Reinforcement Learning-Based Dynamic Pricing Strategy for Life Insurance in a Low-Interest Rate Environment

Zi-Jia Yi*

Central University of Finance and Economics, Beijing, China

*yichao1667@163.com

Abstract. Under the low-interest rate environment, the life insurance industry's interest margin loss risk accumulates continuously, and traditional static and heuristic threshold pricing struggle to balance risk control, customer retention and enterprise benefits. Taking whole life insurance as the research object, this paper constructs an MDP-DQN dynamic pricing model by depicting the pricing sequential decision-making via Markov Decision Process (MDP) and solving the high-dimensional state problem with Deep Q-Network (DQN). Based on 2013-2025 10-year Treasury bond yields, a simulated policy pool is built for three pricing strategies comparison. The results show that compared with the traditional static pricing strategy, the model reduces interest margin loss by 64.3% and increases new business value by 29.6%; compared with the heuristic threshold pricing strategy, it cuts loss by 54.5% and boosts value by 65.3%, with customer churn rate stabilized within 5%, providing an effective technical solution for the industry to cope with low-interest rate risks.

Keywords: Low-Interest Rate Environment, Dynamic Life Insurance Pricing, MDP, DQN, Interest Margin Loss

1 Introduction

After the 2008 global financial crisis, low interest rates have characterized the global macro economy, with China's market interest rates also declining. From 2013 to 2025, the average 10-year Treasury bond yield was only 3.2%, persistently below the life insurance industry's historical guaranteed interest rate range (3.5%-4.0%), creating significant asset-liability mismatch pressure. CBIRC data [1] shows that the industry's average interest margin loss rate reached 0.8% in 2023, exceeding 1.5% for some SMEs. Accumulated interest margin loss risks have become a core bottleneck for the industry's high-quality development, making pricing strategy optimization an urgent need [2].

The life insurance industry's current common pricing strategies (traditional static pricing and heuristic threshold rule pricing) have obvious flaws [3]: static pricing sets a one-time lifetime premium rate, risking cumulative spread loss amid falling rates and missed gains when rates rise; heuristic threshold pricing only adjusts mechanically by

D. Magni et al. (eds.), *Proceedings of the 2026 3rd International Conference on Applied Economics, Management Science and Social Development (AEMSS 2026)*, Advances in Economics, Business and Management Research 389,
https://doi.org/10.2991/978-94-6239-672-2_50

preset thresholds, with strong subjectivity and failure to balance spread loss control, customer retention and enterprise profits.

Based on the above situation and problems, this paper constructs a life insurance dynamic pricing model under a low-interest-rate environment using Markov Decision Process (MDP)'s sequential decision-making advantage and Deep Q-Network (DQN)'s autonomous optimization capability. MDP depicts the pricing sequential decision-making process, while DQN solves the high-dimensional state space bottleneck, enabling market interest rate-linked dynamic pricing of life insurance products and providing technical support for the industry's stable development.

2 Dynamic Pricing Model for Life Insurance

2.1 Problem Description and Assumptions

In an environment where market interest rates remain persistently low and continue to decline, life insurance products with statically determined fixed premiums are highly vulnerable to increased interest margin losses due to falling interest rates[4] [5]. If insurance companies raise premiums, customer attrition will occur, and the decline in premium income will further exacerbate the drop in earnings, forming a vicious cycle of "accumulated Interest margin losses – customer attrition – declining earnings", which poses significant risks to the operation of insurance companies.

Therefore, the accurate determination of life insurance premiums based on market interest rates needs to balance multiple factors such as Interest margin loss control, customer acceptance, and the long-term profitability of insurance companies [6]. This is a sequential decision-making problem of multi-objective collaborative optimization. The establishment of a dynamic life insurance pricing model that adjusts with market interest rates is of great significance for stabilizing the life insurance market and activating data elements [7].

The core of life insurance product pricing lies in effectively balancing risk control, customer retention, and profit improvement, ensuring that life insurance premiums can accurately cover future liability claims, operating expenses, profits, and risk reserves under the market interest rate environment, so as to safeguard the profitability of insurance companies and protect the interests of policyholders[8].

