the predictors. This approach was taken in this study. Specifically, the ML algorithms' predictions were compared to those predictions made using a maximum likelihood logistic regression.

3. Empirical Model Flow

3.1. Research Purpose and Analysis Structure

The overarching purpose of this study was to determine which modeling technique offers the best prediction rate when describing the presence of an emergency fund. As noted above, this study employed and compared various ML algorithms. A four-step analytical process was used, and the steps are described below.

Step 1: Find the best parameters across the various ML algorithms

Multiple sub-algorithms exist within nearly all ML algorithms (Naïve Baynes is an exception). For instance, in terms of *k*NN, the Euclidean method and the Manhattan method can be used to measure distance. For Gradient Boosting, four sub-algorithms are widely used: categorical, Extreme, Extreme with random forest, and scikit-learn. In the case of SVM, the kernel can be assumed to be linear, polynomial, RBF, or sigmoid. Three sub-algorithms exist for SGD (i.e., elastic, lasso, and ridge). At this step of the analytical process, each sub-algorithm was tested. For the conventional analysis (i.e., logistic regression), three types of feature selection were utilized (i.e., all variables, forward stepwise selection, and backward stepwise selection).

In addition to sub-algorithms, each ML algorithm can be affected by internal settings (i.e., parameter settings). Based on the parameter setting, the same algorithm may exhibit different degrees of performance robustness [80]. To account for this possibility, this study tested different parameters for each algorithm. For *k*NN, normally, the number of neighbors can affect classification performance. Therefore, different numbers of neighbors (i.e., from 1 to 100) were employed and compared to find the best tuning for the *k*NN algorithm. Regarding Gradient Boosting, the learning rate may affect the algorithm's performance. As such, various learning rate settings (i.e., 0.10, 0.15, 0.20, 0.25, and 0.30) were employed and compared to find the best application. For SVM, cost values are known to affect classification performance. To account for this, different cost values (i.e., 0.10, 1.00, 5.00, 10.00, 50.00, and 100.00) were employed and compared. It is also known that in terms of SGD, the learning rate may affect the algorithm's performance. To deal with this possibility, various learning rate settings (i.e., 0.001, 0.005, 0.010, 0.050, and 1.000) were employed and compared. For NN, the number of neurons can change the algorithm's performance. Therefore, different settings of neurons (i.e., 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, and 100) were utilized and compared to find the best performance outcome. As shown in Figure 1 (Part A and Part B and Line a), the first step in the analysis involved selecting the best performing sub-algorithms and the best tuning for each algorithm.

Step 2: Find the best ML prediction algorithm among the various ML algorithms

It is important to note that assuming that one specific ML algorithm will ever show a dominant performance across predictions and classifications is unrealistic. Rather, by the topical issue type and the predictive dataset's nature, diverse ML algorithms can be expected to show better/worse prediction and classification performance [27]. Given the binary feature of the dependent variable in this study, various classification algorithms were selected, as explained above. As shown in part A with line b in Figure 1, the second step in the analytical process involved finding the optimal ML algorithm from the selected six ML algorithms. The best prediction performance was selected as the most appropriate for use within the dataset.

Step 3: Check whether ML accuracies are higher than those offered by a conventional analysis

Even if a selected ML algorithm shows excellent performance across tested ML algorithms, the prediction function may actually offer a lower level of prediction when compared to a conventional analytical technique like logistic regression. Therefore, the third step involves comparing the prediction performance of the selected ML algorithm and the conventional analysis (see parts A and B with line b, Figure 1).

Step 4: Determine which factors are associated with holding an emergency fund

Assuming the selected ML algorithm performs better than the conventional analysis, the influencing rank of input factors can be found by evaluating algorithm outcomes. The influencing rank can be viewed similarly to the significant variable list from a regression model, or the rank can differ. By checking the similarity or differences between the rank of influencing factors (ML algorithm) and the significant factors (logistic regression), it is possible to establish variable importance and possible linkages across variables that can then be examined at a later date. This step in the analytical process is crucial because some variables that emerge from an ML algorithm may not be significant in a traditional sense. Therefore, as shown in Figure 1 (line c for both parts A and B), the final step involves checking the variable list generated from the ML algorithms and the logistic analysis.

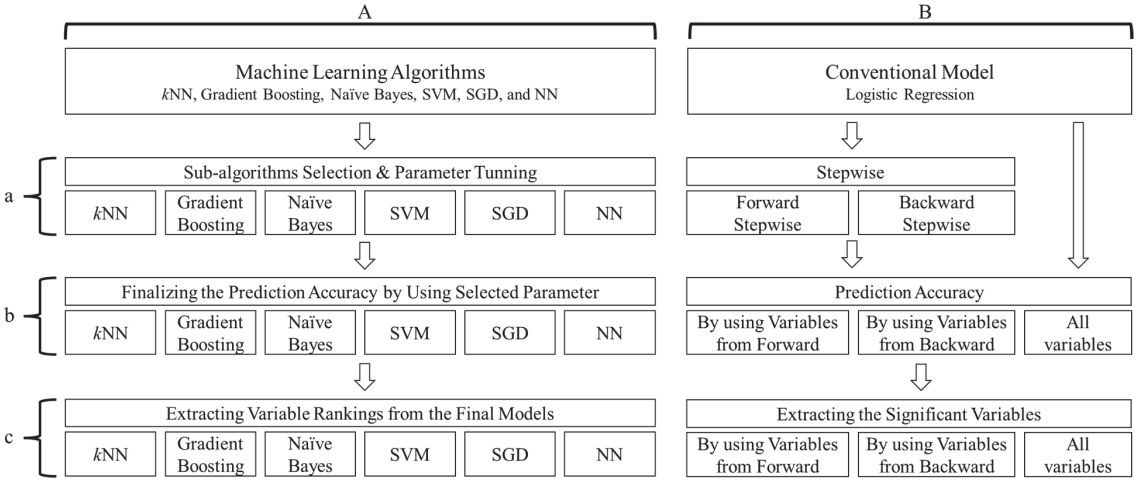

Figure 1. Analytic Structure for the Research (Abbreviations: *k*NN, *k*-Nearest Neighbor; NN, Neural Networks; SGD, Stochastic Gradient Descent; SVM, Support Vector Machine).

3.2. Analytic ML and the Conventional Analysis Process

Each ML algorithm test was conducted by dividing the sample into a training dataset and a test dataset. As shown in Figure 2, using the training dataset, each ML algorithm was used to identify the best prediction model. Data were split into training and testing datasets using a 50:50 random split ratio. As noted by Joseph [81], the split ratio varies by study and typically ranges from 80:20 division, 70:30, 60:40, and 50:50. The literature shows a conspicuous absence of definitive guidelines delineating the optimal or preferred data split ratio for a given dataset. As such, based on the comparatively small size of the dataset used in this study, the research team concluded that a 50:50 ratio was appropriate (see also [82,83]). Moreover, this ratio split allowed for robust validation of the data (i.e., k-fold validation). After a model was identified, the test dataset was utilized to validate the results from the test. If the model still showed a robust prediction outcome, the model was defined as optimal. The Python with Orange 3 visualization tool was used for all the tests. The conventional analysis utilized a similar procedure. A logistic regression model was

estimated utilizing the training dataset. Results were validated using the test dataset. Stata 17.0 was used to estimate the models.

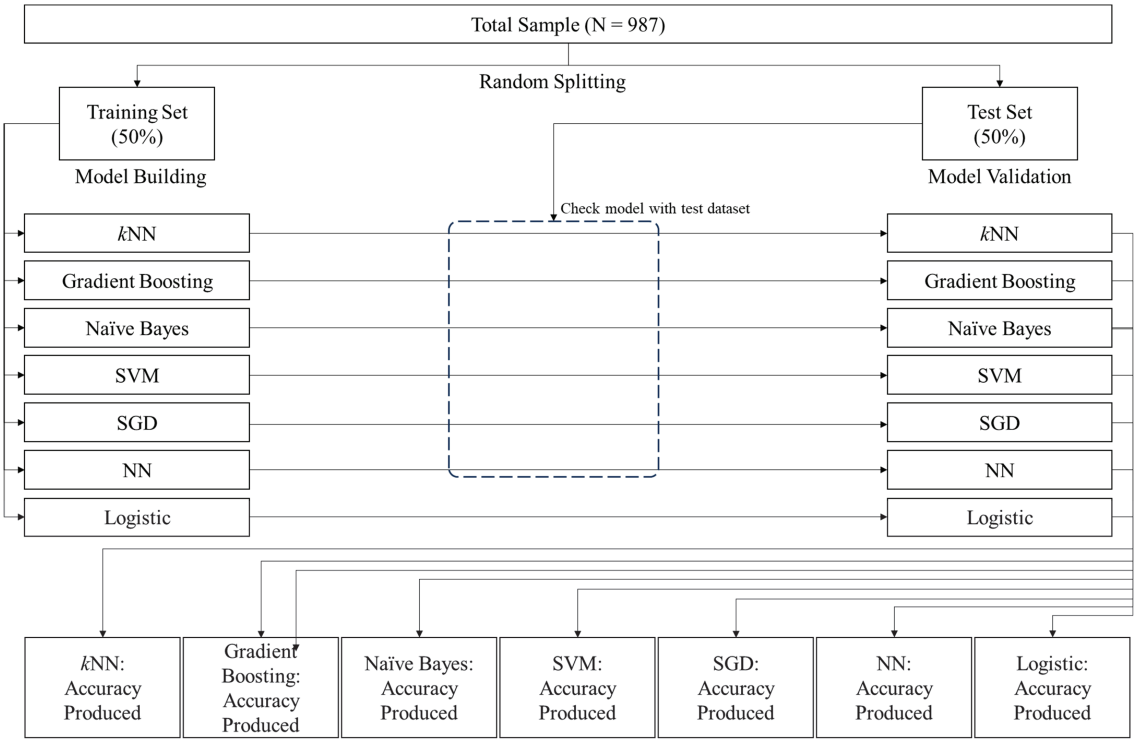

Figure 2. Analytic Process with ML Algorithms and Logistic Regression.

3.3. The Accuracy Estimation Method

To measure prediction accuracy, a receiver operator characteristics curve (ROC curve) and the area under the ROC curve (AUC) methodological approaches were utilized. An ROC curve is produced using two inputs: a true positive (TP) rate and a false positive (FP) rate [84]. The TP rate is calculated as the ratio between positive (i.e., correct) classifications and total positives. The FP rate is calculated using the ratio between negative (i.e., incorrect) classifications and total negatives. This indicates a more precise estimate when the TP rate is close to 1.00. The approach is also more precise when the FP rate is close to zero. An ROC curve shows the TP rate on the vertical axis and the FP rate on the horizontal axis. When an ROC curve shows a convex shape upward to the left, the accuracy is considered to be more precise. Additionally, the area under the curve is called the AUC, which indicates the power of the ROC (i.e., measured as 0.00 to 1.00) [44]. If the ROC curve has a vertical axis with a TP rate (i.e., zero to 1.00) and a horizontal axis with a FP rate (i.e., zero to 1.00), the area can be calculated from zero (zero times zero) to 1.00 (one times one).

3.4. The Factor Ranking Method

In Step 4, the rank of variables, in terms of prediction, is represented numerically (i.e., RReliefF). Whereas predictors in a logistic analysis can be evaluated using significance/insignificance estimates and marginal effects (i.e., coefficients), identifying high-ranking predictors using ML algorithms is more complex. For example, in the case of NN, all input variables connect to the outcome variable through neurons. Multiple weights are connected between a particular input variable and the outcome variable. There is not a specific number. As such, the evaluation of ML algorithms tends to focus on the complex

combinations of input factors and the effects of variables on an outcome variable instead of the unique association between an input variable and the outcome variable.

For this study, variable ranks were identified using RReliefF. RReliefF is an advanced version of Relief [85] and ReliefF [86], which are generally accepted attribute estimators. Relief is the baseline of RReliefF. Robnik-Šikonja and Kononenko [87] introduced RReliefF, which was developed from Relief. The diff function, as shown below, can be used to better understand the baseline of RReliefF. The diff function is used to measure the distance among instances, which can be used to identify the nearest neighbors [87]. Equation (18) is used for categorical attributes, and Equation (19) is for continuous attributes:

$$diff(A,\ I_1,\ I_2) = \begin{cases} 0; value\ (A,\ I_1) = value\ (A, I_2) \\ 1;\ otherwise \end{cases} \tag{18}$$

$$diff(A,\ I_1,\ I_2) = \frac{|value\ (A,\ I_1) - value\ (A, I_2)|}{\max(A) - \min(A)} \tag{19}$$

These equations are used when investigating a dataset that comprises multiple examples, denoted as I_1, I_2, ..., I_n, situated within an instance space. Each example is characterized by a set of attributes, represented as A_i, where attributes are associated with each example. By using the diff function, the weight (W) of attribute A can be estimated as Relief by following Equation (20) [86]

$$W[A] = P(diff.value\ of\ A \mid nearest\ \ instance\ from\ diff.class) - P(diff.value\ of\ A \mid nearest\ \ instance\ from\ same\ class) \tag{20}$$

Based on the fundamental Relief framework, regressional ReliefF was introduced using Equation (21):

$$W[A] = \frac{P(diff.response \mid diff.value\ of\ A\ and\ nearest\ intances)P(diff.value\ of\ A \mid nearest\ instances)}{P(diff.response \mid nearest\ instances)} - \frac{(1-P(diff.response \mid diff.value\ of\ A\ and\ nearest\ intances))P(diff.value\ of\ A \mid nearest\ instances)}{1-P(diff.response \mid nearest\ instances)} \tag{21}$$

Compared to other attribute estimators (e.g., the root mean of squared error and mean absolute error), the RReliefF estimator uses a factor measured by considering interactions with other factors. RReliefF measures a factor's estimator contextually. A higher RReliefF number for a specific variable indicates that the factor is expected to predict the outcome with better (optimized) performance. Therefore, in this study, RReliefF was used to check the factors' ranking.

4. Data and Measurement

4.1. Data

Data were collected in 2021 using an online survey distributed in the United States. A survey agency invited 5900 consumer households to participate in this study; 1000 respondents answered all the questions; however, 13 respondents provided inaccurate information (e.g., reporting two years old for their age), which resulted in a useable sample of 987. Descriptive information for the sample is shown in Appendix A Table A1.

4.2. Measurement

The outcome variable was whether a respondent held an emergency fund or not. The variable was coded dichotomously (Have = 1; Not have = 0) based on an answer to the following question, "*Have you set aside emergency or rainy day funds that would cover your expenses for three months, in case of sickness, job loss, economic downturn, or other emergencies?*".

The input variables (i.e., predictors) were split into the following five categories in alignment with [88] and [89]: (a) financial statements and resources, (b) financial literacy and education, (c) psychological factors, (d) demographic factors, and (e) COVID-associated factors (used to account for the period of data collection).

The following binary-coded variables comprised the financial statements and resources category: (a) have auto loan or not; (b) have student loan or not; (c) have farm loan or not; (d) have equity loan or not; (e) have mortgage loan or not; (f) own house or not; (g) have saving account or not; (h) have checking account or not; (i) own term life insurance or not; (j) own whole life insurance or not; (k) ever use payday loan; and (l) have health insurance or not. In addition, a categorical variable was included to account for the possibility of receiving financial advice for making financial decisions (i.e., 1 = have; 2 = do not know; 3 = no). Finally, a respondent's physical distance from their closest financial professional was asked and coded as follows: 1 = less than 5 miles; 2 = 5 to 10 miles; 3 = 10 to 20 miles; 4 = 20 to 50 miles; 5 = over 50 miles; and 6 = n/a or do not know.

Three variables comprised the financial literacy and education category: (a) had financial courses in high school (1 = Yes; 0 = otherwise); (b) had financial courses in college (1 = Yes; 0 = otherwise); and (c) objective financial literacy. The objective financial literacy variable was based on answers to three true/false questions [90], resulting in scores that could range from 0 (no correct answers) to 3 (all correct answers).

The psychological factors category was comprised of the following variables: (a) financial risk tolerance; (b) financial satisfaction; (c) financial stress; (d) financial self-efficacy; (e) locus of control; (f) life satisfaction; (g) the Rosenberg self-esteem scale; and (h) job insecurity. Financial risk tolerance was assessed using the Grable and Lytton's risk-taking propensity scale [91]. Scores ranged from 13 to 42. Financial satisfaction was measured using seven items on a five-point scale (min = 7; max = 35) (see [92]). Financial stress was measured using 24 items on a five-point scale (min = 24; max = 120) (see [88]). Financial self-efficacy was measured using six items, also on a five-point scale (min = 6; max = 30) (see [93]). Locus of control was measured using seven items on a five-point scale (min = 7; max = 35) (see [94]). Higher scores were representative of an external locus of control. Life satisfaction was measured using seven items on a seven-point scale (min = 5; max = 35) (see [95]). Self-esteem was measured with Rosenberg's 10-item scale that was assessed using a four-point scale (see [96]). Finally, job insecurity was measured using seven items on a five-point scale (min = 7; max = 35) (see [97]).

Demographic factors included (a) a variable representing the region of the country where a respondent lived, (b) work status, (c) agricultural working status, (d) education level, (e) marital status, (f) gender, (g) age, (h) whether a respondent lived in an urban area, (i) ethnicity, (j) income level, (k) number of children in a respondent's household, and (l) perceived health status. The region represented a respondent's state of residence. Work status was coded categorically as 1 = Full-Time; 2 = Part-Time; 3 = Self-Employed; 4 = Homemaker; 5 = Full-Time Student; and 6 = Not Working. Agriculture working status was coded as a categorical variable (1 = farm; 2 = ranch; 3 = agri-business; and 4 = not working in agriculture). Education level was coded categorically as 1 = high school or lower; 2 = some college; 3 = college; and 4 = postgraduate. Gender was coded as female or otherwise. Marital status was coded as a binary variable (i.e., single or otherwise). Age was measured in years. Living in an urban area was coded categorically as follows: 1 = urbanized area of 50,000 or more people; 2 = suburban area, near urbanized area with at least 2500 and less than 50,000 people; and 3 = rural area, all population, housing, and territory not included within any urban areas). Ethnicity was coded as a categorical variable, where 1 = White or Caucasian; 2 = Hispanic or Latino/a; 3 = Black or African American; 4 = Asian; 5 = Pacific Islander/Native American or Alaskan Native; and 6 = Other. Income level was coded categorically as 1 = Less than USD 15,000; 2 = USD 15,000 to USD 25,000; 3 = USD 25,000 to USD 35,000; 4 = USD 35,000 to USD 50,000; 5 = USD 50,000 to USD 75,000; 6 = USD 75,000 to USD 100,000; 7 = USD 100,000 to USD 150,000; and 8 = Over USD 150,000. The number of children living in a respondent's household was measured as a reported number. Finally, the perceived health status of a respondent was measured as a categorical variable (i.e., 1 = Excellent; 2 = Good; 3 = Fair; and 4 = Poor).

Finally, COVID factors were measured with items that asked how a respondent was affected by the COVID-19 virus and pandemic, how long a respondent expected the COVID-19 pandemic to last, and the receipt and timing of a stimulus check. The following items were used to evaluate perceptions of the COVID-19 pandemic: (a) how a respondent's financial situation was affected by COVID-19; (b) how a respondent's health condition was affected by COVID-19; (c) how a respondent's general well-being was affected by COVID-19; and (d) how a respondent's work–life balance was affected by COVID-19. Answers were coded as 1 = almost no impact to 4 = serious impact. Perceptions about the duration of the pandemic were assessed by asking if (a) my financial situation will get better, get worse, or stay the same in three months; (b) my financial situation will get better, get worse, or stay the same in six months; or (c) my financial situation will get better, get worse, or stay same in one year. Answers were coded as 1 = get better; 2 = get worse; or 3 = stay the same. The timing of receiving a stimulus check was measured nominally as 1 = get stimulus check in April; 2 = get stimulus check in May; 3 = get stimulus check in June; 4 = get stimulus check in July; 5 = get stimulus check after July; 6 = do not know; 7 = do not want to answer; 8 = had not received stimulus check yet; and 9 = not eligible for a stimulus check.

5. Results

5.1. Identify the Best Parameters among the Various ML Algorithms

The first step in the ML analyses began by finding the best parameters and tuning the algorithms. Across the six ML algorithms, various parameters were tested and tuned. The tuning procedure is shown in Appendix B.

5.2. Results for Step 2: Find the Best ML Prediction Method among the Various ML Algorithms

It was determined that *k*NN and NN overfit the data somewhat. For example, the prediction accuracy (AUC) of both algorithms were strong when the models were built; however, the prediction accuracy was weakened when tested. Gradient Boosting offered the best performance with categorical consideration and a learning rate of 0.10 (see Table 1). However, *k*NN and SVM were still robust. Figure 3 shows the selected algorithms' ROC curves from the six ML algorithms.

Table 1. Prediction Accuracy Comparison across ML Algorithms.

ML	Selected Algorithm	Selected Parameter	Training	Test
*k*NN		Neighbor = 6	1.000	0.844
Gradient Boosting	Categorical	L.R. = 0.10	0.988	0.849
Naïve Bayes			0.871	0.818
SVM	Sigmoid	cost = 0.10	0.836	0.826
SGD	Lasso/Ridge	L.R. = 0.001	0.919	0.802
NN		Neuron = 30	1.000	0.793

Abbreviation: L.R., learning rate.

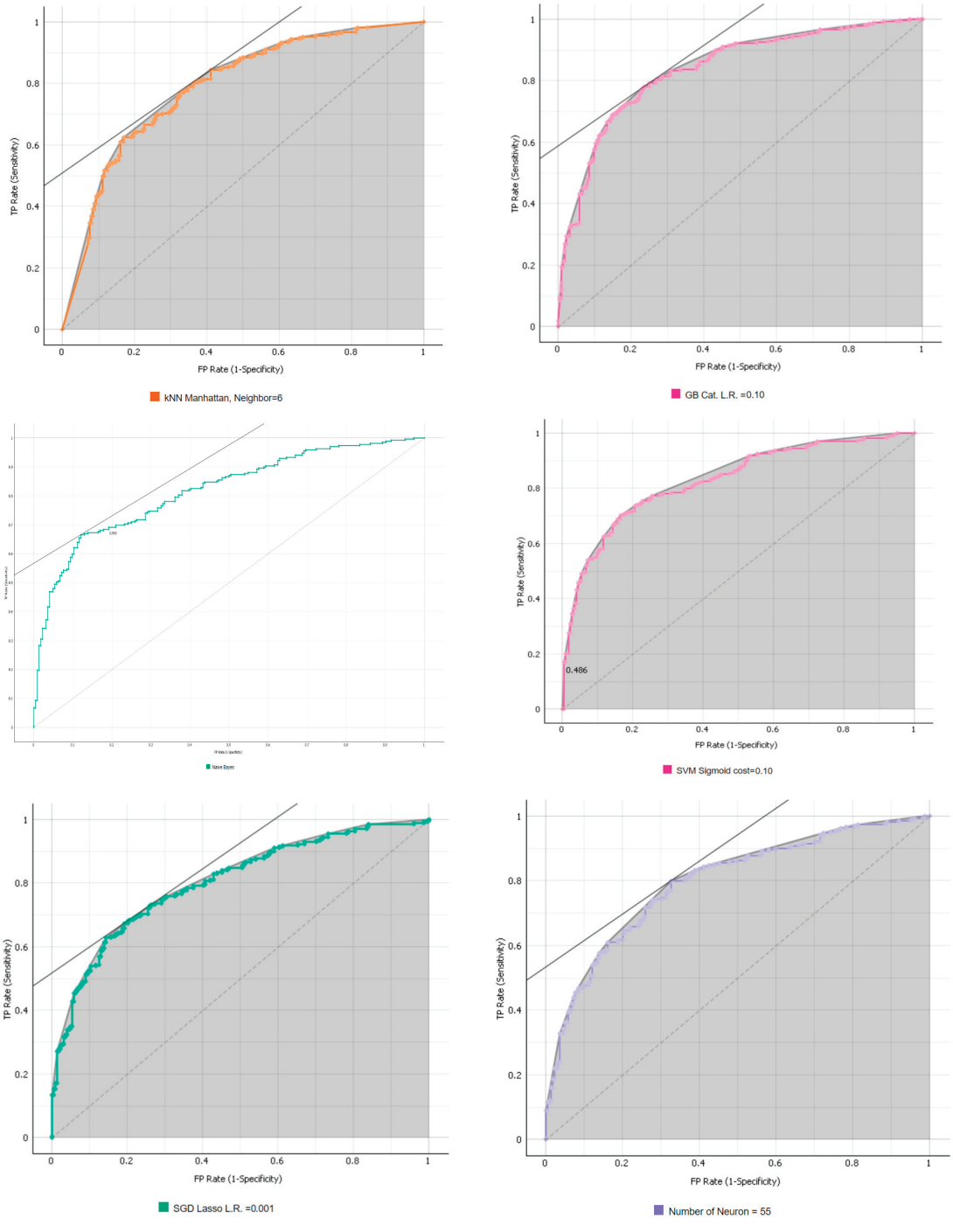

Figure 3. ROC Curves from the Best Predictions from Six ML Algorithms.

5.3. Results for Step 3: Check Whether the Accuracy of the ML Algorithms Is Higher Than the Accuracy Offered by a Logistic Regression

Table 2 shows the results from the logistic regression. As shown in Table 2, none of the variables had a significant effect in describing whether a respondent held an emergency fund. However, when the variables were added using a stepwise variable selection approach, several variables (i.e., savings account, mortgage loan, whole life insurance, no access to financial advisor, financial course in high school, financial satisfaction, financial self-efficacy, life satisfaction, number of children, and financial situation during the COVID-19 pandemic) were observed to be statistically significant.

Table 2. Logistic Regression Results (n = 475, 50% Random Splitting).

Variables	Logistic Regression with All Variables		Logistic Regression with Forward Stepwise		Logistic Regression with Backward Stepwise	
	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*
Auto loan	0.40	0.46				
Student loan	−0.61	0.47				
Farm loan	−0.04	0.80				
Equity loan	0.22	0.66				
Mortgage loan	−1.48	0.51			−0.69 *	0.31
Own house	0.50	0.51				
Saving acct.	−1.86	0.51	−1.40 ***	0.29	−1.28 ***	0.30
Checking acct.	−0.33	0.57				
Term L.I.	−0.08	0.41				
Whole L.I.	−1.02	0.51	−0.90 **	0.33	−0.81 *	0.34
FA do not know	−1.27	0.64				
FA no	−1.82	0.53	−1.12 ***	0.28	−1.14 ***	0.28
Payday loan	−0.66	0.56				
Health insurance	0.61	0.52				
FP Dist. 10 miles	0.49	0.56				
FP Dist. 20 miles	1.06	0.61				
FP Dist. 50 miles	1.08	0.91				
FP Dist. Over 50	1.40	1.19				
FP Dist. na	−0.21	0.57	−0.70 *	0.29	−0.80 **	0.30
Fin course in H.S.	−0.81	0.45	−1.01 **	0.29	−0.95 **	0.30
Fin course in Col.	−0.47	0.53				
Obj. Fin Knw.	−0.08	0.21				
Fin R.T.	0.04	0.05				
Fin Satisfaction	0.09	0.04	0.07 **	0.02	0.06 *	0.03
Fin Stress	0.02	0.01				
Fin Self-efficacy	−0.19	0.06			−0.08 *	0.03
L.O.C.	−0.05	0.05				
S.W.L.S.	0.08	0.03	0.08 ***	0.02	0.08 ***	0.02
Self-esteem	0.01	0.05				
Job insecurity	0.05	0.04				
WS Part-time	0.20	0.71				
WS Self-empl.	1.31	0.70				
WS Homemaker	−1.36	1.00				
WS Full stud.	0.28	0.82				
WS Not working	0.11	0.58				
Agri. Work	0.92	1.67				
Agri. R.Busi.	−0.77	1.02				
Agri. No.	−0.03	0.90				
Ed AA	0.44	0.50				
Ed BA	0.93	0.55				
Ed Grad.	0.66	0.74				
Single	0.22	0.45				
Female	0.12	0.41				
Age	0.02	0.02				

Table 2. *Cont.*

Variables	Logistic Regression with All Variables		Logistic Regression with Forward Stepwise		Logistic Regression with Backward Stepwise	
Suburban	0.42	0.44				
Rural	0.95	0.59				
Ethn. Hispanic	0.16	0.59				
Ethn. Black	−0.22	0.52				
Ethn. Asian	0.42	0.55				
Ethn. Pacific	−0.13	1.07				
Ethn. Others	−0.98	0.87				
Inc. 15 k to 25 k	−0.71	0.68				
Inc. 25 k to 35 k	−1.12	0.70				
Inc. 35 k to 50 k	−1.09	0.73				
Inc. 50 k to 75 k	−0.27	0.72				
Inc. 75 k to 100 k	−0.95	0.86				
Inc. 100 k to 150 k	−1.27	0.85				
Inc. > 150 k	1.26	1.27				
No. of Child	−0.66	0.20	−0.26 *	0.12	−0.27 *	0.12
Hth. Good	−0.24	0.49				
Hth. Fair	−1.17	0.68				
Hth. Poor	0.36	1.23				
Fin Situation	−0.48	0.23	−0.32 *	0.13		
H.Situation	−0.10	0.26				
WB.Situation	0.03	0.28				
Work. Situation	0.32	0.26				
3 months expect	−0.31	0.29				
6 months expect	0.14	0.27				
1 year expect	0.31	0.25				
Stim. May	0.58	0.77				
Stim. Jun.	−1.24	0.91				
Stim. Jul.	0.94	0.99				
Stim. Aft. Jul.	−0.74	0.66				
Stim. Dk	−0.72	0.72				
Stim. Na	−0.62	1.06				
Stim. No get	−1.10	0.74				
Stim. No elig.	−0.44	0.81				
Constant	8.47	3.90	3.93 ***	1.06	5.28 ***	1.25
R^2	0.54		0.41		0.41	
F	352.60		264.57 ***		268.99 ***	

Note. Reference group for auto loan, student loan, farm loan, equity loan, mortgage loan, own house, saving account, checking account, term life insurance, whole life insurance, financial course from high school, financial course from college are those who do not have them; male is the reference group for gender; ever had financial advice before is the reference group for experience of financial advice; distance to the accessible financial profession within 5 miles is the reference group for accessibility of financial professionals; full-time working status is the reference group for working status; working on a farm is the reference group for agriculture working status; high school or lower degree is the reference group for education level; living in urban area is the reference group for urban/suburban/rural living; lower than USD 15,000 is the reference group for income level; excellent health status is the reference group for health status; reference group for stimulus check is receiving stimulus check in April; the results for region (i.e., states) were omitted because the number of states and territories is too large to report while the sample size per location is too small. Significance level: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Based on a sample size of 477, ROC graphs and AUCs (i.e., predictions made from the test dataset) are shown in Figure 4. The predictions resembled convex curves. The upper left ROC was made when all variables were included in the prediction; the lower left ROC was estimated when backward stepwise was utilized; the right upper ROC was made when forward stepwise was utilized.

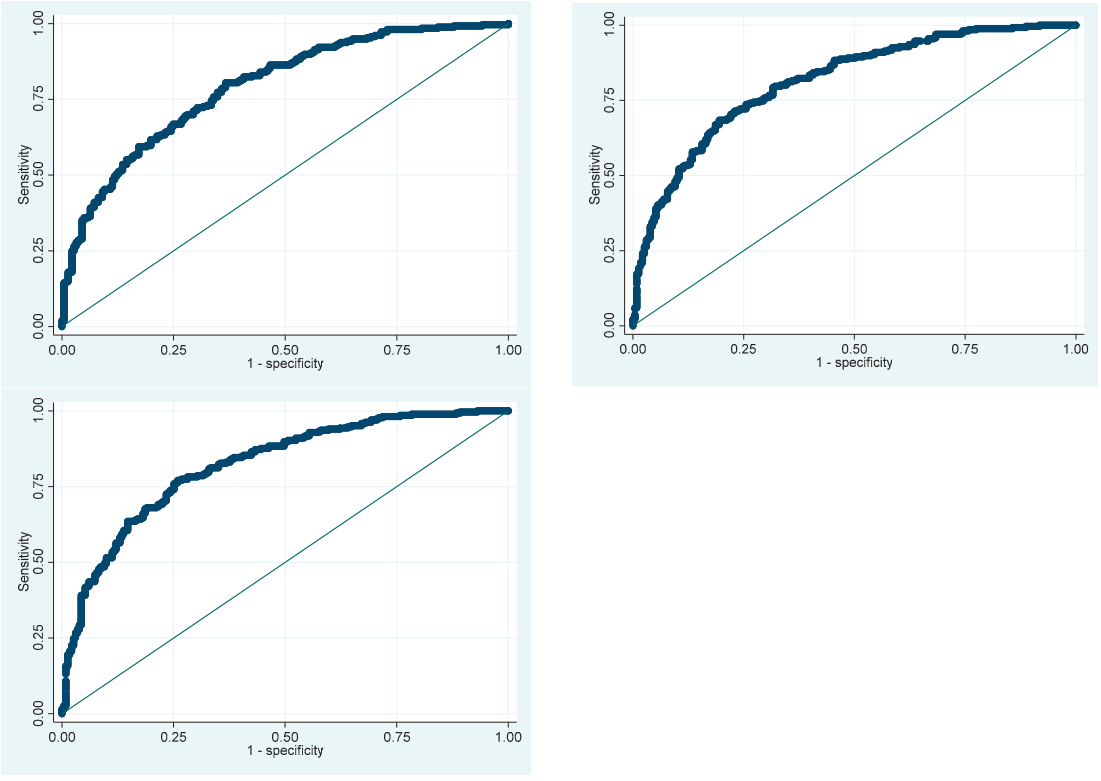

Figure 4. ROC Curves Based on Logistic Regression Modeling.

As shown in Table 3, AUC was under 0.800, which was lower than the ML AUC predictions. Even the worst performing ML exhibited a better AUC (i.e., 0.793 when ML was NN) compared to results from the logistic regression models (i.e., 0.754 when the variable list was determined via backward stepwise variable selection). This means that conventional analysis is proper when the research goal involves identifying significant variables; however, when the research goal involves maximizing prediction performance, ML algorithms provide a more robust insight into behavior (i.e., prediction accuracy can be maximized using ML techniques).

Table 3. AUC Comparison between ML Algorithms and Logistic Predictions.

ML	AUC from Test	Logistic Regression	AUC from Test
*k*NN	0.844	With all variables	0.703
Gradient Boosting	0.849	Forward stepwise	0.741
Naïve Bayes	0.818	Backward stepwise	0.754
SVM	0.826		
SGD	0.802		
NN	0.793		

Table 3 indicates that machine learning (ML) offers more (i.e., efficient) predictive performance than a logistic regression methodology. However, this does not necessarily mean that ML provides a better explanation. As previously explained, ML has the advantage of making better predictions by including more variables, as it incorporates the covariances inherent in each variable into a prediction. This means that some important features with

higher prediction weights are selected based on the covariance with other features. On the other hand, generalized linear models like logistic regression exclude covariances other than the unique covariance between an outcome and input variables. Traditional regression techniques focus on finding precise explanations for individual variables. This ultimately leads to an increase in explanatory power but a decrease in predictive power. Therefore, the results shown in Table 3 signify an improvement in the predictive power of ML but do not necessarily mean that the explanatory power of individual variables has improved.

For example, when looking at Table 2 (i.e., results from the logistic regression), variables that have a significant relationship with holding an emergency fund are easily identified. Most of these variables, including a household's financial situation, number of children, and holding a savings account, match with what has been reported in the previous literature. The explanatory power of these variables remains valid. However, Table 4 shows how different variables influenced these predictive performances. When comparing Tables 2 and 4, it becomes apparent that variables that were significant in Table 2 do not always have high predictive weights in Table 4. This indicates that in the case of the important variables shown in Table 4, various variables, as assumed by complex system science models and ecological system theory, contribute to better predictions. Therefore, the high predictive power in Table 3 and the variable rankings in Table 4 can play a role in identifying variables that conventional analyses, such as logistic regression, may overlook conceptually or theoretically. While ML may provide high predictive power, variables that were not statistically significant in the logistic regression (e.g., region, education level, financial self-efficacy, having a financial advisor, and farm loan) should be reconsidered as potentially important variables based on their high predictive weights, despite being overlooked in previous studies.

Table 4. Variable Rankings from Six ML Algorithms.

	*k*NN Accuracy Rank = 2	RF	GB Accuracy Rank = 1	RF	Naïve Bayes Accuracy Rank = 4	RF	SVM Accuracy Rank = 3	RF	SGD Accuracy Rank = 5	RF	NN Accuracy Rank = 6	RF
1	Region	0.090	Education level	0.110	Fin Self-efficacy	0.075	Fin Course in Col.	0.176	Ever FA	0.128	Fin Course in Col.	0.136
2	Equity loan	0.080	Fin Course in Col.	0.104	Farm loan	0.070	Education level	0.158	Fin Course in Col.	0.108	Farm loan	0.134
3	Farm loan	0.076	Whole L.I.	0.102	Ever FA	0.069	Whole L.I.	0.158	Fin Course in H.S.	0.080	Ever FA	0.117
4	Fin Course in Col.	0.072	Region	0.089	Checking acct.	0.062	Farm loan	0.144	Single	0.078	Equity loan	0.102
5	Fin Course in H.S.	0.070	Ever FA	0.079	Fin Satisfaction	0.057	S.W.L.S.	0.115	Fin Satisfaction	0.074	Whole L.I.	0.088
6	Single	0.064	Farm loan	0.062	Region	0.054	Fin Satisfaction	0.112	Own house	0.072	Student loan	0.086
7	Ever FA.	0.061	Fin Satisfaction	0.061	Saving acct.	0.046	Ever FA	0.109	Gender	0.070	Payday loan	0.082
8	Education level	0.060	Gender	0.056	S.W.L.S.	0.044	Fin Stress	0.101	Farm loan	0.068	Education level	0.080
9	S.W.L.S.	0.054	Single	0.054	Payday loan	0.042	Fin Course in H.S.	0.092	Fin Self-efficacy	0.061	Fin Satisfaction	0.072
10	Payday loan	0.048	Fin Self-efficacy	0.053	Income level	0.040	Payday loan	0.088	S.W.L.S.	0.058	Term L.I.	0.064
11	Term L.I.	0.040	Income level	0.051	Age	0.035	Single	0.088	Fin Stress	0.057	S.W.L.S.	0.055
12	Fin Satisfaction	0.036	Mortgage loan	0.048	1 year expect	0.033	Agri. Work. Type	0.087	Dist. To. FP	0.046	Agri. Work. Type	0.051
13	Mortgage loan	0.034	Fin Stress	0.044	Fin Stress	0.028	Fin Self-efficacy	0.081	Obj. Fin Knw.	0.045	Auto loan	0.048
14	Health status	0.032	Own house	0.042	Education level	0.028	Term L.I.	0.076	Mortgage loan	0.044	Fin Self-efficacy	0.047
15	Fin Situation	0.031	Saving acct.	0.040	Stimulus	0.027	Checking acct.	0.070	Student loan	0.040	Fin Stress	0.046
16	Gender	0.028	Dist. To. FP	0.039	Fin Course in H.S.	0.026	Own house	0.070	Term L.I.	0.040	Saving acct.	0.044
17	Auto loan	0.028	Obj. Fin Knw.	0.035	WB.Situation	0.023	Fin Situation	0.067	Payday loan	0.034	Single	0.038
18	Income level	0.025	Equity loan	0.034	Equity loan	0.022	Health status	0.066	Agri. Work. Type	0.033	Ethnic	0.032
19	Fin Self-efficacy	0.023	6 months expect	0.032	Dist. To. FP	0.021	Work status	0.065	Region	0.033	WB.Situation	0.032
20	H.Situation	0.023	S.W.L.S.	0.031	Work status	0.019	Equity loan	0.064	Job insecurity	0.032	Fin Course in H.S.	0.032
21	Student loan	0.022	Job insecurity	0.031	Agri. Work. Type	0.019	H.Situation	0.059	Equity loan	0.032	Income	0.031
22	1 year expect	0.021	Term L.I.	0.030	L.O.C.	0.019	WB.Situation	0.059	Saving acct.	0.032	Checking acct.	0.030
23	Urban type	0.020	Agri. Work. Type	0.028	Auto loan	0.016	3 months expect	0.055	L.O.C.	0.031	Self-esteem	0.028
24	Agri. Work. Type	0.019	Fin Course in H.S.	0.024	6 months expect	0.016	L.O.C.	0.047	Education level	0.030	Work. Situation	0.026
25	Self-esteem	0.017	Ethnic	0.024	Health status	0.015	Obj. Fin Knw.	0.043	Age	0.030	Region	0.026
26	Fin Stress	0.014	Health status	0.022	Term L.I.	0.014	Work. Situation	0.041	WB.Situation	0.029	Fin Situation	0.023
27	Saving acct.	0.014	Payday loan	0.022	Fin R.T.	0.012	Stimulus	0.040	Income level	0.026	L.O.C.	0.023
28	Job insecurity	0.013	Auto loan	0.022	Obj. Fin Knw.	0.011	Income level	0.040	Self-esteem	0.025	Job insecurity	0.021
29	6 months expect	0.013	H.Situation	0.020	Single	0.010	Health insurance	0.040	Work status	0.024	Gender	0.020
30	Obj. Fin Knw.	0.012	L.O.C.	0.018	Urban	0.010	6 months expect	0.039	Health status	0.023	Mortgage loan	0.016
31	Work status	0.010	Age	0.016	H.Situation	0.010	Job insecurity	0.037	Urban type	0.021	Work status	0.013
32	Age	0.009	1 year expect	0.014	Self-esteem	0.007	Age	0.034	Health insurance	0.016	Age	0.010
33	Ethnic	0.008	Self-esteem	0.012	Fin Situation	0.007	Mortgage loan	0.034	1 year expect	0.015	Health status	0.009
34	Own house	0.006	No. of Child	0.005	Own house	0.006	Self-esteem	0.032	H.Situation	0.014	Fin R.T.	0.008
35	Health insurance	0.006	Fin R.T.	0.005	Job insecurity	0.003	Region	0.031	Fin Situation	0.011	H.Situation	0.008
36	L.O.C.	0.005	Checking acct.	0.004	Work. Situation	0.003	Student loan	0.030	Checking acct.	0.010	3 months expect	0.007
37	Stimulus	0.005	Fin Situation	0.003	Student loan	0.000	Saving acct.	0.028	Fin R.T.	0.009	Dist. To. FP	0.007
38	No. of Child	0.003	Work. Situation	0.000	No. of Child	−0.001	Auto loan	0.026	Auto loan	0.008	1 year expect	0.001
39	Whole L.I.	0.000	WB.Situation	−0.004	Ethnic	−0.009	Dist. To. FP	0.010	Work. Situation	0.008	No. of Child	0.000
40	Fin R.T.	−0.002	3 months expect	−0.005	3 months expect	−0.010	No. of Child	0.009	Stimulus	0.005	Own house	0.000
41	WB. Situation	−0.003	Health insurance	−0.006	Gender	−0.012	1 year expect	0.007	6 months expect	0.004	Obj. Fin Knw.	−0.002
42	Checking acct.	−0.012	Student loan	−0.010	Health insurance	−0.014	Gender	0.004	Whole L.I.	0.004	Urban type	−0.004
43	3 months expect	−0.025	Work status	−0.021	Mortgage loan	−0.018	Ethnic	−0.002	No. of Child	−0.004	Stimulus	−0.005
44	Work. Situation	−0.029	Urban type	−0.021	Whole L.I.	−0.024	Fin R.T.	−0.004	Ethnic	−0.013	Health insurance	−0.014
45	Dist. To FP.	−0.034	Stimulus	−0.030	Fin Course in Col.	−0.024	Urban type	−0.019	3 months expect	−0.020	6 months expect	−0.017

Abbreviations: Agri. Work. Type, agricultural working status; Dist. To. FP, distance to the financial professionals; Ever FA, ever have financial advice; GB, Gradient Boosting; RF, RReliefF; other abbreviations are same as shown in Table 2.

5.4. Results for Step 4: Determine Which Factors Are Associated with Holding an Emergency Fund

Table 4 shows the ranking importance of the best fitting ML algorithm (i.e., Gradient Boosting) across the variables evaluated in this study (i.e., RReliefF). Education level and having completed a financial course while in college ranked highly. This implies that educational attainment is important in helping someone gauge the need for an emergency fund. In addition, this indicates that promoting financial education, both in formal academic settings and through specialized courses, can be an effective strategy when encouraging individuals to (a) recognize the importance of emergency funds and (b) take proactive steps to establish emergency savings. Policy makers and educators should consider expanding financial education programs to enhance financial preparedness.

In addition, some financial-related psychological factors (i.e., financial satisfaction, financial self-efficacy, and financial stress) were found to be important. This implies that these factors are associated with holding an emergency fund. Financial institutions, financial service providers, and financial educators should incorporate psychological aspects into their financial literacy and counseling programs. Fostering financial satisfaction and self-efficacy while addressing financial stress is likely to help individuals develop positive emergency fund attitudes and behaviors.

Interestingly, COVID-19-related factors were not particularly important predictors in the model. This suggests that households are unlikely to change their emergency fund saving behavior even in the context of situational influences like a challenging economic situation.

Although Gradient Boosting was deemed to be the best model, the other ML algorithms produced comparable results. For instance, owning whole life insurance was an important variable when describing who holds an emergency fund across the model. This indicates that those who own whole life insurance are more concerned about their future self and the financial welfare of other household members (i.e., individuals who own life insurance generally exhibit a heightened awareness of their long-term financial security and the financial well-being of their family). Financial service providers can use this insight to emphasize the importance of comprehensive financial planning, including both insurance and emergency fund considerations. Similarly, educational factors (i.e., education level, completing a financial course in high school, or a financial course while in college) were found to be important predictors across the ML algorithms.

The ML results differed in significant ways from the logistic regression estimates. Compared to the Gradient Boosting model, taking a financial course in college and financial stress were unimportant in the logistic regression. Even so, there were some similarities. For instance, owning whole life insurance, taking a financial course in high school, and financial satisfaction ranked highly across the models. This indicates theoretical connections between these variables and holding an emergency fund. This study illustrates that combining insights from different analytical approaches can lead to a more comprehensive understanding and effective promotion of emergency fund savings.

6. Discussion

ML and big data analytical techniques have, over the past decade, garnered increasing attention among researchers, educators, and policy makers as a way to obtain deeper insight into social science phenomena. This study adds to the growing consumer studies methodological literature by illustrating how ML techniques can be applied to assessing household consumer attitudes and behaviors and how ML methods can improve prediction rates.

