Innovative Approaches to Pension Planning: Generative Models to Multi-Agent Systems in a Heterogeneous Environment

Fatih Özhamaratlı

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

of

University College London.

Department of Computer Science
University College London

January 22, 2024
I, Fatih Özhamarath, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.
Abstract

The shift from Defined Benefit (DB) to Defined Contribution (DC) pension schemes, along with changing demographics and work patterns, has made the task of planning and modelling pension ecosystems more complex. This research focuses on modelling the pension ecosystem using novel methodologies to plan for retirement finances. Initially, an exploratory analysis of the factors shaping the pension ecosystem was conducted, focusing on age and income dynamics; subsequently, a generative model utilising a joint distribution of age and income was devised. Reinforcement Learning (RL) and Deep Neural Networks were applied to train the lifetime portfolio optimisation and saving strategy selection policies for contributors with varying age and income trajectories. Calibrated Agent-based models (ABMs) of the pension environment were used to accommodate increasing heterogeneity. In the next phase, pension modelling was explored using an RL-optimised Multi-Agent System of the Pension Ecosystem, involving different actors. This approach enables our model to exceed the limits of hard-coded environment dynamics by allowing endogenous market dynamics and interactions between agents as part of the trained model. This research demonstrates how a set of increasingly complex methodologies can effectively address the challenges of a complex and heterogeneous pension environment, providing valuable insights for financial planning and policy-making. This makes the transition from a “one-size-fits-all” approach to personalised solutions possible.
Impact Statement

The research presented in this doctoral thesis impacts the domains of financial planning and policy-making. It explores the relationship between age and income, as well as the governing dynamics, which are crucial to understand and account for in order to address the problems caused by demographic shifts (Chapter 4).

Addressing the urgent need for heterogeneous strategies in the pension ecosystem is critical due to the paradigm shift from Defined Benefit (DB) to Defined Contribution (DC) schemes, and due to the emergence of diversified working modes and varied income trajectories. This research represents a shift from the prevalent “one-size-fits-all” approach, advocating for heterogeneous investment and saving strategies that reflect individuals’ realities such as their occupations. This is achieved through utilising deep reinforcement learning for investment and savings strategies that can learn complex and non-linear policies addressing an increasingly complex economic environment (Chapters 5 and 6).

In this research, the innovative employment of Multi-Agent Systems, enhanced with Deep Reinforcement Learning in an efficient and structured way, marks a novel approach to financial modelling. This methodology captures the agent interactions within the financial ecosystem and captures market dynamics endogenously. Bridging MARL with Financial Agent-Based Models is a novel approach. The approach extends beyond pension systems and provides a robust, adaptable tool for scenario analysis and policy design for planning and forecasting in the financial domain, where the interactions between agents are fundamental (Chapter 7).
Acknowledgements

Embarking on this journey has been one of the most profound and challenging experiences of my life, and it would not have been possible to navigate without the support of various individuals whose encouragement and guidance have been invaluable.

First and foremost, I want to express my deepest appreciation to my advisor, Prof. Paolo Barucca, whose expertise, understanding, and patience added considerably to my graduate experience. Your scholarly guidance and insightful criticisms. Your role as a mentor was instrumental in this academic journey.

I extend heartfelt thanks to my secondary supervisors, Prof. Giacomo Livan and Prof. Fabio Caccioli. Their invaluable insights, detailed feedback, and critical yet constructive critiques have been instrumental in shaping my research.

I wish to extend my gratitude to the examiners present at my First Year Viva and Transfer Viva, whose feedback was instrumental in refining my research focus. The constructive criticism and objective questions posed by Prof. Jun Wang, Prof. Paul Ormerod and Prof. Giacomo Livan were beneficial. I extend my special gratitude to the examiners Prof. Jiahua Xu, and Prof. Nicola Perra for the valuable discussions throughout my viva.

I am grateful to my co-authors Dr. Oleg Kitov from Cambridge University and Prof. Paolo Barucca, whose brilliant insights, rigorous critiques, and broad knowledge immensely enriched my work.

My colleagues and fellow researchers of the Financial Computing and Analytics Group at University College London, as well as colleagues of various groups of the Computer Science Department that I shared office at University College London
Acknowledgements

have contributed immensely to my personal and professional time. The group has been a source of friendships as well as good advice and collaboration.

I extend my sincere gratitude to Prof. Treleaven, Prof. Wang, and all other people who had conversations with me regarding my research for their invaluable input and stimulating conversations related to it. The discussions we’ve had have been profoundly insightful and have helped shape my work in meaningful ways. Thank you for engaging with me and my research with such interest and depth. I thank ConceptionX programme for supporting me by providing networking opportunities with professionals from the industry who had insights that I benefited during my research.

I want to express my deepest gratitude to my family for their continuous support and belief in me. They have been there throughout my entire journey, giving me energy and resolve through their encouragement. I also extend heartfelt thanks to my family for their patience and understanding through this long journey.
Publications

Chapters 4 and 6 of this thesis were initially submitted as individual manuscripts that have been peer-reviewed and published.


## Contents

1 Introduction 20

1.1 Introduction ......................................................... 20

1.1.1 Origins and Significance of the Topic .......................... 20

1.1.2 Primary Research Questions .................................. 22

1.1.3 Approaches for Addressing the Research Questions ......... 23

2 Background 25

2.1 Background .......................................................... 25

2.1.1 Foundational Concepts and Motivations Behind the Interest 25

2.1.2 Other Approaches .............................................. 26

2.1.3 Bridging the Research Gaps: Impacts and Beneficiaries ... 27

2.1.4 Rationale for the Chosen Approach .......................... 27

2.1.5 Essential Concepts for the Non-Specialist .................. 29

2.1.6 Success of Reinforcement Learning and Deep Neural Networks in Complex Modelling ...................... 32

3 Literature Review 34

3.1 General Overview of Literature Landscape ..................... 34

3.2 Pension Ecosystem .................................................. 36

3.2.1 Government ...................................................... 45

3.2.2 Pension Funds, From DB to DC, and Saving Problem .... 49

3.2.3 Markets .......................................................... 59

3.2.4 Preferences ..................................................... 64
3.2.5 Consumption .................................................. 69
3.2.6 Data .............................................................. 70

3.3 Agent Based Models and Deep RL ......................... 71
  3.3.1 ABM .......................................................... 72
  3.3.2 ABM with ML ............................................... 78
  3.3.3 RL ............................................................. 79
  3.3.4 MARL ........................................................ 83
  3.3.5 Efficient Computing Framework ......................... 85
  3.3.6 The Approach ............................................... 87

4 A Generative Model for Age and Income Distribution 89
  4.1 Introduction ..................................................... 89
  4.2 Methods .......................................................... 92
    4.2.1 Defining Income and Age Dynamics ...................... 93
    4.2.2 Data .......................................................... 93
  4.3 Data Processing and Calibration ............................ 96
    4.3.1 Fitting Distributions ...................................... 97
    4.3.2 Estimation for Generalised Method of Moments (GMM) .. 98
    4.3.3 Estimation of Least Squares for Micro Data ............ 98
  4.4 The Generative Model .......................................... 98
  4.5 Parameters ...................................................... 99
    4.5.1 GMM ........................................................ 99
    4.5.2 LSM by Individual Transitions .......................... 100
    4.5.3 Wave-Specific Analysis .................................. 101
    4.5.4 A Simple Pension System ................................ 102
  4.6 Discussion ....................................................... 102
    4.6.1 Interpretation .............................................. 103
    4.6.2 Conclusions ............................................... 106
  4.7 Supplementary Material ...................................... 114
    4.7.1 Variables of BHPS Dataset ............................... 114
    4.7.2 Model Calibration .......................................... 114
## Contents

4.7.3 Deriving the Update Equations .......................... 115  
4.7.4 Supplementary Plots of the USA .......................... 119  
4.7.5 BHPS - JDFs of Age and Income for Observed and Simu-
lated Data (LSM) ........................................ 122

5 An AI Approach for Portfolio Allocation 126  
5.1 Introduction ............................................. 126  
5.2 Portfolio Allocation and Consumption ...................... 127  
5.3 RL .................................................. 131  
5.4 Case Study ............................................. 138  
5.4.1 Merton’s Model ....................................... 138  
5.4.2 RL Model ............................................ 138  
5.5 Conclusion ............................................. 139  
5.6 Appendix: RL Trained Model Card ......................... 140

6 Heterogeneous Retirement Savings Strategy Selection with Reinforce-
ment Learning 141  
6.1 Introduction ............................................. 141  
6.2 Model .................................................. 147  
6.2.1 Optimisation Problem ................................. 148  
6.2.2 Training the Model with RL .......................... 151  
6.2.3 Agent and Environment Cycle ......................... 153  
6.2.4 Deep Policy Network for Optimal Saving, Investment and  
Liquidity .................................................. 155  
6.2.5 Behavioural Parameters of Agents ...................... 156  
6.3 The Environment ....................................... 158  
6.3.1 The Graph and Synthetic Population ................. 159  
6.3.2 Simulation Processes ................................. 160  
6.3.3 Scaling ............................................. 161  
6.4 Results ............................................... 162  
6.4.1 Labour, Income, Consumption and Wealth .......... 163
6.4.2 Saving Profiles ........................................... 165
6.4.3 Portfolio Allocation ...................................... 168
6.4.4 Discussion Regarding the Previous Work and Limitations of the Model .............................................. 176
6.5 Conclusion ...................................................... 177
6.6 Appendix ....................................................... 180
6.6.1 Appendix: Model Card ..................................... 180
6.6.2 Appendix: Neural Architecture .......................... 181
6.6.3 Appendix: Graph Plot ..................................... 182
6.6.4 Appendix: Cross-Sectional Analysis .................... 183
6.6.5 Appendix: Effects of Behavioural Parameters .......... 185
6.6.6 Appendix: Tables ......................................... 190
6.6.7 Appendix: Raw Plots ..................................... 193

7 Multi-Agent Financial Systems with RL: A Pension Ecosystem Case 195
7.1 Introduction ..................................................... 195
7.2 Design Choices of Financial Model .......................... 199
7.2.1 Actors of Ecosystem and Interactions .................... 199
7.2.2 Mechanism of Interactions ............................... 201
7.2.3 Alignment with Finance Community ..................... 203
7.2.4 Leontief Production Function ............................. 205
7.3 Architecture of Simulation .................................... 205
7.4 Training ........................................................ 208
7.5 Simulation Results ............................................. 211
7.5.1 Simulation Configuration and Initialisation ............. 211
7.5.2 Analysis of Loss Functions ............................... 211
7.5.3 Consumer Behaviour ...................................... 212
7.5.4 Business Dynamics ...................................... 214
7.5.5 Socio-Economic Indicators and Gini Index Analysis .... 215
7.6 Challenges .................................................... 216
7.6.1 Rewards .................................................. 216
7.6.2 Input to ML Models ........................................ 217
7.6.3 Calibration to Real-World ................................. 217
7.6.4 Parallelisation and Optimisation .......................... 217
7.6.5 Training Complexities ..................................... 218
7.6.6 Operational Challenges ................................... 219
7.6.7 Theoretical Challenges .................................... 220
7.7 Conclusion ...................................................... 221
7.8 Appendix ....................................................... 222

8 General Conclusions ........................................... 224
  8.1 Summary of the Study ....................................... 224
  8.2 Broader Contributions ...................................... 226
  8.3 Limitations and Future Work ............................... 226

Bibliography ....................................................... 228
List of Figures

3.1 Tree of Concepts .............................................. 37

4.1 All Years Pooled Age and Income Joint Distribution Function for
UK and USA ...................................................... 107
(a) UK Labour JDF .............................................. 107
(b) UK Total JDF ................................................ 107
(c) USA Labour JDF .............................................. 107
(d) USA Total JDF ................................................ 107

4.2 Population Pyramid for the UK Income ....................... 108
(a) UK Pyramid in 1991 ........................................ 108
(b) UK Pyramid Wave in 2008 ............................... 108

4.3 PDF Plots of the Labour Income Bins of Population .... 108
(a) PDF of UK Labour Income ............................... 108
(b) PDF of USA Labour Income ............................. 108

4.4 $q_a, \sigma_a$ and $\mu_a$ Plots for UK Labour Income .... 109
(a) $q_a$ with GMM ............................................. 109
(b) $\sigma_a$ with GMM .......................................... 109
(c) $\mu_a$ with GMM ............................................ 109
(d) $q_a$ with LSM .............................................. 109
(e) $\sigma_a$ with LSM .......................................... 109
(f) $\mu_a$ with LSM ............................................ 109

4.5 UK Labour Data Observed and Simulation All Years Pooled JDF
between Ages 25 and 55 ...................................... 110
(a) Observed Statistics ........................................ 110
List of Figures

4.6 UK Labour Data Observed and Simulation Statistics . . . . . . . . . 111
   (a) Observed Statistics . . . . . . . . . . . . . . . . . . . . . . 111
   (b) Simulation Statistics with GMM Estimation . . . . . . . . . 111
   (c) Simulation Statistics with LSM . . . . . . . . . . . . . . . 111

4.7 JDF Plots of Simulation for UK Labour Income . . . . . . . . . . . . . 112
   (a) 1995 with GMM Estimation . . . . . . . . . . . . . . . . . 112
   (b) 2005 with GMM Estimation . . . . . . . . . . . . . . . . . 112
   (c) 1995 with LSM Estimation . . . . . . . . . . . . . . . . . . 112
   (d) 2005 with LSM Estimation . . . . . . . . . . . . . . . . . . 112

4.8 $q_a$, $\sigma_a$ and $\mu_a$ Confidence Interval for UK Data LSM Estimation . . 113
   (a) $q_a$ with LSM . . . . . . . . . . . . . . . . . . . . . . . . 113
   (b) $\sigma_a$ with LSM . . . . . . . . . . . . . . . . . . . . . . 113
   (c) $\mu_a$ with LSM . . . . . . . . . . . . . . . . . . . . . . . . 113

4.9 UK Inflow Outflow Plot of a Simple Pension System . . . . . . . . . . . 113
   (a) Inflow & Outflow . . . . . . . . . . . . . . . . . . . . . . . 113

4.10 Population Pyramid for the USA Income between Ages 15-100 . . . . . . . 119
   (a) USA Pyramid in 1991 . . . . . . . . . . . . . . . . . . . . 119
   (b) USA Pyramid in 2008 . . . . . . . . . . . . . . . . . . . . . 119

4.11 $q_a$, $\sigma_a$ and $\mu_a$ Plots for USA Labour Income . . . . . . . . . . . 120
   (a) $q_a$ with GMM . . . . . . . . . . . . . . . . . . . . . . . . 120
   (b) $\sigma_a$ with GMM . . . . . . . . . . . . . . . . . . . . . . 120
   (c) $\mu_a$ with GMM . . . . . . . . . . . . . . . . . . . . . . . 120

4.12 USA Labour Data Observed and Simulation Statistics . . . . . . . . . . 121
   (a) Observed Statistics . . . . . . . . . . . . . . . . . . . . . . 121
   (b) Simulation Statistics with GMM . . . . . . . . . . . . . . . 121

4.13 JDF for 1991-1994 . . . . . . . . . . . . . . . . . . . . . . . . . . . 122
   (a) 1991 JDF of Observed Data . . . . . . . . . . . . . . . . . 122
   (b) 1991 JDF of Sim Data . . . . . . . . . . . . . . . . . . . . 122
List of Figures

4.14 JDF for 1995-1999 ........................................ 123
(a) 1995 JDF of Observed Data. .......................... 123
(b) 1995 JDF of Sim Data ................................. 123
(c) 1996 JDF of Observed Data. .......................... 123
(d) 1996 JDF of Sim Data ................................. 123
(e) 1997 JDF of Observed Data. .......................... 123
(f) 1997 JDF of Sim Data ................................. 123
(g) 1998 JDF of Observed Data. .......................... 123
(h) 1998 JDF of Sim Data ................................. 123
(i) 1999 JDF of Observed Data. .......................... 123
(j) 1999 JDF of Sim Data ................................. 123

4.15 JDF for 2000-2004 ........................................ 124
(a) 2000 JDF of Observed Data. .......................... 124
(b) 2000 JDF of Sim Data ................................. 124
(c) 2001 JDF of Observed Data. .......................... 124
(d) 2001 JDF of Sim Data ................................. 124
(e) 2002 JDF of Observed Data. .......................... 124
(f) 2002 JDF of Sim Data ................................. 124
(g) 2003 JDF of Observed Data. .......................... 124
(h) 2003 JDF of Sim Data ................................. 124
(i) 2004 JDF of Observed Data. .......................... 124
(j) 2004 JDF of Sim Data ................................. 124

4.16 JDF for 2005-2008 ........................................ 125
(a) 2005 JDF of Observed Data. .......................... 125
List of Figures

(b) 2005 JDF of Sim Data .......................... 125
(c) 2006 JDF of Observed Data. .................. 125
(d) 2006 JDF of Sim Data .......................... 125
(e) 2007 JDF of Observed Data. .................. 125
(f) 2007 JDF of Sim Data .......................... 125
(g) 2008 JDF of Observed Data. .................. 125
(h) 2008 JDF of Sim Data .......................... 125

5.1 Consumption Fraction of Wealth .................. 139

6.1 Agent and Environment ........................ 154
6.2 Policy Model .................................... 155
6.3 Wealth, Consumption and Labour Income vs Age Plot ......... 164
   (a) Initial Retirement Income According to OECD as 80% of
       Labour Income ................................. 164
   (b) Retirement Income as Constant Minimum Consumption .... 164
6.4 Liquid and Non-Liquid Asset Amounts ............... 166
   (a) Liquid Asset ................................. 166
   (b) Non-Liquid Asset ............................. 166
6.5 Mean of Income and Unemployment by Occupation at Week ...... 167
   (a) Mean income ................................. 167
   (b) Unemployment ............................... 167
6.6 Saving Rate by Occupation ........................ 169
   (a) Week ........................................... 169
   (b) Total Asset Amount .......................... 169
6.7 Non-Liquid Asset Share vs Total Asset Amount and Age ....... 170
   (a) Total Asset Amount .......................... 170
   (b) Age ......................................... 170
6.8 3d Surface Plot of Share of Non-Liquid Assets, Total Asset Wealth,
    Non-Liquid Asset Investment Rate .......................... 171
6.9 Non Liquid Investment Rate by Occupation .................. 172
List of Figures

(a) Week ................................................................. 172
(b) Total Asset Amount ............................................. 172

6.10 Neural Architecture ........................................... 181

6.11 Graph Plot ......................................................... 182

6.12 Liquid and Non-Liquid Investment Rate by Occupation at Week . . 183
   (a) Liquid Investment Rate ....................................... 183
   (b) Non-Liquid Investment Rate ................................. 183

6.13 Liquid and Non-Liquid Investment Rate by Occupation at Amount . 183
   (a) Liquid Investment Rate ....................................... 183
   (b) Non-Liquid Investment Rate ................................. 183

6.14 Liquid and Non-Liquid Assets by Occupation at Total Amount ... 184
   (a) Liquid Asset ..................................................... 184
   (b) Non-Liquid Asset .............................................. 184

6.15 Liquid and Non-Liquid Assets by Occupation at Age .............. 184
   (a) Liquid Asset ..................................................... 184
   (b) Non-Liquid Asset .............................................. 184

6.16 Total Asset, Non-Liquid Asset, Liquid Asset at Week by Consumption Utility Factor .......................... 187
   (a) Liquid Asset ..................................................... 187
   (b) Non-Liquid Asset .............................................. 187
   (c) Total Asset ..................................................... 187

6.17 3D Scatter Plot of Each Indicator Relative to the Behavioural Parameters of the Agents ............................. 188

6.18 3D Surface Plot of Share of Non-Liquid Assets, Total Asset Wealth and Non-Liquid Asset Investment Rate ......................... 189
   (a) Management ..................................................... 189
   (b) Healthcare Practitioners and Technical ...................... 189
   (c) Farming, Fishing and Forestry ............................... 189
   (d) Life, Physical and Social Science ............................ 189
   (e) All Occupations ............................................... 189
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.19</td>
<td>Mean Income and Unemployment by Occupation at Week</td>
<td>193</td>
</tr>
<tr>
<td>(a)</td>
<td>Mean Income</td>
<td>193</td>
</tr>
<tr>
<td>(b)</td>
<td>Unemployment</td>
<td>193</td>
</tr>
<tr>
<td>6.20</td>
<td>Saving Rate by Occupation</td>
<td>193</td>
</tr>
<tr>
<td>(a)</td>
<td>Week</td>
<td>193</td>
</tr>
<tr>
<td>(b)</td>
<td>Total Asset Amount</td>
<td>193</td>
</tr>
<tr>
<td>6.21</td>
<td>3d Surface Plot (Raw Values)</td>
<td>194</td>
</tr>
<tr>
<td>7.1</td>
<td>Agents and Environment Diagram</td>
<td>201</td>
</tr>
<tr>
<td>7.2</td>
<td>The Trade Module</td>
<td>203</td>
</tr>
<tr>
<td>7.3</td>
<td>Person and Business Entities and Respective Agents</td>
<td>203</td>
</tr>
<tr>
<td>7.4</td>
<td>Single Computational Graph</td>
<td>206</td>
</tr>
<tr>
<td>7.5</td>
<td>Combined Figures</td>
<td>208</td>
</tr>
<tr>
<td>7.6</td>
<td>Residual Rewards Propagating</td>
<td>211</td>
</tr>
<tr>
<td>7.7</td>
<td>Consumption Patterns</td>
<td>212</td>
</tr>
<tr>
<td>7.8</td>
<td>Asset Distribution</td>
<td>213</td>
</tr>
<tr>
<td>7.9</td>
<td>Analysis of Risk Behaviour among Consumers</td>
<td>213</td>
</tr>
<tr>
<td>7.10</td>
<td>GDP Proxy</td>
<td>213</td>
</tr>
<tr>
<td>7.11</td>
<td>Inflation Proxy</td>
<td>214</td>
</tr>
<tr>
<td>7.13</td>
<td>Time Series of Gini Index</td>
<td>216</td>
</tr>
</tbody>
</table>
List of Tables

4.1 Description of Variables from the British Household Panel Survey
   Spanning Years 1991-2008 ......................................... 114

5.1 Model Parameters ................................................... 140

6.1 Occupation and Age vs Share of Non-Liquid Investments for
   Wealth Quartiles .................................................... 175

6.2 Parameters .......................................................... 180

6.3 Occupation vs Rates ................................................ 190

6.4 Age vs Rates ........................................................ 191

6.5 Saving Rate by Occupation and Age .............................. 192

7.1 Key Decisions for Each Entity in the Model .................... 201

7.2 Input-Output Matrix of Sectors .................................. 222

7.3 Consumption Vector ............................................... 222

7.4 Model Card Parameters ............................................ 223
Chapter 1

Introduction

In this section, I discuss the importance of pension planning in the present-day landscape and outline the primary research questions of my study. I also identify the novel methodologies that address pension planning challenges and elaborate on the advantages these approaches have over classical methods.

1.1 Introduction

1.1.1 Origins and Significance of the Topic

The transition from Defined Benefit (DB) to Defined Contribution (DC) pension schemes has been a significant factor. Contributors invest a percentage of their income into DB pension schemes. These schemes define the pension benefit according to a rule, such as a certain multiplier of the employee’s final income or the average income during their employment. The pension schemes invest these contributions based on forecasts for contributions and pension payments. If there is a deficit, guarantors such as the employer finance it, meaning the risk is predominantly borne by the guarantor. In contrast, Defined Contribution pension schemes do not promise a pension benefit based on income. Instead, they function as a pot to accumulate savings, with the pension fund investing these savings during the employee’s working years and retirement. Upon retirement, the pensioner has several options: purchasing an annuity that provides a monthly income until death, gradually drawing from the pot, or withdrawing all the savings at once. In Defined Contribution pension schemes, the financial risk, investment risk and longevity risk
lie with the contributors; there is no scheme sponsor to bail out the pensioners. This transition has increased individual responsibilities regarding pension contributions and investment management. As responsibility shifts towards individuals, there is an urgent need for advanced models and methodologies to help individuals better manage their pension finances [1, 2, 3]. In addition to this, demographic changes have been playing a considerable role. As life expectancy increases, people need more financial resources to support themselves post-retirement. Simultaneously, the working-age population that supports these pensions is decreasing, creating a demographic imbalance that necessitates significant planning and novel solutions.

Another pressing challenge in the pension ecosystem is the issue of increasingly diverse income trajectories. With the rapidly growing gig economy, increased part-time work, and more frequent career shifts, retirement financial planning has become significantly more complicated. Modelling these various income paths is essential in ensuring a stable financial future for individuals in their retirement years.

Currently, in general, Defined Contribution (DC) pension schemes are Target Date Funds that determine your savings’ portfolio allocation based on your projected retirement year, utilising your age as the primary data point for lifetime portfolio allocation strategy selection [4, 5]. This approach tends to overly expose individuals to the same risks as their income, as per their profiles. It overlooks critical variables such as income history, profession, sector of employment, career progression, lifestyle choices, and risk tolerance, which can significantly influence investment strategies and outcomes. Consequently, individuals may face greater financial risk by being exposed to investment risks highly correlated with their income risks, which can potentially affect their financial stability post-retirement. There is also an immediate need to develop savings and investment strategies resilient to environmental changes, paradigm shifts such as automation, and shocks such as pandemics, especially in a non-stationary environment. Creating robust strategies requires accurate modelling of these fluctuations, a critical feature of a resilient pension system. Emerging societal and economic shifts have both motivated this research and increased its importance.
1.1.2 Primary Research Questions

The primary research questions addressed in this work include:

1. How can we develop a model for the pension ecosystem that accurately represents the complex dynamics of real-world financial systems, the heterogeneity of career paths and income trajectories, and the dynamic nature of asset allocation strategies?

2. How can we utilise historical income trajectories to provide a more nuanced understanding of each individual’s financial behaviour for better financial planning for retirement?

3. How can we account for the multi-actor nature of the pension ecosystem, including the complex interaction between individuals, businesses, and the government, in the model?

4. How can we bridge the existing gap between Agent-Based Models (ABM) in finance and Multi-Agent Reinforcement Learning (MARL) to create a methodology that captures the complex dynamics of the financial ecosystem, where the micro-interactions between agents can be calibrated to reflect phenomena observed at the macro scale?

5. How can we leverage economic models that rely on optimisation methods in econometrics and develop a new model for the pension ecosystem that processes the complex underlying dynamics?

6. How can we design and implement a computationally efficient, parallelised MARL within an ABM framework that can be made accessible and usable to the broader economic and finance community?

7. How can we ensure that the model is robust against non-stationary dynamics and black swan events?
1.1.3 Approaches for Addressing the Research Questions

In order to address the challenges of the pension ecosystem, this thesis introduces a set of models that gradually increase in complexity to address the needs of the problem it is focusing on. Initially, the thesis focuses on modelling the fundamental principles of age and income relationships, abstracted from other factors. This foundation then evolves into a simulator calibrated to panel and population surveys, capturing the dynamics of assets and income. In this simulator, employees use Deep Reinforcement Learning (RL) within an Agent-Based Model (ABM) to optimise savings and investment behaviour. The research then evolves into a more sophisticated ABM optimised with Multi-Agent Deep Reinforcement Learning. The challenges of pension planning in today’s complex and heterogeneous financial environment can be addressed by applying novel ABM and ML methodologies of increasing complexity.

In developing a multi-agent model for pension optimisation, my objective is to craft a tool that navigates through the nuanced dynamics of the pension system, rather than constructing an all-encompassing agent-based model of the entire economy. Given that my expertise is not grounded in economics, my approach centres on developing methodologies and assembling diverse agents, each designed to address specific questions and collectively build an ecosystem reflective of the pension system. This method enables the analysis of interactions among agents and liberates us from the restrictions of a hardcoded simulator, in which the simulator itself becomes the limiting factor of our model. Thus, the model is tailored to learn optimally suitable policies for various heterogeneous profiles, within the bounds set by a simulator calibrated with primary or secondary statistical data. It’s worth noting that traditional models may not adeptly navigate cyclic properties, heterogeneity, and unexpected “black swan” events, nor adapt to an environment that does not remain static. This approach facilitates a more dynamic and adaptive exploration of the interacting factors within pension systems, allowing for a robust analysis of various scenarios and outcomes. I believe that in the future, such an approach will be an important toolkit for analysing, planning for, and optimising financial phe-
1.1. INTRODUCTION

nomens, where the dynamics manifest themselves through the interactions between the parts of a system, such as varying actors of an environment.

This research bridges traditional econometric models on lifetime portfolio optimisation and consumption decision with a set of innovative machine learning and agent based models applied to pension planning setting. It represents a new step building on the traditional econometric models, that can be applied to address environments with more complex investment and income dynamics incorporating heterogeneous agents, such complexity and heterogeneity can not be captured by the classical econometric models. This thesis bridges existing inter-temporal portfolio allocation and consumption decision methodology with ABM and ML. Furthermore, this research incorporates a multi actor setting of varying types of participants in the economy and allows the market dynamics to be manifested endogenously, which provides an environment allowing interactions between agents and makes capturing phenomena such as fluctuations or cycles possible. A moonshot side goal of the research is a pathway to modelling financial systems by training agents to take actions, that collectively manifest a market ecosystem that reflects similar dynamics that we observe in the real world, without the need to have micro-data of individual actions. Strengths and challenges of the introduced novel methodologies bridging ABM, ML and RL; are highlighted in each step. The performance of the model will be compared with traditional financial planning models. The implications of the findings for individuals, businesses, government, and policymakers will also be discussed. By filling these gaps in the research, I aim to develop a more effective and comprehensive model for pension ecosystem modelling and lifetime portfolio optimisation. This research will potentially benefit individuals planning for retirement, pension fund managers, businesses, government, and policymakers. By accurately representing the complex dynamics and heterogeneity of real-world financial systems, and by developing robust strategies against changing conditions and black swan events, we can contribute to the financial well-being of individuals post-retirement, the stability of pension fund management, and the overall resilience of our financial system.
Chapter 2

Background

In this section, I explain the foundational concepts of my research, by discussing the classical approaches to pension planning while highlighting their limitations. Furthermore, I outline how the multidisciplinary focus of the research is bridging the gap between economics and artificial intelligence, constituting a significant step towards innovative solutions in financial planning for retirement.

2.1 Background

2.1.1 Foundational Concepts and Motivations Behind the Interest

Pension schemes play a significant role in ensuring the financial well-being of individuals post-retirement. Over the past few decades, a shift from Defined Benefit (DB) to Defined Contribution (DC) pension schemes has occurred globally, placing the responsibility for pension contributions and investment management on individuals [2, 6]. This change, coupled with an increase in life expectancy, a decrease in the working-age population, and the rise in diverse income trajectories due to the growing gig economy, part-time work, and career shifts, has made the planning and modelling of pension ecosystems a complex yet paramount task [3]. The successful application of Reinforcement Learning (RL) and Deep Neural Networks (DNN) across various domains has demonstrated their potential as deep non-linear approximators [7]. RL has excelled in mastering complex strategies in games, while Multi-Agent Reinforcement Learning (MARL) has effectively simulated complex
interactions among multiple agents [8]. DNNs, specifically transformers in Large Language Models (LLMs) [9], and diffusion models in image generation [10] have demonstrated their remarkable capacity to model intricate data distributions and generate high-quality, realistic outputs. The combined success of these methodologies suggests their great potential in modelling intricate financial systems, like the pension ecosystem, capturing non-linear relationships, adapting to changes, and learning from interactions.

2.1.2 Other Approaches

In the history of modelling pension ecosystems and optimising lifetime portfolios, numerous approaches have been taken by a variety of researchers. In the initial stages, the research by Harry Markowitz introduced the Modern Portfolio Theory (MPT) in 1952 [11], which focuses on the creation of an efficient portfolio through the diversification of investments. Markowitz’s work is the foundation for portfolio optimisation, paving the way for future research on the topic.

Further exploration of this topic led to the development of the life-cycle model by Modigliani [12], which proposes that individuals aim to maintain stable consumption throughout their lifetime. Modigliani’s research assumes that individuals save during their working years, when they have a higher income, and then gradually spend their savings during their retirement years.

Expanding on the life-cycle model, Merton’s Optimal Portfolio Theorem [13] incorporated uncertainty into the model and allowed for continuous decision-making. Merton’s model calculates utility through the Constant Relative Risk Aversion (CRRA) function [14], and suggests that an investor should consume a constant proportion of their remaining wealth at each point in time to maximise their lifetime utility.

As the complexity of the models increased, Gomes et al. [15] integrated consumption, portfolio decisions, non-tradable labour income, and borrowing constraints into a more enhanced life-cycle model. The model revealed that labour income acts as a substitute for risk-free asset holdings, and investing in equities results in a decrease in wealth over the lifetime of an individual. This study reinforced
the common financial advice to shift towards safer assets as individuals age.

Building on these models, Campanale et al. [16] introduced liquidity differences across assets into the life-cycle model by considering transaction costs and the different risk properties of financial assets. They were able to generate stock allocation patterns over age and wealth that were consistent with observed household behaviour.

2.1.3 Bridging the Research Gaps: Impacts and Beneficiaries

Moreover, traditional models typically overlook certain potential risks specific to different groups of pensioners and base decisions primarily on current income, without considering historical income trajectories. Lastly, these models may not effectively withstand black-swan events, as they lack the ability to adapt to changing conditions.

The purpose of this research is to address these gaps in pension ecosystem modelling and lifetime portfolio optimisation by introducing a set of novel approaches that utilise Reinforcement Learning (RL) with Deep Neural Networks. This approach, termed Multi-Agent Reinforcement Learning (MARL), allows us to model complex dynamics, heterogeneity, and dynamic strategies effectively. It provides a more nuanced understanding of each individual’s financial behaviour by considering their historical income trajectories. Furthermore, it can identify and mitigate potential risks, scale and calibrate to different scenarios, and account for the multi-actor nature of the pension ecosystem. Most importantly, the RL-based models are robust against non-stationary dynamics and black-swan events due to their inherent ability to learn and adapt in different situations.

2.1.4 Rationale for the Chosen Approach

1. Complex Dynamics of Real-world Financial Systems: Reinforcement Learning (RL) offers an advancement over classical methods as it models the complex dynamics of real-world financial systems more accurately. It captures non-linear relationships and changes over time that cannot be effectively represented using traditional mathematical formulations.
2. **Heterogeneity of Career Paths and Income Trajectories**: RL’s capacity to handle the heterogeneity of career paths and income trajectories is a significant advantage over traditional models that often assume homogeneity across profiles.

3. **Dynamic Asset Allocation Strategies**: RL supports dynamic asset allocation strategies due to its real-time learning and adaptation to changing conditions.

4. **Mitigation of Potential Risks**: RL algorithms can identify and mitigate potential risks specific to certain groups of pensioners, which traditional models often fail to address effectively.

5. **Use of Historical Income Trajectories**: In contrast to classical models that base decisions only on current income, RL models also take into account historical income trajectories. This provides a more nuanced understanding of each individual’s financial behaviour.

6. **Scalability and Calibration to Different Scenarios**: The models introduced in the text are scalable and can be calibrated to different scenarios, making them suitable for a wide range of pension fund management goals and constraints.

7. **Multi-Actor Nature of the Pension Ecosystem**: The RL models can account for the multi-actor nature of the pension ecosystem, including people working towards retirement, businesses, and government. This complex interaction between agents is key to developing robust multi-agent models of the pension ecosystem.

8. **Robustness to Black-Swan Events**: The ability to learn and adapt in different situations makes the RL models robust against non-stationary dynamics and black-swan events, which is an important aspect of pension fund management strategies.
2.1.5 Essential Concepts for the Non-Specialist

2.1.5.1 Portfolio Optimisation

Pension ecosystems and optimisation of lifetime portfolios have primarily been modelled using traditional economic theories. These include Modern Portfolio Theory (MPT) by Harry Markowitz [11], Modigliani’s life-cycle model [12], Merton’s Optimal Portfolio Theorem [13], and advanced life-cycle models developed by Gomes et al. [15] and Campanale et al. [16]. Despite their significant contribution to understanding retirement financial planning, these models often lack the complexity to capture real-world financial systems, the heterogeneity of career paths and income trajectories, and the dynamic nature of asset allocation strategies.

**Markowitz’s Portfolio Theory** seeks to minimise portfolio variance \( \sigma^2 = w' \Sigma w \), where \( w \) represents the portfolio weights vector and \( \Sigma \) is the covariance matrix of returns. It is subject to:

- The expected portfolio return being at least at a desired level \( r^* \), i.e., \( E(r) = w' \mu \) where \( \mu \) is the vector of expected asset returns.

- The sum of total weights equals 1.

In these equations, \( N \) is the number of assets in the portfolio.

**Modigliani’s Life-Cycle Hypothesis** can be summarised by:

1. *The Lifetime Budget Constraint:* Present value equivalence of consumption and resources over all periods, i.e., \( \sum \frac{C_t}{(1+r)^t} = \sum \frac{R_t}{(1+r)^t} \).

2. *Consumption Smoothing:* Constant consumption over time, i.e., \( C_t = C \).

3. *Saving and Dissaving:* Savings as the difference between resources and consumption at each period, i.e., \( S_t = R_t - C_t \).

In these equations, \( C_t, R_t, r, \) and \( S_t \) denote consumption, resources, interest rate, and savings at time \( t \) respectively.
Merton’s portfolio optimisation model maximises an investor’s utility of terminal wealth and consumption via optimal dynamic consumption and investment plan. This model assumes a two-asset case: a risky and a riskless asset. The dynamics of the investor’s wealth \(dW(t) = (rW(t) + \pi(t)(\mu - r)W(t))dt - c(t)dt + \pi(t)\sigma W(t)dB(t)\) involve:

- Wealth \(W(t)\),
- Consumption \(c(t)\),
- Risky investment proportion \(\pi(t)\),
- Brownian motion \(B(t)\).

The total wealth is the sum of wealth invested in both assets, the risky asset’s dynamics are given by: \(dP(t) = P(t)(\mu dt + \sigma dZ(t))\) and risk-free asset’s dynamics: \(dY(t) = Y(t)rdt\). The wealth invested in the risky asset is: \(X(t) = \pi(t)W(t)P(t)\).

The model aims to maximise the utility calculated via the Constant Relative Risk Aversion (CRRA) function:

\[
E\left[ \hat{\tau} e^{-\rho s}u(c_s)ds + \epsilon^T e^{-\rho T}u(W_T) \right]
\]

subject to certain constraints on consumption and stock allocation.

Merton deduces the constant optimal proportion of wealth invested in the risky asset \(\pi^* = (\mu - r)/(A\sigma^2)\) and in risk-free asset: \(1 - \pi^* = 1 - [(\mu - r)/(A\sigma^2)]\). The optimal consumption rate is \(c^* = W(1 - e^{-\rho T})/T\), with \(W\) as initial wealth, \(T\) as planning horizon, and \(\rho\) as time preference rate.

