Individual insurance choice: A stochastic control approach

by

Wenyuan Li

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Actuarial Science

Waterloo, Ontario, Canada, 2023

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner: **Qihe Tang**
Professor, School of Risk and Actuarial Studies,
University of New South Wales

Supervisors: **Mary R. Hardy**
Professor, Department of Statistics and Actuarial Science
University of Waterloo

Ken Seng Tan
Professor, Nanyang Business School,
Nanyang Technological University

Pengyu Wei
Assistant Professor, Nanyang Business School,
Nanyang Technological University

Internal Members: **Alexander Schied**
Professor, Department of Statistics and Actuarial Science,
University of Waterloo

Tony Wirjanto
Professor, Department of Statistics and Actuarial Science,
University of Waterloo

Internal-External Member: **Ken Vetzal**
Associate Professor, School of Accounting and Finance,
University of Waterloo

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

Chapter 2 is published in the Insurance: Mathematics and Economics under the paper title: *Demand for non-life insurance under habit formation*, by Wenyuan Li, Ken Seng Tan, and Pengyu Wei (Li et al. (2021)).

Abstract

This thesis applies the stochastic control approach to study the optimal insurance strategy for three problems. The first problem studies the optimal non-life insurance for an individual exhibiting internal habit formation in a life-cycle model. We show that the optimal indemnity is deductible under the expected premium principle. Under the additional assumption of exponential utility functions, we obtain the optimal strategies explicitly and find that habit formation reduces insurance coverage. Our model offers a potential explanation for the global underinsurance phenomenon. Some numerical examples and sensitivity analysis are presented to highlight our theoretical results.

The second problem analyzes the optimal defined-contribution (DC) pension management under stochastic interest rates and expected inflation. Besides financial risk, we consider the mortality risk before retirement and introduce life insurance to the pension portfolio. We formulate this pension management problem by a Hamilton-Jacobi-Bellman (HJB) equation, derive its explicit solution, show the explicit solution's global existence, and prove the verification theorem. Our numerical research reveals that the pension member's demand for life insurance exhibits a hump shape with age and a "double top" pattern for the real short rates and expected inflation (high demand when the real short rates and expected inflation are both high or both low). These demand patterns are caused by the combined effects of the components in the optimal insurance strategy.

The third problem is constrained portfolio optimization in a generalized life-cycle model. The individual with a stochastic income manages a portfolio consisting of stocks, a bond, and life insurance to maximize his or her consumption level, death benefit, and terminal wealth. Meanwhile, the individual faces a convex-set trading constraint, of which the non-tradeable asset constraint, no short-selling constraint, and no borrowing constraint are special cases. Following Cuoco (1997), we build the artificial markets to derive the dual problem and prove the existence of the original problem. With additional discussions, we extend his uniformly bounded assumption on the interest rate to an almost surely finite expectation condition and enlarge his uniformly bounded assumption on the income process to a bounded expectation condition. Moreover, we propose a dual control neural network approach to compute tight lower and upper bounds for the original problem, which can be utilized in more general cases than the simulation of artificial markets strategies (SAMS) approach in Bick et al. (2013). Finally, we conclude that when considering the trading constraints, the individual will reduce his or her demand for life insurance.

Acknowledgements

During my Ph.D. study, I watched an American baseball film called "Moneyball". It talked about Billy Beane, the team manager of the Oakland Athletics, who only had one-third of his opponent's budget but bought undervalued players based on the sabermetric approach. As a result, he built a legendary team who consecutively won 20 games and broke a record in the 103 years of American League baseball.

In my research career, my supervisor Professor Ken Seng Tan is like Billy Beane to me. He dug me out of the city of Chengdu, allowed me to explore my potential, and provided great support for my research and job seeking. During our discussions, he always encourages me to think independently with an "entrepreneurial spirit" and not follow others, which is a valuable treasure for inspiring my future path. I still can't believe how destiny treated him cruelly and took him away from us so early. He is such a sun-shine person who always brings happiness to the surrounding people and never hesitates to help others. With Professor Tan's help, I fortunately met my supervisor, Professor Pengyu Wei, who finally became my close friend and mentor in research and life. Professor Wei opened a new gate for me. Under his guidance, I could enrich my research style from numerical analysis to mathematical analysis and understood many technical papers I couldn't read before. Moreover, in life, he acted like a big brother who gave me plenty of suggestions and helped me live smoothly. I am very grateful to my supervisor Professor Mary R. Hardy, who kindly guided my research with her expert opinions. During the discussion, she constantly criticized my model with her rich empirical experience and challenged my thought on the purpose of the model, which helped me a lot in proceeding with meaningful research. As time goes by, I hope to foster a critical and sharpened mind close to hers.

I would like to thank my committee members, Professor Alexander Schied and Professor Tony Wirjanto, for taking their time and effort to read my thesis, asking inspiring questions during the proposal, and making insightful comments at the final stage. I would like to thank Professor Qihe Tang for being willing to be my external examiner despite his busy schedule and for giving me valuable suggestions on revising the thesis. I also would like to thank my internal-external member, Professor Ken Vetzal, for reading my thesis and providing me with the beneficial advice from a financial perspective.

In my academic life, I would like to express my special thanks to Professor Jingtang Ma, who introduced me to the research field in the fourth year of my undergraduate. He trained my research ability in mathematical finance during my master's period and encouraged me to pursue a Ph.D. in Canada.

During my study in Canada, I would like to thank Professor Alexander Schied for teaching me the Theory of Probability and the Extreme Value Theory courses, which strengthened my ability in probability analysis. I would like to express my sincere thanks to Professor David Saunders, who wrote me many recommendation letters from the Hickman Scholarship application to the job market. I have learned a lot from his Finance I and Finance II courses, even though I previously majored in financial mathematics. I would like to thank Professor Bin Li for teaching me the Stochastic Control course and always choosing me as his teaching assistant on the Loss Model. I would like to thank Professor Ruodu Wang, who introduced me to his Risk Management course and seminars. He led us to play board games every weekend and have a lot of fun together. Moreover, I also would like to thank Professor Yingli Qin for teaching me Estimation and Hypothesis Testing and never getting bored when I asked stupid questions during her office hour. Furthermore, I would like to thank Ms. Mary Lou Dufton and Mr. Greg Preston, who provided tremendous administrative and IT helps, which made my life much easier.

Reflecting on my Ph.D. journey, I can't forget how I met so many lovely friends at UWaterloo. We fought the comprehensive exams together, took challenging courses, talked about mathematics and poems while eating hotspots, and explored the beauty of Ontario in different seasons. Here, I would like to express my sincere thanks to my friends: Xiyue Han, Chi-Kuang Yeh, Yechao Meng, Xiaoxue Deng, Yuyu Chen, Zijia Wang, Fangya Mao, Zhaohan Sun, Ce Yang, Ruihong Jiang, Qiuqi Wang, Mingren Yin, Yang Liu, and Xia Han. With your company, my Ph.D. life was not like a nerd locked in prison but a painting full of colors. I would also thank my senior friends: Xiaobai Zhu, Yunran Wei, Jingyi Cao, Danqiao Guo, Yuqian Zhao, Lv Chen, and Zhiyi Shen. Thank you for taking me to the gym, giving me suggestions while seeking jobs, and broadening my horizon to the industry. Although I was not very social in my last year, I still made many junior friends and impressed by their energy and skills: Haoming Shi, Yuanyuan Zhang, Jiayue Zhang, Benxuan Shi, Yiping Guo, Chi Zhang, Zhenzhen Huang, Liyuan Lin, Xintong Li, Ziyue Shi, Yimiao Zhao and Zhanyi Jiao. For junior friends, I would say UWaterloo is yours! If you have a dream, go for it, and don't hesitate. I wish you all a bright future and success in your career.

Lastly, I would like to thank my parents for supporting me through this long journey. In my pursuit of a Ph.D. degree, missing and solitary were always their companions. I hope we can reunite soon and enjoy a good time together.

Dedication

In memory of my supervisor, Professor Ken Seng Tan.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

"The essence of finance is the value exchange crossing time and space."

This is a brief definition provided by Professor Zhiwu Chen in his book "The Logic of Finance" Chen (2009), which reveals the nature of financial practices.

Among all the financial practices, insurance is unique for its protection purpose. In a typical insuring procedure, the individual pays a reasonable premium in exchange for the protection of the future loss. The compensation can really smooth big losses as in people's lives and help individuals out of hopeless situations. However, choosing insurance unwisely could lead to serious results, such as over-insurance (insurance amount is in excess of insured object's fair value) and under-insurance (insurance amount is not enough to cover the risk). According to Lloyd's (2018), in 2018, the insurance gap for developed countries and emerging nations are, respectively, the US $2.5 billion and US $160 billion. Therefore, it is of great importance to study the optimal insurance strategy for different contracts.

Generally speaking, insurance contracts can fall into two categories: short-term insurance and long-term insurance. A short-term insurance contract refers to a contract with a period of one year or less. It includes non-life insurance (e.g., automobile insurance, homeowners insurance, etc.) and health insurance (e.g., medical insurance, dental insurance, etc.). A long-term insurance contract is a contract with a period of more than one year. It covers life insurance, annuities, and pension plan.

In the field of short-term insurance, one popular research topic is the optimal choice of the non-life insurance contract. The problem is to determine the indemnity function, which measures the amount of money paid to the individual when the loss occurs. Early work has shown that the optimal indemnity is deductible insurance under the expected premium principle in the single-period setting (see Arrow, 1963; Mossin, 1968; Raviv, 1979). This result has been well extended to the dynamic case under various settings (see Moore and Young, 2006; Perera, 2010, 2013; Steffensen and Thøgersen, 2019; Touzi, 2000; Zou and Cadenillas, 2014). However, most of this literature assumes that the individual's preference follows an additively time-separable utility. In practice, this assumption is not realistic. The individual's preference may not only depend on instantaneous consumption but also be correlated with past habits. Therefore, it is of great interest to study the impact of previous consumption on the individual's insurance strategy.

Chapter 2 aims to study optimal insurance when considering past consumption in dynamic settings. We assume an individual allocates his or her wealth between consumption, saving, and insurance to hedge property losses throughout their lifetime. Habit formation is introduced to measure the previous consumption and relax the time separability of the individual's preference. Most financial literature has shown that the individual will reduce their investment in the risky asset to maintain a reasonable habit level (see Constantinides, 1990; Munk, 2008; Polkovnichenko, 2006; Sundaresan, 1989). However, the effect on the insurance strategy has not been clearly studied in the actuarial field. We introduce habit formation and derive that the optimal insurance is still a deductible under the expected premium principle in a general sense. Then, under the exponential utility, the explicit solution is derived if consumption utility is defined as a function of the difference between current consumption and habit formation. Some interesting corollaries are derived with the solution. First, the optimal deductible is decreasing throughout the time, implying that the individual would purchase more insurance if he or she ages. Second, the insurance demand is reduced after considering habit formation, which shows a similar phenomenon to the risky-asset demand in financial literature. Finally, a rigorous verification theorem is proved to guarantee the optimality of the candidate solution and strategies.

Motivated by the popularity of proportional insurance, we also study the demand of this contract under habit formation. More comprehensive results are derived in this case. First, the demand for proportional insurance increases with age. Second, the habit formation reduces the optimal proportion of loss insured. Third, the individual would buy no proportional insurance in the early years under some conditions (e.g., the insurance is too expensive or the loss is not severe). Compared with deductible insurance, proportional insurance is only a sub-optimal solution under the expected premium principle. A natural question is to examine the welfare loss from choosing the sub-optimal solution. We define the welfare loss as the minimum wealth compensated so that the individual is willing to choose the sub-optimal solution. After rigorous analysis, we find the welfare loss from proportional insurance is relatively small compared to that from no insurance.

Chapter 2 draws a conclusion in sharp contrast to that of Ben-Arab et al. (1996), who also study the optimal insurance demand under habit formation. We show that insurance demand is reduced when considering habit formation, while they obtain the opposite result. After close comparison, it turns out that the choice of utility function contributes to the conflicting result. Our result is highly consistent with the global under-insurance phenomenon, and thus habit formation could potentially offer a plausible explanation for the insurance gap worldwide.

In the realm of long-term insurance, one of the major challenges is the pension management problem. Most pension funds can be classified into two schemes: defined-benefit (DB) pension schemes and defined-contribution (DC) pension schemes. Generally speaking, a DB pension plan provides the individual with a guaranteed retirement benefit based on his or her earnings history, tenure of service, and age; the DC pension plan collects a predetermined contribution in the accumulation period and returns the retirement benefit related to the investment earnings. There has been a global shift from DB pension plans to DC pension plans. According to the OECD report, in 2019, less than 50% of pension assets are managed in DB schemes in 28 out of 33 reporting jurisdictions (see OECD, 2020).

Compared with other funds, the long-term horizon is one of the key features of DC pension management. The accumulation period of a DC pension usually lasts for 20-40 years, and the individual is exposed to various risks before retirement age. In general, the risk to the DC participant can be classified into two categories: financial risk and mortality risk. Financial risk is the risk caused by the changing economic indicators and financial environment. There is a large amount of literature studying optimal DC management under financial risk. The results show that a failure to hedge the time-varying interest rate and inflation rate could lead to a large loss of individual's purchasing power after retirement (see Battocchio and Menoncin, 2004; Boulier et al., 2001; Chen et al., 2017; Han and Hung, 2012). The mortality risk is the risk that an individual may die before the retirement age. According to the latest life tables, the probability of a 22-year old dies before age 65 is respectively 15.23% in the United States and 9.52% in Canada, which is not negligible to the pension managers. The current literature focuses on two types of pre-mature death benefit clauses: the return of premiums (see Bian et al., 2018; He and Liang, 2013; Li et al., 2017; Sun et al., 2016) and the return of account value (see Blake et al., 2008; Konicz and Mulvey, 2015; Wu and Zeng, 2015; Yao et al., 2014). In practice, most DC pension plans would return the account value to the pre-mature dead individuals since the investment revenue in DC plan is also the inheritable estate (see Publication 575 IRS (2019) of 401(k) plan in the United States and other examples in Canada (RBC, 2020; Sun Life Financial, 2017)).

Chapter 3 considers a DC pension management problem in a complete market. The financial market follows a two-factor model proposed by Koijen et al. (2011), which permits time variation in real interest rates, inflation rates, and risk premiums. The pension participant allocates his or her wealth among a stock index, nominal and inflation-linked bonds, and a nominal cash account. Besides financial instruments, the individual can also buy life insurance to hedge his or her mortality risk before retirement. The pre-mature death benefit consists of the DC account value and the payment from life insurance.

We formulate this pension management problem as a utility maximization problem and derive the corresponding Hamilton-Jacobi-Bellman (HJB) equation. An explicit solution is derived under the constant relative risk aversion (CRRA) utility, and its global existence can be proved based on the related matrix Riccati equation. Furthermore, rigorous verification theorems are provided in different ranges of utility's risk-aversion coefficient. In the numerical research, we estimate the model parameters by the Kalman filter method and acquire rich conclusions on individual's insurance demand. More specifically, in the dynamic analysis, we find that the pension plan member's demand for life insurance follows a hump shape, and peaks in old age. In the static analysis, we find the individual's demand for life insurance exhibits a "double top" shape for the real short rate and expected inflation. In general, Chapter 3 constructs a DC account resembling the variable annuity with endogenously determined time-varying death benefits. It relaxes the limitations on the variable annuity's death benefits and can inspire more innovations in designing the new actuarial products.

The other hot topic in long-term insurance is the individual's demand for life insurance under trading constraints. In the real market, the individual faces many trading restrictions from stock brokers and exchanges. These trading constraints influence individuals' investment behavior and affect their earnings and insurance demand.

The existing literature formulates this problem into the constrained portfolio optimization problem. It considers the trading constraints, such as non-tradable assets (incomplete market), no short-selling constraint, no borrowing constraint, etc., and adjusts the ideal model to a more realistic market model. Some seminal papers solve this problem by dual control approach (see Cvitanić and Karatzas, 1992; Karatzas et al., 1991). They build a group of artificial markets and fulfill the trading constraints by manipulating the drift terms of bonds and stocks. He and Pages (1993) add the labor income to the problem and consider the investment with no borrowing constraint. Cuoco (1997) extends Cvitanić and Karatzas (1992) to the case with stochastic income and absorbs He and Pages (1993)'s work (no-borrowing constraint) as special cases. For more recent works and research, we refer to Bick et al. (2013); Chabakauri (2013); Haugh et al. (2006); Jin and Zhang (2013); Kamma and Pelsser (2022); Larsen and Žitković (2013); Mostovyi and Sîrbu (2020). In the actuarial science field. Zeng et al. (2016) extend He and Pages (1993)'s work and study the wealth-constraint effect on the individual's demand for life insurance. Dong and

Zheng (2019) study the optimal defined contribution pension management under short-selling constraints and portfolio insurance. Hambel et al. (2022) build a group of artificial insurance markets to solve a life-cycle model with unhedgeable biometric shocks. However, most existing actuarial literature only focuses on one or two trading constraints, and a general framework is lacking in studies of life-cycle investment.

Chapter 4 studies the constrained portfolio optimization problem in a generalized life cycle model. The individual has a stochastic income and aims to find the optimal trading and insurance strategies to maximize their consumption, bequest, and terminal wealth levels. Inspired by Cuoco (1997)'s framework, we use a non-empty, closed, and convex set to describe the trading strategy, which contains non-tradeable assets, no short-selling constraint, and no borrowing constraint as special cases. Furthermore, we use the "relaxation projection" technique in Levin (1976) to prove the existence of the primal problem. Compared to the Cuoco (1997)'s framework, we relax his uniformly bounded assumptions on the interest rate and income process. In Chapter 4, we only assume the expected exponential integral of the interest rate's absolute value is finite and derive a weaker condition for the income process.

In the numerical research, we propose a dual control neural network approach and compare it with the simulation of artificial markets strategies (SAMS) approach from Bick et al. (2013). We find that the two approaches perform closely when the risk-free interest rate, stock appreciation rate, and volatility are all constant. If the stock appreciation rate follows a perturbation in time, then the SAMS approach is inadequate to solve the problem, but the dual control neural network approach still works well. Lastly, according to the numerical study, both approaches show that the individual will reduce their demand for life insurance in the presence of trading constraints. To the best of our knowledge, this is the first application of the neural network to compute the constrained portfolio optimization problem. It can inspire future work using the neural network to study the optimal investment under realistic market situations.

Chapter 2

Demand for non-life insurance under habit formation

2.1 Introduction

An important problem in actuarial science and insurance economics is to examine the optimal design of insurance. The problem is to determine the so-called indemnity function, i.e., the amount of monetary compensation as a function of the loss, to maximize the insured's utility. It has been well established in the single-period setting that the optimal indemnity is deductible if the expected premium principle is used in the insurance pricing; see the seminal papers by Arrow (1963); Mossin (1968); Raviv (1979). The optimal insurance problem has also been extended to the dynamic setting, see, for example, Briys (1986); Moore and Young (2006); Perera (2010, 2013); Steffensen and Thøgersen (2019); Touzi (2000); Zou and Cadenillas (2014) among others, which show the optimality of certain insurance contracts, mostly deductibles, in various settings. Most of these papers assume that the economic agent's preference is described by an additively time-separable utility function. While this assumption simplifies the analysis, it is not realistic as it assumes that instantaneous satisfaction depends on only instantaneous consumption. In practice, it is conceivable that the economic agent's past consumption can be important. To reflect the past consumption habit, an alternative model is habit formation, which relaxes the time separability of preferences and thus allows the economic agent's current utility to depend on past consumption.

The theory of habit formation has been widely used in finance. Sundaresan (1989) establishes a model where a consumer's utility depends on their consumption history.

He finds the fraction of the wealth invested in the risky asset is no longer a constant but an increasing concave function of wealth, lower than the fraction in Merton (1971). Constantinides (1990) shows that habit formation drives a wedge between the coefficients of relative risk aversion, and the individual will have less risky investment when considering habit formation. Polkovnichenko (2006) constructs a discrete-time life-cycle model with stochastic uninsurable labor income and finds that the allocation to stocks declines with habit because higher habit is harder to maintain. Munk (2008) studies the optimal behavior under a general, possibly non-Markov market price process and concludes that bonds and cash are better investment objects than stocks when the individual needs to ensure a habit level. In particular, habit formation has been successfully used in analyzing asset prices (Abel, 1990; Constantinides, 1990; Detemple and Zapatero, 1991; Sundaresan, 1989).

This chapter contributes to the optimal insurance design problem in the dynamic context under habit formation. We propose a life-cycle model incorporating internal habit formation in the preference to study the economic agent's optimal consumption, saving, and demand for insurance. In particular, we assume that the economic agent's instantaneous utility function depends on not only instantaneous consumption but also a consumption habit defined by the weighted average of past consumption. We show that the optimal indemnity is still deductible insurance if the expected premium principle is used. We further make simplifying assumptions that the economic agent derives utility from the difference between consumption and habit and that the utility function is of the exponential type. We obtain optimal strategies in closed-form and find that the optimal deductible is decreasing in age, implying that the individual gradually increases his or her insurance coverage. Moreover, the presence of habit formation reduces the insurance coverage.

Motivated by the fact that proportional insurance is prevalent in the market and a commonly imposed condition in the study of dynamic optimal insurance, we provide additional analysis assuming an individual can only purchase proportional insurance. Under the more restrictive model setting, we conclude the following: First, the optimal proportional insurance coverage increases with age. Second, habit formation reduces the optimal proportion of losses insured. Third, economic agent may completely opt out of the insurance market and optimally choose to self-insure, especially during early ages.

In our model, the expected premium principle is used in insurance pricing and thus the optimal insurance is deductible insurance. This motivates us to quantify the welfare loss from suboptimal strategies such as no insurance coverage or proportional insurance. We define the welfare loss as the additional wealth that an individual who follows a suboptimal strategy needs to hold, in order to yield the same level of expected utility of the optimal strategy. We find that the welfare loss from proportional insurance is relatively small compared to that from no insurance coverage.

Our result is closely related to Ben-Arab et al. (1996) who also study optimal insurance demand under habit formation. Our model differs from Ben-Arab et al. (1996) in many aspects. First, Ben-Arab et al. (1996) assume that the loss size is given by the total wealth at the time of the loss, while we model the individual risk as a compound Poisson process which is independent of the wealth. Second, Ben-Arab et al. (1996) focus on proportional insurance under the expected premium principle while we allow the economic agent to choose from a general class of indemnity functions. In fact, we show that the optimal strategy is deductible. Third, Ben-Arab et al. (1996) use power utility while our explicit solutions are obtained under exponential utility which allows us to abstract away from the wealth effect. The differences in the assumptions lead to totally different results. Ben-Arab et al. (1996) show that habit formation increases proportional insurance coverage. In contrast, we show that the insurance coverage of the individual, either deductible or proportional insurance, is reduced by habit formation. This phenomenon is of particular interest in that it is more consistent with empirical evidence. The recent report by Lloyd's (2018) indicates that there exists an insurance gap worldwide. For example, in 2018 the insurance gap for developed countries and emerging nations are, respectively, US \$2.5 billion and US \$160 billion. Our results suggest that habit formation in our proposed life-cycle model offers a potential explanation for the global underinsurance phenomenon.

The rest of the chapter is organized as follows. Section 2.2 introduces the economic setting and Section 2.3 formulates the general optimization problem. Section 2.4 solves the model analytically under the exponential utility. Section 2.5 quantifies welfare losses from suboptimal strategies. Section 2.6 presents numerical examples and Section 2.7 concludes the chapter. All proofs are relegated to Appendix A.

2.2 Model

We consider a finite horizon life-cycle model similar to those in Ben-Arab et al. (1996); Moore and Young (2006); Steffensen and Thøgersen (2019). We assume there is an economic agent endowed with initial wealth x_0. He or she consumes at the rate $\{c_t\}_{0\le t\le T}$ and invests the remaining wealth into a risk-free asset that earns an interest rate r.

Following Moore and Young (2006); Steffensen and Thøgersen (2019), we assume the individual is exposed to a risk which is modeled by a compound Poisson process

$$A_t = \sum_{i=1}^{N_t} Y_i, \tag{2.1}$$

where $\{N_t\}_{0\leq t\leq T}$ is a Poisson process with intensity λ representing the number of losses up to time t and the Y_is are loss sizes. It is assumed that Y_is are independent and identically distributed with distribution F on $(0,\infty)$, and that they are independent of $\{N_t\}_{0\leq t\leq T}$. This formulation is slightly different from that in Ben-Arab et al. (1996) where the loss size is assumed to be equal to the total wealth.

As in Moore and Young (2006), the individual can purchase per-claim insurance to reduce the risk exposure. Denoting by I_t the individual's indemnity at time t, the insurance pays $I_t(Y)$ if the individual suffers a loss of Y at time t. We assume $0 \leq I_t(Y) \leq Y$ to exclude over-insurance. Assume that the insurance premium is payable continuously at the rate $\lambda(1+\theta)E[I_t(Y)]$, where the insurer prices the insurance risk by the expected value principle and θ is the safety loading. Then, the wealth of the individual evolves according to

$$dX_t = rX_t dt - c_t dt - \lambda(1+\theta)E[I_t(Y)]dt - R_t(Y)dN_t, \quad X_0 = x_0, \tag{2.2}$$

where $R_t(Y) = Y - I_t(Y)$ is the retention function. Our focus is the impact of habit formation on the individual's demand for non-life insurance and thus we abstract from other factors such as stock investment that may complicate the analysis. There is no essential difficulty to carry out the analysis to include equity investment.

In contrast to Steffensen and Thøgersen (2019) who assume the time-horizon is infinite and thus the indemnity is time-invariant, we consider a finite-horizon problem and assume that the economic agent can adjust the indemnity continuously. This assumption is common in practice as consumer's insurance coverage varies according to the age and that an individual typically renews the policy, say, on an annual basis. Our formulation also allows us to examine the time pattern of the individual's coverage and how habit formation can change the time pattern.

Our model also differs from that in Ben-Arab et al. (1996) who assume that the economic agent can only purchase proportional insurance. In contrast, we assume that the insured can choose the optimal insurance strategy from a large class of indemnity functions. This allows us to investigate the optimal design of the personal non-life insurance.

2.3 Optimization problem

Following Ben-Arab et al. (1996); Boyle et al. (2022); Constantinides (1990); Kraft et al. (2017); Sundaresan (1989), we assume the economic agent's preference exhibits internal habit formation. Define the habit level at time t as

$$h_t = h_0 e^{-\beta t} + \alpha \int_0^t e^{-\beta(t-s)} c_s ds,$$

or, equivalently,

$$dh_t = (\alpha c_t - \beta h_t)dt, \tag{2.3}$$

where h_0, α and β are non-negative constants, and h_0 is the initial habit level. The habit level is a multiple of the weighted average of past consumption rates with the weights being exponentially decreasing so that the recent consumption rates have greater emphasis. β measures the persistence of past consumption and low β implies high persistence. α is a scaling parameter that measures the intensity of consumption habits. As α increases, the habit places more emphasis on the history of consumption. It is assumed that $\beta > \alpha$ to ensure that the habit level will decline when the investor consumes at the habit level.

The individual chooses $\{c_t, I_t\}_{0 \leq t \leq T}$ to maximize his or her discounted expected utility of consumption and terminal wealth (bequest)

$$\sup_{c,I} E\left[\int_0^T e^{-\delta t} U_1(c_t, h_t)dt + e^{-\delta T} U_2(X_T)\right],$$

where δ is the subjective discount rate, $U_1(c, h)$ is the utility function of instantaneous consumption, and $U_2(x)$ is the utility function of bequest. We assume that $\partial U_1/\partial c > 0$, $\partial^2 U_1/\partial c^2 < 0$, $\partial U_1/\partial h < 0$, $\partial U_2/\partial x > 0$, and $\partial^2 U_2/\partial x^2 < 0$. Similarly to Polkovnichenko (2006), we posit that the utility of bequest does not depend on the consumption habit. Naturally, this means at the terminal time, only the bequest of the wealth takes place.

We use dynamic programming to solve the individual's optimization problem. Denote by V the indirect utility function (value function)

$$V(t, x, h) = \sup_{c,I} E\left[\int_t^T e^{-\delta(s-t)} U_1(c_s, h_s)ds + e^{-\delta(T-t)} U_2(X_T)\Big| X_t = x, h_t = h\right].$$

V solves the following Hamilton-Jacobi-Bellman (HJB) equation

$$\begin{aligned} &V_t - (\lambda + \delta)V + rxV_x - \beta h V_h + \sup_c\{U_1(c, h) - cV_x + \alpha c V_h\} \\ &\quad + \sup_I\{-\lambda(1+\theta)E[I_t(Y)]V_x + \lambda E[V(t, x - Y + I_t(Y), h)]\} = 0 \end{aligned} \tag{2.4}$$

with the boundary condition $V(T, x, h) = U_2(x)$.

The first-order conditions with respect to (w.r.t.) c and I lead to, respectively,

$$\frac{\partial U_1(c, h)}{\partial c} - V_x + \alpha V_h = 0,$$

and

$$(1 + \theta)V_x = V_x(t, x - Y + I(Y), h).$$

In view of the "no overinsurance" condition $0 \leq I(Y) \leq Y$, we hypothesize that the optimal indemnity is given by

$$I_t^*(Y) = [Y - (x - V_x^{-1}(t, (1 + \theta)V_x, h))]^+,$$

where $[x]^+ = \max(x, 0)$ and $V_x^{-1}(t, y, h)$ is the inverse function of $V_x(t, x, h)$ w.r.t. x.

The following proposition summarizes the above results.

Proposition 2.3.1. *Suppose that $V(t, x, h) \in C^{1,1,1}$ solves the HJB equation (2.4). The optimal consumption c_t^* satisfies*

$$\frac{\partial U_1(c_t^*, h)}{\partial c} - V_x + \alpha V_h = 0, \tag{2.5}$$

and the optimal indemnity is given by

$$I_t^*(Y) = [Y - (x - V_x^{-1}(t, (1 + \theta)V_x(t, x, h), h))]^+. \tag{2.6}$$

Proposition 2.3.1 states that the optimal insurance in the presence of habit formation is deductible under the expected value principle. In the early years, many papers derive this result in the static sense (see Arrow, 1963; Borch, 1975; Mossin, 1968; Raviv, 1979) and Moore and Young (2006) extends the idea to the continuous time model under the per-claim insurance setting. The optimality of deductible insurance is established in Moore and Young (2006) in the absence of habit formation. Therefore, Proposition 2.3.1 generalizes the results in Moore and Young (2006) to the case of habit formation.

2.4 Optimal Insurance under Exponential Utility

Proposition 2.3.1 in the preceding section establishes the optimality of deductible insurance under habit formation. By assuming that the individual's utility function is given by the exponential utility function, in this section, we derive the optimal demand for insurance explicitly. This, in turn, provides additional insights into the demand for insurance.

Inspired by Sundaresan (1989), we assume

$$U_1(c,h) = -\frac{1}{\gamma}e^{-\gamma(c-h)}, \quad U_2(x) = -\frac{\omega}{\gamma}e^{-\gamma x}, \tag{2.7}$$

where $\gamma > 0$ is the Arrow-Pratt coefficient of absolute risk aversion (Pratt, 1964) and $\omega > 0$ measures the strength of the bequest.[1] We emphasize that the exponential utility is a utility which exhibits a constant absolute risk aversion (CARA) and that such utility has been widely used in the life-cycle literature, see, for example, Merton (1971), Caballero (1991), Wang (2006), and Wang (2009), among others. In particular, Sundaresan (1989) and Angelini (2009) study optimal consumption under habit formation and CARA utility. Moore and Young (2006) and Steffensen and Thøgersen (2019) analyze optimal insurance choice under CARA utility but without habit formation.

2.4.1 Optimal policies

The following proposition presents the optimal consumption and insurance strategies in the presence of habit formation.

Proposition 2.4.1. *Suppose that the utility functions are given by (2.7). The candidate solution to (2.4) is given by*

$$\phi(t,x,h) = -\frac{1}{\gamma}e^{-\gamma(a(t)x+b(t)h+g(t))}, \tag{2.8}$$

the optimal indemnity is

$$I_t^*(Y) = \big[Y - d(t)\big]^+, \tag{2.9}$$

and the optimal consumption is

$$c_t^* = -\frac{1}{\gamma}\ln[a(t) - \alpha b(t)] + a(t)X_t^* + (b(t)+1)h_t^* + g(t), \tag{2.10}$$

[1] The utility of bequest is necessary to prevent terminal wealth from dropping to negative infinity when CARA utility functions are employed and the time horizon is finite.

where X_t^* is the optimal wealth process in the presence of habit formation and h_t^* is the optimal consumption habit process.
Here,

$$
\begin{aligned}
a(t) &= 1/J(t), \\
b(t) &= G(t)/J(t), \\
G(t) &= (e^{-(r+\beta-\alpha)(T-t)} - 1)/(r+\beta-\alpha), \\
J(t) &= \int_t^T (1-\alpha G(s))e^{-r(s-t)}ds + e^{-r(T-t)} \\
&= \frac{r+\beta}{(r+\beta-\alpha)r} + \left(1 - \frac{r+\beta}{(r+\beta-\alpha)r} - \frac{\alpha}{(r+\beta-\alpha)(\beta-\alpha)}\right)e^{-r(T-t)} \\
&\quad + \frac{\alpha}{(r+\beta-\alpha)(\beta-\alpha)}e^{-(r+\beta-\alpha)(T-t)}, \\
d(t) &= \frac{\ln(1+\theta)}{\gamma a(t)} = \frac{\ln(1+\theta)}{\gamma}J(t),
\end{aligned}
$$

and

$$
\begin{aligned}
g(t) &= -\frac{\ln w}{\gamma}e^{-\int_t^T a(s)-\alpha b(s)ds} \\
&\quad + \int_t^T e^{-\int_t^u a(s)-\alpha b(s)ds}\left\{\frac{\lambda+\delta}{\gamma} + [\ln(a(u)-\alpha b(u)) - 1]\frac{a(u)-\alpha b(u)}{\gamma}\right. \\
&\quad -\lambda(1+\theta)\left[\int_{\frac{\ln(1+\theta)}{\gamma a(u)}}^{\infty} ydF(y) - \frac{\ln(1+\theta)}{\gamma a(u)}\overline{F}\left(\frac{\ln(1+\theta)}{\gamma a(u)}\right)\right]a(u) \\
&\quad \left. -\frac{\lambda}{\gamma}\left[(1+\theta)\overline{F}\Big(\frac{\ln(1+\theta)}{\gamma a(u)}\Big) + \int_0^{\frac{\ln(1+\theta)}{\gamma a(u)}} e^{\gamma a(u)y}dF(y)\right]\right\} du,
\end{aligned} \tag{2.11}
$$

where $\overline{F} = 1 - F$ is the tail distribution function of Y.

Before proving the verification theorem, we first give the following definition of the admissible set $\mathcal{A}$.

Definition 2.4.1. *We say that a strategy (c, I) belongs to the admissible set $\mathcal{A}$ if*

1. *(2.2) has a unique strong solution X_t;*
2. *The indemnity $I_t(Y)$ satisfies $0 \le I_t(Y) \le Y$;*

3. $E[\exp\{C_1 X_t + C_2 h_t + C_3 c_t\}] < \infty$*, for any bounded constants* C_1*,* C_2*, and* C_3*.*

4. $E[X_t^2 + h_t^2 + c_t^2] < \infty$

Then inspired by the Theorem 3.1. in Øksendal and Sulem (2007), we have the following verification theorem.

Theorem 2.4.1. *The candidate solution given by (2.8) satisfies*

$$\phi(t, x, h) = V(t, x, h) \quad \text{for all } (t, x, h) \in [0, T] \times \mathbb{R}^2, \tag{2.12}$$

and the strategy (c^*, I^*) *given by (2.10) and (2.9) is the optimal consumption and insurance strategy.*

Proposition 2.4.1 states that the optimal consumption depends on not only wealth but also consumption habit; in contrast, the optimal insurance is deductible but the deductible level $d(t)$ is a deterministic function of t. The optimal deductible is independent of wealth as the exponential utility exhibits CARA and the loss process is not related to wealth. This independence has been widely observed under exponential utility, see, for instance, Moore and Young (2006) and Steffensen and Thøgersen (2019). Moreover, the optimal deductible $d(t)$ is decreasing in risk aversion γ and increasing in the safety loading (premium loading). In other words, the economic agent increases their insurance coverage as he or she becomes more risk-averse, but reduces the coverage when the insurance becomes more expensive.

It is also of interest to investigate how the optimal deductible varies according to time. As shown in the following corollary, the optimal deductible is decreasing in age. In other words, the individual gradually increases the insurance coverage as time approaches the planning horizon. This implies that the individual is becoming increasingly more risk-averse towards the insurable risk as he or she ages. We will provide an explanation through the Arrow-Pratt absolute risk aversion in the next subsection.

Corollary 2.4.1. *Assume that* $\beta > \alpha \geq 0$ *and* $0 < r < 1$*. The optimal deductible* $d(t)$ *is strictly decreasing in* $[0, T]$*.*

We next consider the optimal consumption and insurance for an individual without habit formation. This model serves as a benchmark and allows for a better comparison with optimal deductible obtained in models with and without habit formation.

Proposition 2.4.2. *Suppose that the utility functions are given by (2.7) and* $\alpha = \beta = h_0 = 0$*. For the "no habit" agent, the optimal value function is*

$$\widetilde{V}(t, x) = -\frac{1}{\gamma} e^{-\gamma(\widetilde{a}(t)x + \widetilde{g}(t))},$$

the optimal indemnity is

$$\widetilde{I}_t(Y) = \Big[Y - \widetilde{d}(t)\Big]^+,$$

and the optimal consumption is

$$\widetilde{c}_t = -\frac{\ln(\widetilde{a}(t))}{\gamma} + \widetilde{a}(t)\widetilde{X}_t + \widetilde{g}(t),$$

where $\widetilde{X}_t$ is the optimal wealth process in the absence of habit formation. Here,

$$\begin{aligned}
\widetilde{a}(t) &= 1/\widetilde{J}(t),\\
\widetilde{J}(t) &= e^{-r(T-t)}\frac{r-1}{r} + \frac{1}{r},\\
\widetilde{d}(t) &= \frac{\ln(1+\theta)}{\gamma\widetilde{a}(t)},\\
\widetilde{g}(t) &= -\frac{\ln w}{\gamma}e^{-\int_t^T \widetilde{a}(s)ds}\\
&\quad + \int_t^T e^{-\int_t^u \widetilde{a}(s)ds}\Bigg\{\frac{\lambda+\delta}{\gamma} + [\ln(\widetilde{a}(u)) - 1]\frac{\widetilde{a}(u)}{\gamma}\\
&\quad -\lambda(1+\theta)\Bigg[\int_{\frac{\ln(1+\theta)}{\gamma\widetilde{a}(u)}}^{\infty} ydF(y) - \frac{\ln(1+\theta)}{\gamma\widetilde{a}(u)}\overline{F}\Bigg(\frac{\ln(1+\theta)}{\gamma\widetilde{a}(u)}\Bigg)\Bigg]\widetilde{a}(u)\\
&\quad -\frac{\lambda}{\gamma}\Bigg[(1+\theta)\overline{F}\Big(\frac{\ln(1+\theta)}{\gamma\widetilde{a}(u)}\Big) + \int_0^{\frac{\ln(1+\theta)}{\gamma\widetilde{a}(u)}} e^{\gamma\widetilde{a}(u)y}dF(y)\Bigg]\Bigg\}\,du. \qquad (2.13)
\end{aligned}$$

The following corollary presents the properties of the "no habit" agent's optimal deductible.

Corollary 2.4.2. *For the "no habit" agent, the optimal deductible $\widetilde{d}(t)$ is strictly decreasing in $[0, T)$. Moreover, $\widetilde{d}(t) < d(t)$ for $t \in [0, T)$ and $\widetilde{d}(T) = d(T) = \frac{\ln(1+\theta)}{\gamma}$, where $d(t)$ is the optimal deductible in the presence of habit formation.*

This corollary attests that the individual gradually increases their insurance coverage even in the absence of habit formation. Moreover, the "no habit" agent optimally chooses a lower level of deductible than the individual with habit formation at each point in time. In other words, the presence of habit formation reduces the individual's insurance coverage. We point out that this result is in sharp contrast to Ben-Arab et al. (1996) who show that the level of (proportional) insurance purchased is higher with habit formation than without habit formation. In the next subsection, we will provide a formal explanation accounting for the puzzling phenomenon. It can be shown that even if we consider consumption habit in the bequest utility (which follows the general setting in Boyle et al. (2022)), a similar phenomenon still presents. This implies that the "no consumption habit" in the bequest utility is not the contributing factor for this phenomenon.

To conclude this subsection, we emphasize that the usefulness of our model can be further highlighted by its potential ability to explain the prevailing underinsurance phenomenon as pointed out in (Lloyd's, 2018). Some well-known explanations include affordability, education of risk, significant economic growth, and changing risk landscape. Our results suggest that habit formation may be another contributing factor.

2.4.2 Optimal proportional insurance

We have shown that the optimal insurance must be deductible insurance under the expected premium principle. Because proportional insurance is also common in the market, we re-examine the optimal insurance problem but assuming the individual is restricted to purchase only proportional insurance. In other words, the economic agent now chooses the proportion of the loss insured instead of the indemnity function. This formulation allows us to better contrast our results with those in Ben-Arab et al. (1996).

Proposition 2.4.3. *Suppose that the individual can only purchase proportional insurance, i.e. $I_t(Y) = p(t)Y$, where $p(t) \in [0,1]$ is the proportion of loss covered by the insurer at time t, should a loss incur at that time, and that the utility functions are given by (2.7). The optimal value function is*

$$V^p(t,x,h) = -\frac{1}{\gamma}e^{-\gamma(a(t)x+b(t)h+g^p(t))},$$

the optimal consumption is

$$c_t^p = -\frac{1}{\gamma}\ln[a(t)-\alpha b(t)] + a(t)X_t^p + (b(t)+1)h_t^p + g^p(t),$$

where X_t^p is the optimal wealth process of the proportional insurance in the presence of habit formation and h_t^p is the optimal consumption habit process.

The optimal proportion $p^*(t)$ is the unique solution to

$$E[e^{\gamma a(t)(1-p^*(t))Y}Y] = (1+\theta)E[Y], \quad \textit{if } E[e^{\gamma a(t)Y}Y] > (1+\theta)E[Y],$$

and $p^*(t) = 0$ otherwise.

The functions $a(t)$ and $b(t)$ are as in Proposition 2.4.1, and $g^p(t)$ takes the following form

$$\begin{aligned} g^p(t) = &-\frac{\ln w}{\gamma} e^{-\int_t^T a(s)-\alpha b(s) ds} \\ &+ \int_t^T e^{-\int_t^u a(s)-\alpha b(s) ds} \left\{ \frac{\lambda+\delta}{\gamma} + [\ln(a(u)-\alpha b(u)) - 1] \frac{a(u)-\alpha b(u)}{\gamma} \right. \\ &\left. -\lambda(1+\theta)p^*(u)E[Y]a(u) - \frac{\lambda}{\gamma}E[e^{\gamma a(u)(1-p^*(u))Y}] \right\} du. \end{aligned} \tag{2.14}$$

Similarly to the case of deductible insurance, the optimal proportion of loss insured is increasing in risk aversion γ and decreasing in the safety loading (premium loading) as long as the individual purchases the proportional insurance.

The following corollary concerns the time pattern of the optimal proportion.

Corollary 2.4.3. *Assume that $\beta > \alpha \geq 0$ and $0 < r < 1$. There are three cases.*

1. *If $E[e^{\gamma a(0)Y}Y] > (1+\theta)E[Y]$, $p^*(t)$ is positive and strictly increasing in $[0,T]$.*
2. *If $E[e^{\gamma a(0)Y}Y] \leq (1+\theta)E[Y]$ and $E[e^{\gamma a(T)Y}Y] > (1+\theta)E[Y]$, then there exists a unique $t_0 \in [0,T)$ such that $E[e^{\gamma a(t_0)Y}Y] = (1+\theta)E[Y]$. For $t \in [0,t_0]$, $p^*(t) = 0$. For $t \in (t_0,T]$, $p^*(t)$ is positive and strictly increasing in $[0,T]$.*
3. *If $E[e^{\gamma a(T)Y}Y] \leq (1+\theta)E[Y]$, $p^*(t) = 0$ for all $t \in [0,T]$.*

Similarly to the case of deductible insurance, the optimal proportion $p^*(t)$ is non-decreasing in age and the individual becomes increasingly risk-averse towards the insurable risk. However, when the individual is restricted to proportional insurance, it is sometimes optimal for the young not to buy any insurance. If an individual chooses not to buy any insurance at an early age, then he or she will purchase insurance only if he or she is old enough. In the extreme case, the individual will find it optimal not to purchase any insurance throughout life. In particular, when θ, the safety loading in the premium principle, is large enough, the economic agent finds the insurance is too costly and prefers to self-insure.

To better illustrate the effects of habit formation on the economic agent's demand for proportional insurance, we next consider the optimization problem without habit formation.

Proposition 2.4.4. *Suppose that the individual can only purchase proportional insurance, i.e.* $I_t(Y) = p(t)Y$*, where* $p(t) \in [0,1]$ *is the proportion of the loss covered by the insurer at time* t*, should a loss incur at that time, and that the utility functions are given by (2.7) and* $\alpha = \beta = h_0 = 0$*. For the "no habit" agent, the optimal value function is*

$$\widetilde{V}^p(t,x) = -\frac{1}{\gamma} e^{-\gamma(\widetilde{a}(t)x + \widetilde{g}^p(t))},$$

the optimal consumption is

$$\widetilde{c}_t^p = -\frac{\ln(\widetilde{a}(t))}{\gamma} + \widetilde{a}(t)\widetilde{X}_t^p + \widetilde{g}^p(t),$$

where $\widetilde{X}_t^p$ *is the optimal wealth process of the proportional insurance in the absence of habit formation.*
The optimal indemnity proportion $\widetilde{p}^*(t)$ *is the unique solution to*

$$E[e^{\gamma\widetilde{a}(t)(1-\widetilde{p}^*(t))Y} Y] = (1+\theta)E[Y], \ \text{if } E[e^{\gamma\widetilde{a}(t)Y} Y] > (1+\theta)E[Y], \tag{2.15}$$

and $\widetilde{p}^*(t) = 0$ *otherwise.*

The functions $\widetilde{a}(t)$ *is as in Proposition 2.4.2 and* $\widetilde{g}^p(t)$ *takes the following form*

$$\begin{aligned} \widetilde{g}^p(t) \quad = \quad & -\frac{\ln w}{\gamma} e^{-\int_t^T \widetilde{a}(s)ds} \\ & + \int_t^T e^{-\int_t^u \widetilde{a}(s)ds} \left\{ \frac{\lambda+\delta}{\gamma} + [\ln(\widetilde{a}(u)) - 1] \frac{\widetilde{a}(u)}{\gamma} \right. \\ & \left. -\lambda(1+\theta)\widetilde{p}^*(u)E[Y]\widetilde{a}(u) - \frac{\lambda}{\gamma} E[e^{\gamma\widetilde{a}(u)(1-\widetilde{p}^*(u))Y}] \right\} du, \end{aligned} \tag{2.16}$$

The following corollary presents the properties of the "no habit" agent's optimal proportional insurance.

Corollary 2.4.4. *Assume that* $0 < r < 1$*.* $\widetilde{p}^*(t)$ *is non-decreasing and there are three cases.*

1. *If* $E[e^{\gamma\widetilde{a}(0)Y} Y] > (1+\theta)E[Y]$*,* $\widetilde{p}^*(t)$ *is positive and strictly increasing in* $[0,T]$*.*

2. *If* $E[e^{\gamma\widetilde{a}(0)Y}Y] \leq (1+\theta)E[Y]$ *and* $E[e^{\gamma\widetilde{a}(T)Y}Y] > (1+\theta)E[Y]$*, then there exists a unique* $\widetilde{t}_0 \in [0,T)$ *such that* $E[e^{\gamma\widetilde{a}(\widetilde{t}_0)Y}Y] = (1+\theta)E[Y]$*. For* $t \in [0,\widetilde{t}_0]$*,* $\widetilde{p}^*(t) = 0$*. For* $t \in (\widetilde{t}_0, T]$*,* $\widetilde{p}^*(t)$ *is positive and strictly increasing in* $[0,T]$*.*

3. *If* $E[e^{\gamma\widetilde{a}(T)Y}Y] \leq (1+\theta)E[Y]$*,* $\widetilde{p}^*(t) = 0$ *for all* $t \in [0,T]$*.*

Moreover, $\widetilde{p}^*(T) = p^*(T)$*. For* $t \in [0,T)$*,* $\widetilde{p}^*(t) \geq p^*(t)$ *and the inequality is strict whenever* $\widetilde{p}^*(t) > 0$ *or equivalently* $E[e^{\gamma\widetilde{a}(t)Y}Y] > (1+\theta)E[Y]$*.*

There are several interesting observations in Corollary 2.4.4. First, the individual gradually increases the proportional insurance coverage even in the absence of habit formation, which is consistent with the setting of deductible insurance. Second, the "no habit" agent may also optimally not purchase any insurance if he or she is constrained to proportional insurance. Third, the "no habit" agent's proportional insurance coverage is no less than that with habit formation. Moreover, habit formation strictly reduces the individual's proportional insurance coverage whenever he or she would have engaged in proportional insurance without habit formation. This result is in sharp contrast to Ben-Arab et al. (1996) as previously observed for the deductible case. We now formally address this puzzling phenomenon.

Recall that there are some material differences (in terms of problem formulation and model setup) between the model of Ben-Arab et al. (1996) and the present chapter. Some of these differences include the assumptions on the form of insurance, loss modeling, utility function, and objective function with or without bequest. It turns out that a possible reason accounting for the conflicting insurance demand phenomenon for an agent with and without habit formation lies in the choice of the utility function. This can be explained via the Arrow-Pratt measure of risk aversion.

For exponential utility under the optimal deductible insurance, the Arrow-Pratt absolute risk aversion for the value functions (the indirect utility functions) from Propositions 2.4.1 and 2.4.2 can be shown to have the following representations:

$$
\begin{aligned}
A(x) &= -\frac{V_{xx}}{V_x} = \gamma a(t), \\
\widetilde{A}(x) &= -\frac{\widetilde{V}_{xx}}{\widetilde{V}_x} = \gamma \widetilde{a}(t).
\end{aligned}
$$

From Corollaries 2.4.1 and 2.4.2, we know $a(t)$ and $\widetilde{a}(t)$ are increasing w.r.t. time since $J(t)$ is decreasing w.r.t. time and $a(t) = 1/J(t)$ (similar reasoning for $\widetilde{a}(t)$). Together with the assumption that $\gamma > 0$, the increasing property of the Arrow-Pratt absolute risk aversion implies that the individual will need more insurance coverage as he or she ages. Moreover, the relation $a(t) < \widetilde{a}(t)$ from Corollary 2.4.2 leads to $A(x) < \widetilde{A}(x)$, i.e. compared to the "no habit" agent, the agent with habit formation exhibits lower risk aversion, which in turn translates to lower insurance coverage. Similar results hold for the proportional insurance.

For the model of Ben-Arab et al. (1996) with power utility functions, the authors derive an explicit representation for the Arrow-Pratt relative risk aversion of the indirect utility function (see Ben-Arab et al., 1996, Equation (17)). In this case, the Arrow-Pratt relative risk aversion under habit formation is increasing over time and is higher than the corresponding Arrow-Pratt relative risk aversion for the "no habit" agent. These properties justify that the agent becomes more risk averse as he or she ages, and more importantly, the higher Arrow-Pratt relative risk aversion for the habit formation agent, as compared to the "no habit" agent, induces greater insurance coverage, thereby reconciling the conflicting results of the model of Ben-Arab et al. (1996) and the model in this chapter.

To conclude this section, we would like to relate our results to the prevailing global underinsurance documented in Lloyd's (2018). We have shown in Corollary 2.4.2 that the presence of habit formation reduces insurance coverage when the economic agent is free to choose the form of the indemnity. Corollary 2.4.4 indicates that this result still holds even if the economic agent is confined to only proportional insurance. A comparison between Corollaries 2.4.1, 2.4.2 and Corollaries 2.4.3, 2.4.4 reveals another interesting fact. The individual who purchases only proportional insurance may optimally choose not to maintain any insurance coverage while he or she would maintain at least some level of coverage in a complete insurance market in which any form of insurance contract can be offered. Therefore, the inability of consumers to choose customized insurance contracts due to the incompleteness of the insurance market or the lack of bargaining power may be another contributing factor to global underinsurance, especially in the developing countries.

2.5 Welfare loss from suboptimal strategies

We have shown that the optimal insurance contract is deductible insurance if customization is allowed and the expected premium principle is used. Therefore, any individual who chooses non-deductible insurance, such as proportional insurance or no insurance coverage, will have a lower level of utility and suffers a welfare loss from suboptimal strategies. In this section, we quantify such welfare losses.

2.5.1 Loss from proportional insurance

Under habit formation, the value function of the economic agent who chooses a deductible contract is $V(t,x,h)$, which is the maximum utility he or she could achieve. If the individual is restricted to proportional insurance, the maximum utility is $V^p(t,x,h)$. Clearly, $V(t,x,h) \geq V^p(t,x,h)$. Calculating the difference of value functions $V^p(t,x,h) - V(t,x,h)$ provides one way of quantifying the utility loss from using proportional insurance. In this chapter, we consider another way of analyzing the welfare loss. We do this via the principle of certainty equivalence as in Liu and Pan (2003), Larsen and Munk (2012), Xue et al. (2019), and Tan et al. (2020).

Let z_p be the certainty-equivalent wealth of welfare loss that satisfies the following equation:

$$V(t,x,h) = V^p(t,x+z_p,h). \tag{2.17}$$

In other words, z_p is the additional wealth that the individual who relies only on proportional insurance needs to hold, in order to yield the same level of expected utility as the individual who purchases deductible insurance, with the same age, wealth and consumption habit. Therefore, the incremental value z_p can be interpreted as the "value" of deductible insurance over proportional insurance. In general, the welfare loss z_p should depend on (t,x,h) as people with different ages, levels of wealth and consumption habit could have different losses. However, Proposition 2.5.1 below asserts that z_p is a function of t and does not depend on wealth or habit. This is a consequence of using exponential utility which ignores the wealth effect.

Proposition 2.5.1.

$$\begin{aligned} z_p(t) \;=\; & \frac{1}{a(t)} \int_t^T e^{-\int_t^u a(s)-\alpha b(s)ds} \Bigg\{ \lambda(1+\theta)p(u)E[Y]a(u) + \frac{\lambda}{\gamma} E[e^{\gamma a(u)(1-p(u))Y}] \\ & -\lambda(1+\theta)\left[\int_{\frac{\ln(1+\theta)}{\gamma a(u)}}^{\infty} dF(y) - \frac{\ln(1+\theta)}{\gamma a(u)} \overline{F}\left(\frac{\ln(1+\theta)}{\gamma a(u)}\right)\right] a(u) \\ & -\frac{\lambda}{\gamma}\left[(1+\theta)\overline{F}\Big(\frac{\ln(1+\theta)}{\gamma a(u)}\Big) + \int_0^{\frac{\ln(1+\theta)}{\gamma a(u)}} e^{\gamma a(u) y} dF(y)\right]\Bigg\} du, \end{aligned} \tag{2.18}$$

where $a(t)$ is given by (A.10).

Steffensen and Thøgersen (2019) define the welfare loss z_p through

$$V(t,x-z_p,h) = V^p(t,x,h).$$

Due to the absence of the wealth effect, it is easy to see that this definition is actually identical to (2.17).

Given the complex structure of (2.18), we will not attempt to illustrate here how the welfare loss depends on various parameters. Instead, we provide numerical illustrations in Section 2.6.

2.5.2 Loss from no insurance

We are also interested in the welfare loss if the individual does not have access to the insurance market. To this end, we need to solve the individual's optimization problem assuming $I(Y) \equiv 0$.

Proposition 2.5.2. *Suppose that the individual does not have access to the insurance market, i.e. $I_t(Y) \equiv 0$, and that the utility functions are given by (2.7). The optimal value function is*

$$V^0(t,x,h) = -\frac{1}{\gamma} e^{-\gamma(a(t)x + b(t)h + g^0(t))},$$

and the optimal consumption is

$$c_t^0 = -\frac{1}{\gamma}\ln[a(t) - \alpha b(t)] + a(t)X_t^0 + (b(t)+1)h_t^0 + g^0(t),$$

where the functions $a(t)$ and $b(t)$ are as in Proposition 2.4.1, and $g^0(t)$ takes the form

$$\begin{aligned} g^0(t) &= -\frac{\ln w}{\gamma} e^{-\int_t^T a(s)-\alpha b(s) ds} \\ &\quad + \int_t^T e^{-\int_t^u a(s)-\alpha b(s) ds}\left\{\frac{\lambda+\delta}{\gamma} + [\ln(a(u) - \alpha b(u)) - 1]\frac{a(u)-\alpha b(u)}{\gamma}\right. \\ &\quad \left. -\frac{\lambda}{\gamma}\int_0^\infty e^{\gamma a(u) y} dF(y)\right\} du, . \end{aligned} \tag{2.19}$$

It is obvious that $V(t,x,h) \geq V^p(t,x,h) \geq V^0(t,x,h)$. Let z_0 be the certainty-equivalent wealth of welfare loss that satisfies the following equation:

$$V(t,x,h) = V^0(t, x + z_0, h). \tag{2.20}$$

In other words, z_0 is the additional wealth the individual without access to the insurance market needs to hold, in order to yield the same level of expected utility of the individual in a complete insurance market, with the same age, wealth and consumption habit. Therefore, the incremental value z_0 measures the "value" of deductible over no insurance coverage.

Proposition 2.5.3.

$$\begin{aligned} z_0(t) &= \frac{1}{a(t)}\int_t^T e^{-\int_t^u a(s)-\alpha b(s)ds}\Bigg\{\frac{\lambda}{\gamma}\int_0^\infty e^{\gamma a(u)y}dF(y) \\ &\quad -\lambda(1+\theta)\Bigg[\int_{\frac{\ln(1+\theta)}{\gamma a(u)}}^\infty ydF(y)-\frac{\ln(1+\theta)}{\gamma a(u)}\overline{F}\Bigg(\frac{\ln(1+\theta)}{\gamma a(u)}\Bigg)\Bigg]a(u) \\ &\quad -\frac{\lambda}{\gamma}\Bigg[(1+\theta)\overline{F}\Big(\frac{\ln(1+\theta)}{\gamma a(u)}\Big)+\int_0^{\frac{\ln(1+\theta)}{\gamma a(u)}} e^{\gamma a(u)y}dF(y)\Bigg]\Bigg\}du. \end{aligned} \tag{2.21}$$

We claim that the linearity holds between the welfare losses with different decisions; i.e. the difference between the welfare loss from proportional insurance and that from no insurance is the welfare loss between the two decisions. From the definitions of our welfare losses, we have

$$V(t,x,h)=V^p(t,x+z_p,h)=V^0(t,x+z_0,h).$$

Denote $\bar{x}=x+z_p$, then $x=\bar{x}-z_p$ and

$$V^p(t,\bar{x},h)=V^0(t,\bar{x}+z_0-z_p,h).$$

We see that the linearity of the welfare losses holds between optimal and suboptimal decisions.

In the next section, we will illustrate the welfare losses through numerical examples.

2.6 Numerical illustrations

In this section, we present several numerical examples to highlight the findings of our proposed life-cycled model. We divide this section into two parts. The first part focuses on the effects of habit formation on the demand for insurance. We make comparisons between the optimal deductible and proportional insurance with and without habit formation. In the second part, we analyze welfare losses arising from suboptimal strategies. We also perform a sensitivity analysis of welfare losses with respect to various parameters.

We assume that the loss size Y follows an exponential distribution with mean $1/\eta$. We set the parameters according to Table 2.1 unless otherwise stated. The parameters we have assumed are similar to those in Kraft et al. (2017); Steffensen and Thøgersen (2019).

Table 2.1: **Parameters for the base scenario**

Loss frequency λ	0.100	Subjective discount rate δ	0.100
Loss severity η	1.000	Risk aversion γ	0.500
Premium loading θ	0.250	Strength of bequest w	0.500
Habit intensity α	0.100	Risk free rate r	0.020
Habit persistence β	0.174	Terminal time T	40.000

2.6.1 Optimal deductible and optimal proportional insurance

Our first focus is on the demand for insurance. Figure 2.1 displays the optimal deductible in various settings. Consistent with Corollaries 2.4.1 and 2.4.2, the optimal deductible is decreasing in age, regardless of habit formation, and habit formation increases the level of deductible. Moreover, the optimal deductible is increasing in α and decreasing in β. As the interest rate increases from 0.02 to 0.1, the optimal deductible decreases.

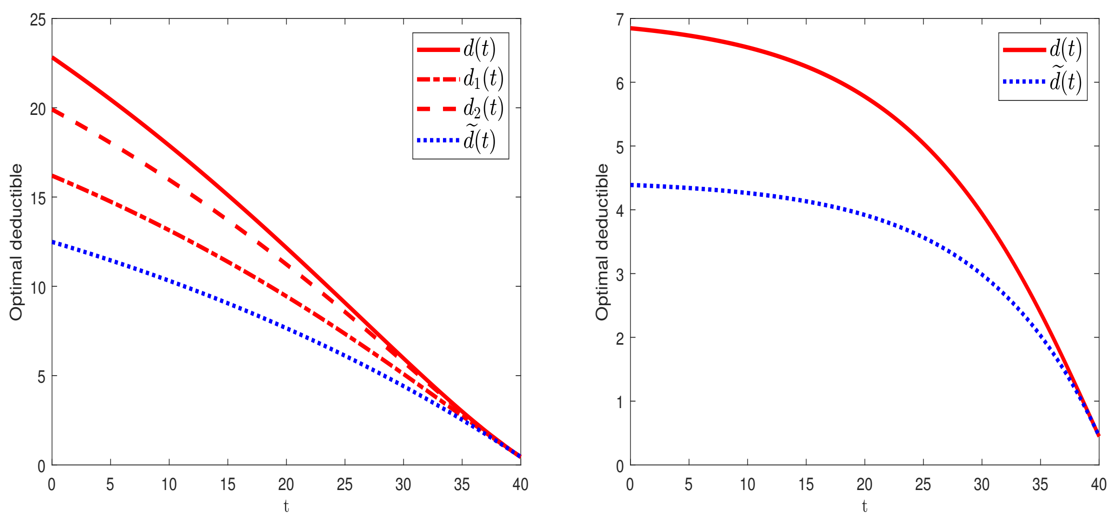


Figure 2.1: Optimal deductibles with and without consumption habit. $d(t)$ (the solid line) is the optimal deductible with consumption habit under $\alpha = 0.1$, $\beta = 0.174$. $d_1(t)$ (the dash-dot line) is the optimal deductible with consumption habit under $\alpha = 0.05$, $\beta = 0.174$. $d_2(t)$ (the dashed line) is the optimal deductible with consumption habit under $\alpha = 0.1$, $\beta = 0.224$. $\widetilde{d}(t)$ (the dotted line) is the optimal deductible without consumption habit. The risk free rate r equals 0.02 in the left figure and 0.1 in the right figure.

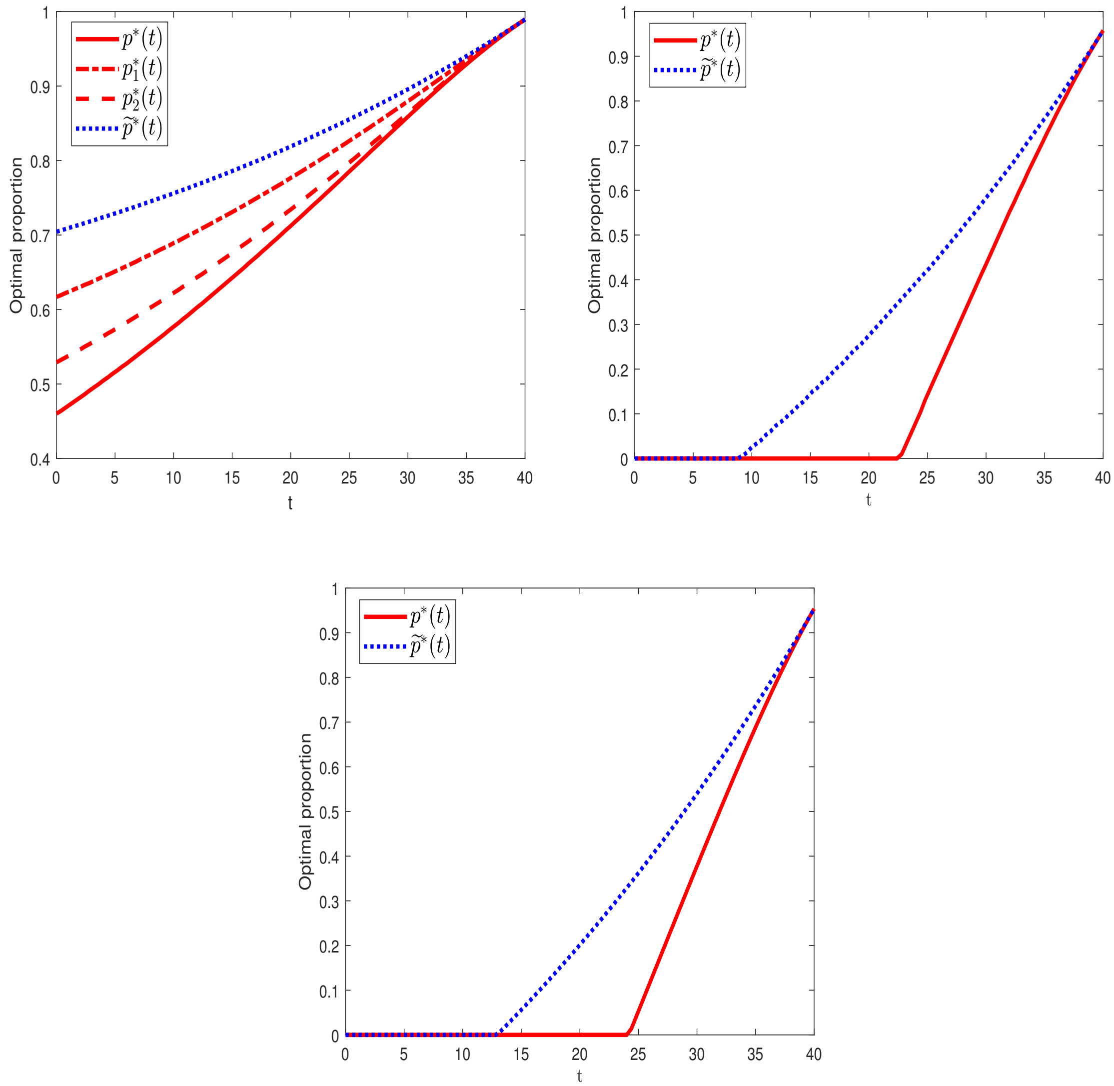


Figure 2.2: Optimal proportions with and without consumption habit. $p^*(t)$ (the solid line) is the optimal proportion with consumption habit $\alpha = 0.1$, $\beta = 0.174$. $p_1^*(t)$ (the dash-dot line) is the optimal proportion with consumption habit under $\alpha = 0.05$, $\beta = 0.174$. $p_2^*(t)$ (the dashed line) is the optimal proportion with consumption habit under $\alpha = 0.1$, $\beta = 0.224$. $\widetilde{p}^*(t)$ (the dotted line) is the optimal proportion without consumption habit. $\eta = 0.05$, $\theta = 0.25$ in the first figure. $\eta = 0.2$, $\theta = 0.25$ in the second figure. And $\eta = 0.05$, $\theta = 2.5$ in the third figure.