The outcome variable in this study was whether a household held an emergency fund, which was used to indicate a household's degree of financial preparedness. The existing financial ratio literature is relatively consistent in reporting that those who hold emergency savings share a common demographic profile [3,4]. They tend to have high income, are more educated, and have greater wealth. It is important to note, however, that nearly all profiles reported in the literature were constructed using traditional methodologies,

primarily regression techniques. At the outset of this paper, it was hypothesized that while existing profiles may remain valid, other variables might also be influential in describing who does and does not hold emergency savings. Traditional regression modeling techniques do not account for hidden layers between and among variables. While it is possible to create moderation and mediation models, to do so with large data is nearly impossible when the constraints associated with regression modeling are applied. This study's methodological approach dealt with this issue by showing that when prediction or profiling is the main purpose of a study, ML algorithms can provide a more nuanced insight into consumer behavior compared to more commonly used statistical analysis techniques [7,16].

This study compared and tested several ML algorithms to determine which offers the most robust prediction rate. The ML algorithm outputs were compared to estimates derived from logistic regression models. Several takeaways emerged from these analyses. First, those using ML techniques must know that parameter tuning is not optional. Incorrect parameter tuning results in lowered prediction and classification rates. Those who adopt ML algorithms in consumer studies should consider this point and compare tuning performance when conceptualizing studies. Second, sub-algorithms should be considered. Using an incorrect sub-algorithm will almost always lower prediction and classification validity. Third, when evaluating ML algorithm outputs, it is important to remember that ML algorithms do not show marginal effects. Instead, ML algorithms provide a ranked ordering of predictors. As such, the interpretation of an ML analysis should not be considered deterministic. Instead, the interpretation of an ML output needs to be conceptualized as more in line with an explorative introduction.

In this study, Gradient Boosting, *k*NN, and SVM were found to provide the most robust degrees of prediction and classification. Gradient Boosting offered the best prediction rate, which aligns with what others have reported in the literature (e.g., [9,10,15,44]). Gradient boosting is an ensemble modeling technique that integrates classification and regression methods [42,43]. The ensemble of classification and regression estimation works well when optimizing prediction accuracy [31] and minimizing error levels [44]. What is particularly interesting in this study is that income and wealth—factors generally considered the most descriptive of financial preparedness—were not highly ranked in the Gradient Boosting algorithm, nor with *k*NN or SVM. This insight differs from what is generally shown using regression techniques [3]. However, educational factors and the existence of financial obligations were more important. It appears that a consumer must possess the financial literacy to anticipate the need for emergency savings, formulate a plan to build an emergency fund, and implement the plan. The consumer must also have an objective reason to hold emergency fund assets. The existence of loans is one reason a consumer may opt to hold assets in an emergency fund. Likewise, a consumer needs to hold an attitudinal disposition that values one's future self or the well-being of household members. The consistently high ranking of life insurance in the ML algorithms suggests that the ability to plan for the future is an important characteristic among those holding emergency fund assets. The region variable in the *k*NN model is worthy of future research. The variable represents the state where a respondent resided at the time of the survey. It appears that some consumers are more likely than others to take financial preparedness steps. Specifically, those living in rural areas who also hold existing debt, are predicted to be more likely to hold an emergency fund.

This study represents a noteworthy advancement in consumer studies literature, particularly in the domains of personal finance and financial planning. This paper illustrates the value of ML techniques when predicting behavior. While numerous researchers have utilized ML methodologies with social science datasets (e.g., [9–15]), these efforts have sometimes suffered from limitations, such as their inability to comprehensively compare diverse ML methods or their focus on non-household factors. This means that the practical relevance of findings about household financial management has notable limitations. This

paper is one of the few studies to comprehensively analyze the nuances associated with holding an emergency fund at the household level.

Another significant contribution of this paper is the expanded scope of variables that were used to predict holding an emergency fund. Rather than rely on a limited set of preexisting variables as described in the literature (i.e., primarily financial factors and sociodemographic attributes) (e.g., [3,4]), this study introduced a broader range of variables, including financial education, psychological aspects, COVID-19-related factors, distance to financial service providers, and holding various types of loans. This approach aligns well with ML's capacity to leverage multiple variables [16], potentially unveiling overlooked variables that could significantly contribute to understanding the dynamics of emergency fund management.

Moreover, this study departs from the prevailing practice of assuming linear relationships between and among variables. The ML technique uses a pattern recognition and classification approach, making it possible to move beyond traditional linear assumptions. To achieve this, six distinct ML algorithms were employed as complex systems science models. The application of these algorithms allowed for a comprehensive investigation of the potential contributions of ML to the field of consumer studies. Notably, each ML algorithm underwent meticulous parameter tuning and calibration, extending beyond algorithmic utilization to demonstrate the application of ML techniques to address complex questions. The comprehensive approach in this study underscores the commitment to advancing the understanding of emergency fund management dynamics and enhancing the practical applicability of ML in consumer studies.

In summary, the results from this study advance the methodological body of literature for those working in the consumer studies field. This study shows that ML algorithms can be used to improve predictions and classifications of consumer attitudes and behaviors. Future research should align the results from this study with existing models and profiles of those who hold emergency savings. Information from such studies can be used by financial educators, consumer advocates, and policy makers when helping households achieve greater levels of financial preparedness.

7. Conclusions

This study is noteworthy in making significant theoretical, practical, and methodological contributions to consumer studies. The theoretical contribution lies in its application of ML techniques to the study of household financial decision making. Unlike traditional linear models, this study used a pattern recognition and classification methodology, shedding light on the intricate complexities underlying emergency fund management. The findings from this study challenge conventional beliefs by highlighting the importance of financial literacy, financial obligations, and a positive attitude towards future financial well-being as key factors in predicting who is more likely to hold emergency savings, with income and wealth taking a secondary role.

On a practical level, findings from this study underscore the critical importance of parameter tuning and sub-algorithm selection when employing ML techniques in consumer studies. This paper offers valuable insights into the use of ML algorithms when predicting and classifying consumer attitudes and behaviors, which can have direct applications for financial service providers, financial educators, consumer advocates, and policy makers. Moreover, this study expands the spectrum of variables considered, incorporating financial education, psychological factors, COVID-19-related variables, and others, thereby enhancing the predictive capacity of models to understand the dynamics of emergency fund management.

Even in the context of these significant contributions, limitations need to be acknowledged. ML techniques, while improving prediction rates, do not readily provide straightforward marginal effects. Thus, some researchers use ML algorithms as a starting point in identifying key variables for use in secondary models. While this study evaluated six robust ML algorithms, including Gradient Boosting, *k*NN, and SVM, further research is needed

to determine when one particular approach should be used to address a specific research question. Further advanced ML algorithms, such as Generative Adversarial Network, Recurrent Neural Network, or Convolutional Neural Network, should be evaluated in future studies. In the context of this study, additional research is needed to decipher regional variations in holding an emergency fund. Future studies should also aim to integrate the findings with existing models and profiles of emergency savings holders. Doing so will contribute to a better understanding of the financial preparedness of households. In addition, the current ML algorithms are all well-known algorithms. Even in the context of these limitations and opportunities for future work, this study advances the consumer studies methodological landscape by showcasing how ML techniques can enrich the field's comprehension of consumer attitudes and behaviors, particularly within the context of holding an emergency fund.

Author Contributions: Conceptualization, W.H. and E.K.; methodology, W.H. and E.K.; software, W.H. and E.K.; validation, W.H., E.K., E.J.K. and J.E.G.; formal analysis, W.H.; investigation, E.K.; data curation, E.J.K.; writing—original draft preparation, W.H., E.K., E.J.K. and J.E.G.; writing—review and editing, W.H., E.K., E.J.K. and J.E.G.; supervision, W.H. and J.E.G.; funding acquisition, W.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the USDA National Institute of Food and Agriculture, Hatch project 1017028 and Multistates project 1019968.

Data Availability Statement: The research dataset can be obtained upon a proper request.

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

Appendix A

Table A1. Descriptive Table (N = 987).

Category	Variable	Frequency	Percentage	Mean	SD
Outcome	Em. Fund (=Have)	538	54.51%		
	Auto loan (=Have)	355	35.97%		
	Student loan (=Have)	307	31.10%		
	Farm loan (=Have)	156	15.81%		
	Equity loan (=Have)	181	18.34%		
	Mortgage loan (=Have)	320	32.42%		
	Own house	487	49.34%		
	Saving acct.	650	65.86%		
	Checking acct.	807	81.76%		
	Term L.I.	418	42.35%		
	Whole L.I.	289	29.28%		
Financial Factors	FA have	330	33.43%		
	FA do not know	143	14.19%		
	FA no	514	52.08%		
	Payday loan	274	27.76%		
	Health insurance	776	78.62%		
	FP Dist. 5 miles	216	21.88%		
	FP Dist. 10 miles	229	22.29%		
	FP Dist. 20 miles	140	14.18%		
	FP Dist. 50 miles	67	6.79%		
	FP Dist. Over 50	44	4.46%		
	FP Dist. na	300	30.40%		
	Fin course in H.S. (=Have)	363	36.78%		
Financial Education	Fin course in Col. (=Have)	296	29.99%		

Table A1. *Cont.*

Category	Variable	Frequency	Percentage	Mean	SD
	Obj. Fin Knw.			1.56	1.00
	Fin R.T.			22.70	4.71
	Fin Satisfaction			22.54	7.31
	Fin Stress			66.95	27.71
Psych.	Fin Self-efficacy			15.59	5.22
Factors	L.O.C.			18.57	6.27
	S.W.L.S.			21.56	8.73
	Self-esteem			28.38	5.05
	Job insecurity			19.69	4.55
	WS Full-time	396	40.12%		
	WS Part-time	93	9.42%		
	WS Self-empl.	80	8.11%		
	WS Homemaker	59	5.98%		
	WS Full stud.	78	7.90%		
	WS Not working	281	28.47%		
	Agri. Farm	113	11.45%		
	Agri. Ranch	21	2.13%		
	Agri. R.Busi	66	6.69%		
	Agri. No	787	79.74%		
	Ed High	279	28.27%		
	Ed AA	269	27.25%		
	Ed BA	269	27.25%		
	Ed Grad.	170	17.22%		
	Single	503	50.96%		
	Female	501	50.76%		
	Age			38.86	15.29
	Urban	419	42.45%		
	Suburban	396	40.12%		
Demo.	Rural	172	17.43%		
Factors	Ethn. White	357	36.17%		
	Ethn. Hispanic	135	13.68%		
	Ethn. Black	250	25.33%		
	Ethn. Asian	149	15.10%		
	Ethn. Pacific	38	3.85%		
	Ethn. Others	58	5.88%		
	Inc. < 15 k	175	17.73%		
	Inc. 15 k to 25 k	118	11.96%		
	Inc. 25 k to 35 k	138	13.98%		
	Inc. 35 k to 50 k	127	12.87%		
	Inc. 50 k to 75 k	148	14.99%		
	Inc. 75 k to 100 k	98	9.93%		
	Inc. 100 k to 150 k	110	11.14%		
	Inc. > 150 k	73	7.40%		
	No. of Child			0.74	1.08
	Hth. Excellent	280	28.37%		
	Hth. Good	468	47.42%		
	Hth. Fair	190	19.25%		
	Hth. Poor	49	4.96%		
	Region	-	-		

Table A1. *Cont.*

Category	Variable	Frequency	Percentage	Mean	SD
	Fin Situation			2.33	1.08
	H.Situation			2.00	1.05
	WB.Situation			2.29	1.07
	Work. Situation			2.27	1.09
	3 months expect			2.06	0.90
	6 months expect			1.91	0.89
	1 year expect			1.72	0.88
C-19 Factors	Stim. Apr.	164	16.62%		
	Stim. May.	101	10.23%		
	Stim. Jun.	78	7.90%		
	Stim. Jul.	61	6.18%		
	Stim. Aft. Jul.	159	16.11%		
	Stim. Dk	133	13.48%		
	Stim. Na	39	3.95%		
	Stim. No get	129	13.07%		
	Stim. Not elig.	123	12.46%		

Abbreviation: Em. Fund, emergency fund; acct., account; L.I., life insurance; FA have, ever have financial advice; FA do not know, not knowing whether have financial advice; FA no, never have financial advice; FP Dist. 5 miles, financial professionals are accessible within 5 miles; FP Dist. 10 miles, financial professionals are accessible within 10 miles; FP Dist. 20 miles, financial professionals are accessible within 20 miles; FP Dist. 50 miles, financial professionals are accessible within 50 miles; FP Dist. Over 50, financial professionals are accessible over 50 miles; FP Dist. na, the accessibility of financial professionals is not known; Fin course in H.S., financial course from high school; Fin course in Col., financial course from college; Obj. Fin Knw., objective financial knowledge; Psych. Factors, psychological factors; Fin R.T., financial risk tolerance; Fin Satisfaction, financial satisfaction; Fin Stress, financial stress; Fin Self-efficacy, financial self-efficacy; L.O.C., locus of control; S.W.L.S., satisfaction with life scale; Demo., demographic; WS Full-time, working status as full-time worker; WS Part-time, working status as part-time worker; WS Self-empl., working status as self-employed; WS Homemaker, working status as homemaker; WS Full stud., working status as full- time student; WS Not working, working status as not working; Agri. Farm, working in agriculture as farm worker; Agri. Ranch, working in agriculture as ranch worker; Agri. R.Busi., working in agriculture as rural business; Agri. No., not working in agriculture; Ed High, education level as high school or lower; Ed AA, some college with associate degree; Ed BA, college with Bachelors' degree; Ed Grad., education level as graduate or higher degree; Ethn. White, ethnic group as White or Caucasian; Ethn. Hispanic, ethnic group as Hispanic or Latino(a); Ethn. Black, ethnic group as black or African American; Ethn. Asian, ethnic group as Asian; Ethn. Pacific, ethnic group as Pacific Islander, Native American, or Alaskan Native; Ethn. Others, ethnic group as others; Inc. < 15 k, income level as lower than USD 15,000; Inc. 15 k to 25 k, income level between USD 15,000 and USD 25,000; Inc. 25 k to 35 k, income level between USD 25,000 and USD 35,000; Inc. 35 k to 50 k, income level between USD 35,000 and USD 50,000; Inc. 50 k to 75 k, income level between USD 50,000 and USD 75,000; Inc. 75 k to 100 k, income level between USD 75,000 and USD 100,000; Inc. 100 k to 150 k, income level between USD 100,000 and USD 150,000; Inc. > 150 k, income level over USD 150,000; # Child, number of children in a household; Hth Excellent, health status as excellent health status; Hth Good, health status as good health status; Hth Fair, health status as fair health status; Hth Poor, health status as poor health status; C-19 Factors, COVID-19 factors; Fin Situation, the financial situation affected by COVID-19; H.Situation, the health situation affected by COVID-19; WB.Situation, general well-being affected by COVID-19; Work. Situation, work-balance affected by COVID-19; 3 months expect, the expected financial situation in 3 months; 6 months expect, the expected financial situation in 6 months; 1 year expect, the expected financial situation in 1 year; Stim. Apr., getting stimulus check in April; Stim. May., getting stimulus check in May; Stim. Jun., getting stimulus check in June; Stim. Jul., getting stimulus check in July; Stim. Aft. Jul., getting stimulus check after July; Stim. Dk, do not know whether get stimulus check or not; Stim. Na, do not want to answer; Stim. No get, the respondent did not get stimulus check; Stim. Not elig., the respondent is not eligible to get stimulus check.

Appendix B

ML Tuning: Identify the Best Parameters among the Various ML Algorithms

Tables A2–A7 and Figures A1–A6 show each ML algorithm's accuracy given the constraints of each algorithm's settings. In the case of *k*NN, both the Euclidean and Manhattan models showed robust predictions in the training dataset. However, when the models were checked using the test dataset, the Manhattan distance algorithm exhibited a better prediction rate. Regarding parameter tuning, the Manhattan model showed the best performance when there were three to eight neighbors. It was determined that the best parameter distance was six (6).

Table A2. Algorithm and Parameter Selection—*k*NN.

	Training		Test	
Number of Neighbors	**Euclidean AUC**	**Manhattan AUC**	**Euclidean AUC**	**Manhattan AUC**
1	1.000	1.000	0.686	0.835
2	1.000	1.000	0.742	0.834
3	1.000	1.000	0.754	0.840
4	1.000	1.000	0.775	0.840
5	1.000	1.000	0.779	0.840
6	1.000	1.000	0.785	0.844
7	1.000	1.000	0.786	0.838
8	1.000	1.000	0.786	0.842
9	1.000	1.000	0.786	0.838
10	1.000	1.000	0.794	0.836
20	1.000	1.000	0.809	0.828
30	1.000	1.000	0.810	0.825
40	1.000	1.000	0.807	0.818
50	1.000	1.000	0.811	0.809
60	1.000	1.000	0.802	0.806
70	1.000	1.000	0.803	0.799
80	1.000	1.000	0.801	0.776
90	1.000	1.000	0.799	0.708
100	1.000	1.000	0.795	0.834

Note. AUC represents the prediction accuracy of the model. AUC ranges in value from 0.00 to 1.00, and the higher the AUC, the better the model predicts. Abbreviation: AUC, area under the curve.

Figure A1 shows the representative ROC curves for *k*NN. The upper left graph is the ROC graph for the Euclidean model with 30 neighbors; the left lower graph is the ROC graph for the Euclidean model with 50 neighbors; the right upper graph is the ROC graph for Manhattan model with six neighbors; the lower right graph is the ROC graph for Manhattan model with eight neighbors. The dark section under the curve is the area used to calculate AUC. As shown in Figure A1, the ROC curves were convex, indicating that *k*NN performed well in prediction. The AUC was maximized when *k*NN was performed using the Manhattan model with six neighbors.

In the case of Gradient Boosting, the four sub-algorithms exhibited prediction robustness with the training dataset. However, when the algorithms were checked using the test dataset, categorical Gradient Boosting showed better prediction. Regarding parameter tuning, categorical Gradient Boosting showed the best performance when the learning rate was 0.10, as shown in Table A3.

Table A3. Algorithm and Parameter Selection—Gradient Boosting.

	Training				Test			
L.R.	**Cat. AUC**	**Ext. AUC**	**Ext. RF AUC**	**Scikit AUC**	**Cat. AUC**	**Ext. AUC**	**Ext. RF AUC**	**Scikit AUC**
0.10	0.988	1.000	1.000	0.968	0.849	0.842	0.842	0.836
0.15	1.000	1.000	1.000	0.981	0.835	0.840	0.840	0.838
0.20	0.998	1.000	1.000	0.985	0.827	0.840	0.840	0.842
0.25	1.000	1.000	1.000	0.991	0.838	0.834	0.834	0.833
0.30	0.999	1.000	1.000	0.994	0.833	0.838	0.838	0.829

Abbreviation: Cat., Categorical Gradient Boosting; Ext., Extreme Gradient Boosting; Ext. RF, Extreme Gradient Boosting with random forest; L.R., learning rate; Scikit, Scikit version of Gradient Boosting.

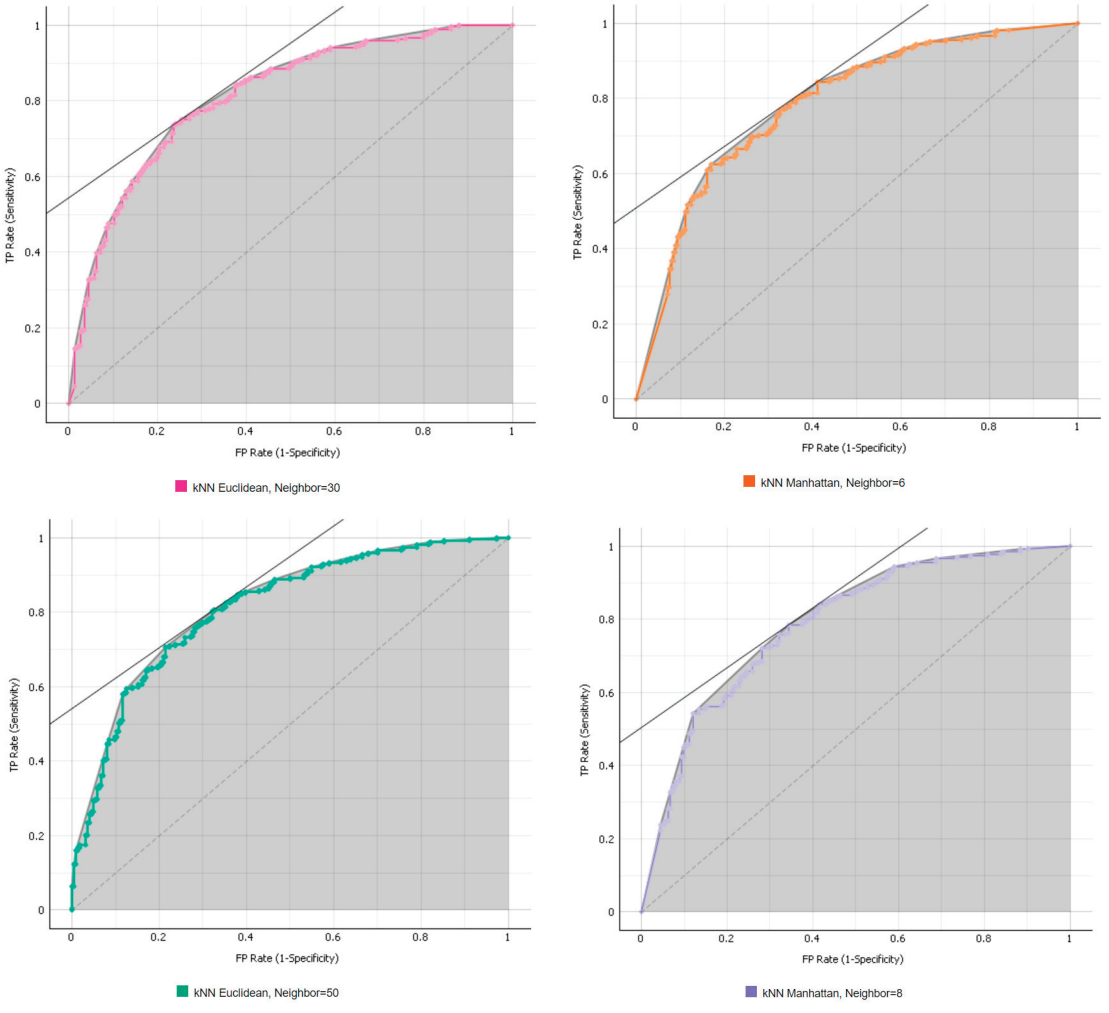

Figure A1. ROC Curves for Algorithm and Parameter Selection—*k*NN.

Figure A2 shows the representative ROC curves for the Gradient Boosting algorithms. The upper left graph is the ROC illustration for Categorical Gradient Boosting with a learning rate of 0.10; the left lower graph is the ROC graph for Extreme Gradient Boosting with a learning rate of 0.10; the right upper graph is the ROC graph for Extreme Gradient Boosting with random forest with a learning rate of 0.10; the lower right graph is the ROC graph for Scikit Gradient Boosting with a learning rate of 0.10. AUC was calculated using the dark area under the curve. As shown in Figure A2, the ROC curves were convex, suggesting that prediction was robust with Gradient Boosting. The AUC was the largest when Categorical Gradient Boosting was performed with a learning rate of 0.10.

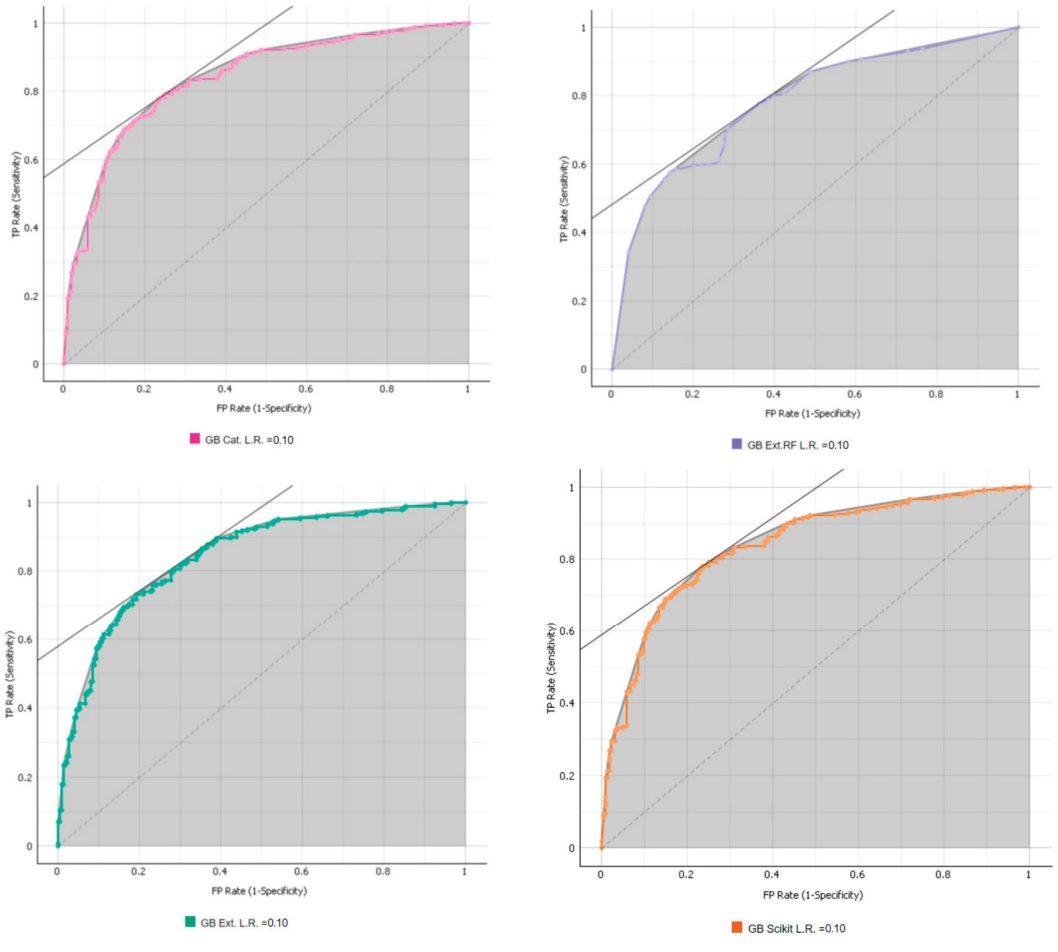

Figure A2. ROC Curve for Algorithm and Parameter Selection—Gradient Boosting.

There are no comparable sub-algorithms and parameter tuning estimates in the case of Naïve Bayes. Table A4 and Figure A3 show the Naïve Bayes' AUC and ROC curves. The dark area under the curve is the area used to estimate AUC.

Table A4. Algorithm and Parameter Selection—Naïve Bayes.

Training AUC	Test AUC
0.871	0.818

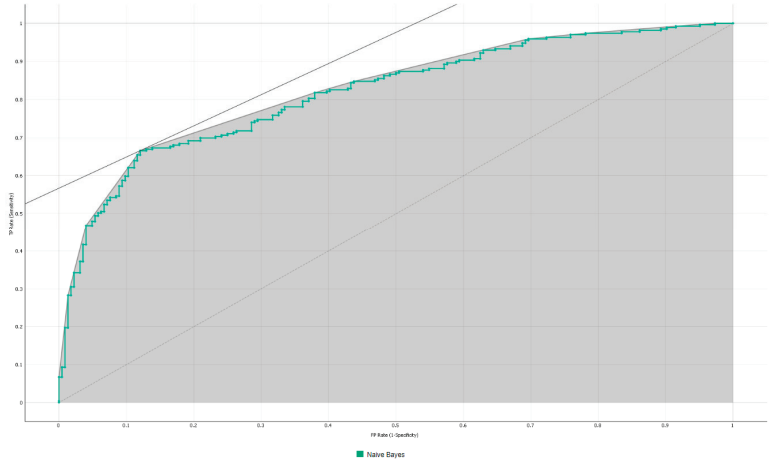

Figure A3. ROC Curve for Algorithm and Parameter Selection—Naïve Bayes.

Table A5 shows the Support Vector Machine (SVM) algorithm accuracy. In the case of SVM, the Radial Basis Function (RBF) kernel model exhibited the best prediction with the training dataset. As an optimal parameter setting, the cost was set between 5 and 100. However, when the algorithm was checked using the test dataset, optimal performance by RBF was overfit (i.e., better performance in training but worse performance when tested). It was determined that the sigmoid model was better in terms of prediction (i.e., the outcomes were similar between the training (0.836) and the test (0.826) datasets). The sigmoid kernel model with cost = 0.10 showed stable prediction (i.e., no overfitting issue) and optimal performance.

Table A5. Algorithm and Parameter Selection—SVM.

	Training				Test			
c	Linear AUC	Poly. AUC	RBF AUC	Sigmoid AUC	Linear AUC	Poly. AUC	RBF AUC	Sigmoid AUC
0.10	0.584	0.944	0.901	0.836	0.442	0.822	0.812	0.826
1.00	0.754	0.982	0.969	0.774	0.719	0.778	0.825	0.773
5.00	0.754	0.977	0.997	0.769	0.720	0.762	0.784	0.747
10.00	0.754	0.977	0.996	0.765	0.720	0.762	0.803	0.738
50.00	0.754	0.977	0.996	0.759	0.720	0.762	0.803	0.733
100.00	0.754	0.977	0.996	0.754	0.280	0.762	0.803	0.729

Abbreviation: c, cost; Linear, SVM with linear kernel; Poly., SVM with polynomial kernel; RBF, SVM with radial based function kernel; Sigmoid, SVM with sigmoid kernel.

Figure A4 shows the representative ROC curves for SVM. The upper left graph is the ROC graph for the linear SVM with a cost of 0.10; the left lower graph is the ROC graph for the polynomial SVM with a cost of 0.10; the right upper graph is the ROC graph for the RBF SVM with a cost of 0.10; the lower right graph is the ROC graph for the sigmoid SVM with a cost of 0.10. The dark area under the curve was used to calculate AUC. As shown in Figure A4, the ROC curves were convex, indicating that three of the SVMs performed well in prediction. When SVM was performed using a linear assumption, the prediction was suboptimal, as indicated by the concave graph. The AUC was optimized when SVM was performed with sigmoid with a cost of 0.10.

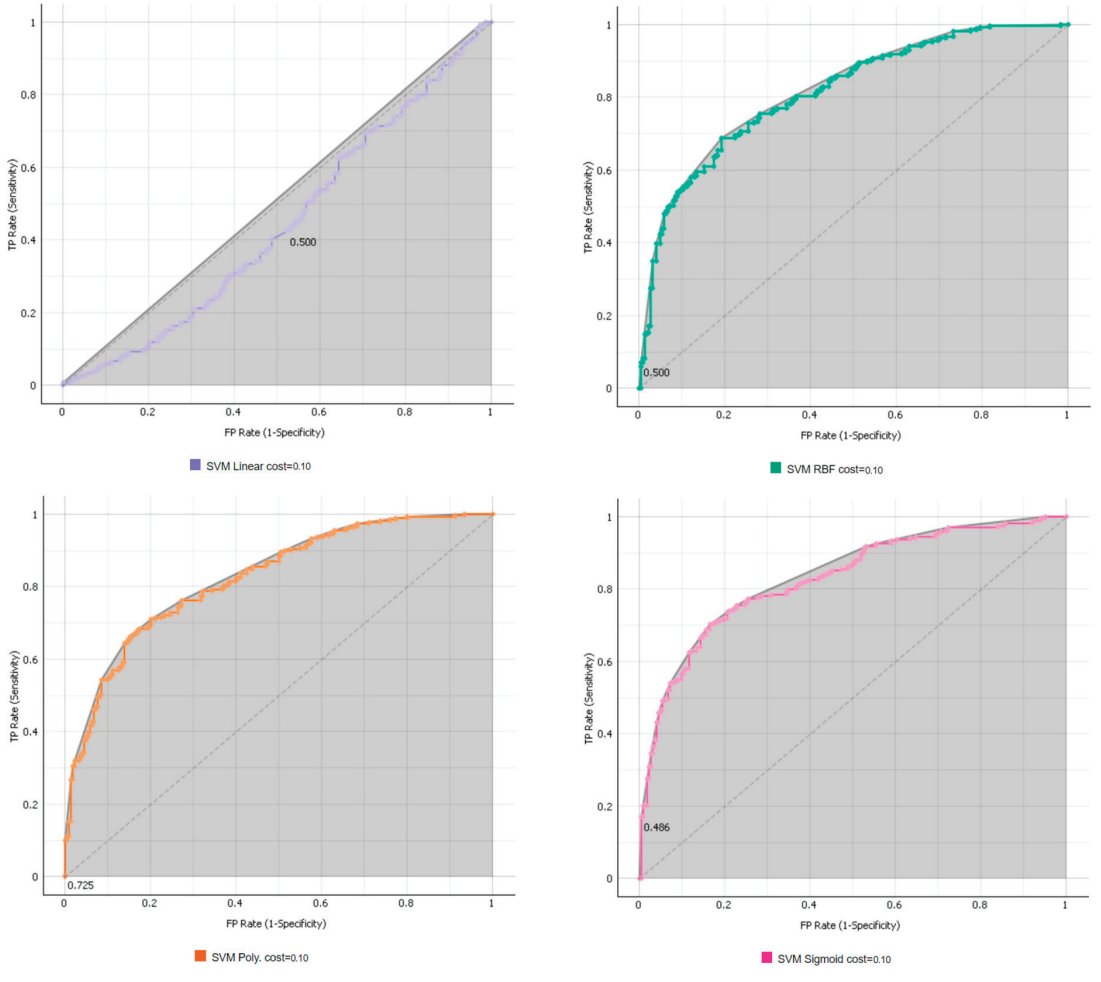

Figure A4. ROC Curve for Algorithm and Parameter Selection—SVM.

In the Stochastic Gradient Descent (SGD) shown in Table A6, reasonably good prediction rates were observed under three assumptions in the training dataset with learning rates of 0.001 and 0.005. However, when the algorithms were checked, the learning rate of 0.001 showed the best level of prediction. The type of assumption used when modeling did not lead to significant differences between the models as long as the learning rate remained at 0.001.

Table A6. Algorithm and Parameter Selection—SGD.

	Training			Test		
L.R.	Elastic AUC	Lasso AUC	Ridge AUC	Elastic AUC	Lasso AUC	Ridge AUC
0.001	0.919	0.919	0.919	0.801	0.802	0.802
0.005	0.924	0.924	0.924	0.790	0.786	0.785
0.010	0.923	0.922	0.922	0.778	0.780	0.787
0.050	0.896	0.896	0.895	0.713	0.759	0.770
0.100	0.870	0.890	0.877	0.759	0.774	0.659

Abbreviation: L.R., learning rate.

Figure A5 shows the representative ROC curves for SGD. The upper left graph is the ROC graph for lasso SGD with a learning rate of 0.001; the left lower graph is the ROC graph for ridge SGD with a learning rate of 0.001; the right upper graph is the ROC graph for lasso SGD with a learning rate of 0.05; the lower right graph is the ROC graph for elastic SGD with a learning rate of 0.001. As with the other analyses, the dark area under the curve was used to calculate AUC. As shown in Figure A5, the ROC curves were convex, indicating that each SGD performed well in prediction. The AUC was the largest when SGD was performed, with lasso/ridge with a learning rate of 0.001.

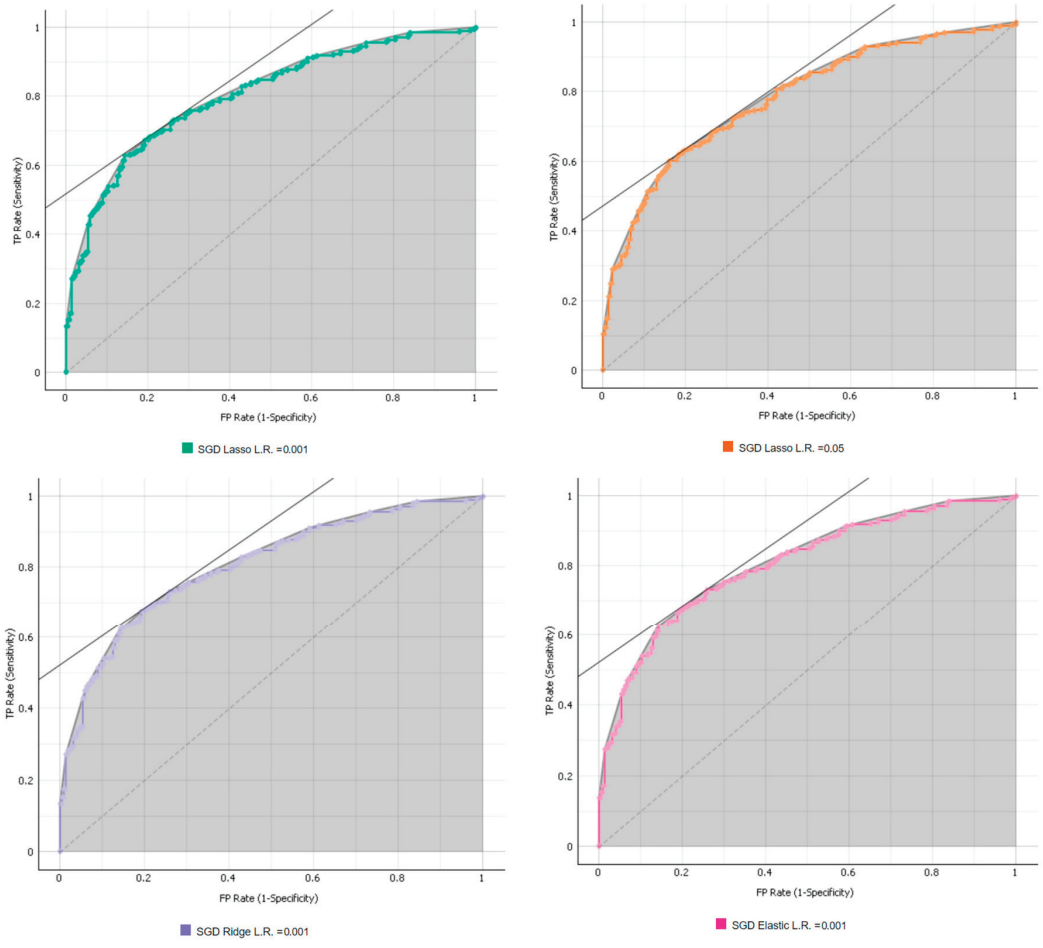

Figure A5. ROC Curve for Algorithm and Parameter Selection—SGD.

The best NN algorithm was identified in the training dataset when the number of neurons was over 15. However, in the test dataset, NN showed the best performance when the number of neurons was 30, 35, 55, and 60. The optimal number of neurons, as shown in Table A7, was 30.

Figure A6 shows the representative ROC curves for NN. The upper left graph is the ROC graph for NN with one neuron; the left lower graph is the ROC graph for NN with 50 neurons; the right upper graph is the ROC graph for NN with 30 neurons; the lower right graph is the ROC graph for NN with 100 neurons. AUC was estimated by examining the dark area under the curve. As shown in Figure A6, the ROC curves were convex, indicating

that the NN algorithms performed well in prediction. The AUC was the largest when NN was performed with 30 neurons.

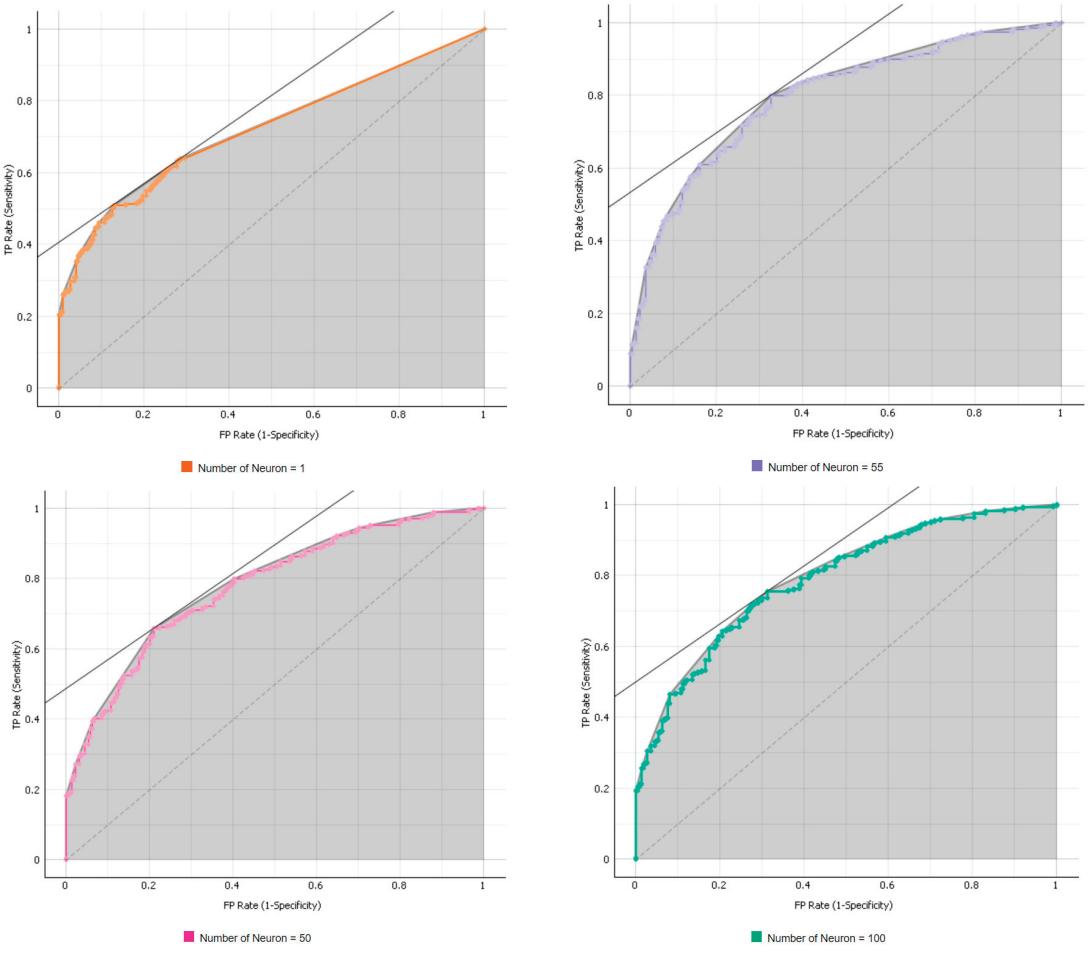

Figure A6. ROC Curve for Algorithm and Parameter Selection—NN.

Table A7. Algorithm and Parameter Selection—NN.

