Deep Reinforcement Learning for Asset Liability Management of a Pension Fund

Micha Villing

July 5th, 2023

Introducing Micha Villing

- Senior Actuary for c-alm AG[1] and Azenes AG[2]
- Finishing up my PHD at the University of St. Gallen

- Joint work with Dr. des. Cédric Müller and Manuel Sager
  - Master Thesis of Manuel Sager
  - Part of Dr. des. Müller and my PHD Thesis

- Real world application for the closed fund of Publica

[1] micha.villing@c-alm.ch

[2] micha.villing@azenes.ch

Agenda

Introduction

Pension Funds

Reinforcement Learning

Application

Conclusion

Asset Liability Managment

The Society of Actuaries defines asset liability management as …

"… the ongoing process of formulating, implementing, monitoring and revising strategies related to assets and liabilities to achieve an organization's financial objectives, given the organization's risk tolerances and other constraints".

Luckner, W. R., Abbott, M. C., Backus, J. E., Benedetti, S., Bergman, D., Cox, S. H., Smith Jr, P. L. (2003). Asset-liability management. Society of Actuaries.

Motivation

Pension funds

Total assets held exceeding USD 35 trillion by the end of the year 2020[1]

Millions of people depend on pension funds for their retirement

Decision makers have to …

… secure the pensions of retirees now

… secure the pensions of workers that retire soon

… secure the pensions of workers that retire 40 years in the future

Decisions makers must find asset allocations and liability side parameters that helps them achieve their objectives.

[1]See the annual study by the OECD:

https://www.oecd.org/finance/private-pensions/Pension-Funds-in-Figures-2021.pdf

In a Nutshell

What is the optimal dynamic risk allocation for a closed pension fund in run-off?

We use deep reinforcement learning to solve the above ALM problem.

Agenda

Introduction

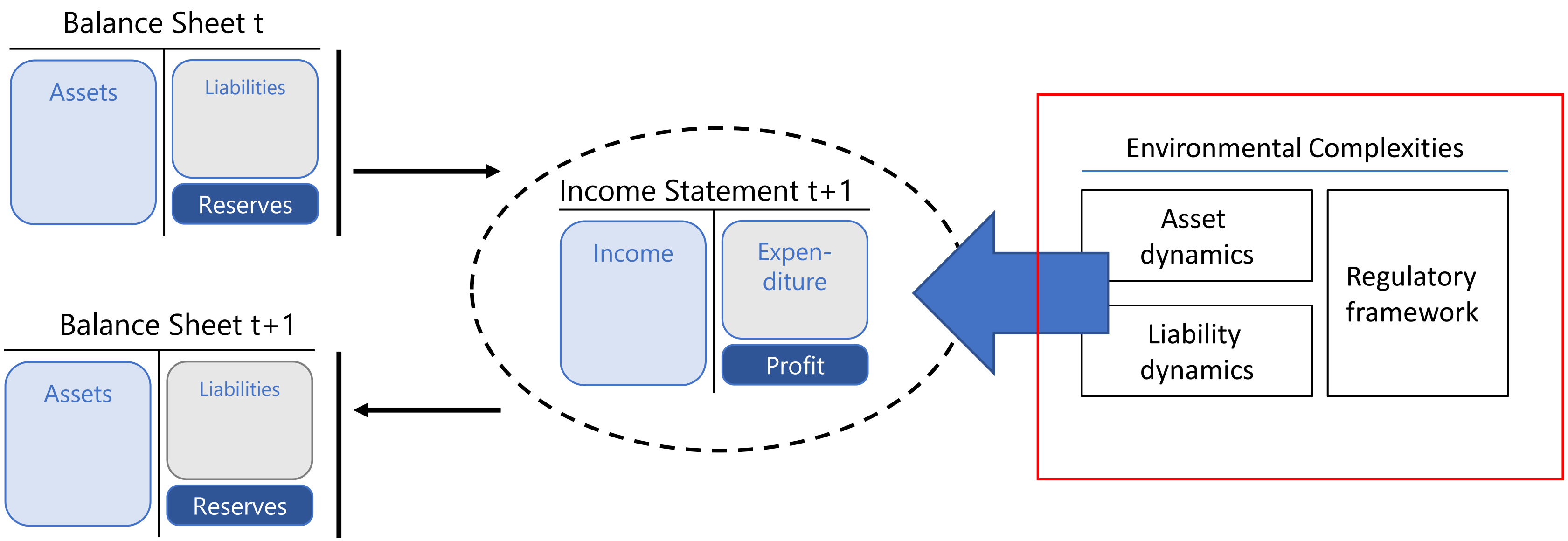
Pension Funds

Reinforcement Learning

Application

Conclusion

Modell of a Pension Fund Projection



- State: Initial balance sheet and income statement items
- Actions: *Will be explained later*
- Environment: Development of the financial markets and the liabilities side (insured portfolio)
- Reward: *Will be explained later*
- State: New balance sheet and income statement items

Environmental Complexities

Liabilities

Initial population

Population development

Valuation

Benefit catalogue

Assets

Asset space

Asset process

Asset allocation

Transaction costs

Regulation

Asset regulation

Valuation regulation

Coverage regulation

Goals of a Pension Fund

The main goal is ***not*** to maximize shareholder value after having accounted for all the benefits for the insured (in contrast to privately organized life insurers)!

A pension fund as a solidaric construct ***should*** maximize the utility of its insured over its whole existence. This entails various complexities:

- Insured have different characteristics
- Decisions taken now have consequences far into the future
- Fluctuation between pension funds with different benefit plans
- How to structure participation in over funding ...
- ... how to finance shortfalls in under funding?

The Problem and the Solution

Problem	Solution: (Deep) Reinforcement Learning
• Pension funds are exposed to multidimensional, intertemporal optimization problems for optimal decision-making. • Classical models need highly simplified assumptions to find solutions.	• The model interacts freely with the environment and learns from the accumulated experience. • A combination with Deep Learning enables the solution of complicated tasks with little prior knowledge.

