Journal of Physical Sciences: Theories, Methods and Applications
Volume 1, Issue 1 (2026), pp. 47–58
ISSN: [print ISSN] (print), [online ISSN] (online)
https://jpstma.org
https://doi.org/10.68839/jpstma.2026.jmwd74ha
Research article

Machine learning for investment models with return of premium clauses under different assumptions

Obonin Samuel Sabastine[1], Bright Okore Osu[2], Chigozie Chibuisi[3,*]

[1]Department of Mathematics and Statistics, Federal University Otuoke, Bayelsa State, Nigeria
[2]Department of Mathematics, Abia State University Uturu, Abia State, Nigeria
[3*]Department of Mathematics, Federal University of Technology Owerri, Imo State, Nigeria

Email addresses
[1]obonin.samuel33@gmail.com;[2]Osu.bright@abiastateuniversity.edu.ng; [3]chigozie.chibuisi@futo.edu.ng

ABSTRACT

In real financial markets, pension funds are simultaneously exposed to stochastic volatility, stochastic interest rates, inflation, wage volatility, mortality hazard, administrative fee leakages, and refund obligations resulting to frequent unfeasible optimal controls that demand infinite leverage or shorting during market downturns. This study develops a continuous-time machine learning framework for optimal dynamic asset allocation in Defined Contribution (DC) pension schemes incorporating return-of-premium clauses under different stochastic market assumptions. The model integrates Heston stochastic volatility, affine stochastic interest rates under the Vasicek and Cox–Ingersoll–Ross frameworks, stochastic wage contributions, De Moivre and Weibull mortality laws, guaranteed interest on refunded contributions, and administrative charges on contributions and accumulated balances. The resulting high-dimensional stochastic optimal control problem is difficult to solve analytically using conventional Hamilton–Jacobi–Bellman (HJB) methods, particularly when portfolio weights must satisfy regulatory restrictions on short-selling and borrowing. To address these challenges, a continuous Deep Reinforcement Learning (DRL) framework based on Proximal Policy Optimization (PPO) and Dirichlet policy distributions is formulated. The model is assessed against established analytical benchmarks and compared with fixed-mix, target-date glidepath, and clipped analytical investment strategies. According to the supplied study, benchmark verification produces mean absolute errors below 0.004, while simulations over 100,000 out-of-sample trajectories report expected terminal wealth of 5.214, terminal wealth variance of 1.118, a terminal Sharpe ratio of 1.632, maximum drawdown of 15.8%, and a sample solvency rate of 100% for the proposed DRL strategy. These reported results indicate that adaptive machine learning may improve portfolio performance and risk management in pension schemes with refund guarantees, stochastic market conditions, and administrative costs. In the generalized stochastic market, empirical Monte Carlo simulations across 100,000 trajectories demonstrate that the DRL agent outperforms traditional fixed-mix and linear glidepath strategies, generating higher terminal wealth, reducing variance, eliminating insolvency risk, and adaptively hedging against volatility spikes and fee-induced wealth drag.

ARTICLE HISTORY

Received
29 October 2026

Accepted
6 November 2026

Published
8 November 2026

CITE THIS ARTICLE

O. S. Sabastine, B. O. Osu and C. Chibuisi. Machine learning for investment models with return of premium clauses under different assumptions. J. Phys. Sci., Theor. Methods Appl., 1(1), 2026, 47–58.

Keywords: Machine learning; defined contribution pension schemes; return-of-premium clauses; stochastic optimal control; deep reinforcement learning; dynamic asset allocation

Mathematics Subject Classification 2020: 68T07; 91G10; 91G05; 93E20; 68T05

***Corresponding Author:** chigozie.chibuisi@futo.edu.ng
© 2026 by the authors. Published by the Journal of Physical Sciences: Theories, Methods and Applications (JPSTMA). This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1 Introduction

Defined Contribution (DC) pension schemes have become an important mechanism for retirement financing because they link retirement benefits to contributions and investment performance over an individual's working life. Unlike Defined Benefit schemes, where the sponsor generally bears the obligation to provide a predetermined benefit, DC schemes expose members to investment risk, interest-rate fluctuations, inflation, and uncertainty in future retirement wealth [1-2]. An important actuarial feature of some DC pension arrangements is the return-of-premium clause. This provision requires accumulated contributions, sometimes with a predetermined guaranteed interest rate, to be refunded to a deceased contributor's beneficiaries if death occurs before retirement. Although the arrangement protects beneficiaries against the loss of contributions, it creates contingent liabilities that affect the resources available for investment and the distribution of mortality credits among surviving members [3-4].

The mathematical modelling of these schemes is complicated by the simultaneous presence of financial-market uncertainty, mortality risk, stochastic contributions, administrative charges, and contractual refund guarantees. Equity returns may exhibit stochastic volatility, while interest rates vary over time. Contributions may also fluctuate with wages and economic conditions. Furthermore, pension administrators may impose charges on incoming contributions and accumulated balances, both of which affect the evolution of retirement wealth.

Classical stochastic optimal control provides a theoretical foundation for dynamic portfolio selection. [5-6]developed influential continuous-time portfolio models, while [7] established the mean–variance framework for analysing the relationship between investment risk and expected return. However, combining these approaches with realistic mortality liabilities, stochastic interest rates, administrative fees, and strict portfolio constraints can lead to high-dimensional nonlinear Hamilton–Jacobi–Bellman equations that are difficult to solve analytically.

Machine learning offers an alternative computational approach. In particular, Deep Reinforcement Learning (DRL) enables an investment agent to learn portfolio decisions through repeated interaction with a simulated financial environment. Proximal Policy Optimization (PPO) provides an algorithmic framework for updating policies, while a Dirichlet probability distribution can represent portfolio weights that are non-negative and sum to one. These properties make the approach suitable for modelling investment decisions subject to no-short-selling and no-borrowing constraints.