Number of Neuron	Training AUC	Test AUC
1	0.843	0.720
5	0.958	0.791
10	0.994	0.781
15	1.000	0.790
20	1.000	0.779
25	1.000	0.779
30	1.000	0.799
35	1.000	0.786
40	1.000	0.783

Table A7. *Cont.*

Number of Neuron	Training AUC	Test AUC
45	1.000	0.776
50	1.000	0.776
55	1.000	0.793
60	1.000	0.781
65	1.000	0.787
70	1.000	0.780
75	1.000	0.768
80	1.000	0.785
85	1.000	0.783
90	1.000	0.780
95	1.000	0.790
100	1.000	0.787

References

1. Bronfenbrenner, U. Toward an experimental ecology of human development. *Am. Psychol.* **1977**, *32*, 513–531. [CrossRef]
2. Salignac, F.; Hamilton, M.; Noone, J.; Marjolin, A.; Muir, K. Conceptualizing financial wellbeing: An ecological life-course approach. *J. Happiness Stud.* **2020**, *21*, 1581–1602. [CrossRef]
3. Despard, M.R.; Friedline, T.; Martin-West, S. Why do households lack emergency savings? The role of financial capability. *J. Fam. Econ. Issues* **2020**, *41*, 542–557. [CrossRef]
4. Gjertson, L. Emergency Saving and Household Hardship. *J. Fam. Econ. Issues* **2016**, *37*, 1–17. [CrossRef]
5. Wang, W.; Cui, Z.; Chen, R.; Wang, Y.; Zhao, X. Regression Analysis of Clustered Panel Count Data with Additive Mean Models. Statistical Papers. Advanced Online Publication. 2023. Available online: https://link.springer.com/article/10.1007/s00362-023-01511-3#citeas (accessed on 1 November 2023).
6. Heo, W. *The Demand for Life Insurance: Dynamic Ecological Systemic Theory Using Machine Learning Techniques*; Springer: Berlin/Heidelberg, Germany, 2020.
7. Luo, C.; Shen, L.; Xu, A. Modelling and estimation of system reliability under dynamic operating environments and lifetime ordering constraints. *Reliab. Eng. Syst. Saf.* **2022**, *218 Pt A*, 108136. [CrossRef]
8. Jordan, M.I.; Mitchell, T.M. Machine learning: Trends, perspectives, and prospects. *Science* **2015**, *349*, 255–260. [CrossRef]
9. Carmona, P.; Climent, F.; Momparler, A. Predicting failure in the U.S. banking sector: An extreme gradient boosting approach. *Int. Rev. Econ. Financ.* **2019**, *61*, 304–323. [CrossRef]
10. Guelman, L. Gradient boosting trees for auto insurance loss cost modeling and prediction. *Experts Syst. Appl.* **2012**, *39*, 3659–3667. [CrossRef]
11. Heo, W.; Lee, J.M.; Park, N.; Grable, J.E. Using artificial neural network techniques to improve the description and prediction of household financial ratios. *J. Behav. Exp. Financ.* **2020**, *25*, 100273. [CrossRef]
12. Jadhav, S.; He, H.; Jenkins, K. Information gain directed genetic algorithm wrapper feature selection for credit rating. *Appl. Soft Comput.* **2018**, *69*, 541–553. [CrossRef]
13. Kalai, R.; Ramesh, R.; Sundararajan, K. Machine Learning Models for Predictive Analytics in Personal Finance. In *Modeling, Simulation and Optimization*; Das, B., Patgiri, R., Bandyopadhyay, S., Balas, V.E., Eds.; Smart Innovation, Systems and Technologies; Springer: Singapore, 2022; Volume 292.
14. Viaene, S.; Derrig, R.A.; Dedene, G. A case study of applying boosting Naïve Bayes to claim fraud diagnosis. *IEEE Trans. Knowl. Data Eng.* **2004**, *16*, 612–620. [CrossRef]
15. Zhang, Y.; Haghni, A. A gradient boosting method to improve travel time predictions. *Transp. Res. Part C-Emerg. Technol.* **2015**, *58 Pt B*, 308–324. [CrossRef]
16. Zhou, L.; Pan, S.; Wang, J.; Vasilakos, A.V. Machine learning on big data: Opportunities and challenges. *Neurocomputing* **2017**, *237*, 350–361. [CrossRef]
17. Harness, N.; Diosdado, L. Household financial ratios. In *De Gruyter Handbook of Personal Finance*; Grable, J.E., Chatterjee, S., Eds.; De Gruyter: Berlin, Germany, 2022; pp. 171–188.
18. Johnson, D.P.; Widdows, R. Emergency fund levels of households. In Proceedings of the 31st Annual Conference of the American Council on Consumer Interests, Fort Worth, TX, USA, 27–30 March 1985; pp. 235–241.
19. Lytton, R.H.; Garman, E.T.; Porter, N. How to use financial ratios when advising clients. *J. Financ. Couns. Plan.* **1991**, *2*, 3–23.
20. Prather, C.G.; Hanna, S. Ratio analysis of personal financial statements: Household norms. In Proceedings of the Association for Financial Counseling and Planning Education; Edmondsson, M.E., Perch, K.L., Eds.; AFCPE: Westerville, OH, USA, 1987; pp. 80–88.
21. Greninger, S.A.; Hampton, V.L.; Kim, K.A.; Achacoso, J.A. Ratios and benchmarks for measuring the financial well-being of families and individuals. *Financ. Serv. Rev.* **1996**, *5*, 57–70. [CrossRef]

22. Bi, L.; Montalto, C.P. Emergency funds and alternative forms of saving. *Financ. Serv. Rev.* **2004**, *13*, 93–109.
23. Hanna, S.; Fan, J.X.; Change, Y.R. Optimal life cycle savings. *J. Financ. Couns. Plan.* **1995**, *6*, 1–16.
24. Cagetti, M. Wealth accumulation over the life cycle and precautionary saving? *Rev. Econ. Stat.* **2003**, *80*, 410–419. [CrossRef]
25. Kudyba, S.; Kwatinetz, M. Introduction to the big data era. In *Big Data, Mining, and Analytics*; Kudyba, S., Ed.; CRC Press and Taylor and Francis: Boca Raton, FL, USA, 2014; pp. 1–15.
26. Thompson, W. Data mining methods and the rise of big data. In *Big Data, Mining, and Analytics*; Kudyba, S., Ed.; CRC Press and Taylor and Francis: Boca Raton, FL, USA, 2014; pp. 71–101.
27. Sarker, I.H. Machine learning: Algorithms, real-World applications and research directions. *SN Comput. Sci.* **2021**, *2*, 160. [CrossRef]
28. Abiodun, O.I.; Jantan, A.; Omolara, A.E.; Dada, K.V.; Mohamed, N.A.; Arshard, H. State-of-the-art in artificial neural network applications: A survey. *Heliyon* **2018**, *4*, e00938. [CrossRef]
29. Demsar, J.; Curk, T.; Erjavec, A.; Gorup, C.; Hočevar, T.; Milutinovič, M.; Možina, M.; Polajnar, M.; Toplak, M.; Starič, A.; et al. Orange: Data mining toolbox in Python. *J. Mach. Learn. Res.* **2013**, *14*, 2349–2353.
30. Pisner, D.A.; Schnyer, D.M. Chapter 6—Support vector machine. In *Machine Learning*; Mechelli, A., Vieira, S., Eds.; Academic Press: Cambridge, MA, USA, 2020; pp. 101–121.
31. Rudin, C.; Daubechies, I.; Schapire, R. Fin The dynamics of AdaBoost: Cyclic behavior and convergence of margins. *J. Mach. Learn. Res.* **2004**, *5*, 1557–1595.
32. Suthaharan, S. Support Vector Machine. In *Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning*; Suthaharan, S., Ed.; Springer: New York, NY, USA, 2016; pp. 207–235.
33. Meng, Y.; Li, X.; Zheng, X.; Wu, F.; Sun, X.; Zhang, T.; Li, J. Fast Nearest Neighbor Machine Translation. *arXiv* **2021**, arXiv:2105.14528.
34. Wu, X.; Kumar, V.; Quinlan, J.R.; Ghosh, J.; Yang, Q.; Motoda, H.; McLachlan, G.J.; Ng, A.; Liu, B.; Philip, S.Y.; et al. Top 10 algorithms in data mining. *Knowl. Inf. Syst.* **2008**, *14*, 1–37. [CrossRef]
35. Triguero, I.; Garcia-Gil, D.; Maillo, J.; Luengo, J.; Garcia, S.; Herrera, F. Transforming big data into smart data: An insight on the use of the k-nearest neighbor algorithms to obtain quality data. *WIREs Data Min. Knowl. Discov.* **2018**, *9*, e1289. [CrossRef]
36. Fix, E.; Hodges, J.L. Discriminatory analysis. Nonparametric discrimination: Consistency properties. *Int. Stat. Rev. Rev. Int. De Stat.* **1989**, *57*, 238–247. [CrossRef]
37. Singh, A.; Yadav, A.; Rana, A. K-means with three different distance metrics. *Int. J. Comput. Appl.* **2013**, *67*, 13–17. [CrossRef]
38. Östermark, R. A fuzzy vector valued KNN-algorithm for automatic outlier detection. *Appl. Soft Comput.* **2009**, *9*, 1263–1272. [CrossRef]
39. Maede, N. A comparison of the accuracy of short-term foreign exchange forecasting methods. *Int. J. Forecast.* **2002**, *18*, 67–83. [CrossRef]
40. Phongmekin, A.; Jarumaneeroj, P. Classification Models for Stock's Performance Prediction: A Case Study of Finance Sector in the Stock Exchange of Thailand. In Proceedings of the 2018 International Conference on Engineering, Applied Sciences, and Technology (ICEAST), Phuket, Thailand, 4–7 July 2018; pp. 1–4.
41. Breiman, L. *Arcing the Edge*; Technical Report 486; Statistics Department, University of California at Berkeley: Berkeley, CA, USA, 1997.
42. Friedman, J.H. Greedy function approximation: A Gradient Boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
43. Sagi, O.; Rokach, L. Ensemble learning: Survey. *WIREs Data Min. Knowl. Discov.* **2017**, *8*, e1249. [CrossRef]
44. Chang, Y.; Chang, K.; Wu, G. Application of eXtreme gradient boosting trees in the construction of credit risk assessment models for financial institutions. *Appl. Soft Comput.* **2018**, *73*, 914–920. [CrossRef]
45. Liu, J.; Wu, C.; Li, Y. Improving financial distress prediction using financial network-based information and GA-based Gradient Boosting model. *Comput. Econ.* **2017**, *53*, 851–872. [CrossRef]
46. Dorogush, A.V.; Ershov, V.; Gulin, A. CatBoost: Gradient Boosting with Categorical Features Support. *arXiv* **2018**, arXiv:1810.11363v1. [CrossRef]
47. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
48. Hand, D.J.; Yu, K. Idiot's Bayes—Not so stupid after all? *Int. Stat. Rev.* **2001**, *69*, 385–398.
49. Lowd, D.; Domingos, P. Naïve Bayes models for probability estimation. In Proceedings of the ICML '05: Proceedings of the 22nd International Conference on Machine Learning, Bonn, Germany, 7–11 August 2005; pp. 529–536.
50. Zhang, H. Exploring conditions for the optimality of Naïve Bayes. *Int. J. Pattern Recognit. Artif. Intell.* **2005**, *19*, 183–198. [CrossRef]
51. Yang, F. An implementation of Naïve Bayes classifier. In Proceedings of the 2018 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 12–14 December 2018; pp. 301–306.
52. Deng, Q. Detection of fraudulent financial statements based on Naïve Bayes classifier. In Proceedings of the 2010 5th International Conference on Computer Science and Education, Hefei, China, 24–27 August 2010; pp. 1032–1035.
53. Shihavuddin, A.S.M.; Ambia, M.N.; Arefin, M.M.N.; Hossain, M.; Anwar, A. Prediction of stock price analyzing the online financial news using Naïve Bayes classifier and local economic trends. In Proceedings of the 2010 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE), Chengdu, China, 20–22 August 2010; pp. V4-22–V4-26.
54. Noble, W.S. What is a support vector machine? *Nat. Biotechnol.* **2006**, *24*, 1565–1567. [CrossRef]

55. Yu, L.; Yao, X.; Wang, S.; Lai, K.K. Credit risk evaluation using a weighted least squares SVM classifier with design of experiment for parameter selection. *Expert Syst. Appl.* **2011**, *38*, 15392–15399. [CrossRef]
56. Chen, F.; Li, F. Combination of feature selection approaches with SVM in credit scoring. *Expert Syst. Appl.* **2010**, *37*, 4902–4909. [CrossRef]
57. Chen, W.; Du, Y. Using neural networks and data mining techniques for the financial distress prediction model. *Expert Syst. Appl.* **2009**, *36*, 4075–4086. [CrossRef]
58. Baesens, B.; Van Gestel, T.; Viaene, S.; Stepanova, M.; Suykens, J.; Vanthienen, J. Benchmarking state-of-the-art classification algorithms for credit scoring. *J. Oper. Res. Soc.* **2003**, *54*, 627–635. [CrossRef]
59. Yang, Y. Adaptive credit scoring with kernel learning methods. *Eur. J. Oper. Res.* **2007**, *183*, 1521–1536. [CrossRef]
60. Kim, K.; Ahn, H. A corporate credit rating model using multi-class support vector machines with an ordinal pairwise partitioning approach. *Comput. Oper. Res.* **2012**, *39*, 1800–1811. [CrossRef]
61. Chaudhuri, A.; De, K. Fuzzy support vector machine for bankruptcy prediction. *Appl. Soft Comput.* **2011**, *11*, 2472–2486. [CrossRef]
62. Chen, L.; Hsiao, H. Feature selection to diagnose a business crisis by using a real Ga-based support vector machine: An empirical study. *Expert Syst. Appl.* **2008**, *35*, 1145–1155. [CrossRef]
63. Hsieh, T.; Hsiao, H.; Yeh, W. Mining financial distress trend data using penalty guided support vector machines based on hybrid of particle swarm optimization and artificial bee colony algorithms. *Neurocomputing* **2012**, *82*, 196–206. [CrossRef]
64. Amari, S. A theory of adaptive pattern classifiers. *IEEE Trans. Electron. Comput.* **1967**, *EC-16*, 299–307. [CrossRef]
65. Amari, S. Backpropagation and stochastic gradient descent method. *Neurocomputing* **1993**, *5*, 185–196. [CrossRef]
66. Ketkar, N. Stochastic Gradient Descent. In *Deep Learning with Python*; Apress: Berkeley, CA, USA, 2017.
67. Song, S.; Chaudhuri, K.; Sarwate, A.D. Stochastic gradient descent with differentially private updates. In Proceedings of the 2013 IEEE Global Conference on Signal and Information Processing, Austin, TX, USA, 3–5 December 2013; pp. 245–248.
68. Newton, D.; Pasupathy, R.; Yousefian, F. Recent trends in stochastic gradient decent for machine learning and big data. In Proceedings of the 2018 Winter Simulation Conference (WSC), Gothenburg, Sweden, 9–12 December 2018; pp. 366–380.
69. Deepa, N.; Prabadevi, B.; Maddikunta, P.K.; Gadekallu, T.R.; Baker, T.; Khan, M.A.; Tariq, U. An AI-based intelligent system for healthcare analysis using Ridge-Adaline Stochastic Gradient Descent Classifier. *J. Supercomput.* **2020**, *77*, 1998–2017. [CrossRef]
70. Zou, H.; Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B* **2005**, *67*, 301–320. [CrossRef]
71. Matías, J.M.; Vaamonde, A.; Taboada, J.; González-Manteiga, W. Support vector machines and gradient boosting for graphical estimation of a slate deposit. *Stoch. Environ. Res. Risk Assess.* **2004**, *18*, 309–323. [CrossRef]
72. Moisen, G.G.; Freeman, E.A.; Blackard, J.A.; Frescino, T.S.; Zimmermann, N.E.; Edward, T.C., Jr. Predicting tree species presence and basal area in Utah: A comparison of stochastic gradient boosting, generalized additive models, and tree-based methods. *Ecol. Model.* **2006**, *199*, 176–187. [CrossRef]
73. Baum, E.B. Neural nets for economics. In *The Economy as an Evolving Complex System, Proceedings of the Evolutionary Paths of the Global Economy Workshop, Sante Fe, NM, USA, 8–18 September 1987*; Anderson, P., Arrow, K., Pindes, D., Eds.; Addison-Wesley: Reading, MA, USA, 1988; pp. 33–48.
74. Kirkos, E.; Spathis, C.; Manolopoulos, Y. Data mining techniques for the detection of fraudulent financial statement. *Expert Syst. Appl.* **2007**, *32*, 995–1003. [CrossRef]
75. Cerullo, M.J.; Cerullo, V. Using neural networks to predict financial reporting fraud: Part 1. *Comput. Fraud. Secur.* **1999**, *5*, 14–17.
76. Dorronsoro, J.R.; Ginel Fin Sgnchez, C.; Cruz, C.S. Neural fraud detection in credit card operations. *IEEE Trans. Neural Netw.* **1997**, *8*, 827–834. [CrossRef]
77. Chauhan, N.; Ravi, V.; Chandra, D.K. Differential evolution trained wavelet neural networks: Application to bankruptcy prediction in banks. *Expert Syst. Appl.* **2009**, *36*, 7659–7665. [CrossRef]
78. Iturriaga, F.J.L.; Sanz, I.P. Bankruptcy visualization and prediction using neural networks: A study of U.S. commercial banks. *Expert Syst. Appl.* **2015**, *42*, 2857–2869. [CrossRef]
79. Menard, S. *Applied Logistic Regression Analysis*, 2nd ed.; Sage Publications: Thousand Oaks, CA, USA, 2002.
80. Arcuri, A.; Fraser, G. Parameter tuning or default values? An empirical investigation in search-based software engineering. *Empir. Softw. Eng.* **2013**, *18*, 594–623. [CrossRef]
81. Joseph, V.R. Optimal ratio for data splitting. *Stat. Anal. Data Min.* **2022**, *15*, 531–538. [CrossRef]
82. Afendras, G.; Markatou, M. Optimality of training/test size and resampling effectiveness in cross-validation. *J. Stat. Plan. Inference* **2019**, *199*, 286–301. [CrossRef]
83. Picard, R.R.; Berk, K.N. Data Splitting. *Am. Stat.* **1990**, *44*, 140–147.
84. Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* **2006**, *27*, 861–874. [CrossRef]
85. Kira, K.; Rendell, L.A. A practical approach to feature selection. In *Machine Learning: Proceedings of International Conference (ICML'92)*; Sleeman, D., Edwards, P., Eds.; Morgan Kaufmann: Burlington, MA, USA, 1992; pp. 249–256.
86. Kononenko, I. Estimating attributes: Analysis and extensions of Relief. In *Machine Learning: ECML-94*; De Raedt, L., Bergadano, F., Eds.; Springer: Berlin/Heidelberg, Germany, 1994; pp. 171–182.
87. Robnik-Šikonja, M.; Kononenko, I. Theoretical and empirical analysis of ReliefF and RReliefF. *Mach. Learn.* **2003**, *53*, 23–69. [CrossRef]
88. Heo, W.; Cho, S.; Lee, P. APR Financial Stress Scale: Development and Validation of a Multidimensional Measurement. *J. Financ. Ther.* **2020**, *11*, 2. [CrossRef]

89. Xiao, J.J.; Ahn, S.Y.; Serido, J.; Shim, S. Earlier financial literacy and later financial behavior of college students. *Int. J. Consum. Stud.* **2014**, *38*, 593–601. [CrossRef]
90. Lusardi, A. Financial literacy and the need for financial education: Evidence and implications. *Swiss J. Econ. Stat.* **2019**, *155*, 1. [CrossRef]
91. Grable, J.E.; Lytton, R.H. Financial risk tolerance revisited: The development of a risk assessment instrument. *Financ. Serv. Rev.* **1999**, *8*, 163–191. [CrossRef]
92. Loibl, C.; Hira, T.K. Self-directed financial learning and financial satisfaction. *J. Financ. Couns. Plan.* **2005**, *16*, 11–22.
93. Lown, J.M. Development and validation of a financial self-efficacy scale. *J. Financ. Couns. Plan.* **2011**, *22*, 54–63.
94. Perry, V.G.; Morris, M.D. Who is in control? The role of self-perception, knowledge, and income in explaining consumer financial Behavior. *J. Consum. Aff.* **2005**, *39*, 299–313. [CrossRef]
95. Diener, E.; Emmons, R.A.; Larsen, R.J.; Griffin, S. The satisfaction with life scale. *J. Personal. Assess.* **1985**, *49*, 71–75. [CrossRef] [PubMed]
96. Rosenberg, M. *Society and the Adolescent Self-Image*; Princeton University Press: Princeton, NJ, USA, 1965.
97. Hellgren, J.; Sverke, M.; Isaksson, K. A two-dimensional approach to job insecurity: Consequences for employee attitudes and well-being. *Eur. J. Work. Organ. Psychol.* **1999**, *8*, 179–195. [CrossRef]

mathematics

MDPI

Article

Robust Portfolio Choice under the Modified Constant Elasticity of Variance

Wei Li Fan † and Marcos Escobar Anel *,†

Department of Statistical and Actuarial Sciences, University of Western Ontario, London, ON N6A 5B7, Canada; wfan54@uwo.ca

* Correspondence: marcos.escobar@uwo.ca

† These authors contributed equally to this work.

Abstract: This study investigates ambiguity aversion within the framework of a utility-maximizing investor under a modified constant-elasticity-of-volatility (M-CEV) model for the underlying asset. We derive closed-form solutions of a non-affine type for the optimal allocation and value function via a Cauchy problem. This work generalizes previous results in non-ambiguous settings by extending existing work to Hyperbolic Absolute Risk Aversion utility (HARA), correcting some typos in the literature for Constant Relative Risk Aversion utility (CRRA). Helpful details and derivations are also included in the manuscript.

Keywords: M-CEV model; expected utility; HARA; ambiguity-aversion; Cauchy problem

MSC: 62P05; 91B70

JEL Classification: C61

Citation: Fan, W.L.; Anel, M.E. Robust Portfolio Choice under the Modified Constant Elasticity of Variance. *Mathematics* **2024**, *12*, 440. https://doi.org/10.3390/math12030440

Academic Editors: Jing Yao, Xiang Hu and Jingchao Li

Received: 29 November 2023
Revised: 16 January 2024
Accepted: 25 January 2024
Published: 30 January 2024

1. Introduction

Portfolio optimization in continuous time is a constantly evolving line of research, with both practitioners and academics actively seeking more realistic objective functions, improved models, and refined stylized facts to enhance decision-making. All of these efforts are essential to keep pace with the growing complexity of financial markets and social interactions.

In this study, we embrace the widely utilized framework of expected utility theory (EUT) as the objective function for investors. EUT provides a fruitful ground for analytical solutions and is, therefore, more easily interpretable, as pioneered in the seminal work by Merton [1]. This early achievement relied on simple geometric Brownian motion (GBM) to describe the underlying asset price with considerations only for risk aversion. Although many extensions have been considered in the literature, the vast majority involve more than one source of risk in explaining asset prices. Examples include stochastic volatility models and the addition of jumps, as evidenced in references [2,3]. However, incorporating multiple risk sources often leads to inaccuracies in the estimation and evaluation methodologies, thereby partially affecting their benefits. To preserve the simplicity of a single source of risk while generalizing the GBM, this study uses a new member of the family of CEV models (see the seminal work of Cox in [4]); the so-called modified constant elasticity of variance (M-CEV). This model, explored for pricing purposes in the reference [5], has been adapted to expected utility optimization in the recent work of Muravei in [6]. Additionally, refer to the references [7,8] for portfolio optimization results on two other types of CEV models: CEV and LVO-CEV.

The primary innovation in our work lies in considering ambiguity in the model, thereby introducing ambiguity aversion in the decision-making of the investor. The assumption of model ambiguity can be traced back to the experimental studies of Ellsberg [9]

and more recently the work in [10], who demonstrate that individuals are averse to ambiguity (unknown probability) in addition to their well-known aversion to risk. A crucial framework for analytical solutions was presented in Maenhout [11], leading to a Hamilton-Jacobi-Bellman (HJBI) representation of the solution in the GBM case. Most extensions of this work consider multiple sources of risk (e.g., the reference [12] for stochastic volatility and jumps). This study is the first to extend the analysis to a member of the CEV family of models.

For clarity, we list the main contributions of our work as follows:

1. We solve the expected utility portfolio problem for a HARA (Hyperbolic Absolute Risk Aversion) investor in the absence of ambiguity aversion. This solution can be reinterpreted as a constant proportion portfolio insurance (CPPI), a strategy widely employed in the financial industry (see the reference [13]).
2. As a by-product of solving HARA, we identified and rectified a few typos in the literature regarding the solution for the embedded Constant Relative Risk Aversion (CRRA) case.
3. We find closed-form solutions for the optimal allocation, optimal reference model, value function, and optimal wealth process for an investor exhibiting both risk and ambiguity aversion within HARA utilities. Numerical and empirical studies with our findings will be conducted in a follow-up paper.

The remainder of this paper is organized as follows. Section 2 outlines the mathematical and financial settings and formulates the problem. In Section 3, we present the derived solutions and delve into details of various embedded cases that hold significance for readers. The appendix provides additional details that are organized into subsections related to the main proposition.

2. Problem Formulation

We consider the general price process S_t given by the following stochastic differential equation (SDE):

$$\frac{dS_t}{S_t} = \left(r + \bar{\lambda}\sigma S_t^{\psi}\right)\mathrm{d}t + \sigma S_t^{\beta}\mathrm{d}Z_t,\ s_0 > 0, \tag{1}$$

where r and $\bar{\lambda}$ are positive real numbers and s_0 is the initial asset price value. In this general setting, S_t^{β} represents the volatility of the risky asset return, and β is the elasticity of variance with respect to the stock price. We assume $\psi = 2\beta$ and $\beta \neq 0$, hence Equation (1) becomes the M-CEV model (see the manuscripts [6,14]).

We assume a market consisting of money market account B and risky asset S_t (such as a stock). These prices follow the following dynamics:

$$\frac{\mathrm{d}B_t}{B_t} = r\mathrm{d}t, \tag{2}$$

$$\frac{\mathrm{d}S_t}{S_t} = \left(r + \bar{\lambda}\sigma S_t^{2\beta}\right)\mathrm{d}t + \sigma S_t^{\beta}\mathrm{d}Z_t,\ s_0 > 0, \tag{3}$$

where r is a constant risk-free interest rate. As explained in the references [5,6,8], the M-CEV allows for a non-zero probability of the underlying touching zero (default) if $\beta > 1$, which makes it realistic for pricing and portfolio problems.

We refer to model (3) as the reference model. Our investors face uncertainty regarding the probability distribution associated with the reference model and consider a set of plausible alternative models when making investment decisions. Specifically, the investors are uncertain about the distribution of S. This formulation enables us to capture the uncertainty regarding the drift in stock prices and the prices of risk in the stock market.

Let $e := e_t^s$ denote an $\mathbb{R}$-valued $\mathcal{F}_t$-progressively measurable process. The Radon–Nikodym derivative process is defined as follows:

$$\zeta_t^e = \mathbb{E}\left[\frac{\mathrm{d}\mathbb{P}^e}{\mathrm{d}\mathbb{P}}|\mathcal{F}_t\right] = \exp\left(-\int_0^t \frac{(e_\tau^s)^2}{2}\mathrm{d}\tau + e_\tau \mathrm{d}Z\tau\right). \tag{4}$$

According to Girsanov's theorem, the process can be expressed as

$$\widetilde{Z}_t = \int_0^t e_\tau^s \mathrm{d}\tau + Z_t. \tag{5}$$

We denote the set of all $\mathcal{F}_t$-progressively measurable processes such that the process given by (4) is a well-defined Radon–Nikodym derivative process as $\mathcal{E}[0,T]$. This process, denoted by $\widetilde{Z}_t$, represents the Wiener process under probability measure $\mathbb{P}^e$. The reference model is generally regarded as the most accurate representation of available data for an agent. However, viable alternative models pose a challenge in terms of statistical differentiation from reference models. Alternative models were generated via the perturbation process e_t^s as follows:

$$\frac{dS_t}{S_t} = \left(r + \bar{\lambda}\sigma S_t^{2\beta} - e_t^s \sigma S_t^{\beta}\right)\mathrm{d}t + \sigma S_t^{\beta}\mathrm{d}Z_t. \tag{6}$$

In the model setup, it is important to acknowledge that this ambiguity stems from an investor's inability to capture the expected returns precisely in the probability laws governing the stock price process. This assumption aligns with the perspectives presented in seminar studies, such as the work by Merton [15], and more recently shown in Tables 1 and 2 in reference [8] on the large standard errors in estimating the parameter $\bar{\lambda}$.

Let π represent the investment strategy applied from time t to time T in model (3). We define space $\mathcal{U}[0,T]$ as the set of admissible strategies π that satisfy the following conditions:

1. π is an $\mathcal{F}_t$-progressively measurable process.
2. Under π, the wealth X_t of the investor remains non-negative for $t \in [0,T]$.
3. The integrability conditions necessary for the expectation operator in (7) to be well-defined are satisfied.

We consider an investor with a preference for Hyperbolic Absolute Risk Aversion (HARA) utility on terminal wealth with risk-aversion level γ. The utility function for terminal wealth X_T is defined as:

$$u(X_T) = \frac{(X_T - F)^\gamma}{\gamma}.$$

The goal is to examine an ambiguous agent with HARA utility who aims to construct an investment strategy for the time interval $[0,T]$ that maximizes the expected utility for terminal wealth X_T. In line with this objective, we define the reward functional realized by the investor when selecting an alternative model specified by e as follows:

$$w^e(x,t;\pi,c) = \mathbb{E}_{x,t}^{\mathbb{P}^e}\left[\frac{(X_T - F)^\gamma}{\gamma}\right], \tag{7}$$

where $-\infty < \gamma < 1$, $\gamma \neq 0$, and denotes the constant relative aversion parameter. Then, the indirect utility function is given by:

$$J(x,t) = \sup_{\pi\in\mathcal{U}} \inf_{e\in\mathcal{E}[t,T]} \left(w^e(x,t;\pi) + \mathbb{E}^{\mathbb{P}^e}\left[\int_t^T \frac{(e_\tau^s)^2}{2\Psi(\tau, X_\tau)}\mathrm{d}\tau\right]\right), \tag{8}$$

where space $\mathcal{U}$ consists of admissible controls $\{\pi_t\}_{t\in[0,T]}$ with $\pi_t \in \mathbb{R}$ and satisfies the standard conditions.

According to reference [16], investors consider alternative models that are statistically challenging to distinguish from reference models. To address this issue, the value function incorporates a penalty term designed to discourage significant deviations from the reference model. This penalty term, which appears as the last expectation term in the problem (8), is computed based on relative entropy. The perturbations e_t^s in the penalty term were scaled by $\Psi(\tau, X_\tau)$. Function Ψ reflects the level of ambiguity or uncertainty associated with the models and is used to adjust the penalties accordingly. Higher values of Ψ indicate smaller penalties for deviating from the reference model, signifying greater uncertainty on the part of the investor.

By incorporating the penalty term, the investor considers both model ambiguity and the associated diffusion risk. This approach allows investors to make decisions that weigh the potential benefits of alternative models while accounting for the risks and uncertainties involved.

For the sake of analytical tractability, we maintain the assumption proposed in reference [11], that the ambiguity aversion parameter ϕ is related to the value function $J(x,t)$ and the risk aversion level γ by the expression

$$\Psi = \frac{\phi}{\gamma J(x,t)}, \tag{9}$$

where $\phi > 0$ denotes the level of ambiguity aversion. This assumption allows us to incorporate the degree of ambiguity into the model and analyze its impact on investors' decision-making processes.

In summary, the value function considers investors' preferences for terminal wealth while also considering the uncertainty associated with alternative models. The penalty term within the value function discourages significant deviation from the reference model. Scaling factor Ψ is a key component that reflects investors' level of ambiguity aversion. Higher values of Ψ indicate greater uncertainty and ambiguity regarding the alternative models, which corresponds to smaller penalties for deviations from the reference model. By incorporating this penalty term and adjusting it based on the level of ambiguity, investors can make decisions that balance their preferences with the associated uncertainties in the alternative models.

3. Optimal Investment Strategies

In this section, we address the problem described in (8) by employing the stochastic control approach for the M-CEV model with ambiguity. Our objective is to derive closed-form solutions to the investment problem. The solution will provide insights into the impact of ambiguity aversion levels on investors' decision-making processes.

3.1. The Hamilton-Jacobi-Bellman-Isaacs (HJBI) Equation

Assuming that π_t represents the fraction of wealth invested in stock S_t, the remaining portion of wealth $(1-\pi_t)$ is invested in the risk-free money account with a constant interest rate r. For mathematical benefits, our focus is on the investor's position in the stock, denoted by ψ_t (i.e., the number of units of the asset held). Then, investor wealth X_t is governed by the following stochastic differential equation (SDE) derived from a self-financing condition:

$$\begin{aligned} dX_t &= X_t\left[\pi_t \frac{\mathrm{d}S_t}{S_t} + (1-\pi_t) r \mathrm{d}t\right] \\ &= X_t\left(r + \pi_t \bar{\lambda}\sigma S_t^{2\beta} - \pi_t e_t^s \sigma S_t^{\beta}\right)\mathrm{d}t + \pi_t \sigma S_t^{\beta} X_t \mathrm{d}Z_t. \end{aligned} \tag{10}$$

Since $\psi_t = \pi_t \frac{X_t}{S_t}$, we can express Equation (10) as follows:

$$\begin{aligned} dX_t &= \psi_t \mathrm{d}S_t + r(X_t - \psi_t S_t)\mathrm{d}t \\ &= \Big[\psi_t r S_t + \psi_t \bar{\lambda}\sigma S_t^{2\beta+1} - \psi_t e_t^s \sigma S_t^{\beta+1} + r(X_t - \psi_t S_t)\Big]\mathrm{d}t + \psi_t \sigma S_t^{\beta+1}\mathrm{d}Z_t \\ &= \Big(rX_t + \psi_t \bar{\lambda}\sigma S_t^{2\beta+1} - \psi_t e_t^s \sigma S_t^{\beta+1}\Big)\mathrm{d}t + \psi_t \sigma S_t^{\beta+1}\mathrm{d}Z_t. \end{aligned} \tag{11}$$

Consequently, the expression for the value function (8) should be modified to:

$$J(X,S,t) = \sup_{\psi\in\mathcal{U}} \inf_{e\in\mathcal{E}[t,T]} \left(w^e(X,S,t;\psi) + \mathbb{E}^{\mathbb{P}^e}\left[\int_t^T \frac{(e_\tau^s)^2}{2\Psi(\tau, X_\tau, S_\tau)}\mathrm{d}\tau\right]\right), \tag{12}$$

where $J(X,S,t)$ satisfies a HJBI equation.

Define

$$\theta_t := \pi_t \sigma S_t^{\beta}, \tag{13}$$

where θ_t represents the portfolio exposure to the fundamental risk factor Z_t, and π_t is the portfolio weight that determines the allocation of wealth in the M-CEV model. Expression (13) illustrates how the investor's desired exposure to the risk factor is calculated, depending on the portfolio weight π_t and the volatility of the asset return σ, which is scaled by S_t^{β}. The portfolio weight determines the portion of investor wealth allocated to the risky asset, thereby influencing the overall exposure to the risk factor. As before, we can represent (13) as:

$$\begin{aligned} \theta_t &:= \frac{\psi_t S_t}{X_t}\sigma S_t^{\beta} \\ &= \psi_t X_t^{-1}\sigma S_t^{\beta+1}. \end{aligned} \tag{14}$$

This means that we can also state that the number of units of the asset dictates the allocation of the investor's wealth to the risky asset, thus impacting overall exposure. To simplify our analysis, we shift our focus to these exposures rather than investor position ψ in stock. As a result, the transformation of model (11) is given by:

$$\begin{aligned} dX_t &= \Big(rX_t + \psi_t \bar{\lambda}\sigma S_t^{2\beta+1} - \psi_t e_t^s \sigma S_t^{\beta+1}\Big)\mathrm{d}t + \psi_t \sigma S_t^{\beta+1}\mathrm{d}Z_t \\ &= X_t\Big(r + \bar{\lambda}\theta_t S_t^{\beta} - e_t^s\theta_t\Big)\mathrm{d}t + \theta_t X_t \mathrm{d}Z_t. \end{aligned} \tag{15}$$

Accordingly, the value function (12) satisfies the Hamilton (Jacobi)-Isaacs equation:

$$\begin{aligned} \sup_{\theta}\inf_{e^s}\Big\{ & J_t + x\Big(r + \bar{\lambda}S^{\beta}\theta - e^s\theta\Big)J_x + \frac{1}{2}x^2\theta^2 J_{xx} \\ & + \Big(rS + \bar{\lambda}\sigma S^{2\beta+1} - e_t^s\sigma S^{\beta+1}\Big)J_s + \frac{1}{2}\sigma^2 S^{2\beta+2}J_{ss} + x\sigma S^{\beta+1}\theta J_{xs} + \frac{(e^s)^2}{2\Psi}\Big\} = 0, \end{aligned} \tag{16}$$

where J_t, J_x, J_s, J_{xx}, J_{ss}, and J_{xs} denote the partial derivatives of the first and second order with respect to time, stock and wealth, respectively. Moreover, $\Psi = \frac{\phi}{\gamma J}$ (see (9)), where ϕ is a positive ambiguity-aversion parameter.

We first find the infimum, denoted as e^{s_*}, and then differentiate (16) to obtain the optimal exposure as:

$$\Big\{\theta^* \quad = \frac{e^{s_*}J_x - \bar{\lambda}S^{\beta}J_x - \sigma S^{\beta+1}J_{xs}}{xJ_{xx}}.$$

The solution to (16) is provided in the next section.

3.2. Closed-Form Solutions for Hara Utility

To obtain the optimal strategy, we solve HJBI Equation (16) (for additional details, refer to Appendix A). The main proposition is presented below.

Proposition 1. *Assume that $\beta < 0$, $\phi > 0$, and $\phi \neq \gamma - 1$. The solution to Problem (16) is as follows:*

1. *The indirect utility function of an ambiguity and risk averse investor is given by*

$$J(x,S,t) = \frac{\left(x - Fe^{-r(T-t)}\right)^{\gamma}}{\gamma} g^{1/\alpha}(S, T-t), \tag{17}$$

where $\alpha = \frac{\gamma-\phi}{\gamma(\phi-\gamma+1)}$, and we have the following properties:

(a) Function $g(S, T-t)$ is determined by solving the Cauchy problem:

$$\begin{cases} \mathcal{L}g \equiv -g_t + \frac{1}{2}\sigma^2 S^{2\beta+2} g_{ss} + \alpha S\left(r\frac{\gamma(\phi-\gamma+1)}{\gamma-\phi} + \frac{\bar{\lambda}\gamma\sigma S^{2\beta}}{\gamma-\phi}\right)g_s + \frac{\alpha\gamma\left(\bar{\lambda}S^{\beta}\right)^2}{2(\phi-\gamma+1)}g + \alpha\gamma r g = 0, \\ g(S,0) = 1 \text{ when } t = T. \end{cases} \tag{18}$$

(b) Solution $g(S, T-t)$ can be represented as:

$$g(S,T-t) = e^{R\tilde{\tau}+zB(\tilde{\tau})} D^{\delta}(\tilde{\tau}) \frac{\Gamma(\eta-\delta+1/2)}{\Gamma(1+2\eta)} e^{-\frac{z}{2}A(\tilde{\tau})} (zA(\tilde{\tau}))^{\delta} M_{\delta,\eta}(zA(\tilde{\tau})), \tag{19}$$

where $z = \omega S^{-2\beta}$, $\tilde{\tau} = \sigma^2\beta^2\omega(T-t)$, $\Gamma(x)$ is the gamma function, $M_{\delta,\eta}(x)$ is the Whittaker function with parameters

$$\delta = -\frac{1}{2} - \frac{1}{2\beta}\left(\frac{1}{2} - \frac{\alpha\bar{\lambda}\gamma}{\sigma(\gamma-\phi)}\right), \quad \eta = \sqrt{\left(\delta + \frac{1}{2}\right)^2 + \frac{\alpha\gamma\bar{\lambda}^2}{4\sigma^2\beta^2(\gamma-\phi-1)}}.$$

Meanwhile, the remaining constants and functions are given by

$$\omega = \tfrac{r}{\sigma^2|\beta|}, \;\; Q = \tfrac{r}{\sigma^2\beta\omega}, \;\; R = \tfrac{r\alpha\gamma}{\sigma^2\beta^2\omega} - 2Q\delta, \tag{20}$$

$$A(\tilde{\tau}) = \tfrac{1}{2\sinh^2\frac{\tilde{\tau}}{2}(\coth\frac{\tilde{\tau}}{2}+Q)}, \;\; B(\tilde{\tau}) = \tfrac{Q^2-1}{2(\coth\frac{\tilde{\tau}}{2}+Q)}, \;\; D(\tilde{\tau}) = \tfrac{Q^2-1}{4A(\tilde{\tau})B(\tilde{\tau})}. \tag{21}$$

2. *The optimal exposure to the risk factor Z_t is*

$$\theta^* = \frac{\left(x - Fe^{-r(T-t)}\right)S^{\beta}}{x}\left[\frac{\bar{\lambda}}{\phi-\gamma+1} + \sigma S\left(B(\tilde{\tau}) + \frac{\delta+\eta+\frac{1}{2}}{z}\frac{M_{\delta+1,\eta}\left(zA(\tilde{\tau})\right)}{M_{\delta,\eta}\left(zA(\tilde{\tau})\right)}\right)\frac{dz}{dS}\right], \tag{22}$$

which is equivalent to writing

$$\theta^* = \frac{(x - Fe^{-r(T-t)})S^{\beta}}{x}\left[\frac{\bar{\lambda}}{\phi-\gamma+1} - 2\sigma\omega\beta S^{-2\beta}B(\tilde{\tau}) - 2\sigma\beta\left(\delta+\eta+\frac{1}{2}\right)\frac{M_{\delta+1,\eta}\left(\omega A(\tilde{\tau})S^{-2\beta}\right)}{M_{\delta,\eta}\left(\omega A(\tilde{\tau})S^{-2\beta}\right)}\right]. \tag{23}$$

3. *The worst-case measure is determined by:*

$$e^{s_*} = \frac{\phi\bar{\lambda}S^{\beta}}{\phi-\gamma+1} + \frac{\phi\sigma S^{\beta+1}}{\gamma-\phi}\left[B(\tilde{\tau}) + \frac{\delta+\eta+\frac{1}{2}}{z}\frac{M_{\delta+1,\eta}\left(zA(\tilde{\tau})\right)}{M_{\delta,\eta}\left(zA(\tilde{\tau})\right)}\right]\frac{dz}{dS}, \tag{24}$$

which can also be written as follows:

$$e^{s_*} = \tfrac{\phi\bar{\lambda}S^{\beta}}{\phi-\gamma+1} + \tfrac{\phi\sigma S^{\beta}}{\phi-\gamma}\left[-2\omega\beta S^{-2\beta}B(\tilde{\tau}) - 2\beta(\delta+\eta+\tfrac{1}{2})\tfrac{M_{\delta+1,\eta}\left(\omega A(\tilde{\tau})S^{-2\beta}\right)}{M_{\delta,\eta}\left(\omega A(\tilde{\tau})S^{-2\beta}\right)}\right]. \tag{25}$$

Proof. The proof is divided into several steps, all of which are presented in Appendix A.

Appendix A.1 derives the solution to the HJBI equation up to the Cauchy representation of g.

Appendix A.2 details the solution to the Cauchy equation. In particular, Appendix A.2.1 provides the scaling transformations needed to obtain an equation for a new function h. Appendix A.2.2 finds the PDE for h. Appendix A.2.3 applies a Laplace transform to solve the PDE for h. Appendix A.2.4 combines all results to obtain g. Lastly, Appendix A.2.5 computes a ratio involving derivatives of g needed for the next step.

Appendix A.3 uses the previous results to derive the optimal exposure and worst-case measure, denoted as θ^* and e^{s_*}, respectively, which are dedicated to the Whittaker function. □

As indicated by reference [6], in the absence of ambiguity and for CRRA, it is crucial to emphasize that assuming a risk-free interest rate r of zero leads to the following simplifications:

$$\omega = z = \bar{\tau} = 0,$$
$$Q = R = 0.$$

Meanwhile, the limit values from the expressions (21) are provided by:

$$\begin{aligned} &\lim_{\omega \longrightarrow 0} D(\tilde{\tau}) = 1, \\ &\lim_{\omega \longrightarrow 0} zB(\tilde{\tau}) = 0, \\ &\lim_{\omega \longrightarrow 0} R\tilde{\tau} = 0, \\ \varphi(S,t) \quad &:= \lim_{\omega \longrightarrow 0} zA(\tilde{\tau}) = \tfrac{1}{2}\sigma^{-2}\beta^{-2}S^{-2\beta}(T-t)^{-1}. \end{aligned}$$

Consequently, function $g(S,t)$ in Proposition 1 can be simplified, resulting in the following corollary.

Corollary 1. *Assume that the risk-free interest rate of $r = 0$, $\beta < 0$, $\phi > 0$, and $\phi \neq \gamma - 1$.*

1. The indirect utility function is given by:

$$J(x,S,t) = \frac{(x-F)^\gamma}{\gamma} g^{1/\alpha}(S, T-t), \quad \alpha = \frac{\gamma - \phi}{\gamma(\phi - \gamma + 1)}, \tag{26}$$

and we have the following properties:

(a) Function $g(S, T-t)$ solves the Cauchy problem

$$\begin{cases} \mathcal{L}g \equiv -g_t + \frac{1}{2}\sigma^2 S^{2\beta+2} g_{ss} + \frac{\alpha\bar{\lambda}\gamma\sigma S^{2\beta+1}}{\gamma-\phi} g_s + \frac{\alpha\gamma\left(\bar{\lambda}S^\beta\right)^2}{2(\phi-\gamma+1)} g = 0, \\ g(S,0) = 1. \end{cases} \tag{27}$$

(b) Solution $g(S, T-t)$ is given by

$$g(S, T-t) = \frac{\Gamma(\eta - \delta + 1/2)}{\Gamma(1+2\eta)} e^{-\frac{1}{2}\varphi(S,t)} \varphi^\delta(S,t) M_{\delta,\eta}(\varphi(S,t)), \tag{28}$$

where $\Gamma(x)$ is the gamma function, $M_{\delta,\eta}(x)$ is the Whittaker function with parameters

$$\delta = -\frac{1}{2} - \frac{1}{2\beta}\left(\frac{1}{2} - \frac{\alpha\bar{\lambda}\gamma}{\sigma(\gamma-\phi)}\right), \quad \eta = \sqrt{\left(\delta + \frac{1}{2}\right)^2 + \frac{\alpha\gamma\bar{\lambda}^2}{4\sigma^2\beta^2(\gamma-\phi-1)}},$$

and there is a limit value

$$\varphi(S,t) = \lim_{\omega \longrightarrow 0} zA(\tilde{\tau}) = \frac{1}{2}\sigma^{-2}\beta^{-2}S^{-2\beta}(T-t)^{-1}. \tag{29}$$

2. *The optimal exposure is determined by:*

$$\theta^* = \frac{(x-F)S^\beta}{x}\left[\frac{\bar{\lambda}}{\phi-\gamma+1} - 2\sigma\beta(\delta+\eta+\frac{1}{2})\frac{M_{\delta+1,\eta}\big(\varphi(S,t)\big)}{M_{\delta,\eta}\big(\varphi(S,t)\big)}\right]. \tag{30}$$

3. *The worst-case measure can be obtained as:*

$$e^{s_*} = \frac{\phi\bar{\lambda}S^\beta}{\phi-\gamma+1} - \frac{2\phi\sigma\beta S^\beta(\delta+\eta+\frac{1}{2})}{\gamma-\phi}\frac{M_{\delta+1,\eta}\big(\varphi(S,t)\big)}{M_{\delta,\eta}\big(\varphi(S,t)\big)}. \tag{31}$$

Proof. The proof is available in Appendix A.4. □

For further simplicity, we set F equal to 0; this would effectively be a CRRA utility setting. Assuming a risk-free interest rate of $r = 0$, we obtain an even simpler case.

Corollary 2. *Consider risk-free interest rates of $r = 0$, $\beta < 0$, $\phi > 0$, and $\phi \neq \gamma - 1$. We have the following properties.*

1. *The indirect utility function of an ambiguity and risk averse investor is given by:*

$$J(x,S,t) = \frac{x^\gamma}{\gamma}g^{1/\alpha}(S,T-t), \quad \alpha = \frac{\gamma-\phi}{\gamma(\phi-\gamma+1)}, \tag{32}$$

where the function $g(S,t)$ is identical to those in parts (a) and (b) of Corollary 1.

2. *The optimal exposure to the risk factor Z_t is determined by:*

$$\theta^* = \frac{\bar{\lambda}S^\beta}{\phi-\gamma+1} - 2\sigma\beta S^\beta(\delta+\eta+\frac{1}{2})\frac{M_{\delta+1,\eta}\big(\varphi(S,t)\big)}{M_{\delta,\eta}\big(\varphi(S,t)\big)}. \tag{33}$$

Proof. The proofs are analogous to those presented in Corollary 1. □

Note that the Cauchy problem and worst-case scenario are the same as those in Corollary 1

3.3. Corrections to the Crra Case

Our work extends the results of the previous study in presented in reference [14] and in the formal publications [6]. However, discrepancies were identified in the Cauchy problem (2.13), as well as in parameters (3.5) and (3.6). This is related to Theorems 1 and 2 in the manuscript [6]. In this section, we address and rectify these inaccuracies by amalgamating the corrections for the aforementioned issues.