Agenda

Introduction

Pension Funds

Reinforcement Learning

Application

Conclusion

Reinforcement Learning – Markov Decision Process (I/II)

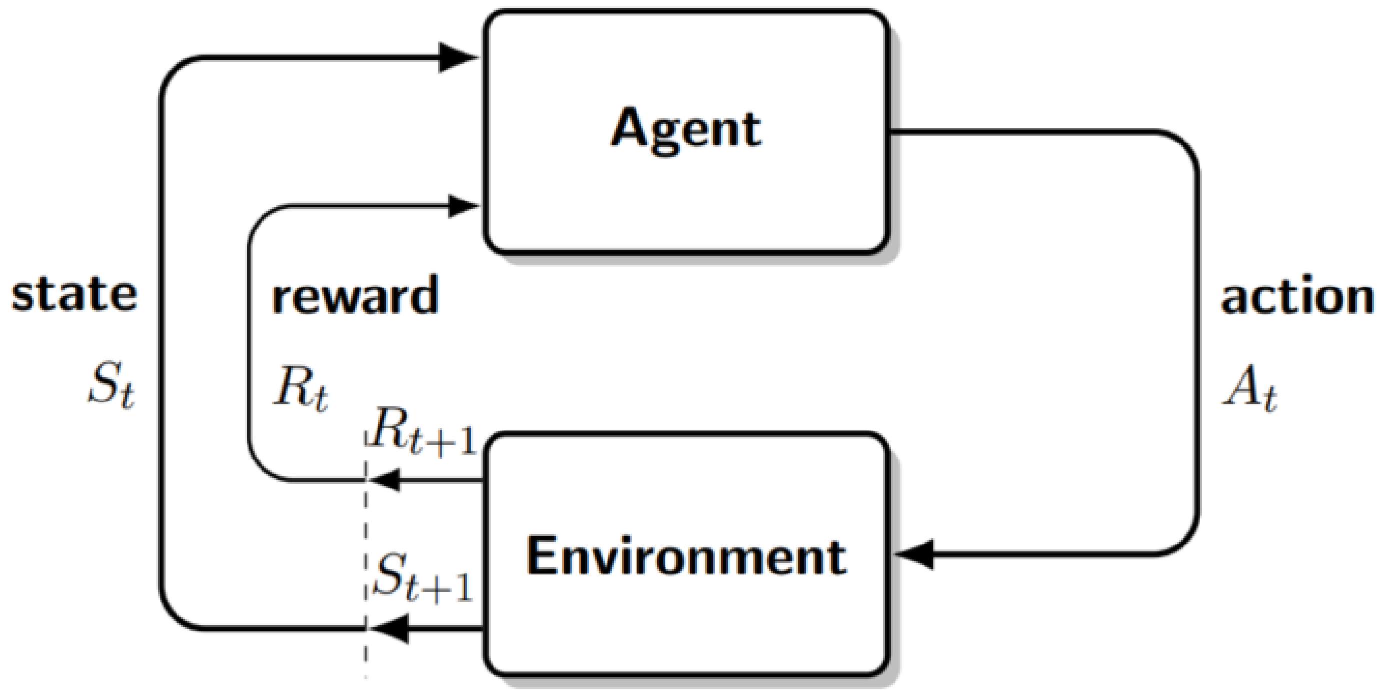
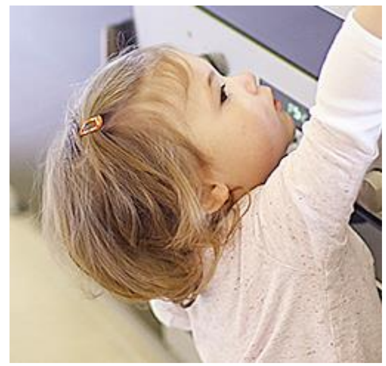


Figure: The agent-environment interaction. Adapted from Sutton and Barto (2018).

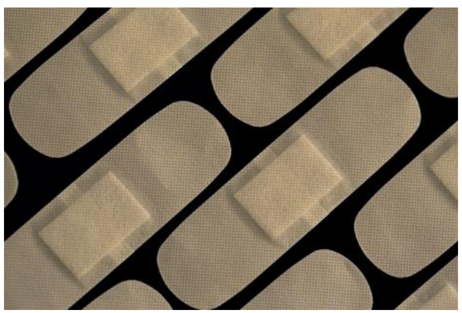
Sutton, R. S., & Barto, A. (2018). Reinforcement learning: An introduction (2nd ed.). Cambridge, MA: The MIT Press.

Agent-Environment Interaction: An Analogy

Agent: Child



(negative) Reward: Pain
State: Plaster on the hand



Action:
«I can touch that»

Environment:
Child burns his hand…

!

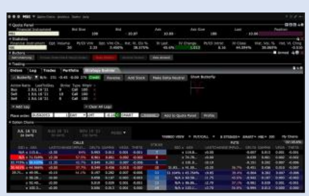
Reinforcement learning – Markov Decision Process (II/II)

Goal: Find a sequence of actions, the **policy $\boldsymbol{\pi}$**, that maximizes the rewards.