Figure 2.2 compares the optimal proportional insurance in different scenarios. As predicted by Corollaries 2.4.3 and 2.4.4, both $p^*(t)$ and $\widetilde{p}^*(t)$ are non-decreasing, and $p^*(t)$ is lower than $\widetilde{p}^*(t)$ before reaching the time horizon. Similarly to the case of deductible insurance, the optimal proportion of losses insured is decreasing in α and increasing in β. When the loss severity increases from 0.01 to 0.2, i.e., the expected loss size decreases, the optimal proportion diminishes or totally disappears especially in the early years. In other words, the individual's willingness to purchase proportional insurance is positively correlated with the expected loss size. As the safety loading θ increases from 0.25 to 2.5, the insurance becomes much more expensive and the insurance demand decreases significantly.

These results suggest that economic agents may opt out of the (proportional) insurance market if the expected loss size is small, or the insurance premium is high, thereby offering potential explanations for global underinsurance.

2.6.2 The welfare loss

Our second focus is on the welfare losses from sub-optimal insurance strategies. Following Steffensen and Thøgersen (2019), we define the relative welfare loss. The net premium for full insurance of the non-life risk of the individual is $\xi = \lambda/\eta$, and relative to this, the relative welfare losses are $z_0(t)/\xi$ and $z_p(t)/\xi$. We plot the relative welfare losses from no insurance and proportional insurance in Figure 2.3. There are several interesting observations.

First, the relative welfare loss from no insurance coverage is generally much larger than that from the proportional insurance which is typically small. This is because, when the insurance demand is high, the individual with proportional insurance can still adjust its risk exposure, although not in an ideal way, while the individual without access to the insurance market has to bear the loss in full.

Second, both relative welfare losses are hump-shaped during the individual's lifetime. Recall that the welfare loss measures the cumulative utility cost of not implementing the optimal insurance strategy until T. Therefore, there are two primary factors contributing to the welfare loss. On the one hand, a larger deviation from the optimal strategy incurs more losses. On the other hand, as the individual approaches the time horizon T, he or she has less to lose from suboptimal strategies. In fact, such losses must be zero at the terminal time. The overall effect is the hump-shape pattern. As illustrated in Figures 2.1 and 2.2, the insurance demand is increasing in age, and the discrepancy between optimal and suboptimal strategies increases. Therefore, the welfare losses first increase with respect to age, reaching the maximum after around 35 years. In the later years, the time effect is more pronounced and the welfare losses decrease rapidly to 0 at the terminal time.

Third, a comparison between the left and right panel of Figure 2.3 reveals the impact of interest rate on the welfare losses. As the interest rate increases from 0.02 to 0.1, both relative welfare losses from no insurance and proportional insurance decrease significantly.

The left panel of Figure 2.3 also illustrates that the welfare losses are relatively insensitive to the habit formation parameters. These results suggest that maintaining appropriate insurance coverage is critical to the welfare of individuals in mid-old ages, especially in an economy with a low interest rate. Moreover, the welfare losses from proportional insurance are relatively small.

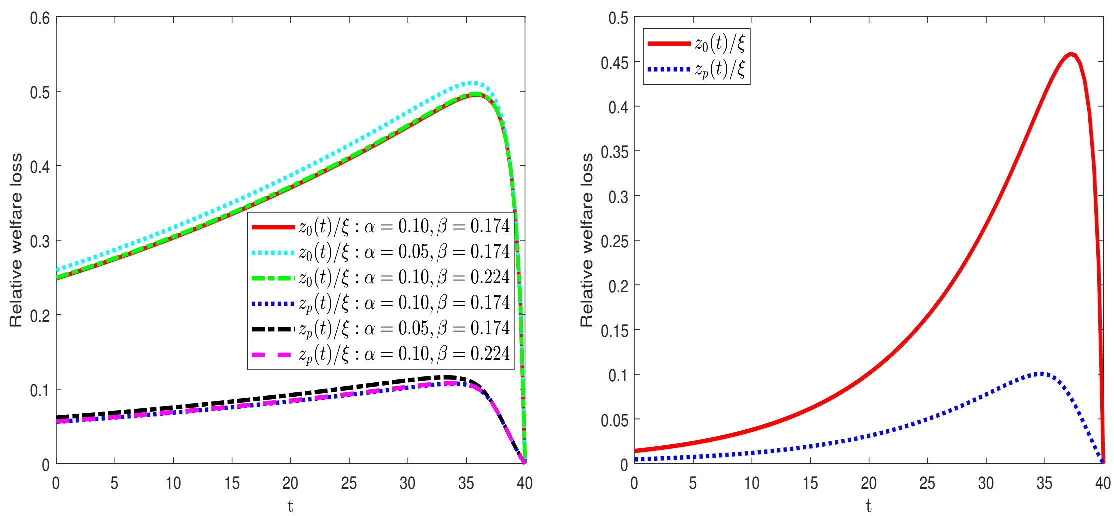


Figure 2.3: The relative welfare losses from no insurance and proportional insurance. $z_0(t)/\xi$ (the solid line) is the relative welfare loss from no insurance. $z_p(t)/\xi$ (the dotted line) is the relative welfare loss from proportional insurance. The risk free rate r equals 0.02 in the left figure and 0.1 in the right figure.

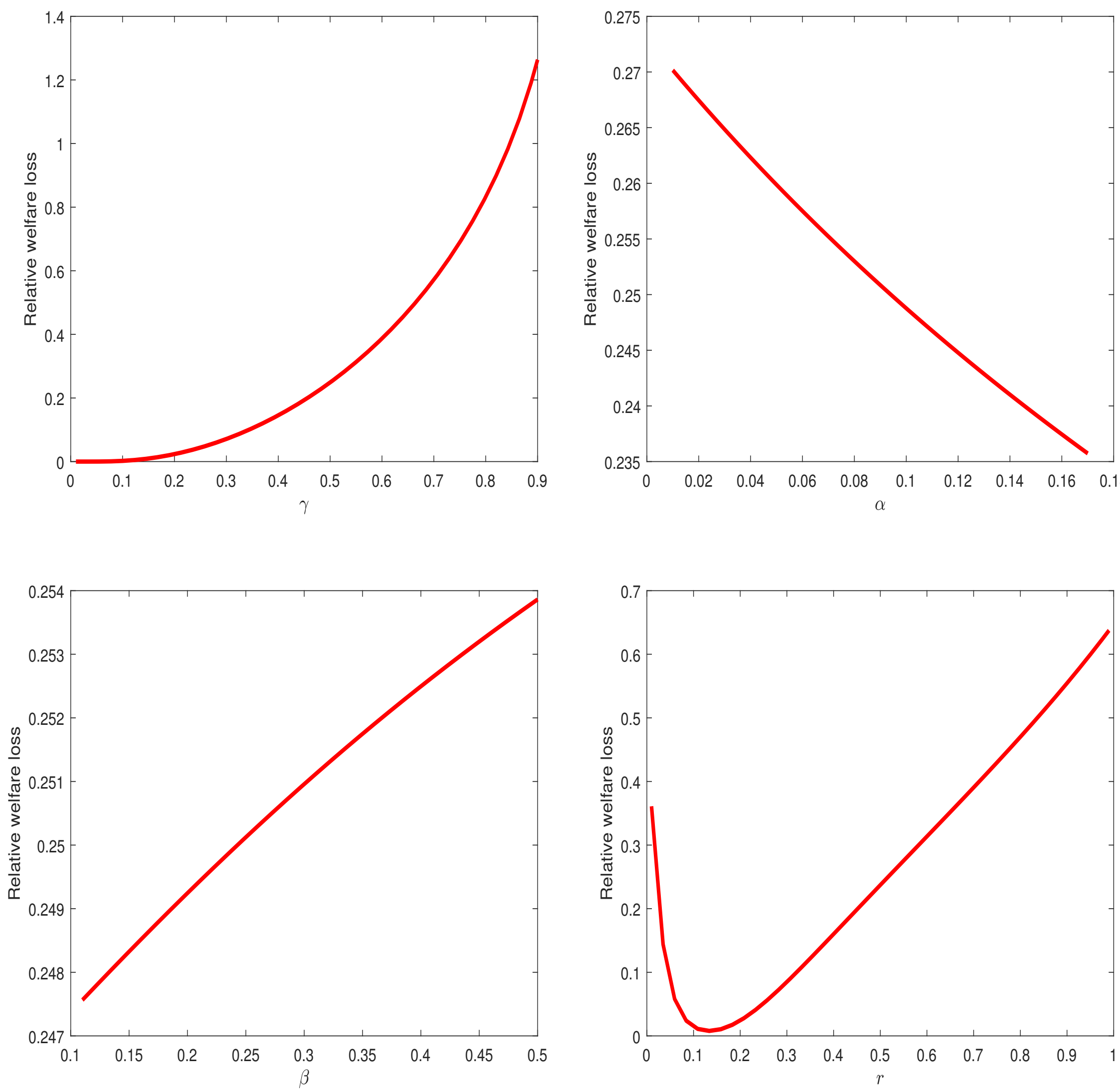
We next turn our attention to the effects of various parameters on the welfare losses. Figure 2.4 plots $z_0(0)/\xi$, the relative welfare loss from no insurance coverage at time 0 against γ, α, β and r, respectively. The welfare loss is increasing in risk aversion γ, which is consistent with intuition. As the individual becomes more risk-averse, the insurance demand rises, and thus the welfare loss from no insurance coverage increases. As confirmed in Figure 2.4, the demand for deductible insurance is decreasing in α and increasing in β.

Therefore, the welfare loss from no insurance is decreasing in α and increasing in β. Last, the welfare loss against the interest rate exhibits a U-shaped relationship. The welfare loss decreases initially as the interest rate increases and reaches its lowest value when the interest rate is around 15%. The welfare loss starts to increase as the interest rate further increases. It should be pointed out that such high interest rates are rather uncommon, especially in the recent economy.

Figure 2.5 plots $z_p(0)/\xi$, the relative welfare loss from proportional insurance at time 0 against γ, α, β and r, respectively. Similarly to $z_0(0)/\xi$, $z_p(0)/\xi$ in increasing in γ, decreasing in α, and increasing in β. However, as the risk-free rate r increases, the welfare loss from proportional insurance first decreases, then increases, and finally decreases. While the welfare loss exhibits a non-linear relationship with the interest rates, it is reasonable to assume that practically the welfare loss will mostly be decreasing due to the low interest rate environment.

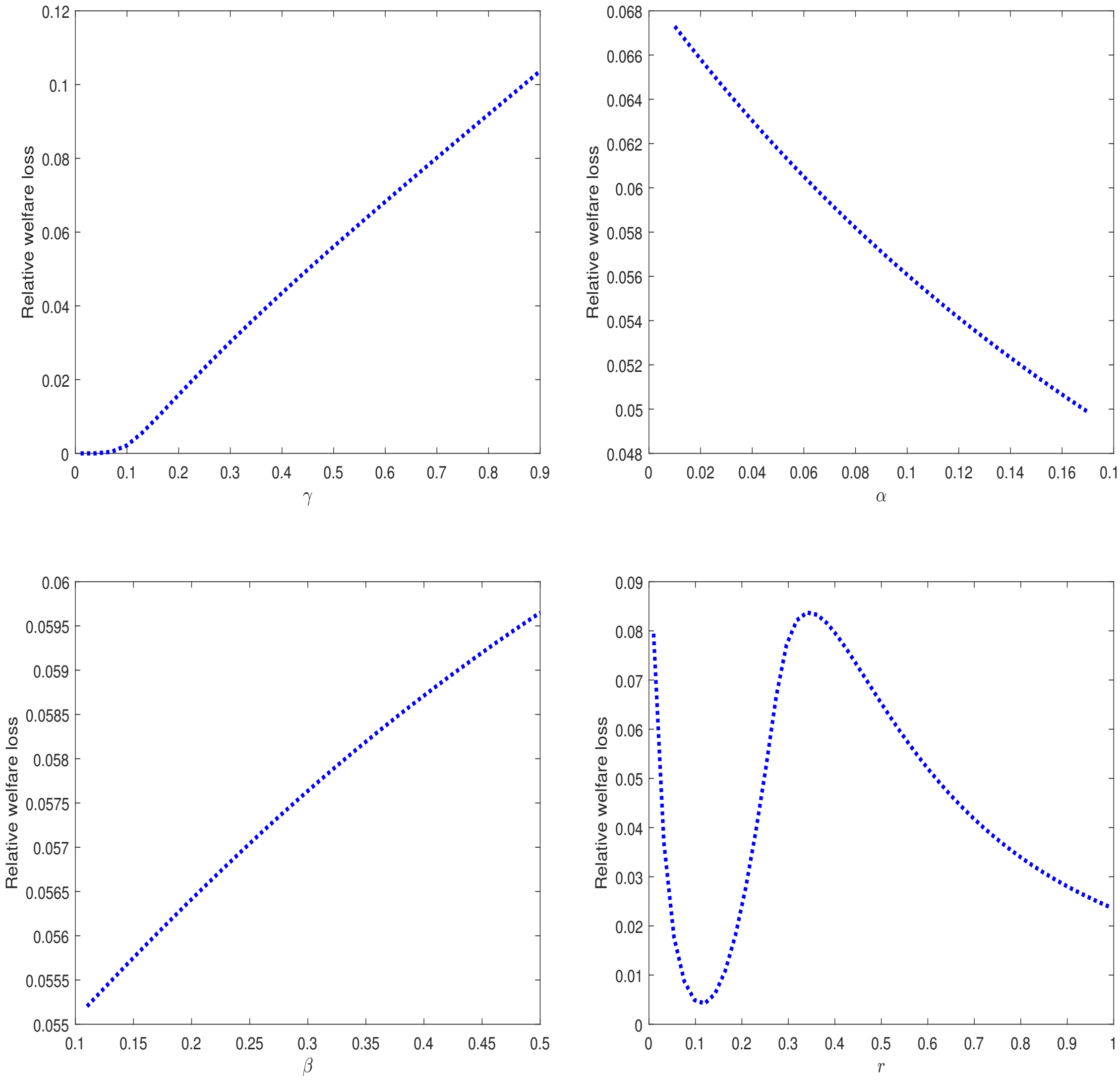
A further comparison between Figures 2.4 and 2.5 reveals that the welfare loss from no insurance coverage is much larger than that from proportional insurance. Moreover, both welfare losses are more sensitive to risk aversion and interest rate than the habit formation parameters.

These results suggest that individuals should maintain sufficient insurance coverage, even if only proportional insurance is available, especially for high risk-averse individual and low interest rate. Moreover, the welfare losses from suboptimal insurance strategies are relatively insensitive to habit formation, although it has a large impact on the economic agent's insurance demand.



.

Figure 2.4: The sensitivity of relative welfare loss $z_0(0)/\xi$ with respect to γ, α, β and r.



.

Figure 2.5: The sensitivity of relative welfare loss $z_p(0)/\xi$ with respect to γ, α, β and r.

2.7 Conclusion

In this chapter, we proposed and solved the optimal insurance problem for an individual exhibiting internal habit formation. Under general utilities, we established that the optimal per-claim insurance must be deductible, provided that the expected principle is used in insurance pricing.

We obtained explicit solutions for individuals who can purchase deductible or proportional insurance under the exponential utility. For both types of insurance, the individual gradually increases insurance coverage as he or she ages. Moreover, the presence of habit formation reduces insurance coverage and that an individual who is restricted to proportional insurance may opt out of the insurance market, especially in early ages. These results suggest that habit formation and incomplete insurance market (such that individuals can only purchase proportional insurance) can partially contribute to explaining the prevailing global underinsurance phenomenon, as documented in (Lloyd's, 2018).

We further used numerical examples to investigate the impact of various parameters on the insurance demand and the welfare losses from suboptimal strategies, such as the individual cannot purchase any insurance or is restricted to only proportional insurance. Although habit formation has a large impact on the demand for insurance, its effects on welfare losses are negligible compared to age, risk-aversion, and the interest rate. Moreover, maintaining enough insurance coverage, even if only proportional insurance is offered, is more important to the welfare of individuals in mid-old ages, with high risk-aversion, and in an economy with a low interest rate.

Chapter 3

Optimal defined-contribution pension management with financial and mortality risks

3.1 Introduction

Due to the increasing population aging trend, pension management has become one of the major challenges in the actuarial field in the past decades. Most pension funds can be classified into two schemes: defined-benefit (DB) pension schemes and defined-contribution (DC) pension schemes in different jurisdictions worldwide. Generally speaking, an individual in a DB pension plan is provided with a guaranteed retirement benefit related to his or her earnings history, tenure of service, and age. By contrast, the individual in a DC pension plan contributes a predetermined amount during the fund's accumulation period and receives the retirement benefit based on the investment earnings. Over past decades, DC pension plans have surpassed DB pension plans in many ways on both demand and supply sides. On the demand side, modern industrial development increases the mobility of workers between employers. Employees could earn less benefit from changing jobs based on the benefit criteria in the DB schemes. On the supply side, individuals tend to live longer with the enhancement of life quality, which adds great pressure to the DB plan costs (see Broadbent et al., 2006). As a result, there has been a rapid shift from DB to DC schemes worldwide. According to the latest OECD's report, less than 50% of pension assets are managed in the DB's scheme in 28 out of 33 reporting jurisdictions in 2019, in some countries (such as countries in Latin America and Central and Eastern Europe) DB plans do not even exist (see OECD, 2020).

Compared with other funds, the long-term horizon is one of the key features of pension plan management. The accumulation period of a DC pension plan usually lasts for 20-40 years and thus heightens the pension manager's concern with respect to time-variations in interest rates and inflation rates. These long-term risks have been comprehensively examined in the literature. Boulier et al. (2001) study a DC pension problem under the Vasicek interest rate model and a minimal annuity guarantee. Battocchio and Menoncin (2004) consider a utility maximization problem of a DC pension member under stochastic interest rate, salary rate, and inflation rate. They highlight adding inflation rate changes the riskless asset into a risky asset, and the lack of tools hedging the inflation risk would lead to a more risky return. Zhang and Ewald (2010) investigate a DC pension management problem under inflation risk where pension contributions can be allocated to a risk-free bond, an index bond, and a stock. Han and Hung (2012) solve an optimal asset allocation of a DC pension plan under the Cox-Ingersoll-Ross interest model and inflation risk. They introduce an inflation indexed bond and find out that risk-averse pension participants concentrate on the inflation bond while aggressive participants would short the inflation bond in the early accumulation period. Some recent studies include Yao et al. (2013); Guan and Liang (2014); Menoncin and Vigna (2017) and Chen et al. (2017).

Besides financial risks, mortality risk is another factor that should be considered in the accumulation period of a DC pension plan. According to the latest life tables, the probability of a 22-year old young adult not surviving to 65 is non-negligible in many countries (see Table 3.1). Therefore, it is of great significance to study the effect of pension member's pre-mature death.

The current literature focuses on two types of death benefit clauses that protect the benefits of DC pension participants. The first type is the return of premiums clause. The pre-mature dead member can leave the beneficiary his or her contributed premium with or without predetermined interest during the accumulation phase (see Bian et al., 2018; He and Liang, 2013; Li et al., 2017; Sun et al., 2016). The other type is the return of account value. The pre-mature dead member can leave the beneficiary his or her pension account value (see Blake et al., 2008; Konicz and Mulvey, 2015; Wu and Zeng, 2015; Yao et al., 2014). In practice, most DC pension plans would return the account value to the individuals if they die before retirement, because the investment revenue is part of the estate, which can be inherited by the designated beneficiary. In the United States, according to Publication 575 (see IRS, 2019), if a 401(k) plan member dies before the beginning date of distribution, the entire account must be distributed within five years or in annual amounts over the lifetime of the designated beneficiary. A similar clause can also be found in Canada (see RBC, 2020; Sun Life Financial, 2017).

Table 3.1: **Mortality rate from age 22 to age 65 in different countries**

Country	Arithmetic average	Male	Female
United States	15.23%	18.89%	11.56%
United Kingdom	10.32%	12.49%	8.14%
Canada	9.52%	11.75%	7.29%
Australia	8.82%	11.10%	6.54%
China	11.04%	14.88%	7.19%
Japan	7.99%	10.63%	5.35%
Korea	7.57%	10.62%	4.53%

The data for each country is based on "2016 Period Life Table for the Social Security area population", "National Life Tables: UK 2017-2019", "Complete Life Tables for Canada 2017-2019", "Life Tables for Australia 2017-2019", "China Life Insurance Mortality Table 2010-2013", "Japanese Life Table No. 22 (JLT22)", "7th Standard Risk Death Rate", respectively.

In this chapter, we consider a DC pension plan management problem in a complete market. The financial market is driven by a two-factor model proposed by Koijen et al. (2011), which contains time-varying real interest rates, inflation rates, and risk premiums. A pension participant is allowed to invest his or her wealth among a stock index, nominal and inflation-linked bonds, and a nominal cash account. Besides financial instruments, the individual is also provided a life insurance to hedge his or her mortality risk of pre-mature death before retirement. If the individual passed away before the retirement age, the beneficiary would receive a death benefit composed of the DC pension account value and the life insurance payment. We formalize this pension problem to a utility maximization problem and derive the corresponding Hamilton-Jacobi-Bellman (HJB) equation based on the dynamic programming principle. The explicit solution is derived under the constant relative risk aversion (CRRA) utility, which is closely related to a matrix Riccati equation. Besides solving the problem, we also derive the conditions ensuring the solution's global existence in the different ranges of utility's risk-aversion coefficient. Moreover, a rigorous verification theorem is proved to guarantee the optimality of the candidate solution and strategies.

We estimate the model parameters by the Kalman filter method and acquire rich conclusions in the numerical research. In the dynamic analysis, we find that the pension member's demand for life insurance follows a hump shape and peaks in old age. This pattern is caused by the four components driving the insurance premiums: the surplus process, the force of mortality, the demand bequest ratio, and the future contributions. The numerical research shows that the surplus process dominates the trend and pulls up the demand for life insurance at an early age. Then, the increasing force of mortality, the decreasing demand bequest ratio, and future contributions dominate the trend and drag down the demand for life insurance. In the static analysis, we find the individual's demand for life insurance exhibits a "double top" shape for the real short rate and expected inflation. In other words, the pension member purchases more life insurance when the real short rate and expected inflation are both extraordinarily high or both extremely low. This phenomenon is caused by the combined effects of the demand bequest ratio and future contributions. The numerical research shows that, among these two components, the demand bequest ratio dominates the life insurance demand throughout the individual's lifetime.

To the best of our knowledge, our work in Chapter 3 is the first time to provide the explicit solution to the optimal life insurance problem in the DC pension management. We make three contributions to the existing literature: First, we first introduce the life insurance to the DC pension portfolios and create a DC account that resembles a variable annuity with endogenously determined time-varying death benefits. Like the DC pension fund, the variable annuity provides the individual with a wide range of investment options. However, its death benefit is exogenously determined as the maximum between the account value and some guaranteed minimum, partially ignoring the individual's bequest demand (see SEC (2009)). This chapter builds up a variable-annuity like DC account with endogenously determined time-varying death benefits. It relaxes the limitations on the variable annuity's death benefits and can inspire more innovations in designing the new actuarial products. Second, we provide rigorous proofs to the explicit solution's global existence and verification theorem under CRRA utility. Due to the lack of global existence, the existing literature under CRRA utility mainly focuses on the case that the risk-aversion coefficient $\gamma > 1$ (see Wang and Li, 2018; Wang et al., 2021). In Chapter 3, we provide a sufficient condition for the global existence of solution under $0 < \gamma < 1$ (see Proposition 3.3.3). Furthermore, inspired by Honda and Kamimura (2011), we prove the verification theorem for our model under both $\gamma > 1$ and $0 < \gamma < 1$. Third, we provide both dynamic and static simulations to reveal the individual's demand for life insurance under CRRA utility. Specifically, our numerical research shows that the individual's life insurance demand exhibits a hump shape with age and a "double top" pattern for the real short rate and expected inflation. It can be a potential guide for the DC pension member to hedge their mortality risk at the old age or under extremely bad market scenarios.

The rest of the chapter is organized as follows. Section 3.2 introduces the economic settings of the financial market and insurance market. Section 3.3 presents the whole process of setting the preference, solving the problem, deducing the solution's global existence, and proving the verification theorem. Section 3.4 estimates the model parameters from the real market and conducts the numerical research of the solution. Section 3.5 concludes. All proofs are relegated to Appendix B.

3.2 Economic setting

3.2.1 Financial market

Finding the common factors is the most efficient way to describe various economic variables in the financial market. Therefore, motivated by Koijen et al. (2011), we use a two-factor model to describe the time variations in five different economic variables. Specifically, they are real short rates, expected inflation, stock appreciation rates, nominal short rates, and risk premiums. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a filtered complete probability space. Then the financial risk is described by Z_t, a four-dimensional vector of independent Brownian motions, which is adapted to the filtration $\mathbb{F} := \{\mathcal{F}_t\}_{t\in[0,T]}$ and come from the real short rate, the expected inflation, the commodity price index, and the equity index.

We assume that the real short rate is driven by a single factor, X_1,

$$r_t = \delta_r + X_{1,t}, \quad \delta_r > 0,$$

where δ_r is the long term average of real short rate. The expected inflation is affine in a second factor, X_2,

$$\pi_t^e = \delta_{\pi^e} + X_{2,t}, \quad \delta_{\pi^e} > 0,$$

where δ_{π^e} is the long term average of expected inflation. Furthermore, the two factors satisfy the following Ornstein-Uhlenbeck process

$$dX_t = -K_X X_t dt + \Sigma_X dZ_t, \tag{3.1}$$

where $X_t = (X_{1,t}, X_{2,t})^\top$, $K_X = diag(\kappa_1, \kappa_2)$, $\kappa_i > 0$, $i = 1, 2$, $\Sigma_X = (\sigma_1, \sigma_2)^\top$, $\sigma_i \in \mathbb{R}^4$, $i = 1, 2$.

Next, we assume that the realized inflation is given by

$$\frac{d\Pi_t}{\Pi_t} = \pi_t^e dt + \sigma_\Pi^\top dZ_t, \ \Pi_0 = 1,$$

where Π_t denotes the level of the (commodity) price index at time t and $\sigma_\Pi \in \mathbb{R}^4$.

The equity index S_t satisfies the following dynamics

$$\frac{dS_t}{S_t} = \mu_t dt + \sigma_S^\top dZ_t,$$

where $\mu_t = R_t + \mu_0 + \mu_1^\top X_t$ is the stock appreciation rate and R_t denotes the instantaneous nominal short rate that is derived in (3.4) below. For identification purposes, we assume the volatility matrix $(\sigma_1, \sigma_2, \sigma_\Pi, \sigma_S)^\top$ is lower triangular.

We assume that the nominal state price density ϕ satisfies

$$\frac{d\phi_t}{\phi_t} = -R_t dt - \Lambda_t^\top dZ_t, \ \ \phi_0 = 1, \tag{3.2}$$

in which the market prices of risk, Λ_t, are affine in the term-structure variables, i.e.,

$$\Lambda_t = \Lambda_0 + \Lambda_1 X_t. \tag{3.3}$$

Then, we follow Koijen et al. (2011) to impose restrictions on Λ_0 and Λ_1

$$\Lambda_0 = \begin{pmatrix} \Lambda_{0(1)} \\ \Lambda_{0(2)} \\ 0 \\ \Lambda_{0(4)} \end{pmatrix}, \ \ \Lambda_1 = \begin{pmatrix} \Lambda_{1(1,1)} & 0 \\ 0 & \Lambda_{1(2,2)} \\ 0 & 0 \\ \Lambda_{1(4,1)} & \Lambda_{1(4,2)} \end{pmatrix},$$

with $\sigma_S^\top \Lambda_0 = \mu_0^\top$ and $\sigma_S^\top \Lambda_1 = \mu_1^\top$. The real state price density $\phi_t^R = \phi_t \Pi_t$ then satisfies

$$\begin{aligned} \frac{d\phi_t^R}{\phi_t^R} &= -(R_t - \pi_t^e + \sigma_\Pi^\top \Lambda_t)dt - (\Lambda_t^\top - \sigma_\Pi^\top)dZ_t \\ &= -r_t dt - (\Lambda_t^\top - \sigma_\Pi^\top)dZ_t, \ \ \phi_0^R = 1, \end{aligned}$$

which implies for the instantaneous nominal short rate

$$R_t = \delta_R + (\iota_2^\top - \sigma_\Pi^\top \Lambda_1)X_t, \tag{3.4}$$

where $\delta_R = \delta_r + \delta_{\pi^e} - \sigma_\Pi^\top \Lambda_0$ and $\iota_2 = (1,1)^\top$.

Finally, we present the prices of nominal and inflation-linked bonds. The derivation is standard in the literature (e.g. Duffie and Kan, 1996). The time-t price of a nominal bond with maturity s is

$$P(X_t, t, s) = \exp\{A_0(s-t) + [A_1(s-t)]^\top X_t\},$$

where A_0 and A_1 satisfy the following ODE system

$$\frac{\partial A_0(\tau)}{\partial \tau} = \frac{1}{2}[A_1(\tau)]^\top \Sigma_X \Sigma_X^\top A_1(\tau) - [A_1(\tau)]^\top \Sigma_X \Lambda_0 - \delta_R,\ A_0(0) = 0, \tag{3.5}$$

$$\frac{\partial A_1(\tau)}{\partial \tau} = -[K_X^\top + \Lambda_1^\top \Sigma_X^\top] A_1(\tau) - \iota_2 + \Lambda_1^\top \sigma_\Pi,\ A_1(0) = 0. \tag{3.6}$$

In addition, the dynamics of $P(X_t, t, s)$ satisfy

$$\frac{dP(X_t, t, s)}{P(X_t, t, s)} = \{R_t + [A_1(s-t)]^\top \Sigma_X \Lambda_t\} dt + [A_1(s-t)]^\top \Sigma_X dZ_t.$$

Similarly, the time-t real price of an inflation-linked bond with maturity s is

$$P^R(X_t, t, s) = \exp\{A_0^R(s-t) + [A_1^R(s-t)]^\top X_t\},$$

where A_0^R and A_1^R satisfy the ODE system

$$\begin{aligned}
&\frac{\partial A_0^R(\tau)}{\partial \tau} = \frac{1}{2}[A_1^R(\tau)]^\top \Sigma_X \Sigma_X^\top A_1^R(\tau) - [A_1^R(\tau)]^\top \Sigma_X (\Lambda_0 - \sigma_\Pi) - \delta_r, \\
&\frac{\partial A_1^R(\tau)}{\partial \tau} = -(K_X^\top + \Lambda_1^\top \Sigma_X^\top) A_1^R(\tau) - e_1, \\
&A_0^R(0) = A_1^R(0) = 0,
\end{aligned}$$

in which e_i represents the i-th unit vector in $\mathbb{R}^2$. Then, the nominal price of the inflation-linked bond $\Pi_t P^R(X_t, t, s)$ satisfies

$$\frac{d(\Pi_t P^R(X_t, t, s))}{\Pi_t P^R(X_t, t, s)} = \{R_t + [A_1^R(s-t)]^\top \Sigma_X \Lambda_t + \sigma_\Pi^\top \Lambda_t\} dt + \{[A_1^R(s-t)]^\top \Sigma_X + \sigma_\Pi^\top\} dZ_t.$$

3.2.2 Mortality

In this subsection, we introduce mortality risk. Denote by T_x the future lifetime of an individual aged x, a non-negative random variable independent of the financial market (i.e., T_x is independent of the filtration $\mathbb{F}$ in the financial market). This independence assumption helps us to simplify the analysis and obtain an explicit solution. However, the relationship between the individual's mortality and the financial market is complex in the real world and still controversial in the existing literature. Some studies show that the individual's mortality rate has a pro-cyclical relationship with the business cycle, i.e., the individual has a low mortality rate when the economy is in a depression (see Ballester et al., 2019; Cervini-Plá and Vall-Castelló, 2021; Haaland and Telle, 2015; Ruhm, 2000; Stevens et al., 2015). By contrast, the other group of research reveals a counter-cyclical pattern between the individual's mortality rate with the business cycle (see Halliday, 2014; Lam and Piérard, 2017; McInerney and Mellor, 2012; Ruhm, 2003; Stevens et al., 2015). There is also a third group of empirical studies conclude that mortality rate is not significantly related to the macro-economic situations (see Brüning and Thuilliez, 2019; Ruhm, 2015). For future work, we recommend using a stochastic force of mortality model to describe the business cycle effect on the individual's mortality rate (see Huang et al., 2012; Qian et al., 2010; Shen and Siu, 2013; Zhou et al., 2022).

With the independence assumption above, we define the following probabilities

$$\begin{aligned}
{}_tp_x &= \mathbb{P}[T_x > t], \\
{}_tq_x &= \mathbb{P}[T_x \leq t] = 1 - {}_tp_x, \\
\lim_{t\to\infty} {}_tp_x &= 0, \lim_{t\to\infty} {}_tq_x = 1,
\end{aligned}$$

where ${}_tp_x$ is the probability that the individual alive at age x survives to at least age $x+t$ and ${}_tq_x$ is the probability that the individual dies before age $x+t$. In actuarial science, it is common to work with the instantaneous force of mortality (or hazard rate)

$$\mu_{x+t} = \frac{1}{{}_tp_x}\frac{d}{dt}{}_tq_x = -\frac{1}{{}_tp_x}\frac{d}{dt}{}_tp_x,$$

and we have

$$\begin{aligned}
{}_tp_x &= \exp\left\{-\int_0^t \mu_{x+s} ds\right\}, \\
{}_tq_x &= \int_0^t {}_sp_x \mu_{x+s} ds.
\end{aligned}$$

The probability density function of T_x is then given by

$$f_{T_x}(t) = {}_tp_x \mu_{x+t}, \text{ for } t > 0. \tag{3.7}$$

3.2.3 Wealth process

The individual (pensioner) enters the DC pension plan at age x at time 0, and retires at time T (so the retirement age is $x+T$). Before retirement or death, the individual contributes a fixed percentage of labor income continuously to the fund. Following Koijen et al. (2011), we assume that real labor income is deterministic. Thus, the real contribution rate, $C_t = C_t^{\$}\Pi_t^{-1}$, satisfies

$$\frac{dC_t}{C_t} = g_t^R dt, \quad 0 \le t < T \wedge T_x, \tag{3.8}$$

where $\wedge$ means taking the minimum of two variables, $C_t^{\$}$ is the nominal contribution rate and g_t^R is the growth rate of the real contribution rate (which is also the growth rate of labor income).

During the accumulation period, the individual allocates his or her wealth dynamically to the stock index, two nominal bonds, and an inflation-linked bond. In particular, for the purchase of bonds, the individual employs the "rolling bond" strategy proposed by Boulier et al. (2001). Denote by α_t the proportions of wealth invested in these assets at time t. The rest of wealth is invested in the cash account. In addition to investment, the individual can purchase life insurance in the DC account to manage the mortality risk. Suppose the individual pays the insurance premium at a rate of $I_t^{\$}$ (in nominal terms) continuously to the insurer while alive. If the individual dies at time $T_x = t$ prior to retirement, then his or her beneficiary receives the death benefit (the face value of life insurance) $I_t^{\$}/\mu_{x+t}$ in addition to the account balance. The individual's DC account balance then evolves according to the following equation

$$dW_t = W_t(\alpha_t^\top \Sigma \Lambda_t + R_t)dt + C_t^{\$}dt + W_t \alpha_t^\top \Sigma dZ_t - I_t^{\$}dt, \quad 0 \le t < T \wedge T_x,$$

where $W_0 = 0$ and Σ is the volatility matrix of the tradable assets. We assume the two nominal bonds have maturities T_1 and T_2, and the inflation-linked bond has maturity T_3. Consequently,

$$\Sigma = \begin{pmatrix} [A_1(T_1)]^\top \Sigma_X \\ [A_1(T_2)]^\top \Sigma_X \\ [A_1^R(T_3)]^\top \Sigma_X + \sigma_\Pi^\top \\ \sigma_S^\top \end{pmatrix}.$$

We can then derive the dynamics of the real wealth $W_t^R = W_t/\Pi_t$

$$dW_t^R = W_t^R[r_t + (\alpha_t^\top \Sigma - \sigma_\Pi^\top)(\Lambda_t - \sigma_\Pi)]dt + C_t dt + W_t^R(\alpha_t^\top \Sigma - \sigma_\Pi^\top)dZ_t - I_t dt, \tag{3.9}$$

where $0 \le t < T \wedge T_x$, $W_0^R = 0$, and $I_t = I_t^{\$}/\Pi_t$ is the real insurance premium rate.

If the individual dies prior to retirement, then the death benefit is added to the account balance

$$W_t^R = W_{t-}^R + \frac{I_t}{\mu_{x+t}}, \text{ if } T_x = t < T.$$

In our model, the financial market is complete, which means any payoff can be replicated by the products in this market. Specifically, we have four independent Brownian motions Z_t and four different products (two nominal bonds, one inflation-linked bond, and one equity index). Then, any single Brownian motion in Z_t can be replicated with these four products, and so does any payoff. We need this complete market assumption to derive the explicit solution and enhance the tractability of our analysis.

3.2.4 Preference

We assume the individual chooses investment and insurance strategies (α, I) to maximize the expected utility of account balance at retirement or death, whichever is earlier, i.e.

$$\sup_{\alpha,I} E[U(W^R_{T\wedge T_x})].$$

Because T_x is independent of financial risks, we can show

$$\begin{aligned}
& \sup_{\alpha,I} E[U(W^R_{T\wedge T_x})] \\
= & \sup_{\alpha,I} E\left[\int_0^T f_{T_x}(t)U\left(W^R_t + \frac{I_t}{\mu_{x+t}}\right)dt + \int_T^\infty f_{T_x}(t)U(W^R_T)dt\right] \\
= & \sup_{\alpha,I} E\left[\int_0^T {}_tp_x\mu_{x+t}U\left(W^R_t + \frac{I_t}{\mu_{x+t}}\right)dt + {}_Tp_xU(W^R_T)\right]. \qquad (3.10)
\end{aligned}$$

3.3 Optimization problem

3.3.1 Dynamic programming

Following Deelstra et al. (2003), we introduce the surplus process $W^{\widetilde{C}}_t$ to enhance the tractability of our analysis

$$W^{\widetilde{C}}_t = W^R_t + \widetilde{C}(t, X_t), \qquad (3.11)$$

where $\widetilde{C}(t, X_t)$ is the accumulated discounted contribution rate

$$\widetilde{C}(t, X_t) = \int_t^T {}_{s-t}p_{x+t}P^R(X_t, t, s)C_s ds. \qquad (3.12)$$

Next, by Ito's formula, we have

$$\begin{aligned} d\widetilde{C}(t,X_t) &= -C_t dt + (r_t+\mu_{x+t})\widetilde{C}(t,X_t)dt + \frac{\partial \widetilde{C}(t,X_t)}{\partial X}\Sigma_X(\Lambda_t-\sigma_\Pi)dt \\ &\quad + \frac{\partial \widetilde{C}(t,X_t)}{\partial X}\Sigma_X dZ_t. \end{aligned} \tag{3.13}$$

Assume that there exists a ξ_t such that

$$\begin{aligned} d\widetilde{C}(t,X_t) &= -C_t dt + \widetilde{C}(t,X_t)[r_t + (\xi_t^\top\Sigma - \sigma_\Pi^\top)(\Lambda_t-\sigma_\Pi)]dt + \mu_{x+t}\widetilde{C}(t,X_t)dt \\ &\quad + \widetilde{C}(t,X_t)(\xi_t^\top\Sigma-\sigma_\Pi^\top)dZ_t, \end{aligned} \tag{3.14}$$

then we obtain ξ by comparing the terms in (3.13) and (3.14)

$$\xi_t = \frac{1}{\widetilde{C}(t,X_t)}(\Sigma^\top)^{-1}\Sigma_X^\top \frac{\partial \widetilde{C}(t,X_t)}{\partial X^\top} + (\Sigma^\top)^{-1}\sigma_\Pi. \tag{3.15}$$

Furthermore, adding (3.9) and (3.14), we derive the SDE for surplus process

$$\begin{aligned} dW_t^{\widetilde{C}} &= dW_t^R + d\widetilde{C}(t,X_t) \\ &= W_t^{\widetilde{C}}\{r_t + (\beta_t^\top\Sigma-\sigma_\Pi^\top)(\Lambda_t-\sigma_\Pi)\}dt + W_t^{\widetilde{C}}(\beta_t^\top\Sigma-\sigma_\Pi^\top)dZ_t \\ &\quad + \mu_{x+t}\widetilde{C}(t,X_t)dt - I_t dt, \end{aligned} \tag{3.16}$$

where $0 \le t < T \wedge T_x$ and $\beta_t^\top = [W_t^R\alpha_t^\top + \widetilde{C}(t,X_t)\xi_t^\top]/W_t^{\widetilde{C}}$. The SDE (3.16) models the investment in the financial market, and the purchase of life insurance with premium $-\mu_{x+t}\widetilde{C}(t,X_t) + I_t$. When the individual dies before the retirement, the surplus process has the following jump.

$$W_t^{\widetilde{C}} = W_{t-}^{\widetilde{C}} - \widetilde{C}(t,X_t) + \frac{I_t}{\mu_{x+t}}, \text{ if } T_x = t < T. \tag{3.17}$$

Then, by definition (3.11), and given that $W_T^{\widetilde{C}} = W_T^R$ at T, the objective function (3.10) can be transformed to

$$\begin{aligned} &\sup_{\alpha,I} E\left[\int_0^T {}_tp_x\mu_{x+t}U\left(W_t^R + \frac{I_t}{\mu_{x+t}}\right)dt + {}_Tp_xU(W_T^R)\right] \\ =\ &\sup_{\beta,I} E\left[\int_0^T {}_tp_x\mu_{x+t}U\left(W_t^{\widetilde{C}} - \widetilde{C}(t,X_t) + \frac{I_t}{\mu_{x+t}}\right)dt + {}_Tp_xU(W_T^{\widetilde{C}})\right]. \end{aligned}$$

Define the value function

$$
\begin{aligned}
V(t, w^{\widetilde{C}}, X) &= \sup_{\beta, I} E\bigg[\int_t^T {}_{s-t}p_{x+t}\mu_{x+s} U\bigg(W_s^{\widetilde{C}} - \widetilde{C}(s, X_s) + \frac{I_s}{\mu_{x+s}}\bigg) ds \\
&\quad + {}_{T-t}p_{x+t} U(W_T^{\widetilde{C}}) | W_t^{\widetilde{C}} = w^{\widetilde{C}}, X_t = X\bigg],
\end{aligned}
$$

then by the dynamic programming principle and Ito's formula, we can derive the following HJB equation

$$
\sup_{\beta_t, I_t} \left\{ \mu_{x+t} U\left(w^{\widetilde{C}} - \widetilde{C}(t, X) + \frac{I_t}{\mu_{x+t}}\right) - \mu_{x+t} V(t, w^{\widetilde{C}}, X) + \mathcal{D}^{\beta, I} V(t, w^{\widetilde{C}}, X) \right\} = 0, \tag{3.18}
$$

where

$$
\begin{aligned}
\mathcal{D}^{\beta, I} V(t, w^{\widetilde{C}}, X) &= \frac{\partial V}{\partial t} + \frac{\partial V}{\partial w^{\widetilde{C}}} \{ w^{\widetilde{C}} [r_t + (\beta_t^\top \Sigma - \sigma_\Pi^\top)(\Lambda_t - \sigma_\Pi)] + \mu_{x+t} \widetilde{C}(t, X) - I_t \} \\
&\quad - \frac{\partial V}{\partial X} K_X X + \frac{1}{2} \frac{\partial^2 V}{(\partial w^{\widetilde{C}})^2} (w^{\widetilde{C}})^2 (\beta_t^\top \Sigma \Sigma^\top \beta_t - 2\beta_t^\top \Sigma \sigma_\Pi + \sigma_\Pi^\top \sigma_\Pi) \\
&\quad + w^{\widetilde{C}} (\beta_t^\top \Sigma - \sigma_\Pi^\top) \Sigma_X^\top \frac{\partial^2 V}{\partial w^{\widetilde{C}} \partial X^\top} + \frac{1}{2} \mathrm{Tr}\bigg(\Sigma_X^\top \frac{\partial^2 V}{\partial X^\top \partial X} \Sigma \bigg).
\end{aligned}
$$

Optimizing with respect to β_t and I_t, we have

$$
\beta_t^* = -\frac{(\Sigma^\top)^{-1}}{w^{\widetilde{C}} \frac{\partial^2 V}{(\partial w^{\widetilde{C}})^2}} \left[\frac{\partial V}{\partial w^{\widetilde{C}}} (\Lambda_t - \sigma_\Pi) + \Sigma_X^\top \frac{\partial^2 V}{\partial w^{\widetilde{C}} \partial X^\top} \right] + (\Sigma^T)^{-1} \sigma_\Pi, \tag{3.19}
$$

$$
I_t^* = \mu_{x+t} (U')^{-1} \left(\frac{\partial V}{\partial w^{\widetilde{C}}} \right) - \mu_{x+t} (W_t^{\widetilde{C}})^* + \mu_{x+t} \widetilde{C}(t, X_t). \tag{3.20}
$$

Substituting (3.19) and (3.20) into (3.18), we simplify the HJB equation to the following form

$$
\begin{aligned}
0 &= \mu_{x+t} U\bigg((U')^{-1} \bigg(\frac{\partial V}{\partial w^{\widetilde{C}}} \bigg) \bigg) - \mu_{x+t} V(t, w^{\widetilde{C}}, X) + \frac{\partial V}{\partial t} - \frac{\partial V}{\partial X} K_X X \\
&\quad + \frac{\partial V}{\partial w^{\widetilde{C}}} \bigg[(r_t + \mu_{x+t}) w^{\widetilde{C}} - \mu_{x+t} (U')^{-1} \bigg(\frac{\partial V}{\partial w^{\widetilde{C}}} \bigg) \bigg] \\
&\quad - \frac{1}{2 \frac{\partial^2 V}{(\partial w^{\widetilde{C}})^2}} \left[\frac{\partial V}{\partial w^{\widetilde{C}}} (\Lambda_t^\top - \sigma_\Pi^\top) + \frac{\partial^2 V}{\partial w^{\widetilde{C}} \partial X} \Sigma_X \right] \left[\frac{\partial V}{\partial w^{\widetilde{C}}} (\Lambda_t - \sigma_\Pi) + \Sigma_X^\top \frac{\partial^2 V}{\partial w^{\widetilde{C}} \partial X^\top} \right] \\
&\quad + \frac{1}{2} \mathrm{Tr}\bigg(\Sigma_X^\top \frac{\partial^2 V}{\partial X^\top \partial X} \Sigma_X \bigg).
\end{aligned} \tag{3.21}
$$

Before moving to the next subsection, we revisit the formulations ((3.16) and (3.17)) and find that the individual's DC account resembles a variable annuity with endogenously determined time-varying death benefits. It relaxes the limitation of the variable annuity's exogenously determined death benefits (defined as the maximum between the account value and some guaranteed minimum) and cares more about the individual's bequest demand. Therefore, we can highlight the usefulness of our study by its potential ability to influence new actuarial products hedging the individual's mortality risk. For more details of variable annuities, see SEC (2009).

3.3.2 Solution to the power utility

Inspired by Koijen et al. (2011), we solve the optimal problem under the power utility.

$$U(x) = \frac{x^{1-\gamma}}{1-\gamma},$$

where $\gamma > 0$ and $\gamma \neq 1$ is the Arrow-Pratt coefficient of relative risk aversion.

Proposition 3.3.1. *The candidate solution $G(t, W_t^{\widetilde{C}}, X_t)$ to the value function $V(t, W_t^{\widetilde{C}}, X_t)$ in (3.21) is given by*

$$G(t, W_t^{\widetilde{C}}, X_t) = \frac{1}{1-\gamma}(W_t^{\widetilde{C}})^{1-\gamma} f_1(t, X_t)^{\gamma}, \tag{3.22}$$

where

$$f_1(t, X_t) = \int_t^T {}_{s-t}p_{x+t}\mu_{x+s} f(X_t, s-t)ds + {}_{T-t}p_{x+t} f(X_t, T-t), \tag{3.23}$$

$$f(X_t, \tau) = \exp\left[\Gamma_0(\tau) + \Gamma_1^\top(\tau)X_t + \frac{1}{2}X_t^\top \Gamma_2(\tau) X_t\right], \ \tau \in [0, T-t]. \tag{3.24}$$

Functions $\Gamma_0(\tau) \in \mathbb{R}$, $\Gamma_1(\tau) \in \mathbb{R}^2$ and $\Gamma_2(\tau) \in \mathbb{R}^2 \times \mathbb{R}^2$ are given by the following ODE system

$$\frac{\partial \Gamma_2(\tau)}{\partial \tau} - \Gamma_2(\tau) Z_2 \Gamma_2(\tau) - Z_1^\top \Gamma_2(\tau) - \Gamma_2(\tau) Z_1 - Z_0 = 0, \tag{3.25}$$

$$\frac{\partial \Gamma_1(\tau)}{\partial \tau} - \Gamma_2(\tau) B_2 \Gamma_1(\tau) - \Gamma_2(\tau) B_{11} - B_{12}\Gamma_1(\tau) - B_0 = 0, \tag{3.26}$$

$$\frac{\partial \Gamma_0(\tau)}{\partial \tau} - \Gamma_1^\top(\tau) D_2 \Gamma_1(\tau) - \Gamma_1^\top(\tau) D_1 - \frac{1}{2} Tr\{\Sigma_X^\top \Gamma_2(\tau) \Sigma_X\} - D_0 = 0, \tag{3.27}$$

$$\Gamma_2(0) = \Gamma_1(0) = \Gamma_0(0) = 0.$$

in which

$$Z_2 = \Sigma_X \Sigma_X^\top, Z_1 = \frac{1-\gamma}{\gamma}\Sigma_X \Lambda_1 - K_X, Z_0 = \frac{1-\gamma}{\gamma^2}\Lambda_1^\top \Lambda_1,$$
$$B_2 = Z_2, B_{11} = \frac{1-\gamma}{\gamma}\Sigma_X(\Lambda_0 - \sigma_\Pi), B_{12} = Z_1^\top, B_0 = \frac{1-\gamma}{\gamma^2}\Lambda_1^\top(\Lambda_0 - \sigma_\Pi) + \frac{1-\gamma}{\gamma}e_1,$$
$$D_2 = \frac{1}{2}Z_2, D_1 = B_{11}, D_0 = \frac{1-\gamma}{\gamma}\delta_r + \frac{1-\gamma}{2\gamma^2}(\Lambda_0^\top - \sigma_\Pi^\top)(\Lambda_0 - \sigma_\Pi).$$

The candidate strategies are given by

$$\beta_t^* = \frac{(\Sigma^\top)^{-1}}{\gamma}(\Lambda_t - \sigma_\Pi) + (\Sigma^\top)^{-1}\Sigma_X^\top \frac{1}{f_1(t, X_t)} \frac{\partial f_1(t, X_t)}{\partial X^\top} + (\Sigma^\top)^{-1}\sigma_\Pi, \tag{3.28}$$

$$I_t^* = \mu_{x+t}\frac{(W_t^{\widetilde{C}})^*}{f_1(t, X_t)} - \mu_{x+t}(W_t^{\widetilde{C}})^* + \mu_{x+t}\widetilde{C}(t, X_t). \tag{3.29}$$

Next, we need to prove the candidate solution's global existence and verification theorem.

3.3.3 The global existence of candidate solution $G(t, W_t^{\widetilde{C}}, X_t)$

Among the ODEs determining the candidate solution, (3.26) and (3.27) are linear ODEs. Their solutions are unique and exist globally (see Theorem 1.1.1. in Abou-Kandil et al. (2012)). However, the ODE (3.25) is a Hermitian matrix Riccati differential equation (HRDE), and we need a special treatment to prove its existence. HRDE (3.25) has the following matrix representation

$$\begin{aligned} \frac{\partial \Gamma_2(\tau)}{\partial \tau} &= (\widetilde{I}_2, \Gamma_2(\tau)) J H(\tau) \begin{pmatrix} \widetilde{I}_2 \\ \Gamma_2(\tau) \end{pmatrix} \\ &:= \mathcal{H}(\Gamma_2; H), \tau \in [0, T], \end{aligned} \tag{3.30}$$

where $\widetilde{I}_2$ is the 2nd-order identity matrix,

$$J := \begin{pmatrix} 0_{2\times 2} & \widetilde{I}_2 \\ -\widetilde{I}_2 & 0_{2\times 2} \end{pmatrix} \in \mathbb{R}^2 \times \mathbb{R}^2,$$

and

$$H := \begin{pmatrix} -Z_1 & -Z_2 \\ Z_0 & Z_1^\top \end{pmatrix} \in \mathbb{R}^2 \times \mathbb{R}^2, \tag{3.31}$$

which is called Hamiltonian matrix.

The global existence of HRDE (3.30) largely depends on the relative risk aversion coefficient γ. We divide the proof into two cases, $\gamma > 1$ and $0 < \gamma < 1$. For $\gamma > 1$, we can use the comparison theorem of HRDE to prove the following proposition.

Proposition 3.3.2. *For $\gamma > 1$, if $\Sigma_X\Sigma_X^\top > 0$ and $\Lambda_1^\top\Lambda_1 > 0$, then the solution to (3.25) exists and stays negative definite in $(0, T]$. Here, for matrix, "$>$" means positive definite.*

Since $\Gamma_2(\tau)$ exists and $\Gamma_2(\tau) < 0$ for $\forall\tau \in (0, T]$, the existence of (3.22) is a direct result.

For $0 < \gamma < 1$, we can prove the existence of (3.25) by Radon's Lemma with additional conditions. Denote $(Q, P)^\top$ as a solution to the linear system of differential equations

$$\frac{d}{d\tau}\begin{pmatrix} Q(\tau) \\ P(\tau) \end{pmatrix} = H \begin{pmatrix} Q(\tau) \\ P(\tau) \end{pmatrix}, Q(0) = \widetilde{I}_2, P(0) = \Gamma_2(0)Q(0) = 0. \tag{3.32}$$

By Radon's Lemma (see Theorem 3.1.1. in Abou-Kandil et al. (2012)), we can represent the solution to (3.25) as $\Gamma_2(\tau) = P(\tau)/Q(\tau)$. Next, we only need $\Gamma_2(\tau) < 0$ to guarantee the candidate solution's global existence. For the tractability, we follow Abou-Kandil et al. (2012) and assume H is diagonalizable, i.e. there exists a 4-dimensional basis of eigenvectors

$$v_1, ..., v_4 \in \mathbb{C}^4,$$

where $\mathbb{C}^4$ denotes the complex vector space of 4×1 complex vectors, and the corresponding eigenvalues are $\lambda_1, ..., \lambda_4$ sorted by their real parts

$$\mathcal{R}(\lambda_1) \leq \mathcal{R}(\lambda_2) \leq \mathcal{R}(\lambda_3) \leq \mathcal{R}(\lambda_4).$$

Denote $V = (v_1, ..., v_4) \in \mathbb{C}^{4\times 4}$, in which $\mathbb{C}^{4\times 4}$ denotes the complex vector space of 4×4 complex matrices, then we have that the solution to (3.32) satisfies

$$\begin{pmatrix} Q(\tau) \\ P(\tau) \end{pmatrix} = Ve^{\Delta\tau}V^{-1}\begin{pmatrix} Q(0) \\ P(0) \end{pmatrix} = Ve^{\Delta\tau}V^{-1}\begin{pmatrix} \widetilde{I}_2 \\ 0 \end{pmatrix}, \tag{3.33}$$

where $\Delta := V^{-1}HV = \text{diag}(\lambda_1, ..., \lambda_4)$.

Furthermore, define

$$f_\lambda(\lambda) = |\lambda\widetilde{I}_4 - H| = \lambda^4 + b\lambda^3 + c\lambda^2 + d\lambda + j, \tag{3.34}$$

we can finally prove the following proposition.

Proposition 3.3.3. *For $0 < \gamma < 1$, if*

$$\widetilde{\Delta} > 0,\ \ q < 0,\ \ s < \frac{q^2}{4}, \tag{3.35}$$

$$det|Q(\tau)| \neq 0 \ \textit{and}\ P(\tau)/Q(\tau) < 0 \ \textit{for}\ \forall \tau \in (0, T], \tag{3.36}$$

then the solution to (3.25) exists and stays negative definite in $(0, T]$, where $det|\cdot|$ is the determinant of a matrix. The expressions of $\widetilde{\Delta}$, q, and s are given in Appendix B.3.

Since $\Gamma_2(\tau)$ exists and $\Gamma_2(\tau) < 0$ for $\forall \tau \in (0, T]$, the existence of (3.22) is a direct result.

3.3.4 The verification theorem

Korn and Kraft (2004) point out that the coefficients of the wealth process may not be subject to the Lipschitz and linear growth condition in the stochastic opportunity set (such as stochastic interest settings, stochastic volatility settings, and stochastic market price of risk settings). Moreover, without checking sufficient conditions, solving the problem directly by the stochastic control approach could lead to unreasonable results. Therefore, we need to prove the verification theorem for the candidate solution.

Before moving forward, we first prove the following lemma, which will play a crucial role in proving the verification theorem

Lemma 3.3.1. *Assume a n-dimensional stochastic process $\widetilde{X}_t$ is driven by a m-dimensional Brownian motion $\widetilde{Z}$*

$$\begin{aligned} d\widetilde{X}_t &= \mu(t, \widetilde{X}_t)dt + \sigma(t)d\widetilde{Z}_t, \\ \widetilde{X}_0 &= \widetilde{x}_0, \end{aligned}$$

where $\widetilde{x}_0$ is a constant n-dimensional vector, $\mu(t, \widetilde{X})$ is a Borel function and $\sigma(t)$ a continuous function

$$\begin{aligned} \mu(t, \widetilde{X}) &: (0, \infty) \times \mathbb{R}^n \to \mathbb{R}^n, \\ \sigma(t) &: (0, \infty) \to \mathbb{R}^n \times \mathbb{R}^m, \end{aligned}$$

satisfying

$$||\mu(t, \widetilde{X}_t) - \mu(t, \widetilde{Y}_t)||_2 \leq k||\widetilde{X}_t - \widetilde{Y}_t||_2,$$
$$||\mu(\cdot, 0)||_2 + ||\sigma(\cdot)||_2 \in L^2(0, T; \mathbb{R}), \forall T > 0,$$

where $||\cdot||_2$ *is the Euclidean norm and* $L^2(0,T;\mathbb{R})$ *represents the set of Lebesgue measurable functions* $\psi : [0,T] \to \mathbb{R}$*, such that* $\int_0^T |\psi(t)|^2 dt < \infty$*.*

If a stochastic process $\widetilde{g}(t, \widetilde{X}_t)$*,* $\widetilde{g} : [0,T] \times \mathbb{R}^n \to \mathbb{R}^n$*, grows linearly with respect to* $\widetilde{X}_t$ *(namely,* $||\widetilde{g}(t,\widetilde{X}_t)||_2 \leq c_0 + c_1||\widetilde{X}_t||_2$ *for some positive constants* c_0 *and* c_1*), then we have*

$$E[\mathcal{E}(T,\widetilde{g})] = 1,$$

where

$$\mathcal{E}(t,\widetilde{g}) := \exp\left\{ \int_0^t [\widetilde{g}(s,\widetilde{X}_s)]^\top d\widetilde{Z}_s - \frac{1}{2}\int_0^t ||\widetilde{g}(s,\widetilde{X}_s)||_2^2 ds \right\}.$$

Finally, after adequate preparations, we can prove the verification theorem following the approach proposed by Honda and Kamimura (2011). We divide the proofs into two cases $0 < \gamma < 1$ and $\gamma > 1$.

Case $0 < \gamma < 1$

Define the admissible set as

$$\mathcal{A}_\gamma(0,T) := \left\{ (\beta, I) \;\middle|\; \begin{array}{c} (\beta, I) \text{ such that } W_t^{\widetilde{C}} > 0, \\ \text{and SDE (3.16) has a unique strong solution.} \end{array} \right\}, \tag{3.37}$$

we can derive the following verification theorem.

Proposition 3.3.4. *For* $0 < \gamma < 1$*, under the parameters settings in Proposition 3.3.3, the candidate solution* $G(t, W_t^{\widetilde{C}}, X_t)$ *exists in* $[0,T]$ *and satisfies* $G(t, W_t^{\widetilde{C}}, X_t) = V(t, W_t^{\widetilde{C}}, X_t)$*. The strategy* (β^*, I^*) *given by (3.28) and (3.29) is the optimal portfolio and insurance strategy.*

Case $\gamma > 1$

When $\gamma > 1$, the power utility is not bounded from below. Thus, we can not use Fatou's lemma in our proof. To prove the verification theorem, we restrict the admissible set as

$$\mathcal{A}_\gamma(0,T) := \left\{ (\beta, I) \;\middle|\; \begin{array}{c} \beta(t, X_t) : [0,T] \times \mathbb{R}^2 \to \mathbb{R}^4 \\ \text{grows linearly with respect to } X_t, \\ \text{and SDE (3.16) has a unique strong solution.} \end{array} \right\}, \tag{3.38}$$

then we can derive the following verification theorem.

Proposition 3.3.5. *For $\gamma > 1$, under the parameters settings in Propostion 3.3.2, the candidate solution $G(t, W_t^{\widetilde{C}}, X_t)$ exists in $[0, T]$ and satisfies $G(t, W_t^{\widetilde{C}}, X_t) = V(t, W_t^{\widetilde{C}}, X_t)$. The strategy (β^*, I^*) given by (3.28) and (3.29) is the optimal portfolio and insurance strategy.*

3.4 Numerical research

3.4.1 Model estimation and calibration

For the financial market, we use monthly U.S. data from June 1961 to December 2020 to estimate the parameters. We use zero-coupon nominal yields from Gürkaynak et al. (2007) with eight maturities: three months, six months, one year, two years, three years, five years, seven years, and ten years. The realized inflation index is obtained from CRSP's Consumer Price Index for All Urban Consumers (CPI-U NSA index). The equity index is based on the CRSP's value-weighted NYSE/Amex/Nasdaq index, which includes dividend payments.

We use a Kalman filter to estimate the parameters (see Appendix B.7 for details) and present the results in Table 3.2, Figure 3.1, and Figure 3.2. Similarly to Koijen et al. (2011), we have $\kappa_1 > \kappa_2$, which means that expected inflation is more persistent than the real short rate. For the innovations, we capture the negative correlation between the real short rate and expected inflation ($\sigma_{2(1)} < 0$). For the equity index process, we find the risk premium is decreasing with the real short rate and expected inflation ($\mu_{1(1)}, \mu_{1(2)} < 0$). Moreover, the unconditional price of risk, Λ_0, is negative for the real short rate and expected inflation but positive for the equity index. Finally, all the parameters in the conditional price of risk, Λ_1, are negative, which means the price of risk is decreasing with two factors X_t.

Figure 3.1 shows the estimated short rates and expected inflation. We see that expected inflation increases rapidly in the 1970s, reaches a historical high during the 1980s, and bounces back in the 1990s, which are the reflections of three oil crises. Then, it slowly decreases in the new century, which has a trend to fall into deflation. For the estimated nominal short rate, we observe that it rises before each financial crisis. The logic is that when the market is booming, an expectation of aggressive tightening monetary policy will be priced in the market, which causes considerable increases in the treasury yield rates. Then, the Federal Reserve will raise the federal funds rate to ensure this expectation. Inversely, when the financial crisis happens, the Federal Reserve will lower the federal funds rate to protect the market. Under such a system, we see that the real short rate

Table 3.2: **Estimation results for the financial market**

Parameters	Estimate	Standard Error	t-statistics	p-value
Average short rate & average expected inflation				
δ_r	0.01256	0.00631	1.99170	0.04641
δ_R	0.05166	0.01336	3.86548	0.00011
δ_{π^e}	0.03879	0.01351	2.87213	0.00408
Two-factor process				
κ_1	0.62591	0.12697	4.92944	0.00000
κ_2	0.19710	0.06846	2.87895	0.00399
$\sigma_{1(1)}$	0.02056	0.00167	12.30697	0.00000
$\sigma_{2(1)}$	-0.00665	0.00156	-4.26618	0.00002
$\sigma_{2(2)}$	0.01476	0.00043	34.59594	0.00000
Realized inflation process				
$\sigma_{\Pi(1)}$	0.00033	0.00050	0.65170	0.51460
$\sigma_{\Pi(2)}$	0.00181	0.00058	3.12965	0.00175
$\sigma_{\Pi(3)}$	0.01286	0.00034	37.87790	0.00000
Equity index process				
μ_0	0.04660	0.03428	1.35949	0.17400
$\mu_{1(1)}$	-1.97908	0.99525	-1.98852	0.04676
$\mu_{1(2)}$	-1.41777	0.54581	-2.59754	0.00939
$\sigma_{S(1)}$	-0.02016	0.00559	-3.60962	0.00031
$\sigma_{S(2)}$	-0.01799	0.00643	-2.79609	0.00517
$\sigma_{S(3)}$	-0.00799	0.00588	-1.35872	0.17424
$\sigma_{S(4)}$	0.15400	0.00301	51.22595	0.00000
Prices of risk of real short rate, inflation, and equity				
$\Lambda_{0(1)}$	-0.00390	—	—	—
$\Lambda_{0(2)}$	-0.17056	0.08355	-2.04136	0.04122
$\Lambda_{0(4)}$	0.28216	—	—	—
$\Lambda_{1(1,1)}$	-9.92622	6.97494	-1.42313	0.15470
$\Lambda_{1(2,2)}$	-9.98032	4.65404	-2.14444	0.03200
$\Lambda_{1(4,1)}$	-14.15060	—	—	—
$\Lambda_{1(4,2)}$	-10.37218	—	—	—

The parameters in the table are annualized. The ones with standard error are directly estimated. The ones without standard error are obtained by solving three equations: $\delta_R = \delta_r + \delta_{\pi^e} - \sigma_\Pi^\top \Lambda_0$, $\sigma_S^\top \Lambda_0 = \mu_0$, $\sigma_S^\top \Lambda_1 = \mu_1$. More details can be found in Appendix B.7.

is controlled in a range and exhibits a significantly mean-reverting pattern. For more details of the monetary policy, we refer to Fed (2001); Fed (2008); and Fed (2020). Lastly, during the pandemic period in 2020, we observe a large drop in the expected inflation due to consumers' shrinking demands. As a result, the real short rate rises even though the nominal short rate is close to zero.