Corollary 3. *(Correction) For the M-CEV model, the value function $J(x,S,t)$ is given by:*

$$J(x,S,t) = \frac{x^\gamma}{\gamma}f^{\frac{1}{\delta}}(S,t), \quad \delta = \frac{1}{1-\gamma}.$$

1. *The function f solves the Cauchy problem,*

$$\begin{cases} \mathcal{L}f(S,t) \equiv f_t + \frac{a^2S^{2\beta+2}}{2}f_{ss} + \delta S\big(\alpha-\gamma r + ca^2S^{2\beta}\big)f_s + \frac{\delta(\delta-1)}{2a^2}\big[(\alpha-r)S^{-\beta} + ca^2S^\beta\big]^2 f + r\gamma\delta f = 0, \\ f(S,T) = 1. \end{cases} \tag{34}$$

2. *The solution of boundary-value problem (34) is the function*

$$f(S,t) = e^{R\tau+zB(\tau)}D^\lambda(\tau)\frac{\Gamma(\eta-\lambda+\frac{1}{2})}{\Gamma(1+2\eta)}e^{-\frac{zA(\tau)}{2}}(zA(\tau))^\lambda M_{\lambda,\eta}(zA(\tau)), \tag{35}$$

where $z = \Lambda S^{-2\beta}$, $\tau = a^2\beta^2\Lambda(T-t)$, $\Gamma(z)$ *is the gamma function,* $M_{\lambda,\eta}(z)$ *is the Whittaker function with parameters*

$$\lambda = -\frac{1}{2} - \frac{1}{2\beta}\left(\frac{1}{2} - \delta c\right), \quad \eta = \sqrt{\left(\lambda + \frac{1}{2}\right)^2 + \frac{\delta(1-\delta)c^2}{4\beta^2}}.$$

The remaining constants and functions are given by:

$$\Lambda = \tfrac{\sqrt{\delta}}{a^2|\beta|}\sqrt{\alpha^2 - \gamma r^2},\ \ Q = \tfrac{\delta(\alpha-\gamma r)}{\Lambda a^2\beta},\ \ R = \tfrac{r\gamma\delta}{a^2\beta^2\Lambda} - 2Q\lambda - \tfrac{\delta(1-\delta)(\alpha-r)c}{\Lambda a^2\beta^2},$$
$$A(\tau) = \tfrac{1}{2\sinh^2\tau(\coth\tau+Q)},\ \ B(\tau) = \tfrac{Q^2-1}{2(\coth\tau+Q)},\ \ D(\tau) = \tfrac{Q^2-1}{4A(\tau)B(\tau)}.$$

3. *The optimal investor strategy is:*

$$\pi^*(X,S,t) = X\left[\delta\tfrac{\alpha-r+ca^2S^{2\beta}}{a^2S^{2\beta+1}} + \left(B(\tau) + \tfrac{\left(\lambda+\eta+\frac{1}{2}\right)M_{\lambda+1,\eta}\left(A(\tau)z\right)}{zM_{\lambda,\eta}\left(A(\tau)z\right)}\right)\tfrac{dz}{dS}\right]. \tag{36}$$

Based on the remark in reference [6], and assuming that $\alpha = r\sqrt{\gamma}$, it is easy to see that $\alpha^2 - \gamma r^2 = 0$. Consequently, $\Lambda = \frac{\sqrt{\delta}}{a^2|\beta|}\sqrt{\alpha^2 - \gamma r^2} = 0$. Therefore, the corrected outcomes are as follows:

Corollary 4. *If* $\alpha = r\sqrt{\gamma}$, *Formulas (35) and (36) are simplified. In this case, the following properties exist:*

1.
$$\Lambda = \tau = z = 0,$$

and the limit values are

$$\lim_{\Lambda\longrightarrow 0} D(\tau) = 1,\ \ \theta(S,t) = \lim_{\Lambda\longrightarrow 0} zA(\tau) = 0.5a^{-2}\beta^{-2}S^{-2\beta}(T-t)^{-1},$$
$$\Omega(S,t) = \lim_{\Lambda\longrightarrow 0} zB(\tau) = \tfrac{1}{2S^{2\beta}}\tfrac{\delta^2(\alpha-\gamma r)^2(T-t)}{a^2\left[1+\beta\delta(\alpha-\gamma r)(T-t)\right]},$$
$$\Psi(t) = \lim_{\Lambda\longrightarrow 0} R\tau = r\gamma\delta(T-t) - 2\delta\beta\lambda(\alpha-\gamma r)(T-t) - \tfrac{\delta(1-\delta)c(\alpha-r)(T-t)}{a^2}.$$

2. *The solution* $f(S,t)$ *is given by:*

$$f(S,t) = \frac{\Gamma(\eta-\lambda+\frac{1}{2})}{\Gamma(1+2\eta)}e^{\Psi(t)+\Omega(S,t)-\frac{1}{2}\theta(S,t)}\theta^{\lambda}(S,t)M_{\lambda,\eta}(\theta(S,t)). \tag{37}$$

3. *The optimal policy is:*

$$\pi^*(X,S,t) = X\left[\delta\frac{\alpha-r+ca^2S^{2\beta}}{a^2S^{2\beta+1}} - \frac{2\beta\left(\lambda+\eta+\frac{1}{2}\right)}{S}\frac{M_{\lambda+1,\eta}\left(\theta(S,t)\right)}{M_{\lambda,\eta}\left(\theta(S,t)\right)}\right].$$

4. Discussion

Our findings reveal closed-form solutions to all elements of interest to a financial investor: optimal investment allocation on the risky asset, optimal alternative model due to ambiguity, optimal wealth process evolution, and finally, the expression for the value function (i.e., the value of the objective function at the optimal). This is very rare outside exponentially linear structures, which are also known as affine solutions.

Our choice of model, the M-CEV, has the advantage of keeping only one source of risk, and therefore, a lower parametric space; this is ideal for practitioners, always searching for a combination of realism and simplicity.

Our methodology to find the solution involved many transformations of the original PDE problem (from the HJBI equation), resulting in analytical expressions for a non-trivial Cauchy problem. The details provided in our work can serve as a foundation for future

research in dealing with complex PDEs and boundary conditions. Thanks to the detailed analysis, we could detect and correct problems in previously known solutions to embedded problems (M-CEV in the absence of ambiguity with a CRRA utility).

This work has the potential for many extensions and applications. First, other CEV models in the literature have provided solutions (in some cases, approximations) to the embedded problem of CRRA with no ambiguity. The methodology developed here can be used to solve such problems analytically, while extending to HARA and ambiguity. Once this is achieved, an empirical comparison would help shed light on the importance of the various models in the presence/absence of ambiguity as well as in the presence/absence of a floor on wealth (HARA versus CRRA).

5. Ethical Compliance

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Author Contributions: W.L.F.: Conceptualization, methodology, validation, formal analysis, investigation, writing—original draft preparation, writing—review and editing, project administration. M.E.A.: Conceptualization, methodology, validation, formal analysis, investigation, writing—original draft preparation, writing—review and editing, supervision, project administration. All authors have read and agreed to the published version of the manuscript.

Funding: This study received no external funding.

Data Availability Statement: The study did not require data.

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

Appendix A. Proof of Proposition 1

Appendix A.1. Proof of Proposition 1 up to the Cauchy Equation

We now solve the HJBI Equation (16) to determine the optimal investment strategy in the M-CEV model. To facilitate the solution later, we parameterize the change in measure, defined as

$$\epsilon_t := \frac{e_t^s}{S_t^\beta}.$$

The modified SDEs (6) and (15) can be rewritten as

$$\frac{dS_t}{S_t} = \left[r + (\bar{\lambda} - \epsilon_t)\sigma S_t^{2\beta}\right]\mathrm{d}t + \sigma S_t^\beta \mathrm{d}Z_t, \tag{A1}$$

$$\begin{aligned}\mathrm{d}X_t &= X_t\Big(r + \bar{\lambda}\theta_t S_t^\beta - \epsilon_t\theta_t S_t^\beta\Big)\mathrm{d}t + \theta_t X_t \mathrm{d}Z_t \\ &= X_t\Big(r + (\bar{\lambda} - \epsilon_t)\theta_t S_t^\beta\Big)\mathrm{d}t + \theta_t X_t \mathrm{d}Z_t.\end{aligned} \tag{A2}$$

Accordingly, the HJBI Equation (16) becomes

$$\begin{aligned}&\sup_\theta \inf_\epsilon \Big\{ J_t + x\Big(r + \bar{\lambda}S^\beta\theta - \epsilon S^\beta\theta\Big)J_x + \frac{1}{2}x^2\theta^2 J_{xx} \\ &+ \Big(rS + \bar{\lambda}\sigma S^{2\beta+1} - \epsilon\sigma S^{2\beta+1}\Big)J_s + \frac{1}{2}\sigma^2 S^{2\beta+2}J_{ss} + x\sigma S^{\beta+1}\theta J_{xs} + \frac{(\epsilon S^\beta)^2}{2\Psi}\Big\} = 0.\end{aligned} \tag{A3}$$

First and foremost, we begin by solving the minimization problem as follows:

$$\begin{aligned}&- xS^\beta\theta J_x - \sigma S^{2\beta+1}J_s + \frac{\epsilon S^{2\beta}}{\Psi} = 0 \\ &\Longrightarrow \epsilon^* = \Psi\big(x\theta S^{-\beta}J_x + \sigma S J_s\big).\end{aligned} \tag{A4}$$

Substituting the value of ϵ^* into Equation (A3) yields

$$\begin{aligned}
&\sup_{\theta}\Big\{J_t+\Big(xr+x\bar{\lambda}S^{\beta}\theta-x\Psi\big(x\theta S^{-\beta}J_x+\sigma SJ_s\big)S^{\beta}\theta\Big)J_x+\frac{1}{2}x^2\theta^2J_{xx}\\
&+\Big[rS+\bar{\lambda}\sigma S^{2\beta+1}-\Psi\big(x\theta S^{-\beta}J_x+\sigma SJ_s\big)\sigma S^{2\beta+1}\Big]J_s+\frac{1}{2}\sigma^2S^{2\beta+2}J_{ss}+x\sigma S^{\beta+1}\theta J_{xs}\\
&+\frac{\Psi^2\big(x\theta S^{-\beta}J_x+\sigma SJ_s\big)^2S^{2\beta}}{2\Psi}\Big\}=0\\
\Longrightarrow&\sup_{\theta}\Big\{J_t+xrJ_x+x\bar{\lambda}S^{\beta}\theta J_x-\Psi x^2\theta^2J_x^2-\Psi x\sigma\theta S^{\beta+1}J_sJ_x+\frac{1}{2}x^2\theta^2J_{xx}\\
&+rSJ_s+\bar{\lambda}\sigma S^{2\beta+1}J_s-\Psi x\sigma\theta S^{\beta+1}J_xJ_s-\Psi\sigma^2S^{2\beta+2}J_s^2+\frac{1}{2}\sigma^2S^{2\beta+2}J_{ss}+x\sigma S^{\beta+1}\theta J_{xs}\\
&+\frac{\Psi}{2}\Big(x^2\theta^2J_x^2+\sigma^2S^{2\beta+2}J_s^2+2x\theta\sigma S^{\beta+1}J_xJ_s\Big)\Big\}=0. \qquad \text{(A5)}
\end{aligned}$$

Applying the Bellman principle, the value function satisfies the HJBI equation with the terminal condition $J(x,T)=\frac{(x-F)^{\gamma}}{\gamma}$. Since $\Psi=\frac{\phi}{\gamma J}$ where $\phi>0$ (as shown in Formula (9)), substituting this into Equation (A5) results in

$$\begin{aligned}
&\sup_{\theta}\Big\{J_t+xrJ_x+x\bar{\lambda}S^{\beta}\theta J_x-\frac{1}{2}\frac{\phi}{\gamma J}x^2\theta^2J_x^2-\frac{\phi}{\gamma J}x\sigma\theta S^{\beta+1}J_sJ_x+\frac{1}{2}x^2\theta^2J_{xx}\\
&+rSJ_s+\bar{\lambda}\sigma S^{2\beta+1}J_s-\frac{1}{2}\frac{\phi}{\gamma J}\sigma^2S^{2\beta+2}J_s^2+\frac{1}{2}\sigma^2S^{2\beta+2}J_{ss}+x\sigma S^{\beta+1}\theta J_{xs}\Big\}=0. \qquad \text{(A6)}
\end{aligned}$$

Hence, the first-order conditions lead to

$$\begin{aligned}
&x\bar{\lambda}S^{\beta}J_x-\frac{\phi}{\gamma J}x^2J_x^2\theta-\frac{\phi}{\gamma J}x\sigma S^{\beta+1}J_sJ_x+x^2J_{xx}\theta+x\sigma S^{\beta+1}J_{xs}=0\\
\Longleftrightarrow\ &\theta=\frac{x\bar{\lambda}S^{\beta}J_x-\frac{\phi}{\gamma J}x\sigma S^{\beta+1}J_sJ_x+x\sigma S^{\beta+1}J_{xs}}{\frac{\phi}{\gamma J}x^2J_x^2-x^2J_{xx}}\\
&=\frac{\gamma\bar{\lambda}S^{\beta}J_xJ-\phi\sigma S^{\beta+1}J_sJ_x+\gamma\sigma S^{\beta+1}J_{xs}J}{\phi xJ_x^2-\gamma xJ_{xx}J}. \qquad \text{(A7)}
\end{aligned}$$

Now, substituting back into Equation (A6) gives

$$\begin{aligned}
&J_t+xrJ_x+\frac{\gamma\bar{\lambda}^2S^{2\beta}J_xJ-\phi\bar{\lambda}\sigma S^{2\beta+1}J_sJ_x+\gamma\bar{\lambda}\sigma S^{2\beta+1}J_{xs}J}{\phi J_x^2-\gamma J_{xx}J}J_x\\
&-\frac{1}{2}\frac{\phi}{\gamma J}\left(\frac{\gamma\bar{\lambda}S^{\beta}J_xJ-\phi\sigma S^{\beta+1}J_sJ_x+\gamma\sigma S^{\beta+1}J_{xs}J}{\phi J_x^2-\gamma J_{xx}J}\right)^2J_x^2\\
&-\frac{\phi}{\gamma J}\frac{\gamma\bar{\lambda}\sigma S^{2\beta+1}J_xJ-\phi\sigma^2S^{2\beta+2}J_sJ_x+\gamma\sigma^2S^{2\beta+2}J_{xs}J}{\phi J_x^2-\gamma J_{xx}J}J_sJ_x\\
&+\frac{1}{2}\left(\frac{\gamma\bar{\lambda}S^{\beta}J_xJ-\phi\sigma S^{\beta+1}J_sJ_x+\gamma\sigma S^{\beta+1}J_{xs}J}{\phi J_x^2-\gamma J_{xx}J}\right)^2J_{xx}+rSJ_s+\bar{\lambda}\sigma S^{2\beta+1}J_s-\frac{1}{2}\frac{\phi}{\gamma J}\sigma^2S^{2\beta+2}J_s^2\\
&+\frac{1}{2}\sigma^2S^{2\beta+2}J_{ss}+\frac{\gamma\bar{\lambda}\sigma S^{2\beta+1}J_xJ-\phi\sigma^2S^{2\beta+2}J_sJ_x+\gamma\sigma^2S^{2\beta+2}J_{xs}J}{\phi J_x^2-\gamma J_{xx}J}J_{xs}=0. \qquad \text{(A8)}
\end{aligned}$$

To find the solution, we employ the separation ansatz

$$J(x,S,t)=\frac{\big(x-Fe^{-r(T-t)}\big)^{\gamma}}{\gamma}g^{1/\alpha}(S,T-t),\quad \alpha=\frac{\gamma-\phi}{\gamma(\phi-\gamma+1)}, \qquad \text{(A9)}$$

where $\phi \neq \gamma - 1$, and the function $g(S, T-t)$ satisfies the boundary condition $g(S,0) = 1$ when $t = T$.

Thus, the derivatives of the function $J(x, S, t)$ are computed as

$$\begin{aligned}
J_t &= -rFe^{-r(T-t)}\Big(x - Fe^{-r(T-t)}\Big)^{\gamma-1} g^{1/\alpha}(S, T-t) \\
&- \alpha^{-1}\frac{g_t(S,T-t)}{g(S,T-t)}\frac{\big(x - Fe^{-r(T-t)}\big)^{\gamma}}{\gamma} g^{1/\alpha}(S, T-t) \\
&= \left(\frac{-rFe^{-r(T-t)}\gamma}{x - Fe^{-r(T-t)}} - \alpha^{-1}\frac{g_t}{g}\right) J, \\
J_x &= \Big(x - Fe^{-r(T-t)}\Big)^{\gamma-1} g^{1/\alpha}(S, T-t) = \frac{\gamma}{x - Fe^{-r(T-t)}} J, \\
J_{xx} &= (\gamma - 1)\Big(x - Fe^{-r(T-t)}\Big)^{\gamma-2} g^{1/\alpha}(S, T-t) = \frac{\gamma(\gamma-1)}{\big(x - Fe^{-r(T-t)}\big)^2} J, \\
J_s &= \alpha^{-1}\frac{g_s(S,T-t)}{g(S,T-t)}\frac{\big(x - Fe^{-r(T-t)}\big)^{\phi-\gamma}}{\phi - \gamma} g^{1/\alpha}(S, T-t) = \alpha^{-1}\frac{g_s}{g} J, \\
J_{ss} &= \alpha^{-1}\frac{g_s}{g} J_s + \alpha^{-1}\left(\frac{g_{ss}}{g} - \frac{g_s^2}{g^2}\right) J = \alpha^{-2}\frac{g_s^2}{g^2} J + \alpha^{-1}\left(\frac{g_{ss}}{g} - \frac{g_s^2}{g^2}\right) J \\
&= \left[(\alpha^{-2} - \alpha^{-1})\frac{g_s^2}{g^2} + \alpha^{-1}\frac{g_{ss}}{g}\right] J, \\
J_{xs} &= \frac{\gamma}{x - Fe^{-r(T-t)}} J_s = \alpha^{-1}\frac{g_s}{g}\left(\frac{\gamma}{x - Fe^{-r(T-t)}}\right) J.
\end{aligned}$$

Referring to (A7), we denote the A and B as shown below for ease of computation:

$$\begin{aligned}
A &:= \gamma\bar{\lambda} S^{\beta} J_x J - \phi\sigma S^{\beta+1} J_s J_x + \gamma\sigma S^{\beta+1} J_{xs} J, \\
B &:= \phi x J_x^2 - \gamma x J_{xx} J.
\end{aligned}$$

It follows that

$$\begin{aligned}
A &= \gamma\bar{\lambda} S^{\beta} J_x J - \phi\sigma S^{\beta+1} J_s J_x + \gamma\sigma S^{\beta+1} J_{xs} J \\
&= \left(\tfrac{\gamma}{x - Fe^{-r(T-t)}}\right)\left(\gamma\bar{\lambda} - \phi\sigma\alpha^{-1} S\tfrac{g_s}{g} + \gamma\sigma\alpha^{-1} S\tfrac{g_s}{g}\right) S^{\beta} J^2, \quad \text{and} \\
B &= \phi x J_x^2 - \gamma x J_{xx} J \\
&= \tfrac{\gamma^2\phi x - \gamma^2 x(\gamma-1)}{\left(x - Fe^{-r(T-t)}\right)^2} J^2,
\end{aligned}$$

which implies that the optimal exposure to the risk factor Z_t is given by

$$\begin{aligned}
\theta^* &= \frac{\gamma\bar{\lambda} S^{\beta} J_x J - \phi\sigma S^{\beta+1} J_s J_x + \gamma\sigma S^{\beta+1} J_{xs} J}{\phi x J_x^2 - \gamma x J_{xx} J} = \frac{A}{B} \\
&= \frac{\left(\tfrac{\gamma}{x - Fe^{-r(T-t)}}\right)\left(\gamma\bar{\lambda} - \phi\sigma\alpha^{-1} S\tfrac{g_s}{g} + \gamma\sigma\alpha^{-1} S\tfrac{g_s}{g}\right) S^{\beta} J^2}{\tfrac{\gamma^2\phi x - \gamma^2 x(\gamma-1)}{\left(x - Fe^{-r(T-t)}\right)^2} J^2} \\
&= \frac{\left[\gamma\bar{\lambda} S^{\beta} - (\phi - \gamma)\sigma\alpha^{-1} S^{\beta+1}\tfrac{g_s}{g}\right]}{\gamma(\phi - \gamma + 1)} \frac{\left(x - Fe^{-r(T-t)}\right)}{x}.
\end{aligned} \tag{A10}$$

Next, we substitute (A9) and its corresponding partial differential equations into Equation (A8) to derive (only the key steps are shown).

$$
\begin{aligned}
&\left(\frac{-rFe^{-r(T-t)}\gamma}{x-Fe^{-r(T-t)}}-\alpha^{-1}\frac{g_t}{g}\right)J+\frac{\gamma xr}{\left(x-Fe^{-r(T-t)}\right)}J\\
&+\left[\frac{\gamma\bar{\lambda}^2S^{2\beta}\frac{\gamma}{\left(x-Fe^{-r(T-t)}\right)}J^2-(\phi-\gamma)\bar{\lambda}\sigma\alpha^{-1}S^{2\beta+1}\frac{g_s}{g}\frac{\gamma}{\left(x-Fe^{-r(T-t)}\right)}J^2}{\phi\frac{\gamma^2}{\left(x-Fe^{-r(T-t)}\right)^2}J^2-\frac{\gamma^2(\gamma-1)}{\left(x-Fe^{-r(T-t)}\right)^2}J^2}\right]\frac{\gamma}{\left(x-Fe^{-r(T-t)}\right)}J\\
&-\frac{\phi}{2\gamma J}\left[\frac{\gamma\bar{\lambda}S^{\beta}\frac{\gamma}{\left(x-Fe^{-r(T-t)}\right)}J^2-(\phi-\gamma)\sigma\alpha^{-1}S^{\beta+1}\frac{g_s}{g}\frac{\gamma}{\left(x-Fe^{-r(T-t)}\right)}J^2}{\phi\frac{\gamma^2}{\left(x-Fe^{-r(T-t)}\right)^2}J^2-\frac{\gamma^2(\gamma-1)}{\left(x-Fe^{-r(T-t)}\right)^2}J^2}\right]^2\frac{\gamma^2}{\left(x-Fe^{-r(T-t)}\right)^2}J^2\\
&-\frac{\phi}{\gamma J}\left[\frac{\gamma\bar{\lambda}\sigma S^{2\beta+1}\frac{\gamma}{\left(x-Fe^{-r(T-t)}\right)}J^2-(\phi-\gamma)\sigma^2\alpha^{-1}S^{2\beta+2}\frac{g_s}{g}\frac{\gamma}{\left(x-Fe^{-r(T-t)}\right)}J^2}{\phi\frac{\gamma^2}{\left(x-Fe^{-r(T-t)}\right)^2}J^2-\frac{\gamma^2(\gamma-1)}{\left(x-Fe^{-r(T-t)}\right)^2}J^2}\right]\frac{\alpha^{-1}\frac{g_s}{g}\gamma}{\left(x-Fe^{-r(T-t)}\right)}J^2\\
&+\frac{1}{2}\left[\frac{\gamma\bar{\lambda}S^{\beta}\frac{\gamma}{\left(x-Fe^{-r(T-t)}\right)}J^2-(\phi-\gamma)\sigma\alpha^{-1}S^{\beta+1}\frac{g_s}{g}\frac{\gamma}{\left(x-Fe^{-r(T-t)}\right)}J^2}{\phi\frac{\gamma^2}{\left(x-Fe^{-r(T-t)}\right)^2}J^2-\frac{\gamma^2(\gamma-1)}{\left(x-Fe^{-r(T-t)}\right)^2}J^2}\right]^2\frac{\gamma(\gamma-1)}{\left(x-Fe^{-r(T-t)}\right)^2}J\\
&+rS\alpha^{-1}\frac{g_s}{g}J+\bar{\lambda}\sigma\alpha^{-1}S^{2\beta+1}\frac{g_s}{g}J-\frac{1}{2}\frac{\phi}{\gamma J}\sigma^2\alpha^{-2}S^{2\beta+2}\frac{g_s^2}{g^2}J^2+\frac{1}{2}\sigma^2S^{2\beta+2}\left[\left(\alpha^{-2}-\alpha^{-1}\right)\frac{g_s^2}{g^2}+\alpha^{-1}\frac{g_{ss}}{g}\right]J\\
&+\left[\frac{\gamma\bar{\lambda}\sigma S^{2\beta+1}\frac{\gamma}{\left(x-Fe^{-r(T-t)}\right)}J^2-(\phi-\gamma)\sigma^2\alpha^{-1}S^{2\beta+2}\frac{g_s}{g}\frac{\gamma}{\left(x-Fe^{-r(T-t)}\right)}J^2}{\phi\frac{\gamma^2}{\left(x-Fe^{-r(T-t)}\right)^2}J^2-\frac{\gamma^2(\gamma-1)}{\left(x-Fe^{-r(T-t)}\right)^2}J^2}\right]\frac{\alpha^{-1}\frac{g_s}{g}\gamma}{\left(x-Fe^{-r(T-t)}\right)}J=0\\
\Longrightarrow\ &\left(\frac{-rFe^{-r(T-t)}\gamma}{x-Fe^{-r(T-t)}}-\alpha^{-1}\frac{g_t}{g}\right)+\frac{\gamma xr}{\left(x-Fe^{-r(T-t)}\right)}\\
&+\left[\frac{\gamma\bar{\lambda}^2S^{2\beta}\frac{\gamma}{\left(x-Fe^{-r(T-t)}\right)}-(\phi-\gamma)\bar{\lambda}\sigma\alpha^{-1}S^{2\beta+1}\frac{g_s}{g}\frac{\gamma}{\left(x-Fe^{-r(T-t)}\right)}}{\phi\frac{\gamma^2}{\left(x-Fe^{-r(T-t)}\right)^2}-\frac{\gamma^2(\gamma-1)}{\left(x-Fe^{-r(T-t)}\right)^2}}\right]\frac{\gamma}{\left(x-Fe^{-r(T-t)}\right)}\\
&-\frac{\phi}{2\gamma}\left[\frac{\gamma\bar{\lambda}S^{\beta}\frac{\gamma}{\left(x-Fe^{-r(T-t)}\right)}-(\phi-\gamma)\sigma\alpha^{-1}S^{\beta+1}\frac{g_s}{g}\frac{\gamma}{\left(x-Fe^{-r(T-t)}\right)}}{\phi\frac{\gamma^2}{\left(x-Fe^{-r(T-t)}\right)^2}-\frac{\gamma^2(\gamma-1)}{\left(x-Fe^{-r(T-t)}\right)^2}}\right]^2\frac{\gamma^2}{\left(x-Fe^{-r(T-t)}\right)^2}\\
&-\frac{\phi}{\gamma}\left[\frac{\gamma\bar{\lambda}\sigma S^{2\beta+1}\frac{\gamma}{\left(x-Fe^{-r(T-t)}\right)}-(\phi-\gamma)\sigma^2\alpha^{-1}S^{2\beta+2}\frac{g_s}{g}\frac{\gamma}{\left(x-Fe^{-r(T-t)}\right)}}{\phi\frac{\gamma^2}{\left(x-Fe^{-r(T-t)}\right)^2}-\frac{\gamma^2(\gamma-1)}{\left(x-Fe^{-r(T-t)}\right)^2}}\right]\frac{\alpha^{-1}\frac{g_s}{g}\gamma}{\left(x-Fe^{-r(T-t)}\right)}\\
&+\frac{1}{2}\left[\frac{\gamma\bar{\lambda}S^{\beta}\frac{\gamma}{\left(x-Fe^{-r(T-t)}\right)}-(\phi-\gamma)\sigma\alpha^{-1}S^{\beta+1}\frac{g_s}{g}\frac{\gamma}{\left(x-Fe^{-r(T-t)}\right)}}{\phi\frac{\gamma^2}{\left(x-Fe^{-r(T-t)}\right)^2}-\frac{\gamma^2(\gamma-1)}{\left(x-Fe^{-r(T-t)}\right)^2}}\right]^2\frac{\gamma(\gamma-1)}{\left(x-Fe^{-r(T-t)}\right)^2}\\
&+rS\alpha^{-1}\frac{g_s}{g}+\bar{\lambda}\sigma\alpha^{-1}S^{2\beta+1}\frac{g_s}{g}-\frac{1}{2}\frac{\phi}{\gamma}\sigma^2\alpha^{-2}S^{2\beta+2}\frac{g_s^2}{g^2}+\frac{1}{2}\sigma^2S^{2\beta+2}\left[\left(\alpha^{-2}-\alpha^{-1}\right)\frac{g_s^2}{g^2}+\alpha^{-1}\frac{g_{ss}}{g}\right]\\
&+\left[\frac{\gamma\bar{\lambda}\sigma S^{2\beta+1}\frac{\gamma}{\left(x-Fe^{-r(T-t)}\right)}-(\phi-\gamma)\sigma^2\alpha^{-1}S^{2\beta+2}\frac{g_s}{g}\frac{\gamma}{\left(x-Fe^{-r(T-t)}\right)}}{\phi\frac{\gamma^2}{\left(x-Fe^{-r(T-t)}\right)^2}-\frac{\gamma^2(\gamma-1)}{\left(x-Fe^{-r(T-t)}\right)^2}}\right]\frac{\alpha^{-1}\frac{g_s}{g}\gamma}{\left(x-Fe^{-r(T-t)}\right)}=0\\
\Longrightarrow\ &-\alpha^{-1}\frac{g_t}{g}+\gamma r+\frac{\gamma\bar{\lambda}^2S^{2\beta}-(\phi-\gamma)\bar{\lambda}\sigma\alpha^{-1}S^{2\beta+1}\frac{g_s}{g}}{\phi-\gamma+1}-\frac{\phi}{2\gamma}\left[\frac{\gamma\bar{\lambda}S^{\beta}-(\phi-\gamma)\sigma\alpha^{-1}S^{\beta+1}\frac{g_s}{g}}{\phi-\gamma+1}\right]^2\\
&-\frac{\phi}{\gamma}\left[\frac{\gamma\bar{\lambda}\sigma\alpha^{-1}S^{2\beta+1}\frac{g_s}{g}-(\phi-\gamma)\sigma^2\alpha^{-2}S^{2\beta+2}\frac{g_s^2}{g^2}}{\phi-\gamma+1}\right]+\frac{\gamma-1}{2\gamma}\left[\frac{\gamma\bar{\lambda}S^{\beta}-(\phi-\gamma)\sigma\alpha^{-1}S^{\beta+1}\frac{g_s}{g}}{\phi-\gamma+1}\right]^2\\
&+\left(rS\alpha^{-1}+\bar{\lambda}\sigma\alpha^{-1}S^{2\beta+1}\right)\frac{g_s}{g}-\frac{1}{2}\left(\frac{\phi}{\gamma}\sigma^2\alpha^{-2}S^{2\beta+2}-\sigma^2\alpha^{-2}S^{2\beta+2}+\sigma^2\alpha^{-1}S^{2\beta+2}\right)\frac{g_s^2}{g^2}\\
&+\frac{1}{2}\sigma^2\alpha^{-1}S^{2\beta+2}\frac{g_{ss}}{g}+\frac{\gamma\bar{\lambda}\sigma\alpha^{-1}S^{2\beta+1}\frac{g_s}{g}-(\phi-\gamma)\sigma^2\alpha^{-2}S^{2\beta+2}\frac{g_s^2}{g^2}}{\phi-\gamma+1}=0\\
\Longrightarrow\ &-\alpha^{-1}\frac{g_t}{g}+\gamma r+\left(rS\alpha^{-1}+\bar{\lambda}\sigma\alpha^{-1}S^{2\beta+1}\right)\frac{g_s}{g}+\frac{1}{2}\sigma^2\alpha^{-1}S^{2\beta+2}\frac{g_{ss}}{g}\\
&+\left(\gamma^2\bar{\lambda}^2S^{2\beta}+(\gamma-\phi)^2\sigma^2\alpha^{-2}S^{2\beta+2}\frac{g_s^2}{g^2}+2\gamma\bar{\lambda}(\gamma-\phi)\sigma\alpha^{-1}S^{\beta+1}\frac{g_s}{g}\right)\frac{1}{2\gamma(\phi-\gamma+1)}\\
&-\frac{1}{2}\left(\frac{\phi}{\gamma}\sigma^2\alpha^{-2}S^{2\beta+2}-\sigma^2\alpha^{-2}S^{2\beta+2}+\sigma^2\alpha^{-1}S^{2\beta+2}\right)\frac{g_s^2}{g^2}=0\\
\Longrightarrow\ &-\frac{g_t}{g}+\frac{1}{2}\alpha\sigma^2\alpha^{-1}S^{2\beta+2}\frac{g_{ss}}{g}+\alpha\left(rS\alpha^{-1}+\frac{\bar{\lambda}\sigma\alpha^{-1}S^{2\beta+1}}{\phi-\gamma+1}\right)\frac{g_s}{g}+\alpha\frac{\gamma\left(\bar{\lambda}S^{\beta}\right)^2}{2(\phi-\gamma+1)}+\alpha\gamma r\\
&+\alpha\left[\frac{(\gamma-\phi)\sigma^2\alpha^{-2}S^{2\beta+2}}{2\gamma(\phi-\gamma+1)}-\frac{1}{2}\sigma^2\alpha^{-1}S^{2\beta+2}\right]\frac{g_s^2}{g^2}=0
\end{aligned}
$$

$$\Longrightarrow -g_t+\frac{1}{2}\sigma^2S^{2\beta+2}g_{ss}+\alpha S\Big(r\frac{\gamma(\phi-\gamma+1)}{\gamma-\phi}+\frac{\bar{\lambda}\sigma\gamma S^{2\beta}}{\gamma-\phi}\Big)g_s+\frac{\alpha\gamma\big(\bar{\lambda}S^{\beta}\big)^2}{2(\phi-\gamma+1)}g+\alpha\gamma rg=0, \quad \text{(A11)}$$

where $\phi-\gamma+1\neq 0$, and the boundary condition is $g(S,0)=1$.

Notably, when $\phi=0$, Equation (A11) is simplified to the following form:

$$-g_t+\frac{1}{2}\sigma^2S^{2\beta+2}g_{ss}+\alpha S\Big(r-\gamma r+\bar{\lambda}\sigma S^{2\beta}\Big)g_s+\frac{\alpha(\alpha-1)}{2}\big(\bar{\lambda}S^{\beta}\big)^2g+r\gamma\alpha g=0 \quad \text{(A12)}$$

where $1-\gamma\neq 0$. This corresponds to the corrected Cauchy problem (34) in Corollary 3 which was also documented in the reference [14]. The following notations are used from the source to our notation: $a=\sigma$, $\delta=\alpha$, $\alpha=r$, and $c=\frac{\bar{\lambda}}{\sigma}$. Thus, our robustness problem can be reduced to a non-robust scenario within the context of the M-CEV model by forcing $e_t=0$ or equivalently $\phi=0$.

In accordance with Theorem 2 from the reference [6], we can establish that function $g(S,T-t)$ satisfies the Cauchy problem defined by Equation (A11) when the terminal function is set as $g(S,0)=1$. To facilitate a connection to the source, we can define $f(S,t)=g(S,T-t)$. The Cauchy problem (A11) can be rewritten as

$$\begin{cases}\mathcal{L}f(S,t)\equiv f_t+\frac{1}{2}\sigma^2S^{2\beta+2}f_{ss}+\alpha S\Big(r\frac{\gamma(\phi-\gamma+1)}{\gamma-\phi}+\frac{\bar{\lambda}\gamma\sigma S^{2\beta}}{\gamma-\phi}\Big)f_s+\frac{\alpha\gamma\left(\bar{\lambda}S^{\beta}\right)^2}{2(\phi-\gamma+1)}f+\alpha\gamma rf=0,\\ f(S,T)=1.\end{cases} \quad \text{(A13)}$$

Appendix A.2. Solving Cauchy Problem (A13) for the M-CEV Case with Ambiguity

Appendix A.2.1. Scaling Transformation

In Appendix A.1, we derive the Cauchy problem (A13). Assume that $f(S,t)$ is the solution to Cauchy problem (A13). To simplify our analysis, we introduce scaled space and inverse variables, denoted as z and $\tilde{\tau}$, respectively. At first, let

$$\begin{aligned}&z=\omega S^x\Longrightarrow S=\big(\tfrac{z}{\omega}\big)^{\frac{1}{x}},\ \ \omega=?,\ \ \tilde{\tau}=B(T-t)\Longrightarrow t=T-\tfrac{\tilde{\tau}}{B},\\&\tfrac{\partial z}{\partial S}=x\omega S^{x-1},\ \ \tfrac{\partial\tilde{\tau}}{\partial t}=-B,\\&\tfrac{\partial}{\partial t}f=\tfrac{\partial}{\partial\tilde{\tau}}f\tfrac{\partial\tilde{\tau}}{\partial t}=-B\tfrac{\partial}{\partial\tilde{\tau}}f,\\&\tfrac{\partial}{\partial S}f=\tfrac{\partial}{\partial z}f\tfrac{\partial z}{\partial S}=x\omega S^{x-1}\tfrac{\partial}{\partial z}f=x\omega\big(\tfrac{z}{\omega}\big)^{1-\frac{1}{x}}\tfrac{\partial}{\partial z}f=x\omega^{\frac{1}{x}}z^{1-\frac{1}{x}}\tfrac{\partial}{\partial z}f,\\&\tfrac{\partial^2}{\partial S^2}f=x\omega^{\frac{1}{x}}\Big(z^{1-\frac{1}{x}}\tfrac{\partial}{\partial z}f\Big)_S=x\omega^{\frac{1}{x}}\Big[z^{1-\frac{1}{x}}\Big(\tfrac{\partial}{\partial z}f\Big)_S+\tfrac{\partial}{\partial z}f\Big(z^{1-\frac{1}{x}}\Big)_S\Big]\\&=x\omega^{\frac{1}{x}}\Big[z^{1-\frac{1}{x}}\tfrac{\partial^2}{\partial z^2}f\tfrac{\partial z}{\partial S}+\tfrac{\partial}{\partial z}f\Big(1-\tfrac{1}{x}\Big)z^{1-\frac{1}{x}-1}\tfrac{\partial z}{\partial S}\Big]\\&=x^2\omega^{\frac{2}{x}}z^{2-\frac{2}{x}}\tfrac{\partial^2}{\partial z^2}f+\big(x^2-x\big)\omega^{\frac{2}{x}}z^{1-\frac{2}{x}}\tfrac{\partial}{\partial z}f.\end{aligned}$$

According to (A13), the scaling transformed PDE of $f(S,t)$ is given by (only the key steps are shown):

$$\begin{aligned}&\tfrac{\partial}{\partial t}f(S,t)+\tfrac{1}{2}\sigma^2S^{2\beta+2}\tfrac{\partial^2}{\partial S^2}f(S,t)+\alpha S\Big(r\alpha^{-1}+\tfrac{\bar{\lambda}\gamma\sigma}{\gamma-\phi}S^{2\beta}\Big)\tfrac{\partial}{\partial S}f(S,t)\\&+\tfrac{\alpha\gamma\bar{\lambda}^2}{2(\phi-\gamma+1)}S^{2\beta}f(S,t)+\alpha\gamma rf(S,t)=0\\\Longrightarrow\ \ &(-B)\tfrac{\partial}{\partial\tilde{\tau}}f\Big(\big(\tfrac{z}{\omega}\big)^{\frac{1}{x}},T-\tfrac{\tilde{\tau}}{B}\Big)+\tfrac{\sigma^2}{2}\big(\tfrac{z}{\omega}\big)^{\frac{1}{x}(2\beta+2)}x^2\omega^{\frac{2}{x}}z^{2-\frac{2}{x}}\tfrac{\partial^2}{\partial z^2}f\Big(\big(\tfrac{z}{\omega}\big)^{\frac{1}{x}},T-\tfrac{\tilde{\tau}}{B}\Big)\\&+\tfrac{\sigma^2}{2}\big(\tfrac{z}{\omega}\big)^{\frac{1}{x}(2\beta+2)}\big(x^2-x\big)\omega^{\frac{2}{x}}z^{1-\frac{2}{x}}\tfrac{\partial}{\partial z}f\Big(\big(\tfrac{z}{\omega}\big)^{\frac{1}{x}},T-\tfrac{\tilde{\tau}}{B}\Big)\\&+\Big[\alpha\big(\tfrac{z}{\omega}\big)^{\frac{1}{x}}r\alpha^{-1}+\alpha\big(\tfrac{z}{\omega}\big)^{\frac{1}{x}}\tfrac{\bar{\lambda}\gamma\sigma}{\gamma-\phi}\big(\tfrac{z}{\omega}\big)^{\frac{2\beta}{x}}\Big]x\omega^{\frac{1}{x}}z^{1-\frac{1}{x}}\tfrac{\partial}{\partial z}f\Big(\big(\tfrac{z}{\omega}\big)^{\frac{1}{x}},T-\tfrac{\tilde{\tau}}{B}\Big)\\&+\tfrac{\alpha\gamma\bar{\lambda}^2}{2(\phi-\gamma+1)}\big(\tfrac{z}{\omega}\big)^{\frac{2\beta}{x}}f\Big(\big(\tfrac{z}{\omega}\big)^{\frac{1}{x}},T-\tfrac{\tilde{\tau}}{B}\Big)+\alpha\gamma rf\Big(\big(\tfrac{z}{\omega}\big)^{\frac{1}{x}},T-\tfrac{\tilde{\tau}}{B}\Big)=0\end{aligned}$$

$$\begin{aligned}
\Longrightarrow & B\frac{\partial}{\partial\tilde{\tau}}f\left(\left(\frac{z}{\omega}\right)^{\frac{1}{x}},T-\frac{\tilde{\tau}}{B}\right)-\frac{\sigma^2}{2}x^2\omega^{\frac{2}{x}-\frac{1}{x}(2\beta+2)}z^{\frac{1}{x}(2\beta+2)+2-\frac{2}{x}}\frac{\partial^2}{\partial z^2}f\left(\left(\frac{z}{\omega}\right)^{\frac{1}{x}},T-\frac{\tilde{\tau}}{B}\right)\\
&-\frac{\sigma^2}{2}\left(x^2-x\right)\omega^{\frac{2}{x}-\frac{1}{x}(2\beta+2)}z^{\frac{1}{x}(2\beta+2)+1-\frac{2}{x}}\frac{\partial}{\partial z}f\left(\left(\frac{z}{\omega}\right)^{\frac{1}{x}},T-\frac{\tilde{\tau}}{B}\right)\\
&-\left(rxz+x\omega^{-\frac{2\beta}{x}}z^{1+\frac{2\beta}{x}}\frac{\alpha\bar{\lambda}\gamma\sigma}{\gamma-\phi}\right)\frac{\partial}{\partial z}f\left(\left(\frac{z}{\omega}\right)^{\frac{1}{x}},T-\frac{\tilde{\tau}}{B}\right)\\
&-\frac{\alpha\gamma\bar{\lambda}^2}{2(\phi-\gamma+1)}\omega^{-\frac{2\beta}{x}}z^{\frac{2\beta}{x}}f\left(\left(\frac{z}{\omega}\right)^{\frac{1}{x}},T-\frac{\tilde{\tau}}{B}\right)-\alpha\gamma rf\left(\left(\frac{z}{\omega}\right)^{\frac{1}{x}},T-\frac{\tilde{\tau}}{B}\right)=0. \qquad \text{(A14)}
\end{aligned}$$

Denote

$$F(z,\tilde{\tau}):=f\left(\left(\frac{z}{\omega}\right)^{\frac{1}{x}},T-\frac{\tilde{\tau}}{B}\right),$$

and then Equation (A14) can be rewritten as

$$\begin{aligned}
&\frac{\partial}{\partial\tilde{\tau}}F(z,\tilde{\tau})-\underbrace{\frac{\sigma^2}{2B}x^2\omega^{-\frac{2\beta}{x}}z^{\frac{2\beta}{x}+2}}_{term\,2}\frac{\partial^2}{\partial z^2}F(z,\tilde{\tau})\\
&-\underbrace{\left[\frac{\sigma^2}{2B}\left(x^2-x\right)\omega^{-\frac{2\beta}{x}}z^{\frac{2\beta}{x}+1}+\frac{1}{B}\left(rxz+x\omega^{-\frac{2\beta}{x}}z^{1+\frac{2\beta}{x}}\frac{\alpha\bar{\lambda}\gamma\sigma}{\gamma-\phi}\right)\right]}_{term\,3}\frac{\partial}{\partial z}F(z,\tilde{\tau})\\
&-\underbrace{\left[\frac{\alpha\gamma\bar{\lambda}^2}{2B(\phi-\gamma+1)}\omega^{-\frac{2\beta}{x}}z^{\frac{2\beta}{x}}+\frac{\alpha\gamma r}{B}\right]}_{term\,4}F(z,\tilde{\tau})=0. \qquad \text{(A15)}
\end{aligned}$$

To make Equation (A15) simplest, we assume

$$x=-2\beta$$

so that $z^{\frac{2\beta}{x}+1}=z^{-1+1}=z^0=1$ in *term* 3. Thus, the spatial variables can be scaled as

$$z=\omega S^{-2\beta}=\frac{\omega}{S^{2\beta}},$$

where ω is an unknown parameter. We will determine it later.

As a result, Equation (A15) can be rewritten as

$$\begin{aligned}
&\frac{\partial}{\partial\tilde{\tau}}F(z,\tilde{\tau})-\underbrace{\frac{2\sigma^2\beta^2\omega}{B}z}_{term\,2}\frac{\partial^2}{\partial z^2}F(z,\tilde{\tau})-\underbrace{\left[\frac{\sigma^2\left(4\beta^2+2\beta\right)}{2B}\omega-\frac{2\beta rz}{B}-\frac{2\beta\omega\alpha\bar{\lambda}\gamma\sigma}{B(\gamma-\phi)}\right]}_{term\,3}\frac{\partial}{\partial z}F(z,\tilde{\tau})\\
&-\underbrace{\left[\frac{\alpha\gamma\bar{\lambda}^2}{2B(\phi-\gamma+1)}\omega z^{-1}+\frac{\alpha\gamma r}{B}\right]}_{term\,4}F(z,\tilde{\tau})=0. \qquad \text{(A16)}
\end{aligned}$$

To simplify Equation (A16), we assume $B=\sigma^2\beta^2\omega$. That is, we can scale the inverse time variable as follows:

$$\tilde{\tau}=\sigma^2\beta^2\omega(T-t).$$

Therefore, the PDE for (A16) can be expressed as

$$\frac{\partial}{\partial\tilde{\tau}}F(z,\tilde{\tau})-2z\frac{\partial^2}{\partial z^2}F(z,\tilde{\tau})-\left[\frac{2\beta+1}{\beta}-\frac{2r}{\sigma^2\beta\omega}z-\frac{2\alpha\bar{\lambda}\gamma}{\sigma\beta(\gamma-\phi)}\right]\frac{\partial}{\partial z}F(z,\tilde{\tau})$$
$$-\left[\frac{\alpha\gamma\bar{\lambda}^2}{2\sigma^2\beta^2(\phi-\gamma+1)}z^{-1}+\frac{\alpha\gamma r}{\sigma^2\beta^2\omega}\right]F(z,\tilde{\tau})=0. \tag{A17}$$

We determined two scaled variables, z and $\tilde{\tau}$. However, we have not yet specified parameter ω. We require this degree of freedom to simplify the PDEs.

Appendix A.2.2. Finding the Representation of $f(S,t)$ and the PDE of the Function h

In Appendix A.2.1, we discuss how to determine the two scaled variables z and $\tilde{\tau}$. For convenience, we use these two scaled space and inverse time variables, i.e.,

$$z=\frac{\omega}{S^{2\beta}},\quad \tilde{\tau}=\sigma^2\beta^2\omega(T-t),\ \omega=?$$

in the process of solving the Cauchy problem (A13).

