Advances in Applied Mathematics 应用数学进展, 2025, 14(4), 135-142
Published Online April 2025 in Hans. https://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2025.144146

Hans汉斯

基于Actor-Critic强化学习的再保险与投资问题

欧　萍，卢相刚*

广东工业大学数学与统计学院，广东　广州

收稿日期：2025年3月3日；录用日期：2025年3月26日；发布日期：2025年4月8日

摘　要

本文研究了具有普通理赔和巨灾理赔两个业务线的保险公司的最优再保险和投资策略。它假设公司购买风险资产受到随机市场影响，索赔的时间和规模受到随机因素的影响，同时考虑金融市场和保险市场之间的共同冲击。建立了一个优化准则，以最大化在有限时间范围内保险公司的累积财富效用。然后，利用动态规划原理和Itô公式，我们推导了Hamilton-Jacobi-Bellman (HJB)方程。由于扩散过程和状态切换的复杂性，很难找到精确的解，因此本文采用了数值方法(Actor-Critic强化学习算法)。最后，我们给出一个数值例子。

关键词

再保险，投资，制度转换，强化学习

Reinsurance and Investment Problems Based on Actor-Critic Reinforcement Learning

Ping Ou, Xianggang Lu*

School of Mathematics and Statistics, Guangdong University of Technology, Guangzhou Guangdong

Received: Mar. 3rd, 2025; accepted: Mar. 26th, 2025; published: Apr. 8th, 2025

Abstract

This paper studies the optimal reinsurance and investment strategies of insurance companies with two business lines: general claims and catastrophic claims. It assumes that the company's purchase of risky assets is influenced by random market factors, and the timing and scale of claims are affected by random factors, while considering the joint impact between the financial market and the

*通讯作者。

文章引用: 欧萍, 卢相刚. 基于 Actor-Critic 强化学习的再保险与投资问题[J]. 应用数学进展, 2025, 14(4): 135-142.
DOI: 10.12677/aam.2025.144146

insurance market. An optimization criterion has been established to maximize the cumulative wealth utility of insurance companies within a limited time frame. Then, using the principles of dynamic programming and the Itô formula, we derived the Hamilton Jacobi Bellman (HJB) equation. Due to the complexity of the diffusion process and state switching, it is difficult to find an exact solution, so this paper adopts a numerical method (Actor-Critic reinforcement learning algorithm). Finally, we provide a numerical example.

Keywords

Reinsurance, Investment, Regime-Switching, Reinforcement Learning

1. 引言

随着金融市场的复杂性和不确定性不断增加，保险公司在再保险和投资策略优化方面受到越来越大的挑战。强化学习(Reinforcement Learning, RL)作为一种数据驱动的优化方法，逐渐在金融领域得到了广泛关注[1]。近年来，强化学习已被成功地应用于解决各种复杂的决策问题。它不仅可以解决离散时间的优化问题，也可以解决连续时间的优化问题。例如，Wang 和 Zhou 等人[2]在 RL 框架下研究了连续时间和空间中的探索性随机控制问题，Jia 和 Zhou 等人[3]在演员－评论家 RL 框架下研究了连续时间和空间中的政策梯度(PG)。此外，Zhou 等人[4]提出了一种基于 Actor-Critic 的数值方法，用于求解高维 Hamilton-Jacobi-Bellman (HJB)方程，通过学习确定性策略和策略梯度，显著提升了样本利用率并优化了控制效果。

在再保险和投资策略优化领域，强化学习的应用展现出巨大潜力。通过与环境的实时交互，Actor-Critic 算法能够动态学习最优的再保险比例和投资策略，从而实现风险与收益的平衡。Jin 等人[5]提出了一种混合强化学习框架，将再保险决策与投资组合优化相结合，显著提升了保险公司的风险调整收益。Li 等人[6]则在随机波动和市场切换环境下，利用强化学习和马尔可夫链切换模型，优化了再保险和投资策略，以最小化破产概率。受到这些研究的启发，我们构建了一个基于 Actor-Critic 算法的再保险与投资模型，将最优比例再保险和投资问题建模为马尔可夫决策过程(Markov Decision Process, MDP)，并设计了相应的状态空间、动作空间和奖励函数。本文旨在填补现有研究的空白，提出一种基于 Actor-Critic 强化学习框架的最优再保险与投资问题，以解决保险公司在有限时间范围内的累计财富效用最大化问题。