Interest margin loss is the core risk in life insurance pricing under a low-interest-rate environment, and its strict mathematical definition is as follows:

$$L = \sum_{t=1}^{T} \left(r_{guar} - r_t\right) PV_t \tag{1}$$

Where: L denotes the total Interest margin loss, r_{guar} represents the predetermined interest rate promised by the life insurance company, r_t is the market interest rate in period t, PV_t stands for the present value of premiums in period t, PV_t stands for the present value of premiums in period t, and T indicates the policy duration.

The core optimization objective of life insurance dynamic pricing is to minimize the long-term expected interest margin loss [9], and its mathematical expression is:

$$\mathbb{E}[L] = \mathbb{E}\left[\sum_{t=1}^{\infty} \gamma^t \left(r_{guar} - r_t\right) PV_t\right] \tag{2}$$

Where: $\mathbb{E}[\cdot]$ is the expectation operator, $\gamma \in (0, 1)$ is the discount factor.

This paper focuses on the dynamic pricing of whole life insurance, which operates during the policy term according to the following procedure:

The policy term T is divided into equal annual pricing periods, referred to as renewal periods. At the end of each renewal period t, the insurer evaluates the premium p_{t+1} for the next period $t+1$ based on the market interest rate. The premium is adjusted and published when the renewal period transitions, and the process repeats until the end of the policy term.

Based on the practical conditions of the life insurance industry and the rationality of model solution, the following assumptions are proposed for this process:

(1) The insurer accrues operating expenses at a fixed proportion of premium income, with a constant expense ratio, and no residual value upon policy termination.

(2) Premium fluctuations affect the policyholders' willingness to insure: a premium increase reduces the insurance uptake rate, and a decrease raises it.

(3) Policyholders' renewal depends on the premium and their own reservation price, where the reservation price follows a normal distribution.

(4) The risk-free market interest rate follows a mean-reverting process.

(5) There are no abnormal events such as malicious surrender or premium default, and policyholders pay premiums steadily.

2.2 Markov Decision Process of Life Insurance Dynamic Pricing

Markov Decision Process is a classical time-series decision framework in reinforcement learning[10]. This paper models the dynamic pricing problem of life insurance as a discrete finite Markov Decision Process, which consists of a five-tuple$<S,A,R,P,\gamma>$, where S is the state space, A is the action space, R is the reward function, P is the state transition probability function, and γ is the discount factor.

(1) State Space S: $S=[s_1,s_2,\ldots,s_T]$, where s_t represents the environmental state at time t, composed of four core variables: market interest rate r_t (based on the 10-year Treasury bond yield in this paper, %), solvency ratio sol_t(%), average customer risk level cr_t (divided into levels 1-5, with level 1 being the lowest), current premium rate p_t (¥5,000/unit).

(2) Action Space A: The set of rate adjustment actions executable by the insurance company. The action at taken by the dynamic pricing model in period t is the premium rate p_t for the period $[t,t+1)$, where $p_t \in A$. In accordance with common industry practice, the rate adjustment range is controlled within ±0.5%. A is a discrete rate set with intervals determined by product characteristics and market demand, and p_0 is the benchmark premium rate.

(3) Reward Function R: The immediate reward r_{t+1} obtained by the dynamic pricing model when taking action at in state s_t , which represents the net corporate income generated by policy renewal in period t. The net income integrates interest margin income, premium income, claim expenses, operating costs, and customer churn penalty, expressed as:

$$r_{t+1} = (1 - \omega) \times p_t \times d_t - C_f - L_t - \lambda \times loss_t \quad (3)$$

where: ω=15% (referring to the comprehensive expense ratio of life insurance companies in 2023: 14.34%), p_t is the premium rate in period t,d_t is the actual policy purchase volume in period t, C_f is the fixed operating cost per period, L_t is the expected claim expense, λ is the customer churn penalty coefficient, $loss_t$ is the customer churn rate in period t (referring to the comprehensive surrender rate of life insurance companies in Q4 2024; the portion exceeding 5% is doubly penalized)

(4) State Transition Probability Function P: The probability $P(s_{t+1}|s_t,a_t)$ of transitioning from state s_t to s_{t+1} when taking action a_t. Due to high uncertainty from interest rate fluctuations and the impact of rate adjustments on policy volume, accurate modeling is difficult. This paper uses deep reinforcement learning to estimate the state transition probabilities without prior knowledge.