We plot the fitted yield curves for both 3-year and 10-year nominal bonds in Figure 3.2, which are the two nominal bonds used in the later simulations. The max absolute error between the actual and the fitted yield curves is 0.74% for 3-year nominal bonds and 0.76% for 10-year nominal bonds. The results above reassure the accuracy of our Kalman-filter estimation.

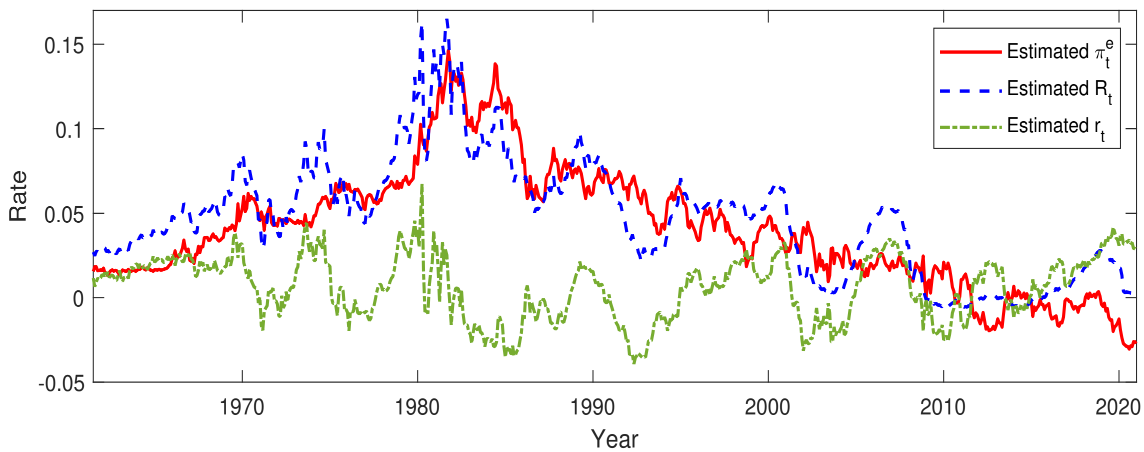
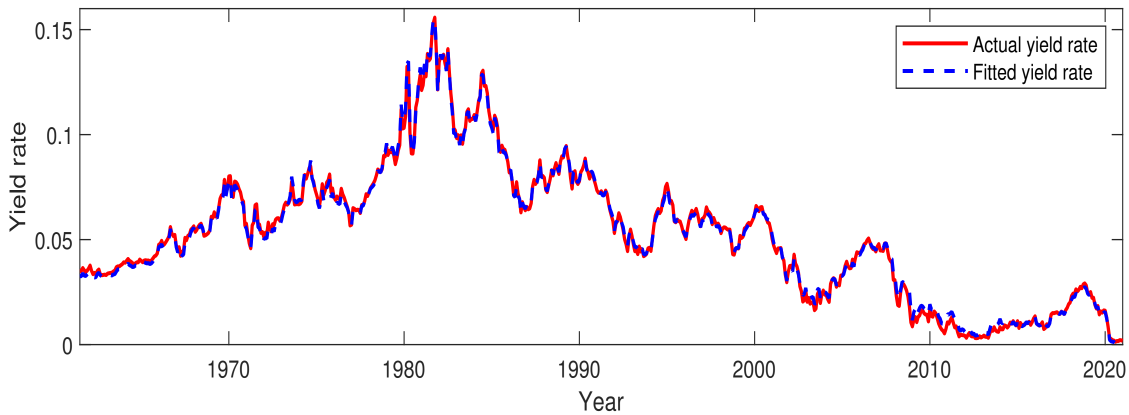


Figure 3.1: Estimated short rate and expected inflation process. The solid line is the estimated expected inflation π_t^e. The dashed line is the estimated nominal short rate R_t. The dash-dotted line is the estimated real short rate r_t.

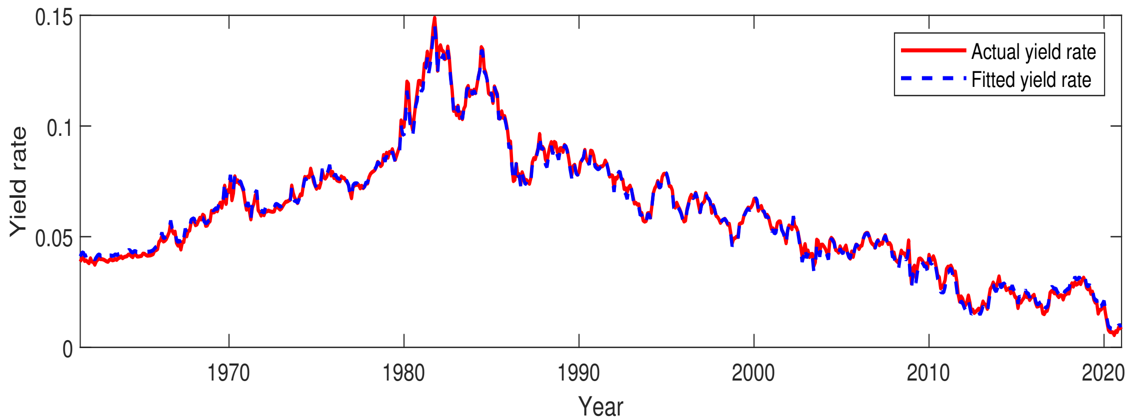
We assume that the pension member enters at age 22 and retires at age 66, which means $T = 44$. The pension member allocates his or her wealth between the 3-year nominal bonds, 10-year nominal bonds, 10-year inflation-linked bonds, the equity index, and cash ($T_1 = 3, T_2 = T_3 = 10$). Besides investing in the financial market, the pension member also purchases life insurance and aims to maximize his or her expected utility at the first time of death and retirement.

Similarly to Koijen et al. (2011), we suppose that the growth rate g_t^R in the real contribution rate (3.8) follows

$$g_t^R = 0.1682 - 0.00646(22 + t) + 0.00006(22 + t)^2, \tag{3.39}$$



(a) Yield rate of 3-year nominal bonds



(b) Yield rate of 10-year nominal bonds

Figure 3.2: Actual and fitted yield curves for 3-year and 10-year nominal bonds. The solid line is the actual yield curve. The dashed line is the fitted yield curve. The max absolute error between the actual and the fitted yield curves is 0.74% for 3-year nominal bonds and 0.76% for 10-year nominal bonds.

which corresponds to an individual with a high school education in the estimates of Cocco et al. (2005) and Munk and Sørensen (2010). The initial real contribution rate C_0 is set to be \$1 k USD.

For the individual's mortality rate, we use the U.S. data of males in "2017 Period Life Table for the Social Security area population". Following Forfar et al. (1988), we set the force of mortality μ_x as a general form of Gompertz-Makeham approach

$$\mu_x = GM_a^{s_1,s_2}(x) = \sum_{i=1}^{s_1} a_i(x-22)^{i-1} + \exp\left\{\sum_{i=s_1+1}^{s_1+s_2} a_i(x-22)^{i-s_1-1}\right\}, \quad 22 \le x \le 67,$$

where the change of location $x-22$ is used to improve the significance of parameters. After some trials, we find a few parameters start to become insignificant when estimating $GM_a^{4,0}$ or $GM_a^{0,5}$. Therefore, we test all the combinations of s_1 and s_2 in 3×4 and pick up the $GM_a^{s_1,s_2}$ model with the lowest Bayesian information criterion (BIC) with all parameters are significant. Table 3.3 shows the estimation results for the force of mortality. We plot the corresponding actual and fitted curves for the survival probability in Figure 3.3. The max absolute error between two curves is 5.85×10^{-4}.

Table 3.3: **Estimation results for the force of mortality**

Model	$GM_a^{3,3}(x)$	BIC	254255.08
Parameters	a_1	a_2	a_3
Values	-1.196773×10^{-3}	-1.406588×10^{-4}	-1.568144×10^{-5}
Standard error	2.51×10^{-5}	1.72×10^{-5}	5.06×10^{-7}
t-statistics	-47.72	-8.20	-30.98
Parameters	a_4	a_5	a_6
Values	-5.956450×10^{0}	9.006499×10^{-2}	-4.710629×10^{-4}
Standard error	2.18×10^{-2}	8.93×10^{-4}	1.40×10^{-5}
t-statistics	-273.26	100.83	-33.73

3.4.2 The change of optimal strategies with age

In this section, we show how the individual's optimal strategies change with age. We use the Monte-Carlo method to do the numerical research, using 10,000,000 simulations and a time-step is one year. We plot the expected annual optimal investment strategies $E[\beta_t^*]$ and insurance strategy $E[I_t^*]$ in Figure 3.4 and Figure 3.5 .

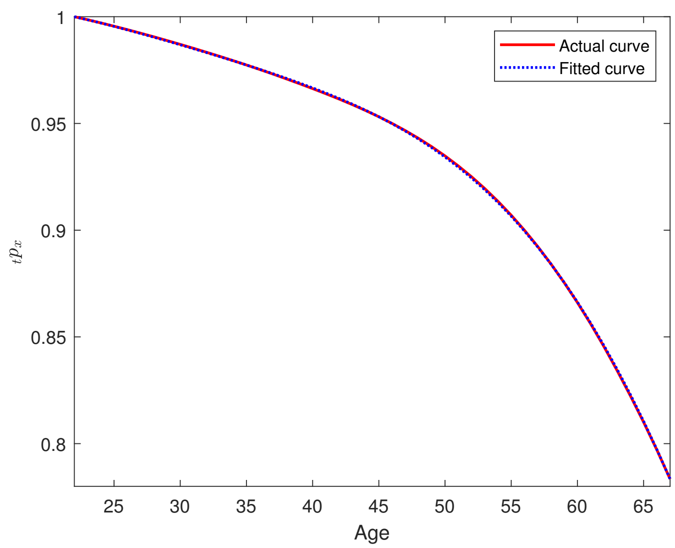


Figure 3.3: Actual and fitted survival probabilities for individual at age 22 (x=22). The solid line is the actual survival probability curve. The dotted line is the fitted survival probability curve. The max absolute error between two curves is 5.85×10^{-4}.

For the investment strategies, the first two graphs in Figure 3.4 illustrate that the individual's allocations on 3-year nominal bonds and 10-year nominal bonds are subject to hump shapes. Specifically, the individual decreases the absolute exposure of these two bonds around age 60. Then, up to retirement, the individual shorts 3-year nominal bonds and holds more 10-year nominal bonds. For the other two financial instruments, the third graph depicts that the individual holds a constant proportion of his or her wealth on 10-year inflation-linked bonds. The individual makes such a decision due to the protection feature of inflation-linked bonds. The fourth graph portrays that the individual holds a smaller proportion of stocks compared with bonds, and his or her preference toward stocks is insensitive to time. Furthermore, it should be mentioned that the third and fourth graphs are based on the expectations, which does not mean individual's allocations stay constant in each path. The sensitivity analysis of the individual's allocations can be found in the next subsection.

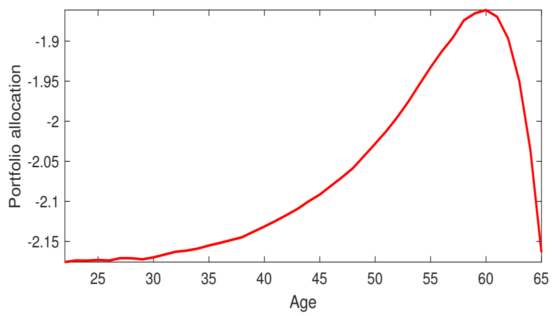
For the insurance strategy, the first graph in Figure 3.5 shows the expected insurance premium paid by the individual. It has a hump shape and reaches its peak at age 59. The second graph displays the expected insurance face value $E[I_t^*]/\mu_{x+t}$. We observe that it also follows a hump shape and reaches its peak at age 50. The last four graphs exhibit the four components in (3.29) that are highly related to the insurance premium's hump shape (the force of mortality μ_{x+t}, the optimal surplus process $(W_t^{\widetilde{C}})^*$, the demand bequest ratio $1/f_1(t, X_t)$, and the future contributions $\widetilde{C}(t, X_t)$). Specifically, the third graph plots the increasing pattern of the expected optimal surplus process $E[(W_t^{\widetilde{C}})^*]$. The larger the surplus process, the larger the purchasing power of the life insurance. As a result, an increasing expected optimal surplus process means an increasing purchase of life insurance. The forth graph presents the increasing feature of the force of mortality μ_{x+t}. Due to the life insurance payment I_t/μ_{x+t}, we see that the larger force of mortality μ_{x+t}, the less attractive the life insurance. Therefore, an increasing force of mortality implies a decreasing demand for life insurance with age. Before moving to the last two graphs, we first recollect the terms in (3.29) and obtain

$$\frac{1}{f_1(t, X_t)} = \frac{(W_t^R)^* + I_t^*/\mu_{x+t}}{(W_t^R)^* + \widetilde{C}(t, X_t)},$$

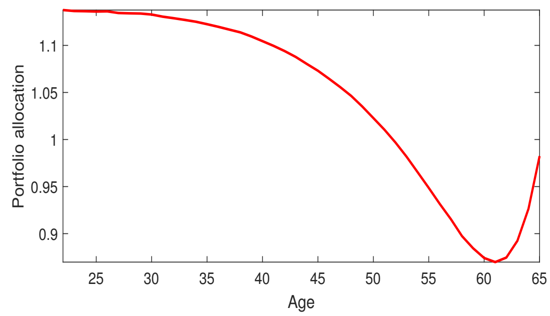
where the right-hand side measures the individual's bequest demand over the current wealth level and future contributions. Therefore, we call $1/f_1(t, X_t)$ the demand bequest ratio. By the control variable method, we see that I_t^* increases with $1/f_1(t, X_t)$ when other components are fixed. That is to say, the larger the demand bequest ratio, the greater the purchase of life insurance. A similar result also applies to the future contributions $\widetilde{C}(t, X_t)$. The optimal insurance premium I_t^* increases with $\widetilde{C}(t, X_t)$ when other components are fixed. Next, when we look back at the last two graphs, we see that the decreasing expected demand bequest ratio $E[1/f_1(t, X_t)]$ and future contributions $E[\widetilde{C}(t, X_t)]$ are just two forces reducing the life insurance purchase. From all the above, we can finally explain the insurance premium's hump shape in the following way. In the individual's early age, the increasing $E[(W_t^{\widetilde{C}})^*]$ dominates the trend. As a result, the increasing purchasing power pulls up the demand for life insurance. In the individual's old age, the increasing μ_{x+t}, decreasing $E[1/f_1(t, X_t)]$, and decreasing $E[\widetilde{C}(t, X_t)]$ dominate the trend. Consequently, these three components drag down the demand for life insurance. Moreover, we observe that the insurance face value's peak comes earlier than the insurance premium's. This early peak is caused by the large increment of the force of mortality μ_{x+t} in old age.

3.4.3 The change of optimal strategies following two factors X_t

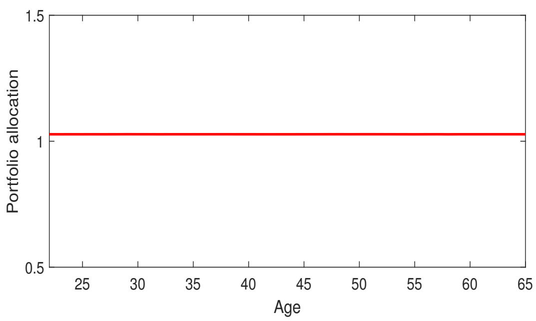
This section conducts the static analysis of optimal strategies with two factors X_t. For all the figures in this section, we set the range of X_1 as $[-0.0736, 0.0736]$ and the range of X_2 as $[-0.1032, 0.1032]$, which cover 8 standard deviations of $X_{1,T}$ and $X_{2,T}$, respectively.



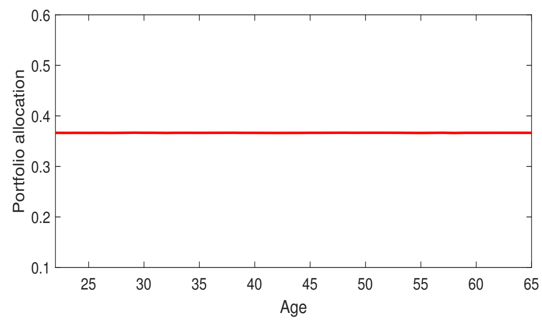
(a) 3-year nominal bonds



(b) 10-year nominal bonds

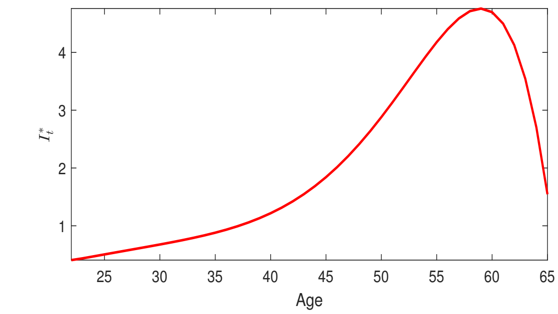


(c) 10-year inflation-linked bonds

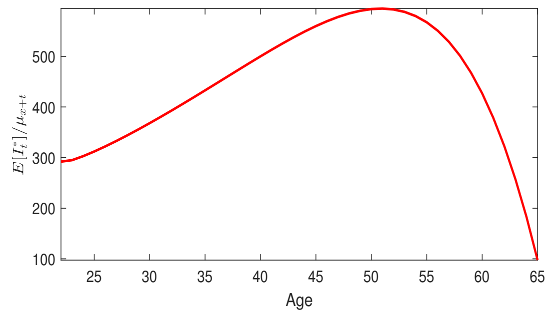


(d) Stocks

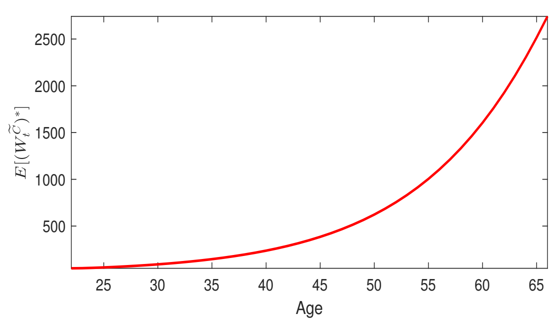
Figure 3.4: Expected annual optimal investment strategies. The risk-aversion coefficient for the individual is $\gamma = 5$. The individual enters the DC pension plan at age 22 and retires at 66. We use the Monte-Carlo method for simulations. The path number is 10,000,000, and the time-step is one year.



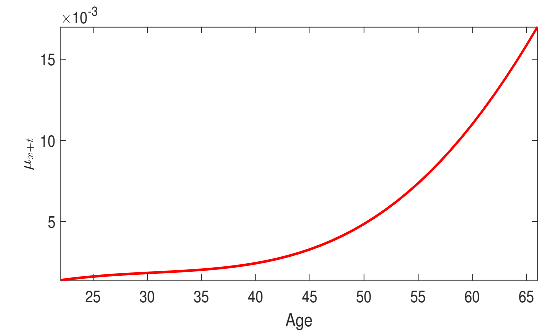
(a) Insurance premium



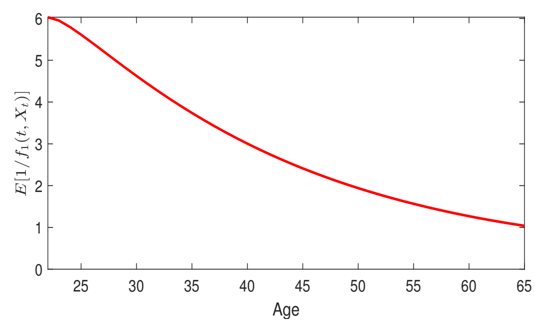
(b) Insurance face value



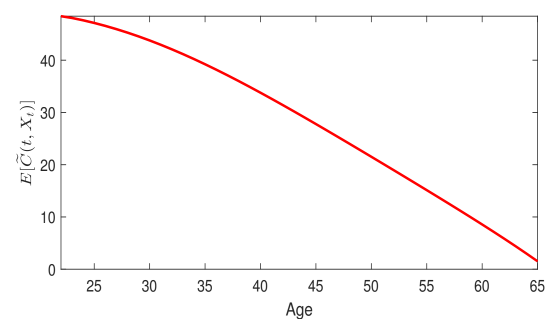
(c) Surplus process



(d) Force of mortality



(e) Demand bequest ratio



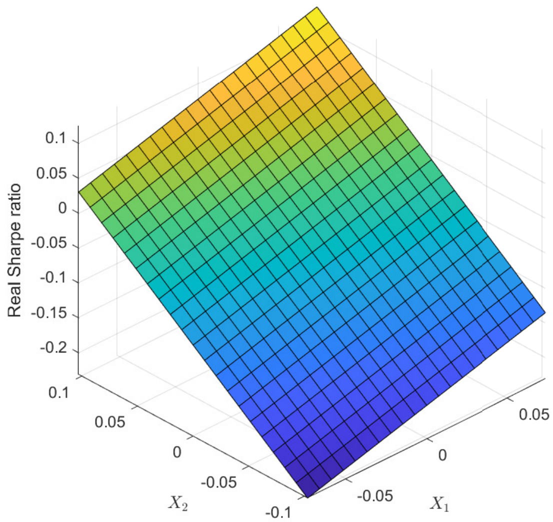
(f) Future contributions

Figure 3.5: Expected annual optimal insurance strategy and its components. The risk-aversion coefficient for the individual is $\gamma = 5$. The individual enters the DC pension plan at age 22 and retires at 66. We use the Monte-Carlo method for simulations. The path number is 10,000,000, and the time-step is one year.

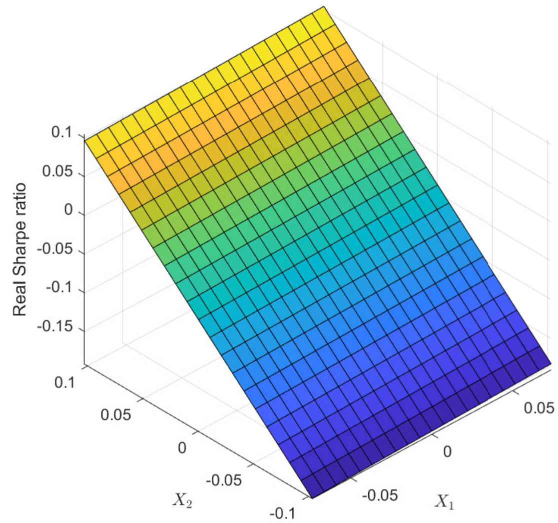
First of all, we draw the real Sharpe ratios for different financial instruments concerning two factors X_t in Figure 3.6. For 3-year nominal bonds, we see that its real Sharpe ratio significantly increases with X_1 and X_2. For 10-year nominal bonds, its real Sharpe ratio slightly increases with real short-rate factor X_1 but sharply increases with inflation factor X_2. The comparison between above two nominal bonds shows that long-term nominal bonds are more sensitive to the inflation risk than short-term nominal bonds. Contrary to the nominal bonds, the real Sharpe ratio of 10-year inflation-linked bonds increases with real short-rate factor X_1 but stays constant with the inflation factor X_2. This pattern is due to the protection design of inflation-linked bonds. Finally, the real Sharpe ratio of stocks decreases dramatically for both factors, which shows the erosion effects on stocks' real risk premium from both the real short rate and expected inflation.

With such an overview of the real Sharpe ratio, we can look into the individual's investment strategies. Figure 3.7 shows the pension member's investment strategy on two factors. Among all three kinds of bonds, we see that the individual set the 3-year nominal bonds as his or her priority. The individual allocates a high ratio of wealth on 3-year nominal bonds and significantly changes its proportion with respect to two factors. Specifically, when the real short-rate factor X_1 increases, the individual sells 10-year nominal and 10-year inflation-linked bonds to purchase more 3-year nominal bonds. When the inflation factor X_2 increases, the individual purchases more 3-year nominal bonds, buys a few 10-year nominal bonds, and sells the inflation-linked bonds. Lastly, the individual sells the stocks when the real short-rate factor X_1 increases or the inflation factor X_2 increases. This decision is caused by the decrements of stocks' real Sharpe ratio when two factors increase.

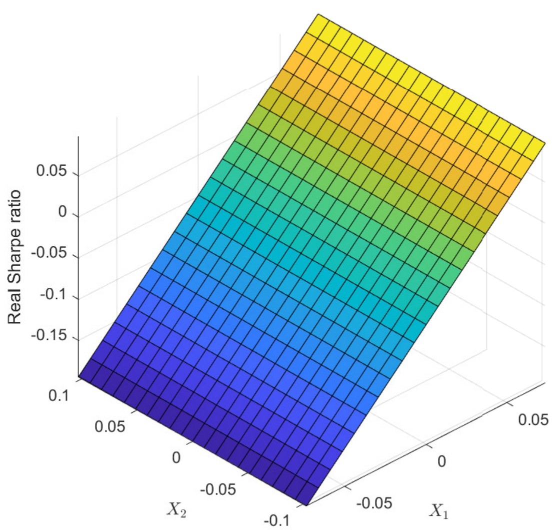
Figure 3.8 reveals that the individual's demand for life insurance follows a "double top" pattern. In other words, the individual purchases more life insurance when the real short rate and expected inflation are both extraordinarily high or both extremely low. This phenomenon is due to the combined effects of the two components in the optimal insurance strategy (3.29). One component is the demand bequest ratio $1/f_1(t, X_t)$. Since $\Gamma_2(\tau) < 0$ (guaranteed by Proposition 3.3.2 and 3.3.3), we have $f(t, X_t)$ in $f_1(t, X_t)$ (see (3.23) and (3.24)) follows a quadratic form opening downwards for X_t on the exponential. Therefore, the demand bequest ratio $1/f_1(t, X_t)$ exhibits a "double top" pattern. The other component is the future contributions $\widetilde{C}(t, X_t)$. It decreases with the real short-rate factor X_1 and stays constant with the inflation factor X_2. We plot the insurance premium with these two components in three ages (22, 59, and 65) in Figure 3.8. At each age, the optimal surplus process in the insurance premium (see expression (3.29)) is set to be the expected surplus process $E[(W_t^{\widetilde{C}})^*]$, which is 48.42, 1462.00, and 2516.00 respectively. This shows that the bequest effect dominates the insurance demand throughout the individual's



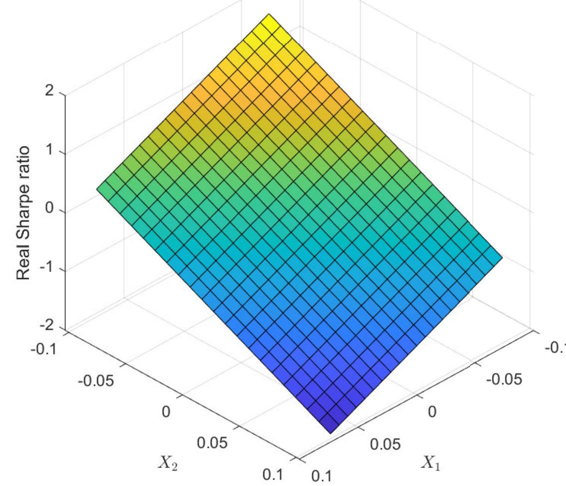
(a) 3-year nominal bonds



(b) 10-year nominal bonds

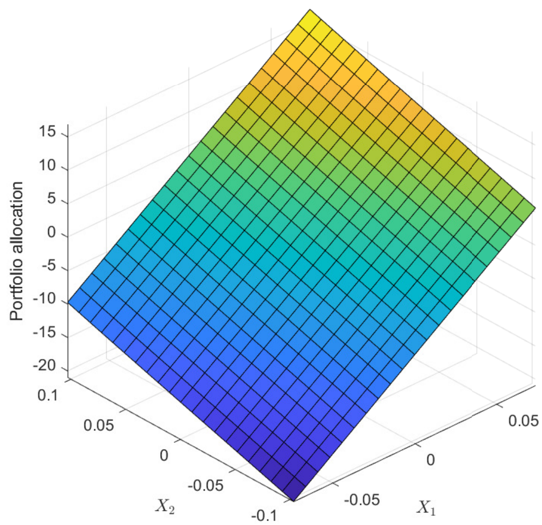


(c) 10-year inflation-linked bonds

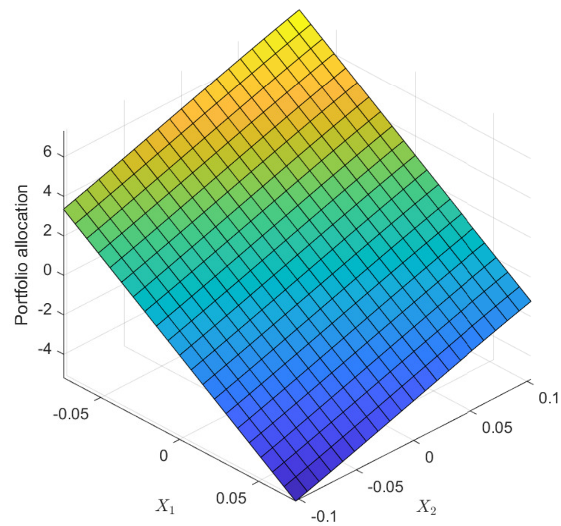


(d) Stocks

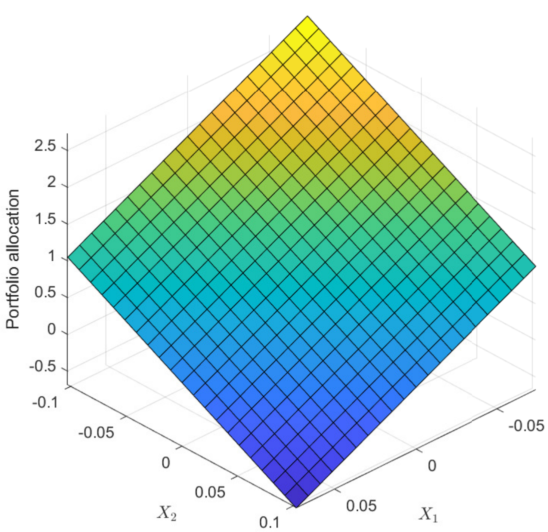
Figure 3.6: Real Sharpe ratios for financial instruments with respect to two factors X_t. The graphs take order in 3-year nominal bonds, 10-year nominal bonds, 10-year inflation-linked bonds, and stocks.



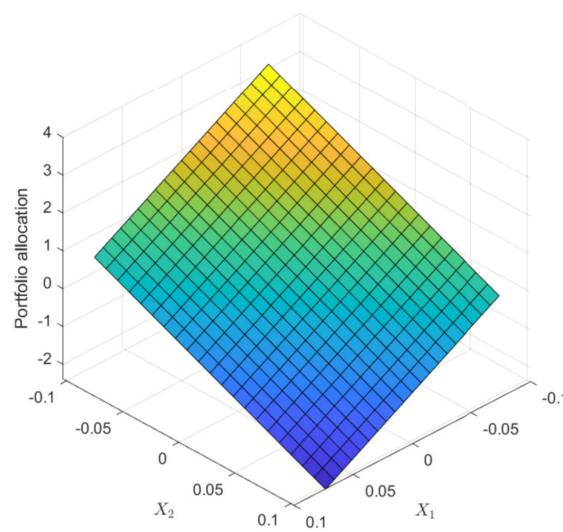
(a) 3-year nominal bonds



(b) 10-year nominal bonds



(c) 10-year inflation-linked bonds



(d) Stocks

Figure 3.7: Individual's investment strategy at $t = T/2$ with respect to two factors X_t. The graphs take order in 3-year nominal bonds, 10-year nominal bonds, 10-year inflation-linked bonds, and stocks.

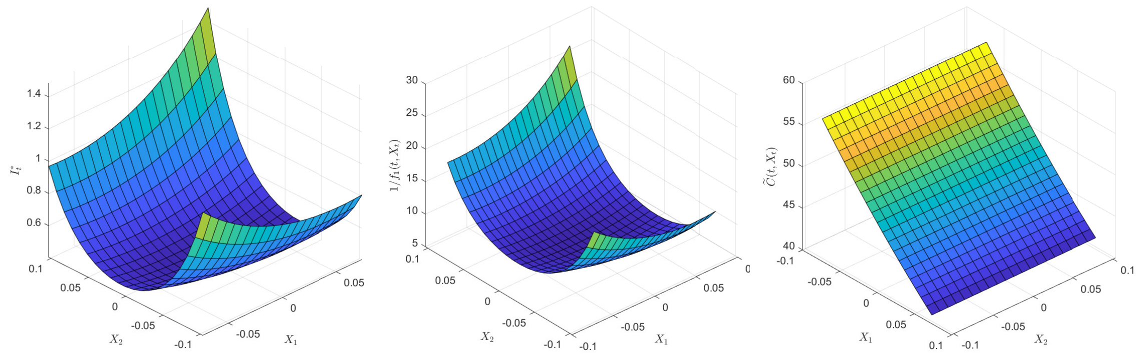
lifetime when comparing $E[(W_t^{\widetilde{C}})^*]/f_1(t, X_t)$ with $\widetilde{C}(t, X_t)$.

3.5 Conclusion

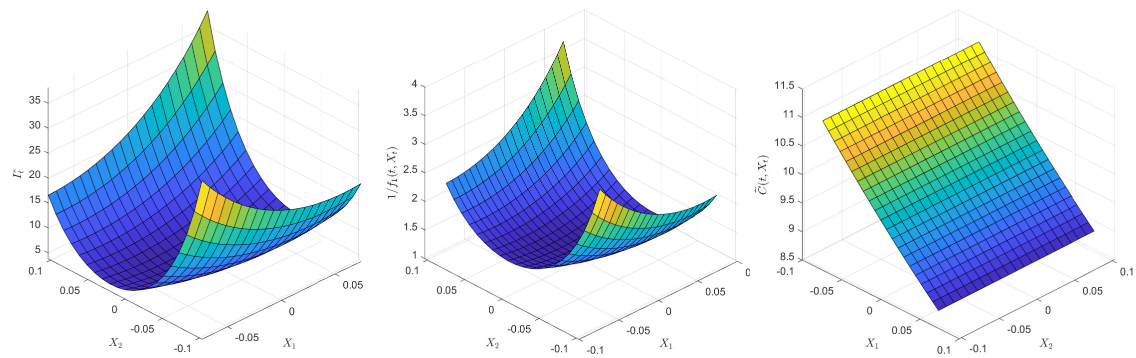
In Chapter 3, we consider a DC pension plan management problem under the two-factor model proposed by Koijen et al. (2011). The financial market is assumed to be complete and contains time-varying real interest rates, inflation rates, and risk premiums. A pension member allocates his or her wealth among a stock index, nominal and inflation-liked bonds, and a nominal cash account. In addition, the pension member can also purchase life insurance to hedge the mortality risk before retirement.

We formulate this pension management problem by an HJB equation and derive its explicit solution under the CRRA utility. To complete the analysis, we also prove the explicit solution's global existence and verification theorem. Besides technical proofs, we use the Kalman filter method to calibrate our model with the real market data. Finally, both dynamic and static simulations are provided to study the pension member's investment strategy and insurance demand.

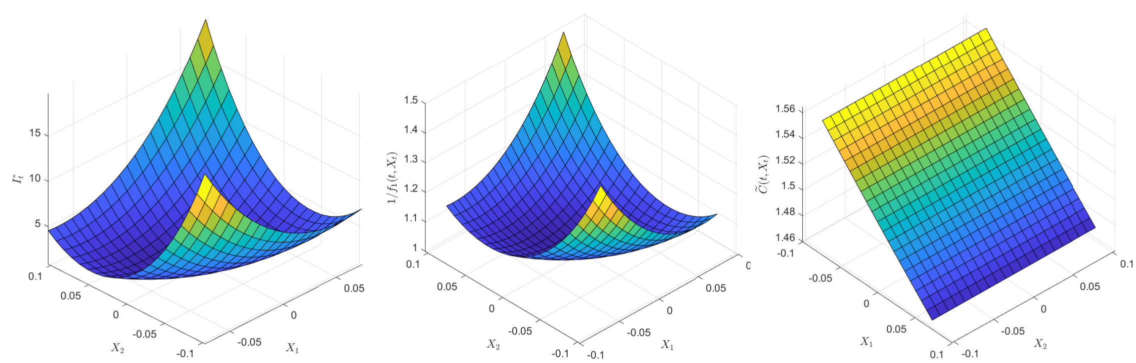
Our numerical research shows that the individual's demand for life insurance exhibits a hump shape with age and a "double top" pattern for two factors. To be specific, the individual purchases more insurance in the old age before retirement or the extreme market scenarios that the real short rate and expected inflation are both high or both low. These behaviors are caused by the combined effects of the components in the optimal insurance premium. Furthermore, our model builds a DC account that resembles a variable annuity with endogenously determined time-varying death benefits. It relaxes the constraints on variable annuity's death benefits and can inspire more innovations in creating new actuarial products.



(a) Insurance premium (b) Demand bequest ratio (c) Future contributions



(d) Insurance premium (e) Demand bequest ratio (f) Future contributions



(g) Insurance premium (h) Demand bequest ratio (i) Future contributions

Figure 3.8: Individual's optimal insurance with respect to two factors X_t. The figures in the same row share the same age, 22, 59, and 65, respectively. In each row, the left figure is the optimal insurance premium I_t^*. The middle figure is the demand bequest ratio $1/f_1(t, X_t)$. The right figure is the future contributions $\widetilde{C}(t, X_t)$.

Chapter 4

Constrained portfolio optimization in a life-cycle model

4.1 Introduction

The constrained portfolio optimization problem is an extension of the classical portfolio allocation problem. It considers trading constraints, such as non-tradable assets (incomplete market), no short-selling constraint, no borrowing constraint, etc., and hence adjusts the ideal model to a more realistic market model. Compared to the classical problem, the constrained problem does not always have an explicit solution. The incompleteness caused by the trading constraints removes the uniqueness of the martingale measure and leaves the traditional martingale approach inadequate.

Several seminal papers generalize the martingale approach via the convex duality method. Karatzas et al. (1991) propose a "fictitious completion" method to deal with the portfolio optimization problem in the incomplete market. They introduce additional stocks and build a "fictitious" complete market. By manipulating the drift term of these additional stocks, they can guarantee that the individual will not invest in them in the original complete market. Cvitanić and Karatzas (1992) study a general constrained portfolio problem in which the proportion invested in risky asset π belongs to a non-empty, closed, and convex set K. By a dual control method, they construct a group of artificial markets that can invest without trading constraints, which provides the upper bounds of the primal problem. Finally, they prove the optimal strategy under the smallest artificial market is the optimal strategy feasible for the primal problem. Their framework contains an incomplete market, no short-selling, and no-borrowing constraints as special cases. He and Pages (1993) add labor income to the constrained portfolio optimization problem. They use a dual control approach and transform a no-borrowing problem into a variational inequality in the dual space. Several examples of deterministic labor income have been studied in their paper. Cuoco (1997) extends Cvitanić and Karatzas (1992) to the case with stochastic income. He focuses on the optimal amount instead of the optimal proportion allocating among the assets and includes He and Pages (1993)'s work (no-borrowing constraint) as special cases. For more recent work, we refer to Bick et al. (2013); Chabakauri (2013); Haugh et al. (2006); Jin and Zhang (2013); Kamma and Pelsser (2022); Larsen and Žitković (2013); Mostovyi and Sîrbu (2020).

In the actuarial science field, more and more researchers apply the constrained portfolio optimization problem to deal with trading constraints and unhedgeable health shocks in an individual's lifetime investment. Zeng et al. (2016) extend He and Pages (1993)'s work to the actuarial field and study the wealth-constraint effect on the life insurance purchase. Dong and Zheng (2019) use a dual control method to study the optimal defined contribution pension management under short-selling constraints and portfolio insurance. Hambel et al. (2022) build a group of artificial insurance markets to solve a life-cycle model with unhedgeable biometric shocks. However, most existing actuarial literature only focuses on one or two trading constraints, and a general framework is lacking in the content of studying the life-cycle investment.

This chapter considers a constrained portfolio optimization problem in a generalized life cycle model. The individual has a stochastic income and aims to find the optimal trading and insurance strategies to maximize his or her expected consumption utility plus bequest utility and terminal wealth utility. Inspired by the existing literature, we restrict the trading strategy to a non-empty, closed, and convex set, which contains many trading constraints (non-tradeable asset constraint, no short-selling constraint, no borrowing constraint, portfolio mix constraint) as special cases. Following Cuoco (1997)'s framework, we build a group of artificial markets by adding compensations to the drift terms of stocks and bonds. Due to the lack of uniqueness of martingale measures under trading constraints, we first derive a group of static budget constraints from the individual's wealth process. Then, a dual problem is obtained through the Lagrangian dual control method, which is an upper bound for the primal problem. Furthermore, a one-to-one relationship is proved between the optimal solutions of the primal problem and the dual problem. More specifically, once the optimal solution exists for one problem, the optimal solution for the other problem exists and can be obtained immediately. Lastly, due to the stochastic income process, the dual problem is not convex, which causes great difficulty in proving the existence of optimal strategies by the dual control approach. Fortunately, Levin (1976) uses the "relaxation projection" technique and proves the existence of solution under the non-reflexive spaces. To utilize their theorem, we only need to verify that our objective function is lower semi-continuous and that the trading constraint set is convex, topologically closed, and norm-bounded.

It seems that the dual problem does not play an essential role in proving the existence of the optimal strategies. However, since it is a tight upper bound for the primal problem, minimizing the dual problem provides an excellent approximation to the primal problem. Bick et al. (2013) propose a simulation of artificial markets strategies (SAMS) method to compute the lower and upper bounds of the primal problem. Their artificial market is characterized by the adjustment of the drift terms of stocks and bonds, which is denoted as $v(t)$. They restrict $v(t)$ to be affine in time and minimize the artificial market with affine $v(t)$ to get the lowest upper bound. Finally, a lower bound is obtained by deriving a candidate strategy from the lowest upper bound and substituting the candidate strategy into the wealth process. The deficiency of the SAMS method is apparent. The artificial market is constrained to a subfamily of affine $v(t)$, and the gap between the lower and upper bounds always exists. To overcome this difficulty, we introduce a neural network to study the best form of $v(t)$. We find that when the risk-free interest rate, stock appreciation rate, and volatility are all constant, the SAMS method and neural network performance are very close. If the stock appreciation rate follows a perturbation in time, the SAMS is inadequate to solve the problem, and the gap between the lower and upper bounds is enormous. However, the neural network $v(t)$ can learn the perturbation pattern very well and provides tight lower and upper bounds with a small gap. Last but not least, both methods show that when considering trading constraints, the individual will reduce his or her demand for life insurance.

To the best of our knowledge, this is the first application of neural network to compute the best trading and insurance strategies for a constrained portfolio optimization problem. We make three contributions to the existing literature: First, we study the constrained portfolio optimization problem in a life cycle model with stochastic income and insurance provided. A general dual control framework is constructed, and the existence of the primal problem is proved. Second, we relax the assumptions in Cuoco (1997) and extend their work to a more general case. Cuoco (1997) assumes the interest rate process is uniformly bounded, and the integral of discounted stochastic income is uniformly bounded. In our work, we assume the expected exponential integral of the interest rate's absolute value is finite and gives a weaker condition on the income process, which contains the uniform bounded income process as a special case. Third, we first propose a dual control neural network approach to compute the constrained life cycle model and find that the individual will reduce his or her demand for life insurance when considering the trading constraints. Compared to Bick et al. (2013), our approach can solve more challenging cases, such as the stock return has a perturbation in time. It can inspire future work to use neural network learning the best solution for the constrained portfolio optimization problem.

The rest of the chapter is organized in the following order: Section 4.2 introduces our model settings of the financial market, insurance market, wealth process, preference, and trading constraint set. Section 4.3 explains the construction of the artificial market and derives the static budget constraint for the wealth process. Section 4.4 describes the Lagrangian dual control approach and proves the one-to-one relationship between the primal problem and the dual problem. Section 4.5 proves the existence of the primal problem. Section 4.6 conducts the numerical simulation and compares our algorithm with existing literature. Section 4.7 concludes. All proofs are relegated to Appendix C.

4.2 Model settings

We consider a constrained portfolio optimization problem in a generalized life cycle model. The model contains three important dates, a random death time T_x (defined later), a deterministic retirement time T_R, and a deterministic time horizon of the family T. During the decision period $[0, T \wedge T_x)$, where $T \wedge T_x = \min(T, T_x)$, the individual is allowed to purchase stocks, a bond, and life insurance to improve his or her consumption level, death benefit, and the terminal wealth.

4.2.1 Financial market

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a filtered complete probability space. The financial risk is described by a n-dimensional Brownian motion Z_t adapted to the filtration $\mathbb{F} = \{\mathcal{F}_t\}_{t\in[0,T]}$.

In the financial market, there are $n+1$ assets. The first asset is the bond which is locally risk free and pays no dividends. Its price process is given by

$$B_t = \exp\left(\int_0^t r_s ds\right), \tag{4.1}$$

where r_t is the interest rate process generated by Z_t.

Assumption 4.2.1. *The interest rate process r_t satisfies*

$$E\left[\exp\left(\int_0^T |r_t| dt\right)\right] < \infty,$$

where $|\cdot|$ means the absolute value.

Remark 4.2.1. *Assumption 4.2.1 implies* $\exp\left(\int_0^T |r_t|dt\right) < \infty$ *almost every where. Because the expectation is finite, it implies that the random variable is finite almost everywhere. We directly use this corollary without mention in the appendixes' proofs.*

The price process of the risky assets are $S = (S_1, ..., S_n)$ with a cumulative dividend process $D = (D_1, ..., D_n)$ satisfying the Ito process

$$S_t + D_t = S_0 + \int_0^t I_{S,u}\mu_u du + \int_0^t I_{S,u}\sigma_u dZ_u,$$

where $I_{S,t}$ denotes the $n \times n$ diagonal matrix with element S_t and

$$\int_0^T |I_{S,t}\mu_t|dt + \int_0^T |I_{S,t}\sigma_t|^2 dt < \infty.$$

Assumption 4.2.2. *The volatility matrix* σ_t *satisfies the nondegneracy condition*

$$x^\top \sigma_t \sigma_t^\top x \geq \epsilon |x|^2, P\text{-}a.s.$$

for any $(x, t) \in \mathbb{R}^2 \times [0, T]$ *and* $\epsilon > 0$*. Moreover, denote the market price of risk vector by*

$$\kappa_{0,t} = -\sigma_t^{-1}(\mu_t - r_t \bar{1}_n),$$

where $\bar{1}_n = (1, ..., 1)^\top \in \mathbb{R}^n$*, we assume a Novikov condition*

$$E\left[\exp\left(\frac{1}{2}\int_0^T |\kappa_{0,t}|^2 dt\right)\right] < \infty.$$

in order to ensure the existence of an equivalent martingale measure.

4.2.2 Mortality

Denote by T_x, the future life time of the individual aged x, which is a random variable independent of the filtration $\mathbb{F}$ in the financial market. Then, we can introduce the following actuarial notations

$$\begin{aligned}
{}_t p_x &= \mathbb{P}[T_x > t], \\
{}_t q_x &= \mathbb{P}[T_x \leq t] = 1 - {}_t p_x, \\
\lim_{t\to\infty} {}_t p_x &= 0, \lim_{t\to\infty} {}_t q_x = 1,
\end{aligned}$$

where ${}_tp_x$ is the probability that the individual alive at age x survives to at least age $x+t$, ${}_tq_x$ is the probability that the individual aged x dies before $x+t$. Following actuarial practice, we also define the force of mortality (hazard rate)

$$\lambda_{x+t} = \frac{1}{{}_tp_x}\frac{d}{dt}{}_tq_x = -\frac{1}{{}_tp_x}\frac{d}{dt}{}_tp_x. \tag{4.2}$$

Then, the survival and death probabilities can be rewritten as

$$\begin{aligned} {}_tp_x &= \exp\left\{-\int_0^t \lambda_{x+s}ds\right\}, \\ {}_tq_x &= \int_0^t {}_sp_x\lambda_{x+s}ds. \end{aligned}$$

The probability density function of T_x satisfies

$$f_{T_x}(t) = {}_tp_x\lambda_{x+t}, \text{ for } t > 0.$$

4.2.3 Wealth process

At time 0, the individual at age x starts to manage portfolio until the first time of the death time T_x and the family's time horizon T. Denoted the retirement time as T_R. Before death time T_x and the retirement time $T_R < T$, the individual receives a stochastic non-negative income Y_t generated by Z_t.

Define the trading strategy (α, θ) under the price coefficients $\mathcal{P}(r, \mu, \sigma)$, where α and θ_k represent the money amounts invested at time t in the bond and $k-$th risky asset, respectively. A trading strategy is called admissible if

$$\int_0^T |\alpha_t r_t|dt + \int_0^T |\theta_t^\top \mu_t|dt + \int_0^T |\theta_t^\top \sigma_t|^2 dt < \infty. \tag{4.3}$$

We use Θ to denote the admissible set of trading strategies. Before the individual's death or the family's time horizon, the wealth process satisfies

$$W_t = \alpha_t + \sum_{k=1}^n \theta_{k,t}, \ 0 \le t < \min(T_x, T), \tag{4.4}$$

$$W_t = w_0 + \int_0^t (\alpha_s r_s + \theta_s^\top \mu_s)ds + \int_0^t \theta_s^\top \sigma_s dZ_s - \int_0^t (c_s + I_s - Y_s)ds - C_t, \tag{4.5}$$

$$W_t \ge -K, \ K \in \mathbb{R}^+, \tag{4.6}$$

$$W_T \ge 0, \tag{4.7}$$

where c_t is the consumption rate, I_t is the life insurance premium, and C_t is the free disposal of wealth. Free disposal of wealth is the amount of money the individual chooses not to reinvest up to time t. We show when this free disposal of wealth disappears in Corollary 4.3.1. Equation (4.5) is usually called the "dynamic budget constraint". Equations (4.6) and (4.7) show that the individual is allowed to borrow against the future income but needs to pay the debt at the terminal time. Lastly, equation (4.6) admits a uniform lower bound to eliminate the arbitrage opportunity, such as the doubling strategy in Harrison and Kreps (1979). At the death time T_x, the individual's wealth has a jump from the insurance payment

$$W_{T_x} = W_{T_x-} + \frac{I_{T_x}}{\lambda_{x+T_x}},$$

where λ_t is the force of mortality defined in (4.2).

4.2.4 Preference and feasibility

The individual's objective is to choose an investment and insurance strategy (α, θ, I) to optimize the expected utility of consumption when the individual is alive, the wealth level at the death time, or the terminal wealth at the family's time horizon,

$$\sup_{(\alpha,\theta)\in A, I} E\left[\int_0^T U_1(c_t, t)\mathbb{1}_{\{t<T_x\}}dt + U_2\left(W_{T_x}, T_x\right)\mathbb{1}_{\{T_x<T\}} + U_3(W_T, T)\mathbb{1}_{\{T_x\geq T\}}\right],$$

where A is the portfolio constraint set in $\mathbb{R}^{n+1}$, U_1 is the consumption utility, U_2 is the bequest utility, and U_3 is the terminal utility. We assume all the utilities satisfy the following properties.

Definition 4.2.1. *Utility functions $U_i : (0, \infty) \times [0, T] \to \mathbb{R}, i = 1, 2, 3$ are increasing, strictly concave, and continuously differentiable in its first variable and continuous in the second variable.*

Since the individual's time to death T_x is independent of the filtration $\mathbb{F}$ in the financial market, we have the equivalent preference

$$
\begin{aligned}
& \sup_{(\alpha,\theta)\in A,I} E\left[\int_0^T U_1(c_t,t)\mathbb{1}_{\{t<T_x\}}dt + U_2\left(W_{T_x},T_x\right)\mathbb{1}_{\{T_x<T\}} + U_3(W_T,T)\mathbb{1}_{\{T_x\geq T\}}\right] \\
= & \sup_{(\alpha,\theta)\in A,I} E\left[\int_0^T {}_tp_x U_1(c_t,t)dt + \int_0^T f_{T_x}(t)U_2\left(W_t+\frac{I_t}{\lambda_{x+t}},t\right)dt \right.\\
& \left. + \int_T^\infty f_{T_x}(t)U_3(W_T,T)dt\right] \\
= & \sup_{(\alpha,\theta)\in A,I} E\left[\int_0^T {}_tp_x U_1(c_t,t)dt + \int_0^T {}_tp_x\lambda_{x+t}U_2\left(W_t+\frac{I_t}{\lambda_{x+t}},t\right)dt + {}_Tp_xU_3(W_T,T)\right] \\
:= & \sup_{(\alpha,\theta)\in A,I} E\left[\int_0^T {}_tp_x U_1(c_t,t)dt + \int_0^T {}_tp_x\lambda_{x+t}U_2\left(M_t,t\right)dt + {}_Tp_xU_3(W_T,T)\right], \qquad (4.8)
\end{aligned}
$$

where $M_t = W_t + \frac{I_t}{\lambda_{x+t}}$.

Before moving to the feasibility of strategies, we first define the consumption and bequest set. Consider the set G

$$
G := \left\{(c,M,W_T) : E^{Q_0}\left[\int_0^T |c_t| + |M_t|\,dt + |W_T|\right] < \infty, \text{P-a.s.}\right\}, \qquad (4.9)
$$

where Q_0 is the risk neutral measure such that $dZ_{0,t} = dZ_t - \kappa_{0,t}dt$ is a Brownian motion (see Assumption 4.2.2). Let G_+ denote the orthant of (c,M,W_T) that $c_t \geq 0$, $M_t \geq 0$, and $W_T \geq 0$, then we can define the individual consumption and bequest set G_+^* as the plan $(c,M,W_T) \in G_+$ satisfying

$$
\begin{aligned}
\min\left(E\left[\int_0^T U_1(c_t,t)^+dt\right], E\left[\int_0^T U_1(c_t,t)^-dt\right]\right) < \infty, \\
\min\left(E\left[\int_0^T U_2(M_t,t)^+dt\right], E\left[\int_0^T U_2(M_t,t)^-dt\right]\right) < \infty,
\end{aligned}
$$

and

$$
\min\left(E\left[U_3(W_T,T)^+\right], E\left[U_3(W_T,T)^-\right]\right) < \infty.
$$

Thus, the expectation of utility is well defined in $[-\infty,+\infty]$.

Given price coefficients $\mathcal{P} = (r,\mu,\sigma)$, a consumption and bequest plan $(c,M,W_T) \in G_+^*$ is called "feasible" if there exists an admissible trading strategy $(\alpha,\theta) \in \Theta$ for $\forall t \in [0,T]$, and a non-negative increasing free disposal C satisfying the dynamic budget constraint from (4.4) to (4.7). In addition, the plan $(c, M, W_T) \in G^*_+$ is said to be "A-feasible" if it is feasible and $(\alpha, \theta) \in A$ for $\forall t \in [0, T]$. In both cases, the trading strategy is said to "finance" (c, M, W_T). We use $\mathcal{B}(\mathcal{P}, A)$ to denote the set of A-feasible consumption and bequest plan given the pricing coefficient $\mathcal{P}$.

4.2.5 Portfolio constraint set

We assume that the agent's portfolio (α, θ) is constrained to take values in a portfolio constraint set A, which is a non-empty, closed, and convex subset of $\mathbb{R}^{n+1}$. It can describe various trading constraints such as short-sale prohibitions, non-tradeable asset, or minimal capital requirement. For $v = (v_0, v_-) \in \mathbb{R} \times \mathbb{R}^n$, define

$$\delta(v) = \sup_{(\alpha,\theta)\in A} -(\alpha v_0 + \theta^\top v_-), \tag{4.10}$$

which is the support function of $-A$. This function can easily reach $+\infty$ and hence it is important to define its effective domain as

$$\widetilde{A} = \left\{ v \in \mathbb{R}^{n+1} : \delta(v) < \infty \right\}.$$

In the convex analysis, it is well-known that δ is a positively homogeneous, lower semi-continuous, and proper convex function on $\mathbb{R}^{n+1}$ and $\widetilde{A}$ is a closed convex cone. We assume the support function satisfies the following constraint

Assumption 4.2.3. *The function δ is upper semi-continuous and bounded above on $\widetilde{A}$. Moreover, $v_0 \geq 0$ for all $v \in \widetilde{A}$.*

$v_0 \geq 0$ for all $v \in \widetilde{A}$ is immediately obtained if $(\alpha, 0) \in A$ for any α large enough, i.e., as long as lending and investing nothing in the risky assets is admissible. Moreover, since δ is positively homogeneous and $\widetilde{A}$ is a cone, the function δ bounded above on $\widetilde{A}$ is equivalent to δ being non-positive on $\widetilde{A}$. Specifically, if A is a cone, then $\delta \equiv 0$ on $\widetilde{A}$. Below, we provide some examples of constraint sets A satisfying Assumption 4.2.3, together with the associated support functions and dual sets.

(a) No constraints:

$$\begin{aligned} &A = \mathbb{R}^{n+1}, \\ &\widetilde{A} = \{0\}, \\ &\delta(v) = 0 \text{ for } \forall v \in \widetilde{A}. \end{aligned}$$

This problem is well-studied in Karatzas et al. (1987), Cox and Huang (1989), and Cox and Huang (1991).

(b) Nontradeable assets (incomplete market):

$$\begin{aligned} &A = \{(\alpha, \theta) \in \mathbb{R}^{n+1} : \theta_k = 0,\ k = m+1, ..., n\}, \\ &\widetilde{A} = \{v \in \mathbb{R}^{n+1} : v_k = 0,\ k = 0, ..., m\}, \\ &\delta(v) = 0 \text{ for } v \in \widetilde{A}. \end{aligned}$$

For the case without stochastic income, He and Pearson (1991) and Karatzas et al. (1991) solve the problem using martingale techniques.

(c) Short-sale constraint

$$\begin{aligned} &A = \{(\alpha, \theta) \in \mathbb{R}^{n+1} : \theta_k \geq 0,\ k = m+1, ..., n\}, \\ &\widetilde{A} = \{v \in \mathbb{R}^{n+1} : v_k = 0,\ k = 1, ..., m; v_k \geq 0,\ k = m+1, ..., n\}, \\ &\delta(v) = 0 \text{ for } v \in \widetilde{A}. \end{aligned}$$

Xu and Shreve (1992) study this problem without an income stream.

(d) Buying constraints

$$\begin{aligned} &A = \{(\alpha, \theta) \in \mathbb{R}^{n+1} : \theta_k \leq 0,\ k = m+1, ..., n\}, \\ &\widetilde{A} = \{v \in \mathbb{R}^{n+1} : v_k = 0,\ k = 1, ..., m; v_k \leq 0,\ k = m+1, ..., n\}, \\ &\delta(v) = 0 \text{ for } v \in \widetilde{A}. \end{aligned}$$

(e) Portfolio-mix constraint

$$A = \left\{ (\alpha, \theta) \in \mathbb{R}^{n+1} : \alpha + \sum_{k=1}^{n} \theta_k \geq 0,\ \theta \in D\left(\alpha + \sum_{k=1}^{n} \theta_k\right) \right\},$$

where D is any nonempty, closed, convex subset of $\mathbb{R}^n$ containing the origin,

$$\begin{aligned} &\widetilde{A} = \{v \in \mathbb{R}^{n+1} : v^\top(\alpha, \theta) \geq 0,\ \forall(\alpha, \theta) \in A\}, \\ &\delta(v) = 0 \text{ for } v \in \widetilde{A}. \end{aligned}$$

The problem without an income stream and hence a nonbinding nonnegativity constraint on wealth is examined in Cvitanić and Karatzas (1992).

(f) Minimum capital requirement

$$A = \left\{ (\alpha, \theta) \in \mathbb{R}^{n+1} : \alpha + \sum_{k=1}^{n} \theta_k \geq K \right\},$$

where $K \geq 0$,

$$\widetilde{A} = \{k\bar{1}_n : k \geq 0\},$$
$$\delta(v) = -Kv_0 \text{ for } v \in \widetilde{A}.$$

This constraint covers the special cases such as the "borrowing constraint" which is studied in He and Pages (1993) for $K = 0$ and "portfolio insurance constraint" which is studied in Bardhan (1994) and Basak (1995) for $K > 0$.

(g) Collateral constraints

$$A = \left\{ (\alpha, \theta) \in \mathbb{R}^{n+1} : \Psi_0 \alpha + \sum_{k=1}^{n} \Psi_k \theta_k \geq \gamma (\Psi_0 \alpha^+ + \sum_{k=1}^{n} \Psi_k \theta_k^+) \right\},$$

where $\Psi_k \in [0, 1]$ for $k = 0, 1, ..., n$ denotes the fraction of the amount of asset k can be borrowed using the asset as collateral and $\gamma \in [0, 1]$,

$$\widetilde{A} = \{v \in \mathbb{R}^{n+1} : v^\top (\alpha, \theta) \geq 0, \forall (\alpha, \theta) \in A\},$$
$$\delta(v) = 0 \text{ for } v \in \widetilde{A}.$$

This constraint is introduced by Hindy (1995) who consider the viable pricing operator. Hindy and Huang (1995) study the optimal investment problem in a discrete-time setting in which $\gamma = 0$.

(h) Any combination of above constraints.

4.3 Artificial market and static budget constraint

Following Cuoco (1997), we define the artificial market to solve the constrained portfolio optimization. Given a constraint set A, let $\mathcal{N}$ denote the $\widetilde{A}$ valued process satisfying

$$E\left[\int_0^T |v_t|^2 dt\right] < \infty.$$

For each $v \in \mathcal{N}$, the processes

$$\begin{aligned}
\beta_{v,t} &= \exp\left(-\int_0^t r_s + v_{0,s} ds\right), \\
\kappa_{v,t} &= -\sigma_t^{-1}(\mu_t + v_{-,t} - (r_t + v_{0,t})\bar{1}_n), \\
\xi_{v,t} &= \exp\left(\int_0^t \kappa_{v,s}^\top dZ_s - \frac{1}{2}\int_0^t |\kappa_{v,s}|^2 ds\right), \\
\pi_{v,t} &= \beta_{v,t}\xi_{v,t}, && (4.11) \\
dZ_{v,t} &= dZ_t - \kappa_{v,t} dt, && (4.12)
\end{aligned}$$

define an artificial market $\mathcal{M}_v$, where ξ_v is a strictly positive local martingale. We further use $\mathcal{N}^*$ to denote the subset of elements v in $\mathcal{N}$ for which ξ_v is exactly a martingale. Note that $\mathcal{N}^*$ is nonempty given the Novikov condition and the fact that $\widetilde{A}$ is a cone ensuring that $0 \in \mathcal{N}^*$. Then, each $\pi_{v,t}, v \in \mathcal{N}^*$ can be interpreted as the unique state-price density in a fictitious unconstrained market $\mathcal{M}_v$ with price coefficients $\mathcal{P} = (r + v_0, \mu + v_-, \sigma)$. With the adjustment of drift term by $v = (v_0, v_-)$, the stocks can become more attractive or less attractive compared to the bond. Then, "A-feasible" trading strategies can be built by the change of individual's preference between stocks and the bond. More generally, each $\pi_{v,t}$ with $v \in \mathcal{N}^*$ constitutes an arbitrage-free state-price density in the original economy when the portfolio policies are constrained to be in A, and that the fulfilment of a budget constraint with respect to all of these state-price densities is sufficient to guarantee the A-feasibility.

To satisfy the lower boundedness property (4.6) of wealth process W_t, we add the following assumption to the income process Y_t

Assumption 4.3.1.

$$\sup_{v \in \mathcal{N}^*} E^{Q_v}\left[\int_0^T e^{-\int_0^t r_s + \lambda_{x+s} ds} Y_t dt\right] \le K_y, \tag{4.13}$$

for some positive constant $K_y > 0$.

Assumption 4.3.1 includes the uniformly bounded income case studied in Cuoco (1997).

Next, we show the equivalent static budget constraint of the A-feasible dynamic constraint.

Theorem 4.3.1. *A consumption and bequest plan $(c, M, W_T) \in G_+^*$ is A-feasible if and only if*

$$E^{Q_v}\left[\beta_{v,T}e^{-\int_0^T \lambda_{x+t}dt}W_T + \int_0^T \lambda_{x+t}\beta_{v,t}e^{-\int_0^t \lambda_{x+s}ds}M_t dt + \int_0^T \beta_{v,t}e^{-\int_0^t \lambda_{x+s}ds}(c_t - Y_t)dt\right]$$
$$\leq w_0 + E^{Q_v}\left[\int_0^T \beta_{v,t}e^{-\int_0^t \lambda_{x+s}ds}\delta(v_t)dt\right] \text{ for } \forall v \in \mathcal{N}^*. \tag{4.14}$$

A direct corollary is when the free disposal will disappear.

Corollary 4.3.1. *If there exists a process $v^* \in \mathcal{N}$ such that*

$$\begin{aligned}
&E^{Q_v}\left[\beta_{v,T}e^{-\int_0^T \lambda_{x+t}dt}W_T + \int_0^T \lambda_{x+t}\beta_{v,t}e^{-\int_0^t \lambda_{x+s}ds}M_t dt\right.\\
&\left.+\int_0^T \beta_{v,t}e^{-\int_0^t \lambda_{x+s}ds}(c_t - Y_t - \delta(v_t))dt\right]\\
&\leq E^{Q_{v^*}}\left[\beta_{v^*,T}e^{-\int_0^T \lambda_{x+t}dt}W_T + \int_0^T \lambda_{x+t}\beta_{v^*,t}e^{-\int_0^t \lambda_{x+s}ds}M_t dt\right.\\
&\left.+\int_0^T \beta_{v^*,t}e^{-\int_0^t \lambda_{x+s}ds}(c_t - Y_t - \delta(v_t^*))dt\right]\\
&= w_0
\end{aligned}$$

then (c, M, W_T) is feasible, the optimal wealth is given by

$$W_{v^*,t} = E^{Q_{v^*}}\left[\int_t^T e^{-\int_t^s r_u + v_{0,u}^* + \lambda_{x+u}du}[c_s - Y_s + \lambda_{x+s}M_s - \delta(v_s^*)]ds\right.$$
$$\left.+e^{-\int_t^T r_s + v_{0,s}^* + \lambda_{x+s}ds}W_T|\mathcal{F}_t\right], \tag{4.15}$$

and the optimal free disposal $C_t^* \equiv 0$.