本章的概述如下，第 2 节介绍了一个金融和保险市场模型及其优化问题，其中我们假设环境因素只影响保险业务，状态切换只影响金融市场，灾难性索赔与风险资产价格之间受到共同冲击。第 3 节利用 Actor-Critic 强化学习方法建立适当的数值近似算法并解决随机优化控制问题。第 4 节给出了一个数值例子。

2. 模型与优化问题

基于对再保险和投资现状的深入研究，我们构建了一家保险公司涵盖两条业务线的模型。设 $(\Omega,\mathcal{F},\mathbb{F},\mathbb{P})$ 为滤波概率空间，其中滤波 $\mathbb{F}=\{\mathcal{F}_t\}_{t\geq 0}$ 满足通常条件。

2.1. 财富过程

在保险市场中，我们假设保险公司的盈余过程满足以下方程：

$$\mathrm{d}R_t^{(k),u_t^{(k)}} = \left[c^{(k)}\left(t,Y_t^{(k)}\right) - q^{(k)}\left(t,Y_t^{(k)},u_t^{(k)}\right)\right]\mathrm{d}t - \int_0^{+\infty} z u_t^{(k)} m^{(k)}(\mathrm{d}t,\mathrm{d}z).$$

其中 $R_t^{(k),u_t^{(k)}}, k=1,2, t\in[0,+\infty)$，表示保险公司在两条业务线下的盈余过程，$u_t^{(k)}$ 为 t 时刻保险公司购买再保险时的自留比，$c^{(k)}\left(t,Y_t^{(k)}\right)$ 表示保险保费，$q^{(k)}\left(t,Y_t^{(k)},u_t^{(k)}\right)$ 表示再保险保费。

累计索赔过程 $S=\left\{S_t, t\in[0,\infty)\right\}$ 定义为：

$$S_t = S_t^{(1)} + S_t^{(2)} = \sum_{n=1}^{N_t^{(1)}} Z_n^{(1)} + \sum_{n=1}^{N_t^{(2)}} Z_n^{(2)}.$$

其中，$S_t^{(1)}$ 和 $S_t^{(2)}$ 分别表示保险公司在 $[0,t]$ 内两条业务线(普通理赔和灾难性理赔)的累计索赔规模，$\left\{N_t^{(1)}\right\}_{t\geq 0}$ 和 $\left\{N_t^{(2)}\right\}_{t\geq 0}$ 分别表示两条业务线的独立计数过程，$\left\{Z_n^{(k)}, n\geq 1\right\}, k=1,2$ 表示保险公司的第 n 个索赔金额，是正独立同分布随机变量的序列。

随机因 $Y_t^{(k)}$ 子满足以下方程：

$$\mathrm{d}Y_t^{(k)} = b^{(k)}(t)\mathrm{d}t + a^{(k)}(t)\mathrm{d}W_t^{(k)}, Y_0^{(k)} = y^{(k)}\in\mathbb{R}, k=1,2.$$

其中 $W_t^{(1)}$ 和 $W_t^{(2)}$ 是相互独立的布朗运动，$a^{(k)}(t), b^{(k)}(t)\in\mathbb{R}$ 是可测函数。设 $m^{(k)}(\mathrm{d}t,\mathrm{d}z)$ 代表与 $S_t^{(k)}$ 相关联的随机计数测度。且满足如下关系：

$$m^{(k)}(\mathrm{d}t,\mathrm{d}z) = \sum_{n\in\mathbb{N}} \hat{\delta}_{\left\{\nu_n^{(k)}, Z_n^{(k)}\right\}}(\mathrm{d}t,\mathrm{d}z)\mathrm{I}_{\left\{\nu_n^{(k)}<+\infty\right\}},$$

对于 $k=1,2$，其中 $\hat{\delta}_{\{t,z\}}$ 表示在点 (t,z) 处的 Dirac 测度，$\nu_n^{(k)}$ 表示索赔到达时间(即 $\left\{N_t^{(k)}\right\}_{t\geq 0}$ 的跳跃时间)，z 是索赔大小并且具有 $F^{(k)}(z):\mathbb{R}\to(0,1)$ 的分布。

