Risk Allocation in Hybrid Pension Plans

by

Xiaobai Zhu

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, 2019

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: **An Chen**
Professor, Faculty of Mathematics and Economics
Ulm University

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

David Saunders
Associate Professor, Department of Statistics and Actuarial Science
University of Waterloo

Internal Members: **Ken Seng Tan**
Professor, Department of Statistics and Actuarial Science
University of Waterloo

Chengguo Weng
Associate Professor, Department of Statistics and Actuarial Science
University of Waterloo

Internal-External Member: **James R. Thompson**
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 1 is published in the ASTIN Bulletin: The Journal of the IAA under the paper title *Dynamic Hedging Strategies for Cash Balance Pension Plans*, by Xiaobai Zhu, Mary R. Hardy, and David Saunders (Zhu *et al.* (2018b)).

Chapter 2 is published in the Scandinavian Actuarial Journal under the paper title *Valuation of an Early Exercise Defined Benefit Underpin Hybrid Pension*, by Xiaobai Zhu, Mary R. Hardy, and David Saunders (Zhu *et al.* (2018a).

Abstract

Due to increases in life expectancy and decreases in fertility rates, pension plans around the world are facing sustainability challenges. Pension sponsors are shifting risks back to employees by moving away from Defined Benefit (DB) plans to Defined Contribution (DC) plans. However, since members bear all of the inflation risk, the investment risk, and the longevity risk, there is growing concern about the financial security offered through DC plans. Hybrid pension plans, which combine features of both DC and DB plans, may be able to meet the needs of employer and employee better than DC or DB individually. Numerous hybrid pension designs have been proposed. Examples include Cash Balance plans, DB Underpin plans, Second-Election plans, Collective DC plans and Target Benefit plans. Although the overall hybrid pension market remains small compared to either DC or DB plans, their significance is growing.

Generally speaking, there are four stakeholder groups involved in any occupational pension plans: the pension sponsor, the active employees, the retirees and the regulatory authority. The aim of my thesis is to study different types of risk allocation strategies between stakeholders in the hybrid pension plans.

Three different hybrid designs are considered in this thesis. Chapter 1 studies the risk management of Cash Balance (CB) pension plans, which are the most popular hybrid plans in U.S. We show that the simple delta and gamma hedging is effective in mitigating the interest rate risk embedded in the CB liabilities. Chapter 2 introduces a new pension design based on two existing hybrid plans - the second election option and the DB underpin plan. We demonstrate how our new plan is able to provide more income security to the retirees without incurring large costs on top of the other two existing hybrid plans. Chapter 3 explores the optimal design for intergenerational risk sharing plans. We demonstrate the necessity to incorporate regulatory constraints to ensure the generational fairness. The last chapter concludes the thesis, and proposes future work in hybrid pension designs.

Acknowledgements

Words cannot express my gratitude to my supervisors Professor Mary Hardy and Professor David Saunders. I remember clearly back in the winter of 2013, I felt lost about my future career path. Since then, my research prospect has changed for the better. I have met with my supervisors, and opened the door to the world of research. Their endless encouragement has turned me to a confident researcher, and their inspiring guidance has shaped my research philosophy. The life lessons they taught me will be forever invaluable, and hopefully one day I may pass them to my students.

I would also like to acknowledge the committee members: Professor Ken Seng Tan, Professor Chengguo Weng, Professor James Thompson, Professor An Chen and my Master Thesis committee member Professor Phelim Boyle for all those insightful comments that significantly enhanced my research content. I would also like to mention here, Professor Johnny Li and Kenneth. Thank you for your collaborations that expanded my research area.

I would like to acknowledge the financial support from the Department of Statistics and Actuarial Science, University of Waterloo and National Pension Hub to relieve my financial stress.

I would like to thank our coordinator Mrs. Mary Lou, who helped me gone through all the boring paperwork necessary for graduation. As well, I would like to thank my friends who shared (most time unwillingly) their offices with me and listen to my rubbish research ideas.

Finally, I would like to express my profound gratitude to my parents for their unwavering understanding of my career decisions, and to my wife Feng Man, I know my life became better because of you.

Dedication

This is dedicated to my family.

Table of Contents

List of Tables

List of Figures

Introduction

Over the past few decades, traditional Defined Benefit (DB) and Defined Contribution (DC) plans have revealed their shortcomings in providing adequate and stable retirement income. Many existing DB plans are facing sustainability issues, and current employees are not expecting to receive their promised retirement income in full. On the other hand, members in the DC plans suffer great uncertainty in both account accumulation and decumulation phases. As a result, there is a growing interest in searching for alternative pension structures to control the variability in both contributions and benefits; those plans are generally referred to as Hybrid Pension plans.

The aim of my thesis is to study different types of hybrid pension designs, specifically, Cash Balance pension plans, the Early Exercise DB underpin plans, and a stylized Intergenerational Risk Sharing plan. The topics covered include liability hedging, liability valuation, and optimal pension structure.

Chapter 1 presents the performance of dynamic hedging strategies for Cash Balance (CB) pension plans. The CB design is increasing in popularity in the U.S. market, with more than 12 million participants in 2016. It preserves the portability of DC plan with market risk being shared between employees and sponsor. Our study is motivated by Hardy *et al.* (2014), which studies market consistent valuation of CB plans. Harvey (2012) shows that the common view from the industry that CB plans have low risk is in fact a misbelief. There are substantial interest rate risks associated with CB liabilities, and our task is to explore whether such risks can be managed. In this chapter, we consider delta and gamma hedging strategies, which are simple to implement and are consistent with our valuation technique. We demonstrate their effectiveness in mitigating interest rate risk.

Chapter 2 introduces a new hybrid pension design, the Early Exercise DB Underpin plan. The intuition behind the new pension scheme is inspired by two existing hybrid pension designs: the DB underpin plan and the second election plan. The sponsors of DB underpin plans grant a minimum guarantee on top of a DC plan at the retirement date. Thus, the downside market risk is shared between employees and sponsor. The second election option is a one-time option that allows the members to switch between DC and DB plan at anytime prior to the retirement date. Any transition cost will be borne by the employee. One significant drawback of the second election plan is the possibility of a "lose-lose" situation where the cumulative cost incurred is higher than either the DB or the DC plan but the employees receive less benefit. This motivates us to incorporate a minimum guarantee at the switching time, which is similar to a DB underpin plan with early exercise option. Since the new plan is a natural link between the second election plan and the DB underpin plan, we will jointly study the three plans and analyze their pros and cons.

Chapter 3 explores optimal design for intergenerational risk sharing (IRS) pension plans under which the investment and the longevity risks are shared between active employees and retirees. Most of the IRS plans are in their preliminary discussion stage. They are also known as collective DC plans, conditional indexation plans, defined ambition plans and target benefit plans. Our study is based on a stylized plan suggested by Cui *et al.* (2011). Both contributions and pension rights are adjustable based on the level of pension asset. We formulate an ergodic control problem to ensure the stability of member's lifetime consumption. We provide a natural extension to the original plan with theoretical justification. In addition, we illustrate the importance of including the regulatory requirements for the sake of fairness across different generations.

Chapter 4 presents the possible directions for future research in hybrid pension plans.

Chapter 1

Cash Balance Pension - Dynamic Hedging

1.1 Introduction

Cash Balance (CB) pensions represent the fastest growing pension plan design in the US, according to Kravitz (2018). The number of CB plans has increased from around 3% to 37% of all Defined Benefit (DB) plans since 2001. In 2016 there were 20,452 CB plans in the US, with a total of around 12 million participants.

CB plans are classified and regulated as DB plans but are presented to look very much like Defined Contribution (DC) plans. For example, participants have individual accounts showing their up-to-date accrued benefits. However, an important difference between CB and DC plans is that employee accounts in the CB plan are notional. The assets are not allocated to individuals and the amounts paid into the aggregated plan funds need not be equal to the notional contributions. The total funds invested are generally not equal to the sum of the employee accounts, and are often substantially smaller because of the actuarial funding methods used, and because DB plans are not required to be fully funded (see Hardy *et al.* (2014)). The individual member accounts are notionally accumulated at the specified plan crediting rate, which may be a fixed rate, or a variable rate such as the current yield on government bonds of specified term. In this chapter we focus on the 30-year Treasury crediting rate; it is one of the most common options, (covering around 24% of plans (34% of plans with more than 100 participants), according to Kravitz (2018)).

The original motivation for the CB plan design was to replicate the lower risk properties of DC plans within a DB plan. CB plans started gaining popularity in the mid-1990s;

Niehaus and Yu (2005) explain that in 1990s, companies switching from traditional DB plans favored CB plans over DC, to avoid the high reversion tax applied to excess pension assets. Coronado and Copeland (2004) on the other hand conclude that the conversions from traditional DB to CB were primarily driven by the labor market conditions. In 2015, new regulations came into effect clarifying and extending the permitted range of crediting rate formulae[1].