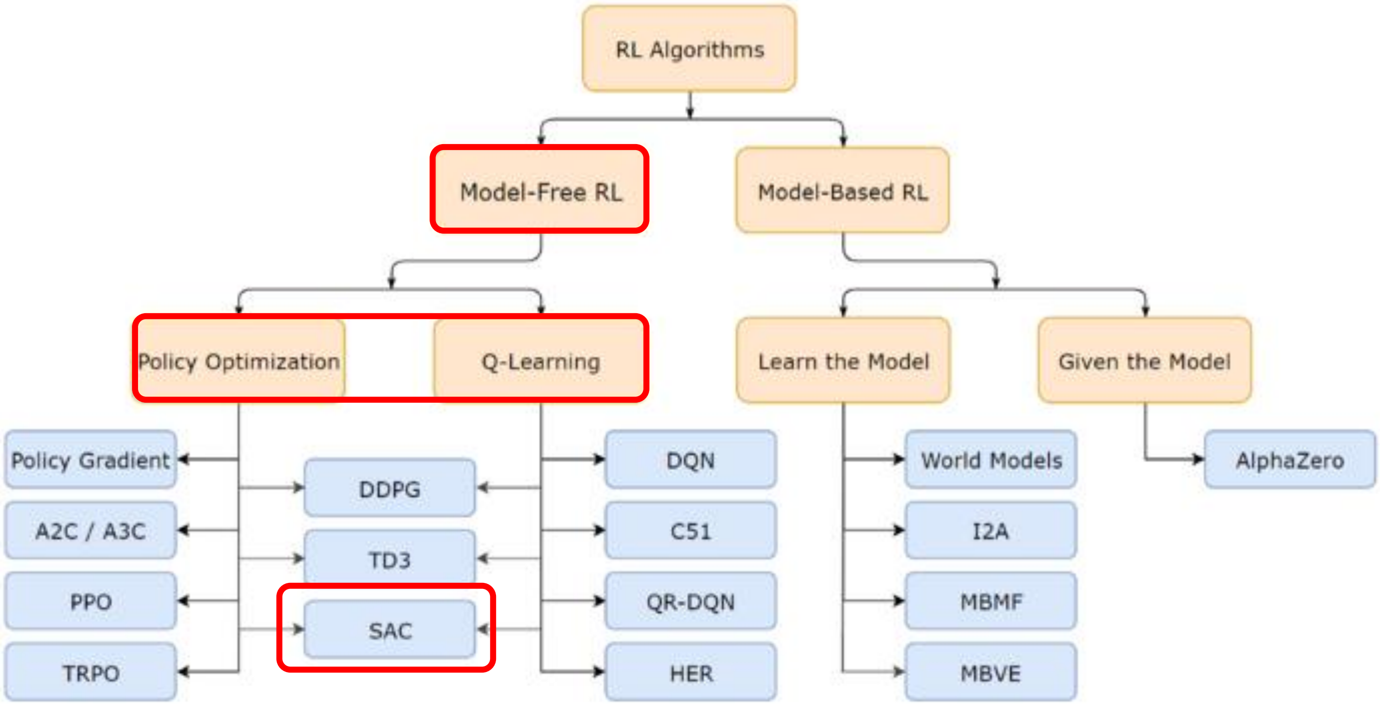
Solution: Let the agent interact with the environment and through trial and error find the best π: **Model-free reinforcement learning**

Side condition: Action/state space large and/or continuous: **Approximate learning with neural networks**

Famous reinforcement learning applications

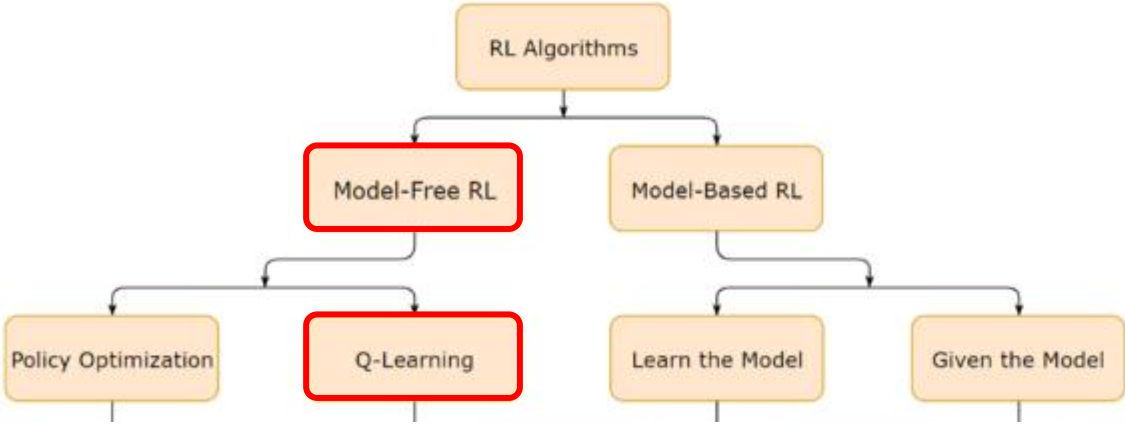
	Environment	Action Space	Rewards
Videogame			
Chess			
Algorithmic Trading			

Which Algorithm to Choose



Taxonomy of algorithms in modern RL by Open AI, Spinning Up. https://spinningup.openai.com/en/latest/spinningup/rl_intro2.html

Q-Learning



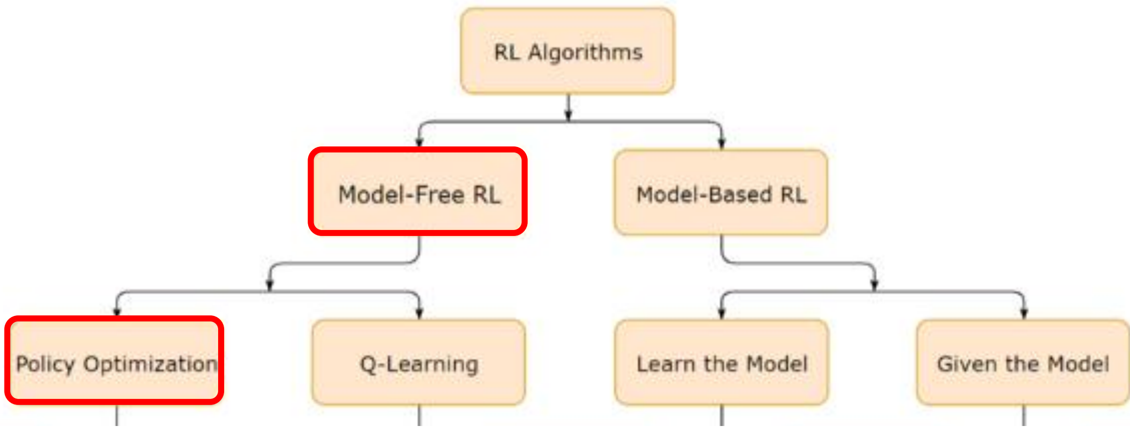
Find a policy that optimizes action-value function Q:

$$Q_\pi^* = \max_\pi E[R_t | s_t = s, a_t = a, \pi].$$

Solution approach:

- Reformulation as Bellman equation
- We use a neural network to approximate $Q(s, a; \theta) \approx Q^*(s, a)$, with θ being the weights of a neural network

Policy Optimization



Now we directly focusing on the policy distribution $\pi_\emptyset$:

- $\pi_\emptyset$ is again a neural network with parameters $\emptyset$
- We are trying to find the right set of weights to maximize $E_{\tau\sim\tau_\emptyset}[(\tau)]$, where τ is a trajectory $(s_t, a_t, s_{t+1} a_{t+1}, \ldots)$.