引理 2.1 假设 $N_t^{(k)}, k=1,2$，是由随机强度 $\lambda^{(k)}\left(t,Y_t^{(k)}\right):[0,+\infty)\times\mathbb{R}\to(0,+\infty)$ 表示的计数过程，$Y_t^{(k)}$ 表示影响索赔到达强度的环境因素，并将测度 $m^{(k)}(\mathrm{d}t,\mathrm{d}z)$ 的对偶可预测投影定义如下：

$$v^{(k)}(\mathrm{d}t,\mathrm{d}z) = \lambda^{(k)}\left(t,Y_t^{(k)}\right)F^{(k)}(\mathrm{d}z)\mathrm{d}t.$$

证明：请参见文献[7]。

在金融市场中，我们假设保险公司将部分盈余投资风险资产，其价格遵循如下几何布朗运动过程：

$$\mathrm{d}P_t = P_t\left[\mu(\alpha(t))\mathrm{d}t + \sigma(\alpha(t))\mathrm{d}W(t) - \int_0^{+\infty} K(t,z)m^{(2)}(\mathrm{d}t,\mathrm{d}z)\right].$$

其中 $\left\{P_t\right\}_{t\geq 0}$ 表示风险资产的价格，对于每一个 $i\in M$，$\mu(\alpha(t))\in\mathbb{R}$ 是风险资产的收益率，$\sigma(\alpha(t))\in(0,+\infty)$ 是风险资产的波动率，$K(t,z):[0,+\infty)\times[0,+\infty)\to[0,1)$ 表示风险资产的跳跃规模 $\left\{W_t\right\}_{t\geq 0}$ 是标准几何布朗运动。在有限空间 $\mathcal{M}=\{1,\cdots,m\}$ 中，连续时间马尔可夫链 $\alpha(t)$ 是时间 t 的随机市场。这里 $\alpha(t)$ 表示经历风险资产过程马尔可夫状态切换的市场状态，由 $Q=\left(q_{ij}\right)\in\mathbb{R}^{m\times m}$ 生成。符号表示如下：

$$P\left\{\alpha(t+\delta)=j \mid \alpha(t)=i, \alpha(s), s\leq t\right\} = \begin{cases} q_{ij}\delta + o(\delta), & \text{if } j\neq i \\ 1+q_{ii}\delta+o(\delta), & \text{if } j=i \end{cases}$$

此外，假设保险公司还将盈余投资于无风险资产，其价格满足以下随机微分方程：

$$\mathrm{d}B_t = r(t)B_t\mathrm{d}t.$$

其中 $\left\{B_t\right\}_{t\geq 0}$ 表示无风险资产价格，$r(t)$ 为无风险资产利率。

设 $\pi_t\in\mathbb{R}^+$ 表示投资于风险资产的盈余总额，即不允许卖空。建立财富过程：

$$\mathrm{d}X_t^{\xi} = \mathrm{d}R_t^{(1),u^{(1)}} + \mathrm{d}R_t^{(2),u^{(2)}} + \pi_t\frac{\mathrm{d}P_t}{P_t} + \left(X_t^{\xi}-\pi_t\right)\frac{\mathrm{d}B_t}{B_t}, \quad X_0^{\xi} = x\in\mathbb{R}^+,$$

其中，$\xi=\left\{\xi_t=\left(u_t^{(1)},u_t^{(2)},\pi_t\right)\right\}_{t\geq 0}$ 是再保险与投资策略。代入化简后，保险公司的财富过程如下：

$$\begin{aligned}\mathrm{d}X_t^{\xi}=&\left[c^{(1)}\left(t,Y_t^{(1)}\right)+c^{(2)}\left(t,Y_t^{(2)}\right)-q^{(1)}\left(t,Y_t^{(1)},u_t^{(1)}\right)-q^{(2)}\left(t,Y_t^{(2)},u_t^{(2)}\right)+\pi_t\left(\mu\left(\alpha(t)\right)-r(t)\right)+X_t^{\xi}r(t)\right]\mathrm{d}t\\&+\sigma\left(\alpha(t)\right)\pi_t\mathrm{d}W(t)-\int_0^{+\infty}zu_t^{(1)}m^{(1)}(\mathrm{d}t,\mathrm{d}z)-\int_0^{+\infty}\left(zu_t^{(2)}+\pi_tK(t,z)\right)m^{(2)}(\mathrm{d}t,\mathrm{d}z).\end{aligned}$$