Until recently, CB plans have been considered by sponsors and most actuaries to be low risk, low volatility plans. The views expressed by the consultants at Kravitz are typical, for example:

"*Cash balance plans remove the interest rate risk that led to constantly changing value of liabilities in traditional defined benefit plans.*" (Kravitz (2018))

More recently, there has been a greater recognition of the inherent risk involved in guaranteeing returns. In a 2015 report, Segal Rogerscasey wrote:

"*Cash balance plans do not represent a low-risk arbitrage opportunity, but rather are a leveraged investment in risk assets, similar to a traditional DB plan.*" (Rogerscasey (2015)).

Hedging the cash balance liability has become a more urgent question as the low interest rate environment has persisted so long that rising costs and risks cannot be ignored. Even the traditional actuarial funding approach, which is highly unresponsive to market conditions, is beginning to generate unexpected strains and volatility.

The major objective of this chapter is to derive and quantitatively explore the efficacy of a hedging strategy based on financial engineering models and principles. We base the hedge on the market valuation of the liability, rather than the actuarial valuation. There are two reasons for this. The first is that the application of traditional actuarial valuation methods to CB plans generates an 'actuarial liability' that is quite unrelated to the short or long term costs. Both Murphy (2001) and Hardy *et al.* (2014) conclude that traditional methods may significantly understate the actual liability for participants who terminate early, and tend to generate losses even where the participant exits at the assumed age. Market valuation methods are more objective than actuarial methods and the valuation and hedging formulae are inextricably linked; the valuation formula can be used to derive the hedging portfolio for a risk. In Hardy *et al.* (2014) explicit valuation formulae were derived in the case where the crediting rate is the k-year spot rate, and where interest crediting applies continuously. Thus, the valuation results give us a starting point for determining a hedge strategy.

[1] Internal Revenue Bulletin, T.D.9743, 2015-48 I.R.B.679, "Transitional Amendments to Satisfy the Market Rate of Return Rules for Hybrid Retirement Plans".

In practice CB crediting rates based on treasury bonds use the yield to maturity (YTM), not spot rates. The same valuation framework that we used for spot rates can be applied to the YTM, but the valuation requires Monte Carlo simulation. Hardy *et al.* (2014) showed that the numerical results using the spot rate valuation formulae were a good approximation for the YTM valuation. In this chapter we show that the same is true for hedging strategies based on spot rates and YTMs. We also use this chapter to extend the valuation formulae in Hardy *et al.* (2014) to the case where the interest crediting is discrete.

Market consistent valuation techniques have been studied for other pension-related liabilities; examples can be found in Boyle and Hardy (2003), Marshall (2011) and Chen and Hardy (2009). Despite the fact that US public and private sector pension plans have not (yet) adopted market consistent valuation, much of the recent pensions literature has discussed the benefit of using market values to assess objectively the funding status of pension plans. Novy-Marx and Rauh (2009) and Biggs (2010) recalculate the funding ratios of US public pensions to market-consistent benchmarks, and conclude that the subjectivity of traditional actuarial techniques significantly understates the liabilities, and overstates the funding rates. Moreover, Biggs and Smetters (2013) outline common misconceptions about using a long-term market rate as the discount rate for pension liabilities. In this chapter, we reinforce the advantage of market consistent valuation from a risk management perspective, as it provides a natural approach not only to the valuation, but also to the appropriate investment strategies to hedge the interest rate risk.

The CB payoff based on treasury bond crediting rates can be viewed as an interest rate derivative. Compare to pricing of interest rate derivatives, hedging tends to be more sensitive to the model risk. Hardy *et al.* (2014) show that the price of the CB payoff evaluated using the one factor Hull-White (HW) model is very close to the price evaluated with the two-factor HW model. However, that may not be the case for the hedge strategy. Several researchers have examined how many factors are required for effective hedging of interest rate risks. Fan *et al.* (2001) studied up to four factor models in the swaption market and conclude that although low-dimensional models are capable of accurately pricing swaptions, they are not sufficient for hedging purposes. Gupta and Subrahmanyam (2005) and Driessen *et al.* (2003) draw a similar conclusion for the cap/floor market. Their work is closer to ours, as they used delta hedging strategies, whereas Fan *et al.* (2001) focus on bucket hedging.

Other research on hedging CB liabilities includes Brown *et al.* (2001), who studied a duration based hedging strategy, and Harvey (2012), who discussed the difficulties in calculating the duration of a CB plan, and outlined several practical investment strategies including credit default swaps, Treasury futures and swaptions, but without supporting theoretical or empirical analysis.

The remainder of the chapter is structured as follows. Section 1.2 introduces the assumptions, notation and models we adopt in this chapter, and also the valuation formulae. Section 1.3 presents the construction of hedging portfolios. Section 1.4 describes the accumulated hedge error, which we use to measure hedge effectiveness. Section 1.5 examines the hedge error arising from discrete rebalancing, without model or parameter risk. In Section 1.6 we consider the impact of model and parameter error by analyzing the hedging effectiveness using real historical interest rates. In Section 1.7 we consider basis risk, by comparing the results using a spot rate crediting rate, which is the more tractable approach, with results using the YTM crediting rate which is what is used in practice. In Section 1.8 we briefly compare the results of hedging with results where the assets are invested in a traditional DB equity/bond portfolio and in Section 1.9 we introduce early exits through a brief example. Section 1.10 concludes.

1.2 Model and Assumptions

In this section we outline the formulae and assumptions used for valuing the cash balance pension liability, following Hardy *et al.* (2014) and Zhu (2015).

1.2.1 The CB Accrued Benefit

This chapter focuses on interest rate risk; in this section we ignore demographic risk, but we will consider early exits in Section 1.9. Treating the individual CB account balance at retirement as a contingent payment, the market value of the liability is the risk neutral expectation of the discounted payoff at termination.

Our valuation approach is accrual based, and so are all the hedging strategies we construct. Under the accruals principle (equivalent to unit credit in a traditional pension plan), the accrued liability is based only on past contributions. Future notional contributions into the plan can be each regarded as a separate derivative which will be evaluated and hedged in the same way as the existing account value at the time of payment. This means that our problem is to hedge the interest rate risk arising from the application of future, unknown crediting rates, to the current notional account value.

Viewing the pension plan as a whole, the hedge portfolio will comprise hedges based on a range of accrued funds and times to exit, but to illustrate we consider a single employee fund, with a T-year term to exit from the valuation date.

We denote by F_t the notional account value at time t, based on a fixed notional account value of F_0 at $t = 0$, and ignoring notional contributions after $t = 0$. The frequency of interest crediting varies between plans, but the most common choice is annual crediting. Let $i^c(s)$ denote the crediting rate applied in the year s to $s+1$. The exit date (which we assume is known in advance) is denoted T. The lump-sum payment at the retirement date is the random variable F_T. For $t = 0, 1, 2, ..., T$ we have

$$F_t = F_0 \prod_{s=0}^{t-1} (1 + i^c(s))$$

For the continuous crediting case we let $r^c(s)$ denote the continuously compounded crediting rate at time s, so that for $0 < t \leq T$

$$F_t = F_0 \, e^{\int_0^t r^c(s)ds}$$

The formulae for valuation and hedging are more straightforward for the continuous crediting case than for the discrete, so in the main part of the chapter we present the formulae and results using continuous crediting. The formulae for the discrete case are derived alongside the continuous versions in the appendices.

As in Hardy *et al.* (2014), we develop our results assuming that the crediting rate at t is equal to the k-year spot rate from the US government bond yield curve at that time. This is a convenient and reasonably accurate approximation to the most common form of crediting rate, that is the yield to maturity on k-year government bonds ((Kravitz, 2018)), with $k = 30$ being the most popular choice, and this is the rate used throughout the numerical illustrations in this chapter.

1.2.2 The Valuation Formula

Let $r(t)$ denote the continuously compounded short rate of interest at time t, let $P(t,T)$ denote the price at t of a zero coupon bond with face value of \$1, which matures at time T, and let $r_k(t)$ denote the k-year spot rate at time t. Then

$$P(t,T) = e^{-r_{T-t}(t)\,(T-t)} = \mathrm{E}_t^Q \left[e^{-\int_t^T r(s)ds} \right] \tag{1.1}$$

where E_t^Q denotes the risk neutral expectation given the information at t.

The market consistent value at time t of the payoff F_T due at time $T \geq t$ is the expected discounted payoff, using the risk neutral measure. Thus, at time $t > 0$, the value is

$$\begin{aligned} V_t &= \mathrm{E}_t^Q \left[F_T e^{-\int_t^T r(s)ds} \right] \\ &= F_t \, \mathrm{E}_t^Q \left[e^{\int_t^T r^c(s) - r(s)ds} \right] \end{aligned}$$

If $r^c(t)$ and $r(t)$ are independent, for example if the crediting rate is a constant, this becomes a simple formula. In the case where $r^c(t)$ is based on treasury rates at t, there is a strong dependence between the crediting rate and the short rate, and so we use a stochastic model of the yield curve for the joint distribution of $r^c(s)$ and $r(s)$.