4.4 Primal problem and dual problem

From Theorem 4.3.1, we can formulate the primal problem with the dynamic budget constraint (4.5) to a problem with static budget constraint (4.14).

$$\sup_{(c,M,W_T)\in G_+^*} J(c,M,W_T)$$

$$\text{s.t. } E^{Q_v}\left[\beta_{v,T}e^{-\int_0^T \lambda_{x+s}ds}W_T + \int_0^T \lambda_{x+t}\beta_{v,t}e^{-\int_0^t \lambda_{x+s}ds}M_t dt \right. \tag{P}$$

$$\left. + \int_0^T \beta_{v,t}e^{-\int_0^t \lambda_{x+s}ds}(c_t - Y_t)dt\right] \le w_0 + E^{Q_v}\left[\int_0^T \beta_{v,t}e^{-\int_0^t \lambda_{x+s}ds}\delta(v_t)dt\right],$$

for $\forall v \in \mathcal{N}^*$, where

$$\begin{aligned} J(c,M,W_T) &= E\left[\int_0^T e^{-\int_0^t \lambda_{x+s}ds}U_1(c_t,t)dt + \int_0^T \lambda_{x+t}e^{-\int_0^t \lambda_{x+s}ds}U_2(M_t,t)dt \right. \\ &\quad \left. + e^{-\int_0^T \lambda_{x+t}dt}U_3(W_T,T)\right]. \end{aligned}$$

Since $0 \in \mathcal{N}^*$, problem (P) can be considered as a convex optimization problem on a closed, norm bounded subset of $L^1(\bar{\lambda} \times Q_0)$, where $\bar{\lambda}$ is the Lebesgue measure on $[0,T]$. However, L^1 spaces are not reflexive so lack compactness. The existing literature circumvents this difficulty using the Lagrangian dual control method. Because the set $\{\pi_v : v \in \mathcal{N}^*\}$ is convex, this suggests the existence of pricing kernel π_{v^*}, a Lagrangian multiplier $\psi^* > 0$ such that $(c^*, M^*, W_T^*, \psi^*, v^*)$ is a saddle point of the Lagrangian

$$\begin{aligned} &\mathcal{L}(c,M,W_T,\psi,v) = \\ &E\left[\int_0^T e^{-\int_0^t \lambda_{x+s}ds}U_1(c_t,t)dt + \int_0^T \lambda_{x+t}e^{-\int_0^t \lambda_{x+s}ds}U_2(M_t,t)dt + e^{-\int_0^T \lambda_{x+t}dt}U_3(W_T,T)\right] \\ &+\psi\left\{w_0 - E\left[\int_0^T \pi_{v,t}e^{-\int_0^t \lambda_{x+s}ds}[c_t + \lambda_t M_t - Y_t - \delta(v_t)]dt + \pi_{v,T}e^{-\int_0^T \lambda_{x+t}dt}W_T\right]\right\}. \end{aligned}$$

Maximizing (c, M, W_T) and minimizing (ψ, v), we derive the dual problem

$$\inf_{(\psi,v)\in(0,\infty)\times\mathcal{N}^*} \widetilde{J}(\psi,v)$$

$$= \inf_{(\psi,v)\in(0,\infty)\times\mathcal{N}^*} E\left[\int_0^T e^{-\int_0^t \lambda_{x+s}ds}\widetilde{U}_1(\psi\pi_{v,t},t)dt + \int_0^T \lambda_{x+t}e^{-\int_0^t \lambda_{x+s}ds}\widetilde{U}_2(\psi\pi_{v,t},t)dt \right. \tag{D}$$

$$\left. + e^{-\int_0^T \lambda_{x+t}dt}\widetilde{U}_3(\psi\pi_{v,T},T) + \psi\left\{w_0 + \int_0^T e^{-\int_0^t \lambda_{x+s}ds}\pi_{v,t}[Y_t + \delta(v_t)]dt\right\}\right],$$

where dual utilities are given by

$$\widetilde{U}_1(z,t) = \sup_{c>0} \{U_1(c,t) - zc\},$$
$$\widetilde{U}_2(z,t) = \sup_{M>0} \{U_2(M,t) - zM\},$$
$$\widetilde{U}_3(z,T) = \sup_{W>0} \{U_3(W,T) - zW\},$$

for $z > 0$ and each $U_i, i = 1,2,3$, satisfies the Inada condition

$$U_i'(0+,t) = \infty, U_i'(\infty,t) = 0+, \text{ for } \forall t \in [0,T], \tag{4.16}$$

in which U_i' is the first order derivative with respect to the first variable.

For $\widetilde{U}_1(z,t),\ z > 0$, by the concavity of U_1, we have a c^* such that

$$\widetilde{U}_1(z,t) = U_1(c^*,t) - zc^*, \tag{4.17}$$

where $U_1'(c^*,t) - z = 0$, i.e. $c^* = U_1'^{-1}(z,t)$, and $U_1'^{-1}(z,t)$ is the inverse of $U'(c,t)$ with respect to the first variable. Next, take the first order derivative with z on both sides of (4.17) and by $U_1'(c^*,t) - z = 0$, we have

$$\frac{\partial \widetilde{U}_1(z,t)}{\partial z} = U_1'(c^*,t)\frac{\partial c^*}{\partial z} - c^* - z\frac{\partial c^*}{\partial z} = -c^*,$$

i.e.

$$c^* = U_1'^{-1}(z,t) = -\frac{\partial \widetilde{U}_1(z,t)}{\partial z}.$$

Define the function $f_i(z,t) = U_i'^{-1}(z,t) = -\frac{\partial}{\partial z}\widetilde{U}_i(z,t), i = 1,2,3$, similarly to the argument above, we have

$$c^* = f_1(z,t), M^* = f_2(z,t), W^* = f_3(z,T). \tag{4.18}$$

Then, by Definition 4.2.1, we can derive the following properties for dual utility.

Lemma 4.4.1. *The dual utilities $\widetilde{U}_i(\cdot,t) : (0,\infty) \to \mathbb{R}, i = 1,2,3$ are strictly decreasing and strictly convex with respect to the first variable. They have the explicit representations*

$$\widetilde{U}_i(z,t) = U_i(f_i(z,t),t) - zf_i(z,t), \text{ where } i = 1,2,3. \tag{4.19}$$

and derivatives $\frac{\partial}{\partial z}\widetilde{U}_i(z,t) = -f_i(z,t) = -U_i'^{-1}(z,t)$. Furthermore,

$$\widetilde{U}_i(0+,t) = U_i(\infty,t), \quad \widetilde{U}_i(\infty,t) = U_i(0+,t).$$

Finally, we can prove the following relationship between Problem (P) and Problem (D).

Theorem 4.4.1. *Assume that $U_i, i = 1, 2, 3$, satisfy the Inada conditions and the following constraint holds*

$$\beta U_i'(x,t) \geq U_i'(\gamma x, t), \quad \forall (x,t) \in (0,\infty) \times [0,T], \tag{4.20}$$

for some constants $\beta \in (0,1)$ and $\gamma \in (0,\infty)$. If there exists a solution (ψ^*, v^*) to the dual problem (D) and

$$E\left[\int_0^T \pi_{v^*,t} e^{-\int_0^t \lambda_{x+s} ds}(f_1(\psi^* \pi_{v^*,t}) + \lambda_{x+t} f_2(\psi^* \pi_{v^*,t}) - Y_t - \delta(v_t^*))dt \right.$$
$$\left. + \pi_{v^*,T} e^{-\int_0^T \lambda_{x+t} dt} f_3(\psi^* \pi_{v^*,T})\right] < \infty, \tag{4.21}$$

then there exists an A-feasible optimal $(c^*, M^*, W_T^*) \in \mathcal{B}(\mathcal{P}, A)$ such that

$$\frac{\partial U_1}{\partial c}(c_t^*, t) = \frac{\partial U_2}{\partial M}(M_t^*, t) = \psi^* \pi_{v^*,t}, \ \frac{\partial U_3}{\partial W}(W_T^*, T) = \psi^* \pi_{v^*,T}, \tag{4.22}$$

for $\forall t \in [0,T]$ and some $\psi^* > 0$. Moreover, the optimal solution (c^*, M^*, W_T^*) satisfies the budget constraint

$$E\left[\int_0^T \pi_{v^*,t} e^{-\int_0^t \lambda_{x+s} ds}(f_1(\psi^* \pi_{v^*,t}) + \lambda_{x+t} f_2(\psi^* \pi_{v^*,t}) - Y_t - \delta(v_t^*))dt \right.$$
$$\left. + \pi_{v^*,T} e^{-\int_0^T \lambda_{x+t} dt} f_3(\psi^* \pi_{v^*,T})\right] = w_0. \tag{4.23}$$

Conversely, if (4.22) and (4.23) hold for some $(\psi^*, v^*) \in (0,\infty) \times \mathcal{N}^*$ and some A-feasible $(c^*, M^*, W_T^*) \in \mathcal{B}(\mathcal{P}, A)$, then (ψ^*, v^*) solves the dual problem.

Furthermore, under each artificial market $\mathcal{M}_v$, we can derive the following corollary for the dual problem (D).

Corollary 4.4.1. *For an arbitrary $v \in \mathcal{N}^*$, there exists a unique optimal ψ_v minimizing $\widetilde{J}(\psi, v)$ such that*

$$\frac{\partial \widetilde{J}(\psi_v, v)}{\partial \psi} = 0.$$

In addition, the optimal wealth under (ψ_v, v) is given by

$$W_{v,t} = E^{Q_v}\left[\int_t^T e^{-\int_t^s r_u + v_{0,u} + \lambda_{x+u} du}[f_1(\psi_v \pi_{v,s}, s) - Y_s + \lambda_{x+s} f_2(\psi_v \pi_{v,s}, s) - \delta(v_s)]ds \right.$$
$$\left. + e^{-\int_t^T r_s + v_{0,s} + \lambda_{x+s} ds} f_3(\psi_v \pi_{v,T}, T)|\mathcal{F}_t\right], \tag{4.24}$$

and the optimal free disposal $C_{v,t}^* \equiv 0$ under (ψ_v, v).

4.5 The existence of primal problem

For the dual problem (D), the difficulty in applying dual control method is that $\widetilde{J}(\psi, v)$ is not convex with respect to v_t unless $Y_t \equiv 0$, $\delta(v_t) \equiv 0$, and the Arrow-Pratt coefficient of risk-aversion is strictly less than 1. If these rather restrictive assumptions are satisfied, the problem can be relaxed by looking for a solution in $(0, \infty) \times \mathcal{N}$ (i.e., by allowing the density process to be a local martingale instead of a martingale), and the existence of a solution to Problem (D) can then be shown using the technique of Cvitanić and Karatzas (1992).

Fortunately, Levin (1976) proves the existence of solution under non-reflexive spaces, which can be applied to deal with the lack of compactness in the set of feasible plan $(c, M, W_T) \in G_+^*$. Next, we prove the existence of the primal problem.

Theorem 4.5.1. *Suppose that*

1. *There exists a* $(c, M, W_T) \in \mathcal{B}(\mathcal{P}, A)$ *with* $J(c, M, W_T) > -\infty$.
2. *Either* $U_i, i = 1, 2, 3$, *are bounded above on* $(0, \infty) \times [0, T]$, *or there exist constants* $k_i \geq 0$, $b_i \in (0, 1)$, *and* $p_i > 1$ *such that*

$$U_i(x, t) \leq k_i(1 + x^{1-b_i}), \ \forall (x, t) \in (0, \infty) \times [0, T], \tag{4.25}$$

and

$$\xi_0^{-1} \in L^{\max(p_1/b_1,\ p_2/b_2,\ p_3/b_3)}(\bar{\lambda} \times Q_0). \tag{4.26}$$

Then the solution to the primal problem (P) *exists.*

4.6 Numerical Analysis

Following the parameter settings in Huang et al. (2008), we assume that an individual is 45 years old at the initial time, retires at the age of 65, and the family stops making investment decisions at the individual's age of 95, so $T_R = 20$ and $T = 50$. The individual's force of mortality follows the Gompertz law

$$\lambda_{x+t} = \frac{1}{9.5} e^{\frac{x+t-86.3}{9.5}}, \ x = 45.$$

Before the first time of the family decision horizon T and death time T_x, the individual is allowed to invest in a bond and a stock

$$\begin{aligned} B_t &= \exp\left(\int_0^t r(u)du\right), \\ S_t + D_t &= S_0 + \int_0^t \mu(u)S_u du + \int_0^t \sigma(u)S_u dZ_u, \end{aligned}$$

where $r(t), \mu(t), \sigma(t)$ are continuous functions of t , $\sigma(t) > 0$ for $t \in [0, T]$, and Z_t is a one-dimensional Brownian motion. Moreover, the individual's income process has no idiosyncratic risk (only has Brownian motion from the financial market)

$$\begin{cases} Y_t &= Y_0 + \mu_Y \int_0^t Y_u du + \sigma_Y \int_0^t Y_u dZ_u, \quad 0 \le t < \min(T_x, T_R), \\ Y_t &= 0, \quad \min(T_x, T_R) \le t \le T, \end{cases} \tag{4.27}$$

where μ_Y and σ_Y are two constants. We consider the portfolio-mix constraint (Part 4.2.5 (e)) with $D = [0, 1]$, then the portfolio constraint set A and its effective domain $\widetilde{A}$ are given by

$$\begin{aligned} A &= \left\{(\alpha, \theta) \in \mathbb{R}^2 : \alpha + \theta \ge 0, \theta \in [0, \alpha + \theta]\right\} && (4.28) \\ &= \left\{(\alpha, \theta) \in \mathbb{R}^2 : \alpha \ge 0, \theta \ge 0\right\}, \\ \widetilde{A} &= \left\{(v_0, v_-) : (\alpha, \theta)(v_0, v_-)^\top \ge 0, \forall (\alpha, \theta) \in A\right\} \\ &= \left\{(v_0, v_-) : v_0 \ge 0, v_- \ge 0\right\}. && (4.29) \end{aligned}$$

As a result, the individual's wealth process (4.4) has the following equivalent form

$$\begin{aligned} W_t &= W_0 + \int_0^t [(r(s) + \lambda_{x+s})W_s + (\mu(s) - r(s))\theta_s]ds + \int_0^t \sigma(s)\theta_s dZ_s \\ &\quad - \int_0^t (c_s + \lambda_{x+s}M_s - Y_s)ds - C_t, \end{aligned} \tag{4.30}$$

where $0 \le t \le \min(T_x, T)$ and $M_t = W_t + \frac{I_t}{\lambda_{x+t}}$.

Inspired by Huang et al. (2008), we set the base model parameters as

$$\begin{aligned} &\widetilde{\delta} = 0.02, \quad \mu_Y = 0.01, \quad \sigma_Y = 0.05, \\ &W_0 = 200.00, \quad Y_0 = 50.00, \quad \gamma_1 = \gamma_2 = \gamma_3 = \gamma = 1.50, \end{aligned} \tag{4.31}$$

and restrict utility into power utility

$$\begin{cases} U_1(c_t,t) = e^{-\widetilde{\delta}t}\frac{c_t^{1-\gamma}}{1-\gamma}, \\ U_2(M_t,t) = e^{-\widetilde{\delta}t}V_B(t,M_t), \\ U_3(W_T,T) = e^{-\widetilde{\delta}T}\frac{W_T^{1-\gamma}}{1-\gamma}, \end{cases}$$

where $V_B(t, M_t)$ is the value function of family investment after the individual dies and the subscript "B" is short for bequest. The same setting for bequest utility can be found in Zeng et al. (2016) and Boyle et al. (2022).

We assume there is no trading constraint after the individual dies, so we can make fair comparisons between the cases with and without constraint when the individual is alive. Thus, the wealth process after individual dies at time $t \in [0, T]$ is

$$\begin{aligned} dW_s &= [r(s)W_s + (\mu(s) - r(s))\theta_s]ds + \sigma(s)\theta_s dZ_s - c_s ds, s \in [t, T], \\ W_t &= M_t. \end{aligned} \tag{4.32}$$

Furthermore, the value function of family investment after individual dies follows

$$V_B(t, W_t) = \sup_{\theta,c} E_t\left[\int_t^T e^{-\widetilde{\delta}(s-t)}\frac{c_t^{1-\gamma}}{1-\gamma}ds + e^{-\widetilde{\delta}(T-t)}\frac{W_T^{1-\gamma}}{1-\gamma}\right], \tag{4.33}$$

where $E_t[\cdot]$ means the conditional expectation on the filtration $\mathcal{F}_t$. Then, under the dynamic programming principle, we can derive the following lemma

Lemma 4.6.1. *The explicit solution of $V_B(t, M_t)$ is given by*

$$V_B(t, W_t) = \frac{1}{1-\gamma}W_t^{1-\gamma}g(t)^\gamma, \tag{4.34}$$

where

$$\begin{aligned} g(t) &= \int_t^T e^{-\frac{\widetilde{\delta}}{\gamma}(s-t)}F_B(s-t, s)ds + e^{-\frac{\widetilde{\delta}}{\gamma}(T-t)}F_B(T-t, T), \\ F_B(\tau, s) &= e^{-\int_0^\tau \frac{\gamma-1}{\gamma}r(s-u)du - \frac{1}{2}\frac{\gamma-1}{\gamma^2}\int_0^\tau \kappa_{0,s-u}^2 du}. \end{aligned} \tag{4.35}$$

Next, we compute the following methods to make comparisons.

- **Method 1: SAMS approach**
  Benchmark from Bick et al. (2013), assume v_t is affine in t, minimize the upper bound, and then compute the lower bound under v_t^*, where v_t^* is the optimal v_t minimizing the upper bound.

- **Method 2: Dual control neural network approach**
  Restrict $v_t = v(t)$ as a neural network of time t, minimize the upper bound, and then compute the lower bound under v_t^*, where v_t^* is the optimal v_t minimizing the upper bound.

Denote $(\alpha_v, \theta_v, c_v, I_v)$ as the general strategy and $((\alpha_v)^*, (\theta_v)^*, (c_v)^*, (I_v)^*)$ as the optimal strategy under the artificial market $\mathcal{M}_v$, then we derive the lower and upper bounds in each method.

- **Explicit upper bound for Method 1 and Method 2**

When $v_t = v(t)$, i.e., v_t is a function of t, we can derive the explicit solution of the upper bound for primal problem (P).

Proposition 4.6.1. *Suppose that $v_t = v(t)$ and $t \in [T_R, T]$, then the upper bound of the primal problem (P) is given by*

$$V_R(t, W_{v,t}; v) = \frac{1}{1-\gamma}\widetilde{F}_1(t, W_{v,t})^{1-\gamma}\widetilde{F}_2(t)^{\gamma}, \tag{4.36}$$

where

$$\begin{aligned}
&\widetilde{F}_1(t, W_{v,t}) = W_{v,t} + \int_t^T e^{-\int_t^s \lambda_{x+u}du}\delta(v_s)F_2(s-t, s)ds,\\
&\widetilde{F}_2(t) = \int_t^T e^{-\int_t^s \lambda_{x+u}du - \frac{\tilde{\delta}}{\gamma}(s-t)}(1+\lambda_{x+s}g(s))F_3(s-t, s)ds\\
&+e^{-\int_t^T \lambda_{x+u}du - \frac{\tilde{\delta}}{\gamma}(T-t)}F_3(T-t, T),\\
&F_2(\tau, s) = e^{-\int_0^\tau r(s-u)+v_0(s-u)du},\\
&F_3(\tau, s) = e^{-\int_0^\tau \frac{\gamma-1}{\gamma}(r(s-u)+v_0(s-u))du - \frac{1}{2}\frac{\gamma-1}{\gamma^2}\int_0^\tau \kappa_{v,s-u}^2du},
\end{aligned}$$

and $g(s)$ follows (4.35). Moreover, the optimal strategies are

$$(\theta_{v,t})^* = \min\left\{\max\left\{-\frac{1}{\gamma\sigma(t)}\widetilde{F}_1(t, W_{v,t})\kappa_{v,t}, 0\right\}, W_{v,t}\right\}, \tag{4.37}$$

$$(c_{v,t})^* = \widetilde{F}_1(t, W_{v,t})/\widetilde{F}_2(t), \quad (M_{v,t})^* = [\widetilde{F}_1(t, W_{v,t})g(t)]/\widetilde{F}_2(t). \tag{4.38}$$

Proposition 4.6.2. *Suppose that $v_t = v(t)$ and $t \in [0, T_R]$, then the upper bound of the primal problem (P) is given by*

$$\widetilde{J}(t, W_{v,t}, Y_t; v) = \frac{1}{1-\gamma}\widetilde{F}_3(t, W_{v,t}, Y_t)^{1-\gamma}\widetilde{F}_2(t)^{\gamma}, \tag{4.39}$$

where

$$\widetilde{F}_3(t, W_{v,t}, Y_t) = W_{v,t} + Y_t \int_t^{T_R} e^{-\int_t^s \lambda_{x+u} du} F_1(s-t, s) ds$$
$$+ \int_t^T e^{-\int_t^s \lambda_{x+u} du} \delta(v(s)) F_2(s-t, s) ds,$$
$$\widetilde{F}_2(t) = \int_t^T e^{-\int_t^s \lambda_{x+u} du - \frac{\widetilde{\delta}}{\gamma}(s-t)} (1 + \lambda_{x+s} g(s)) F_3(s-t, s) ds$$
$$+ e^{-\int_t^T \lambda_{x+u} du - \frac{\widetilde{\delta}}{\gamma}(T-t)} F_3(T-t, T),$$
$$F_1(\tau, s) = e^{\mu_Y \tau + \int_0^\tau -[r(s-u)+v_0(s-u)] + \kappa_{v,s-u} \sigma_Y du},$$
$$F_2(\tau, s) = e^{-\int_0^\tau r(s-u)+v_0(s-u) du},$$
$$F_3(\tau, s) = e^{-\int_0^\tau \frac{\gamma-1}{\gamma}(r(s-u)+v_0(s-u)) du - \frac{1}{2}\frac{\gamma-1}{\gamma^2}\int_0^\tau \kappa_{v,s-u}^2 du},$$

and $g(s)$ follows (4.35). Moreover, the optimal strategies are

$$(\theta_{v,t})^* = \min\left\{\max\left\{-\frac{1}{\gamma\sigma(t)}\widetilde{F}_3(t, W_{v,t}, Y_t)\kappa_{v,t} - \frac{\sigma_Y}{\sigma(t)} Y_t \int_t^{T_R} e^{-\int_t^s \lambda_{x+u} du} F_1(s-t, s) ds, 0\right\}, W_{v,t}\right\}, \quad (4.40)$$

$$(c_{v,t})^* = \widetilde{F}_3(t, W_{v,t}, Y_t)/\widetilde{F}_2(t), \quad (M_{v,t})^* = [\widetilde{F}_3(t, W_{v,t}, Y_t) g(t)]/\widetilde{F}_2(t) \quad (4.41)$$

For Method 1, follow Bick et al. (2013), we separate v_t at the retirement time T_R, i.e.

$$v_t = v(t) = \begin{cases} v^w(t) = (v_0^w(t), v_-^w(t)) = ((a_1 + a_2 t)_+, (a_3 + a_4 t)_+), & 0 \le t < T_R, \\ v^R(t) = (v_0^R(t), v_-^R(t)) = ((a_5 + a_6 t)_+, (a_7 + a_8 t)_+), & T_R \le t \le T, \end{cases} \quad (4.42)$$

where superscript w is short for "working", superscript R is short for "retirement", and $(\cdot)_+$ is the positive part of a function.

For Method 2, we use one neural network (v_0, v_-) with state variable time t to describe v_t. We let the neural network learn the retirement time T_R by itself and therefore do not separate v_t at T_R.

$$v_t = v(t) = (v_0(t), v_-(t)), \quad 0 \le t \le T, \quad (4.43)$$

After minimizing the upper bound $\widetilde{J}(0, W_{v,0}, Y_0; v)$, we obtain the optimal v_t^*. Then, we can define the candidate value function $\overline{J}(t, \overline{W}_{v^*,t}, Y_t; v^*)$ as

$$\overline{J}(t, \overline{W}_{v^*,t}, Y_t; v^*) = E_t \left[\int_t^T e^{-\int_t^s \lambda_{x+u} du - \widetilde{\delta}(s-t)} \frac{((c_{v^*,s})^*)^{1-\gamma}}{1-\gamma} ds \right.$$
$$\left. + \int_t^T \lambda_{x+s} e^{-\int_t^s \lambda_{x+u} du - \widetilde{\delta}(s-t)} \frac{((M_{v^*,s})^*)^{1-\gamma}}{1-\gamma} g(s)^\gamma ds + e^{-\int_t^T \lambda_{x+s} ds - \widetilde{\delta}(T-t)} \frac{((\overline{W}_{v^*,T})^*)^{1-\gamma}}{1-\gamma} \right],$$

where the candidate wealth process $\overline{W}_{v^*,t}$ is driven by the optimal strategies (4.37), (4.38), (4.40), and (4.41)

$$\begin{aligned} d\overline{W}_{v^*,t} &= \{[r(t) + \lambda_{x+t}]\overline{W}_{v^*,t} + (\theta_{v^*,t})^*[\mu(t) - r(t)]\}dt + (\theta_{v^*,t})^*\sigma(t)dZ_t \\ &\quad -[(c_{v^*,t})^* + \lambda_{x+t}(M_{v^*,t})^* - Y_t]dt, \\ \overline{W}_{v^*,0} &= \overline{w}_0. \end{aligned} \tag{4.44}$$

The candidate value function $\overline{J}(t, \overline{W}_{v^*,t}, Y_t; v^*)$ provides a lower bound for the primal Problem (P) because $\theta_{v^*,t}$ satisfies the portfolio constraint set (4.28) and $C_t \equiv 0$ is a sub-strategy for free disposal in (4.30). From all things above, we obtain the tight lower and upper bounds for the primal Problem (P)

$$\overline{J}(0, \overline{W}_{v^*,0}, Y_0; v^*) \leq J(c, M, W_T) \leq \widetilde{J}(0, W_{v^*,0}, Y_0; v^*).$$

Remark 4.6.1. *To avoid the arbitrage opportunity for doubling strategy, we need Y_t to satisfy Assumption 4.3.1 to ensure (4.6). By Ito's formula, we derive*

$$d(\pi_{v,t}Y_t) = \pi_{v,t}Y_t[-(r(t) + v_{0,t}) + \mu_Y + \sigma_Y \kappa_{v,t}]dt + \pi_{v,t}Y_t(\kappa_{v,t} + \sigma_Y)dZ_t. \tag{4.45}$$

Furthermore, we assume that

$$\sigma_Y \leq \sigma(t), \tag{4.46}$$

$$\frac{\mu_Y}{\sigma_Y} \leq \frac{\mu(t)}{\sigma(t)}. \tag{4.47}$$

Together with (4.29), we have the drift term of (4.45)

$$\begin{aligned} & -(r(t) + v_{0,t}) + \mu_Y + \sigma_Y \kappa_{v,t} \\ = & -(r(t) + v_{0,t}) + \mu_Y - \frac{\sigma_Y}{\sigma}(\mu + v_{-,t} - (r + v_{0,t})) \\ = & \left(\frac{\sigma_Y}{\sigma} - 1\right)(r + v_{0,t}) + \mu_Y - \frac{\sigma_Y}{\sigma}(\mu + v_{-,t}) \\ \leq & \ 0. \end{aligned}$$

Thus, $\pi_{v,t}Y_t$ is a non-negative local super-martingale, which is also a super-martingale by Fatou's lemma. Therefore,

$$E[\pi_{v,t}Y_t] \leq Y_0, \tag{4.48}$$

for arbitrary $v \in \mathcal{N}^*$ and $t \in [0,T]$. Finally, Assumption 4.3.1 is a direct result from (4.48). In the numerical examples, we set all the parameters to follow the constraints (4.46) and (4.47).

Furthermore, we also need to check the conditions in Theorem 4.5.1 to guarantee the primal problem's existence. For the power utility with risk aversion coefficient $\gamma > 1$, we have the utility bounded above by 0. Thus, the second condition in Theorem 4.5.1 is satisfied automatically. For the first condition, under $\gamma > 1$, we only need to find a pair of positive A-feasible (c, M, W_T) to avoid $J(c, M, W_T)$ going to negative infinity. Let $\theta_t \equiv 0 \in A$, $r(t) = r > 0$, and $C_t \equiv 0$, we can rewrite the wealth process (4.30) as

$$\frac{dW_t}{W_t} = \left[r + \lambda_{x+t} + \frac{Y_t}{W_t} - \frac{c_t}{W_t} - \lambda_{x+t}\frac{M_t}{W_t}\right] dt, W_0 > 0$$

By choosing

$$c_t = M_t = \frac{1}{2(1+\lambda_{x+t})}\{[r + \lambda_{x+t}]W_t + Y_t\} > 0,$$

we obtain

$$\frac{dW_t}{W_t} = 0.5\left[r + \lambda_{x+t} + \frac{Y_t}{W_t}\right] dt > 0, W_0 > 0.$$

Therefore, we find a positive A-feasible strategy (c, M, W_T) (this strategy is A-feasible because $\theta_t \equiv 0 \in A$) such that $J(c, M, W_T) > -\infty$. Finally, by Theorem 4.5.1, the primal problem's existence is guaranteed.

Example 4.6.1. In this example, we study the case when the risk-free interest rate, stock appreciation rate, and volatility are all constant, i.e., $\mu(t) = 0.07$, $r(t) = 0.02$, and $\sigma(t) = 0.2$.

Table 4.1 shows the lower and upper bounds for each method. We use the default "interior-point" algorithm provided in the Matlab package "fmincon" to minimize the upper bounds in each method.

Method 1 and 2 share a similar explicit upper bound. We use the Trapezoidal rule to compute the double integral in this explicit upper bound, and the number of the time interval is set as 100. Moreover, we apply the quasi-Monte Carlo method to compute the lower bound. The Sobol sequence with the first 4,000 numbers skipped is used to generate the normal random variables. To make fair comparisons, we set all the lower bounds

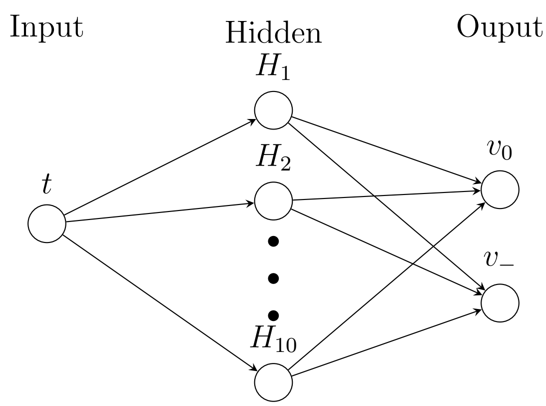


Figure 4.1: Neural network with structure "1-10-2".

with the same path number, 20,000, and the same time interval of 1,000. In addition, we add liquidity constraint that when $\overline{W}_{v^*,t} = 0$, $(c_{v^*,t})^*$ is truncated by $\frac{Y_t}{1+\lambda_{x+t}g(t)}$, then $-[(c_{v^*,t})^* + \lambda_{x+t}(M_{v^*,t})^* - Y_t] = -[1+\lambda_{x+t}g(t)](c_{v^*,t})^* + Y_t \geq 0$ in the wealth process (4.44). In other words, when the wealth equals zero, the consumption and death benefit should not be bigger than the income Y_t.

For Method 1, we randomly choose the initial values for the parameters in (4.42). We sample the initial values for 30 groups, and in each group, we train the affine structure 50 times. Finally, we choose the lowest upper bound among the 30 groups.

For Method 2, we set the structure of neural network v_t as "1-10-2", which means one node (time t) in the input layer, ten nodes in one hidden layer, and two nodes (v_0 and v_-) in the output layer. More specifically, we show the structure of neural network in Figure 4.1. The value of a hidden node is $H_i = f_a(w_i t + b_i)$, $i = 1, 2, ..., 10$, where the $f_a(\cdot)$ is the activation function, w_i is the weight parameter for edge connecting to H_i, and b_i is the bias at the node H_i. In this example, we choose the rectified linear unit (ReLU) function as the activation function, i.e., $f_a(x) = max(0, x)$. The values of the two output nodes are $v_0 = (\sum_{i=1}^{10} w_{i+10}H_i + b_{11})^+$ and $v_- = (\sum_{i=1}^{10} w_{i+20}H_i + b_{12})^+$, where w_{i+10} is the weight parameter for the edge connecting to node v_0, b_{11} is the bias for the node v_0, w_{i+20} is the weight parameter for the edge connecting to node v_-, and b_{12} is the bias for the node v_-. There are 30 edges and 12 biases, and hence 42 parameters wait to be optimized. Similarly to Method 1, we randomly choose the initial values for the weights and bias of neural network (4.43) from a normal distribution with mean 0 and standard deviation 10^{-4}. We sample the initial values for 30 groups, and in each group, we train the neural network 50 times. Finally, we choose the lowest upper bound among the 30 groups.

In Table 4.1, we design three quantities to compare the two methods. The first is the "duality gap". It is defined as the absolute difference between the lower and upper bounds. The second is the "relative gap". It is defined as the absolute ratio of the "duality gap" over the lower bound. The third is "welfare loss". Following Bick et al. (2013), we define the "welfare loss" as the upper bound of the fraction of wealth that an individual would like to through away to get access to an optimal strategy. More specifically, under the market $\mathcal{M}_{v^*}$, it is the proportion L such that the following equation holds for the lower and upper bounds of the value function.

$$\overline{J}(0, \overline{W}_{v^*,0}, Y_0; v^*) = \widetilde{J}(0, W_{v^*,0}[1-L], Y_0[1-L]; v^*).$$

From Proposition 4.6.2 and $\delta(v) = 0$ under portfolio-mix constraint, we have

$$\widetilde{J}(0, W_{v^*,0}[1-L], Y_0[1-L]; v^*) = (1-L)^{1-\gamma}\widetilde{J}(0, W_{v^*,0}, Y_0; v^*).$$

Therefore, the upper bound of welfare loss is

$$L = 1 - \left(\frac{\overline{J}(0, \overline{W}_{v^*,0}, Y_0; v^*)}{\widetilde{J}(0, W_{v^*,0}, Y_0; v^*)}\right)^{\frac{1}{1-\gamma}}. \tag{4.49}$$

From Table 4.1, we see Method 2 slightly beats Method 1 in every aspect: smaller upper bound, bigger lower bound, smaller duality gap, smaller relative gap, and smaller welfare loss. The relative gaps of these two methods are very low, only around 0.2%. Moreover, the welfare losses for both methods are also low at a level of 0.5%.

Figure 4.2 shows the change of the upper bound in each training iteration. We find that the upper bound of Method 2 decreases faster but finally stays at the level close to Method 1. Figure 4.3 reveals that the neural network (4.43) of Method 2 learns a similar result as Method 1. It turns out there is no big difference between the affine structure and the neural network when $\mu(t), \sigma(t), r(t)$ are all constant. Therefore, the results of the two methods in Table 4.1 are quite similar. Figure 4.4 illustrates that when considering the trading constraint, the individual reduces their demand for life insurance. Moreover, the individual's demand for life insurance performs a "spoon shape". Specifically, the expected optimal face value is positive initially because the individual has a large future income to protect. Then, the optimal face value decreases with time t and becomes negative a little earlier than the retirement time $T_R = 20$. This is because the increasing force of mortality makes life insurance less attractive than stocks and bonds (the face value of life insurance is I_t/λ_{x+t}). Finally, the optimal face value increases to 0 at the terminal time.

Table 4.1: **Lower and upper bounds for Example 4.6.1**

	Method 1	Method 2
Structure	Affine	(1-10-2)
Activation function	None	ReLU
Upper bound	-8.4850600	-8.4853506
Lower bound	-8.5064352	-8.5061158
Duality gap	0.0213752	0.0207652
Relative gap	0.2513%	0.2441%
Welfare loss	0.5019%	0.4876%
Time elapsed	7.43 hours	8.31 hours

For the upper bounds of Method 1 and Method 2, the number of time intervals is 100 for the numerical double integral. For the quasi-Monte Carlo simulation of the lower bound in each method, the number of paths is 20,000, and the number of time intervals is 1,000. The structure "(1-10-2)" means that the neural network is chosen as one node (time t) in the input layer, ten nodes in one hidden layer, and two nodes (v_0 and v_-) in the output layer. The "Duality gap" is defined as the absolute difference between the lower and upper bounds. The "Relative gap" is defined as the absolute ratio of the "Duality gap" over the "Lower bound". The "Welfare loss" is defined by (4.49).

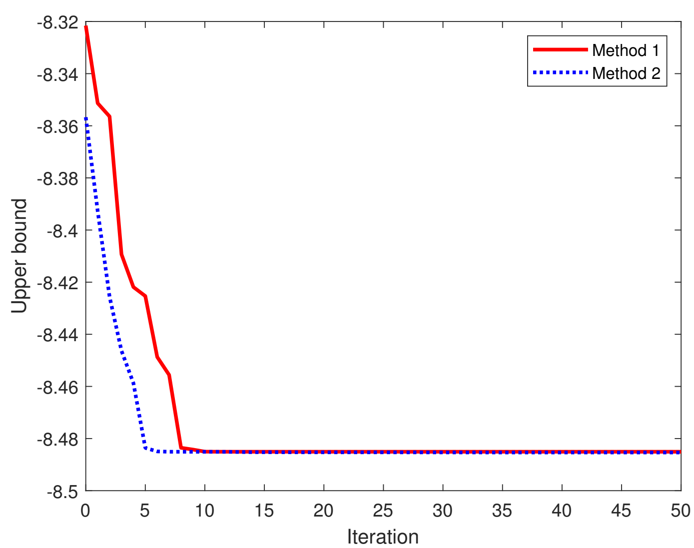


Figure 4.2: Change of upper bound in each training iteration for Example 4.6.1

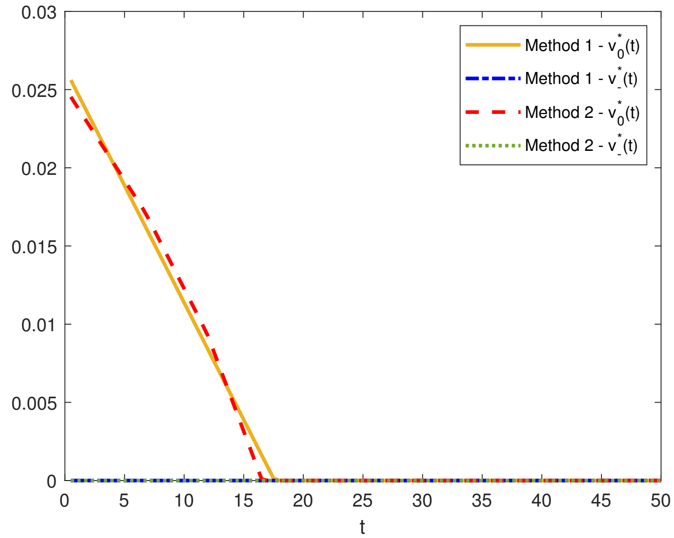


Figure 4.3: Optimal v^* for each method in Example 4.6.1

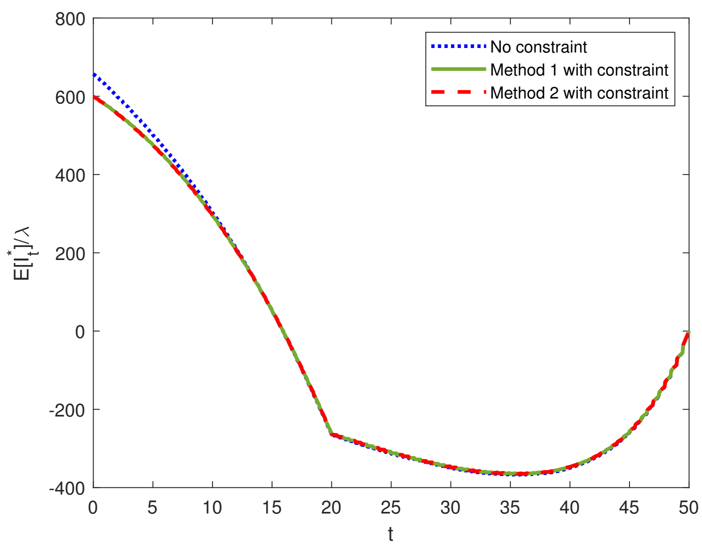


Figure 4.4: Optimal face-value $E[I_t^*]/\lambda_{x+t}$ for each method in Example 4.6.1

Example 4.6.2. In this example, we study the case when the stock appreciation rate has a perturbation, and the risk-free interest rate and volatility are both constant, i.e., $\mu(t) = 0.07 + 0.03\sin(t/2)$, $r(t) = 0.02$, and $\sigma(t) = 0.2$.

Table 4.2 shows the lower and upper bounds for each method. We use the default "interior-point" algorithm provided in the Matlab package "fmincon" to minimize the upper bounds in each method. We use the same accuracy and initial value sampling design for numerical settings as in Example 4.6.1.

From Table 4.2, we see that Method 1 generates a big duality gap of 0.1663950, a relative gap of 1.9828%, and suffers from a large welfare loss of 3.9263%. When we apply Method 2 with "(1-10-2)" structure under the ReLU activation function, the duality gap is slightly improved to 0.0828014, the relative gap decreases to 0.9921%, and the welfare loss falls down to 1.9743%. Lastly, we apply the snake function,

$$\text{Snake}_a := x + \frac{1}{a} sin^2(ax), \tag{4.50}$$

which is an activation function designed to learn the periodic function (see Ziyin et al. (2020)). In the numerical example, we choose $a = 10$. With the same initial values sampling and training iteration following Example 4.6.1, we observe that the snake activation function greatly reduces the duality gap and provides much tighter lower and upper bounds. More specifically, the duality gap shrinks from 0.1663950 to only 0.0230592, the relative gap reduces from 1.9828% to 0.2762%, and the welfare loss decreases from 3.9263% to 0.5516%.

Figure 4.5 shows the change of the upper bound with the training iteration. We see that the three methods decrease at the same rate, but Method 2, with the snake activation function stays lower than the other methods. Figure 4.6 displays each method's learning result, v^*. We observe that Method 1 can not identify the perturbation pattern of drift $\mu(t)$ but only learns $v(t)$ as zig-zag lines. Method 2 with ReLU activation function ($max(0, x)$) under the structure "(1-10-2)" can identify the first period of $\mu(t)$'s perturbation, but not other periods. Finally, Method 2 with Snake activation function (4.50) under structure "(1-10-2)" not only perfectly identifies the perturbation pattern of $\mu(t)$, but also learns the decreasing trend before the retirement time $T_R = 20$. This is the reason why "Method 2 Snake (1-10-2)" outperforms the other methods. Similarly to Figure 4.4, Figure 4.7 also shows that when considering the trading constraint, the individual reduces their demand for life insurance. Moreover, the individual's demand for life insurance also forms a "spoon shape" but has some perturbations after the retirement time $T_R = 20$.

Table 4.2: **Lower and upper bounds for Example 4.6.2**

	Method 1	Method 2	Method 2
Structure	Affine	(1-10-2)	(1-10-2)
Activation function	None	ReLU	Snake
Upper bound	-8.2255790	-8.2633075	-8.3259363
Lower bound	-8.3919740	-8.3461089	-8.3489955
Duality gap	0.1663950	0.0828014	0.0230592
Relative gap	1.9828%	0.9921%	0.2762%
Welfare loss	3.9263%	1.9743%	0.5516%
Time elapsed	7.59 hours	8.82 hours	10.79 hours

The simulation accuracy and terms in this table are the same as those in Table 4.1.

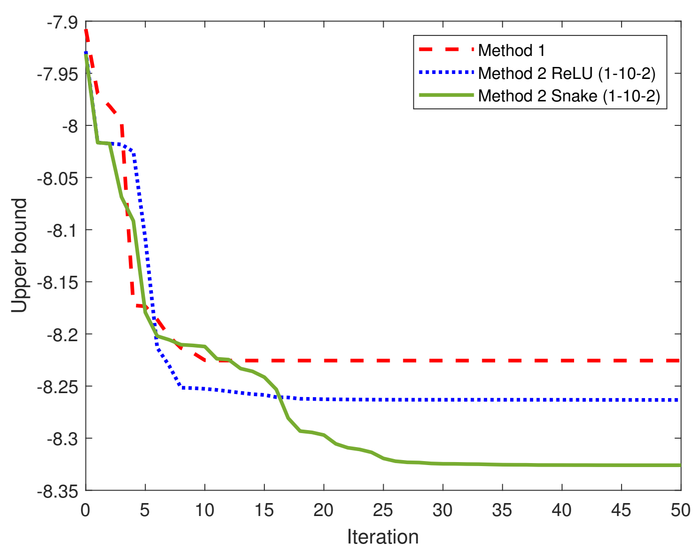
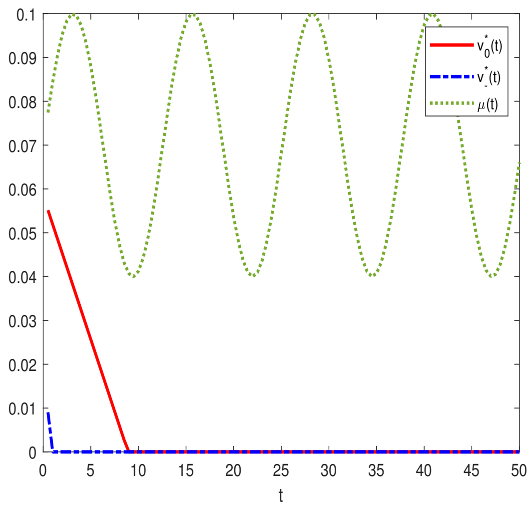
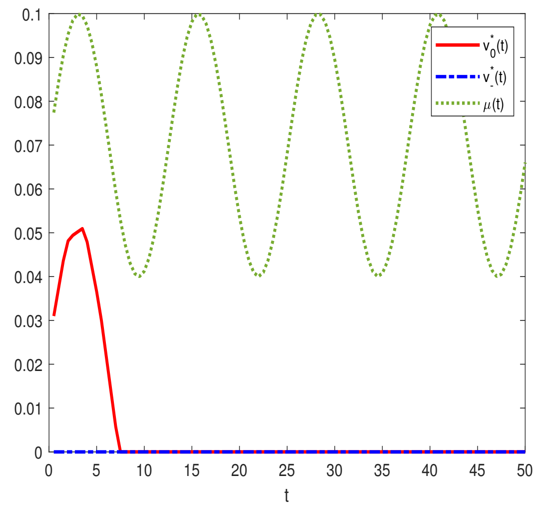


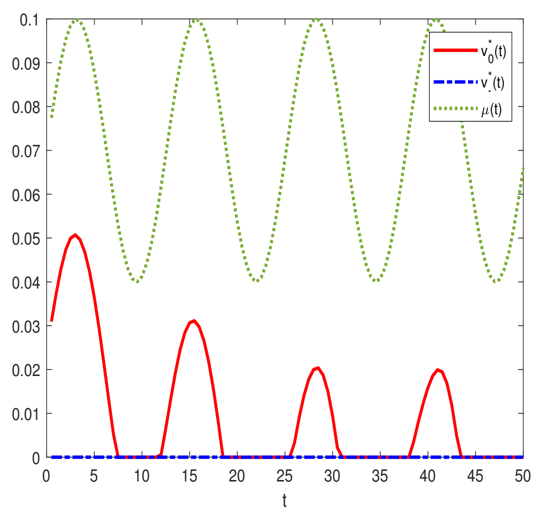
Figure 4.5: Change of upper bound in each training iteration for Example 4.6.2



(a) Method 1



(b) Method 2 ReLU (1-10-2)



(c) Method 2 Snake (1-10-2)

Figure 4.6: Optimal v^* for each method in Example 4.6.2

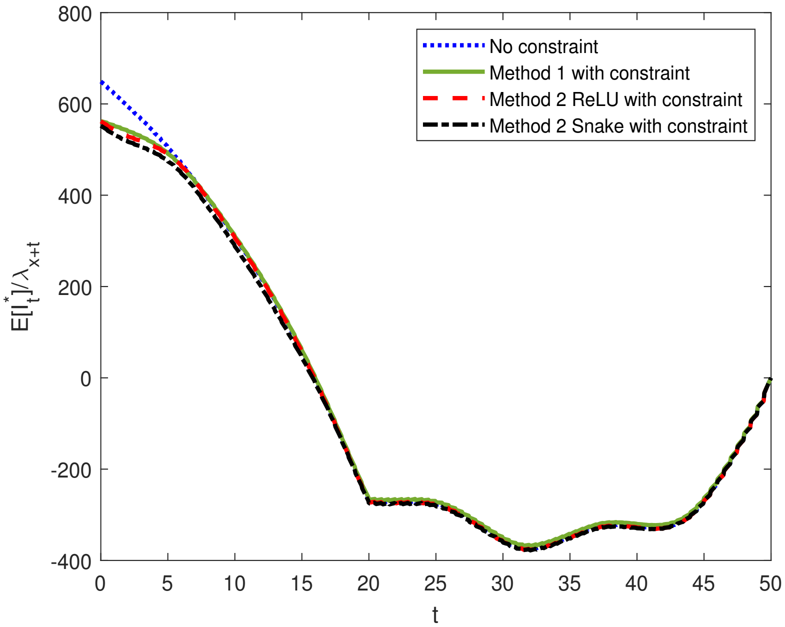


Figure 4.7: Optimal face-value $E[I_t^*]/\lambda_{x+t}$ for each method in Example 4.6.2

4.7 Conclusion

This chapter studies the constrained portfolio optimization problem in a generalized life cycle model. The individual has a stochastic income and allocates his or her wealth among stocks, a bond, and life insurance to optimize consumption, death benefits, and terminal wealth. In addition, the individual's trading strategy is restricted to a non-empty, closed convex set, which contains non-tradeable assets, no short-selling, and no borrowing constraints as special cases.

Following the framework of Cuoco (1997), we first define the artificial markets and change the dynamic budget constraint in the primal problem to a group of static budget constraints in the artificial markets. Then, through the Lagrangian dual control approach, we transfer the primal problem to the dual problem and prove a one-to-one relationship between the optimal solutions of the primal problem and the dual problem. Finally, we use the "relaxation projection" technique (see Levin (1976)) to prove the existence of the primal problem. In Cuoco (1997), the interest rate and income process are both assumed to be uniformly bounded. We extend the interest rate to satisfy a finite expectation constraint and enlarge the income process assumption to a condition containing uniformly bounded case.

To the best of our knowledge, this is the first application of neural networks to the constrained portfolio optimization problem in the life cycle model. We find that when considering the trading constraint, the individual will reduce his or her demand for life insurance. Furthermore, compared with the SAMS approach in Bick et al. (2013), we find that both approaches have a similar performance when interest rate, stock appreciation rate, and volatility are all constant. When the underlying model is more complex (e.g., the stock appreciation rate has a perturbation in time), the SAMS approach is inadequate to provide a tight lower and upper bound, but the neural network approach still works very well. In general, the dual control neural network approach, overcomes the defects of the SAMS approach and can inspire further future work on applying neural networks to study the constrained portfolio optimization problem.

Chapter 5

Conclusions and Future work

This thesis studies optimal insurance strategies for the individual using stochastic control approach. In general, Chapter 2 studies the optimal insurance design for non-life insurance. The individual's loss is described by the compound Poisson process. The topics for Chapter 3 and Chapter 4 are closely related and they study the individual's demand for life insurance. The markets in these two chapters include not only the financial risk generated by Brownian motions but also the individual's mortality risk.

Specifically, in Chapter 2, we propose and solve the optimal insurance problem for an individual exhibiting internal habit formation. Under general utilities, we establish that the optimal per-claim insurance must be deductible insurance, provided that the expected value principle is used in insurance pricing. Under exponential utility, we obtain explicit solutions for individuals who can purchase deductible or proportional insurance. For both types of insurance, the individual gradually increases insurance coverage as he or she ages. Moreover, the presence of habit formation reduces insurance coverage such that an individual who is restricted to proportional insurance may opt out of the insurance market, especially at early ages. These results suggest that habit formation and incomplete insurance markets (such that individuals can only purchase proportional insurance) can partially contribute to explaining the prevailing global underinsurance phenomenon, as documented in (Lloyd's, 2018).

Chapter 3 considers a DC pension plan management problem under the two-factor model proposed by Koijen et al. (2011). We find that an individual's demand for life insurance exhibits a hump shape with age and a "double top" pattern for the two factors. To be specific, the individual purchases more insurance at the old ages before retirement, or in extreme market scenarios in which real short rate and expected inflation are both high or both low. These behaviors are caused by the combined effects of the components in the optimal insurance premium. Furthermore, our model builds a DC account that resembles a variable annuity with endogenously determined time-varying death benefits. It relaxes the constraints on variable annuity's death benefits and can inspire new and innovative actuarial products.

Chapter 4 studies the constrained portfolio optimization problem in a generalized life cycle model. We first propose the dual control neural network approach to study this problem and find that when considering the trading constraint, the individual will reduce his or her demand for life insurance. Compared with the SAMS approach in Bick et al. (2013), we find that the two approaches have a similar performance when interest rate, stock appreciation rate, and volatility are all constant. When things come to a more complex case (e.g., stock appreciation rate has a perturbation in time), the SAMS approach is inadequate to provide a tight lower and upper bound, but the neural network approach still works very well. In general, our pioneering work, the dual control neural network approach, implements the defects of the SAMS approach and can inspire more further future work on applying neural networks to study the constrained portfolio optimization problem.

Lastly, we consider the following future work for the three chapters. First, in Chapter 2, we only study the habit formation effect on the optimal insurance strategy under the expected premium principle. The optimal insurance strategy and the influence of habit formation under other premium principles (such as the variance premium principle, value-at-risk premium principle, Wang's premium principle, etc.) are not known and are worth studying. Meanwhile, for empirical work, we can estimate the parameters for individual consumption habits in different countries and determine whether historical consumption plays a significant role in explaining countries' insurance gaps.

Second, in Chapter 3, we consider the DC pension management without trading constraint. We can use the techniques from Chapter 4 and add the trading constraint (such as no short-selling constraint, no borrowing constraint, and non-negative constraint of life insurance) to formulate a model in a more realistic way. Moreover, we can also use the market data to see whether the insurance sales follow a "hump" shape pattern in time and a "double-top" pattern when the interest rate and expected inflation are both very low or very high.

Third, for Chapter 4, we plan to apply our dual control neural network approach to other classical financial models (such as the regime-switching model, factor model, constant elasticity variance (CEV) model, Heston model, etc.), improve its efficiency, and make it a more reliable algorithm for financial practice. Furthermore, we can also study whether the trading constraints reduce the individual's insurance demand empirically. For example, in the insurance economics, we have an indicator called the insurance penetration rate. It is a quantity measuring a country's insurance sector's development and defined as the ratio of the total insurance premium over the gross domestic product (GDP) in a given year. We can set the insurance penetration rate as the dependent variable and trading constraints as latent variables to study whether a free financial market (with less trading constraints) can boom the insurance market.

References

Abel, A. B. (1990). Asset prices under habit formation and catching up with the Joneses. American Economic Review, 80(2):38–42.

Abou-Kandil, H., Freiling, G., Ionescu, V., and Jank, G. (2012). Matrix Riccati Equations in Control and Systems Theory. Birkhäuser.

Angelini, V. (2009). Consumption and habit formation when time horizon is finite. Economics Letters, 103(2):113–116.

Arrow, K. J. (1963). Uncertainty and the welfare economics of medical care. American Economic Review, 53(5):941–973.

Babbs, S. H. and Nowman, K. B. (1999). Kalman filtering of generalized Vasicek term structure models. Journal of Financial and Quantitative Analysis, 34(1):115–130.

Ballester, J., Robine, J.-M., Herrmann, F. R., and Rodó, X. (2019). Effect of the great recession on regional mortality trends in europe. Nature Communications, 10(1):679.

Bardhan, I. (1994). Consumption and investment under constraints. Journal of Economic Dynamics and Control, 18(5):909–929.

Basak, S. (1995). A general equilibrium model of portfolio insurance. The Review of Financial Studies, 8(4):1059–1090.

Battocchio, P. and Menoncin, F. (2004). Optimal pension management in a stochastic framework. Insurance: Mathematics and Economics, 34(1):79–95.

Ben-Arab, M., Briys, E., and Schlesinger, H. (1996). Habit formation and the demand for insurance. Journal of Risk and Insurance, pages 111–119.

Bensoussan, A. (2004). Stochastic Control of Partially Observable Systems. Cambridge University Press.

Bian, L., Li, Z., and Yao, H. (2018). Pre-commitment and equilibrium investment strategies for the DC pension plan with regime switching and a return of premiums clause. Insurance: Mathematics and Economics, 81:78–94.

Bick, B., Kraft, H., and Munk, C. (2013). Solving constrained consumption–investment problems by simulation of artificial market strategies. Management Science, 59(2):485–503.

Blake, D., Wright, I., and Zhang, Y. (2008). Optimal funding and investment strategies in defined contribution pension plans under Epstein-Zin utility (actuarial research paper no. 186). Faculty of Actuarial Science & Insurance, City University London.

Borch, K. (1975). Optimal insurance arrangements. ASTIN Bulletin: The Journal of the IAA, 8(3):284–290.

Boulier, J.-F., Huang, S., and Taillard, G. (2001). Optimal management under stochastic interest rates: the case of a protected defined contribution pension fund. Insurance: Mathematics and Economics, 28(2):173–189.

Boyle, P., Tan, K. S., Wei, P., and Zhuang, S. C. (2022). Annuity and insurance choice under habit formation. Insurance: Mathematics and Economics, 105:211–237.

Briys, E. (1986). Insurance and consumption: The continuous time case. The Journal of Risk and Insurance, 53(4):718–723.

Broadbent, J., Palumbo, M., and Woodman, E. (2006). The shift from defined benefit to defined contribution pension plans–implications for asset allocation and risk management. Reserve Bank of Australia, Board of Governors of the Federal Reserve System and Bank of Canada, pages 1–54.

Brüning, M. and Thuilliez, J. (2019). Mortality and macroeconomic conditions: what can we learn from france? Demography, 56(5):1747–1764.

Caballero, R. J. (1991). Earnings uncertainty and aggregate wealth accumulation. American Economic Review, pages 859–871.

Cervini-Plá, M. and Vall-Castelló, J. (2021). Business cycle and mortality in spain. The European Journal of Health Economics, 22(8):1289–1299.

Chabakauri, G. (2013). Dynamic equilibrium with two stocks, heterogeneous investors, and portfolio constraints. The Review of Financial Studies, 26(12):3104–3141.

Chen, Z. (2009). The Logic of Finance. Beijing: Modern Press.

Chen, Z., Li, Z., Zeng, Y., and Sun, J. (2017). Asset allocation under loss aversion and minimum performance constraint in a DC pension plan with inflation risk. Insurance: Mathematics and Economics, 75:137–150.

Chicone, C. (2006). Ordinary Differential Equations with Applications, volume 34. Springer Science & Business Media.

Chung, K. L. (2013). Lectures from Markov Processes to Brownian Motion, volume 249. Springer Science & Business Media.

Cocco, J. F., Gomes, F. J., and Maenhout, P. J. (2005). Consumption and portfolio choice over the life cycle. The Review of Financial Studies, 18(2):491–533.

Constantinides, G. M. (1990). Habit formation: A resolution of the equity premium puzzle. Journal of Political Economy, 98(3):519–543.

Cox, J. C. and Huang, C.-f. (1989). Optimal consumption and portfolio policies when asset prices follow a diffusion process. Journal of Economic Theory, 49(1):33–83.

Cox, J. C. and Huang, C.-F. (1991). A variational problem arising in financial economics. Journal of Mathematical Economics, 20(5):465–487.

Cuoco, D. (1997). Optimal consumption and equilibrium prices with portfolio constraints and stochastic income. Journal of Economic Theory, 72(1):33–73.

Cvitanić, J. and Karatzas, I. (1992). Convex duality in constrained portfolio optimization. The Annals of Applied Probability, pages 767–818.

Cvitanić, J. and Karatzas, I. (1993). Hedging contingent claims with constrained portfolios. The Annals of Applied Probability, pages 652–681.

Deelstra, G., Grasselli, M., and Koehl, P.-F. (2003). Optimal investment strategies in the presence of a minimum guarantee. Insurance: Mathematics and Economics, 33(1):189–207.

Dellacherie, C. and Meyer, P.-A. (2011). Probabilities and Potential, C: Potential Theory for Discrete and Continuous Semigroups. Elsevier.

Detemple, J. B. and Zapatero, F. (1991). Asset prices in an exchange economy with habit formation. Econometrica: Journal of the Econometric Society, pages 1633–1657.

Dong, Y. and Zheng, H. (2019). Optimal investment of dc pension plan under short-selling constraints and portfolio insurance. Insurance: Mathematics and Economics, 85:47–59.

Duffie, D. and Kan, R. (1996). A yield-factor model of interest rates. Mathematical Finance, 6(4):379–406.

Fed (2001). Monetary Policy Report to the Congress - February,2001. Federal Reserve System.

Fed (2008). Monetary Policy Report to the Congress - July,2008. Federal Reserve System.

Fed (2020). Monetary Policy Report to the Congress - June,2020. Federal Reserve System.

Forfar, D., McCutcheon, J., and Wilkie, A. (1988). On graduation by mathematical formula. Journal of the Institute of Actuaries (1886-1994), 115(1):1–149.

Guan, G. and Liang, Z. (2014). Optimal management of DC pension plan in a stochastic interest rate and stochastic volatility framework. Insurance: Mathematics and Economics, 57:58–66.

Gürkaynak, R. S., Sack, B., and Wright, J. H. (2007). The US treasury yield curve: 1961 to the present. Journal of Monetary Economics, 54(8):2291–2304.

Haaland, V. F. and Telle, K. (2015). Pro-cyclical mortality across socioeconomic groups and health status. Journal of Health Economics, 39:248–258.

Halliday, T. J. (2014). Unemployment and mortality: Evidence from the psid. Social Science & Medicine, 113:15–22.

Hambel, C., Kraft, H., and Munk, C. (2022). Solving life-cycle problems with biometric risk by artificial insurance markets. Scandinavian Actuarial Journal, 2022(4):307–327.

Han, N.-w. and Hung, M.-w. (2012). Optimal asset allocation for DC pension plans under inflation. Insurance: Mathematics and Economics, 51(1):172–181.

Harrison, J. M. and Kreps, D. M. (1979). Martingales and arbitrage in multiperiod securities markets. Journal of Economic Theory, 20(3):381–408.

Harvey, A. C. (1990). Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge university press.

Haugh, M. B., Kogan, L., and Wang, J. (2006). Evaluating portfolio policies: A duality approach. Operations Research, 54(3):405–418.

He, H. and Pages, H. F. (1993). Labor income, borrowing constraints, and equilibrium asset prices. Economic Theory, 3(4):663–696.

He, H. and Pearson, N. D. (1991). Consumption and portfolio policies with incomplete markets and short-sale constraints: The infinite dimensional case. Journal of Economic Theory, 54(2):259–304.

He, L. and Liang, Z. (2013). Optimal investment strategy for the DC plan with the return of premiums clauses in a mean–variance framework. Insurance: Mathematics and Economics, 53(3):643–649.

Hindy, A. (1995). Viable prices in financial markets with solvency constraints. Journal of Mathematical Economics, 24(2):105–135.

Hindy, A. and Huang, M. (1995). Asset pricing with linear collateral constraints. Unpublished Manuscripts.

Honda, T. and Kamimura, S. (2011). On the verification theorem of dynamic portfolio-consumption problems with stochastic market price of risk. Asia-Pacific Financial Markets, 18(2):151–166.

Huang, H., Milevsky, M. A., and Salisbury, T. S. (2012). Optimal retirement consumption with a stochastic force of mortality. Insurance: Mathematics and Economics, 51(2):282–291.

Huang, H., Milevsky, M. A., and Wang, J. (2008). Portfolio choice and life insurance: The crra case. Journal of Risk and Insurance, 75(4):847–872.

IRS (2019). Publication 575 - Pension and Annuity Income. Internal Revenue Service.

Jacod, J. and Shiryaev, A. (2013). Limit Theorems for Stochastic Processes, volume 288. Springer Science & Business Media.

Jin, X. and Zhang, K. (2013). Dynamic optimal portfolio choice in a jump-diffusion model with investment constraints. Journal of Banking & Finance, 37(5):1733–1746.

Kamma, T. and Pelsser, A. (2022). Near-optimal asset allocation in financial markets with trading constraints. European Journal of Operational Research, 297(2):766–781.

Karatzas, I., Lehoczky, J. P., and Shreve, S. E. (1987). Optimal portfolio and consumption decisions for a “small investor” on a finite horizon. SIAM Journal on Control and Optimization, 25(6):1557–1586.

Karatzas, I., Lehoczky, J. P., Shreve, S. E., and Xu, G.-L. (1991). Martingale and duality methods for utility maximization in an incomplete market. SIAM Journal on Control and optimization, 29(3):702–730.

Koijen, R. S., Nijman, T. E., and Werker, B. J. (2011). Optimal annuity risk management. Review of Finance, 15(4):799–833.

Konicz, A. K. and Mulvey, J. M. (2015). Optimal savings management for individuals with defined contribution pension plans. European Journal of Operational Research, 243(1):233–247.

Korn, R. and Kraft, H. (2004). On the stability of continuous-time portfolio problems with stochastic opportunity set. Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics, 14(3):403–414.

Kraft, H. (2004). Optimal portfolios with stochastic interest rates and defaultable assets. Springer Science & Business Media.

Kraft, H., Munk, C., Seifried, F. T., and Wagner, S. (2017). Consumption habits and humps. Economic Theory, 64(2):305–330.

Lam, J.-P. and Piérard, E. (2017). The time-varying relationship between mortality and business cycles in the usa. Health Economics, 26(2):164–183.

Larsen, K. and Žitković, G. (2013). On utility maximization under convex portfolio constraints. The Annals of Applied Probability, 23(2):665–692.

Larsen, L. S. and Munk, C. (2012). The costs of suboptimal dynamic asset allocation: General results and applications to interest rate risk, stock volatility risk, and growth/value tilts. Journal of Economic Dynamics and Control, 36(2):266–293.

Levin, V. L. (1976). Extremal problems with convex functionals that are lower semicontinuous with respect to convergence in measure. Soviet Math. Dokl., 16(5):1384–1388.