Algorithms: Soft Actor-Critic

Deep Q-learning: Sample efficient **but** unstable hyperparameter tuning

Policy optimization: Stochastic policies, stable **but** suffer from low sample efficiency

Soft Actor-Critic by Haarnoja et al. (2018)

- ”Best-of-both-worlds” and ”state-of-the-art”
- Actor-critic approach: The actor is a policy function with parameters φ and the critic is two (or more) value functions with parameters θ and ψ (all of the functions are neural networks).
- Soft approach: Adds entropy of policy to the objective function, i.e. the agent learns how much exploration it needs with parameter α

Haarnoja, T., Zhou, A., Abbeel, P., & Levine, S. (2018). Soft Actor-Critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In International Conference on Machine Learning (pp. 1861–1870).

Further Technical Considerations

Specialties to stabilize learning and improve convergence:

- Stochastic gradient descent
- Replay buffer
- Double Q-learner

Open AI Gym:

- Offers a wide array of pre-defined environments.
- Standardizes the interface for different environments, allowing easier benchmarking and development.
- Stable Baselines provides solutions to the most state-of-the-art algorithms that can be readily used to train your agent.

Agenda

Introduction

Pension Funds

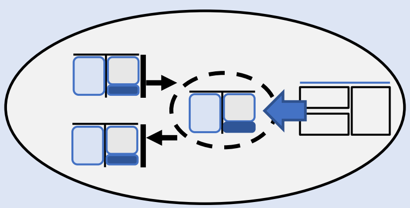
Reinforcement Learning

Application

Conclusion

Environment, Actions, Rewards

Environment



Action Space

...

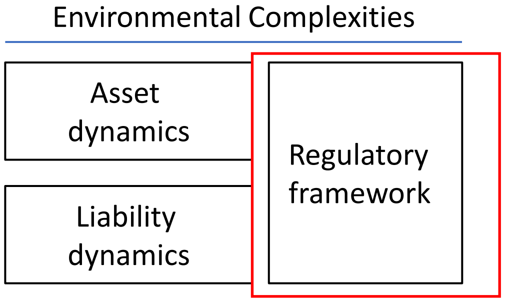
Reward

...

- State: Initial balance sheet and income statement items
- Action: ...
- Environment: Development of the financial markets and the liabilities side (insured portfolio)
- Reward: ...
- State: New balance sheet and income statement items

Regulatory Framework

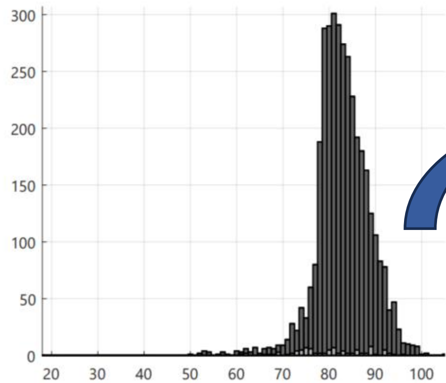
- Closed pension fund in run-off, start coverage 102%:
  - An insurance portfolio (retirees in our case) that only develops according to mortality;
  - There will be no other fluctuation (no new entries, no leaving the pension fund by changing the employer);
  - There are no contribution payments, i.e., no cash-inflows;
  - There are only annuity payments.
- Technical valuation with a fixed interest rate.
- Asset regulation according to BVV 2.
- Underfunding allowed until 95%. A recovery mechanism is enacted if this threshold is reached:
  - The employer covers the gap with recovery payments until 100% is reached;
  - The recovery payments are distributed over three years;
  - The payments are a form of loan to the pension fund (see next bullet).
- If recovery payments have been made, repayments to the employer are conducted if the coverage ratio allows it.



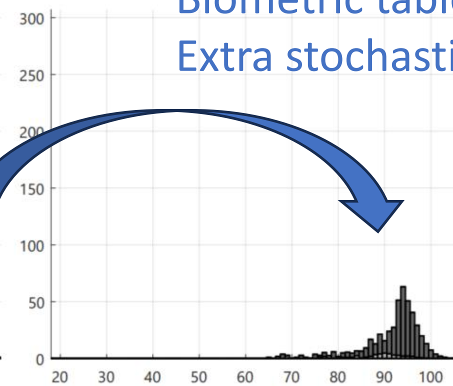
Very important for the specification of the reward

Liabilities

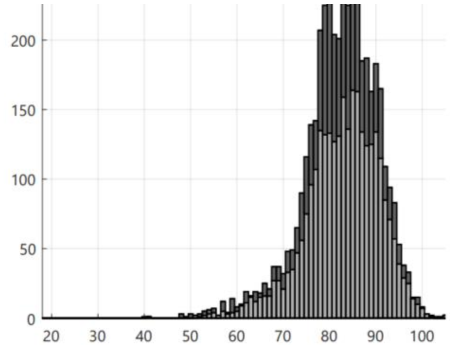
Biometric tables &
Extra stochasticity



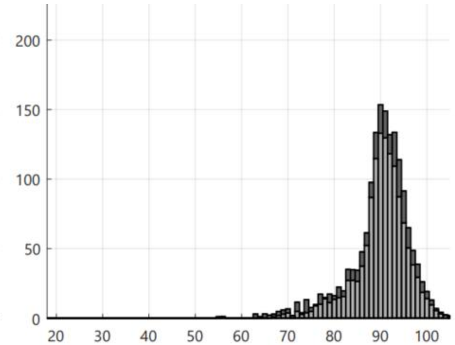
Men: Old age pension (black) and
widow’s pension (grey) 2021