We define the valuation factor $V(t,T)$ as the market value at time t per \$1 in the participant's account balance at time t, where the benefit matures at time $T \geq t$; that is

$$V(t,T) = \mathrm{E}_t^Q \left[e^{\int_t^T r^c(s) - r(s)ds} \right] \tag{1.2}$$

1.2.3 Short Rate Models

We use both the One-Factor and Two-Factor versions of the Hull-White (HW) model. Both of these are well recognized models; both allow a perfect match between the model and market starting yield curve, and both offer convenient analytical tractability[2]. For the one-factor HW model, the instantaneous short rate under the risk-neutral measure has the following SDE:

$$dr(t) = (\theta(t) - ar(t))dt + \sigma dW(t), \qquad r(0) = r_0$$

where $a > 0$, $\sigma > 0$ are constant, $\theta(t)$ is a deterministic function chosen to match the market term structure at the starting date, $W(t)$ is a standard Brownian motion under the Q measure, and r_0 is the observed short rate at time 0.

[2] One disadvantage of these models is that the short rate has a Gaussian distribution, with the possibility of being negative, but Hardy *et al.* (2014) show that this does not have a significant impact on the valuation factor.

There are two common parameterizations for the two-factor HW model. The one we adopt here is commonly referred to as G2++ (Brigo and Mercurio (2007)). The dynamics of the instantaneous short rate under Q are

$$
\begin{aligned}
& r(t) = x(t) + y(t) + \varphi(t), && r(0) = r_0 \\
& dx(t) = -a_1 x(t)dt + \sigma_1 dW_1(t), && x(0) = 0 \\
& dy(t) = -a_2 y(t)dt + \sigma_2 dW_2(t), && y(0) = 0 \\
& dW_1(t)dW_2(t) = \rho dt &&
\end{aligned}
$$

where r_0 is the initial observed short rate, a_1, a_2, σ_1,σ_2 are all positive constants, $\varphi(t)$ is the deterministic function used to match the initial term structure, and $(W_1(t), W_2(t))$ is a two-dimensional Brownian motion under the Q measure, with correlation ρ, $-1 \leq \rho \leq 1$.

The normality assumption of the short rate allows us to derive the valuation factors explicitly under both continuous and discrete crediting settings. The formulae are given in Appendix A. Interested readers can refer to Hardy *et al.* (2014) and Zhu (2015) for more detailed derivations and Brigo and Mercurio (2007) (Ch 3 & 4) for general pricing problems with one- and two- factor HW.

1.3 Constructing the Dynamic Hedging Portfolio

Delta hedging strategy attempts to neutralize the movement of the derivative price relative to the underlying asset. It is constructed naturally from the market-consistent valuation. In the case when delta itself is volatile, an additional Greek, Gamma, should be incorporated. In this section, we develop hedging strategies for a CB plan with crediting rate equal to the 30-year spot rate on Treasuries. We develop and compare three different strategies.

1. A delta hedge under the one-factor Hull-White model (requires two assets).
2. A delta-gamma hedge under the one-factor Hull-White model (requires three assets).
3. A delta hedge under the two-factor Hull-White model (requires three assets).

The hedging instruments we use are a money market account, with return equal to the short rate, and Treasury STRIPs of varying length. In practice, hedging with Treasury STRIPs may be impractical due to the size of the STRIPs market, but the techniques may be adapted to other fixed income securities, such as Treasury bonds, as we show in Section 1.7.2.

1.3.1 Greeks

The delta and gamma for a particular security are defined as the first and second order derivatives respect to the underlying risk factor. Here we provide the delta and gamma for the CB liability (subscript V) and for the zero coupon bond price (subscript B) under the one-factor HW model. In the following, recall that $t > 0$ is the valuation date, $T \geq t$ is the exit date, and k is the assumed term of the spot rate used as a crediting rate under the plan. The explicit solutions for the zero coupon bond price $P(t,T)$ (defined in equation (1.1)) and for the valuation factor $V(t,T)$ (defined in equation (1.2)) are given in Appendix A.

$$\Delta_B(t) = \frac{\partial P(t,T)}{\partial r(t)} = -B(a, T-t)P(t,T) \tag{1.3}$$

$$\Delta_V(t) = \frac{\partial V(t,T)}{\partial r(t)} = -\gamma(a,k)B(a, T-t)V(t,T) \tag{1.4}$$

$$\Gamma_B(t) = \frac{\partial^2 P(t,T)}{\partial r(t)^2} = B(a, T-t)^2 P(t,T) \tag{1.5}$$

$$\Gamma_V(t) = \frac{\partial^2 P(t,T)}{\partial r(t)^2} = \gamma(a,k)^2 B(a, T-t)^2 V(t,T) \tag{1.6}$$

$$\text{where} \quad B(a,s) = \frac{1-e^{-as}}{a}$$
$$\text{and} \quad \gamma(a,k) = 1 - \frac{B(a,k)}{k}$$

Under the two-factor HW model, the deltas for the CB liability and for the zero coupon bond are the first partial derivatives with respect to each stochastic driver (x and y).

$$\Delta_B^x(t) = -B(a_1, T-t)P(t,T)$$
$$\Delta_B^y(t) = -B(a_2, T-t)P(t,T)$$
$$\Delta_V^x(t) = -\gamma(a_1,k)B(a_1, T-t)V(t,T)$$
$$\Delta_V^y(t) = -\gamma(a_2,k)B(a_2, T-t)V(t,T)$$

Analogous formulae for discrete crediting are given in Appendix A.3.

1.3.2 Positions in hedging instruments

The hedging instruments we use for delta hedging under the one-factor HW model are the money market account, with return equal to the short rate, and a zero-coupon bond with maturity T_1.

For delta-gamma hedging under the one-factor HW model and for delta hedging under the two-factor HW model, we need a third asset, so we add another zero-coupon bond with maturity $T_2 \neq T_1$. The position in each instrument can be obtained by solving linear equation(s). For example, for delta hedging under the one-factor HW model, we solve for the position in the T_1-year zero coupon bond ($\Lambda_1^{HW}(t)$) as:

$$\Delta_V(t) = \Delta_B(t,T_1)\Lambda_1^{HW}(t) = -B(a,T-t)P(t,T)\Lambda_1^{HW}(t) \qquad \text{from (1.3)}$$

$$\Rightarrow -B(a,T-t)P(t,T)\Lambda_1^{HW}(t) = -\gamma(a,k)B(a,T-t)V(t,T) \qquad \text{from (1.4)}$$

$$\Rightarrow \Lambda_1^{HW} = \frac{\gamma(a,k)B(a,T-t)V(t,T)}{B(a,T_1-t)P(t,T_1)}$$

Similarly, for delta-gamma hedging with the one-factor HW model, the position in the T_1-year zero coupon bond (Λ_1^{HW}) and the position in the T_2-year zero coupon bond (Λ_2^{HW}), are

$$\Lambda_1^{HW} = \frac{\gamma(a,k)B(a,T-t)(\gamma(a,k)B(a,T-t) - B(a,T_2-t))V(t,T)}{B(a,T_1-t)(B(a,T_1-t) - B(a,T_2-t))P(t,T_1)}$$

$$\Lambda_2^{HW} = \frac{\gamma(a,k)B(a,T-t)(\gamma(a,k)B(a,T-t) - B(a,T_1-t))V(t,T)}{B(a,T_2-t)(B(a,T_2-t) - B(a,T_1-t))P(t,T_2)}$$

For delta-hedging with the two-factor HW model, the position in the T_1-year zero coupon bond (Λ_1^{G2}) and the position in the T_2-year zero coupon bond (Λ_2^{G2}) are

$$\Lambda_1^{G2} = \frac{(\gamma(a_1,k)B(a_1,T-t)B(a_2,T_2-t) - \gamma(a_2,k)B(a_2,T-t)B(a_1,T_2-t))\,V(t,T)}{(B(a_1,T_1-t)B(a_2,T_2-t) - B(a_2,T_1-t)B(a_1,T_2-t))\,P(t,T_1)}$$

$$\Lambda_2^{G2} = \frac{(\gamma(a_1,k)B(a_1,T-t)B(a_2,T_1-t) - \gamma(a_2,k)B(a_2,T-t)B(a_1,T_1-t))\,V(t,T)}{(B(a_1,T_2-t)B(a_2,T_1-t) - B(a_1,T_1-t)B(a_2,T_2-t))\,P(t,T_2)}$$

In each case, the amount invested in the bank account $S(t)$ equals the difference between the liability value and the total value of the position in zero coupon bonds.

For simplicity, in our numerical illustrations, we select the first zero coupon bond duration, T_1 to be the same maturity as the horizon of the liability, and select the longest maturity zero coupon bond available for T_2 (30-year for the U.S. Treasury Strip market). We have performed experiments with other choices of T_1 and T_2 and our main conclusion remains valid.

