A Sequential Decision-Making Framework for Life-Cycle Retirement Pension Optimization Using Deep Reinforcement Learning

Haizhou Cui[1] Bonsoo Koo[1] Anthanasios Pantelous[1] Samuel Vigne[2]

[1]Monash University, Department of Econometrics and Business Statistics

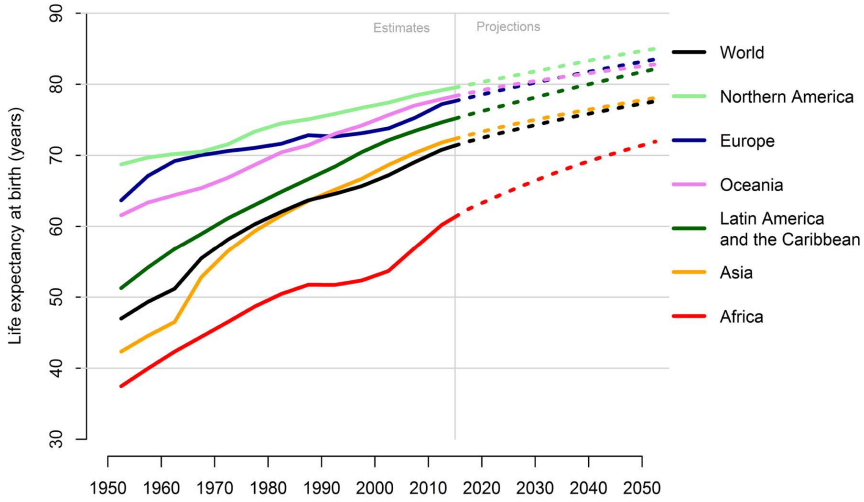
[2]LUISS Business School, Villa Blanc Campus, Rome, Italy

33rd Colloquium on Pensions and Retirement Research

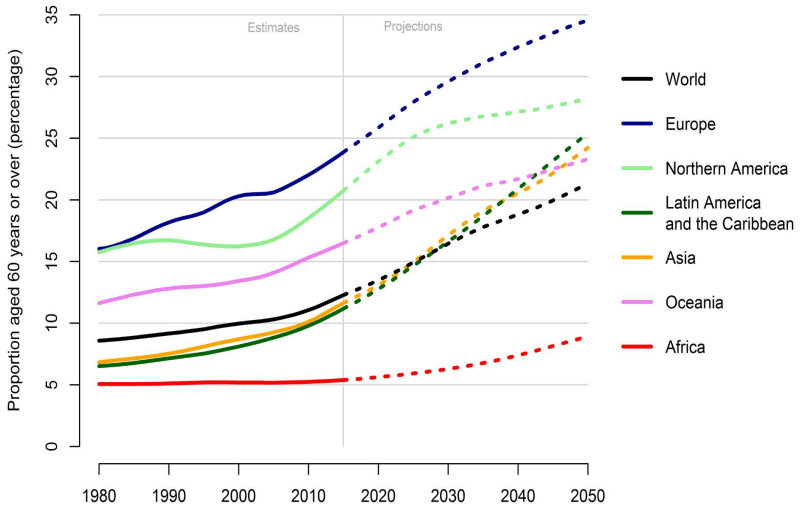
Contents

Global Situation

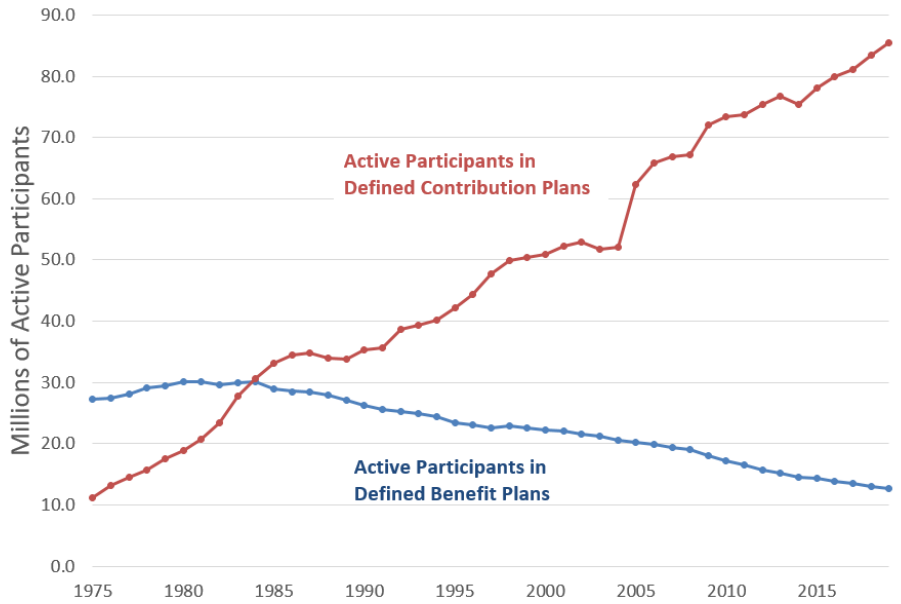
Life Expectancy[1]



Elderly Population[1]



Shift from DB to DC[2]



Source:

[1] WHO [WHO Report]

[2] U.S. Department of Labor, Employee Benefits Security Administration [EBSA Report]

Motivation

The shift from DB to DC means:

- No longer predetermined income stream, but uncertain benefits for retirees
- Challenges:
  - (1) How to make prudent personal retirement wealth management?
  - (2) How much should I consume, invest, or take in reverse mortgage at different retirement stages? How much should I leave as bequest?
  - (3) What retirement products should I buy, when, and how much, given financial risks?
  - (4) How can I address potential risks when making decisions?

Research questions:

Can we have a comprehensive model framework to provide personalized optimal financial decisions for retirement wealth management? Embrace today's AI technology.

Classical Framework

A classical **utility-based life-cycle model** seeks optimal post-retirement decisions to maximize utility. [Koo et al., 2022].

- Pension decisions are made from t_0 (**retirement start**) to T (**terminal age**).
- At each $t = t_0, \ldots, T$, the **control** π_t determines consumption ratio (α_t), investment ratio (δ_t), and reverse mortgage taking ratio (ℓ_t).
- **Financial state** X_t (e.g., liquid wealth W_t, consumption C_t, house H_t and Loan L_t) is realized each period; **Pension entitlement** $P_t = P(X_t)$ depends on it..
- **Reward** $R(X_t, \pi_t)$ reflects utility from either consumption or bequest, given current state and decision.

Classical Framework

Dynamics

- Consumption dynamics:

$$C_t = \alpha_t W_t + \ell_t H_t + P_t \tag{1}$$

- Wealth dynamics:

$$W_{t+1} = (W_t - \alpha_t W_t) e^{\delta_t Z_t + (1-\delta_t) r} \tag{2}$$

- House dynamics:

$$H_{t+1} = H_t e^h \tag{3}$$

- Loan dynamics:

$$L_{t+1} = (L_t + \ell_t H_t) e^k \tag{4}$$

- House growth rate (h), reverse mortgage loan interest rate (k), risk-free rate (r), risk asset return $Z_t \sim N(\mu, \sigma^2)$.
- Bequest:

$$B_T = W_T + H_T - L_T \tag{5}$$

Objective:

- Reward (HARA utility), consumption and bequest:

$$R_t(X_t, \pi_t) = U_c(C_t) = (C_t - \overline{C})^\gamma/\gamma; \quad t < T$$
$$R_T(X_T) = U_B(B_T) = \left(\frac{\theta}{1-\theta}\right)^{1-\gamma}; \quad t = T \tag{6}$$

- Objective function:

$$V^*_{t_0}(X_{t_0}) := \sup_\pi \mathbb{E}^\pi_{t_0}\left[\sum_{t=t_0}^{T-1} \beta^{t-t_0} R_t(X_t, \pi_t) + \beta^{T-t_0} R_T(X_T)\right] \tag{7}$$

Classical Framework

Potential challenges:

- No analytical solution.
- Curse of dimensionality.
- Solved numerically via function approximation and space discretization [Koo et al., 2022]
- Extremely high computational cost.

Limitations:

- Lack of consideration for evolving risk preferences, an important factor influencing decisions [Chen et al., 2024, Wang et al., 2024].
- Neglect influence of dynamic economic environments on decisions.
- Overlook how individual expectations shape financial behavior.

Sequential Framework

To address the aforementioned limitations, we develop a sequential decision-making framework, which is upon the **Markov Decision Process (MDP)**.

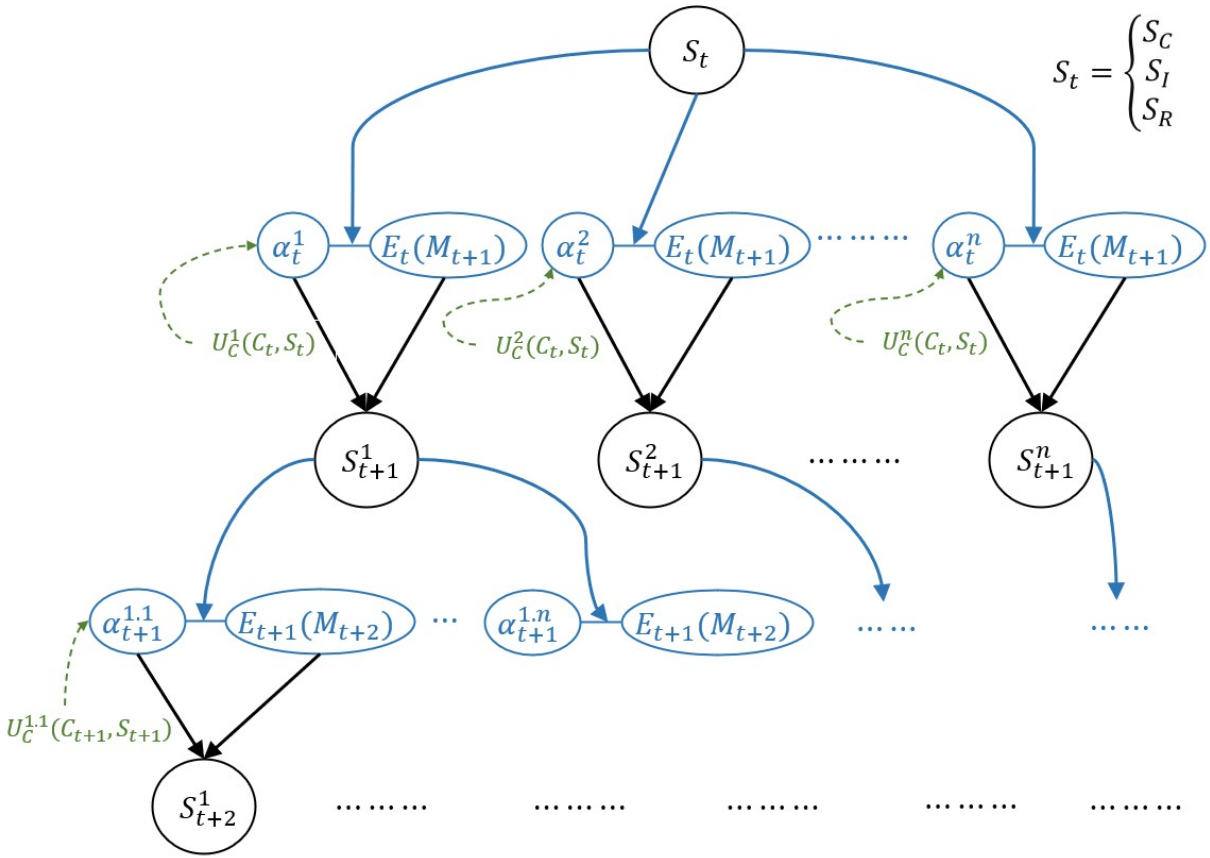


Figure 1: Risk preference state transition defined by Markov Decision Process

Sequential Framework

Risk preference state S_t at t is shaped by:

- Expectation on **economic condition** M_t at t.
- The most recent decision.

$$S_{t+1} = f\left(\mathbb{E}_t(M_{t+1}), \alpha_t\right) = \begin{cases} S_C, & \text{if } C \text{ (Conservative)} \\ S_I, & \text{if } I \text{ (Intermediate)} \\ S_R, & \text{if } R \text{ (Aggressive)} \end{cases} \tag{8}$$

Reflected on **Utility function dynamics** $(\gamma_c < \gamma_i < \gamma_r < 0)$:

$$\gamma_t(S_t) = \begin{cases} \gamma_c, & \text{if } S_t = S_C \\ \gamma_i, & \text{if } S_t = S_I \\ \gamma_r, & \text{if } S_t = S_R \end{cases} \tag{9}$$

Reward at non-terminal point:

$$R_t(X_t, \pi_t, S_t) = U_c(C_t, S_t) = \frac{(C_t - \bar{C})^{\gamma_t(S_t)}}{\gamma_t(S_t)} = \frac{(\alpha_t W_t + \ell_t H_t + P(X_t) - \bar{C})^{\gamma_t(S_t)}}{\gamma_t(S_t)} \tag{10}$$

Sequential Framework

Bequest preference is also dynamic and personalized.

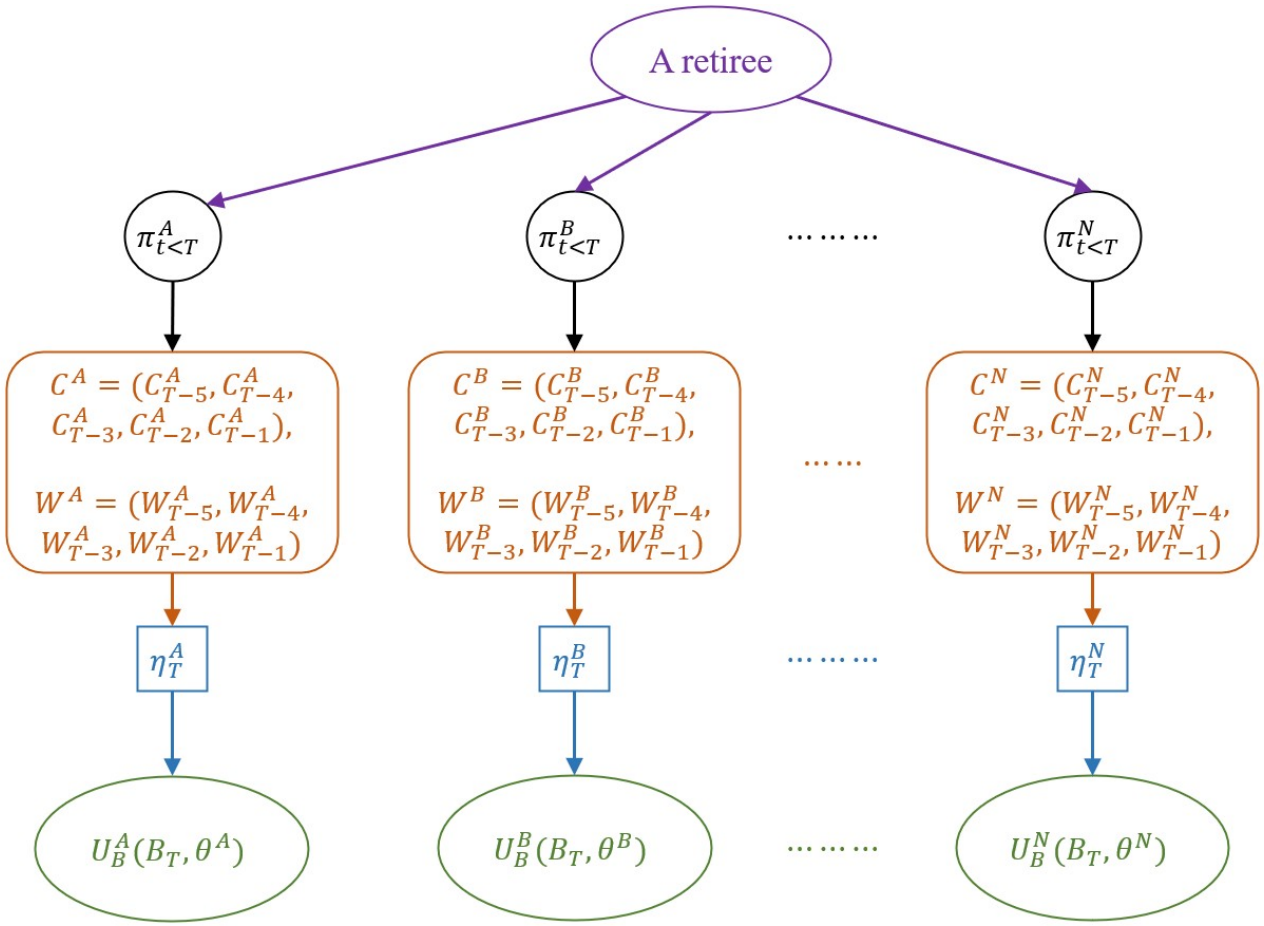


Figure 2: Principle of bequest preference determination.