Li, D., Rong, X., Zhao, H., and Yi, B. (2017). Equilibrium investment strategy for dc pension plan with default risk and return of premiums clauses under CEV model. Insurance: Mathematics and Economics, 72:6–20.

Li, W., Tan, K. S., and Wei, P. (2021). Demand for non-life insurance under habit formation. Insurance: Mathematics and Economics, 101:38–54.

Liu, J. and Pan, J. (2003). Dynamic derivative strategies. Journal of Financial Economics, 69(3):401–430.

Lloyd's (2018). A World at Risk: Closing the Insurance Gap. Lloyd's.

McInerney, M. and Mellor, J. M. (2012). Recessions and seniors' health, health behaviors, and healthcare use: Analysis of the medicare current beneficiary survey. Journal of Health Economics, 31(5):744–751.

Menoncin, F. and Vigna, E. (2017). Mean–variance target-based optimisation for defined contribution pension schemes in a stochastic framework. Insurance: Mathematics and Economics, 76:172–184.

Merton, R. C. (1971). Optimum consumption and portfolio rules in a continuous-time model. Journal of Economic Theory, 3(4):373–413.

Moore, K. S. and Young, V. R. (2006). Optimal insurance in a continuous-time model. Insurance: Mathematics and Economics, 39(1):47–68.

Mossin, J. (1968). Aspects of rational insurance purchasing. Journal of Political Economy, 76(4, Part 1):553–568.

Mostovyi, O. and Sîrbu, M. (2020). Optimal investment and consumption with labor income in incomplete markets. The Annals of Applied Probability, 30(2):747–787.

Munk, C. (2008). Portfolio and consumption choice with stochastic investment opportunities and habit formation in preferences. Journal of Economic Dynamics and Control, 32(11):3560–3589.

Munk, C. and Sørensen, C. (2010). Dynamic asset allocation with stochastic income and interest rates. Journal of Financial Economics, 96(3):433–462.

OECD (2020). Pension Markets in Focus 2020. OECD.

Øksendal, B. and Sulem, A. (2007). Applied Stochastic Control of Jump Diffusions. Springer Science & Business Media.

Perera, R. S. (2010). Optimal consumption, investment and insurance with insurable risk for an investor in a lévy market. Insurance: Mathematics and Economics, 46(3):479–484.

Perera, R. S. (2013). Optimal investment, consumption–leisure, insurance and retirement choice. Annals of Finance, 9(4):689–723.

Polkovnichenko, V. (2006). Life-cycle portfolio choice with additive habit formation preferences and uninsurable labor income risk. Review of Financial Studies, 20(1):83–124.

Pratt, J. W. (1964). Risk aversion in the small and in the large. Econometrica, 32(1/2):122–136.

Qian, L., Wang, W., Wang, R., and Tang, Y. (2010). Valuation of equity-indexed annuity under stochastic mortality and interest rate. Insurance: Mathematics and Economics, 47(2):123–129.

Raviv, A. (1979). The design of an optimal insurance policy. American Economic Review, 69(1):84–96.

RBC (2020). Pensions part 2 - Defined Contribution Plans. Royal Bank of Canada.

Rees, E. (1922). Graphical discussion of the roots of a quartic equation. The American Mathematical Monthly, 29(2):51–55.

Rockafellar, R. T. (1970). Convex analysis princeton university press. Princeton, NJ.

Ruhm, C. J. (2000). Are recessions good for your health? The Quarterly Journal of Economics, 115(2):617–650.

Ruhm, C. J. (2003). Good times make you sick. Journal of Health Economics, 22(4):637–658.

Ruhm, C. J. (2015). Recessions, healthy no more? Journal of Health Economics, 42:17–28.

SEC (2009). Variable Annuities: What You Should Know. U.S. Securities and Exchange Commission.

Shen, Y. and Siu, T. K. (2013). Longevity bond pricing under stochastic interest rate and mortality with regime-switching. Insurance: Mathematics and Economics, 52(1):114–123.

Steffensen, M. and Thøgersen, J. (2019). Personal non-life insurance decisions and the welfare loss from flat deductibles. ASTIN Bulletin: The Journal of the IAA, 49(1):85–116.

Stevens, A. H., Miller, D. L., Page, M. E., and Filipski, M. (2015). The best of times, the worst of times: understanding pro-cyclical mortality. American Economic Journal: Economic Policy, 7(4):279–311.

Sun, J., Li, Z., and Zeng, Y. (2016). Precommitment and equilibrium investment strategies for defined contribution pension plans under a jump–diffusion model. Insurance: Mathematics and Economics, 67:158–172.

Sun Life Financial, I. (2017). A Guide to Purolator Inc. Retirement Program Defined Contribution Pension Plan (DCPP). Sun Life Financial, Inc.

Sundaresan, S. M. (1989). Intertemporally dependent preferences and the volatility of consumption and wealth. Review of Financial Studies, 2(1):73–89.

Tan, K. S., Wei, P., Wei, W., and Zhuang, S. C. (2020). Optimal dynamic reinsurance policies under a generalized Denneberg's absolute deviation principle. European Journal of Operational Research, 282(1):345–362.

Touzi, N. (2000). Optimal insurance demand under marked point processes shocks. Annals of Applied Probability, pages 283–312.

Wang, N. (2006). Generalizing the permanent-income hypothesis: Revisiting Friedman's conjecture on consumption. Journal of Monetary Economics, 53(4):737–752.

Wang, N. (2009). Optimal consumption and asset allocation with unknown income growth. Journal of Monetary Economics, 56(4):524–534.

Wang, P. and Li, Z. (2018). Robust optimal investment strategy for an aam of dc pension plans with stochastic interest rate and stochastic volatility. Insurance: Mathematics and Economics, 80:67–83.

Wang, P., Li, Z., and Sun, J. (2021). Robust portfolio choice for a dc pension plan with inflation risk and mean-reverting risk premium under ambiguity. Optimization, 70(1):191–224.

Wu, H. and Zeng, Y. (2015). Equilibrium investment strategy for defined-contribution pension schemes with generalized mean–variance criterion and mortality risk. Insurance: Mathematics and Economics, 64:396–408.

Xu, G.-L. and Shreve, S. E. (1992). A duality method for optimal consumption and investment under short-selling prohibition. i. general market coefficients. The Annals of Applied Probability, pages 87–112.

Xue, X., Wei, P., and Weng, C. (2019). Derivatives trading for insurers. Insurance: Mathematics and Economics, 84:40–53.

Yao, H., Lai, Y., Ma, Q., and Jian, M. (2014). Asset allocation for a DC pension fund with stochastic income and mortality risk: A multi-period mean–variance framework. Insurance: Mathematics and Economics, 54:84–92.

Yao, H., Yang, Z., and Chen, P. (2013). Markowitz's mean–variance defined contribution pension fund management under inflation: A continuous-time model. Insurance: Mathematics and Economics, 53(3):851–863.

Zeng, X., Carson, J. M., Chen, Q., and Wang, Y. (2016). Optimal life insurance with no-borrowing constraints: duality approach and example. Scandinavian Actuarial Journal, 2016(9):793–816.

Zhang, A. and Ewald, C.-O. (2010). Optimal investment for a pension fund under inflation risk. Mathematical Methods of Operations Research, 71(2):353–369.

Zhou, H., Zhou, K. Q., and Li, X. (2022). Stochastic mortality dynamics driven by mixed fractional brownian motion. Insurance: Mathematics and Economics, 106:218–238.

Ziyin, L., Hartwig, T., and Ueda, M. (2020). Neural networks fail to learn periodic functions and how to fix it. Advances in Neural Information Processing Systems, 33:1583–1594.

Zou, B. and Cadenillas, A. (2014). Explicit solutions of optimal consumption, investment and insurance problems with regime switching. Insurance: Mathematics and Economics, 58:159–167.

APPENDICES

Appendix A

Proofs for Chapter 2

A.1 Proof of Proposition 2.4.1

We conjecture that

$$\phi(t,x,h) = -\frac{1}{\gamma}e^{-\gamma(a(t)x+b(t)h+g(t))}. \tag{A.1}$$

Because $\phi(T,x,h) = U_2(x) = -\frac{\omega}{\gamma}e^{-\gamma x}$, we have the following boundary condition

$$a(T) = 1, \quad b(T) = 0, \quad g(T) = -\frac{\ln(w)}{\gamma}.$$

Suppose $\phi_x(t,x,h) = y$, then the inverse function of ϕ_x w.r.t. x is given by

$$\phi_x^{-1}(t,y,h) = -\frac{1}{\gamma a(t)}\ln\left(\frac{y}{a(t)}\right) - \frac{b(t)}{a(t)}h - \frac{g(t)}{a(t)}.$$

In view of (2.6), the optimal indemnity is given by

$$I_t^*(Y) = \left[Y - \frac{\ln(1+\theta)}{\gamma a(t)}\right]^+,$$

Plugging (A.1) into (2.5), we have

$$e^{-\gamma(c-h)} = [a(t) - \alpha b(t)]e^{-\gamma(a(t)x+b(t)h+g(t))},$$

and the optimal consumption is given by

$$c_t^* = -\frac{1}{\gamma}\ln[a(t)-\alpha b(t)] + a(t)X_t^* + (b(t)+1)h_t^* + g(t). \tag{A.2}$$

Assume the random loss Y has pdf f and tail distribution function $\overline{F} = 1-F$, we have

$$\begin{aligned} E[I_t^*(Y)] &= \int_{\frac{\ln(1+\theta)}{\gamma a(t)}}^{\infty} \left(y - \frac{\ln(1+\theta)}{\gamma a(t)}\right) f(y)dy \\ &= \int_{\frac{\ln(1+\theta)}{\gamma a(t)}}^{\infty} yf(y)dy - \frac{\ln(1+\theta)}{\gamma a(t)}\overline{F}\left(\frac{\ln(1+\theta)}{\gamma a(t)}\right) \end{aligned} \tag{A.3}$$

$$\begin{aligned} E[\phi(t, x-Y+I_t^*(Y), h)] &= E\left[-\frac{1}{\gamma}e^{-\gamma\left[a(t)\left(x-Y+\left(Y-\frac{\ln(1+\theta)}{\gamma a(t)}\right)^+\right)+b(t)h+g(t)\right]}\right] \\ &= -\frac{1}{\gamma}e^{-\gamma\left[a(t)\left(x-\frac{\ln(1+\theta)}{\gamma a(t)}\right)+b(t)h+g(t)\right]}\overline{F}\left(\frac{\ln(1+\theta)}{\gamma a(t)}\right) \\ &\quad -\frac{1}{\gamma}e^{-\gamma\left[a(t)x+b(t)h+g(t)\right]}\int_0^{\frac{\ln(1+\theta)}{\gamma a(t)}} e^{\gamma a(t)y} f(y)dy. \end{aligned} \tag{A.4}$$

Substituting (A.1), (A.2), (A.3) and (A.4) into the HJB equation (2.4), we have

$$\begin{aligned} 0 &= x[a'(t) + ra(t) - (a(t)-\alpha b(t))a(t)] \\ &\quad + h[b'(t) - \beta b(t) - (a(t)-\alpha b(t))(b(t)+1)] \\ &\quad + g'(t) - (a(t)-\alpha b(t))g(t) + \frac{\lambda+\delta}{\gamma} + \{\ln[a(t)-\alpha b(t)] - 1\}\frac{a(t)-\alpha b(t)}{\gamma} \\ &\quad -\lambda(1+\theta)\left[\int_{\frac{\ln(1+\theta)}{\gamma a(t)}}^{\infty} yf(y)dy - \frac{\ln(1+\theta)}{\gamma a(t)}\overline{F}\left(\frac{\ln(1+\theta)}{\gamma a(t)}\right)\right]a(t) \\ &\quad -\frac{\lambda}{\gamma}\left[(1+\theta)\overline{F}\left(\frac{\ln(1+\theta)}{\gamma a(t)}\right) + \int_0^{\frac{\ln(1+\theta)}{\gamma a(t)}} e^{\gamma a(t)y} f(y)dy\right]. \end{aligned}$$

By separation of variables, we arrive at the following ODE system

$$a'(t) + ra(t) - (a(t) - \alpha b(t))a(t) = 0, \quad a(T) = 1, \tag{A.5}$$

$$b'(t) - \beta b(t) - (a(t) - \alpha b(t))(b(t) + 1) = 0, \quad b(T) = 0, \tag{A.6}$$

$$\begin{aligned} & g'(t) - (a(t) - \alpha b(t))g(t) + \frac{\lambda + \delta}{\gamma} + \{\ln[a(t) - \alpha b(t)] - 1\}\frac{a(t) - \alpha b(t)}{\gamma} \\ & -\lambda(1+\theta)\left[\int_{\frac{\ln(1+\theta)}{\gamma a(t)}}^{\infty} yf(y)dy - \frac{\ln(1+\theta)}{\gamma a(t)}\overline{F}\left(\frac{\ln(1+\theta)}{\gamma a(t)}\right)\right]a(t) \\ & -\frac{\lambda}{\gamma}\left[(1+\theta)\overline{F}\Big(\frac{\ln(1+\theta)}{\gamma a(t)}\Big) + \int_0^{\frac{\ln(1+\theta)}{\gamma a(t)}} e^{\gamma a(t)y} f(y)dy\right] = 0, \quad g(T) = -\frac{\ln(w)}{\gamma}. \end{aligned} \tag{A.7}$$

Note that (A.5) and (A.6) form a second-order coupled ODE system, whose solution is not readily available. The following lemma presents explicit solution to (A.5) and (A.6).

Lemma A.1.1. *The solution to the ODE system*

$$a'(t) + ra(t) - (a(t) - \alpha b(t))a(t) = 0, \quad a(T) = 1, \tag{A.8}$$

$$b'(t) - \beta b(t) - (a(t) - \alpha b(t))(b(t) + 1) = 0, \quad b(T) = 0, \tag{A.9}$$

is given by

$$\begin{aligned} a(t) &= 1/J(t), \\ b(t) &= G(t)/J(t), \\ G(t) &= (e^{-(r+\beta-\alpha)(T-t)} - 1)/(r + \beta - \alpha), \\ J(t) &= \int_t^T (1 - \alpha G(s))e^{-r(s-t)}ds + e^{-r(T-t)} \\ &= \frac{r+\beta}{(r+\beta-\alpha)r} + \left(1 - \frac{r+\beta}{(r+\beta-\alpha)r} - \frac{\alpha}{(r+\beta-\alpha)(\beta-\alpha)}\right)e^{-r(T-t)} \\ & \quad + \frac{\alpha}{(r+\beta-\alpha)(\beta-\alpha)}e^{-(r+\beta-\alpha)(T-t)}, \end{aligned} \tag{A.10}$$

Proof. Suppose that $b(t) = G(t)a(t)$ for a deterministic function $G(t)$. Substituting it into (A.8) and (A.9), we have

$$a'(t) + (\alpha G(t) - 1)a^2(t) + ra(t) = 0, \quad a(T) = 1, \tag{A.11}$$

and

$$G'(t)a(t)+G(t)a'(t)-\beta G(t)a(t)-a^2(t)G(t)-a(t)+\alpha G^2(t)a^2(t)+\alpha G(t)a(t)=0,\ \ G(T)=0. \tag{A.12}$$

Multiplying (A.11) by $G(t)$ and plugging it into (A.12), we have

$$G'(t)+(\alpha-\beta-r)G(t)=1,$$

and

$$G(t)=(e^{-(r+\beta-\alpha)(T-t)}-1)/(r+\beta-\alpha).$$

To solve (A.11), we make the transformation $a(t)=1/J(t)$. We have

$$-\frac{J'(t)}{J^2(t)}+(\alpha G(t)-1)\frac{1}{J^2(t)}+\frac{r}{J(t)}=0,\quad J(T)=1.$$

Multiply $J^2(t)$ on both hands sides, we have

$$J'(t)-rJ(t)=\alpha G(t)-1,$$

and

$$J(t)=\int_t^T e^{-r(s-t)}(1-\alpha G(s))ds+e^{-r(T-t)}.$$

□

We now solve the ODE (A.7). We have

$$\begin{aligned}
& d(e^{\int_t^T a(s)-\alpha b(s)ds}g(t)) \\
= \;& -e^{\int_t^T a(s)-\alpha b(s)ds}\left\{\frac{\lambda+\delta}{\gamma}+\{\ln[a(t)-\alpha b(t)]-1\}\frac{a(t)-\alpha b(t)}{\gamma}\right. \\
& -\lambda(1+\theta)\left[\int_{\frac{\ln(1+\theta)}{\gamma a(t)}}^{\infty} yf(y)dy-\frac{\ln(1+\theta)}{\gamma a(t)}\overline{F}\left(\frac{\ln(1+\theta)}{\gamma a(t)}\right)\right]a(t) \\
& \left.-\frac{\lambda}{\gamma}\left[(1+\theta)\overline{F}\Big(\frac{\ln(1+\theta)}{\gamma a(t)}\Big)+\int_0^{\frac{\ln(1+\theta)}{\gamma a(t)}} e^{\gamma a(t)y}f(y)dy\right]\right\},
\end{aligned}$$

and $g(t)$ is given by (2.11).

Finally, it is a simple exercise to show that the technical conditions in Theorem 2.3.1 are indeed satisfied, thereby proving the optimality.

A.2 Proof of Theorem 2.4.1

Proof. We divide the verification proof into three steps:

Step 1: Verify the optimal strategy (c^*, I^*) belongs to the admissible set $\mathcal{A}$.

First, we verify condition 1 in Definite 2.4.1.

Plug (2.10) and (2.9) into (2.2) and (2.3), we have

$$\begin{aligned} d\begin{pmatrix} X_t^* \\ h_t^* \end{pmatrix} &= \begin{pmatrix} r-a(t) & -[b(t)+1] \\ \alpha a(t) & \alpha[b(t)+1]-\beta \end{pmatrix}\begin{pmatrix} X_t^* \\ h_t^* \end{pmatrix} dt \\ &+ \begin{pmatrix} \frac{1}{\gamma}\ln[a(t)-\alpha b(t)] - g(t) - \lambda(1+\theta)\int\limits_{d(t)}^{\infty} \bar{F}_Y(y)dy \\ \alpha g(t) - \frac{\alpha}{\gamma}\ln[a(t)-\alpha b(t)] \end{pmatrix} dt \\ &- \begin{pmatrix} \min(Y,d(t)) & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} dN_t \\ d\widetilde{N}_t \end{pmatrix}, \end{aligned} \tag{A.13}$$

where N_t and $\widetilde{N}_t$ are independent Poisson processes with intensity λ and $\widetilde{\lambda}$. Denote $\boldsymbol{Z_t^*} = (X_t^*, h_t^*)^\top, \boldsymbol{\bar{N}_t} = (N_t, \widetilde{N}_t)^\top$, where $(\cdot)^\top$ means the transpose of vector and matrix, we can rewrite the (A.13) in the new notations

$$d\boldsymbol{Z_t^*} = \boldsymbol{B_t}\boldsymbol{Z_t^*}dt + \boldsymbol{f_1}(t)dt - \boldsymbol{f_2}(t,Y)d\boldsymbol{\bar{N}_t}. \tag{A.14}$$

For the compound poisson process (2.1), denote $\Delta A_t = A_t - A_{t-}$ the jump of A_t, we can define its random measure for any Borel set $U \in \mathbb{R}$

$$\eta_{A_t}(t,U) := \eta_{A_t}(t,U,\widetilde{\omega}) = \sum_{0<s\le t} \mathbb{1}\{\Delta A_s \in U\},$$

where $\widetilde{\omega}$ is a sample in the filtration $\mathcal{F}$ generated by compound Poisson process A_t. Similarly, we can define random measure $\eta_{\widetilde{N}_t}(t,U)$ for the Poisson process $\widetilde{N}_t$. Then, the equation (A.14) has the following equivalent form

$$\begin{aligned} d\boldsymbol{Z_t^*} &= \boldsymbol{B_t}\boldsymbol{Z_t^*}dt + \boldsymbol{f_1}(t)dt - \int_0^\infty \boldsymbol{f_2}(t,Y)\begin{pmatrix} \eta_{A_t}(dt,dy) \\ \eta_{\widetilde{N}_t}(dt,dy) \end{pmatrix} \\ &= \boldsymbol{B_t}\boldsymbol{Z_t^*}dt + \boldsymbol{f_1}(t)dt - \int_0^\infty \boldsymbol{f_2}(t,Y)\begin{pmatrix} \nu(dy)dt \\ \widetilde{\nu}(dy)dt \end{pmatrix} \\ &\quad - \int_0^\infty \boldsymbol{f_2}(t,Y)\begin{pmatrix} \eta_{A_t}(dt,dy) - \nu(dy)dt \\ \eta_{\widetilde{N}_t}(dt,dy) - \widetilde{\nu}(dy)dt \end{pmatrix}, \end{aligned} \tag{A.15}$$

where ν is the Levy measure of $A_t = \sum_{i=1}^{N_t} Y_i$ and $\widetilde{\nu}$ the Levy measure for $\widetilde{N}_t$. For compound Poisson process A_t, we have $\nu(U) = \lambda\mu_Y(U)$ for each Borel subset U of $\mathbb{R}$, where the distribution measure $\mu_Y(U) = P\{Y \in U\}$. For Poisson process $\widetilde{N}_t$ with intensity $\widetilde{\lambda}$, we have $\widetilde{\nu}(U) = \widetilde{\lambda}$ (see Theorem 1.5 and Example 1.6 in Øksendal and Sulem (2007)).

For (A.15), we can obtain it has linear growth

$$
\begin{aligned}
& \left|\left|\boldsymbol{B}_t\boldsymbol{Z_t^*} + \boldsymbol{f_1}(t) - \int_0^\infty \boldsymbol{f_2}(t,Y)\begin{pmatrix}\nu(dy)\\ \widetilde{\nu}(dy)\end{pmatrix}\right|\right|_2^2 + \int_0^\infty [\min(Y,d(t))]^2\nu(dy) \\
= & \left|\left|\boldsymbol{B}_t\boldsymbol{Z_t^*} + \boldsymbol{f_1}(t) - \begin{pmatrix}\lambda E[\min(Y,d(t))]\\ 0\end{pmatrix}\right|\right|_2^2 + \int_0^\infty [\min(Y,d(t))]^2\nu(dy) \\
\leq & \; C_0(1 + ||\boldsymbol{Z_t^*}||_2^2) + \int_0^\infty [\min(Y,d(t))]^2\lambda\mu_Y(dy) \\
\leq & \; C_0(1 + ||\boldsymbol{Z_t^*}||_2^2) + \lambda d^2(t)\int_0^\infty \mu_Y(dy) \\
\leq & \; C_1(1 + ||\boldsymbol{Z_t^*}||_2^2),
\end{aligned}
$$

where $||\cdot||_2$ is the Euclidean norm, and C_0 and C_1 are finite constants. The first inequality is due to the Minkowski inequality and the continuity of $a(t)$, $b(t)$, $d(t)$, and $g(t)$ in $[0,T]$. The third inequality is based on the continuity of $d(t)$ in $[0,T]$.

In addition, we can prove the Lipschitz continuity of (A.15)

$$
\begin{aligned}
& \left|\left|\left[\boldsymbol{B}_t\boldsymbol{Z_{2,t}^*} + \boldsymbol{f_1}(t) - \int_0^\infty \boldsymbol{f_2}(t,Y)\begin{pmatrix}\nu(dy)\\ \widetilde{\nu}(dy)\end{pmatrix}\right]\right.\right. \\
& \left.\left. - \left[\boldsymbol{B}_t\boldsymbol{Z_{1,t}^*} + \boldsymbol{f_1}(t) - \int_0^\infty \boldsymbol{f_2}(t,Y)\begin{pmatrix}\nu(dy)\\ \widetilde{\nu}(dy)\end{pmatrix}\right]\right|\right|_2^2 \\
& + \int_0^\infty |[\min(Y,d(t))]^2 - [\min(Y,d(t))]^2|^2\lambda\mu_Y(dy) \\
= & \; ||\boldsymbol{B}_t\boldsymbol{Z_{2,t}^*} - \boldsymbol{B}_t\boldsymbol{Z_{1,t}^*}||_2^2 \\
\leq & \; C_2||\boldsymbol{Z_{2,t}^*} - \boldsymbol{Z_{1,t}^*}||_2^2.
\end{aligned}
$$

The first inequality follows from the submultiplicative of induced matrix norm in Euclidean space. Then, by Theorem 1.19 in Øksendal and Sulem (2007), there exists a unique cadlag adapted solution Z_t^* such that

$$
E[||Z_t^*||_2^2] < \infty, \text{for all } t. \tag{A.16}
$$

Second, it is obvious that (2.9) satisfies $0 \leq \left[Y - d(t)\right]^+ \leq Y$, i.e. $I_t(Y)$ satisfies the condition 2 in Definite 2.4.1,.

Third, we verify the condition 3 in Definite 2.4.1. The solution to (A.14) is given by

$$\begin{aligned} Z_t^* &= e^{\int_0^t \boldsymbol{B_s} ds} Z_0 + e^{\int_0^t \boldsymbol{B_s} ds} \int_0^t e^{-\int_0^s \boldsymbol{B_u} du} \boldsymbol{f_1}(s) ds \\ & \quad - e^{\int_0^t \boldsymbol{B_s} ds} \int_0^t e^{-\int_0^s \boldsymbol{B_u} du} \boldsymbol{f_2}(s, Y) d\boldsymbol{\bar{N}_s}. \end{aligned} \tag{A.17}$$

Before moving forward, we prove the norm boundedness of the matrix exponentials $e^{\int_0^t \boldsymbol{B_s} ds}$ and $e^{-\int_0^t \boldsymbol{B_s} ds}$.

Lemma A.2.1. *$||e^{\int_0^t \boldsymbol{B_s} ds}||_2$ and $||e^{-\int_0^t \boldsymbol{B_s} ds}||_2$ are uniformly bounded for any $t \in [0, T]$*

Proof. First, we extend the continual domain of $a(t)$ and $b(t)$. From Proposition 2.4.1, we have $J(T) = 1$. By the continuity of function $J(t)$, there exists a small $\epsilon > 0$ such that $J(T + \epsilon) > 0$. Then, $a(t) = 1/J(t)$ and $b(t) = G(t)/J(t)$ exist and are continuous for $t \in (-1, T + \epsilon)$, so does $\boldsymbol{B_t}$.

Therefore, due to the Theorem 2.4 in Chicone (2006), we have $||e^{\int_0^t \boldsymbol{B_s} ds}||_2$ and $||e^{-\int_0^t \boldsymbol{B_s} ds}||_2$ are uniformly bounded for any $t \in [0, T]$. □

Then, for any bounded constants C_1, C_2, we have

$$\begin{aligned} & E[\exp\{C_1 X_t^* + C_2 h_t^* + C_3 c_t^*\}] \\ = \; & E\left[\exp\left\{(C_1 + C_3 a(t), C_2 + C_3[b(t) + 1]) Z_t^* + g(t) - \frac{1}{\gamma} \ln[a(t) - \alpha b(t)]\right\}\right] \\ \leq \; & C_4 E[\exp\{(C_5, C_6) Z_t^*\}] \\ \leq \; & C_4 E\left[\exp\left\{\left|(C_5, C_6)\left(e^{\int_0^t \boldsymbol{B_s} ds} Z_0 + e^{\int_0^t \boldsymbol{B_s} ds} \int_0^t e^{-\int_0^s \boldsymbol{B_u} du} \boldsymbol{f_1}(s) ds\right)\right| \right.\right. \\ & \left.\left. + \left|(C_5, C_6) e^{\int_0^t \boldsymbol{B_s} ds} \int_0^t e^{-\int_0^s \boldsymbol{B_u} du} \boldsymbol{f_2}(s, Y) d\boldsymbol{\bar{N}_s}\right|\right\}\right] \\ \leq \; & C_4 E\left[\exp\left\{||(C_5, C_6)||_2 \left(||e^{\int_0^t \boldsymbol{B_s} ds}||_2 ||Z_0||_2 + ||e^{\int_0^t \boldsymbol{B_s} ds}||_2 \int_0^t ||e^{-\int_0^s \boldsymbol{B_u} du}||_2 \right.\right.\right. \\ & \left.\left.\left. ||\boldsymbol{f_1}(s)||_2 ds\right) + \left|(C_5, C_6) e^{\int_0^t \boldsymbol{B_s} ds} \int_0^t e^{-\int_0^s \boldsymbol{B_u} du} \boldsymbol{f_2}(s, Y) d\boldsymbol{\bar{N}_s}\right|\right\}\right] \end{aligned}$$

$$
\begin{aligned}
&\leq C_7 E\left[\exp\left\{\left|(C_5, C_6)e^{\int_0^t \boldsymbol{B_s} ds}\int_0^t e^{-\int_0^s \boldsymbol{B_u} du}\boldsymbol{f_2}(s,Y)d\bar{\boldsymbol{N}}_{\boldsymbol{s}}\right|\right\}\right] \\
&= C_7 E\left[\exp\left\{\left|(C_5, C_6)e^{\int_0^t \boldsymbol{B_s} ds}\left[\sum_{i=1}^{N_t} e^{-\int_0^{\tau_i} \boldsymbol{B_u} du}\boldsymbol{f_2}(\tau_i,Y)\begin{pmatrix}1\\0\end{pmatrix}\right.\right.\right.\right. \\
&\qquad \left.\left.\left.\left.\sum_{j=1}^{\widetilde{N}_t} e^{-\int_0^{\widetilde{\tau}_j} \boldsymbol{B_u} du}\boldsymbol{f_2}(\widetilde{\tau}_j,Y)\begin{pmatrix}0\\1\end{pmatrix}\right]\right|\right\}\right] \\
&= C_7 E\left[\exp\left\{\left|(C_5, C_6)e^{\int_0^t \boldsymbol{B_s} ds}\sum_{i=1}^{N_t} e^{-\int_0^{\tau_i} \boldsymbol{B_u} du}\boldsymbol{f_2}(\tau_i,Y)\begin{pmatrix}1\\0\end{pmatrix}\right|\right\}\right] \\
&\leq C_7 E\left[\exp\left\{||(C_5, C_6)||_2||e^{\int_0^t \boldsymbol{B_s} ds}||_2\sum_{i=1}^{N_t} ||e^{-\int_0^{\tau_i} \boldsymbol{B_u} du}||_2||\boldsymbol{f_2}(\tau_i,Y)||_2\right\}\right] \\
&\leq C_7 E\left[\exp\left\{C_8\sum_{i=1}^{N_t}\min(Y, d(\tau_i))\right\}\right] \\
&\leq C_7 E\left[\exp\left\{C_8 N_t d(T)\right\}\right] \\
&\leq C_7 E\left[\exp\left\{C_9 N_t\right\}\right] < \infty,
\end{aligned}
$$

where τ_i and $\widetilde{\tau}_i$ are the i-th jump times for N_t and $\widetilde{N}_t$, respectively.

The first equality follows from the substitution of (2.10). The first inequality is due to the continuity of $a(t)$, $b(t)$, and $g(t)$ in $[0,T]$. The second inequality follows from the substitution of (A.17). The third inequality is based on the Cauchy-Schwarz inequality and Minkowski inequality. The forth inequality follows the Lemma A.2.1 and the continuity of $\boldsymbol{f_1}(t)$ in $[0,T]$. The fifth inequality is due to the Cauchy-Schwarz inequality and the Minkowski inequality. The sixth inequality is due to the definition of $f_2(t)$ and the induced matrix norm in Euclidean space. The seventh and eighth inequalities are based on the continuity of $d(t)$ in $[0,T]$. The final finite inequality is based on the existence of moment generating function of Poisson process at each time t.

Forth, we prove the condition 4 in Definite 2.4.1.

$$
\begin{aligned}
& E[(X_t^*)^2 + (h_t^*)^2 + (c_t^*)^2] \\
= \; & E\left[||Z_t^*||_2^2 + \left|(a(t), b(t)+1)Z_t^* + g(t) - \frac{1}{\gamma}\ln[a(t) - \alpha b(t)]\right|^2\right] \\
\leq \; & E[||Z_t^*||_2^2] + C_9 E[1 + |(a(t), b(t)+1)Z_t^*|^2]
\end{aligned}
$$

$$\leq \quad E[||Z_t^*||_2^2] + C_9 E[1 + ||(a(t), b(t) + 1)||_2^2 ||Z_t^*||_2^2] < \infty.$$

The first equality is follows from substitution of (2.10) into. The first inequality is due to the continuity of $g(t)$ and $\ln[a(t) - \alpha b(t)]$ in $[0, T]$ and the inequality $(C_{10} + x)^2 \leq C_{11}(1 + x^2)$ for some constant C_{11}. The last finite inequality is based on (A.16).

Finally, we claim that the optimal strategy (c^*, I^*) belongs to the admissible set $\mathcal{A}$.

Step 2: Verify $V(t, X_t, h_t) \leq \phi(t, X_t, h_t)$ for any $(c, I) \in \mathcal{A}$.

Define the generator

$$\begin{aligned} \mathcal{L}^{c,I}\phi(t,x,h) &= \phi_t - (\lambda + \delta)\phi + rx\phi_x - \beta h \phi_h - c\phi_x + \alpha c \phi_h \\ &\quad -\lambda(1+\theta)E[I_t(Y)]\phi_x + \lambda E[\phi(t, x - Y + I_t(Y), h)]. \end{aligned} \quad \text{(A.18)}$$

We first prove the properties for $\phi(t, X_t, h_t)$.

Lemma A.2.2. *$\phi(t, X_t, h_t)$ has the following properties for any $(c, I) \in \mathcal{A}$*

(a)

$$\mathcal{L}^{c,I}\phi(t,x,h) + U_1(c,h) \leq 0, \quad \text{(A.19)}$$

for all $(t, x, h) \in [0, T] \times \mathbb{R}^2$ and $(c, I) \in \mathcal{A}$,

(b)

$$\phi(T, x, h) \geq U_2(x) \quad \text{for } (x, h) \in \mathbb{R}^2, \quad \text{(A.20)}$$

(c)

$$\begin{aligned} &E\left[e^{-\delta T}|\phi(T, X_T, h_T)| + \int_0^T e^{-\delta t}|\mathcal{L}^{c,I}\phi(t, x_t, h_t)|dt\right] \\ &+ \int_0^T \int_0^\infty e^{-\delta t}|\phi(t, X_t - Y + I_t(Y), h_t) - \phi(t, X_t, h_t)|^2 \nu(dy)dt < \infty, \end{aligned}$$

for all $(c, I) \in \mathcal{A}$, where the Levy measure $\nu(dy) = \lambda \mu_Y(dy)$ for compound Poisson process and the distribution measure $\mu_Y(U) = P\{Y \in U\}$ or each Borel subset U of $\mathbb{R}$.

Proof. Properties (a) and (b) are the direct results of the fact that $\phi(t, x, h)$ satisfies (2.4) with the boundary condition.

With regard to property (c), we divide the proof into three terms. For the first term,

$$
\begin{aligned}
E[e^{-\delta T}|\phi(T, X_T, h_T)|] &= E\left[e^{-\delta T}\left|-\frac{1}{\gamma}e^{-\gamma(a(t)X_t+b(t)h_t+g(t))}\right|\right] \\
&\le C_0 E[e^{C_1 X_t + C_2 h_t}],
\end{aligned}
$$

where C_0, C_1, and C_2 are some bounded constants. The first inequality is based on the continuity of $a(t)$, $b(t)$, and $g(t)$ in $[0, T]$. For the second term,

$$
E\left[\int_0^T e^{-\delta t}|\mathcal{L}^{c,I}\phi(t, x_t, h_t)|dt\right] = \int_0^T e^{-\delta t} E[|\mathcal{L}^{c,I}\phi(t, x_t, h_t)|]dt
$$

$$
\begin{aligned}
&= \int_0^T e^{-\delta t} E\left[\left|e^{-\gamma[a(t)X_t+b(t)h_t+g(t)]} f_3(t, X_t, h_t)\right.\right. \\
&\quad \left.\left.-E\left[\frac{\lambda}{\gamma}e^{-\gamma\{a(t)[X_t-Y+I_t(Y)]+b(t)h_t+g(t)\}}\right]\right|\right] dt \\
&\le \int_0^T e^{-\delta t}\left\{E\left[e^{-\gamma[a(t)X_t+b(t)h_t+g(t)]}|f_3(t, X_t, h_t)|\right]\right. \\
&\quad \left.+E\left[\frac{\lambda}{\gamma}e^{-\gamma\{a(t)[X_t-Y+I_t(Y)]+b(t)h_t+g(t)\}}\right]\right\} dt \\
&\le \int_0^T e^{-\delta t}\left\{E\left[e^{-2\gamma[a(t)X_t+b(t)h_t+g(t)]}\right] E\left[f_3^2(t, X_t, h_t)\right]\right. \\
&\quad \left.+\frac{\lambda}{\gamma}E\left[e^{-2\gamma\{a(t)X_t+b(t)h_t+g(t)\}}\right] E\left[e^{2\gamma a(t)[Y-I_t(Y)]}\right]\right\} dt \\
&\le \int_0^T e^{-\delta t}\left\{C_0 E\left[e^{C_1 X_t + C_2 h_t}\right] E\left[f_3^2(t, X_t, h_t)\right]\right. \\
&\quad \left.+\frac{\lambda}{\gamma}C_0 E\left[e^{C_1 X_t + C_2 h_t}\right] E\left[e^{C_3 \min(Y, d(t))}\right]\right\} dt \\
&\le \int_0^T e^{-\delta t}\left\{C_4 E\left[f_3^2(t, X_t, h_t)\right] + C_5\right\} dt \\
&\le \int_0^T e^{-\delta t}\left\{C_4 E\left[[a'(t) + ra(t)]^2 X_t^2 + [b'(t) - \beta b(t)]^2 h_t^2\right.\right. \\
&\quad \left.\left.+[-a(t) + \alpha b(t)]^2 c_t^2 + C_6^2\right] + C_5\right\} dt \\
&\le \int_0^T e^{-\delta t}\left\{C_7 E\left[X_t^2 + h_t^2 + c_t^2\right] + C_4 C_6^2 + C_5\right\} dt \\
&< \infty,
\end{aligned}
$$

where $f_3(t, X_t, h_t)$ is give by

$$\begin{aligned} f_3(t, X_t, h_t) \quad := \quad & [a'(t) + ra(t)]X_t + [b'(t) - \beta b(t)]h_t + [\alpha b(t) - a(t)]c_t \\ & + g'(t) + \frac{\lambda + \delta}{\gamma} - \lambda(1+\theta)E[I_t(Y)]a(t). \end{aligned}$$

The second inequality is due to the Cauchy-Schwarz inequality. The third inequality is owing to the continuity of $a(t)$, $b(t)$, and $g(t)$ in $[0, T]$ and the condition 2 in in Definition 2.4.1. The forth inequality is based on the condition 3 in Definition 2.4.1 and the continuity of $d(t)$. The fifth inequality is the result of the inequality that $(a_1 + a_2 + a_3 + a_4)^2 \leq 4(a_1^2 + a_2^2 + a_3^2 + a_4^2)$ for any $a_1, a_2, a_3, a_4 \in \mathbb{R}$. The sixth inequality comes from the fact that $a(t)$ and $b(t)$ are the C^1 functions in $[0, T]$ (see Lemma A.1.1). The last finite inequality is determined by the condition 4 in Definition 2.4.1.

For the third term, we have

$$\begin{aligned} & \int_0^T \int_0^\infty e^{-\delta t} |\phi(t, X_t - Y + I_t(Y), h_t) - \phi(t, X_t, h_t)|^2 \nu(dy) dt \\ = \quad & \int_0^T \int_0^\infty e^{-\delta t} |\phi(t, X_t - Y + I_t(Y), h_t) - \phi(t, X_t, h_t)|^2 \lambda \mu_Y(dy) dt \\ \leq \quad & \int_0^T C_0 E[e^{-2\gamma[a(t)(X_t - Y + I_t(Y)) + b(t)h_t + g(t)]}] dt \\ \leq \quad & \int_0^T C_0 E[e^{-4\gamma[a(t)X_t + b(t)h_t + g(t)]}] E[e^{4\gamma(Y - I_t(Y))}] dt \\ \leq \quad & \int_0^T C_1 E[e^{C_2 X_t + C_3 h_t}] E[e^{4\gamma \min(Y, d(t))}] dt \\ < \quad & \infty. \end{aligned}$$

The first equality is due to the definition of Levy measure for the compound Poisson process. The first inequality is based on the fact that $\phi(t, X_t - Y + I_t(Y), h_t) \leq \phi(t, X_t, h_t) \leq 0$, then $|\phi(t, X_t - Y + I_t(Y), h_t) - \phi(t, X_t, h_t)| < |\phi(t, X_t - Y + I_t(Y), h_t)|$. The second inequality holds because of the Cauchy-Schwarz inequality. The third inequality is based on the continuity of $a(t)$, $b(t)$, and $g(t)$ in $[0, T]$ and the condition 2 in in Definition 2.4.1. The last finite inequality holds by the condition 3 in Definition 2.4.1 and the continuity of $d(t)$.

This completes the proof of Lemma A.2.2. □

Moreover, for any $(c, I) \in \mathcal{A}$, we have the following property

Lemma A.2.3.

$$E\left[\int_0^T e^{-\delta u}|U_1(c_u,h_u)|du + e^{-\delta T}|U_2(X_T)|\right] < \infty, \ \textit{for } \forall (c,I) \in \mathcal{A}.$$

Proof.

$$\begin{aligned}
& E\left[\int_0^T e^{-\delta u}|U_1(c_u,h_u)|du + e^{-\delta T}|U_2(X_T)|\right] \\
= & \int_0^T e^{-\delta u}E[|U_1(c_u,h_u)|]du + e^{-\delta T}E[|U_2(X_T)|] \\
= & \int_0^T e^{-\delta u}\frac{1}{\gamma}E[\exp\{-\gamma c_u + \gamma h_u\}]du + e^{-\delta T}\frac{\omega}{\gamma}E[\exp\{-\gamma X_T\}] \\
< & \ \infty
\end{aligned}$$

The last finite inequality comes from the condition 3 in Definition 2.4.1. □

Let $(c, I) \in \mathcal{A}$, then by the property (c) in Lemma A.2.2, we can use Dynkin formula (see Theorem 1.24 in Øksendal and Sulem (2007))

$$E_t[e^{-\delta(T-t)}\phi(T,X_T,h_T)] = \phi(t,x,h) + E_t\left[\int_t^T e^{-\delta(u-t)}\mathcal{L}^{c,I}\phi(u,X_u,h_u)du\right],$$

where $E_t[\cdot]$ denotes the conditional expectation $E[\cdot|X_t = x, h_t = h]$. Together with the inequalities (A.19) and (A.20), we have

$$\begin{aligned}
\phi(t,x,h) \ \geq & \ E_t\left[\int_t^T e^{-\delta(u-t)}U_1(c_u,h_u)du\right] + E_t[e^{-\delta(T-t)}\phi(T,X_T,h_T)] \\
\geq & \ E_t\left[\int_t^T e^{-\delta(u-t)}U_1(c_u,h_u)du + e^{-\delta(T-t)}U_2(X_T)\right]. \qquad \text{(A.21)}
\end{aligned}$$

The finiteness of (A.21)'s R.H.S. is given by Lemma A.2.3. Taking the supreme w.r.t (c, I) of the right hand side of the inequality, we have

$$\phi(t,x,h) \geq V(t,x,h) \ \text{ for all } (t,x,h) \in [0,T]\times\mathbb{R}^2. \qquad \text{(A.22)}$$

This completes the proof of Step 2.

Step 3: Verify $V(t, X_t, h_t) = \phi(t, X_t, h_t)$ for the optimal strategy (c^*, I^*).

Under the strategy (c^*, I^*), we have a direct result from Proposition 2.4.1

$$\mathcal{L}^{c^*,I^*}\phi(t,x,h) + U_1(c,h) = 0. \tag{A.23}$$

Then, (A.21) becomes

$$\phi(t,x,h) = E_t\left[\int_t^T e^{-\delta(u-t)}U_1(c_u^*, h_u^*)du + e^{-\delta(T-t)}U_2(X_T^*)\right] \le V(t,x,h) \tag{A.24}$$

for all $(t, x, h) \in [0, T] \times \mathbb{R}^2$. Combining (A.22) and (A.24), we arrive at (2.12). Moreover, (c^*, I^*) given by (2.10) and (2.9) is the optimal consumption and insurance strategy.

□

A.3 Proof of Corollary 2.4.1

The first-order derivative of $J(t)$ is

$$J'(t) = \left(-\frac{r+\beta}{r+\beta-\alpha} - \frac{\alpha r}{(r+\beta-\alpha)(\beta-\alpha)} + r\right)e^{-r(T-t)} + \frac{\alpha}{\beta-\alpha}e^{-(r+\beta-\alpha)(T-t)}.$$

$J'(t) < 0$ is equivalent to

$$-\frac{r+\beta}{r+\beta-\alpha} - \frac{\alpha r}{(r+\beta-\alpha)(\beta-\alpha)} + r + \frac{\alpha}{\beta-\alpha}e^{-(\beta-\alpha)(T-t)} < 0.$$

Because $\beta > \alpha$ and $0 < r < 1$, we have

$$(\beta-\alpha)(r-1)(r+\beta-\alpha) < 0,$$

or equivalently

$$-\frac{r+\beta}{r+\beta-\alpha} - \frac{\alpha r}{(r+\beta-\alpha)(\beta-\alpha)} + r + \frac{\alpha}{\beta-\alpha} < 0.$$

We then have

$$-\frac{r+\beta}{r+\beta-\alpha} - \frac{\alpha r}{(r+\beta-\alpha)(\beta-\alpha)} + r + \frac{\alpha}{\beta-\alpha}e^{-(\beta-\alpha)(T-t)} < 0,$$

and consequently $J'(t) < 0$ for any $t \in [0, T]$.

A.4 Proof of Propostion 2.4.2

Setting $\alpha = \beta = h_0 = 0$ in (2.4), we have the HJB equation for the "no habit" agent

$$\begin{aligned}&\widetilde{V}_t - (\lambda+\delta)\widetilde{V} + rx\widetilde{V}_x + \sup_c\{U_1(c,0) - c\widetilde{V}_x\}\\&\quad + \sup_I\{-\lambda(1+\theta)E[I_t(Y)]\widetilde{V}_x + \lambda E[\widetilde{V}(t, x - Y + I_t(Y))]\} = 0,\end{aligned} \tag{A.25}$$

with the boundary condition

$$\widetilde{V}(T,x) = U_2(x) = -\frac{\omega}{\gamma}e^{-\gamma x}.$$

Note that $\alpha = \beta = h_0 = 0$ implies that $h_t = 0$ for any $t \in [0,T]$, and thus the "no habit" agent's value function does not depend on h.

We make the ansatz

$$\widetilde{V}(t,x) = -\frac{1}{\gamma}e^{-\gamma(\widetilde{a}(t)x + \widetilde{g}(t))}. \tag{A.26}$$

We immediately have

$$\widetilde{a}(T) = 1, \quad \widetilde{b}(T) = 0, \quad \widetilde{g}(T) = -\frac{\ln(w)}{\gamma}.$$

The first-order condition with respect to c gives

$$\frac{\partial U_1(\widetilde{c},0)}{\partial c} - \widetilde{V}_x = 0. \tag{A.27}$$

Plugging (A.26) into (A.27), we arrive at the optimal consumption

$$\widetilde{c}_t = -\frac{\ln(\widetilde{a}(t))}{\gamma} + \widetilde{a}(t)\widetilde{X}_t + \widetilde{g}(t). \tag{A.28}$$

The first-order condition with respect to I implies that

$$(1+\theta)\widetilde{V}_x(t,x) = \widetilde{V}_x(t, x - Y + I(Y)).$$

Because it is required that $0 \le I(Y) \le Y$, the optimal indemnity follows

$$\widetilde{I}_t(Y) = [Y - (x - \widetilde{V}_x^{-1}(t, (1+\theta)\widetilde{V}_x(t,x)))]^+, \tag{A.29}$$

where $\widetilde{V}_x^{-1}(t,y)$ is the inverse function of $\widetilde{V}_x(t,x)$ w.r.t. x.

Plugging (A.26) into (A.29), we arrive at the optimal indemnity

$$\widetilde{I}_t(Y) = \left[Y - \frac{\ln(1+\theta)}{\gamma\widetilde{a}(t)}\right]^+. \tag{A.30}$$

Substituting (A.26), (A.28) and (A.30) into (A.25), we have

$$\begin{aligned}
0 \;=\;& x[(\widetilde{a}(t))' + r\widetilde{a}(t) - (\widetilde{a}(t))^2] \\
&+(\widetilde{g}(t))' - \widetilde{a}(t)\widetilde{g}(t) + \frac{\lambda+\delta}{\gamma} + \{\ln[\widetilde{a}(t)] - 1\}\frac{\widetilde{a}(t)}{\gamma} \\
&-\lambda(1+\theta)\left[\int_{\frac{\ln(1+\theta)}{\gamma\widetilde{a}(t)}}^{\infty} yf(y)dy - \frac{\ln(1+\theta)}{\gamma\widetilde{a}(t)}\overline{F}\left(\frac{\ln(1+\theta)}{\gamma\widetilde{a}(t)}\right)\right]\widetilde{a}(t) \\
&-\frac{\lambda}{\gamma}\left[(1+\theta)\overline{F}\Big(\frac{\ln(1+\theta)}{\gamma\widetilde{a}(t)}\Big) + \int_0^{\frac{\ln(1+\theta)}{\gamma\widetilde{a}(t)}} e^{\gamma\widetilde{a}(t)y} f(y)dy\right].
\end{aligned}$$

By separation of variables, we have the following ODE system

$$(\widetilde{a}(t))' + r\widetilde{a}(t) - (\widetilde{a}(t))^2 = 0, \quad \widetilde{a}(T) = 1, \tag{A.31}$$

$$\begin{aligned}
&(\widetilde{g}(t))' - \widetilde{a}(t)\widetilde{g}(t) + \frac{\lambda+\delta}{\gamma} + \{\ln[\widetilde{a}(t)] - 1\}\frac{\widetilde{a}(t)}{\gamma} \\
&-\lambda(1+\theta)\left[\int_{\frac{\ln(1+\theta)}{\gamma\widetilde{a}(t)}}^{\infty} yf(y)dy - \frac{\ln(1+\theta)}{\gamma\widetilde{a}(t)}\overline{F}\left(\frac{\ln(1+\theta)}{\gamma\widetilde{a}(t)}\right)\right]\widetilde{a}(t) \\
&-\frac{\lambda}{\gamma}\left[(1+\theta)\overline{F}\Big(\frac{\ln(1+\theta)}{\gamma\widetilde{a}(t)}\Big) + \int_0^{\frac{\ln(1+\theta)}{\gamma\widetilde{a}(t)}} e^{\gamma\widetilde{a}(t)y} f(y)dy\right] = 0, \quad \widetilde{g}(T) = -\frac{\ln(w)}{\gamma}
\end{aligned} \tag{A.32}$$

Assuming $\widetilde{a}(t) = 1/\widetilde{J}(t)$ and substituting it into (A.31), we have

$$(\widetilde{J}(t))' = r\widetilde{J}(t) - 1,$$

and

$$\widetilde{J}(t) = e^{-r(T-t)}\frac{r-1}{r} + \frac{1}{r}.$$

For the ODE (A.32), we have

$$
\begin{aligned}
& d(e^{\int_t^T \widetilde{a}(s)ds}\widetilde{g}(t)) \\
= \; & -e^{\int_t^T \widetilde{a}(s)ds}\left\{\frac{\lambda+\delta}{\gamma} + \{\ln[\widetilde{a}(t)] - 1\}\frac{\widetilde{a}(t)}{\gamma}\right. \\
& -\lambda(1+\theta)\left[\int_{\frac{\ln(1+\theta)}{\gamma\widetilde{a}(t)}}^{\infty} yf(y)dy - \frac{\ln(1+\theta)}{\gamma\widetilde{a}(t)}\overline{F}\left(\frac{\ln(1+\theta)}{\gamma\widetilde{a}(t)}\right)\right]\widetilde{a}(t) \\
& \left. -\frac{\lambda}{\gamma}\left[(1+\theta)\overline{F}\Big(\frac{\ln(1+\theta)}{\gamma\widetilde{a}(t)}\Big) + \int_0^{\frac{\ln(1+\theta)}{\gamma\widetilde{a}(t)}} e^{\gamma\widetilde{a}(t)y} f(y)dy\right]\right\},
\end{aligned}
$$

whose solution is given by (2.13).

A.5 Proof of Corollary 2.4.2

The first claim follows from

$$(\widetilde{d}(t))' = \frac{\ln(1+\theta)}{\gamma}e^{-r(T-t)}(r-1) < 0,$$

for all $t \in [0, T]$.

For the second claim, we want to compare $\widetilde{d}(t)$ with $d(t)$. At initial time, we have

$$
\begin{aligned}
d(0) - \widetilde{d}(0) &= \frac{\ln(1+\theta)}{\gamma}(J(0) - \widetilde{J}(0)) \\
&= \frac{\ln(1+\theta)}{\gamma}\left\{\frac{r+\beta}{(r+\beta-\alpha)r} + \left(-\frac{r+\beta}{(r+\beta-\alpha)r} - \frac{\alpha}{(r+\beta-\alpha)(\beta-\alpha)}\right.\right. \\
&\quad \left.\left. +1\right)e^{-rT} + \frac{\alpha}{(r+\beta-\alpha)(\beta-\alpha)}e^{-(r+\beta-\alpha)T} - \frac{r-1}{r}e^{-rT} - \frac{1}{r}\right\} \\
&= \frac{\ln(1+\theta)}{\gamma}\left\{\frac{\alpha}{(r+\beta-\alpha)r} - \frac{\alpha}{r(\beta-\alpha)}e^{-rT} + \frac{\alpha}{(r+\beta-\alpha)(\beta-\alpha)}\right. \\
&\quad \left. e^{-(r+\beta-\alpha)T}\right\}
\end{aligned}
$$

To prove $d(0) - \widetilde{d}(0) > 0$, it is equivalent to show

$$
\begin{aligned}
\frac{\alpha}{(r+\beta-\alpha)r} - \frac{\alpha}{r(\beta-\alpha)}e^{-rT} + \frac{\alpha}{(r+\beta-\alpha)(\beta-\alpha)}e^{-(r+\beta-\alpha)T} > 0, \\
\beta - \alpha - (r+\beta-\alpha)e^{-rT} + re^{-(r+\beta-\alpha)T} > 0,
\end{aligned}
$$

where we have used the fact that $\beta > \alpha \geq 0$ and $0 < r < 1$.

Define $f_r(r) := \beta - \alpha - (r + \beta - \alpha)e^{-rT} + re^{-(r+\beta-\alpha)T}$. It is easy to see

$$\begin{aligned}
f_r(0) &= \beta - \alpha - (\beta - \alpha) = 0 \\
f_r'(r) &= -e^{-rT} + (r + \beta - \alpha)Te^{-rT} + e^{-(r+\beta-\alpha)T} - rTe^{-(r+\beta-\alpha)T} \\
&= [(r + \beta - \alpha)T - 1]e^{-rT} + (1 - rT)e^{-(r+\beta-\alpha)T} \\
&= e^{-rT}[(r + \beta - \alpha)T - 1 + (1 - rT)e^{-(\beta-\alpha)T}] \\
&= e^{-rT}[(\beta - \alpha)T - 1 + e^{-(\beta-\alpha)T} + rT(1 - e^{-(\beta-\alpha)T})] \\
&> e^{-rT}rT(1 - e^{-(\beta-\alpha)T}) \\
&> 0,
\end{aligned}$$

where we have used the inequality $e^{-x} > 1 - x$ for $x > 0$. Thus, we have $f_r(r) > 0$ for $r \in [0, 1)$. i.e. $d(0) - \widetilde{d}(0) > 0$ for $\beta > \alpha \geq 0$ and $0 < r < 1$.

Moreover, at the terminal time, we have

$$\begin{aligned}
d(T) - \widetilde{d}(T) &= \frac{\ln(1 + \theta)}{\gamma}(J(T) - J_0(T)) \\
&= \frac{\ln(1 + \theta)}{\gamma}(1 - 1) \\
&= 0.
\end{aligned}$$

We now compare the first order derivatives. We have

$$\begin{aligned}
& d'(t) - (\widetilde{d}(t))' \\
&= \frac{\ln(1 + \theta)}{\gamma}(J'(t) - (\widetilde{J}(t))') \\
&= \frac{\ln(1 + \theta)}{\gamma}\Bigg\{\Bigg(- \frac{r + \beta}{r + \beta - \alpha} - \frac{\alpha r}{(r + \beta - \alpha)(\beta - \alpha)} + r\Bigg)e^{-r(T-t)} \\
&\quad + \frac{\alpha}{\beta - \alpha}e^{-(r+\beta-\alpha)(T-t)} - (r - 1)e^{-r(T-t)}\Bigg\} \\
&= \frac{\ln(1 + \theta)}{\gamma}\Bigg\{\Bigg(- \frac{\alpha}{r + \beta - \alpha} - \frac{\alpha r}{(r + \beta - \alpha)(\beta - \alpha)}\Bigg)e^{-r(T-t)} + \frac{\alpha}{\beta - \alpha}e^{-(r+\beta-\alpha)(T-t)}\Bigg\}.
\end{aligned}$$

To show $d'(t) - (\widetilde{d}(t))' < 0$, it is equivalent to prove

$$-\alpha(\beta - \alpha) - \alpha r + \alpha(r + \beta - \alpha)e^{-(\beta-\alpha)(T-t)} < 0,$$

or

$$\alpha(r+\beta-\alpha)(e^{-(\beta-\alpha)(T-t)}-1)<0,$$

which is indeed satisfied for $t \in [0, T)$.

Therefore, $d'(t) - (\widetilde{d}(t))' < 0$ for $t \in [0, T)$, and $d'(T) - (\widetilde{d}(T))' = 0$. The rest of the proof follows from $d(0) - \widetilde{d}(0) > 0$ and $d(T) - \widetilde{d}(T) = 0$.

A.6 Proof of Proposition 2.4.3

Substituting $I(Y) = pY$ into (2.4), we have

$$\begin{aligned} &V_t^p - (\lambda+\delta)V^p + rxV_x^p - \beta h V_h^p + \sup_c\{U_1(c,h) - cV_x^p + \alpha c V_h^p\} \\ &\quad + \sup_p\{-\lambda(1+\theta)pE[Y]V_x^p + \lambda E[V^p(t, x-(1-p)Y, h)]\} = 0, \end{aligned} \tag{A.33}$$

with the boundary condition

$$V^p(T,x,h) = U_2(x) = -\frac{\omega}{\gamma}e^{-\gamma x}.$$

We make the ansatz

$$V^p(t,x,h) = -\frac{1}{\gamma}e^{-\gamma(a^p(t)x + b^p(t)h + g^p(t))}. \tag{A.34}$$

We immediately have

$$a^p(T) = 1, \quad b^p(T) = 0, \quad g^p(T) = -\frac{\ln(w)}{\gamma}.$$

The first-order condition with respect to c implies that

$$\frac{\partial U_1(c,h)}{\partial c} - V_x^p + \alpha V_h^p = 0. \tag{A.35}$$

Plugging (A.34) into (A.35), we have

$$c_t^p = -\frac{1}{\gamma}\ln[a^p(t) - \alpha b^p(t)] + a^p(t)X_t^p + (b^p(t)+1)h_t^p + g^p(t). \tag{A.36}$$

Substituting (A.34) into (A.33), we have the following optimization problem for p

$$\sup_p \left\{ -\lambda(1+\theta)pE[Y]e^{-\gamma(a^p(t)x+b^p(t)h+g^p(t))}a^p(t) - \frac{\lambda}{\gamma}E[e^{-\gamma(a^p(t)(x-(1-p)Y)+b^p(t)h+g^p(t))}] \right\},$$

which is equivalent to

$$\sup_p \left\{ -\lambda(1+\theta)pE[Y]a^p(t) - \frac{\lambda}{\gamma}E[e^{\gamma a^p(t)(1-p)Y}] \right\}.$$

Define $f_p(p) = -\lambda(1+\theta)pE[Y]a^p(t) - \frac{\lambda}{\gamma}E[e^{\gamma a^p(t)(1-p)Y}]$. We have

$$\begin{aligned} f_p'(p) &= \lambda a^p(t)\{E[Ye^{\gamma a^p(t)(1-p)Y}] - (1+\theta)E[Y]\}, \\ f_p''(p) &= -\lambda\gamma(a^p(t))^2E[Y^2e^{\gamma a^p(t)(1-p)Y}] < 0, \end{aligned}$$

and

$$\begin{aligned} f_p'(0) &= \lambda a^p(t)\{E[Ye^{\gamma a^p(t)Y}] - (1+\theta)E[Y]\}, \\ f_p'(1) &= -\lambda\theta a^p(t)E[Y] < 0. \end{aligned}$$

Because $f_p''(p) < 0$, $\forall p$, $f_p'(p)$ is a decreasing function. Moreover, $f_p'(1) = \lambda a^p(t)\{E[Y] - (1+\theta)E[Y]\} < 0$ and the monotonicity of $f_p(p)$ only depends on the sign of $f_p'(0)$. If $f_p'(0) > 0$, $f_p'(p)$ decreases from a positive number to a negative number in $[0,1]$, and there exists a unique $p^* \in [0,1]$ such that $f_p'(p^*) = 0$. If $f_p'(0) \leq 0$, $f_p'(p)$ is always negative for $p \in (0,\infty]$ and $f_p(p)$ attains its maximum at $p = 0$. Therefore, if $f_p'(0) > 0$, we have the unique optimal solution $p^*(t)$ satisfying

$$E[e^{\gamma a(t)(1-p^*(t))Y}Y] = (1+\theta)E[Y]. \tag{A.37}$$

Otherwise, we have $p^* = 0$.

Plugging (A.36) and (A.34) into (A.33), we have

$$\begin{aligned} 0 = \; & x[(a^p(t))' + ra^p(t) - (a^p(t) - \alpha b^p(t))a^p(t)] \\ & + h[(b^p(t))' - \beta b^p(t) - (a^p(t) - \alpha b^p(t))(b^p(t)+1)] \\ & + (g^p(t))' - (a^p(t) - \alpha b^p(t))g^p(t) + \frac{\lambda+\delta}{\gamma} + \{\ln[a^p(t) - \alpha b^p(t)] - 1\}\frac{a^p(t) - \alpha b^p(t)}{\gamma} \\ & - \lambda(1+\theta)p^*(t)E[Y]a^p(t) - \frac{\lambda}{\gamma}E[e^{\gamma a^p(t)(1-p^*(t))Y}]. \end{aligned}$$

From separation of variables, we arrive at the following ODE system

$$\begin{aligned}
&(a^p(t))' + ra^p(t) - (a^p(t) - \alpha b^p(t))a^p(t) = 0, \quad a^p(T) = 1,\\
&(b^p(t))' - \beta b^p(t) - (a^p(t) - \alpha b^p(t))(b^p(t) + 1) = 0, \quad b^p(T) = 0,\\
&(g^p(t))' - (a^p(t) - \alpha b^p(t))g^p(t) + \frac{\lambda + \delta}{\gamma} + \{\ln[a^p(t) - \alpha b^p(t)] - 1\}\frac{a^p(t) - \alpha b^p(t)}{\gamma}\\
&-\lambda(1+\theta)p^*(t)E[Y]a^p(t) - \frac{\lambda}{\gamma}E[e^{\gamma a^p(t)(1-p^*(t))Y}], \quad g^p(T) = -\frac{\ln(w)}{\gamma}.
\end{aligned} \tag{A.38}$$

Note that $a^p(t)$ and $b^p(t)$ satisfy the same ODEs as (A.5) and (A.6), we claim that $a^p(t) = a(t)$ and $b^p(t) = b(t)$ for all $t \in [0, T]$. We only need to solve the ODE (A.38) for $g^p(t)$. After simplifying, we have

$$\begin{aligned}
d(e^{\int_t^T a^p(s) - \alpha b^p(s) ds} g^p(t)) = -e^{\int_t^T a^p(s) - \alpha b^p(s) ds} \Bigg\{ &\frac{\lambda + \delta}{\gamma} + \{\ln[a^p(t) - \alpha b^p(t)] - 1\}\frac{a^p(t) - \alpha b^p(t)}{\gamma}\\
&-\lambda(1+\theta)p^*(t)E[Y]a^p(t) - \frac{\lambda}{\gamma}E[e^{\gamma a^p(t)(1-p^*(t))Y}] \Bigg\},
\end{aligned}$$

whose solution is given by (2.14).

A.7 Proof of Corollary 2.4.3

From the Proof of Corollary 2.4.1, $J(t)$ is positive and strictly decreasing and thus $a(t) = 1/J(t)$ is strictly increasing in $[0, T]$. Moreover, $E[e^{\gamma a(t)Y}Y]$ is strictly increasing in $[0, T]$. There are three cases.

1. If $E[e^{\gamma a(0)Y}Y] > (1+\theta)E[Y]$, then $E[e^{\gamma a(t)Y}Y] > (1+\theta)E[Y]$ for any $t \in [0, T]$. Therefore, there is always a unique positive solution $p^*(t)$ to the equation (A.37). Because the right hand side of (A.37) is constant, $a(t)(1-p^*(t))$ is constant. Together with the fact that $a(t)$ is strictly increasing, we have $p^*(t)$ is positive and strictly increasing in $[0, T]$.

2. If $E[e^{\gamma a(0)Y}Y] \leq (1+\theta)E[Y]$ and $E[e^{\gamma a(T)Y}Y] > (1+\theta)E[Y]$, then there exists a unique $t_0 \in [0, T)$ such that $E[e^{\gamma a(t_0)Y}Y] = (1+\theta)E[Y]$. For $t \in [0, t_0]$, $E[e^{\gamma a(t)Y}Y] \leq (1+\theta)E[Y]$ and $p^*(t) = 0$. For $t \in (t_0, T]$, $E[e^{\gamma a(t)Y}Y] > (1+\theta)E[Y]$ and $p^*(t)$ is positive and strictly increasing in $[0, T]$. Moreover, $p^*(t)$ is continuous at time t_0.

3. If $E[e^{\gamma a(T)Y}Y] \leq (1+\theta)E[Y]$, then $E[e^{\gamma a(t)Y}Y] \leq (1+\theta)E[Y]$ and $p^*(t) = 0$ for all $t \in [0, T]$.