1.3.3 Estimating the Parameters

In both Hardy *et al.* (2014) and Zhu (2015), the same parameters are used. For the one-factor HW model, we used $\alpha = 0.02$, $\sigma = 0.006$, and for the two-factor model, we used

$$a_1 = 0.055, \qquad a_2 = 0.108, \qquad \sigma_1 = 0.032, \qquad \sigma_2 = 0.044, \qquad \rho = -0.9999$$

These parameters indicate a long term, unconditional standard deviation for the short rate of 3% per year, and for the 30-year rate of around 2.3% per year. Notice that the correlation ρ is close to negative one, which implies that the two-factor model is in some sense very close to an one-factor model. However, this one-factor model is not equivalent to the one-factor HW whenever $a_1 \neq a_2$ (see Brigo and Mercurio (2007)). Gurrieri *et al.* (2009), Gupta and Subrahmanyam (2005) and Brigo and Mercurio (2007) all note that ρ may become highly negative depending on the calibration instruments and the calibration dates. The parameter values shown above imply a long-term unconditional standard deviation for the short rate that is close to experience over the past 30-40 years.

In Section 1.5, we will assess the effectiveness of delta and delta-gamma hedging strategies by applying them to CB liabilities maturing in the period 2005-2015. It would not be appropriate to use the parameters above to examine the hypothetical effectiveness of the hedging strategies over this testing period, as the parameters were derived using that same data. To test whether the hedge strategy would have been effective, we should only use information available at the time that hedge portfolio is constructed. The parameters used should be consistent with the data available at the assumed initial valuation; we also consider the possibility that the parameters could be updated between the initial valuation and the termination date.

The most common choices of market derivatives used for calibration of interest rate models are caps and swaptions. We use swaptions as they contain more information on the correlations between forward rates, which is critical in our application (see Brigo and Mercurio (2007)). The choice of the number of swaptions used in the calibration strongly affects the parameter values (especially on a_1 and a_2). For example, Gurrieri *et al.* (2009) compare the calibration results using three sets of swaptions for the one-factor HW with time-varying mean reversion and volatility. We use at-the-money swaptions with expiration dates between one year and the maturity of the CB liability. The tenor of the swaptions are chosen to be 30 years, with quarterly payments[3]. We set the model parameters by minimizing the sum of the relative squared difference between the theoretical swaption price and the market price, that is, minimizing

$$obj = \sum_{i=1}^{n} \left(\frac{\text{model implied swaption price}_i - \text{market swaption price}_i}{\text{market swaption price}_i} \right)^2$$

For the two-factor model, to distinguish between the parameters of the two stochastic drivers, we set constraints as

$$0 < a_2 < a_1 < 1, \qquad 0 < \sigma_1, \sigma_2 < 0.5, \qquad -1 < \rho < 1$$

The upper bounds for the parameters a and σ ensure the optimization does not result in highly unrealistic values, the choice of $\sigma < 0.5$ is arbitrarily chosen with the prior belief that the actual volatility is much less than 0.5.

There exist closed-form solutions for swaption prices under the one and two factor HW models. However, the calibration is computationally expensive. Here we adopt the approximation proposed by Schrager and Pelsser (2006), which greatly simplifies the formulae for swaption prices with sufficient accuracy (for parameter values in our study, the relative difference in the swaption prices is at most 2%).

[3] The swaption data are quoted as a Black-volatility matrix and are obtained from Bloomberg.

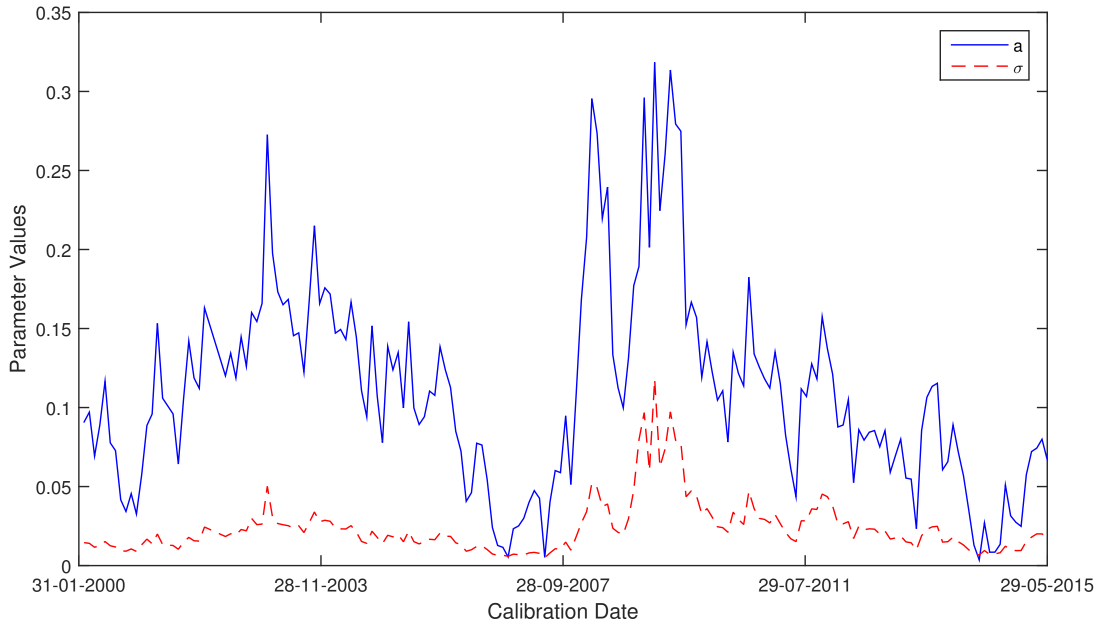


Figure 1.1: Calibrated parameters for the one-factor HW based on swaption prices, 2000-2015.

Figures 1.1 and 1.2 present the parameters that calibrated using the swaption data on each date. At first glance, the estimated parameters for the two-factor HW model do not match with the estimated parameters for the one-factor HW model. Indeed, the variance of future short rates under these two sets of parameters differs significantly. The reason is that the swaptions we used for calibration are focused on the long rates over the next five years. In fact, the two sets of parameters imply a close match for the variance of spot rates with long maturity in the next one to five years. Also, notice that all calibrated parameters are somewhat volatile, especially for the two-factor HW model (where parameters are often close to the constraints). This phenomenon has been observed elsewhere in the literature (for example, Enev (2011) and Gurrieri *et al.* (2009)). In Section 1.6 we will apply the hedging strategies derived from these models and parameters to real world interest rate paths; we will demonstrate that in spite of the volatility in the implied parameters, the hedging strategies can nevertheless be quite effective.

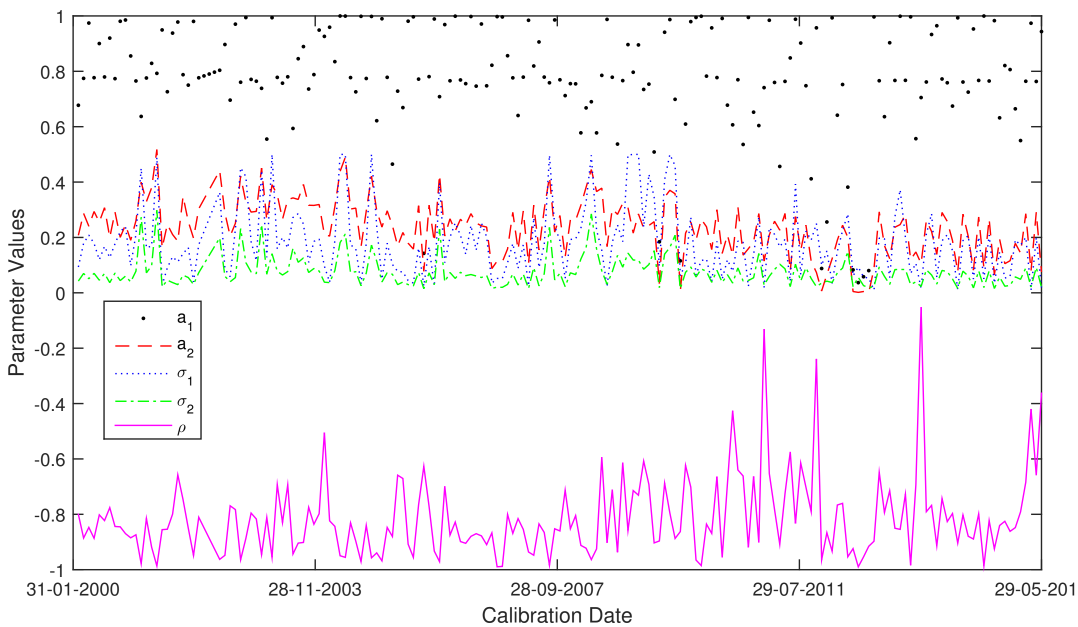


Figure 1.2: Calibrated Parameters for Two-Factor HW (volatility and correlation only) based on swaption prices, 2000-2015.