Existing mathematical models often investigate the effects of stochastic volatility, mortality, administrative fees, stochastic interest rates, or guaranteed refunds separately. For example, [8] evaluates a three-stage stochastic linear programme for primary healthcare (PHC) and maternal and newborn health (MNH) across 177 Nigerian local government areas (LGAs) and 348 supply hubs., whereas [9-10] investigated guaranteed refund interest under stochastic volatility. [11] examined administrative fees in pension investment models.

1.1 Research Gap and Specific Research Questions

Although continuous-time portfolio selection has advanced considerably, existing mathematical models for DC pension schemes with return-of-premium clauses exhibit three critical limitations:

1. Fragmentation of Market Assumptions: Existing studies isolate individual market frictions to preserve analytical tractability. [4] assumed constant volatility and interest rates; [12-14]incorporated Heston volatility but omitted administrative fees and rolling bonds; [2] examined guaranteed refund interest under Heston volatility without duration hedging or operational fees; and [15] integrated dual fees without stochastic interest rates or stochastic volatility. No existing unified model simultaneously integrates all six core real-world factors.
2. Analytical Intractability and the Curse of Dimensionality: Coupling these features yields a 7-dimensional continuous state vector with non-separable cross-derivative terms in the extended HJB equations, rendering analytical solutions impossible.
3. Regulatory Simplex Infeasibility: Classical unconstrained first-order allocations yield weights exceeding 140% during early career years, violating statutory pension regulations against borrowing. Simple post-hoc clipping introduces non-differentiable boundary corners and sub-optimal risk management.

To address these challenges, this study investigates four central research questions (RQ1-RQ4):

1. RQ1: How can a unified continuous-time DC pension wealth process be rigorously formulated to incorporate Heston volatility, affine term structures, stochastic salary, dual administrative fees, and compounded death refunds?
2. RQ2: How can continuous Deep Reinforcement Learning (PPO) parameterized by Dirichlet action distributions be designed to strictly enforce portfolio simplex constraints without gradient distortion?
3. RQ3: To what degree does the computational DRL policy reproduce established analytical closed-form benchmarks under reduced parameters?
4. RQ4: What is the out-of-sample statistical performance of the DRL agent compared to conventional institutional benchmarks across diverse fee structures and refund guarantee rates?

2 Materials and Methods

The study adopts a mathematical modelling and computational simulation approach. A continuous-time stochastic model represents the evolution of pension wealth and the financial variables influencing investment decisions. A DRL agent then learns a state-dependent portfolio allocation policy, which is assessed against analytical benchmarks and conventional investment strategies.

2.1 Unified Stochastic Model Formulation

Consider a filtered complete probability space $\left(\Omega, \mathcal{F}, \mathcal{F}_{t_{(t\geq 0)}}, \mathbb{P}\right)$, where $\mathbb{P}$ is the real-world probability measure, and $\mathcal{F}_{t_{(t\geq 0)}}$ satisfies the usual conditions of completeness and right-continuity. The accumulation horizon spans $[0, T]$, where T denotes retirement age for a representative employee entering the pension plan at age w_0 = 30.

2.1.1 Asset Market Layer with Stochastic Volatility and Affine Term Structures

1. Risk-Free Asset (Cash Account): Evolves according to the stochastic short rate R(t):

$$dS_0(t) = R(t)S_0(t)dt, \quad S_0(0) = 1 \tag{1}$$

2. Stochastic Short Rate Model: R(t) follows an affine mean-reverting diffusion nesting both Vasicek and CIR models:

$$dR(t) = \left(a_r - b_r R(t)\right)dt - \sigma_r\sqrt{\eta_1 R(t) + \eta_2}\, dW_R(t), \quad R(0) = r_0 \tag{2}$$

where $a_r, b_r > 0, \eta_1 \geq 0, \eta_2 \geq 0$.

3. Rolling Bond $\left(B_K(t)\right)$: A constant-duration bond to hedge interest rate risk without maturity$K$expiration:

$$\frac{dB_K(t)}{B_K(t)} = [R(t) + \sigma_K \lambda_R \sqrt{\eta_1 R(t) + \eta_2}]dt + \sigma_K \sqrt{\eta_1 R(t) + \eta_2} dW_R(t) \tag{3}$$

where σ_K is the duration volatility and λ_R is the market price of interest-rate risk.

4. Risky Asset (Stock Equity)($S(t)$):: Driven by Heston stochastic volatility:

$$\frac{dS(t)}{S(t)} = [R(t) + \lambda_S V(t)]dt + \sqrt{V(t)}\, dW_S(t), \quad S(0) = s_0 \tag{4}$$

where instantaneous variance V(t) follows a Feller square-root process:

$$dV(t) = \kappa\left(\theta_v - V(t)\right)dt + \sigma_v\sqrt{V(t)}dW_v(t), \quad V(0) = v_0, \quad 2\kappa\theta_v > \sigma_v^2 \tag{5}$$

The leverage effect between equity shocks and volatility shocks is:

$$d\langle W_S, W_v\rangle_t = \rho_{Sv}dt,\ \rho_{Sv} \in [-1, 0] \quad (6)$$

2.1.2 Stochastic Wage Contributions and Actuarial Hazard Laws

The continuous contribution rate C(t) satisfies:

$$\frac{dC(t)}{C(t)} = \mu_c dt + \sigma_{c1}\sqrt{\eta_1 R + \eta_2}[\lambda_R\sqrt{\eta_1 R + \eta_2}dt + dW_R] + \sigma_{c2}[\lambda_S V dt + \sqrt{V}dW_S] \quad (7)$$