A.8 Proof of Proposition 2.4.4

Setting $\alpha = \beta = h_0 = 0$ and $I(Y) = pY$ in the HJB equation (2.4), we have

$$\begin{aligned} &\widetilde{V}_t^p - (\lambda+\delta)\widetilde{V}^p + rx\widetilde{V}_x^p + \sup_c\{U_1(c,0) - c\widetilde{V}_x^p\} \\ &\quad + \sup_p\{-\lambda(1+\theta)pE[Y]\widetilde{V}_x^p + \lambda E[\widetilde{V}^p(t, x-(1-p)Y)]\} = 0, \end{aligned} \tag{A.39}$$

with the boundary condition

$$\widetilde{V}^p(T,x) = U_2(x) = -\frac{\omega}{\gamma}e^{-\gamma x}.$$

We make the ansatz

$$\widetilde{V}^p(t,x) = -\frac{1}{\gamma}e^{-\gamma(\widetilde{a}^p(t)x+\widetilde{g}^p(t))}. \tag{A.40}$$

We immediately have

$$\widetilde{a}^p(T) = 1, \quad \widetilde{b}^p(T) = 0, \quad \widetilde{g}^p(T) = -\frac{\ln(w)}{\gamma}.$$

The first-order condition with respect to c gives

$$\frac{\partial U_1(\widetilde{c}^p, 0)}{\partial c} - \widetilde{V}_x^p = 0. \tag{A.41}$$

Plugging (A.40) into (A.41), we arrive at the optimal consumption

$$\widetilde{c}_t^p = -\frac{\ln(\widetilde{a}^p(t))}{\gamma} + \widetilde{a}^p(t)\widetilde{X}_t^p + \widetilde{g}^p(t). \tag{A.42}$$

Substituting (A.40) into (A.39), we have the following optimization for p

$$\sup_p\Big\{-\lambda(1+\theta)pE[Y]e^{-\gamma(\widetilde{a}^p(t)x+\widetilde{b}^p(t)h+\widetilde{g}^p(t))}\widetilde{a}^p(t) - \frac{\lambda}{\gamma}E[e^{-\gamma(\widetilde{a}^p(t)(x-(1-p)Y)+\widetilde{b}^p(t)h+\widetilde{g}^p(t))}]\Big\},$$

which is equivalent to

$$\sup_p\Big\{-\lambda(1+\theta)pE[Y]\widetilde{a}^p(t) - \frac{\lambda}{\gamma}E[e^{\gamma\widetilde{a}^p(t)(1-p)Y}]\Big\}.$$

Define $f_{\widetilde{p}}(p) = -\lambda(1+\theta)pE[Y]\widetilde{a}^p(t) - \frac{\lambda}{\gamma}E[e^{\gamma\widetilde{a}^p(t)(1-p)Y}]$. Following the similar argument for $f_p(p)$ as in Proposition 2.4.3, we have that if $f'_{\widetilde{p}}(0) > 0$, then the unique solution $\widetilde{p}^*$ satisfies (2.15), and $\widetilde{p}^* = 0$ otherwise.

Plugging (A.42) and(A.40) into (A.39), we have

$$\begin{aligned} 0 &= x[(\widetilde{a}^p(t))' + r\widetilde{a}^p(t) - (\widetilde{a}^p(t))^2] \\ &\quad + (\widetilde{g}^p(t))' - \widetilde{a}^p(t)\widetilde{g}^p(t) + \frac{\lambda+\delta}{\gamma} + \{\ln[\widetilde{a}^p(t)] - 1\}\frac{\widetilde{a}^p(t)}{\gamma} \\ &\quad - \lambda(1+\theta)\widetilde{p}^* E[Y]\widetilde{a}^p(t) - \frac{\lambda}{\gamma}E[e^{\gamma\widetilde{a}^p(t)(1-\widetilde{p}^*)Y}]. \end{aligned}$$

By separation of variables, we have the following ODE system

$$\begin{aligned} &(\widetilde{a}^p(t))' + r\widetilde{a}^p(t) - (\widetilde{a}^p(t))^2 = 0, \quad \widetilde{a}^p(T) = 1, \\ &(\widetilde{g}^p(t))' - \widetilde{a}^p(t)\widetilde{g}^p(t) + \frac{\lambda+\delta}{\gamma} + \{\ln[\widetilde{a}^p(t)] - 1\}\frac{\widetilde{a}^p(t)}{\gamma} \\ &-\lambda(1+\theta)\widetilde{p}^*(t)E[Y]\widetilde{a}^p(t) - \frac{\lambda}{\gamma}E[e^{\gamma\widetilde{a}^p(t)(1-\widetilde{p}^*(t))Y}], \quad \widetilde{g}^p(T) = -\frac{\ln(w)}{\gamma}. \end{aligned} \tag{A.43}$$

Note that $\widetilde{a}^p(t)$ satisfies the same ODE as (A.31), we claim that $\widetilde{a}^p(t) = \widetilde{a}(t)$ for all $t \in [0,T]$. We only need to solve the ODE (A.43) for $\widetilde{g}^p(t)$. We have

$$\begin{aligned} d(e^{\int_t^T \widetilde{a}^p(s)ds}\widetilde{g}^p(t)) &= -e^{\int_t^T \widetilde{a}^p(s)ds}\left\{\frac{\lambda+\delta}{\gamma} + \{\ln[\widetilde{a}^p(t)] - 1\}\frac{\widetilde{a}^p(t)}{\gamma}\right. \\ &\quad \left. -\lambda(1+\theta)\widetilde{p}^*(t)E[Y]\widetilde{a}^p(t) - \frac{\lambda}{\gamma}E[e^{\gamma\widetilde{a}^p(t)(1-\widetilde{p}^*(t))Y}]\right\}, \end{aligned}$$

whose solution is given by (2.16).

A.9 Proof of Corollary 2.4.4

From the proof of Corollary 2.4.2, we know that $\widetilde{a}(t) = 1/\widetilde{J}(t)$ is strictly increasing in $[0,T]$. The proof for the monotonicity of $\widetilde{p}^*(t)$ is similar to that of Corollary 2.4.3.

In the following part, we compare the $\widetilde{p}^*(t)$ with $p^*(t)$. Because $\widetilde{a}(T) = a(T) = 1$, we have $\widetilde{p}^*(T) = p^*(T)$.

For $t \in [0,T)$, because $J(t) > \widetilde{J}(t)$ due to Proposition 2.4.2, we have $a(t) = 1/J(t) < 1/\widetilde{J}(t) = \widetilde{a}(t)$. Therefore, $E[e^{\gamma a(t)Y}Y] < E[e^{\gamma\widetilde{a}(t)Y}Y]$ for any $t \in [0,T]$. There are three cases.

1. If $(1+\theta)E[Y] < E[e^{\gamma a(t)Y}Y] < E[e^{\gamma \widetilde{a}(t)Y}Y]$, then

$$E[e^{\gamma a(t)(1-p^*(t))Y}Y] = (1+\theta)E[Y] = E[e^{\gamma \widetilde{a}(t)(1-\widetilde{p}^*(t))Y}Y].$$

   Because $a(t) < \widetilde{a}(t)$, $0 < p^*(t) < \widetilde{p}^*(t)$.

2. If $E[e^{\gamma a(t)Y}Y] \le (1+\theta)E[Y] < E[e^{\gamma \widetilde{a}(t)Y}Y]$, then $p^*(t) = 0 < \widetilde{p}^*(t)$.

3. If $E[e^{\gamma a(t)Y}Y] < E[e^{\gamma \widetilde{a}(t)Y}Y] \le (1+\theta)E[Y]$, then $p^*(t) = \widetilde{p}^*(t) = 0$.

A.10 Proof of Proposition 2.5.1

From (2.17), we have

$$-\frac{1}{\gamma}e^{-\gamma(a(t)x+b(t)h+g(t))} = -\frac{1}{\gamma}e^{-\gamma(a(t)(x+z_p)+b(t)h+g^p(t))}.$$

Solving for z_p, we have

$$z_p(t) = \frac{g(t)-g^p(t)}{a(t)}.$$

Substituting $g(t)$ and $g^p(t)$, we have $z_p(t)$ follows (2.18).

A.11 Proof of Proposition 2.5.2

Substituting $I(Y) = 0$ into (2.4), we have

$$\begin{aligned} &V_t^0 - (\lambda+\delta)V^0 + rxV_x^0 - \beta h V_h^0 \\ &\quad + \sup_c\{U_1(c,h) - cV_x^0 + \alpha c V_h^0\} + \lambda E[V^0(t,x-Y,h)] = 0, \end{aligned} \tag{A.44}$$

with the boundary condition

$$V^0(T,x,h) = U_2(x) = -\frac{\omega}{\gamma}e^{-\gamma x}.$$

We make the ansatz

$$V^0(t,x,h) = -\frac{1}{\gamma}e^{-\gamma(a^0(t)x+b^0(t)h+g^0(t))}. \tag{A.45}$$

We immediately have

$$a^0(T) = 1, \quad b^0(T) = 0, \quad g^0(T) = -\frac{\ln(w)}{\gamma}.$$

The first-order condition with respect to c gives

$$\frac{\partial U_1(c^0, h)}{\partial c} - V_x^0 + \alpha V_h^0 = 0,$$

and we have

$$c_t^0 = -\frac{1}{\gamma} \ln[a^0(t) - \alpha b^0(t)] + a^0(t) X_t^0 + (b^0(t) + 1) h_t^0 + g^0(t). \tag{A.46}$$

Substituting (A.45) and (A.46) into (A.44), we have

$$\begin{aligned} 0 \;=\;& x[(a^0(t))' + r a^0(t) - (a^0(t) - \alpha b^0(t)) a^0(t)] \\ & + h[(b^0(t))' - \beta b^0(t) - (a^0(t) - \alpha b^0(t))(b^0(t) + 1)] \\ & + (g^0(t))' - (a^0(t) - \alpha b^0(t)) g^0(t) + \frac{\lambda + \delta}{\gamma} + \{\ln[a^0(t) - \alpha b^0(t)] - 1\} \frac{a^0(t) - \alpha b^0(t)}{\gamma} \\ & - \frac{\lambda}{\gamma} \int_0^\infty e^{\gamma a^0(t) y} dF(y). \end{aligned}$$

By separation of variables, we have the following ODE system

$$\begin{aligned} & (a^0(t))' + r a^0(t) - (a^0(t) - \alpha b^0(t)) a^0(t) = 0, \quad a^0(T) = 1, \\ & (b^0(t))' - \beta b^0(t) - (a^0(t) - \alpha b^0(t))(b^0(t) + 1) = 0, \quad b^0(T) = 0, \\ & (g^0(t))' - (a^0(t) - \alpha b^0(t)) g^0(t) + \frac{\lambda + \delta}{\gamma} + \{\ln[a^0(t) - \alpha b^0(t)] - 1\} \frac{a^0(t) - \alpha b^0(t)}{\gamma} \\ & - \frac{\lambda}{\gamma} \int_0^\infty e^{\gamma a^0(t) y} f(y) dy = 0, \quad g^0(T) = -\frac{\ln(w)}{\gamma}. \end{aligned} \tag{A.47}$$

Note that $a^0(t)$ and $b^0(t)$ satisfy the same ODEs as (A.5) and (A.6), we claim that $a^0(t) = a(t)$ and $b^0(t) = b(t)$ for all $t \in [0, T]$. We only need to solve the ODE (A.47) for $g^0(t)$. After simplifying, we have

$$\begin{aligned} & d(e^{\int_t^T a^0(s) - \alpha b^0(s) ds} g^0(t)) \\ = \;& -e^{\int_t^T a^0(s) - \alpha b^0(s) ds} \left\{ \frac{\lambda + \delta}{\gamma} + \{\ln[a^0(t) - \alpha b^0(t)] - 1\} \frac{a^0(t) - \alpha b^0(t)}{\gamma} \right. \\ & \left. - \frac{\lambda}{\gamma} \int_0^\infty e^{\gamma a^0(t) y} f(y) dy \right\}, \end{aligned}$$

whose solution is given by (2.19).

A.12 Proof of Proposition 2.5.3

From (2.20), we have

$$-\frac{1}{\gamma}e^{-\gamma(a(t)x+b(t)h+g(t))} = -\frac{1}{\gamma}e^{-\gamma(a(t)(x+z_0)+b(t)h+g^0(t))}.$$

Solving for z_0, we have

$$z_0(t) = \frac{g(t) - g^0(t)}{a^0(t)}.$$

Substituting $g(t)$ and $g^0(t)$, we have $z_0(t)$ follows (2.21).

Appendix B

Proofs for Chapter 3

B.1 Proof of Proposition 3.3.1

Proof. We substitute the candidate solution $G(t, W_t^{\widetilde{C}}, X_t)$ into HJB equation (3.21) to verify the result. The derivatives of candidate solution are given by

$$\begin{aligned}
\frac{\partial G}{\partial t} &= \frac{\gamma}{1-\gamma}\left(\frac{f_1}{W_t^{\widetilde{C}}}\right)^{\gamma-1}\frac{\partial f_1}{\partial t}, \\
\frac{\partial G}{\partial w^{\widetilde{C}}} &= \left(\frac{f_1}{W_t^{\widetilde{C}}}\right)^{\gamma}, \\
\frac{\partial G}{\partial X^\top} &= \frac{\gamma}{1-\gamma}\left(\frac{f_1}{W_t^{\widetilde{C}}}\right)^{\gamma-1}\frac{\partial f_1}{\partial X^\top}, \\
\frac{\partial^2 G}{(\partial w^{\widetilde{C}})^2} &= -\gamma (W_t^{\widetilde{C}})^{-\gamma-1} f_1^{\gamma}, \\
\frac{\partial^2 G}{\partial w^{\widetilde{C}}\partial X^\top} &= \gamma (W_t^{\widetilde{C}})^{-\gamma} f_1^{\gamma-1}\frac{\partial f_1}{\partial X^\top}, \\
\frac{\partial^2 G}{\partial X^\top \partial X} &= -\gamma (w^{\widetilde{C}})^{1-\gamma} f_1^{\gamma-2}\frac{\partial f_1}{\partial X^\top}\frac{\partial f_1}{\partial X} + \frac{\gamma}{1-\gamma}(w^{\widetilde{C}})^{1-\gamma} f_1^{\gamma-1}\frac{\partial^2 f_1}{\partial X^\top \partial X}.
\end{aligned}$$

Plug these derivatives into the right-hand side of (3.21), we have

$$(W_t^{\widetilde{C}})^{1-\gamma} f_1^{\gamma-1}\left\{\frac{\mu_{x+t}\gamma}{1-\gamma}(1-f_1) + \frac{\gamma}{1-\gamma}\frac{\partial f_1}{\partial t} + (\delta_r + e_1^\top X_t) f_1 - \frac{\gamma}{1-\gamma}\frac{\partial f_1}{\partial X} K_X X_t\right.$$

$$
+\frac{1}{2\gamma}\left[f_1(\Lambda_0^\top + X_t^\top\Lambda_1^\top - \sigma_\Pi^\top)(\Lambda_0 + \Lambda_1 X_t - \sigma_\Pi) + 2\gamma\frac{\partial f_1}{\partial X}\Sigma_X(\Lambda_0 + \Lambda_1 X_t - \sigma_\Pi)\right]
$$

$$
+\frac{1}{2}\frac{\gamma}{1-\gamma}\mathrm{Tr}\left(\Sigma_X^\top\frac{\partial^2 f_1}{\partial X^\top\partial X}\Sigma_X\right)\Bigg\}. \tag{B.1}
$$

After some tedious calculation, we simplify (B.1) to the following form

$$
-\frac{\gamma}{1-\gamma}(W_t^{\widetilde{C}})^{1-\gamma}f_1^{\gamma-1}\quad\Big\{\int_t^T e^{-\int_t^s \mu_{x+u}du}\mu_{x+s}f(X_t,s-t)f_2(X_t,s-t)ds
$$

$$
+e^{-\int_t^T \mu_{x+s}ds}f(X_t,T-t)f_2(X_t,T-t)\Big\}, \tag{B.2}
$$

where

$$
\begin{aligned}
f_2(X_t,\tau) \quad = \quad & \frac{\partial\Gamma_0(\tau)}{\partial\tau} - \frac{1}{2}[\Gamma_1(\tau)]^\top\Sigma_X\Sigma_X^\top\Gamma_1(\tau) - \frac{1-\gamma}{\gamma}[\Gamma_1(\tau)]^\top\Sigma_X(\Lambda_0-\sigma_\Pi) \\
& -\frac{1}{2}\mathrm{Tr}\{\Sigma_X^\top\Gamma_2(\tau)\Sigma_X\} - \frac{1-\gamma}{\gamma}\delta_r - \frac{1-\gamma}{2\gamma^2}(\Lambda_0^\top-\sigma_\Pi^\top)(\Lambda_0-\sigma_\Pi) \\
& +X_t^\top\left\{\frac{\partial\Gamma_1(\tau)}{\partial\tau} - \Gamma_2(\tau)\Sigma_X\Sigma_X^\top\Gamma_1(\tau) - \left[\frac{1-\gamma}{\gamma}\Lambda_1^\top\Sigma_X^\top - K_X^\top\right]\Gamma_1(\tau)\right. \\
& \left.-\frac{1-\gamma}{\gamma}\Gamma_2(\tau)\Sigma_X(\Lambda_0-\sigma_\Pi) - \frac{1-\gamma}{\gamma^2}\Lambda_1^\top(\Lambda_0-\sigma_\Pi) - \frac{1-\gamma}{\gamma}e_1\right\} \\
& +\frac{1}{2}X_t^\top\left\{\frac{\partial\Gamma_2(\tau)}{\partial\tau} - \Gamma_2(\tau)\left[\frac{1-\gamma}{\gamma}\Sigma_X\Lambda_1 - K_X\right]\right. \\
& \left.-\left[\frac{1-\gamma}{\gamma}\Lambda_1^\top\Sigma_X^\top - K_X^\top\right]\Gamma_2(\tau) - \Gamma_2(\tau)\Sigma_X\Sigma_X^\top\Gamma_2(\tau) - \frac{1-\gamma}{\gamma^2}\Lambda_1^\top\Lambda_1\right\}X_t.
\end{aligned}
$$

Substitute ODEs (3.25)-(3.27) into (B.2), we have (B.2) equals zero. Therefore, $G(t, W_t^R, X_t)$ is the candidate solution to HJB (3.21). Finally, plug $G(t, W_t^R, X_t)$ into (3.19) and (3.20), we can derive the optimal strategies (3.28) and (3.29). □

B.2 Proof of Proposition 3.3.2

We extend the methods in Theorem 4.1.4. and Theorem 4.1.6. of Abou-Kandil et al. (2012) from their backward HRDE with the terminal value to our forward HRDE case with the initial value. Moreover, we restrict their comparison theorem from a semi-definite matrix case to a definite matrix case.

To begin with, we prove the following lemmas to deduce the comparison theorem.

Lemma B.2.1. *$\frac{\partial Y(t)}{\partial t} \geq$ (or $>$) 0 for $Y \in \mathbb{C}^{n\times n}$ and $t \in [t_1, t_2]$, implies $Y(t_2) \geq$ (or $>$) $Y(t_1)$.*

Proof. With $h(t, y) = y^\top Y(t)y$ for $(t, y) \in [t_1, t_2] \times \mathbb{C}^n$, we have

$$\frac{\partial h(t, y)}{\partial t} \geq (or >) 0,$$

for $t \in [t_1, t_2]$ and $y \neq 0$. Then, the mean value theorem shows

$$0 \leq (or < 0) h(t_2, y) - h(t_1, y) = y^\top [Y(t_2) - Y(t_1)]y,$$

for $y \neq 0$, which proves the lemma. □

Lemma B.2.2. *If Y is a Hermitian solution of the Lyapunov differential inequality*

$$\frac{\partial Y}{\partial t} > A(t)Y + Y A^\top(t), t \in [0, T], \tag{B.3}$$

then $Y(0) \geq 0$ implies $Y(t) > 0$ on $(0, T]$.

Proof. Define $\Phi_{-A^\top}$, the fundamental matrix of $-A^\top$, by property (see Theorem 1.1.1. in Abou-Kandil et al. (2012)), we have

$$\frac{\partial}{\partial t}\Phi_{-A^\top}(t, \tau) = -A^\top(t)\Phi_{-A^\top}(t, \tau),$$

and

$$\frac{\partial}{\partial t}\Phi^\top_{-A^\top}(t, \tau) = -\Phi^\top_{-A^\top}(t, \tau)A(t),$$

where $\Phi_{-A^\top}(t, t) = \widetilde{I}_n$, $t, \tau \in [0, T]$, $0 \leq t \leq \tau \leq T$.

Let $P(t,\tau) := \Phi^{\top}_{-A^{\top}}(t,\tau)Y(t)\Phi_{-A^{\top}}(t,\tau)$, then we infer from $Y(0) \geq 0$ that

$$P(0,\tau) \geq 0, \text{ for } t \leq \tau.$$

Fix $t \leq \tau$, we have

$$\frac{\partial P(t,\tau)}{\partial t} = \Phi^{\top}_{-A^{\top}}(t,\tau)\Big\{ - A(t)Y(t) + \frac{\partial Y(t)}{\partial t} - Y(t)A^{\top}(t)\Big\}\Phi_{-A^{\top}}(t,\tau) > 0,$$

then by (B.3) and Lemma B.2.1, we have $P(t,\tau)$ is strictly increasing from 0 to τ with respect to t. Since $P(0,\tau) \geq 0$, we have

$$P(t,\tau) > 0, \text{for } 0 < t \leq \tau \leq T.$$

Let τ move to t, we have

$$Y(t) = P(t,t) > 0, \text{ for } t \in (0,T],$$

which completes the proof. □

Consider two Hamiltonian matrices

$$H_i(t) = \begin{pmatrix} -\widetilde{A}_i(t) & \widetilde{S}_i(t) \\ \widetilde{Q}_i(t) & \widetilde{A}_i^{\top}(t) \end{pmatrix}, \ i = 1,2,$$

where $\widetilde{A}_i, \widetilde{S}_i, \widetilde{Q}_i : \mathbb{R} \to \mathbb{C}^{n\times n}$, $\widetilde{S}_i(t) = \widetilde{S}_i^{\top}(t)$ and $\widetilde{Q}_i(t) = \widetilde{Q}_i^{\top}(t)$. Define

$$J(t) = \begin{pmatrix} 0 & \widetilde{I}_n \\ -\widetilde{I}_n & 0, \end{pmatrix},$$

where $\widetilde{I}_n$ is the n-th order identity matrix.

Then, we have two Hermitian Riccati differential equations can be written as

$$\begin{aligned} \frac{\partial Y_i(t)}{\partial t} &= \widetilde{A}_i^{\top}(t)Y(t) + Y(t)\widetilde{A}_i(t) + \widetilde{Q}_i(t) - Y(t)\widetilde{S}_i(t)Y(t) \\ &= (\widetilde{I}_n, Y)JH_i(t)\begin{pmatrix} \widetilde{I}_n \\ Y \end{pmatrix} = \mathcal{H}(Y; H_i). \end{aligned}$$

Lemma B.2.3. *(**Comparison theorem for HRDEs**) Suppose that the matrix functions $H_i, i = 1, 2$ be piecewise continuous and locally bounded on $[0, T]$. If $Y_i, i = 1, 2$ are on $[0, T]$ the solutions of*

$$\frac{\partial Y_i(t)}{\partial t} = \widetilde{A}_i^\top(t)Y_i(t) + Y_i(t)\widetilde{A}_i(t) + \widetilde{Q}_i(t) - Y_i(t)\widetilde{S}_i(t)Y_i(t),$$

with

$$\begin{array}{rcl} Y_1(0) & \leq & Y_2(0), \\ JH_1(t) & < & JH_2(t), \ \textit{for } t \in [0, T], \end{array}$$

then $Y_1(t) < Y_2(t)$ for $t \in (0, T]$.

Proof. Denote $\bar{Y} = Y_2 - Y_1$, then

$$\begin{array}{rcl} \dfrac{\partial \bar{Y}}{\partial t} & = & \widetilde{A}_2^\top Y_2 + Y_2\widetilde{A}_2 + \widetilde{Q}_2 - Y_2\widetilde{S}_2Y_2 \\ & & -\widetilde{A}_1^\top Y_1 - Y_1\widetilde{A}_1 - \widetilde{Q}_1 + Y_1\widetilde{S}_1Y_1 \\ & = & -(Y_2 - Y_1)\widetilde{S}_2(Y_2 - Y_1) - Y_1\widetilde{S}_2(Y_2 - Y_1) - (Y_2 - Y_1)\widetilde{S}_2Y_1 + \widetilde{A}_2^\top(Y_2 - Y_1) \\ & & +(Y_2 - Y_1)\widetilde{A}_2 + [(\widetilde{A}_2 - \widetilde{A}_1)^\top Y_1 + Y_1(\widetilde{A}_2 - \widetilde{A}_1) + (\widetilde{Q}_2 - \widetilde{Q}_1) - Y_1(\widetilde{S}_2 - \widetilde{S}_1)Y_1] \\ & = & \widetilde{A}\bar{Y} + \bar{Y}\widetilde{A}^\top + \mathcal{H}(Y_1; H_2 - H_1), \end{array}$$

with $\widetilde{A} = -\frac{1}{2}\bar{Y}\widetilde{S}_2 - Y_1\widetilde{S}_2 + \widetilde{A}_2^\top$. Since $JH_1(t) < JH_2(t)$, we have $\mathcal{H}(Y_1; H_2 - H_1) > 0$, then

$$\frac{\partial \bar{Y}(t)}{\partial t} > \widetilde{A}\bar{Y} + \bar{Y}\widetilde{A}, \quad \bar{Y}(0) \geq 0.$$

By Lemma B.2.2, we have $\bar{Y}(t) := Y_2(t) - Y_1(t) > 0$ for $t \in (0, T]$, which completes the proof. □

Lemma B.2.4. *Suppose that the HRDE*

$$\frac{\partial Y}{\partial t} = \widetilde{A}^\top(t)Y + Y\widetilde{A} + \widetilde{Q}(t) - Y\widetilde{S}(t)Y, \ Y(0) = 0,$$

has piecewise continuous and locally bounded coefficients, if $\widetilde{S}(t), \widetilde{Q}(t) < 0$ for $t \in [0, T]$, then the unique solution Y exists for $t \in [0, T]$ and

$$\widetilde{Y}(t) < Y(t) < 0, \ \textit{for } t \in (0, T].$$

Here, $\widetilde{Y}$ is the solution of

$$\frac{\partial \widetilde{Y}}{\partial t} = \widetilde{A}^\top(t)\widetilde{Y} + \widetilde{Y}\widetilde{A}(t) + 2\widetilde{Q}(t), \ \widetilde{Y}(0) = 0.$$

Proof. Define the 3rd HRDE with respect to $Y_0(t)$

$$\frac{\partial Y_0}{\partial t} = \widetilde{A}^\top(t)Y_0 + Y_0\widetilde{A}(t) - 2Y_0\widetilde{S}(t)Y_0,\ Y_0(0) = 0.$$

For $Y(t)$, $\widetilde{Y}(t)$ and $Y_0(t)$, we have the following matrices respectively

$$JH(t) = \begin{pmatrix} \widetilde{Q}(t) & \widetilde{A}^\top(t) \\ \widetilde{A}(t) & -\widetilde{S}(t) \end{pmatrix}, J\widetilde{H}(t) = \begin{pmatrix} 2\widetilde{Q}(t) & \widetilde{A}^\top(t) \\ \widetilde{A}(t) & 0 \end{pmatrix}, JH_0(t) = \begin{pmatrix} 0 & \widetilde{A}^\top(t) \\ \widetilde{A}(t) & -2\widetilde{S}(t) \end{pmatrix},$$

Since $\widetilde{S}(t), \widetilde{Q}(t) < 0$ for $t \in [0, T]$, we have

$$\widetilde{Y}(0) = Y(0), \quad J\widetilde{H} < JH.$$

Then, by Lemma B.2.3, we obtain $\widetilde{Y}(t) < Y(t)$ for $t \in (0, T]$. Similarly, between $Y(t)$ and $Y_0(t)$, we have $Y(0) = Y_0(0)$ and $JH < JH_0$. Therefore, $Y(t) < Y_0(t) = 0$ for $t \in (0, T]$.

In general,

$$\widetilde{Y}(t) < Y(t) < 0 \text{ for } t \in (0, T].$$

Since $\widetilde{Y}(t)$ satisfies a linear ODE, the boundness of $Y(t)$ is guaranteed. □

Due to $\gamma > 1$, $\Sigma_X\Sigma_X^\top > 0$, and $\Lambda_1^\top\Lambda_1 > 0$, we have $Z_0 < 0$ and $-Z_2 < 0$ for any $t \in [0, T]$. By Lemma B.2.4, the solution to ODE (3.30) exists in $[0, T]$, so does ODE (3.25).

B.3 Proof of Proposition 3.3.3

Substituting $y = \lambda - \frac{b}{4}$ into (3.34), we derive the reduced form

$$f_y(y) = y^4 + qy^2 + ry + s,$$

where

$$\begin{aligned} q &= \frac{8c - 3b^2}{8}, \\ r &= \frac{b^3 - 4bc + 8d}{8}, \\ s &= \frac{-3b^4 + 256j - 64bd + 16b^2c}{256}. \end{aligned}$$

Moreover, the discriminant of $f_y(y)$ is given by

$$\widetilde{\Delta} = -4q^3r^2 - 27r^4 + 256s^3 + 16q^4s + 144qr^2s - 128q^2s^2.$$

According to Rees (1922), once conditions (3.35) satisfied, (3.34) has 4 distinct real roots, i.e. the Hamiltonian matrix H has 4 different real eigen-values, which guarantees its diagonalizability and the full rank of its eigen-vector matrix V.

Finally, by Radon's lemma (see Theorem 3.1.1. in Abou-Kandil et al. (2012)), we have $\Gamma_2(\tau) = P(\tau)/Q(\tau)$ and the existence and negative definiteness of $\Gamma_2(\tau)$ from (3.36).

B.4 Proof of Lemma 3.3.1

Proof. Inspired by the Lemma 4.1.1. in Bensoussan (2004), we extend their result to the case where $\widetilde{X}_t$ and $\mathcal{E}(t, \widetilde{g})$ share the same Brownian motion.

First, we prove the following bounded result

$$E[\mathcal{E}(t, \widetilde{g})||\widetilde{X}_t||_2^2] < \infty. \tag{B.4}$$

By Ito's formula, we have

$$\begin{aligned} d||\widetilde{X}_t||_2^2 &= d(\widetilde{X}_t^\top \widetilde{X}_t) \\ &= 2\widetilde{X}_t^\top \sigma(t) d\widetilde{Z}_t + \{2\widetilde{X}_t^\top \mu(t, \widetilde{X}_t) + \mathrm{Tr}[\sigma^\top(t)\sigma(t)]\}dt, \\ d(\mathcal{E}(t, \widetilde{g})||\widetilde{X}_t||_2^2) &= ||\widetilde{X}_t||_2^2 \mathcal{E}(t, \widetilde{g})\widetilde{g}^\top(t, \widetilde{X}_t) d\widetilde{Z}_t + 2\mathcal{E}(t, \widetilde{g})\widetilde{X}_t^\top \sigma(t) d\widetilde{Z}_t \\ &\quad + \mathcal{E}(t, \widetilde{g})\{2\widetilde{X}_t^\top \mu(t, \widetilde{X}_t) + \mathrm{Tr}[\sigma^\top(t)\sigma(t)] + 2\widetilde{X}_t^\top \sigma(t)\widetilde{g}(t, \widetilde{X}_t)\}dt. \end{aligned}$$

Then, for $\frac{\mathcal{E}(t,\widetilde{g})||\widetilde{X}_t||_2^2}{1+\epsilon\mathcal{E}(t,\widetilde{g})||\widetilde{X}_t||_2^2}, \epsilon \geq 0$, we can derive its differential

$$d\left(\frac{\mathcal{E}(t, \widetilde{g})||\widetilde{X}_t||_2^2}{1 + \epsilon\mathcal{E}(t, \widetilde{g})||\widetilde{X}_t||_2^2}\right) = \frac{d(\mathcal{E}(t, \widetilde{g})||\widetilde{X}_t||_2^2)}{[1 + \epsilon\mathcal{E}(t, \widetilde{g})||\widetilde{X}_t||_2^2]^2} - \frac{\epsilon d(\mathcal{E}(t, \widetilde{g})||\widetilde{X}_t||_2^2) d(\mathcal{E}(t, \widetilde{g})||\widetilde{X}_t||_2^2)}{[1 + \epsilon\mathcal{E}(t, \widetilde{g})||\widetilde{X}_t||_2^2]^3}.$$

Thus,

$$d\left(\frac{\mathcal{E}(t, \widetilde{g})||\widetilde{X}_t||_2^2}{1 + \epsilon\mathcal{E}(t, \widetilde{g})||\widetilde{X}_t||_2^2}\right) \leq \frac{d(\mathcal{E}(t, \widetilde{g})||\widetilde{X}_t||_2^2)}{1 + \epsilon\mathcal{E}(t, \widetilde{g})||\widetilde{X}_t||_2^2}. \tag{B.5}$$

Integrate both sides of (B.5) over $(0,t)$, take expectation and then take the first-order derivative with respect to t, we have

$$\begin{aligned}\frac{d}{dt}E\left[\frac{\mathcal{E}(t,\widetilde{g})||\widetilde{X}_t||_2^2}{1+\epsilon\mathcal{E}(t,\widetilde{g})||\widetilde{X}_t||_2^2}\right] &\leq E\left\{\frac{\mathcal{E}(t,\widetilde{g})}{1+\epsilon\mathcal{E}(t,\widetilde{g})||\widetilde{X}_t||_2^2}\left[2\widetilde{X}_t^\top\mu(t,\widetilde{X}_t)+\mathrm{Tr}[\sigma^\top(t)\sigma(t)]\right.\right.\\ &\left.\left.+2\widetilde{X}_t^\top\sigma(t)\widetilde{g}(t,\widetilde{X}_t)\right]\right\}. \end{aligned} \tag{B.6}$$

For each term in (B.6), we can derive the following estimates

$$\begin{aligned}2\widetilde{X}_t^\top\mu(t,\widetilde{X}_t) &\leq 2|\widetilde{X}_t^\top\mu(t,\widetilde{X}_t)|\\ &\leq 2||\widetilde{X}_t||_2||\mu(t,\widetilde{X}_t)||_2\\ &\leq 2||\widetilde{X}_t||_2(c_0+c_1||\widetilde{X}_t||_2)\\ &\leq c_2+c_3||\widetilde{X}_t||_2^2,\\ \mathrm{Tr}[\sigma^\top(t)\sigma(t)] &\leq c_4,\\ 2\widetilde{X}_t^\top\sigma(t)\widetilde{g}(t,\widetilde{X}_t) &\leq 2|\widetilde{X}_t^\top\sigma(t)\widetilde{g}(t,\widetilde{X}_t)|\\ &\leq 2||\widetilde{X}_t^\top||_2||\sigma(t)||_2||\widetilde{g}(t,\widetilde{X}_t)||_2\\ &\leq 2||\widetilde{X}_t||_2(c_5+c_6||\widetilde{X}_t||_2)\\ &\leq c_7+c_8||\widetilde{X}_t||_2^2,\end{aligned}$$

where c_0 to c_8 are some constants, and the boundedness of $\mathrm{Tr}[\sigma^\top(t)\sigma(t)]$ is due to the continuity of $\sigma(t)$ over $[0,T]$. Therefore, (B.6) is bounded by

$$\begin{aligned}\frac{d}{dt}E\left[\frac{\mathcal{E}(t,\widetilde{g})||\widetilde{X}_t||_2^2}{1+\epsilon\mathcal{E}(t,\widetilde{g})||\widetilde{X}_t||_2^2}\right] &\leq E\left\{\frac{\mathcal{E}(t,\widetilde{g})(c_9+c_{10}||\widetilde{X}_t||_2^2)}{1+\epsilon\mathcal{E}(t,\widetilde{g})||\widetilde{X}_t||_2^2}\right\}\\ &\leq c_{10}E\left\{\frac{\mathcal{E}(t,\widetilde{g})||\widetilde{X}_t||_2^2}{1+\epsilon\mathcal{E}(t,\widetilde{g})||\widetilde{X}_t||_2^2}\right\}+c_9E\left\{\mathcal{E}(t,\widetilde{g})\right\}\\ &\leq c_{10}E\left\{\frac{\mathcal{E}(t,\widetilde{g})||\widetilde{X}_t||_2^2}{1+\epsilon\mathcal{E}(t,\widetilde{g})||\widetilde{X}_t||_2^2}\right\}+c_9,\end{aligned} \tag{B.7}$$

where c_9 and c_{10} are some constants. The third inequality holds because $\mathcal{E}(t,\widetilde{g})$ is a super-martingale such that $E\left\{\mathcal{E}(t,\widetilde{g})\right\}\leq 1$. And by Gronwall inequality, we derive the boundedness of $E\left[\frac{\mathcal{E}(t,\widetilde{g})||\widetilde{X}_t||_2^2}{1+\epsilon\mathcal{E}(t,\widetilde{g})||\widetilde{X}_t||_2^2}\right]$. Finally, since $\frac{\mathcal{E}(t,\widetilde{g})||\widetilde{X}_t||_2^2}{1+\epsilon\mathcal{E}(t,\widetilde{g})||\widetilde{X}_t||_2^2}\geq 0$ and $\epsilon\geq 0$, we can derive

(B.4) by Fatou's lemma

$$E[\mathcal{E}(t,\widetilde{g})||\widetilde{X}_t||_2^2] \leq \lim_{\epsilon\to 0} E\left[\frac{\mathcal{E}(t,\widetilde{g})||\widetilde{X}_t||_2^2}{1+\epsilon\mathcal{E}(t,\widetilde{g})||\widetilde{X}_t||_2^2}\right] < \infty.$$

Next, we construct the structure $\frac{\mathcal{E}(t,\widetilde{g})}{1+\epsilon\mathcal{E}(t,\widetilde{g})}$, $\epsilon \geq 0$, and take its differential

$$d\frac{\mathcal{E}(t,\widetilde{g})}{1+\epsilon\mathcal{E}(t,\widetilde{g})} = \frac{d(\mathcal{E}(t,\widetilde{g}))}{[1+\epsilon\mathcal{E}(t,\widetilde{g})]^2} - \frac{\epsilon\mathcal{E}^2(t,\widetilde{g})}{[1+\epsilon\mathcal{E}(t,\widetilde{g})]^3}\widetilde{g}^\top(t,\widetilde{X}_t)\widetilde{g}(t,\widetilde{X}_t)dt. \tag{B.8}$$

Integrate (B.8) from 0 to t and take expectation, we have

$$E\left[\frac{\mathcal{E}(t,\widetilde{g})}{1+\epsilon\mathcal{E}(t,\widetilde{g})}\right] = \frac{1}{1+\epsilon} - E\left[\int_0^t \frac{\epsilon\mathcal{E}^2(s,\widetilde{g})}{[1+\epsilon\mathcal{E}(s,\widetilde{g})]^3}\widetilde{g}^\top(s,\widetilde{X}_s)\widetilde{g}(s,\widetilde{X}_s)ds\right],\ \epsilon \geq 0. \tag{B.9}$$

For the integrand in the right-hand side of (B.9), we can control it by

$$\begin{aligned}
\frac{\epsilon\mathcal{E}^2(s,\widetilde{g})}{[1+\epsilon\mathcal{E}(s,\widetilde{g})]^3}\widetilde{g}^\top(s,\widetilde{X}_s)\widetilde{g}(s,\widetilde{X}_s) &= \frac{\epsilon\mathcal{E}(s,\widetilde{g})}{1+\epsilon\mathcal{E}(s,\widetilde{g})}\frac{1}{[1+\epsilon\mathcal{E}(s,\widetilde{g})]^2}\mathcal{E}(s,\widetilde{g})\widetilde{g}^\top(s,\widetilde{X}_s)\widetilde{g}(s,\widetilde{X}_s) \\
&\leq \mathcal{E}(s,\widetilde{g})\widetilde{g}^\top(s,\widetilde{X}_s)\widetilde{g}(s,\widetilde{X}_s) \\
&\leq \mathcal{E}(s,\widetilde{g})(c_{11}+c_{12}||\widetilde{X}_s||_2)^2 \\
&\leq \mathcal{E}(s,\widetilde{g})[c_{13}+c_{14}||\widetilde{X}_s||_2^2],
\end{aligned}$$

where c_{11} to c_{14} are some constants. Then, from (B.4) and $E\{\mathcal{E}(s,\widetilde{g})\} \leq 1$, we obtain

$$E\{\mathcal{E}(s,\widetilde{g})[c_{13}+c_{14}||\widetilde{X}_s||_2^2]\} < \infty.$$

Finally, we can use Lebesgue's dominated convergence theorem to the right-hand side of (B.9) when $\epsilon \to 0$. Moreover, $\frac{\mathcal{E}(t,\widetilde{g})}{1+\epsilon\mathcal{E}(t,\widetilde{g})}$ is increasing as $\epsilon \to 0$, the monotone convergence theorem can be applied to the left-hand side of (B.9).

In general, when $\epsilon \to 0$, (B.9) implies $E\{\mathcal{E}(t,\widetilde{g})\} = 1$, which completes the proof.

□

B.5 Proof of Proposition 3.3.4

Proof. The parameter settings in Proposition 3.3.3 guarantees the global existence of HRDE (3.25), and thus ensures the global existence of $G(t, W_t^{\widetilde{C}}, X_t)$ in (3.22). Then, we make the following preparations before the proof of verification theorem.

To begin with, for any $(\beta_t, I_t) \in \mathcal{A}_\gamma(0,T)$, we define the value process

$$\begin{aligned} g^{\beta,I}(s, W_s^{\widetilde{C}}, X_s) \quad := \quad & \int_t^s {}_{u-t}p_{x+t}\mu_{x+u} U\left(W_u^{\widetilde{C}} - \widetilde{C}(u, X_u) + \frac{I_u}{\mu_{x+u}}\right) du \\ & + {}_{s-t}p_{x+t} G(s, W_s^{\widetilde{C}}, X_s), s \in [t, T]. \end{aligned} \tag{B.10}$$

By Ito's formula, we have

$$\begin{aligned} dg^{\beta,I}(s, W_s^{\widetilde{C}}, X_s) \quad = \quad & {}_{s-t}p_{x+t}\Bigg\{\mu_{x+s} U\left(W_s^{\widetilde{C}} - \widetilde{C}(s, X_s) + \frac{I_s}{\mu_{x+s}}\right) - \mu_{x+s} G(s, W_s^{\widetilde{C}}, X_s) \\ & + \mathcal{D}^{\beta,I} G(s, W_s^{\widetilde{C}}, X_s)\Bigg\} ds + g^{\beta,I}(s, W_s^{\widetilde{C}}, X_s) h^{\beta,I}(s, W_s^{\widetilde{C}}, X_s) dZ_s, \end{aligned} \tag{B.11}$$

where

$$\begin{aligned} \mathcal{D}^{\beta,I} G(s, W_s^{\widetilde{C}}, X_s) \quad = \quad & \frac{\partial G}{\partial s} + \frac{\partial G}{\partial W^{\widetilde{C}}}\{W_s^{\widetilde{C}}[r_s + (\beta_s^\top \Sigma - \sigma_\Pi^\top)(\Lambda_s - \sigma_\Pi)] + \mu_{x+s}\widetilde{C}(s, X_s) - I_s\} \\ & - \frac{\partial G}{\partial X} K_X X_s + \frac{1}{2}\frac{\partial^2 G}{\partial (W^{\widetilde{C}})^2}(W_s^{\widetilde{C}})^2 (\beta_s^\top \Sigma - \sigma_\Pi^\top)(\Sigma^\top \beta_s - \sigma_\Pi) \\ & + W_s^{\widetilde{C}} \frac{\partial^2 G}{\partial X \partial W^{\widetilde{C}}} \Sigma_X (\Sigma^\top \beta_s - \sigma_\Pi) + \frac{1}{2}\mathrm{Tr}\left(\Sigma_X^\top \frac{\partial^2 G}{\partial X^\top \partial X}\Sigma_X\right), \\ h^{\beta,I}(s, W_s^{\widetilde{C}}, X_s) \quad = \quad & \frac{{}_{s-t}p_{x+t} G(s, W_s^{\widetilde{C}}, X_s)}{g^{\beta,I}(s, W_s^{\widetilde{C}}, X_s)}\left[(1-\gamma)(\beta_s^\top \Sigma - \sigma_\Pi^\top) + \gamma \frac{1}{f_1}\frac{\partial f_1}{\partial X}\Sigma_X\right]. \end{aligned} \tag{B.12}$$

Next, fix $(t, w^R, X) \in [0,T] \times [0,\infty) \times \mathbb{R}^2$ and denote the conditional expectation of the value process as

$$\begin{aligned} & J(t, w^{\widetilde{C}}, X) := \\ & E_{t,w^{\widetilde{C}},X}\left[\int_t^T {}_{s-t}p_{x+t}\mu_{x+s} U\left(W_s^{\widetilde{C}} - \widetilde{C}(s, X_s) + \frac{I_s}{\mu_{x+s}}\right) ds + {}_{T-t}p_{x+t} U(W_T^{\widetilde{C}})\right], \end{aligned} \tag{B.13}$$

where $E_{t,w^{\widetilde{C}},X}[\cdot]$ is short for $E[\cdot | W_t^{\widetilde{C}} = w^{\widetilde{C}}, X_t = X]$. Then, we have

$$V(t, W_t^{\widetilde{C}}, X_t) = \sup_{(\beta,I)\in\mathcal{A}_\gamma(0,T)} J(t, W_t^{\widetilde{C}}, X_t). \tag{B.14}$$

Finally, we can prove the verification theorem by the following three steps:

Step 1: Verify the optimal strategy (β^*, I^*) belongs to the admissible set $\mathcal{A}_\gamma(0,T)$.

Substitute (3.28) and (3.29) into (3.16), we have

$$d(W_t^{\widetilde{C}})^* = (W_t^{\widetilde{C}})^* \left\{ r_t + \mu_{x+t}\left(1 - \frac{1}{f_1(t,X_t)}\right) + (\eta_t)^\top(\Lambda_t - \sigma_\Pi) \right\} dt + (W_t^{\widetilde{C}})^*(\eta_t)^\top dZ_t, \tag{B.15}$$

where $(\eta_t)^\top = \frac{1}{\gamma}(\Lambda_t^\top - \sigma_\Pi^\top) + \frac{1}{f_1}\frac{\partial f_1}{\partial X}\Sigma_X$. Then

$$\begin{aligned}(W_t^{\widetilde{C}})^* &= W_0^{\widetilde{C}} \exp\left\{ \int_0^t \left[r_s + \mu_{x+s}\left(1 - \frac{1}{f_1(s,X_s)}\right) + \eta_s^\top(\Lambda_s - \sigma_\Pi) \right.\right. \\ &\quad \left.\left. -\frac{1}{2}\eta_s^\top \eta_s \right] ds + \int_0^t \eta_s^\top dZ_s \right\} > 0. \end{aligned} \tag{B.16}$$

Since the drift term and volatility term of SDE (B.15) are almost surely sample continuous, then on any bounded interval $[0,t]$, $t \in [0,T]$, we have

$$\int_0^t \left| r_s + \mu_{x+s}\left(1 - \tfrac{1}{f_1(s,X_s)}\right) + (\eta_s)^\top(\Lambda_s - \sigma_\Pi) \right| ds < \infty,$$
$$\int_0^t (\eta_s)^\top \eta_s ds < \infty.$$

By Proposition 1.1 in Kraft (2004), we derive SDE (3.16) has a unique strong solution under (β^*, I^*).

From all things above, we show that $(\beta_t^*, I_t^*) \in \mathcal{A}_\gamma(0,T)$.

Step 2: Verify $J(t, W_t^{\widetilde{C}}, X_t) \le G(t, W_t^{\widetilde{C}}, X_t)$ for any $(\beta, I) \in \mathcal{A}_\gamma(0,T)$.

Define

$$\Psi(s) := \int_t^s ||g^{\beta,I}(u, W_u^{\widetilde{C}}, X_u) h^{\beta,I}(u, W_u^{\widetilde{C}}, X_u)||_2^2 du,$$

and $\tau_n := T \wedge \inf\{s \in [t,T] | \Psi(s) \ge n\}, n \in \mathbb{N}$. For $s \in [t, \tau_n]$, we have the stochastic integral $\int_t^s g^{\beta,I}(u, W_u^{\widetilde{C}}, X_u) h^{\beta,I}(u, W_u^{\widetilde{C}}, X_u) dZ(u)$ is a martingale.

Then, following (3.18) and (B.11), we have

$$\begin{aligned} g^{\beta,I}(\tau_n, W_{\tau_n}^{\widetilde{C}}, X_{\tau_n}) &\le g^{\beta,I}(t, W_t^{\widetilde{C}}, X_t) \\ &\quad + \int_t^{\tau_n} g^{\beta,I}(s, W_s^{\widetilde{C}}, X_s) h^{\beta,I}(s, W_s^{\widetilde{C}}, X_s) dZ_s. \end{aligned} \tag{B.17}$$

Since $\lim_{n\to\infty}\tau_n = T$ and $g^{\beta,I}(t, W_t^{\widetilde{C}}, X_t) \geq 0$ for any $t \in [0,T]$ under $0 < \gamma < 1$, we can use the Fatou's lemma and derive the following inequality

$$
\begin{aligned}
& J(t, w^{\widetilde{C}}, X) \\
= \; & E_{t,w^{\widetilde{C}},X}\left[\int_t^T {}_{s-t}p_{x+t}\mu_{x+s}U\left(W_s^{\widetilde{C}} - \widetilde{C}(s, X_s) + \frac{I_s}{\mu_{x+s}}\right)ds + {}_{T-t}p_{x+t}U(W_T^{\widetilde{C}})\right] \\
= \; & E_{t,w^{\widetilde{C}},X}[g^{\beta,I}(T, W_T^{\widetilde{C}}, X_T)] \\
\leq \; & \lim_{n\to\infty} E_{t,w^{\widetilde{C}},X}[g^{\beta,I}(\tau_n, W_{\tau_n}^{\widetilde{C}}, X_{\tau_n})] \\
\leq \; & g^{\beta,I}(t, w^{\widetilde{C}}, X) \\
= \; & G(t, w^{\widetilde{C}}, X), \text{ for } \forall(\beta, I) \in \mathcal{A}_\gamma(0,T),
\end{aligned} \tag{B.18}
$$

where the first inequality is by the Fatou's lemma, the second inequality is deduced by taking conditional expectation on both sides of (B.17), and the last equality holds by the definition of value process (B.10).

Step 3: Verify $V(t, W_t^{\widetilde{C}}, X_t) = G(t, W_t^{\widetilde{C}}, X_t)$ under the optimal strategy (β^*, I^*).

Since (β_t^*, I_t^*) maximizes the HJB (3.18) and $G(t, W_t^{\widetilde{C}}, X_t)$ is the solution to (3.18), we have

$$
dg^{\beta^*,I^*}(s, (W_s^{\widetilde{C}})^*, X_s) = g^{\beta^*,I^*}(s, (W_s^{\widetilde{C}})^*, X_s)h^{\beta^*,I^*}(s, (W_s^{\widetilde{C}})^*, X_s)^\top dZ_s, s \in [t,T],
$$

where

$$
h^{\beta^*,I^*}(s, (W_s^{\widetilde{C}})^*, X_s) = \frac{{}_{s-t}p_{x+t}G(s, (W_s^{\widetilde{C}})^*, X_s)}{g^{\beta^*,I^*}(s, (W_s^{\widetilde{C}})^*, X_s)}\left[\frac{1-\gamma}{\gamma}(\Lambda_s^\top - \sigma_\Pi^\top) + \frac{1}{f_1}\frac{\partial f_1}{\partial X}\Sigma_X\right]. \tag{B.19}
$$

Solving it by Ito's formula, we have the optimal value process

$$
g^{\beta^*,I^*}(s', (W_{s'}^{\widetilde{C}})^*, X_{s'}) = g^{\beta^*,I^*}(s, (W_s^{\widetilde{C}})^*, X_s)\frac{\mathcal{E}(s', h^{\beta^*,I^*})}{\mathcal{E}(s, h^{\beta^*,I^*})}, s' \in [s,T], \tag{B.20}
$$

where

$$
\mathcal{E}(t,h) := \exp\left\{\int_0^t h(s, (W_s^{\widetilde{C}})^*, X_s)^\top dZ_s - \frac{1}{2}\int_0^t ||h(s, (W_s^{\widetilde{C}})^*, X_s)||_2^2 ds\right\}.
$$

If h^{β^*,I^*} satisfies the linear growth condition, we can use Lemma 3.3.1 to prove $E[\mathcal{E}(t,h)] = 1$ and thus $\mathcal{E}(t,h)$ is a martingale. We split (B.19) into the following parts to prove it satisfies linear growth when $0 < \gamma < 1$.

(a) Under $0 < \gamma < 1$, we have $G(s, (W_s^{\widetilde{C}})^*, X_s) > 0$ for $\forall s \in [t, T]$. Since $U'(0) = +\infty, 0 < \gamma < 1$, we obtain $(W_u^{\widetilde{C}})^* - \widetilde{C}(u, X_u) + \frac{I_u^*}{\mu_{x+u}} > 0$ and $U\bigg((W_u^{\widetilde{C}})^* - \widetilde{C}(u, X_u) + \frac{I_u^*}{\mu_{x+u}}\bigg) > 0$ for $\forall u \in [t, s]$. Then, the following inequality holds by the definition of the value process (B.10)

$$\left| \frac{{}_{s-t}p_{x+t} G(s, (W_s^{\widetilde{C}})^*, X_s)}{g^{\beta^*, I^*}(s, (W_s^{\widetilde{C}})^*, X_s)} \right| \leq 1, \text{ for } \forall s \in [t, T].$$

(b) Since $\Lambda_s = \Lambda_0 + \Lambda_1 X_s$, a direct result is

$$\left|\left| \frac{1-\gamma}{\gamma} (\Lambda_s^\top - \sigma_\Pi^\top) \right|\right|_2 \leq c_0(1 + ||X_s||_2), \forall s \in [t, T],$$

for some constant c_0

(c) For $\frac{1}{f_1}\frac{\partial f_1}{\partial X}\Sigma_X$ term, we have

$$\begin{aligned} \frac{\partial f_1(s, X_s)}{\partial X} &= \int_s^T {}_{u-s}p_{x+s}\mu_{x+u} f(X_s, u-s)[\Gamma_1^\top(u-s) + X_s^\top \Gamma_2(u-s)]du \\ &\quad + {}_{T-s}p_{x+s} f(X_s, T-s)[\Gamma_1^\top(T-s) + X_s^\top \Gamma_2(T-s)]. \end{aligned}$$

Under $0 < \gamma < 1$, once $\Gamma_1(\tau)$ and $\Gamma_2(\tau)$ exist in $[0, T]$, i.e. under the setting of Proposition 3.3.3, we can obtain

$$\left|\left| \frac{\partial f_1}{\partial X} \right|\right|_2 \leq f_1 c_1(1 + ||X_s||_2),$$

for some constant c_1. Then

$$\left|\left| \frac{1}{f_1}\frac{\partial f_1}{\partial X} \right|\right|_2 \leq \frac{1}{f_1}\left|\left| \frac{\partial f_1}{\partial X} \right|\right|_2 \leq c_1(1 + ||X_s||_2).$$

In general, we have proved $h^{\beta^*, I^*}(s, (W_s^{\widetilde{C}})^*, X_s)$ is subject to linear growth with respect to X_t. Therefore, by Lemma 3.3.1, $\mathcal{E}(t, h)$ is a martingale. Then, by

(B.20), we have the inequality for $G(t, w^{\widetilde{C}}, X)$

$$\begin{aligned}
& V(t, w^{\widetilde{C}}, X) \\
\geq\ & E_{t,w^{\widetilde{C}},X}\left[\int_t^T {}_{s-t}p_{x+t}\mu_{x+s}U\left((W_s^{\widetilde{C}})^* - \widetilde{C}(s, X_s) + \frac{I_s^*}{\mu_{x+s}}\right)ds \right. \\
& \left. + {}_{T-t}p_{x+t}U((W_T^{\widetilde{C}})^*)\right] \\
=\ & E_{t,w^{\widetilde{C}},X}[g^{\beta^*,I^*}(T, (W_T^{\widetilde{C}})^*, X_T)] \\
=\ & E_{t,w^{\widetilde{C}},X}\left[g^{\beta^*,I^*}(t, w^{\widetilde{C}}, X)\frac{\mathcal{E}(T, h^{\beta^*,I^*})}{\mathcal{E}(t, h^{\beta^*,I^*})}\right] \\
=\ & G(t, w^{\widetilde{C}}, X).
\end{aligned} \tag{B.21}$$

Combining (B.18), (B.21), and (B.14), we show that $G(t, W_t^{\widetilde{C}}, X_t) = V(t, W_t^{\widetilde{C}}, X_t)$, and (β^*, I^*) given by (3.28) and (3.29) is the optimal portfolio and insurance strategy.

□

B.6 Proof of Proposition 3.3.5

Proof. For $\gamma > 1$, Proposition 3.3.2 guarantees the global existence of HRDE (3.25), and thus ensures the global existence of $G(t, W_t^{\widetilde{C}}, X_t)$ in (3.22). Then, similarly to Appendix B.5, we follow three steps to prove the verification theorem.

Step 1: Verify the optimal strategy (β^*, I^*) belongs to the admissible set $\mathcal{A}_\gamma(0, T)$.

Recall from (3.28)

$$\beta_t^* = \frac{(\Sigma^\top)^{-1}}{\gamma}(\Lambda_t - \sigma_\Pi) + (\Sigma^\top)^{-1}\Sigma_X^\top \frac{1}{f_1}\frac{\partial f_1}{\partial X^\top} + (\Sigma^\top)^{-1}\sigma_\Pi.$$

The terms $(\Lambda_t - \sigma_\Pi)$ and $\frac{1}{f_1}\frac{\partial f_1}{\partial X^\top}$ satisfy linear growth according to the proofs in Step 3: (b) and (c) of Appendix B.5. Moreover, similar to the Step 1 in Appendix B.5, we can show that SDE (3.16) has a unique strong solution under (β^*, I^*). From all things above, we have $(\beta_t^*, I_t^*) \in \mathcal{A}_\gamma(0, T)$.

Step 2: Verify $J(t, W_t^{\widetilde{C}}, X_t) \leq G(t, W_t^{\widetilde{C}}, X_t)$ for any $(\beta, I) \in \mathcal{A}_\gamma(0,T)$, where $J(t, W_t^{\widetilde{C}}, X_t)$ is given by (B.13).

Following (3.18) and (B.11), we have

$$g^{\beta,I}(T, W_T^{\widetilde{C}}, X_T) \leq g^{\beta,I}(t, W_t^{\widetilde{C}}, X_t)\frac{\mathcal{E}(T, h^{\beta,I})}{\mathcal{E}(t, h^{\beta,I})}, \tag{B.22}$$

Recall from (B.12), $h^{\beta,I}(s, W_s^{\widetilde{C}}, X_s)$, $s \in [t,T]$ is given by

$$h^{\beta,I}(s, W_s^{\widetilde{C}}, X_s) = \frac{{}_{s-t}p_{x+t}G(s, W_s^{\widetilde{C}}, X_s)}{g^{\beta,I}(s, W_s^{\widetilde{C}}, X_s)}\left[(1-\gamma)(\beta_s^\top \Sigma - \sigma_\Pi^\top) + \gamma\frac{1}{f_1}\frac{\partial f_1}{\partial X}\Sigma_X\right].$$

We split (B.12) into the following parts to prove it satisfies linear growth with respect to X_t when $\gamma > 1$.

(a) Under $\gamma > 1$, we have $G(s, W_s^{\widetilde{C}}, X_s) < 0$ for $\forall s \in [t,T]$. Since $U'(0) = +\infty$, for a reasonable individual, we have $W_u^{\widetilde{C}} - \widetilde{C}(u, X_u) + \frac{I_u}{\mu_{x+u}} > 0$ and $U\left(W_u^{\widetilde{C}} - \widetilde{C}(u, X_u) + \frac{I_u}{\mu_{x+u}}\right) < 0$ for $\forall u \in [t,s]$. Then, the following inequality holds by the definition of the value process (B.10)

$$\left|\frac{{}_{s-t}p_{x+t}G(s, W_s^{\widetilde{C}}, X_s)}{g^{\beta,I}(s, W_s^{\widetilde{C}}, X_s)}\right| \leq 1, \text{ for } \forall s \in [t,T].$$

(b) Since $(\beta_t, I_t) \in \mathcal{A}_\gamma(0,T)$, by the definition of $\mathcal{A}_\gamma(0,T)$, we have $(1-\gamma)(\beta_s^\top \Sigma - \sigma_\Pi^\top)$ follows a linear growth with respect to X_t.

(c) For $\frac{1}{f_1}\frac{\partial f_1}{\partial X}\Sigma_X$ term, we have

$$\begin{aligned}\frac{\partial f_1(s, X_s)}{\partial X} &= \int_s^T {}_{u-s}p_{x+s}\mu_{x+u}f(X_s, u-s)[\Gamma_1^\top(u-s) + X_s^\top \Gamma_2(u-s)]du \\ &\quad + {}_{T-s}p_{x+s}f(X_s, T-s)[\Gamma_1^\top(T-s) + X_s^\top\Gamma_2(T-s)].\end{aligned}$$

Under $\gamma > 1$, once $\Gamma_1(\tau)$ and $\Gamma_2(\tau)$ exist in $[0,T]$, i.e. under the setting of Proposition 3.3.3, we can obtain

$$\left|\left|\frac{\partial f_1}{\partial X}\right|\right|_2 \leq f_1 c_0(1 + ||X_s||_2),$$

for some constant c_0. Then

$$\left|\left|\frac{1}{f_1}\frac{\partial f_1}{\partial X}\right|\right|_2 \leq \frac{1}{f_1}\left|\left|\frac{\partial f_1}{\partial X}\right|\right|_2 \leq c_0(1 + ||X_s||_2).$$

In general, we have proved $h^{\beta,I}(s, W_s^{\widetilde{C}}, X_s)$ is subject to linear growth with respect to X_t. Therefore, by Lemma 3.3.1, $\mathcal{E}(t,h)$ is a martingale. Hence,

$$\begin{aligned}
& J(t, w^{\widetilde{C}}, X) \\
= \ & E_{t,w^{\widetilde{C}},X}\left[\int_t^T {}_{s-t}p_{x+t}\mu_{x+s}U\left(W_s^{\widetilde{C}} - \widetilde{C}(s,X_s) + \frac{I_s}{\mu_{x+s}}\right)ds + {}_{T-t}p_{x+t}U(W_T^{\widetilde{C}})\right] \\
= \ & E_{t,w^{\widetilde{C}},X}[g^{\beta,I}(T, W_T^{\widetilde{C}}, X_T)] \\
\leq \ & E_{t,w^{\widetilde{C}},X}\left[g^{\beta,I}(t, w^{\widetilde{C}}, X)\frac{\mathcal{E}(T,h^{\beta,I})}{\mathcal{E}(t,h^{\beta,I})}\right] \\
= \ & G(t, w^{\widetilde{C}}, X), \ \text{for } \forall(\beta, I) \in \mathcal{A}_\gamma(0,T).
\end{aligned} \tag{B.23}$$

Step 3: Verify $V(t, W_t^{\widetilde{C}}, X_t) = G(t, W_t^{\widetilde{C}}, X_t)$ under the optimal strategy (β^*, I^*).

The proof of this step is the same as that in Step 3 of Appendix B.5 except for part (a): Under $\gamma > 1$, we have $G(s, (W_s^{\widetilde{C}})^*, X_s) < 0$ for $\forall s \in [t,T]$. Since $U'(0) = +\infty, \gamma > 1$, we have $(W_u^{\widetilde{C}})^* - \widetilde{C}(u, X_u) + \frac{I_u^*}{\mu_{x+u}} > 0$ and $U\left((W_u^{\widetilde{C}})^* - \widetilde{C}(u,X_u) + \frac{I_u^*}{\mu_{x+u}}\right) < 0$ for $\forall u \in [t,s]$. Then, the following inequality holds by the definition of the value process (B.10)

$$\left|\frac{{}_{s-t}p_{x+t}G(s,(W_s^{\widetilde{C}})^*, X_s)}{g^{\beta^*,I^*}(s,(W_s^{\widetilde{C}})^*, X_s)}\right| \leq 1, \ \text{for } \forall s \in [t,T].$$

These complete the proof. $\square$

B.7 Estimation details for financial market

Denote $K_t = (X_{1,t}, X_{2,t}, \log \Pi_t, \log S_t)^\top$, then the underlying states in the financial market are given by

$$dK_t = (\theta_0 + \theta_1 K_t)dt + \Sigma_K dZ_t,$$

where

$$\theta_0 = \begin{pmatrix} 0_{2\times 1} \\ \delta_{\pi^e} - \frac{1}{2}\sigma_\Pi^\top\sigma_\Pi \\ \delta_R + \mu_0 - \frac{1}{2}\sigma_S^\top\sigma_S \end{pmatrix}, \theta_1 = \begin{pmatrix} -K_X & 0_{2\times 2} \\ e_2^\top & 0_{1\times 2} \\ \iota_2^\top - \sigma_\Pi^\top\Lambda_1 + \mu_1^\top & 0_{1\times 2} \end{pmatrix}, \Sigma_K = \begin{pmatrix} \Sigma_X \\ \sigma_\Pi^\top \\ \sigma_S^\top \end{pmatrix},$$

e_i represents the i-th unit vector in $\mathbb{R}^2$ and $\iota_2 = (1,1)^\top$. By Ito's formula, the transition equation for states follows

$$K_{t+\Delta t} = \Upsilon_1 + \Psi_1 K_t + \epsilon_{t+\Delta t}, \quad \epsilon_{t+\Delta t} \overset{i.i.d.}{\sim} N(0_{4\times 1}, \Sigma_\epsilon), \tag{B.24}$$

where

$$\Upsilon_1 = \int_0^{\Delta t} e^{\theta_1(\Delta t - s)}\theta_0 ds,\ \ \Psi_1 = e^{\theta_1 \Delta t},\ \ \Sigma_\epsilon = \int_0^{\Delta t} e^{\theta_1(\Delta t - s)}\Sigma_K \Sigma_K^\top (e^{\theta_1(\Delta t - s)})^\top ds.$$

For monthly data, we set $\Delta t = \frac{1}{12}$. Every month, there are 10 observations in the financial market: inflation index, equity index, and yield rate of nominal zero-coupon bonds with 8 maturities. Following Koijen et al. (2011), we also assume that the yield rates are observed with independent errors. Let $R^Y(t,\tau_i), i = 1,2,...,8$ denote the yield rates of nominal zero-coupon bonds at time t with maturity $\tau_i, i = 1,2,...,8$, then we have the measurement equation for the states

$$L_t = \Upsilon_2 + \Psi_2 K_t + \eta_t,\ \ \eta_t \overset{i.i.d.}{\sim} N(0_{10\times 1}, \Sigma_\eta), \tag{B.25}$$

where $L_t = (R^Y(t,\tau_i)_{i=1,2,...,8}, \log \Pi_t, \log S_t)^\top$ is the observation vector. Moreover, the coefficients in (B.25) are

$$\Upsilon_2 = \begin{pmatrix} -A_0(\tau_1)/\tau_1 \\ \vdots \\ -A_0(\tau_8)/\tau_8 \\ 0_{2\times 1} \end{pmatrix},\ \ \Psi_2 = \begin{pmatrix} -A_1^\top(\tau_1)/\tau_1 & 0_{1\times 2} \\ \vdots & \vdots \\ -A_1^\top(\tau_8)/\tau_8 & 0_{1\times 2} \\ 0_{2\times 2} & \widetilde{I}_2 \end{pmatrix},\ \ \Sigma_\eta = \begin{pmatrix} \chi_1 & & & & \\ & \ddots & & & \\ & & \chi_8 & & \\ & & & 0 & \\ & & & & 0 \end{pmatrix},$$

where A_0 and A_1 are given by (3.5) and (3.6) respectively, $\widetilde{I}_2$ is the 2nd-order identity matrix, and $\chi_i, i = 1,2,...,8$ are the constants to be estimated. Finally, with the transition equation (B.24) and the measurement equation (B.25), we can use the Kalman filter method to estimate the parameters. For more details, we refer to Babbs and Nowman (1999) and Harvey (1990).