2.1.5.2 Utility and Preferences

Bernoulli’s foundational work for expected utility theory provides an essential understanding for investment decisions. If we denote wealth by \(x\) and its utility by \(U(x)\), the expected utility \(E[U(x)]\) can be calculated as

\[
E[U(x)] = \sum_{i=1}^{n} p_i U(x_i),
\]

(2.1)
where $p_i$ represents the probabilities of the possible outcomes $x_i$. The preference of $A$ over $B$ is determined if $E[U(A)] > E[U(B)]$. A risk-averse investor would favour a certain outcome with the same expected return due to the concavity of the utility function, indicating a preference for stability.

Building upon this foundation, Pratt [14] introduced measures of risk aversion, $r(x) = -\frac{U''(x)}{U'(x)}$ and $r^a(x) = xr(x)$, which provided a connection between risk aversion, utility functions, and investment strategies. Pratt also proposed measures of absolute and relative risk aversion:

\begin{align*}
A &= -\frac{U''(x)}{U'(x)}, \\
R &= -\frac{cU''(x)}{U'(x)},
\end{align*}

(2.2)

(2.3)

where $U''(x)$ and $U'(x)$ are the second and first derivatives of the utility function respectively. Higher (lower) values of $R$ signify high (low) levels of risk aversion.

The Constant Relative Risk Aversion (CRRA) utility function is expressed as:

\begin{align*}
U(c) = \frac{c^{1-\rho}}{(1-\rho)},
\end{align*}

(2.4)

where $c$ symbolises consumption and $\rho$ is the coefficient of relative risk aversion.

Epstein and Zin made further refinements to these concepts by introducing recursive utility functions. Their work made it possible to separately specify risk aversion and intertemporal substitution, which in turn facilitates a more profound understanding of consumption and asset returns over time [17]. Their utility function is formulated as:

\begin{align*}
U_t = [(1-\beta)c_t^\rho + \beta \mu_t(U_{t+1})^\rho]^{1/\rho},
\end{align*}

(2.5)

with $\mu_t(U_{t+1}) = [E_tU_{t+1}^{\alpha}]^{1/\alpha}$. This expression allows for separate specification of the elasticity of intertemporal substitution and risk aversion.
2.1.5.3 Pensions and Prospect Theory

Within the pensions context, a study by the UK Department of Work and Pension [18] argued for the consideration of complex interactions between incentives and behaviours when adjusting policies. The report emphasised the importance of incorporating behavioural factors into traditional economic models to develop more effective policies. Key behavioural factors influencing savings include inertia and defaults, framing, mental accounting, loss aversion, heuristics, and social norms. The report suggested a range of interventions such as simplification, reframing, and understanding motivations like loss aversion.

Kahneman and Tversky’s Prospect Theory models decision-making under risk using psychological principles. The theory asserts that individuals perceive outcomes as gains or losses relative to a reference point, with losses having a larger impact than equivalent gains. Other contributing factors include diminishing sensitivity to the magnitude of gains/losses and the tendency to overweight small probabilities while underweighting large ones. These factors can lead to behaviours such as gambling and over-insurance [19]. Despite its widespread application across various fields and ongoing influence, there are still unanswered questions regarding the definition of gains/losses and its general applicability outside experimental settings [20].

2.1.6 Success of Reinforcement Learning and Deep Neural Networks in Complex Modelling

The unprecedented success of Reinforcement Learning (RL) and Deep Neural Networks (DNN) in modelling intricate systems, particularly in games and complex simulations, has underscored their potential as effective deep non-linear approximators [21].

For instance, RL algorithms have triumphed in a variety of games, from classic board games such as chess and Go to more dynamic and interactive environments like StarCraft II and Dota 2 [8]. These victories are significant not only because they are an example of the capability of RL to master the strategic, tactical, and reactive aspects of these games, but also because they illustrate the adaptability of
RL in a series of applications with diverse complexities and dimensions.

Simultaneously, Multi-Agent Reinforcement Learning (MARL) has been instrumental in creating more complex models [22]. MARL, by design, simulates multiple decision-making entities or “agents” interacting with each other, making it ideal for complex simulations, such as those involving social and economic systems. Its successes in modelling intricate multiplayer games serve as compelling proof of its potential in economic modelling, including pension ecosystem modelling.

The effectiveness of RL and MARL owes much to the rapid advancements in Deep Neural Networks (DNNs) [7]. DNNs, such as transformers, have exhibited exceptional capabilities in the domain of Large Language Models (LLMs). When combined with Reinforcement Learning from Human Feedback (RLHF) [23], these transformer models have been able to generate human-like responses, demonstrating remarkable understanding and generation of natural language.

Similarly, the use of diffusion models in the field of image generation has been transformative [10]. These models, powered by DNNs, have been able to generate high-quality, realistic images by learning from vast datasets of real images. This reflects the remarkable potential of DNNs as powerful non-linear function approximators capable of capturing and modelling complex data distributions.

The wide-ranging success of RL, MARL, and DNNs, across different fields and applications, underscores their promise as tools for modelling the complexity inherent in financial systems, such as the pension ecosystem. Their ability to capture non-linear relationships, adapt to changes, and learn from interactions makes them uniquely suited to address the challenges presented by the rapidly changing and increasingly complex financial landscape.
Chapter 3

Literature Review

In this section, I explain academic works relevant to pension planning, providing a comprehensive overview of the varied literature landscape. The review covers key studies on the pension ecosystem, consumers’ preferences and profiles, as well as panel and population surveys. It includes a thorough analysis of Agent-Based Models and Deep Reinforcement Learning. This section not only reflects the breadth of research in these areas but also identifies the important works and traditional theories, stating existing knowledge gaps that this study aims to address.

3.1 General Overview of Literature Landscape

The literature landscape of this research spans multiple disciplines. At the core, this research focuses on the financial problem the pension ecosystem, which is quite comprehensive by itself. The ecosystem includes individuals with different profiles, such as people of various ages and occupations, the income dynamics during the workforce, consumption and savings decisions, shocks that affect the income, lifelong investment strategies, taxes, markets and regulations. The classical methodology to model economic and financial dynamics is econometric modelling, where the relationship between different components of the ecosystem is assumed to be captured by simple mathematical equations. Although this school of thought proved to be useful for communicating simple financial principles and practical solutions with wider audiences, the limits of classical econometric methodology were known, but it was perceived that there could be no alternative due to the technical limita-
tions of problem-solving tools, which were closed form solutions to a set of simple equations and numerical approximators with limited computing capacities.

Agent based modelling of financial systems has offered a more comprehensive way to capture multiple components of complex financial ecosystems such as pensions, where the phenomena such as income dynamics do not have to be hardcoded as assumptions in the shape of mathematical equations, but they manifest as part of the interactions between agents of the model. Agent based models have been adapted and utilised to multiple fields, but they usually share the same perils of classical econometric models, where relationships and interactions between agents are hard-coded as simple mathematical equations. Such a model, although allowing for investigation of emergent features from simple rules, lacked the power to capture the complex relationship and dynamics of environments that were modelled. In this research, Multi-Agent Deep Reinforcement Learning is used to automatically learn the relationships and nature of interactions between agents. In such a configuration, our main anchors are paths of interactions and physical principles that are being hardcoded and we use the aggregate statistics that are being observed in the real world to calibrate our model. The deep neural networks enable higher level of heterogeneity to be captured, where we can capture more heterogeneous strategies and calibrate for complex behaviour.

In the first part of the literature review, the current challenges of the pension ecosystem are covered in light of reports from various national and international institutions, and the econometric models capturing lifetime portfolio optimisation and saving strategy selection are introduced. The components of the pension ecosystem such as people, businesses, markets, funds, government, central bank and the relationship between those are explained.

In the second part of the literature review, the methodologies other than econometric modelling are investigated, where micro simulations, agent based models, efficient computation infrastructures, machine learning algorithms, further capabilities with Deep Learning, and capabilities of reinforcement learning, and settings with multi-agent reinforcement learning algorithms are explored.
3.2. PENSION ECOSYSTEM

The hierarchical tree of concepts that are being covered in this research can be found on 3.1.

3.2 Pension Ecosystem

There has been extensive reporting regarding the state of the pension ecosystem, the empirical evidence on the savings data, and how critical measures such as replacement rate, pension contribution distribution, party of responsibility and investment strategies change reflect the challenges regarding the retirement finances. In this section, the state of the pension ecosystem is explored to understand why optimising lifetime saving and investment strategies is vital for pensions and how an agent based model that can capture the heterogeneity and complexity of this problem can address the challenges of the pension ecosystem.

The Pensions and Lifetime Savings Association publish reports on retirement; their analysis [24] suggests that automatic enrolment will deliver a real improvement in retirement outcomes, but there is a significant gap in ensuring adequate retirement income. The report suggests that “despite this progress, as things stand, many individuals are not on target to attain the Pensions Commission’s definition of an adequate retirement income – 67% of pre-retirement income for a median earner. This is not, in the main, the result of the current policy’s failure to deliver its objectives. The target for statutory minimum contributions, 8% of qualifying earnings, was intended to achieve a replacement rate of around 45%, with the remaining 15-22% being made up of additional voluntary contributions. Rather, many people are not on track to achieve an adequate retirement income due to a combination of past developments, such as the gradual decline of DB pensions from the mid-1990s; the failure of attempts to stimulate voluntary saving in the 1990s and early 2000s; rising longevity; and the impact of relatively poor market conditions.” The report mentions significant pitfalls of the current system and labels the decline of Defined Benefit Pension Schemes, poor market conditions and rising longevity as significant factors contributing to the deteriorating pension system.

The researchers [3] also have scrutinised the retirement period of the pen-
3.2. PENSION ECOSYSTEM
3.2. PENSION ECOSYSTEM

Pensioners and distinguished three phases, namely the Independent Phase, the Decline Phase, and the Dependent Phase. The Independent Phase is the initial part of retirement where the individual has only minimal limitations, the Decline Phase is the next phase where the individual experiences at least one mild physical limitation and might experience a decline in physical and cognitive capabilities. The Dependent Phase is the last phase where at least one severe physical limitation, which means that the individual might need substantial support and maybe to move into care facility; this phase also increases the risk of social isolation. Understanding the phases and needs of retirees guides us to plan their financial requirements arising from not a simple desire to consume more but a necessity of declining health and need for support, which also puts in some cases pressure on state finances in addition to personal retirement finances.

Significant findings [25] about the current dynamics shaping future retirement incomes are listed as follows: “The decline in private sector Defined Benefit (DB) provision, reductions in the proportion of future income from State Pensions, and the increase of casual working, mean that Generation X is likely to reach retirement with less income from sustainable sources than those in older generations. A decrease in house purchases among this cohort, a greater likelihood of indebtedness and an increase in the likelihood of the need to provide or receive care at older ages means that those in Generation X are likely to have higher expenditure needs on average than older cohorts, which will further reduce their disposable income and make it more difficult to achieve a suitable standard of living in retirement.” The findings regarding the nature of the employment and governing market dynamics for Millennials are as follows: “Millennials are less likely to reach retirement with DB entitlement but will have greater DC savings as a result of benefiting for longer from automatic enrolment. Millennials are most likely of all generations to work casually or be self-employed but could benefit from future policies designed to assist those outside of the full-time employed workforce to save for retirement. Millennials are the least likely of any generation to reach retirement owning their own home outright, though future policy or economic changes could change the preva-
ence of house buying or the way that benefits are used to support those renting in retirement.”

A similar picture is drawn by the report [26] from PLSA: “Many current retirees have very little wealth and there is a broad division between the amount of income and assets possessed by different groups within the existing generation of retirees. Future generations of retirees are, however, much less likely to have sufficient assets to generate an adequate retirement income. The decline of Defined Benefit (DB) schemes, low Defined Contribution (DC) pension contributions and lower levels of home ownership will reduce the wealth available to them in later life. Nevertheless, 51% of savers (13.6 million people) are unlikely to meet the Pension Commission’s Target Replacement Rate (TRR) (£19,162 for a median earner in 2017) with this level of contributions. The degree to which savers are on track to meet their TRR differs by generation. Of those who will achieve their TRR, the majority in each generation have some DB entitlement. Nevertheless, for the increasing proportion of pension savers who have only DC pensions, the percentage of those likely to meet their TRR is lower (3%).”

The Organisation for Economic Co-operation and Development publishes extensive and significant reports [27], [1], [2]; where they compare different countries and thoroughly investigate the global pension ecosystem and vital issues.

A critical review of these reports is as follows: Pension Ecosystem should be analysed and studied as a whole. It is a complex system, which is constituted by Market Dynamics, Employers and Employees as Contributors, Pensioners, Financial Markets, Pension Funds and Intermediaries, Policy Makers and Regulatory Bodies. Modelling the pension ecosystem in an ABM will require these actors to be modelled fully (Contributors) or partially (Pension Funds) or at least accounted for (Financial Markets). The research mostly focuses on modelling the contributor behaviour of withdrawal and contributions with a perspective of individually personalised income trajectory inference. Pension Funds are required to maintain certain amounts of liquidity at all times. Furthermore, currently the cash flow forecasting for pension payments as liabilities is the prominent planning tool of pension funds,
external shocks to financial markets or financial crashes, are affecting the fund’s ability to address these liabilities. According to the Liability Driven Investment Strategy, the pensioners’ pension income is a liability to the pension fund. Cash flow hedge, according to Dedicated Portfolio Theory, complements this investment strategy. Pension Funds are utilising an Integrated Risk Management framework for assessing the risk associated with their assets, liabilities and investment strategies.

This phenomenon is also observed by COVID-19, some people were raiding retirement accounts amid COVID. Pensioners were draining pots, and this will have an effect on lower future cumulative pot and earnings. External shocks are particularly damaging towards individual funded DC schemes, especially if a shock occurs towards the end of the period. There is a clear trend of transition from defined benefit schemes to defined contribution schemes. DB Schemes have scheme sponsors as ultimate guarantors of the schemes. Many DB schemes deficit require their sponsors to bail out. Employers prefer DC schemes because the risk and responsibility of managing funds, longevity risk, and market risks are transferred to individuals in DC schemes.

Furthermore, employer contribution decreased from 16.5% at DB to 6.5% at DC. Not only the pension fund’s underlying assets are affected, but the pensioners’ behaviour of withdrawal is also affected by external shocks, such as a crisis that might result in early withdrawal of financial assets from the pension pot due to urgent needs, psychological effects, or concerns regarding financial system. The effects of the external shocks on the contributor and pensioner behaviour change according to the demography, profession and background of the individual. There are three ways to use funds in pension funds during retirement: At the time of retirement a person can purchase an annuity with pension funds, and it will transfer the longevity risk to the annuity provider, but usually, the annuities are expensive relative to their monthly returns, a second way is fixed rate withdrawal, as withdrawing a few per cent each month, the third way is withdrawing a fixed amount each month. The pensioners are free to withdraw even their full savings from a pension pot, but efficient taxation requires a balanced withdrawal behaviour. The post-retirement
support system in the UK consists of basic state benefits, secondary benefits from the state, pension credit, universal credit and occupational pension benefits, defined pension benefits, and defined contribution pension benefits, private pensions.

In the UK, the unwritten rule of Triple Lock is applied to the basic state pension; the triple lock means the higher of the 2.5%, inflation rate or average earnings growth is used to increase and update the basic state benefits each year. The universal provision is part of the public safety net and cannot be borne by the private sector. Universal basic income and basic state benefits are the last resort and safety net for post-retirement. A significant problem of our time is the low interest rates, which means pensioner savings cannot be invested in relatively higher yield safe assets. The important question in the pension system is: Who finances my retirement? Are my past savings financing my retirement income, or are the savings of current contributors financing my retirement? In the second case, it means that the population pyramid and the ratio of contributors/pensioners are critical for the self-sustainability of the system. Many state-funded and sponsored pension plans are experiencing the consequences of the changes in the population age pyramid. This is the most striking effect of changing demography on the pension system. The rising longevity (long life expectancy and pension period) and increasing dependency ratio are the outcomes of the changes in the demographic context. Another reality in the pension ecosystem is low savings in general. ONS Statistics 60% of the private sector workforce is not saving towards a pension as of 2012. The behavioural context of this problem is that the savings context is not strong, but retirement is not a period people would like to go through with insufficient funds. OECD defines the reasonable income at retirement as 70% of the wage. Historically, the relationship between Pension Funds, pensioners, and contributors has been marked by scepticism. Issues of trust within companies are “toxic”. Additionally, the system’s opacity, coupled with hidden costs and charges, further fuels this mistrust. The high quality and profitable management of the pension funds are paramount because if poor quality management absorbs the people’s savings, it will lead to high levels of opt-outs.
There has been increasing criticism regarding the investment strategies of pension funds, and there are fundamental challenges to pension funds regarding how they generate wealth. In principle, the critics claim that pension funds are funds with very long term horizons, they should act so as well. Today many funds are actively trading with short term horizons. Managing a pension fund with a short term horizon forces fund managers to try to make gains in the short term by participating in active trading with more than necessary frequency, and this can cause substantial transaction costs, which means less future holdings. A new approach suggests the funds to invest with long term vision in long term projects and should participate in the long term investments that also create positive value with environmental and social benefits. The belief that “The system is too complex for individuals to handle by themselves” is causing some individuals to withdraw from the system. Retiring people are choosing a lump sum instead of an annuity due to distrust. As part of efforts to help pensioners and contributors to track their assets divided into various pension schemes during their employment, a new tool was introduced in 2017. The Pension Dashboards act as a platform to aggregate all of an individual’s pension data, and this is part of policy-making efforts to make the system more transparent and accessible. Pension Dashboards are not yet fully functional, but the integration and participation to these dashboards are increasing. The efforts for multi-agent modelling of the pension ecosystem must account for various variables of the ecosystem. First of all, the model should account for the fragmented nature of the pension provision in retirement; this is a combination of public and private pension supplementary schemes. The efforts of the research focus on DC scheme contribution, while acknowledging the state benefits. The DB schemes are out of the scope of this research. A key feature that should be investigated is the improvement of the design of financial incentives for promoting savings for retirement. A non-negligible part of the efforts to raise savings is the cost of the pension schemes during the accumulation phase, the cumulative effect of these costs is significant. Improving pension incomes requires consideration of behavioural biases and limited financial knowledge of the individuals. A system-wide factor affecting saving
behaviour is the mortality differences across socio-economic groups and the implications of outliving the pension funds.

Pension Fund Management involves costs of investment and administration activities. The asymmetric information and behavioural biases are causing the market mechanisms not to be sufficient to align charges, and this raises the actual costs. Simple interventions such as automatic enrolment can contribute significantly towards saving for post-retirement. Some of the structural solutions that can be investigated by an ABM are the adjustments to the pricing regulations of pension Funds, structural adjustments to the access to saving instruments, and investigating gains of a transparent system. An important aspect of the efforts to improve the system should focus on flexibility for various social, and economic groups according to the life expectancy and income trajectories. OECD encourages a mixed approach of PAYG, funded, public and private pensions; the consumption smoothing can be achieved via raising the contributions to PAYG, public or private pension schemes. Furthermore, the redistribution can be achieved via a mandatory PAYG pension system. An agent-based model of the pension ecosystem is a great way to assess the objectives of consumption smoothing and redistribution. Other primary design features of pension systems are balancing mandatory and voluntary participation, defining the sources of benefits by accumulated assets or current contributions of the workforce, the decision of assigning the responsibility and risks to which actors (DB vs DC). Such a system should also balance and track the pension system objectives and the associated risks; a balance is required for multiple and competing objectives, and this could be achieved by sustaining financial security, ensuring “Adequate” retirement income and constituting a financially sustainable system.

It is clear from the literature that the current and future pension ecosystem is a complex one, filled with challenges and changing dynamics that threaten the adequacy of retirement income. The complexity arises from the diversity in demographics, differences in career trajectories, variability in income levels, as well as uncertainties in life expectancy and market returns. Such complex systems are well-suited to be modelled using Agent-Based Models (ABMs) due to their ability
to simulate interactions between a multitude of heterogeneous agents, under different scenarios and capture emergent behaviours.

This thesis demonstrates a set of novel methodologies to address the challenges in the pension ecosystem.

The demographical shifts identified by the OECD, IMF and governmental reports introduced in the previous parts indicate a significant challenge of planning pensions for an ageing society. The “A generative model for age and income distribution” section of the thesis investigates income distribution and its relationship with age, identifies a stable joint distribution function, and calibrates it using panel data, which can be used to forecast and plan future income trajectories as well as pension savings.

The transition from Defined Benefit to Defined Contribution Pension Schemes means that the responsibility for financing retirement is now on the shoulders of employees, and current data reflect that the DC funds are failing to match the contribution levels of the DB funds. Furthermore, the increasingly turbulent environment that introduces different kinds of shocks such as COVID or automation and increasingly heterogeneous income trajectories mean that there is a very urgent need for more capable models that can provide savings and investment strategies for an increasingly complex environment and for increasingly heterogeneous income and consumption trajectories. The section “Heterogeneous Retirement Savings Strategy Selection with Reinforcement Learning” provides a model that utilises deep neural networks for representing a complex policy function for savings and investment decisions, and uses RL to learn optimal investment policies in an environment that is calibrated for heterogeneous income trajectories and market dynamics.

In the Next Sections: The pension ecosystem is composed of different parties, the next sections will scrutinise the actors of the ecosystem such as the Government, Businesses, and Employees working towards retirement. The markets manifest by the interaction of these actors and the known phenomena such as business cycles, input output relationship between actors, and exogenous/endogenous shocks. The concepts governing retirement finances such as Lifetime Savings and Investment
3.2. PENSION ECOSYSTEM

Decisions, Income and Consumption Dynamics, as well as Preferences will be introduced.

3.2.1 Government

In this section, the role of government is explored. The government has a vital role in the pension ecosystem with the multiple functions it holds. The regulation of pension contributions is the most direct way the government can intervene in retirement finances, these regulations mostly manifest themselves as minimum contribution limits for the employee and employer. Another level of government intervention is taxation structures that incentivise savings towards retirement. A softer toolkit that governments use is the default pension system participation with opt-out options, which affects employee saving behaviour. A key role of the government is providing pension credits or state pensions that acts as a minimum pension income for people who have difficulty financing retirement.

Regulating how pension funds can be invested and stipulating conditions for risk and liability responsibilities are other ways the government intervenes to protect retirement finances from excessive risk and shocks.

A different perspective can be provided from the fiscal and monetary policies of government which affect the savings and investments of the contributors, as well as the retirees that are considering annuities, and financial shocks that retirees can experience.

One of the overlooked toolkits and responsibilities that the government has is the financial education of the employees and retirees to make optimal decisions with their respective risk profiles.

The agent based model of a pension ecosystem capturing also the government as an actor making a diverse set of decisions affecting the pension ecosystem in general is the next step for being able to model and optimise the pension savings, in an increasingly complex and diverse setting.

Government policies, regulations, and tax rules around pensions have made decisions about how to best access retirement savings increasingly complex over recent years. For example, the introduction of pension freedoms in 2015 gave peo-
3.2. PENSION ECOSYSTEM

People more flexibility but also more responsibility in deciding how to utilise their Defined Contribution savings [3] according to a report from the Pensions Policy Institute (PPI). To help improve outcomes, the Financial Conduct Authority (FCA) is introducing investment pathways to guide people towards appropriate investment strategies when accessing drawdown. Nevertheless, many people still struggle to make optimal choices about their savings. This suggests an ongoing need to balance flexibility with soft defaults, guidance and advice to support decision-making. Meanwhile, safety nets like pension credit remain important for protecting incomes in retirement, especially for those with limited resources, though uptake of these benefits remains low. Overall, government policy has shifted towards greater pension freedom and choice, but this has increased the complexity of decisions people face, requiring a robust policy response to help citizens make the most of their retirement savings.

The COVID-19 pandemic has highlighted how government policies and regulations around pensions and benefits can have a disproportionate impact on underpensioned groups [25]. While palliative measures like the furlough scheme helped minimise immediate income loss, extra support may be needed to aid economic recovery among those facing greater barriers to securing employment. Sufficient working-age benefits are imperative to protect financial resilience too. With underpensioned groups more reliant on the State Pension, changes to uprating policies could also unequally affect them. Meanwhile, reforms to pension tax relief, which currently favours higher earners, could better redistribute this to lower income groups and improve their retirement outcomes. Overall, policymakers must carefully consider equality impacts on underpensioned groups in order to effectively support their retirement security.

Government policies and regulations have a major influence on pension behaviours and outcomes [28]. For example, 401(k) participation is far higher than IRA participation, likely due to auto-enrolment and payroll deduction in 401(k)s. Nondiscrimination testing requires equitable tax treatment, yet in 2003, lower-paid workers contributed 6% on average versus 9% for higher earners. When given
a choice between a lump sum and an annuity worth 17.5-19.8% in returns, 92% of enlisted personnel and 52% of officers took the lump sum, forfeiting $1.7 billion in value. Furthermore, limiting investment options to 2-4 funds versus 10 funds boosted 401(k) participation rates by 1.5-2 times in one study, suggesting that too much choice reduces participation. In summary, regulations and tax policies strongly shape participation, contribution rates, investment behaviours and retirement options. Policymakers must weigh participant freedom against the need for structure, given behavioural tendencies that lead to suboptimal outcomes.

Tax incentives play a pivotal role in stimulating the growth of private pensions by permitting tax-deferred compensation [29]. Estimates suggest workers can reduce lifetime tax liabilities by 20-40% by exploiting pension tax deferral optimally. Nevertheless, non-tax factors like productivity benefits from enhanced employee retention also affect pension coverage. Empirical studies find a 0.4 percentage point increase in coverage rates for every 1 percentage point rise in marginal tax rates. Pension tax incentives may also indirectly influence personal saving by changing pension types offered – for instance, 401(k) plans likely partially substituted for conventional defined benefit pensions. Additionally, government deficit policies influencing national saving can alter the generational impact of pension tax incentives. Thus, both tax and non-tax policies shape pension formation, while the effect on private saving hinges on whether pension saving displaces other savings.

Taylor’s research [30] dives into how new studies on monetary policy rules - how we adjust interest rates in response to things like inflation or economic output - can be used in real-world decisions. But, it’s not always practical to follow these rules to the letter.

So, Taylor presents two ways to use these rules. One way is to follow a specific rule proposed in the paper. It suggests that we should increase interest rates when either inflation or economic output go beyond a certain limit. Interestingly, it was noted that this rule was quite similar to what the Federal Reserve actually did from 1987 to 1992.

The other way is to not use a specific formula, but to stick to the main idea
behind policy rules. In other words, when it looks like inflation is going up, increase rates. When a recession might be on the horizon, cut them down. How much to change rates is a judgment call.

Taylor then gives two examples to show these methods:

- The 1990 oil price shock: Normally, the rule would say to increase rates. But, looking closer, it seemed the price rise was temporary, so they didn’t change the rates.

- The 1990 bond yields: Rates didn’t change here either, because the increase was due to Germany uniting, not inflation.

A key point about the rule proposed in the paper is how it shows interest rates changing in response to inflation and economic output. The rule says that for every 1% increase in inflation or output beyond a certain level, the target interest rate should go up by 0.5%. Taylor explains that giving equal weight to changes in inflation and output is a common theme in these policy rules. We might not know exactly how much the rates should respond, but this rule shows that it’s better for rates to react systematically to economic conditions rather than being set randomly. The two examples show how we can apply the rule flexibly in real situations.

The government is an essential actor in the pension ecosystem and affects pensions through a wide range of mechanisms and toolkits, among these are the mandatory contribution rules for employees and employers, the default opt-in participation regulations which has a behavioural impact on participation, the tax incentives to accumulate wealth towards the pensions, the poverty alleviation measures such as pension credits, and special measures such as furlough schemes that shield the people towards the shocks such as global pandemics. These different aspects have an impact on the pension ecosystem, and an agent based model is suitable for incorporating varying mechanisms and observing their impact and system wide interactions. The thesis implements agent base models of the pension system in differing complexity levels in the following sections.
3.2.2 Pension Funds, From DB to DC, and Saving Problem

There are reports especially focusing on the effects of transitioning from DB to DC, [31] covers the area as well, Sweeting states that: “the cost of employing a member of the defined benefit pension scheme has outpaced the cost of employing someone in a defined contribution arrangement by 1.1% of earnings per annum from 1995 to 2015. As at March 2015, the total cost of accrual was 25.4% of earnings. Looking only at changes in interest rates, the estimated cost of accrual has risen to 36.7% of earnings as at September 2016. If the current 2.5% LPI increases to pensions in payment were removed and replaced with conditional indexation, the cost of accrual for the employer would fall back to 27.0%. Nevertheless, this is still more than three times the maximum level of auto-enrolment contributions that will be required these peaks at 8% of earnings in 2019. It is also nearly seven times the current average level of contributions, which in 2015 stood at 4.0% of earnings. If contribution rates to defined contribution arrangements do not rise, a large proportion of the population will reach retirement with inadequate retirement savings.”

The research on global landscape of occupational pensions [32] identifies that there has been a major transformation over the past few decades. Defined benefit (DB) pension plans, once the predominant choice, are gradually giving way to defined contribution (DC) pensions. The trend is anticipated to hasten due to the latest regulatory and accounting reforms in the pension sectors across various countries.

One of the most significant implications of this shift is that investment risks are moving away from the corporate sector and into the hands of households. As households gain more exposure to financial markets, retirement income becomes subject to increased volatility. Interestingly, this trend is not only observed in countries with mature pension systems but also in emerging markets adopting new pension reforms.

Multiple reasons have been identified for the transition from DB to DC pension plans. Long-term factors such as increased workforce mobility due to demographic and industrial changes have been significant in driving this shift. The trend is particularly pronounced in the U.S., where mobile workers prefer DC pensions due to
the lack of portability of DB benefits. In recent years, the acceleration towards DC plans has been attributed to factors such as persistent pension underfunding, decline in long-term interest rates, market-based accounting, increasing regulatory burden, uncertainty, and the recognition of longevity’s impact on plan costs.

However, the shift towards DC pensions is not without its benefits. DC plans favour labour market mobility by reducing accrual risk, i.e., the risk of losing a significant portion of expected benefits if these are not transferable from one employer to another. It’s important to note, though, that such a shift also reallocates investment risk from corporations to households, which may have implications for financial stability.

Analysing aggregate pension sector data from Australia, Canada, and the U.S., similar asset allocations are noted for both DB and DC plans. A key difference is that DC plans tend to hold a larger share of assets in mutual funds, whereas DB plans lean towards directly held securities. While this reflects the investment strategies of the two plan types, it’s worth noting that the risk management practices of households do not always align with optimal strategies. Research shows inertia and myopia concerning retirement decisions, which can potentially undermine the ability of DC plans to provide retirement security.

The clear trend of transition from defined benefit to defined contribution schemes has been identified by OECD reports as well [6]. DB schemes, prevalent in the past, involve employers serving as ultimate guarantors and bailing out funds in case of deficits. In contrast, DC schemes transfer the risks and responsibilities of managing funds, such as longevity risk and market risks, to the contributors themselves [33]. Notably, DC schemes in the UK have lower average contribution rates (5.1%) compared to DB schemes (28.5%) [33]. Given this transition, it becomes crucial to model the evolving pension ecosystem accurately. Traditional econometric approaches, which have dominated the literature in the field, do not fully capture the complexities and dynamics of the system.

Research on known pension fund strategies [34] identified the characteristics differentiating defined benefit funds and defined contribution funds as follows,
meanwhile the investment strategies for defined contribution plans are similar to an individual’s investment strategies, such that the guiding principle is efficient diversification for maximising returns given the desired risk exposure, where the only major difference is the tax advantage. The defined benefit plans on the other hand focus on immunisation strategies that the managers try to hedge current liabilities to pensioners and insure the accruing liabilities to current employees who will be future retirees.

A different dimension that is not focal to my research but still has a characteristically similar case is the public/state pensions. The research of [35] devises a normative and positive model for public pension fund asset allocation. It suggested that the socially optimal policy is to hedge risk from market value changes in the pension fund’s liabilities. Nevertheless, the paper identified conflicts arising from the career concerns of pension fund managers, observing a tendency of managers to allocate assets based on the performance of peer pension funds, not to immunise the plan’s liabilities. The empirical study of 125 state pension plans from 2000 to 2009 showed that funds took on more asset-liability “tracking error” risk following declines in relative performance. Tracking error volatility was also found to be higher for funds that chose a high discount rate for their liabilities and those with a higher proportion of participant members on their Boards of Trustees. The authors also argued that misleading accounting standards, divorced from financial theory, encourage portfolio choices to deviate from liability immunising strategies, causing asset-liability mismatches that could burden taxpayers in declining economic conditions.

3.2.2.1 Pension Portfolio

In this section, the portfolio optimisation literature is explored. Although portfolio optimisation is a topic that has been explored starting from modern times, there is rich literature on portfolio optimisation getting gradually more complex, it started with single time step portfolio selection problem and expanded to multi time-step portfolio selection with various income dynamics modelling. The classical econometric models have reached their plateau due to the increasing heterogeneity of
income trajectories, the complexity of the asset dynamics, which can’t be captured by classical methods that assume simple mathematical equations for modelling asset dynamics, which in fact should be modelled as manifestation of interaction between actors. Furthermore, the restrictiveness of rationality assumptions in classical econometrics ignoring the behavioural differences is a limiting factor. ABMs augmented with DNNs as approximators that can imitate the complex interactions between actors are more suitable to capture the complexity and heterogeneity of the pension ecosystem and its constituents.

**Portfolio Optimisation:** Harry Markowitz's Portfolio Selection Theory introduced in 1952 [11], also known as Modern Portfolio Theory (MPT), is an essential methodology for investors to construct their portfolios, which can be regarded as a significant cornerstone for the literature of portfolio optimisation. The central idea behind MPT is that investors should not only focus on maximising returns, but also consider the level of risk associated with each investment. According to Markowitz, investors should aim to achieve the highest possible expected return for a given level of risk. He introduced the concept of an "efficient frontier", which represents the set of optimal portfolios that offer the highest expected return for a particular level of risk. By diversifying their investments across multiple asset classes, investors can reduce the overall risk of their portfolio while still maintaining a reasonable level of expected return. Mathematically, Markowitz’s model uses standard deviation as a measure of risk and formulates the problem as a quadratic optimisation one and uses historical data to estimate returns and risk. A portfolio with less than perfectly positively correlated assets can have less risk than the weighted average risk of the individual assets. This is the benefit of diversification, and it’s how Markowitz’s portfolio theory minimises portfolio variance.

Modigliani’s research [12] explores the implications of the life-cycle model, which proposes that individuals strive to achieve a stable level of consumption throughout their lifetimes. The model assumes that household consumption and saving decisions are based on a deliberate attempt to balance current expenditures with future needs, taking into account factors such as income fluctuations and ex-
3.2. PENSION ECOSYSTEM

pected longevity. Specifically, the model assumes that individuals should save during their working years, when they have higher earnings potential, and then dissave (spend down their savings) during their retirement years. To test these hypotheses, Modigliani develops a stationary economy model, assuming constant population size and per capita income over time. He then extends his analysis to include non-stationary models that allow for changes in population growth rates and income patterns.

Merton’s work builds upon Modigliani’s foundational research on the life-cycle hypothesis by incorporating uncertainty in investment returns and allowing for continuous decision making. By extending Modigliani’s deterministic framework to account for these factors, Merton developed a comprehensive microeconomic model that applies the principles of consumption smoothing to financial investment decisions. Merton’s model represents a bridge between macroeconomic theories of consumption and individual financial decision-making under conditions of market uncertainty.

A Lifetime Portfolio Allocation and Consumption Decision Problem by Merton: Merton’s Optimal Portfolio Theorem was developed by Robert C. Merton [13]. It extends the work of Harry Markowitz [36], who introduced the concept of portfolio optimisation using mean-variance analysis. Merton’s approach focuses on continuous-time portfolio optimisation for an investor with a utility function.

The main result of Merton’s Optimal Portfolio Theorem states that an investor should hold a combination of a risk-free asset and a risky asset (or a portfolio of risky assets) in proportions that maximise the expected utility of their terminal wealth. Merton’s framework also provides formulas to compute the optimal proportions of these assets in the investor’s portfolio.

The investor’s problem is to choose a dynamic consumption and investment plan that maximises their expected utility, subject to the wealth dynamics, where asset prices follow a standard Brownian motion.

The utility is calculated by the Constant Relative Risk Aversion (CRRA) [14] function.
Merton derived the optimal proportion of wealth invested in the risky asset $\pi^*$ is constant over time and depends on the investor’s risk aversion, the risk-free rate, the expected return on the risky asset, and the risk associated with the risky asset.

Merton’s model suggests that an investor should consume a constant proportion of their remaining wealth at each point in time to maximise their lifetime utility.

The investor’s optimal consumption plan depends on their initial wealth, time preference rate, and planning horizon. In general, the investor will consume more if they have a higher initial wealth, a higher time preference rate, or a shorter planning horizon. Conversely, they will consume less if they have a lower initial wealth, a lower time preference rate, or a longer planning horizon.

Following Merton’s work, the subsequent models became more complicated, in [15] Gomes et al. utilise a sophisticated life-cycle model that takes into account consumption, portfolio decisions, non-tradable labour income, and borrowing constraints. The study finds that labour income acts as a substitute for risk-free asset holdings, and that investing in equities leads to a reduction in wealth over the lifetime of an individual. Neglecting labour income leads to substantial utility losses while disregarding its risk leads to lower costs unless a catastrophic labour income shock is taken into account.

The presented model contributes to the existing literature on buffer-stock savings by incorporating an asset allocation decision. The utility costs associated with different portfolio rules for individuals with varying characteristics are quantified. Their research explores the implications of endogenous borrowing constraints on the life cycle. It shows that labour income increases the demand for stocks, particularly in the early stages of life, but the risk of a disastrous labour income shock can significantly decrease the average stock allocation.

As a financial takeaway from their results, their study supports the common financial advice to shift portfolio composition towards safer assets with ageing. It also reveals that ignoring only labour income risk leads to lower utility costs than completely ignoring the income source, except in the case of a catastrophic shock.

Campanale et al. [16] introduced the concept of liquidity differences across
assets into the standard life-cycle model of portfolio choice by considering the different risk properties of financial assets. They incorporated transaction costs associated with stocks, leading to an infrequent trading pattern consistent with observed household behaviour. This made it possible for the model to generate stock allocation patterns over age and wealth that are constant or moderately increasing, which is in line with empirical data.

The model demonstrates the advantage of matching the subtly increasing portfolio share pattern across the life-cycle and wealth - conditional on age - without correlating labour earnings and market returns. The presented model is a partial equilibrium established within a life-cycle framework, assuming households have an Epstein-Zin utility function based on a single non-durable consumption good. The model assumes that households cannot insure against earnings shocks due to missing markets, so they resort to using savings to balance consumption amidst earnings variability. Two assets are available - a risk-free, liquid financial asset, and a risk-laden one. Transactions between these two distinct accounts incur a fixed cost, regardless of the traded risky asset quantity.