(5) Discount Factor γ: $\gamma \in [0,1]$, which measures the weight placed on future rewards. Considering the long-term operation characteristics of life insurance, γ=0.99 is set to emphasize long-term Interest margin loss control and profit improvement while balancing short-term net income.

The Markov Decision Process for dynamic pricing of life insurance by insurance companies is shown in Figure 1. At pricing period t, the insurer observes the current state s_t, selects action at to determine the dynamic premium p_t, interacts with the market environment to obtain reward r_{t+1}, and then moves to period t+1.By repeating this interactive process, the insurer continuously adjusts pricing based on environmental feedback, accumulates experience to improve pricing accuracy, and finally obtains the optimal pricing strategy.

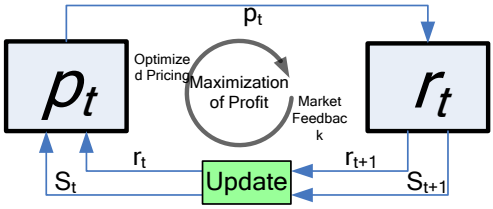

Fig. 1. Markov Decision Process for Dynamic Pricing of Data Products.

The cumulative net corporate income and constraints of life insurance products are as follows:

$$maxZ = \sum_{t=1}^{T} \gamma^{t-1} \{(1 - \omega) \times p_t \times d_t - C_f - L_t - \lambda \times loss_t\}$$

$$\begin{cases} p_t \in A \\ loss_t \leq 5 \\ d_t \geq 0 \\ sol_t \geq 50 \ (t = 1,2,\cdots,T) \end{cases} \quad (4)$$

The symbols and their meanings involved in this paper are presented in Table 1.

Table 1. Parameter symbols and their meanings.

Parameter	Parameter meanings
t	Pricing period,$t \in \{1,2,\ldots\ldots,T\}$
γ	Discount factor
p_t	Premium rate in period t ,$p_t \in A$
C_f	Fixed operating cost per period
r_t	Market interest rate in period t
cr_t	Average customer risk level in period $t \in \{1,2,\ldots,5\}$
λ	Average customer risk level in period
T	Total policy duration
ω	Operating expense ratio
d_t	Actual number of policies purchased in period t
A	Rate set
sol_t	Solvency adequacy ratio in period t
L_t	Expected claim payment in period t
$loss_t$	Customer churn rate in period t

3 Algorithm Design

3.1 Principle of the Classic DQN Algorithm

The classic DQN algorithm was proposed by Mnih et al. in 2015. The core of the algorithm is to combine the perception capability of deep learning with the decision-making capability of reinforcement learning, and use neural networks to approximate the Q-value table, so as to solve the curse of dimensionality of traditional reinforcement learning in high-dimensional state spaces.

In life insurance pricing, the Q-value $Q(s,a|\theta)$ represents the expected long-term cumulative reward of taking the premium adjustment action a under the environmental state s. The larger the Q-value, the better the action can achieve multi-objective balance.

The DQN algorithm uses a neural network to approximate the Q-value table, and takes the mean square error $L(\theta)$ as the loss function. The network parameters θ are updated by gradient descent to realize continuous training of the Q-network, and finally the estimated value approaches the true value. The calculation formula of the loss function is as follows:

$$L(\theta) = E\left\{\left[r + \gamma \max_{a' \in A} Q(s', a'|\theta) - Q(s, a|\theta)\right]^2\right\} \tag{5}$$

where s' and a' are the next state and action corresponding to the current state s and action a, θ denotes the parameters of the target network, r is the immediate reward, γ is the discount factor, and $E\{\cdot\}$ represents the mathematical expectation.