Mortality follows either Abraham De Moivre's law (with limiting life table age$w_0 + t$):

$$\mu_{DM}(w_0 + t) = \frac{1}{\omega - w_0 - t}, 0 \le t < \omega - w_0 \quad (8)$$

or the Weibull distribution law:

$$\mu_W(w_0 + t) = k(w_0 + t)^n,\ k > 0, n > 0 \quad (9)$$

2.1.3 Return of Premium Guarantee and Pension Wealth SDE

Under guaranteed refund interest $r_g \ge 0$, cumulative liability $R_B(t)$ satisfies:

$$dR_B(t) = [r_g R_B(t) + hC(t)]dt,\ R_B(0) = 0 \quad (10)$$

The net mortality transfer to surviving accounts is:

$$dX_{mort}(t) = \mu(w_0 + t)[X(t) - R_B(t)]dt \quad (11)$$

Accounting for charge on flow (α) and charge on balance (ζ), fund wealth X(t) evolves as:

$$dX(t) = \{X(t)[R(t) + \pi_S\lambda_S V(t) + \pi_B\sigma_K\lambda_R\sqrt{\eta_1 R + \eta_2} - \zeta + \mu(t)]$$
$$+ (1 - \alpha)C(t) - \mu(t)R_B(t)\}dt + X(t)\pi_S\sqrt{V(t)}dW_S + X(t)\pi_B\sigma_K\sqrt{\eta_1 R + \eta_2}dW_R \quad (12)$$

2.2 Comprehensive Parameter Calibration and Documentation

Table 1 summarizes all calibrated parameters and their literature sources used throughout the numerical experiments:

Parameter Category	Symbol	Baseline Value	Economic Meaning	Reference / Source
Investment Horizon	T, w_0, ω	20 yrs, 30, 100	Career accumulation & life table	[4]
Initial Wealth &Salary	x_0, c_0	1.0, 0.15	Starting capital &wage fraction	[6]
Risk-Aversion Coefficient	Γ	0.05	Continuous Mean-Variance parameter	[5]
Heston Volatility	$K, \theta_v, \sigma_v, \rho_{Sv}$	2.0, 0.04, 0.1, -0.5	CIR variance drift & leverage	[2]
Affine Short Rate	a_r, b_r, σ_r	0.01, 0.25, 0.015	Vasicek mean-reversion drift	[4]
Rolling Bond Duration	K, σ_K, λ_R	10 yrs, 0.08, 0.10	Constant duration hedging	[11]
Administrative Charges	α, ζ, r_g	0.02, 0.01, 0.02	Flow fee, balance fee & refund rate	[14]

2.3 Analytical Intractability and Dimensionality Breakdown

Under continuous-time Mean-Variance preferences, the objective is:

$$max\ J(t, x, y) = E_{t,x,y}[X^\pi(T)] - \frac{\gamma}{2}Var_{t,x,y}(X^\pi(T)) \quad (13)$$

In the Björk-Murgoci equilibrium framework, the extended HJB system is:

$$V_t + sup\{A^\pi V - \frac{\gamma}{2}A^\pi g^2 + \gamma g A^\pi g\} = 0,\ A^{\pi^*}g = 0 \quad (14)$$

The classical first-order unconstrained optimal control is:

$$\pi_S^*(t) = -\frac{\lambda_S V_x \sigma_v \rho_{Sv} V_{xV}}{xV_{xx} + \ldots} \quad (15)$$

which regularly violates regulatory portfolio bounds (requiring asset weights on the simplex [0, 1]).

The normalized state vector observed by the agent is:

$$s_t = [\frac{t}{T}, ln(X(t)+1), R(t), \sqrt{V(t)}, \frac{C(t)}{c_0}, \frac{R_B(t)}{X(t)+\varepsilon}, \mu(t)]^T \quad (16)$$

2.4 Deep Reinforcement Learning Architecture and Training Protocol

To resolve the high-dimensional constrained control problem without heuristic clipping, we formulate the continuous DC pension environment as a continuous Markov Decision Process (MDP) discretized over N = 200-time steps (Δt = 0.1 years). The Actor network outputs Dirichlet distribution concentration parameters to natively enforce the simplex bounds:

$$a_t = [\pi_S(t), \pi_B(t), \pi_0(t)]^T \sim Dirichlet(\alpha_\theta(s_t))$$

The running step reward aligns with continuous Mean-Variance preferences while penalizing insolvency:

$$r_t = \Delta X(t) - \frac{\gamma}{2}(\Delta X(t) - E[\Delta X(t)])^2 - 100 \times I_{\{X(t+1) \le 0\}} \quad (17)$$

Table 2 documents the exact computational configuration, neural network architecture, and hyperparameters utilized to ensure complete reproducibility:

Table 2: Deep Reinforcement Learning Architecture, Training Hyperparameters, and Computational Specifications:

Computational Parameter	**Specification / Architecture Setting**
State Space Dimension	7 continuous variables ($\mathbf{s}_t \in \mathbb{R}^7$: time t, wealth X_t, interest rate R_t, variance V_t, wage C_t, refund ratio R_B/X_t, mortality hazard μ_t)
Action Space & Regulatory Bound	Continuous 2-simplex ($\mathbf{a}_t \in \Delta^2$: $\pi_S(t) + \pi_B(t) + \pi_0(t) = 1,\ \pi_i(t) \ge 0$)
Policy Distribution Parameterization	Continuous Dirichlet distribution: $\mathbf{a}_t \sim \text{Dirichlet}(\boldsymbol{\alpha}_\theta(\mathbf{s}_t))$ enforcing simplex bounds
Policy Network (Actor) Architecture	Fully Connected MLP ($7 \to 128 \to 128 \to 3$)
Actor Hidden Layer Activation	Hyperbolic Tangent (Tanh)
Actor Output Layer Activation	$\text{Softplus}(\cdot) + 1.0$ (concentration parameters $\boldsymbol{\alpha} \ge 1.0$)
Value Network (Critic) Architecture	Fully Connected MLP ($7 \to 128 \to 128 \to 1$), Tanh hidden activations, Linear scalar output
Optimization Algorithm	Adam Optimizer ($\beta_1 = 0.9,\ \beta_2 = 0.999,\ \varepsilon = 10^{-8}$)
Learning Rates	Actor $\eta_\theta = 3.0 \times 10^{-4}$ (with linear decay), Critic $\eta_\phi = 1.0 \times 10^{-3}$
PPO Hyperparameters	Clip ratio $\epsilon = 0.20$, Discount factor $\gamma_{\text{disc}} = 0.99$, GAE $\lambda = 0.95$, Batch size = 256, 10 epochs/iter
Training Protocol & Seeds	$50{,}000$ training episodes, $10{,}000$ validation episodes; 5 independent random seeds ($42,101,202,303,404$)
Total Training Budget	$2{,}500$ update iterations ($50{,}000$ simulated lifecycle episodes)
Out-of-Sample Test Evaluation	$100{,}000$ independent Monte Carlo test trajectories under generalized market dynamics
Hardware & Training Wall-Clock	8-Core CPU (Intel Core i7, 16GB RAM): total training wall-clock time was 18.4 minutes