The model must be solved numerically, and the solution is described by authors as slow and difficult due to three continuous state variables, two continuous control variables, and a fixed transaction cost breaking the concavity of the objective function. The Campanale model assumes that a person has the freedom to switch between liquid and non-liquid asset types, which is not the case with locked pension savings. In the Campanale et al. model, the most important calibration challenge is the transaction cost, which also includes psychological and non-monetary costs.

The innovative approach that diverges from the previous literature redefines the risk-free asset as a liquid financial asset in the model. Consumption can only be bought with the liquid asset, and a cash-in-advance constraint can be bypassed by incurring a fixed transaction cost between the stock and liquid accounts. This approach allows the model to enhance traditional models that assume entry or participation costs, particularly regarding stock allocation over age and wealth. Despite slight over-predictions in stock share, the model is largely consistent with empirical
3.2. PENSION ECOSYSTEM

population averages. The authors recognise the potential to improve the model by incorporating other risk sources such as marital and health risks, which could decrease the allocation to stocks. They suggest this as a fruitful direction for future research in the field.

Agent Based Modelling (ABM) powered with Deep Learning (DL) and trained with Reinforcement Learning (RL) provides a significant advancement over classical methods such as the models introduced by Merton, Campanale or Gomes. It has the advantage of modelling complex dynamics of real-world financial systems more accurately, capturing non-linear relationships and changes over time that can’t be effectively represented using traditional mathematical formulations. This is essential in situations like the heterogeneity of career paths and income trajectories or dynamic asset allocation strategies.

The RL algorithms enable the model to learn from data and adapt to changing conditions in real time, enabling more robust and responsive strategies. These algorithms can be tuned to identify and mitigate potential risks specific to certain groups of pensioners, thereby creating a more personalised approach. This is an important advance over traditional models, which often assume homogeneity across profiles and fail to address idiosyncratic challenges effectively.

Unlike classical models that make decisions based on only the current income, our model can also take into account historical income trajectories using recurrent neural networks. This provides a more nuanced understanding of each individual’s financial behaviour and can significantly improve decision making in retirement finances.

In addition, the models introduced in this thesis are scalable and can be calibrated to different scenarios, making it suitable for a wide range of pension fund management goals and constraints. It contributes to the development of personalised portfolios, factoring in variables such as age, profession, risk tolerance, and financial goals.

Asset/Liability: Dahlquist et al.’s study [37] investigates optimal default fund composition in defined contribution (DC) pension plans, drawing on comprehensive
Swedish investor data. The authors propose an optimal default asset allocation that could enhance retirement welfare by 1.5% when compared to a standard age-based allocation. The research shows passive investors typically have less equity exposure outside the pension system, with those participating in the stock market having financial wealth equating to 1.4 years of labour income. In contrast, non-participating investors’ wealth equals roughly five months of labour income.

The authors present a model extending the life-cycle portfolio choice models, factoring in variations in income and wealth, and a DC pension account. This model suggests welfare gain from optimal asset allocation can be achieved by following a simple rule of thumb based on age, pension account balance, and stock market participation status. The research found that 40% of active pension investors opting out under the “100-minus-age” rule would favour the default fund if it were invested based on this heuristic. The study underscores the power of asset allocation rules for cost-effective mass customisation, highlighting the importance of readily available information such as pension account balance.

The importance of accounting for rare and potentially catastrophic events, such as the 2008 financial crisis, in pension system modelling is highlighted by recent studies [38]. Black swan events can have significant effects on the financial markets, challenging the assumptions of stationarity in traditional models. The use of agent-based modelling (ABM) and reinforcement learning (RL) techniques in pension modelling can improve the representation of the non-stationary dynamics associated with these events, allowing for more accurate assessments of pension strategies. The 2022 pensions leveraged gilt crisis [39] also demonstrates the importance of considering the interdependence between pension funds and the wider financial system. ABM and RL methods allow for the explicit modelling of these interactions, providing a more comprehensive view of pension ecosystem dynamics. The relevance of modelling pension systems counter-cyclically with the business cycle is emphasised by studies of the Norwegian Sovereign Wealth Fund [40]. Non-stationary market dynamics can have a significant impact on pension fund saving and investment strategies. ABM and RL techniques provide a more realistic repre-
presentation of these dynamics, enabling the development of effective pension models that consider both short-term fluctuations and long-term trends. Traditional econometric methods are often used in pension research, but their reliance on unrealistic assumptions can limit their accuracy. For example, these models may assume that historical patterns can accurately predict future market dynamics, overlooking the impact of unforeseen events.

**Microsimulation of UK Pensions:** The models that are being so far introduced by classical econometrics are utilised to incorporate to micro simulations, and these can be used for policy testing, but these models are usually parametric models, that are comprehensive in coverage of various factors, but limited in functionality to capture the complex relationships and higher order interactions. There are various pension simulators that are being used in the UK and around the world; both governmental bodies and independent researchers investigated these models thoroughly.

Report [41] from the Department of Works and Pensions base their approach on the adequacy of retirement income to be above the poverty threshold and alternatively look at whether savings allow people to maintain the same broad living standards in retirement as during the time in the workforce. The report focuses on Pensim2, which is a dynamic micro-simulation model based on a sample of synthetic individuals, reflecting the characteristics of the population and producing descriptive results about group-specific income inadequacy in aggregate levels. The report also mentions I-Pen, a simplistic, Excel-based model to predict state and private pension income in retirement based on work histories for hypothetical individuals. A subsequent report [42] also draws attention to the incapability of Pensim2 to reflect behavioural impacts of policy changes such as a simplified model, increasing levels of savings and lacks to capture the interactions between different forms of savings that are modelled.

The Department for Work and Pensions (DWP)’s Pensim2 model is a dynamic microsimulation model used to estimate the future distribution of pensioner incomes in the UK. The model starts with large, representative samples of individuals and households and “grows” them through time by simulating relevant life events like
Individual and household incomes are modelled dynamically based on simulated employment, earnings, taxes, benefits, pensions, and savings. This allows analysis of how incomes change over the life cycle and how policy impacts income distributions over time. Panel data is used to capture income mobility.

The savings module models the probability of having financial assets at age 60 and the level of assets using data from the Family Resources Survey. Assets then evolve from age 60 based on British Household Panel Survey data. Limitations of the current modelling are acknowledged, and recommendations are made for incorporating income/substitution effects and using proposed new survey data on wealth.

This thesis provides more adaptive models that can capture complex, heterogeneous income dynamics, and dynamic environments by utilising ABM and Deep RL techniques.

### 3.2.3 Markets

The Markets have a continuous impact on the income dynamics of employees and the asset dynamics, the market dynamics can also be perceived as a manifestation of the interaction between Businesses, People, Government and Central Bank. The classical econometrics attempts to model these relationships with simple mathematical equations expressing the relationship between two or three factors in the market dynamics. But because these are the observed effects of underlying interactions trying to model these relationships in macroeconomic equations although provides some intuition, it is unsuccessful in capturing the complexity of the ecosystem. ABMs powered with deep neural networks as powerful approximators for capturing the interaction dynamics between actors can be much more powerful to capture the causal relationships between these interactions. In order to communicate with the economics community it is important to understand and use simple principles such as business cycles and preferences.

The research by Christiano et. al. [44] is capturing the impact of investor wealth and interest rates on GDP growth. Christiano et al.’s article provides an in-
depth review of recent research on the impact of exogenous monetary policy shocks. The authors emphasise the importance of understanding the effects of these shocks on economic aggregates in order to evaluate the empirical validity of structural economic models and to assess the impact of monetary policy reforms. Despite the lack of a consensus on the specific assumptions needed to identify the impacts of a monetary policy shock, there is a general agreement on the qualitative effects of such a shock. The paper goes on to explain the relationship between the identification assumptions and the inferences about the effects of a monetary policy shock, making it easier for readers and model builders to assess the different claims about the aftermath of a shock. The article is divided into several sections that discuss topics such as potential interpretations of monetary policy shocks, the use of vector autoregressions (VARs) as the primary statistical tool, the issue of recursiveness assumption, the interpretation of estimated monetary policy rules, and the narrative approach to studying the impacts of a shock. The authors propose a method for implementing the third step of the Lucas critique applied to monetary policy shocks, and they emphasise the need for theoretical models to use estimated reaction functions from the policy shock literature. Overall, the article provides a comprehensive review of the recent research on monetary policy shocks and offers valuable insights for both academics and policymakers.

3.2.3.1 Business Cycles

Being aware and accounting for business cycles are vital according to economic theory. Business cycles govern the income and asset dynamics, and accounting for cyclicalities is vital for investment and saving behaviour. The returns fluctuate with the economic climate, potentially leading to higher returns during growth periods and losses during downturns. These cycles also influence a fund’s funding status, with contributions and returns typically increasing during economic expansions, thus improving the ratio of assets to liabilities. And pension funds have the advantage of having a long horizon, which is demonstrated to be beneficial if leveraged by accounting for the business cycles in investment strategies. The study of the business cycle has been around since the late 1800s when classical economists
first began developing models [45]. These early thinkers believed that the economy was self-correcting and that unexpected events or shocks were responsible for fluctuations in factors such as production and employment. Additionally, they assumed that an economy would always adjust to achieve full employment due to its ability to adjust prices and wages quickly. But later scholars such as John Maynard Keynes challenged these assumptions with his theory of aggregate demand [46]. According to this perspective, low levels of demand can result in persistently high levels of unemployment, which underscores the critical role of government intervention through measures such as fiscal stimulus and regulation of interest rates. Finally, in more recent times, monetarist theories have emerged, arguing that changes in the money supply are the primary driver of economic cycles. This view highlights the importance of central bank policies and cautions against mistakes that could destabilise the economy [47].

Kalecki introduced a model [48] where business cycles are endogenously created in the system, where the model is characterised by economic activity driven by the capital investments, and the periodicity is explained by the investments resulting in profits that are used for new investments which compete with older investments, and causing cycles of expansions and depressions. The Real Business Cycle (RBC) theory suggests that the business cycle is driven primarily by real factors such as technological advancements, productivity improvements, and changes in supply rather than demand-side factors like consumption or monetary policies. According to this framework, individuals and firms make rational choices based on available information, leading to economic fluctuations. One aspect of this theory, known as “Time to Build” proposes that investments take time to yield returns, which can introduce delays and fluctuations in output and employment levels.

Later research [49] investigates empirical evidence analyses the real business cycle theory which has been widely used to explain economic fluctuations in the United States since World War II. The researcher decomposes total factor productivity into technological change and other factors such as labour inputs. He finds surprising evidence that technology shocks have a negative impact on productivity.
while non-technological shocks have a positive effect. These results challenge the standard view that technological progress is always associated with higher levels of efficiency and growth. Furthermore, he argues that these results are more compatible with alternative theories that take into account market failures such as imperfect competition or incomplete information. Overall, this paper provides important new insights into how we understand economic fluctuations and suggests new directions for future research.

The recent paper by Daron Acemoglu [50] has made significant contributions towards challenging the traditional view on how small changes in different parts of the economy affect overall growth and stability. His study shows that even seemingly minor shifts in the production process can have large consequences for the entire system due to intersectoral input-output linkages and the asymmetric roles sectors play as suppliers to others. Microeconomic shocks can propagate through the economy and influence aggregate volatility, this suggests that the sectoral structure of an economy can significantly impact aggregate fluctuations.

Research on U.S. Social Security data [51] indicates that labour income has characterising moments that are counter-cyclically exposed to business cycle effects. Researchers investigate [40] Norwegian Sovereign Wealth fund that invests counter-cyclically with the business cycle, demonstrating the importance of accounting for non-stationary market dynamics in pension models. The research emphasises the importance of portfolio weight rebalancing during counter-cyclical investment strategies. Following research [52] introduces cyclical skewness to the life cycle model of portfolio choices, countering the view that human capital always increases equity demand. This approach explains observed patterns such as low stock market participation among young households, increasing equity shares until retirement, and less investment by renters. The concept of cyclical skewness, highlighting simultaneous losses during recessions, also refines preference parameter estimates, although the author indicated that countercyclical labour income risk has a limited influence on aggregate equity demand.

Further research in the field [53] article provides a detailed analysis of aggre-
gate price dynamics within an overlapping generations (OLG) model. It primarily focuses on the impact of retirement age on macroeconomic fluctuations, indicating that when individuals can choose their retirement age, the volatility of aggregate prices decreases. The research hinges on two assumptions: an economy producing a single non-storable good and a demographic structure consisting of OLG with finitely-lived individuals. By contrasting mandatory and chosen retirement ages, d’Albis creates two distinct economic environments and examines their responses to price fluctuations.

Agent Based Models that can capture the business cycles as emergent and endogenous properties of the model have the potential for a better understanding of how individual agents’ behaviours and interactions lead to business cycles and a potential to understand the underlying mechanisms that drive the business cycles. Such a model can be used to train agents who can learn to plan according to the cyclic nature of markets. In this thesis, the deep RL allows the agents to learn policy functions of micro interactions that lead to the observation of macro phenomena that are observed in the real world, this is one of the differentiating aspects of the models introduced in this thesis.

**IO:** In this PhD thesis it is attempted to model the interactions at the most fundamental and “real” level, such that the nature of interactions are calibrated to either behavioural principles or physical rules, such as input output dynamics, which characterise what kind of different sectoral inputs are required to output one unit of specific sectoral output. The research on Input-Output relationships of businesses can assist if incorporated into the ABM. An in-depth analysis of the economic impacts of simultaneous supply and demand shocks due to the COVID-19 pandemic should focus on utilizing a dynamic approach incorporating input-output (IO) modelling techniques [54]. This methodology allows for consideration of multiple interconnected industries simultaneously, providing insight into how initial disruptions propagate throughout various sectors. In particular, the study’s findings suggest that network effects can magnify the severity of initial shocks, necessitating careful consideration of policies such as rationing to prevent shortages from causing long-term
3.2. PENSION ECOSYSTEM

damage to the economy.

Task-specific human capital is an important form of human capital that differs from general purpose human capital in that it is specific to individual tasks performed at work instead of being broadly applicable across many jobs within a company. This insight builds upon previous work by Gary Becker who studied general purpose and firm-specific human capital. The study [55] shows that task-specific human capital can have just as much impact on productivity as general purpose human capital and may be more useful due to its wide range of applications.

3.2.4 Preferences

The expected utility theory [56] assumes that investors do not merely maximise their expected return, rather they maximise their expected utility.

Under this assumption, risk-averse individuals would choose a certain outcome over a risky one with the same expected return. This reflects their preference for stability and predictability.

Pratt’s research [14] explores the complex relationship between utility functions for money and risk aversion. He introduces a measure of local risk aversion, 
\[ r(x) = \frac{\mu''(x)}{\mu'(x)} \]
arguing that neither the second derivative of the utility function nor the graph’s curvature offer suitable measurements. The author defines risk in terms of both the risk premium or insurance premium for an arbitrary risk and the proportion of total assets.

Pratt also introduces the concept of decreasing risk aversion. In his analysis, a decision maker exhibits decreasing risk aversion if his cash equivalent increases with his assets, and his risk premium and the amount he would pay for insurance decrease accordingly.

Pratt’s work focused more on individual decision-making under uncertainty. He developed the measure of absolute risk aversion [14], which quantifies an individual’s preference for a certain outcome over a risky one. Pratt introduced absolute and relative versions of risk aversion according to the type of difference in wealth.

Epstein and Zin [17] introduce a novel class of recursive utility functions that uniquely disentangles risk attitudes from the degree of intertemporal substi-
tutability, permitting an in-depth understanding of consumption and asset returns behaviour over time. Their utility functions permit preferences over intertemporal consumption lotteries, enabling analysts to distinguish between risk preferences and time preferences, an innovative concept not addressed by conventional expected utility models. This is a central point in the context of asset returns, where systematic risk is determined by the covariance with both the return to the market portfolio and consumption growth.

A more recent study [18] focusing on preferences in the context of pensions by Department of Work and Pension in the UK concluded that the interaction between incentives and behaviour is complex and is not fully understood. Therefore, the impact on savings of such measures as changes in tax relief is difficult to assess with robustness, and different studies over time and in different countries, and indeed some in the same country, have tended to be contradictory. This indicates that fine tuning policy set for optimal pensions is a fine-grained work.

The paper discusses how traditional economic incentives like tax relief on savings assume people behave rationally, while behavioural incentives acknowledge that people don’t always behave rationally. Behavioural incentives are based on observed attitudes, mindsets, and behaviours around saving.

Key behavioural factors that influence savings behaviour discussed in the paper include:

- Inertia and defaults - Automatic enrolment into savings schemes relies on inertia to increase participation. Making the default option simple and allowing procrastination on opting out can be effective.

- Framing and presentation - How incentives are framed influences take up e.g. a match is more appealing than an equivalent tax credit. The timing and perceived impact of saving decisions also matter.

- Mental accounting - How people treat different sources of income and compartmentalise decisions affects saving.

- Loss aversion - People are motivated to avoid losing benefits. This can deter
3.2. PENSION ECOSYSTEM

• Heuristics and rules of thumb - Simpler pension systems reduce complexity and the need for difficult forecasting.

• Social norms - Having any pension at all is seen as sufficient by many due to norms passed down generations.

The paper concludes that integrating behavioural factors into traditional economic models is important for effective policy. Relying solely on “minor economic incentives” is unlikely to be effective. A range of interventions including defaults, framing, simplification and understanding motivations like loss aversion is required.

Reinforcement learning methodologies applied in this thesis are suitable for incorporating in the model or providing analogies with respect to the utility functions introduced in the econometric theories. This interoperability enhances the communication of the model with wider audiences and furthermore, it provides an additional flexibility to entangle the time preference and the risk aversion.

3.2.4.1 Income

Income Dynamics are a critical part of the problem of pension savings, and assumptions about income dynamics have an impact on the saving and consumption strategies. Classical econometric models have simplifying assumptions, and these models got more complex and heterogeneous over time, but still lack the level of complexity and heterogeneity that ABMs with DNNs can offer. Initial research modelling income dynamics utilised a first-order Markov process [57] to model the time-evolution of wages, but it lacks the mechanism to link individual wage dynamics to the time-evolution of the distribution of wages.

The following research by Shorrocks utilised a second-order Markov process [58], but it fails to capture the relationship between individual wage dynamics and time-evolution of the distribution of wages. Furthermore, Shorrocks investigates an econometric methodology [59] to deal with life cycle earnings and mobility among discrete earnings classes. The paper utilises a co-variance structure model for modelling individual characteristics with linear regression. Probit model focuses on
poverty and transition properties but lacks a relationship between individual wage
dynamics and time-evolution of the distribution of wages.

The researchers [60] use National Longitudinal Survey of Youth to investigate
the mobility through the wage and earning distributions. The estimates of tran-
sition probability in wage quantiles are determined depending on various factors
such as education, experience and age; later research [61] presented a new regres-
sion method to evaluate the effects of changes in the distribution of the explanatory
variables of the quantised marginals of an outcome variable, namely defines an un-
conditional quantile regression.

The Persistence of income is also investigated by the reports of governmental
bodies [62]. One report from the Department of Works and Pension in the UK
provides data on Income Dynamics, examines the changes in household incomes
in the UK and empirically draws attention to the persistence of low income across
location, gender, ethnicity and social status.

Guvenen explains [63] that “Restricted Income Profiles” (RIP) process holds
that individuals are subject to large and very persistent shocks, while facing similar
life-cycle income profiles. Alternatively, he investigates the concept of the “Hetero-
genous Income Profiles” (HIP) process, which holds that individuals are subject
to income shocks with modest persistence, while facing individual-specific income
profiles. The Heterogeneous Income Profiles (HIP) model depicts a world where
individuals are experiencing shocks with modest persistence but face life-cycle pro-
files that are individual-specific and introduce variability. The alternative model that
he names the Restricted Income Profiles (RIP) model depicts a world where individ-
uals are experiencing extremely persistent and nearly random-walk shocks, while
facing similar income profiles that may be conditional on some other observable
factors.

A later study [64] uses extensive panel data of earnings histories to scrutinise
the evolution of individual labour earnings over the life cycle. They find out a
high kurtosis referring to the fact that individuals generally are subjected to small
earnings shocks and only a minority is subjected to significantly large shocks. The
same research reflects that the statistical properties vary significantly over the life cycle and income levels. Furthermore, the researchers estimate impulse response functions of earnings shocks, and reveal differing characteristics governing income groups such as high-income individuals are subject to positive transitory shocks in contrast to persistent negative-shocks, and low-income individuals experience positive shocks that are persistent but experience transitory negative shocks.

Recent work [65] examines the impact of the COVID-19 pandemic on the retirement prospects of underpensioned groups in the UK, defined as those at higher risk of inadequate retirement outcomes including women, ethnic minorities, disabled people, carers, multiple job holders, and the self-employed. It finds that underpensioned groups have been disproportionately affected by negative economic impacts during the pandemic due to existing labour market inequalities.

Specifically, underpensioned groups were more likely to experience unemployment, furloughing, and income loss during the pandemic. This may widen pension gaps if underpensioned groups face continued difficulties rebuilding careers and finances post-pandemic. Policy recommendations include expanding automatic enrolment, increasing contribution rates, and providing greater employment flexibility to support underpensioned groups. The long-term impacts on retirement inequality will depend on the pace of economic recovery and policy responses to address systemic labour market disadvantages.

The article provides useful evidence on how economic crises exacerbate existing inequalities for marginalised groups. Understanding these dynamics can inform policy approaches to building more equitable and resilient pension systems. The rise of the so-called gig economy has sparked debates about implications for workers, firms, and economic measurement [66]. New technology and business models are perceived to be fuelling growth in non-traditional work arrangements like gig work. Different data sources provide conflicting evidence on trends. Household surveys since the mid-1990s show flat or declining self-employment, which is the category where gig workers should appear. This contrasts with administrative tax data showing growth in self-employment. Understanding this divergence between
survey and tax data is an important puzzle.

Both household surveys and tax data have limitations in fully capturing informal and non-traditional work arrangements. Analysis of matched individual-level survey and tax records helps reconcile the discrepancy. It shows the tax data is capturing growing self-employment that is not reported in household surveys. Much of the growth in the tax-based measure comes from individuals with self-employment income who do not mention it when responding to household surveys. The number of people reporting self-employment income in surveys but not showing it in tax data is smaller and has been flat.

There are opportunities to improve economic measurement related to gig and non-traditional work. Household surveys could add periodic modules with more probing questions. Tax data, employer surveys, and private sector data sources could be better utilised. Integrated data sets that combine multiple sources would provide more insight than any single source alone.

This thesis focuses on individual-specific income profiles rather than generic averages, and the model considers unique aspects such as profession, age, and personal behavioural parameters, creating a more tailored understanding of income trajectories. These abilities to capture and learn optimal saving and investment strategies for heterogeneous profile properties such as distinct occupation groups and age brackets set it apart from the classical income dynamics literature.

### 3.2.5 Consumption

Consumption planning is a core part of retirement and lifelong finance planning. The consumption trajectories between individuals can differ greatly due to their profiles, which can be affected by the family stance, profession, mode of working, and age. The classical econometrics fail to capture the complexity and heterogeneity of the consumption dynamics.

**Lifetime Consumption:** A life-cycle model of consumption was developed by Modigliani and Brumberg [67], which states that individuals aim to maintain a consistent level of consumption throughout their lifetime. The model is based on utility maximisation by individuals subject to their lifetime budget constraints. The cur-
rent consumption of an individual is assumed to be proportional to the present value of their total resources over their remaining lifetime, which consists of current net worth plus the present value of expected future labour income. After aggregating across individuals, the model yields an aggregate consumption function where current consumption is a function of current labour income, expected future labour income, and current net worth.

Cocco et al. develop a life cycle model [68] with uninsurable labour income risk and borrowing constraints to study optimal consumption and portfolio choice. The labour income process before retirement has a deterministic component based on age $t$ and characteristics $Z_{it}$, as well as persistent $v_{it}$ and transitory $\epsilon_{it}$ shocks where, the dynamic portfolio problem is solved via backward induction. Their results indicate that labour income acts more as a substitute for bonds than equities.

### 3.2.6 Data

An array of data sources was applied in different phases of the research project, each contributing unique insights and data points. Utilising a comprehensive approach, the research incorporates both longitudinal and cross-sectional datasets, allowing it to provide an in-depth exploration of socioeconomic and demographic variables.

These datasets are used for the calibration of the models and benchmarking the results.

The British Household Panel Survey (BHPS) [69], spanning from 1991 to 2008, offers longitudinal data essential to this research. The wealth of data from BHPS allows the research to track changes in individuals’ behaviours and household compositions over time.

A critical companion to the BHPS data is the guide [70], which explains the methodologies for handling data issues like topcoding, attrition, and post-stratification. The handling of these issues is critical in ensuring accurate interpretation of data, particularly in the analysis phase of the research.

Contrasting the longitudinal nature of the BHPS, the Current Population Survey (CPS) from the Integrated Public Use Microdata Series (IPUMS) [71] in the USA provides cross-sectional data. Its unique characteristic of excluding self-
employed income from the labour income data enriches the diversity of data in the research and is particularly considered during the comparative analysis phase.

The population pyramid from the ONS [72] is integrated into the research during the demographic analysis phase. It serves as a visual representation tool to help pinpoint trends like an ageing population.

Furthermore, the US Census Data and the summarised tables of labour and earnings related statistics [73] and Poverty Threshold Statistics [74] are employed in examining patterns of income, employment, and poverty in the United States. These data sources come into play in the later stages of the research, assisting in the correlation and comparison of demographic and economic variables, as well as setting minimum consumption levels according to formal poverty statistics.

Lastly, the US Actuarial Life Table [75] is used for the mortality calibration during the research. Providing mortality rates and life expectancy data, it enables the research to evaluate health outcomes in relation to socio-economic variables.

3.3 Agent Based Models and Deep RL

The econometric models have been under criticism for their significant downsides. Classical econometric models, while valuable in various contexts, have fallback due to their foundational assumptions and methodological constraints. Among these, the assumption of linearity, causality issues, the assumption of independence, and their static modelling approach stand out.

Robert Lucas Jr. introduces a critical argument that has since come to be known as the Lucas critique [76]. Lucas challenges the use of econometric forecasting models for quantitative policy evaluation. According to him, the features leading to successful short-term forecasting do not correlate with quantitative policy evaluation, leading him to suggest that the “theory of economic policy” based on these models needs major revision. Lucas claims that simulations using these models can provide no useful information regarding the actual consequences of alternative economic policies.

The actual behaviour of economic variables under a new policy cannot be ac-
Lucas cautions his readers that any change in policy will systematically alter the structure of econometric models. While this might have occasional significance for short-term forecasting, it becomes fundamental for policy evaluation. He claims that comparing the effects of alternative policy rules using current macroeconometric models is invalid regardless of the performance of these models over the sample period or in ex-ante short-term forecasting.

### 3.3.1 ABM

There are various applications of the agent-based modelling of financial systems; among the most popular ones are the ones supported by central banks and can be accepted as also a source of credibility for such models. Bank of England is a prominent backer of agent-based models and applied it to various subjects. But there is a gap in pension ecosystem modelling, that can capture the heterogeneous and complex dynamics.

Several researchers including Bank of England staff and academics researched An agent-based model of the UK housing market [77] to study the effects of macroprudential policy on the housing market; such an approach helped them to account for heterogeneity in the housing market by modelling the individual behaviour and interactions between the actors such as first-time buyers, homeowners, buy-to-let investors, and renters. The agent-based modelling made it possible not only to capture the individual characteristics of the subcomponents but also the collective behaviour of a complex system. The results presented the relationship between the housing booms and busts, price cycles, volatility and the macroprudential policy. The success of the paper drew more attention to the topic in the Bank of England, and a series of papers applying agent-based modelling to various subjects followed up. A few of these papers will be presented below.

One of the remarkable properties of agent-based modelling is the methodology [78] of bottom-up reasoning and building; the paper investigates the strengths and fall-backs of agent-based modelling of the economy. The results suggest that agent-based modelling provides significant success in explaining emergent behaviour and
3.3. AGENT BASED MODELS AND DEEP RL

heterogeneity, reflecting stylised facts and realistic behaviour, exploring possibilities, and describing complexity, non-linearity, and multiple equilibria. On the other hand, the weaknesses are the adverse outcomes of the great freedom and flexibility can be problematic such that assumptions can have high impacts on the results. Lucas critique has significant weight in the agent-based models such that the trade-off between optimal agent behaviour and simple, explainable model is a concern by designing of ABMs. Another weakness of the system is generalisation, then it is trained for, such as using a model for the bond market in the housing market; and the last challenge argued by the paper is the difficulty of calibrating and interpreting agent-based models. The paper presents the strengths and weaknesses of agent-based models in finance and provides examples; it is a clear guide for agent-based financial modelling.

The research [79] focuses on the role of high-frequency traders in flash episodes in electronic financial markets. The researchers construct an agent-based model of a market for a financial asset in which trading occurs through a central limit order book. The model consists of heterogeneous agents with different trading strategies and frequencies and is calibrated to high-frequency time series data. The outcome spotlighted the role of procyclical behaviour in flash episodes and investigated the effectiveness of countermeasures such as circuit breakers. The results reflect empirical evidence and provide tentative theoretical foundations.

The paper [80] presents a structural framework for the development of system-wide financial stress tests with multiple interacting contagion, amplification channels and heterogeneous financial institutions. The framework consists of financial institutions, contracts, markets, constraints, and behaviour. The paper differs from the other research in the sense that the scope of the paper is comprehensive, and the focus is not only the actions of a single group of actors but the behaviour of various actors with different motives and constraints, which constitutes a much more complex model, with motives such as contagion, amplification and buffers for absorption. The paper also provides the software that it developed as open-sourced to the research community, and it contains clues for the design perspective of a
financial agent-based model.

The paper [81] presents an agent-based simulation of the catastrophe insurance and reinsurance industry and uses it to study the problem of risk model homogeneity. The model incorporates the balance sheets of insurance firms, who collect premiums from clients in return for ensuring them against intermittent, heavy-tailed risks and the results demonstrate that using too few models increases the risk of non-payment and default while lowering profits for the industry as a whole. The paper provides the first step towards a simulation model of the insurance industry for testing policies and strategies for better capital management. In that sense, the fields of insurance, as insuring against risks contain some analogies to pensions and the risks associated; but these analogies are more suitable for the pension industry rather than the contributors and pensioners.

The paper [82] presents an agent-based model of households and their savings behaviour constituting macroeconomy, the households in the model are myopically rational and belong to a social network; this approach contrasts with the classical economic theory of households as rational utility maximisers. In the model, the households are affected by their neighbour’s saving behaviour and divided into two classes as poor and rich households with respectively low and high savings rates. The paper explores a system that substitutes utility maximisation for behaviourally grounded decision making, by which behaviour propagation time and characteristics are decisive for the performance of the model. The Oxford INET paper investigates a related phenomenon to this research, but the focus of the paper is rather social interactions and behavioural effects. The subject of household saving is also vital for the proposed ABM of the Pension Ecosystem. There are ideas such as behaviour propagation between households that might be incorporated in the ABM for Pension Ecosystem, but the vital part of income trajectories is missing at the Oxford INET paper.

The [83] paper models the economics and epidemiology of diverse scenarios for the phased restart of the UK economy during COVID-19. It uses input-output data and HIS Markit survey to understand necessary industry inputs. The model
reveals how social distancing influences supply, demand, and economic output, and how different industries impact disease transmission. Six re-opening scenarios were investigated, suggesting a viable compromise between economic growth and a minimal increase in disease transmission rates. The study also shows the usefulness of agent-based modelling in finance, hinting at potential integration in pension ecosystem modelling.

The [84] paper leverages an agent-based model to predict the supply and demand shocks in the US economy due to COVID-19, focusing on specific industries and occupations. It uses existing research to estimate the Remote Labour Index and model COVID-19 impacts. Despite the exclusion of indirect industry effects, a significant limitation, the findings provide valuable insight into the workforce module of the pension ecosystem model.

The paper [85] investigates the methodological problems of empirical validation in agent-based models in economics and how these are currently being tackled. The recurring motives identified in ABM are the bottom-up perspective of the models, heterogeneity of the agents, bounded rationality of the agents encapsulated in models, and direct networked interactions between agents. The problems identified in the paper are lack of cohesion and structure among ABM models focusing on the economy, lack of comparability between models, lack of standard techniques for constructing and analysing ABMs, problematic relationship with ABMs and empirical data and stylised fact centric evaluation methods. Some of the issues stated in the paper are as follows: “Is a realist methodology appropriate? Why should empirical validation be the primary basis for accepting or rejecting a model? Do other tests of model validation exist other than the reproduction of stylised facts? If we do proceed down the path of empirical validation, then how should we relate and calibrate the construction of parameters, initial conditions, and stochastic variability in AB models to the existing empirical data? Which classes of empirically observed objects do we actually want to replicate? What are the consequences, for the explanatory power of a model, if the stylised facts are actually unconditional objects that only indicate properties of stationary distributions and, hence, do not provide
information on the dynamics of the stochastic processes that generated them?"

The paper [86] seeks to establish a framework for directing a society of simple, specialised, self-interested agents to solve what traditionally are posed as monolithic single-agent sequential decision problems. The paper builds on the hypothesis that the optimal solution for the global objective coincides with a Nash equilibrium strategy profile of the agents optimising their own local objectives. The researchers derive a class of decentralised reinforcement learning algorithms that are broadly applicable not only to standard reinforcement learning but also for selecting options in semi-MDPs and dynamically composing computation graphs.

The paper [87] resembles a guideline for researchers on how to build agent-based computational stock markets. The main areas outlined are the key design decisions, the definition of utilities, actions, the time perspective and strategies for the agents, the modelling of the securities, the concept of evolution in financial ABM, and the problem of calibration and benchmarking.

The Eurace@Unibi model [88] is based on the agent-based macroeconomic simulation platform developed within the EURACE project. It is a representation of a closed macroeconomic model with a spatial structure, where it provides a micro-founded macroeconomic model that can be used as a unified framework for policy analysis in different economic policy areas and for the examination of generic macroeconomic research questions. The model is very detailed and large.

A comprehensive literature survey [89] identifies the current landscape, challenges and opportunities for ABM in Finance:

**Landscape of Agent-Based Modelling in Finance:** The literature review [89] takes a comprehensive look at the current state of Agent-Based Modelling (ABM) in finance, examining its origins, methodologies, applications, and impacts. ABM is seen as a modern computational methodology, empowering researchers to view economics and finance from a different perspective.

In finance, ABMs have been used to analyse a diverse set of phenomena. From explaining clustered volatility and fat tails, to modelling systemic risk, to examining the effects of NASDAQ decimalisation, and housing markets, ABMs have displayed
significant potential. The review also refers to a growing trend in ABM to move beyond the traditional Walrasian model, by incorporating institutional emergence, and looking at economies as many-level systems. In policy-making too, ABMs are gaining traction, as they offer a unique opportunity to test the potential impacts of different strategies in a controlled, risk-free environment.

**Challenges in Agent-Based Modelling in Finance:** However, as with any emerging field, ABM faces significant challenges. The literature review highlights two main issues: the curse of dimensionality and the problem of forecasting. The former refers to the exponential increase in computational complexity as the number of agents and their possible states increases. The latter arises due to the inherent unpredictability and randomness that exist in financial markets, which makes it difficult to forecast future states accurately using ABM.

The review also notes the challenges associated with the development of user-friendly software tools and the necessity of parallel execution to handle large-scale models. The lack of standardised methodologies and community models presents an additional hurdle for the widespread adoption and advancement of ABM.

**Opportunities in Agent-Based Modelling in Finance:** Despite these challenges, the literature review argues that there are numerous opportunities for ABM in finance. One key opportunity lies in the integration of micro-data, including social network data, which can provide a more detailed and accurate representation of the behaviour of individual agents and their interactions.

The possibility of moving to large-scale and full-scale models represents another opportunity. As computational power continues to increase, the feasibility of creating highly detailed, large-scale models of financial systems also increases. This could provide unprecedented insights into the workings of financial markets and institutions.

Furthermore, the review points to the potential for ABM to contribute to a new kind of economics, termed nanoeconomics. This approach would aim to understand the economy from the bottom up, focusing on the interactions of individual agents and the emergence of macro phenomena from these micro-level interactions.
Finally, the review identifies the development of new kinds of ABMs for economics and finance as a promising area for future exploration. By incorporating innovative agent types and interaction mechanisms, these new models could offer fresh perspectives on complex economic and financial issues.

In conclusion, while ABM faces significant challenges, the potential benefits that it offers make it an exciting and promising field for future research in economics and finance.

3.3.2 ABM with ML

Incorporation of ML and DNN into ABM has been a new development from the financial ABM perspective, but the incorporation of ML and RL into ABM has been done extensively in the scope of games where agents in games have agencies and process their environment and act by utilising the DNN, as it will be explored in the following sections.

The paper [90] utilises AI assisted deep reinforcement learning and implements an agent-based model for addressing the needs of socio-economic challenges introduced by designing and testing economic policies. In the paper modules called social planners are trained to discover tax policies in dynamic economies that can effectively trade-off economic equality and productivity. A two-level deep reinforcement learning approach is applied to learn dynamic tax policies, based on economic simulations in which both agents and a government learn and adapt. The researchers claim that a data-driven approach does not make use of economic modelling assumptions and learns from observational data alone. Four main contributions that the model claims to accomplish are economic simulation environment that features competitive pressures and market dynamics. Validation is done by demonstrating that baseline tax systems perform in a way that is consistent with economic theory, including in regard to learned agent behaviours and specialisations. The second contribution of the paper is the demonstration of how AI-driven tax policies improve the trade-off between equality and productivity. The third contribution of the paper is the observation of several emergent features such as AI-driven tax policies are qualitatively different from baselines, setting a higher top tax rate and higher
net subsidies for low incomes. The paper is inspiring and novel by craft full fusing an agent-based model of the financial ecosystem, with deep reinforcement learning utilising agents. The research presented in this paper includes methodologies and ideas that can be benefited from for developing an agent-based model of pension ecosystem with personalised savings strategies. The paper proposes a model that can be adapted and developed for a much more complex environment.

The research [91] proposes a novel way to communicate and coordinate agents in multi-agent systems. Multi-Agent Reinforcement Learning (MARL), through rewarding agents for having causal influence over other agents’ actions. At each timestep, alternate actions that it could have taken and their effect on the behaviour of other agents are evaluated, actions that lead to bigger changes in other agents’ behaviour are considered influential and rewarded. The researchers present empirical results reflecting that influence leads to enhanced coordination and communication in challenging social dilemma environments, dramatically increasing the learning curves of the deep RL agents and leading to more meaningful learned communication protocols in a way that the training and execution can be done in distributed systems. The potential existence of social dilemmas in the context of the pension ecosystem is an area of interest for the research, the relationship between employer contribution and employee contribution, contributor behaviour of increased saving leading to consumption decrease, which might lead to a decrease in demand might constitute some of the social dilemmas in the ecosystem.

3.3.3 RL

Reinforcement Learning (RL) has gained significant interest in recent years due to its potential to autonomously learn complex tasks. It has rapidly evolved, with multiple advancements marking this journey.