To improve training stability and convergence, the DQN algorithm adopts the experience replay mechanism and the fixed target network technique. The experience replay mechanism stores the interaction experience$<s,a,r,s'>$in a replay buffer. During

training, a batch of experience samples is randomly selected to train the Q-network and update its parameters, which breaks the correlation between samples, avoids over fitting, and improves sample utilization efficiency. The fixed target network technique constructs a target network with the same structure as the evaluation Q-network, and periodically copies parameters from the Q-network to update the target network parameters. This reduces the oscillation of target values and solves the problem of weak training stability.

3.2 Life Insurance Dynamic Pricing Algorithm Based on DQN

Combining the MDP model for life insurance pricing and the classic DQN principle, this paper designs the algorithm flow for life insurance dynamic pricing, and the basic process is shown in Figure 2.

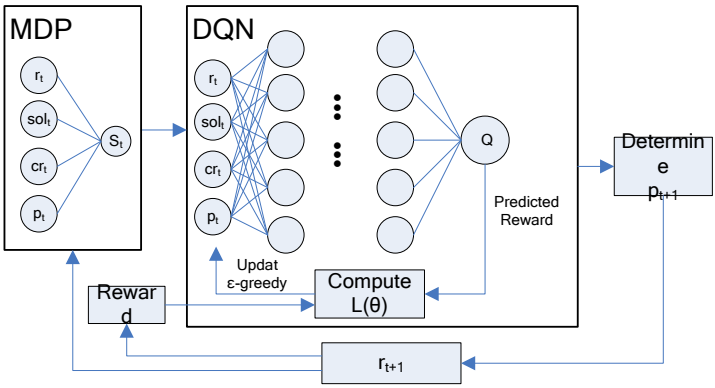

Fig. 2. Flowchart of the Life Insurance Dynamic Pricing Algorithm.

The core logic of the life insurance dynamic pricing algorithm flow is "initialization – interactive training – iterative optimization – policy output", and the specific steps are as follows:

Step 1: Initialization. Initialize core parameters such as total policy duration T, discount factor γ, experience replay buffer capacity, training batch size, target network update steps, learning rate, and exploration rate ε. Also initialize the initial environment state s_1 at pricing period t=1 (initial market interest rate, initial solvency ratio, initial average customer risk level, and initial premium rate).

Step 2: Termination judgment. Set iteration step t=t+1. If t>T, terminate training and output the optimal pricing strategy; otherwise, proceed to the next step.

Step 3: Action selection. The ε-greedy exploration-exploitation strategy is adopted. A random number $\eta \in [0,1]$ is generated. If $\eta \leq \varepsilon$, an action is randomly selected for unknown strategy exploration. If $\eta > \varepsilon$, the current state is input into the evaluation network, and the action with the maximum Q-value is selected based on existing experience to determine the current premium adjustment action.

Step 4: Experience collection. Execute the selected premium adjustment action a_t to obtain the period-t premium p_t. Interact with the pricing environment to compute the current immediate reward r and obtain the next state s'. Store the transition $<s,a,r,s'>$ into the experience replay buffer.

Step 5: Network training. When the number of samples in the experience buffer reaches the preset batch size, a batch of samples is randomly sampled. The Q-estimation is computed via the evaluation network, and the Q-target is computed via the target network. The loss function is constructed, and the Adam optimizer is used to update the parameters of the evaluation network.

Step 6: Target network update. After completing the preset number of training steps, copy the current parameters of the evaluation network to the target network. Fix the target network parameters for a period to reduce the oscillation of Q-target values and improve training stability.

Step 7: State and period update. Update the current state to the next state, i.e., $s=s'$, and set $t=t+1$. Return to Step 2 and repeat until the end of T.

Step 8: Optimal policy output. After training, for each possible environment state s, input it into the evaluation network and select the premium adjustment action with the maximum Q-value. This forms the optimal premium adjustment rules under different states for practical application by life insurance companies.