1.4 Hedging Performance Metrics

Under certain theoretical assumptions, a dynamic hedging strategy with no transaction costs can perfectly replicate the liability. However, this is impossible in practice, and the difference between the liability and the replicating instrument value is the hedge error. The hedge error comes from two main sources: the fact that hedges are rebalanced discretely, not continuously as the theory requires, and from the difference between the model and parameters assumed, and the actual real world interest rate process. There are also additional potential errors from transaction costs, but these are not very significant here because fees for trading government bonds are assumed to be very low. See Zhu (2015) for more details.

In Section 1.5, we quantify the discrete hedge error by considering weekly, monthly and annual re-balancing, where the monthly case serves as the benchmark, and where we eliminate the impact of model and parameter risk, by using the hedging model, adapted for the market price of risk, to generate the simulated "real world" interest paths.

In Section 1.6 we examine the model and parameter risk by testing the hedging strategies derived from our models, using historical paths of crediting rates and yield curves, assuming crediting rates are treasury spot rates.

The approach in Section 1.6 introduces basis risk, from the fact that we are assuming spot crediting rates, where the real liability is the yield to maturity (YTM). In Section 1.7 we investigate the basis risk by considering the difference between the spot rate hedge and the YTM liability.

1.4.1 Accumulated Maturity Hedging Error (MHE)

We let $\mathcal{H}(t)$ denote the value of the hedge portfolio at time t, given the participant remains in the plan. By definition, on rebalancing dates we have $\mathcal{H}(t) = V(t) = F_t V(t,T)$. Let $\mathcal{H}(t^-)$ denote the value of the hedge portfolio at time t immediately before re-balancing, and we let $\mathcal{E}(t)$ denote the hedge error at time t.

$$\mathcal{E}(t) = V(t) - \mathcal{H}(t^-)$$

Assuming that at each re-balancing date, we invest/borrow the amount equal to the hedge error at the risk-free rate, we get the accumulated hedge error at maturity (MHE)

$$\text{MHE} = \sum_{t=1}^{T} \mathcal{E}(t) e^{\int_t^T r(s)ds}$$

Here we want to mention that MHE is not the only reasonable metric to measure the hedging effective, but MHE has the interpretation advantage at maturity date since it is naturally developed from a valid hedging strategy (risk-free investment of $\mathcal{E}(t)$). It is also possible to examine the performance from the starting date, which means to discount the hedge error at each rebalancing date to the present value, see Hardy (2003) for details.

1.5 Evaluating discrete rebalancing hedging errors

The distribution of hedge errors arising from discrete rebalancing can be obtained by generating a large number of random sample paths for the future crediting rate under the real-world probability measure, then calculating the maturity hedge error for each sample path, based on a notional fund of \$1000 at the start of the simulation.

We use 10,000 paths under a P-measure model for future interest rates that is a simple shift of the assumed risk neutral model, to allow for the market price of risk. For details see Björk (2009). All other model assumptions are the same for the random paths and for the hedging and valuation model, so we are isolating the effect of discrete rebalancing without considering model, parameter or basis risk, which are all considered in subsequent sections.

Specific assumptions for both this section and Section 1.6 are:

- The crediting rate is the $k = 30$-year treasury spot rate
- We are hedging the payoff from a CB plan, based on the notional account $F_0 = \$1000$ at the initial valuation date, and assuming the participant exits $T = 5$ years later. We ignore additional notional contributions between the valuation date and the exit date.
- The hedge instruments at t for the one-factor HW model with delta hedging, are the money market and a pure discount bond with $T_1 = T - t$ years remaining.
- The hedge instruments at t for the one-factor HW model with delta-gamma hedging, and for the two factor HW model with delta hedging, are the money market, a pure discount bond with $T_1 = T - t$ years remaining, and another pure discount bond with $T_2 = 30 - t$.

For the market price of risk, we follow the work of PricewaterhouseCoopers (2014), which assumed a constant market price of risk, denoted λ, estimated through historical rates from 2001 to 2012. The dynamics of the short rate under the real world measure for the one-factor HW model are

$$dr(t) = (\theta(t) - ar(t) + \lambda\sigma)dt + \sigma dW^P(t)$$

We evaluate λ using historical monthly returns (from 1990 to 2015) on zero coupon bonds with maturities varying from 5 years to 30 years. The estimates we obtained range from 0.46 to 0.6. We assume that 0.5 would be a reasonable approximation for λ, and since we use 0.006 for the volatility term, the drift adjustment would be 0.003.

Figures 1.3 and 1.4 display the distribution of maturity hedge errors under simulation using the one-factor HW model, where $F_0 = \$1000$. The starting date of the plan (or the initial term structure) is chosen as at 2009-02-27 (we find that the maturity hedge error is not very sensitive to the initial term structure). Immediately, we observe that increasing the frequency of the re-balancing will reduce the overall hedge loss, as we expect. Also, for delta hedging, the effect of switching from annual re-balancing to monthly is much greater than from monthly to weekly (in both the magnitude and the shape of the distribution). Most importantly, even for annual re-balancing, the maturity hedge error for a 5-year plan is less than 0.2% for the worst case, which is insignificant compared with other sources of error (as we shall illustrate in Section 1.6). In terms of terminal funding level, a monthly hedge frequency replicates the maturity benefit with error less than 0.01%.

We conclude that the impact of discrete hedging error is relatively minor, and in the following sections where we consider model, parameter, and basis risk we may assume monthly hedging does not significantly impact the ultimate hedging errors.

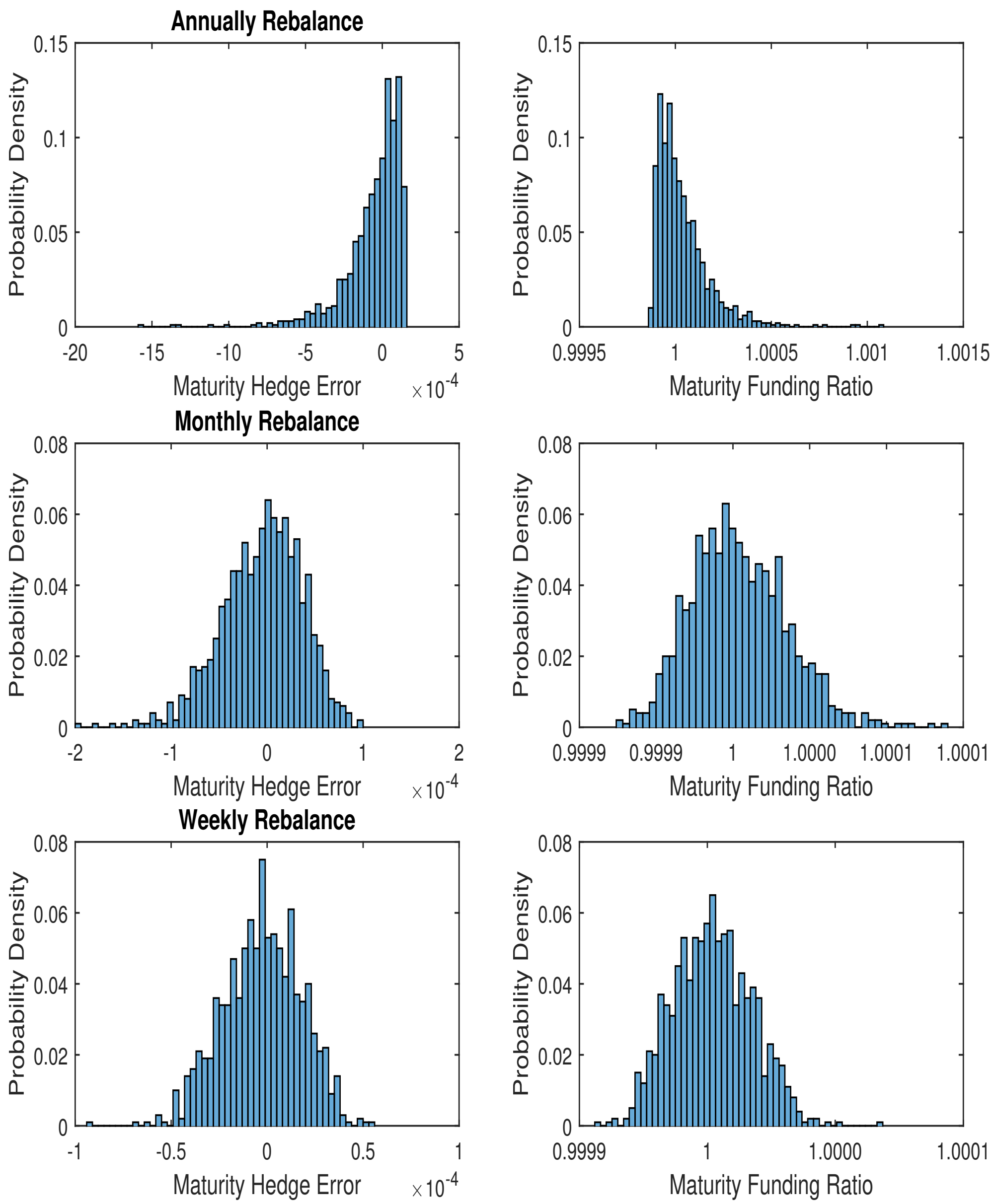


Figure 1.3: Simulation Results for Delta Hedging (One Factor HW), plan started at 2009-02-27, initial account value $F_0 = 1000$. Annual (top), monthly (middle) and weekly (bottom) rebalancing.