Mnih et al.’s pioneering work in combining Deep Learning (DL) with RL presented the first model that successfully learned control policies directly from high-dimensional sensory inputs [92]. Their Deep Q-Network (DQN) algorithm allowed an agent to learn to play video games to a superhuman level, using only the raw pixels and game score as inputs. This integration of DL with RL, now termed Deep
Reinforcement Learning, has proven to be extremely potent, enabling RL to deal with high-dimensional, complex state spaces.

In another development, Konda and Tsitsiklis focused on Actor-Critic methods, a type of RL algorithm that utilises two components: an actor that controls how our agent behaves, and a critic that measures how good the action taken is [93]. The actor-critic model offers a balance between high bias, low variance methods (like value iteration) and low bias, high variance methods (like policy iteration). This approach has significantly improved the efficiency of learning algorithms and their convergence properties, being particularly effective in continuous action spaces.

With the use of more powerful computational resources, the application of RL in virtual simulations has been explored [94]. This method allows agents to be trained in a simulated environment, which can then be transferred to real-world applications.

Another significant advancement in the field of RL is the introduction of curriculum learning [95]. Inspired by the concept of learning from easy to hard tasks in human education, curriculum learning involves designing a sequence of learning tasks that leads to more effective or faster learning. This approach has been especially useful in complex tasks, where direct learning is computationally expensive or practically infeasible.

The field of RL has seen significant advancements, increasingly complex approximators and the utilisation of more compute resources will lead to new progress in the future.

The ability to train a non-linear approximator with feedback from the environment is a very good toolkit to calibrate the agent behaviour according to various myopic, intertemporal, and global criteria. In this thesis different econometric utilities, intertemporal or global rewards need to be used. Having a simulator makes the use of RL possible. Economic problems are suitable to be modelled as Agent Based Model simulators where the system manifests itself by the interactions and decision of the agents.
3.3.3.1 Deep RL

Deep Reinforcement Learning (Deep RL), a field at the intersection of Deep Learning (DL) and Reinforcement Learning (RL), has demonstrated remarkable capabilities in various tasks, from game playing to autonomous driving, robotics, and beyond. A key advantage of Deep RL is its ability to learn directly from raw, high-dimensional inputs in complex environments.

A significant development in Deep RL is the Proximal Policy Optimisation (PPO) algorithm [96]. PPO, introduced by Schulman et al., is an advanced policy optimisation method. It aims to address the challenges associated with training policy gradient methods such as instability and high sensitivity to hyperparameters. PPO accomplishes this by using a novel objective function that imposes a constraint on the policy update at each step, ensuring small, stable updates and thus improving training stability.

The Stable Baselines (SB) library is another important contribution to the field of Deep RL [97]. It is a set of high-quality implementations of reinforcement learning algorithms in Python, offering a unified and easy-to-use interface to RL practitioners. It aims to provide a reliable foundation for RL research and development. The details on the implementation of Stable Baselines have been discussed extensively, demonstrating its effective application and use in a range of Deep RL problems [98].

As machine learning models became more complex and diverse, the need for more sophisticated training environments arose. This led to the development of PettingZoo, a Python library for conducting research in multi-agent reinforcement learning. It offers a Gym-like API, easing the process of creating, training, and evaluating agents in a multi-agent context [99].

One of the most innovative recent trends in Deep RL is meta-learning, often referred to as “learning to learn” [100]. It aims to design models that can learn new skills or adapt to new environments rapidly with a few training examples. This is achieved by training the model on a multitude of learning tasks, such that it learns an internal representation that is useful for learning new tasks. In essence, meta-
learning shifts the focus from learning a specific task to learning a strategy to learn any task.

The work [92] constitutes “first deep learning model to successfully learn control policies directly from high-dimensional sensory input using reinforcement learning. The model is a convolutional neural network, trained with a variant of Q-learning, whose input is raw pixels and whose output is a value function estimating future rewards.” A later work [7] demonstrates a clear path to successfully implement a novel artificial agent, termed a deep Q-network for end to end deep reinforcement learning, without explicitly modelling the environment. The research bridges the divide between high-dimensional sensory inputs and actions. Although the input of the data differs drastically as image vs global overview, individual state and market data; the fully connected intermediary layers and action mapping layers can be utilised by applying in the agent decision process, a striking difference is the multi-agent nature of the pension ecosystem.

Further research [21] in the field demonstrated wide and flexible capabilities of RL algorithms and state-of-the-art methodology by covering the main streams of value-based and policy-based methods, as well as central algorithms in deep reinforcement learning, including the deep Q-network, trust region policy optimisation, and asynchronous advantage actor-critic.

Zhang et al. [101] extend the field of reinforcement learning “by considering the environment is not given, but controllable and learnable through its interaction with the agent at the same time. This extension is motivated by environment design scenarios in the real world, including game design, shopping space design and traffic signal design. The agent derives a policy gradient solution to optimising the parameterised environment. Furthermore, discontinuous environments are addressed by a proposed general generative framework.” The research “aims at learning to design an environment via interacting with an also learnable agent or multiple agents”. The proposed methodology of integrating the simulation parameterisation to the training loop can be achieved due to the design of the simulation environment in my research of Pension Ecosystem, such a methodology can be thought as an
adversarial learning challenge for the agents in challenging market conditions and shocks, and greatly benefit the training to be robust and flexible.

Deep reinforcement learning allows complex behaviour to be learnt from the interaction with the environment; the highly heterogeneous income trajectories of different professions and work modes make the deep reinforcement learning necessary. The other factors covering the asset dynamics and the relationship between asset and income dynamics make this a complex problem that needs a capable and complex approximator.

### 3.3.4 MARL

The field of Multi-Agent Reinforcement Learning (MARL) has shown exciting potential for handling complex, dynamic environments involving multiple interacting entities. However, it comes with its unique set of challenges and advancements.

A critical challenge in MARL is the computational expense and complexity it incurs. The interaction between multiple agents can lead to a non-stationary environment, making the learning process particularly challenging [8]. The work by Samvelyan et al. highlighted these difficulties using the example of the multi-player game StarCraft II, emphasising the need for robust, efficient MARL algorithms to handle the inherent complexity and computational demands.

Despite these challenges, notable advancements have been made in the domain of MARL. A surprising discovery made by Yu et al. revealed that Proximal Policy Optimisation (PPO) can successfully optimise MARL problems in cooperative environments [102]. This finding is significant as it proves that PPO, primarily designed for single-agent settings, can also be adapted for multi-agent scenarios. This expands the scope of PPO’s utility and makes it a viable choice for MARL problems.

A noteworthy breakthrough came with the work of Yang et al., which introduced a MARL methodology designed for handling a high number of agents, on the scale of millions [103]. The methodology addresses the curse of dimensionality and the exponential growth of agent interactions, two critical challenges in large-scale MARL problems. This innovative approach opened new horizons for MARL,
enabling it to scale and adapt to an unprecedented number of interacting agents.

Yang et al. have also provided a comprehensive literature review on MARL, summarising the existing methodologies, the main challenges, and the potential directions for future research [22]. This review has become a valuable resource for researchers in the field, helping them understand the current landscape of MARL and guiding them towards addressing the most pressing issues in the field.

The state of the art multi-agent research [103] introduces a multi-agent reinforcement learning methodology that is designed for a very high number of agents on the scale of millions; the main problem that addressed is the curse of dimensionality and exponential growth of agent interactions. The suggested answer to the problem is called Mean Field Reinforcement Learning, where the interactions within the population of agents are approximated by those between a single agent and the average effect from the overall population of agents. The forces mutually reinforce each other, individual agents’ policies be optimised in accordance with the dynamics of the population, and collectively, the dynamics of the population depend on the decisions of individual policies. This model provides an efficient methodology, which can enable the simulation of the pension ecosystem with millions of contributors and pensioners in a reasonable time frame.

The multi actor nature of the pension ecosystem which consists of people working towards retirement, businesses and government, where the many agents of these actor types interact to manifest asset and income dynamics, make it possible to test and learn in many different situation that can happen in state space, such an interaction between agents makes the problem space very vast, and having a complex approximator that can improve its policy by interaction in different situations instead of simulating all possible states is important to develop multi agent models of the pension ecosystem that are robust, and can reflect various non-stationary dynamics, which in result will help to develop strategies that are robust to black-swan events, and actively adapt to varying environment dynamics, that doesn’t need income or asset dynamics to be hard-coded in.

This thesis differentiates by utilising Deep RL to learn from raw, high-
3.3. AGENT BASED MODELS AND DEEP RL

dimensional inputs in intricate scenarios, as demonstrated by various applications such as game playing, autonomous driving, and robotics, making it an appropriate choice for dealing with diverse income trajectories and asset dynamics [92, 7]. Particularly, PPO provides an advanced policy optimisation method with improved training stability, allowing for robust policy learning and optimisation [96]. Also, the Stable Baselines library further enriches Deep RL through high-quality implementation of reinforcement learning algorithms [97]. An emerging trend in Deep RL, meta-learning or “learning to learn”, offers a promising approach to adapt to new tasks rapidly, a pressing need in the continually evolving pension environment [100]. This becomes especially relevant when learning to design an environment while interacting with also learnable agents, as highlighted by Zhang et al. [101].

The inherently multi-agent nature of the pension ecosystem necessitates the application of MARL, despite its challenges such as computational expense and non-stationary environments [8]. Recent advancements, like the surprising compatibility of PPO with MARL problems [102], and the Mean Field Reinforcement Learning technique addressing the curse of dimensionality [103], equip us with capable methodologies for simulating the pension ecosystem with numerous interacting agents.

### 3.3.5 Efficient Computing Framework

Optimising agents in an agent based model with deep reinforcement learning for various environment states that manifest by the interaction of the agents in the environment requires a large amount of sample. Simulating the environment agent interactions that are usually either DNN inferences for actions or mathematical operations for calculating the resulting micro operations can be vastly parallelised if the hardware such as GPUs are being utilised, it is important that the parallelisation of these operations can be abstracted from the modeller as much as possible, because we should not require our economic modellers to be cable of highly parallelised CUDA coding in C++. JAX is a groundbreaking domain-specific Just-In-Time (JIT) compiler that generates optimised accelerator code from pure Python and Numpy machine learning programs. The idea behind JAX is to leverage famil-
iar programming tools such as Python, Numpy, and Autograd to enable machine learning researchers to harness computing power efficiently and effectively [104]. JAX utilises the XLA compiler infrastructure to generate optimised code for certain subroutines of a program, which are most suitable for acceleration. These optimised subroutines can then be called and orchestrated by arbitrary Python code. Importantly, JAX is compatible with Autograd, supporting both forward- and reverse-mode automatic differentiation of Python functions to arbitrary order.

A fundamental characteristic of JAX is its capability to manage pure-and-statically-composed (PSC) subroutines, orchestrated by dynamic logic. Essentially, these PSC subroutines are the ones that do not have side effects and can be represented as a static data dependency graph on those primitives. Provided the primitives are themselves acceleratable, these PSC routines are excellent candidates for acceleration. The JAX system JIT compiles these PSC subroutines using high-level tracing with the XLA compiler infrastructure [104]. Another critical aspect of JAX’s design is its extensibility. It allows the addition of new primitives that simply need to be annotated with a translation rule, which builds corresponding XLA computations. With this capability, JAX lifts the XLA programming model into Python and enables its use for acceleratable subroutines while still allowing dynamic orchestration.

In summary, JAX [105] presents a powerful tool that integrates the convenience of Python programming, the versatility of Numpy and Autograd, and the efficiency of XLA. It aims to empower machine learning researchers to write high-performance code that can seamlessly scale to leverage accelerators and supercomputers. As a result, JAX plays a significant role in advancing the field of machine learning by addressing the tension between accessibility to machine FLOPs and research-friendly programmability.

Flax is an open-source neural network library designed for flexibility and built on top of JAX, a high-performance JIT compiler [106]. Flax aims to provide an interface that is intuitive to researchers while allowing users to harness the performance capabilities of JAX. With a focus on enabling custom and complex network
architectures, Flax does not hide JAX’s powerful features behind high-level APIs but instead provides users with a clear path to define and manipulate custom models or training loops. It fosters an ecosystem where the sharing and reuse of model components or optimisers are straightforward and promote reproducibility. Therefore, Flax represents a significant milestone in making powerful machine learning tools more accessible and user-friendly for the research community [106]. There is a missing link between the optimised code execution libraries that leverage the parallelised hardware, and the researchers that do economic modelling, where the research needs to be increasingly complex and should be able to incorporate large amounts of data and utilise more computational power to address the needs of the society and maintaining pace with other fields that incorporate more complex models and compute power for more accurate and powerful results. The methodologies introduced in this thesis can be utilised to develop intermediary frameworks for economists to leverage more computational capability with less expertise on parallelised efficient software development.

3.3.6 The Approach

This research’s literary framework encompasses various fields, fundamentally concentrating on the financial issues within the comprehensive pension ecosystem. This ecosystem includes individuals with diverse profiles, like people of different ages and occupations, income fluctuations during employment, decisions on consumption and savings, shocks that influence income, lifelong investment strategies, taxes, markets, and regulations. The conventional approach for modelling economic and financial dynamics is econometric modelling, which uses straightforward mathematical equations to encapsulate the relationship between the ecosystem’s various elements. Despite its usefulness in simplifying complex financial principles for wider audiences, the constraints of classic econometric methodology were acknowledged. However, it was seen as the only feasible approach due to the technical limitations of available problem-solving tools, such as solutions to simple equations and numerical approximators with limited computing power.

Agent-Based modelling has provided a broader approach to capturing the mul-
3.3. AGENT BASED MODELS AND DEEP RL

Multiple components of intricate financial ecosystems like pensions. This approach doesn’t require the hardcoding of phenomena such as income dynamics into mathematical equations. Instead, these dynamics emerge from interactions between model agents. While Agent-Based Models have been adopted and used in various fields, they often fall into the same pitfalls as classic econometric models, in which relationships and interactions between agents are hard-coded as simple mathematical equations. These models, though useful in exploring emergent features from simple rules, struggle to encapsulate the complexity of the modelled environments’ relationships and dynamics.

In this thesis, I first model fundamental principles of age and income relationship isolated from the other factors, then progress to a panel and population survey calibrated simulator of the asset and income dynamics, where employees optimise saving and investor behaviour with Deep RL in ABM, then progress to a more advanced ABM optimised with multi agent based deep reinforcement learning.

Multi-Agent Deep Reinforcement Learning is utilised to learn the relationships and interactions between agents automatically. In this framework, the primary focuses are the interaction paths, hard-coded physical principles, and real-world observed aggregate statistics used to calibrate our model. Deep neural networks allow for a higher level of heterogeneity, capturing more diverse strategies and calibrating for complex behaviour.
Chapter 4

A Generative Model for Age and Income Distribution

In the thesis, I devise a range of models for capturing the dynamics of the pension ecosystem. An exploratory analysis of pension demographics was conducted. I use a generative model to represent age and income dynamics, evolving the age-income population for simulation and forecasting purposes. Age and income are among the foundational tenets of retirement finances.

4.1 Introduction

A universal element of societies is the emergence of hierarchical organisational structures within professions. People develop work experience through time and manage to obtain jobs of increasing responsibility and increasing level of income with time. Hence, it is a natural property of income distribution to be correlated with work experience and age; nevertheless, most income models do not study the relationship between income and age, and consequently between income distribution and demographic changes. This chapter introduces a model of income, dependent on age-specific model parameters and random shocks. The model contributes to the understanding of the relationship between age and income and its dynamics.

Our aim is to compare the estimated parameters in the UK and the US age and income distribution to find out similar characteristics of age and income across states, as well as the contrasting differences. A simple age and income model is
fundamental for the development of a sustainable pension system. The model focuses on the age and income relationship and further factors, such as occupational levels, are not considered. The model is estimated via panel survey data from the UK and population survey data from the USA. The data from panel surveys track the same individuals for the duration of the survey, and the population survey is repeated with different people each wave. The results reflect a clear income-age relationship in the UK and US, a clear structure of the joint distribution characterised by rapidly increasing income at younger ages, followed by income levels stabilising near mean income but spreading till retirement. At this point, the income decreases and concentrates around mean retirement income. The chapter demonstrates a flexible methodology to estimate parameters from population surveys, as well as panel surveys. The chapter provides a simple generative model to evolve age-income population for simulation and forecasting purposes, which can constitute the foundation for future studies of financially sustainable pension systems by providing a benchmark for capturing age and income relationship. The purpose is to have a baseline model simple enough for isolating age and income relationship of income dynamics. Such a model will serve for investigating the properties of a sustainable and balanced pension system. The mean and standard deviation statistics from the panel and population surveys in Fig. 4.6 and Fig. 4.1 from observed panel data and simulation results reflect a clear relationship between age and income. More complex models, which investigate additional factors, and profile heterogeneity of income dynamics are out of the scope of our work.

Previous research on income have been conducted, and the research focused on investigating and explaining wage dynamics. Champernowne explicitly introduces a first-order Markov process to model the time-evolution of wages [107]. Following Markov process path, the validity of the first-order Markov assumption is tested by Shorrocks [108]. Following research introduces a second-order Markov process, yet neither of these works links individual wage dynamics to time-evolution of the distribution of wages [109]. A different approach focused on poverty, which deals
with modelling individual data using linear regression and transitioning to poverty (probit model) [110]. A more comprehensive model incorporating various factors is developed to estimate transition probability in wage quintiles conditioned on various regressors, including education, experience and age [111], furthermore study both intra- and inter-group inequality. The persistence of the low pay state and factors affecting the low pay probability are expressed with a generalised regression model. For modelling low income transitions the previous research uses British Panel Data for the '90s, focusing on the transition probability and state dependence for the poverty status [112] and defines the poverty transition equation as coarse-grained dynamics. Inequality and upwards mobility between quintiles considering gender effects are investigated [113]. The previous models in the literature incorporate numerous external variables, distribution characteristics and functions, such as innovation constants or they are limiting their scope to the investigation of dependence on a single variable [114, 115]. A more recent article by Guvenen investigated a model for which focal variables are the human capital consisting of education, work experience, and idiosyncratic shocks [116], following research modelled male income for studying the impact of labour income taxation policy on inequality [117] The referred life-cycle model’s distribution characteristics of the pre-tax income arise from the differences in the individual’s ability to learn new things and idiosyncratic shocks. Previous research tried to capture the income dynamics with Markov Models, linear autoregressive models, or by relying on econometric toolset such as covariance matrices. We investigate a generative model with an empirical distribution for sustainable age and income relationship in a population; we achieve this via an income evolution model with an age-dependent parameter, estimated from previous population and panel surveys.

The previous research [118] presents a two-class distribution, the majority of the population is described by the exponential function and a small fraction of the population of higher income individuals is described by power law [119]. The BHPS and IPUMS CPS data are top-coded and not suitable for studying the power law at the top, but the majority of the population as reflected in Fig. 4.3 is consistent
with empirically well-established exponential distribution of income [120].

Although there are models that incorporate indirectly the age as years of experience in job for studying income dynamics. There are no studies, to the best of our knowledge, studying the joint distribution of age and income in the scope of income evolution.

In contrast to previous research, our study introduces a dynamic model that describes the income-dependent only on age and previous income. This chapter investigates the stationary property of the income distribution dependent on age. We provide a model in which the mean and variance of income given age are preserved at any time point.

4.2 Methods

We introduce a simple model which focuses on age and income relationship and differs from recent literature by not incorporating other variables such as occupational level, level of education and skill coefficients. The model is stationary, i.e. the mean and variance of income given age are preserved in time. The model is utilised to represent observed panel data for gaining empirical insights regarding age dependent income dynamics and mobility. The calibrated model can be utilised as a simple generative model to evolve an age-income population for simulation purposes and it provides a theoretical background for studies focusing on ageing and pension income of the population. We initially assume the following model, by which $\mu(.)$ and $\sigma(.)$ represent a function of individual-specific age, income, and additional parameters $\theta_i$ or $\lambda_i$, for the sake of generality. $\mu(.)$ is a function capturing mean income characteristics, and $\sigma(.)$ captures the variational characteristics of the income. We consider the following individual income stochastic process for an economic agent $i$ characterised at each time step $t$ by a given age $a$ and income $y$: 

$$y_{i(t+1)} = \mu(a_{i(t+1)}, y_{it}|\theta_i) + \sigma(a_{i(t+1)}, y_{it}|\lambda_i) \eta_{it}$$  

(4.1)

The characterising insights in Fig. 4.1 from the panel data lead to the assumption that the probabilistic step at time $t$ depends only on the age and income of the
4.2. METHODS

4.2.1 Defining Income and Age Dynamics

Earnings of individual \( i \) at the time step \( t \) is denoted as \( Y_{i,t} \) and its logarithm is \( y_{i,t} \).

The parameters that describe the income process are: age-dependent persistence parameter \( q_a \), age-dependent mean \( \mu_a \) and age-dependent standard deviation \( \sigma_a \). The income shock process consists of independent random shock \( \eta^j_{i,t} \) which is normally distributed with mean zero and variance 1, and it is applied to \( \sigma_a \), the model can be defined as follows:

\[
y_{a+1,t+1}^j = q_a y_{a,t}^j + \mu_a + \sigma_a \eta^j_{i,t}
\]

Averaging income \( y \), for individuals who are \( a \) years old, gives \( \bar{y}_a \), which denotes the average income for age group \( a \) across all individuals \( i \) and periods \( t \). Assuming that the age-dependent income profiles are stationary, we can average incomes \( y_{a,t}^j \) across individuals and time to get the following equation:

\[
\bar{y}_{a+1} = q_a \bar{y}_a + \mu_a
\]

where \( \bar{y}_a \) denotes the average income for age group \( a \), taken across all individuals \( i \) and periods \( t \). The following equation can find the estimator for \( \mu_a \):

\[
\mu_a = \bar{y}_{a+1} - q_a \bar{y}_a
\]

The income data from different waves are inflation adjusted to isolate the effects of economic growth.

4.2.2 Data

The British Household Panel Survey (BHPS) [69] from the UK and The Current Population Survey (IPUMS CPS) [71] from the USA are used for estimating the parameters of the model in Eq. 4.2 and comparing the results of simulated data and surveys. The BHPS is a panel survey conducted between 1991 and 2008. For our model we focus on labour income data, which captures wage, salary or self-
employment income. To investigate population characteristics, we also incorporate
other income sources and call it “Total Income”, which additionally captures the
transfers, pensions, grants, aids, state-benefits, dividends, capital income and rents.
BHPS provides individuals specific longitudinal weights for ensuring the representa-
tiveness of the population. Two types of weights are provided with BHPS. The
first wave is weighted for adjusting population marginals at the households and
post-stratified to the population age by sex marginals. Consecutive waves are re-
weighted to take into account sample attrition, variables such as address change,
household region, age, sex, race, employment status, income total and composi-
tion, educational qualifications [70]. Panel Survey is conducted via questionnaires
with tracked individuals of the initial sample. A detailed explanation of the relevant
variables from BHPS dataset can be found on the Supplementary Material. There
is an extension to the sample population in 1999. For the USA, IPUMS CPS is
used, which is annually conducted with different samples each year. In contrast
with BHPS, the Labour Income from IPUMS CPS does not include self-employed
income, and the weights are cross-sectional.

Income distribution, age distribution and income-dependent age distribution
from the surveys are utilised for parameter estimation and further analysis. \( q_a, \mu_a, \sigma_a \) are the key parameters estimated according to the proposed model. Following
investigation and interpretation of the estimated parameters, these parameters are
used to simulate the population’s income transitions. The simulation is initialised
using the panel data from the Wave 1, and the income evolution function on Eq. 4.2
is applied transitively in an iterative approach to the data for simulating successive
waves. The simulated data is plotted, interpreted and finally compared with the
observed data.

Fig. 4.2 reflects the Population Pyramid in the UK, and how the shape evolved
over the 18 years considered. The population pyramid of USA can be found in the
Supplementary Material. The UK population sample from BHPS has a relatively
balanced population with a slight weight towards younger cohorts initially in 1991,
which denotes Wave 1. The UK population gradually got older, and the population
pyramid reflects mass’s shift towards older generations, this shift happened gradually over the years. The US population from CPS reflects a young population in Wave 1 with a notable skew towards younger cohorts, after 17 years the US population loses this property towards younger cohorts and gets significantly older. Both the UK and US population get older and reflect a trend towards an ageing population, which will significantly impact the pension system.

The shape of the population pyramid and its evolution with time from the panel survey reflect an ageing community [72]. JDFs of Total and Labour Income in the UK and USA reflect that, there is a sharp increase between the ages 15-20, which can be interpreted as the beginning of the work-life, transitioning from part-time work to full-time work, and graduation from higher or vocational education. The most significant difference of the UK Total JDF in contrast to the UK Labour JDF is the tail section corresponding to the retired population, which denotes the significant percentage of individuals older than 55. The tail section is relatively concentrated, which can be explained by the state pension benefit levels and mandatory social security system. The US population reflects a surprisingly sparse older cohort for the Total Income data, and the most significant difference to the UK is the relatively lower income levels compared to the wage income. In the US population higher variance spread to a wider band, which might be caused by a non-standard retirement system not supported by strong state pension benefit and mandatory pension schemes during employment.

The comparison for the model simulation and observed data shows common characteristics, as the joint distribution of age and income in logarithmic scale is presented in Fig. 4.1: an initial sharp rise between the entrance to the graph on 16 years old, the amount of 16 years old includes pocket money, allowance and part-time or internship jobs. There is a steep increase in mean and variance between the initial income and income at the age of 20. The increase is sharper for the mean in comparison to the variance. The population’s mass has similar characteristics with near 23k GBP annual income, and for ages between 20-45. Between ages 65-75, there is a significant decrease in income and after 75, the income converges to a
certain mean. The data and the models provide an essential tool to tackle problems related to an ageing population and shocks introduced by technological and political changes.

In compliance with the literature [118], the population is divided as two-classes, and the majority of the population covering low and middle income follows Boltzmann-Gibbs distribution. The observed and simulated populations from the UK and USA are reflecting exponential characteristics for low and mid-income individuals, log-linear PDF plots reflect similar PDF characteristics in Fig. 4.3

In the following sections we will focus on labour income and employed labour population. Total Income covers all of the income streams including transfer income such as pocket-money, labour income, capital income, and pension income; these different streams might be governed by varying dynamics non-uniform across the type of income; so we decided to focus on labour income, which involves the broadest section of the population; with most significant impact. The only other primary source of income in terms of gross value is the capital income, which might be significantly affected by other factors such as inter-generational shifts, market conditions and global financial state. In order to focus on labour income dynamics, the other income sources are left out of our modelling.

### 4.3 Data Processing and Calibration

BHPS provides a vast amount of socioeconomic data for each individual and household participating in the study. The columns of income, age data, the individual’s statistical weight representative of the British population and overall survey with the individual’s intra-wave unique identifier mostly suffice for this chapter’s scope. PID, Wxage12, wFIYR, Wxrwght fields of BHPS are used for each wave.

The Income variable xfiyr is each individual’s annual income, including labour income, benefits, pensions, transfer income, and investment income. Participants were asked according to annual income in the reference year from September in the year prior to the interview until September in which interviewing begins [70]. The income figures are adjusted for inflation, as part of pre-processing. During the
4.3. DATA PROCESSING AND CALIBRATION

Dataset preparation, a floor wage is determined to exclude in labour income, which denotes to excluding part-time and short-employment income. The income data is inflation-adjusted and transformed into log-domain.

IPUMS harmonises the CPS and provides IPUMS CPS micro-data. The IPUMS CPS includes a large spectrum of topics such as demographics and employment, as well as supplemental studies such as the Annual Social and Economic Supplement (ASEC). Each individual can be identified by “CPSIDP”, “INCTOT” and “INCWAGE” correspond to the total income and wage income, and “ASECWGT” denotes the weights derived from ASEC Supplement. The data set is topcoded, and specific codes are used for labelling missing and incorrect data. The ages over 80, 90 and 99 are top-coded till 2004, and after 2004, the top-coding bins are determined as ages 80, 85, 90 and 99 by the panel data collectors [71]. Although this dataset contains high-income individuals, there is top-coding applied, so individuals with very high income are not included.

4.3.1 Fitting Distributions

Estimating the income evolution function parameters is the most critical part of the research, and the decision depends on various factors such as the type of data, bias, and assumptions. Various techniques are investigated, leading to different results, with each having unique strengths and weaknesses.

The first method investigated is Generalised Method of Moments (GMM), which presumes that the first three moments of the income evolution functions provide the necessary information for approximating the underlying generative process. The equations of the first three moments of the income evolution equation can be solved for the parameters $q_a$, $\sigma_a$, $\mu_a$. Both of the BHPS from the UK and the IPUMS CPS from the USA can be used for estimation with Generalised Method of Moments (GMM) with first three moments as reflected in Fig. 4.4 and on the Supplementary Material.

The second method utilises the micro-data from the longitudinal surveys, which tracks the individual for consecutive years. The parameters are approximated to fit the income evolution function using least squares minimisation for the individuals
4.4. THE GENERATIVE MODEL

participating in the studies for consecutive years. The BHPS from the UK is a suitable micro-data consisting of a panel survey, and the survey tracks income of the same individuals over the years.

4.3.2 Estimation for Generalised Method of Moments (GMM)

The three moments of the age income evolution function are utilised to find a polynomial; afterwards, the equation is solved for $q_a$, $\sigma_a$ and $\mu_a$; at this point, a observed solution for parameters is found, but the relationship captures only the dynamics of the first three moments. Calculations can be found in the Supplementary Material.

The statistical variables such as $\bar{y}_a$ are found for each wave and than averaged across waves for finding a one set of stationary variables, which can be used to estimate $q_a$, $\sigma_a$, $\mu_a$. The details and derivations for the GMM estimation technique can be found in the Supplementary Material.

4.3.3 Estimation of Least Squares for Micro Data

The Least Square Method requires that an individual’s income for two consecutive years be existent in the dataset, this restriction is fulfilled by the BHPS, a panel survey, but the CPS IPUMS population survey does not satisfy this condition. The income data from two consecutive years per agent is used to estimate age-specific parameters, which characterise the income evolution function at Eq. 4.2. LSM tries to estimate parameters by fitting the data to the income evolution function.

4.4 The Generative Model

The model can also be used for simulation and forecast, tracking income trajectories of the individuals, providing a bench table for observing the stylised facts and complex properties of the income dynamics. Following the estimation of model parameters, the model is bootstrapped with data from Wave 1 for initialising the simulation. Each individual from Wave 1 is initialised as an agent in our model. According to Age Income Evolution Dynamics Eq. 4.2, the income of an agent is transitively updated at each consecutive wave update. $\eta_a$ provides the random feed, which introduces variability for the income evolution of the agents. At each wave update, a new generation of agents consisting of 25 years old individuals from the
initial wave are injected. Following each wave, distributions corresponding to the state of the simulated population are calculated. A full calibration of the model is shown in the Supplementary Material.

4.5 Parameters

The optimised performance of these three methods are compared and discussed in the following sections. No boundaries are explicitly imposed by LSM estimation. The $\mu$, $\sigma$ and $q$ variables are independent of each other, but the estimation process or data itself can introduce a slight dependence. The GMM estimation technique results in minimal $q$ values near 0, so the estimated parameters approximately resemble an auto-regressive model. However, despite near 0 negligible $q$ values, the $q$ plot has a distinct shape with an increasing trend with a small decrease between 25 and 30, has very different characteristics depending on the estimation method. The GMM estimation method results in minimal $q$ and the $\mu$ reflects the characteristics of $\bar{y}$, which is in compliance with this estimation method’s nature. The $\mu$ value increases at first and then plateaus and slightly decreases near retirement. On the contrary LSM estimation mainly characterises the income with an increasing $q$ parameter, so the $\mu$ parameter has limited effect and reflects a decreasing trend. $\sigma$ values reflect a distinct trend of initially decreasing values with a spike around the age of 34 followed by a stable decrease and noisy plateau with a minor increase towards 55. The LSM with bootstrap is the most accurate estimation method and reflects the characteristics of the model clearly.

4.5.1 GMM

GMM estimation technique approximates the $\mu_a$ values to be consistently around 10 and the $q_a$ values are around 0 with an initial sine-like wave followed by a steady increase. The $\sigma_a$ values are around 0.8 and have a positive trend. $q_a$ values display a positive trend as well. The $\bar{y}_a$ and $std(y)_a$ plots of the simulation is similar to the observed data, but the standard deviation plot is particularly noisy. The JDF of the simulation in Fig. 4.5 is sparse, consistent; but not highly concentrated around mean. Both of these methods depend on assumptions about the dynamics of the
income evolution function. The GMM method assumes that the first three moments of the equation are sufficient for estimating the parameters because they provide a solvable system. However, individual characteristics in an age group such as different income levels and clusters within are lost during the moment estimation.

4.5.2 LSM by Individual Transitions

In order to use LSM for approximating the parameters, one needs the individual income transitions in consecutive years, thus identifying the same individual in consecutive cohorts is necessary and the panel studies such as BHPS satisfy this condition. The age-dependent income evolution function is fitted with individual income transitions of consecutive years, which results in consistent parameter plots and the \( \bar{y}_a \) plot of the simulation reflect similar shape with the observed data in Fig. 4.6. The JDF of the simulation in Fig. 4.7 is able to reflect the dispersion among various clusters better because unlike the other methods heavily depending on the statistics such as mean and standard deviation of the entire age group, the LSM utilises individual-level microdata.

The 95% confidence interval with 2000 bootstrap samples for the estimated parameters from UK microdata by LSM can be found in Fig. 4.8. It is evident from the plots of \( \bar{y}_a \) and \( \text{std}(y_a) \) for the observed and simulated data that the model can capture the characteristics of the income conditional on age distribution. A close investigation of simulations on models calibrated with UK Labour Income Data suggests that the GMM is most successful for reflecting the outcomes with similar mean and standard deviation characteristics of all waves after simulation with 18 waves that were simulated with the parameters \( q_a, \sigma_a, \mu_a \) estimated by the GMM. But LSM reflects the individual trajectories, and JDF more accurately. The results showing the performance of GMM method can be found on the Supplementary Material.

A general analysis of the comparison of joint distribution of age and inflation-adjusted income results in the following plots for weighted observed data and simulated data in Fig. 4.5: JDF of the simulated UK Labour data is in parallel with the expectations for GMM Estimation method, consistent and stable, resembling a
similar shape but not concentrated for the heat regions with intense concentrations in Fig. 4.6 and Fig. 4.6. The main difference between the observed and simulated JDFs is the concentration of the mass of the population between 23 and 50.

4.5.3 Wave-Specific Analysis

The population from wave-1 is used for bootstrapping the simulation and the weights of the individuals are not incorporated to the simulation, because the income evolution Eq. 4.2 is the focus of this chapter, and the main purpose is not the perfect representativeness of the initial wave. The new agent injection on 1999 by panel survey is reason of difference in the UK simulation and observed JDF plots. Although the simulation’s initial state is bootstrapped as the unweighted dataset, starting from the second wave, the JDF of the simulated population resembles the characteristics of the JDF from the panel survey with the weighted population, which reflects that the model is successfully capturing income evolution dynamics.
4.5.4 A Simple Pension System

A financially sustainable pension system can be characterised by the balance between inflow and outflow of funds. The specifics and stability of pension system is out of the scope of this chapter, and needs case specific detailed modelling. For a general demonstration, we assume simple inflow and outflow dynamics (Eq. 4.5 and Eq.4.6), which are derived to represent statistical properties of the savings and consumption. Fig. 4.9 reflects the imbalance between inflow and outflow, which results in a deficit.

Pension is assumed to be annually £16368, in light of the median net income before housing costs for all pensioners from DWP Pensioners Income Series in 2008/2009 [121]. Constant alpha for pension saving rate is selected to be 0.0775 to 0.2, in light of OECD Pension Report statistics [122].

Outflow $O_t$ in a given year $t$ is characterised by constant annual pension amount $p$, and count of people above 65 $c_{a>65}$ is assumed to be pensioner counts.

$$O_t = pc_{a>65} \tag{4.5}$$

Inflow $I_t$ in a given year $t$ is characterised by a constant pension contribution rate $\alpha$ and the total labour income of individuals $y_i$. Here, the summation is taken over all individuals $i$, but only for those whose age $a_i$ is 65 or below.

$$I_t = \alpha \sum_{\forall i | a_i \leq 65} y_i \tag{4.6}$$

The amounts are adjusted for inflation and reflect the 2009 levels. The inflow and outflow plots from our simplified generalisation of the pension system reflect a deficit.

4.6 Discussion

The income evolution Eq. 4.2 of the proposed model consists of the parameters $q_a$, $\mu_a$, $\sigma_a$: the persistence coefficient for the respective age group $q_a$, determines the rate of persistence at a given age.
Age-dependent mean income parameter $\mu_a$ expresses the expected age-specific income evolution mean for the next income and behaves such that if the mean parameter is high the persistence $q_a$ is most likely to be lower. If the mean parameter $\mu_a$ is lower, the persistence parameter is higher which signals a potential widening of the income gap for the population.

$\sigma_a$ captures the variability of the individuals according to conditional distribution and incorporates randomness of the shocks.

The social safety nets, basic pension incomes and the Defined Benefit Pension plans are financed via the working population; the ever-growing unbalance towards ageing cohorts needs careful forecasting and planning. The demographic shift will impact the economy’s functioning in general, introducing a heavy burden to welfare states financing the health and pension of the retired population, which will reflect society as taxes and benefit cuts. The best course of action is forecasting the changes and planning in advance for the future.

4.6.1 Interpretation

The $q_a$ persistence estimated by GMM reflects that UK population reflects an initially high $q_a$ value in youth, followed by a relative decrease, and then a consistent increase. The $q$ values estimated by GMM fluctuate around 0 and minimal. The income persistence variable of individuals is not captured by GMM, which does not utilise panel survey’s tracked individual income micro-data each year.

LSM calibration methodology uses longitudinal information regarding the evolution of an individual’s income, this property of LSM makes it suitable to provide higher $q$ values, which captures the persistence. The GMM calibration methodology does not utilise longitudinal data, which makes it applicable with survey data but results in lower $q$ values. This tells us that we are still able to reproduce consistently the data year by year almost disregarding the past year data. Our analysis suggests that LSM is a more reliable calibration model as using longitudinal information appears to be crucial to capture the income evolution dynamics. Furthermore, the model has the power to capture some heterogeneity, if the parameters fit with LSM; and the income evolution function can preserve the bootstrapped wave heterogene-
ity due to persistence parameter $q$ and the randomness injected with $\sigma$ can provide outlier behaviour after multiple waves of income evolution function updates.

The LSM results in a consistently increasing $q_a$ value by the UK model, with a significant jump between ages 25-30, which corresponds with a $\mu_a$ plot consistently decreasing with a significantly sharper decrease between ages 25-30. $q_a$ and $\mu_a$ corresponding each other in an inverse proportion, especially by significant changes, especially by LSM. There are various examples of parameter effects that can be observed from BHPS dataset.