As per Equation (A17), if denoted as

$$F(z,\tilde{\tau}):=f(S,t)=f\left(\left(\frac{\omega}{z}\right)^{\frac{1}{2\beta}},T-\frac{\tilde{\tau}}{a^2\beta^2\omega}\right).$$

we then re-express (A13) as

$$\begin{cases}\frac{\partial}{\partial\tilde{\tau}}F(z,\tilde{\tau})-2z\frac{\partial^2}{\partial z^2}F(z,\tilde{\tau})-\left[\frac{2\beta+1}{\beta}-\frac{2r}{\sigma^2\beta\omega}z-\frac{2\alpha\bar{\lambda}\gamma}{\sigma\beta(\gamma-\phi)}\right]\frac{\partial}{\partial z}F(z,\tilde{\tau})\\ -\left[\frac{\alpha\gamma\bar{\lambda}^2}{2\sigma^2\beta^2(\phi-\gamma+1)}z^{-1}+\frac{\alpha\gamma r}{\sigma^2\beta^2\omega}\right]F(z,\tilde{\tau})=0,\\ F(z,0)=1.\text{ when }\tilde{\tau}=0.\end{cases} \tag{A18}$$

To solve (A18) with the initial function $F(z,0)=1$, we employ the Laplace transformation method to find its solution. The following steps outline this process:

Step I: Assume that the solution $f(S,t)$ for the Cauchy problem is represented as follows:

$$f(S,t):=F(z,\tilde{\tau})=z^{\delta}\exp\{c+x\tilde{\tau}+yz\}h(z,\tilde{\tau}). \tag{A19}$$

At this point, the parameters δ, c, x and y are currently considered to be unknown, and their expression or values will be determined later.

From (A19) we have

$$\begin{aligned}\frac{\partial}{\partial\tilde{\tau}}F &= xF+z^{\delta}\exp\{c+x\tilde{\tilde{\tau}}+yz\}\frac{\partial}{\partial\tau}h=xF+z^{\delta}\exp\{c+x\tilde{\tau}+yz\}h\frac{1}{h}\frac{\partial}{\partial\tilde{\tau}}h\\ &= \left(x+\frac{h_{\tilde{\tau}}}{h}\right)F,\\ \frac{\partial}{\partial z}F &= z^{\delta}\exp\{c+x\tilde{\tau}+yz\}\frac{\partial}{\partial z}h+yF+\delta z^{\delta-1}\exp\{c+x\tilde{\tau}+yz\}h\\ &= \left(\frac{h_z}{h}+y+\frac{\delta}{z}\right)F,\end{aligned}$$

$$\begin{aligned}\frac{\partial^2}{\partial z^2}F &= \left(\frac{1}{h}\frac{\partial}{\partial z}h+y+\frac{\delta}{z}\right)^2F+\left(\frac{\partial}{\partial z}\left(\frac{1}{h}\frac{\partial}{\partial z}h\right)-\frac{\delta}{z^2}\right)F\\ &= \left[y^2+2y\frac{\delta}{z}+\frac{(\delta^2-\delta)}{z^2}+\left(2y+2\frac{\delta}{z}\right)\frac{h_z}{h}+\frac{h_{zz}}{h}\right]F.\end{aligned}$$

Substituting these into (A18) and dividing by F to yield

$$\left(x+\tfrac{h_{\bar{\tau}}}{h}\right)-2z\left[y^2+2y\tfrac{\delta}{z}+\tfrac{\left(\delta^2-\delta\right)}{z^2}+\left(2y+2\tfrac{\delta}{z}\right)\tfrac{h_z}{h}+\tfrac{h_{zz}}{h}\right]$$
$$-\left[\tfrac{2\beta+1}{\beta}-\tfrac{2r}{\sigma^2\beta\omega}z-\tfrac{2\alpha\bar{\lambda}\gamma}{\sigma\beta(\gamma-\phi)}\right]\left(\tfrac{h_z}{h}+y+\tfrac{\delta}{z}\right)-\left[\tfrac{\alpha\gamma\bar{\lambda}^2}{2\sigma^2\beta^2(\phi-\gamma+1)}z^{-1}+\tfrac{\alpha\gamma r}{\sigma^2\beta^2\omega}\right]=0.$$

Multiplying by h gives:

$$h_{zz}-\frac{1}{2z}\left[\left(\frac{2r}{\sigma^2\beta\omega}-4y\right)z+\frac{2\alpha\bar{\lambda}\gamma}{\sigma\beta(\gamma-\phi)}-\frac{2\beta+1}{\beta}-4\delta\right]h_z \tag{A20}$$

$$+\left[\begin{array}{c}\left(2y^2-\tfrac{2ry}{\sigma^2\beta\omega}\right)\tfrac{1}{2}+\left[2(\delta^2-\delta)+\tfrac{\alpha\gamma\bar{\lambda}^2}{2\sigma^2\beta^2(\phi-\gamma+1)}+\tfrac{(2\beta+1)\delta}{\beta}-\tfrac{2\alpha\bar{\lambda}\gamma\delta}{\sigma\beta(\gamma-\phi)}\right]\tfrac{1}{2z^2}\\ +\left(4y\delta-x+\tfrac{\alpha r\gamma}{\sigma^2\beta^2\omega}+y\tfrac{2\beta+1}{\beta}-\tfrac{2r\delta}{\sigma^2\beta\omega}-y\tfrac{2\alpha\bar{\lambda}\gamma}{\sigma\beta(\gamma-\phi)}\right)\tfrac{1}{2z}\end{array}\right]h=\frac{1}{2z}h_{\bar{\tau}}.$$

Step II: The goal is to cancel the h_z term:
In order to eliminate the term of h_z, we set

$$\left(\frac{2r}{\sigma^2\beta\omega}-4y\right)z+\frac{2\alpha\bar{\lambda}\gamma}{\sigma\beta(\gamma-\phi)}-\frac{2\beta+1}{\beta}-4\delta=0.$$

This can be achieved by choosing appropriate values for δ and Q:

$$\begin{aligned}&\frac{2\alpha\bar{\lambda}\gamma}{\sigma\beta(\gamma-\phi)}-\frac{2\beta+1}{\beta}-4\delta=0\\ \Longrightarrow\ &\delta=\frac{2\alpha\bar{\lambda}\gamma}{4\sigma\beta(\gamma-\phi)}-\frac{2\beta+1}{4\beta}=\frac{\alpha\bar{\lambda}\gamma}{2\sigma\beta(\gamma-\phi)}-\frac{1}{2}-\frac{1}{4\beta}\\ &=-\frac{1}{2}-\frac{1}{4\beta}+\frac{\alpha\bar{\lambda}\gamma}{2\sigma\beta(\gamma-\phi)}\\ \Longrightarrow\ &\delta=-\frac{1}{2}-\frac{1}{2\beta}\left(\frac{1}{2}-\frac{\alpha\bar{\lambda}\gamma}{\sigma(\gamma-\phi)}\right),\end{aligned} \tag{A21}$$

and

$$\begin{aligned}&\frac{2r}{\sigma^2\beta\omega}-4y=0\\ \Longrightarrow\ &y=\frac{2r}{4\sigma^2\beta\omega}=\frac{1}{2}\frac{r}{\sigma^2\beta\omega}.\end{aligned}$$

We denote

$$Q:=\frac{r}{\sigma^2\beta\omega}. \tag{A22}$$

Then,

$$y=\frac{1}{2}Q \tag{A23}$$

Remark A1. *When $\phi = 0$ (i.e., without ambiguity), the parameters δ and Q in (A21) and (A22) are identical to those presented in the reference [6]. Specifically, referring to [6], these two parameters were defined as follows:*

$$\lambda = -\tfrac{1}{2} - \tfrac{1}{2\beta}\left(\tfrac{1}{2} - \delta c\right),$$
$$Q = \tfrac{\delta(\alpha - \gamma r)}{\Lambda \sigma^2 \beta}.$$

It should be noted that the expressions for δ and Q mentioned above correspond to the definitions provided in the reference [6]

We substitute (A21)–(A23), into Equation (A20) to obtain:

$$h_{zz} - \underbrace{\frac{1}{2z}\left[\left(\frac{2r}{\sigma^2\beta\omega} - 2Q\right)z + \frac{2\alpha\bar{\lambda}\gamma}{\sigma\beta(\gamma-\phi)} - \frac{2\beta+1}{\beta} - 4\delta\right]}_{term\,2} h_z$$
$$+\left[\begin{array}{l} \underbrace{\left(\frac{Q^2}{2} - \frac{rQ}{\sigma^2\beta\omega}\right)\frac{1}{2}}_{term\,3} + \underbrace{\left[2(\delta^2-\delta) + \frac{\alpha\gamma\bar{\lambda}^2}{2\sigma^2\beta^2(\phi-\gamma+1)} + \frac{(2\beta+1)\delta}{\beta} - \frac{2\alpha\bar{\lambda}\gamma\delta}{\sigma\beta(\gamma-\phi)}\right]\frac{1}{2z^2}}_{term\,4} \\ \qquad + \underbrace{\left(2Q\delta - x + \frac{\alpha r\gamma}{\sigma^2\beta^2\omega} + Q\frac{2\beta+1}{2\beta} - \frac{2r\delta}{\sigma^2\beta\omega} - Q\frac{2\alpha\bar{\lambda}\gamma}{2\sigma\beta(\gamma-\phi)}\right)\frac{1}{2z}}_{term\,5} \end{array}\right] h = \tfrac{1}{2z}h_{\bar{\tau}}.$$

Based on the previous steps, it is clear that the coefficient of h_z is zero, allowing us to eliminate *term* 2. Therefore, the above equation can be expressed as

$$h_{zz} + \left[\begin{array}{l} \underbrace{\left(\frac{Q^2}{2} - \frac{rQ}{\sigma^2\beta\omega}\right)\frac{1}{2}}_{term\,3} + \underbrace{\left[2(\delta^2-\delta) + \frac{\alpha\gamma\bar{\lambda}^2}{2\sigma^2\beta^2(\phi-\gamma+1)} + \frac{(2\beta+1)\delta}{\beta} - \frac{2\alpha\bar{\lambda}\gamma\delta}{\sigma\beta(\gamma-\phi)}\right]\frac{1}{2z^2}}_{term\,4} \\ \qquad + \underbrace{\left(2Q\delta - x + \frac{\alpha r\gamma}{\sigma^2\beta^2\omega} + Q\frac{2\beta+1}{2\beta} - \frac{2r\delta}{\sigma^2\beta\omega} - Q\frac{2\alpha\bar{\lambda}\gamma}{2\sigma\beta(\gamma-\phi)}\right)\frac{1}{2z}}_{term\,5} \end{array}\right] h = \frac{1}{2z}h_{\bar{\tau}}. \tag{A24}$$

Step III: Now, we have an unknown parameter ω in *term* 3. It has been observed that we can choose an appropriate parameter ω to transform Equation (A24), which governs the function h, into the well-known Whittaker equation. Hence, we assume that *term* 3 is equal to $-\frac{1}{4}$, i.e.,

$$\left(\frac{Q^2}{2} - \frac{rQ}{\sigma^2\beta\omega}\right)\frac{1}{2} = -\frac{1}{4}$$
$$\Longrightarrow \frac{Q^2}{2} - \frac{rQ}{\sigma^2\beta\omega} = -\frac{1}{2}$$
$$\Longrightarrow \left(\frac{Q^2}{2} + \frac{1}{2}\right) - \frac{rQ}{\sigma^2\beta\omega} = 0. \tag{A25}$$

It is known that $Q = \frac{r}{\sigma^2\beta\omega}$ in (A22). Substituting this into Equation (A25) yields

$$\frac{r^2}{2\sigma^4\beta^2\omega^2} + \frac{1}{2} - \frac{r\frac{r}{\sigma^2\beta\omega}}{\sigma^2\beta\omega} = 0$$
$$\Longrightarrow \omega = \frac{r}{\sigma^2 \mid \beta \mid}, \tag{A26}$$

where the parameter ω is determined.

Remark A2. *With the correspondence in notation from the source to our notation $\alpha = r$, $\delta = \alpha$ and $\Lambda = \omega$, the formula for the parameter ω in our M-CEV case is the same as that in the reference [6]. In that study, the Λ is given by*

$$\Lambda = \sqrt{\frac{\delta(\alpha^2 - \gamma r^2)}{a^4 \beta^2}} = \frac{\sqrt{\delta}}{a^2 \mid \beta \mid} \sqrt{\alpha^2 - \gamma r^2}.$$

Step IV: To determine that the coefficient in *term* 4 is appropriate for Equation (A24), we begin by assuming that the coefficient of z^{-2} in *term* 4 is represented by an unknown parameter, denoted as k. That is

$$k = \frac{1}{2}\left(2(\delta^2 - \delta) + \frac{\alpha\gamma\bar{\lambda}^2}{2\sigma^2\beta^2(\phi - \gamma + 1)} + \frac{(2\beta + 1)\delta}{\beta} - \frac{2\alpha\bar{\lambda}\gamma\delta}{\sigma\beta(\gamma - \phi)}\right) \tag{A27}$$

$$= \delta^2 - \delta + \frac{\alpha\gamma\bar{\lambda}^2}{4\sigma^2\beta^2(\phi - \gamma + 1)} + \frac{(2\beta + 1)\delta}{2\beta} - \frac{2\alpha\bar{\lambda}\gamma\delta}{2\sigma\beta(\gamma - \phi)}. \tag{A28}$$

Due to $\delta = \frac{\alpha\bar{\lambda}\gamma}{2\sigma\beta(\gamma - \phi)} - \frac{2\beta+1}{4\beta}$, we have from Equation (A28)

$$\begin{aligned} k &= \delta^2 - \delta - \left[2\frac{\alpha\bar{\lambda}\gamma\delta}{2\sigma\beta(\gamma - \phi)} - 2\frac{(2\beta + 1)}{4\beta}\delta\right] + \frac{\alpha\gamma\bar{\lambda}^2}{4\sigma^2\beta^2(\phi - \gamma + 1)} \\ &= -\delta^2 - \delta + \frac{\alpha\gamma\bar{\lambda}^2}{4\sigma^2\beta^2(\phi - \gamma + 1)} \\ \Longrightarrow \frac{1}{4} - k &= \left(\delta + \frac{1}{2}\right)^2 + \frac{\alpha\gamma\bar{\lambda}^2}{4\sigma^2\beta^2(\gamma - \phi - 1)}. \end{aligned} \tag{A29}$$

To obtain the form of the Whittaker equation, we introduce a new parameter η defined as

$$\begin{aligned} \eta^2 &= \left(\delta + \frac{1}{2}\right)^2 + \frac{\alpha\gamma\bar{\lambda}^2}{4\sigma^2\beta^2(\gamma - \phi - 1)} \\ \Longrightarrow \eta &= \sqrt{\left(\delta + \frac{1}{2}\right)^2 + \frac{\alpha\gamma\bar{\lambda}^2}{4\sigma^2\beta^2(\gamma - \phi - 1)}}. \end{aligned} \tag{A30}$$

Accordingly, we rewrite Equation (A29) as

$$\begin{aligned} &\tfrac{1}{4} - k = \eta^2 \\ \Longrightarrow \quad &k = \tfrac{1}{4} - \eta^2, \end{aligned}$$

which implies that Equation (A27) can be expressed as

$$\frac{1}{2}\left(2(\delta^2 - \delta) + \frac{\alpha\gamma\bar{\lambda}^2}{2\sigma^2\beta^2(\phi - \gamma + 1)} + \frac{(2\beta + 1)\delta}{\beta} - \frac{2\alpha\bar{\lambda}\gamma\delta}{\sigma\beta(\gamma - \phi)}\right) = \frac{1}{4} - \eta^2.$$

Remark A3. *When $\phi = 0$, the parameter η (see Formula (A30)) is different from the reference [6]:*

$$\eta = \sqrt{\left(\lambda + \frac{1}{2}\right)^2 + \frac{\delta(1 - \delta)c^2}{4a^4\beta^2}}.$$

This is because of the difference in the Cauchy problem.

Step V: We aim to make the coefficient of z^{-1} zero in *term* 5 by selecting an appropriate parameter, x in Equation (A24). To determine the appropriate parameter x, we initially assume that the coefficient of *term* 5 is equal to zero, which can be expressed as

$$\begin{aligned}&\left[2Q\delta - x + \frac{\alpha r\gamma}{\sigma^2\beta^2\omega} + Q\frac{2\beta+1}{2\beta} - \frac{2r\delta}{\sigma^2\beta\omega} - Q\frac{2\alpha\bar{\lambda}\gamma}{2\sigma\beta(\gamma-\phi)}\right]\frac{1}{2} = 0\\ \Longrightarrow &x = 2Q\delta + \frac{\alpha\gamma}{\sigma^2\beta^2\omega}r - 2Q\frac{\alpha\bar{\lambda}\gamma}{2\sigma\beta(\gamma-\phi)} + Q\frac{2\beta+1}{2\beta} - 2\frac{r\delta}{\sigma^2\beta\omega}. \qquad \text{(A31)}\end{aligned}$$

From (A21) and (A22), we know that $\delta = \frac{\alpha\bar{\lambda}\gamma}{2\sigma\beta(\gamma-\phi)} - \frac{2\beta+1}{4\beta}$ and $Q = \frac{r}{\sigma^2\beta\omega}$, then Equation (A31) can be written as

$$\begin{aligned}x &= 2Q\delta + \tfrac{\alpha\gamma}{\sigma^2\beta^2\omega}r - 2Q\tfrac{\alpha\bar{\lambda}\gamma}{2\sigma\beta(\gamma-\phi)} + 2Q\tfrac{2\beta+1}{4\beta} - 2\tfrac{r}{\sigma^2\beta\omega}\delta\\ &= 2Q\delta + \tfrac{\alpha\gamma}{\sigma^2\beta^2\omega}r - 2Q\delta - 2Q\delta\\ &= \tfrac{r\alpha\gamma}{\sigma^2\beta^2\omega} - 2Q\delta.\end{aligned}$$

Denote

$$R^* := -2Q\delta. \qquad \text{(A32)}$$

This means

$$x = \frac{r\alpha\gamma}{\sigma^2\beta^2\omega} + R^*.$$

Remark A4. *When $\alpha = r$, the R^* for (A32) is not the same as the reference [6]. We obtained*

$$R^* = -2Q\lambda - \frac{\delta(1-\delta)(\alpha-r)c}{\Lambda a^2\beta^2}.$$

As opposed to

$$R = -2Q\lambda - \frac{\delta(1-\delta)(\alpha-r)c}{\Lambda a^4\beta^2}.$$

Next, we determine the unknown parameter c:

Let $d := c + x\tilde{\tau}$. Because $\tilde{\tau} = \sigma^2\beta^2\omega(T-t)$ and the assumed representation of the solution $F(z,\tilde{\tau})$ is given by

$$\begin{aligned}F(z,\tilde{\tau}) &= z^\delta \exp\{c + x\tilde{\tau} + yz\}h(z,\tilde{\tau})\\ &= z^\delta \exp\{d + yz\}h(z,\tilde{\tau}),\end{aligned}$$

then

$$\begin{aligned}d &= c + x\tilde{\tau} = c + \left(\tfrac{r\alpha\gamma}{\sigma^2\beta^2\omega} + R^*\right)\tilde{\tau}\\ &= c + r\alpha\gamma T - r\alpha\gamma t + R^*\tilde{\tau}.\end{aligned}$$

As a result, the educated guess of c can be given by

$$c = -r\alpha\gamma T,$$

which implies

$$d = -r\alpha\gamma t + R^*\tilde{\tau}.$$

Consequently, we can represent its solution $f(S,t)$ for the Cauchy problem as

$$f(S,t) = z^{\delta}\exp\left\{-r\alpha\gamma t + R^{*}\tilde{\tau} + \frac{Q}{2}z\right\}h(z,\tilde{\tau}), \tag{A33}$$

where $z = \omega S^{-2\beta}$ and $\tilde{\tau} = \sigma^2\beta^2\omega(T-t)$ with parameters

$$\begin{gathered}\omega = \tfrac{r}{\sigma^2|\beta|},\ \ \delta = -\tfrac{1}{2} - \tfrac{1}{2\beta}\left(\tfrac{1}{2} - \tfrac{\alpha\bar{\lambda}\gamma}{\sigma(\gamma-\phi)}\right),\\ \eta = \sqrt{\left(\delta+\tfrac{1}{2}\right)^2 + \tfrac{\alpha\gamma\bar{\lambda}^2}{4\sigma^2\beta^2(\gamma-\phi-1)}},\\ Q = \tfrac{r}{\sigma^2\beta\omega},\ \ R^{*} = -2Q\delta.\end{gathered}$$

After ensuring the five terms in (A24), we express the PDE for (A24) as follows:

$$h_{zz} + \left[-\frac{1}{4} + \frac{\left(\frac{1}{4}-\eta^2\right)}{z^2}\right]h = \frac{1}{2z}h_{\tilde{\tau}} \tag{A34}$$

with initial condition

$$h(z,0) = \frac{F(z,0)}{z^{\delta}\exp\left\{\frac{Q}{2}(z)\right\}} = z^{-\delta}\exp\left\{r\alpha\gamma T - \frac{Q}{2}z\right\}, \tag{A35}$$

which indicates that Equation (A34) is a Whittaker equation. Solving the Whittaker equation is provided below.

Let us denote the corresponding operator as

$$\mathcal{L}_h h(z,\tilde{\tau}) = h_{zz} + \left[-\frac{1}{4} + \left(\frac{1}{4}-\eta^2\right)\frac{1}{z^2}\right]h - \frac{1}{2z}h_{\tilde{\tau}}.$$

Appendix A.2.3. Applying the Laplace Transform $G(x;\zeta)$ to Find the Solution $h(z,\tilde{\tau})$

In Appendix A.2.2, we established the representation (A33) of the function $f(S,t)$ as follows:

$$f(S,t) = z^{\delta}\exp\left\{-r\gamma\delta t + R^{*}\tilde{\tau} + \frac{Q}{2}z\right\}h(z,\tilde{\tau}),$$

where $h(z,\tilde{\tau})$ is the solution to the Cauchy problem in Equation (A34). In the following steps, our objective is to determine the solution for $h(z,\tilde{\tau})$:

Step 1: Find the Laplace transform of $h(z,\tilde{\tau})$, and produce its ODE.

According to the definition of the Laplace transform, we denote the Laplace transform of $h(z,\tilde{\tau})$ by $G(z;\zeta)$. That is,

$$G(z;\zeta) := \mathcal{L}_L h(z,\tilde{\tau}) = \int_0^{\infty} e^{-\zeta\tilde{\tau}}h(z,\tau)d\tilde{\tau}$$

with $Re(\zeta) > 0$.

Using the properties of Laplace transform methods, we know that the transform of a derivative w. r. t. z is a just differentiating the transformed function.

$$\begin{aligned}\mathcal{L}_L h_z(z,\tilde{\tau}) &= \textstyle\int_0^{\infty}\exp\{-\zeta\tau\}h_z(z,\tilde{\tau})d\tau\\ &= \tfrac{d}{dz}G(z;\zeta) = G_z(z;\zeta),\end{aligned}$$

and

$$\begin{aligned}\mathcal{L}_L h_{zz}(z,\tilde{\tau}) &= \textstyle\int_0^{\infty}\exp\{-\zeta\tilde{\tau}\}h_{zz}(z,\tilde{\tau})d\tilde{\tau}\\ &= \tfrac{d^2}{dz^2}G(z;\zeta) = G_{zz}(z;\zeta).\end{aligned}$$

To transform the derivative in $\tilde{\tau}$, we use the common rules (see the reference [17])

$$\mathcal{L}_L h_{\tilde{\tau}}(z,\tilde{\tau}) = \zeta G(z;\zeta) - h(z,0),$$

where

$$h(z,0) = \frac{F(z,0)}{z^{\delta}\exp\left\{\frac{Q}{2}(z)\right\}} = z^{-\delta}\exp\left\{r\alpha\gamma T - \frac{Q}{2}z\right\}$$

Thus, we use the Laplace transform in Equation (A34) and produce the following ODE:

$$\begin{aligned} &G_{zz} + \left[-\frac{1}{4} + \frac{\left(\frac{1}{4}-\eta^2\right)}{z^2}\right]G = \frac{1}{2z}\left(\zeta G - z^{-\delta}\exp\left\{r\alpha\gamma T - \frac{Q}{2}z\right\}\right) \\ \Longrightarrow &G_{zz} + \left[-\frac{1}{4} - \frac{\frac{\zeta}{2}}{z} + \frac{\left(\frac{1}{4}-\eta^2\right)}{z^2}\right]G = -\chi(z) \end{aligned} \tag{A36}$$

where denoting

$$\chi(z) := \frac{1}{2}z^{-1-\delta}\exp\left\{r\alpha\gamma T - \frac{Q}{2}z\right\}. \tag{A37}$$

Step 2: Solve the ODE for $G(z;\zeta)$.
In ODE (A36), the homogeneous equation for G is known as the Whittaker equation:

$$G_{zz} + \left[-\frac{1}{4} + \frac{-\zeta/2}{z} + \frac{\left(\frac{1}{4}-\eta^2\right)}{z^2}\right]G = 0,$$

with two linearly independent solutions, namely the Whittaker functions $M_{-\frac{\zeta}{2},\eta}(z)$ and $W_{-\frac{\zeta}{2},\eta}(z)$ (see the reference [18]) .

Hence, the general solution of the homogeneous equation is given by

$$G_0(z;\zeta) = C_1 M_{-\frac{\zeta}{2},\eta}(z) + C_2 W_{-\frac{\zeta}{2},\eta}(z).$$

Let us return to the non-homogeneous Equation (A36). We seek its solution in the form:

$$G(z;\zeta) = C_1(z) M_{-\frac{\zeta}{2},\eta}(z) + C_2(z) W_{-\frac{\zeta}{2},\eta}(z). \tag{A38}$$

The functions $C_1(z)$ and $C_2(z)$ can be determined from the following system of equations:

$$\begin{cases} C_1'(z) M_{-\frac{\zeta}{2},\eta}(z) + C_2'(z) W_{-\frac{\zeta}{2},\eta}(z) = 0, \\ C_1'(z) M'_{-\frac{\zeta}{2},\eta}(z) + C_2'(z) W'_{-\frac{\zeta}{2},\eta}(z) = -\chi(z). \end{cases} \tag{A39}$$

We express the derivative $C_1'(z)$ from the first equation

$$C_1'(z) = -C_2'(z)\frac{W_{-\frac{\zeta}{2},\eta}(z)}{M_{-\frac{\zeta}{2},\eta}(z)}.$$

Substituting back into the second equation, we find the derivative $C_2'(z)$:

$$\begin{aligned} &-C_2'(z)\frac{W_{-\frac{\zeta}{2},\eta}(z)}{M_{-\frac{\zeta}{2},\eta}(z)}M'_{-\frac{\zeta}{2},\eta}(z)+C_2'(z)W'_{-\frac{\zeta}{2},\eta}(z)=-\chi(z)\\ \Longrightarrow &C_2'(z)=-\frac{\chi(z)M_{-\frac{\zeta}{2},\eta}(z)}{M_{-\frac{\zeta}{2},\eta}(z)W'_{-\frac{\zeta}{2},\eta}(z)-W_{-\frac{\zeta}{2},\eta}(z)M'_{-\frac{\zeta}{2},\eta}(z)}. \end{aligned} \tag{A40}$$

It can be seen that the denominators $M_{-\frac{\zeta}{2},\eta}(z)W'_{-\frac{\zeta}{2},\eta}(z)-W_{-\frac{\zeta}{2},\eta}(z)M'_{-\frac{\zeta}{2},\eta}(z)$ in (A40) are the Wronskians for these two Whittaker functions (see the reference [19]), which is given by

$$\begin{aligned} \mathcal{W}\Big\{M_{-\frac{\zeta}{2},\eta}(z),W_{-\frac{\zeta}{2},\eta}(z)\Big\} &= M_{-\frac{\zeta}{2},\eta}(z)W'_{-\frac{\zeta}{2},\eta}(z)-W_{-\frac{\zeta}{2},\eta}(z)M'_{-\frac{\zeta}{2},\eta}(z)\\ &= -\frac{\Gamma(1+2\eta)}{\Gamma(\frac{1}{2}+\eta+\frac{\zeta}{2})}\neq 0. \end{aligned}$$

Denote

$$\begin{aligned} \Xi\Big(-\frac{\zeta}{2},\eta\Big) &:= -\frac{1}{\mathcal{W}\Big\{M_{-\frac{\zeta}{2},\eta}(z),W_{-\frac{\zeta}{2},\eta}(z)\Big\}}\\ &= \frac{\Gamma(\frac{1}{2}+\eta+\frac{\zeta}{2})}{\Gamma(1+2\eta)}, \end{aligned}$$

and then it follows that

$$M_{-\frac{\zeta}{2},\eta}(z)W'_{-\frac{\zeta}{2}\eta}(z)-W_{-\frac{\zeta}{2},\eta}(z)M'_{-\frac{\zeta}{2},\eta}(z)=-\frac{1}{\Xi\Big(-\frac{\zeta}{2},\eta\Big)}.$$

Thus,

$$\begin{aligned} C_1(z) &= \int_z^\infty \frac{\chi(\psi)W_{-\frac{\zeta}{2},\eta}(\psi)}{\mathcal{W}\Big\{M_{-\frac{\zeta}{2},\eta}(\psi),W_{-\frac{\zeta}{2},\eta}(\psi)\Big\}}\mathrm{d}\psi=-\Xi\Big(-\frac{\zeta}{2},\eta\Big)\int_z^\infty \chi(\psi)W_{-\frac{\zeta}{2},\eta}(\psi)\mathrm{d}\psi,\\ &= \Xi\Big(-\frac{\zeta}{2},\eta\Big)\int_0^z \chi(\psi)W_{-\frac{\zeta}{2},\eta}(\psi)\mathrm{d}\psi\\ C_2(z) &= -\int_z^\infty \frac{\chi(\psi)M_{-\frac{\zeta}{2},\eta}(\psi)}{\mathcal{W}\Big\{M_{-\frac{\zeta}{2},\eta}(\psi),W_{-\frac{\zeta}{2},\eta}(\psi)\Big\}}\mathrm{d}\psi=\Xi\Big(-\frac{\zeta}{2},\eta\Big)\int_z^\infty \chi(\psi)M_{-\frac{\zeta}{2},\eta}(\psi)\mathrm{d}\psi. \end{aligned}$$

From (A38) the solution $G(z;\zeta)$ can be obtained as

$$\begin{aligned} G(z;\zeta) &= C_1(z)M_{-\frac{\zeta}{2},\eta}(z)+C_2(z)W_{-\frac{\zeta}{2},\eta}(z)\\ &= M_{-\frac{\zeta}{2},\eta}(z)\Xi\Big(-\frac{\zeta}{2},\eta\Big)\int_0^z \chi(\psi)W_{-\frac{\zeta}{2},\eta}(\psi)\mathrm{d}\psi+W_{-\frac{\zeta}{2},\eta}(z)\Xi\Big(-\frac{\zeta}{2},\eta\Big)\int_z^\infty \chi(\psi)M_{-\frac{\zeta}{2},\eta}(\psi)\mathrm{d}\psi\\ &= \Xi\Big(-\frac{\zeta}{2},\eta\Big)\Big(M_{-\frac{\zeta}{2},\eta}(z)\int_0^z \chi(\psi)W_{-\frac{\zeta}{2},\eta}(\psi)\mathrm{d}\psi+W_{-\frac{\zeta}{2},\eta}(z)\int_z^\infty \chi(\psi)M_{-\frac{\zeta}{2},\eta}(\psi)\mathrm{d}\psi\Big). \end{aligned} \tag{A41}$$

See the references [20,21] (such as [20]: 6.669.4], the following relationship exists between Whittaker functions and modified Bessel functions:

$$\begin{aligned} \int_0^\infty e^{-\frac{1}{2}(a_1+a_2)t\cosh x}\big[\coth\big(\tfrac{x}{2}\big)\big]^{2\nu}I_{2\mu}\big(t\sqrt{a_1a_2}\sinh x\big)\mathrm{d}x &= \frac{\Gamma(\frac{1}{2}+\mu-\nu)}{t\sqrt{a_1a_2}\Gamma(1+2\mu)}W_{\nu,\mu}(a_1t)M_{\nu,\mu}(a_2t),\\ &\Big[Re\Big(\frac{1}{2}+\mu-\nu\Big)>0,\ Re(\mu)>0,a_1>a_2\Big]. \end{aligned}$$

Hence, when $\nu=-\frac{\zeta}{2}$, $a_1t=\psi$ and $a_2t=z$, we have

$$\begin{aligned}
&\frac{\Gamma(\frac{1}{2}+\eta+\frac{\zeta}{2})}{\sqrt{\psi z}\Gamma(1+2\eta)}W_{-\frac{\zeta}{2},\eta}(\psi)M_{-\frac{\zeta}{2},\eta}(z)=\int_0^\infty e^{-\frac{1}{2}(\psi+z)\cosh\Psi}\left[\coth\left(\frac{\Psi}{2}\right)\right]^{2(-\frac{\zeta}{2})}I_{2\eta}\left(\sqrt{\psi z}\sinh\Psi\right)\mathrm{d}\Psi\\
\Longrightarrow\quad &\Xi\left(-\frac{\zeta}{2},\eta\right)W_{-\frac{\zeta}{2},\eta}(\psi)M_{-\frac{\zeta}{2},\eta}(z)=\sqrt{\psi z}\int_0^\infty e^{-\frac{\psi+z}{2}\cosh\Psi}\left[\tanh^{\zeta}\left(\frac{\Psi}{2}\right)\right]I_{2\eta}\left(\sqrt{\psi z}\sinh\Psi\right)\mathrm{d}\Psi,\\
&\left[Re\left(\frac{1}{2}+\eta+\frac{\zeta}{2})\right)>0,\ Re(\eta)>0,\psi>z\right].
\end{aligned}$$

It follows that

$$\begin{aligned}
G(z;\zeta)&=\Xi\left(-\frac{\zeta}{2},\eta\right)\left(M_{-\frac{\zeta}{2},\eta}(z)\int_0^z\chi(\psi)W_{-\frac{\zeta}{2},\eta}(\psi)\mathrm{d}\psi+W_{-\frac{\zeta}{2},\eta}(z)\int_z^\infty\chi(\psi)M_{-\frac{\zeta}{2},\eta}(\psi)\mathrm{d}\psi\right)\\
&=\Xi\left(-\frac{\zeta}{2},\eta\right)\left(M_{-\frac{\zeta}{2},\eta}(z)\int_0^z\frac{1}{2}\psi^{-1-\delta}e^{r\alpha\gamma T-\frac{Q}{2}\psi}W_{-\frac{\zeta}{2},\eta}(\psi)\mathrm{d}\psi+W_{-\frac{\zeta}{2},\eta}(z)\int_z^\infty\frac{1}{2}\psi^{-1-\delta}e^{r\alpha\gamma T-\frac{Q}{2}z}M_{-\frac{\zeta}{2},\eta}(\psi)\mathrm{d}\psi\right)\\
&=\frac{\sqrt{z}}{2}\left(\int_0^\infty\int_0^\infty\sqrt{\psi}\psi^{-1-\delta}e^{r\alpha\gamma T-\frac{Q}{2}\psi}e^{-\frac{\psi+z}{2}\cosh\Psi}\left[\tanh^{\zeta}\left(\frac{\Psi}{2}\right)\right]I_{2\eta}\left(\sqrt{\psi z}\sinh\Psi\right)\mathrm{d}\psi\mathrm{d}\Psi\right.\\
&=\left.\frac{\sqrt{z}}{2}e^{r\alpha\gamma T}\left(\int_0^\infty\int_0^\infty e^{-\frac{z\cosh\Psi}{2}}\psi^{-\frac{1}{2}-\delta}e^{-\frac{Q+\cosh\Psi}{2}\psi}I_{2\eta}\left(\sqrt{\psi z}\sinh\Psi\right)\tanh^{\zeta}\left(\frac{\Psi}{2}\right)\mathrm{d}\psi\mathrm{d}\Psi.\right.\right.
\end{aligned}\tag{A42}$$

Again, we know the relationship formula (6.643.2) in the reference [20]

$$\begin{aligned}
\int_0^\infty x^{\mu-\frac{1}{2}}e^{-\alpha x}I_{2\nu}(2b\sqrt{x})dx\quad&=\frac{\Gamma(\mu+\nu+\frac{1}{2})}{\Gamma(2\nu+1)}b^{-1}e^{\frac{b^2}{2\alpha}}\alpha^{-\mu}M_{-\mu,\nu}\left(\frac{b^2}{\alpha}\right),\\
&\left[Re\left(\mu+\nu+\frac{1}{2}\right)>0\right].
\end{aligned}$$

Based on (A42), we have

$$\begin{aligned}
\int_0^\infty\psi^{\frac{1}{2}-\delta}e^{-\alpha\psi}I_{2\eta}\left(2\lambda\sqrt{\psi}\right)d\psi&=\Xi(\delta,\eta)\lambda^{-1}e^{\frac{\lambda^2}{2\alpha}}\alpha^{\delta}M_{\delta,\eta}\left(\frac{\lambda^2}{\alpha}\right),\\
&\left[Re\left(\frac{1}{2}-\delta+\eta\right)>0\right].
\end{aligned}\tag{A43}$$

with $\alpha=\frac{\cosh\Psi+Q}{2}$ and $\lambda=\frac{\sqrt{z}\sinh\Psi}{2}$.

From (A42) and (A43), the representation for $G(z;\zeta)$ is given by

$$\begin{aligned}
G(z;\zeta)&=e^{r\alpha\gamma T}\Xi(\delta,\eta)\frac{\sqrt{z}}{2}\left(\int_0^\infty e^{-\frac{z\cosh\Psi}{2}}\lambda^{-1}e^{\frac{\lambda^2}{2\alpha}}\alpha^{\delta}M_{\delta,\eta}\left(\frac{\lambda^2}{\alpha}\right)\tanh^{\zeta}\left(\frac{\Psi}{2}\right)\mathrm{d}\Psi\right.\\
&=\left.e^{r\alpha\gamma T}\Xi(\delta,\eta)\left(\int_0^\infty e^{-\frac{z\cosh\Psi}{2}+\frac{z\sinh^2(\Psi)}{4(\cosh\Psi+Q)}}\tanh^{\zeta}\left(\frac{\Psi}{2}\right)\left(\frac{\cosh\Psi+Q}{2}\right)^{\delta}M_{\delta,\eta}\left(\frac{z\sinh^2\Psi}{2(\cosh\Psi+Q)}\right)\frac{\mathrm{d}\Psi}{\sinh\Psi}.\right.\right.
\end{aligned}\tag{A44}$$

Appendix A.2.4. Finding the Solution $f(S,t)$ for the Cauchy Problem

We achieved the function $G(z;\zeta)$ as shown in (A44) in Appendix A.2.3. For convenience, it can also be written as

$$\begin{aligned}
&e^{\frac{Qz}{2}}G(z;\zeta)\\
&=e^{r\alpha\gamma T}\Xi(\delta,\eta)\left(\int_0^\infty e^{\frac{Qz}{2}}e^{-\frac{z\cosh\Psi}{2}+2\frac{z\sinh^2(\Psi)}{4(\cosh\Psi+Q)}}e^{-\frac{z\sinh^2(\Psi)}{4(\cosh\Psi+Q)}}\tanh^{\zeta}\left(\frac{\Psi}{2}\right)\left(\frac{\cosh\Psi+Q}{2}\right)^{\delta}M_{\delta,\eta}\left(\frac{z\sinh^2\Psi}{2(\cosh\Psi+Q)}\right)\frac{\mathrm{d}\Psi}{\sinh\Psi}.\right.
\end{aligned}\tag{A45}$$

Let $Y(\Psi):=\frac{Q}{2}-\frac{\cosh\Psi}{2}+2\frac{\sinh^2(\Psi)}{4(\cosh\Psi+Q)}$, and we simplify $Y(\Psi)$ to obtain

$$\begin{aligned}
Y(\Psi)\quad&=\frac{Q}{2}+\frac{\sinh^2(\Psi)}{2(\cosh\Psi+Q)}-\frac{\cosh\Psi}{2}\\
&=\frac{Q^2-1}{2(\cosh\Psi+Q)}.
\end{aligned}$$

Thus, Equation (A45) can be expressed as

$$z^{\delta}e^{\frac{Qz}{2}}G(z;\zeta)=e^{r\alpha\gamma T}\Xi(\delta,\eta)\left(\int_0^{\infty}e^{zY(\Psi)}Z^{\delta}(\Psi)\tanh^{\zeta}\left(\frac{\Psi}{2}\right)(zI(\Psi))^{\lambda}e^{-\frac{zI(\Psi)}{2}}M_{\lambda,\eta}(zI(\Psi))\frac{\mathrm{d}\Psi}{\sinh\Psi},\right. \tag{A46}$$

where

$$I(\Psi)=\frac{\sinh^2\Psi}{2(\cosh\Psi+Q)},\ \ Y(\Psi)=\frac{Q^2-1}{2(\cosh\Psi+Q)},\ \ Z(\Psi)=\frac{(\cosh\Psi+Q)^2}{\sinh^2\Psi}.$$

Let the integration variable $v=\ln\tanh\left(\frac{\Psi}{2}\right)$. Since

$$\begin{aligned}
\mathrm{d}v &= \tfrac{1}{\tanh\left(\frac{\Psi}{2}\right)}\tfrac{\mathrm{d}}{\mathrm{d}\left(\frac{\Psi}{2}\right)}\tanh\left(\tfrac{\Psi}{2}\right)\tfrac{\mathrm{d}}{\mathrm{d}\Psi}\left(\tfrac{\Psi}{2}\right)\\
&= \tfrac{\cosh\Psi+1}{\sinh\Psi}\tfrac{1}{2\cosh^2\left(\frac{\Psi}{2}\right)}\mathrm{d}\Psi,\\
e^{v} &= \tanh\left(\tfrac{\Psi}{2}\right)=\tfrac{\sinh\Psi}{\cosh\Psi+1}=\tfrac{\cosh\Psi-1}{\sinh\Psi},\\
e^{-v} &= \left(\tanh\left(\tfrac{\Psi}{2}\right)\right)^{-1}=\tfrac{\cosh\Psi+1}{\sinh\Psi}=\tfrac{\sinh\Psi}{\cosh\Psi-1},\\
e^{-v}-e^{v} &= \tfrac{\sinh\Psi}{\cosh\Psi-1}-\tfrac{\cosh\Psi-1}{\sinh\Psi}=\tfrac{2(\cosh\Psi-1)}{(\cosh\Psi-1)\sinh\Psi}=\tfrac{2}{\sinh\Psi},\\
e^{v}+e^{-v} &= \tfrac{\cosh\Psi-1}{\sinh\Psi}+\tfrac{\sinh\Psi}{\cosh\Psi-1}=\tfrac{2\cosh^2\Psi-2\cosh\Psi}{(\cosh\Psi-1)\sinh\Psi}=\tfrac{2\cosh\Psi}{\sinh\Psi},
\end{aligned}$$

it follows that

$$\begin{aligned}
\mathrm{d}v &= \tfrac{\cosh\Psi+1}{\sinh\Psi}\tfrac{1}{2\cosh^2\left(\frac{\Psi}{2}\right)}\mathrm{d}\Psi=\tfrac{\cosh\Psi+1}{\sinh\Psi}\tfrac{1}{2\frac{\cosh\Psi+1}{2}}\mathrm{d}\Psi\\
&= \tfrac{\cosh\Psi+1}{\sinh\Psi}\tfrac{1}{\cosh\Psi+1}\mathrm{d}\Psi=\tfrac{\mathrm{d}\Psi}{\sinh\Psi},\\
\tfrac{1}{\sinh(-v)} &= -\tfrac{1}{\sinh(v)}=\tfrac{2}{e^{-v}-e^{v}}=\tfrac{2}{\frac{2}{\sinh\Psi}}=\sinh\Psi\\
\Longrightarrow \sinh\Psi &= \tfrac{1}{\sinh(-v)},\\
\coth(-v)=-\coth v &= \tfrac{e^{v}+e^{-v}}{e^{-v}-e^{v}}=\tfrac{\frac{2\cosh\Psi}{\sinh\Psi}}{\frac{2}{\sinh\Psi}}=\cosh\Psi\\
\Longrightarrow \cosh\Psi &= \coth(-v).
\end{aligned}$$

And yet,

$$\begin{aligned}
I(\Psi) &= \tfrac{\sinh^2\Psi}{2(\cosh\Psi+Q)}=\tfrac{\frac{1}{\sinh^2(-v)}}{2(\coth(-v)+Q)}=\tfrac{1}{2\sinh^2(-v)(\coth(-v)+Q)},\\
Y(\Psi) &= \tfrac{Q^2-1}{2(\cosh\Psi+Q)}=\tfrac{Q^2-1}{2(\coth(-v)+Q)},\\
Z(\Psi) &= \tfrac{(\cosh\Psi+Q)^2}{\sinh^2\Psi}=\tfrac{4(\coth(-v)+Q)(\coth(-v)+Q)}{4\frac{1}{\sinh^2(-v)}}\\
&= \tfrac{1}{4}\left(\tfrac{Q^2-1}{\frac{1}{2\sinh^2(-v)(\cosh(-v)+Q)}\frac{Q^2-1}{2(\cosh(-v)+Q)}}\right)\\
&= \tfrac{Q^2-1}{4A(-v)B(-v)}.
\end{aligned}$$

Then, denote

$$\begin{aligned}
A(-v) &:= \tfrac{1}{2\sinh^2(-v)(\coth(-v)+Q)},\\
B(-v) &= \tfrac{Q^2-1}{2(\coth(-v)+Q)},\\
D(-v) &:= \tfrac{Q^2-1}{4A(-v)B(-v)}.
\end{aligned}$$

Subsequently, introducing the integration variable $v=\ln\tanh\left(\frac{\Psi}{2}\right)$ results in

$$\begin{aligned}
&\frac{\mathrm{d}\Psi}{\sinh\Psi}=\mathrm{d}v,\ \ \sinh\Psi=\frac{1}{\sinh(-v)},\ \ \cosh\Psi=\coth(-v),\\
&I(\Psi)=A(-v),\ \ Y(\Psi)=B(-v),\ \ Z(\Psi)=D(-v),
\end{aligned} \tag{A47}$$

and

$$\begin{aligned}\tanh^{\zeta}\left(\frac{\Psi}{2}\right) &= \frac{(\cosh\Psi-1)^{\zeta}}{\sinh^{\zeta}(\Psi)} = \frac{(\coth(-v)-1)^{\zeta}}{\frac{1}{\sinh^{\zeta}(-v)}}\\ &= [\sinh(-v)(\coth(-v)-1)]^{\zeta}\\ &= \left(\frac{e^{-v}+e^{v}-e^{-v}+e^{v}}{2}\right)^{\zeta} = e^{\zeta v}. \end{aligned} \tag{A48}$$

By (A47) and (A48) we rewrite Equation (A46) as

$$\begin{aligned} & z^{\delta}e^{\frac{Qz}{2}}G(z;\zeta) = e^{r\alpha\gamma T}\Xi(\delta,\eta)\int_0^{\infty}e^{zY(\Psi)}Z^{\delta}(\Psi)\tanh^{\zeta}\left(\tfrac{\Psi}{2}\right)(zI(\Psi))^{\delta}e^{-\frac{zI(\Psi)}{2}}M_{\delta,\eta}(zI(\Psi))\tfrac{\mathrm{d}\Psi}{\sinh\Psi}\\ \Longrightarrow\quad & z^{\delta}e^{\frac{Qz}{2}}G(z;\zeta) = e^{r\alpha\gamma T}\Xi(\delta,\eta)\int_{-\infty}^{0}e^{zB(-v)-\frac{zA(-v)}{2}+\zeta v}D^{\delta}(-v)(zA(-v))^{\delta}M_{\delta,\eta}(zA(-v))\mathrm{d}v. \end{aligned}$$

Applying the following inverse Laplace transform

$$\begin{aligned} h(z,\tilde{\tau}) &= \mathcal{L}^{-1}\{G(z;\zeta)\}(\tilde{\tau}) = \tfrac{1}{2\pi i}\lim_{T\longrightarrow\infty}\int_{N-iT}^{N+iT}e^{\zeta\tilde{\tau}}G(z;\zeta)\mathrm{d}\zeta\\ &= \tfrac{1}{2\pi i}\int_{N-i\infty}^{N+i\infty}e^{\zeta\tilde{\tau}}G(z;\zeta)\mathrm{d}\zeta\\ &= z^{-\delta}e^{r\alpha\gamma T-\frac{Qz}{2}}\tfrac{\Xi(\delta,\eta)}{2\pi i}\int_{N-i\infty}^{N+i\infty}\int_{-\infty}^{0}e^{zB(-v)-\frac{zA(-v)}{2}+\zeta v+\zeta\tilde{\tau}}D^{\delta}(-v)(zA(-v))^{\delta}M_{\delta,\eta}(zA(-v))\mathrm{d}v\mathrm{d}\zeta \end{aligned}$$

to the function (A33), we obtain

$$\begin{aligned} f(S,t) &= z^{\delta}e^{-r\alpha\gamma t+R^{*}\tilde{\tau}+\frac{Q}{2}z}h(z,\tilde{\tau})\\ &= \frac{\Xi(\delta,\eta)}{2\pi i}e^{r\alpha\gamma(T-t)+R^{*}\tilde{\tau}}\int_{N-i\infty}^{N+i\infty}\int_{-\infty}^{0}e^{zB(-v)-\frac{zA(-v)}{2}+\zeta(v+\tilde{\tau})}(zD(-v)A(-v))^{\delta}M_{\delta,\eta}(zA(-v))\mathrm{d}v\mathrm{d}\zeta, \end{aligned} \tag{A49}$$

where N is chosen such that all singularities of the integrand expression are to the left of the straight line $(N-i\infty, N+i\infty)$ in the complex plane.