2.2. 优化目标

下面我们介绍了随机最优控制问题。保险公司的目标是使其财富的期望累积效用最大化，即最大化以下优化问题

$$J\left(s,x,y^{(1)},y^{(2)},i,\xi\right)=E_{s,x,y^{(1)},y^{(2)},i}\left[\int_s^T\mathrm{e}^{-\rho(t-s)}U\left(X_t^{\xi}\right)\mathrm{d}t+h\left(X_T^{\xi}\right)\right],$$

其中，$U(x)=1-\mathrm{e}^{-\gamma x}$ 表示保险公司的指数偏好，$h(x)$ 代表终端函数，ρ 代表贴现系数，γ 为保险公司的相对风险厌恶系数，s 为初始时间，$s\in[0,\infty)$，T 为保险合同终止时间。

定义可容许控制集 $\pi_t\in\mathbb{R}^+$，最优控制问题的值函数定义为：

$$V\left(s,x,y^{(1)},y^{(2)},i\right)=\sup_{\xi\in\Xi}J\left(s,x,y^{(1)},y^{(2)},i,\xi\right).$$

由动态规划原理和伊藤公式，得出相应的 HJB 方程如下：

$$\begin{aligned}0=&V_s+\sup_{\xi\in U}\left\{\left(c^{(1)}\left(s,y^{(1)}\right)+c^{(2)}\left(s,y^{(2)}\right)-q^{(1)}\left(s,y^{(1)},u^{(1)}\right)-q^{(2)}\left(s,y^{(2)},u^{(2)}\right)\right.\\&+\pi\left(\mu(i)-r(s)\right)+xr(s)\Big)V_x+b^{(1)}(s)V_{y^{(1)}}+b^{(2)}(s)V_{y^{(2)}}+\frac{1}{2}\sigma(i)^2\pi^2V_{x^2}\\&+\frac{1}{2}\left(a^{(1)}\right)^2(s)V_{y^{(1)^2}}+\frac{1}{2}\left(a^{(2)}\right)^2(s)V_{y^{(2)^2}}+\sum_{j\neq i}q_{ij}\left(V\left(s,x,y^{(1)},y^{(2)},j\right)-V\left(s,x,y^{(1)},y^{(2)},i\right)\right)\\&+\int_0^x\left(V\left(s,x-zu^{(1)},y^{(1)},y^{(2)},i\right)-V\left(s,x,y^{(1)},y^{(2)},i\right)\right)\lambda^{(1)}\left(s,y^{(1)}\right)F^{(1)}(\mathrm{d}z)\\&+\int_0^x\left(V\left(s,x-zu^{(2)}-\pi K\left(\alpha(s),z\right),y^{(1)},y^{(2)},i\right)-V\left(s,x,y^{(1)},y^{(2)},i\right)\right)\lambda^{(2)}\left(s,y^{(2)}\right)F^{(2)}(\mathrm{d}z)\\&\left.-\rho V\left(s,x,y^{(1)},y^{(2)},i\right)+U(x)\right\},\end{aligned}$$

其中 $V_{\mathcal{I}}=\dfrac{\partial V\left(s,x,y^{(1)},y^{(2)},i\right)}{\partial\mathcal{I}}$ 和 $V_{\mathcal{I}^2}=\dfrac{\partial^2V\left(s,x,y^{(1)},y^{(2)},i\right)}{\partial\mathcal{I}^2}$，$\mathcal{I}=s,x,y^{(1)},y^{(2)}$。

通常，如果 V 足够光滑，我们可以用动态规划原理将其描述为 HJB 方程的解。然而，上述 HJB 方程的显式解不容易获得，因此我们采用数值求解方法。

3. 数值算法

基于强化学习中的演员－评论家(Actor-Critic)框架，可以同时求解价值函数和策略函数，针对评论家的策略评估(Policy Evaluation, PE)和针对演员的策略迭代(Policy Iteration, PI)。目前，在 Actor-Critic 框架下已经开发了许多算法，例如文献[8] [9]。