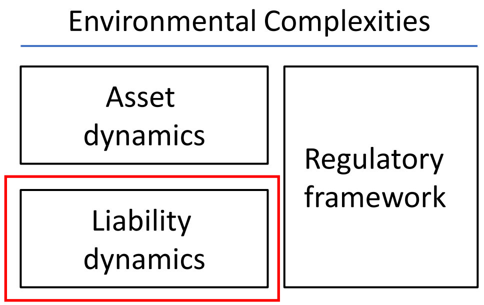
Men: Old age pension (black) and
widow’s pension (grey) 2036



Women: Old age pension (black)
and widow’s pension (grey) 2021



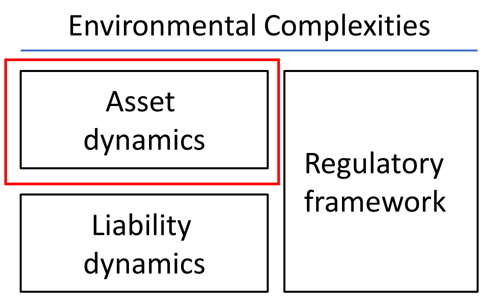
Women: Old age pension (black)
and widow’s pension (grey) 2036



Assets

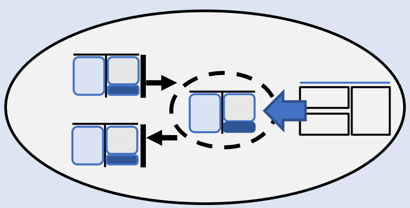
- Generating two types of portfolios:
    - “Risky” portfolio: Includes stocks (SPI & MSCÎ), starting allocation 30%
    - “Low risk” portfolio: Includes bonds (SBI), starting allocation 70%
    - Transaction fees
    - Maximal portfolio reallocation of 15%-points
- Return process follows a multivariate Normal distribution with assumptions for …
    - … expected returns
    - … expected variance
    - … expected correlation

Note: The asset process is chosen to be very basic to produce a foundational study. In the real-world application and in our further research, we increase the complexity.



Environment, Actions, Rewards

Environment



Action Space

Choose appropriate risk level

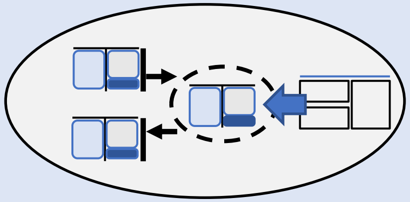
Reward

…

- State: Initial balance sheet and income statement items
- Actions: Risk level. *Note: We do not optimize specific asset classes.*
- Environment: Development of the financial markets and the liabilities side (insured portfolio)
- Reward: E.g. solvency of the pension fund
- State: New balance sheet and income statement items

Environment, Actions, Rewards

Environment



Action Space

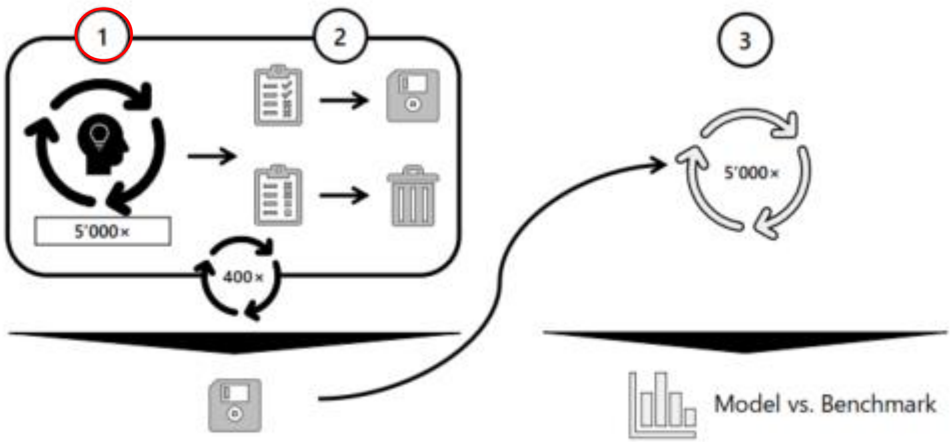
Choose appropriate risk level

Reward

Minimize the net-contribution of the employer

- State: Initial balance sheet and income statement items
- Actions: Risk level. *Note: We do not optimize specific asset classes.*
- Environment: Development of the financial markets and the liabilities side (insured portfolio)
- Reward: With the policy we want to achieve the lowest net-contribution (recovery payments – back payments)
- State: New balance sheet and income statement items

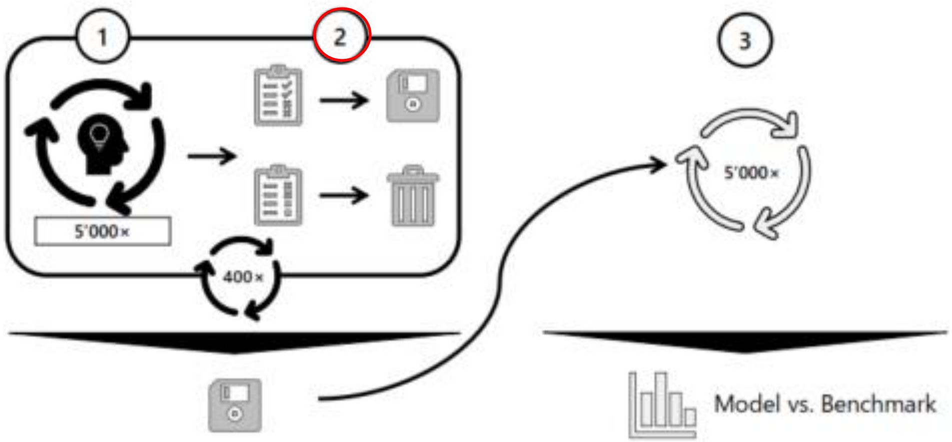
Training Process (I/II)



1

The Agent ...
... makes decisions, ...
... receives rewards, ...
... observes new states, ...
... until the end time T is reached.
This results in an episode.