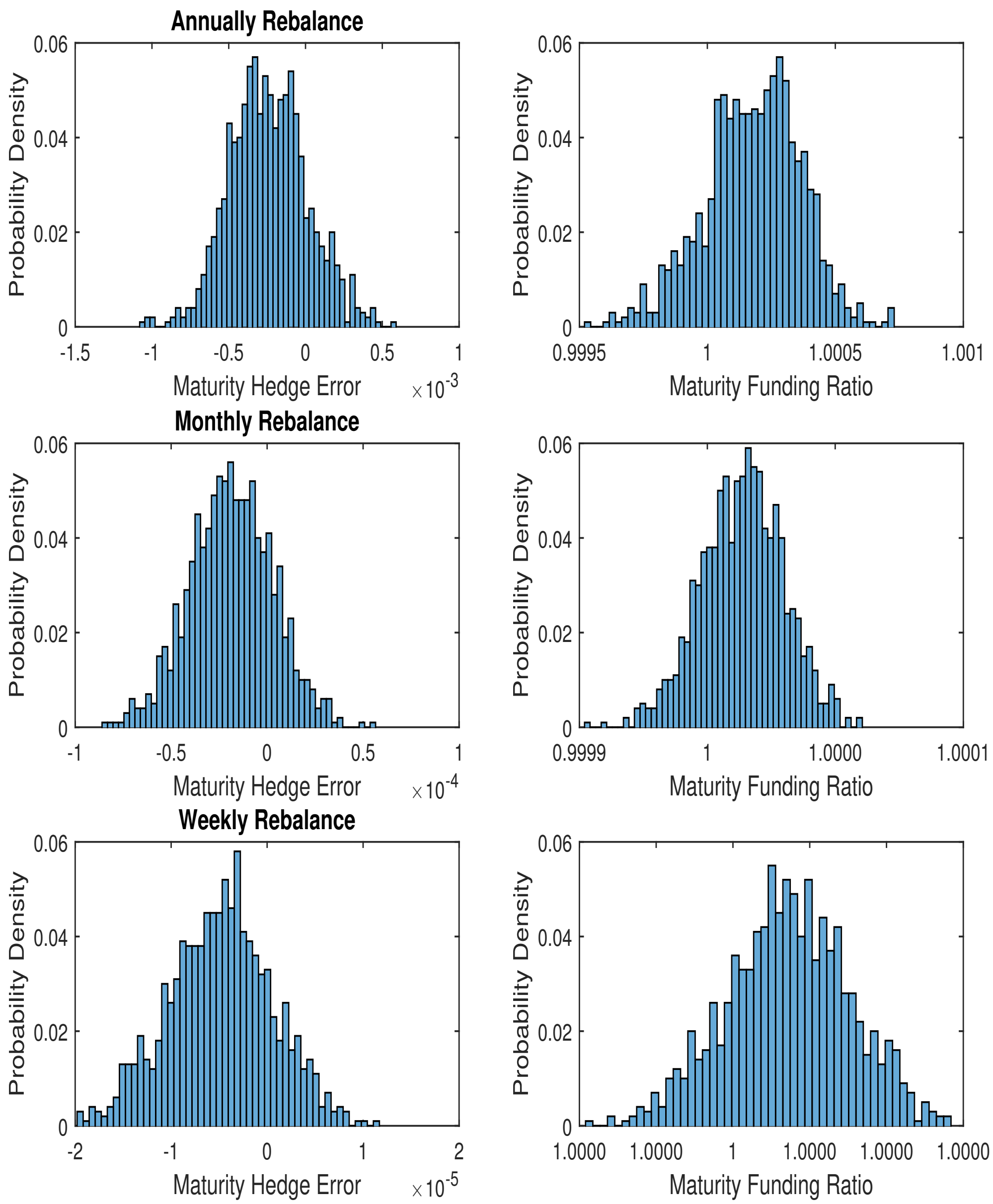


Figure 1.4: Simulation Results for Gamma Hedging (One Factor HW), plan started at 2009-02-27, initial account value $F_0 = 1000$. Annual (top), monthly (middle) and weekly (bottom) rebalancing.

1.6 Evaluating model and parameter error using the hypothetical historical hedging performance

In this section we consider the same CB plan as in Section 1.5, but instead of using randomly generated 5-year paths for short rates and crediting rates, we will use a succession of 5-year paths from the past 15 years, with end dates ranging from 2005 to 2015. That is, we assume the plan trustees hedge the accrued liability (i.e. past notional accumulated contributions) assuming all members leave at the end of 5-years from the valuation date[4]. We assume monthly rebalancing, based on the results of Section 1.5 which indicate that this will be sufficiently frequent.

To model the potential impact of different ways of updating the parameters, we show the results using three different approaches

(i) Constant parameters

(ii) Parameters calibrated at the plan starting date, and then maintained through the five year term

(iii) Parameters re-calibrated at each hedging date.

Recalibrating will, in principle, affect both the hedge and the valuation factors, but the impact is very small. In Figure 1.5 we present the valuation factors under different parameter calibrations and under the different models (one factor or two factors). The graph demonstrates that the valuation factors are extremely close, meaning that they are not very dependent on the parameters or on the number of factors.

In Figures 1.6 to 1.8 we show results for each of the different hedging portfolios, with different approaches to parameter recalibration. For easier comparison, the graphs are all shown on the same scale. Figure 1.6 shows the results for the delta strategy, using the one-factor HW model. Figure 1.7 shows the delta-gamma strategy hedging result, still with the one-factor HW model, and Figure 1.8 shows the delta strategy hedging result using the two-factor HW model.

From Figure 1.6 we see that the accumulated hedge errors are quite variable, and the impact of re-calibrating the model is quite significant. Using constant parameters, the MHE ranges from around +11% (in 2005, 2012) to −4% (in 2008-9) of the terminal liability (which ranges from around \$1180 to \$1320). Using parameters that are set at the

[4] This is somewhat analogous to a partly projected approach to traditional pension valuation.

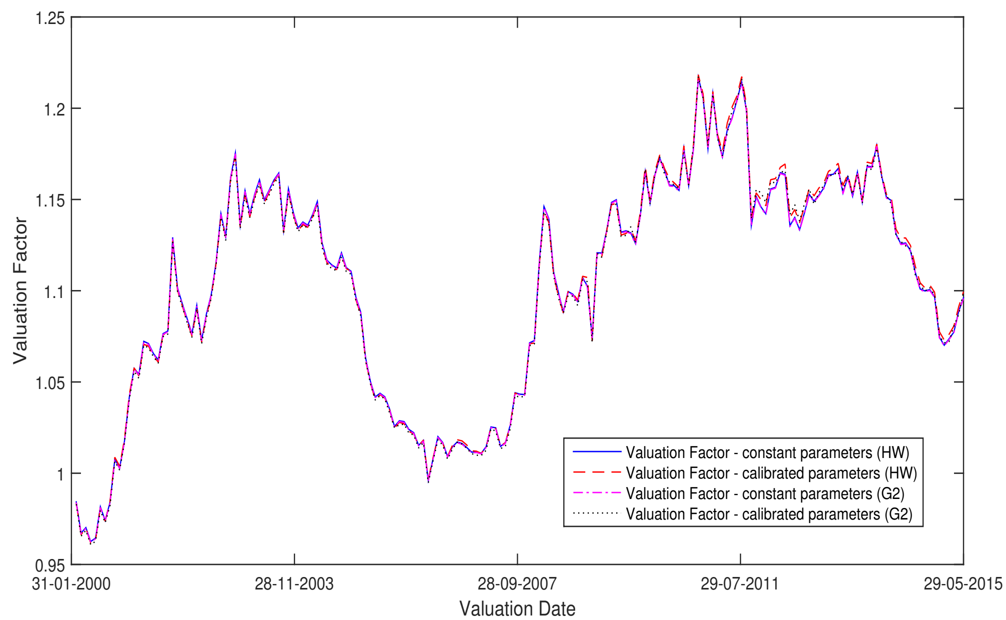


Figure 1.5: Valuation Factors evaluated using different parameter calibration approaches, for Hull-White models using one factor (labelled HW) and two factors (labelled G2).

start of each 5-year term, the errors range from +12% (in 2012) to −4% (in 2008-9). Using parameters that are recalibrated monthly, the higher end of the range is reduced to around +6% of the terminal liability.