Additionally, as we know, by analytic continuation of the Fourier transform, the Laplace transform of the delta-function satisfies

$$\int_0^{\infty}\delta(t-a)e^{-st}\mathrm{d}t = e^{-sa}.$$

By using the inverse Laplace transform, the Dirac delta function for the M-CEV case with ambiguity is given by

$$\frac{1}{2\pi i}\int_{N-i\infty}^{N+i\infty}e^{z\zeta}\mathrm{d}\zeta = \delta(z),$$

and changing the order of integration in (A49) yields

$$\begin{aligned} f(S,t) &= \frac{\Xi(\delta,\eta)}{2\pi i}e^{r\alpha\gamma(T-t)+R^{*}\tilde{\tau}}\int_{N-i\infty}^{N+i\infty}\int_{-\infty}^{0}e^{zB(-v)-\frac{zA(-v)}{2}}e^{\zeta(v+\tilde{\tau})}(zD(-v)A(-v))^{\delta}M_{\delta,\eta}(zA(-v))\mathrm{d}v\mathrm{d}\zeta\\ &= \Xi(\delta,\eta)e^{r\alpha\gamma(T-t)+R^{*}\tilde{\tau}}\int_{-\infty}^{0}\delta((-v)-\tilde{\tau})e^{zB(-v)-\frac{zA(-v)}{2}}(zD(-v)A(-v))^{\delta}M_{\delta,\eta}(zA(-v))\mathrm{d}(-v). \end{aligned} \tag{A50}$$

Note that $\tilde{\tau}\geq 0$, the integration interval can be extended to the entire line. Using the properties of the delta-function

$$\int_{-\infty}^{\infty}\delta(\zeta-z)g(\zeta)\mathrm{d}\zeta = g(z), \tag{A51}$$

we express (A50) as

$$\begin{aligned} f(S,t) &= \Xi(\delta,\eta)e^{r\alpha\gamma(T-t)+R\tilde{\tau}}e^{zB(\tilde{\tau})-\frac{zA(\tilde{\tau})}{2}}(zD(\tilde{\tau})A(\tilde{\tau}))^{\delta}M_{\delta,\eta}(zA(\tilde{\tau})) \\ &= e^{r\alpha\gamma(T-t)+R^*\tilde{\tau}+zB(\tilde{\tau})}D^{\delta}(\tilde{\tau})\frac{\Gamma(\eta-\delta+\frac{1}{2})}{\Gamma(1+2\eta)}e^{-\frac{zA(\tilde{\tau})}{2}}(zA(\tilde{\tau}))^{\delta}M_{\delta,\eta}(zA(\tilde{\tau})). \end{aligned} \tag{A52}$$

Due to $T-t=\frac{\tilde{\tau}}{\sigma^2\beta^2\omega}$, the expression (A52) can be written as

$$\begin{aligned} f(S,t) &= e^{\frac{r\alpha\gamma}{\sigma^2\beta^2\omega}\tilde{\tau}+R^*\tilde{\tau}+zB(\tilde{\tau})}D^{\delta}(\tilde{\tau})\tfrac{\Gamma(\eta-\delta+\frac{1}{2})}{\Gamma(1+2\eta)}e^{-\frac{zA(\tilde{\tau})}{2}}(zA(\tilde{\tau}))^{\delta}M_{\delta,\eta}(zA(\tilde{\tau})) \\ &= e^{\left(\frac{r\alpha\gamma}{\sigma^2\beta^2\omega}+R^*\right)\tilde{\tau}+zB(\tilde{\tau})}D^{\delta}(\tilde{\tau})\tfrac{\Gamma(\eta-\delta+\frac{1}{2})}{\Gamma(1+2\eta)}e^{-\frac{zA(\tilde{\tau})}{2}}(zA(\tilde{\tau}))^{\delta}M_{\delta,\eta}(zA(\tilde{\tau})). \end{aligned}$$

Denote

$$R := \frac{r\alpha\gamma}{\sigma^2\beta^2\omega}+R^*.$$

Again since (see (A32))

$$R^* = -2Q\delta,$$

it follows that

$$R = \frac{r\alpha\gamma}{\sigma^2\beta^2\omega}-2Q\delta. \tag{A53}$$

Therefore, we obtain the solution $f(S,t)$ for Cauchy problem (A13) as follows:

$$f(S,t) = e^{R\tilde{\tau}+zB(\tilde{\tau})}D^{\delta}(\tilde{\tau})\frac{\Gamma(\eta-\delta+\frac{1}{2})}{\Gamma(1+2\eta)}e^{-\frac{zA(\tilde{\tau})}{2}}(zA(\tilde{\tau}))^{\delta}M_{\delta,\eta}(zA(\tilde{\tau})), \tag{A54}$$

where $z=\omega S^{-2\beta}$, $\tilde{\tau}=\sigma^2\beta^2\omega(T-t)$, $\Gamma(z)$ is the gamma function, $M_{\delta,\eta}(z)$ is the Whittaker function with parameters

$$\delta = -\frac{1}{2}-\frac{1}{2\beta}\left(\frac{1}{2}-\frac{\alpha\bar{\lambda}\gamma}{\sigma(\gamma-\phi)}\right), \quad \eta = \sqrt{\left(\delta+\frac{1}{2}\right)^2+\frac{\alpha\gamma\bar{\lambda}^2}{4\sigma^2\beta^2(\gamma-\phi-1)}}.$$

The remaining constants and functions are given by

$$\begin{aligned} \omega &= \tfrac{r}{\sigma^2|\beta|}, \ Q=\tfrac{r}{\sigma^2\beta\omega}, \ R=\tfrac{r\alpha\gamma}{\sigma^2\beta^2\omega}-2Q\delta, \\ A(\tilde{\tau}) &= \tfrac{1}{2\sinh^2\tilde{\tau}(\coth\tilde{\tau}+Q)}, \ B(\tilde{\tau})=\tfrac{Q^2-1}{2(\coth\tilde{\tau}+Q)}, \ D(\tilde{\tau})=\tfrac{Q^2-1}{4A(\tilde{\tau})B(\tilde{\tau})}. \end{aligned}$$

To compute the optimal exposure, we must determine the ratio $\frac{f_S}{f}$ in the next section.

Appendix A.2.5. Calculating the Ratio $\frac{f_S(S,t)}{f(S,t)}$

As documented in the reference [22], the ratio $\frac{f_s}{f}$ can be determined using the differential rules for Whittaker functions, i.e.,

$$\left(z\frac{\mathrm{d}}{\mathrm{d}z}z\right)^n\left(e^{-\frac{z}{2}}z^{k-1}M_{k,\mu}(z)\right) = \frac{\Gamma(\mu+k+n+\frac{1}{2})}{\Gamma(\mu+k+\frac{1}{2})}e^{-\frac{z}{2}}z^{k+n-1}M_{k+n,\mu}(z).$$

When $n=1$, it follows that

$$\begin{aligned} & \left(z\tfrac{\mathrm{d}}{\mathrm{d}z}z\right)\left(e^{-\frac{z}{2}}z^{k-1}M_{k,\mu}(z)\right) = \tfrac{\Gamma(\mu+k+\frac{1}{2}+1)}{\Gamma(\mu+k+\frac{1}{2})}e^{-\frac{z}{2}}z^{k+1-1}M_{k+1,\mu}(z) \\ \Longrightarrow \quad & \left(z\tfrac{\mathrm{d}}{\mathrm{d}z}\right)\left(e^{-\frac{z}{2}}z^{k-1+1}M_{k,\mu}(z)\right) = \tfrac{(\mu+k+\frac{1}{2})\Gamma(\mu+k+\frac{1}{2})}{\Gamma(\mu+k+\frac{1}{2})}e^{-\frac{z}{2}}z^{k}M_{k+1,\mu}(z) \\ \Longrightarrow \quad & \tfrac{\mathrm{d}}{\mathrm{d}z}\left(e^{-\frac{z}{2}}z^{k}M_{k,\mu}(z)\right) = \left(\mu+k+\tfrac{1}{2}\right)e^{-\frac{z}{2}}z^{k-1}M_{k+1,\mu}(z). \end{aligned}$$

In the previous section, we derived the solution $f(S,t)$ that satisfies the following expression (that is, function (A54)):

$$f(S,t) = e^{R\tilde{\tau}+zB(\tilde{\tau})}D^{\delta}(\tilde{\tau})\frac{\Gamma(\eta-\delta+\frac{1}{2})}{\Gamma(1+2\eta)}e^{-\frac{z}{2}A(\tilde{\tau})}(zA(\tilde{\tau}))^{\delta}M_{\delta,\eta}(zA(\tilde{\tau})),$$

where $\Gamma(z)$ is the Euler gamma function, $M_{\lambda,\eta}(z)$ is the Whittaker M-function, and functions $A(\tilde{\tau})$, $B(\tilde{\tau})$, and $D(\tilde{\tau})$ are obtained in Appendix A.2.4; namely,

$$\begin{aligned} A(\tilde{\tau}) &= \frac{1}{2\sin h^2(\tilde{\tau})\left(\coth(\tilde{\tau})+Q\right)}, \\ B(\tilde{\tau}) &= \frac{Q^2-1}{2\left(\coth(\tilde{\tau})+Q\right)}, \quad D(\tilde{\tau}) = \frac{Q^2-1}{4A(\tilde{\tau})B(\tilde{\tau})}. \end{aligned}$$

For convenience in calculations, denote

$$G := D^{\delta}(\tilde{\tau})\frac{\Gamma(\eta-\lambda+\frac{1}{2})}{\Gamma(1+2\eta)}$$

and thus, we rewrite the function (A54) as

$$f(S,t) = Ge^{R\tilde{\tau}+zB(\tilde{\tau})}e^{-\frac{z}{2}A(\tilde{\tau})}(zA(\tilde{\tau}))^{\delta}M_{\delta,\eta}(zA(\tilde{\tau})). \tag{A55}$$

The derivative of the function $f(S,t)$ with respect to S is given by

$$\begin{aligned} f_s(S,t) &= G\left[\frac{\partial e^{R\tilde{\tau}+zB(\tilde{\tau})}}{\partial S}\left(e^{-\frac{z}{2}A(\tilde{\tau})}(zA(\tilde{\tau}))^{\delta}M_{\delta,\eta}(zA(\tilde{\tau}))\right) + e^{R\tilde{\tau}+zB(\tilde{\tau})}\frac{\partial}{\partial S}\left(e^{-\frac{z}{2}A(\tilde{\tau})}(zA(\tilde{\tau}))^{\delta}M_{\delta,\eta}(zA(\tilde{\tau}))\right)\right] \\ &= Ge^{R\tilde{\tau}+zB(\tilde{\tau})}e^{-\frac{z}{2}A(\tilde{\tau})}(zA(\tilde{\tau}))^{\delta}M_{\delta,\eta}(zA(\tilde{\tau}))\left[B(\tilde{\tau}) + \frac{\left(\eta+\delta+\frac{1}{2}\right)}{z}\frac{M_{\delta+1,\eta}(zA(\tilde{\tau}))}{M_{\delta,\eta}(zA(\tilde{\tau}))}\right]\frac{\mathrm{dz}}{\mathrm{dS}}. \end{aligned}$$

Consequently,

$$\begin{aligned} \frac{g_s}{g} &:= \frac{g_s(S,T-t)}{g(S,T-t)} = \frac{f_s(S,t)}{f(S,t)} \\ &= \frac{Ge^{R\tilde{\tau}+zB(\tilde{\tau})}e^{-\frac{z}{2}A(\tilde{\tau})}(zA(\tilde{\tau}))^{\delta}M_{\delta,\eta}(zA(\tilde{\tau}))\left[B(\tilde{\tau}) + \frac{\left(\eta+\delta+\frac{1}{2}\right)}{z}\frac{M_{\delta+1,\eta}(zA(\tilde{\tau}))}{M_{\delta,\eta}(zA(\tilde{\tau}))}\right]\frac{\mathrm{dz}}{\mathrm{dS}}}{Ge^{R\tilde{\tau}+zB(\tau)}e^{-\frac{z}{2}A(\tilde{\tau})}(zA(\tilde{\tau}))^{\delta}M_{\delta,\eta}(zA(\tilde{\tau}))} \\ &= \left[B(\tilde{\tau}) + \frac{\delta+\eta+\frac{1}{2}}{z}\frac{M_{\delta+1,\eta}\left(zA(\tilde{\tau})\right)}{M_{\delta,\eta}\left(zA(\tilde{\tau})\right)}\right]\frac{\mathrm{dz}}{\mathrm{dS}}. \end{aligned} \tag{A56}$$

Appendix A.3. The Optimal Exposure and Worse-Case Measure

Building upon the preceding analysis, we derive the following optimal exposure from (A10):

$$\theta^* = \frac{\bar{\lambda}S^{\beta}}{(\phi-\gamma+1)}\frac{\left(x-Fe^{-r(T-t)}\right)}{x} + \sigma S^{\beta+1}\frac{\left(x-Fe^{-r(T-t)}\right)}{x}\frac{g_s}{g}. \tag{A57}$$

Substituting $\frac{\mathrm{dz}}{\mathrm{dS}} = -2\beta\omega S^{-2\beta-1}$ and (A56) into (A57), we can determine the portfolio's optimal exposure to the fundamental risk factor Z_t as follows:

$$\begin{aligned}\theta^* &= \frac{\bar{\lambda}S^{\beta}}{(\phi-\gamma+1)}\frac{\left(x-Fe^{-r(T-t)}\right)}{x}+\sigma S^{\beta+1}\frac{\left(x-Fe^{-r(T-t)}\right)}{x}\left[B(\tilde{\tau})+\frac{\delta+\eta+\frac{1}{2}}{z}\frac{M_{\delta+1,\eta}\left(zA(\tilde{\tau})\right)}{M_{\delta,\eta}\left(zA(\tilde{\tau})\right)}\right]\frac{\mathrm{d}z}{\mathrm{d}S}\\ &= \frac{\bar{\lambda}S^{\beta}}{(\phi-\gamma+1)}\frac{\left(x-Fe^{-r(T-t)}\right)}{x}-2\sigma\omega\beta S^{-\beta}\left[B(\tilde{\tau})+\frac{\delta+\eta+\frac{1}{2}}{z}\frac{M_{\delta+1,\eta}\left(zA(\tilde{\tau})\right)}{M_{\delta,\eta}\left(zA(\tilde{\tau})\right)}\right]\frac{\left(x-Fe^{-r(T-t)}\right)}{x}\\ &= \frac{x-Fe^{-r(T-t)}}{x}\left[\frac{\bar{\lambda}S^{\beta}}{\phi-\gamma+1}-2\sigma\omega\beta S^{-\beta}B(\tilde{\tau})-2\sigma\beta S^{\beta}\left(\delta+\eta+\frac{1}{2}\right)\frac{M_{\delta+1,\eta}\left(\omega A(\tilde{\tau})S^{-2\beta}\right)}{M_{\delta,\eta}\left(\omega A(\tilde{\tau})S^{-2\beta}\right)}\right]. \qquad \text{(A58)}\end{aligned}$$

By Formulas (A4) and (A57), we have

$$\begin{aligned}\epsilon^* &= \frac{\phi}{\gamma J}\left[xS^{-\beta}\left(\frac{\bar{\lambda}S^{\beta}}{(\phi-\gamma+1)}\frac{\left(x-Fe^{-r(T-t)}\right)}{x}+\sigma S^{\beta+1}\frac{\left(x-Fe^{-r(T-t)}\right)}{x}\frac{g_s}{g}\right)\frac{\gamma}{\left(x-Fe^{-r(T-t)}\right)}J+\sigma S\alpha^{-1}\frac{g_s}{g}J\right]\\ &= \frac{\phi\bar{\lambda}}{\phi-\gamma+1}+\frac{\phi\sigma S}{\gamma-\phi}\frac{g_s}{g}. \qquad \text{(A59)}\end{aligned}$$

We substitute (A56) into (A59) to obtain

$$\begin{aligned}\epsilon^* &= \tfrac{\phi\bar{\lambda}}{\phi-\gamma+1}+\tfrac{\phi\sigma S}{\gamma-\phi}\left[B(\tilde{\tau})+\tfrac{\delta+\eta+\frac{1}{2}}{z}\tfrac{M_{\delta+1,\eta}\left(zA(\tilde{\tau})\right)}{M_{\delta,\eta}\left(zA(\tilde{\tau})\right)}\right]\tfrac{\mathrm{d}z}{\mathrm{d}S}\\ &= \tfrac{\phi\bar{\lambda}}{\phi-\gamma+1}+\tfrac{\phi\sigma S}{\gamma-\phi}\left[-2\omega\beta S^{-2\beta-1}B(\tilde{\tau})-\tfrac{2\beta(\delta+\eta+\frac{1}{2})}{S}\tfrac{M_{\delta+1,\eta}\left(\omega A(\tilde{\tau})S^{-2\beta}\right)}{M_{\delta,\eta}\left(\omega A(\tilde{\tau})S^{-2\beta}\right)}\right].\end{aligned}$$

As a result, the worse-case measure is given by

$$\begin{aligned}e^{s_*} &= \epsilon^* S^{\beta}\\ &= \frac{\phi\bar{\lambda}S^{\beta}}{\phi-\gamma+1}+\frac{\phi\sigma S^{\beta+1}}{\gamma-\phi}\left[B(\tilde{\tau})+\frac{\delta+\eta+\frac{1}{2}}{z}\frac{M_{\delta+1,\eta}\left(zA(\tilde{\tau})\right)}{M_{\delta,\eta}\left(zA(\tilde{\tau})\right)}\right]\frac{\mathrm{d}z}{\mathrm{d}S} \qquad \text{(A60)}\\ &= \frac{\phi\bar{\lambda}S^{\beta}}{\phi-\gamma+1}+\frac{\phi\sigma S^{\beta}}{\phi-\gamma}\left[-2\omega\beta S^{-2\beta}B(\tilde{\tau})-2\beta\left(\delta+\eta+\frac{1}{2}\right)\frac{M_{\delta+1,\eta}\left(\omega A(\tilde{\tau})S^{-2\beta}\right)}{M_{\delta,\eta}\left(\omega A(\tilde{\tau})S^{-2\beta}\right)}\right] \qquad \text{(A61)}\end{aligned}$$

Appendix A.4. Special Case: Zero Interest-Free Rate

By assuming an interest-free rate, $r=0$, we can simplify the solution $f(S,t)$ and the optimal exposure θ^*. In this special case, we have $\omega=\tilde{\tau}=z=0$, and the constants Q and R are both zero. The limit values are

(a)

$$\begin{aligned}\lim_{\omega\longrightarrow 0}D(\tilde{\tau}) &= \lim_{\omega\longrightarrow 0}\frac{Q^2-1}{4A(\tau)B(\tilde{\tau})}=\lim_{\omega\longrightarrow 0}\frac{-1}{4\frac{1}{2\sinh^2\tilde{\tau}(\coth\tilde{\tau})}\frac{-1}{2(\coth\tilde{\tau})}}\\ &= \lim_{\omega\longrightarrow 0}\cosh^2\tilde{\tau}=\left(\frac{e^0+1}{2e^0}\right)^2\\ &= \left(\frac{1+1}{2}\right)^2=1. \qquad \text{(A62)}\end{aligned}$$

(b)

$$\begin{aligned}\varphi(S,t) &= \lim_{\omega\longrightarrow 0}zA(\tilde{\tau})=\lim_{\omega\longrightarrow 0}\frac{z}{2\sinh^2\tilde{\tau}(\coth\tilde{\tau}+Q)}=\lim_{\omega\longrightarrow 0}\frac{\omega S^{-2\beta}}{2\sinh^2\tilde{\tau}\coth\tilde{\tau}}\\ &= \lim_{\omega\longrightarrow 0}\frac{S^{-2\beta}}{2\sigma^2\beta^2(T-t)\cosh^2\tilde{\tau}+2\sigma^2\beta^2(T-t)\sinh^2\tilde{\tau}}\\ &= \tfrac{1}{2}\sigma^{-2}\beta^{-2}S^{-2\beta}(T-t)^{-1},\end{aligned}$$

because of $\sinh\tilde{\tau} = \frac{e^{\tilde{\tau}} - e^{-\tilde{\tau}}}{2}$.

(c)

$$\begin{aligned}
\Omega(S,t) &= \lim_{\omega\longrightarrow 0} zB(\tilde{\tau}) = \lim_{\omega\longrightarrow 0} \frac{z(Q^2-1)}{2(\coth\tilde{\tau}+Q)} = \lim_{\omega\longrightarrow 0} \frac{\omega S^{-2\beta}(\frac{r^2}{\sigma^4\beta^2\omega^2}-1)}{2(\coth\tilde{\tau}+\frac{r}{\sigma^2\beta\omega})} \\
&= \lim_{\omega\longrightarrow 0} \frac{(-2\omega\sigma^4\beta^2 S^{-2\beta})\sinh\tilde{\tau} + (S^{-2\beta}r^2 - \omega^2\sigma^4\beta^2 S^{-2\beta})\sigma^2\beta^2(T-t)\cosh\tilde{\tau}}{2\sigma^4\beta^2\cosh\tilde{\tau} + 2\sigma^4\beta^2\sigma^2\beta^2(T-t)\omega\sinh\tilde{\tau} + 2r\sigma^2\beta\sigma^2\beta^2(T-t)\cosh\tilde{\tau}} \\
&= \lim_{\omega\longrightarrow 0} -\left(\frac{2r\sigma^2|\beta|S^{-2\beta}\sinh\tilde{\tau}}{2\sigma^2\cosh\tilde{\tau} + 2\sigma^4\beta^2(T-t)\omega\sinh\tilde{\tau} + 2r\sigma^2\beta(T-t)\cosh\tilde{\tau}}\right) \\
&= \frac{0}{2\sigma^2} = 0.
\end{aligned}$$

(d)

$$\begin{aligned}
\Psi(t) = \lim_{\omega\longrightarrow 0} R\tilde{\tau} &= \lim_{\omega\longrightarrow 0}\left(\frac{r\gamma\alpha}{\sigma^2\beta^2\omega} - 2Q\delta\right)\sigma^2\beta^2\omega(T-t) \\
&= \lim_{\omega\longrightarrow 0}\left[\frac{r\gamma\alpha}{\sigma^2\beta^2\omega}\sigma^2\beta^2\omega(T-t) - 2\frac{r\delta}{\omega\beta\sigma^2}\sigma^2\beta^2\omega(T-t)\right] \\
&= \lim_{\omega\longrightarrow 0}\left[r\gamma\alpha(T-t) - 2r\delta\beta(T-t)\right] \\
&= 0. \qquad \text{(A63)}
\end{aligned}$$

Thus, the solution $f(S,t)$ is given by

$$\begin{aligned}
f(S,t) &= e^{R\tilde{\tau}+zB(\tilde{\tau})}D^{\delta}(\tilde{\tau})\frac{\Gamma(\eta-\delta+\frac{1}{2})}{\Gamma(1+2\eta)}e^{-\frac{zA(\tilde{\tau})}{2}}\left(zA(\tilde{\tau})\right)^{\delta}M_{\delta,\eta}\left(zA(\tilde{\tau})\right) \\
&= \frac{\Gamma\left(\eta-\delta+\frac{1}{2}\right)}{\Gamma(1+2\eta)}e^{\Psi(t)+\Omega(S,t)-\frac{\varphi(S,t)}{2}}\varphi^{\delta}(S,t)M_{\delta,\eta}\left(\varphi(S,t)\right) \\
&= \Xi(\delta,\eta)e^{-\frac{\varphi(S,t)}{2}}\varphi^{\delta}(S,t)M_{\delta,\eta}\left(\varphi(S,t)\right).
\end{aligned}$$

Then,

$$\begin{aligned}
\frac{g_s}{g} &= \frac{f_s(S,t)}{f(S,t)} \\
&= \frac{\Xi(\delta,\eta)e^{-\frac{\varphi(S,t)}{2}}\varphi^{\delta}(S,t)M_{\delta,\eta}\left(\varphi(S,t)\right)\left[B(\tilde{\tau}) + \frac{\left(\eta+\delta+\frac{1}{2}\right)}{z}\frac{M_{\delta+1,\eta}(\varphi(S,t))}{M_{\delta,\eta}(\varphi(S,t))}\right]\frac{\mathrm{dz}}{\mathrm{dS}}}{\Xi(\delta,\eta)e^{-\frac{\varphi(S,t)}{2}}\varphi^{\delta}(S,t)M_{\delta,\eta}(\varphi(S,t))} \\
&= -\frac{2\beta\left(\delta+\eta+\frac{1}{2}\right)}{S}\frac{M_{\delta+1,\eta}\left(\varphi(S,t)\right)}{M_{\delta,\eta}\left(\varphi(S,t)\right)}
\end{aligned}$$

and thereby, if $r = 0$, we have

$$\begin{aligned}
\theta^* &= \frac{\bar{\lambda}S^{\beta}}{(\phi-\gamma+1)}\frac{(x-F)}{x} + \sigma S^{\beta+1}\frac{(x-F)}{x}\frac{g_s}{g} \\
&= \frac{(x-F)S^{\beta}}{x}\left[\frac{\bar{\lambda}}{\phi-\gamma+1} - 2\sigma\beta\left(\delta+\eta+\frac{1}{2}\right)\frac{M_{\delta+1,\eta}\left(\varphi(S,t)\right)}{M_{\delta,\eta}\left(\varphi(S,t)\right)}\right]. \qquad \text{(A64)}
\end{aligned}$$

Meanwhile,

$$\begin{aligned}
\epsilon^* &= \frac{\phi\bar{\lambda}}{\phi-\gamma+1} + \frac{\phi\sigma S}{\gamma-\phi}\frac{g_s}{g} \\
&= \frac{\phi\bar{\lambda}}{\phi-\gamma+1} - \frac{2\phi\sigma\beta\left(\delta+\eta+\frac{1}{2}\right)}{\gamma-\phi}\frac{M_{\delta+1,\eta}\left(\varphi(S,t)\right)}{M_{\delta,\eta}\left(\varphi(S,t)\right)},
\end{aligned}$$

which implies that the worse case is given by

$$\begin{aligned} e^{s_*} &= \epsilon^* S^\beta \\ &= \frac{\phi\bar{\lambda}S^\beta}{\phi-\gamma+1} - \frac{2\phi\sigma\beta S^\beta\left(\delta+\eta+\frac{1}{2}\right)}{\gamma-\phi}\frac{M_{\delta+1,\eta}\left(\varphi(S,t)\right)}{M_{\delta,\eta}\left(\varphi(S,t)\right)}. \end{aligned} \tag{A65}$$

References

1. Merton, R.C. Lifetime portfolio selection under uncertainty: The continuous-time case. *Rev. Econ. Stat.* **1969**, *51*, 247–257. [CrossRef]
2. Kraft, H. Optimal portfolios and Heston's stochastic volatility model: An explicit solution for power utility. *Quant. Financ.* **2005**, *5*, 303–313. [CrossRef]
3. Liu, J.; Pan, J. Dynamic derivative strategies. *J. Financ. Econ.* **2003**, *69*, 401–430. [CrossRef]
4. Cox, J. *Notes on Option Pricing in Constant Elasticity of Variance Diffusions*; Technical Report, Stanford University Working Paper; Stanford University: Stanford, CA, USA, 1975.
5. Heath, D.; Platen, E. Consistent pricing and hedging for a modified constant elasticity of variance model. *Quant. Financ.* **2002**, *2*, 459. [CrossRef]
6. Muravei, D.L. Optimal portfolio management in a modified constant elasticity of variance model. *Comput. Math. Model.* **2018**, *29*, 110–119. [CrossRef]
7. Gao, J. Optimal portfolios for DC pension plans under a CEV model. *Insur. Math. Econ.* **2009**, *44*, 479–490. [CrossRef]
8. Escobar-Anel, M.; Fan, W.L. A New Type of CEV Model: Properties, Comparison and Application to Portfolio Optimization. SSRN. Available online: https://ssrn.com/abstract=4645366 (accessed on 24 May 2023).
9. Ellsberg, D. Risk, ambiguity, and the savage axioms. *Q. J. Econ.* **1961**, *75*, 643–669. [CrossRef]
10. Bossaerts, P.; Ghirardato, P.; Guarnaschelli, S.; Zame, W.R. Ambiguity in asset markets: Theory and experiment. *Rev. Financ. Stud.* **2010**, *23*, 1325–1359. [CrossRef]
11. Maenhout, P.J. Robust portfolio rules and asset pricing. *Rev. Financ. Stud.* **2004**, *17*, 951–983. [CrossRef]
12. Escobar, M.; Ferrando, S.; Rubtsov, A. Robust portfolio choice with derivative trading under stochastic volatility. *J. Bank. Financ.* **2015**, *61*, 142–157. [CrossRef]
13. Escobar-Anel, M.; Keller, M.; Zagst, R. Optimal HARA investments with terminal VaR constraints. *Adv. Oper. Res.* **2022**, *2022*, 6357701. [CrossRef]
14. Muravey, D. Optimal investment problem with M-CEV model: Closed form solution and applications to the algorithmic trading. *arXiv* **2017**, arXiv:1703.01574.
15. Merton, R.C. On estimating the expected return on the market: An exploratory investigation. *J. Financ. Econ.* **1980**, *8*, 323–361. [CrossRef]
16. Anderson, E.W.; Hansen, L.P.; Sargent, T.J. A quartet of semigroups for model specification, robustness, prices of risk, and model detection. *J. Eur. Econ. Assoc.* **2003**, *1*, 68–123. [CrossRef]
17. Schiff, J.L. *The Laplace Transform: Theory and Applications*, 1st ed.; Springer: Auckland, New Zealand, 1999.
18. Nikiforov, A.F.; Uvarov, V.B. *Special Functions of Mathematical Physics*; Nauka: Moscow, Russia, 1984.
19. Olver, F.; Lozier, D.; Boisvert, R.; Clark, C. *The NIST Handbook of Mathematical Functions*; Cambridge University Press: New York, NY, USA, 2010. Available online: https://dlmf.nist.gov/13.14 (accessed on 28 November 2023).
20. Gradshteyn, I.S.; Ryzhik, I.M. *Table of Integrals, Series, and Products*, 8th ed.; Elsevier Imprint—Academic Press: Cambridge, MA, USA, 2015.
21. Gradshteyn, I.S.; Ryzhik, I.M. *Tables of Integrals, Dums, Series, and Products*; Nauka: Moscow, Russia, 1963.
22. Abramowitz, M.; Stegun, I.A. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*; Dover Publications: Washington, DC, USA, 2023.

mathematics

MDPI

Article

Measurement and Forecasting of Systemic Risk: A Vine Copula Grouped-CoES Approach

Huiting Duan, Jinghu Yu and Linxiao Wei *

College of Science, Wuhan University of Technology, Wuhan 430070, China; htduan@whut.edu.cn (H.D.); jhyu@whut.edu.cn (J.Y.)
* Correspondence: lxwei@whut.edu.cn

Abstract: Measuring systemic risk plays an important role in financial risk management to control systemic risk. By means of a vine copula grouped-CoES method, this paper aims to measure the systemic risk of Chinese financial markets. The empirical study indicates that the banking industry has a low risk and a strong ability to resist risks, but also contributes the most of the systemic risk. On the other hand, insurance companies and securities have high ES but low ΔCoES, indicating their low risk tolerance and small contribution to the systemic risk. Furthermore, this study employs a sliding window in Monte Carlo simulation to forecast systemic risk. The findings of this paper suggest that different types of financial industries should adopt different systemic risk measures.

Keywords: vine copula grouped model; CoES; systemic risk; rolling of Monte Carlo simulation

MSC: 91B30; 91B28; 90C48

1. Introduction

A risk measure is to evaluate the risk of financial positions. Artzner [1] first introduced the coherent risk measures by an axiomatic approach. Later, Föllmer and Schied [2] and Frittelli and Rosazza Gianin [3] introduced a broader class of convex risk measures. Risk measures have been extensively studied in the literature. For a comprehensive literature overview, we refer to Föllmer and Schied [4]. At the same time, multivariate risk measures were initiated by Burgert and Rüschendorf [5], see also Wei and Hu [6]. A multivariate risk measure is to evaluate not only the risks of components of a portfolio separately, but also the joint risk of the portfolio caused by the possible dependence of components. For a comprehensive literature overview, we refer to Rüschendorf [7].

While univariate and multivariate risk measures are blooming, systemic risk measures have been attracting more and more researchers' attention. Chen et al. [8] first studied systemic risk measures by an axiomatic approach. Kromer et al. [9] further studied systemic risk measures on general measurable spaces. A systemic risk measure is to evaluate the risk of a whole financial system which consists of finitely many financial institutions. As a simple risk measure, value at risk (VaR) has been commonly adopted by financial institutions to evaluate the risk of financial positions. However, VaR is not sensitive to extreme events. By an extreme event, we mean an event that has a very small probability of occurring but has a huge potential loss. To accurately measure the systemic risk, Adrian and Brunnermeier [10] first proposed CoVaR (Conditional Value at Risk) based on VaR and also initiated the concept of CoES. CoVaR measures the risk spillover effect from a single institution to other financial institutions and the financial system in an extreme financial situation. CoVaR has been widely used since it was proposed. Many researchers choose to calculate CoVaR by quantile regression or the GARCH model. Using CoVaR and quantile regression, Bai and Shi [11] studied the impact of the risk of each financial institution on the systemic risk at different time periods, where the financial institution includes banks, securities, trust, and insurance companies. Based on CoVaR, Zhu et al. [12] introduced

Citation: Duan, H.; Yu, J.; Wei, L. Measurement and Forecasting of Systemic Risk: A Vine Copula Grouped-CoES Approach. *Mathematics* **2024**, *12*, 1233. https://doi.org/10.3390/math12081233

Academic Editors: Jing Yao, Xiang Hu and Jingchao Li

Received: 6 March 2024
Revised: 16 April 2024
Accepted: 17 April 2024
Published: 19 April 2024

state variables to simulate the time-varying risk, and studied the systemic risk of banking, insurance and securities industries via quantile regression. By constructing a multivariate GARCH model, Girardi and Ergün [13] estimated the systemic risk contribution of four financial industry groups consisting of a large number of institutions. Zhang et al. [14] measured the systemic risk using CoES and quantile regression. By constructing a CoES model and using quantile regression, Cui [15] evaluated the impact of risk of each financial institution on the financial system. Note, that either the GARCH model or the quantile regression can only describe the linear risk spillover, and hence can not describe the nonlinear risk spillover.

In order to describe the structure of dependence among financial institutions, copula is widely used in the GARCH model. Copula proposed by Sklar [16] is an effective tool to describe the dependence between financial assets. The Copula-CoES model can enable a more comprehensive assessment of systemic risk and enhance the accuracy of measuring systemic risk. However, the CoES is not used in dependency-based systemic risk measures in the literature. Currently, studies on measuring systematic risk based on copula still mainly focus on CoVaR. Yang et al. [17] employed the copula-CoVaR to calculate risk spillover between corporate and bank sector bonds. Li [18] constructed an ARFIMA-APARCH-GPD-SKST marginal distribution combined with a copula-CoES model to measure the risk spillover effect of the index of Chinese pillar industries and the CSI 300 index. Using an MSGARCH-Mixture copula model combined with CoVaR and CoES models, Li et al. [19] measured the risk spillover between off-shore and on-shore RMB interbank lending rates. In the above studies, the copula is limited to two variables. On the other hand, in order to evaluate the systemic risk of the financial system, multivariate copula are needed. Bedford and Cooke [20] introduced the so-called vine copula to describe the possible dependence structure between financial institutions. By constructing an R-vine-copula-CoVaR model, Lin et al. [21] measured the risk spillover effects between the international crude oil market, the international gold market, the U.S. stock market, the Chinese stock market and the foreign exchange market. Shahzad et al. [22] implemented a C-vine copula-CoVaR model to analyze the downside and upside spillover effects, systemic and tail dependence risks of the DJ World Islamic (DJWI) and DJ World Islamic Financial (DJWIF) indices. Based on the GARCH-R vine copula-CoVaR model, Zhang et al. [23] constructed the direct spillover matrix of systemic risk and further explored the indirect spillover path through R-vine. Zhu et al. [24] utilized an R-vine copula-CoES to measure the risk spillover effects among the carbon markets of Guangdong, Hubei, and Shenzhen.

Although the traditional vine copula can better describe the dependence of variables, the traditional vine copula does not reflect the mixed operation. That is, the traditional one can only consider the whole financial market as a whole, and ignore the different dependency structures for different financial industries. A copula-based grouped model proposed by Zhou et al. [25] groups the basic risks. Based on this model, the aggregated risk faced by a financial body under mixed operations can be measured. By dividing the industries, the copula-based grouped model effectively reduces the dimensionality and improves the accuracy of the dependency description. However, Zhou et al. [25] only divided the risk factors into two categories, and each category has only two basic risks. In reality, there are far more than the two categories of either financial industry or the number of financial institutions within each industry. To address this issue, Chen and Hao [26] proposed a vine copula-grouped model to describe the structure of interdependence within the financial market and demonstrated the advantage of the vine copula-grouped model. After that, based on the vine copula grouped model, Chen and Hao [27] constructed a mean-CVaR model to study the optimal portfolio selection. In addition, Hao and Chen [28] measured the systemic risk of financial markets by constructing a vine copula grouped-CoVaR model.

This paper investigates systemic risk measured by CoES in the Chinese financial market. To be precise, we combine the GJR-GARCH model and the vine copula grouped

model. This combination enables us to describe the dependency structure among three key industries: banking, securities and insurance. These three industries are of a high degree of mixed operation. Specifically, while constructing the marginal distribution, we account for volatility asymmetry and leverage effects by utilizing the GJR-GARCH(1,1) model. This allows us to accurately characterize the distribution of the data and better understand the underlying dynamics of the system. Moreover, this paper employs the sliding window algorithm to estimate the parameters of the dependency model. Monte Carlo simulation is performed to calculate the VaR and ES of each financial industry and the whole financial system. Finally, CoES is used to measure the systemic risk. Our results show that the GJR-Vine Copula grouped-CoES can enhance the accuracy of measuring financial systemic risk.

2. Methodology

In this section, we introduce the models and methods used in this paper, including the AR-GJR-GARCH model for constructing the marginal distribution and the distribution obeyed by the standardized residuals to be selected, the vine copula grouped model for describing the dependence structure, the definition of the risk measures VaR, ES and CoES, and the rolling Monte Carlo method for calculating the risk measures.

2.1. Marginal Distribution Modeling

Since financial time series are usually characterized by conditional fat-tailed, non-normality, skewed distribution, leverage effect, and volatility clustering, many studies have employed AR-GARCH models to capture these features. However, in the GARCH model, historical data affect future volatility in the form of squares, thus the effect of increase or decrease on future volatility is the same. In 1993, Glosten et al. [29] showed that the same degree of positive news and negative news have significantly different effects on the volatility of financial assets, i.e., there is a leverage effect. We know that the negative shocks can lead to an increase in leverage, and thus increases risk. Therefore, we choose the GJR-GARCH model to capture the asymmetry of volatility.

It has been shown that during modeling the volatility of returns, using an excessively high model order makes parameter estimation difficult, and does not provide significant practical meaning. In the first-order model, since the current value indirectly contains all the historical information in the past, it has high accuracy, and thus is also close to the prediction results of higher-order models (Lamoureux and Lastrapes [30], Lin [21]). Therefore, we use the AR(1)-GJR-GARCH(1,1) model to describe the marginal distributions. After filtering the logarithmic returns by the AR(1)-GJR-GARCH(1,1) model, we select the student distribution, the skewed student distribution, the generalized error distribution (GED), and the skewed generalized error distribution (SGED) as the candidate distributions of the standard deviation. The AR(1)-GJR-GARCH(1,1) model can be represented as follows:

$$r_t = \mu + cr_{t-1} + \varepsilon_t, \tag{1}$$

$$\varepsilon_t = \sigma_t Z_t, \tag{2}$$

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + \gamma\varepsilon_{t-1}^2 I(\varepsilon_{t-1} < 0), \tag{3}$$

where r_t and r_{t-1} are the log returns at day t and day $t-1$, respectively, μ is the conditional mean of the log returns r_t, ε_t is the residual, σ_t^2 is the conditional variance of ε_t, and Z_t is the standard Gaussian residual. $I(\varepsilon_{t-1} < 0)$ is the indicator function of the event of $\varepsilon_{t-1} < 0$. γ is an asymmetric parameter to measure the leverage effect. When $\gamma > 0$, it indicates a negative leverage effect, while $\gamma < 0$ indicates a positive leverage effect. When $\gamma = 0$, GJR degenerates to a GARCH model. Due to the significant non-normality of the financial time series, we abandon the normal distribution. Instead, we use the four distributions aforementioned to describe the distribution of the normalized residual series.

2.2. Joint Distribution Modeling

2.2.1. R Vine Copula

In 2001, Bedford and Cooke [20] proposed a regular vine copula model to model the dependency of assets through the graph theory. Vine copula uses pair copula as the base module to construct multidimensional models, which can compensate for the deficiency of traditional multivariate copula in portraying the flexibility of interdependence. Therefore, vine copula is often used to describe the dependence among high-dimensional variables and has significant superiority in portraying the risk contagion relationship among high-dimensional assets. Compared to the C-vine and D-vine, the R-vine is constructed based on the actual dependence of each edge, which makes the R-vine more flexible in describing the dependence of assets. The vine copula function decomposes the traditional multivariate copula function into a series of binary copula. The vine structure consists of nodes, edges and trees. Each level of the tree consists of edges with two nodes, where the two nodes of each edge can be described by the copula function. In fact, the different vine structures are different decompositions of the multidimensional copula density function. For an n-component R-vine model, there are $n-1$ trees and n nodes. According to Bedford and Cooke [20], an n-dimensional R-vine density function can be expressed as follows:

$$\begin{aligned} & f(x_1, x_2, \cdots, x_n) \\ &= \prod_{k=1}^{n} f_k(x_k) \prod_{k=1}^{n-1} \prod_{e \in \mathrm{E}_k} c_{j(e),i(e)|D(e)} \left(F\left(x_{j(e)} \mid x_{D(e)} \right), F\left(x_{i(e)} \mid x_{D(e)} \right) \right), \end{aligned} \tag{4}$$

where $E_k = \{e_{k1}, \ldots, e_{kk}\}$ is the set of all edges of the k-th tree, $j(e)$ and $c(e)$ are the two nodes connecting edge e, $D(e)$ is the condition set, $c_{j(e),c(e)|D(e)}$ is the copula density function corresponding to e, and $F(\cdot \mid \cdot)$ is the conditional distribution function.

2.2.2. Vine Copula Grouped Model

Accurately describing the dependence structure is a prerequisite for accurately measuring systemic risk. Taking into account that the financial institutions may belong to different industries, Chen and Hao [26] proposed a vine copula grouped model to describe the dependence among financial assets. They first divided the financial institutions based on their respective industries and then used the vine copula model to describe the dependence structure among financial institutions belonging to the same industry and the dependence structure between different industries, respectively. In this process, the asset return of one industry can be obtained from the weighted sum of the asset returns of financial institutions in the industry. Then, the asset returns of each industry are treated as the new variables. Finally, the asset return of the whole financial system is obtained from the weighted sum of these new variables. According to Chen and Hao [26], the structure of the vine copula grouped model can be shown in Figure 1:

$$S(C_X)\begin{cases} X_1(C_{X_1})\begin{cases} X_{11}(F_{11}) \\ \vdots \\ X_{1n_1}(F_{1n_1}) \end{cases} \\ \vdots \\ X_N(C_{X_N})\begin{cases} X_{N1}(F_{N1}) \\ \vdots \\ X_{Nn_N}(F_{Nn_N}) \end{cases} \end{cases}$$

Figure 1. The structure of vine copula grouped model.

where $X_{i1}, \cdots, X_{in_i}$ are the asset returns of the financial institutions and $F_{i1}, \cdots, F_{in_i}$ are the marginal distribution functions corresponding to each financial institution in the i-th industry group for $1 \leqslant i \leqslant N$. $X_i, i = 1, \cdots, N$ is the asset return of the i-th industry group, $S = \sum_{i=1}^{N} \sum_{k=1}^{n_i} w_{ik} X_{ik}$ is the asset return of the whole financial system, w_{ik} is the weight of the k-th financial institution in the i-th industry. $C_{X_i}, i = 1, \cdots, N$ is the vine copula within the i-th industry group, respectively, and C_X is the vine copula between the whole industry groups.

In most of the aforementioned literature, a vine copula affects the dependence among all the financial institutions since all the financial institutions are considered as a whole. However, from the viewpoint of practice, the dependence of the financial institutions belonging to the same industry is not necessarily the same as the one of the financial institutions belonging to different industries. In 2016, Zhou et al. [25] divided the aggregated risk faced by all the financial institutions into several groups according to the different kinds of industries. Therefore, the vine copula grouped model can not only reduce the dimensionality to make the structure clearer but also has more practical significance.