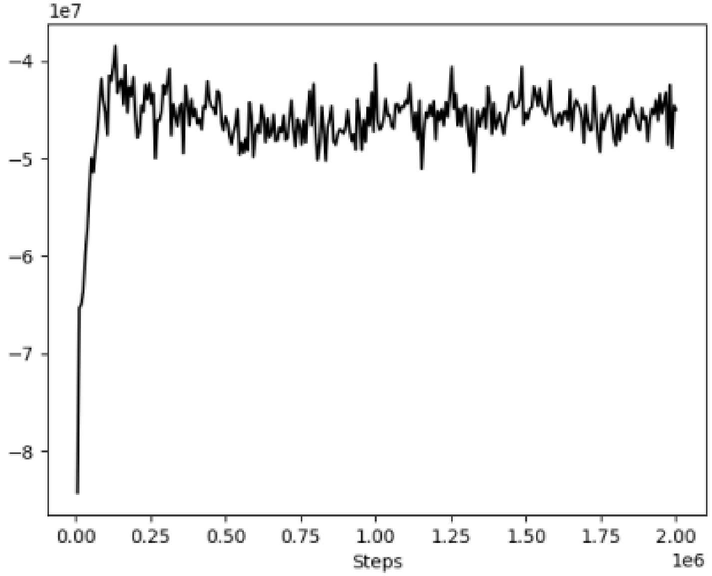
Training Process (II/II)



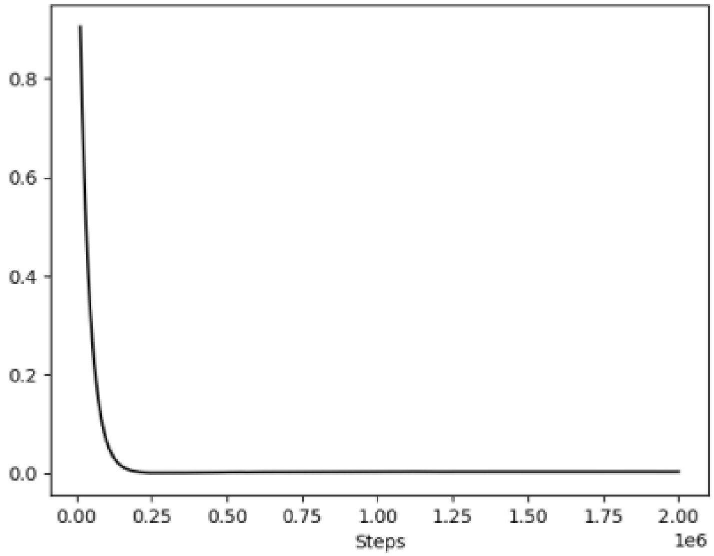
2

The Learner ...
... Lets the agent play in N episodes, ...
... collects the experience of the scenarios, ...
... trains the agent to make better decisions
(i.e., optimizes the parameters of the NN) ...
... and saves the new policy if it performs better then the last saved policy.

Training Results

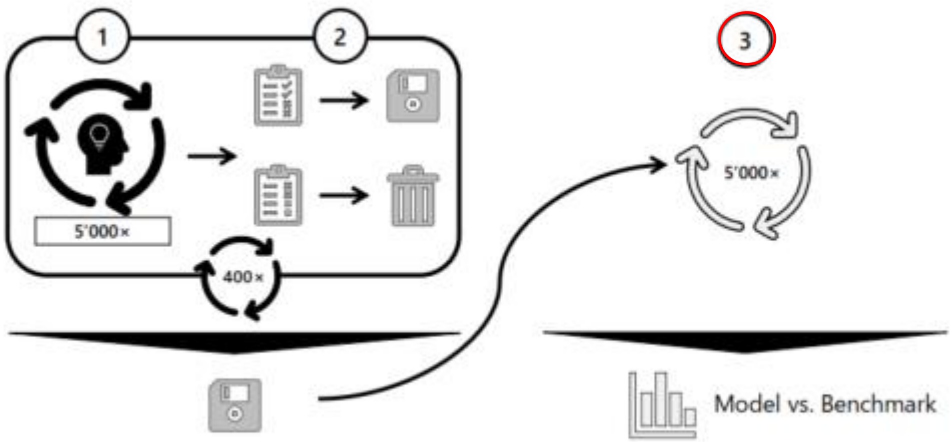


Mean reward of episodes



Entropy coefficient

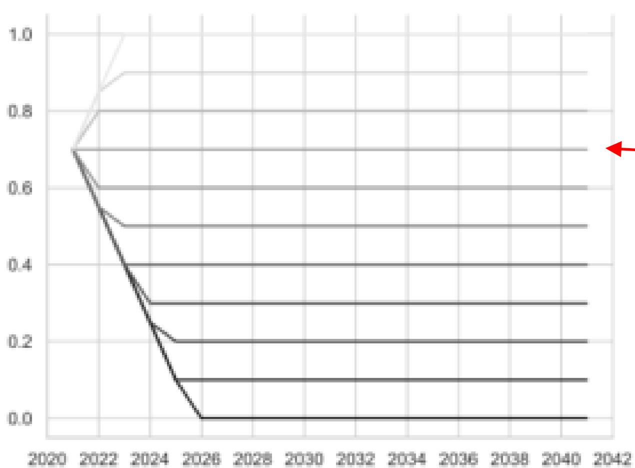
Generating the Output



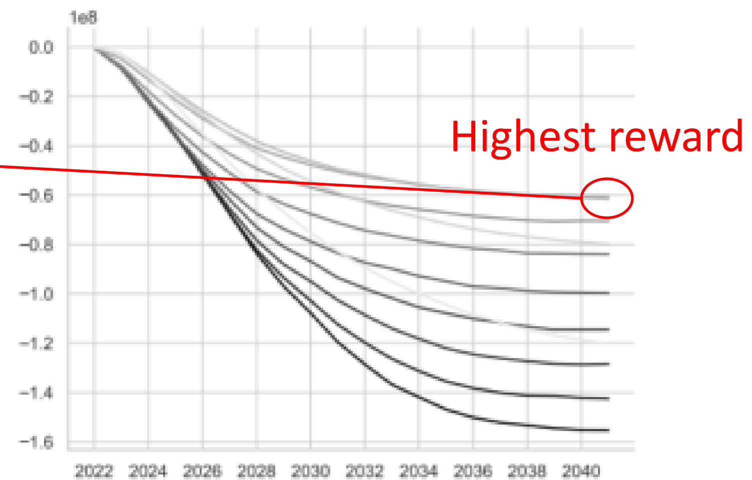
3

We let the best policy run on 5'000 new scenarios and ...
... compare them to benchmark models ...
... analyze the economic behavior of the agent in these scenarios.