3.1. 策略评估

在本节中，我们开发相应的 PE 程序，并使用函数逼近方法来获得价值函数的估计。现在让我们将连续设置中的 TD 误差定义为：

$$\begin{aligned}TD^{\xi}=&\int_s^T \mathrm{e}^{-\rho(t-s)}U\left(X_t^{\xi}\right)\mathrm{d}t+\mathrm{e}^{-\rho(T-s)}V^{\xi}\left(T,x_T,y_T^{(1)},y_T^{(2)},\alpha_T\right)\\&-\int_s^T \mathrm{e}^{-\rho(t-s)}V_x^{\xi}\sigma\left(\alpha(t)\right)\pi_t\mathrm{d}W_t-\int_s^T \mathrm{e}^{-\rho(t-s)}V_{y^{(1)}}^{\xi}a^{(1)}(t)\mathrm{d}W_t^{(1)}\\&-\int_s^T \mathrm{e}^{-\rho(t-s)}V_{y^{(2)}}^{\xi}a^{(2)}(t)\mathrm{d}W_t^{(2)}-V^{\xi}\left(s,x,y^{(1)},y^{(2)},i\right).\end{aligned}$$

我们进一步为评论家定义以下损失函数：

$$\begin{aligned}L(\theta_V,\theta_G)&=E_{s,z,i}\left[\frac{1}{2}\left(TD^{\xi}\right)^2\right]\\&=E_{s,z,i}\left[\frac{1}{2}\left(\int_s^T \mathrm{e}^{-\rho(t-s)}U\left(X_t^{\xi}\right)\mathrm{d}t+\mathrm{e}^{-\rho(\tau-s)}V^{\xi}\left(\tau,x,y^{(1)},y^{(2)},i;\theta_V\right)\right.\right.\\&\quad-\int_s^T \mathrm{e}^{-\rho(t-s)}V_x^{\xi}\left(t,x,y^{(1)},y^{(2)},i;\theta_G\right)\sigma\left(\alpha(t)\right)\pi_t\mathrm{d}W_t\\&\quad-\int_s^T \mathrm{e}^{-\rho(t-s)}V_{y^{(1)}}^{\xi}\left(t,x,y^{(1)},y^{(2)},i;\theta_G\right)a^{(1)}(t)\mathrm{d}W_t^{(1)}\\&\quad\left.\left.-\int_s^T \mathrm{e}^{-\rho(t-s)}V_{y^{(2)}}^{\xi}\left(t,x,y^{(1)},y^{(2)},i;\theta_G\right)a^{(2)}(t)\mathrm{d}W_t^{(2)}-V^{\xi}\left(s,x,y^{(1)},y^{(2)},i;\theta_V\right)\right)^2\right].\end{aligned}$$

3.2. 策略梯度

在本小节中，我们介绍演员部分，并使用策略梯度来改进演员的策略。回忆我们的最优控制问题并利用动态规划原理，对于最优值函数 V，我们有

$$V\left(s,x,y^{(1)},y^{(2)},i\right)=\sup_{\xi\in\Xi}E_{s,x,y^{(1)},y^{(2)},i,\xi}\left[\int_s^T \mathrm{e}^{-\rho(t-s)}U\left(X_t^{\xi}\right)\mathrm{d}t+\mathrm{e}^{-\rho(T-s)}V\left(T,x_T,y_T^{(1)},y_T^{(2)},\alpha_T\right)\right].$$

其中，ξ 是最优控制，即控制 ξ 应该最大化右侧的功能函数。定义 $\tau=\inf\left\{t\geq 0,X(t)<0\right\}$ 为破产时间，考虑停止时间 τ，我们可以使用以下目标函数来定义演员(Actor)的目标

$$\begin{aligned}&J\left(s,x,y^{(1)},y^{(2)},i,\xi\right)\\&=E_{s,x,y^{(1)},y^{(2)},i,\xi}\left[\int_s^{T\wedge\tau} \mathrm{e}^{-\rho(t-s)}U\left(X_t^{\xi}\right)\mathrm{d}t+\mathrm{e}^{-\rho(T\wedge\tau-s)}V\left(T\wedge\tau,x_{T\wedge\tau},y_{T\wedge\tau}^{(1)},y_{T\wedge\tau}^{(2)},\alpha_{T\wedge\tau}\right)\right].\end{aligned}$$