4 Verification Experiments

4.1 Experimental Environment Setup

To verify the effectiveness and superiority of the proposed life insurance dynamic pricing model, this paper establishes a numerical experimental environment, sets model parameters combined with practical industry data, constructs a simulated policy pool, and determines baseline strategies and evaluation metrics, so as to ensure the authenticity of the comparative validation.

1. Experimental Data and Parameter Setting

The experimental data include real market interest rate data and simulated policy pool data. Market interest rate data adopt monthly 10-year Treasury bond yields from 2013 to 2025, with a total of 156 samples and an average yield of 3.2%. In this paper, the value ranges from 2.5% to 4.6%, each period is set as 1 month, and T=60 months, covering typical scenarios such as downward, stable, and upward interest rate movements. The simulated policy pool contains 10,000 policies, with policyholders aged 25–55 (average 40 years old) and customer risk levels from 1 to 5 (60% low-risk, 30% medium-risk, 10% high-risk). The reservation price follows an exponential distribution with parameter v=5000. Combined with the reality of the life insurance industry and algorithm requirements, the specific experimental parameters are set as shown in Table 2.

Table 2. Experimental Parameter Table.

Parameter	Value	Parameter	Value
Total policy duration T	60 months	Experience replay buffer capacity	100000
Operating expense ratio ω	15%	Pricing action set A	¥{4800,4850,…,5200}/unit
Fixed operating cost C_f	¥50000	Learning rate	0.0005
Initial number of potential customers	1000 people	Discount factor γ	0.99
Customer decay coefficient	1000 people / month	Exploration rate ε	0.3
Rate increase impact coefficient χ^1	20	Rate decrease impact coefficient χ^2	30
Reservation price parameter v	5000	Customer churn penalty coefficient λ	10000
Target network update steps	10	Training batch size	64
Initial solvency adequacy ratio	120%	benchmark premium rate	¥5,000/unit

Note: The number of potential customers $N_t=N_1-d\times t$,the rate increase impact function $h_1(p_t, p_{t-1}) = \sqrt{p_t - p_{t-1}}$,and the rate decrease impact function $h_2(p_t, p_{t-1}) = \sqrt{p_{t-1} - p_t}$; the actual policy purchase volume $d_t=N_t\times(1-F(pt))$.

2. Experimental Baselines and Evaluation Metrics

To verify the superiority of the life insurance dynamic pricing model proposed in this paper, a comparative verification is conducted with the mainstream static pricing strategy and heuristic threshold-based pricing strategy in the industry by setting multi-dimensional indicators. The verification baselines and evaluation metrics are given as follows:

(1) Verification Baseline 1: Traditional static pricing strategy. A fixed premium of ¥5000/unit is determined when the policy is signed, without any adjustment during the policy duration. Only the current interest margin is considered, while interest rate fluctuations and customer churn are ignored.

(2) Verification Baseline 2: Heuristic threshold-based pricing strategy. Preset interest rate threshold with a lower limit of 3% and an upper limit of 4%. The premium is increased by ¥50/unit when r_t<3%, decreased by ¥50/unit when r_t>4%, and kept unchanged when 3%≤r_t≤4%.

(3) Proposed model: MDP-DQN dynamic pricing strategy. The model is constructed based on MDP, and the classic DQN is adopted to solve the optimal premium adjustment policy.

Three core dimensions are selected as evaluation metrics:

- Long-term expected loss on interest margin: measures risk control capability; the smaller the value, the better.
- Customer churn rate: measures customer retention performance, which should be controlled within 5%.

— Value of new business: measures enterprise profitability; the larger the value, the better.

4.2 Experimental Results and Analysis

The experiment was conducted under the above environment and parameters. The MDP-DQN dynamic pricing model was trained repeatedly for 10 times. To ensure the reliability of the results, the average value of the 10 training runs was adopted. The experimental results are shown in Table 3.

Table 3. Comparative Experimental Results.