One example of $q_a$ effect is the upwards mobility of age-group between 25 and 30, which is reflected by the increasing $q_a$ values and sharp increase observed on the joint-distribution plot. This effect can be due to finishing higher education and internships, in addition few years of experience, which results in a widening of income scissors. This change in mobility is healthy for the economy and does not represent a negative effect. One assumption should be researched further; if either this initial difference in mobility might limit of people with lower income for upwards mobility.

An example of the $\sigma_a$ mobility is the age group of 30-35, which is reflected by a locally sharp increase of $\sigma_a$ values. Such mobility reflects a bidirectional movement of income for individuals, and such a variation might arise from the short-time employment, interruption of employment for education, temporary jobs and most importantly this mobility might be caused by the initial differentiation according to the education of individuals such as higher education or vocational education. This window represents an increase in the variation of the income.

In general, the shape of the distribution can be explained in three periods; the first period is the introduction to employment and teenagers, which represents income from part-time and temporary jobs at the beginning and start of full-time employment it sharply increases in Fig. 4.1.

The age group of 25-55 denotes the main productive era of economic life, and the income reflects a high dispersion. All of the factors and random shocks act together and result in a dispersed but consistent distribution. Mobility wise this
era provides opportunities for upwards mobility and possesses downwards mobility risks. At the end of this period, income tends to decrease slightly, which reflects a decrease in productivity. Another limiting factor is the minimum wage and state benefits, which introduces a lower bound envelope for the mass. Income sources and affecting factors of individuals in this era vary greatly, which results in the widest dispersion in the entire life-span. Some of the factors are education, social strata, adaptability to innovation, total-work hours per week, experience and expertness, seniority of the jobs and ageism. The third and final era represents the exit from the workforce and retirement, and temporary or part-time jobs for low-income old individuals. The income decreases gradually as the number of individuals exiting workforce increases with time, the income stabilises, and variation decreases significantly. Income in this era is relatively low, and the source is usually pension benefits, state support or temporary jobs. This model’s outcomes can be used for various purposes; the most apparent fields for drawing consequences and planning are the works on inequality and mobility depending on age. Characteristics of workforce entrance, work-efficiency of individuals per age, the structure of the society, pension system, income stability, and the taxation system are the most obvious fields.

In the chapter, two main estimation techniques are investigated, and the corresponding results from the simulated waves are presented. The first estimation method investigated is GMM Estimation. The income regions appear smoothed and spread. The second estimation method investigated is LSM, it utilises the microdata and is suitable for capturing an agent’s income evolution. The JDFs from the simulated waves have the most similar mean characteristics to the observed data.

The LSM evidently performs better by utilizing longitudinal microdata; the GMM estimation method can be applied to both population and panel surveys, provides feasible distributions but with unrealistic modelling of an agent’s individual income trajectory.
4.6.2 Conclusions

We demonstrated (1) a clear income-age relationship, which is reflected by the data from BHPS and IPSUM CPS, as well as simulations. (2) a clear structure of the joint age-income distribution in both the UK and USA. (3) a flexible methodology to estimate parameters from population surveys, as well as panel surveys. (4) a simple generative model to evolve the age-income population with real constraints for evaluating general policy scenarios, that is agnostic about occupation levels.

The model can be interpreted as delivering a premise that the information of an individual’s experience and education can be encapsulated by income. Although in early career, the income dynamics are governed by the initial difference at the level of education and profession; the main dynamics governing income transitioning can be reduced to the relationship between income and age, which collectively encapsulate education and experience. These premises can be leveraged for developing simplified models for evaluating mobility, inequality, welfare state, and pensions.

The proposed model focuses on the evolution of age and income population and the chapter successfully demonstrates a simple model that can be calibrated for age and income that can be used as a backbone for forecasting income and planning pensions. Understanding the dynamics and having the ability to forecast the age and income population is the key to the design of financially sustainable pension systems.

There are different dimensions for future work: one of the dimensions is injecting random shocks to the distributions itself, which can be in the form of new population injection or withdrawal, as well as tuning the $\eta_{it}$ with various means for simulating a global or regional shock, such as pandemics or mass migration. Stress-testing the age and income distribution for different labour market scenarios could lead to relevant policy implications. The second dimension for future work is modifying the simulation system to estimate parameters on the fly, and provide a more adaptive and granular version of the simulation system. The third dimension for future work is incorporating data encompassing more years and more countries and with a higher resolution in time to investigate the role of multiple economic
factors for short, medium and long time horizons.

**Figures**

Figure 4.1: All Years Pooled Age and Income Joint Distribution Function for UK and USA between Ages 15-100 where labour and total income heatmap reflect unique shape. Income after retirement can be observed on Total Income JDFs which include pension and capital income, the data from USA is top-coded which is reflected by stripes of concentrated concentration around ages 80, 85 and 90.
4.6. DISCUSSION

Figure 4.2: Population Pyramid for the UK Income between Ages 15-100 reflecting changes of age distribution in 18 years, which reflects mostly an ageing population

(a) UK Pyramid in 1991

(b) UK Pyramid Wave in 2008

Figure 4.3: PDF Plots of the labour income bins of population and corresponding exponential distributions after 18 Simulated Waves, which confirms the exponential behaviour for low and mid-income individuals

(a) PDF of UK Labour Income

(b) PDF of USA Labour Income
4.6. DISCUSSION

Figure 4.4: $q_a$, $\sigma_a$ and $\mu_a$ Plots for UK Labour Income, the parameters are estimated with two different methodologies of Generalised Method of Moments and Least Squares Method. The LSM utilises longitudinal data, which provides the capability to estimate substantially high $q$ values, which represent the persistence level of an individual’s income, GMM capture much lower $q$ values, due to its incapability of utilising longitudinal data.
4.6. DISCUSSION

(a) Observed Statistics

(b) Simulation Statistics with GMM Estimation

(c) Simulation Statistics with LSM

Figure 4.5: UK Labour Data Observed and Simulation All Years Pooled JDF between ages 25 and 55, the stripe of concentration at last column is due to concatenation of last two ages to fit the plot
Figure 4.6: UK Labour Data Observed and Simulation Statistics representing average and standard deviation of the survey data and simulated population, which reflect similar shape with the statistics from survey data, further insight can be found on PDF plots.
4.6. DISCUSSION

Figure 4.7: JDF Plots of Simulation for UK Labour Income

(a) 1995 with GMM Estimation
(b) 2005 with GMM Estimation
(c) 1995 with LSM Estimation
(d) 2005 with LSM Estimation
4.6. DISCUSSION

Figure 4.8: $q_a$, $\sigma_a$ and $\mu_a$ Confidence Interval for UK Data LSM Estimation reflecting bootstrapped parameter estimation values and the robustness of the estimation method

Figure 4.9: UK Inflow Outflow Plot of our Simple Pension System reflecting inflow of pension savings (assuming savings rate of 0.0775) from contributors who are still in workforce, and outflow of funds to pensioners (assuming weekly pension of £308)
### 4.7 Supplementary Material

#### 4.7.1 Variables of BHPS Dataset

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>pid</td>
<td>Unique ID Describing an Individual</td>
</tr>
<tr>
<td>wAGE12</td>
<td>Age of Individual on 1st of December</td>
</tr>
<tr>
<td>wFIYR</td>
<td>Total annual income including labour income, benefits, pensions, transfer income, and investments</td>
</tr>
<tr>
<td>wFIYRL</td>
<td>Annual labour income</td>
</tr>
<tr>
<td>wXRWGHT</td>
<td>A cross-sectional respondent weight</td>
</tr>
<tr>
<td>wave</td>
<td>BHPS Wave Number between 1-18, Wave 1 denotes to 1991</td>
</tr>
</tbody>
</table>

**Table 4.1:** Description of Variables from the British Household Panel Survey Spanning Years 1991-2008

#### 4.7.2 Model Calibration

We can define the mean and standard deviation of income at a given age $a$ as following:

\[
\langle y_{a,t} \rangle, \text{std}(y_a) \tag{4.7}
\]

\[
\langle y_{a,t} \rangle = \bar{y}_a \tag{4.8}
\]

The standard deviation and mean has the following relation with the squared average of incomes:

\[
\langle (y_{a,t})^2 \rangle - \langle \bar{y}_a \rangle^2 = (\text{std}(y_a))^2 \tag{4.9}
\]

$\eta_a$ has characteristics of the standard normal distribution:

\[
\langle \eta_a \rangle = 0 \tag{4.10}
\]

\[
\langle \eta_a^2 \rangle = 1 \tag{4.11}
\]

\[
\langle \eta_a^3 \rangle = 0 \tag{4.12}
\]

Squaring both sides of income evolution equation 4.2 results in following distribution:

\[
(y_{a+1,t+1})^2 = (q_a y_{a,t} + \mu_a + \sigma_a \eta_t)^2 \tag{4.13}
\]
Eq.(4.9) can be formalised as:

\[
(\bar{y}_a)^2 + (std(y_a))^2 = \langle (y^i_{a,t})^2 \rangle \tag{4.14}
\]

Placing Eq. 4.14 for \(a + 1\) and Eq. 4.13 results in following equation:

\[
(\bar{y}_{a+1})^2 + (std(y_{a+1}))^2 = \langle (q_{a+1}y^i_{a,t} + \mu_a + \sigma_a \eta^i_t)^2 \rangle \tag{4.15}
\]

Expanding the right side of the equation results in:

\[
(\bar{y}_{a+1})^2 + (std(y_{a+1}))^2 = \langle (q_{a+1}y^i_{a,t})^2 + (\mu_a + \sigma_a \eta^i_t)^2 + 2(q_{a+1}y^i_{a,t})(\mu_a + \sigma_a \eta^i_t) \rangle \tag{4.16}
\]

\[
(\bar{y}_{a+1})^2 + (std(y_{a+1}))^2 = \langle (q_{a+1}y^i_{a,t})^2 + (\mu_a + \sigma_a \eta^i_t)^2 + 2(q_{a+1}y^i_{a,t})(\mu_a + \sigma_a \eta^i_t) \rangle \tag{4.17}
\]

\[
= \langle (q_{a+1}y^i_{a,t})^2 + (\mu_a)^2 + (\sigma_a \eta^i_t)^2 + 2(\mu_a \sigma_a \eta^i_t) + 2(q_{a+1}y^i_{a,t})(\mu_a + (q_{a+1}y^i_{a,t}) \sigma_a \eta^i_t) \rangle \tag{4.18}
\]

Averaging the equation by using Eq.(4.14), Eq.(4.10) and Eq.(4.11).

\[
(\bar{y}_{a+1})^2 + (std(y_{a+1}))^2 = (q_a)^2 (\bar{y}_a)^2 + (std(y_a))^2 + (\mu_a)^2 + (\sigma_a)^2 + 2q_a \mu_a \bar{y}_a \tag{4.19}
\]

### 4.7.3 Deriving the Update Equations

For clarity \((\bar{y}_a)^2 + (std(y_a))^2\) is expressed as \((\Delta_a)^2\). The number of parameters can be reduced to 2 using the third parameter of Eq. 4.19 by expressing \(\mu_a\) as \(\bar{y}_{a+1} - q_a \bar{y}_a\) according to Eq. 4.4:

\[
(\Delta_{a+1})^2 = (q_a)^2 (\Delta_a)^2 + (\mu_a)^2 + (\sigma_a)^2 + 2q_a \mu_a \bar{y}_a \tag{4.20}
\]

\[
(\Delta_{a+1})^2 = (q_a)^2 (\Delta_a)^2 + (\bar{y}_{a+1} - q_a \bar{y}_a)^2 + (\sigma_a)^2 + 2q_a (\bar{y}_{a+1} - q_a \bar{y}_a) \bar{y}_a \tag{4.21}
\]
unpacking $\Delta$:

\[
(y_{a+1})^2 + (std(y_{a+1}))^2 = q_a^2 ((\bar{y}_a)^2 + (std(y_a))^2) + (\bar{y}_{a+1})^2 + (q_a\bar{y}_a)^2 - 2 (\bar{y}_{a+1}q_a\bar{y}_a) + \sigma_a^2 + 2q_a\bar{y}_a\bar{y}_{a+1} - 2(q_a)^2(\bar{y}_a)^2
\]  

(4.22)

expressions at both sides of the equation cancel each other and simplify as follows:

\[
(std(y_{a+1}))^2 = q_a^2 (std(y_a))^2 + (\sigma_a)^2
\]  

(4.23)

solving in quadratic equation form:

\[
0 = q_a^2 (std(y_a))^2 + (\sigma_a)^2 - (std(y_{a+1}))^2
\]  

(4.24)

for \( - (\sigma_a)^2 ((\bar{\sigma}_a)^2 - (\sigma_{a+1})^2) > 0 \) and \( (\sigma_a)^2 > 0 \), \( \tilde{q} \) values can be solved as follows:

\[
\tilde{q}_1^a = \sqrt{- (\sigma_a)^2 ((\bar{\sigma}_a)^2 - (\sigma_{a+1})^2)} / (\sigma_a)^2
\]  

(4.25)

\[
\tilde{q}_2^a = -\sqrt{- (\sigma_a)^2 ((\bar{\sigma}_a)^2 - (\sigma_{a+1})^2)} / (\sigma_a)^2
\]  

(4.26)

Following equations are used in the method of GMM:

Using unnormalised unstandardised third moment of the Equation 4.2

\[
E[(y_{a+1})^3] = E[(q_a y_a + \mu_a + \sigma_a \eta)^3]
\]  

(4.27)

Expanding the cube equation

\[
E[(y_{a+1})^3] = \]

\[
E[(q_a y_a)^3 + (\mu_a)^3 + (\sigma_a \eta)^3 + 6(q_a y_a \mu_a \sigma_a \eta) + 3(q_a y_a)^2 \sigma_a + 3(q_a \sigma_a)^2 \mu_a + 3(\sigma_a)^2 q_a \eta] + (4.28)
\]
Using Eq. 4.12, \((\sigma_y a^3)^3, \eta^3\) equals zero

\[
E \left[ (y_{a+1})^3 \right] =
(q_a)^3 E \left[ (y_a)^3 \right] + (\mu_{a})^3 + 3 (q_a)^2 \mu_a E \left[ (y_a)^2 \right] + 3 (\mu_{a})^2 q_a E [y_a] + 3 (\sigma_{a})^2 \mu_a + 3 (\sigma_{a})^2 q_a E [y_a]
\]

\[E \left[ (y_{a+1})^3 \right] =
(q_a)^3 E \left[ (y_a)^3 \right] + (\mu_{a})^3 + 3 \mu_a \left( (q_a)^2 E \left[ (y_a)^2 \right] + (\sigma_{a})^2 \right) + 3 q_a E [y_a] \left( (\mu_{a})^2 + (\sigma_{a})^2 \right)
\]

Expressing \((\sigma_{a})^2\) from Eq. 4.23 in terms of \(q_a\)

\[
E \left[ (y_{a+1})^3 \right] =
(q_a)^3 E \left[ (y_a)^3 \right] + (\mu_{a})^3 + 3 \mu_a \left( (q_a)^2 E \left[ (y_a)^2 \right] + (\mu_{a})^2 + (\sigma_{a})^2 \right)
\]

Replacing \(E [y_a] = \bar{y}_a\) and \(E \left[ (y_a)^2 \right] = (std(y_a))^2 + (\bar{y}_a)^2\) from Eq. 4.9

\[
E \left[ (y_{a+1})^3 \right] =
(q_a)^3 E \left[ (y_a)^3 \right] + (\mu_{a})^3 + 3 \mu_a (q_a)^2 (\bar{y}_a)^2
\]

\[+ 3 \mu_a (std(y_{a+1}))^2 + 3 q_a \bar{y}_a (\mu_{a})^2 + 3 \mu_a \bar{y}_a (std(y_{a+1}))^2 - 3 q_a^3 \bar{y}_a (std(y_a))^2
\]

Expressing \(\mu_{a}\) from Eq. 4.4 in terms of \(q_a\)

\[
E \left[ (y_{a+1})^3 \right] =
(q_a)^3 E \left[ (y_a)^3 \right] + (\bar{y}_{a+1} - q_a \bar{y}_a)^3 + 3 (\bar{y}_{a+1} - q_a \bar{y}_a) (std(y_{a+1}))^2
\]

\[+ 3 q_a \bar{y}_a (\bar{y}_{a+1} - q_a \bar{y}_a)^2 + 3 q_a \bar{y}_a (std(y_{a+1}))^2 - 3 q_a^3 \bar{y}_a (std(y_a))^2
\]
4.7. SUPPLEMENTARY MATERIAL

Expressing in the form of cubic polynomial equation of \( q_a \)

\[
0 = (q_a)^3 \left( E \left[ (y_a)^3 \right] - (\bar{y}_a)^3 - 3\bar{y}_a \left( std(y_a) \right)^2 \right) + (q_a)^2 \left( 3\bar{y}_{a+1} (\bar{y}_a)^2 - 6(\bar{y}_a)^2 \bar{y}_{a+1} \right) + (q_a) \left( 3(\mu_{a+1})^2 \bar{y}_a - 3\bar{y}_a \left( std(y_{a+1}) \right)^2 + 3\bar{y}_a (\bar{y}_{a+1})^2 + 3\bar{y}_a \left( std(y_{a+1}) \right)^2 \right) + (\bar{y}_{a+1})^3 + 3\bar{y}_{a+1} \left( std(y_{a+1}) \right)^2 - E \left[ (y_{a+1})^3 \right]
\]

(4.34)

This equation can be solved for \( q_a \) corresponding each age group. Cardano solution for cubic equations guarantees single real root to exist, the other two complex roots that Cardano solution provides are not used. Both of the \( \sigma_a = std(y_a) \) and GMM estimation techniques can use the following equations for determining the \( \mu_a \) and \( \sigma_a \): For \((q_a)_1\) and \((q_a)_2\) according to Eq. 4.4:

\[
\mu_a = \bar{y}_{a+1} - q_a\bar{y}_a \quad (4.35)
\]

The \( \sigma_a^2 \) can also be expressed in terms of \( q_a \), using Eq. 4.4:

\[
\sigma_a^2 = (\Delta_{a+1})^2 - q_a^2(\Delta_a)^2 - (\bar{y}_{a+1} - q_a\mu)^2 - 2q_a(\bar{y}_{a+1} - q_a\bar{y}_a)\bar{y}_a \quad (4.36)
\]
4.7.4 Supplementary Plots of the USA

Figure 4.10: Population Pyramid for the USA Income between Ages 15-100 reflecting changes of age distribution in 18 years, which reflects mostly an ageing population.
Figure 4.11: $q_\alpha$, $\sigma_\alpha$ and $\mu_\alpha$ Plots for USA Labour Income
Figure 4.12: USA Labour Data Observed and Simulation Statistics
4.7.5 BHPS - JDFs of Age and Income for Observed and Simulated Data (LSM)

Figure 4.13: JDF for 1991-1994
4.7. SUPPLEMENTARY MATERIAL

(a) 1995 JDF of Observed Data.  
(b) 1995 JDF of Sim Data

(c) 1996 JDF of Observed Data.  
(d) 1996 JDF of Sim Data

(e) 1997 JDF of Observed Data.  
(f) 1997 JDF of Sim Data

(g) 1998 JDF of Observed Data.  
(h) 1998 JDF of Sim Data

(i) 1999 JDF of Observed Data.  
(j) 1999 JDF of Sim Data

Figure 4.14: JDF for 1995-1999
Figure 4.15: JDF for 2000-2004

(a) 2000 JDF of Observed Data.
(b) 2000 JDF of Sim Data
(c) 2001 JDF of Observed Data.
(d) 2001 JDF of Sim Data
(e) 2002 JDF of Observed Data.
(f) 2002 JDF of Sim Data
(g) 2003 JDF of Observed Data.
(h) 2003 JDF of Sim Data
(i) 2004 JDF of Observed Data.
(j) 2004 JDF of Sim Data
Figure 4.16: JDF for 2005-2008
Chapter 5

An AI Approach for Portfolio Allocation

The previous chapter introduced a generative model of age and income. In this chapter, a simple agent-based model of pensions is trained with deep reinforcement learning to learn optimal savings and investment strategy selection in Merton’s lifetime portfolio allocation and consumption decision problem setting, depending on age and income trajectories. Here, it is demonstrated that this classical problem can be captured with a deep RL model. In future chapters, it will be developed to include profile heterogeneity and a more complex savings and investment setting with varying liquidity restrictions.

5.1 Introduction

Traditional econometric models for pension savings, such as Merton’s [13], often make simplifying assumptions regarding labour income, investment returns, and liquidity requirements. Nevertheless, subsequent research has revealed the need for a more comprehensive understanding of pension dynamics, accounting for factors such as liquidity constraints [16], labour income fluctuations, and asset return fluctuations [123]. These factors are influenced by complex interactions between businesses, individuals, and central bank decisions, which are focal points for devising effective pension investment strategies.

To address the limitations of traditional econometric models, we propose an
agent-based model (ABM) of the pension ecosystem, leveraging Deep Reinforcement Learning to optimise investment strategies by accounting for the endogenous dynamics of the pension environment [68]. Our proposed model generates synthetic and highly heterogeneous income trajectories, enabling the development of more inclusive saving strategies targeting a broader spectrum of the population. Additionally, our approach learns meta-strategies for contributor agents that are robust to changes in environment dynamics, setting it apart from existing models that make stationary assumptions regarding employment and market dynamics.

Pension funds are expected to invest with a long-term strategic vision, avoiding the effects of financial crises and vulnerability to low probability, high impact black swan events, such as the 2008 financial crisis [38] or the 2022 pensions leveraged gilt crisis [39]. Examples of successful investment strategies, such as the Norwegian Sovereign Wealth fund that invests counter-cyclically with the business cycle, demonstrate the importance of accounting for non-stationary market dynamics in pension models [40].

A recent work approached the modelling of financial systems in a simple game environment that [90] has demonstrated the potential for applying RL in game theory with applications in finance.

This approach has the potential to not only advance the understanding of pension savings optimisation but also to equip investors and policymakers with powerful tools for managing pension investment strategies in a dynamic and competitive environment.

5.2 A Lifetime Portfolio Allocation and Consumption Decision Problem

Merton’s Optimal Portfolio Theorem was developed by Robert C. Merton [13]. It extends the work of Harry Markowitz [36], who introduced the concept of portfolio optimisation using mean-variance analysis. Merton’s approach focuses on continuous-time portfolio optimisation for an investor with a utility function.

The main result of Merton’s Optimal Portfolio Theorem states that an investor
should hold a combination of a risk-free asset and a risky asset (or a portfolio of risky assets) in proportions that maximise the expected utility of their terminal wealth. Merton’s framework also provides formulas to compute the optimal proportions of these assets in the investor’s portfolio.

Given a General Budget Equation for representing wealth:

\[
W(t) = \left[ \sum_{i=1}^{m} w_i(t_0) \frac{X_i(t)}{X_i(t_0)} \right] \cdot [W(t_0) - C(t_0) h] \tag{5.1}
\]

Merton and literature commonly focus on two-asset cases consisting of a risky asset and a riskless asset. The dynamic optimisation problem of Merton’s portfolio optimisation can be expressed as following:

The investor’s problem is to choose a dynamic consumption and investment plan \((c(t), \pi(t))\) that maximises their expected utility, subject to the wealth dynamics:

\[
dW(t) = (rW(t) + \pi(t)(\mu - r)W(t)) dt - c(t) dt + \pi(t)\sigma W(t) dB(t) \tag{5.2}
\]

Here, \(W(t)\) is the investor’s wealth at time \(t\), \(c(t)\) is the consumption at time \(t\), \(\pi(t)\) is the proportion of wealth invested in the risky asset at time \(t\), and \(B(t)\) is a standard Brownian motion.

In Merton’s optimal portfolio problem, the asset dynamics for a single risky asset and a risk-free asset are often modelled using geometric Brownian motion. Let’s denote the price of the risky asset at time \(t\) as \(P(t)\) and the wealth invested in the risky asset as \(X(t)\). The dynamics of the risky asset can be expressed as:

\[
dP(t) = P(t)(\mu dt + \sigma dZ(t)) \tag{5.3}
\]

where \(\mu\) is the expected return, \(\sigma\) is the volatility, and \(dZ(t)\) is the increment of the standard Brownian motion at time \(t\).

The risk-free asset is assumed to have a constant return \(R_f\), and the wealth invested in the risk-free asset is denoted as \(Y(t)\). The dynamics of the risk-free asset
can be expressed as:

\[ dY(t) = Y(t)R_f dt \]  

(5.4)

The total wealth of the investor, \( W(t) \), is the sum of the wealth invested in the risky asset and the risk-free asset:

\[ W(t) = X(t) + Y(t) \]  

(5.5)

Considering the asset dynamics and the wealth equation, Merton’s problem is to find an optimal investment strategy that maximises the expected utility of terminal wealth and consumption over the investment horizon \([0, T]\).

The relationship between \( X(t) \), the wealth invested in the risky asset, and \( P(t) \), the price of the risky asset, can be expressed as:

\[ X(t) = \pi(t)W(t)P(t) \]  

(5.6)

where \( \pi_t \) is the proportion of the investor’s wealth allocated to the risky asset at time \( t \), and \( W(t) \) represents the total wealth of the investor at time \( t \). The investor’s decision is to find the optimal allocation \( \pi_t \) to maximise their expected utility of terminal wealth and consumption over the investment horizon.

The utility is calculated by Constant Relative Risk Aversion (CRRA) [14] function:

\[ u(c_t, \eta) = crra(c, \eta) = \begin{cases} \frac{c^{1-\eta}-1}{1-\eta} & \eta \geq 0, \eta \neq 1 \\ \ln(c) & \eta = 1 \end{cases} \]  

(5.7)

The constraint on consumption and stock allocation:

\[ c_t \geq 0, \quad \pi_t \text{ unrestricted} \]

The objective that we would like to maximise is:

\[ \max E \left[ \int_0^T e^{-\rho s}u(c_s) ds + \epsilon^T e^{-\rho T} u(W_T) \right] \]  

(5.8)
5.2. PORTFOLIO ALLOCATION AND CONSUMPTION

We can express the problem in terms of the value function $V(W, t)$, which is the maximum expected utility that can be achieved from time $t$ onwards, given the current wealth level $W$, where $B(W_T) = e^{rT}u(W_T)$ and $\varepsilon$ parameterise the desired level of bequest, and Merton assumes $0 < \varepsilon << 1$.

\[
V(W, t) = \max_{c, \pi} E \left[ B(W_T) + \int_t^T e^{-\rho(s-t)} U(c_s) ds \right] W_t = W \tag{5.9}
\]

where $U(c)$ is the utility function, $\rho$ is the rate of time preference, and the expectation is taken over the possible paths of wealth $W$ from time $t$ to the investment horizon $T$.

The Hamilton-Jacobi-Bellman (HJB) equation is a partial differential equation (PDE) that characterises the optimal value function of the dynamic optimisation problem is:

\[
\frac{\partial V}{\partial t} + \max_{c(t), \pi(t)} \left\{ (rW + \pi(\mu - r)W - c) \frac{\partial V}{\partial W} \right\} + \left\{ 0.5(\pi \sigma W)^2 \frac{\partial^2 V}{\partial W^2} \right\} = 0 \tag{5.10}
\]

The HJB equation can be solved using the method of characteristics or the method of undetermined coefficients. The solution yields the optimal consumption and investment policies $(c^*_t, \pi^*_t)$ as well as the optimal value function $V^*(W, t)$. Merton derived the optimal proportion of wealth invested in the risky asset $\pi^*$ is constant over time and depends on the investor’s risk aversion, the risk-free rate, the expected return on the risky asset, and the risk associated with the risky asset. This constant proportion is given as [124]:

\[
\pi^* = \frac{(\mu - r)}{(A\sigma^2)} \tag{5.11}
\]

Here, $A$ is the Arrow-Pratt coefficient of absolute risk aversion, which depends on the utility function.

The optimal proportion of wealth invested in the risk-free asset $(1 - \pi^*)$ is given by:
The optimal consumption rate $c^*$ is given by:

$$c^* = \frac{W(1-e^{-\rho T})}{T}$$

(5.13)

Here, $W$ is the investor’s initial wealth, $T$ is the planning horizon, and $\rho$ is the investor’s time preference rate or the rate at which they discount future consumption. The time preference rate is determined by the investor’s utility function and their level of risk aversion.

Merton’s model suggests that an investor should consume a constant proportion of their remaining wealth at each point in time to maximise their lifetime utility. This constant proportion is given by:

$$\frac{c^*}{W} = \frac{(1-e^{-\rho T})}{T}$$

(5.14)

The investor’s optimal consumption plan depends on their initial wealth, time preference rate, and planning horizon. In general, the investor will consume more if they have a higher initial wealth, a higher time preference rate, or a shorter planning horizon. Conversely, they will consume less if they have a lower initial wealth, a lower time preference rate, or a longer planning horizon.

5.3 RL

Reinforcement Learning (RL) can be applied to the portfolio optimisation problem as an alternative to the traditional dynamic optimisation methods such as Merton’s Optimal Portfolio Theorem. In the context of portfolio optimisation, the RL approach involves an agent learning the optimal investment strategy by interacting with the financial market environment.

We define the portfolio optimisation problem as an MDP by specifying the state space, action space, and reward function.

The state space represents the information available to the agent for
decision-making. Let $W_t$ denote the investor’s wealth at time $t$. Let $R_t = (r_1(t), r_2(t), \ldots, r_N(t))$ be the vector of returns for $N$ risky assets at time $t$. We also include any relevant financial information or macroeconomic indicators, represented as $F_t = (f_1(t), f_2(t), \ldots, f_M(t))$. The state at time $t$ can be represented as a vector:

$$s_t = (W_t, R_t, F_t) \quad (5.15)$$

The action space corresponds to the investment decisions made by the agent. In the context of portfolio optimisation with consumption, an action $a_t$ consists of the allocation of the investor’s wealth to the $N$ risky assets, the risk-free asset, and the consumption. The action at time $t$ can be represented as a vector:

$$a_t = (a_1(t), a_2(t), \ldots, a_N(t), c(t)) \quad (5.16)$$

where $a_i(t)$ is the proportion of the investor’s wealth allocated to asset $i$ at time $t$, $0 \leq a_i(t) \leq 1$ for all $i$, and $c(t)$ is the proportion of the investor’s wealth allocated to consumption at time $t$. The constraint $\sum_{i=1}^{N} a_i(t) + c(t) = 1$ ensures that the entire wealth is allocated between the assets and consumption. The action space $A$ can be either discrete or continuous, depending on the desired granularity of the investment decisions and will also have an impact on the decision of the RL training algorithm.

The reward function represents the feedback the agent receives based on its actions. In the context of portfolio optimisation with consumption, we define the reward function as the utility of consumption at time $t$, given by a utility function $U(c(t))$, which is constituted of CRRA utility and final utility from the bequest.

The wealth dynamics can be expressed as:

$$W_t = (W_{t-1} - c(t)W_{t-1}) \left[ r_f(t)(1 - \sum_{i=1}^{N} a_i(t)) + \sum_{i=1}^{N} a_i(t)r_i(t) \right] \quad (5.17)$$

The reward function $R$ is defined as:
The objective of the agent in this MDP formulation is to learn a policy $\pi(s)$ that maps states to actions, such that the expected cumulative reward (i.e. the expected utility of consumption) over the investment horizon is maximised.

In the context of portfolio optimisation, we can apply the PPO algorithm to learn an optimal investment strategy that maximises the expected cumulative reward, i.e. the expected utility of consumption over the investment horizon.

PPO [96] is an actor-critic algorithm that trains two neural networks: an actor network, which represents the policy $\pi(s_t; \theta)$, and a critic network, which estimates the state value function $V(s_t; \omega)$. Here, $\theta$ and $\omega$ denote the parameters of the actor and critic networks, respectively.

The policy network $\pi(s_t; \theta)$ outputs a probability distribution over the action space (portfolio allocations and consumption), and the critic network $V(s_t; \omega)$ estimates the expected cumulative reward given the current state. The PPO algorithm updates the policy and value networks by optimising the following objectives:

For the actor network, the clipped surrogate objective function $L(\theta)$ is optimised:

$$L(\theta) = E_t \left[ \min \left( r_t(\theta)A_t, \text{clip} \left( r_t(\theta), 1 - \varepsilon, 1 + \varepsilon \right)A_t \right) \right]$$

(5.19)

where $p_t(\theta)$ is the probability ratio given by $\frac{\pi_\theta(a_t|s_t)}{\pi_{\theta,\text{old}}(a_t|s_t)}$, $A_t$ is the advantage function, and $\varepsilon$ is a hyperparameter controlling the size of the trust region.

For the critic network, the mean squared error (MSE) loss function $L(\omega)$ is optimised:

$$L(\omega) = E_t \left[ (V(s_t; \omega) - G_t)^2 \right]$$

(5.20)

where $G_t$ is the observed return (cumulative discounted reward) for the state $s_t$.

The advantage function $A_t$ quantifies the relative value of taking action $a_t$ in state $s_t$ compared to the average value of that state. It can be computed as the
difference between the observed return $G_t$ and the estimated state value $V(s_t; \omega)$:

$$A_t = G_t - V(s_t; \omega)$$  \hspace{1cm} (5.21)

To calculate $G_t$, we can use the discounted sum of future rewards:

$$G_t = \sum_{k=0}^{T-t-1} \gamma^k R(s_{t+k}, a_{t+k}, s_{t+k+1})$$  \hspace{1cm} (5.22)

where $T$ is the investment horizon and $\gamma$ is the discount factor.

In the PPO training process for portfolio optimisation, the actor and critic networks are initialised with parameters $\theta$ and $\omega$, respectively, and both networks are updated through gradient-based optimisation.

To express the analogy between classical Merton portfolio optimisation and reinforcement learning (RL) portfolio optimisation, we can compare their respective optimisation objectives and methods.

Merton’s problem is to maximise the expected utility of terminal wealth at time $T$. The investor’s preferences are represented by a utility function $U(W)$.

Where $W(t)$ is the investor’s wealth at time $t$, $c(t)$ is the consumption at time $t$, $\pi(t)$ is the proportion of wealth invested in the risky asset at time $t$, $r$ is the risk-free rate, $\mu$ is the expected return on the risky asset, $\sigma$ is the volatility of the risky asset, and $B(t)$ is a standard Brownian motion.

In the RL portfolio optimisation context, we can formulate the problem as a Markov Decision Process (MDP) with the following components: state space $S$, action space $A$, and reward function $R$.

The objective in the RL portfolio optimisation is to learn a policy $\pi(s)$ that maps states to actions, such that the expected cumulative reward over the investment horizon is maximised, where $\gamma$ is the discount factor that balances the importance of immediate and future rewards.

The analogy between classical Merton portfolio optimisation and RL portfolio optimisation can be summarised as follows:

Both approaches aim to optimise an investor’s portfolio based on the dynamics
of the financial market. Merton’s framework focuses on maximising the expected utility of terminal wealth, while the RL framework focuses on maximising the expected cumulative reward. Merton’s framework relies on continuous-time stochastic calculus and dynamic optimisation, whereas the RL framework uses discrete-time MDPs and reinforcement learning algorithms. Both approaches can incorporate multiple sources of information, such as investor’s wealth and asset returns, in their formulations.

If we consider the direct equivalence between the two problems, in both Merton’s problem and the RL problem, the decision variable is the proportion of wealth allocated to risky assets \( \pi^* \) in Merton’s problem and \( a_t \) in the RL problem. Merton’s solution gives an explicit formula for the optimal allocation, while the RL problem aims to learn a policy that produces similar allocations. The objective of maximising expected utility in Merton’s problem corresponds to maximising expected cumulative reward in the RL problem. In both cases, the goal is to find the best trade-off between consumption and wealth growth. If the RL algorithm can learn the optimal policy, the actions produced by this policy should be close to Merton’s optimal allocation \( \pi^* \).

Reinforcement Learning (RL) offers several benefits when applied to portfolio optimisation compared to traditional methods. Key advantages include adaptability, as RL algorithms can learn and adjust to evolving market conditions, while traditional methods may require manual intervention. The model-free nature of RL allows for effective strategy discovery even with unknown or complex market dynamics.

RL algorithms can handle large state and action spaces, making them suitable for multi-asset optimisation with complex constraints. Deep Reinforcement Learning (DRL) techniques further enhance this capability by leveraging neural networks as function approximators. RL algorithms also balance exploration and exploitation, enabling agents to discover new strategies while maximising returns.

Robustness to noise is another advantage of RL, as it can incorporate techniques such as experience replay or target networks to counteract financial data
noise. RL can also be easily integrated with other machine learning, statistical, or econometric techniques for hybrid strategies.

In Merton’s framework, the optimal investment strategy is derived based on the assumption of geometric Brownian motion for asset prices, which may not fully capture the actual market dynamics. In contrast, RL-based portfolio optimisation does not rely on a specific model of the underlying market dynamics, making it more adaptable to changing market conditions and new information. By incorporating a rich set of state variables, the RL agent can learn the underlying relationships between various financial and market indicators and their effects on asset returns.

The state space in an RL-based portfolio optimisation problem can include various financial and economic indicators, such as asset prices, macroeconomic indicators, financial market data, investor-specific information, and investment constraints. This rich set of information allows the agent to capture the complex relationship between these factors and the asset returns, enabling it to learn a more realistic and rational investment strategy.

Furthermore, the ability of RL algorithms to balance exploration and exploitation throughout the learning process allows the agent to discover new investment strategies while also capitalising on the best-known strategies to maximise returns. This makes RL particularly well-suited to portfolio optimisation problems with large state and action spaces, where traditional optimisation techniques may struggle to find a global optimum.

Deep Reinforcement Learning (DRL) techniques, which combine RL with deep learning, can further enhance the ability to handle high-dimensional state spaces by using neural networks as function approximators. This allows the agent to learn complex non-linear relationships between the state variables and the optimal investment and consumption decisions.

Reinforcement Learning (RL) has proven to be an effective approach for portfolio optimisation due to its flexibility in handling complex state and action spaces and its adaptability to changing market conditions. An additional advantage of RL in this context is its ability to balance various goals, such as maximising an in-
individual’s rational utility, while also considering other objectives like exposure to Environmental, Social, and Governance (ESG) goals or liquidity constraints.

The reward feedback mechanism in RL enables the incorporation of multiple goals into the portfolio optimisation problem. By designing a reward function that reflects these goals, the RL agent learns to make investment decisions that simultaneously satisfy multiple objectives. This allows the agent to strike a balance between different goals without compromising the performance of the investment strategy in general.

By adjusting the Lagrange multipliers, the RL agent can learn an optimal investment strategy that balances the primary objective of maximising expected utility with the secondary objectives related to ESG exposure and liquidity. This approach enables the agent to consider multiple goals in its investment decisions, resulting in a more comprehensive and well-rounded portfolio strategy.

For example, in a portfolio optimisation problem where the primary objective is to maximise the expected utility of terminal wealth and consumption, additional goals such as ESG exposure or liquidity requirements [125] can be introduced as constraints. These constraints can be incorporated into the reward function through a Lagrangian relaxation approach, where a set of Lagrange multipliers is associated with each constraint.