我们将函数导数近似为

$$\nabla J=E_{s,x,y^{(1)},y^{(2)},i,\xi}\left[\int_s^{T\wedge\tau} \mathrm{e}^{-\rho(t-s)}\gamma\mathrm{e}^{-\gamma x_t}\frac{\delta X_t}{\delta\theta_{\xi}}\mathrm{d}t+\mathrm{e}^{-\rho(T\wedge\tau-s)}\left(\frac{\delta V}{\delta x}\frac{\delta X_{T\wedge\tau}}{\delta\theta_{\xi}}+\frac{\delta V}{\delta y^{(1)}}\frac{\delta Y_{T\wedge\tau}^{(1)}}{\delta\theta_{\xi}}+\frac{\delta V}{\delta y^{(2)}}\frac{\delta Y_{T\wedge\tau}^{(2)}}{\delta\theta_{\xi}}\right)\right].$$

4. 数值算法

在本节中，我们通过一个基本的例子和敏感性分析，深入探讨了如何找到最优的比例再保险和投资策略使得在有限时间范围内保险公司的累计财富效用最大化问题。

4.1. 一个基本的例子

在这一部分，我们设置了一组合理的参数进行数值实验。让 $\delta=0.001$，$\mu_1=0.1$，$\mu_2=0.2$，$\sigma_1=0.1$，$\sigma_2=0.2$，$Q=\begin{pmatrix}-2 & 2\\ 1 & -1\end{pmatrix}$，$r=0.05$，$\rho=0.1$，$\gamma=0.5$，$b^{(1)}(t)=0.8$，$b^{(2)}(t)=0.4$，$a^{(1)}(t)=0.2$，$a^{(2)}(t)=0.1$，假设索赔的到达强度为

$$\begin{cases}\lambda^{(1)}\left(t,y^{(1)}\right)=\lambda(t)\,f_1\left(y^{(1)}\right),\\ \lambda^{(2)}\left(t,y^{(2)}\right)=\lambda(t)\,f_2\left(y^{(2)}\right),\end{cases}$$

其中 $\lambda(t)=2t+1$， $f_1\left(y^{(1)}\right)=y^{(1)}+1\in(1,2)$， $f_2\left(y^{(2)}\right)=y^{(2)}\in(0,1)$， $y^{(1)},y^{(2)}\in[0,1]$。设 $F^{(1)}(z)=1-\mathrm{e}^{-az}$，$F^{(2)}(z)=1-\mathrm{e}^{-bz}$， $b>a>0$， $z>0$ 设 $\theta=0.3$， $\theta_R=0.4$， $a=2$， $b=5$， $E\left[Z^{(1)}\right]=1/2$， $E\left[Z^{(2)}\right]=1/5$。在期望值原则下，我们有

$$\begin{cases}c^{(1)}\left(t,y^{(1)}\right)=\left(1+\theta^{(1)}\right)E\left[Z^{(1)}\right]\lambda^{(1)}\left(t,y^{(1)}\right),\\ c^{(2)}\left(t,y^{(2)}\right)=\left(1+\theta^{(2)}\right)E\left[Z^{(2)}\right]\lambda^{(2)}\left(t,y^{(2)}\right),\\ q^{(1)}\left(t,y^{(1)},u^{(1)}\right)=\left(1+\theta_R^{(1)}\right)E\left[Z^{(1)}\right]\left(1-u^{(1)}\right)\lambda^{(1)}\left(t,y^{(1)}\right),\\ q^{(2)}\left(t,y^{(2)},u^{(2)}\right)=\left(1+\theta_R^{(2)}\right)E\left[Z^{(2)}\right]\left(1-u^{(2)}\right)\lambda^{(2)}\left(t,y^{(2)}\right).\end{cases}$$