2.3. The Definitions of VaR, ES and CoES

Let $(\Omega, \mathscr{F}, P)$ be a fixed probability space, and X be a random variable that represents the loss of a financial institution. VaR (Value at risk) is an important risk measure, which refers to the maximum expected loss within a certain confidence level over a certain period of time. The definition of VaR at the confidence level $1 - c$ can be given as:

$$\mathrm{VaR}_{c,t}(X) = -inf\{r \in \mathbb{R} : P(X \leqslant r \mid \Omega_{t-1}) > c\}, \tag{5}$$

where Ω_{t-1} is the set of information at the moment $t - 1$.

VaR can be used widely in different markets to measure the risk of positions, and provides a numerical value to quantify the potential loss. However, it does not satisfy the subadditivity of the coherent risk measure proposed by Artzner et al. [1]. More seriously, VaR only focuses on the extreme losses corresponding to the specified confidence level, while it ignores the severity of losses beyond the VaR level. In 2002, Acerbi et al. [31] proposed the expected shortfall (ES), which measures the average loss exceeding VaR under a certain confidence level. More importantly, ES satisfies the subadditivity, thus it is a coherent risk measure. In recent decades, the ES has replaced VaR as the measure metric to determine the minimum capital requirements for the financial market by the Basel Accord. ES is defined as:

$$\mathrm{ES}_c(X) = E[X \mid X < \mathrm{VaR}_{c,t}(X)] = \int_{-\infty}^{\mathrm{VaR}_{c,t}(X)} x f_X(x) dx, \tag{6}$$

where $f_X(x)$ is the probability density function of the risk position X, and $1 - c$ is the confidence level.

It is well known that once an individual institution is exposed to a crisis, systemic risk may occur. Both VaR and ES are difficult to accurately measure systemic risk. In 2016, Adrian and Brunnermeier [10] proposed CoVaR to measure the systemic risk, and proposed the initial idea of CoES. While CoVaR only focuses on the single quantile of the loss random variable, CoES pays more attention to tail average loss. Thus, CoES can be used as a risk measure to take into account the maximum tail average loss. $\mathrm{CoVaR}_c^{j|i}$ can be denoted by the VaR of the j-th financial industry is conditional on some event $\mathbb{C}(X^i)$ of the i-th financial industry. Its definition is given by

$$\Pr\left(X^j \leqslant \mathrm{CoVaR}_c^{j|i} \mid \mathbb{C}\left(X^i\right)\right) = c. \tag{7}$$

The calculation formula for $\mathrm{CoVaR}_c^{j|i}$ is:

$$\mathrm{CoVaR}_c^{j|X^i \leqslant \mathrm{VaR}_c^i} = \mathrm{VaR}_c^j \mid \mathrm{VaR}_c^i = \hat{\alpha}_c^i + \hat{\beta}_c^i \, \mathrm{VaR}_c^i, \tag{8}$$

where $\hat{\alpha}_c^i$ and $\hat{\beta}_c^i$ are quantile regression coefficients, and $1 - c$ is the confidence level of VaR.

As Adrian and Brunnermeier [10] illustrated, CoVaR can be adopted for the co-expected shortfall (CoES). $\mathrm{CoES}_c^{j|i}$ represents the expected shortfall of industry j (or system) due to industry i being under an abnormal extreme risk. $\mathrm{CoES}_c^{j|i}$ is defined by the expectation over the c-tail of the conditional probability distribution:

$$\mathrm{CoES}_c^{j|i} = E\left[X^j \mid X^j \leqslant \mathrm{CoVaR}_c^{j|i}, \mathbb{C}\left(X^i\right)\right]. \tag{9}$$

The contribution of the i-th financial industry to the j-th financial industry is denoted by $\Delta\mathrm{CoES}_c^{j|i}$, where $\Delta\mathrm{CoES}_c^{j|i}$ represents the difference between the ES of the j-th financial industry j (or system) under the condition of distress in industry i (c) and the ES of industry j (or system) when industry i is in a normal state (c = 50%).

$\Delta\mathrm{CoES}_c^{j|i}$ is the difference between the ES of industry j (or system) conditional on the distress of industry i and the ES of industry j (or system) when industry i is in a normal state (c = 50%). $\Delta\mathrm{CoES}_c^{j|i}$ is defined as

$$\Delta\,\mathrm{CoES}_c^{j|i} = E\left[X^j \mid X^j \leqslant \mathrm{CoVaR}_c^{j|i}, \mathbb{C}\left(X^i\right)\right] - E\left[X^j \mid X^j \leqslant \mathrm{CoVaR}_{50\%}^{j|i}\right]. \tag{10}$$

2.4. Estimation Methods

2.4.1. CoES with Vine Copula Grouped Model

Monte Carlo simulation plays a significant role in the measurement of financial risk. When calculating CoES, CoVaR is first computed through Equation (8), followed by utilizing Equation (9) to calculate CoES. In this paper, the vine copula grouped model is used to describe the dependency among financial institutions (or financial industries). Unlike the traditional Monte Carlo simulation method which is based on the vine copula, in this paper, we utilize a vine copula-grouped model. Therefore, during the Monte Carlo simulation it is necessary to distinguish between intra-group copulas and inter-group copulas for multiple simulations. The specific steps are as follows:

Step 1: The Monte Carlo method is employed to simulate random numbers within the range (0,1) according to the n-dimensional vine copula. The probability integration inverse transformation is applied to obtain the sequence of normalized residuals based on the distribution followed by the normalized residuals of each marginal distribution. Then the residuals of each series are obtained by Equation (2), and thus the return series of each institution are obtained by Equation (1). Finally, the industry (or system) return rates are obtained by the weighted sum of the simulated rates of return.

Step 2: Sort the obtained return rates of the industries in an increasing order. From Equations (5) and (6), the VaR value of the industry (or system) is equal to the generated random numbers m multiplied by the selected significance level. ES value of the industry (or system) is equal to the average of the return rates smaller than the VaR.

Step 3: The CoVaR is calculated by quantile regression, and the CoES is calculated by CoVaR according to Equation (9). To obtain more accurate and robust results, the above steps can be repeated several times and then averaged.

2.4.2. Rolling of Monte Carlo Simulation Based on a Vine Copula Grouped Model

In this paper, we use the rolling of Monte Carlo simulation to simulate the marginal distributions of the vine copula grouped model, and then calculate VaR, ES and CoES. In simple terms, calculating the returns of the financial industries and the financial system in the rolling Monte Carlo simulation is to repeat the above method multiple times. With the rolling Monte Carlo method, we can obtain multiple sets of data, which is more advantageous for analyzing systemic risk and observing changes in risk measures more clearly. At the same time, the data obtained from frequent forecasting are more consistent with the current state of the financial markets. The specific steps are as follows:

Step 1: We divide the overall sample (t = 1, 2, 3, . . . , 1460) into an "estimation sample" and "prediction sample" where the first 1300 samples are selected as "estimation sample" and the last 160 samples are "prediction sample". After constructing the marginal distributions for the entire population, we utilize the data from t = 1, 2, 3, . . . , 1300 as the first estimation sample to build a vine copula, and estimate the parameters using the probability integral transformation (PIT) series.

Step 2: By the above Monte Carlo simulation based on the vine copula grouped model with the simulated parameters, we obtain the value of VaR, ES, CoVaR and CoES.

Step 3: Keeping the length of the estimated sample interval, and shifting the estimated sample interval backward by one day. That is, using data from t = 2, 3, 4, . . . , 1301 as the second rolling estimation sample interval and repeating the Monte Carlo simulations again. Thus, we can obtain the simulated return rates for the 1301st day and the risk measures.

Step 4: Repeat the above steps until the simulated return rate and risk measures for the last day are obtained.

3. Data and Descriptive Statistics

We conduct an empirical study on the systemic risk of the Chinese financial industry based on the multiple financial institutions belonging to different financial industries. To ensure a certain level of representativeness of the selected financial institutions and to consider data availability, we selected 20 financial institutions from the industry classification of the China Securities Regulatory Commission in 2012, including 10 banking institutions, 7 securities institutions, and 3 insurance institutions (see Table 1). We obtain the daily closing prices of the selected financial institutions from the CSMAR database, covering the period from 13 October 2016 to 14 October 2022. After removing the unmatched data among the daily closing prices of the 20 financial institutions, we obtain a total of 1460 observations. To ensure the stationarity of the data, we used logarithmic returns as the variable. The calculation formula for logarithmic returns is as follows:

$$X_{i,t} = \ln \mathrm{P}_{i,t} - \ln \mathrm{P}_{i,t-1}, \tag{11}$$

where P_{it} is the daily closing price of stock i at time t.

Table 1. Financial institutions selected in each industry.

Industry	Financial Institutions
Banking	Industrial and Commercial Bank of China (601398), China Construction Bank (601939), Agricultural Bank of China (601288), Bank of China (601988), Bank of Communications (601328), China Merchants Bank (600036), Industrial Bank (601166), China Citic Bank (601998), China Minsheng Bank (600016), China Everbright Bank (601818)
Securities	Citic Securities (600030), Huatai Securities (601688), Guotai Junan (601211), China Merchants Securities (600999), Haitong Securities (600837), GF Securities (000776), Guosen Securities (002736)
Insurance	China Life (601628), Ping An Insurance (601318), China Pacific Insurance Company (601601)

Notes: The numbers in ( ) is the stock code of each financial institution.

After applying the logarithm transformation to the samples, we obtain a total of 1460 sets of logarithmic returns. Due to space limitations, only the descriptive statistics of industry returns are presented here. The data in Table 2 represent the descriptive statistical characteristics of the logarithmic returns for each financial industry. Table 2 shows that the skewness of all financial industry returns is non-zero, and the excess kurtosis is greater than zero, indicating the presence of the typical "peak and fat-tailed" distribution. Specifically, the mean of logarithmic returns for every industry is close to zero, and the standard deviation of the banking industry is smaller than that of the securities and insurance industries. Therefore, the banking industry is relatively more stable compared to the other two industries, and this is consistent with the general perception of the Chinese financial market. The securities industry has the highest maximum value and the lowest minimum value among the three industries. This is because stock prices experience significant

increases during bull markets, and decreases during bear markets, which leads to high volatility. In contrast, the banking industry has the largest minimum extreme value and the smallest maximum extreme value, as bank stocks exhibit stable fluctuations regardless of whether the overall market is in a bull or bear market. Even in a bear market, the decline in bank stocks is relatively smaller compared to the overall market, reflecting the stability of bank stocks. Besides, the skewness of all three financial industries is greater than 0, which means they all exhibit a right-skewed distribution. This suggests that positive returns are more likely to be observed in the selected time period for all three financial industries.

Except for the descriptive statistics, Table 2 also shows the test statistics for the normality test, autocorrelation test, ARCH effect test and stationarity test. The daily logarithmic returns for all financial industries significantly reject the assumption of normality. The results of the L-BQ test indicate the autocorrelation in the three industries, with the banking industry exhibiting the strongest autocorrelation. The ARCH-LM test results demonstrate significant volatility clustering in all industries. Furthermore, to avoid spurious regression, we conduct a stationarity test on the log returns of each industry. The results of the ADF test are all significant, indicating that the data series are stationary, which ensures the stability of the proposed model.

Table 2. Financial institutions selected in each industry.

	Banking	**Securities**	**Insurance**
Mean	−0.000088	−0.000222	0.000048
Std	0.010453	0.017625	0.018187
Max	0.081269	0.095291	0.092121
Min	−0.064432	−0.105246	−0.087806
Kurtosis	5.743234	5.528226	2.325904
Skewness	0.336056	0.445978	0.281523
J-B	1170.4 ***	1584.6 ***	265.97 ***
Q (15)	42.877 ***	30.104 **	27.483 **
LM (5)	65.612 ***	55.915 ***	55.669 ***
ADF	−11.512 ***	−10.998 ***	−11.491 ***

*** Indicate significance at 1% level. ** Indicate significance at 5% level.

4. Empirical Results

In this section, we conduct an empirical analysis based on the aforementioned methodology. First, we estimate the marginal distribution of each financial institution (or industry) with an AR(1)-GJR-GARCH(1,1) model and choose the distribution for the standardized residuals of each institution's (or industry's) returns based on the maximum likelihood estimation. Then, we divide the standardized residual series of the samples into estimation and prediction samples. We perform a rolling Monte Carlo estimation of the model based on the vine copula grouped model to calculate the values of the prediction interval for VaR, ES, CoVaR, CoES and ΔCoES.

4.1. Constructing Marginal Model

First, focusing on the industry internally, descriptive statistics of financial institutions within each industry indicate that the industry returns exhibit common characteristics of financial data, including non-normality, autocorrelation and volatility clustering. Therefore, after estimating the institutional returns using the AR(1)-GJR-GARCH(1,1) model, we fit the standardized residuals with Student's t distribution, skewed Student's t distribution, generalized error distribution and skewed generalized error distribution, respectively. Then, we select the distribution with the maximum likelihood value based on the maximum likelihood criterion. Due to the large number of institutions, only the distributions of the standardized residuals are shown in Table 3. From Table 3, we can observe that the standardized residuals of most financial institutions follow the skewed Student t-distribution or the skewed generalized error distribution, while only the standardized residuals of the logarithmic returns of Agricultural Bank of China follow the generalized

error distribution. Therefore, we can conclude that, except for the Agricultural Bank of China, the residual series of other financial institutions exhibit heavy tails and asymmetry.

Table 3. Distribution of standardized residuals by financial institution.

	sstd	**ged**	**sged**
Banking	601398, 601939, 601328, 601166, 601998, 600016, 601818	601288	601988, 600036
Securities	600030, 601688, 601211, 600999, 600837, 000776, 002736	-	-
Insurance	601628	-	601318, 601601

Notes: For simplicity, each financial institution is represented by its stock code.

After grouping the financial institutions within each industry, we obtain the logarithmic returns of each financial industry by weighting the logarithmic return of the financial institutions within each industry. Based on the descriptive statistical characteristics of each financial industry provided in the previous section, we can understand the heavy-tailed and asymmetric characteristics of the financial industry returns. Then, after estimating the industry returns using the AR(1)-GJR-GARCH(1,1) model, we fit the standardized residual series with the four candidate distributions mentioned above, and the parameter estimation results are shown in Table 4. According to the leverage parameter estimation results of the GJR-GARCH, the leverage parameters for the banking industry and securities industry are both less than 0, indicating that positive news has a greater impact on the banking and securities industries compared to negative news. In terms of the magnitude of the coefficients, positive news has a greater impact on the banking industry than on the securities industry. On the other hand, the leverage coefficient for the insurance industry is greater than 0, suggesting that negative news has a greater impact on the insurance industry than positive news. The last row in Table 4 displays the selected distribution for the standardized residuals based on the maximum likelihood criterion. We can observe that the standardized residuals of the banking and securities industries follow a skewed Student's t distribution, while the standardized residuals of the insurance industry follow a skewed generalized error distribution. Overall, most parameters are significant with high likelihood function estimates, indicating reasonable estimation of the marginal distributions.

Table 4. Parameter estimation results of the marginal distribution models.

	Banking	**Securities**	**Insurance**
μ	0.0001 (0.0002)	−0.0002 (0.0003)	0.0003 (0.0005)
c	−0.0092 (0.0261)	−0.0681 *** (0.0226)	−0.0333 (0.0252)
ω	0.0000 *** (0.0000)	0.0000 (0.0000)	0.0000 ** (0.0000)
α	0.1144 *** (0.0203)	0.0555 *** (0.0144)	0.0589 *** (0.0138)
β	0.8359 *** (0.0182)	0.9423 *** (0.0134)	0.9160 *** (0.0000)
γ	−0.0292 (0.0349)	−0.0052 (0.0192)	0.0132 (0.6225)
skew	1.0887 *** (0.0378)	1.1035 *** (0.0384)	1.1124 *** (0.0348)
shape	4.5762 *** (0.5386)	3.2217 *** (0.3045)	1.2184 *** (0.0593)
LL	4744.27	4049.919	3877.209
Z_t	sstd	sstd	sged

*** Indicate significance at 1% level. ** Indicate significance at 5% level.

4.2. Constructing Vine Copula Grouped Model

After modeling the marginal distribution of each return series with the AR(1)-GJR-GARCH(1,1) model, we obtain the standardized residual sequences. Then, we estimate the skewness parameters and degrees of freedom parameters for each standardized residual sequence according to the corresponding candidate distribution. We perform probability

integral transforms based on the obtained parameters. The transformed sequences followed a uniform distribution on (0,1), which serve as the input variables for the vine copula grouped model. We construct intra-group vine copula with PIT (Probability Integral Transform) sequences of financial institutions with t = 1, 2, 3, . . . , 1300 within each industry separately. Thus, we built inter-group vine copulas using the PIT sequences of each financial industry. Consequently, the vine copula grouped model is obtained for the first rolling estimation. By keeping the estimation interval fixed and shifting it one day backward at a time, we obtain the vine copula combination grouped for the second rolling estimation. Repeat the above steps until the last day. We can obtain all the vine copula grouped models. It is worth noting that only the vine structure of the first rolling estimation is shown here, and the vine copula among financial industries may not follow the same structure in subsequent vine copula estimations.

In constructing the vine copula, we choose the flexible R-vine copula to describe the dependency structure among the returns of financial institutions within groups and the dependency structure across financial industries. From Figure 2, we observe the inter-industry dependency structure generated by the first rolling estimate, where this three-dimensional R-vine comprises two trees. T_1 has two edges (C_{12}, C_{13}), and the two edges represent the dependence between the banking and securities industries and the dependence between the banking and insurance industries, respectively. T_2 has one edge ($C_{23|1}$), representing the conditional dependence relationship between the securities industry and the insurance industry conditioned on the banking industry.

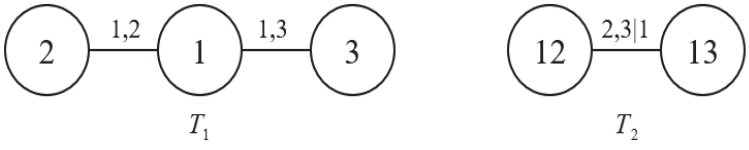

Figure 2. R-vine structure among industries.

In Table 5, we present the results of the inter-industry vine copula estimation from the first rolling estimation. The parameters are estimated according to the AIC criterion and the maximum likelihood estimation method. Based on the estimated parameters of the inter-industry vine copula, in the first estimation interval, the dependence among the three industries is dominated by the banking industry, and the dependence relationships between industries are described by t copulas. This indicates that there is fat-tailedness and symmetry in the dependence relationships among the industries. In addition, the one between the banking and insurance industries is stronger than the one between the banking and the securities industry. This suggests that the dependence relationship between the banking industry and the insurance industry is stronger than the dependence relationship between the banking industry and the securities industry.

Table 5. Estimation of vine copula among industries.

Tree	Edge	Copula	Par	Par2
T_1	(1,2)	t	0.57	7.49
	(3,1)	t	0.73	4.84
T_2	(3,2;1)	t	0.29	8.57

4.3. CoES Results

In this section, we first analyze the systemic risk among different financial industries and then proceed to examine the risk spillover of each financial industry to the whole financial system. We calculate the VaR and ES of each industry, as well as CoVaR, CoES

and ΔCoES of the systemic risk measures at 97.5% confidence level and 99% confidence level, respectively.

4.3.1. The CoES between Financial Industries

After constructing the dependency model, we use the Monte Carlo method to calculate risk measures such as VaR, ES, CoES and ΔCoES. The results of the first rolling estimate of the risk measures are presented in Table 6. Taking the banking industry as an example, to calculate the risk values for the banking industry, we employ the aforementioned Monte Carlo method to simulate 10,000 sets of 10-dimensional return rate sequences and calculate the weighted sum. This yields 10,000 sets of industry returns. Then, we sort the industry returns and calculate the VaR and ES according to the corresponding significance levels. Similarly, VaR and ES values for other industries are calculated. Once the VaR of each financial industry was calculated, we can use Equations (8) and (9) to calculate the systemic risk between each pair of industries separately.

According to Table 6, we know that at the same significance level, the banking industry has the lowest inherent risk, while the insurance industry has the highest inherent risk. Specifically, the banking industry has the smallest VaR and ES, which implies that the banking industry faces the least potential loss when the entire financial system experiences negative shocks. This means that the banking industry is relatively more stable compared to other sectors when facing adverse events. At the same significance level, the ES for all three industries is greater than VaR. The reason is that VaR ignores extreme risks, potentially leading to an underestimation of the actual risk. Additionally, the lower the significance level, the larger the gap between VaR and ES. Comparing the ES and CoES of each industry, we find that the risk of all industries is less than the systemic risk, indicating that the risk exposure of each industry is greater than the systemic risk of the industry. Furthermore, as the significance level decreases, the difference between the individual inherent risk and systemic risk becomes larger. Therefore, systemic risk in the financial market should receive more attention.

Table 6. VaR, ES, CoVaR, CoES, ΔCoES for industries.

Industry	c	VaR	ES	CoVaR	CoES	Δ CoES
Banking	2.5%	−0.0195	−0.0283			
	1%	−0.0266	−0.0372			
Banking to Securities	2.5%			−0.0502	−0.0734	−0.0631
	1%			−0.0647	−0.0966	−0.0863
Banking to Insurance	2.5%			−0.0438	−0.0562	−0.0428
	1%			−0.0607	−0.0750	−0.0617
Securities	2.5%	−0.0294	−0.0439			
	1%	−0.0407	−0.0590			
Securities to Banking	2.5%			−0.0255	−0.0367	−0.0297
	1%			−0.0360	−0.0511	−0.0441
Securities to Insurance	2.5%			−0.0423	−0.0547	−0.0413
	1%			−0.0590	−0.0732	−0.0598
Insurance	2.5%	−0.0362	−0.0475			
	1%	−0.0461	−0.0584			
Insurance to Banking	2.5%			−0.0289	−0.0416	−0.0346
	1%			−0.0386	−0.0546	−0.0476
Insurance to Securities	2.5%			−0.0477	−0.0698	−0.0595
	1%			−0.0680	−0.1022	−0.0919

Similarly, CoES exceeds CoVaR for all three industries. Additionally, as the significance level decreases, the disparity between CoVaR and CoES widens. This suggests that CoVaR may underestimate systemic risk, and such underestimation can lead to significant losses. Therefore, in this paper, CoES is used as a measure of systemic risk. In addition to having the lowest inherent risk, the banking industry has the highest systemic risk, indicated

by the inverse relationship between CoES and ES. More specifically, we find that the risk spillover effect from banking to securities is stronger than from insurance to securities. Similarly, the risk spillover effect from banking to insurance is stronger than from securities to insurance. Therefore, in the first estimation interval the banking industry plays a major role in the financial industry linkages, which is consistent with the constructed vine copula results. The CoES values of the banking industry to the other two industries are greater than 0.05, and the ΔCoES values are greater than 0.04 at the significance level of 2.5%, which is a significant contribution to inter-industry systemic risk. On the other hand, the risk spillover measures from other industries to the banking industry are relatively smaller, with CoES ranging from 0.03 to 0.05 at a significance level of 2.5%. Compared to the risk spillover from the securities industry to the banking industry, the risk spillover from the insurance industry to the banking industry is greater, which is consistent with the higher dependency between the insurance and banking industries mentioned earlier.

4.3.2. The CoES between the Financial System and the Financial Industries

We perform Monte Carlo simulations to obtain the financial system returns. After obtaining the simulated returns for the system, we rank the returns to calculate the VaR and ES of the financial system. Then, using the VaR of each industry, the systemic risk CoVaR and CoES of each financial industry are calculated by Equations (8) and (9).

In Table 7, we present the first rolling estimate of the financial industry's risk spillover to the financial system in terms of the CoES measure at significance levels of 2.5% and 1%. From Table 7, we can see that banking has the highest risk spillover effect on the financial system, followed by insurance and finally securities. At the 2.5% significance level, the banking industry's risk contribution ΔCoES to the financial system exceeds 0.04, while those of the insurance and securities industries are below 0.04. In addition, the insurance industry's risk contribution to the financial system is greater than the securities industry's risk contribution to the financial system. The findings align with the conclusions by Zhang et al. [14] and Cui [15].

Table 7. CoVaR, CoES, ΔCoES for financial industries and financial system.

Industry to System	c	CoVaR	CoES	ΔCoES
Banking to System	2.5%	−0.0423	−0.0554	−0.0442
	1%	−0.0531	−0.0699	−0.0588
Securities to System	2.5%	−0.0318	−0.0405	−0.0303
	1%	−0.0461	−0.0566	−0.0464
Insurance to System	2.5%	−0.0348	−0.0479	−0.0396
	1%	−0.0472	−0.0639	−0.0556

4.4. Dynamic CoES

4.4.1. The Dynamic CoES among Industries

In this section, we combine the vine copula grouped model and the rolling Monte Carlo method to calculate the out-of-sample ES of the financial industries and the CoES between industries over a period of 160 days, which can be displayed visually in Figures 3–5.

In a two-by-two comparison of ES and two-way CoES across industries, it can be observed that the fluctuation trends of ES and CoES are consistent for each financial industry. However, there are large differences in the magnitude of the fluctuations. In terms of individual industry risk during the forecast period, the banking industry has the lowest value of risk, while the securities and insurance industries have similar levels of risk. From the perspective of the fluctuation of risk, the banking industry has the smallest fluctuation in its own risk, followed by the securities industry, and the insurance industry has the largest fluctuation in risk, which is in accordance with the order of the standard deviation of the three industries' logarithmic returns. Most of the time, the systemic risk of all industries exceeds their individual risk values. In other words, the risk exposure of the

financial system as a whole is greater than the financial industry's own risk. This situation can be attributed to the high interdependence within the financial system, especially considering that the selected three financial industries are highly integrated within the Chinese financial system.

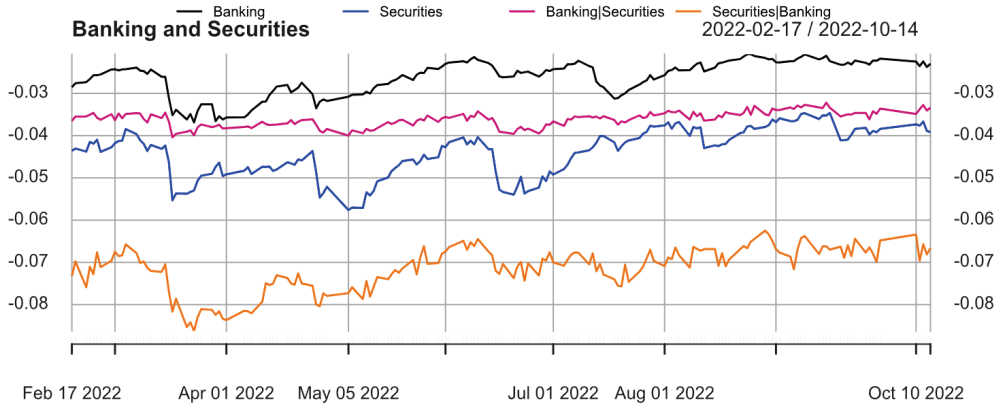

Figure 3. The CoES between banking and securities.

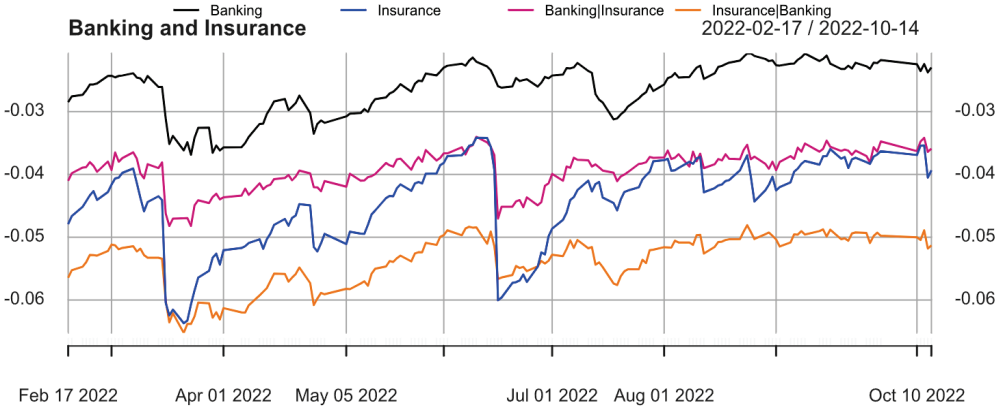

Figure 4. The CoES between banking and insurance.

From the perspective of risk spillover between industries, the banking industry is the least affected by other industries. Moreover, the risk spillover effects on the banking industry are smaller than the risks faced by financially distressed industries, and the magnitude of the fluctuation in the risk spillover effect on the banking industry is smaller than that of the financially distressed industry. This indicates that the risk spillover effects of other industries in financial distress do not affect the banking industry to a greater extent. The reason for the above situation is that the size of banks occupies a dominant position in the Chinese financial system. In general, the risk spillover effect of the insurance industry to the securities industry is greater than the risk spillover effect of the securities industry to the insurance industry.

This suggests that the risk spillover effects generated by financially distressed industries do not have a significant impact on the banking industry. We can observe that the fluctuations in risk spillover effects from the banking industry to other industries are highly similar to the fluctuations in their own risks. Therefore, once the banking industry experiences a crisis, it can easily influence other industries. The underlying reason for this phenomenon might be the banking industry's overwhelming dominance in scale within the

Chinese financial system. The difference in risk spillover volatility between the securities industry and the insurance industry is relatively small. However, in general, the securities industry has a greater risk spillover effect on the insurance industry.

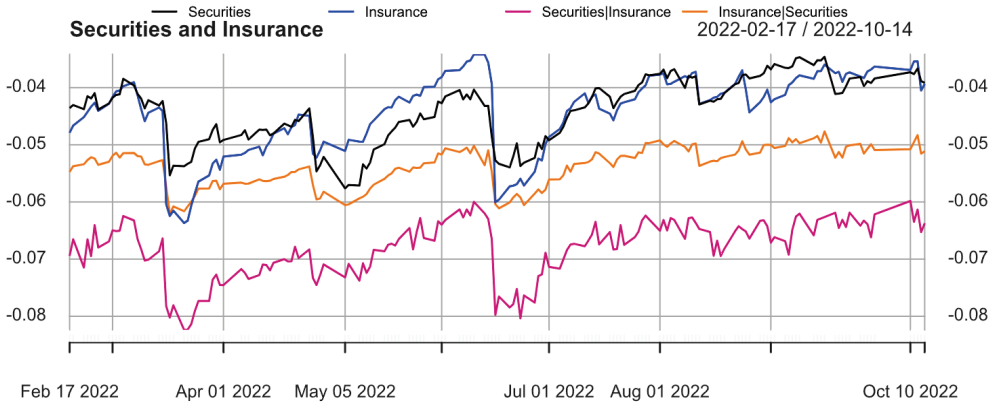

Figure 5. The CoES between securities and insurance.

4.4.2. The Dynamic CoES between Industries and System

In Figure 6, we present the dynamic CoES of each financial industry to the financial system throughout the forecast interval. Consistent with the low ES and high CoES characteristics of the banking industry in the previous section, this indicates that the banking industry demonstrates greater resilience to risk autonomously. However, once the banking industry is in financial distress, it could generate a large risk spillover to the whole financial system. We observe that the banking industry demonstrates the largest risk spillover effect on the whole financial system in comparison to other financial industries. This observation is likely related to the strong development and dominance of the banking industry in the Chinese financial market. Furthermore, the risk spillover effects of the three financial industries on the financial system have the same trend, while the banking industry has the largest change in risk spillover effects on the financial system, followed by the insurance industry and finally the securities industry. Despite exhibiting a high level of their own risk and weaker resistance to risk, we observe that the securities industry contributes less to the systemic risk of the whole financial system when it experiences a crisis itself.

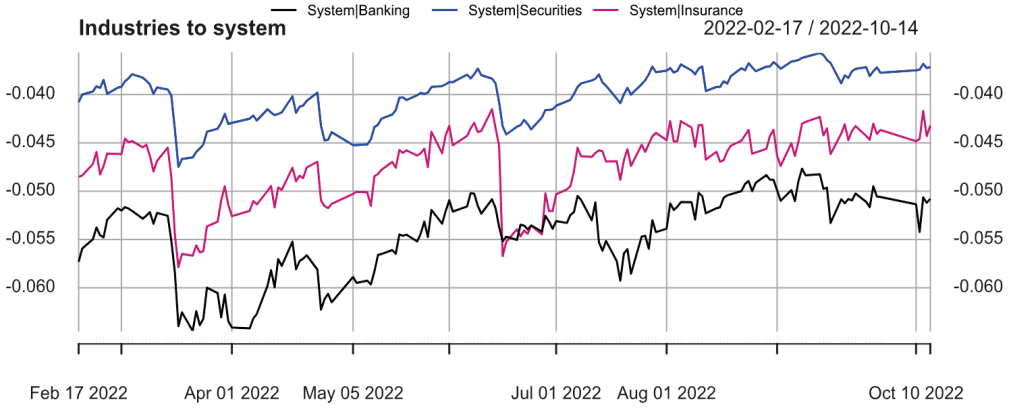

Figure 6. The CoES between industries and system.

4.5. A Comparison between Dynamic CoVaR and Dynamic CoES

In this section, we compare out-of-sample dynamic CoVaR and dynamic CoES. The specific arrangement involves comparing dynamic CoVaR and dynamic CoES across industries first, followed by comparing the impact of different industries on the financial system's dynamic CoVaR and dynamic CoES.

4.5.1. The CoVaR and CoES between Industries

In Figures 7–9, we present the comparison between the dynamic CoES and dynamic CoVaR among financial industries at a significance level of 2.5%. It is evident from the graphs that the measured values of CoES which account for the presence of extreme risks are significantly higher than CoVaR.

In terms of the magnitude of changes in the indicators, CoES exhibits a larger variation compared to CoVaR. By comparing CoVaR and through the comparison of CoVaR and CoES, it is evident that CoVaR has the flaw of underestimating risk spillover effects. Underestimating these effects may lead to erroneous risk management decisions, whereby risks are further amplified through inter-industry dependencies, ultimately causing the entire financial system to fall into crisis and affecting the stability and development of the country and society.

Specifically, the difference between CoVaR and CoES for the spillover effects from the securities industry to the banking industry is relatively small, while the difference is slightly larger for the spillover effects from the insurance industry to the banking industry. This is because the insurance industry can experience extreme losses due to natural disasters, which also have a significant impact on the banking industry. Additionally, for the banking industry to be less affected by spillover risks, the magnitude of variation in CoES is comparable to CoVaR, once again indicating the strong resilience of the banking industry to extreme risk spillovers. Regarding spillover risks between the insurance and securities industries, the magnitudes of CoVaR are similar, but there is a significant difference in the magnitudes of CoES. The insurance industry's CoES for spillover to the securities industry is noticeably larger, primarily due to the insurance industry's susceptibility to natural disasters. CoVaR fails to account for this extreme risk spillover phenomenon, resulting in comparable measurements and underestimation of spillover effects.

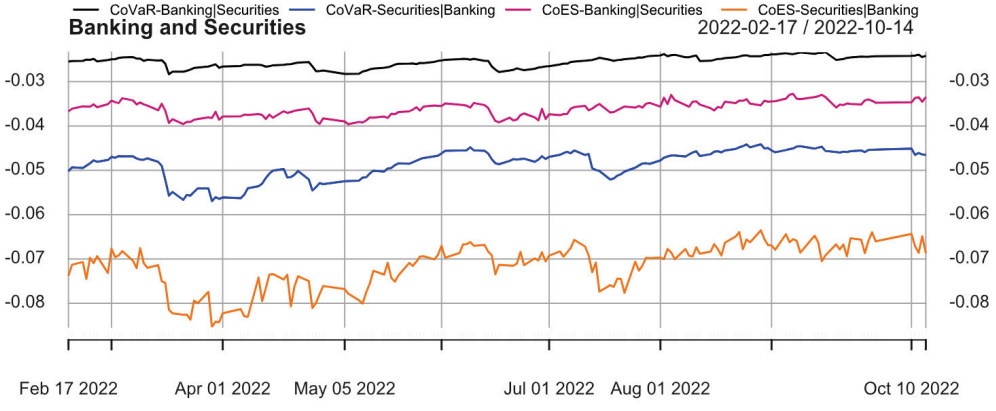

Figure 7. The CoVaR and CoES between banking and securities.

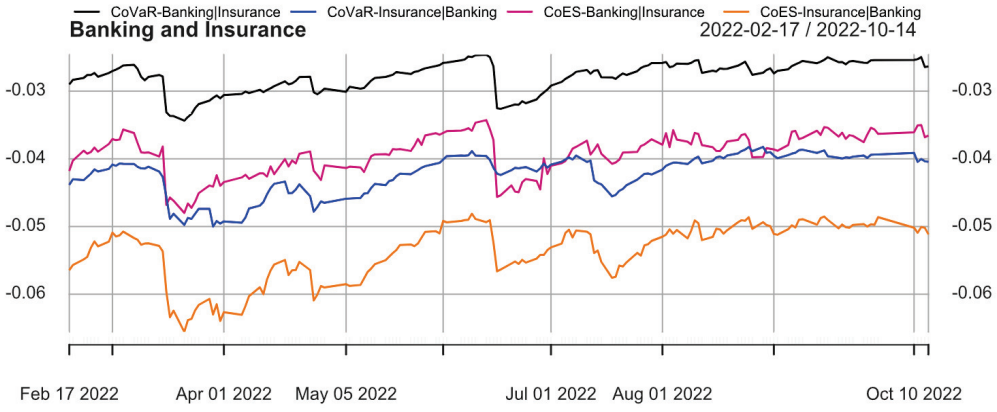

Figure 8. The CoVaR and CoES between banking and insurance.

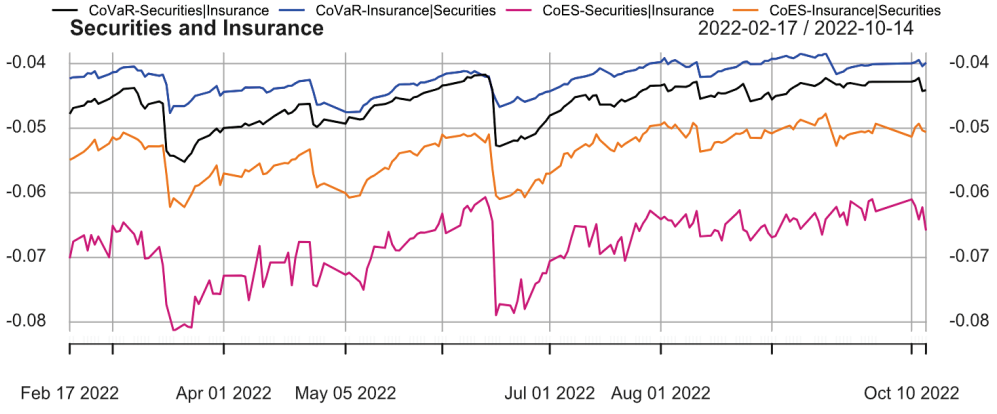

Figure 9. The CoVaR and CoES between securities and insurance.

4.5.2. The CoVaR and CoES between Industries and System

In Figure 10, this paper presents a comparison of CoVaR and CoES for risk spillover effects of different financial industries on the financial system. The figure clearly shows that the CoES values of different financial industries on the financial system are greater than their CoVaR values. Overall, due to the presence of extreme losses, the fluctuations in CoES measurements are slightly larger than those in CoVaR, but they are still relatively robust and can be used to measure financial risk spillover effects.

In terms of the differences in indicator values, the banking industry exhibits the greatest disparity between CoVaR and CoES values for risk spillover to the financial system, followed by the insurance industry, and finally the securities industry. This is because the banking industry plays a dominant role in the entire Chinese financial market, and the extreme loss situation in the banking industry can have a significant impact on overall financial risk. However, CoVaR overlooks these extreme losses, leading to an underestimation of financial risk spillover effects. The insurance industry, due to its business characteristics and its role in providing insurance to other financial industries to share risks, leads to increased risk spillover to other industries. If the insurance industry experiences extreme losses, it can affect the risk transfer to other industries to a certain extent, thereby impacting the entire financial system and causing disruptions.

Overall, CoVaR underestimates risk spillover effects due to its neglect of extreme risks. However, the existence of extreme risks can potentially plunge the entire financial system

into crisis. Thus, underestimating financial risk spillover effects can affect the accuracy and effectiveness of risk management decisions.

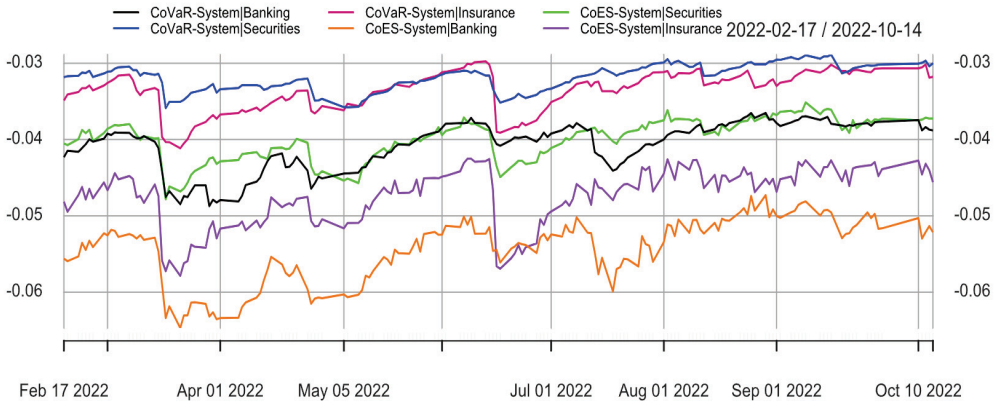

Figure 10. The CoVaR and CoES for financial industries on the financial system.

5. Conclusions

With the rapid development of the Chinese financial markets, there has been an increasing level of dependency among financial industries within the financial system. In this paper, using Monte Carlo simulation based on the dependency structure of the financial system described by the vine copula grouped model, we calculate the VaR, ES, CoVaR and CoES of the Chinese financial markets. We draw the following conclusions: First, judging from the risks faced by the financial industry itself, the banking industry exhibits the lowest level of risk while the securities industry and the insurance industry face nearly equal levels of risk. However, most of the time, the insurance industry faces greater risk. Second, the banking industry makes the greatest contribution to systemic risk, signifying the most significant risk spillover effect on other financial industries and the whole financial system, and thus makes itself the primary systemic risk driver. Third, the insurance industry's risk fluctuation is the largest of the three industries which is consistent with the insurance industry having the largest standard deviation of log returns.

Author Contributions: Formal analysis, H.D.; writing—original draft, H.D.; writing—review and editing, J.Y.; supervision, L.W. All authors have read and agreed to the published version of the manuscript.

Funding: Supported by the National Natural Science Foundation of China (No:12001411) and the Fundamental Research Funds for the Central Universities of China (No:2021IVB024).

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found at: https://data.csmar.com.

Acknowledgments: The authors are very grateful to the Editor and the anonymous referees for their valuable comments and suggestions which led to the present greatly improved version of the manuscript.

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

References

1. Artzner, P.; Delbaen, F.; Eber, J.; Heath, D. Coherent Measures of Risk. *Math. Financ.* **1999**, *9*, 203–228. [CrossRef]
2. Föllmer, H.; Schied, A. Convex measures of risk and trading constraints. *Financ. Stoch.* **2002**, *6*, 429–447. [CrossRef]
3. Frittelli, M.; Rosazza, G.E. Putting order in risk measures. *J. Bank. Financ.* **2002**, *26*, 1473–1486. [CrossRef]
4. Föllmer, H.; Schied, A. *Stochastic Finance: An Introduction in Discrete Time*; Walter de Gruyter: Berlin, Germany, 2011.
5. Burgert, C.; Rüschendorf, L. Consistent risk measures for portfolio vectors. *Insur. Math. Econ.* **2006**, *38*, 289–297. [CrossRef]
6. Wei, L.; Hu, Y. Coherent and convex risk measures for portfolios with applications. *Stat. Probab. Lett.* **2014**, *90*, 114–120.