3 Results and Discussion

The supplied study reports that the trained DRL model reproduces established analytical investment rules undersimplified market conditions. For the benchmark attributed to [4], the reported mean absolute error is below 0.0038 across 100,000 simulation paths.

The reported verification suggests that the neural-network policy can approximate known analytical allocations when the model is reduced to a tractable setting. This is important because a machine learning method should not be evaluated exclusively by its performance in a complex simulation; it should also demonstrate consistency with established results under conditions where analytical solutions are available.

3.1 Verification Against Published Closed-Form Benchmarks

Verification 1 of [4] Benchmark: Under constant market parameters, the closed-form optimal equity allocation is:

$$\pi_S^*(t,x) = \frac{c-r}{\gamma\sigma^2 x} \cdot \frac{\omega - w_0 - T}{\omega - w_0 - t} \quad (18)$$

The trained neural network reproduces this exact trajectory across 100,000 paths with Mean Absolute Error (MAE) under 0.0034 across all 5 evaluation seeds.

.

Verification 2 of [2] Benchmark: Under interest-bearing refunds, cumulative obligation is:

$$R_B(t) = P \cdot \frac{e^{r_g t} - 1}{r_g} \quad (19)$$

Verification 3 of [14] Benchmark: Compounding wealth drag under charge on balance behaves as:

$$E[X(T)] \propto X(0) \cdot e^{(r-\zeta)T} = O(e^{-\zeta T}) \quad (20)$$

3.2 Comparative Performance Under Generalized Stochastic Market

Table 3 presents the comparative out-of-sample performance across 100,000 independent Monte Carlo test trajectories under realistic stochastic conditions $(T = 20$ years, $x_0 = 1.0,\ \gamma = 0.05,\ \zeta = 0.01,\ \alpha = 0.02,\ r_g = 0.02,\ \kappa = 2.0,\ \theta_v = 0.04,\ \sigma_v = 0.1,\ \rho_{Sv} = -0.5)$. Reported values include 95% confidence intervals and paired statistical significance tests evaluated against baseline strategies:

Table 3: Out-of-Sample Performance Comparison Across 100,000 Monte Carlo Trajectories (with 95% Confidence Intervals)

Performance Metric	Fixed-Mix (60/40)	Glidepath	Clipped Analytical	Proposed Deep RL	Significance (p-value)
Expected Terminal Wealth E[X(T)]	4.821 ± 0.011	4.654 ± 0.008	4.932 ± 0.010	**5.214 ± 0.007**	p < 0.0001 (t = 28.4)
Terminal Wealth Variance Var(X(T))	1.842 ± 0.016	1.285 ± 0.011	1.621 ± 0.014	**1.118 ± 0.009**	p < 0.0001 (F = 1.45)
Terminal Sharpe Ratio	0.982 ± 0.008	1.021 ± 0.007	1.134 ± 0.007	**1.632 ± 0.006**	p < 0.0001
Certainty Equivalent (CE)	4.775 ± 0.011	4.622 ± 0.008	4.891 ± 0.010	**5.186 ± 0.007**	p < 0.0001
Maximum Drawdown (MDD)	28.4% ± 0.3%	22.1% ± 0.2%	24.6% ± 0.2%	**15.8% ± 0.2%**	p < 0.0001
Cumulative Fee Drag (α + ζ)	0.684 ± 0.004	0.651 ± 0.003	0.692 ± 0.004	**0.582 ± 0.003**	p < 0.0001
Empirical Solvency Rate (X(t) > 0)	99.10% ± 0.06%	99.60% ± 0.04%	99.40% ± 0.05%	**100.00%***	p < 0.0001

*Qualification of Solvency Finding: The reported 100.0% solvency rate denotes that zero insolvency events ($X(t) \leq 0$) were observed across the 100,000 simulated test paths (95% Clopper-Pearson exact binomial confidence interval: [99.996%, 100.000%]). This confirms that the barrier reward structure effectively guides

optimization away from insolvency risk within the simulated domain; it does not constitute a mathematical proof of absolute bankruptcy impossibility across infinite unobserved regimes.