设学习率 0.01，时间间隔 0.001，3 个隐藏层，128 个隐藏层维度，200 次迭代，批量 500，我们用 python 软件计算得到如下图。

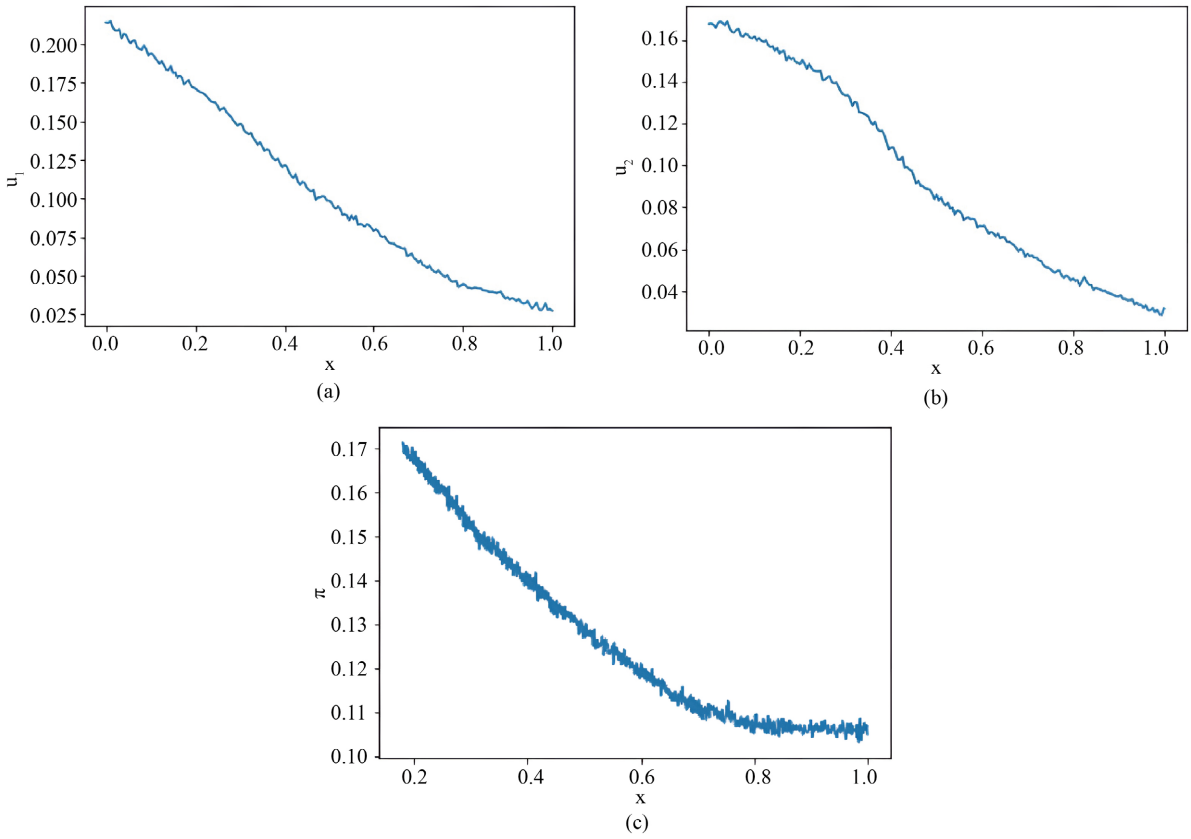

Figure 1. The control u_1, u_2 and π
图 1. 控制 u_1、u_2 和 π

在图 1 中，我们可以看到，在上述数据设置下，两项保险业务的保险自留比例随着财富增长而降低，投资于风险资产的金额也随之降低。保险公司在期望原则下，选择较低的自留比例，即再保险比例上升，

把索赔风险分担出去。这样做，虽然再保险保费上升，但是保险公司需要付出自己部分的索赔金额就越少了，总体而言财富是在累积上升，符合目标函数的假设。因此，对于普通索赔和灾难索赔到来时，保险公司面临着巨大的赔款，此时会选择把索赔风险分担出去，即购买再保险业务，从而再保险比例上升，保险自留比例下降。那么对于保险公司的索赔损失、购买再保险费用就能够从获取的保费中弥补回来，并且总财富在上升。同时，对于灾难性保险的自留比例 u_2 比普通保险的自留比例 u_1 要小，说明灾难性索赔支付的金额更加巨大，保险公司更倾向于将其风险分担出去。

此外，我们可以看见由于资产市场的价格不稳定，造成投资于风险资产的金额部分会减少到一定值后趋于稳定，也就是说，保险公司在动荡的金融市场中会选择较为稳定的无风险资产，以此维持财富的增加。