Sequential Framework

Bequest preference is shaped by propensity to bequest the offspring

- Bequest motivation is proxied by the ratio of cumulative liquid wealth to consumption in the final periods of life (e.g., last five years) [Kraft et al., 2022].

$$\eta_T = \frac{\sum_{i=1}^{5} W_{T-i}}{\sum_{i=1}^{5} C_{T-i}} \tag{11}$$

- Liquid wealth at the end of life and cumulative late-life consumption result from the control sequence $\pi_t(\alpha_t, \delta_t, \ell_t)$.
- Reflected on bequest utility function form ($0 < \theta_S < \theta_M < \theta_L < 1$):

$$\theta_T(\eta_T) = \begin{cases} \theta_S, & \text{if } 0 < \eta_T \leq \eta_S, \quad S \text{ (Small)} \\ \theta_M, & \text{if } \eta_S < \eta_T \leq \eta_M, \quad M \text{ (Medium)} \\ \theta_L, & \text{if } \eta_M < \eta_T \leq 1, \quad L \text{ (Large)} \end{cases} \tag{12}$$

Sequential Framework

Reward at terminal point:

$$R_T(X_T, \eta_T, S_T) = U_B(B_T, \eta_T, S_T) = \left(\frac{\theta_T(\eta_T)}{1-\theta_T(\eta_T)}\right)^{1-\gamma_T(S_T)} \cdot \frac{B_T^{\gamma_T(S_T)}}{\gamma_T(S_T)} \tag{13}$$

(where $B_T = W_T + H_T - L_T$ is the terminal bequest.)

Final objective function:

$$V_{t_0}^* = \sup_{\pi} \mathbb{E}^{\pi_{t_0}} \left[\beta^{T-t_0} R_T(X_T, \eta_T, S_T) + \sum_{t=t_0}^{T-1} \beta^{t-t_0} R_t(X_t, \pi_t, S_t)\right] \tag{14}$$

- Increased model complexity and computational cost.
- Severe curse of dimensionality due to expanded state space.
- Numerical methods fail to obtain optimal controls.

Deep reinforcement learning method

Deep reinforcement learning algorithm as a Solution Strategy for Control Optimization:

- High computational efficiency in solving complex control tasks [Wang et al., 2023]
- Strong adaptability to dynamic, high-dimensional environments [Powell, 2022]
- Naturally handles continuous state and action spaces (e.g., deterministic policy gradient) [Silver et al., 2014, Lillicrap et al., 2019]
- Demonstrated robustness in financial optimization applications [Zhang et al., 2024]

Deep reinforcement learning method

Deep Deterministic Policy Gradient (DDPG) Architecture for Life-Cycle Optimization

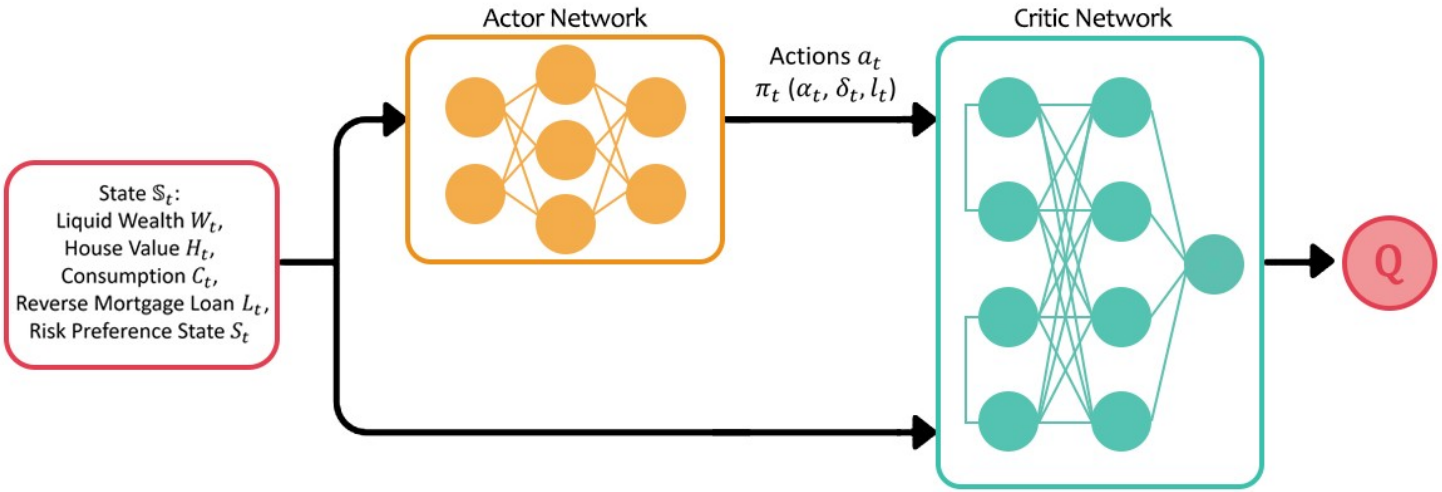


Figure 3: Actor-Critic network system of DDPG algorithm for the life-cycle model with sequential framework.

$$Q(\mathbb{S}, a) = \mathbb{E}\left[\sum_{\tau=t_0}^{T} \beta^{T-\tau} R_\tau \mid \mathbb{S}_0 = \mathbb{S}, a_0 = a\right] \tag{15}$$

$$a^* = \pi^*(\alpha, \delta, \ell) = \mu^*(\mathbb{S}) = \arg\max_a Q(\mathbb{S}, a) \tag{16}$$

A replay buffer stores past experiences for mini-batch training during the learning loop until convergence.

Deep reinforcement learning method

- Critic Network – (Map $(\mathbb{S}, a) \Rightarrow Q$)

$$
\begin{aligned}
G_t &= R_t + \beta Q'_\omega(\mathbb{S}_{t+1}, \mu'_\lambda(\mathbb{S}_{t+1})) \\
L(\omega) &= \mathbb{E}_{(S_t, a_t, R_t, S_{t+1}) \sim v_{\mu_\lambda}} \left[(Q_\omega(\mathbb{S}_t, a_t) - G_t)^2 \right] \\
\omega^* &= \min_\omega \left(Q_\omega(\mathbb{S}, \mu_\lambda(\mathbb{S})) \right) \\
\omega &\leftarrow \omega - \eta_c \nabla_\omega L(\omega)
\end{aligned} \tag{17}
$$

- Actor Network – (Map $\mathbb{S} \Rightarrow a$) – Deterministic policy gradient theorem

$$
\begin{aligned}
Q(\mathbb{S}, a) \sim \mathcal{J}(\mu_\lambda) &= \int_{\mathcal{S}} \nu^{\mu_\lambda}(\mathbb{S}) R(\mathbb{S}, \mu_\lambda(\mathbb{S})) d\mathbb{S} = \mathbb{E}_{\mathbb{S} \sim \nu^{\mu_\lambda}} \left[R(\mathbb{S}, \mu_\lambda(\mathbb{S})) \right] \\
\nabla_\lambda \mathcal{J}(\mu_\lambda) &= \mathbb{E}_{\mathbb{S} \sim \nu^{\mu_\lambda}} \left[\nabla_\lambda \mu_\lambda(\mathbb{S}) \nabla_a Q(\mathbb{S}, a) \Big|_{a = \mu_\lambda(\mathbb{S})} \right] \\
\lambda^* &= \arg\max_\lambda Q_{\omega^*}(\mathbb{S}, \mu_\lambda(\mathbb{S})) \sim \arg\max_\lambda \mathcal{J}(\mu_\lambda) \\
\lambda &\leftarrow \lambda + \eta_A \nabla_\lambda \mathcal{J}(\mu_\lambda)
\end{aligned} \tag{18}
$$

- With λ^* and ω^* the optimal Actor-Critic network system will output the optimal action sequence given any state and profile.

Dynamic economic environment and expectation

- Financial state $X_t = (C_t, W_t, H_t, L_t)$ evolves in dynamic economic environment M_t
- Conditional expectation as a key determinant of risk-preference state transition:

$$E_t(M_{t+1}) = f_{\text{LSTM}}(\mathcal{H}_t), \quad \mathcal{H}_t = \{M_t, M_{t-1}, M_{t-2}, \dots\} \tag{19}$$

- Risk preference state transition principle, by statistical learning:

$$S_t = \begin{cases} S_C, & \text{if } 0 < \alpha_t \leq \alpha_C, \quad C \text{ (Conservative)} \\ S_I, & \text{if } \alpha_C < \alpha_t \leq \alpha_I, \quad I \text{ (Intermediate)} \\ S_R, & \text{if } \alpha_I < \alpha_t \leq 1, \quad R \text{ (Aggressive)} \end{cases} \tag{20}$$

$$S_t = SVC(M_t, \alpha_t) \tag{21}$$

- Risk preference state transition determination in MDP:

$$S_{t+1} = SVC(E_t(M_{t+1}), \alpha_t) = SVC(f_{\text{LSTM}}(\mathcal{H}_t), \alpha_t) \tag{22}$$

Application

- **Data source for risk preference statistical learning:**
  - ABS Survey of Income and Housing (SIH)
  - Household Expenditure Survey (HES): 2003–2004, 2009–2010, 2015–2016
- **Model calibration:**
  - Macroeconomic: New Keynesian DSGE model
  - Financial: CAPM
  - Estimation method: Maximum Likelihood Estimation (MLE)

The calibrated dynamic system is:

$$
\begin{aligned}
y_t &= \mathbb{E}_t[y_{t+1}] - (1/\xi) r_t + u_t^{IS}, \\
p_t &= \varphi \mathbb{E}_t[p_{t+1}] + \kappa y_t + u_t^{p}, \\
r_t &= \phi_p \mathbb{E}_t[p_{t+1}] + \phi_y \mathbb{E}_t[y_{t+1}] + u_t^{MP}, \\
Z_t &= r_t + \beta_0 (Z_t - r_t) + \varepsilon_t
\end{aligned}
$$

Notes: y_t: GDP growth; p_t: inflation (WDI); r_t: interest rate (CRT from RBA); Z_t: expected return (ASX 200)

Application

- Applied to Australian pension rules: Age Pension (AP) entitlement rules and Pension Loan Scheme (PLS) principle for reverse mortgage.

$$P(X_t) = \max\left(0, \min\left(P^d_{\max}, \min\left(P_{A,t}, P_{I,t}\right)\right)\right)$$

$$\ell_t \in \left[0,\ \max\left(0, \min\left(\tilde{L}_t - L_t,\ 1.5P^d_{\max} - P(X_t)\right)\right)/H_t\right] \tag{23}$$

- Calibration results for DSGE and CAPM model:

ξ	φ	κ	ϕ_p	ϕ_y	β_0
5.0000	0.9000	0.1018	1.0000	0.6393	1.5081

Table 1: DSGE and CAPM model parameters calibrated based on real Australian economic data

- Wealth evolution based on the calibrated dynamic economic sequence.

$$W_{t+1} = (1 - \alpha_t)\, W_t \cdot (1 + \delta_t Z_t + (1 - \delta_t) r_t) \tag{24}$$

$\eta_S = 2$, $\eta_M = 3$; $h = 0.0177$ and $e^k = 1.026$ [Koo et al., 2022]

Benchmark Case

- Profile: $X_{t_0} = (W_{t_0}, H_{t_0}, L_{t_0}) = (\text{A\$360k}, \text{A\$1.5 million}, 0)$
- Model parameters based on the calibration result of [Koo et al., 2022] with modification

Θ	γ_C	γ_I	γ_R	θ_S	θ_M	θ_L	$\bar{C}$ (A\$10,000)
Value	-2.3	-1.99	-1.7	0.78	0.85	0.92	3.21

Table 2: Utility parameters for the sequential life-cycle model of benchmark case