In Figure 1.7 we see that adding the gamma hedge to the one-factor model has not significantly improved the hedge accuracy in the constant parameter case, but has improved the accuracy where the parameters are regularly recalibrated.

In contrast, Figure 1.8 shows that the hedge using the two-factor model is quite robust to the parameter variability, and overall losses are contained in a range from close to 0% up to around 5% of the terminal liability.

In Figure 1.9 we show the MHEs for each model, where the hedges were determined using constant parameters. The most immediate result is that the two factor model appears to create a much more reliable hedge than the one factor, even though the valuation results were shown in Hardy *et al.* (2014) to be very insensitive to the number of factors. However, the conclusion is less clear when we compare results allowing for recalibration of the model at each hedging date, as shown in Figure 1.10. Here we see that, although the variability

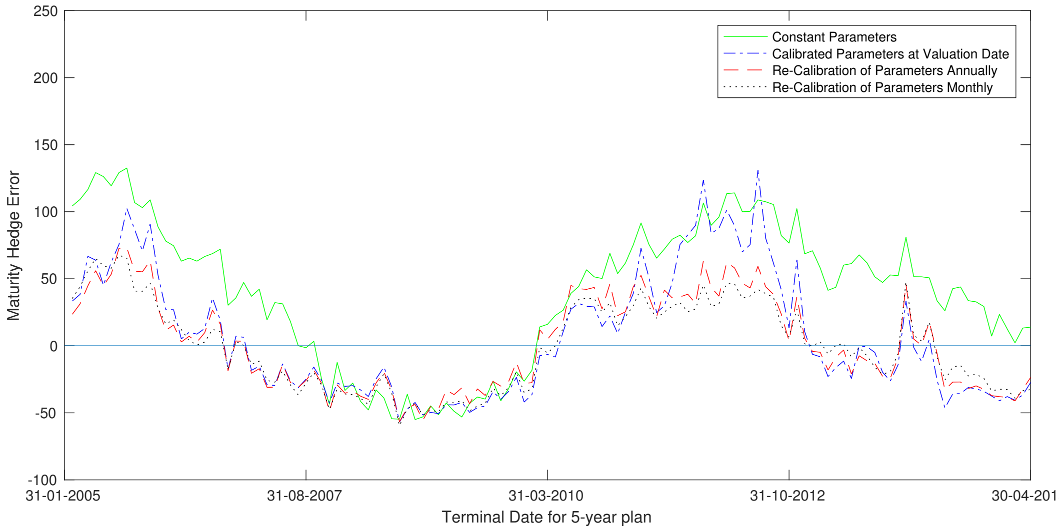


Figure 1.6: Maturity Hedge Loss for CB plan, 5-year horizon, maturing at 2005-2015, with initial account value $F_0 = 1000$, Delta-Hedging with the one-factor HW Model, using different parameter calibration approaches.

of hedging errors in the two-factor case is less than either of the one factor cases, the upper end of the errors are similar – though generated by different market conditions. The one-factor model does worst through the 2001-2006 period, with declining spreads between long and short rates. The two factor model is less sensitive to the changing market conditions up to the crisis; the worst performance covers the period from around 2008 to 2013. We note also from this graph that the addition of the gamma hedge in the one factor case does seem to generate improved hedging accuracy.

1.7 Crediting with the yield to maturity

As mentioned in previous sections, in practice a treasury based crediting rate in a CB plan would use the YTM on treasuries, not the spot rate. This means that there is basis risk in the analysis in Section 1.6, arising from the difference between the assumed hedge liability and the actual hedge liability. The reason for using spot rates is tractability; using YTMs, all the hedging portfolios must be determined using repeated Monte Carlo simulation. In this section we examine the magnitude of the basis risk by considering the

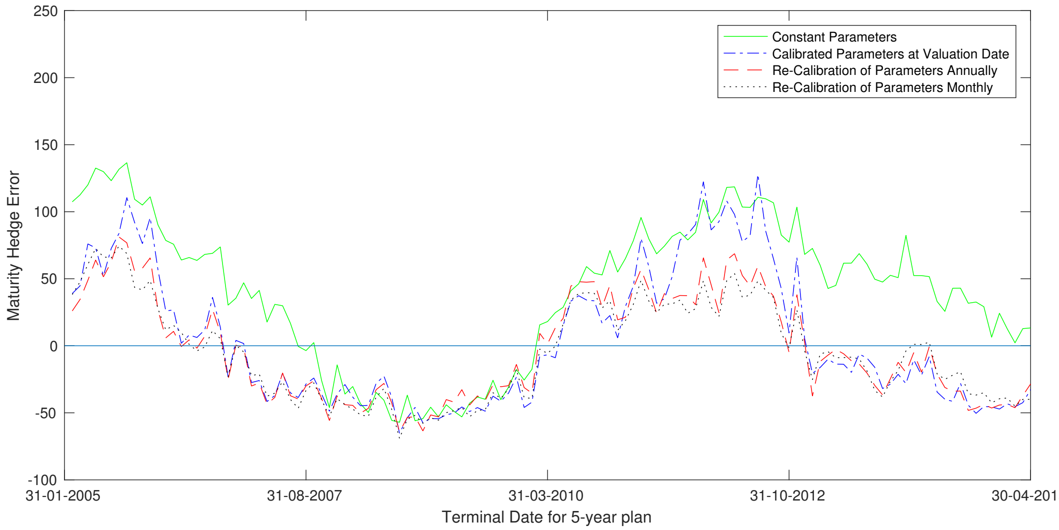


Figure 1.7: Maturity Hedge Loss for CB plan, 5-year horizon, maturing at 2005-2015, with initial account value $F_0 = 1000$, Gamma-Hedging with the one-factor HW Model, using different parameter calibration approaches.

difference between hedging spot crediting rates and hedging YTM crediting rates, with all other assumptions as in Section 1.6; specifically, 30-year Treasury Bond crediting rates, 5-year hedging horizon, with Hull-White one and two factor interest rate models.

1.7.1 Valuation by Control Variate Method

Unfortunately, there is no closed-form solution available, thus, numerical methods such as the Monte Carlo method must be implemented. Here we denote $y_k(t)$ as the k-year YTM at time t, which is the annual coupon rate payable semi-annually on a new k-year bond issued at par at time t:

$$y_k(t) = \frac{2(1 - P(t, t+k))}{\sum_{u=1}^{2k} P(t, t+\frac{u}{2})}$$

Due to the close relationship between the YTM and the spot rate, the valuation factor for a CB plan using the spot rate can be set as a control variate in the Monte Carlo simulation for the YTM crediting rate valuation. To measure the effectiveness of the control variate,

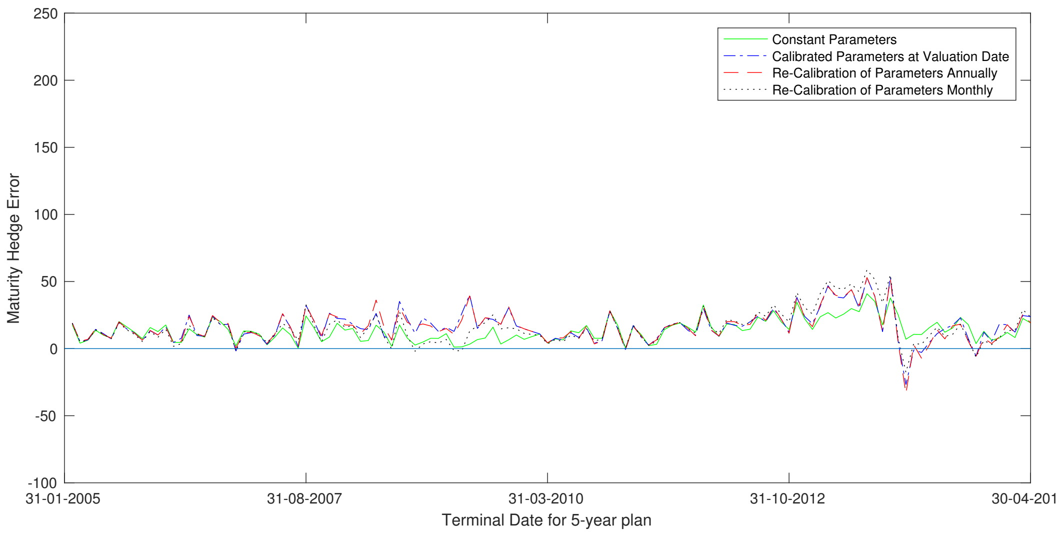


Figure 1.8: Maturity Hedge Loss for CB plan, 5-year horizon, maturing at 2005-2015, with initial account value $F_0 = 1000$, Delta-Hedging with the two-factor HW Model, using different parameter calibration approaches.

we use the variance reduction ratio (VRR), which is defined as

$$\text{VRR} = \frac{\text{Var(naive estimator)}}{\text{Var(estimator with control variate(s))}}$$

Figure 1.11 displays the VRRs for different valuation dates. The effectiveness of the control variate depends on the correlation between the spot rate and the YTM. When the yield curve is relatively flat, the YTM is close to the long-term spot rate, and the VRR is almost 15,000. Even when the yield curve is fairly steep the VRR is above 5000. This implies that the control variate reduces at least 99.98% of the variance of the naive Monte Carlo method.