Appendix C

Proofs for Chapter 4

C.1 Proof of Theorem 4.3.1

Proof. "Only if" part: "$\Rightarrow$"
By Ito's formula and equation (4.5), we have

$$
\begin{aligned}
&d(\beta_{v,t}e^{-\int_0^t \lambda_{x+s}ds}W_t) \\
&= \beta_{v,t}e^{-\int_0^t \lambda_{x+s}ds}(-v_{0,t}\alpha_t dt - \theta_t^\top v_{-,t}dt + \theta_t^\top \sigma_t dZ_{v,t} - c_t dt - \lambda_{x+t}M_t dt + Y_t dt - dC_t).
\end{aligned}
$$

Integrate on both hands sides, we obtain the inequality

$$
\begin{aligned}
&\beta_{v,t}e^{-\int_0^t \lambda_{x+s}ds}W_t - w_0 + \int_0^t \lambda_{x+s}\beta_{v,s}e^{-\int_0^t \lambda_{x+s}ds}M_s ds + \int_0^t \beta_{v,s}e^{-\int_0^t \lambda_{x+s}ds}(c_s - Y_s)ds \\
&\leq \int_0^t \beta_{v,s}e^{-\int_0^s \lambda_{x+u}du}\left[-(\alpha_s, \theta_s^\top)\begin{pmatrix} v_{0,s} \\ v_{-,s}\end{pmatrix}\right]ds + \int_0^t \beta_{v,s}e^{-\int_0^s \lambda_{x+u}du}\theta_s^\top \sigma_s dZ_{v,s}. \qquad \text{(C.1)}
\end{aligned}
$$

Moreover, by the definition of supporting function (4.10), together with the inequality (C.1), we arrive at the following inequality

$$
\begin{aligned}
&\beta_{v,t}e^{-\int_0^t \lambda_{x+s}ds}W_t - w_0 + \int_0^t \lambda_{x+s}\beta_{v,s}e^{-\int_0^t \lambda_{x+s}ds}M_s ds + \int_0^t \beta_{v,s}e^{-\int_0^t \lambda_{x+s}ds}(c_s - Y_s)ds \\
&\leq \int_0^t \beta_{v,s}e^{-\int_0^s \lambda_{x+u}du}\delta(v_s)ds + \int_0^t \beta_{v,s}e^{-\int_0^s \lambda_{x+u}du}\theta_s^\top \sigma_s dZ_{v,s}. \qquad \text{(C.2)}
\end{aligned}
$$

Define the stopping time $\tau_n = T \wedge \inf\{t \in [0,T] : \int_0^t |\theta_s^\top \sigma_s|^2 ds \geq n\}$ for $n \in \mathbb{N}_+$ and $\inf(\emptyset) = \infty$. Since the stochastic integral in (C.2) is a Q_v martingale in $[0, \tau_n]$, we have

$$E^{Q_v}\left[\beta_{v,\tau_n} e^{-\int_0^{\tau_n} \lambda_{x+s} ds} W_{\tau_n} + \int_0^{\tau_n} \lambda_{x+t} \beta_{v,t} e^{-\int_0^t \lambda_{x+s} ds} M_t dt + \int_0^{\tau_n} \beta_{v,t} e^{-\int_0^t \lambda_{x+s} ds}(c_t - Y_t) dt\right]$$
$$\leq w_0 + E^{Q_v}\left[\int_0^{\tau_n} \beta_{v,t} e^{-\int_0^t \lambda_{x+s} ds} \delta(v_t) dt\right]. \tag{C.3}$$

By the definition of admissible strategy (4.3), we have $\tau_n \nearrow T$ when $n \to \infty$. Because of $v_0 \geq 0$ in Assumption 4.2.3 and (4.13), we have the boundedness of the income process

$$E^{Q_v}\left[\int_0^T \beta_{v,t} e^{-\int_0^t \lambda_{x+s} ds} Y_t dt\right] \leq E^{Q_v}\left[\int_0^T \beta_{0,t} e^{-\int_0^t \lambda_{x+s} ds} Y_t dt\right] \leq K_y.$$

Therefore, the following equality holds by the monotone convergence theorem

$$\lim_{n\to\infty} E^{Q_v}\left[\int_0^{\tau_n} \beta_{v,t} e^{-\int_0^t \lambda_{x+s} ds}(c_t - Y_t) dt\right] = E^{Q_v}\left[\int_0^T \beta_{v,t} e^{-\int_0^t \lambda_{x+s} ds}(c_t - Y_t) dt\right].$$

According to Assumption 4.2.3, $\delta(v)$ is bounded above. Then, by the monotone convergence theorem, we have

$$\lim_{n\to\infty} E^{Q_v}\left[\int_0^{\tau_n} \beta_{v,t} e^{-\int_0^t \lambda_{x+s} ds} \delta(v_t) dt\right] = E^{Q_v}\left[\int_0^T \beta_{v,t} e^{-\int_0^t \lambda_{x+s} ds} \delta(v_t) dt\right].$$

As for the wealth term in (C.3), we derive from (4.6) and Assumption 4.2.1

$$(\beta_{v,\tau_n} e^{-\int_0^{\tau_n} \lambda_{x+s} ds} W_{\tau_n})^- \leq (\beta_{0,\tau_n} e^{-\int_0^{\tau_n} \lambda_{x+s} ds} W_{\tau_n})^- \leq K \exp\left(\int_0^T r_t^- dt\right) < \infty, \text{ P-a.s.}$$

for all n. Then, by Assumption 4.2.1 , we can use Fatou's lemma to show

$$\liminf_{n\to\infty} E^{Q_v}[\beta_{v,\tau_n} e^{-\int_0^{\tau_n} \lambda_{x+s} ds} W_{\tau_n}] \geq E^{Q_v}[\beta_{v,T} e^{-\int_0^T \lambda_{x+s} ds} W_T] \geq 0.$$

Finally, we derive

$$
\begin{aligned}
&E^{Q_v}\left[\beta_{v,T}e^{-\int_0^T \lambda_{x+t}dt}W_T + \int_0^T \lambda_{x+t}\beta_{v,t}e^{-\int_0^t \lambda_{x+s}ds}M_t dt + \int_0^T \beta_{v,t}e^{-\int_0^t \lambda_{x+s}ds}(c_t - Y_t)dt\right] \\
&\le \liminf_{n\to\infty} E^{Q_v}\left[\beta_{v,\tau_n}e^{-\int_0^{\tau_n} \lambda_{x+t}dt}W_{\tau_n} + \int_0^{\tau_n} \lambda_{x+t}\beta_{v,t}e^{-\int_0^t \lambda_{x+s}ds}M_t dt\right. \\
&\left.+ \int_0^{\tau_n} \beta_{v,t}e^{-\int_0^t \lambda_{x+s}ds}(c_t - Y_t)dt\right] \\
&\le w_0 + \liminf_{n\to\infty} E^{Q_v}\left[\int_0^{\tau_n} \beta_{v,t}e^{-\int_0^t \lambda_{x+s}ds}\delta(v_t)dt\right] \\
&= w_0 + E^{Q_v}\left[\int_0^T \beta_{v,t}e^{-\int_0^t \lambda_{x+s}ds}\delta(v_t)dt\right],
\end{aligned}
$$

where the second inequality comes from inequality (C.3). This completes the proof of "only if" part.

Next, we prove the "if" part: "$\Leftarrow$"
To show the inverse, we use $\mathcal{T}$ to denote the set of stopping times τ with $\tau \le T$, and for $\forall \tau \in \mathcal{T}$, define

$$
\begin{aligned}
\widehat{W}_\tau = \sup_{v\in\mathcal{N}^*} E^{Q_v}&\left[\int_\tau^T e^{-\int_\tau^t r_s+v_{0,s}+\lambda_{x+s}ds}[c_t - Y_t + \lambda_{x+t}M_t - \delta(v_t)]dt\right. \\
&\left.+e^{-\int_\tau^T r_s+v_{0,s}+\lambda_{x+s}ds}W_T|\mathcal{F}_\tau\right]. \qquad \text{(C.4)}
\end{aligned}
$$

Since $(c, M, W_T) \in G_+^*$, Assumption 4.2.3, and Assumption 4.3.1, we have

$$
\widehat{W}_\tau \quad \ge \quad -\sup_{v\in\mathcal{N}^*} E^{Q_v}\left[\int_\tau^T e^{-\int_\tau^t r_s+v_{0,s}+\lambda_{x+s}ds}Y_t dt\middle|\mathcal{F}_\tau\right] \ge -K_y, \qquad \text{(C.5)}
$$

which satisfies lower boundedness condition (4.6) of wealth process. Follow the same discussion in Cvitanić and Karatzas (1993), we have $\widehat{W}_t$ satisfies the dynamic programming principle

$$
\begin{aligned}
\widehat{W}_{\tau_1} = \sup_{v\in\mathcal{N}^*} E^{Q_v}&\left[\int_{\tau_1}^{\tau_2} e^{-\int_{\tau_1}^t r_s+v_{0,s}+\lambda_{x+s}ds}[c_t - Y_t + \lambda_{x+t}M_t - \delta(v_t)]dt\right. \\
&\left.+e^{-\int_{\tau_1}^{\tau_2} r_s+v_{0,s}+\lambda_{x+s}ds}\widehat{W}_{\tau_2}\middle|\mathcal{F}_{\tau_1}\right], \qquad \text{(C.6)}
\end{aligned}
$$

for all $\tau_1 \leq \tau_2$, $\tau_1, \tau_2 \in \mathcal{T}$. Setting $\tau_1 = t$, $\tau_2 = T$, and cancel out the supreme operator in (C.6), we derive

$$H_{v,t} = \beta_{v,t} e^{-\int_0^t \lambda_{x+s} ds} \widehat{W}_t + \int_0^t \beta_{v,s} e^{-\int_0^s \lambda_{x+u} du} [c_s - Y_s + \lambda_{x+s} M_s - \delta(v_s)] ds \tag{C.7}$$

is a Q_v-supermartingale for all $v \in \mathcal{N}^*$. By the Doob decomposition (see Theorem VII.12 in Dellacherie and Meyer (2011)) and the martingale representation theorem, for each $v \in \mathcal{N}^*$ there exists an increasing real valued process A_v and a $\mathbb{R}^n$-valued process Ψ_v with $\int_0^T |\Psi_{v,t}|^2 dt < \infty$ such that

$$H_{v,t} = \widehat{W}_0 + \int_0^t \Psi_{v,s}^\top dZ_{v,s} - A_{v,t}. \tag{C.8}$$

By the definition of $H_{v,t}$ (C.7), we have

$$\begin{aligned}
&\beta_{v,t}^{-1} e^{\int_0^t \lambda_{x+s} ds} \left(H_{v,t} - \int_0^t \beta_{v,s} e^{-\int_0^s \lambda_{x+u} du} [c_s - Y_s + \lambda_{x+s} M_s - \delta(v_s)] ds \right) \\
&= \widehat{W}_t = \beta_{0,t}^{-1} e^{\int_0^t \lambda_{x+s} ds} \left(H_{0,t} - \int_0^t \beta_{0,s} e^{-\int_0^s \lambda_{x+u} du} [c_s - Y_s + \lambda_{x+s} M_s] ds \right).
\end{aligned}$$

Then using Ito's formula and change of measure (4.12), we drive

$$\begin{aligned}
d\widehat{W}_t &= d\left[\beta_{v,t}^{-1} e^{\int_0^t \lambda_{x+s} ds} \left(H_{v,t} - \int_0^t \beta_{v,s} e^{-\int_0^s \lambda_{x+u} du} [c_s - Y_s + \lambda_{x+s} M_s - \delta(v_s)] ds \right) \right] \\
&= (r_t + v_{0,t} + \lambda_{x+t}) \widehat{W}_t dt + \beta_{v,t}^{-1} e^{\int_0^t \lambda_{x+s} ds} \Psi_{v,t}^\top [dZ_t + \sigma_t^{-1} (\mu_t + v_{-,t} - (r_t + v_{0,t}) \bar{1}_n) dt] \\
&- \beta_{v,t}^{-1} e^{\int_0^t \lambda_{x+s} ds} dA_{v,t} - [c_t - Y_t + \lambda_{x+t} M_t - \delta(v_t)] dt,
\end{aligned} \tag{C.9}$$

and

$$\begin{aligned}
d\widehat{W}_t &= d\left[\beta_{0,t}^{-1} e^{\int_0^t \lambda_{x+s} ds} \left(H_{0,t} - \int_0^t \beta_{0,s} e^{-\int_0^s \lambda_{x+u} du} [c_s - Y_s + \lambda_{x+s} M_s] ds \right) \right] \\
&= (r_t + \lambda_{x+t}) \widehat{W}_t dt + \beta_{0,t}^{-1} e^{\int_0^t \lambda_{x+s} ds} \Psi_{0,t}^\top [dZ_t + \sigma_t^{-1} (\mu_t - r_t \bar{1}_n) dt] \\
&- \beta_{0,t}^{-1} e^{\int_0^t \lambda_{x+s} ds} dA_{0,t} - [c_t - Y_t + \lambda_{x+t} M_t] dt.
\end{aligned} \tag{C.10}$$

Compare (C.9) and (C.10), we have

$$\beta_{v,t}^{-1} \Psi_{v,t}^\top = \beta_{0,t}^{-1} \Psi_{0,t}^\top, \tag{C.11}$$

$$\begin{aligned}
&\int_0^t \{ v_{0,s} \widehat{W}_s + \beta_{v,s}^{-1} e^{\int_0^s \lambda_{x+u} du} \Psi_{v,s}^\top \sigma_s^{-1} [v_{-,s} - v_{0,s} \bar{1}_n] + \delta(v_s) \} ds \\
&- \int_0^t \beta_{v,s}^{-1} e^{\int_0^s \lambda_{x+u} du} dA_{v,s} = - \int_0^t \beta_{0,s}^{-1} e^{\int_0^s \lambda_{x+u} du} dA_{0,s},
\end{aligned} \tag{C.12}$$

for all $v \in \mathcal{N}^*$ and all $t \in [0, T]$. Let

$$\theta_t^\top = \beta_{0,t}^{-1} e^{\int_0^t \lambda_{x+s} ds} \Psi_{0,t}^\top \sigma_t^{-1}, \ \alpha_t = \widehat{W}_t - \theta_t^\top \bar{1}_n \tag{C.13}$$

Substitute them into (C.10) and integrate, we derive

$$\begin{aligned}\widehat{W}_t &= w_0 + \int_0^t (\alpha_s r_s + \theta_s^\top \mu_s) ds + \int_0^t \theta_s^\top \sigma_s dZ_s - \int_0^t (c_s + I_s - Y_s) ds \\ &\quad -(w_0 - \widehat{W}_0 + \int_0^t e^{\int_0^s r_u + \lambda_{x+u} du} dA_{0,s}) \\ &:= w_0 + \int_0^t (\alpha_s r_s + \theta_s^\top \mu_s) ds + \int_0^t \theta_s^\top \sigma_s dZ_s - \int_0^t (c_s + I_s - Y_s) ds - C_t,\end{aligned}$$

which is the same as the dynamic budget constraint (4.5) and C_t is the free disposal equals

$$C_t = w_0 - \widehat{W}_0 + \int_0^t e^{\int_0^s r_u + \lambda_{x+u} du} dA_{0,s}.$$

Finally, we only need to prove $(\alpha, \theta) \in A$ for the trading strategy (C.13). Substituting $\widehat{W}_t = \alpha_t + \theta_t^\top \bar{1}_n$ into (C.12), we can derive

$$\int_0^t \alpha_s v_{0,s} + \theta_s^\top v_{-,s} + \delta(v_s) ds + \int_0^t \beta_{0,t}^{-1} dA_{0,s} = \int_0^t \beta_{v,s}^{-1} e^{\int_0^s \lambda_{x+u} du} dA_{v,s} \geq 0.$$

Since $v \in \mathcal{N}^*$ is arbitrage, $\widetilde{A}$ is a convex cone, and δ is positive homogeneous, if there exists some (v_0, v_-) such that $\alpha_s v_{0,s} + \theta_s^\top v_{-,s} + \delta(v_s) < 0$, then $\alpha_s b v_{0,s} + \theta_s^\top b v_{-,s} + \delta(b v_s)$ can be any negative number for $b > 0$, which contradicts

$$\int_0^t \alpha_s v_{0,s} + \theta_s^\top v_{-,s} + \delta(v_s) ds + \int_0^t \beta_{0,t}^{-1} dA_{0,s} \geq 0.$$

Therefore, there exists a set E having full $(\bar{\lambda} \times P)$ measure (where $(\bar{\lambda} \times P)$ is product measure on $[0, T] \times \Omega$) such that

$$\delta(v) + \alpha(t, \omega) v_0 + \theta(t, \omega)^\top v_- \geq 0, \forall (t, \omega) \in E, v \in \widetilde{A}.$$

(see Step 3 of Theorem 9.1 in Cvitanić and Karatzas (1992)). By Theorem 13.1 in Rockafellar (1970), we derive $(\alpha, \theta) \in A$, $(\bar{\lambda} \times P)$-a.s. $\square$

C.2 Proof of Corollary 4.3.1

Proof. The proof is similar to the "⇐" part of Appendix C.1. According to the formula of $W_{v^*,t}$ in (4.15), we obtain

$$H_{v^*,t} = \beta_{v^*,t} e^{-\int_0^t \lambda_{x+s} ds} W_{v^*,t} + \int_0^t \beta_{v^*,s} e^{-\int_0^s \lambda_{x+u} du} [c_s - Y_s + \lambda_{x+s} M_s - \delta(v_s^*)] ds \quad \text{(C.14)}$$

is a Q_v-martingale for $v^* \in \mathcal{N}^*$. Then by martingale presentation theorem, there exists a $\mathbb{R}^n$-valued process Ψ_v with $\int_0^T |\Psi_{v,t}|^2 dt < \infty$, such that

$$H_{v^*,t} = W_{v^*,0} + \int_0^t \Psi_{v^*,s}^\top dZ_{v^*,s}. \quad \text{(C.15)}$$

Substitute (C.15) into (C.14), we derive

$$\begin{aligned} W_{v^*,t} &= \beta_{v^*,t}^{-1} e^{\int_0^t \lambda_{x+s} ds} \left\{ H_{v^*,t} - \int_0^t \beta_{v^*,s} e^{-\int_0^s \lambda_{x+u} du} [c_s - Y_s + \lambda_{x+s} M_s - \delta(v_s^*)] ds \right\} \\ &= \beta_{v^*,t}^{-1} e^{\int_0^t \lambda_{x+s} ds} \left\{ W_{v^*,0} + \int_0^t \Psi_{v^*,s}^\top dZ_{v^*,s} \right. \\ &\quad \left. - \int_0^t \beta_{v^*,s} e^{-\int_0^s \lambda_{x+u} du} [c_s - Y_s + \lambda_{x+s} M_s - \delta(v_s^*)] ds \right\}. \end{aligned}$$

By Ito's formula and change of measure (4.12), we obtain

$$\begin{aligned} dW_{v^*,t} &= (r_t + v_{0,t}^* + \lambda_{x+t}) W_{v^*,t} dt \\ &\quad + \beta_{v^*,t}^{-1} e^{\int_0^t \lambda_{x+s} ds} \Psi_{v^*,t}^\top [dZ_t + \sigma_t^{-1} (\mu_t + v_{-,t}^* - (r_t + v_{0,t}^*) \bar{1}_n) dt] \\ &\quad - [c_t - Y_t + \lambda_{x+t} M_t - \delta(v_t^*)] dt. \end{aligned} \quad \text{(C.16)}$$

Since $\Psi_{v^*,t}^\top = \beta_{v^*,t} e^{-\int_0^t \lambda_{x+s} ds} \theta_t^\top \sigma_t$ and $M_t = W_t + \frac{I_t}{\lambda_{x+t}}$, (C.16) can be simplified to

$$dW_{v^*,t} = (r_t \alpha_t + \theta_t^\top \mu_t) dt + [\alpha_t v_{0,t}^* + \theta_t^\top v_{-,t}^* + \delta(v_t^*)] dt + \theta_t^\top \sigma_t dZ_t - (c_t + I_t - Y_t) dt,$$

which has no free disposal. Next, we only need to prove

1. $(\alpha_t, \theta_t) \in A$.
2. $\alpha_t v_{0,t}^* + \theta_t^\top v_{-,t}^* + \delta(v_t^*) = 0$, $\bar{\lambda} \times P$-a.s.

Before moving forward, we first fix an arbitrary $v \in \mathcal{N}$ and define

$$\zeta_t = \int_0^t (v_{0,s}^* - v_{0,s})ds + \int_0^t (v_{-,s}^* - v_{-,s} - (v_{0,s}^* - v_{0,s})\bar{1}_n)^\top \sigma_s^{-1} dZ_{v^*,s},$$

also the sequence of stopping times

$$\begin{aligned}\tau_n = T \wedge \inf \Big\{ & t \in [0,T] : |\zeta_t| + |\pi_{v^*,t}| + |W_{v^*,t}| \geq n, \\ \text{or} & \int_0^t |\theta_s^\top \sigma_s|^2 ds \geq n, \\ \text{or} & \int_0^t |v_{0,s}^* - v_{0,s}| ds \geq n, \\ \text{or} & \int_0^t |\sigma_s^{-1}(v_{-,s}^* - v_{-,s} - (v_{0,s}^* - v_{0,s})\bar{1}_n)|^2 ds \geq n \Big\}.\end{aligned}$$

Then $\tau_n \nearrow T$ almost everywhere. To conduct the calculus of variations, we add a perturbation $v_t \in \mathcal{N}$ to the optimal v_t^* and define

$$v_{\epsilon,n,t} = v_t^* + \epsilon(v_t - v_t^*)\mathbb{1}_{\{t \leq \tau_n\}} \text{ for } \epsilon \in (0,1).$$

By the convexity of $\widetilde{A}$, we have $v_{\epsilon,n} \in \mathcal{N}$, and the pricing kernel under $v_{\epsilon,n,t}$ is given by

$$\begin{aligned}\pi_{v_{\epsilon,n,t}} &= \pi_{v^*,t} \exp\left(\epsilon \zeta_{t\wedge\tau_n} - \frac{\epsilon^2}{2}\int_0^{t\wedge\tau_n} |\sigma_s^{-1}(v_{-,s}^* - v_{-,s} - (v_{0,s}^* - v_{0,s})\bar{1}_n)|^2 ds\right) \\ &:= \pi_{v^*,t} \exp\left(\epsilon \zeta_{t\wedge\tau_n} - \frac{\epsilon^2}{2}\int_0^{t\wedge\tau_n} K_s^2 ds\right).\end{aligned}$$

Together with the definition of stopping times τ_n, we have

$$\begin{aligned} e^{-2\epsilon n}\pi_{v^*,t} \leq \pi_{v_{\epsilon,n},t} \leq e^{2\epsilon n}\pi_{v^*,t}, \\ e^{-3\epsilon n}\xi_{v^*,t} \leq \xi_{v_{\epsilon,n},t} \leq e^{3\epsilon n}\xi_{v^*,t}. \end{aligned} \tag{C.17}$$

Therefore, $\xi_{v_{\epsilon,n}}$ is of class D, and hence $v_{\epsilon,n} \in \mathcal{N}^*$ (see Proposition I.1.47 in Jacod and Shiryaev (2013)). Define two wealth processes

$$\begin{aligned} W_n(\epsilon) &= E\left[\int_0^T e^{-\int_0^t \lambda_{x+s}ds}\pi_{v_{\epsilon,n},t}[c_t - Y_t + \lambda_{x+t}M_t - \delta(v_{\epsilon,n,t})]dt + e^{-\int_0^T \lambda_{x+s}ds}\pi_{v_{\epsilon,n},T}W_T\right] \\ W_n(0) &= E\left[\int_0^T e^{-\int_0^t \lambda_{x+s}ds}\pi_{v^*,t}[c_t - Y_t + \lambda_{x+t}M_t - \delta(v_t^*)]dt + e^{-\int_0^T \lambda_{x+s}ds}\pi_{v^*,T}W_T\right]. \end{aligned}$$

From inequality (C.17), we derive

$$\begin{aligned}
&\left| e^{-\int_0^t \lambda_{x+s} ds} \frac{\pi_{v_{\epsilon,n},t} - \pi_{v^*,t}}{\epsilon} (c_t - Y_t + \lambda_{x+t} M_t - \delta(v_t^*)) \right| \\
&\leq \bar{K}_n \pi_{v^*,t} (c_t + Y_t + \lambda_{x+t} M_t - \delta(v_t^*)), \\
&e^{-\int_0^T \lambda_{x+s} ds} W_T \left| \frac{\pi_{v_{\epsilon,n},T} - \pi_{v^*,T}}{\epsilon} \right| \leq \bar{K}_n \pi_{v^*,T} W_T,
\end{aligned}$$

where

$$\bar{K}_n = \sup_{\epsilon \in (0,1)} \frac{e^{2\epsilon n} - 1}{\epsilon} < \infty.$$

Moreover, for the supporting function, we have

$$\pi_{v_{\epsilon,n},t} [\delta(v_t^*) - \delta(v_t)]^- \leq -e^{2n} \pi_{v^*,t} \delta(v_t^*).$$

Then by Lebesgue's dominated convergence theorem, convexity of $\delta(v)$, and Fatou's lemma, we have

$$\begin{aligned}
&\lim_{\epsilon \searrow 0} \frac{W_n(\epsilon) - W_n(0)}{\epsilon} \\
&= \lim_{\epsilon \searrow 0} E \left[\int_0^T e^{-\int_0^t \lambda_{x+s} ds} \frac{\pi_{v_{\epsilon,n},t} - \pi_{v^*,t}}{\epsilon} (c_t - Y_t + \lambda_{x+t} M_t) dt \right. \\
&\left. + \int_0^T e^{-\int_0^t \lambda_{x+s} ds} \frac{-\pi_{v_{\epsilon,n},t} \delta(v_{\epsilon,n,t}) + \pi_{v^*,t} \delta(v_t^*)}{\epsilon} dt + e^{-\int_0^T \lambda_{x+t} dt} W_T \frac{\pi_{v_{\epsilon,n},T} - \pi_{v^*,T}}{\epsilon} \right] \\
&= \lim_{\epsilon \searrow 0} E \left[\int_0^T e^{-\int_0^t \lambda_{x+s} ds} \frac{1}{\epsilon} \left\{ (\pi_{v_{\epsilon,n},t} - \pi_{v^*,t})(c_t - Y_t + \lambda_{x+t} M_t) \right. \right. \\
&\left. + \pi_{v^*,t} \delta(v_t^*) - \pi_{v_{\epsilon,n},t} \delta(v_t^*) + \pi_{v_{\epsilon,n},t} \delta(v_t^*) - \pi_{v_{\epsilon,n},t} \delta(v_{\epsilon,n,t}) \right\} dt \\
&\left. + e^{-\int_0^T \lambda_{x+s} ds} W_T \frac{\pi_{v_{\epsilon,n},T} - \pi_{v^*,T}}{\epsilon} \right] \\
&= \lim_{\epsilon \searrow 0} E \left[\int_0^T e^{-\int_0^t \lambda_{x+s} ds} \frac{1}{\epsilon} (\pi_{v_{\epsilon,n},t} - \pi_{v^*,t})(c_t - Y_t + \lambda_{x+t} M_t - \delta(v_t^*)) dt \right. \\
&\left. + e^{-\int_0^T \lambda_{x+s} ds} W_T \frac{\pi_{v_{\epsilon,n},T} - \pi_{v^*,T}}{\epsilon} \right] \\
&+ \lim_{\epsilon \searrow 0} E \left[\int_0^T \frac{1}{\epsilon} e^{-\int_0^t \lambda_{x+s} ds} \pi_{v_{\epsilon,n},t} \{ \delta(v_t^*) - \delta(v_t^* + \epsilon(v_t - v_t^*) \mathbb{1}_{\{t \leq \tau_n\}}) \} dt \right] \\
&\geq \lim_{\epsilon \searrow 0} E \left[\int_0^T e^{-\int_0^t \lambda_{x+s} ds} \pi_{v^*,t} (c_t - Y_t + \lambda_{x+t} M_t - \delta(v_t^*)) \frac{1}{\epsilon} \left(e^{\epsilon \zeta_{t \wedge \tau_n} - \frac{\epsilon^2}{2} \int_0^{t \wedge \tau_n} |K_s|^2 ds} - 1 \right) dt \right.
\end{aligned}$$

$$
\begin{aligned}
&+e^{-\int_0^T \lambda_{x+t}dt}\pi_{v^*,T}W_T\frac{1}{\epsilon}\left(e^{\epsilon\zeta_{T\wedge\tau_n}-\frac{\epsilon^2}{2}\int_0^{T\wedge\tau_n}|K_s|^2ds}-1\right)\bigg]\\
&+\lim_{\epsilon\searrow 0}E\left[\int_0^{\tau_n}\frac{1}{\epsilon}e^{-\int_0^t\lambda_{x+s}ds}\pi_{v_{\epsilon,n},t}\{\delta(v_t^*)-(1-\epsilon)\delta(v_t^*)-\epsilon\delta(v_t)\}dt\right]\\
&=E\left[\int_0^T e^{-\int_0^t\lambda_{x+s}ds}(c_t-Y_t+\lambda_{x+t}M_t-\delta(v_t^*))\pi_{v^*,t}\zeta_{t\wedge\tau_n}dt\right.\\
&\left.+e^{-\int_0^T\lambda_{x+t}dt}W_T\pi_{v^*,T}\zeta_{T\wedge\tau_n}\right]+E\left[\int_0^{\tau_n}e^{-\int_0^t\lambda_{x+s}ds}\pi_{v^*,t}[\delta(v_t^*)-\delta(v_t)]dt\right]\\
&=E\left[\int_0^{\tau_n}e^{-\int_0^t\lambda_{x+s}ds}(c_t-Y_t+\lambda_{x+t}M_t-\delta(v_t^*))\pi_{v^*,t}\zeta_t dt\right.\\
&\left.+\int_{\tau_n}^T e^{-\int_0^t\lambda_{x+s}ds}(c_t-Y_t+\lambda_{x+t}M_t-\delta(v_t^*))\pi_{v^*,t}dt\zeta_{\tau_n}+\zeta_{\tau_n}e^{-\int_0^T\lambda_{x+s}ds}\pi_{v^*,T}W_T\right]\\
&+E\left[\int_0^{\tau_n}e^{-\int_0^t\lambda_{x+s}ds}\pi_{v^*,t}[\delta(v_t^*)-\delta(v_t)]dt\right]\\
&=E\left[\int_0^{\tau_n}e^{-\int_0^t\lambda_{x+s}ds}(c_t-Y_t+\lambda_{x+t}M_t-\delta(v_t^*))\pi_{v^*,t}\zeta_t dt\right.\\
&\left.+\zeta_{\tau_n}\pi_{v^*,\tau_n}e^{-\int_0^{\tau_n}\lambda_{x+s}ds}W_{v^*,\tau_n}\right]+E\left[\int_0^{\tau_n}e^{-\int_0^t\lambda_{x+s}ds}\pi_{v^*,t}[\delta(v_t^*)-\delta(v_t)]dt\right].
\end{aligned}
\tag{C.18}
$$

For $t\le\tau_n$, by Ito's formula, we have

$$
\begin{aligned}
&\beta_{v^*,t}\zeta_t e^{-\int_0^t\lambda_{x+s}ds}W_t+\int_0^t e^{-\int_0^s\lambda_{x+u}du}[c_s-Y_s+\lambda_{x+s}M_s-\delta(v_s^*)]\beta_{v^*,s}\zeta_s ds\\
&=\int_0^t\beta_{v^*,s}e^{-\int_0^s\lambda_{x+u}du}[\alpha_s(v_{0,s}^*-v_{0,s})+\theta_s^\top(v_{-,s}^*-v_{-,s})]ds\\
&+\int_0^t\beta_{v^*,s}e^{-\int_0^s\lambda_{x+u}du}\left[\zeta_s\theta_s^\top\sigma_s\right.\\
&\left.+W_{v^*,s}(v_{-,s}^*-v_{-,s}-(v_{0,s}^*-v_{0,s})\bar{1}_n)^\top\sigma_s^{-1}\right]dZ_{v^*,s}.
\end{aligned}
\tag{C.19}
$$

Plug (C.19) into (C.18), we drive

$$
\begin{aligned}
&\lim_{\epsilon\searrow 0}\frac{W_n(\epsilon)-W_n(0)}{\epsilon}\\
&\ge E\left[\int_0^{\tau_n}e^{-\int_0^t\lambda_{x+s}ds}(c_t-Y_t+\lambda_{x+t}M_t-\delta(v_t^*))\pi_{v^*,t}\zeta_t dt\right.\\
&\left.+\zeta_{\tau_n}\pi_{v^*,\tau_n}e^{-\int_0^{\tau_n}\lambda_{x+s}ds}W_{v^*,\tau_n}\right]+E\left[\int_0^{\tau_n}e^{-\int_0^t\lambda_{x+s}ds}\pi_{v^*,t}[\delta(v_t^*)-\delta(v_t)]dt\right]
\end{aligned}
$$

$$= E\left[\int_0^{\tau_n} e^{-\int_0^t \lambda_{x+s}ds}\pi_{v^*,t}\{\alpha_t(v^*_{0,t} - v_{0,t}) + \theta_t^\top(v^*_{-,t} - v_{-,t}) + \delta(v^*_t) - \delta(v_t)\}dt\right] \tag{C.20}$$

Let $v = v^* + \rho$, $\rho \in \mathcal{N}$, since $\widetilde{A}$ is a convex cone, we have $v \in \mathcal{N}$. Substitute $v = v^* + \rho$ into (C.20), we have

$$E\left[\int_0^{\tau_n} e^{-\int_0^t \lambda_{x+s}ds}\pi_{v^*,t}[\alpha_t\rho_{0,t} + \theta_t^\top\rho_{-,t} + \delta(\rho_t)]dt\right] \geq 0.$$

Since $\rho \in \mathcal{N}$ is arbitrage, this implies the existence of a set $\mathcal{D}$ having full $(\bar{\lambda} \times P)$ measure that

$$\alpha(t,\omega)v_0 + \theta^\top(t,\omega)v_- + \delta(v) \geq 0,\ \forall(t,\omega) \in \mathcal{D},\ v \in \widetilde{A}. \tag{C.21}$$

From Theorem 13.1 in Rockafellar (1970), it implies

$$(\alpha_t, \theta_t) \in A,\ (\bar{\lambda} \times P)\text{-a.s.}$$

Let $v \equiv 0$, we have

$$0 \geq E\left[\int_0^{\tau_n} e^{-\int_0^t \lambda_{x+s}ds}\pi_{v^*,t}[\alpha_t v^*_{0,t} + \theta_t^\top v^*_{-,t} + \delta(v^*_t)]dt\right],$$

together with (C.21), we have

$$\alpha_t v^*_{0,t} + \theta_t^\top v^*_{-,t} + \delta(v^*_t) = 0, \bar{\lambda} \times P\text{-a.s.}$$

Finally, since $(c, M, W_T) \in G^*_+$, income constraint (4.13), and Assumption 4.2.3, we have $W_{v^*,t}$ bounded below. Moreover, the optimal wealth $W_{v^*,t}$ satisfies $W_{v^*,0} = w_0$ and $W_{v^*,T} = W_T$.

From all things above, we have proved that (c, M, W_T) is feasible, which completes the proof. □

C.3 Proof of Lemma 4.4.1

Proof. By the definition of $\widetilde{U}_1$, we have

$$\widetilde{U}_1(z,t) = \sup_{c\geq 0}\{U_1(c,t) - zc\} = U_1(c^*,t) - zc^*, \quad z > 0$$

where c^* is the optimal consumption satisfying

$$U_1'(c^*,t) - z = 0, z > 0. \tag{C.22}$$

Then, we have $c^* > 0$ because U_1 satisfies Inada condition (4.16), U_1 is strictly concave with the first variable by Definition 4.2.1, and $z > 0$. Moreover, by (C.22), we have the optimal c^* is a function of z. Next, by the law of implicit differentiation, we can derive the first-order and second-order partial derivatives of $\widetilde{U}_1$ with respect to z

$$\frac{\partial \widetilde{U}_1(z,t)}{\partial z} = U_1'(c^*,t)\frac{\partial c^*}{\partial z} - c^* - z\frac{\partial c^*}{\partial z} = -c^* < 0, \tag{C.23}$$

$$\frac{\partial^2 \widetilde{U}_1(z,t)}{\partial z^2} = -\frac{\partial c^*}{\partial z} = -\frac{\partial U_1'^{-1}(z,t)}{\partial z} = -\frac{1}{U_1''(U_1'^{-1}(z,t),t)} = -\frac{1}{U_1''(c^*,t)} > 0. \tag{C.24}$$

Therefore, $\widetilde{U}_1(z,t)$ is strictly decreasing and strictly convex in its first variable. The same arguments can be applied to $\widetilde{U}_2$ and $\widetilde{U}_3$.

The representation (4.19) is a direct result by substituting c^* in (4.18) into (4.17). The same arguments are for $\widetilde{U}_2$ and $\widetilde{U}_3$.

For $i = 1, 2, 3$, by the Inada condition (4.16)

$$U_i'(0+,t) = \infty, U_i'(\infty,t) = 0+, \text{ for } \forall t \in [0,T],$$

we have

$$U_i'^{-1}(0+,t) = \infty, U_i'^{-1}(\infty,t) = 0+, \text{ for } \forall t \in [0,T].$$

i.e.

$$f_i(0+,t) = \infty, f_i(\infty,t) = 0+, \text{ for } \forall t \in [0,T].$$

When z goes to infinity, we have

$$\begin{aligned}
\widetilde{U}_i(\infty,t) &\leq U_i(f_i(\infty,t),t) = U_i(0+,t) \\
\widetilde{U}_i(\infty,t) &\geq \lim_{z\to\infty}\left[U\left(\frac{\epsilon}{z},t\right) - \epsilon\right] = U_i(0+,t) - \epsilon, \forall \epsilon > 0.
\end{aligned}$$

Therefore, $\widetilde{U}_i(\infty, t) = U_i(0+, t)$.
The inverse transform from the dual utility to the primal utility is

$$U_i(x,t) = \inf_{y>0}[\widetilde{U}_i(y,t) + xy] = \widetilde{U}_i(U_i'(x,t),t) + xU_i'(x,t).$$

Next, we can derive

$$\begin{aligned} U_i(\infty,t) &\geq \widetilde{U}_i(U_i'(\infty,t),t) = \widetilde{U}_i(0+,t) \\ U_i(\infty,t) &\leq \lim_{x\to\infty}\left[\widetilde{U}_i\left(\frac{\epsilon}{x},t\right) + \epsilon\right] = \widetilde{U}_i(0+,t) + \epsilon,\ \forall\epsilon > 0. \end{aligned}$$

Thus, $\widetilde{U}_i(0+, t) = U_i(\infty, t)$, which completes the proof. □

C.4 Proof of Theorem 4.4.1

Proof. Assume that $(\psi^*, v^*) \in (0,\infty) \times \mathcal{N}^*$ solves Problem (D) and constraint (4.20) holds. To prove (c^*, M^*, W_T^*) in (4.21) is A-feasible optimal, we need to check two things:

1. $J(c^*, M^*, W_T^*) \geq J(c, M, W_T)$ for $\forall(c, M, W_T) \in \mathcal{B}(\mathcal{P}, A)$,
2. $(c^*, M^*, W_T^*) \in \mathcal{B}(\mathcal{P}, A)$.

We divide the proof into three steps.

Step 1: Applying $f_i(\cdot, t)$ on both hands sides of (4.20), we have for $\forall\beta \in (0,\infty)$, $\gamma \in (0,\infty)$,

$$f_i(\beta y, t) \leq \gamma f_i(y,t),\ i = 1,2,3,\ \forall(y,t) \in (0,\infty) \times [0,T]. \tag{C.25}$$

By Assumption 4.2.3, supporting function δ is bounded above on $\widetilde{A}$, then (C.25) and (4.21) imply

$$
\begin{aligned}
&E\left[\int_0^T \pi_{v^*,t} e^{-\int_0^t \lambda_{x+s}ds}(f_1(\psi\pi_{v^*,t}) + \lambda_{x+t} f_2(\psi\pi_{v^*,t}))dt\right.\\
&\left.+\pi_{v^*,T} e^{-\int_0^T \lambda_{x+t}dt} f_3(\psi\pi_{v^*,T})\right]\\
&\leq E\left[\int_0^T \pi_{v^*,t} e^{-\int_0^t \lambda_{x+s}ds}\left[f_1\left(\frac{\psi}{\psi^*}\psi^*\pi_{v^*,t}\right) + \lambda_{x+t} f_2\left(\frac{\psi}{\psi^*}\psi^*\pi_{v^*,t}\right)\right]dt\right.\\
&\left.+\pi_{v^*,T} e^{-\int_0^T \lambda_{x+t}dt} f_3\left(\frac{\psi}{\psi^*}\psi^*\pi_{v^*,T}\right)\right]\\
&\leq c_0 E\left[\int_0^T \pi_{v^*,t} e^{-\int_0^t \lambda_{x+s}ds}\left[f_1\left(\psi^*\pi_{v^*,t}\right) + \lambda_{x+t} f_2\left(\psi^*\pi_{v^*,t}\right)\right]dt\right.\\
&\left.+\pi_{v^*,T} e^{-\int_0^T \lambda_{x+t}dt} f_3\left(\psi^*\pi_{v^*,T}\right)\right]\\
&< \infty,
\end{aligned}
$$

for a constant $c_0 \in (0, \infty)$ and $\forall \psi \in (0, \infty)$. By the optimality of ψ^*, we have

$$
\begin{aligned}
0 &= \lim_{\epsilon\to 0} \frac{\widetilde{J}(\psi^* + \epsilon, v^*) - \widetilde{J}(\psi^*, v^*)}{\epsilon}\\
&= E\left\{\int_0^T e^{-\int_0^t \lambda_{x+s}ds} \lim_{\epsilon\to 0}\left\{\frac{\widetilde{U}_1((\psi^* + \epsilon)\pi_{v^*,t}, t) - \widetilde{U}_1(\psi^*\pi_{v^*,t}, t)}{\epsilon}\right.\right.\\
&\quad \left. +\lambda_{x+t}\left[\frac{\widetilde{U}_2((\psi^* + \epsilon)\pi_{v^*,t}, t) - \widetilde{U}_2(\psi^*\pi_{v^*,t}, t)}{\epsilon}\right]\right\} dt\\
&\quad e^{-\int_0^T \lambda_{x+t}dt} \lim_{\epsilon\to 0} \frac{\widetilde{U}_3((\psi^* + \epsilon)\pi_{v^*,T}, T) - \widetilde{U}_3(\psi^*\pi_{v^*,T}, T)}{\epsilon}\\
&\quad \left. +w_0 + \int_0^T e^{-\int_0^t \lambda_{x+s}ds}\pi_{v^*,t}[Y_t + \delta(v_t)]dt\right\}\\
&= w_0 - E\left[\int_0^T e^{-\int_0^t \lambda_{x+s}ds}\pi_{v^*,t}(c_t^* + \lambda_{x+t}M_t^* - Y_t - \delta(v_t^*))dt\right.\\
&\quad \left. +e^{-\int_0^T \lambda_{x+t}dt} W_T^*\right]. \qquad \text{(C.26)}
\end{aligned}
$$

The second equality comes from Lebesgue's dominated convergence theorem, where

$$
\left|\frac{\widetilde{U}_i((\psi^* + \epsilon)\pi_{v^*,t}, t) - \widetilde{U}_i(\psi^*\pi_{v^*,t}, t)}{\epsilon}\right|
$$

$$
\begin{aligned}
&\leq \frac{\widetilde{U}_i((\psi^* - |\epsilon|)\pi_{v^*,t}, t) - \widetilde{U}_i(\psi^* \pi_{v^*,t})}{|\epsilon|} \\
&\leq \pi_{v^*,t} f((\psi^* - |\epsilon|)\pi_{v^*,t}, t) \\
&\leq \pi_{v^*,t} f((\psi^*/2)\pi_{v^*,t}, t),
\end{aligned}
$$

for $|\epsilon| < \frac{\psi^*}{2}$. These inequalities are based on the fact that $\widetilde{U}_i$ is decreasing and convex, hence $f(z,t) = -\frac{\partial \widetilde{U}_i}{\partial z}$ is also decreasing. By the concavity of $U_i, i = 1,2,3$, we have

$$
U_i(f_i(z,t),t) - U_i(c,t) \geq z[f_i(z,t) - c], \ \forall c > 0, z > 0,
$$

together with the static budget constraint (4.14) and (C.26), the following equality holds

$$
\begin{aligned}
&J(c^*, M^*, W_T^*) - J(c, M, W_T) \\
&= \psi^* E\left[\int_0^T e^{-\int_0^t \lambda_{x+s} ds} \pi_{v^*,t}[c_t^* - c_t + \lambda_{x+t}(M_t^* - M_t)]dt \right. \\
&\left. + e^{-\int_0^T \lambda_{x+t} dt} \pi_{v^*,T}(W_T^* - W_T)\right] \geq 0.
\end{aligned}
$$

Then, the optimality of (c^*, M^*, W_T^*) is proved.

Step 2: By the continuity of f_i and $\pi_{v^*,t}$, it is clear that

$$
\int_0^T c_t^* + M_t^* dt + W_T^* < \infty, \text{P-a.s.}
$$

Moreover, from the inequality

$$
U_1(1,t) - z \leq \max_{c \geq 0}\{U_1(c,t) - zc\} = U_1(f_1(z,t),t) - z f_1(z,t),
$$

we have

$$
E\left[\int_0^T U_1(c_t^*, t)^- dt\right] \leq \int_0^T U_1(1,t)^- dt + \psi^* E\left[\int_0^T \pi_{v^*,t} dt\right] < \infty.
$$

Similar to $U_2(M_t^*, t)^-$ and $U_2(W_T^*, T)^-$. Therefore, $(c^*, M^*, W_T^*) \in G_+^*$. Next, we only need to show there exists a $(\alpha, \theta) \in A$ financing (c^*, M^*, W_T^*).

Define the wealth process W_t by

$$
\begin{aligned}
W_t &= (\pi_{v^*,t} \cdot {}_t p_x)^{-1} E\left[\int_t^T \pi_{v^*,s} \cdot {}_s p_x [c_s^* + \lambda_{x+s} M_s^* - Y_s - \delta(v_s^*)] ds \right. \\
&\quad \left. + \pi_{v^*,T} \cdot {}_T p_x W_T^* | \mathcal{F}_t\right] \\
&= (\beta_{v^*,t} \cdot {}_t p_x)^{-1} E^{Q_v}\left[\int_t^T \beta_{v^*,s} \cdot {}_s p_x [c_s^* + \lambda_{x+s} M_s^* - Y_s - \delta(v_s^*)] ds \right. \\
&\quad \left. + \beta_{v^*,T} \cdot {}_T p_x W_T^* | \mathcal{F}_t\right],
\end{aligned}
$$

then by (4.13) and (4.21), we have the expectation in W_t is finite. Moreover, $W_T = W_T^*$, W_t is bounded below by (4.13) and Assumption 4.2.3, and $W_0 = w_0$ by (C.26). Next, by using martingale representation theorem, there exists a process Ψ with $\int_0^T |\Psi_t|^2 dt < \infty$ a.s. such that

$$\beta_{v^*,t} \cdot {}_tp_x W_t + \int_0^t \beta_{v^*,s} \cdot {}_sp_x [c_s^* + \lambda_{x+s} M_s^* - Y_s - \delta(v_s^*)] ds = w_0 + \int_0^t \Psi_s^\top dZ_{v^*,s}. \tag{C.27}$$

Define the trading strategy $(\alpha, \theta) \in \Theta$ by

$$\theta_t^\top = (\beta_{v^*,t} \cdot {}_tp_x)^{-1} \Psi_t^\top \sigma_t^{-1}, \alpha_t = W_t - \theta_t^\top \bar{1}_n.$$

Using (C.27), we derive

$$W_t = (\beta_{v^*,t} \cdot {}_tp_x)^{-1} \left[w_0 + \int_0^t \Psi_s^\top dZ_{v^*,s} - \int_0^t \beta_{v^*,s} \cdot {}_sp_x (c_s^* + \lambda_{x+s} M_s^* - Y_s - \delta(v_s^*)) ds \right]$$

By Ito's formula, W_t satisfies following SDE

$$dW_t = (r_t \alpha_t + \theta_t^\top \mu_t) dt + [v_{0,t}^* \alpha_t + \theta_t^\top v_{-,t}^* + \delta(v_t^*)] dt + \theta_t^\top \sigma_t dZ_t - (c_t^* + I_t^* - Y_t) dt \tag{C.28}$$

Comparing (C.28) with (4.5), we only need to verify

1. $$(\alpha_t, \theta_t) \in A, \quad (\bar{\lambda} \times P)\text{-a.s.} \tag{C.29}$$

2. $$\alpha_t v_{0,t}^* + \theta_t^\top v_{-,t}^* + \delta(v_t^*) = 0, \quad (\bar{\lambda} \times P)\text{-a.s.} \tag{C.30}$$

Fix an arbitrary $v \in \mathcal{N}$ and define the process

$$\zeta_t = \int_0^t (v_{0,s}^* - v_{0,s}) ds + \int_0^t [v_{-,s}^* - v_{-,s} - (v_{0,s}^* - v_{0,s}) \bar{1}_n]^\top \sigma_s^{-1} dZ_{v^*,s}, \tag{C.31}$$

and the sequence of stopping times

$$\begin{aligned} \tau_n \quad = \quad & T \wedge \inf \{ t \in [0,T] : |\zeta_t| + |\pi_{v^*,t}| + |W_t| \geq n, \\ & \text{or} \int_0^t |\theta_s^\top \sigma_s|^2 ds \geq n, \\ & \text{or} \int_0^t |v_{0,s}^* - v_{0,s}| ds \geq n, \\ & \text{or} \int_0^t |\sigma_s^{-1} [v_{-,s}^* - v_{-,s} - (v_{0,s}^* - v_{0,s}) \bar{1}_n]|^2 ds \geq n \Big\}. \end{aligned}$$

Then $\tau_n \nearrow T$ almost surely. Next, define

$$v_{\epsilon,n,t} = v_t^* + \epsilon(v_t - v_t^*)\mathbb{1}_{\{t\le\tau_n\}} \text{ for } \epsilon \in (0,1),$$

then by the convexity of $\widetilde{A}$, $v_{\epsilon,n} \in \mathcal{N}$. Furthermore, the pricing kernel under $v_{\epsilon,n}$ is given by

$$\pi_{v_{\epsilon,n},t} \quad = \quad \pi_{v^*,t} \exp\left(\epsilon\zeta_{t\wedge\tau_n} - \frac{\epsilon^2}{2}\int_0^{t\wedge\tau_n} |\sigma_s^{-1}[v_{-,s}^* - v_{-,s} - (v_{0,s}^* - v_{0,s})\bar{1}_n]|^2 ds\right).$$

Then, by the definition of stopping times τ_n, we have

$$e^{-2\epsilon n}\pi_{v^*,t} \le \pi_{v_{\epsilon,n},t} \le e^{2\epsilon n}\pi_{v^*,t},$$
$$e^{-3\epsilon n}\xi_{v^*,t} \le \xi_{v_{\epsilon,n},t} \le e^{3\epsilon n}\xi_{v^*,t}.$$

Therefore, $\xi_{v_{\epsilon,n}}$ is of class D and hence $v_{\epsilon,n} \in \mathcal{N}^*$ by Proposition I.1.47 in Jacod and Shiryaev (2013). Before moving forward, we first claim the following lemma

Lemma C.4.1. *For $\forall v \in \mathcal{N}$,*

$$\lim_{\epsilon\searrow 0} \frac{\widetilde{J}(\psi^*, v^*) - \widetilde{J}(\psi^*, v_{\epsilon,n})}{\epsilon} \ge$$
$$\psi^* E\left[\int_0^{\tau_n} \pi_{v^*,t}\cdot {}_tp_x[\alpha_t(v_{0,t}^* - v_{0,t}) + \theta_t^\top(v_{-,t}^* - v_{-,t}) + \delta(v_t^*) - \delta(v_t)]dt\right]. \quad \text{(C.32)}$$

Proof. First, we can derive

$$\left|\frac{\widetilde{U}_1(\psi^*\pi_{v^*,t},t) - \widetilde{U}_1(\psi^*\pi_{v_{\epsilon,n},t},t)}{\epsilon} + \frac{\widetilde{U}_2(\psi^*\pi_{v^*,t},t) - \widetilde{U}_2(\psi^*\pi_{v_{\epsilon,n},t},t)}{\epsilon}\right.$$
$$\left. + \frac{\widetilde{U}_3(\psi^*\pi_{v^*,T},T) - \widetilde{U}_3(\psi^*\pi_{v_{\epsilon,n},T},T)}{\epsilon} + \psi^*[Y_t + \delta(v_t^*)]\frac{\pi_{v^*,t} - \pi_{v_{\epsilon,n},t}}{\epsilon}\right|$$
$$\le \frac{1}{\epsilon}[f_1(\psi^* e^{-2\epsilon n}\pi_{v^*,t}) + f_2(\psi^* e^{-2\epsilon n}\pi_{v^*,t}) + f_3(\psi^* e^{-2\epsilon n}\pi_{v^*,T})]\psi^*|\pi_{v^*,t} - \pi_{v_{\epsilon,n},t}|$$
$$+\psi^*\pi_{v^*,t}\frac{Y_t - \delta(v_t^*)}{\epsilon}\left|\frac{\pi_{v_{\epsilon,n},t}}{\pi_{v^*,t}} - 1\right|$$
$$= \frac{\psi^*\pi_{v^*,t}}{\epsilon}[f_1(\psi^* e^{-2\epsilon n}\pi_{v^*,t}) + f_2(\psi^* e^{-2\epsilon n}\pi_{v^*,t}) + f_3(\psi^* e^{-2\epsilon n}\pi_{v^*,T})$$
$$+Y_t - \delta(v_t^*)]\left|\frac{\pi_{v_{\epsilon,n},t}}{\pi_{v^*,t}} - 1\right|$$
$$\le \psi^*\bar{K}_n\pi_{v^*,t}[f_1(\psi^* e^{-2\epsilon n}\pi_{v^*,t}) + f_2(\psi^* e^{-2\epsilon n}\pi_{v^*,t}) + f_3(\psi^* e^{-2\epsilon n}\pi_{v^*,T})$$
$$+Y_t - \delta(v_t^*)],$$

where $\bar{K}_n = \sup_{\epsilon\in(0,1)} \frac{e^{2\epsilon n}-1}{\epsilon} < \infty$. Moreover,

$$\pi_{v_{\epsilon,n},t}(\delta(v_t^*) - \delta(v_t))^- \leq -e^{2n}\pi_{v^*,t}\delta(v_t^*).$$

Then, by (4.21), (4.23), mean value theorem, Lebesgue's dominated convergence theorem, the convexity of supporting function δ, and Fatou's Lemma, we derive

$$\begin{aligned}
&\lim_{\epsilon\searrow 0} \frac{\widetilde{J}(\psi^*, v^*) - \widetilde{J}(\psi^*, v_{\epsilon,n})}{\epsilon} \\
= \;& \lim_{\epsilon\searrow 0} E\left\{\int_0^T e^{-\int_0^t \lambda_{x+s}ds} \frac{\widetilde{U}_1(\psi^*\pi_{v^*,t}, t) - \widetilde{U}_1(\psi^*\pi_{v_{\epsilon,n},t}, t)}{\epsilon} dt \right.\\
&+ \int_0^T \lambda_{x+t} e^{-\int_0^t \lambda_{x+s}ds} \frac{\widetilde{U}_2(\psi^*\pi_{v^*,t}, t) - \widetilde{U}_2(\psi^*\pi_{v_{\epsilon,n},t}, t)}{\epsilon} dt \\
&+ e^{-\int_0^T \lambda_{x+t}dt} \frac{\widetilde{U}_3(\psi^*\pi_{v^*,T}, T) - \widetilde{U}_3(\psi^*\pi_{v_{\epsilon,n},T}, T)}{\epsilon} \\
&+ \psi^* \int_0^T e^{-\int_0^t \lambda_{x+s}ds} [Y_t + \delta(v_t^*)] \frac{\pi_{v^*,t} - \pi_{v_{\epsilon,n},t}}{\epsilon} dt \\
&\left.+ \psi^* \int_0^T e^{-\int_0^t \lambda_{x+s}ds} \pi_{v_{\epsilon,n},t} \frac{\delta(v_t^*) - \delta(v_{\epsilon,n,t})}{\epsilon} dt \right\} \\
= \;& \lim_{\epsilon\searrow 0} E\left[\int_0^T e^{-\int_0^t \lambda_{x+s}ds} f_1(\psi^*\widetilde{\pi}_t, t)\psi^* \frac{\pi_{v_{\epsilon,n},t} - \pi_{v^*,t}}{\epsilon} dt \right.\\
&+ \int_0^T \lambda_{x+t} e^{-\int_0^t \lambda_{x+s}ds} f_2(\psi^*\widetilde{\pi}_t, t)\psi^* \frac{\pi_{v_{\epsilon,n},t} - \pi_{v^*,t}}{\epsilon} dt \\
&+ e^{-\int_0^T \lambda_{x+t}dt} f_3(\psi^*\widetilde{\pi}_T, T)\psi^* \frac{\pi_{v_{\epsilon,n},T} - \pi_{v^*,T}}{\epsilon} \\
&- \psi^* \int_0^T e^{-\int_0^t \lambda_{x+s}ds} [Y_t + \delta(v_t^*)] \frac{\pi_{v^*,t} - \pi_{v_{\epsilon,n},t}}{\epsilon} dt \\
&\left.+ \psi^* \int_0^T e^{-\int_0^t \lambda_{x+s}ds} \pi_{v_{\epsilon,n},t} \frac{\delta(v_t^*) - \delta(v_{\epsilon,n,t})}{\epsilon} dt \right] \\
= \;& \lim_{\epsilon\searrow 0} \psi^* E\left[\int_0^T e^{-\int_0^t \lambda_{x+s}ds} [f_1(\psi^*\widetilde{\pi}_t, t) + \lambda_{x+t} f_2(\psi^*\widetilde{\pi}_t, t) - [Y_t + \delta(v_t^*)]] \pi_{v^*,t} \right.\\
&\frac{e^{\epsilon\zeta_{t\wedge\tau_n} - \frac{\epsilon^2}{2}\int_0^{t\wedge\tau_n} |K_s|^2 ds} - 1}{\epsilon} dt
\end{aligned}$$

$$
\begin{aligned}
& +e^{-\int_0^T \lambda_{x+t}dt} f_3(\psi^* \widetilde{\pi}_T, T)\pi_{v^*,T} \frac{e^{\epsilon\zeta_{T\wedge\tau_n} - \frac{\epsilon^2}{2}\int_0^{T\wedge\tau_n}|K_s|^2 ds} - 1}{\epsilon}\Bigg] \\
& + \lim_{\epsilon\searrow 0} \psi^* E\left[\int_0^T e^{-\int_0^t \lambda_{x+s}ds} \pi_{v_{\epsilon,n},t} \frac{\delta(v_t^*) - \delta(v_{\epsilon,n,t})}{\epsilon} dt\right] \\
= \; & \psi^* E\Bigg[\int_0^T e^{-\int_0^t \lambda_{x+s}ds}[c_t^* + \lambda_{x+t}M_t^* - Y_t - \delta(v_t^*)]\pi_{v^*,t}\zeta_{t\wedge\tau_n} dt \\
& +e^{-\int_0^T \lambda_{x+t}dt} W_T^* \pi_{v^*,T}\zeta_{T\wedge\tau_n}\Bigg] \\
& + \lim_{\epsilon\searrow 0} \psi^* E\left[\int_0^T e^{-\int_0^t \lambda_{x+s}ds} \pi_{v_{\epsilon,n},t} \frac{\delta(v_t^*) - \delta(v_t^* + \epsilon(v_t - v_t^*)\mathbb{1}_{\{t\le\tau_n\}})}{\epsilon} dt\right] \\
\geq \; & \psi^* E\Bigg[\int_0^{\tau_n} e^{-\int_0^t \lambda_{x+s}ds}[c_t^* + \lambda_{x+t}M_t^* - Y_t - \delta(v_t^*)]\pi_{v^*,t}\zeta_t dt \\
& + \int_{\tau_n}^T e^{-\int_0^t \lambda_{x+s}ds}[c_t^* + \lambda_{x+t}M_t^* - Y_t - \delta(v_t^*)]\pi_{v^*,t}\zeta_{\tau_n} dt \\
& +e^{-\int_0^T \lambda_{x+s}ds} W_T^* \pi_{v_{v^*,T}\zeta_{\tau_n}}\Bigg] \\
& + \lim_{\epsilon\searrow 0} \psi^* E\left[\int_0^{\tau_n} e^{-\int_0^t \lambda_{x+s}ds} \pi_{v_{\epsilon,n},t} \frac{\delta(v_t^*) - \epsilon\delta(v_t) - (1-\epsilon)\delta(v_t^*)}{\epsilon} dt\right] \\
= \; & \psi^* E\Bigg[\int_0^{\tau_n} e^{-\int_0^t \lambda_{x+s}ds}[c_t^* + \lambda_{x+t}M_t^* - Y_t - \delta(v_t^*)]\pi_{v^*,t}\zeta_t dt \\
& +\pi_{v^*,\tau_n}\zeta_{\tau_n} e^{-\int_0^{\tau_n} \lambda_{x+s}ds} W_{\tau_n}\Bigg] \\
& +\psi^* E\left[\int_0^{\tau_n} e^{-\int_0^t \lambda_{x+s}ds} \pi_{v^*,t}[\delta(v_t^*) - \delta(v_t)]dt\right], \qquad \text{(C.33)}
\end{aligned}
$$

where the second equality comes from mean value theorem and

$$
\widetilde{\pi}_t \in [\min(\pi_{v_{\epsilon,n},t}, \pi_{v^*,t}), \max(\pi_{v_{\epsilon,n},t}, \pi_{v^*,t})].
$$

By (C.31) and Ito's formula, the first term in (C.33) satisfies the following SDE for $t \in [0, \tau_n]$

$$
\begin{aligned}
& d\left(\int_0^t e^{-\int_0^s \lambda_{x+u}du}[c_s^* + \lambda_{x+s}M_s^* - Y_s - \delta(v_s^*)]\beta_{v^*,s}\zeta_s ds + \beta_{v^*,t}\zeta_t e^{-\int_0^t \lambda_{x+s}ds} W_t\right) \\
= \; & \beta_{v^*,t} e^{-\int_0^t \lambda_{x+s}ds}\{W_t[v_{-,t}^* - v_{-,t} - (v_{0,t}^* - v_{0,t})\bar{1}_n]^\top \sigma_t^{-1} + \zeta_t\theta_t^\top\sigma_t\} dZ_{v^*,t} \\
& +\beta_{v^*,t} e^{-\int_0^t \lambda_{x+s}ds}[\alpha_t(v_{0,t}^* - v_{0,t}) + \theta_t^\top(v_{-,t}^* - v_{-,t})]dt, \qquad \text{(C.34)}
\end{aligned}
$$

which has the integral form

$$\begin{aligned}
& \int_0^{\tau_n} e^{-\int_0^t \lambda_{x+s}ds}[c_t^* + \lambda_{x+t}M_t^* - Y_t - \delta(v_t^*)]\beta_{v^*,t}\zeta_t dt + \beta_{v^*,\tau_n}\zeta_{\tau_n} e^{-\int_0^{\tau_n}\lambda_{x+t}dt}W_{\tau_n} \\
= & \int_0^{\tau_n} \beta_{v^*,t} e^{-\int_0^t \lambda_{x+s}ds}\{W_t[v_{-,t}^* - v_{-,t} - (v_{0,t}^* - v_{0,t})\bar{1}_n]^\top \sigma_t^{-1} + \zeta_t\theta_t^\top\sigma_t\}dZ_{v^*,t} \\
& + \int_0^{\tau_n} \beta_{v^*,t} e^{-\int_0^t \lambda_{x+s}ds}\{\alpha_t(v_{0,t}^* - v_{0,t}) + \theta_t^\top(v_{-,t}^* - v_{-,t})\}dt. \qquad \text{(C.35)}
\end{aligned}$$

Recall the definition of τ_n, the stochastic integral in (C.35) is a Q_{v^*} martingale, then we have

$$\begin{aligned}
& E\left[\int_0^{\tau_n} e^{-\int_0^t \lambda_{x+s}ds}[c_t^* + \lambda_{x+t}M_t^* - Y_t - \delta(v_t^*)]\pi_{v^*,t}\zeta_t dt + \pi_{v^*,\tau_n}\zeta_{\tau_n} e^{-\int_0^{\tau_n}\lambda_{x+s}ds}W_{\tau_n}\right] \\
= & E\left[\int_0^{\tau_n} \pi_{v^*,t} e^{-\int_0^t \lambda_{x+s}ds}[\alpha_t(v_{0,t}^* - v_{0,t}) + \theta_t^\top(v_{-,t}^* - v_{-,t})]dt\right] \qquad \text{(C.36)}
\end{aligned}$$

Substitute (C.36) into (C.33), we finish proving (C.32). □

In Lemma C.4.1, the left hand side of (C.32) is non-positive, so is the right hand side. Let $v = v^* + \rho,\ \rho \in \mathcal{N}$, since $\widetilde{A}$ is a convex cone, then $v \in \mathcal{N}$. Substitute v into (C.32), we have

$$\begin{aligned}
0 \ \geq\ & E\left[\int_0^{\tau_n} \pi_{v^*,t} e^{-\int_0^t \lambda_{x+s}ds}[-\alpha_t\rho_{0,t} - \theta_t^\top\rho_{-,t} + \delta(v_t^*) - \delta(v_t^* + \rho_t)]dt\right] \\
\geq\ & E\left[\int_0^{\tau_n} \pi_{v^*,t} e^{-\int_0^t \lambda_{x+s}ds}[-\alpha_t\rho_{0,t} - \theta_t^\top\rho_{-,t} - \delta(\rho_t)]dt\right]. \qquad \text{(C.37)}
\end{aligned}$$

where the second inequality comes from the sub-additivity of $\delta(v)$. Therefore, we obtain

$$\alpha_t\rho_{0,t} + \theta_t^\top\rho_{-,t} + \delta(\rho_t) \geq 0,\ \bar{\lambda} \times P\text{-a.s.} \qquad \text{(C.38)}$$

Inequality (C.38) implies for every $v \in \widetilde{A}$,

$$\alpha_t v_0 + \theta_t^\top v_- + \delta(v) \geq 0,\ \forall (t,\omega) \in D_v,$$

where $D_v \subset [0,T] \times \Omega$ is a set of full product measure, so is $D \triangleq \bigcap_{v \in \widetilde{A} \bigcap \mathbb{Q}^{n+1}} D_v$ that the following inequality holds

$$\alpha_t v_0 + \theta_t^\top v_- + \delta(v) \geq 0,\ \forall (t,\omega) \in D, v \in \widetilde{A}.$$

By Theorem 13.1 in Rockafellar (1970), we have proved (C.29).