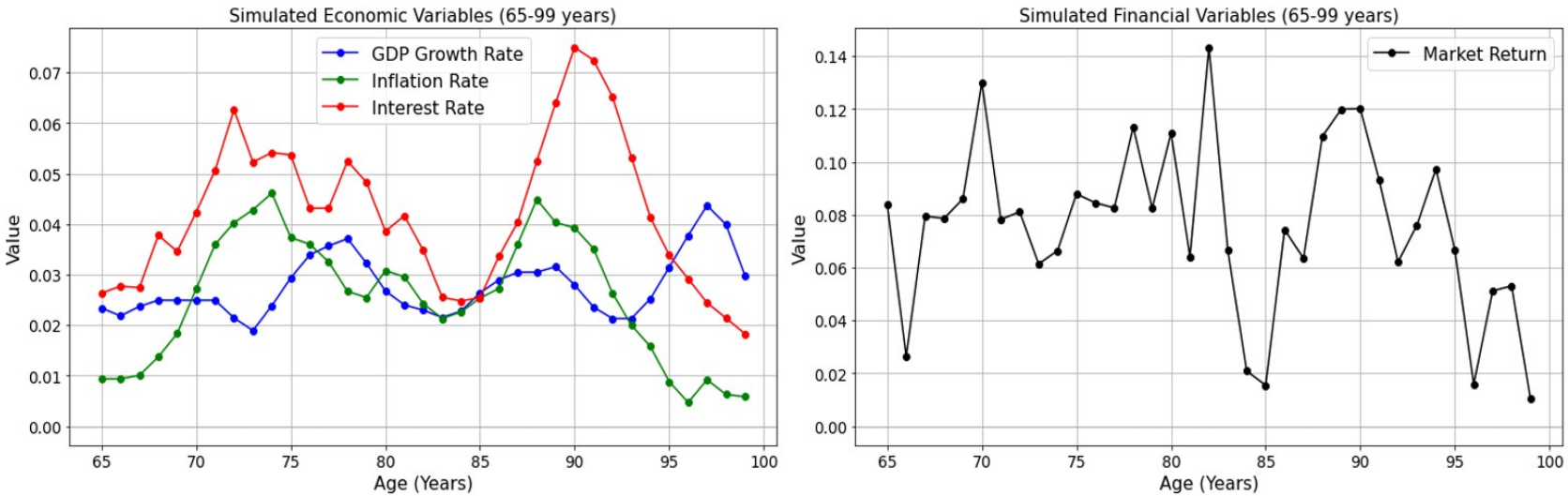
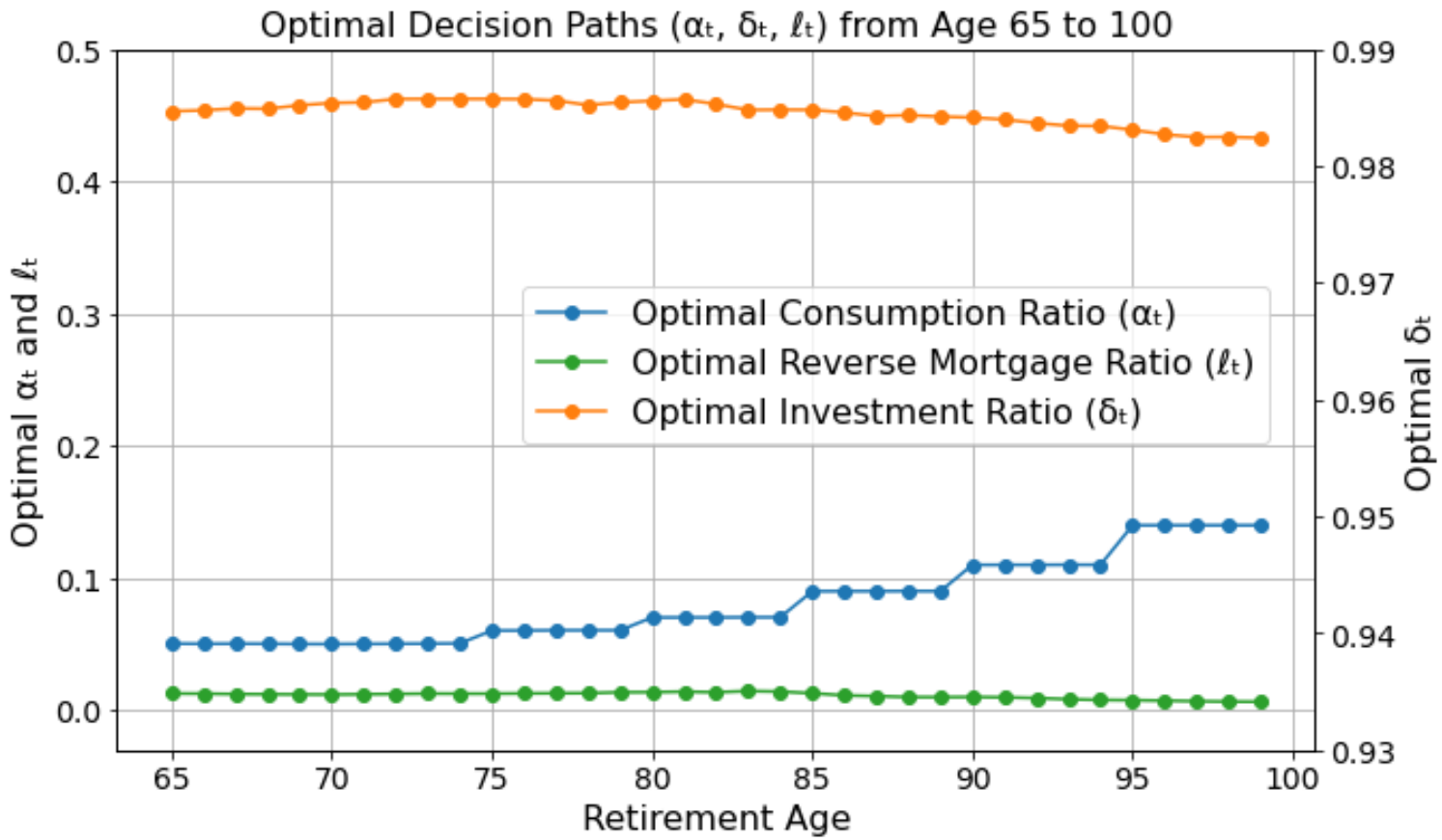


Figure 4: Dynamic economic environment for retiree decision-making benchmark case

Benchmark Case

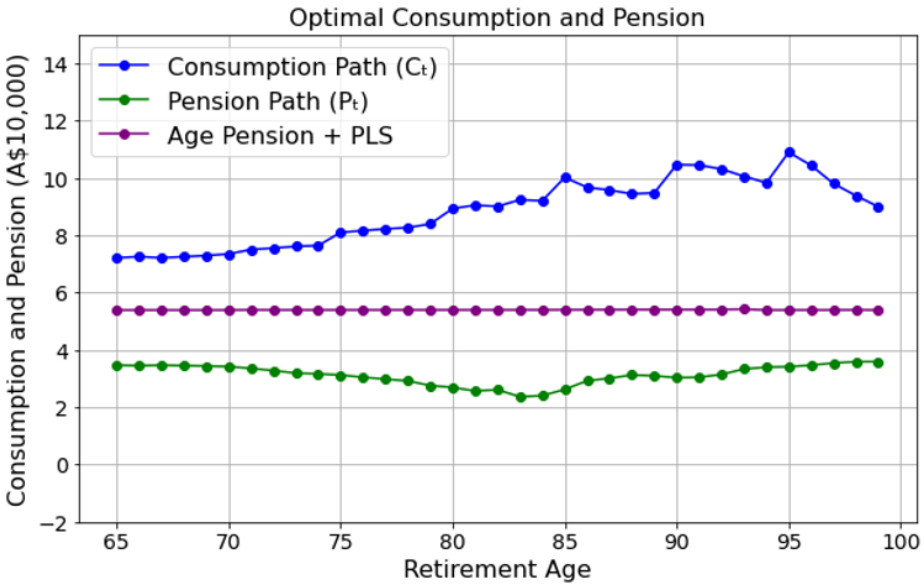
Optimal control $\pi_t = (\alpha_t, \delta_t, \ell_t)$ – Stable and steady pattern; life-cycle utility maximization

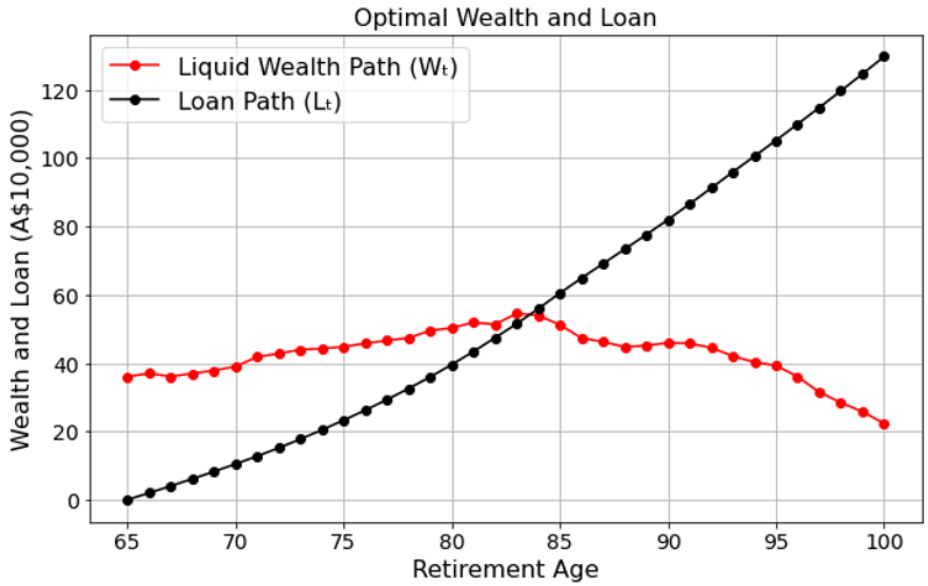


Benchmark Case

Optimal Consumption, Pension, Wealth and Loan path:

- Stable and steady consumption–wealth paths;
- Consistent with life-cycle utility maximization;
- Avoids large utility-reducing fluctuations [Modigliani, 1986]
- Moderate increase in consumption over time
- Late-stage consumption decline for bequest motive

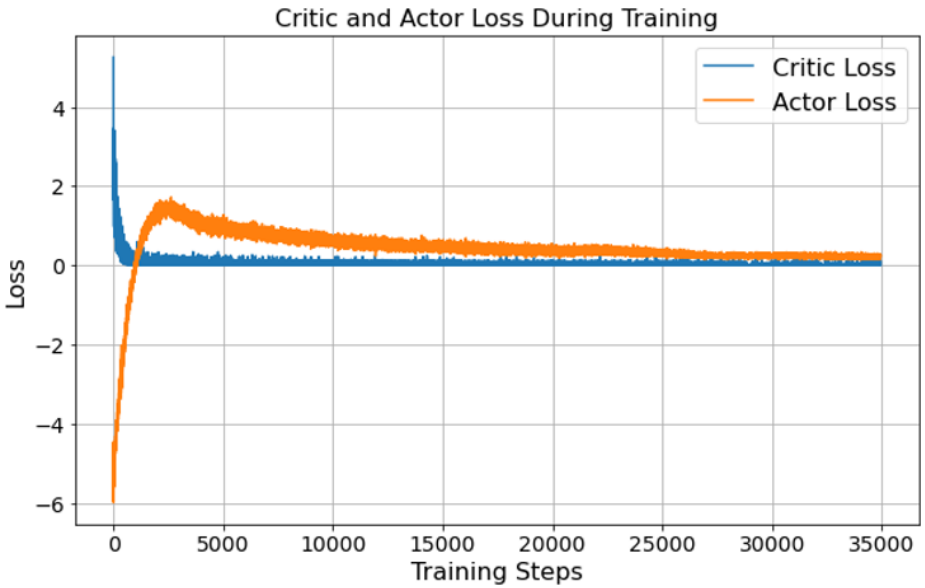


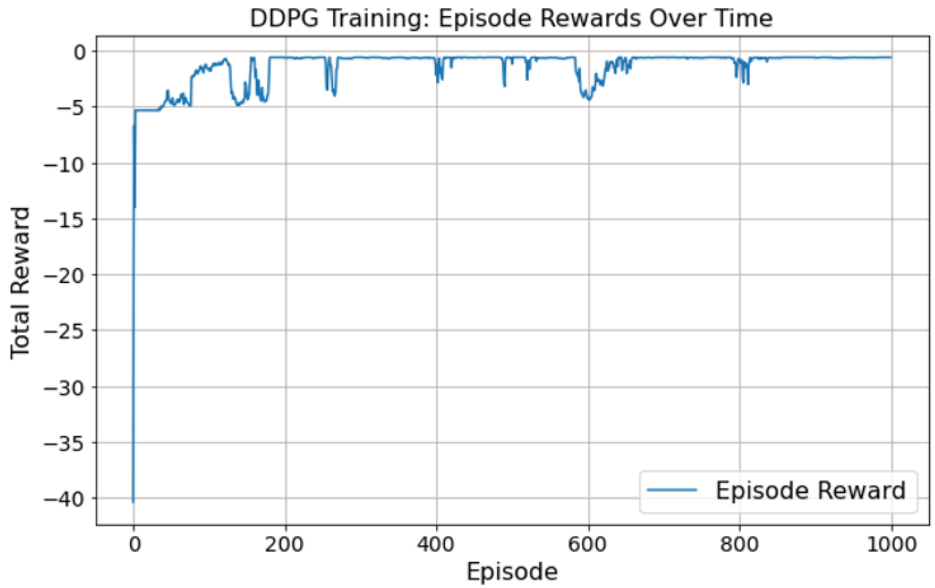


Benchmark Case

Accuracy and convergence of model neural network system:

- **Critic network loss converges to 0**; the relationship between state-action pair $(\mathbb{S}, a)$ and Reward Q is well approximated.
- **Actor network loss sees convergence**; always producing an optimal policy $a = \pi(\alpha, \delta, \ell)$ for each state $\mathbb{S}$; updated via gradient ascent: $\nabla_\lambda \mathcal{J}(\mu_\lambda)$
- Episode reward approaches the **maximum achievable reward** after several iterations





Classical framework

Comparison with Classical (Static) Framework

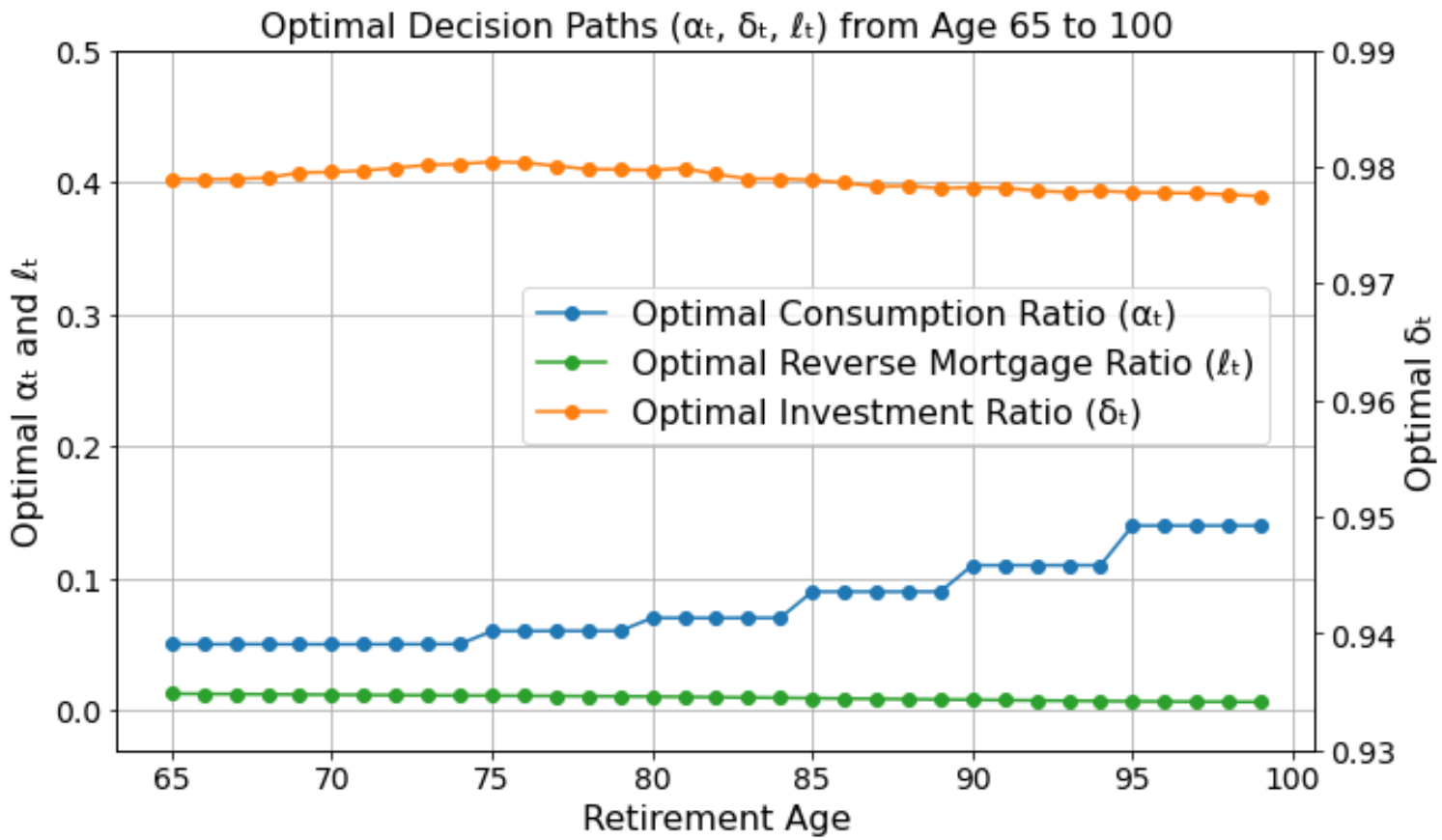
- **Classical (non-sequential)** framework applied using the same DDPG algorithm
- **Same agent profile and economic setting** as in the benchmark case $X_{t_0} = (W_{t_0}, H_{t_0}, L_{t_0}) = (\text{A\$360k}, \text{A\$1.5 million}, 0)$
- **Still produces stable and optimal outcomes**, though differing in sequential patterns
- **Emphasizes the critical role** of sequential modeling in life-cycle decision-making
- Model parameters [Koo et al., 2022]:

Θ	γ	θ	μ	σ	r	h	e^k
Value	-1.99	0.92	0.0684	0.0328	0.0374	0.0177	1.026

Table 3: Utility parameters and economic indicators applied to classical framework