Using DRL models allows for flexible, personalised estimation of lifetime consumption and investment choices, while accounting for heterogeneous income dynamics. This helps in simulating different economic scenarios, providing insights into risks and opportunities, and supporting informed decision-making. Despite challenges such as computational expense and data quality, RL offers a promising approach to portfolio optimisation and financial applications.

During training, the policy and value networks for each agent are updated using stochastic gradient descent with multiple epochs of mini-batch updates. The use of GAE and the PPO clipping objective helps to stabilise the learning process and improve sample efficiency.
5.4 Case Study

A simple case to compare Merton’s model and RL model can be designed based on a stylised financial market with two assets: a risk-free asset (e.g., a government bond) and a risky asset (e.g., a stock index). The objective is to maximise the expected utility of lifetime wealth for each investor agent.

5.4.1 Merton’s Model

In this scenario, each investor agent uses Merton’s optimal portfolio selection model to allocate their wealth among the three assets. The model assumes constant investment opportunities and a stationary environment. The agents’ investment decisions are based on the risk-return characteristics of the assets and their own risk preferences. The performance of each agent’s investment strategy can be compared in terms of utility and the final wealth achieved. The results would show that Merton’s model provides a solid baseline for investment strategies, but its assumption of a stationary environment and constant investment opportunities may lead to suboptimal investment decisions in dynamic markets.

5.4.2 RL Model

Each investor agent uses reinforcement learning algorithms, such as Proximal Policy Optimisation (PPO), to optimise their investment strategies. The RL model allows agents to learn and adapt to changes in the market dynamics by interacting with the environment. The agents are rewarded based on the utility of their wealth, and their investment decisions are updated based on the learned policy. The performance of each agent’s investment strategy can be compared to Merton’s model in terms of utility and the final wealth achieved. The results would demonstrate that the RL model with PPO is capable of adapting to changing market conditions and potentially outperforming Merton’s model in terms of utility and final wealth achieved. However, the RL model may still be limited in capturing the complex interactions between investor agents, businesses, and the central bank.
5.5 Conclusion

In this work, I defined lifetime portfolio allocation and consumption decision problem, and compared the methodologies of Merton, with reinforcement learning of the policy model, which will be expanded to capture more complex dynamics and relationships in the following sections, that can lead to more comprehensive and inclusive policy sets.

**Figure 5.1:** Consumption fraction of wealth decided by learnt policy is in compliance with Merton’s model
Proximal Policy Optimisation (PPO) is used, with a curriculum with three steps, where initially invalid actions such as negative consumption are penalised with a penalty of 87.78, then for the third step, the penalty is decreased to 10. Initialisations during the first step were at random time-step, at later stages simulation is always started at first time-step. The simulation was for $T=10$.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Time Steps</td>
<td>70 million (10M + 20M + 40M)</td>
</tr>
<tr>
<td>Number of Steps per Environment Interaction</td>
<td>128</td>
</tr>
<tr>
<td>Discount Factor (gamma)</td>
<td>0.99</td>
</tr>
<tr>
<td>GAE Parameter (gae._lambda)</td>
<td>0.95, 0.95, 0.99</td>
</tr>
<tr>
<td>Learning Rate</td>
<td>0.00025</td>
</tr>
<tr>
<td>Entropy Coefficient (ent._coef)</td>
<td>0.01</td>
</tr>
<tr>
<td>Clipping Range for Policy (clip_range)</td>
<td>0.2</td>
</tr>
<tr>
<td>Number of Training Epochs per Update (n_epochs)</td>
<td>4</td>
</tr>
<tr>
<td>Clipping Range for Value Function (clip_range_vf)</td>
<td>None (i.e. not used)</td>
</tr>
<tr>
<td>Max Gradient Norm (max_grad_norm)</td>
<td>0.5</td>
</tr>
<tr>
<td>Use State-Dependent Exploration (use_sde)</td>
<td>False</td>
</tr>
<tr>
<td>Value Function Coefficient (vf._coef)</td>
<td>0.5</td>
</tr>
<tr>
<td>Batch Size (batch_size)</td>
<td>4096 (128*32)</td>
</tr>
</tbody>
</table>

Table 5.1: Model Parameters
Chapter 6

Heterogeneous Retirement Savings
Strategy Selection with Reinforcement Learning

In this section, we build on the foundations of the agent-based model with deep reinforcement learning introduced in the previous section by augmenting the model and simulator to include profile heterogeneity and introducing a liquidity and investment setting suitable for pensions. It is successfully demonstrated that the model learns optimal saving and investment strategies according to profile and a complex investment environment with liquidity constraints. The only limit to the capacity of such a model is the comprehensiveness and complexity of the simulator. In the following section, a multi-agent model of different actors in the pension environment is introduced to capture market dynamics as an endogenous feature, marking the next stage in a computationally efficient and flexible methodology.

6.1 Introduction

Retirement financing has been experiencing a clear transition trend from defined benefit (DB) schemes to defined contribution (DC) schemes, as reported by the [6]. DB schemes require scheme sponsors as ultimate guarantors, which can bail out funds in case of a deficit. Employers prefer DC schemes, because the risk and responsibility of managing funds, longevity risk, and market risks are transferred to
contributors in DC schemes. Furthermore, the contribution rates in DC schemes in the UK are on average significantly less, 5.1%, in comparison to DB average contributions of 28.5% [33]. The effects of economic shocks during the accumulation phase are critical, some people were raiding retirement accounts amid COVID-19. Under-pensioned groups [65] faced significant wage shocks, and this also affected their future cumulative wealth and earnings. Exceptional government policies were critical to alleviate the effects of COVID-19 on pension savings and wages but a significant shock with effects to the labour market could not be avoided. It has become apparent how different professions can be affected differently by economic shocks, bringing attention to the role of profile heterogeneity also in the context of pension management. For instance, the rise of the gig economy [66] and irregular workforce participation modes enable more flexible work-life conditions, but introduce larger variations to income trajectories due to the lack of guaranteed income streams.

Previous research has addressed the income distribution and its relationship with age [126], which can be used to quantify the effects of demographic shifts and an ageing population on income. The increasing heterogeneity of career paths and income trajectories requires addressing the questions of how much to save in a more consistent way, as well as how to allocate the savings between spendable liquid investments and non-liquid retirement investments. The foundations of the theories presented in the following section are based on the life cycle hypothesis of saving by [67], which states that individuals aim to maintain a consistent level of consumption throughout their lifetime. In the literature the life-cycle models of income, consumption, and portfolio allocation have been analysed with various perspectives. Samuelson approached lifetime portfolio selection [127] in the context of dynamic stochastic programming in discrete time and solved the multi-period generalisation corresponding to lifetime planning of consumption and investment decisions. Merton formulated the continuous-time version [124] of the same approach for portfolio selection under uncertainty. Later he extended these results [13] to more general utility functions, price behaviour assumptions, and for income generated also from non-capital gains sources. A comprehensive study [68] proposes a
life-cycle model of consumption and portfolio choice, as a temporal portfolio optimisation problem where labour income is assumed to be a risk-free asset, and where the portfolio choice is calibrated with real-world data. [68] present a model where risky income is invested in either risky asset or riskless asset, both are liquid and can be used for consumption, they model the income process explicitly and analytically; they solve the optimal portfolio allocation problem at a given age by numerical solution of their model with backwards induction. A following study by [16] presents a model, which includes an explicit formulation of the income process; it differs from previous research by introducing liquidity friction to risky asset, by charging an excess cost if consumption is financed through the risky asset. The model must be solved numerically, and the solution is described by authors as slow and difficulty, due to three continuous state variables, two continuous control variables, and a fixed transaction cost breaking the concavity of the objective function. The Campanale model assumes that a person has the freedom to switch between liquid and non-liquid asset types, which is not the case with locked pension savings. Campanale et al. use dynamic programming to optimise the [17] preference utility of a household, given specific labour income process consisting of deterministic $G(t)$ of third-order polynomial, and idiosyncratic shock. In the Campanale et al. model, the most important calibration challenge is the transaction cost, which also includes psychological and non-monetary costs.

Further studies focus on liquid and non-liquid retirement savings accounts where liquidity is constrained by introducing cost to liquidate retirement savings [37] and [16]. Previous research fails to address the heterogeneity of contributor profiles and falls short of addressing the idiosyncratic challenges of avoiding consumption crisis during unemployment periods and saving an adequate pension pot for retirement.

Advances in agent-based modelling of complex financial systems, increased computational power, and advances in techniques for optimising agent behaviour in complex environments motivated our investigation. In particular, a deep learning approach for addressing an economic optimisation problem is introduced in the
model called AI Economist by [90]. It uses AI-assisted deep reinforcement learning and implements an agent-based model to address the needs of socioeconomic challenges introduced by designing and testing economic policies, where modules called social planners are trained to discover tax policies in dynamic economies that can effectively trade off economic equality and productivity. A two-level deep reinforcement learning approach is applied to learn dynamic tax policies, based on economic simulations in which both agents and a government learn and adapt.

In this chapter, we introduce a simple model of contributor agents, who decide how much to save and how to allocate the savings, this decision is affected by state variables, specific behavioural parameters and by the information flow in the peer network. Agents decide and optimise their allocation strategy using a deep neural network trained with reinforcement learning. We introduce a simple simulation environment for the agents; which encapsulates employment and income dynamics. Our research bridges a gap between agent-based modelling of the pension system and deep reinforcement learning for finance.

We provide results from agents trained with a state of the art learning methodology and implementing agent-specific optimal behaviour with high granularity for heterogeneous profiles. The model is dynamic, scalable, and can be calibrated to different scenarios. The results show that the balance between near-term consumption safety and retirement savings can be achieved by profile-specific allocation strategies.

RL algorithms are able to learn from data and adapt to changing conditions that can not be expressed with simple mathematical formulations, which means they can be more flexible and responsive to changes and non-linear dynamics. Our model is suitable for tailoring to specific pension fund management goals and constraints. Our model contributes to the development of personalised portfolios, which can factor in profile heterogeneity of age, profession, risk tolerance and financial goals. The model can be trained to mitigate potential risks such as market volatility, labour risk, and changes to geopolitical conditions, as well as sustainability goals. RL algorithm can be trained to identify and mitigate potential risks that are specific to
certain groups of pensioners. All of these can be achieved by incorporating relevant property in to simulation dynamics, and training the same model with the new simulator. Such a model is also adaptive to changes in the market conditions and can be used for dynamic asset allocation strategies.

The recurrent nature of our deep neural network model makes it possible not only to provide good saving and pension investment decisions at any time given profile and current data of the agent, but also it makes possible to capture historical income trajectory via the recurrent embedding, which is a great difference with available models [16], [15] and [124], where the decisions are made by processing current income but not the trajectory. Our recurrent neural network powered policy model can also learn the dynamics of heterogeneous income trajectories, which is a great progress towards more capable decision making of retirement finances.

Our framework is suitable for incorporating extensive behavioural modelling and parameterisation of the agents. It captures the effect of information transmission [50], emphasises consumption sensitivity against negative shocks, as well as covering utility perception [20].

In addition, our model makes it possible for contributors to account for occupation-specific dynamics of life-time income trajectories, which in turn makes it possible to prepare against profile-specific income shocks by allocating savings to cash buffer at the right time frames of their lives.

Our research represents a significant first step to model pension finances in an agent-based model with deep reinforcement learning which permits modelling configurations with increased complexity and realism, in the chapter we presented a simple two-asset version with simple environment dynamics.

In the following sections, in order to evaluate the performance of the proposed deep reinforcement learning model, we conducted a series of simulations using synthetic data. The simulations were designed to mimic the income trajectories of different occupation groups and to test the ability of the model to determine optimal saving and investment strategies for these scenarios.

We first generated synthetic data for a range of occupation groups, including
6.1. INTRODUCTION

low-income, medium-income, and high-income groups. The income trajectories for each group were generated using age-dependent income dynamics, with different growth rates and volatility levels for each group. We also included random shocks to the income streams, such as sudden decreases or increases in income due to economic events or changes in employment status.

Next, we used the proposed deep reinforcement learning model to train agents belonging to different occupation groups. The agents were trained with the objective of maximising long-term wealth while taking into account the age-dependent income dynamics and the income shocks. We incorporated behavioural parameters for each agent, such as risk aversion, shock sensitivity, and individuality factor, in order to make the model more realistic and to capture the different decision-making styles of individuals. This is the first time this has been done for a pension ecosystem.

Once the agents were trained, we ran a series of simulations to evaluate their performance. In each simulation, the agents were given an initial wealth level and were required to make decisions about how much to save and invest in each time period, based on their current income and the expected future income. We measured the performance of each agent by tracking their cumulative wealth over time and comparing it to the optimal wealth that could be achieved given the same income streams, as well as their ability to sustain themselves during unemployment periods.

Overall, our simulations showed that the proposed deep reinforcement learning model was able to accurately capture the profession and age-dependent income dynamics, and that it was able to learn optimal saving and investment strategies for the different occupation groups for the first time. The agents were able to maximise their long-term wealth while taking into account the income volatility, liquidity and the trade-off between immediate consumption and future savings. These results demonstrate the power of the proposed model for tackling the challenges of personalised retirement planning.

Our model is able to account for the unique income profiles and decision-making styles of each individual, rather than focusing on average or typical income
6.2. Model

We introduce a simple model where the agents interact with the simulation environment and optimise their savings behaviour. Dynamics of asset prices are features of the simulation environment, and various dynamics can be used, which provides flexibility. For our simulations, we proceed with simple assumptions of constant return rates for each asset class. Endowment dynamics is not hard-coded into the system, the investment behaviour of our agents at each time step Fig. 6.1, which is governed by a deep recurrent neural network, determines the agent specific endowment dynamics. These neural networks are trained by reward outcomes from interactions of agents with the environment.

During each cycle, agents observe the environment in which they are situated; they choose to allocate their income between consumption, liquid and non-liquid assets.

Each agent has a heterogeneous profile reflecting the occupation and demo-
6.2. MODEL

graphic characteristics; these characteristics are determinants of the unique income and consumption trajectories of each agent. Agents also have characteristic behavioural parameters such as shock sensitivity, consumption utility, and peer-influence factor; which affect the way agents perceive the world and assign value to their stances. In particular, agents are bootstrapped in a social graph which is used for the transmission of information, such as employment status.

Each month, agents receive their income according to their employment. Simulated employment and market dynamics, such as asset return rates, are exogenous and provided by the modeller according to empirical observations. The employment dynamics are dependent on heterogeneous profiles (occupation and demography) and include the new employment of unemployed agents.

The agent first decides how much to save and how much to consume, and secondly the agent allocates the saved amount among a liquid asset and non-liquid asset towards pension savings, each with different return rates. In order to make this financial decision, the agent’s profile, income, behavioural parameters, and peer information observed from the own social network are given as input to a deep policy network.

Deep reinforcement learning and parallel simulation of nearly 30000 agents in 100M timesteps is used for training the deep policy network. The policy network learns an optimal saving and investment strategy for pension savings, avoiding a consumption crisis due to insufficient liquid savings during unemployment.

6.2.1 Optimisation Problem

In the literature optimal consumption, and investment problem has been expressed as a Bellman value function of consumption and assets optimised by dynamic programming [16]. Each agent receives an income governed by simulation’s state transition dynamics $\mathcal{T}$, and makes a consumption and investment decision according to a policy $\pi$, that results in a perceived reward for the agent that can be formulated as:

$$ r_{i,t} = u(c_{i,t}, \eta) + \Delta x - \psi \chi(c_{i,t} - x_{i,t}^{liquid}) - \zeta \chi(m - c_{i,t})(m - c_{i,t}) $$

(6.1)
6.2. MODEL

\( u(c_{i,t}, \eta) \) denotes the utility from consumption and \( \Delta x \) denotes the change of wealth at current time step \( t \), with respect to \( t - 1 \). A penalty of \( \psi \) for not being able to finance current consumption \( c_{i,t} \) with liquid savings \( x_{i,t}^{\text{liquid}} \) is applied by unit step function \( \chi \). Which can also be related with the concept of borrowing constraint in the finance literature, in our case such a constraint would be applied as a Lagrangian relaxation.

Agent is penalised by \( \zeta \) for not being able to consume the minimum consumption amount \( m \), the penalty is proportional to consumption deficit. where constant relative risk aversion function (CRRA) defines the utility from consumption [14] with \( \eta \) as degree of non-linearity:

\[
\begin{align*}
    u(c_{i,t}, \eta) &= \text{crra}(l, \eta) = \\
    &\begin{cases} 
        l^{1-\frac{\eta-1}{1-\eta}} & \eta \geq 0, \eta \neq 1 \\
        \ln(l) & \eta = 1
    \end{cases}
\end{align*}
\]

(6.2)

Reinforcement learning is reliant on feedback from the environment, strict rules need to be communicated mostly via the reward signal, which makes penalisation necessary in some cases. If the agent is unemployed or allocated insufficient funds to fulfil minimum consumption required by the modeller, then the liquid funds are used to finance consumption. If the funds are insufficient, a consumption crisis occurs, which impacts rewards negatively with a consumption crisis penalty. If the agent consumes a lesser percentage then it is required to finance at least minimum consumption amount, then there is an invalid action penalty.

We can further augment the rewards with agent specific parameters, to augment the effects of negative changes. The negative utility difference is augmented with an agent’s shock perception modifier, in order to amplify the negative shocks according to the agent’s behavioural parameter \( \kappa \).

\[
    f(\Delta, \kappa) = \begin{cases} 
        1 & \text{if } \Delta \geq 0 \\
        e^{\kappa} & \text{if } \Delta < 0
    \end{cases}
\]

(6.3)

Which can be used as a function of the reward excluding penalties. The updated
6.2. MODEL

reward can be defined as:

\[ r_{shaped_i}^{1} = f(u(c_{i,t}, \eta) + \Delta x, \kappa) - \psi \chi(c_{i,t} - x_{i,t}^{liquid}) - \zeta \chi(m - c_{i,t})(m - c_{i,t}) \]  

(6.4)

We can shape the reward to incorporate additional relaxed constraints to improve training stability of the neural networks, one such modification can be applied to the penalty of the consumption decision leading to consumption insufficiency, we should penalise the agent only if the current income is exceeding the minimum consumption amount, which means we do not penalise the policy network \( \pi \) for something that it is not in control of, because the simulation \( T \) is in control of the income. The updated formula can be defined as:

\[ r_{shaped}^{2} = u(c_{i,t}, \eta) + \Delta x - \psi \chi(c_{i,t} - x_{i,t}^{liquid}) - \zeta \chi(m - c_{i,t})(m - c_{i,t})(m_e - m) \]  

(6.5)

Agents try to maximise the discounted rewards that they receive during the simulation:

\[ \max_{\theta} E_{a_i \sim \pi(\theta) \sim T} \left[ \sum_{t=0}^{T} \gamma^t r_{i,t} \right] \]  

(6.6)

The goal is to maximise the expectation of the \( \gamma \) discounted reward \( r_{i,t} \) over the entire epoch of \( T \) periods, which denotes entire epoch of \( T \) months. These rewards are determined according to the income that they get, which is determined by the simulation \( T \), and their decisions \( a_{i,t} \) following the policy \( \pi \). The state of the environment is updated according to \( T(s_{t+1}|s_t, a_t) \). Agents maximise their \( \gamma \) discounted expected return for time periods 0 to \( T \), which denotes each month, depending on the agent state \( s_{i,t} \), and the policy parameter \( \theta_i \).

\[ a_{i,t} \sim \pi(\theta) : \pi(a_{i,t}|s_{i,t}, \theta) \]  

(6.7)

Our policy function \( \pi \) is a deep neural network with weight parameters \( \theta \), that gets the agent specific state \( s_{i,t} \) as input.

We are looking to find an optimal parameter \( \theta^* \) for our policy function \( \pi \) that
maximises the expected return of discounted rewards.

\[
\theta^* = \arg\max_{\theta} E_{a_i \sim \pi_\theta, s' \sim \mathcal{F}} \left[ \sum_{t=0}^{T} \gamma^t r_{i,t} \right]
\]  

(6.8)

We calibrated the simulation with Census Data and trained a deep recurrent neural network for policy estimation.

### 6.2.2 Training the Model with RL

Rewards from the environment are used to make the probabilities outputted by the \( \pi(a_{i,t} | \theta) \) policy function more accurate, we accomplish this by back-propagating the gradients of the objective function to optimise the \( \theta \) parameters. Reinforcement learning uses feedback from the environment to optimise the weights of the model towards more accurate estimation, it is achieved by defining an objective function to maximise or a loss function to minimise. In this chapter, we use a policy optimisation technique. In our case there are two networks one is policy network and the other is value network, the value network is used during training of policy network, such an architecture is called Actor Critic Models [93]. The policy network is responsible for selecting actions by generating action probabilities, and value network is used during training to evaluate the goodness of each selected action.

The agents select an action \( a_{i,t} \) according to the policy \( \pi(a_{i,t} | s_{i,t}, \theta) \) at a given state \( s_t \), these actions are saving and portfolio allocation decisions, these decisions can result in changes in the agents’ wealth and current consumption, the environment calculates a reward \( r_{i,t} \) according to chosen reward functions described in Section 2.1. The rewards \( r_{i,t} \) at the end of each time-step are used to calculate the estimated advantages \( \hat{A}_t \) during an entire epoch, these advantages are used to optimise the policy network and the value network \( V_\theta \). The model is trained with Proximal Policy Optimisation method, during the value function is clipped and advantages are normalised, standard Stable Baselines implementation of [97] PPO2 algorithm is used, which is based on OpenAI PPO2 Algorithm [96]:

\[
L^{CLIP}(\theta) = \hat{E}_t \left[ \min(p_t(\theta)\hat{A}_t, clip(p_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t) \right]
\]  

(6.9)
6.2. MODEL

Where the $\theta$ is the policy parameter, $\hat{E}_t$ denotes the empirical expectation, $\epsilon$ is a hyperparameter of the clipped surrogate objective of the actor, and $p_t$ is probability ratio under the new and old actor policies:

$$p_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta,old}(a_t|s_t)} \quad (6.10)$$

Advantage estimations $\hat{A}_t$ are calculated with truncated version of Generalised Advantage Estimation (GAE) [128] for $T$ timesteps, where $V(s_t)$ is value function of the critic, and $r_t$ denotes reward at time-step $t$, and $\gamma$ denotes the discount factor:

$$\hat{A}_t = \delta_t + (\gamma \lambda) \delta_{t+1} + \ldots + (\gamma \lambda)^{T-t} \delta_{T-1} \quad (6.11)$$

$$\delta_t = r_t + \gamma V_{\theta_t}(s_{t+1}) - V_{\theta_t}(s_t) \quad (6.12)$$

where for bootstrapping:

$$V_{\theta_t}(s_{t=0}) = 0 \quad (6.13)$$

The value function of the critic is clipped with the same $\epsilon$ hyperparameter of the actor, to constitute a loss function that is minimised [98] where $V_{\text{target}}$ is the sum of advantage and value:

$$L^V(\theta) = \max \left[ (V_{\theta_t} - V_{\text{target}})^2, (\text{clip}(V_{\theta_t}, V_{\theta_{t-1}} - \epsilon, V_{\theta_{t-1}} + \epsilon) - V_{\text{target}})^2 \right] \quad (6.14)$$

as:

$$V_{\text{target}} = \hat{A}_t + V_{\theta_t}(s_t) \quad (6.15)$$

The composite objective function constitutes the actor’s clipped surrogate objective function, the clipped squared error loss of the critic’s value function and $S$ an entropy bonus as described in [96]:

$$L^{CLIP+V+S}(\theta) = \hat{E}_t \left[ L_t^{CLIP}(\theta) - c_1 L_t^V(\theta) + c_2 S_\pi(\theta) \right] \quad (6.16)$$
where entropy [129] is defined over action probabilities for $n$ actions given a state as:

$$S[\pi](s) = -\sum_{i=1}^{n} \pi(a_i) \log_e \pi(a_i|s)$$ (6.17)

Each epoch is simulated and the advantage estimations are calculated the model is trained with the composite objective function and stochastic gradient updates with Adam optimiser [130].

6.2.3 Agent and Environment Cycle

In order for the simulation to be integrated with existing frameworks, the AEC(Agent Environment Cycle) [99] is followed also to provide a standardised GYM-Like API. The simulations are vectorised and run in parallel. For the purpose of this research, the simulations are conducted in parallel utilising 32 processors, where each processor runs a cohort of more than a thousand agents. For each time step, all of the agents observe and act simultaneously.

Agents observe the environment; these observations include information regarding the market, graph, and agent’s own state including occupation, age, income, and wealth.

The agent action $a_{i,t}$ is shaped by policy $\pi_i$, during learning the reward $r_{i,t}$ for the agent is the sum of total discounted utility and penalty for consumption crisis, which denotes the situation where the agent cannot finance its consumption $c_{i,t}$ governed by consumption dynamic $C$.

The actions are percentage choices between consumption and savings, and investment choices between pension orientated non-liquid funds and liquid funds that can be used at any time to finance consumption, these funds have a vital function especially during the times of unemployment.

Agent behaviour is shaped by influences from peers, individuality factor, consumption utility, and shock response characteristics.

The agent policies are modelled with a deep neural network, which takes as
input agent specific observations and a hidden-state:

\[ a_{i,t} \sim \pi_{\theta} : \pi(a_{i,t} | s_{i,t} = (o_{network}^{i,t}, o_{agent}^{i,t}, o_{market}^{i,t}, h_{i,t}, \theta)) \quad (6.18) \]

The parameter variable \( \theta \) is not agent specific, but common for all contributor agents and the hidden state is updated during action inference of the policy network. Where the state \( s_{i,t} \) constitutes of observations of the agent:

- \( o_{network}^{i,t} \): Observation of the network
- \( o_{agent}^{i,t} \): Observation of own behavioural factors, income, and resources.
- \( o_{market}^{i,t} \): Observation of the market
- \( h_{i,t} \): Hidden state. The updating of hidden state can be interpreted as agents updating their Risk Profile given observations and previous state. And in the future the hidden-state can be used as Risk Profile embedding.

The action space is as follows:

- \( a_{i,t}^{\text{save}} \): Decides to save \( x\% \) (and consuming \( (100 - x)\% \))
- \( a_{i,t}^{\text{liquid}} \): Decide to allocate \( y\% \) to liquid asset \( x\% \) (and allocating the non-liquid asset \( (100 - x)\% \)).

Saving and liquidity percentages are discretised into bins such as \([0, 0.25, 0.5, 0.75, 1]\) in the model.

The full list of variables can be found in the Appendix.

\[ \]

Figure 6.1: Agent and Environment
6.2.4 Deep Policy Network for Optimal Saving, Investment and Liquidity

Agent observations are expressed as a single vector that comprises the concatenation of agent, market and graph vectors. The observation vector is passed through the deep neural network towards the LSTM [131], which updates the agent’s hidden state and outputs a vector for the next layer, which is softmaxed to output a vector representing the action probabilities. A single policy network is trained for all the actions: the action can be as follows “(‘C25’, ‘L75’)”, where “C25” means consume 25% and save 75%; “L75” means allocate 75% of your savings to liquid asset, and 25% of your saving to non-liquid asset.

The hidden states from the model can be thought of as Risk Profile Embedding, which is updated by observations and processing the Agent Profile with the observed environment and shocks via a Deep Neural Architecture that can be found in Fig. 6.2. Reinforcement Learning is used for adjusting the allocation profile according to the Risk Profile Embedding also expressed as a hidden state. At each time step, the agent decides to allocate the income among consumption, savings, and investment classes. This is accomplished by a deep neural network constituted of several layers of a feedforward neural network and a LSTM, which is responsible for acting as the memory of agents. The details of the neural architecture can be...
found in the Appendix.

There is a single action space unifying the choices of consumption and liquidity preference, which means that there are not two different networks for different decisions but one unified network which represents the collection of actions such as “(‘C25’, ‘L75’)”. Setting the reward function for the agents is arguably the trickiest part of the training process, different reward function structures can give spurious and unintended conclusions, which makes the hyper-parameter tuning for the penalties paramount. Failing to tune the penalties results in unintended shortcuts that obstruct the main goal of optimising agent behaviour in an understandable and meaningful way.

After retirement, agents do not act according to their policy networks, but according to the desired retirement pension target such as 80% of labour income being pension income, or receiving a constant pension amount. These time-steps are still used for advantage estimation calculation that spans all of the epochs and for training the hidden-state evolution weighs of the LSTM, which means after training although we do not use the policy output of the LSTM, we do train the hidden state update weighs.

### 6.2.5 Behavioural Parameters of Agents

For modelling behaviour, we base our parameterisation on the approach in [132], where the authors investigated the applicability of the Theoretical Domains Framework outside clinical uses for cross-disciplinary implementation and other research on behaviour change, and provided a simplified version containing 14 domains and 84 component constructs. The Theoretical Domains Framework includes many factors and reports on pension behaviour tend to focus on few factors; for the scope of our research we chose three factors:

- **Consumption Utility**: How do they value current consumption? An agent specific consumption utility multiplier factor

- **Shock Response Characteristics**: How do they respond to the shock? A factor reflecting how sharp agents react to the shock and how drastic they are
• Individuality Factor: How are they being affected by each others beliefs and decisions.

In our simulations each agent has constant risk-aversion parameter $\eta_i$ that is randomly assigned at the beginning, but our model allows the risk-aversion to vary during the simulation and be fed as input to the decision module. Variations of risk-aversion parameter could be used to capture external effects of risk-aversion, which are not captured by simulation captured profile properties such as age, profession, wealth, etc. Some agents are optimistic and underestimate the severity of the shocks, and some agents are pessimistic and overestimate the effect of the shocks. The shock sensitivity factor $\kappa_{i,t}$ is a multiplier of the perceived shock effect, which is normalised for agents of the same occupation. It can be assigned from a normal distribution, can be controlled for experimentation, or fed from empirical report. The agents are affected by the peers and the shocks experienced (if $z_{i,t-1} = 0$). The observations are informative for the closer agents on the graph, and becomes less informative for other agents with weaker connection on the graph. The shocks that affect agents are also weighted with the shock-sensitiveness parameter. In our simulations peer effects are limited to observation of a shock propagating through the peer network, which provides a signal to adjust own behaviour, well before the shock potentially affects the agent, in the presented simulations only the peer effects of income groups of low, mid and high are captured. In this chapter the shocks are not in focus, so the graph structure is simplistic and changes in the income are only governed by age and profession; in more complex simulations we can use peer observation to adjust the agent’s own behaviour well before a shock, whose propagation can be represented on a graph such as disease, automation, or supply chain shocks, potentially reach the agent. The behaviour parameters that are introduced in this section are kept fixed during the entire simulation.
6.3 The Environment

At each time step, the environment operations are executed first. Agent environment operations are executed as follows: first, the market dynamics is executed, which ensures that assets are gaining value according to the calculated interest rates determined by the modeller. Secondly, essential population dynamics are executed such as the ageing of agents, and agents are removed from the system according to the age-specific death probability. The Retirement Process checks if any new agents are required to retire due to age. If an agent retires, their retirement pension is calculated as a rate of their previous consumption at employment, according to the recommended guidelines of the OECD that refers to ideal pension income being 80% of labour income, which is used as initial pension income. An alternative that is investigated is having a constant pension income such as a minimum consumption amount. If an agent is retired, then the agent collects pension from non-liquid pension fund that they contributed during employment life. The agents that are not retired are processed to determine stochastically if they will lose their employment and, if so, for how long they will stay unemployed according to the unemployment duration distribution dependent on the occupation and age. Unemployed agents are assigned new income at their new jobs according to the income distribution depending on occupation and age. These distributions are fed as quantile distribution tables to simulation. The employed agents receive their salary at each month according to their predetermined income.

The agents decide how to allocate their income between consuming and saving, and decide to allocate the saved amount in liquid and riskless assets, or non-liquid and low-risk assets. The decision is shaped by learnt policy and observations, which include the market dynamics, information regarding actions and information from peers, and considering the agent’s own profile. We aim to demonstrate the capability of the model to capture long-horizon decisions such as investing in illiquid pension funds. Our model is flexible to broaden the asset classes to include risky but high returns assets like stocks, but for our demonstration we wanted to focus on the decision of an individual to allocate the income to pension savings that is unreachable
6.3. THE ENVIRONMENT

by the individual till retirement but known to have robust returns due to professional
and diversified management. The other asset that is captured is liquidity which is
known to have only minor return, but is necessary to finance immediate needs such
as periods of unemployment. The focus is not on the optimal asset allocation of a
fund among assets, but on the investment decision of a person into pension funds or
liquidity.

For simulations, we made a narrow assumption based on a very small return
rate to the liquid asset, and a small but larger return rate to the non-liquid asset
which can be assumed as pension fund investment. The model allows agents to
be trained for different asset return rates, but the focus is on profile heterogeneity
and not asset return rates so the training assumed asset return rates fixed with the
parameters reflected on the model card.

6.3.1 The Graph and Synthetic Population

A synthetic representative population is used for the initialising agent population.
Information such as age, income, profession, education level, and other relevant
background information are included. For the purposes of this research, only a
very simple graph is used to demonstrate the capability and flexibility of the model
to augment it with graph structure. This capability makes it possible for further
research to incorporate structures that allow differing communication structures.

We assume the employee network consists of three communities divided by
income level low, medium, and high. The three communities have significant intra-
community interaction but limited intercommunity interaction. The graph choice is
based on the idea that geographical and social networks are also characterised by so-
cioeconomic clusters, and the choice of three communities with income levels is the
simplification of the socioeconomic network. The synthetic database is generated
according to the basic insights from the surveys. Later investigation could incor-
porate survey data to bootstrap the population and investigate geographical graph,
potential social network data, and known network structures to model connections
between agents.

Observation of graphs can be done in several ways; a simplistic way to do this
is modelling information transmission between each agent and its vicinity, i.e. the first and second neighbours, including the transmission of employment information. A more advanced graph observation might be modelled as transmission of not just employment information, but also incorporating additional information such as occupation and the income or consumption data; moreover, the near-neighbour graph can be represented with state-of-the-art graph embedding methodology. $A_{a,b}$ is 1 if there is an edge $a \rightarrow b$ and 0 if there is no edge between two agents of indices $a$ and $b$, $\delta(x,y)$ is the Kronecker Delta $E_{i,t}$ denotes the current earnings, $\iota$ individuality factor. We can formulate a simple information transmission from the immediate vicinity of neighbours and their neighbours as:

$$o_{a,t}^{\text{network}} = \sum_b \left[ \sum_c A_{b,c} \delta(E_{c,t}, 0) \right] + \sum_b A_{a,b} \delta(E_{b,t}, 0) \quad (6.19)$$

Observations from the network are augmented with the agent specific individuality factor. The simplest case is using the individuality factor as a multiplier to the observation:

$$o_{a,t}^{\text{network, perceived}} = o_{a,t}^{\text{network}} \iota_a \quad (6.20)$$

For the purpose of experimentation and investigation of the model, a synthetic but representative population can provide both fidelity and flexibility in a controlled environment. As a design choice for the synthetic population network, we include three clusters, which can be thought of as three neighbourhoods; these neighbourhoods possess nodes with three different income groups: high, medium, and low income. Each node is connected to its own neighbourhood node, and the neighbourhoods are connected to each other with specified weights. Agents are bootstrapped with one of the general occupation groups, occupation-specific incomes, employment status, and ages derived from USA Census Data [73]. Census data are used to generate the synthetic agent population.  

### 6.3.2 Simulation Processes

The simulation is initialised by bootstrapping the agent population and processes. During each time step, the simulation dynamics such as getting income, and getting
employed if vacant, are applied first, then the agent decides to allocate income for consumption or saving, and decides to save by investing in liquid assets, which can be liquidised easily during unemployment, or non-liquid assets which are towards a future retirement, but usually have a better return. Agents are bound by constraints such as the need to consume a minimum amount determined in light of government statistics [74] that determines a minimum consumption per individual.

The occupation-specific income for new employment is determined according to the summary tables from the USA Census Data. The tables reflect the quantile breakdown, and the agents are probabilistically assigned to one of the income quantiles.

The unemployment events and employment processes are explicitly modelled and calibrated with the US Census Data [73]. The probability of unemployment and the duration of unemployment are determined according to the summary tables of the US Census.

Retirement age and retirement income can be accounted for in the system. For the sake of simplicity, initial simulations neglect the retirement period, by only focusing on the contribution period; but the system is later extended to cover the retirement period. Retirement income is defined as a fraction of the last income, fractional retirement income is recommended by international institutions, and this methodology is often also used in the literature [68].

The agent death probabilities are modelled using the Actuarial Life Table [75] in order to make the model comparable with existing models in the literature.

### 6.3.3 Scaling

The agent observations are continuously scaled and standardised, with an online methodology. This is due to the fact that the training dataset is generated continuously during simulation and the distribution of the observed dataset is not known in advance at the start of the simulation, but it can be learnt to an extent after several epochs, and these learnt scales can be utilised in following training and inference as well. The relevant agent variables ("OCC_CODE", "income", "consumption_utility_factor", "shock_sensitivity_factor", "individuality_factor",...
6.4 Results

In this study, we adopted a robust approach to gauge the quality of the model fit within the RL paradigm. During training, the accumulated rewards served as an intrinsic metric to track the agent’s progress. Specifically, a steady uptick in rewards over iterations is a positive indication of the agent mastering its interactions with the environment. Post-training, our evaluation focused on contrasting stylised facts derived from the simulated data with empirical evidence and established literature. Stylised facts refer to characteristic patterns and properties that align with real-world observations. Figures 6.3 and 6.7 are particularly noteworthy, where trajectories depicting wealth, consumption, and labour income with respect to age, as well as the non-liquid asset share concerning total asset amount and age, show a striking resemblance to the findings of [68]. Additionally, aggregate statistics presented in Table 1, such as occupation and age versus the share of non-liquid investments for wealth quartiles, were compared with results from [16] with high Transaction Costs (TC). These comparisons are critical in determining the model’s capability to accurately replicate the inherent dynamics of the real-world system.

We look at longitudinal trajectory plots and strategy breakdowns per total asset size, which provide granular information regarding the differences between occupations. These plots can capture various scenarios such as differences between early career and mid career saving rate strategies among various occupations; which provide more tailored strategies for short-term consumption security and healthy long-term pension finances. 22 parallel initially identical cohorts are simulated for 1000
weeks of agent-time in order to generate resulting tables and plots, which results in 40M agent time-step samples.

6.4.1 Labour, Income, Consumption and Wealth

Fig. 6.3 reflects a similar shape of average simulated income, consumption, and wealth accumulation and decrease over the life cycle compared to [68]. The simulated income trajectory is a reflection of the observed data, which is used for calibration of the environment, and the shape of decrease by retirement age is due to the retirement income being defined as a fraction of the last income, which then gradually decreases. The consumption trajectory during the work-life reflects the saving choices of the population. The agent saves during work-life for financing potential unemployment periods and for retirement finances. The pension income and consumption at the retirement age of 65 converge to the determined retirement income percentage of 80% of latest salary. The data becomes noisy for older ages of 80, which might be due to a significantly smaller sample size.