7. Rüschendorf, L. *Mathematical Risk Analysis*; Springer: Berlin/Heidelberg, Germany, 2013.
8. Chen, C.; Iyengar, G.; Moallemi, C.C. An axiomatic approach to systemic risk. *Manag. Sci.* **2013**, *59*, 1373–1388. [CrossRef]
9. Kromer, E.; Overbeck, L.; Zilch, K. Systemic risk measures on general measurable spaces. *Math. Methods Oper. Res.* **2016**, *84*, 323–357. [CrossRef]
10. Adrian, T.; Brunnermeier, M.K. CoVaR. *Am. Econ. Rev.* **2016**, *106*, 1705–1741. [CrossRef]
11. Bai, X.; Shi, D. Measurement of the systemic risk of China's financial system. *Stud. Int. Financ.* **2014**, *326*, 75–85. (In Chinese)
12. Zhu, B.; Zhou, X.; Liu, X.; Wang, H.; He, K.; Wang, P. Exploring the risk spillover effects among China's pilot carbon markets: A regular vine copula-CoES approach. *J. Clean. Prod.* **2020**, *242*, 118455. [CrossRef]
13. Girardi, G.; Ergun, A.T. Systemic risk measurement: Multivariate GARCH estimation of CoVaR. *J. Bank. Financ.* **2013**, *37*, 3169–3180. [CrossRef]
14. Zhang, B.; Wang, S.; Wei, Y.; Zhao, X. A measure of financial systemic risk in China based on the CoES model. *Syst.-Eng.-Theory Pract.* **2018**, *38*, 565–575. (In Chinese)
15. Cui, J. Measurement of systematic financial risk based on CoES model. *Stat. Decis.* **2019**, *2019*, 148–151. (In Chinese)
16. Sklar, C. Fonctions de répartition à n dimensions et leurs marges. *Publ. Del'Institut Stat. L'Universit'E Paris* **1959**, *8*, 229–231.
17. Yang, L.; Yang, L.; Ho, K.; Hamori, S. Dependence structures and risk spillover in China's credit bond market: A copula and CoVaR approach. *J. Asian Econ.* **2020**, *68*, 101200. [CrossRef]
18. Li, Q. Study on the risk of spillover effect between capital market, pillar industry and exchange rate based on Copula-CoES model. *J. Ind. Eng. Eng. Manag.* **2021**, *35*, 12–26. (In Chinese)
19. Li, Q.; He, F.; He, J.; Dong, Y.; Wang, Y. A study on risk spillover effects between offshore and onshore RMB interest rates based on MSGARCH-Mixture Copula Model. *J. Stat. Inf.* **2022**, *37*, 75–86. (In Chinese)
20. Bedford, T.; Cooke, R.M. Probability Density Decomposition for Conditionally Dependent Random Variables Modeled by Vines. *Ann. Math. Artif. Intell.* **2001**, *32*, 245–268. [CrossRef]
21. Lin, Y. Stylized facts, extreme value theory and a study of dynamic risk measurement in financial markets. *Rev. Invest. Stud.* **2012**, *31*, 41–56. (In Chinese)
22. Shahzad, S.J.; Arreola-Hernandez, J.; Bekiros, S.; Shahbaz, M.; Kayani, G.M. A systemic risk analysis of Islamic equity markets using vine copula and delta CoVaR modeling. *J. Int. Financ. Mark. Inst. Money* **2018**, *56*, 104–127. [CrossRef]
23. Zhang, X.; Zhang, T.; Lee, C.-C. The path of financial risk spillover in the stock market based on the R-vine-copula model. *Phys. Stat. Mech. Its Appl.* **2022**, *600*, 127470. [CrossRef]
24. Zhu, N.; Wang, X. A measure of systemic risk in Chinese financial markets—A CoVaR model based on quantile regression. *Shanghai Financ.* **2017**, *2017*, 50–55. (In Chinese)
25. Zhou, Q.; Chen, Z.; Ming, R. Copula-based grouped risk aggregation under mixed operation. *Appl. Math.* **2016**, *61*, 103–120. [CrossRef]
26. Chen, Z.; Hao, X. A study on risk measurement of financial market based on vine copula grouped model. *Stat. Res.* **2018**, *35*, 77–84. (In Chinese)
27. Chen, Z.; Hao, X. Optimizing the risk of stock market based on vine copula grouped model. *J. Bus. Econ.* **2018**, *2018*, 89–97. (In Chinese)
28. Hao, X.; Chen, Z. Systemic risk in Chinese financial industries: A vine copula grouped CoVaR approach. *Econ.-Res.-Ekon. IstražIvanja* **2021**, *35*, 2747–2763. [CrossRef]
29. Glosten, L.R.; Jagannathan, R.; Runkle, D.E. On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *J. Financ.* **1993**, *48*, 1779–1801. [CrossRef]
30. Lamoureux, C.G.; Lastrapes, W.D. Forecasting Stock-Return Variance: Toward an Understanding of Stochastic Implied Volatilities. *Rev. Financ. Stud.* **1993**, *6*, 293–326. [CrossRef]
31. Acerbi, C.; Tasche, D. Expected shortfall: A natural coherent alternative to value at risk. *Econ. Notes* **2002**, *31*, 379–388. [CrossRef]

mathematics

MDPI

Article

Transient Analysis for a Queuing System with Impatient Customers and Its Applications to the Pricing Strategy of a Video Website

Qihui Bu

School of Science, Nanjing University of Posts and Telecommunications, Nanjing 210023, China; buqihui@njupt.edu.cn

Abstract: In this paper, we consider a queuing system with impatient customers, which includes infinite servers and two types of customers. During the service process, Type-1 customers may leave the system or upgrade to be Type-2 customers due to their impatience. By solving the partial differential equations, we obtain the generating functions of the transient distribution of the queue length, and many stationary performance measures are further derived. Then, as an application, we formulate an expected profit function for a video website, and maximize it by determining the optimal pricing strategy. Finally, numerical examples are provided to demonstrate the impacts of parameters on the optimal website profit.

Keywords: queuing system; impatient customers; optimization; pricing strategy

MSC: 60K30; 68M20; 90B22

1. Introduction

Queuing systems with impatient customers have been studied extensively, and have applications in a wide range of areas, such as call centers. In such systems, customers may become impatient and leave the system during their waiting time or service processes. Barrer [1,2] firstly considered customers' psychology in queuing systems. He analyzed $M/M/1$ and $M/M/c$ queuing systems with constant impatience time, and obtained the stationary distributions of the queue length, respectively. Since then, more and more scholars have focused on the queuing systems with impatient customers. Phung-Duc [3] studied an Markovian multi-server retrial queue, in which a blocked customer has two opportunities for abandonment. The tail asymptotics formulae for the joint stationary distribution of the number of customers in the system and for those in the orbit are obtained. Kim and Kim [4] studied a single queuing system in which customers wait for service for a fixed time, and if the time is expired, customers leave the system instantly. By constructing the age process, they derived the stationary distribution of the queue length, the loss probability and the waiting time distribution. Adan et al. [5] investigated a queuing system with two classes of impatient customers, and obtained system performance measures by introducing a virtual waiting time process instead of a queue length process. These systems with impatient customers are exemplified by studies like [6–9] and others.

With the development of technology, many stochastic service systems in real life operates with large service capacity so that all customers in the system can access to service immediately upon their arrival. Such service systems are regarded as the queuing systems with an infinite number of servers and are investigated in many studies. Jackson and Aspden [10] studied a finite number of $M/M/\infty$ queues in series and proposed a novel method to derive the time dependent solution of a multistage nonqueuing process. Sato et al. [11] constructed an $M/D/\infty$ queuing model to characterize a direct-sequence spread spectrum multiple access unslotted ALOHA with fixed packet length. Shi et al. [12]

Citation: Bu, Q. Transient Analysis for a Queuing System with Impatient Customers and Its Applications to the Pricing Strategy of a Video Website. *Mathematics* **2024**, *12*, 2030. https://doi.org/10.3390/math12132030

Academic Editors: Jing Yao, Xiang Hu and Jingchao Li

Received: 31 May 2024
Revised: 21 June 2024
Accepted: 25 June 2024
Published: 29 June 2024

took the online service as a background, and considered a queuing system with free experience service. The closed form of the expected numbers of informed and uninformed customers in steady-state are derived by solving nonhomogeneous linear partial differential equations. Hassin and Ravner [13] studied the probabilistic properties of a queuing system with infinite servers by considering the overflows from subsystems with finite servers.

Recently, scholars prefer to study from an economic perspective rather than classical queuing system performance analyses. Game theory and pricing theory are two main problems in economic queuing studies. Burnetas and Economou [14] discussed customers' equilibrium strategies in an Markovian queue model with four different levels of system information. Economou and Kanta [15] studied the equilibrium balking strategies in an observable $M/M/1$ queuing system with breakdowns and repairs. Recently, more studies about the equilibrium strategies in queuing models can be found in [16–20]. Additionally, the pricing problem is an important research subject in economic queuing. Lee and Ward [21] considered how to jointly set the static price and capacity to maximize the steady state mean profit in a $GI/GI/1$ queue with a high rate of prospective customer arrivals. Bai et al. [22] proposed a queuing model with server sharing and determine an optimal admission policy to maximize managers' profit function.

However, few papers focus on the pricing problem of two-sided markets based on the queuing theory. Rochet and Tirole [23], Armstrong [24] studied the pricing strategy for a two-sided market from the point of externality. Most studies are developed and explored based on these two papers. Zeithammer and Thomadsen [25] analyzed price and quality competition in a vertically differentiated duopoly in which consumers have a preference for variety. See references [26–29] for more knowledge about two-sided markets. In this paper, we take the video website as an application to analyze the profit function of website and maximize it by determining optimal pricing strategy.

The rest of this paper is organized as follows. The transient and steady state probability generating function of the queue length and several stationary performance measures are derived in Section 2. Then, we study the optimal pricing strategy of a website profit function in Section 3, and discuss the impacts of the customers' potential value and advertisement negative effect on the optimal pricing strategy. Finally, numerical examples are provided to illustrate the impacts of parameters, such as advertising time and membership reward gaps, on the optimal website profit in Section 4.

2. Transient and Stationary State Analysis

We first describe the basic setting of our queuing system, the working mechanism of which is presented in Figure 1.

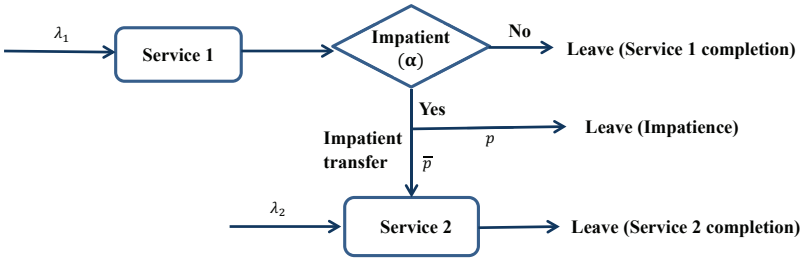

Figure 1. Working mechanism of this system.

- **Arrival:** There are two types of customers, Type-1 and Type-2, and they arrive the system according to the Possion processes with rate λ_1 and λ_2, respectively.
- **Service:** Infinite severs are provided for each type customers, and the service time of Type-1 (Type-2) customers is assumed to be an exponential random variable with mean value $1/\mu_1$ ($1/\mu_2$).

- **Impatience:** A Type-1 customer may become impatient during their service process, and the time between two successive impatience events are supposed to follow the exponential distribution with mean value $1/\alpha$. The impatient customer leaves the system directly with probability p, or otherwise upgrades to be a Type-2 customer.

From the description of the model, the state of the system at time t can be described by the stochastic process $\{(N_1(t), N_2(t)), t \geq 0\}$, where $N_k(t)$ is the number of Type-k customers in the system at time t, $k = 1, 2$. The corresponding state space is given by

$$\Omega = \{0, 1, 2 \cdots\} \times \{0, 1, 2 \cdots\}.$$

Define the following joint probabilities that there are i Type-1 customers and j Type-2 customers in the system at time t,

$$P_{i,j}(t) = P\{N_1(t) = i, N_2(t) = j\}, \tag{1}$$

where $i, j \in \{0, 1, 2 \cdots\}$.

According to the forward and backward Kolmogorov differential equations, we have:

1. $N_2(t) = 0$,

$$\begin{aligned} P'_{0,0}(t) &= -(\lambda_1 + \lambda_2)P_{0,0}(t) + (\mu_1 + \alpha p)P_{1,0}(t) + \mu_2 P_{0,1}(t), & (2)\\ P'_{i,0}(t) &= -(\lambda_1 + \lambda_2 + i\mu_1 + i\alpha)P_{i,0}(t) + \lambda_1 P_{i-1,0}(t) \\ &\quad +(i+1)(\mu_1 + \alpha p)P_{i+1,0}(t) + \mu_2 P_{i,1}(t), & (3) \end{aligned}$$

where $i = 1, 2, \cdots$.

2. $N_2(t) = j, j = 1, 2, \cdots$,

$$\begin{aligned} P'_{0,j}(t) &= -(j\mu_2 + \lambda_1 + \lambda_2)P_{0,j}(t) + \lambda_2 P_{0,j-1}(t) + \alpha\overline{p}P_{1,j-1}(t) \\ &\quad +(\mu_1 + \alpha p)P_{1,j}(t) + (j+1)\mu_2 P_{0,j+1}(t), & (4)\\ P'_{i,j}(t) &= -(j\mu_2 + \lambda_1 + \lambda_2 + i\mu_1 + i\alpha)P_{i,j}(t) + \lambda_2 P_{i,j-1}(t) + (i+1)\alpha\overline{p}P_{i+1,j-1}(t) \\ &\quad +\lambda_1 P_{i-1,j}(t) + (i+1)(\mu_1 + \alpha p)P_{i+1,j}(t) + (j+1)\mu_2 P_{i,j+1}(t), & (5) \end{aligned}$$

where $i = 1, 2, \cdots$.

We define the partial generating functions as follows:

$$\begin{aligned} g_j(x,t) &= \sum_{i=0}^{\infty} P_{i,j}(t)x^i, \quad j = 0, 1, 2, \cdots, & (6)\\ G(x,y,t) &= \sum_{j=0}^{\infty} g_j(x,t)y^j. & (7) \end{aligned}$$

Multiplying (3) by x^i, summing over all possible values of i, and combining with (2), we obtain the following result:

$$\frac{\partial g_0(x,t)}{\partial t} = -(\lambda_1 + \lambda_2 - \lambda_1 x)g_0(x,t) + (\mu_1 + \alpha p - \alpha x - \mu_1 x)\frac{\partial g_0(x,t)}{\partial x} + \mu_2 g_1(x,t). \tag{8}$$

Multiplying (5) by x^i, summing over all possible values of i, and combining with (4), we arrive at

$$\begin{aligned} \frac{\partial g_j(x,t)}{\partial t} &= -(j\mu_2 + \lambda_1 + \lambda_2 - \lambda_1 x)g_j(x,t) + \lambda_2 g_{j-1}(x,t) + (j+1)\mu_2 g_{j+1}(x,t) \\ &\quad +\alpha\overline{p}\frac{\partial g_{j-1}(x,t)}{\partial x} + (\mu_1 + \alpha p - \alpha x - \mu_1 x)\frac{\partial g_j(x,t)}{\partial x}, & (9) \end{aligned}$$

where $j = 1, 2, \cdots$.

Multiplying (9) by y^j and summing over all j, together with (8), we have

$$\begin{aligned}(\lambda_1+\lambda_2-\lambda_1 x-\lambda_2 y)G(x,y,t) &= -\frac{\partial G(x,y,t)}{\partial t}+\mu_2(1-y)\frac{\partial G(x,y,t)}{\partial y} \\ &\quad +(\mu_1+\alpha p-\alpha x-\mu_1 x+\alpha\overline{p}y)\frac{\partial G(x,y,t)}{\partial x}.\end{aligned} \tag{10}$$

Equation (10) is a first-order linear partial differential equation. Its characteristic equations are given by

$$\frac{dt}{-1}=\frac{dx}{\mu_1+\alpha p-\alpha x-\mu_1 x+\alpha\overline{p}y}=\frac{dy}{\mu_2(1-y)}=\frac{dG(x,y,t)}{(\lambda_1+\lambda_2-\lambda_1 x-\lambda_2 y)G(x,y,t)} \tag{11}$$

Deriving from the first three items in (11), we have

$$\xi = (1-y)e^{-\mu_2 t}, \tag{12}$$

$$\eta = \left[x-1-\frac{\alpha\overline{p}}{\mu_2-\mu_1-\alpha}(1-y)\right]e^{-(\mu_1+\alpha)t}. \tag{13}$$

Putting (12) and (13) into (10), we have the following equation related to ξ and η:

$$\frac{\partial G(\xi,\eta,t)}{\partial t}=\left[\lambda_1\left(\eta e^{(\mu_1+\alpha)t}+\frac{\alpha\overline{p}}{\mu_2-\mu_1-\alpha}\xi e^{\mu_2 t}\right)-\lambda_2\xi e^{\mu_2 t}\right]G(\xi,\eta,t). \tag{14}$$

Integrating the above equation leads to

$$G(\xi,\eta,t)=C(\xi,\eta)exp\left\{\left(\frac{\alpha\overline{p}\lambda_1}{\mu_2(\mu_2-\mu_1-\alpha)}-\frac{\lambda_2}{\mu_2}\right)\xi e^{\mu_2 t}+\frac{\lambda_1}{\mu_1+\alpha}\eta e^{(\mu_1+\alpha)t}\right\}, \tag{15}$$

where C is an undefined binary function and $C(\xi,\eta)$ is the function value when ξ and η are taken as the independent variables.

According to the boundary condition $G(x,y,0)=x^k y^b$, which means that there are k customers of Type-1 and b customers of Type-2 in the system at the beginning, we obtain

$$C(\vartheta,\gamma)\exp\left\{\left(-\frac{\lambda_2}{\mu_2}+\frac{\lambda_1\alpha\overline{p}}{\mu_2(\mu_2-\mu_1-\alpha)}\right)\vartheta+\frac{\lambda_1}{\mu_1+\alpha}\gamma\right\}=x^k y^b, \tag{16}$$

where ϑ and γ are the values when $t=0$ in (12) and (13), i.e.,

$$\vartheta = 1-y, \tag{17}$$

$$\gamma = x-1-\frac{\alpha\overline{p}}{\mu_2-\mu_1-\alpha}(1-y). \tag{18}$$

Then, we have the formula for the function C,

$$\begin{aligned}C(\vartheta,\gamma) &= (1-\vartheta)^b(\gamma+1+\frac{\alpha\overline{p}}{\mu_2-\mu_1+\alpha}\vartheta)^k \\ &\quad \times\exp\left\{\left(\frac{\lambda_2}{\mu_2}-\frac{\lambda_1\alpha\overline{p}}{\mu_2(\mu_2-\mu_1+\alpha)}\right)\vartheta-\frac{\lambda_1}{\mu_1+\alpha}\gamma\right\},\end{aligned} \tag{19}$$

and

$$\begin{aligned}C(\xi,\eta) &= (1-\xi)^b(\eta+1+\frac{\alpha\overline{p}}{\mu_2-\mu_1+\alpha}\xi)^k \\ &\quad \times\exp\left\{\left(\frac{\lambda_2}{\mu_2}-\frac{\lambda_1\alpha\overline{p}}{\mu_2(\mu_2-\mu_1+\alpha)}\right)\xi-\frac{\lambda_1}{\mu_1+\alpha}\eta\right\}.\end{aligned} \tag{20}$$

Thus,

$$G(x,y,t) = C(\xi,\eta)\exp\left\{\left(-\frac{\lambda_2}{\mu_2}-\frac{\alpha\overline{p}\lambda_1}{\mu_2(\mu_1+\alpha)}\right)(1-y)+\frac{\lambda_1}{\mu_1+\alpha}(x-1)\right\}. \quad (21)$$

Theorem 1. *The probability generating function of the two-dimensional Markov chain* $\{(N_1(t), N_2(t)), t \geq 0\}$*, in the transient state is*

$$G(x,y,t) = C(\xi,\eta)\exp\left\{\left(-\frac{\lambda_2}{\mu_2}-\frac{\alpha\overline{p}\lambda_1}{\mu_2(\mu_1+\alpha)}\right)(1-y)+\frac{\lambda_1}{\mu_1+\alpha}(x-1)\right\}, \quad (22)$$

where

$$\begin{aligned} C(\xi,\eta) &= (1-\xi)^b(\eta+1+\frac{\alpha\overline{p}}{\mu_2-\mu_1+\alpha}\xi)^k \\ &\quad\times\exp\left\{\left(\frac{\lambda_2}{\mu_2}-\frac{\lambda_1\alpha\overline{p}}{\mu_2(\mu_2-\mu_1+\alpha)}\right)\xi-\frac{\lambda_1}{\mu_1+\alpha}\eta\right\}. \end{aligned} \quad (23)$$

Furthermore, noting that there are always enough servers for arriving customers, and there is no congestion and waiting, this queuing system is always stable. Let $t \to \infty$ in (2)–(5). We can obtain steady-state equations and compute the probability generating function with the normalization condition, which is provided by the following theorems.

Theorem 2. *The probability generating function of the two-dimensional Markov chain* $\{(N_1(t), N_2(t)), t \geq 0\}$ *in the steady state is*

$$Q(x,y) = \exp\left\{-\left(\frac{\alpha\overline{p}\lambda_1}{\mu_2(\mu_1+\alpha)}+\frac{\lambda_2}{\mu_2}\right)(1-y)-\frac{\lambda_1}{\mu_1+\alpha}(1-x)\right\}. \quad (24)$$

Theorem 3. *The stationary distribution of this model is given as follows:*

$$\pi_{i,j} = \frac{a_i}{i!}\frac{b_j}{j!}\exp\left\{-\frac{\alpha\overline{p}\lambda_1}{\mu_2(\mu_1+\alpha)}-\frac{\lambda_2}{\mu_2}-\frac{\lambda_1}{\mu_1+\alpha}\right\}, \quad (25)$$

where $a_i = \left(\frac{\lambda_1}{\mu_1+\alpha}\right)^i, b_j = \left(\frac{\alpha\overline{p}\lambda_1}{\mu_2(\mu_1+\alpha)}+\frac{\lambda_2}{\mu_2}\right)^j$.

Corollary 1. *There are i Type-*1 *customers in the system with probability* P_{1i}*,*

$$P_{1i} = \frac{a_i}{i!}\exp\left\{-\frac{\lambda_1}{\mu_1+\alpha}\right\}.$$

*There are j Type-*2 *customers in the system with probability* P_{2j}*,*

$$P_{2j} = \frac{b_j}{j!}\exp\left\{-\frac{\alpha\overline{p}\lambda_1}{\mu_2(\mu_1+\alpha)}-\frac{\lambda_2}{\mu_2}\right\}.$$

There are no customers in the system with probability P_{00}*,*

$$P_{00} = \exp\left\{-\frac{\alpha\overline{p}\lambda_1}{\mu_2(\mu_1+\alpha)}-\frac{\lambda_2}{\mu_2}-\frac{\lambda_1}{\mu_1+\alpha}\right\}.$$

Corollary 2. *The mean number of Type-*1 *customers in the system is*

$$E_1 = \frac{\lambda_1}{\mu_1+\alpha}.$$

The mean number of Type-2 customers in the system is

$$E_2 = \frac{\alpha\overline{p}\lambda_1}{\mu_2(\mu_1+\alpha)} + \frac{\lambda_2}{\mu_2}.$$

Remark 1. *E_1 can be seen as the mean queue length of an $M/M/\infty$ queue with arrival rate λ_1 and service rate $\mu_1+\alpha$. E_2 can be explained by two terms: the first term is the average number of customers who are transferred from impatient Type-1 customers, and the second term means the mean queue length of an $M/M/\infty$ queue with arrival rate λ_2 and service rate μ_2.*

3. Application: The Optimal Pricing Strategy of a Video Website

For a video website, viewers can be divided into two types, free memberships and premium memberships, which correspond to Type-1 and Type-2 in the queuing system. The viewing time is supposed to be customers' service time. Seeing that free memberships need to watch advertisements in addition to programs, we assume $\mu_1 < \mu_2$. Annoyed by advertisements and other factors, free memberships may leave the system or upgrade to be premium memberships.

We formulate the profit function of a website based on the queuing system performances, and study the pricing strategy to optimize the profit function. Specifically, the website collects membership fees from premium memberships, who can watch the program directly, and charges advertisement fees from advertisers. Collections from membership fees are related to both the population of premium memberships and membership fees, so too high a membership fee with a small number of premium memberships or too low a membership fee with a large population may both lead to the deviation from the maximal website profit. For the price charged for advertisers, a high price may block the entry of advertisers, while a low price goes against the purpose of maximizing website profit. Thus, an appropriate pricing strategy should be designed to maximize the overall profit of the video website.

We first define some parameters as follows:

- P: the price charged from advertisers per unit time.
- s: the price charged from per premium member, i.e., the membership fee.
- R_k: the utility obtained by a Type-k customer when they leave the system with service completed.
- θ: the degree of each customer's satisfaction for the website content.
- δ: the cost of the negative effect on a Type-1 customer per unit advertising time.
- β: the potential value of each Type-1 customer for advertisers.
- c_k: the service cost per type-k customer.

When customers are going to enter the website, the expected utility functions of Type-1 and Type-2 customers are assumed to be

$$u_1 = \theta R_1 - \delta\left(\frac{1}{\mu_1} - \frac{1}{\mu_2}\right), \tag{26}$$

$$u_2 = \theta R_2 - s, \tag{27}$$

where $\frac{1}{\mu_1} - \frac{1}{\mu_2}$ is the average advertising time.

Additionally, the utility is assumed to be zero for balking customers. In order to describing the customers' comment heterogeneity, we assume that θ is a uniform random variable in $[0,1]$. In Figure 2, when the preference for website content is θ_1, customers are indifference to balking or visiting the website. Similarly, there is no difference between having a free or premium membership for a customer with the preference value θ_2 for the website content.

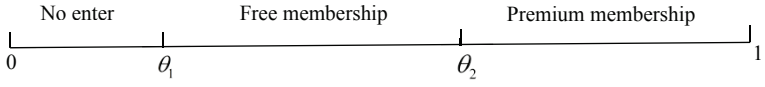

Figure 2. The degree of satisfaction.

Thus, we have the following equations for θ_1 and θ_2:

$$\theta_1 R_1 - \delta\left(\frac{1}{\mu_1} - \frac{1}{\mu_2}\right) = 0, \tag{28}$$

$$\theta_2 R_1 - \delta\left(\frac{1}{\mu_1} - \frac{1}{\mu_2}\right) = \theta_2 R_2 - s, \tag{29}$$

and solving the above equations yields

$$\theta_1 = \frac{\delta\left(\frac{1}{\mu_1} - \frac{1}{\mu_2}\right)}{R_1}, \tag{30}$$

$$\theta_2 = \frac{s - \delta\left(\frac{1}{\mu_1} - \frac{1}{\mu_2}\right)}{R_2 - R_1}. \tag{31}$$

Thus, a new visitor may balk with probability θ_1, visit as a free membership with probability $\theta_2 - \theta_1$ and obtain a premium membership with probability $1 - \theta_2$. That is to say, when the total potential visitors arrive following a Poisson process with rate λ, the Type-1 and Type-2 customers arrive following a Poisson process with rate $\lambda_1 = (\theta_2 - \theta_1)\lambda$, $\lambda_2 = (1 - \theta_2)\lambda$, respectively.

The utility function of advertisers is as follows:

$$u_{ad} = \beta E_1 - P\left(\frac{1}{\mu_1} - \frac{1}{\mu_2}\right). \tag{32}$$

Then, we have the profit function of this video website,

$$\begin{cases} \max\limits_{s,P} W = P + s\lambda_2 - c_1 E_1 - c_2 E_2, \\ \text{subject to} \\ \quad 0 \le \theta_1 \le \theta_2 \le 1, \\ \quad \beta E_1 - P\left(\frac{1}{\mu_1} - \frac{1}{\mu_2}\right) \ge 0. \end{cases} \tag{33}$$

Furthermore, the optimal pricing strategy can be obtained by solving problem (33).

Theorem 4. *For the profit function model* (33),

$$\begin{cases} \min\limits_{s,P} W = -P + As^2 - Bs - C, \\ \textit{subject to} \\ \quad P - k_1 s + b_1 \le 0, \\ \quad -(k_2 s + b_2) + \theta_1 \le 0, \\ \quad k_2 s + b_2 - 1 \le 0, \end{cases} \tag{34}$$

the optimal pricing strategy (s^*, P^*) *is given by*

$$s^* = \frac{B + k_1}{2A},$$

$$P^* = k_1 \frac{B + k_1}{2A} + b_1,$$

where

$$\begin{aligned}
A &= k_2\lambda,\\
B &= \lambda(1-b_2)-\frac{c_1k_2\lambda}{\mu_1+\alpha}-\frac{c_2\alpha\bar{p}k_2\lambda}{\mu_2(\mu_1+\alpha)}+\frac{c_2k_2\lambda}{\mu_2},\\
C &= \left(-\frac{\lambda c_1}{\mu_1+\alpha}-\frac{c_2\alpha\bar{p}\lambda}{\mu_2(\mu_1+\alpha)}\right)(b_2-\theta_1)-\frac{c_2\lambda}{\mu_2}(1-b_2),\\
k_1 &= \frac{\beta\lambda}{\mu_1+\alpha}\frac{\mu_1\mu_2}{\mu_2-\mu_1}k_2,\\
b_1 &= -\frac{\beta\lambda}{\mu_1+\alpha}\frac{\mu_1\mu_2}{\mu_2-\mu_1}b_2,\\
k_2 &= \frac{1}{R_2-R_1},\\
b_2 &= -\frac{\delta}{R_2-R_1}(\frac{1}{\mu_1}-\frac{1}{\mu_2}).
\end{aligned}$$

Proof. According to the Karush–Kuhn–Tucker condition, we have the following Lagrange function and necessary conditions for the optimal solution:

$$\begin{aligned}
&\mathcal{L}=(-P+As^2-Bs-C)+\gamma_1(P-k_1s+b_1)+\gamma_2(-(k_2s+b_2)+\theta_1)\\
&+\gamma_3(k_2s+b_2-1),\\
&\bigtriangledown_s\mathcal{L}=2As-B-\gamma_1k_1-\gamma_2k_2+\gamma_3k_2=0, &(35)\\
&\bigtriangledown_P\mathcal{L}=-1+\gamma_1=0, &(36)\\
&\gamma_1(P-k_1s+b_1)=0, &(37)\\
&\gamma_2(-(k_2s+b_2)+\theta_1)=0, &(38)\\
&\gamma_3(k_2s+b_2-1)=0. &(39)
\end{aligned}$$

It is easy to see that $\gamma_1=1$ and $P=k_1s-b_1$ from Equations (36) and (37). Then, we discussed the KKT point in the following situations:

(1) When $\gamma_2=0$ and $\gamma_3=0$, solving Equations (35)–(39) yields the following possible optimal solution:

$$\begin{aligned}
s_1 &= \frac{B+k_1}{2A},\\
P_1 &= k_1\frac{B+k_1}{2A}-b_1.
\end{aligned}$$

(2) When $\gamma_2=0$ and $k_2s+b_2-1=0$, combining with the KKT condition, we obtain an alternative optimal solution:

$$\begin{aligned}
s_2 &= \frac{1-b_2}{k_2},\\
P_2 &= k_1\frac{1-b_2}{k_2}-b_1.
\end{aligned}$$

(3) When $-(k_2s+b_2)+\theta_1=0$ and $\gamma_3=0$, another optional optimal solution is given by

$$\begin{aligned}
s_3 &= \frac{\theta_1-b_2}{k_2},\\
P_3 &= k_1\frac{\theta_1-b_2}{k_2}-b_1.
\end{aligned}$$

Next, we need to find the optimal solution from three candidates. According to condition (37), we have $P = k_1 s - b_1$, and substituting it into the profit function, we obtain

$$W = -As^2 + (B + k_1)s + C - b_1, \tag{40}$$

based on which it is obvious that the optimal solution for this profit function model is (s_1, P_1). □

Corollary 3.

$$\begin{aligned} \frac{\partial s^*}{\partial \delta} &= \frac{1}{2}\left(\frac{1}{\mu_1} - \frac{1}{\mu_2}\right) > 0, \\ \frac{\partial s^*}{\partial \beta} &= \frac{\lambda k_2}{2A} > 0. \end{aligned}$$

Remark 2. *It follows from Corollary 3 that when the negative effect on free memberships becomes larger, the membership fee increases. This implies that when customers become more annoyed with advertisements and more prone to skip them, the website will exploit this to charge a higher membership fee.*

When the potential value of a free membership for advertisers increases, the website tends to increase the membership fee. The reason is that the website can provide more free memberships for advertisers with a higher membership fee, and optimize its profit by charging a high price for advertisers.

Corollary 4.

$$\begin{aligned} \frac{\partial P^*}{\partial \delta} &= \frac{k_1}{2}\left(\frac{1}{\mu_1} - \frac{1}{\mu_2}\right) + \frac{\beta\lambda}{(R_2 - R_1)(\mu_1 + \alpha)} > 0, \\ \frac{\partial P^*}{\partial \beta} &= \frac{\lambda\mu_1\mu_2 k_2 s^*}{(\mu_1 + \alpha)(\mu_2 - \mu_1)} + \frac{k_1}{2} - \frac{\lambda\mu_1\mu_2 b_2}{(\mu_1 + \alpha)(\mu_2 - \mu_1)} > 0. \end{aligned}$$

Remark 3. *When the negative effect on free memberships becomes larger, the website will increase the cost of advertisers to protect memberships' utilities. With the increasing potential value of a free membership, the website will set a higher price of advertisement to maximize their own revenue.*

Corollary 5. *The optimal profit W^* is a quadratic function with respect to both δ and β, the symmetry axes of which are on the right and left side of the origin, respectively. The relationship of advertising negative effect (free memberships' potential value) and optimal website profit is shown in Figure 3 (Figure 4).*

Remark 4. *According to Corollary 5, it is easy to observe that the website profit is a convex function in terms of advertising's negative effect on free memberships, which means that when δ is small, as the advertising negative effect increases, the number of free memberships decreases, and then the website profit decreases. However, when δ reaches a certain value, more and more viewers may choose to be premium members with an increasing negative advertising effect, thus the website gains more from membership fees and the website profit increases. In addition, the increasing of free memberships' potential value leads to a higher price for advertisements, thus the website's profit increases.*

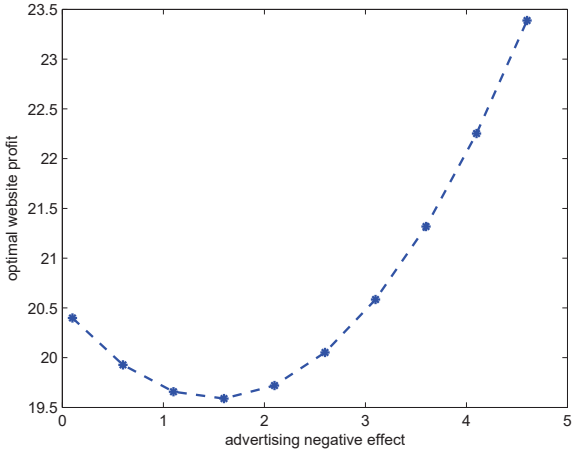

Figure 3. The optimal website profit versus advertising negative effect.

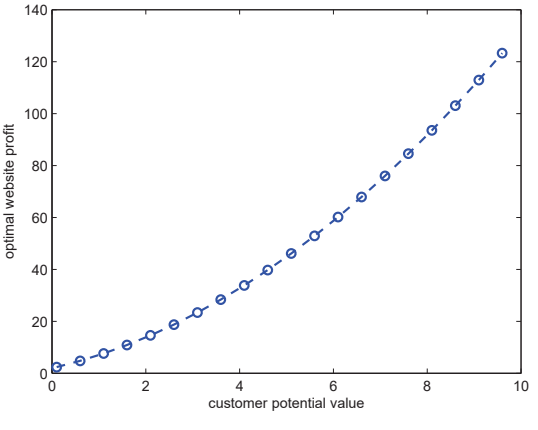

Figure 4. The optimal website profit versus free membership's potential value.

4. Numerical Examples

In this section, we will present the impacts of parameters on the optimal website profit using figures, and try to give some management insights.

Recall our basic setting, where $1/\mu_2$ is the average program time, and $1/\mu_1$ is the average program and advertising time; thus, we suppose that $1/\mu_1 - 1/\mu_2$ is the average advertising time. In Figure 5, when the advertising time is short, the optimal website profit declines quickly, as the advertising time becomes longer. We find that a sharp decline in the number of free memberships, resulting from the increasing advertising time, leads to the decrease in advertising fees and website profits. When the profit hits the bottom, it rebounds. As the advertising time increases, the number of premium memberships grows gradually, which results in an increase in website profit.

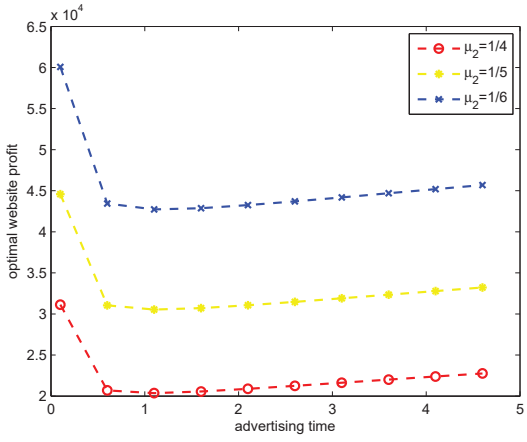

Figure 5. The optimal website profit versus advertising time for $\lambda = 50$, $\alpha = 0.6$, $p = 0.5$, $\delta = 0.2$, $\beta = 1$, $c_1 = 4$, $c_2 = 6$, $R_1 = 1$, $R_2 = 2$.

Figure 6 shows the impact of the reward gap between free and premium memberships on the website's profit. It is clear that what mainly affects the optimal profit is the difference between the two rewards instead of the specific value of the two rewards. When there is little difference between the two rewards, viewers prefer free memberships, and advertising revenue is a major part of website profit. As the gap increases, the number of free memberships declines rapidly and the advertising revenue falls, as well as the website profit. When the reward gap reaches a certain value, it is worthy for viewers to upgrade to be premium memberships at some appropriate membership fee. Thus, the membership fee becomes the major part of website profit, which rebounds and grows with the increasing reward gap.

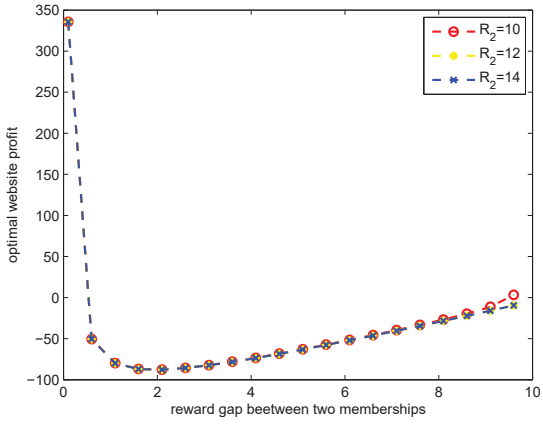

Figure 6. The optimal website profit versus reward gap for $\lambda = 50$, $\mu_1 = 1$, $\mu_2 = 5$, $\alpha = 0.6$, $p = 0.5$, $\delta = 0.2$, $\beta = 1$, $c_1 = 1$, $c_2 = 4$.

The impact of the impatience rate on the optimal website profit is provided by Figure 7. With the impatience rate increasing, the number of free memberships decreases, which leads to a fall in the price for advertisers and a slight growth in the number of premium memberships. In addition, from Figure 7, for a fixed impatience rate, a larger arrival rate λ corresponds to a larger optimal website profit, which implies more potential viewers and means more profit.

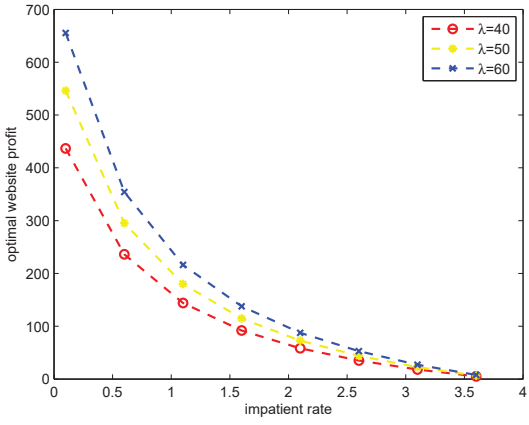

Figure 7. The optimal website profit versus impatience rate for $\mu_1 = 1$, $\mu_2 = 3$, $p = 0.5$, $\delta = 0.2$, $\beta = 1$, $c_1 = 4$, $c_2 = 6$, $R_1 = 1$, $R_2 = 2$.

Observing Figure 8, we find that when the service cost gap is small, the optimal website profit decreases with the increasing service cost gap. While the gap increases, the number of premium memberships and the corresponding service costs should be controlled to maximize website profit. Thus, a higher membership fee leads to more free memberships and a high price for advertisers. And then, as shown in Figure 8, the optimal website profit grows with the increasing service cost gap, when the gap is more than a certain value.

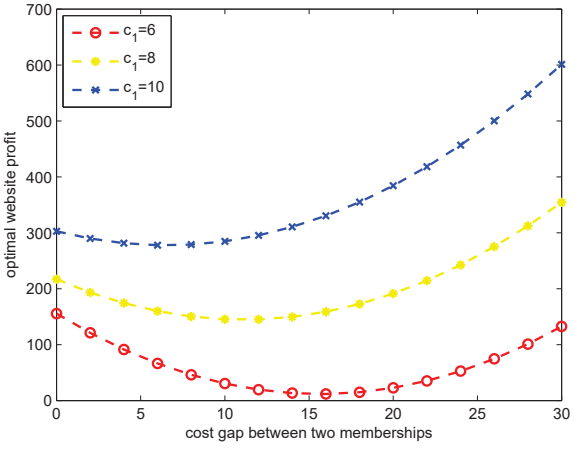

Figure 8. The optimal website profit versus cost gap between two memberships for $\lambda = 50$, $\mu_1 = 3$, $\mu_2 = 5$, $\alpha = 0.6$, $p = 0.5$, $\delta = 0.2$, $\beta = 1$, $R_1 = 1$, $R_2 = 2$.

5. Conclusions

In this paper, we investigated a queuing system with two types of customers and infinite servers. Transient and steady state probability generation functions are obtained, based on which several performance measures are derived. As an application of the queuing system, we build an optimal price decision model for a video website based on the system performances. The impacts of negative advertising effects and customers' potential value on the optimal pricing strategy are investigated, and numerical examples are presented to illustrate the impacts of parameters, such as advertising time and reward

gap, on the optimal website profit. To obtain the optimal profit of the website, the prices for the advertisers and premium members need to be designed delicately. The optimal profit is very sensitive when the advertising time is short, and as the time gets longer, the optimal profit increases steadily. Increasing the reward gap between the two types of customers is helpful for obtaining high profits from the website. The optimal pricing for video websites is very complex, and what we worked out in this paper is limited, and there are many problems can be further discussed. For example, we can consider the pricing problem in a system where the serving devices have limited service capacity, then more congestion and complex impatient problems should be included for investigation. Moreover, the service for the free memberships can be discussed more finely, which means the service time can be divided into two parts, namely the advertising time and the video time.

Funding: This research was funded by NUPTSF (No. NY220160).

Data Availability Statement: All data required for this research are included within the paper.

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

References

1. Barrer, D.Y. Queueing with impatient customers and indifferent clerks. *Oper. Res.* **1957**, *5*, 644–649. [CrossRef]
2. Barrer, D.Y. Queueing with impatient customers and ordered service. *Oper. Res.* **1957**, *5*, 650–656. [CrossRef]
3. Phung-Duc, T. Asymptotic analysis for Markovian queues with two types of nonpersistent retrial customers. *Appl. Math. Comput.* **2015**, *265*, 768–784. [CrossRef]
4. Kim, B.; Kim, J. A single server queue with Markov modulated service rates and impatient customers. *Perform. Eval.* **2015**, *83*, 1–15. [CrossRef]
5. Adan, I.; Hathaway, B.; Kulkarni, V.G. On first-come, first-served queues with two classes of impatient customers. *Queueing Syst.* **2019**, *91*, 113–142. [CrossRef]
6. Vijayashree, K.; VAmbika, K. An M/M/1 queue subject to differentiated vacation with partial interruption and customer impatience. *Qual. Technol. Quant. Manag.* **2021**, *17*, 657–682.
7. Dong, J.; Ibrahim, R. SRPT scheduling discipline in many-server queues with impatient customers. *Manag. Sci.* **2021**, *67*, 7708–7718. [CrossRef]
8. Yin, M.; Yan, M.; Guo, Y.; Liu, M. Analysis of a Pre-Emptive Two-Priority Queuing System with Impatient Customers and Heterogeneous Servers. *Mathematics* **2023**, *11*, 3878. [CrossRef]
9. Dudin, A.N.; Chakravarthy, S.R.; Dudin, S.A.; Dudina, O.S. Queueing system with server breakdowns and individual customer abandonment. *Qual. Technol. Quant. Manag.* **2024**, *21*, 441–460. [CrossRef]
10. Jackson, R.R.P.; Aspden, P. A transient solution to the multistage Poisson queueing system with infinite servers. *Oper. Res.* **1980**, *28*, 618–622. [CrossRef]
11. Sato, T.; Okada, H.; Yamazato, T.; Katayama, M.; Ogawa, A. Throughput analysis of DS/SSMA unslotted ALOHA system with fixed packet length. *IEEE J. Sel. Areas Commun.* **1996**, *14*, 750–756. [CrossRef]
12. Shi, Y.; Li, X.; Fan, P. Optimization of an M/M/∞ queueing system with free experience service. *Asia-Pac. J. Oper. Res.* **2016**, *33*, 1650051. [CrossRef]
13. Hassin, R.; Ravner, L. Delay-Minimizing Capacity Allocation in an Infinite Server-Queueing System. *Stoch. Syst.* **2019**, *9*, 27–46. [CrossRef]
14. Burnetas, A.; Economou, A. Equilibrium customer strategies in a single server Markovian queue with setup times. *Queueing Syst.* **2007**, *56*, 213–228. [CrossRef]
15. Economou, A.; Kanta, S. Equilibrium balking strategies in the observable single-server queue with breakdowns and repairs. *Oper. Res. Lett.* **2008**, *36*, 696–699. [CrossRef]
16. Sun, W.; Li, S.; Cheng, E. Equilibrium and optimal balking strategies of customers in Markovian queues with multiple vacations and N-policy. *Appl. Math. Model.* **2016**, *40*, 284–301. [CrossRef]
17. Yu, S.; Liu, Z.; Wu, J. Equilibrium strategies of the unobservable M/M/1 queue with balking and delayed repairs. *Appl. Math. Comput.* **2016**, *290*, 56–65. [CrossRef]
18. Economou, A.; Manou, A. Strategic behavior in an observable fluid queue with an alternating service process. *Eur. J. Oper. Res.* **2016**, *254*, 148–160. [CrossRef]
19. Hassin, R.; Roet-Green, R. Cascade equilibrium strategies in a two-server queueing system with inspection cost. *Eur. J. Oper. Res.* **2018**, *267*, 1014–1026. [CrossRef]
20. Lee, D.H. Equilibrium balking strategies in Markovian queues with a single working vacation and vacation interruption. *Qual. Technol. Quant. Manag.* **2019**, *16*, 355–376. [CrossRef]
21. Lee, C.; Ward, A.R. Optimal pricing and capacity sizing for the GI/GI/1 queue. *Oper. Res. Lett.* **2014**, *42*, 527–531. [CrossRef]

22. Bai, J.; So, K.C.; Tang, C. A queueing model for managing small projects under uncertainties. *Eur. J. Oper. Res.* **2016**, *253*, 777–790. [CrossRef]
23. Rochet, J.C.; Tirole, J. Platform competition in two-sided markets. *J. Eur. Econ. Assoc.* **2003**, *1*, 990–1029. [CrossRef]
24. Armstrong, M.; Wright, J. Two-Sided Markets with Multihoming and Exclusive Dealing. Idei Working Paper Diw. 2004. Available online: https://www.diw.de/documents/dokumentenarchiv/17/42281/2004-480-v01.pdf (accessed on 1 July 2004).
25. Zeithammer, R.; Thomadsen, R. Vertical differentiation with variety-seeking consumers. *Manag. Sci.* **2013**, *59*, 390–401. [CrossRef]
26. Manuszak, M.D.; Krzysztof, K. The Impact of Price Controls in Two-Sided Markets: Evidence from US Debit Card Interchange Fee Regulation. 2017. Available online: https://doi.org/10.17016/FEDS.2017.074 (accessed on 4 July 2017).
27. Rochet, J.C.; Tirole, J. Two-sided markets: A progress report. *Rand J. Econ.* **2006**, *37*, 645–667. [CrossRef]
28. Armstrong, M.; Wright, J. Two-sided markets, competitive bottlenecks and exclusive contracts. *Econ. Theory* **2007**, *32*, 353–380. [CrossRef]
29. Wang, S.; Chen, H.; Wu, D. Regulating platform competition in two-sided markets under the O2O era. *Int. J. Prod. Econ.* **2019**, *215*, 131–143. [CrossRef]

MDPI AG
Grosspeteranlage 5
4052 Basel
Switzerland
Tel.: +41 61 683 77 34

Mathematics Editorial Office
E-mail: mathematics@mdpi.com
www.mdpi.com/journal/mathematics

Academic Open
Access Publishing

mdpi.com

ISBN 978-3-7258-1730-6