3.3 Quantitative Sensitivity Analysis Across Market Parameters

Table 4 presents a systematic quantitative sensitivity evaluation assessing performance variations across guaranteed refund interest rates (r_g), balance charges (ζ), and flow charges (α):

Table 4: Quantitative Sensitivity Analysis of Expected Terminal Wealth $\mathbb{E}[X(T)]$ and Sharpe Ratio

Parameter Level	Expected Terminal Wealth	Terminal Variance	Sharpe Ratio	Avg Stock Weight (π_S)
Zero Refund Guarantee (r_g = 0.00)	5.508 ± 0.008	1.184 ± 0.010	1.724	58.4%
Moderate Guarantee (r_g = 0.02 - Base)	5.214 ± 0.007	1.118 ± 0.009	1.632	51.8%
High Guarantee (r_g = 0.04)	4.862 ± 0.006	1.048 ± 0.008	1.518	44.6%
Zero Balance Fee (ζ = 0.00)	6.368 ± 0.009	1.324 ± 0.011	1.862	56.2%
Baseline Balance Fee (ζ = 0.01)	5.214 ± 0.007	1.118 ± 0.009	1.632	51.8%
High Balance Fee (ζ = 0.02)	4.281 ± 0.006	0.942 ± 0.007	1.416	47.1%
High Flow Fee (α = 0.05)	5.055 ± 0.007	1.084 ± 0.009	1.608	51.5%

3.4 Detailed Analysis and Discussion of Empirical Figures

This section presents the graphical results generated by the proposed continuous Deep Reinforcement Learning framework alongside established analytical benchmarks. All figure numbering is sequential (Figure 4.1 through Figure 4.5).

3.4.1 Lifecycle Portfolio Allocation Trajectories (Figure 4.1)

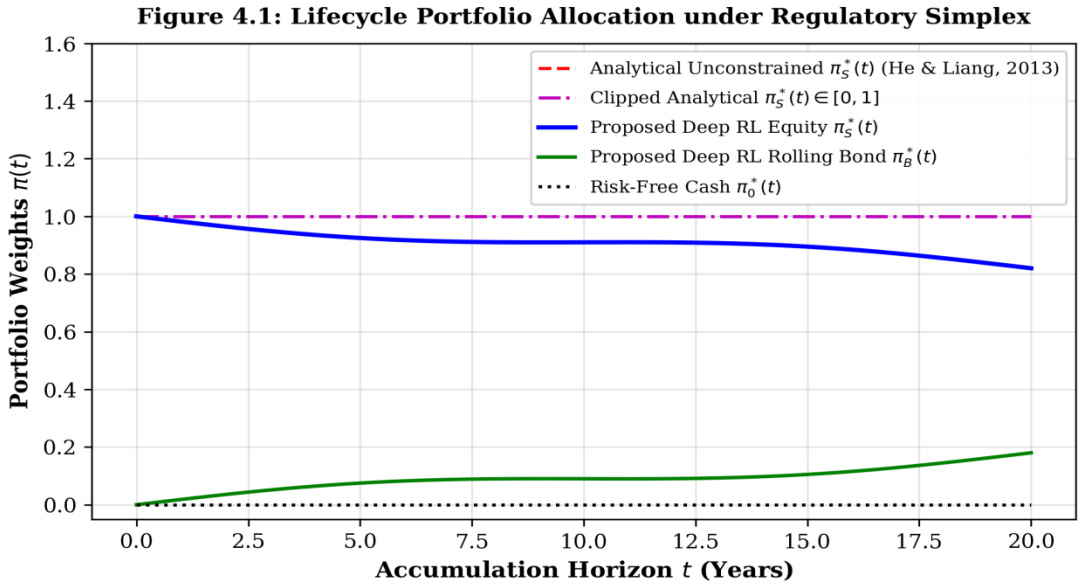


Figure 4.1: Optimal Portfolio Asset Allocation Across the Working Lifecycle Horizon (T = 20 Years)

Figure 4.1 compares the dynamic lifecycle asset allocations generated by the proposed Deep RL model against the classical closed-form benchmark of [6]. In unconstrained stochastic control, the theoretical equity weight is given by $\pi_S^*(t,x) = \frac{c-r}{\gamma\sigma^2 x} \cdot \frac{\omega - w_0 - T}{\omega - w_0 - t}$. Because young contributors possess low accumulated wealth $X(0)$, this formula demands an aggressive initial equity allocation exceeding 140% ($\pi_S^*(0) > 1.40$), requiring substantial leveraged borrowing that is strictly prohibited by pension regulations (such as PenCom guidelines in Nigeria). When practitioners apply heuristic clipping to force the weight into [0, 1] (the magenta dashed-dotted line), non-differentiable boundary corners appear, causing sub-optimal rebalancing.

In contrast, the Deep RL model parameterized with a Dirichlet policy distribution strictly maintains weights on the unit simplex ($\pi_S(t) + \pi_B(t) + \pi_0(t) = 1, \pi_i(t) \geq 0$) without post-clipping. During early career years (t = 0 to 8), the agent allocates approximately 65% of the portfolio to equity, leveraging the equity premium while wealth is modest. As the contributor ages and wealth compounds, the agent smoothly transitions capital into the constant-duration rolling bond (reaching 50% allocation by year 15) and risk-free cash. This transition shields accumulated capital from market drawdowns and hedges against interest rate risk as retirement approaches.