Benchmark – Best Constant Action (BCA)

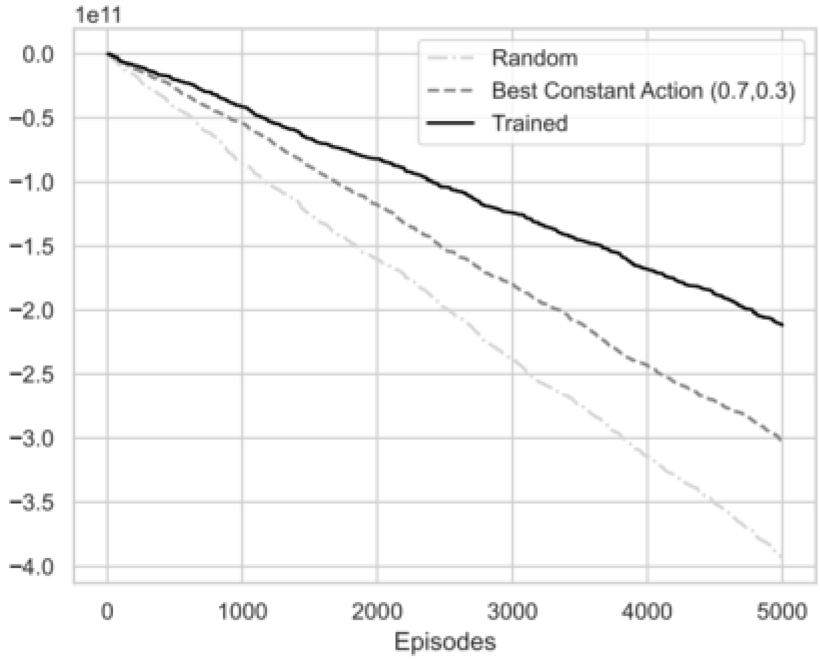


Constant action in ten percentage point steps



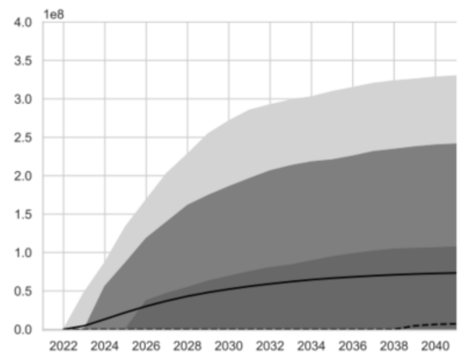
Mean cumulative reward over the episodes

Comparison over Test Scenarios

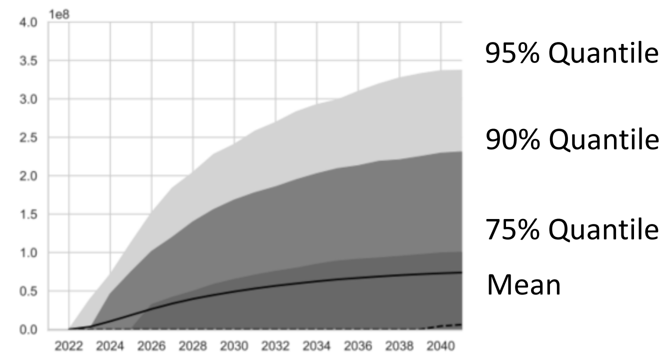


Cumulative rewards over scenarios

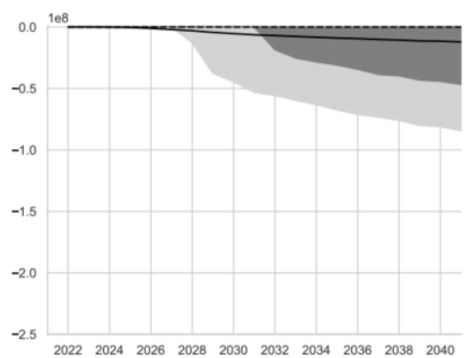
Recovery and Back Payments



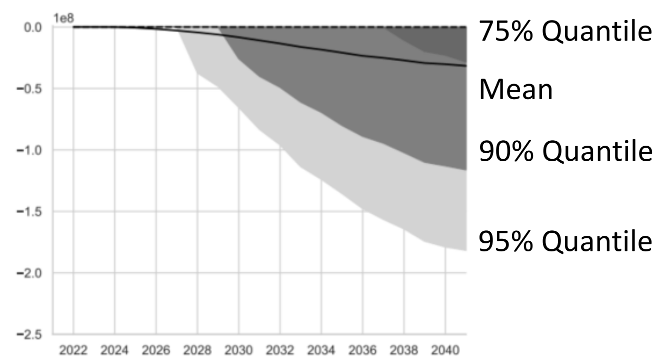
BCA: Cum. rec. pay. [in 100 mill.]



Agent: Cum. rec. pay. [in 100 mill.]

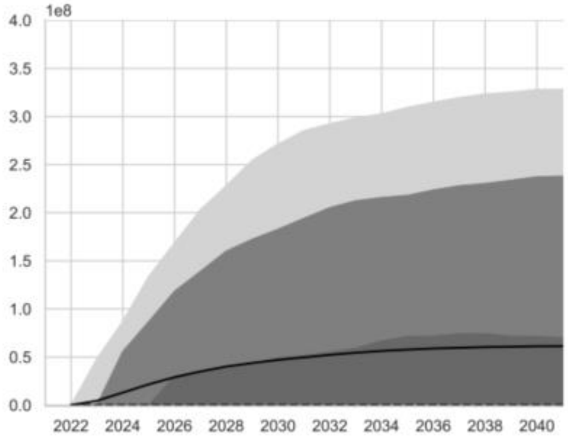


BCA: Cum. back pay. [in 100 mill.]