4.2. 敏感性分析

我们分别采取不同的厌恶系数 γ 和索赔速率 $\lambda(t)$ 来研究对值函数的影响。

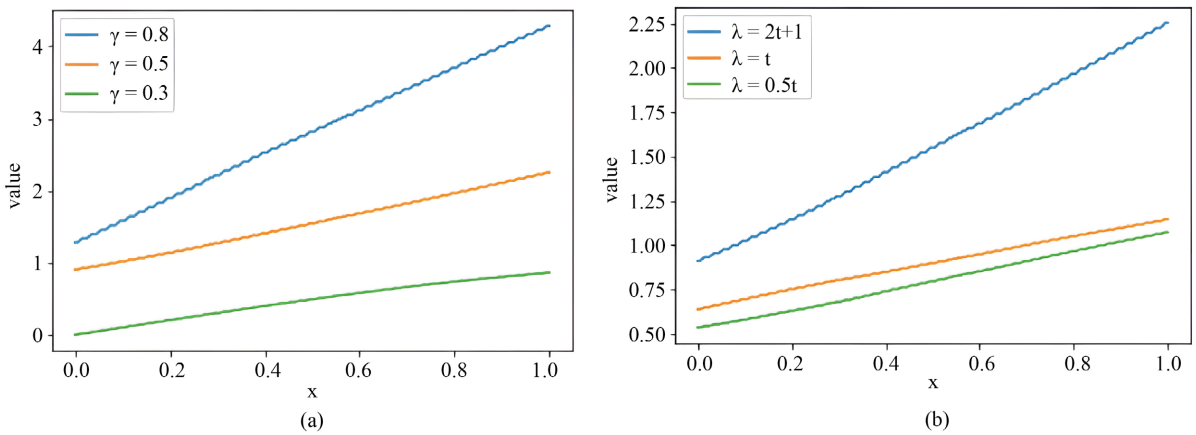

Figure 2. The influence of different aversion coefficients γ and claim rates $\lambda(t)$ on the value function

图 2. 不同的厌恶系数 γ 和索赔速率 $\lambda(t)$ 对值函数的影响

在图 2(a)中，我们看到随着厌恶系数 γ 的增长，值函数也在不断增加。在图 2(b)中，随着索赔速率 $\lambda(t)$ 的增加，值函数也在不断上涨。可能原因如下：第一，在效用函数 U 中，U 是关于厌恶系数的增函数，当越大时，效用函数也越大，那么值函数就会增长。第二，索赔速率 $\lambda(t)$ 影响索赔到达强度，进而影响保费的收入，当索赔速率 $\lambda(t)$ 越大时，两条保险业务线的保费金额也会增加，这给保险公司带来较大的财富。索赔金额虽然也在增长，但是保险公司通过调整再保险策略，把风险分担出去，使得利润还是正增长，从而保持了财富的增长。总之，这两个参数都会对值函数产生比较大的影响。有效利用参数，能够使效用最大化。

参考文献

[1] Sutton, R.S. and Barto, A.G. (1998) Reinforcement Learning: An Introduction. MIT Press.

[2] Wang, H., Zariphopoulou, T. and Zhou, X.Y. (2020) Reinforcement Learning in Continuous Time and Space: A Stochastic Control Approach. *Journal of Machine Learning Research*, **21**, 1-34.

[3] Jia, Y. and Zhou, X. (2021) Policy Gradient and Actor-Critic Learning in Continuous Time and Space: Theory and Algorithms. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3969101

[4] Zhou, M., Han, J. and Lu, J. (2021) Actor-Critic Method for High Dimensional Static Hamilton-Jacobi-Bellman Partial

Differential Equations Based on Neural Networks. *SIAM Journal on Scientific Computing*, **43**, A4043-A4066. https://doi.org/10.1137/21m1402303

[5] Jin, Z., Yang, H. and Yin, G. (2021) A Hybrid Deep Learning Method for Optimal Insurance Strategies: Algorithms and Convergence Analysis. *Insurance*: *Mathematics and Economics*, **96**, 262-275. https://doi.org/10.1016/j.insmatheco.2020.11.012

[6] Li, L. and Qiu, Z. (2025) Time-Consistent Robust Investment-Reinsurance Strategy with Common Shock Dependence under CEV Model. *PLOS ONE*, **20**, e0316649. https://doi.org/10.1371/journal.pone.0316649

[7] Brémaud, P. (1981) Point Processes and Queues. Springer.

[8] Pik, J., Chan, E.P., Broad, J., *et al*. (2025) Hands-On AI Trading with Python, Quant Connect, and AWS. Wiley.

[9] Liu, Y., Zhang, K., Basar, T., *et al*. (2020) An Improved Analysis of (Variance-Reduced) Policy Gradient and Natural Policy Gradient Methods. *Advances in Neural Information Processing Systems*, **33**, 7624-7636.