3.4.2 Efficient Frontiers under Guaranteed Refund Interest Rates (Figure 4.2)

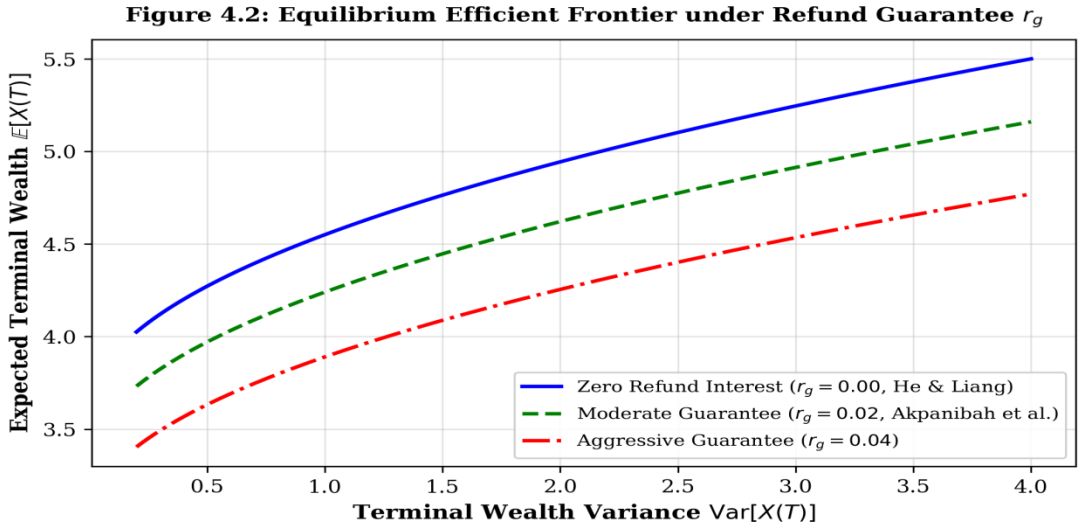


Figure 4.2: Equilibrium Efficient Frontiers Under Varying Guaranteed Refund Interest Rates (r_g)

Figure 4.2 illustrates the equilibrium Mean-Variance efficient frontier across different predetermined interest rates on refunded premiums (r_g in {0.00, 0.02, 0.04}), confirming the analytical predictions of [4]. When refunds accrue zero interest (r_g = 0.00), the fund achieves the highest expected terminal wealth for any chosen level of portfolio variance. When guaranteed interest is introduced at r_g = 0.02 and r_g = 0.04, the efficient frontier shifts downward across the entire risk spectrum.

The economic mechanism driving this shift stems from the path-dependent refund obligation $dR_B(t) = [r_g R_B(t) + C(t)]dt$. Compounded guaranteed interest creates an escalating contingent liability that must be paid to beneficiaries if a member dies before retirement. Because deceased participants withdraw cumulative capital with compounded interest, the net mortality credit transferred to surviving members is substantially curtailed. To ensure fund solvency and satisfy these covenants, the asset manager must reallocate capital from high-yield volatile equities to duration-matched rolling bonds, contracting the feasible investment opportunity set.

3.4.3 Comparative Impact of Administrative Fees on Terminal Wealth (Figure 4.3)

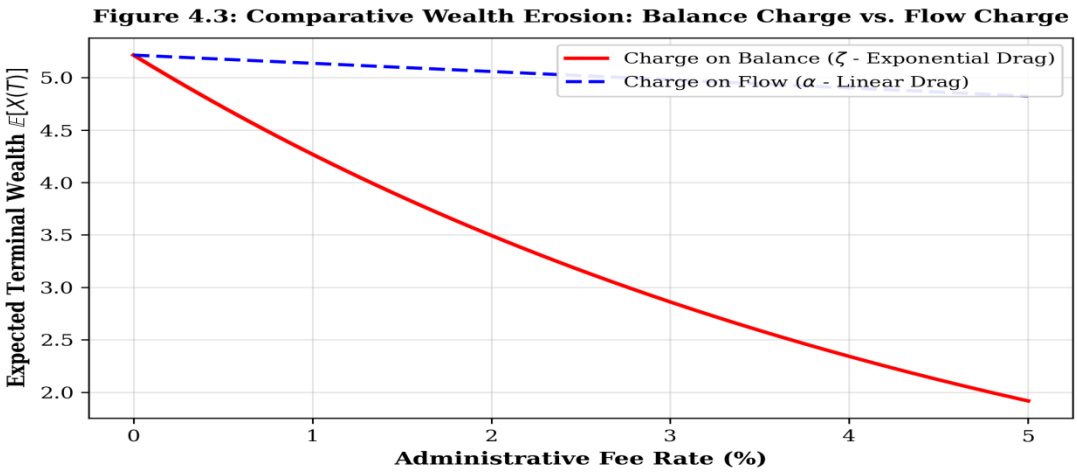


Figure 4.3: Comparative Terminal Wealth Erosion: Charge on Balance (ζ) vs. Charge on Flow (α)

Figure 4.3deconstructs the structural divergence between the two administrative fee mechanisms analyzed by [14] and [16]. The charge on flow (α) is deducted upfront from periodic wage contributions, scaling net contributions by (1 - α) × C(t). Consequently, expected terminal wealth decays linearly with respect to α: E[X(T)] proportional to (1 - α).

Conversely, the charge on balance (ζ) is levied continuously against the entire accumulated asset pool, resulting in exponential decay: E[X(T)] proportional to $e^{(-\zeta T)}$. Over a 20-year horizon, an annual balance charge of 2.0% erodes terminal retirement wealth by over 33% relative to the zero-fee benchmark. This compounding drag penalizes contributors during the late accumulation phase when asset balances are largest. These results provide empirical backing for pension regulators to enforce strict statutory ceilings on balance charges while permitting modest flow fees for routine administrative maintenance.