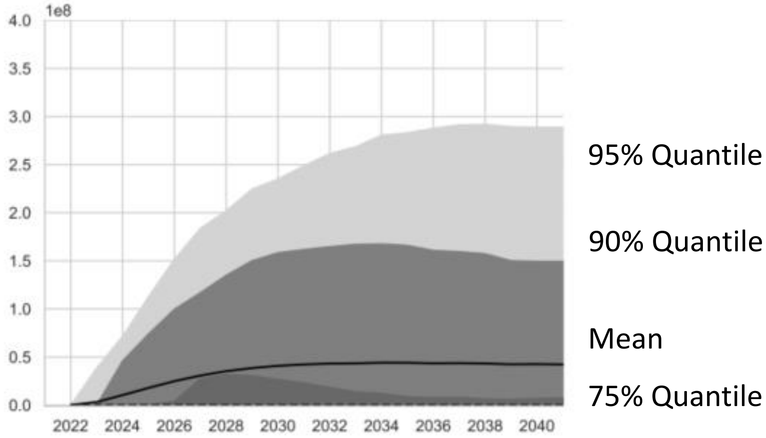


Agent: Cum. back pay. [in 100 mill.]

Net Payments

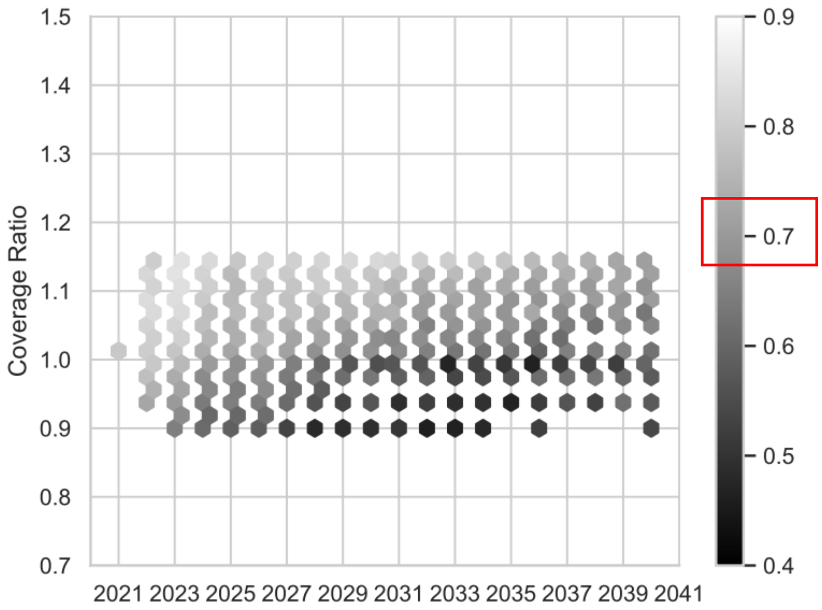


BCA: Cum. net pay. [in 100 mill.]



Agent: Cum. net pay. [in 100 mill.]

The Policy

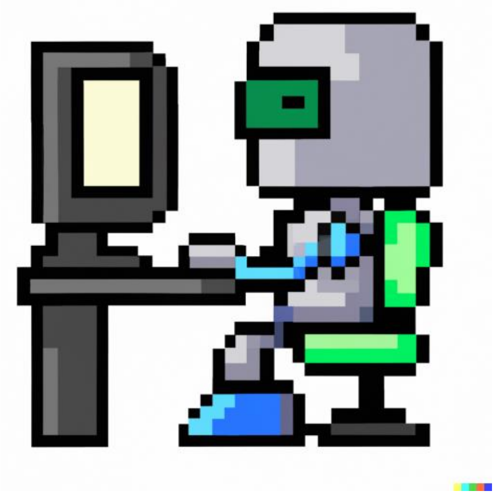


Low risk

Best constant action

High risk

What to do with these results



Agenda

Introduction

Pension Funds

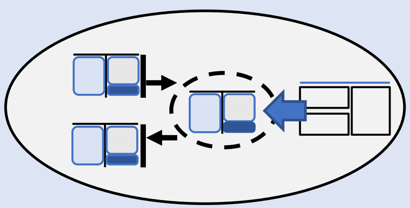
Reinforcement Learning

Application

Conclusion

Where to Go

Environment



Action Space

...

Reward

...

Environmental Complexities

Asset dynamics

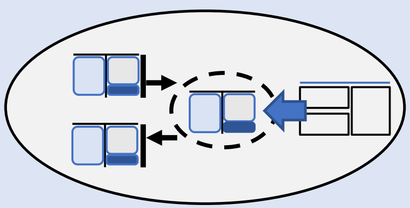
Liability dynamics

Regulatory framework

- Open pension funds
- Life insurer
- Non-life insurer
- Energy companies

Where to Go

Environment



Action Space

...

Reward

...

- Other types of processes (e.g., mean reverting)
- Other types of assets
- Duration- and/or Cash Flow matching throughout

Environmental Complexities

Asset dynamics

Liability dynamics

Regulatory framework

Where to Go

Environment	Action Space	Reward
...	???	...

- Optimal participation mechanism
- Optimal recovery mechanism
- More finely tuned asset allocation

Where to Go

Environment	Action Space	Reward
…	…	???

- Solvency ratio
- Welfare (of insured), participation mechanism
- Time spent below 100%
- Minimize Risk

Conclusion

- The proposed model is flexible and can be applied to different problems:
  - We define an arbitrarily complex environment.
  - We define an agent and its possible actions.
  - We define a reward function that rewards or punishes the agent based on its decisions.
  - The agent trains its behavior in countless trials.
  - The agent's behavior is assessed by benchmark comparison.

- Liability cannot be delegated to agent, but the presented study can be a valuable tool for consultants and experts.

Q & A