Moreover, set $v \equiv 0$, (C.32) implies

$$E\left[\int_0^{\tau_n} \pi_{v^*,t} e^{-\int_0^t \lambda_{x+s} ds}[\alpha_t v_{0,t}^* + \theta_t^\top v_{-,t}^* + \delta(v_t^*)]dt\right] \leq 0. \tag{C.39}$$

Set $\rho = v^*$ in (C.37), we have

$$E\left[\int_0^{\tau_n} \pi_{v^*,t} e^{-\int_0^t \lambda_{x+s} ds}[\alpha_t v_{0,t}^* + \theta_t^\top v_{-,t}^* + \delta(v_t^*)]dt\right] \geq 0. \tag{C.40}$$

Finally, we can conclude from (C.39) and (C.40) that

$$\alpha_t v_{0,t}^* + \theta_t^\top v_{-,t}^* + \delta(v_t^*) = 0, (\bar{\lambda} \times P)\text{-a.s.},$$

i.e. (C.30) is verified. This completes the proof of one direction.

Conversely, due to the convexity of $\widetilde{U}_i$, we have

$$\widetilde{U}_i(z,t) \geq \widetilde{U}_i(x,t) + f_i(x,t)(x-z), \quad i = 1,2,3. \tag{C.41}$$

Then, the dual problem, $\widetilde{J}(\psi, v)$ satisfies

$$\begin{aligned}
\widetilde{J}(\psi, v) &= E\left\{\int_0^T e^{-\int_0^t \lambda_{x+s} ds} \widetilde{U}_1(\psi \pi_{v,t}, t)dt + \int_0^T \lambda_{x+t} e^{-\int_0^t \lambda_{x+s} ds} \widetilde{U}_2(\psi \pi_{v,t}, t)dt\right. \\
&\quad + e^{-\int_0^T \lambda_{x+t} dt} \widetilde{U}_3(\psi \pi_{v,T}, T) \\
&\quad \left. + \psi\left[w_0 + \int_0^T e^{-\int_0^t \lambda_{x+s} ds} \pi_{v,t}[Y_t + \delta(v_t)]dt\right]\right\} \\
&\geq E\left\{\int_0^T e^{-\int_0^t \lambda_{x+s} ds}[\widetilde{U}_1(\psi^* \pi_{v^*,t}, t) + c_t^*(\psi^* \pi_{v^*,t} - \psi \pi_{v,t})]dt\right. \\
&\quad \int_0^T \lambda_{x+t} e^{-\int_0^t \lambda_{x+s} ds}[\widetilde{U}_2(\psi^* \pi_{v^*,t}, t) + M_t^*(\psi^* \pi_{v^*,t} - \psi \pi_{v,t})]dt \\
&\quad + + e^{-\int_0^T \lambda_{x+t} dt}[\widetilde{U}_3(\psi^* \pi_{v^*,T}, T) + W_T^*(\psi^* \pi_{v^*,T} - \psi \pi_{v,T})] \\
&\quad \left. + \psi\left\{w_0 + \int_0^T e^{-\int_0^t \lambda_{x+s} ds} \pi_{v,t}[Y_t + \delta(v_t)]dt\right\}\right\} \\
&= E\left\{\int_0^T e^{-\int_0^t \lambda_{x+s} ds}[\widetilde{U}_1(\psi^* \pi_{v^*,t}, t) + c_t^* \psi^* \pi_{v^*,t}]dt\right.
\end{aligned}$$

$$
\begin{aligned}
&+\int_0^T \lambda_{x+t} e^{-\int_0^t \lambda_{x+s} ds} [\widetilde{U}_2(\psi^* \pi_{v^*,t}, t) + M_t^* \psi^* \pi_{v^*,t}] dt \\
&+ e^{-\int_0^T \lambda_{x+t} dt} [\widetilde{U}_3(\psi^* \pi_{v^*,T}, T) + W_T^* \psi^* \pi_{v^*,T}] \\
&+ \psi \left\{ w_0 + \int_0^T e^{-\int_0^t \lambda_{x+s} ds} \pi_{v,t} [Y_t + \delta(v_t) - c_t^* - \lambda_{x+t} M_t^*] dt \right. \\
&\left. - e^{-\int_0^T \lambda_{x+t} dt} \pi_{v,T} W_T^* \right\} \Big\} \\
\geq\ & E \left\{ \int_0^T e^{-\int_0^t \lambda_{x+s} ds} [\widetilde{U}_1(\psi^* \pi_{v^*,t}, t) + c_t^* \psi^* \pi_{v^*,t}] dt \right. \\
&+ \int_0^T \lambda_{x+t} e^{-\int_0^t \lambda_{x+s} ds} [\widetilde{U}_2(\psi^* \pi_{v^*,t}, t) + M_t^* \psi^* \pi_{v^*,t}] dt \\
&\left. + e^{-\int_0^T \lambda_{x+t} dt} [\widetilde{U}_3(\psi^* \pi_{v^*,T}, T) + W_T^* \psi^* \pi_{v^*,T}] \right\} \\
=\ & E \left\{ \int_0^T e^{-\int_0^t \lambda_{x+s} ds} [\widetilde{U}_1(\psi^* \pi_{v^*,t}, t) + \lambda_{x+t} \widetilde{U}_2(\psi^* \pi_{v^*,t}, t)] dt \right. \\
&+ e^{-\int_0^T \lambda_{x+t} dt} \widetilde{U}_3(\psi^* \pi_{v^*,T}, T) \\
&\left. + \psi^* \left[\int_0^T e^{-\int_0^t \lambda_{x+s} ds} \pi_{v^*,t} (c_t^* + \lambda_{x+t} M_t^*) dt + e^{-\int_0^T \lambda_{x+t} dt} \pi_{v^*,T} W_T^* \right] \right\} \\
=\ & E \left\{ \int_0^T e^{-\int_0^t \lambda_{x+s} ds} [\widetilde{U}_1(\psi^* \pi_{v^*,t}, t) + \lambda_{x+t} \widetilde{U}_2(\psi^* \pi_{v^*,t}, t)] dt \right. \\
&+ e^{-\int_0^T \lambda_{x+t} dt} \widetilde{U}_3(\psi^* \pi_{v^*,T}, T) \\
&\left. + \psi^* \left[w_0 + \int_0^T e^{-\int_0^t \lambda_{x+s} ds} \pi_{v^*,t} [Y_t + \delta(v_t^*)] dt \right] \right\} \\
=\ & \widetilde{J}(\psi^*, v^*),
\end{aligned}
$$

where the first inequality is based on the inequality (C.41), the second inequality holds true because of static budget constraint (4.14). The above inequality shows (ψ^*, v^*) is the solution to Problem (D), which completes the whole proof of the current theorem. □

C.5 Proof of Corollary 4.4.1

Proof. From the dual problem (D), we can obtain the following first-order partial derivative

$$\begin{aligned}\frac{\partial \widetilde{J}(\psi, v)}{\partial \psi} &= E\left[\int_0^T e^{-\int_0^t \lambda_{x+s}ds}\widetilde{U}_1'(\psi\pi_{v,t}, t)\pi_{v,t}dt + \int_0^T \lambda_{x+t}e^{-\int_0^t \lambda_{x+s}ds}\widetilde{U}_2'(\psi\pi_{v,t}, t)\pi_{v,t}dt\right. \\ &\left. +e^{-\int_0^T \lambda_{x+t}dt}\widetilde{U}_3'(\psi\pi_{v,T}, T)\pi_{v,T}\right] + \left\{w_0 + E\left[\int_0^T e^{-\int_0^t \lambda_{x+s}ds}\pi_{v,t}[Y_t + \delta(v_t)]dt\right]\right\},\end{aligned}$$

where $\widetilde{U}_i'(z, t)$, $i = 1, 2, 3$, are the first-order partial derivatives of dual utilities in its first variables.

For dual utility $\widetilde{U}_1(z, t)$, based on (C.22) and (C.23), we derive

$$\frac{\partial \widetilde{U}_1(z, t)}{\partial z} = -c^* = -U_1'^{-1}(z, t). \tag{C.42}$$

Together with the Inada condition (4.16), we obtain

$$\widetilde{U}_1'(0+, t) = -\infty,\ \widetilde{U}_1'(\infty, t) = 0,\ \text{for } \forall t \in [0, T].$$

In addition, by (C.24), we have $\widetilde{U}_1'(z, t)$ increase from $-\infty$ to 0 when z moves from $0+$ to ∞. The same arguments can be applied to $\widetilde{U}_2'(z, t)$ and $\widetilde{U}_3'(z, t)$.

In addition, since $\pi_{v,t} > 0$ and $w_0 + E\left[\int_0^T e^{-\int_0^t \lambda_{x+s}ds}\pi_{v,t}[Y_t + \delta(v_t)]dt\right] > 0$, we can always find a unique $\psi^v > 0$ such that

$$\frac{\partial \widetilde{J}(\psi^v, v)}{\partial \psi} = 0. \tag{C.43}$$

Finally, because $\widetilde{J}(\psi, v)$ is convex in ψ, the zero point ψ^v of $\frac{\partial \widetilde{J}(\psi,v)}{\partial \psi}$ minimizes $\widetilde{J}(\psi, v)$ under a given v. Lastly, by (4.18) and (C.43), we find the optimal strategy under (ψ^v, v) satisfies the following static budget constraint

$$\begin{aligned}&E\left[\int_0^T \pi_{v,t}e^{-\int_0^t \lambda_{x+s}ds}[f_1(\psi_v\pi_{v,t}) + \lambda_{x+t}f_2(\psi_v\pi_{v,t})]dt\right. \\ &\left. +\pi_{v,T}e^{-\int_0^T \lambda_{x+t}dt}f_3(\psi_v\pi_{v,T})\right] = w_0 + E\left[\int_0^T \pi_{v,t}[Y_t + \delta(v_t)]dt\right]\end{aligned}$$

.

From this static budget constraint, we can define the optimal wealth following (4.24), which is a martingale. Therefore, the optimal free disposal equals zero. □

C.6 Proof of Theorem 4.5.1

Proof. Due to the result from Levin (1976), we have the following lemma

Lemma C.6.1. *Let $F : L^1(S, \Sigma, \mu; X) \to \mathbb{R} \bigcup \{+\infty\}$ be a convex functional where (S, Σ, μ) is a measure space with μ finite and non-negative, Σ complete, X is a reflexive Banach space, and $L^1(S, \Sigma, \mu; X)$ denotes the set of Lebesgue measure functions: $\Psi : S \to X$, such that $\int_S |\Psi| d\mu < \infty$. If F is lower semi-continuous in the topology τ of convergence in measure, then it attains a minimum on any convex set $\mathcal{K} \subset L^1(S, \Sigma, \mu; X)$ that is τ-closed and norm-bounded.*

Proof. See Theorem 1 in Levin (1976). □

Before going to the final proof, we make the following preparations. Let $\mathcal{D}$ denotes the σ-field generated by the progressively measurable processes, $\mathcal{L}^*$ denotes the class of $(\bar{\lambda} \times Q_0)$-null sets in $\mathcal{B}([0, T]) \times \mathcal{F}$, and $\mathcal{D}^* = \sigma(\mathcal{D} \bigcup \mathcal{L}^*)$ denotes the smallest σ-field containing $\mathcal{D}$ and $\mathcal{L}^*$. Then, we have the following lemma

Lemma C.6.2.

1. *$\mathcal{D}^* = \{A \in \mathcal{B}([0, T]) \times \mathcal{F} : \exists B \in \mathcal{D}$ s.t. $A \Delta B \in \mathcal{L}^*\}$, where $A \Delta B$ denotes the symmetric difference of A and B, defined by $A \Delta B = (A \setminus B) \bigcup (B \setminus A)$.*
2. *Suppose $Y : [0, T] \times \Omega \to \mathbb{R}^n$ is $(\mathcal{B}([0, T]) \times \mathcal{F})$-measurable. Then Y is $\mathcal{D}^*$-measurable if and only if there exists a progressive process $\widetilde{Y}$ such that $Y = \widetilde{Y}$, $(\bar{\lambda} \times Q_0)$-a.s.*

Proof. See Page 59-60 in Chung (2013). □

The first part of Lemma C.6.2 implies $\mathcal{D}^*$ is complete. Using $L^1(\bar{\lambda} \times Q_0; \mathbb{R}^n) = L^1([0, T] \times \Omega, \mathcal{D}^*, \bar{\lambda} \times Q_0; \mathbb{R}^n)$ to denote the set of $\mathcal{D}^*$-measurable integrable process, the second part of Lemma C.6.2 implies if $(c, M, W_T) \in L^1(\bar{\lambda} \times Q_0; \mathbb{R}^3)$, then there exists equivalent version of $(c, M, W_T) \in L^1(\bar{\lambda} \times Q_0; \mathbb{R}^3)$ that is progressive measurable.

Denote the discounted control variables $\widetilde{c}_t = e^{-\int_0^t r_s^+ ds} c_t$, $\widetilde{M}_t = e^{-\int_0^t r_s^+ ds} M_t$, and $\widetilde{W}_T = e^{-\int_0^T r_t^+ dt} W_T$, where r_t^+ denotes the positive part of interest rate, then we can rewrite the consumption and bequest set (4.9) as

$$\widetilde{G} := \left\{ (\widetilde{c}, \widetilde{M}, \widetilde{W}_T) : E^{Q_0} \left[\int_0^T \left| e^{\int_0^t r_s^+ ds} \widetilde{c}_t \right| + \left| e^{\int_0^t r_s^+ ds} \widetilde{M}_t \right| dt + \left| e^{\int_0^T r_t^+ dt} \widetilde{W}_T \right| \right] < \infty \right\}. \quad \text{(C.44)}$$

By the definition of (C.44), once $(\widetilde{c}, \widetilde{M}, \widetilde{W}_T) \in \widetilde{G}$, then $(\widetilde{c}, \widetilde{M}, \widetilde{W}_T) \in L^1(\bar{\lambda} \times Q_0; \mathbb{R}^3)$. Denote the non-negative orthant of $\widetilde{G}$ as $\widetilde{G}_+$, then we use $\widetilde{G}_+^*$ to represent $(\widetilde{c}, \widetilde{M}, \widetilde{W}_T) \in \widetilde{G}_+$ such that

$$\min\left\{E\left[\int_0^T U_1\left(e^{\int_0^t r_s^+ ds}\widetilde{c}_t, t\right)^+ dt\right], E\left[\int_0^T U_1\left(e^{\int_0^t r_s^+ ds}\widetilde{c}_t, t\right)^- dt\right]\right\} < \infty, \quad \text{(C.45)}$$

$$\min\left\{E\left[\int_0^T U_2\left(e^{\int_0^t r_s^+ ds}\widetilde{M}_t, t\right)^+ dt\right], E\left[\int_0^T U_2\left(e^{\int_0^t r_s^+ ds}\widetilde{M}_t, t\right)^- dt\right]\right\} < \infty \quad \text{(C.46)}$$

and

$$\min\left\{E\left[U_3\left(e^{\int_0^T r_t^+ dt}\widetilde{W}_T, T\right)^+\right], E\left[U_3\left(e^{\int_0^T r_t^+ dt}\widetilde{W}_T, T\right)^-\right]\right\} < \infty. \quad \text{(C.47)}$$

Moreover, for the discounted wealth, we have $W_T e^{-\int_0^T r_s ds} = W_T e^{-\int_0^T r_s^+ - r_s^- ds} = \widetilde{W}_T e^{\int_0^T r_s^- ds}$, similar to $M_t e^{-\int_0^t r_s ds}$ and $c_t e^{-\int_0^t r_s ds}$. Then, the primal problem (P) can be rewritten as

$$\sup_{(\widetilde{c}, \widetilde{M}, \widetilde{W}_T) \in \widetilde{G}_+^*} J_1(\widetilde{c}, \widetilde{M}, \widetilde{W}_T)$$

$$\text{s.t. } E^{Q_v}\left[e^{-\int_0^T v_{0,s} + \lambda_{x+s} ds}\widetilde{W}_T e^{\int_0^T r_s^- ds} + \int_0^T \lambda_{x+t} e^{-\int_0^t v_{0,s} + \lambda_{x+s} ds}\widetilde{M}_t e^{\int_0^t r_s^- ds} dt\right. \quad (P_1)$$

$$\left. + \int_0^T e^{-\int_0^t v_{0,s} + \lambda_{x+s} ds}\widetilde{c}_t e^{\int_0^t r_s^- ds} dt\right] \leq w_0 + E^{Q_v}\left[\int_0^T \beta_{v,t} e^{-\int_0^t \lambda_{x+s} ds}[Y_t + \delta(v_t)] dt\right],$$

for $\forall v \in \mathcal{N}^*$, where

$$\begin{aligned} J_1(\widetilde{c}, \widetilde{M}, \widetilde{W}_T) &= E\left[\int_0^T e^{-\int_0^t \lambda_{x+s} ds} U_1(e^{\int_0^t r_s^+ ds}\widetilde{c}_t, t) dt\right. \\ &\quad + \int_0^T \lambda_{x+t} e^{-\int_0^t \lambda_{x+s} ds} U_2(e^{\int_0^t r_s^+ ds}\widetilde{M}_t, t) dt \\ &\quad \left. + e^{-\int_0^T \lambda_{x+t} dt} U_3(e^{\int_0^T r_s^+ ds}\widetilde{W}_T, T)\right]. \end{aligned}$$

Since $0 \in \mathcal{N}^*$, we can restrict the existence proof of the problem (P_1) to the existence proof of the following problem

$$\sup_{(\widetilde{c},\widetilde{M},\widetilde{W}_T)\in\mathcal{K}} J_1(\widetilde{c},\widetilde{M},\widetilde{W}_T)$$

$$\text{s.t. } \mathcal{K}=\Bigg\{(\widetilde{c},\widetilde{M},\widetilde{W}_T)\in\widetilde{G}_+^* : E^{Q_0}\Bigg[e^{-\int_0^T\lambda_{x+s}ds}\widetilde{W}_T e^{\int_0^T r_s^- ds}+\int_0^T \lambda_{x+t}e^{-\int_0^t\lambda_{x+s}ds}\widetilde{M}_t e^{\int_0^t r_s^- ds}dt \tag{P_2}$$

$$+\int_0^T e^{-\int_0^t\lambda_{x+s}ds}\widetilde{c}_t e^{\int_0^t r_s^- ds}dt\Bigg]\leq w_0+E^{Q_0}\left[\int_0^T e^{-\int_0^t r_s+\lambda_{x+s}ds}Y_t dt\right]\Bigg\}.$$

$$e^{-\int_0^t r_s^+ ds}\leq e^{-\int_0^t r_s ds}$$

Lemma C.6.3. *Under the assumptions of Theorem 4.5.1, $\mathcal{K}$ is a convex and norm bounded subset of $L^1(\bar{\lambda}\times Q_0;\mathbb{R}^3)$, and topological closed in $(\bar{\lambda}\times Q_0)$-measure.*

Proof. First, since $e^{\int_0^T r_s^- ds}>1$ and the definition of $\mathcal{K}$, we have $(\widetilde{c},\widetilde{M},\widetilde{W}_T)\in L^1(\bar{\lambda}\times Q_0;\mathbb{R}^3)$.

Second, we prove that $\mathcal{K}$ is a convex set.

Specifically, for arbitrary $(\widetilde{c}_{1,t},\widetilde{M}_{1,t},\widetilde{W}_{1,T})\in\mathcal{K}$ and $(\widetilde{c}_{2,t},\widetilde{M}_{2,t},\widetilde{W}_{2,T})\in\mathcal{K}$, we need to prove $(\lambda\widetilde{c}_{1,t}+(1-\lambda)\widetilde{c}_{2,t},\lambda\widetilde{M}_{1,t}+(1-\lambda)\widetilde{M}_{2,t},\lambda\widetilde{W}_{1,T}+(1-\lambda)\widetilde{W}_{2,T}),\lambda\in[0,1]$ satisfies the static budget constraint under Q_0 and belongs to $\widetilde{G}_+^*$. The static budget constraint is easy to verify

$$\begin{aligned}
&E^{Q_0}\Bigg[\int_0^T[\lambda e^{-\int_0^t\lambda_{x+s}ds}\widetilde{c}_{1,t}e^{\int_0^t r_s^- ds}+(1-\lambda)e^{-\int_0^t\lambda_{x+s}ds}\widetilde{c}_{2,t}e^{\int_0^t r_s^- ds}]\\
&+[\lambda\lambda_{x+t}e^{-\int_0^t\lambda_{x+s}ds}\widetilde{M}_{1,t}e^{\int_0^t r_s^- ds}+(1-\lambda)\lambda_{x+t}e^{-\int_0^t\lambda_{x+s}ds}\widetilde{M}_{2,t}e^{\int_0^t r_s^- ds}]dt\\
&+\lambda e^{-\int_0^T\lambda_{x+s}ds}\widetilde{W}_{1,T}e^{\int_0^T r_s^- ds}+(1-\lambda)e^{-\int_0^T\lambda_{x+s}ds}\widetilde{W}_{2,T}e^{\int_0^T r_s^- ds}\Bigg]\\
&\leq[\lambda+(1-\lambda)]\left\{w_0+E^{Q_0}\left[\int_0^T e^{-\int_0^t r_s+\lambda_{x+s}ds}Y_t dt\right]\right\}.
\end{aligned}$$

Next, we check $(\lambda\widetilde{c}_{1,t}+(1-\lambda)\widetilde{c}_{2,t},\lambda\widetilde{M}_{1,t}+(1-\lambda)\widetilde{M}_{2,t},\lambda\widetilde{W}_{1,T}+(1-\lambda)\widetilde{W}_{2,T})\in\widetilde{G}$.

$$
\begin{aligned}
&E^{Q_0}\left[\int_0^T e^{\int_0^t r_s^+ ds}\left|\lambda\widetilde{c}_{1,t}+(1-\lambda)\widetilde{c}_{2,t}\right| + e^{\int_0^t r_s^+ ds}\left|\lambda\widetilde{M}_{1,t}+(1-\lambda)\widetilde{M}_{2,t}\right| dt\right.\\
&\left.+e^{\int_0^T r_s^+ ds}\left|\lambda\widetilde{W}_{1,T}+(1-\lambda)\widetilde{W}_{2,T}\right|\right]\\
&\le \lambda E^{Q_0}\left[\int_0^T e^{\int_0^t r_s^+ ds}\widetilde{c}_{1,t}+e^{\int_0^t r_s^+ ds}\widetilde{M}_{1,t}dt+e^{\int_0^T r_s^+ ds}\widetilde{W}_{1,T}\right]\\
&+(1-\lambda)E^{Q_0}\left[\int_0^T e^{\int_0^t r_s^+ ds}\widetilde{c}_{2,t}+e^{\int_0^t r_s^+ ds}\widetilde{M}_{2,t}dt+e^{\int_0^T r_s^+ ds}\widetilde{W}_{2,T}\right]\\
&<\infty.
\end{aligned}
$$

The last inequality holds true because $(\widetilde{c}_{1,t},\widetilde{M}_{1,t},\widetilde{W}_{1,T})\in\mathcal{K}$ and $(\widetilde{c}_{2,t},\widetilde{M}_{2,t},\widetilde{W}_{2,T})\in\mathcal{K}$. Finally, we prove $(\lambda\widetilde{c}_{1,t}+(1-\lambda)\widetilde{c}_{2,t},\lambda\widetilde{M}_{1,t}+(1-\lambda)\widetilde{M}_{2,t},\lambda\widetilde{W}_{1,T}+(1-\lambda)\widetilde{W}_{2,T})\in\widetilde{G}^*_+$. For the consumption process, we have

$$
\begin{aligned}
&E\left[\int_0^T U_1(e^{\int_0^t r_s^+ ds}(\lambda\widetilde{c}_{1,t}+(1-\lambda)\widetilde{c}_{2,t}),t)^+dt\right]\\
&\le kE\left\{\int_0^T\left[1+\left(\lambda e^{\int_0^t r_s^+ ds}\widetilde{c}_{1,t}+(1-\lambda)e^{\int_0^t r_s^+ ds}\widetilde{c}_{2,t}\right)^{1-b_1}\right]dt\right\}\\
&=kT+kE^{Q_0}\left[\int_0^T \xi_{0,t}^{-1}\left(\lambda e^{\int_0^t r_s^+ ds}\widetilde{c}_{1,t}+(1-\lambda)e^{\int_0^t r_s^+ ds}\widetilde{c}_{2,t}\right)^{1-b_1}dt\right]\\
&\le kT+k\left\{E^{Q_0}\left[\int_0^T \xi_{0,t}^{-1/b_1}dt\right]\right\}^{b_1}\left\{E^{Q_0}\left[\int_0^T \lambda e^{\int_0^t r_s^+ ds}\widetilde{c}_{1,t}+(1-\lambda)e^{\int_0^t r_s^+ ds}\widetilde{c}_{2,t}dt\right]\right\}^{1-b_1}\\
&<\infty.
\end{aligned}
$$

The first inequality comes from (4.25). The second inequality is due to the Holder's inequality. The last inequality is because $\widetilde{c}_{1,t}\in\widetilde{G}$, $\widetilde{c}_{2,t}\in\widetilde{G}$, and (4.26). Similar proofs for $U_2(e^{\int_0^t r_s^+ ds}\widetilde{M}_t,t)$ and $U_3(e^{\int_0^T r_t^+ dt}\widetilde{W}_T,T)$. Therefore, $\mathcal{K}$ is a convex set.

Second, we verify $\mathcal{K}$ is norm bounded in $L^1(\bar{\lambda}\times Q_0;\mathbb{R}^3)$.

Due to the continuity of the deterministic force of mortality λ_{x+t}, Assumption 4.3.1, $e^{\int_0^T r_s^- ds}>1$, and the static budget constraint in $\mathcal{K}$, we derive

$$
E^{Q_0}\left[\widetilde{W}_T+\int_0^T\widetilde{M}_t dt+\int_0^T\widetilde{c}_t dt\right]\le K_0,
$$

where K_0 is some positive constant.

Third, we check set $\mathcal{K}$ is topological closed in $(\bar{\lambda} \times Q_0)$-measure.

To be specific, we need to prove if an arbitrary sequence $(\widetilde{c}_{n,t}, \widetilde{M}_{n,t}, \widetilde{W}_{n,T}) \in \mathcal{K}$ converges to $(\widetilde{c}_{\infty,t}, \widetilde{M}_{\infty,t}, \widetilde{W}_{\infty,T})$, then $(\widetilde{c}_{\infty,t}, \widetilde{M}_{\infty,t}, \widetilde{W}_{\infty,T}) \in \mathcal{K}$.

First, we check $(\widetilde{c}_{\infty,t}, \widetilde{M}_{\infty,t}, \widetilde{W}_{\infty,T})$ satisfy the static budget constraint in $\mathcal{K}$. Since the non-negative orthant of (c, M, W_T) is closed, then by Fatou's lemma, we obtain

$$
\begin{aligned}
& E^{Q_0}\left[e^{-\int_0^T \lambda_{x+t}dt}\widetilde{W}_{\infty,T}e^{\int_0^T r_s^- ds} + \int_0^T \lambda_{x+t}e^{-\int_0^t \lambda_{x+s}ds}\widetilde{M}_{\infty,t}e^{\int_0^t r_s^- ds}dt \right.\\
& \left. + \int_0^T e^{-\int_0^t \lambda_{x+s}ds}\widetilde{c}_{\infty,t}e^{\int_0^t r_s^- ds}dt\right] \\
& \leq \lim_{n\to\infty} E^{Q_0}\left[e^{-\int_0^T \lambda_{x+t}dt}\widetilde{W}_{n,T}e^{\int_0^T r_s^- ds} + \int_0^T \lambda_{x+t}e^{-\int_0^t \lambda_{x+s}ds}\widetilde{M}_{n,t}e^{\int_0^t r_s^- ds}dt \right.\\
& \left. + \int_0^T e^{-\int_0^t \lambda_{x+s}ds}\widetilde{c}_{n,t}e^{\int_0^t r_s^- ds}dt\right] \\
& \leq w_0 + E^{Q_0}\left[\int_0^T e^{-\int_0^t r_s+\lambda_{x+s}ds}Y_t dt\right] \\
& \leq K_1,
\end{aligned}
$$

where K_1 is some positive constant. The first inequality is based on the Fatou's lemma. The second inequality is because $(\widetilde{c}_{n,t}, \widetilde{M}_{n,t}, \widetilde{W}_{n,T}) \in \mathcal{K}$. The third inequality is by Assumption 4.3.1.

Second, we claim that $(\widetilde{c}_{\infty,t}, \widetilde{M}_{\infty,t}, \widetilde{W}_{\infty,T}) \in \widetilde{G}$, i.e.

$$
E^{Q_0}\left[\int_0^T \left|e^{\int_0^t r_s^+ ds}\widetilde{c}_{\infty,t}\right| + \left|e^{\int_0^t r_s^+ ds}\widetilde{M}_{\infty,t}\right| dt + \left|e^{\int_0^T r_t^+ dt}\widetilde{W}_{\infty,T}\right|\right] < \infty.
$$

This is because

$$
\left(e^{\int_0^t r_s^+ ds}\widetilde{c}_{n,t}, e^{\int_0^t r_s^+ ds}\widetilde{M}_{n,t}, e^{\int_0^T r_t^+ dt}\widetilde{W}_{n,T}\right) \in L^1_+(\bar{\lambda} \times Q_0; \mathbb{R}^3),
$$

and the completeness of $L^1_+(\bar{\lambda} \times Q_0; \mathbb{R}^3)$.

Third, we verify that $(\widetilde{c}_{\infty,t}, \widetilde{M}_{\infty,t}, \widetilde{W}_{\infty,T}) \in \widetilde{G}_+^*$. Since $\left(e^{\int_0^t r_s^+ ds}\widetilde{c}_{\infty,t}\right) \in L_+^1(\bar{\lambda} \times Q_0)$, we have

$$
\begin{aligned}
& E\left[\int_0^T U_1\left(e^{\int_0^t r_s^+ ds}\widetilde{c}_{\infty,t}, t\right)^+ dt\right] \le kE\left[\int_0^T \left(1 + e^{(1-b_1)\int_0^t r_s^+ ds}\widetilde{c}_{\infty,t}^{1-b_1}\right) dt\right] \\
& \le kT + kE^{Q_0}\left[\int_0^T \xi_{0,t}^{-1} e^{(1-b_1)\int_0^t r_s^+ ds}\widetilde{c}_{\infty,t}^{1-b_1} dt\right] \\
& \le kT + k\left\{E^{Q_0}\left[\int_0^T \xi_{0,t}^{-1/b_1} dt\right]\right\}^{b_1}\left\{E^{Q_0}\left[\int_0^T e^{\int_0^t r_s^+ ds}\widetilde{c}_{\infty,t} dt\right]\right\}^{1-b_1} \\
& < \infty.
\end{aligned}
\tag{C.48}
$$

Similar proofs for $U_2(e^{\int_0^t r_s^+ ds}\widetilde{M}_t, t)$ and $U_3(e^{\int_0^T r_t^+ dt}\widetilde{W}_T, T)$. Therefore, $\mathcal{K}$ is topological closed in $(\bar{\lambda} \times Q_0)$-measure. This completes the whole proof of Lemma C.6.3. □

Lemma C.6.4. *Under the assumptions of Theorem 4.5.1, J_1 is bounded above on $\mathcal{K}$ and upper semicontinuous with respect to convergence in $\bar{\lambda} \times Q_0$-measure, which means for any $\{(\widetilde{c}_n, \widetilde{M}_n, \widetilde{W}_{T,n})\} \in \mathcal{K}$ and $(\widetilde{c}, \widetilde{M}, \widetilde{W}_T) \in L^1(\bar{\lambda} \times Q_0; \mathbb{R}^3)$, if $(\widetilde{c}_n, \widetilde{M}_n, \widetilde{W}_{T,n}) \to (\widetilde{c}, \widetilde{M}, \widetilde{W}_T)$ in measure, then*

$$J_1(\widetilde{c}, \widetilde{M}, \widetilde{W}_T) \ge \limsup_{n\to\infty} J_1(\widetilde{c}_n, \widetilde{M}_n, \widetilde{W}_{T,n})$$

Proof. By the definition of $\mathcal{K}$, we have J_1 bounded above on $\mathcal{K}$ from (C.48) for any $\left(e^{\int_0^t r_s^+ ds}\widetilde{c}_t, e^{\int_0^t r_s^+ ds}\widetilde{M}_t, e^{\int_0^T r_t^+ dt}\widetilde{W}_T\right) \in L_+^1(\bar{\lambda} \times Q_0; \mathbb{R}^3)$, and the fact that $\mathcal{K}$ is bounded in $L^1(\bar{\lambda} \times Q_0)$-norm. Next, we assume that $J_1(\widetilde{c}, \widetilde{M}, \widetilde{W}_T)$ is not upper semi-continuous on $\mathcal{K}$. Then, there exists a constant α such that

$$J_1(\widetilde{c}, \widetilde{M}, \widetilde{W}_T) < \alpha \le J_1(\widetilde{c}_n, \widetilde{M}_n, \widetilde{W}_{T,n}) \text{ for all } n, \tag{C.49}$$

where $\{(\widetilde{c}_n, \widetilde{M}_n, \widetilde{W}_{T,n})\} \subset \mathcal{K}$ and $(\widetilde{c}, \widetilde{M}, \widetilde{W}_T) \subset \mathcal{K}$, and $(\widetilde{c}_n, \widetilde{M}_n, \widetilde{W}_{T,n}) \to (\widetilde{c}, \widetilde{M}, \widetilde{W}_T)$ in measure. Taking a subsequence, we can assume $(\widetilde{c}_n, \widetilde{M}_n, \widetilde{W}_{T,n}) \to (\widetilde{c}, \widetilde{M}, \widetilde{W}_T)$ almost everywhere. Then, we prove that the family

$$
\begin{aligned}
&\left\{e^{-\int_0^t \lambda_{x+s} ds}\xi_{0,t}^{-1} U_1\left(e^{\int_0^t r_s^+ ds}\widetilde{c}_{n,t}, t\right)^+, \lambda_{x+t} e^{-\int_0^t \lambda_{x+s} ds}\xi_{0,t}^{-1} U_2\left(e^{\int_0^t r_s^+ ds}\widetilde{M}_{n,t}, t\right)^+,\right. \\
&\left. e^{-\int_0^T \lambda_{x+t} dt}\xi_{0,T}^{-1} U_3\left(e^{\int_0^T r_t^+ dt}\widetilde{W}_{n,T}, T\right)^+\right\}
\end{aligned}
$$

is uniformly integrable. For $\{e^{-\int_0^t \lambda_{x+s}ds}\xi_{0,t}^{-1}U_1(e^{\int_0^t r_s^+ ds}\widetilde{c}_{n,t},t)^+\}$, since $U_1(e^{\int_0^t r_s^+ ds}\widetilde{c}_{n,t},t)^+ \leq k_1[1+(e^{\int_0^t r_s^+ ds}\widetilde{c}_{n,t})^{1-b_1}]$, we only need to prove

$$\sup_n E^{Q_0}\left[\int_0^T (\xi_{0,t}^{-1}(e^{\int_0^t r_s^+ ds}\widetilde{c}_{n,t})^{1-b_1})^{\widehat{p}_1}dt\right] < \infty, \text{ for some } \widehat{p}_1 > 1. \tag{C.50}$$

Taking $\widehat{p}_1 = \frac{p_1}{b_1+p_1(1-b_1)}$, where $b_1 \in (0,1)$, $p_1 > 1$, then by Holder's inequality, we have

$$\begin{aligned}
&E^{Q_0}\left[\int_0^T \xi_{0,t}^{-\widehat{p}_1}(e^{\int_0^t r_s^+ ds}\widetilde{c}_{n,t})^{\widehat{p}_1(1-b_1)}dt\right] \\
&\leq \left\{E^{Q_0}\left[\int_0^T \xi_{0,t}^{-\widehat{p}_1/(1-\widehat{p}_1(1-b_1))}dt\right]\right\}^{1-\widehat{p}_1(1-b_1)} \left\{E^{Q_0}\left[\int_0^T e^{\int_0^t r_s^+ ds}\widetilde{c}_{n,t}dt\right]\right\}^{\widehat{p}_1(1-b_1)} \\
&= \left\{E^{Q_0}\left[\int_0^T \xi_{0,t}^{-p_1/b_1}dt\right]\right\}^{1-\widehat{p}_1(1-b_1)} \left\{E^{Q_0}\left[\int_0^T e^{\int_0^t r_s^+ ds}\widetilde{c}_{n,t}dt\right]\right\}^{\widehat{p}_1(1-b_1)} \\
&< \infty.
\end{aligned}$$

The first inequality comes from Holder's inequality. The second inequality is due to (4.26), and $\widetilde{c}_{n,t} \in \mathcal{K}$ so that $\widetilde{c}_{n,t}$ satisfies (C.44). Similar proofs for $\lambda_{x+t}e^{-\int_0^t \lambda_{x+s}ds}\xi_{0,t}^{-1}U_2(e^{\int_0^t r_s^+ ds}\widetilde{M}_{n,t},t)^+$ and $e^{-\int_0^T \lambda_{x+t}dt}\xi_{0,T}^{-1}U_3(e^{\int_0^T r_t^+ dt}\widetilde{W}_{n,T},T)^+$. Since J_1 is bounded above (see Lemma C.6.4), following Fatou's lemma, we obtain

$$\begin{aligned}
J_1(\widetilde{c},\widetilde{M},\widetilde{W}_T) &= E^{Q_0}\left[\int_0^T e^{-\int_0^t \lambda_{x+s}ds}\xi_{0,t}^{-1}U_1\left(e^{\int_0^t r_s^+ ds}\widetilde{c}_t,t\right)dt\right. \\
&\quad + \int_0^T \lambda_{x+t}e^{-\int_0^t \lambda_{x+s}ds}\xi_{0,t}^{-1}U_2\left(e^{\int_0^t r_s^+ ds}\widetilde{M}_t,t\right)dt \\
&\quad \left. + e^{-\int_0^T \lambda_{x+t}dt}\xi_{0,T}^{-1}U_3(e^{\int_0^T r_t^+ dt}\widetilde{W}_T,T)\right] \\
&\geq \limsup_{n\to\infty} E^{Q_0}\left[\int_0^T e^{-\int_0^t \lambda_{x+s}ds}\xi_{0,t}^{-1}U_1\left(e^{\int_0^t r_s^+ ds}\widetilde{c}_{n,t},t\right)dt\right. \\
&\quad + \int_0^T \lambda_{x+t}e^{-\int_0^t \lambda_{x+s}ds}\xi_{0,t}^{-1}U_2\left(e^{\int_0^t r_s^+ ds}\widetilde{M}_{n,t},t\right)dt \\
&\quad \left. + e^{-\int_0^T \lambda_{x+t}dt}\xi_{0,T}^{-1}U_3(e^{\int_0^T r_t^+ dt}\widetilde{W}_{n,T},T)\right] \\
&= \limsup_{n\to\infty} J_1(\widetilde{c}_n,\widetilde{M}_n,\widetilde{W}_{n,T}),
\end{aligned} \tag{C.51}$$

which contradicts (C.49). Therefore, $J_1(\widetilde{c},\widetilde{M},\widetilde{W}_T)$ is upper semi-continuous. □

With all the lemmas above, we can finally prove Theorem 4.5.1. Define the map $J_2 : L^1(\bar{\lambda} \times Q_0; \mathbb{R}^3) \to \mathbb{R} \cup \{+\infty\}$ as

$$J_2(\widetilde{c}, \widetilde{M}, \widetilde{W}_T) = \begin{cases} -J_1(\widetilde{c}, \widetilde{M}, \widetilde{W}_T), & \text{if } (\widetilde{c}, \widetilde{M}, \widetilde{W}_T) \in \mathcal{K}; \\ +\infty, & \text{otherwise.} \end{cases}$$

Then, Lemma C.6.4 and concavity of J_1 prove J_2 is convex and lower semi-continuous in measure. Lemma C.6.3 shows $\mathcal{K}$ is a convex and norm bounded subset of $L^1(\bar{\lambda} \times Q_0; \mathbb{R}^3)$, and topological closed in $(\bar{\lambda} \times Q_0)$-measure. Moreover, $\mathbb{R}^3$ is a reflexive Banach space.

Finally, following Lemma C.6.1 and the fact $J_2(\widetilde{c}, \widetilde{M}, \widetilde{W}_T) < \infty$ for some $(\widetilde{c}, \widetilde{M}, \widetilde{W}_T) \in \mathcal{K}$, there exists a $(\widetilde{c}^*, \widetilde{M}^*, \widetilde{W}_T^*) \in \mathcal{K}$ such that $J_2(\widetilde{c}^*, \widetilde{M}^*, \widetilde{W}_T^*) \leq J_2(\widetilde{c}, \widetilde{M}, \widetilde{W}_T)$ for $\forall (\widetilde{c}, \widetilde{M}, \widetilde{W}_T) \in L^1(\bar{\lambda} \times Q_0; \mathbb{R}^3)$. This shows $(\widetilde{c}^*, \widetilde{M}^*, \widetilde{W}_T^*)$ solves the primal problem. □

C.7 Proof of Lemma 4.6.1

By the definitions (4.32) and (4.33), we can apply dynamic programming principle to derive the following Hamilton–Jacobi–Bellman(HJB) equation

$$0 = -\widetilde{\delta} V_B(t, W_t) + \frac{\partial V_B}{\partial t} + \frac{\partial V_B}{\partial W} r(t) W_t - \frac{1}{2}\kappa_{0,t}^2 \left(\frac{\partial V_B}{\partial W}\right)^2 / \frac{\partial^2 V_B}{\partial W^2} + \frac{\gamma}{1-\gamma}\left(\frac{\partial V_B}{\partial W}\right)^{-\frac{1-\gamma}{\gamma}} \tag{C.52}$$

From (4.34), we can derive the following derivatives

$$\begin{aligned} \frac{\partial V_B}{\partial t} &= -\frac{\gamma}{1-\gamma} W_t^{1-\gamma} F_B(t)^{\gamma-1} \\ &\quad + \frac{\gamma}{1-\gamma} W_t^{1-\gamma} F_B(t)^{\gamma} \left\{ \frac{\widetilde{\delta}}{\gamma} + \frac{\gamma-1}{\gamma} r(t) + \frac{1}{2}\frac{\gamma-1}{\gamma^2}\kappa_{0,t}^2 \right\} \\ \frac{\partial V_B}{\partial W} &= W_t^{-\gamma} F_B(t)^{\gamma} \\ \frac{\partial^2 V_B}{\partial W^2} &= -\gamma W_t^{-\gamma-1} F_B(t)^{\gamma} \end{aligned}$$

Substitute these derivatives into (C.52), we have

$$
\begin{aligned}
&-\frac{\widetilde{\delta}}{1-\gamma}W_t^{1-\gamma}F_B(t)^{\gamma}-\frac{\gamma}{1-\gamma}W_t^{1-\gamma}F_B(t)^{\gamma-1}\\
&+\frac{\gamma}{1-\gamma}W_t^{1-\gamma}F_B(t)^{\gamma}\left[\frac{\widetilde{\delta}}{\gamma}+\frac{\gamma-1}{\gamma}r(t)+\frac{1}{2}\frac{\gamma-1}{\gamma_2^2}\kappa_{0,t}^2\right]\\
&+W_t^{-\gamma}F_B(t)^{\gamma}r(t)W_t+\frac{1}{2\gamma}W_t^{1-\gamma}F_B(t)^{\gamma}\kappa_{0,t}^2+\frac{\gamma}{1-\gamma}W_t^{1-\gamma}F_B(t)^{\gamma-1},
\end{aligned}
$$

which equals zero. Therefore, (4.34) is the explicit solution to (C.52).

C.8 Proof of Proposition 4.6.1

Proof. First, we denote $(\alpha_v,\theta_v,c_v,I_v)$ as the general strategy and $((\alpha_v)^*,(\theta_v)^*,(c_v)^*,(I_v)^*)$ as the optimal strategy under artificial market $\mathcal{M}_v$. Then, according to the optimal wealth $W_{v,t}$ in (4.24), we can restrict the static budget constraint to the following form

$$
\begin{aligned}
W_{v,t}=E^{Q_v}&\left[\int_t^T e^{-\int_t^s r(u)+v_0(u)+\lambda_{x+u}du}[c_{v,s}-Y_s+\lambda_{x+s}M_{v,s}-\delta(v(s))]ds\right.\\
&\left.+e^{-\int_t^T r(s)+v_0(s)+\lambda_{x+s}ds}W_{v,T}|\mathcal{F}_t\right].
\end{aligned}
$$

Therefore,

$$
H_{v,t}=\beta_{v,t}e^{-\int_0^t\lambda_{x+s}ds}W_{v,t}+\int_0^t\beta_{v,s}e^{-\int_0^s\lambda_{x+u}du}[c_{v,s}-Y_s+\lambda_{x+s}M_{v,s}-\delta(v(s))]ds \tag{C.53}
$$

is a Q_v-martingale for $v\in\mathcal{N}^*$. Next, by martingale presentation theorem, there exists a $\mathbb{R}$-valued process Ψ_v with $\int_0^T|\Psi_{v,t}|^2dt<\infty$, such that

$$
H_{v,t}=W_{v,0}+\int_0^t\Psi_{v,s}dZ_{v,s}. \tag{C.54}
$$

Substitute (C.54) into (C.53), we derive

$$
\begin{aligned}
W_{v,t}&=\beta_{v,t}^{-1}e^{\int_0^t\lambda_{x+s}ds}\left\{H_{v,t}-\int_0^t\beta_{v,s}e^{-\int_0^s\lambda_{x+u}du}[c_{v,s}-Y_s+\lambda_{x+s}M_{v,s}-\delta(v(s))]ds\right\}\\
&=\beta_{v,t}^{-1}e^{\int_0^t\lambda_{x+s}ds}\left\{W_{v,0}+\int_0^t\Psi_{v,s}dZ_{v,s}\right.\\
&\qquad\left.-\int_0^t\beta_{v,s}e^{-\int_0^s\lambda_{x+u}du}[c_{v,s}-Y_s+\lambda_{x+s}M_{v,s}-\delta(v(s))]ds\right\}.
\end{aligned}
$$

By Ito's formula and change of measure (4.12), we obtain

$$\begin{aligned} dW_{v,t} &= (r(t) + v_0(t) + \lambda_{x+t})W_{v,t}dt \\ &\quad + \beta_{v,t}^{-1} e^{\int_0^t \lambda_{x+s}ds} \Psi_{v,t}[dZ_t + \sigma^{-1}(t)(\mu(t) + v_-(t) - (r(t) + v_0(t)))dt] \\ &\quad - [c_{v,t} - Y_t + \lambda_{x+t}M_{v,t} - \delta(v(t))]dt. \end{aligned} \tag{C.55}$$

If we choose $\Psi_{v,t} = \beta_{v,t} e^{-\int_0^t \lambda_{x+s}ds} \sigma(t)\theta_{v,t}$ and rewrite $M_{v,t} = W_{v,t} + \frac{I_{v,t}}{\lambda_{x+t}}$, then (C.55) can be simplified to

$$\begin{aligned} dW_{v,t} &= [r(t)\alpha_{v,t} + \theta_{v,t}\mu(t)]dt + [\alpha_{v,t}v_0(t) + \theta_{v,t}v_-(t) + \delta(v(t))]dt \qquad \text{(C.56)} \\ &\quad + \sigma(t)\theta_{v,t}dZ_t - (c_{v,t} + I_{v,t} - Y_t)dt, \\ W_{v,0} &= w_0, (\alpha_v, \theta_v) \in \mathbb{R}^2. \end{aligned}$$

which has no free disposal. Here, we enlarge the domain of (α_v, θ_v) to $\mathbb{R}^2$ because $(\alpha_v, \theta_v) \in A$ (see (4.28)) is not guaranteed. By the definition (4.10), we have $v_0(t)\alpha_{v,t} + v_-(t)\theta_{v,t} + \delta(v(t)) \geq 0$ for $(\alpha_v, \theta_v) \in A$. Therefore, the wealth process (C.56) is bigger and equal to the wealth process (4.30) almost surely for $(\alpha_v, \theta_v) \in A$. Moreover, since $A \subset \mathbb{R}^2$, optimizing the objective function $J(c_v, M_v, W_{v,T})$ under the wealth process (C.56) with $(\alpha_v, \theta_v) \in \mathbb{R}^2$ provides an upper bound for the optimal objective function $J(c_v, M_v, W_{v,T})$ under the wealth process (4.30) with $(\alpha_v, \theta_v) \in A$. In other words, the expected utility of an individual who invests freely following (C.56) under artificial market $\mathcal{M}_v$ provides an upper bound for the primal problem. That is how we find the upper bound. For $t \in [T_R, T]$, SDE (C.56) equals

$$\begin{aligned} dW_{v,t} &= \{\alpha_{v,t}[r(t) + v_0(t)] + \theta_{v,t}[\mu(t) + v_-(t)]\}dt + \sigma(t)\theta_{v,t}dZ_t - (c_{v,t} + I_{v,t} - \delta(v(t)))dt \\ &= \{[r(t) + \lambda_{x+t} + v_0(t)]W_{v,t} + \theta_{v,t}[\mu(t) + v_-(t) - (r(t) + v_0(t))]\}dt + \theta_{v,t}\sigma(t)dZ_t \\ &\quad - [c_{v,t} + \lambda_{x+t}M_{v,t} - \delta(v(t))]dt. \end{aligned}$$

Define the value function $\widetilde{J}_R(t, W_{v,t}; v)$ as

$$\begin{aligned} \widetilde{J}_R(t, W_{v,t}; v) &= \sup_{\theta_v, c_v, M_v} E_t \left[\int_t^T e^{-\int_t^s \lambda_{x+u}du - \widetilde{\delta}(s-t)} \frac{(c_{v,s})^{1-\gamma}}{1-\gamma} ds \right. \\ &\quad + \int_t^T \lambda_{x+s} e^{-\int_t^s \lambda_{x+u}du - \widetilde{\delta}(s-t)} \frac{(M_{v,s})^{1-\gamma}}{1-\gamma} g(s)^\gamma ds \\ &\quad \left. + e^{-\int_t^T \lambda_{x+u}du - \widetilde{\delta}(T-t)} \frac{(W_{v,T})^{1-\gamma}}{1-\gamma} \right]. \end{aligned}$$

By the dynamic programming principal, we derive the HJB equation

$$\begin{aligned} 0 &= -(\lambda_{x+t}+\widetilde{\delta})\widetilde{J}_R(t,W_{v,t};v)+\frac{\partial \widetilde{J}_R}{\partial t}+\frac{\partial \widetilde{J}_R}{\partial W_v}[(r(t)+\lambda_{x+t}+v_0(t))W_{v,t}+\delta(v(t))] \\ & -\frac{1}{2[\sigma(t)]^2\frac{\partial^2 \widetilde{J}_R}{\partial (W_v)^2}}\left(\frac{\partial \widetilde{J}_R}{\partial W_v}\right)^2[\mu(t)+v_-(t)-(r(t)+v_0(t))]^2 \\ & +\frac{\gamma}{1-\gamma}[1+\lambda_{x+t}g(t)]\left(\frac{\partial \widetilde{J}_R}{\partial W_v}\right)^{\frac{\gamma-1}{\gamma}}, \end{aligned} \tag{C.57}$$

together with the optimal strategies

$$(\theta_{v,t})^* = \min\left\{\max\left\{\frac{\kappa_{v,t}}{\sigma(t)\frac{\partial^2 \widetilde{J}_R}{\partial (W_v)^2}}\frac{\partial \widetilde{J}_R}{\partial W_v},0\right\},W_{v,t}\right\}, \tag{C.58}$$

$$(c_{v,t})^* = \left(\frac{\partial \widetilde{J}_R}{\partial W_v}\right)^{-\frac{1}{\gamma}}, \quad (M_{v,t})^* = \left(\frac{\partial \widetilde{J}_R}{\partial W_v}\right)^{-\frac{1}{\gamma}} g(t). \tag{C.59}$$

For (4.36), we can derive the following derivatives

$$\begin{aligned} \frac{\partial \widetilde{J}_R}{\partial t} &= \widetilde{F}_1(t,W_{v,t})^{-\gamma}\{-\delta(v(t)) \\ & +[r(t)+v_0(t)+\lambda_{x+t}]\int_t^T e^{-\int_t^s \lambda_{x+u}du}\delta(v(s))F_2(s-t,s)ds\Big\}\widetilde{F}_2(t)^\gamma \\ & +\frac{\gamma}{1-\gamma}\widetilde{F}_1(t,W_{v,t})^{1-\gamma}\widetilde{F}_2(t)^{\gamma-1}\{-[1+\lambda_{x+t}g(t)] \\ & +\left[\lambda_{x+t}+\frac{\widetilde{\delta}}{\gamma}+\frac{\gamma-1}{\gamma}(r(t)+v_0(t))+\frac{1}{2}\frac{\gamma-1}{\gamma^2}\kappa_{v,t}^2\right]\widetilde{F}_2(t)\Big\}, \\ \frac{\partial \widetilde{J}_R}{\partial W_v} &= \widetilde{F}_1(t,W_{v,t})^{-\gamma}\widetilde{F}_2(t)^\gamma \\ \frac{\partial^2 \widetilde{J}_R}{\partial (W_v)^2} &= -\gamma\widetilde{F}_1(t,W_{v,t})^{-\gamma-1}\widetilde{F}_2(t)^\gamma. \end{aligned}$$

Plug these derivatives into the equation (C.57), we have

$$-(\lambda_{x+t}+\widetilde{\delta})\frac{1}{1-\gamma}\widetilde{F}_1(t,W_{v,t})^{1-\gamma}\widetilde{F}_2(t)^\gamma$$

$$
\begin{aligned}
&+\widetilde{F}_1(t,W_{v,t})^{-\gamma}\{-\delta(v(t)) \\
&+[r(t)+v_0(t)+\lambda_{x+t}]\int_t^T e^{-\int_t^s \lambda_{x+u}du}\delta(v(s))F_2(s-t,s)ds\Big\}\widetilde{F}_2(t)^\gamma \\
&+\frac{\gamma}{1-\gamma}\widetilde{F}_1(t,W_{v,t})^{1-\gamma}\widetilde{F}_2(t)^{\gamma-1}\Big\{-[1+\lambda_{x+t}g(t)]+\Big[\lambda_{x+t}+\frac{\widetilde{\delta}}{\gamma}+\frac{\gamma-1}{\gamma}(r(t)+v_0(t)) \\
&+\frac{1}{2}\frac{\gamma-1}{\gamma^2}\kappa_{v,t}^2\Big]\widetilde{F}_2(t)\Big\}+\widetilde{F}_1(t,W_{v,t})^{-\gamma}\widetilde{F}_2(t)^\gamma[(r(t)+\lambda_{x+t}+v_0(t))W_{v,t}+\delta(v(t))] \\
&+\frac{1}{2\gamma[\sigma(t)]^2\widetilde{F}_1(t,W_{v,t})^{-\gamma-1}\widetilde{F}_2(t)^\gamma}\widetilde{F}_1(t,W_{v,t})^{-2\gamma}\widetilde{F}_2(t)^{2\gamma}[\mu(t)+v_-(t)-(r(t)+v_0(t))]^2 \\
&+\frac{\gamma}{1-\gamma}\widetilde{F}_1(t,W_{v,t})^{1-\gamma}\widetilde{F}_2(t)^{\gamma-1}(1+\lambda_{x+t}). \qquad \text{(C.60)}
\end{aligned}
$$

After tedious calculation, we simplify (C.60) to the following form

$$
\begin{aligned}
&\widetilde{F}_1(t,W_{v,t})^{1-\gamma}\widetilde{F}_2(t)^\gamma\Big\{-\frac{1}{1-\gamma}(\lambda_{x+t}+\widetilde{\delta})+\frac{\gamma}{1-\gamma}\lambda_{x+t}+\frac{\widetilde{\delta}}{1-\gamma}-[r(t)+v_0(t)] \\
&-\frac{1}{2\gamma}\kappa_{v,t}^2+\frac{1}{2\gamma}\kappa_{v,t}^2+r(t)+v_0(t)+\lambda_{x+t}\Big\},
\end{aligned}
$$

which equals zero. Therefore, the value function $\widetilde{J}_R(t,W_{v,t};v)$ is the solution to (C.57). Moreover, substitute (4.36) into the optimal strategies (C.58) and (C.59), we obtain (4.37) and (4.38). □

C.9 Proof of Proposition 4.6.2

Proof. For $t\in[0,T_R]$, SDE (C.56) equals

$$
\begin{aligned}
dW_{v,t} &= [\alpha_{v,t}(r(t)+v_0(t))+\theta_{v,t}(\mu(t)+v_-(t))]dt+\sigma(t)\theta_{v,t}dZ_t \\
&\quad -[c_{v,t}+I_{v,t}-Y_t-\delta(v(t))]dt \\
&= \{(r(t)+\lambda_{x+t}+v_0(t))W_{v,t}+\theta_{v,t}[\mu(t)+v_-(t)-(r(t)+v_0(t))]\}dt+\theta_{v,t}\sigma(t)dZ_t \\
&\quad -[c_{v,t}+\lambda_{x+t}M_{v,t}-Y_t-\delta(v(t))]dt.
\end{aligned}
$$

Define the value function $\widetilde{J}(t, W_{v,t}, Y_t; v)$ as

$$\begin{aligned}\widetilde{J}(t, W_{v,t}, Y_t; v) &= \sup_{\theta_v, c_v, M_v} E_t\left[\int_t^{T_R} e^{-\int_t^s \lambda_{x+u}du-\widetilde{\delta}(s-t)}\frac{(c_{v,s})^{1-\gamma}}{1-\gamma}ds\right.\\ &\quad+\int_t^{T_R}\lambda_{x+s}e^{-\int_t^s \lambda_{x+u}du-\widetilde{\delta}(s-t)}\frac{(M_{v,s})^{1-\gamma}}{1-\gamma}g(s)ds\\ &\quad\left.+e^{-\int_t^{T_R}\lambda_{x+u}du-\widetilde{\delta}(T_R-t)}J_R(T_R, W_{v,T_R}; v)\right].\end{aligned}$$

By the dynamic programming principal, we derive the HJB equation

$$\begin{aligned}&0=-(\lambda_{x+t}+\widetilde{\delta})\widetilde{J}(t, W_{v,t}, Y_t; v)+\frac{\partial\widetilde{J}}{\partial t}+\frac{\partial\widetilde{J}}{\partial W_v}[(r(t)+\lambda_{x+t}+v_0(t))W_{v,t}+Y_t+\delta(v(t))]\\ &+\frac{\partial\widetilde{J}}{\partial Y}\mu_Y Y_t+\frac{1}{2}\frac{\partial^2\widetilde{J}}{\partial Y^2}\sigma_Y^2 Y_t^2-\frac{1}{2\frac{\partial^2\widetilde{J}}{\partial(W_v)^2}}\left(\frac{\partial\widetilde{J}}{\partial W_v}\kappa_{v,t}-\frac{\partial^2\widetilde{J}}{\partial W_v\partial Y}\sigma_Y Y_t\right)^2\\ &+\frac{\gamma}{1-\gamma}[1+\lambda_{x+t}g(t)]\left(\frac{\partial\widetilde{J}}{\partial W_v}\right)^{\frac{\gamma-1}{\gamma}},\\ &\widetilde{J}(T_R, W_{v,T_R}, Y_{T_R}; v)=\widetilde{J}_R(T_R, W_{v,T_R}; v),\end{aligned}\tag{C.61}$$

together with the optimal strategies

$$(\theta_{v,t})^* = \min\left\{\max\left\{\frac{1}{\sigma(t)\frac{\partial^2\widetilde{J}}{\partial(W_v)^2}}\left(\frac{\partial\widetilde{J}}{\partial W_v}\kappa_{v,t}-\frac{\partial^2\widetilde{J}}{\partial W_v\partial Y}\sigma_Y Y_t\right), 0\right\}, W_{v,t}\right\},\tag{C.62}$$

$$(c_{v,t})^* = \left(\frac{\partial\widetilde{J}}{\partial W_v}\right)^{-\frac{1}{\gamma}}, \quad (M_{v,t})^* = \left(\frac{\partial\widetilde{J}}{\partial W_v}\right)^{-\frac{1}{\gamma}} g(t).\tag{C.63}$$

For (4.39), we can obtain the following derivatives

$$\begin{aligned}\frac{\partial\widetilde{J}}{\partial t} &= \widetilde{F}_3(t, W_{v,t}, Y_t)^{-\gamma}\widetilde{F}_2(t)^{\gamma}\{-Y_t-\delta(v(t))\\ &\quad -Y_t(\mu_Y+\kappa_{v,t}\sigma_Y)\int_t^{T_R} e^{-\int_t^s\lambda_{x+u}du}F_1(s-t, s)ds\\ &\quad +(r(t)+v_0(t)+\lambda_{x+t})(\widetilde{F}_3(t, W_{v,t}, Y_t)-W_{v,t})\Big\}\\ &\quad +\frac{\gamma}{1-\gamma}\widetilde{F}_3(t, W_{v,t}, Y_t)^{1-\gamma}\widetilde{F}_2(t)^{\gamma-1}\{-(1+\lambda_{x+t}g(t))\end{aligned}$$

$$
\begin{aligned}
& +\widetilde{F}_2(t)\left[\lambda_{x+t}+\frac{\widetilde{\delta}}{\gamma}+\frac{\gamma-1}{\gamma}(r(t)+v_0(t))\right. \\
& \left.\left.+\frac{1}{2}\frac{\gamma-1}{\gamma^2}\kappa_{v,t}^2\right]\right\}, \\
\frac{\partial \widetilde{J}}{\partial W_v} &= \widetilde{F}_3(t,W_{v,t},Y_t)^{-\gamma}\widetilde{F}_2(t)^{\gamma}, \\
\frac{\partial^2 \widetilde{J}}{\partial (W_v)^2} &= -\gamma\widetilde{F}_3(t,W_{v,t},Y_t)^{-\gamma-1}\widetilde{F}_2(t)^{\gamma}, \\
\frac{\partial \widetilde{J}}{\partial Y} &= \widetilde{F}_3(t,W_{v,t},Y_t)^{-\gamma}\widetilde{F}_2(t)^{\gamma}\int_t^{T_R} e^{-\int_t^s \lambda_{x+u}du}F_1(s-t,s)ds, \\
\frac{\partial^2 \widetilde{J}}{\partial Y^2} &= -\gamma\widetilde{F}_3(t,W_{v,t},Y_t)^{-\gamma-1}\widetilde{F}_2(t)^{\gamma}\left(\int_t^{T_R} e^{-\int_t^s \lambda_{x+u}du}F_1(s-t,s)ds\right)^2, \\
\frac{\partial^2 \widetilde{J}}{\partial W_v \partial Y} &= -\gamma\widetilde{F}_3(t,W_{v,t},Y_t)^{-\gamma-1}\widetilde{F}_2(t)^{\gamma}\int_t^{T_R} e^{-\int_t^s \lambda_{x+u}du}F_1(s-t,s)ds.
\end{aligned}
$$

Plug these derivatives into the HJB equation (C.61) and simplify it, we have

$$
\begin{aligned}
& -(\lambda_{x+t}+\widetilde{\delta})\frac{1}{1-\gamma}\widetilde{F}_3(t,W_{v,t},Y_t)^{1-\gamma}\widetilde{F}_2(t)^{\gamma}+\widetilde{F}_3(t,W_{v,t},Y_t)^{1-\gamma}\widetilde{F}_2(t)^{\gamma}(r(t)+v_0(t)+\lambda_{x+t}) \\
& -\frac{\gamma}{1-\gamma}\widetilde{F}_3(t,W_{v,t},Y_t)^{1-\gamma}\widetilde{F}_2(t)^{\gamma-1}[1+\lambda_{x+t}g(t)] \\
& +\frac{\gamma}{1-\gamma}\widetilde{F}_3(t,W_{v,t},Y_t)^{1-\gamma}\widetilde{F}_2(t)^{\gamma}\left[\lambda_{x+t}+\frac{\widetilde{\delta}}{\gamma}+\frac{\gamma-1}{\gamma}(r(t)+v_0(t))+\frac{1}{2}\frac{\gamma-1}{\gamma^2}\kappa_{v,t}^2\right] \\
& +\frac{1}{2\gamma}\widetilde{F}_3(t,W_{v,t},Y_t)^{1-\gamma}\widetilde{F}_2(t)^{\gamma}\kappa_{v,t}^2+\frac{\gamma}{1-\gamma}[1+\lambda_{x+t}g(t)]\widetilde{F}_3(t,W_{v,t},Y_t)^{1-\gamma}\widetilde{F}_2(t)^{\gamma-1},
\end{aligned}
$$

which equals zero. Therefore, the value function $J(t,W_{v,t},Y_t;v)$ is the solution to (C.61). Moreover, substitute (4.39) into (C.62) and (C.63), we obtain the optimal strategies (4.40) and (4.41). □