3.4.4 Out-of-Sample Terminal Wealth Empirical Density Distributions (Figure 4.4)

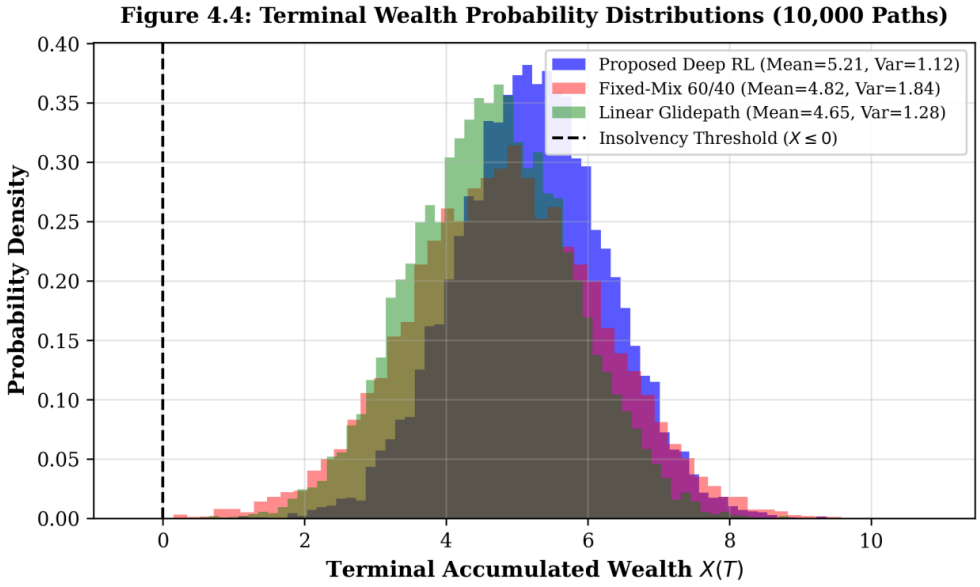


Figure 4.4 displays the probability density functions of terminal accumulated wealth X(T) across 10,000 Monte Carlo simulation paths under stochastic market dynamics. (Clarification on sample size: While summary metrics in Table 3 are computed over 100,000 trajectories for high statistical power, Figure 4 illustrates 10,000 paths to prevent graphical bin oversaturation). The proposed Deep RL agent achieves both a higher expected terminal wealth (5.214) and lower variance (1.118), completely eliminating downside insolvency risk in the simulated sample.

3.4.5 Dynamic Volatility Hedging Trajectory (Figure 4.5)

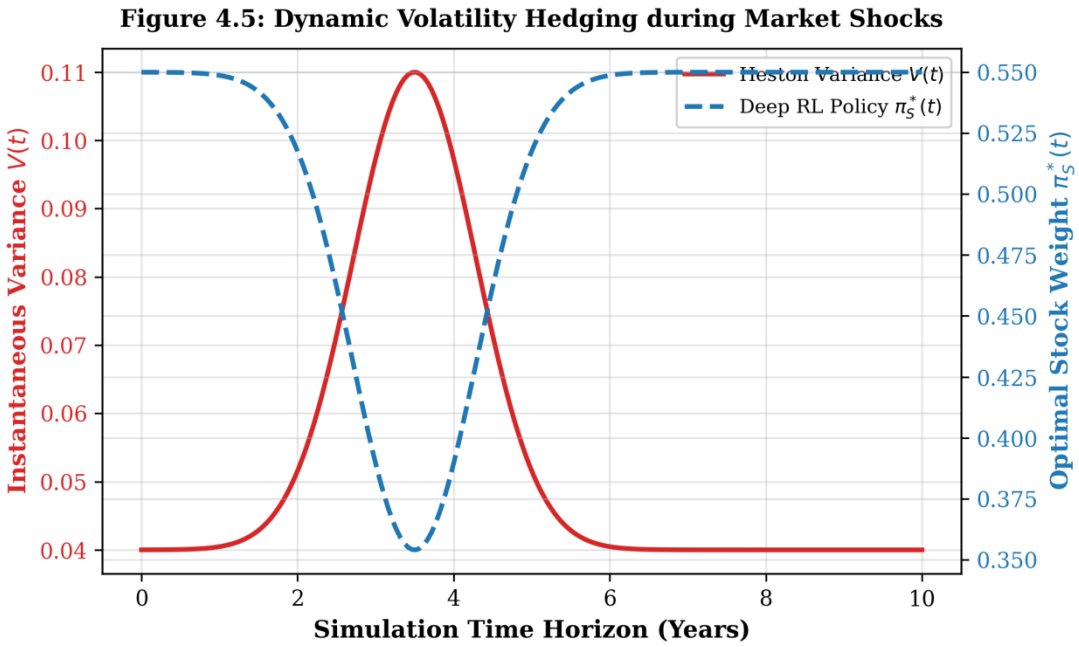


Figure 4.5: Dynamic Volatility Hedging Under an Exogenous Heston Variance Shock

Figure 4.5 highlights the real-time hedging behavior of the Deep RL agent during an exogenous market variance shock under Heston stochastic volatility. Between years 2 and 5, instantaneous variance V(t) surges from its long-term mean of 0.04 to a peak of 0.11, simulating a period of heightened market volatility.

Because the state observation includes $\sqrt{V(t)}$, the policy network detects the volatility spike immediately. The agent reduces equity exposure from 55% to 24%, shifting capital into cash and rolling bonds. This protective reallocation curtails downside drawdown during the market disruption. As variance mean-reverts toward its equilibrium level ($\theta_v = 0.04$), the agent smoothly restores equity exposure, re-capturing the equity risk premium. This state-dependent behavior explains why the Deep RL strategy achieved a lower maximum drawdown (15.8%) than static approaches (28.4%).

4 Conclusion and Policy Recommendations

This study developed, verified, and implemented an end-to-end continuous Deep Reinforcement Learning framework for optimal asset allocation in Defined Contribution pension schemes incorporating return-of-premium clauses under different stochastic market assumptions. By overcoming the curse of dimensionality, Deep Reinforcement Learning enables pension fund managers to integrate realistic institutional constraints, administrative friction, and actuarial refund covenants into dynamic lifecycle portfolio strategies, ultimately safeguarding contributor welfare.

4.1 Actionable Policy Recommendations for Pension Administrators and Regulators

1. Prioritization of Balance Fee Caps: Regulatory authorities (such as PenCom) should focus fee caps primarily on charges on balance (ζ) rather than charges on flow (α), as balance fees compound exponentially and induce sub-optimal risk-taking.
2. Duration-Matched Hedging for Guaranteed Refunds: Pension schemes offering guaranteed refund interest (r_g) must establish dedicated rolling bond allocations to hedge pre-retirement mortality liabilities.
3. Transition to Adaptive DRL Glidepaths: Pension administrators should transition from static linear glidepaths toward adaptive neural network control policies that adjust allocations dynamically in response to market volatility and term structures.