The rewards of agents during the simulation can be decomposed into two periods; the first period is the labour participation part, where the agent works and gets an income according to income dynamics; in this period the policy inference module \( \pi_\theta \) will make decisions of consumption and portfolio allocation and get a reward as a result of the current and previous actions, these rewards are used for determining the advantages for training the model. The second period is the retirement period, where the agent makes decisions by the pre-determined conditions of the modeller; these pre-determined conditions can be having a constant pension, or having a pension denoting a certain percentage of labour income, the consumption decision is pre-determined, and there is no portfolio allocation decision during retirement; the retirement income and retirement consumption are in other terms hyperparameters or constraints that are given to our model. But during the second period agent still gets a reward, which is used for advantage estimation and also for training the recurrent neural network, where the embedding is still updated and the
6.4. RESULTS

(a) Initial Retirement Income According to OECD as 80% of Labour Income

(b) Retirement Income as Constant Minimum Consumption

Figure 6.3: Wealth, Consumption and Labour Income vs Age Plot

During retirement the pension income is supposed to come from pension savings that have been non-liquid during work-life, but if the pension savings are depleted any liquid savings can be used to finance the retirement income in Fig. 6.4. An interesting outcome of mandating pension income at retirement to be 80% of employment income is comparatively lower consumption during employment,
which might not be desirable; but our optimiser was forced to high saving rates due to 80% mandate, which is stipulated by literature, detailed information can be found in previous sections focusing on literature. One alternative that is investigated is the constant pension income at retirement, where the pensioner gets minimum consumption amount as pension during retirement, which results in much smoother pension savings withdrawal as reflected in Fig. 6.3/b. The results indicate that OECD targets are difficult to reach for significant part of the population.

In Fig. 6.3, we present two contrasting scenarios that depict the consumption patterns of individuals before and after retirement. In Fig. 6.3/a, the model is trained with an initial retirement consumption target set at 80% of the final income earned during employment, following the OECD guidelines. This represents a relatively high consumption aspiration upon retirement. The model simulates conservative consumption behaviour throughout the working years, emphasising saving and investing, in order to meet this substantial retirement target. This is evident from the sharp increase in consumption at the age of 65, which is the transition point from employment to retirement. Conversely, Fig. 6.3/b illustrates a scenario where a more lenient retirement consumption target is set. Here, the target is a constant consumption level slightly above the minimum necessary consumption amount. This lower retirement target leads the model to learn a policy wherein consumption during the employment years is markedly higher since a smaller budget is required to meet the retirement consumption goal. There might be various solutions to this problem that are out of the scope of this chapter, such as easing pension level mandate, or government contributions, or higher returns of investment. The presented results on profile heterogeneity are based on the simulation conducted in parallel to OECD target of labour income’s 80% as pension income.

### 6.4.2 Saving Profiles

The evolution of occupational income in a time frame of nearly 20 years in Fig. 6.5 reflect different characteristics for each occupational group, occupations such as “Sales and Related” and “Transportation and Material Moving” reflect significantly lower mean income with lower variance characteristics; on the contrary, occupa-
6.4. RESULTS

(a) Liquid Asset

(b) Non-Liquid Asset

Figure 6.4: Liquid and non-liquid asset amounts by occupation at age, where only a selection of occupations are depicted on plot for clear visibility. The different characteristics of occupation groups are reflected by plots.

...
6.4. RESULTS

Figure 6.5: Mean of income and unemployment by occupation at week ts; the values are smoothed by 30-weeks moving average and only a selection of occupations are depicted on plot for clear visibility. The different characteristics of occupation groups are reflected by plots.
6.4. RESULTS

The savings profiles in Fig. 6.6 reflect heterogeneous characteristics, where at the same total wealth the saving rate differs greatly, which can be due to different income levels and unemployment risks of occupations. The saving rate plot shows increasing noise at higher wealth levels near 10M, and a much clearer trajectory at lower wealth. An interesting insight is that at the lowest wealth levels, all occupations display similar saving rates. Minimum consumption requirement has a direct consequence of lower saving rates by occupations with low income occupations such as “Farming, Fishing and Forestry”, “Building and Grounds Cleaning and Maintenance”, “Personal Care and Services”, “Food Preparation and Serving Related” occupations have very low saving rate due to their difficulties to finance minimum consumption. Some general patterns can be identified, such as lower income occupations tend to have lower saving rates, but it does not imply that income itself can explain saving decisions; as we can observe varying saving rates among “Healthcare Practitioners”, “Legal Professionals”, and “Business and Financial Operations”.

6.4.3 Portfolio Allocation

The results of our model are in line with existing literature on the relationship between the share of non-liquid assets and age distribution. As shown in Fig. 6.7, our model exhibits similar patterns and rates as those found in other studies.

In particular, our model’s results are comparable to those of [16], who also differentiate non-liquid and liquid assets with transaction costs for switching between them. Furthermore, the similarity is particularly strong when the transaction costs are high.

Additionally, our model’s results on the share of non-liquid asset according to total current wealth also reflect a similar shape of an initial increase followed by a plateau. This concurs with the findings of [16].

The relationship between the share of non-liquid assets and age as inferred from our model is consistent with existing literature, as well as the empirical data presented by [16]. Furthermore, the representation of this relationship in our model is further nuanced, as demonstrated in Fig. 6.8, which reflects a more heterogeneous
6.4. RESULTS

(a) Week

Figure 6.6: Saving rate by occupation at week ts and saving rate by occupation at amount capped at 10M, the values are smoothed by 30-data points moving average and only a selection of occupations are depicted on plot for clear visibility.

(b) Total Asset Amount

The results of this study suggest that consumption and non-liquid investment decisions should not be based solely on total assets at a specific point in time, as is commonly studied in the literature. Instead, our analysis suggests that these decisions should also take into account the unique income trajectories of individuals as determined by their occupation and age. This highlights the importance of incorporating the heterogeneity of individuals and their specific economic conditions into the analysis of consumption and investment decisions. This is reflected in the findings presented in Tables 6.3, 6.5, and the 3D plot on Figure 6.18 that are in the relationship with a greater level of granularity compared to previous literature.
Our model provides a comprehensive representation of income, consumption, and wealth dynamics, as well as portfolio allocation strategies that are suitable for a wide range of heterogeneity and income processes. Furthermore, the level of granularity our model provides is higher than most models in the literature, allowing for a more precise understanding of the investment and consumption decisions made by individuals across different demographic groups.

In summary, the results of our model are in line with existing literature regarding the relationship between the share of non-liquid assets and age. However, our model goes further by providing a more detailed representation of this relationship, which is suitable for a wide range of heterogeneity and income processes. The granularity of our model also allows for a more precise understanding of the invest-
6.4. RESULTS

ment and consumption decisions made by individuals across different demographic groups.

A limitation of the model is the presence of a high level of noise in the portfolio allocation 3d surface depicted in Fig. 6.8, which may be a result of the increased complexity of the model. The empty areas on the plot indicate that individuals with higher total wealth tend to have a higher share of non-liquid assets in their portfolios, which is likely due to the higher potential returns associated with these assets and the fact that wealthy individuals have a greater amount of cash buffers as liquidity to finance their consumption during periods of unemployment. It’s worth to note that, a different model that stipulates higher minimum consumption levels for individuals with higher wealth might lead to some changes in the plot, but the plot is consistent with empirical data and the characteristics of the model.

Figure 6.8: 3d Surface Plot of share of non-liquid assets in x-axis, with respect to total asset wealth in y-axis, and corresponding decision of non-liquid asset investment rate in z-axis, the values are smoothed with 9 weeks moving average for clearer visibility

Contrasting general saving rate and the non-liquid investment rate characteris-
6.4. RESULTS

Figure 6.9: Non-liquid investment rate by occupation at amount capped at 10M, the values are smoothed by 30-data points moving average and only a selection of occupations are depicted on plot for clear visibility.

tics of occupations with respect to total assets results in interesting findings. The non-liquid investment rate by total asset among occupations diverges less than the saving rate by total asset, but still the characteristically differentiating investment strategies are evident in Fig. 6.9. We also observe a noteworthy increase in non-liquid investment rates among production occupations and a stark decrease in farming, fishing, and forestry occupations in relation to total wealth. We believe that these conspicuous shifts, particularly in marginal cases, are likely influenced by outliers present in our dataset. It is conceivable that within lower-income professions such as farming, fishing, and forestry, there exists a small fraction of individuals
who have accrued a significant wealth, standing as outliers within their occupational groups. The model’s interpretation of these outliers can be twofold. First, due to the scarcity of training samples representing high wealth within these occupations, the model may extrapolate and learn policies that seem unexpected or non-intuitive, culminating in the steep decline depicted in Fig. 6.9. Alternatively, the model could be capturing genuine characteristics of these outliers, but the limitations in our dataset render us unable to provide a conclusive explanation. Saving rate by total asset generally increases for all occupations with more assets, with exponential like increase, then it plateaus and slightly varies with noise. Saving rates by highest total asset amounts fluctuate greatly, which might be due to different dynamics governing their decisions such as capital income or behavioural parameters weighing more themselves rather than income being the determinant of the decisions.

Our analysis reveals that the proportion of non-liquid investments in relation to total assets is notably higher for individuals in low-income occupations, with the exception being for those with very high levels of total assets where high-income occupations may surpass low-income occupations in terms of non-liquid investment rate. This disparity can be attributed to the fact that low-income individuals have a greater need for liquidity in order to meet short-term consumption needs during periods of unemployment. This finding can be taken into account by policymakers in formulating policies aimed at mitigating risks faced by low-income workers, such as providing unemployment benefits or increasing early-career pension contributions from the government.

Our research makes a significant advance by focusing on the distinction between risky non-liquid savings, such as endowments to defined contribution pension funds, and riskless liquid savings, which can be used to finance immediate consumption. This approach departs from the previous literature in finance, which only focuses on the dichotomy between risky or riskless assets without liquidity constraints.

The distinction between liquid and non-liquid savings offers a more nuanced
understanding of consumption and saving decisions made by individuals. It also allows for a detailed examination of how factors such as income and occupation influence these decisions and how they might inform policy design aimed at promoting financial stability for all individuals.

In addition, the focus on the difference between liquid and non-liquid savings offers new insights into how investors evaluate the risk-return trade-off. It takes into account that the risks associated with non-liquid assets may be different from those of liquid assets, which is a crucial departure from standard portfolio optimisation.

Furthermore, this research also aligns with the principle of utility maximisation, where individuals make choices that maximise their satisfaction or happiness. The research highlights how individuals from different occupation groups, income levels, and ages differ in their choice of investments and the proportion of liquid vs non-liquid savings. This aligns with the principle that individuals will make choices based on their specific circumstances.

Additionally, in order to account for the potential negative consequences of not being able to finance immediate consumption, our model incorporates penalties for such failures in its analysis. These penalties help to accurately reflect the real-world consequences of not having sufficient liquidity, and are an important aspect of the model’s overall representation of the consumption and saving decisions made by individuals. We also include the parameterisation of negative income shocks and their effect on consumption and investment behaviour. This allows us to account for the impact of unexpected events such as job loss or economic downturns on individual financial situations and behaviours.

A comparison of our model’s results with those of Campanale et al. presented in Table 6.1 under the assumption of high transaction costs illustrates that our model generally results in a higher proportion of non-liquid investments in total asset portfolios, with some exceptions where Campanale et al. identify a similarly high non-liquid asset share. Furthermore, our analysis reflects the substantial variations in non-liquid asset shares in relation to income quartile and age group, which vary significantly across occupation groups.
### Table 6.1: Occupation and Age vs Share of Non-Liquid Investments for Wealth Quartiles

The results from Campanale et al. with high Transaction Costs (TC) are used for comparison, in our model there are no transfers between non-liquid and liquid assets before retirement so high transaction cost results are relatively compatible with our model. There is a margin of error of 0.02
The use of a deep reinforcement learning model allows for a more flexible and personalized approach to estimating lifetime consumption and investment choices. Additionally, the focus on heterogeneous income trajectories allows the model to better reflect the diversity of economic conditions experienced by individuals in different occupation groups and at different stages of their lives. The proposed model generates consumption and retirement saving strategies that account for heterogeneous income dynamics specific to an individual’s occupation and age.

6.4.4 Discussion Regarding the Previous Work and Limitations of the Model

The key contributions of previous work in the field of dynamic optimisation for retirement planning include the development of life-cycle models of income, consumption, and portfolio allocation, as well as the incorporation of behavioural parameters and the consideration of individual differences in decision-making. Previous work has also reflected the importance of incorporating age-dependent income dynamics and the potential for income shocks in order to capture the complexities of real-world income trajectories. Overall, while previous work has made important contributions to the field of retirement planning, there is still a need for more realistic and flexible models that can capture the unique income profiles and decision-making styles of individuals. The proposed deep reinforcement learning model addresses some of these limitations by allowing for the incorporation of individual behavioural parameters and by providing a dynamic optimisation approach that can adapt to changes in income and investment options over time according to their heterogeneous profiles.

The model assumes that agents have access to accurate and up-to-date information about their current income and the expected future income, as well as information about the different investment options and their potential returns. One of the main limitations of the proposed model is that it relies on the availability of accurate and comprehensive data about individual income profiles and investment options. Without access to high-quality data, the model may not be able to accurately capture the unique income profiles and decision-making styles of individuals. Another
limitation of the model is that it assumes that agents are able to make rational and optimal decisions about their saving and investment strategies, which may not always be the case in the real world. The heterogeneity and flexibility of the trained policy models are bounded by the dynamics of the simulation; the models can be as complex and heterogeneous as the simulation itself.

6.5 Conclusion

We modelled a pension ecosystem, where heterogeneous contributors make consumption and investment decisions with Deep RL, which advances available models by providing better granularity and accounting for profile heterogeneity.

We provide a novel methodology to optimise agent behaviour for consumption and investment between pension savings and liquid cash buffer, which is flexible and can be calibrated to work in various scenarios and capture agent heterogeneity. Our model does not need an explicit formulation of the income process and can work with empirical data.

Our research represents a first example of end-to-end modelling of pension ecosystem, it provides a general model to optimise the behaviour for heterogeneous contributors in a dynamic environment. We introduce a single-actor RL model of pension environment, which constitutes a significant step towards multi-actor RL modelling of the pension ecosystem. We successfully devised optimal contributor portfolio allocation strategies between non-liquid pension savings and liquid cash buffer, as well as optimal consumption decision, which can be calibrated with the behavioural parameters of agents. We accomplish this by minimising the consumption crisis periods of agents and maximising retirement savings.

One of the main limitations of previous work is that it has often relied on simplifying assumptions, such as the assumption of a constant risk-free rate of return or the assumption of a constant level of volatility for all individuals. Another limitation is that previous work has often focused on average or typical income trajectories, rather than accounting for the diversity and complexity of individual income profiles. Finally, previous work has often relied on static optimisation tech-
niques, which do not account for the dynamic nature of retirement planning and the potential for changes in income and investment options over time.

One of the key benefits of our deep reinforcement learning model is its ability to simulate different economic scenarios and evaluate the effects on individuals’ saving and investment strategies. This can be useful for policymakers and financial advisers who want to understand how different economic conditions, such as market fluctuations or changes in income levels, can impact individuals’ retirement savings. By simulating these scenarios, our model can provide insights into the potential risks and opportunities that individuals may face, and help them make more informed decisions about how to manage their retirement savings. In addition, our model can be easily adapted to incorporate new data and changes in economic conditions, making it a valuable tool for ongoing analysis and decision making in the field of retirement finance.

The development of models adaptable to diverse policy scenarios, such as varying retirement age regulations and incentive schemes, can require substantial computational resources. Extending these models to address different sets of government policies is a topic left for future research.

Overall, our simulations showed that the deep reinforcement learning model was able to capture the effects of occupation and age on income dynamics, and that it was able to learn optimal saving and investment strategies for the different occupation groups. The agents were able to maximise their long-term wealth while taking into account the income volatility and the trade-off between immediate consumption and future savings. These results demonstrate the value of our model for providing personalised recommendations for individual saving and investment decisions, taking into account the unique income profiles of different occupation groups.

In conclusion, the proposed deep reinforcement learning model is a novel and effective approach for addressing the challenges of retirement planning. By incorporating individual behavioural parameters and using a dynamic optimisation approach, the model is able to capture the unique income profiles and decision-
making styles of individuals, providing more personalised and realistic recommendations for saving and investment decisions. The extensive simulations conducted using synthetic data demonstrated that the model was able to capture the effects of occupation and age on income dynamics and to learn optimal saving and investment strategies for different occupation groups. These results provide strong evidence that the proposed model is able to provide accurate and effective recommendations for individual saving and investment decisions. Overall, the proposed model represents an important contribution to the field of retirement planning and has the potential to provide valuable insights and guidance for individuals looking to plan for their retirement.
### Appendix

#### 6.6.1 Appendix: Model Card

<table>
<thead>
<tr>
<th>Simulation Parameters</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parallel Environment Count</td>
<td>32</td>
</tr>
<tr>
<td>Income Calibration Data</td>
<td>USA CPS 2019 Median weekly earnings</td>
</tr>
<tr>
<td>Unemployment Duration Data</td>
<td>USA CPS 2019 Unemployment duration table</td>
</tr>
<tr>
<td>use_min_max_scaler</td>
<td>1</td>
</tr>
<tr>
<td>time steps</td>
<td>1000</td>
</tr>
<tr>
<td>consumption_crisis_penalty</td>
<td>100000</td>
</tr>
<tr>
<td>invalid_action_penalty_modifier</td>
<td>1000</td>
</tr>
<tr>
<td>retirement_age</td>
<td>65</td>
</tr>
<tr>
<td>retirement_salary_multiplier</td>
<td>0.8</td>
</tr>
<tr>
<td>death_rate</td>
<td>USA SSA Actuarial Life Table</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Agent States</th>
<th>Value</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Market Parameters</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>monthly_market_interest_rate</td>
<td>0</td>
</tr>
<tr>
<td>CPI</td>
<td>0</td>
</tr>
<tr>
<td>monthly_non_liquid_asset_return_rate</td>
<td>0.0125</td>
</tr>
<tr>
<td>monthly_liquid_asset_return_rate</td>
<td>0.0025</td>
</tr>
<tr>
<td>monthly_minimum_consumption</td>
<td>1073 (2021 USA Poverty Guidelines)</td>
</tr>
<tr>
<td>monthly_minimum_wage</td>
<td>1160</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ML Parameters</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>batch_size</td>
<td>14656</td>
</tr>
<tr>
<td>$c_1$</td>
<td>0.5</td>
</tr>
<tr>
<td>$c_2$</td>
<td>0.01</td>
</tr>
<tr>
<td>$e$</td>
<td>$1e^{-5}$</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.99</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>0.95</td>
</tr>
<tr>
<td>$n_{lstm}$</td>
<td>128</td>
</tr>
<tr>
<td>$n_{steps}$</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 6.2: Parameters</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parallel Environment Count</td>
<td>32</td>
</tr>
<tr>
<td>Income Calibration Data</td>
<td>USA CPS 2019 Median weekly earnings</td>
</tr>
<tr>
<td>Unemployment Duration Data</td>
<td>USA CPS 2019 Unemployment duration table</td>
</tr>
<tr>
<td>use_min_max_scaler</td>
<td>1</td>
</tr>
<tr>
<td>time steps</td>
<td>1000</td>
</tr>
<tr>
<td>consumption_crisis_penalty</td>
<td>100000</td>
</tr>
<tr>
<td>invalid_action_penalty_modifier</td>
<td>1000</td>
</tr>
<tr>
<td>retirement_age</td>
<td>65</td>
</tr>
<tr>
<td>retirement_salary_multiplier</td>
<td>0.8</td>
</tr>
<tr>
<td>death_rate</td>
<td>USA SSA Actuarial Life Table</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Agent States</th>
<th>Value</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Market Parameters</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>monthly_market_interest_rate</td>
<td>0</td>
</tr>
<tr>
<td>CPI</td>
<td>0</td>
</tr>
<tr>
<td>monthly_non_liquid_asset_return_rate</td>
<td>0.0125</td>
</tr>
<tr>
<td>monthly_liquid_asset_return_rate</td>
<td>0.0025</td>
</tr>
<tr>
<td>monthly_minimum_consumption</td>
<td>1073 (2021 USA Poverty Guidelines)</td>
</tr>
<tr>
<td>monthly_minimum_wage</td>
<td>1160</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ML Parameters</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>batch_size</td>
<td>14656</td>
</tr>
<tr>
<td>$c_1$</td>
<td>0.5</td>
</tr>
<tr>
<td>$c_2$</td>
<td>0.01</td>
</tr>
<tr>
<td>$e$</td>
<td>$1e^{-5}$</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.99</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>0.95</td>
</tr>
<tr>
<td>$n_{lstm}$</td>
<td>128</td>
</tr>
<tr>
<td>$n_{steps}$</td>
<td>1</td>
</tr>
</tbody>
</table>
6.6.2 Appendix: Neural Architecture

Figure 6.10: Neural Architecture
6.6.3 Appendix: Graph Plot

**Figure 6.11:** Graph Plot where occupations are reflected with colours and income is reflected with the size of nodes, the graph consists of three sub-groups representing three neighbourhoods with three differing income levels as high mid and low income, central nodes of each group are representing the neighbourhoods.
6.6.4 Appendix: Cross-Sectional Analysis

Fig. 6.12: Liquid and Non-Liquid Investment Rate by Occupation at Week ts

Fig. 6.13 reflects the relationship between liquid investment rate and wealth for amounts less than 5M USD. 3 distinctive behaviours are observable, one is Computer and Mathematical Occupations, which start at the lowest liquid investment rate, the other group represents the majority of the occupations representing most low- and mid-income occupations, which start at nearly 35% but then lower their liquid investment rates when the total wealth increases, the third group consist mostly of the high income occupations such as Management Occupations which increase their liquid investment rate with total asset increase until nearly 500K USD, at that point they start to decrease their liquid investment rate.

Fig. 6.14 shows that the increase in liquid assets slows with increasing total wealth, which reflects the fact that the need for security buffer savings decrease and
reward of non-liquid asset is higher. On the contrary, the increase of non-liquid assets with respect to the total wealth increase speeds up at higher amounts and converges to a stable linear trajectory.

The distribution of assets with respect to age in Fig. 6.4 highly differentiates according to the occupation. For example, Management and Legal Occupations have the highest value of assets while Farming, Fisheries and Food Preparation Occupations has the lowest level of assets. Asset differentiation with respect to age depends heavily on the occupation type, some occupations show great variations in the income asset values while other occupations provide minimal savings opportunities, due to the income being merely sufficient to finance consumption.
6.6.5 Appendix: Effects of Behavioural Parameters

In order to further refine the behavioural parameterisation of agents in the proposed deep reinforcement learning model, we introduced three additional factors: consumption utility, individuality, and shock sensitivity. These factors capture additional aspects of individual decision-making styles and allow for personalised and realistic recommendations for saving and investment decisions.

The consumption utility factor captures an individual’s preference for immediate consumption versus future savings. This factor is similar to time preference, but it takes into account not only the individual’s focus on the present or the future, but also their overall utility or enjoyment from consuming goods and services. Individuals with a high consumption utility value are more focused on enjoying the present and tend to prioritise immediate consumption over long-term savings, while individuals with a low consumption utility value are more focused on the future and tend to prioritise long-term savings over immediate consumption.

The individuality factor captures an individual’s willingness to deviate from the average or typical behaviour of their peers. Individuals with a high individuality value are more likely to make unique or unconventional decisions, while individuals with a low individuality value are more likely to conform to the average or typical behaviour of their peers. This factor allows the model to capture the diversity of individual decision-making styles and to account for individuals who may be more likely to take risks or to make unconventional investment decisions.

The shock sensitivity factor captures an individual’s sensitivity to sudden negative changes or shocks to their income. Individuals with a high shock sensitivity value are more likely to be affected by income shocks and may be more conservative in their investment decisions as a result, while individuals with a low shock sensitivity value are less likely to be affected by income shocks and may be more willing to take on risky investments. This factor allows the model to capture the effects of income volatility on individual decision-making and to provide more personalised recommendations for saving and investment decisions in the face of income shocks.

Incorporating these three additional factors into the behavioural parameterisa-
tion of agents allows the proposed deep reinforcement learning model to capture a wider range of individual decision-making styles and to provide personalised and realistic recommendations for saving and investment decisions. This allows the model to better reflect the diversity and complexity of individual preferences and to provide more tailored and effective recommendations for retirement planning.

These factors capture the behaviour of agents, and they impact how agents perceive, understand, and act in their environment. The consumption utility factor is necessary for quantifying how agents value immediate consumption, which can be interpreted as level of consumerism, or temporal preference and eagerness. The shock sensitivity factor is a parameter helpful for capturing the agent’s perception of the consumption change, which can amplify the effects of the changes and force agents to avoid abrupt changes, and an alternative interpretation can be as risk aversion modifier that augments the utility. The individuality factor models the level of influence an agent’s social network exerts on the agent. This is achieved by factoring in the information transmitted from the neighbourhood. The increase in the liquid assets reflects a linear increase; on the contrary, the increase of non-liquid assets is exponential due to the interest income of the assets. The distribution of outcomes reflects heterogeneous characteristics according to behavioural parameters and the relationship between parameters and outcomes are non-linear.
Figure 6.16: Total Asset, Non-Liquid Asset, Liquid Asset at Week $t$ by Consumption Utility Factor
Figure 6.17: 3D Scatter Plot of each indicator relative to the behavioural parameters of the agents, where dark blue indicates lower values and light yellow indicates higher values, which reflect how the parameters are effecting the values such as accumulated assets, investment rates or share of non-liquid assets. The income vs parameters plot is provided for convenience, the income itself is not affected by the behavioural parameters.
Figure 6.18: 3D Surface plot of share of non-liquid assets in x-axis, with respect to total asset wealth in y-axis, and corresponding decision of non-liquid asset investment rate in z-axis, the values are smoothed with 9 weeks moving average for clearer visibility
### Appendix: Tables

<table>
<thead>
<tr>
<th>Occupation Title</th>
<th>Saving Rate</th>
<th>Non-Liquid Investment Rate</th>
<th>Share of Non-Liquid Investments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Architecture and Engineering</td>
<td>0.57</td>
<td>0.60</td>
<td>0.73</td>
</tr>
<tr>
<td>Arts, Design, Entertainment, Sports, and Media</td>
<td>0.42</td>
<td>0.58</td>
<td>0.75</td>
</tr>
<tr>
<td>Building and Grounds Cleaning and Maintenance</td>
<td>0.20</td>
<td>0.57</td>
<td>0.77</td>
</tr>
<tr>
<td>Business and Financial Operations</td>
<td>0.51</td>
<td>0.55</td>
<td>0.71</td>
</tr>
<tr>
<td>Community and Social Service</td>
<td>0.36</td>
<td>0.62</td>
<td>0.77</td>
</tr>
<tr>
<td>Computer and Mathematical</td>
<td>0.59</td>
<td>0.64</td>
<td>0.76</td>
</tr>
<tr>
<td>Construction and Extraction</td>
<td>0.37</td>
<td>0.64</td>
<td>0.78</td>
</tr>
<tr>
<td>Education, Training, and Library</td>
<td>0.40</td>
<td>0.57</td>
<td>0.74</td>
</tr>
<tr>
<td>Farming, Fishing, and Forestry</td>
<td>0.16</td>
<td>0.56</td>
<td>0.84</td>
</tr>
<tr>
<td>Food Preparation and Serving Related</td>
<td>0.21</td>
<td>0.57</td>
<td>0.79</td>
</tr>
<tr>
<td>Healthcare Practitioners and Technical</td>
<td>0.53</td>
<td>0.57</td>
<td>0.73</td>
</tr>
<tr>
<td>Healthcare Support</td>
<td>0.21</td>
<td>0.57</td>
<td>0.77</td>
</tr>
<tr>
<td>Installation, Maintenance, and Repair</td>
<td>0.38</td>
<td>0.63</td>
<td>0.78</td>
</tr>
<tr>
<td>Legal</td>
<td>0.51</td>
<td>0.55</td>
<td>0.70</td>
</tr>
<tr>
<td>Life, Physical, and Social Science</td>
<td>0.47</td>
<td>0.60</td>
<td>0.73</td>
</tr>
<tr>
<td>Management</td>
<td>0.59</td>
<td>0.58</td>
<td>0.72</td>
</tr>
<tr>
<td>Office and Administrative Support</td>
<td>0.27</td>
<td>0.60</td>
<td>0.77</td>
</tr>
<tr>
<td>Personal Care and Service</td>
<td>0.20</td>
<td>0.57</td>
<td>0.77</td>
</tr>
<tr>
<td>Production</td>
<td>0.26</td>
<td>0.60</td>
<td>0.78</td>
</tr>
<tr>
<td>Protective Service</td>
<td>0.39</td>
<td>0.62</td>
<td>0.77</td>
</tr>
<tr>
<td>Sales and Related</td>
<td>0.28</td>
<td>0.57</td>
<td>0.76</td>
</tr>
<tr>
<td>Transportation and Material Moving</td>
<td>0.26</td>
<td>0.61</td>
<td>0.78</td>
</tr>
</tbody>
</table>

**Table 6.3:** Occupation vs Rates, the saving rate denotes the average monthly saving rate of the members of each occupation, the non-liquid investment rate denotes the average of the decided rate of allocating monthly savings to non-liquid investments for each occupation, the share of non-liquid investments denotes the share of non-liquid assets with respect to all of the investments averaged for each occupation.
Table 6.4: Age vs Rates: Consumption rates are defined as consumption amount divided to income. Consumption rates are compared to literature by extracting values from plots of [68], their research differs by our work such that the income values exclude contributions towards pension income, and savings are used as a mean to finance consumption deficit especially during retirement, so during retirement there are positive consumption rates, which mean that the pension deficit is financed by spending savings. This definition difference causes consumption rates to be much higher. There is a margin of error of 0.01.
<table>
<thead>
<tr>
<th>Occupation</th>
<th>20-30</th>
<th>30-40</th>
<th>40-50</th>
<th>50-60</th>
</tr>
</thead>
<tbody>
<tr>
<td>Architecture and Engineering</td>
<td>0.66</td>
<td>0.68</td>
<td>0.65</td>
<td>0.66</td>
</tr>
<tr>
<td>Arts, Design, Entertainment, Sports, and Media</td>
<td>0.46</td>
<td>0.44</td>
<td>0.44</td>
<td>0.48</td>
</tr>
<tr>
<td>Building and Grounds Cleaning and Maintenance</td>
<td>0.23</td>
<td>0.23</td>
<td>0.23</td>
<td>0.23</td>
</tr>
<tr>
<td>Business and Financial Operations</td>
<td>0.55</td>
<td>0.60</td>
<td>0.59</td>
<td>0.59</td>
</tr>
<tr>
<td>Community and Social Service</td>
<td>0.44</td>
<td>0.42</td>
<td>0.43</td>
<td>0.41</td>
</tr>
<tr>
<td>Computer and Mathematical</td>
<td>0.63</td>
<td>0.67</td>
<td>0.67</td>
<td>0.67</td>
</tr>
<tr>
<td>Construction and Extraction</td>
<td>0.41</td>
<td>0.43</td>
<td>0.43</td>
<td>0.42</td>
</tr>
<tr>
<td>Education, Training, and Library</td>
<td>0.45</td>
<td>0.47</td>
<td>0.47</td>
<td>0.43</td>
</tr>
<tr>
<td>Farming, Fishing, and Forestry</td>
<td>0.18</td>
<td>0.18</td>
<td>0.17</td>
<td>0.18</td>
</tr>
<tr>
<td>Food Preparation and Serving Related</td>
<td>0.23</td>
<td>0.23</td>
<td>0.23</td>
<td>0.23</td>
</tr>
<tr>
<td>Healthcare Practitioners and Technical</td>
<td>0.59</td>
<td>0.61</td>
<td>0.59</td>
<td>0.57</td>
</tr>
<tr>
<td>Healthcare Support</td>
<td>0.23</td>
<td>0.23</td>
<td>0.23</td>
<td>0.23</td>
</tr>
<tr>
<td>Installation, Maintenance, and Repair</td>
<td>0.44</td>
<td>0.42</td>
<td>0.43</td>
<td>0.41</td>
</tr>
<tr>
<td>Legal</td>
<td>0.60</td>
<td>0.58</td>
<td>0.59</td>
<td>0.57</td>
</tr>
<tr>
<td>Life, Physical, and Social Science</td>
<td>0.56</td>
<td>0.54</td>
<td>0.55</td>
<td>0.53</td>
</tr>
<tr>
<td>Management</td>
<td>0.68</td>
<td>0.69</td>
<td>0.69</td>
<td>0.68</td>
</tr>
<tr>
<td>Office and Administrative Support</td>
<td>0.31</td>
<td>0.31</td>
<td>0.31</td>
<td>0.31</td>
</tr>
<tr>
<td>Personal Care and Service</td>
<td>0.23</td>
<td>0.23</td>
<td>0.23</td>
<td>0.23</td>
</tr>
<tr>
<td>Production</td>
<td>0.30</td>
<td>0.30</td>
<td>0.31</td>
<td>0.30</td>
</tr>
<tr>
<td>Protective Service</td>
<td>0.42</td>
<td>0.43</td>
<td>0.44</td>
<td>0.42</td>
</tr>
<tr>
<td>Sales and Related</td>
<td>0.33</td>
<td>0.31</td>
<td>0.32</td>
<td>0.31</td>
</tr>
<tr>
<td>Transportation and Material Moving</td>
<td>0.30</td>
<td>0.30</td>
<td>0.31</td>
<td>0.29</td>
</tr>
</tbody>
</table>

*Table 6.5: Saving Rate by Occupation and Age*
6.6.7 Appendix: Raw Plots

Figure 6.19: Mean Income and Unemployment by Occupation at Week ts

(a) Mean Income  (b) Unemployment

Figure 6.20: Saving Rate by Occupation at Week ts and Saving Rate by Occupation at Amount Capped at 10M

(a) Week  (b) Total Asset Amount
Figure 6.21: 3d Surface Plot of share of non-liquid assets in x-axis, with respect to total asset wealth in y-axis, and corresponding decision of non-liquid asset investment rate in z-axis
In previous chapters, we devised and demonstrated a series of models to capture the generative dynamics of age and income, as well as an agent-based model of the pension system. In this model, each individual optimises their saving and investment strategies according to their heterogeneous profiles, considering various liquidity restrictions and asset dynamics appropriate for the pension system. To capture market dynamics as an endogenous feature, we have now developed a multi-agent model of the pension ecosystem involving multiple actors. The endogenous market dynamics result from the interactions among agents, and all actors are collectively trained with Multi-Agent Reinforcement Learning (MARL). The goal is to explore the complex and interactive workings of the actors within pension systems, rather than developing a detailed model of the entire economy.

7.1 Introduction

The problem of pension savings has been extensively researched in economics, with Merton [13] first formulating the problem using an econometric approach that assumed non-insurable labour income, constant income risk and investment returns, and no liquidity requirements. However, subsequent research has revealed the need for a more comprehensive understanding of pension dynamics [68], including the introduction of liquidity constraints [16]. Factors such as labour income fluctuations
and asset return fluctuations [123]. These factors are influenced by interactions between businesses and individuals as well as central bank decisions, which are critical to consider in pension investment strategies. Despite the importance of these factors, econometric models often assume them to be constant or to follow predetermined linear models. An agent-based model (ABM) of the pension ecosystem, using multi-agent reinforcement learning (MARL) to optimise investment strategies, can address the limitations of traditional econometric models by accounting for the endogenous dynamics of the pension environment.

Developing a multi-agent model tailored for pension optimisation, my aim is to navigate the complex dynamics of pension systems instead of creating a comprehensive agent-based model of the whole economy. While examples of broad economic models exist (e.g., by Doyne Farmer’s group in Oxford [84], Andrea Roventini’s works [133], Eurace@Unibi model [88]), my model, not rooted in extensive economic expertise, prioritises developing methods and assembling varied agents to specifically explore and represent the methodology to use deep RL to optimise a multi-agent pension system. This enables analysing agent interactions without being constrained by a hardcoded simulator, facilitating the learning of optimal policies for diverse profiles using calibrated simulators. Traditional models may struggle with cyclicality, heterogeneity, “black swan” events, and non-static environments, hence, this dynamic and adaptive model offers robust scenario analysis and outcome exploration. This approach, foreseeably, will become a crucial toolkit for analysing and optimising financial phenomena, focusing on system interactions among varied environmental actors.

Here I take a step to move from a single agent model where one agent learns optimal actions with respect to environment/simulation, such a simulation is usually calibrated according to observed first or second order statistics, and has strict limitations. In a single agent model, the policy function that is being trained can only be as comprehensive and heterogeneous as the environment dynamics are modelled, which in the case of econometrics the first or second order statistics of the variables of interest, or a simple polynomial for representing a process. I take a step
from a single agent model to multi-agent model, where the environment dynamics manifest themselves as a result of agent interactions between each other, where the policy functions of the agents are being trained. Multi-agent based modelling and training of a model introduce its own methodological and technical complexity and challenges, which are discussed in the following sections.

The proposed model generates synthetic and heterogeneous income trajectories that can be used to devise inclusive savings strategies for a broader population. Furthermore, this research presents a novel contribution by developing a multi-agent model that is robust to changes in environmental dynamics, which distinguishes it from existing econometric models that rely on stationary assumptions about employment and market dynamics. By emphasising the multi-stationary nature of the model, researchers can not only analyse first-order effects but also capture emergent properties arising from the interactions within the system, as described by [82]. This approach allows for better responsiveness to paradigm shifts and black swan events and accounts for the consequences of heterogeneous profiles among pension investors.

Pension funds are meant to invest with a long term strategic vision, to avoid the effect of financial crises and vulnerability to low probability high impact black swan events, as observed in 2008 financial crisis [38] or the 2022 pensions leveraged gilt crisis [39] that effected pension funds that are in general investing with short term vision. Yet there are examples of successful investment strategies such as Norwegian Sovereign Wealth fund that invests counter-cyclically with business cycle [40].

Research on U.S. Social Security data indicates that the labour income has characterising moments that are counter-cyclically exposed to business cycle effects [51]. Cascaded effects of investment decisions [134] and supply chain shock propagation [54] result in non-stationary market dynamics, which violates the premises of general pension models assuming stationary income risk and asset return dynamics.

The proposed ABM of the pension ecosystem addresses these limitations by using a MARL approach to optimise investment strategies. However, implementing
such a model is not without its challenges. Deep Reinforcement Learning has been successful in accomplishing complex tasks [135], but multi-agent deep reinforcement learning is still a computationally expensive solution [8]. The challenges of MARL include non-stationarity of the environment, combinatorial complexity, difficulties arising when a number of agents are greater than 2, and multidimensional learning objectives as stated in a comprehensive review of MARL [22]. Despite these challenges, recent research has shown that carefully trained Proximal Policy Optimisation can perform successfully for optimising MARL problems in cooperative environments [102].

Applying reinforcement learning for game theory with applications in finance was recently explored in AI Economist [90].