Pricing Strategy	Interest margin loss(¥)	Customer churn rate(%)	NBV(¥)
Traditional static pricing strategy	1523,000	6.8	846,200
Heuristic threshold-based pricing strategy	128,700	5.7	512,800
MDP-DQN	54,500	3.9	1096,700

According to the experimental results:

(1) Risk control capability: The long-term expected Interest margin loss of the proposed model is ¥545,000, which is 64.3% lower than the ¥1.523 million of the traditional static pricing strategy and 54.5% lower than the ¥1.287 million of the heuristic threshold-based pricing strategy. The results show that the MDP-DQN model can dynamically adjust the premium according to multiple current states, achieve precise adaptation to interest rate fluctuations, and effectively control the risk of Interest margin loss.

(2) Customer retention performance: The customer churn rate of the proposed model is 3.9%, which is stably within the reasonable industry threshold of 5%. It is lower than the 6.8% of the static pricing strategy and 5.7% of the heuristic threshold-based pricing strategy. This is because the model incorporates customer churn into the reward function and avoids excessive premium adjustments through a penalty mechanism, balancing risk control and customer acceptance.

(3) Profit improvement capability: The value of new business of the proposed model is ¥10.967 million, which is 29.6% higher than that of the static pricing strategy (¥8.462 million) and 65.3% higher than that of the heuristic threshold-based pricing strategy (¥5.128 million). This indicates that the model can maximize the long-term benefits of the enterprise on the basis of controlling risks and retaining customers, and realize the collaborative optimization of the three objectives.

5 Conclusion

This paper focuses on the core pain points of life insurance pricing under the low-interest rate background, this paper carries out research on dynamic life insurance pricing strategies based on reinforcement learning and constructs an MDP-DQN pric-

ing model. Through Markov Decision Process (MDP), life insurance pricing is transformed into a discrete finite decision-making problem, and the state space and multi-objective reward function including core variables such as market interest rate and solvency ratio are clarified; combined with the Deep Q-Network (DQN) algorithm, a dynamic pricing process is designed to effectively solve the bottleneck of high-dimensional state solution. Numerical experiment verification shows that the MDP-DQN pricing model constructed in this paper is superior to the traditional static pricing strategy and the heuristic threshold rule pricing strategy in three aspects: interest margin loss control, customer retention and revenue improvement, realizing the coordinated optimization of risk, retention and revenue, and providing reliable technical support for the stable development of the life insurance industry.

References

1. State Administration of Financial Supervision: 2023 Insurance Industry Fund Utilization Report (2024)
2. Ma Xin, et al.: Research and Discussion on Life Insurance Asset-Liability Management Under the Low-Interest Rate Environment. Shanghai Insurance (06), 55-57(2025)
3. Wang Qing: Asset-Liability Management Under the Low-Interest Rate Environment. Insurance Theory&Practice (05),34-58(2022)
4. Xu Zheng, Gao Lei: Research on Asset-Liability Management of Insurance Companies Under the Low-Interest Rate Environment. Hainan Finance (12), 53-57(2020)
5. Tang Shi: Thoughts on Investment Strategies in the Low-Interest Rate Environment from the Perspective of Asset-Liability Management. China Insurance (11),8-12(2020)
6. Ling Xiuli: Research on Asset-Liability Management of Insurance Companies Under the Low-Interest Rate Environment. China Insurance (11),13-18(2020)
7. Li Yanlin, Feng Maohan: Research on Asset-Liability Management and Large-Scale Asset Allocation of Insurance Institutions Under the Low-Interest Rate Environment. China Bond (09),53-58(2025)
8. Zhang Hao, Zhang Xiao:The Financial Risk Control of the Supply Chain of E-commerce Platform Based on Markov Model. Journal of Yunnan University of Finance and Economics 33 (02),118-126(2017)
9. Wang Xin, Wang Fang: A Review of Dynamic Pricing Strategy Based on Reinforcement Learning. Computer Applications and Software (12),1-6(2019)
10. Shen Junxin, Wang Yashi: Research on Dynamic Pricing of Data Products Based on Deep Reinforcement Learning . Systems Engineering (05),1-15(2025)