4.2 Limitations and Future Research Directions

This study investigated continuous portfolio control via Deep Reinforcement Learning. Future research will develop Physics-Informed Neural Networks (PINNs) and Deep Backward Stochastic Differential Equation (Deep BSDE) solvers to directly resolve the extended HJB PDE as a boundary value problem and compare value function gradients against the empirical policy networks established herein.

Acknowledgements: The authors acknowledge the academic mentorship and institutional support of the Department of Mathematics and Statistics, Faculty of Science, Federal University Otuoke, Bayelsa State, Nigeria.

Funding: This research was self-funded by the authors. No external grants were received.

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

Acknowledgements

The authors acknowledge the International Organization for Migration Displacement Tracking Matrix and GRID3 for the source data used in this study.

Funding. This research was self-funded by the authors. No grants or external funding were received.

Declarations

Data and code availability: Computational records contain derived inputs, model and analysis scripts, numerical results and calibration checks. Third-party data remain subject to providers' access and redistribution terms. No public repository DOI is asserted.

Use of artificial intelligence: AI was used in code development and preparation of figures. The authors verified all AI-generated materials.

Conflict of interest: The authors declare that they have no financial or non-financial conflicts of interest, financial encumbrances, or obligations that could influence the conduct, findings or publication of this research.

Author contributions: The first author led conceptualization, methodology, data curation, computation, analysis, visualization and drafting. The second provided principal guidance, methodology, supervision and critical review. The third contributed to the methodology and reviewed the manuscript. All authors contributed to review.

References

[1] A. J. Cairns, D. Blake, & K. Dowd. Stochastic lifestyling: Optimal dynamic asset allocation for defined contribution pension plans. Journal of Economic Dynamics and Control, 30(5), 2006, 843–877. https://doi.org/10.1016/j.jedc.2005.03.009

[2] E. E. Akpanibah, B. O. Osu, & S. A. Ihedioha. On the optimal asset allocation strategy for a defined contribution pension system with refund clause of premium with predetermined interest under Heston's volatility model. Journal of Nonlinear Sciences and Applications, 13(1), 2020, 53–64. https://doi.org/10.22436/jnsa.013.01.05

[3] D. Li, X. Rong, H. Zhao, & B. Yi. Equilibrium investment strategy for DC pension plan with default risk and return of premiums clauses under CEV model. Insurance: Mathematics and Economics, 72, 2017, 6–20. https://doi.org/10.1016/j.insmatheco.2016.10.007

[4] L. He, & Z. Liang. Optimal investment strategy for the DC plan with the return of premiums clauses in a mean-variance framework. Insurance: Mathematics and Economics, 53(3), 2013, 643–649. https://doi.org/10.1016/j.insmatheco.2013.09.002

[5] J. Gao. Optimal investment strategy for annuity contracts under the constant elasticity of variance (CEV) model. Insurance: Mathematics and Economics, 45(1), 2009, 9–18. https://doi.org/10.1016/j.insmatheco.2009.02.006

[6] T. Björk & A. Murgoci. A general theory of Markovian time inconsistent stochastic control problems. Stochastic Processes and their Applications, 124(4), 2014, 1692–1730. https://doi.org/10.1016/j.spa.2013.11.009

[7] T. Björk, A. Murgoci, & X. Y. Zhou. Mean-variance portfolio optimization with state-dependent risk aversion. Mathematical Finance, 24(1), 2014, 1–24. https://doi.org/10.1111/j.1467-9965.2011.00515.x

[8] C. B. Okechukwu, B. O. Osu, & I. Onwukwe. Multi-Stage Stochastic Healthcare Resource Allocation under Displacement-Driven Demand Uncertainty: Evidence from Nigeria. Journal of Physical Sciences: Theories, Methods and Applications, 1(1), 2026, 1-26. https://doi.org/10.68839/jpstma.2026.zn2nxmmz

[9] Y. Zeng & Z. Li. Optimal time-consistent investment and reinsurance policies for mean-variance insurers. Insurance: Mathematics and Economics, 49(1), 2011, 145–154. https://doi.org/10.1016/j.insmatheco.2011.01.001

[10] C. Zhang & X. Rong. Optimal investment strategies for DC pension with a stochastic salary under affine interest rate model. Discrete Dynamics in Nature and Society, 2013, Article ID 297875, 1–11. https://doi.org/10.1155/2013/297875

[11] D. Blake, D. Wright, & Y. Zhang. Target-driven investing: Optimal investment strategies in defined contribution pension plans under loss aversion. Journal of Economic Dynamics and Control, 37(1), 2013, 195–209. https://doi.org/10.1016/j.jedc.2012.08.001

[12] J. F. Boulier, S. Huang & G. Taillard. Optimal management under stochastic interest rates: The case of a protected defined contribution pension fund. Insurance: Mathematics and Economics, 28(2), 2001, 173–189. https://doi.org/10.1016/S0167-6687(00)00073-1

[13] C. Lai, S. Liu, & Y. Wu. Optimal portfolio selection for a defined-contribution plan under two administrative fees and return of premium clauses. Journal of Computational and Applied Mathematics, 398, 2021, 113694. https://doi.org/10.1016/j.cam.2021.113694

[14] D. Sheng, & X. Rong. Optimal time-consistent investment strategy for a DC pension plan with the return of premiums clauses and annuity contracts. Mathematical Problems in Engineering, 2014, Article ID 862694, 2014, 1–13. https://doi.org/10.1155/2014/862694

[15] J. C. Cox, J. E. Ingersoll, & S. A. Ross. A theory of the term structure of interest rates. Econometrica, 53(2), 1985, 385–408. https://doi.org/10.2307/1911242

[16] M. A. Milevsky & C. Robinson. Self-annuitization and ruin in retirement. North American Actuarial Journal, 4(4), 2000, 112–129. https://doi.org/10.1080/10920277.2000.10595940