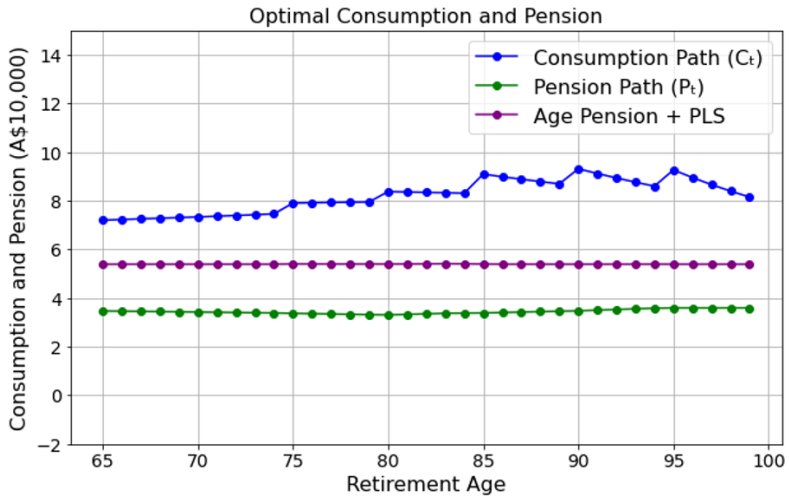
Classical framework

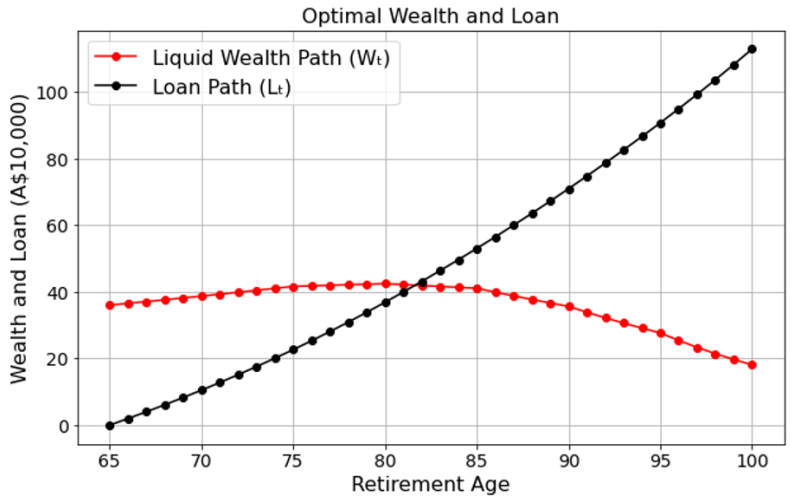
Optimal control $\pi_t = (\alpha_t, \delta_t, \ell_t)$



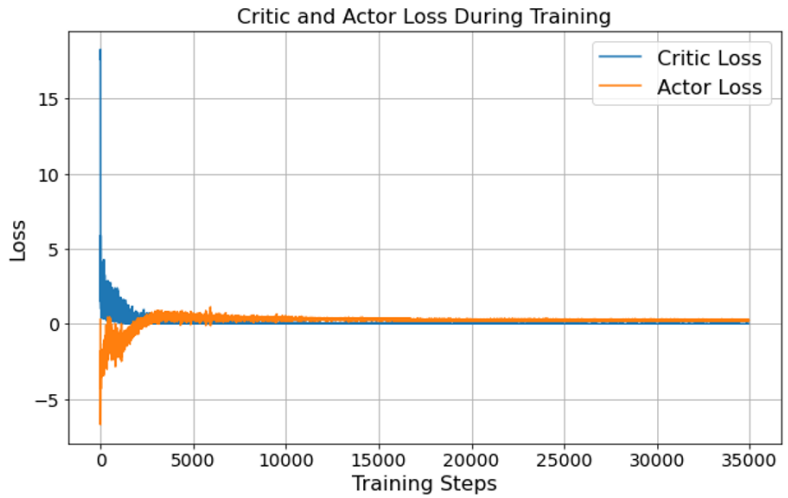
Classical framework

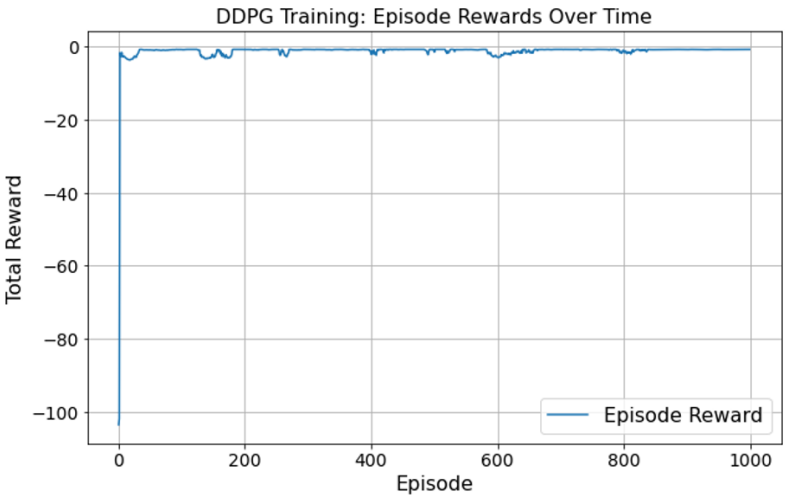
- Optimal Consumption, Pension, Wealth and Loan path:





- Accuracy and convergence of model neural network system:





Application to different profile

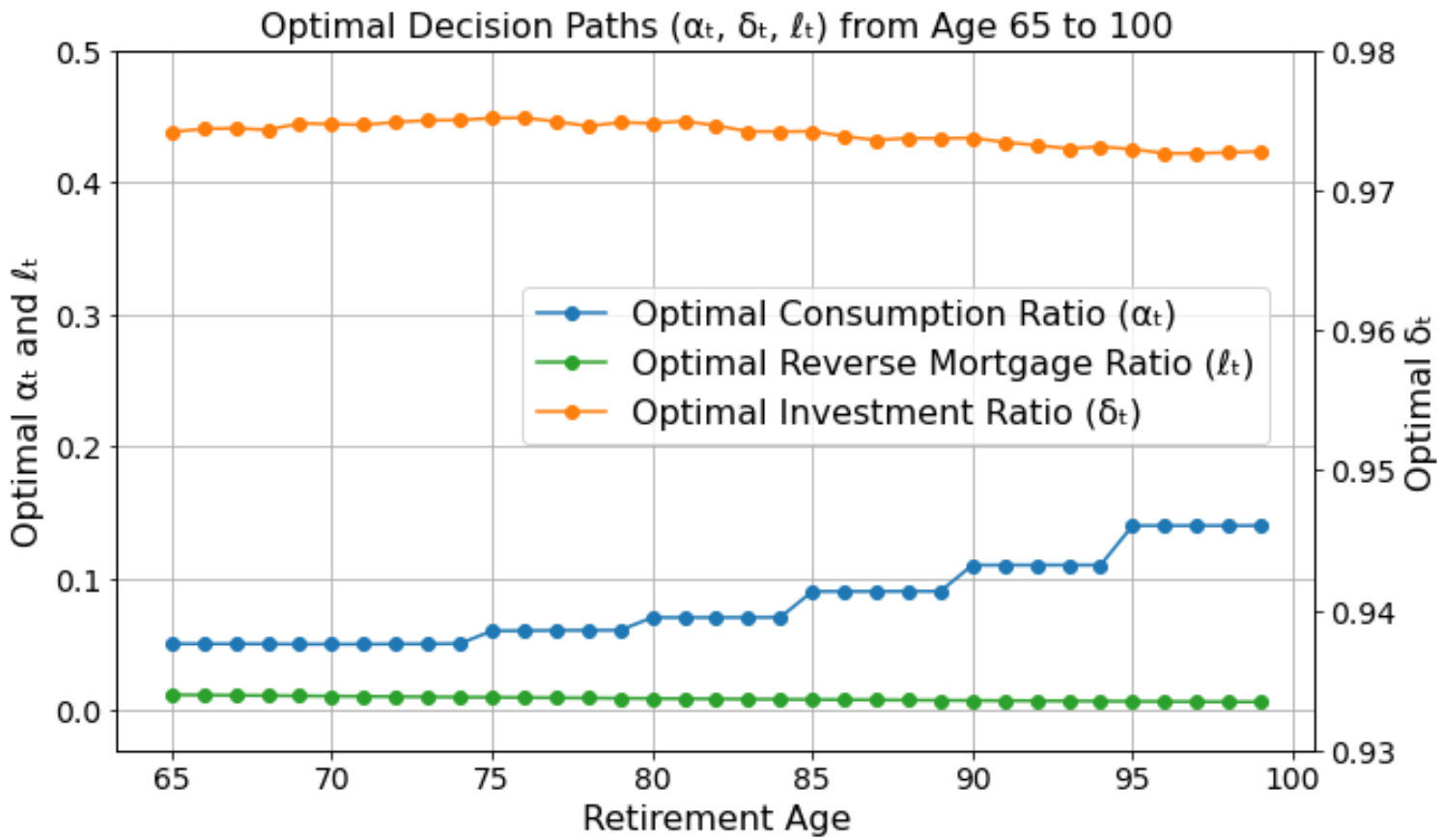
Retiree agent of different profile:

- Profile of low starting DC savings and large home
  $X_{t_0} = (W_{t_0}, H_{t_0}, L_{t_0}) = (\text{A\$146k},\ \text{A\$1.5 million},\ 0)$
- Profile of high starting DC savings and large home
  $X_{t_0} = (W_{t_0}, H_{t_0}, L_{t_0}) = (\text{A\$680k},\ \text{A\$1.5 million},\ 0)$
- Profile of low home asset
  $X_{t_0} = (W_{t_0}, H_{t_0}, L_{t_0}) = (\text{A\$360k}, \text{A\$1.0 million}, 0)$
  $X_{t_0} = (W_{t_0}, H_{t_0}, L_{t_0}) = (\text{A\$360k},\ \text{A\$500k},\ 0)$

(1) Same economic scenario as the benchmark case.
(2) Same model parameters.

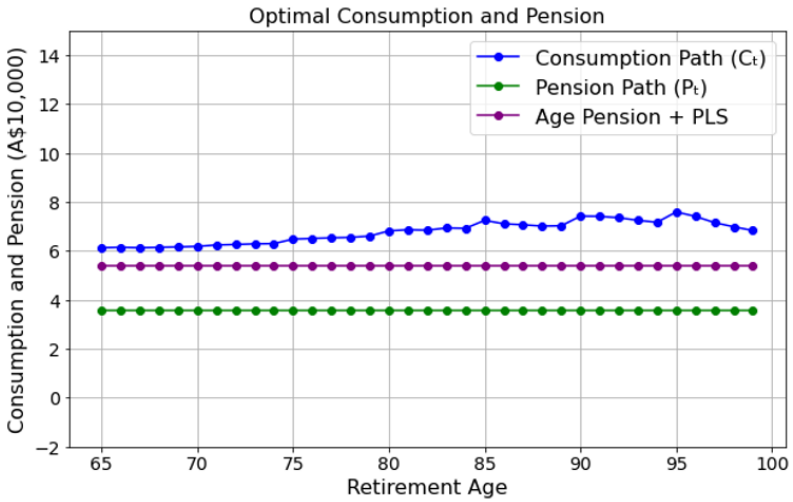
Profile of low starting DC savings and large home

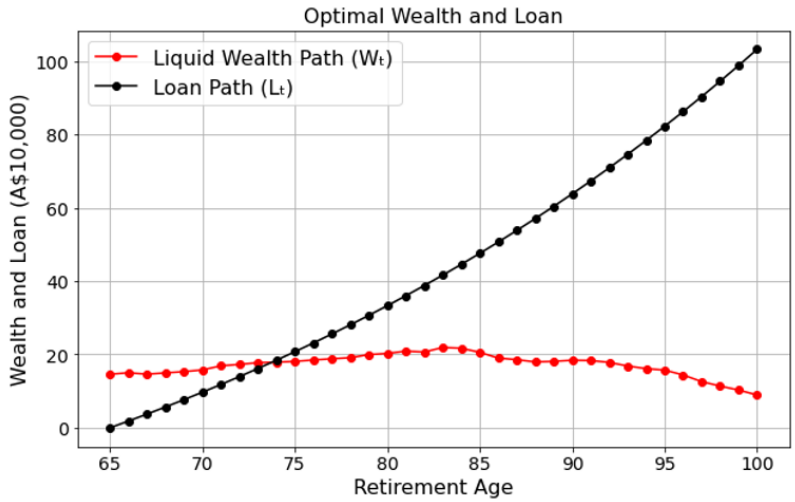
Optimal control $\pi_t = (\alpha_t, \delta_t, \ell_t)$:



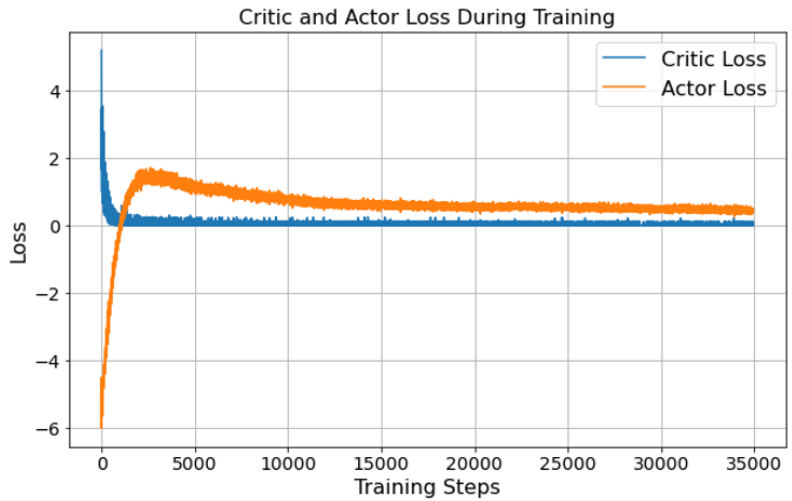
Profile of low starting DC savings and large home

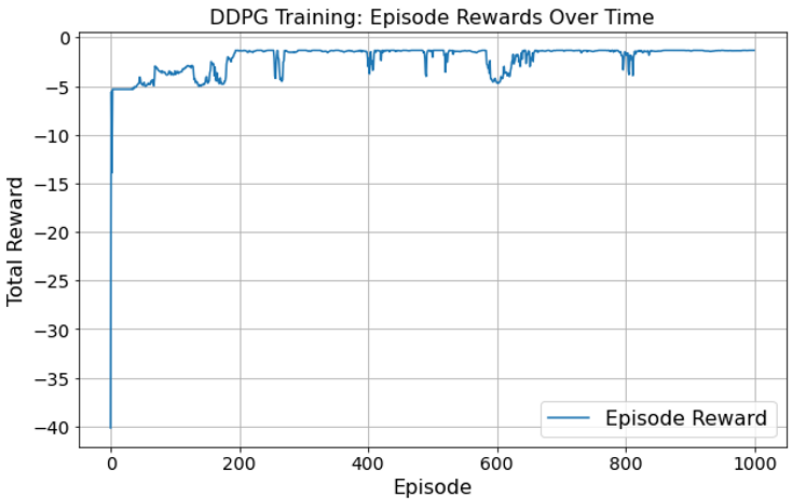
- Optimal Consumption, Pension, Wealth and Loan path:





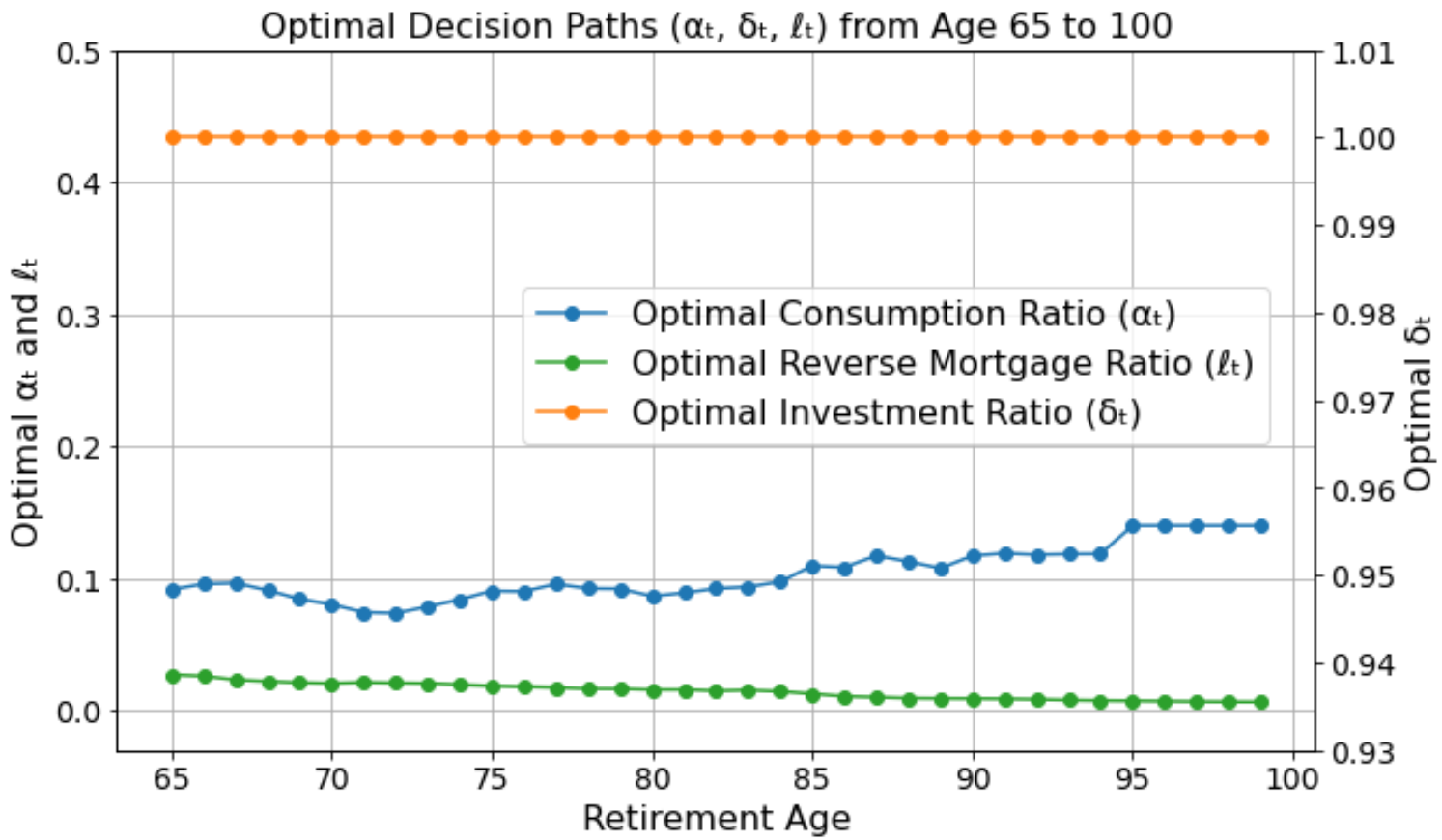
- Accuracy and convergence of model neural network system:





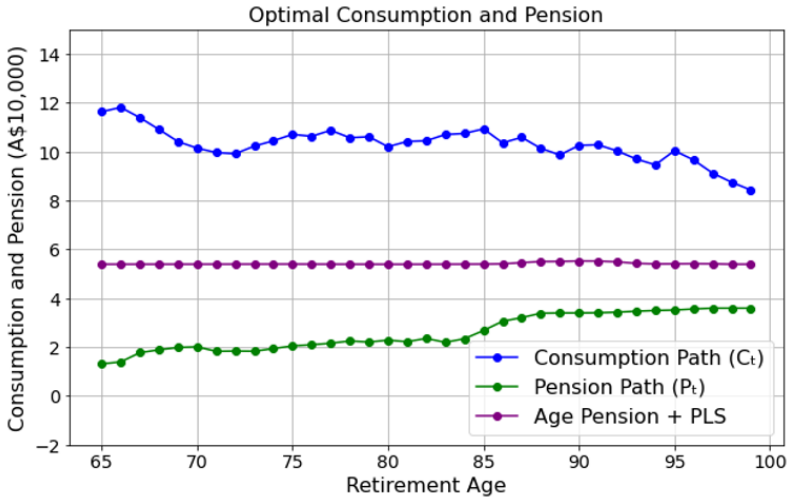
Profile of high starting DC savings and large home

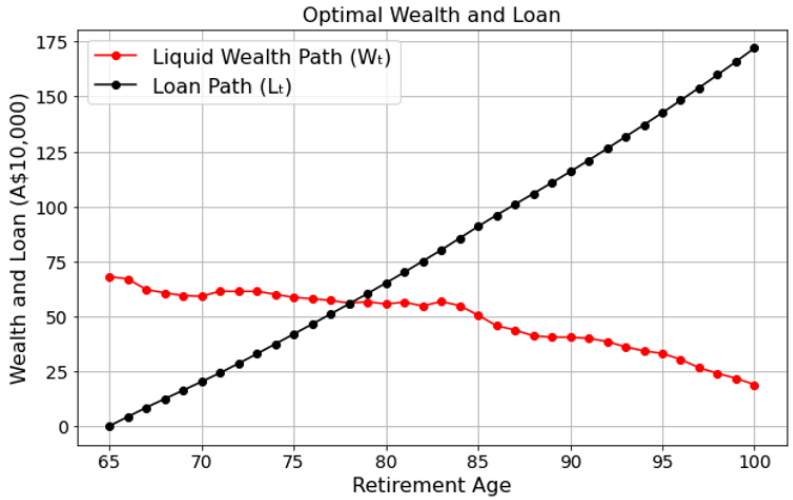
Optimal control $\pi_t = (\alpha_t, \delta_t, \ell_t)$:



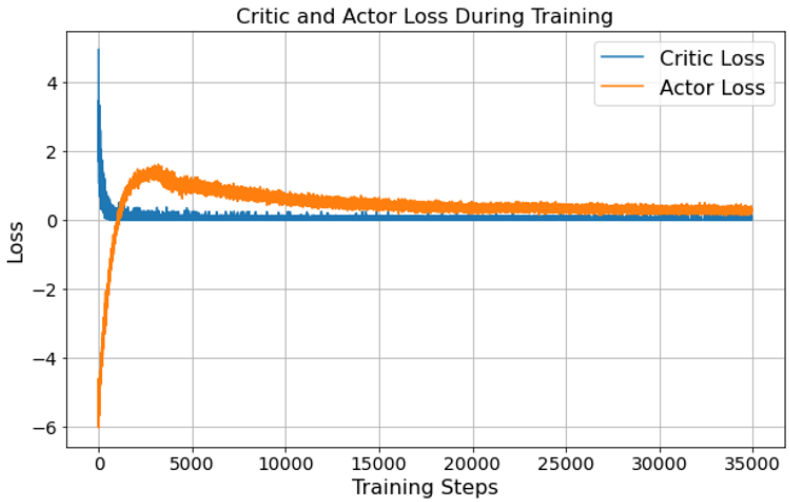
Profile of high starting DC savings and large home

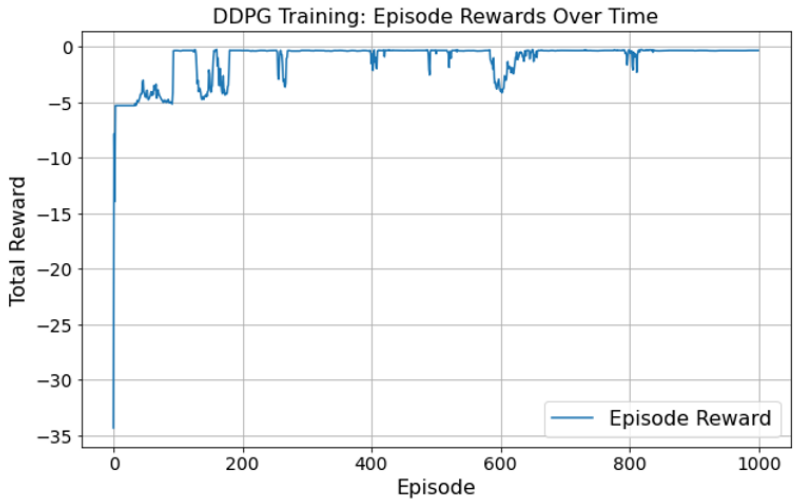
- Optimal Consumption, Pension, Wealth and Loan path:





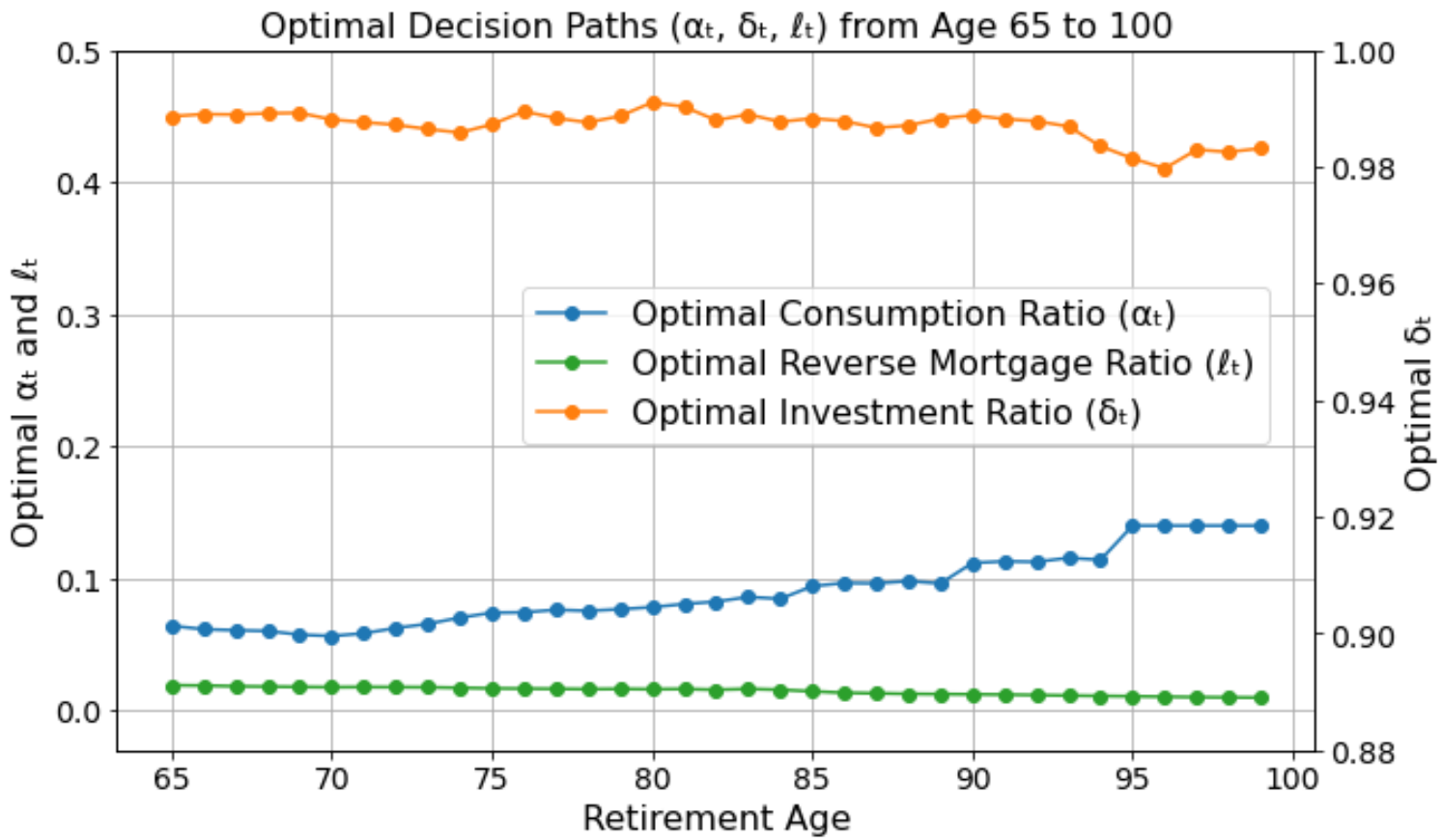
- Accuracy and convergence of model neural network system:





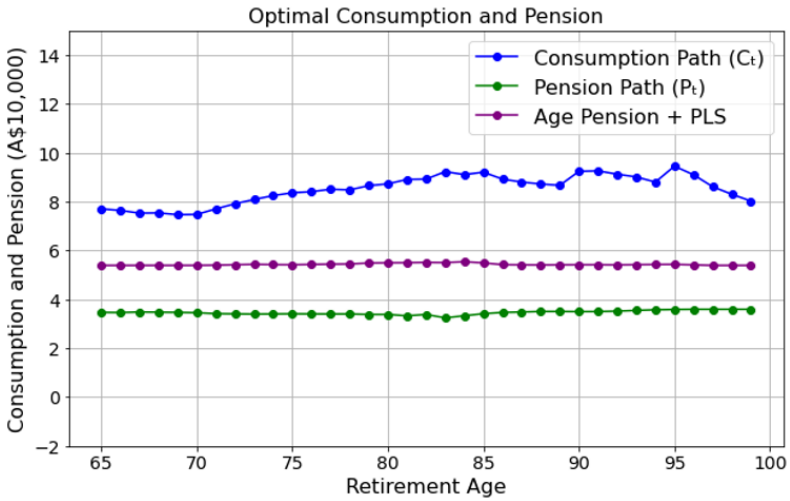
Profile of low home asset (A$360k, A$1.0 million, 0)

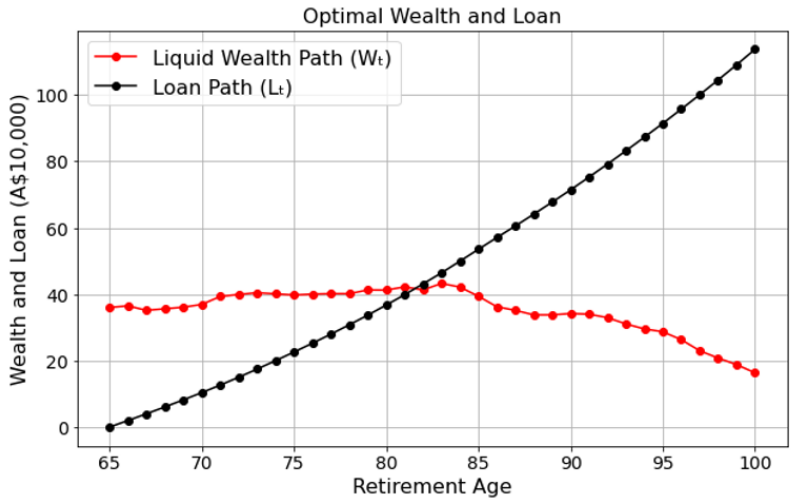
Optimal control $\pi_t = (\alpha_t, \delta_t, \ell_t)$:



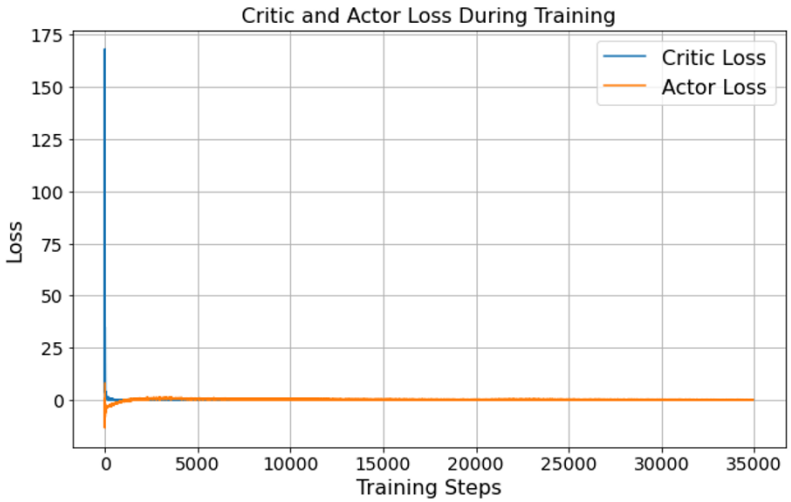
Profile of low home asset (A$360k, A$1.0 million, 0)

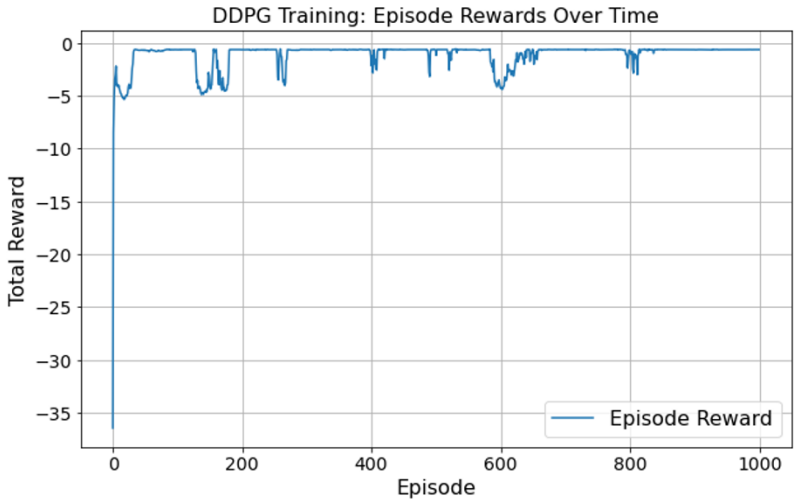
- Optimal Consumption, Pension, Wealth and Loan path:





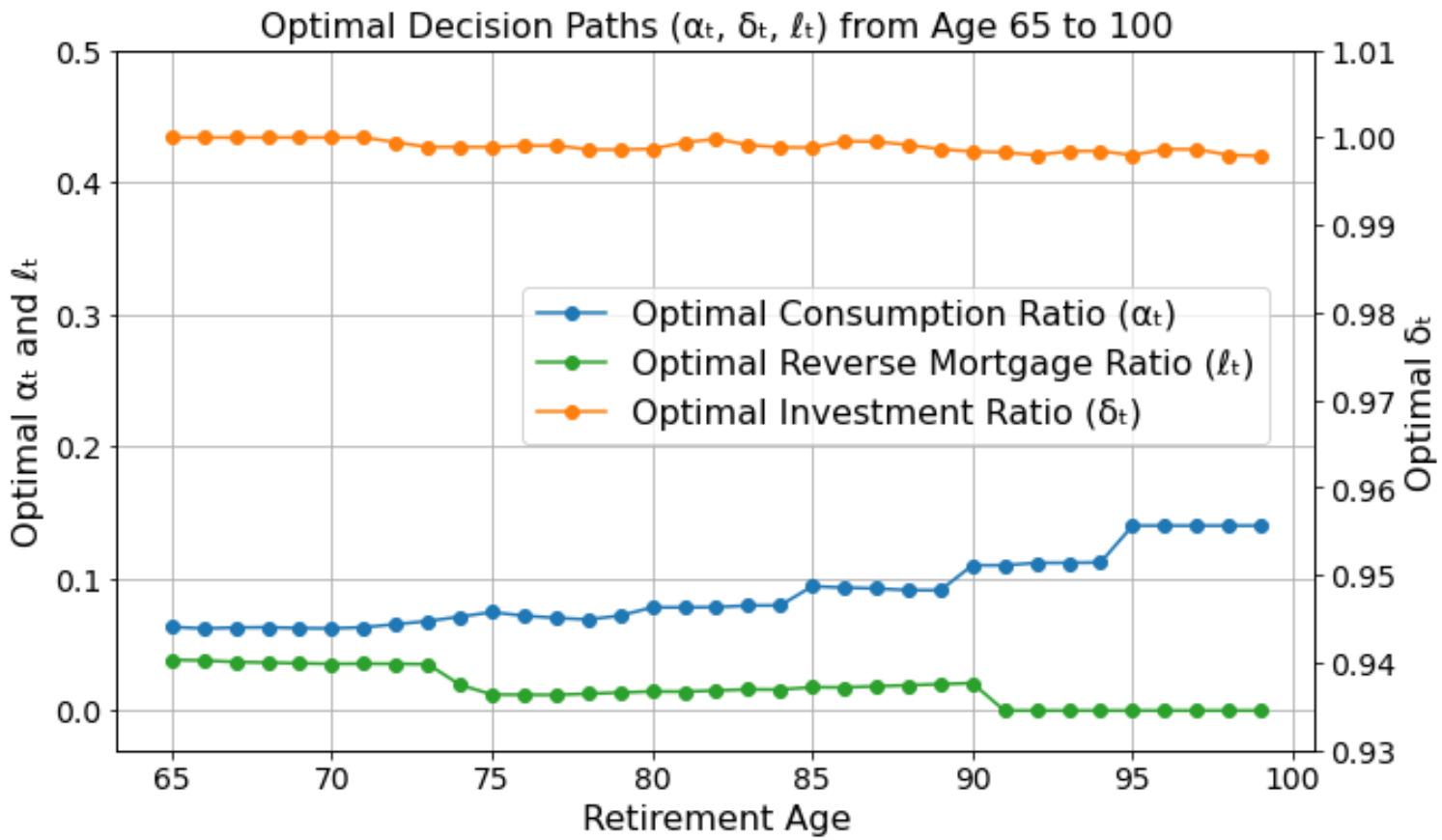
- Accuracy and convergence of model neural network system:





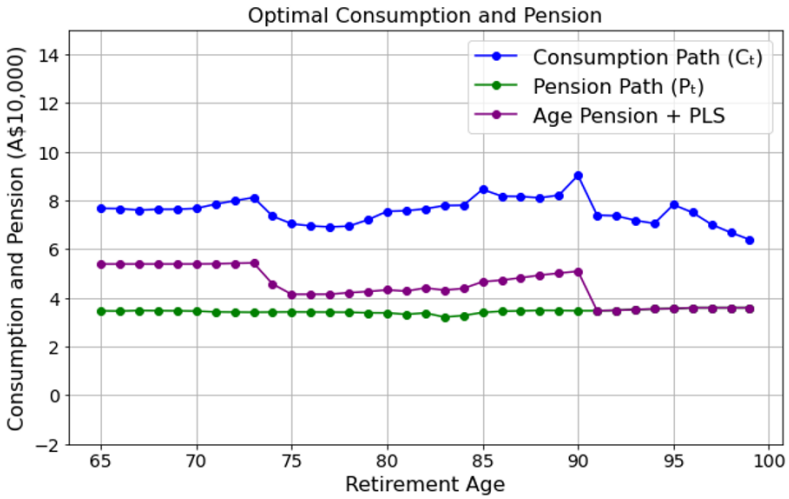
Profile of low home asset (A$360k, A$500k, 0)

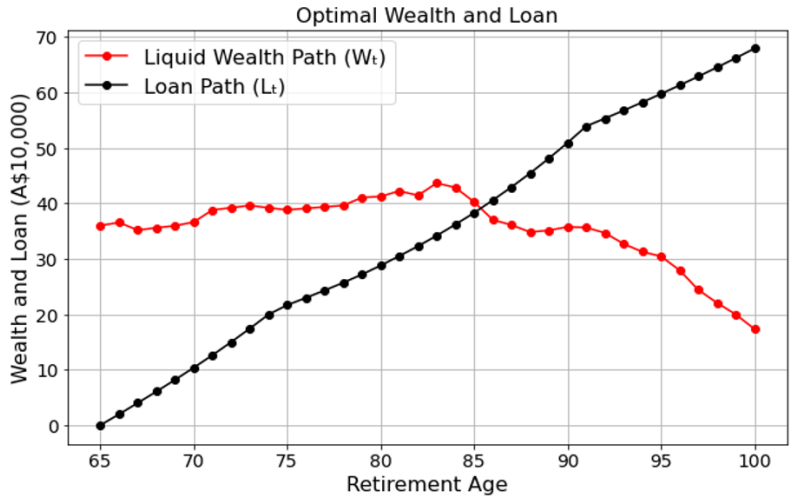
Optimal control $\pi_t = (\alpha_t, \delta_t, \ell_t)$:



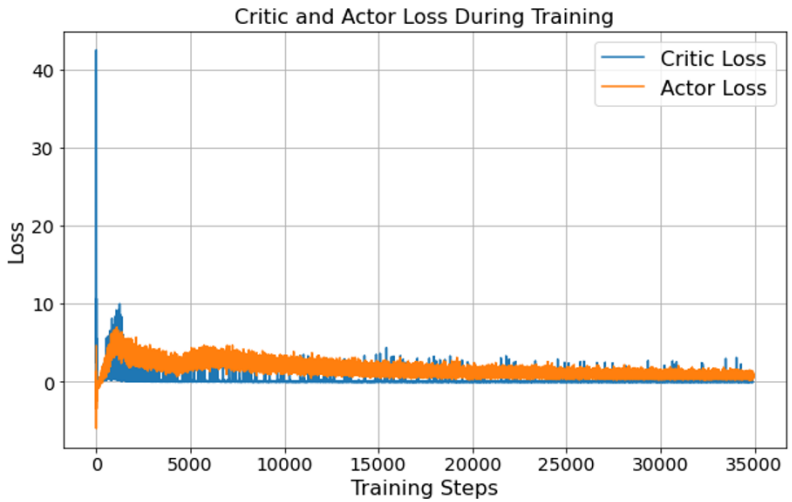
Profile of low home asset (A$360k, A$500k, 0)

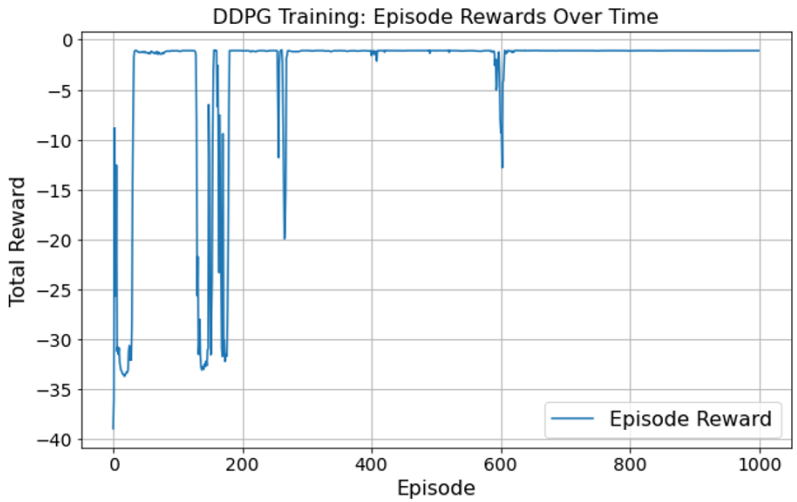
- Optimal Consumption, Pension, Wealth and Loan path:





- Accuracy and convergence of model neural network system:





Application to different profile

Across all retiree profiles:

- Generated policies yield stable and reliable results.
- Utility is consistently maximized.
- Basic consumption, liquidity and bequest needs are safeguarded.

In the last low-asset profile case:

- Greater volatility in consumption/control paths but still within acceptable level.
- Slower convergence and higher training cost due to limited asset endowment.
- Less stable optimal paths $\Rightarrow$ mainly driven by asset-poor trait, particularly low house asset value.
- Asset constraints reduce flexibility in managing the PLS.
- Still ensures basic consumption and liquidity under personal financial stress.

Application to different economic scenario

Optimistic scenario:

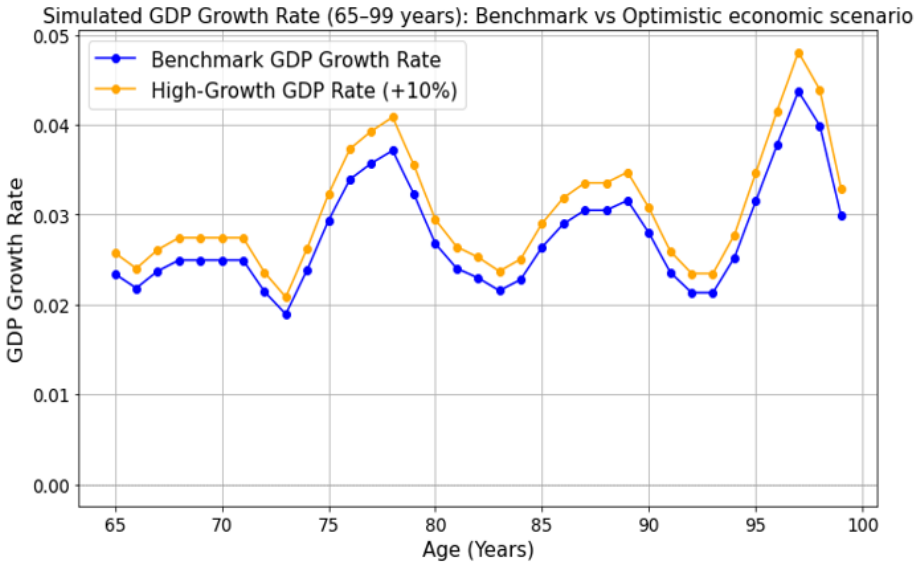
- Higher GDP Growth
- Lower inflation
- Higher mean market return
- Lower market volatility
- Eased interest rate

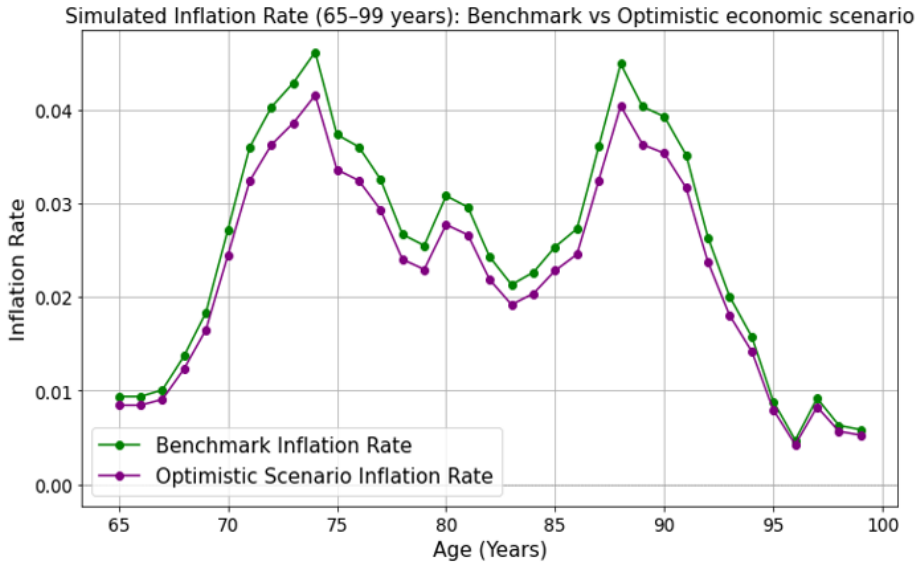
Pessimistic scenario:

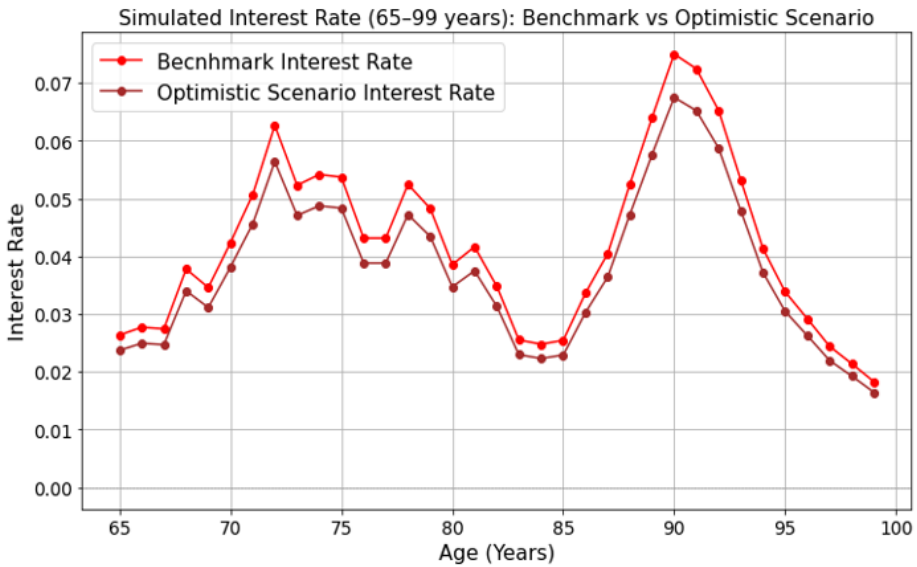
- Lower GDP Growth
- Higher inflation
- Lower mean market return (negative return, sometimes)
- Higher market volatility
- Contracted interest rate

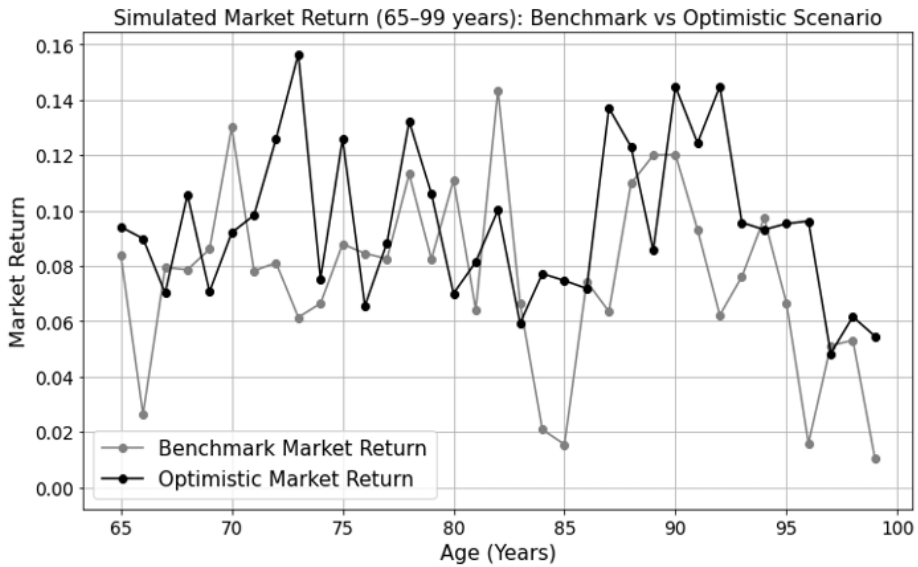
Same retiree profile, model parameter as the benchmark case.

Optimistic economic scenario



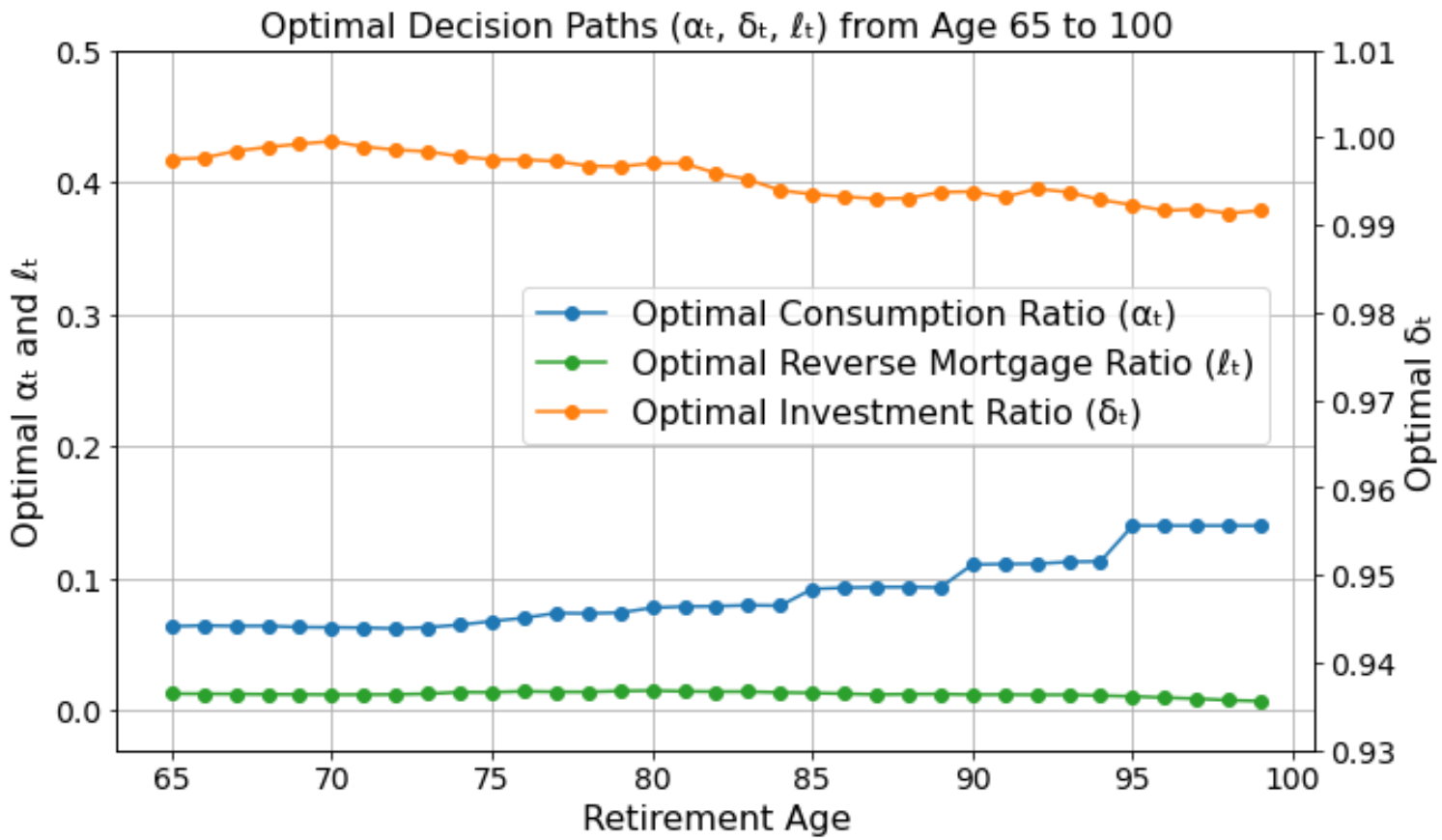






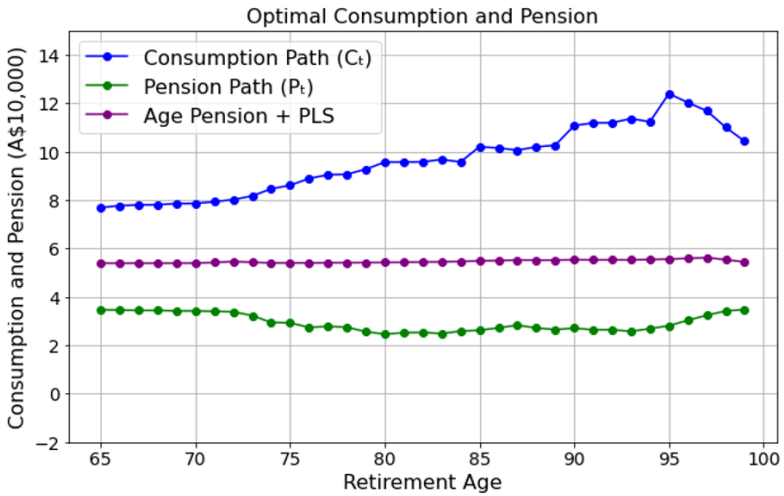
Optimistic economic scenario

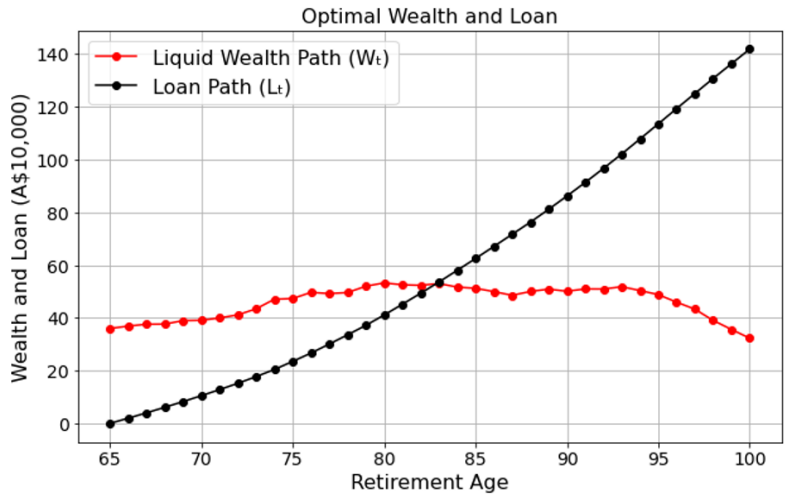
Optimal control $\pi_t = (\alpha_t, \delta_t, \ell_t)$:



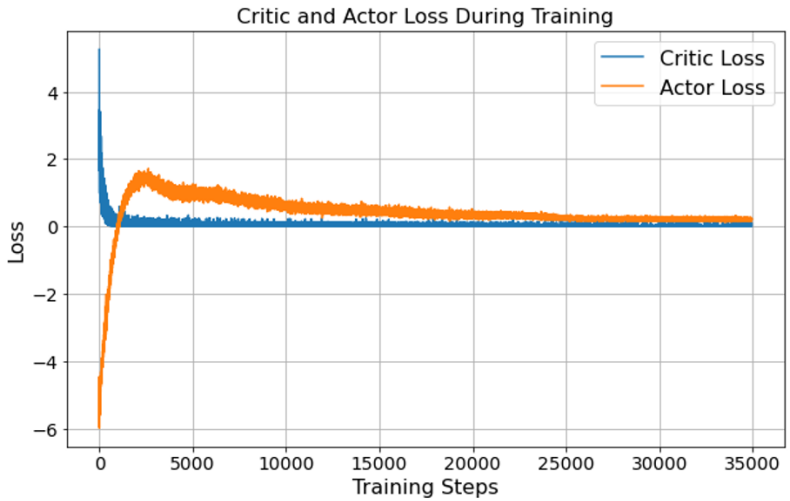
Optimistic economic scenario

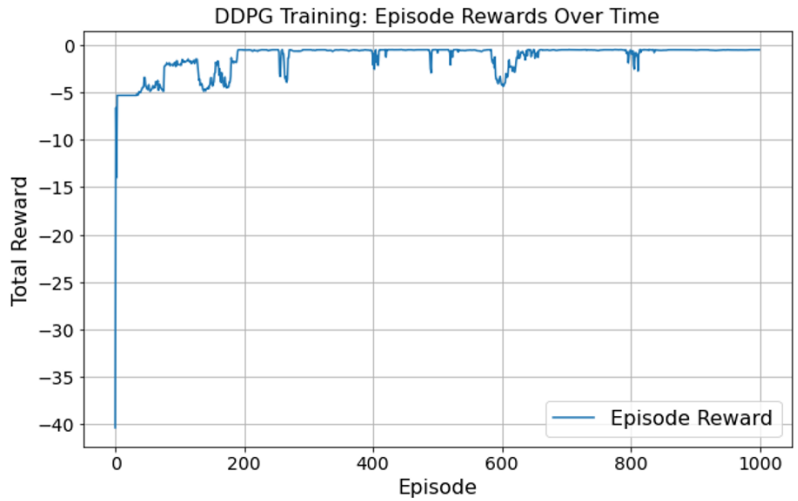
- Optimal Consumption, Pension, Wealth and Loan path:





- Accuracy and convergence of model neural network system:



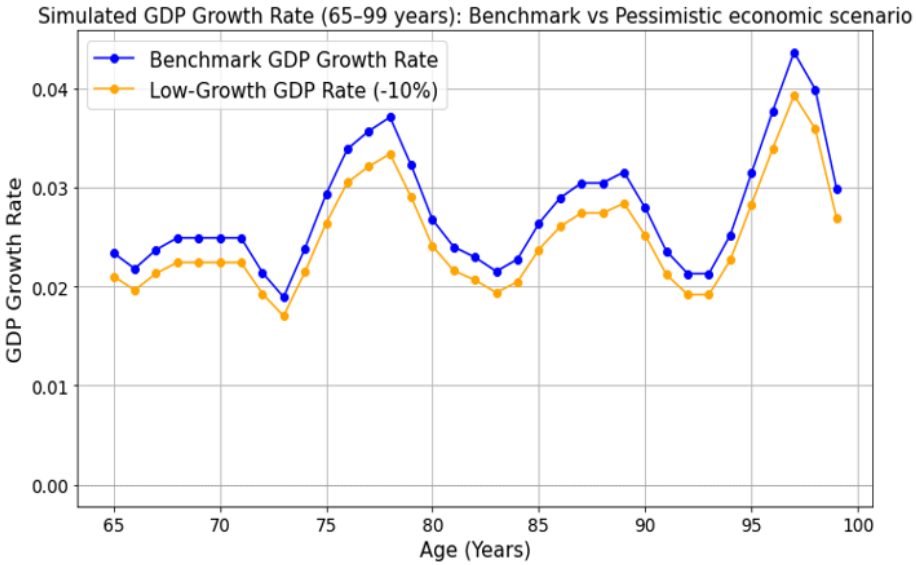


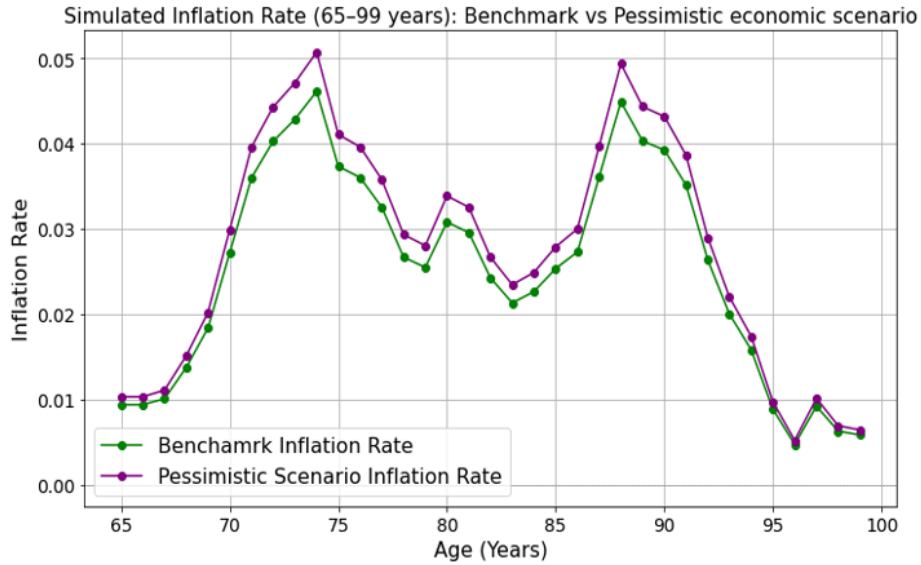
Optimistic economic scenario

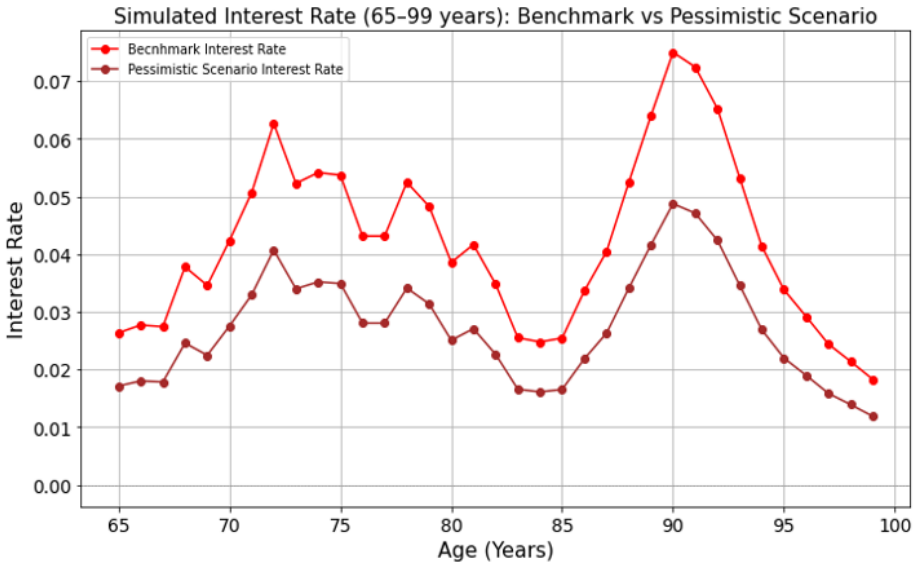
Result interpretation:

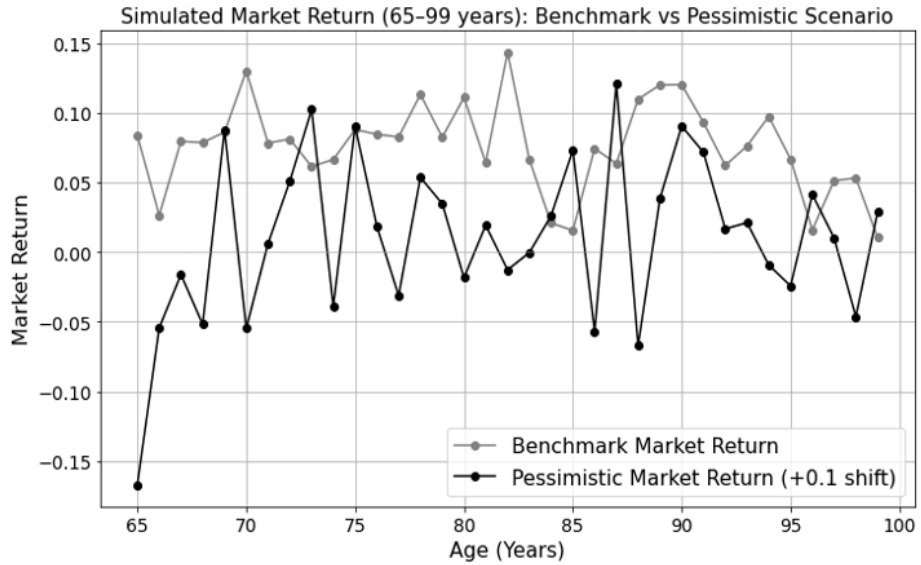
- **Higher consumption growth** than benchmark case, reflecting improved optimism.
- **Effective liquidity use**: wealth is neither over-consumed nor excessively hoarded.
- **Paths remain stable and robust** even under favorable market conditions.

Pessimistic economic scenario



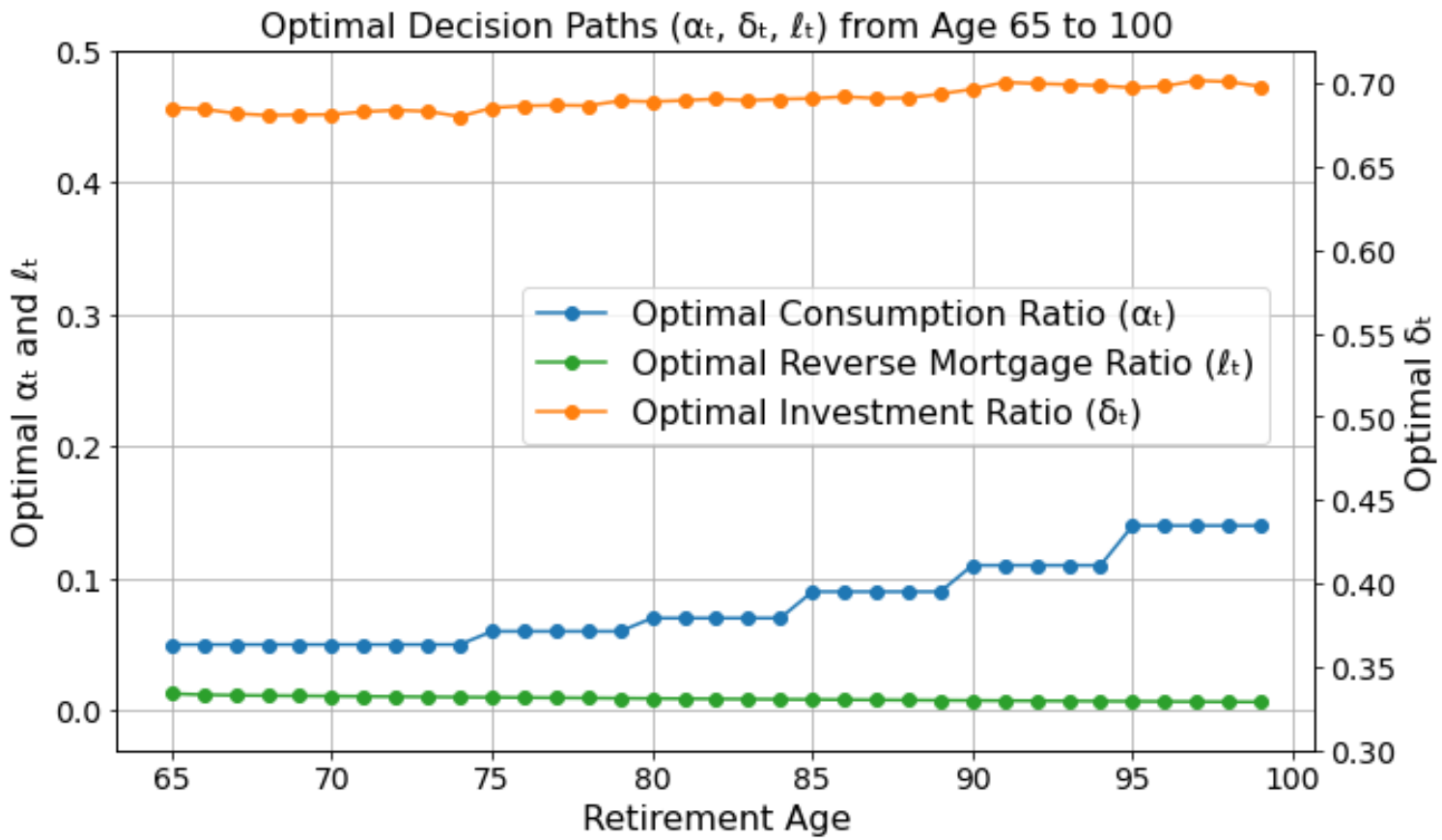






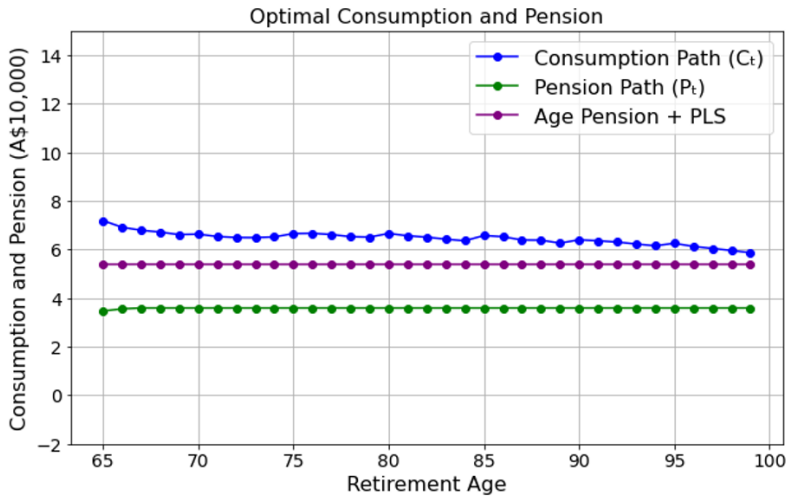
Pessimistic economic scenario

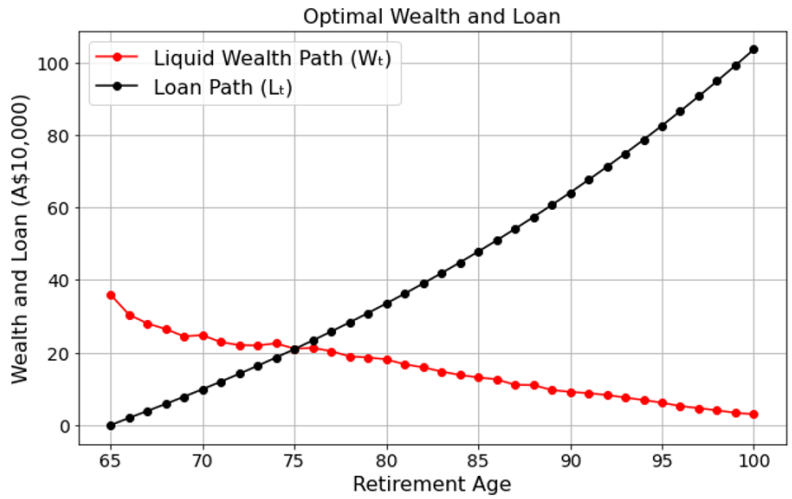
Optimal control $\pi_t = (\alpha_t, \delta_t, \ell_t)$:



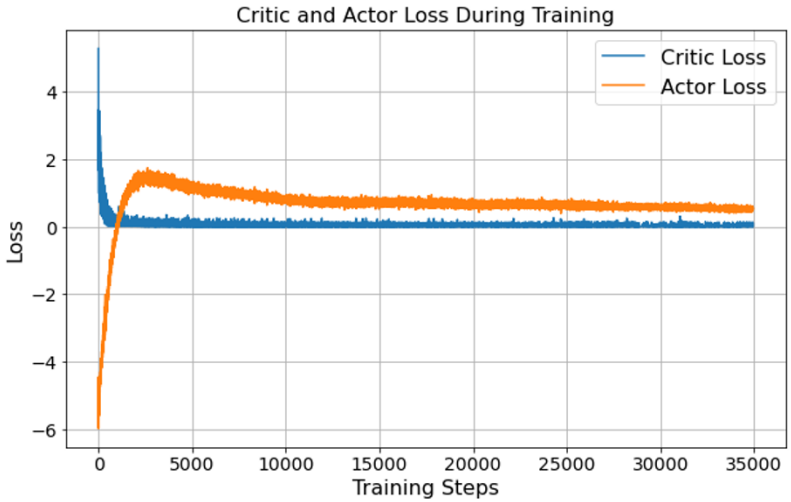
Pessimistic economic scenario

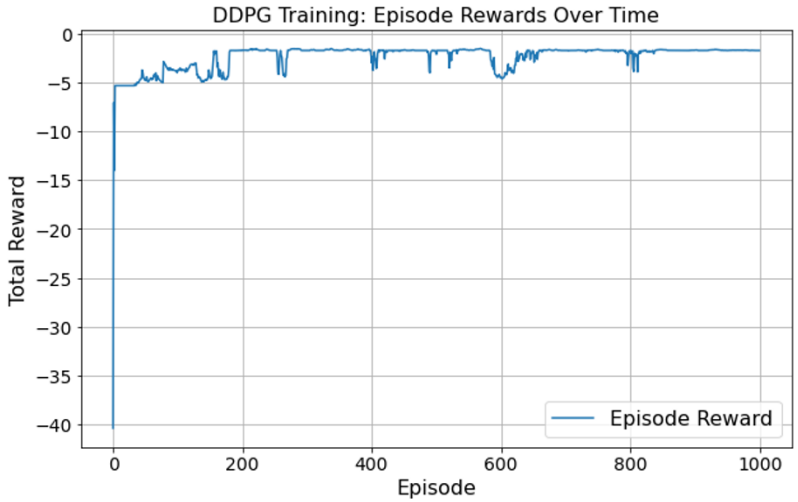
- Optimal Consumption, Pension, Wealth and Loan path:





- Accuracy and convergence of model neural network system:





Pessimistic economic scenario

Result interpretation:

- Gradual consumption decline is expected under pessimistic conditions.
- Basic needs and utility are preserved despite shrinking consumption.
- Optimal policy adopts **slow, controlled reductions**, avoiding sharp drops in living standards (consumption).
- **Liquidity buffer maintained**, ensuring financial resilience.
- Decision and state paths remain **stable and robust**.

Summary

- Provides optimal pension strategies under uncertainty in a DC scheme.
- Enables effective and personalized decision-making.
- Ensures **stable and reliable** paths that maximize life-cycle utility.
- Captures risk preferences & economic dynamics, and conditional expectations; better reflect real-world behavior.
- Leverages deep reinforcement learning to solve high-dimensional, continuous control problems beyond numerical methods.
- An important step forward for addressing more complex decision-making models.
- Demonstrates the potential of data science and AI in pension optimization.
- This integrated framework enriches the existing decision model system.

The End

Thanks!

Reference I

Chen, S., Luo, D., and Yao, H. (2024).
Optimal investor life cycle decisions with time-inconsistent preferences.
Journal of Banking & Finance, 161:107115.

Koo, B., Pantelous, A. A., and Wang, Y. (2022).
Novel utility-based life cycle models to optimise income in retirement.
European Journal of Operational Research, 299(1):346–361.

Kraft, H., Munk, C., and Weiss, F. (2022).
Bequest motives in consumption-portfolio decisions with recursive utility.
Journal of Banking & Finance, 138:106428.

Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., and Wierstra, D. (2019).
Continuous control with deep reinforcement learning.
Conference Paper, Cornell University Library, ISSN: 2331-8422.

Reference II


Modigliani, F. (1986).
Life cycle, individual thrift, and the wealth of nations.
The American Economic Review, 76(3):297–313.


Powell, W. B. (2022).
Reinforcement Learning and Stochastic Optimization: A Unified Framework for Sequential Decisions.
John Wiley & Sons.


Silver, D., Lever, G., Heess, N., Degris, T., Wierstra, D., and Riedmiller, M. (2014).
Deterministic policy gradient algorithms.
In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 387–395. PMLR.

Reference III


Wang, H., Hu, S., Siu, T. K., Wang, R., and Wang, N. (2024).
Life-cycle planning with cev model and time-inconsistent preferences.
International Review of Economics & Finance, 96:103517.


Wang, W., Li, B., Luo, X., and Wang, X. (2023).
Deep reinforcement learning for sequential targeting.
Management Science, 69(9):5439–5460.


Zhang, J., Wan, C., Chen, M., and Liu, H. (2024).
An efficient reinforcement learning approach for goal-based wealth management.
Expert Systems with Applications, 237:121578.