Recent advances in software architectures that bridge the gap between mathematical formulations with GPU accelerated Just-in-Time (JIT) executed codes by [136], enable scientists to express the fundamental mathematical operations governing interactions between agents and simulation dynamics as APIs similar to NumPy [137] that are easy to use for mathematical expressions, without need to factor the code for batches, and distributed execution.

A significant part of the reinforcement learning literature evolved around computer games [138], or virtual simulations [94]. Although games provide a flexible environment that can be used to collect large samples, analysis of such games is relatively simplistic in comparison to the modelling needed for interpreting the financial ecosystem. To analyse financial phenomena, a system needs not only to provide a general test bench for MARL, but also to be able to integrate financial and economic indicators, and to benchmark against real-world observations, e.g. for analysing cooperation, competition and behavioural heterogeneity in an environment. Financial systems, given the constraints of their marketplaces, are generally well suited to be captured by well-defined interactions between agents that can be expressed as mathematical equations. Such a system is well suited to leverage accelerators such as GPUs and TPUs. In this way, a high number of samples can be generated in an interactive environment that is energetically and computationally
efficient in comparison to classical CPU multi-processing loads for games.

7.2 Design Choices of Financial Model

Algorithm 1: Simulation Loop Overview

Initialise: BusinessEntity, IndividualEntity, CBEntity, MarketView, and RewardCalculator;
N ← 0;
while N < maxSimulationStepCount do

  with MarketView and RewardCalculator capturing data from entities

    Business Entities: Borrow Choice
    Market Dynamics: Execution
    Business Entities: B2B Trade
    Business and Individual Entities: Employment
    Business Entities: Production
    Business and Individual Entities: B2C Trade
    Individual Entities: Consumption
    Individual Entities: Investment Choice
    Individual Entities: Investment Choice
    Rewards: Feedback
    Training: Step

  end

N ← N + 1;
end

7.2.1 Actors of Ecosystem and Interactions

The multi-agent reinforcement learning model presented in this research simulates a pension ecosystem consisting of two main actors: Individuals and Businesses. Businesses trade with each other and produce goods (inventory) using a Sectoral IO matrix and a chosen production function. Businesses also engage in trade with individuals for their labour.

Individuals, as the primary contributors to the pension system, make investment decisions that affect their pension savings. They also consume inventory produced by the businesses. The model simulates the impact of endogenous and exogenous shocks, business cycles, and policy decisions on the behaviour of these individuals. Additionally, the model generates synthetic and highly diverse income trajectories to provide a more realistic representation of the population, which can be used to develop more inclusive savings strategies.
In terms of investment decisions, the model assumes that asset returns are correlated with the estimated fundamental values of companies, which are estimated based on the market trading value of their inventory. The model also uses meta-strategies for contributor agents that are robust to changes in the environment dynamics, as opposed to traditional econometric models that assume stationary employment and market dynamics.

The proposed model in this research utilises a simulation loop to simulate the interactions between the different actors in the pension ecosystem. The actors in our ecosystem are the Individuals and Businesses. The simulation begins by initialising various entities that represent these actors, including the Business Entity, Individual Entities, CB Entities, Market View, and Reward Calculator.

Inside the simulation loop 1, several operations are performed in a specific order. The Market View and Reward Calculator capture data from the entities, allowing them to calculate the price statistics for actors to make informed decisions about the market and rewards that are being used for reinforcement learning. Market dynamics are then executed, simulating the interactions between the different actors in the ecosystem.

The Business Entities engage in Business-to-Business (B2B) trade and both Businesses and Individuals engage in employment. Business Entities also produce goods (inventory) and both Businesses and Individuals engage in Business-to-Consumer (B2C) trade. Individual Entities then consume the inventory and make investment choices. The Reward Calculator provides feedback and the training step is executed. The variable N is then incremented by 1 and the loop continues until the maximum number of simulation steps is reached.

The algorithm reflects the order of operations in the simulation, which is designed to simulate the interactions between the different actors in the pension ecosystem and how they affect the investment and saving behaviours of individuals. It is a key aspect of the proposed model, as it allows for the simulation of various scenarios and the impact of different factors on the ecosystem as a whole.

The choices of the agents in the pension ecosystem make are crucial to its
7.2. DESIGN CHOICES OF FINANCIAL MODEL

Figure 7.1: Agents and Environment Diagram

functioning and are as follows:

<table>
<thead>
<tr>
<th>Entity</th>
<th>Key Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business</td>
<td>Borrowing choices</td>
</tr>
<tr>
<td></td>
<td>Trading activities</td>
</tr>
<tr>
<td></td>
<td>Production capacity utilization</td>
</tr>
<tr>
<td>Individual</td>
<td>Employment</td>
</tr>
<tr>
<td></td>
<td>B2C trade activities</td>
</tr>
<tr>
<td></td>
<td>Saving Decision</td>
</tr>
<tr>
<td></td>
<td>Investment and portfolio allocation</td>
</tr>
</tbody>
</table>

Table 7.1: Key decisions for each entity in the model

Table 7.1 presents the key decision-making factors for various entities in the financial ecosystem. These factors play a crucial role in shaping the dynamics of the pension ecosystem and should be considered when modelling agent interactions in our multi-agent reinforcement learning framework.

It is important to note that the choices made by each agent have a direct impact on the overall functioning of the pension ecosystem. Businesses must balance production and debt management, while individuals must consider their consumption, savings, and investment decisions. Prices and market dynamics manifest from the interactions between agents.

7.2.2 Mechanism of Interactions

There are three types of operations, firstly an interaction of a trade nature where two parties actively participate by making an offer and the other party taking decisions based on the offer. Secondly, there are choice operations where an agent
makes a choice by itself, on issues such as setting interest rates, deciding how much percentage to utilise for production capacity, or an individual making an investment decision. Thirdly the simulation dynamics such as charging the owed interest per month to businesses, where no active decision is made but operations relevant to simulation dynamics are executed.

Trade operations consist of two components: an Offer and a Decision. In this example, a deep neural network is utilised to take in an agent’s own embedding, along with market data, and output a vector that specifies the inventory being offered and the requested amount of cash. Although the trade module can also handle barter transactions, cash-only transactions are preferred for simplicity as they allow for easy market view coupling. This means that the average prices of inventory transactions can be used to update the market table, which acts as a signal for businesses to guide their decisions and shape rewards for RL. The Decision network uses the embedding of the decision-making party, the offer, and the market table as input to make a decision on whether to accept the offer. Following policy inference model of [90], agents share the trained policy inference parameter $\theta$ and each agent has their own state $h_{i,t}$, and state is updated with each policy inference:

$$a_{i,t}^{offer} \sim \pi(o_{agent}^{agent}, o_{market}^{market}; \theta)$$ (7.1)

$$a_{i,t}^{decision} \sim \pi(o_{agent}^{agent}, o_{offer}^{offer}, o_{market}^{market}; \theta)$$ (7.2)

Choice operations (ii) get the relevant agent’s embedding and market information to make a decision via deep neural network.

$$a_{i,t}^{choice} \sim \pi(o_{agent}^{agent}, o_{market}^{market}, h_{i,t}; \theta)$$ (7.3)

A detailed overview of which agencies each entity has can be found in Fig. 7.3
7.2. DESIGN CHOICES OF FINANCIAL MODEL

Figure 7.2: The Trade Module includes Offer and Decision Models that handle the inference on trade operations. The Offer Model creates an Offer Vector, which is made up of two parts: first half represents the inventory that the offering party will provide, and the second half represents the inventory that the accepting party will provide. The Decision Model then produces a decision rate value, indicating whether or not to proceed with the trade, with 0 meaning no trade and 1 meaning full acceptance of the trade offer.

Figure 7.3: There are two kinds of entities Person and Business, there are thousands of instances of these entities. Each entity has multiple agents, these agents are responsible for specific tasks and instances of entities learn shared neural network weights per agent. For a trade agent there are two neural networks one for the offering party one for deciding party, for a choice agent there is only one set of neural network. So in this research there are 10 different neural networks that are being trained.

7.2.3 Alignment with Finance Community

To make the simulation results more accessible to the broader financial community, we have taken steps to align our model with familiar financial terminology and concepts found in lifetime consumption and portfolio selection literature.
Our choices include the use of utility functions, which are widely studied in finance and help to express trade-offs between immediate and future returns in terms of intertemporal preferences [13].

The model’s calibration is grounded in observable phenomena through the use of the input-output technology matrix as the only hard-encoded market dynamic. This approach facilitates comparisons between various countries with differing input-output technology matrices.

Regarding asset endowment dynamics, two configurations are explored. The first is cash and risky vs riskless asset dynamic [68], which can be calibrated to have the same mean and variance as other portfolio allocation models in the literature:

$$R_{t+1} - R_f = \mu_{t+1} + \sigma_{t+1}$$

(7.4)

where $R_f$ is the risk-free return rate, and the risky asset return $R_{t+1}$ is defined by $R_f$ and excess returns characterised by $\mu_{t+1}$ and $\sigma_{t+1}$. Another configuration that differentiates from the literature that can provide analysis of the higher-order effects would be coupling asset returns to the revenues or valuation of the companies being simulated in the model, such an approach can provide new insights.

For investigating different countries, the simulations must be bootstrapped with populations reflecting the respective sectoral distributions of countries.

The parallelism between the discounted rewards used in reinforcement learning [96] and time-discounting in the multi-step portfolio optimisation makes this methodology suitable to communicate the financial interpretation of an agent-based model. Time discounting can be classically found in financial literature incorporated into utility functions such as CRRA preferences [13] or the Bellman optimisation equations [68].

Another potential improvement involves incorporating heterogeneous employee profiles by breaking down labour into various professional groups, each with their own pricing. This heterogeneity can be integrated into the input-output production matrix, as seen in [139]. Alternatively, differentiated labour values can reflect the talent distribution of the population, which may result in differing incomes for
individuals with varying talent, education, or experience levels [55].

### 7.2.4 Leontief Production Function

Table 7.2 reflects the Input-Output Matrix of Sectors excluding \textit{inv.cash}. The matrix represents the inter-dependencies among various sectors. For instance, values in the first row illustrate the dependency of the \textit{inv.farm} sector on inputs from other sectors. Leontief production function is used for production and consumption operations, which is an expression of the principle of minimums:

\[
q = \min \left( \frac{z_1}{a}, \frac{z_2}{b}, \frac{z_3}{c}, \frac{z_4}{d}, \frac{z_5}{e} \right)
\]  

(7.5)

The consumption rates for different sectors can be observed in Table 7.3 in the Appendix. Leontief production function is applied for the consumption as well, where the consumption bucket follows the law of minimums.

### 7.3 Architecture of Simulation

JAX [105] provides a unified framework where both the training and inference of deep neural networks, as well as the execution of simulation operations, can be done by transforming simple expressions into vectorised operations that can utilise hardware accelerators. Furthermore, the auto-differentiation property provides flexible functionality for training deep neural network models. The JAX ecosystem provides Flax [106] as a Deep Neural Networks library. Such a flexible framework ensures the flexibility to build a framework that can work with variable inputs. The input flexibility is important to scale and sophisticate the models with minor interventions to the code and underlying structure, and to simplify the process of feeding the system with more granular and richer data sets.

Developing an efficient system that leverages low-level hardware functionalities and accurately models a financial system or any other complex system is a challenging task. Often, researchers or companies must prioritise one aspect due to increased complexity, resource, and time constraints. A framework and methodology that allows for rigorous modelling of a financial ecosystem without the need to
worry about efficient code execution at every step of development enables resources to focus on the model, optimisation algorithms, and sensitivity analysis of the ABM results.

This research introduces a framework and methodology to test financial agent-based models and simulations that can be optimised with various optimisation techniques including deep RL in a structured way, where the dynamics of the system are coded for one agent as simple python functions, and the execution at the backend will be JIT and hardware accelerated by GPUs, TPUs and CPUs. Another advantage is decoupling the development of code responsible for parallelism, from the code responsible for the simulation dynamics; this helps the modellers to deal with the complexity of modelling financial systems, parallelising it for efficient execution via hardware accelerators, and developing complex optimisation algorithms.

One way to simultaneously take advantage of accelerated JIT execution of RL and financial models in our ABM is expressing them on a single computational graph, where simulation operations, reward calculation and optimisation algorithms run on a single network. The advantages of a single computational graph come with a great cost of high development complexity, error-proneness and not being flexible for adapting to various reward mixing methodologies. In contrast, a modular and decoupled methodology for the execution of various operations provides advantages, such as flexibility by reward mixing and ease of development of simulation and training environment. This can be accomplished without jeopardising the computational efficiency of accelerated JIT under a single framework.

![Figure 7.4: Single Computational Graph](image-url)
7.3. ARCHITECTURE OF SIMULATION

- **SimDirector**: SimDirector module is responsible for calling the operations regarding the simulation at the highest level, rewards computation and training. It is the module where the system is orchestrated and where, with slight modifications, different simulation and training flows can be achieved, such as accessing the intermediate states of the simulation for reward mixing purposes, or changing the order of the simulation operations.

- **Reward Calculator**: RewardCalculator is a buffer module where the implementation of desired complex reward mixing, storage of state transitions and rewards from various episodes, as well as parallel running simulation instances can be achieved.

- **Entity**: The Entity module is the wrapper around the stack module and agent modules. Each entity that has multiple underlying units, must have a Stack module attached. Entities can have multiple agents, regarding different areas of interaction.

- **Stack**: The Stack module stores the matrix for data of multiple units, such as a collection of businesses, but also provides an API to access units as single objects. It is the module that provides duality of hardware accelerated matrix backend data type and a familiar Object Oriented Programming like API.

- **Agent**: Agent is the module that encapsulates the underlying Model module responsible for ML inference and training at low-level, and Operation Executor module is responsible for executing the environment actions/dynamics for updating the environment state, once the ML module made a decision. Agent module governs the loop for dividing, matching and shuffling of units in operations which require the interaction of the same kind of entities, such as businesses trading with each other.

- **Model**: It is the module where underlying neural networks are implemented and called for forward inference and backward backpropagation operations on neural networks, as well as batch normalisation.
7.4 Training

Multi-agent systems face stationarity challenges compared to single-agent reinforcement learning. The dynamics agents learn actively change during training [22]. Various methodologies address this issue, with PPO, which shows strong empirical performance, being chosen for our multi-agent pension model [102].

---

**Figure 7.5: Combined Figures**

- **Operation Executor**: the Operation Executor module is responsible for the execution of JIT functions defined in Operation module for batches of randomly chosen units or indexed batches, by applying the operations on Stack module.

- **Operation**: the Operation module is where simulation dynamics are implemented as stateless Python functions conforming to the stateless function format of JAX that is required for vectorisation and JIT. The functions are implemented without explicitly implementing by factoring in the batch dimensions or parallel execution.

- **View**: the View module is where the calculated statistics information from the underlying entities such as the latest prices of assets being traded in the simulation or by agents designated information such as interest rate can be stored.

- **Logger**: the Logger module is responsible for calculating relevant statistics that can be outputted as plots or used for rewards.
PPO [140] was preferred over pure MADDPG [141] due to its on-policy nature, simplifying learning by updating policies based on recent experiences, enhancing sample efficiency. Real-world economic agents typically lack global information access, relying on local observations. Accurate financial models should consider this limited information availability.

MADDPG, while effective in some settings, might not suit financial systems modelling due to its assumption of full state and action space access during training. This could lead to over-optimisation and unrealistic models. PPO, however, optimises local policies based on an agent’s experiences, offering a more realistic representation of market dynamics.

Considering real-world limitations and the importance of local information, PPO is a more suitable choice for financial system modelling. Agents can be trained as meta-learners, acting optimally in different dynamics. Possible approaches include using a latent recurrent state to capture historical experiences or employing a meta-learner methodology [100].

For the actor network, the clipped surrogate objective function $L(\theta)$ is optimised:

$$L(\theta) = E_t \left[ \min \left( r_t(\theta) A_t, \text{clip} \left( r_t(\theta), 1 - \epsilon, 1 + \epsilon \right) A_t \right) \right] \quad (7.6)$$

where $r_t(\theta)$ is the probability ratio given by $\frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)}$, $A_t$ is the advantage function, and $\epsilon$ is a hyperparameter controlling the size of the trust region.

For the critic network, the mean squared error (MSE) loss function $L(\omega)$ is optimised:

$$L(\omega) = E_t \left[ (V(s_t; \omega) - G_t)^2 \right] \quad (7.7)$$

where $G_t$ is the observed return (cumulative discounted reward) for the state $s_t$. The advantage function $A_t$ quantifies the relative value of taking action $a_t$ in state $s_t$ compared to the average value of that state. It can be computed as the difference between the observed return $G_t$ and the estimated state value $V(s_t; \omega)$.
7.4. TRAINING

\[ A_t = G_t - V(s_t; \omega) \]  

(7.8)

To calculate \( G_t \), we can use the discounted sum of future rewards:

\[ G_t = \sum_{k=0}^{T-t-1} \gamma^k R(s_{t+k}, a_{t+k}, s_{t+k+1}) \]  

(7.9)

Training multiple models each governing different aspects for agents such as borrowing decisions, investment decisions and trade decisions is challenging, one way to deal with the complexity is curriculum learning, where the different functionalities of the agents are being enabled and trained incrementally, which can also help to tackle with the training instability problem of reward attribution from the outcomes of multiple machine learning modules of the agents in a temporal setting. Curriculum learning is proven to be beneficial for developing increasingly complex capabilities [95] can greatly speed up and improve the success of the training. Entities such as “Business” and “Person” have multiple agents learning a policy function for the task at hand, one aspect is attributing reward of the actions, which in our case can be modelled relatively more simplistically and transparently due to the nature of financial systems, where we have values for all assets and utility differences can be used for constructing the advantages. This still doesn’t change the fact that multiple choices are being made at the same step and their consequences are propagating to multiple time steps which introduces not only multi-agent reward attribution of the same entity or among two kinds of entities but also intertemporal reward attribution challenges. One example was the consumption of people being insufficient not because of the choice of consuming less percentage, but because the B2C trade decisions by businesses and people end up the people with less then desirable inventory levels. Such challenges can be addressed with reward shaping by providing a residual serial flow of rewards between agents as seen in Fig. 7.6, or by using different training algorithms that can be augmented where it allows a mixture of value functions of individual agents, to be trained collectively [142, 143]. RL is sample hungry, Deep RL is sample hungrier, and the Deep MARL is the
hungriest, because the multiple steps with 10s of thousands of agents interacting with each other constitute a single training epoch, and the more sample efficient derivatives of the algorithms are not always available, because after each training epoch the environment dynamics change due to different kinds of agents changing behaviour, this is the non-stationarity problem. One potential solution might be innovative scheduling ideas if training a single agent by sustaining the policy functions of other agents constant and utilising value based methods with Replay Buffer and importance sampling, which can potentially be more sample efficient.

Figure 7.6: Residual Rewards Propagating

7.5 Simulation Results

7.5.1 Simulation Configuration and Initialisation

In the simulation presented in the following section the interest rates are assumed to be constant, and the business and people populations are randomly initialised.

The hyperparameters and configuration parameters employed for the simulation and neural networks for reinforcement learning are stated in Appendix on Table 7.4. These parameters play a critical role in shaping the behaviour and results of the machine learning models integrated into the simulation.

7.5.2 Analysis of Loss Functions

Analysis of loss functions indicates that different agents can have various learning patterns, even if they belong to the same type of agent type such as choice and trade agents. The training of nearly 35000 training epochs, where each epoch has 10 steps, results in a steady but slow improvement. This training run took a couple of days on a PC with 16 cores and 3080TI as GPU.
7.5.3 Consumer Behaviour

7.5.3.1 Consumption Patterns

The consumption plot in Fig. 7.7 reflects a surge in consumption at the beginning and a decrease in consumption at the following time steps. This decrease can be explained by the lack of inventory in terms of the law of minimum, where the people lack some of the necessary inventories to consume the consumption bucket. Which means this behaviour is not characterised by the deficiency in consumption policy function, but signals a potential problem covering the B2C trade.

![Figure 7.7: Consumption patterns of consumers over time.](image)

7.5.3.2 Asset Distribution

Asset distribution plot in Fig. 7.8 reflects a population where people invest heavily in risky asset with higher return rates and cash with high liquidity to consume, and the riskless asset is not preferred as getting richer.

7.5.3.3 Risk Behaviour (Risky vs Riskless vs Cash)

The plot reflecting the assets of the people in Fig. 7.9 reflects two things, one is that our top coding of thresholding and limiting maximum asset by a million is the limiting factor, and the second is the riskless asset reflects asymptotic trajectory despite not being limited by the top threshold.
7.5. SIMULATION RESULTS

7.5.3.4 GDP Proxy Through Business Inventory

Fig. 7.10 reflects mean business inventory value as a proxy of the GDP, this plot reflects that after initial fluctuations the system stabilises in general.

Figure 7.8: Asset Distribution

Figure 7.9: Analysis of Risk Behaviour among Consumers

Figure 7.10: A proxy measure of GDP using business inventory data.
7.5. SIMULATION RESULTS

7.5.3.5 Inflation Proxy Through People’s Consumption Bucket

Inflation can be tracked by looking into the prices of people’s consumption bucket in Fig. 7.11 where the prices fairly decrease and relatively stabilise with a downwards trajectory, one aspect that shouldn’t be forgotten is the consumption is insufficient for the people, and people have surplus assets, it is expected the people to demand more consumption goods(inventory) and cause an increase of the prices.

![Average Consumption Bucket Price](image)

**Figure 7.11:** A proxy measure of inflation using prices of consumption bucket.

7.5.4 Business Dynamics

The plot of prices in Fig. 7.12 for 1000 timesteps reflects the dynamics of the market. The simulation is initialised with random initialisations of business and people populations. The prices are manifested as a result of bilateral trade relationships between entities in the simulation. Three kinds of prices are there prices that consist as a result of B2B trade cycles at each time step, B2C trade prices, and price of labour employment. The prices at the beginning fluctuate greatly, then they get stabilised and reflect minor cyclicalities with fairly stable trajectories. Even some rare but relatively extreme fluctuations after the initial period tend to stabilise which signifies that this is a robust system. One point to address is why the prices are not according to the consumption bucket’s needs. For example in B2C, why the unneeded assets are highly priced, with high total trade value? One possible answer might be the fact that the customers do not learn meaningful behaviour for these irrelevant inventories, causing utility maximisation by maximising the inventories even if it is not highly relevant to consumption. One another explanation might
be the trade cycle scheduling calibration, which indicates that there might not be enough chances for the trades to happen.

Figure 7.12: Time series of B2B, B2C, and Labour Prices

### 7.5.5 Socio-Economic Indicators and Gini Index Analysis

Social equality can be tracked by various measures; one such indicator is the Gini index; a sample of the population is tracked during the entire simulation timesteps, to estimate the Gini index, which reflects characteristics of starting at a higher in-
equality reflecting value and gradually fluctuating and converging around a lower Gini index value.

![Figure 7.13: Time series of Gini index indicating wealth distribution](image)

### 7.6 Challenges

#### 7.6.1 Rewards

Rewards for the trained agents should account for both intertemporal reward attribution and reward attribution in the presence of multiple modules determining actions. Reinforcement learning relies on rewards as training signals. The rewards for each module being optimised can be expressed as a combination of rewards from various action steps, from the agent’s perspective as well as broader statistics, such as overall output, where cooperative action can be rewarded. Social concerns like income inequality have been shown to serve as a component of system-wide rewards in the literature [90]. As a distinguishing feature compared to utility calculation, this research employs Epstein-Zin preferences [144] for utility computation, separating risk aversion and inter-temporal consumption preference parameters for more accurate agent parametrization. Rewards are modelled as advantage estimations of utility differences, aggregated at the reward-constituent level and combined according to a set of hyperparameters in the polynomial $\phi$. The Reward Mixture comprises immediate advantage rewards of the agent after taking action, the reward of the agent following all trading and choice activities within a single loop time-step covering all economic activities, as well as immediate operation relevant global re-
wards such as collective production, and global rewards at the end of the simulation loop time-step addressing global objectives like reducing economic fluctuation amplitude and promoting social welfare:

$$r_{i,t}^{Mix} \sim \phi(r_{i,immediate}, r_{i,f}, r_{global,immediate}, r_{global,f})$$ (7.10)

### 7.6.2 Input to ML Models

Each model that is trained for agents must at least get as input the deciding agent’s own state; further inputs regarding the decision must be provided as input. These inputs can be directly relevant aspects such as a trade offer, or can be general signals such as prices from the market, the information flow structure is chosen to imitate the real world, where agents make decisions from multiple sources, but they are not omniscient, the information flow structure shall be closely related with the communication structure that is being observed in the real world.

### 7.6.3 Calibration to Real-World

Calibrating a deep multi-agent model to the statistics and phenomena that are being observed in the real world can be challenging. The only hard inputted information is the inter-company technology production matrix (input-output) that is being used by business to produce sectoral outputs. Further fine-tuning of the system can be done by training the models with a reward signal reflecting the divergence of the simulated statistics from the real world phenomena.

### 7.6.4 Parallelisation and Optimisation

To further enhance the model’s scalability, we can leverage parallelisation techniques and optimisation algorithms. We can significantly speed up the training phase by distributing the learning process across multiple processors or GPUs. Additionally, optimisation techniques such as experience replay, prioritised sampling, and parameter optimisation can be used to enhance learning efficiency and performance. The proposed multi-agent reinforcement learning model for the pension ecosystem can be adapted to handle large-scale systems with millions of agents through a combination of distributed learning, hierarchical structures, model ab-
straction, and optimisation techniques.

7.6.5 Training Complexities

7.6.5.1 Addressing Stationarity and Training Challenges

Multi-agent systems face stationarity challenges compared to single-agent reinforcement learning. The dynamics agents learn actively change during training [22]. Various methodologies address this issue, with PPO, which shows strong empirical performance, being chosen for our multi-agent pension model [102].

PPO [140] was preferred over pure MADDPG [141] due to its on-policy nature, simplifying learning by updating policies based on recent experiences, enhancing sample efficiency. Real-world economic agents typically lack global information access, relying on local observations. Accurate financial models should consider this limited information availability.

MADDPG, while effective in some settings, might not suit financial systems modelling due to its assumption of full state and action space access during training. This could lead to over-optimisation and unrealistic models. PPO, however, optimises local policies based on an agent’s experiences, offering a more realistic representation of market dynamics.

Considering real-world limitations and the importance of local information, PPO is a more suitable choice for financial system modelling. Agents can be trained as meta-learners, acting optimally in different dynamics. Possible approaches include using a latent recurrent state to capture historical experiences or employing a meta-learner methodology [100].

7.6.5.2 Stateful Agents

The proposed model can be augmented with sequential neural networks, such as LSTMs or transformers, which have proven successful in capturing memory and attention to information [9]. This could help train agents to consider future rewards more effectively.
7.6.5.3 Curricula Learning

Training multiple models each governing different aspects for agents such as borrowing decisions, investment decisions and trade decisions is challenging, one way to deal with the complexity is curriculum learning, where the different functionalities of the agents are being enabled and trained incrementally, which can also help to tackle with the training instability problem of reward attribution from the outcomes of multiple machine learning modules of the agents in a temporal setting. Curriculum learning is proven to be beneficial for developing increasingly complex capabilities [95] can greatly speed up and improve the success of the training.

7.6.6 Operational Challenges

7.6.6.1 Parallelization and Optimisation

One challenge that was faced and revealed during profiling the software during making the computations efficient is the fact that trading action requires randomly batching the populations for various cycles and updating the respective agents of the population takes a significantly longer time, which can be to an extent addressed in the future with aspects like unrolling and more asynchronous operations. The important thing is this operation is a linearly dependent kind of operation where in a trading time-step of multiple cycles, trade at previous cycles should affect the trade decisions of the future cycles. For example, if I have already purchased a specific inventory, I should not decide to do the same again.

7.6.6.2 Ensuring Numerical Stability

A simulation of multiple agents where the decisions are made by neural networks, tends to be very unstable in the initial training runs, so a very comprehensive and careful clipping behaviour is necessary, where the outputs, gradients, and losses are carefully clipped and accounting for the numerical stability issues.

7.6.6.3 Issues with Trade Agent Modules

One other issue that was faced in the trade agent module is the scheduling of the trade cycles, where the businesses are sampled uniformly, which results in a case where the sectors with very few but large players such as the energy sector find
less chance to interact with people, and other businesses such that it can results in fewer trades, and scarcity of the outputs of such sectors. One potential solution is sampling the businesses for trading turns according to the volume of their capital.

### 7.6.6.4 Complexities in Integrating ML with Environment Dynamics

One aspect that made the development cycles challenging is the consecutive interplay between multiple agents, and the consecutive execution of market operations and policy function inferences, which makes the debugging of a problem difficult, in the sense that one needs to both policy inferences of multiple agents and execution of these decisions, and try to understand where the behaviour deviates from the common-wisdom. Consider the flexibility and debugging issues of the ML system, and suggest potential approaches to remedy them.

### 7.6.7 Theoretical Challenges

#### 7.6.7.1 The Balance of Agent Count and Agent Interaction

There are multiple factors that can affect the nature of trades that are being carried, one is the ratio of agent count and number of trading cycles. One question that needs to be further investigated in future research is the trade dynamics that are being manifested varying according to the count of cycles, can different emergent properties be manifested according to the count of cycles?

#### 7.6.7.2 Bootstrapping Populations

One critical issue of calibration and initialisation is the bootstrapping population, the balance between the expectation of various income, wealth and capital distributions in the population in fidelity with the real world, and the dynamics that will be manifested that is governed by the input-output matrix and consumption vectors that are being inputted requires careful consideration, and the impact of these initial, situations and the potential equilibria that is reached after a long run of simulation is not necessarily same or constant. So setting up initialisations is a tricky art, but one might expect that there exists a vast majority of initialisation distributions for
populations that will end up with a relatively stable simulation run after training for fairly long epochs.

7.7 Conclusion

This research introduces a novel approach by exploring the interplay between agents that shape market dynamics. A multi-agent interactive system was developed, allowing for a better representation of real-world dynamics.

In this study, we identified three main kinds of challenges for MARL in pension ecosystems. The training challenges is a central one: calibration of the environment was mainly based on the inter-company technology production matrix (input-output), but further details on the economic environment could make the ecosystem more realistic. Another critical training challenge is the sensitivity with respect to the intertemporal reward attribution [142, 143] during the first phases of the training. Secondly, we have operational challenges that include computational optimisation, as well as numerical stability during training, and the complexity of integrating MARL with a non-stationary financial environment. Finally, we have theoretical challenges, which are related to the modelling of the agents and their interactions, and the interpretation of the transient bootstrapping dynamics and of the long-term general equilibrium dynamics.

This research introduces a multi-agent reinforcement learning model adapted for the pension ecosystem, which identifies optimal saving and investment strategies for contributors. By incorporating multiple agents into the model, we are able to model market shocks, business cycles, and policy initiatives together with contributor dynamics. This not only provides a new approach for synthetic income trajectories but also enables more inclusive and adaptive savings strategies.
## 7.8 Appendix

**Table 7.2:** Input-Output Matrix of Sectors Excluding inv\_cash

<table>
<thead>
<tr>
<th></th>
<th>inv_farm</th>
<th>inv_energy</th>
<th>inv_mining</th>
<th>inv_production</th>
<th>inv_labour</th>
</tr>
</thead>
<tbody>
<tr>
<td>inv_farm</td>
<td>0.010</td>
<td>0.030</td>
<td>0.020</td>
<td>0.001</td>
<td>0.010</td>
</tr>
<tr>
<td>inv_energy</td>
<td>0.001</td>
<td>0.003</td>
<td>0.003</td>
<td>0.030</td>
<td>0.020</td>
</tr>
<tr>
<td>inv_mining</td>
<td>0.010</td>
<td>0.010</td>
<td>0.010</td>
<td>0.010</td>
<td>0.020</td>
</tr>
<tr>
<td>inv_production</td>
<td>0.010</td>
<td>0.010</td>
<td>0.010</td>
<td>0.020</td>
<td>0.015</td>
</tr>
<tr>
<td>inv_labour</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

**Table 7.3:** Consumption Vector

<table>
<thead>
<tr>
<th></th>
<th>inv_farm</th>
<th>inv_energy</th>
<th>inv_mining</th>
<th>inv_production</th>
<th>inv_labour</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consumption</td>
<td>0.200</td>
<td>0.150</td>
<td>0.100</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Parameter</td>
<td>Value</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>--------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sim_director_steps</td>
<td>1001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>epoch steps</td>
<td>3700</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>consumption_choice_agent</td>
<td>{“gamma”: 0.2, “threshold”: 1}</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>interest_rate</td>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>asset_type</td>
<td>“cash_riskless_risky”</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>m_risky_asset</td>
<td>1.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>s_risky_asset</td>
<td>0.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>emp_trade_agent_args</td>
<td>{“executor_args”: {“turn_count”: 50}}</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>trade_agent_args</td>
<td>{“executor_args”: {“turn_count”: 2}}</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b2c_trade_agent_args</td>
<td>{“executor_args”: {“turn_count”: 2}}</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>RL and NN Params</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>model_offer</td>
<td>{“hidden_size”: 32}</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>model_decide</td>
<td>{“hidden_size”: 32}</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rl_learning_rate</td>
<td>0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gamma</td>
<td>0.95</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gae_lambda</td>
<td>0.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>norm_adv</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>norm_ret</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>clip_coef</td>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ent_coef</td>
<td>0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vf_coef</td>
<td>0.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>max_grad_norm</td>
<td>0.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>train_batch_size</td>
<td>125</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n_mini_batches</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 7.4: Model Card Parameters**
Chapter 8

General Conclusions

8.1 Summary of the Study

The core aim of this research has been to devise more nuanced and efficient strategies for pension planning, in response to a complex pension ecosystem affected by various societal and economic factors such as the shifting responsibility from Defined Benefit (DB) to Defined Contribution (DC) schemes, changing demographics, and increasingly diverse income trajectories. To achieve this goal, I built upon the fundamental theories of pension finance, incorporated innovative ABM models with deep reinforcement learning in a computationally efficient way, and introduced innovative models to adapt to real-world scenarios.

This research represents a new approach bridging traditional econometric models with advanced machine learning and agent-based modelling techniques, tailored specifically for pension planning. It overcomes the limitations of classical econometric models by introducing methodologies for capturing investment and income dynamics in environments of higher complexity and heterogeneity, that traditional econometric models alone are incapable of capturing. This research introduces a multi-actor economic setting that enables endogenous market dynamics. By allowing different types of participants to interact within this setting, it becomes possible to capture key economic phenomena, such as market fluctuations and cycles. A moonshot side goal of this research is a pathway to modelling financial systems by training agents to take actions, that collectively manifest market ecosystem with
dynamics similar to those observed in the real world, without the need to have microdata of individual actions. Throughout this research, I have explored and acknowledged the strengths and challenges of combining agent-based modelling, machine learning, and reinforcement learning for pension planning.

This research has laid a foundation for understanding the evolving income dynamics as a function of age and occupation, reflecting the intricate aspects of the real-world income trajectories. It underlined the fundamental differences in the income distribution for different demographics, which is vital in designing more inclusive pension strategies. It also critically recognises the inadequacy of the current DC funds determining portfolio allocation based only on age, and addresses the need for incorporating income history, profession, sector of employment, career progression, lifestyle choices, and risk tolerance.

This research builds on the classical lifetime portfolio optimisation and consumption decision problem, highlights the connection and mapping between classical econometric models, and models trained with deep reinforcement learning, and evolves the introduced models to address the challenges of increasing complexity and heterogeneity. Further, the research has provided an innovation in optimising portfolio and saving strategies by utilising deep reinforcement learning. Through these models, we were able to handle heterogeneous income paths, capture historical income trajectories, and devise dynamic asset allocation strategies. These models are designed to be robust to environmental changes and unexpected shocks, catering to the need for resilience in a non-stationary environment. The classical econometric models can not capture the dynamics of complex environments, the deep reinforcement learning is capable of capturing the increasingly complex and heterogeneous environments.

Lastly, the research has established a robust framework for modelling the multi-actor pension ecosystem using multi-agent reinforcement learning (MARL). This allowed the research to understand and predict the complex interactions among various agents, and to adapt to changing economic scenarios and unexpected events. The inclusion of MARL is a significant step forwards in bridging the existing gap
between Agent-Based Models in finance and MARL, enabling the simulation to capture complex macroeconomic aspects through learning micro-interactions.

### 8.2 Broader Contributions

The implications of this research extend beyond its specific findings. The methodology provides a novel approach for incorporating complex income dynamics and diverse profiles into financial decision-making models. This opens up possibilities for more nuanced financial modelling and strategy development in fields beyond pension finance. It takes into account the rising gig economy, increased part-time work, and more frequent career shifts, making it relevant for a wide range of income trajectories.

Furthermore, the research enhances the understanding of how deep reinforcement learning and multi-agent reinforcement learning can be applied in financial modelling. It bridges the gap between traditional econometric models and the complexities of real-world financial systems, offering a fresh perspective and innovative tools for financial planning.

In conclusion, this research, motivated by the significant changes in pension systems and societal shifts, has the potential to significantly improve the financial security of individuals as they approach retirement, by providing sophisticated tools and methodologies to navigate the complex pension ecosystem. By doing so, it contributes to the broader societal goal of ensuring financial stability and security for all in their post-working years. It sets a promising path for future research to continue refining these models and exploring their applicability to various regions and economies.

### 8.3 Limitations and Future Work

Despite its contributions, this research has certain limitations. First, while the model captures the income dynamics as a function of age and occupation, other factors affecting income, such as education level, region, and industry, were not included. These variables could potentially influence income trajectories and therefore pension savings. Secondly, our reinforcement learning models do not explicitly incor-
porate behavioural biases that can affect investment and saving decisions. For instance, individuals may be prone to myopic loss aversion or overconfidence, which may affect their saving behaviour. Another limiting factor to note is the computational resources that are available, what is possible is strictly bounded by the availability of computational resources.

Future work should aim at addressing these limitations. Expanding the model to include additional variables that can influence income would provide a more comprehensive understanding of income dynamics and lead to improved saving strategies. Incorporating behavioural biases into the models would also allow a more accurate representation of individuals’ decision-making processes, leading to strategies that are better aligned with their actual behaviour. The integration of behavioural finance aspects into reinforcement learning models could be a promising direction for future research. Integrating additional factors into the model and testing for various variations of the existing factors require additional computational resources.

Moreover, future research could explore the application of the proposed model to specific countries or regions, which would help in understanding the global dynamics of pension ecosystems and in developing robust strategies that can be applied across diverse geographical locations.


[97] Ashley Hill, Antonin Raffin, Maximilian Ernestus, Adam Gleave, Anssi Kanervisto, Rene Traore, Prafulla Dhariwal, Christopher Hesse, Oleg Klimov, Alex Nichol, Matthias Plappert, Alec Radford, John Schulman, Szymon


[105] James Bradbury, Roy Frostig, Peter Hawkins, Matthew James Johnson, Chris Leary, Dougal Maclaurin, George Necula, Adam Paszke, Jake VanderPlas,