Figure 1.12 compares the valuation factors using spot crediting rates and YTM crediting rates, assuming constant parameters. We note that spot crediting rates have generally produced larger valuation factors over this period; the difference is usually less than 4%, but in periods where the yield curve is particularly steep it has been as high as 9%.

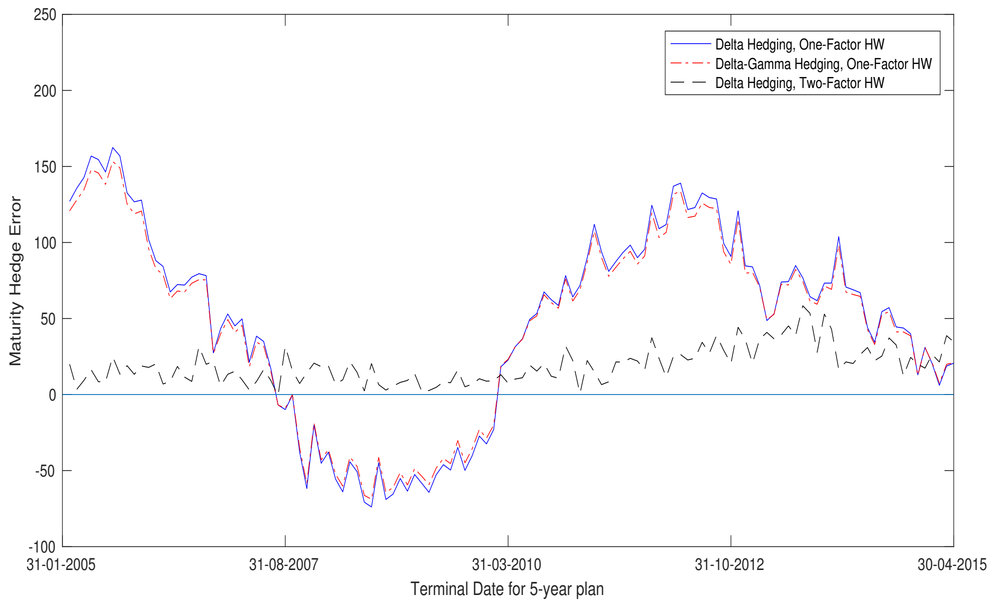


Figure 1.9: MHE for 5-year CB plan maturing at 2005-2015, using constant parameters.

1.7.2 Hedging the YTM crediting rate

To calculate the delta for the YTM valuation factor we adopt the pathwise method (see Glasserman (2013)), and use the delta from equation (1.4) as the control variate. Although the variance reduction is not as effective as the valuation factors, VRR remains above 15 (up to 200) for all scenarios. It is important to point out that although the valuation factors for the spot rate and YTM are close, their Deltas and Gammas may differ significantly. This is because the sensitivity of the short rate to the YTM is quite different from the sensitivity of the short rate to the long spot rate. Therefore, it is important to verify if delta hedging remains effective.

In this section, instead of using treasury STRIPs, we considered more liquid treasury bonds in our portfolio. All other settings follow exactly as in Section 1.6.

Comparing Figures 1.13 and 1.14 with Figures 1.9 and 1.10, we can observe many similarities. The shape of the graphs are almost identical, and, as we expected from the comparison between valuation factors, the absolute value of the MHE is slightly lower for YTM crediting rates. Importantly, we see that, as for the spot rates, the delta hedge applied under the one or two factor HW models appears to offer a stable risk management

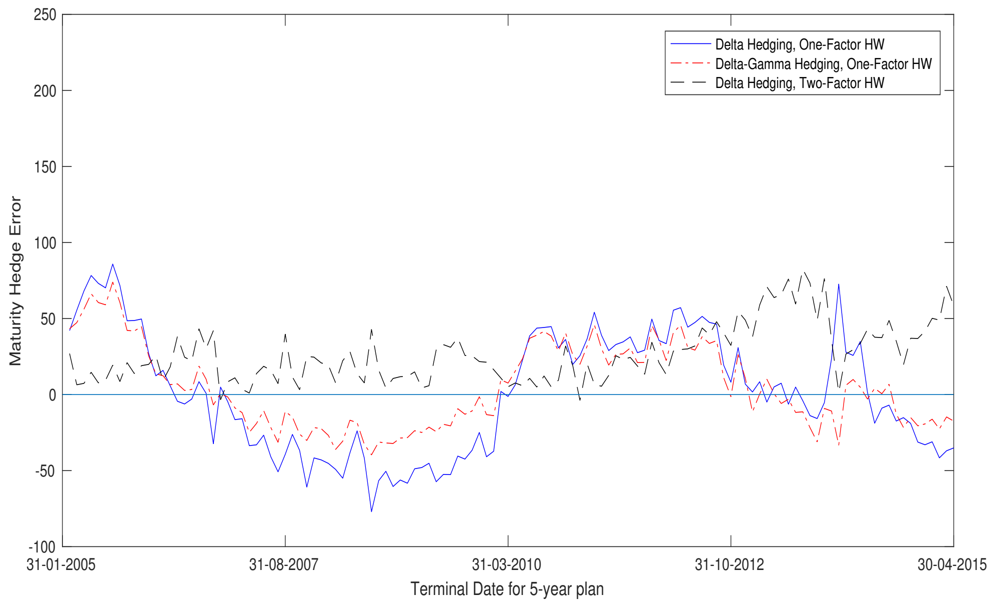


Figure 1.10: MHEs for 5-year CB plan maturing at 2005-2015, parameters re-calibrated monthly.

strategy, with accumulated hedging errors generally less than around 5% of the fund value at maturity.

1.8 Comparing the Delta hedge strategy with a traditional investment strategy

One of the initial attractions of CB plans to employers is that by investing in equities, the employer could pay less than the notional contribution rate (as % of pay), since the additional return on equities would fund the crediting rate and make up the shortfall in contributions (see Gold (2001)).

Using a traditional approach to pension investing for DB plans, trustees might select a simple 60% equity/40% long bond portfolio, and it is interesting to compare the funding deficit or surplus at maturity under this strategy with the case where one of the hedging strategies is adopted. In Figure 1.15, we show the maturity hedging loss (or surplus)

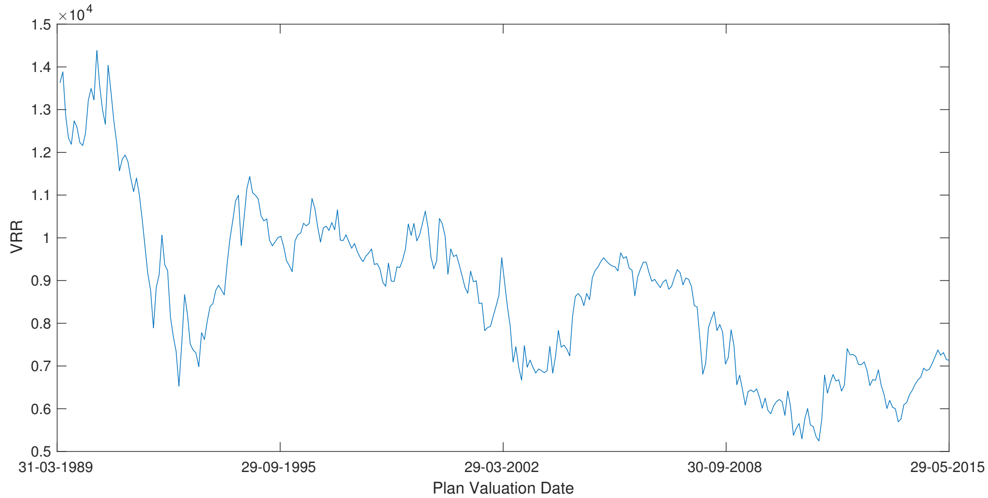


Figure 1.11: VRR for YTM crediting rate valuation under One-Factor HW, using spot rate as the control variable, 10,000 simulations

following a traditional 60/40 equity/long bond strategy, alongside the accumulated maturity hedging error using the two-factor HW hedge. Note that we have had to extend the scale dramatically compared with the hedging graphs above. We have assumed the same starting assets for each 5-year period (specifically, $1000V(t,5)$), although in practice this is significantly more than a traditionally managed CB plan might hold.

This figure clearly demonstrates how the hedged portfolio targets the terminal liability, while the unhedged portfolio may end up vastly over or under funded. On the other hand, the unhedged case does generate some very enticing profits – but this is not a good case for rejecting hedging, as the profits are unlikely to be repeated in the near future. The gains from the equity/bond portfolio are largely generated from the steady period of declining interest rates, which has generated consistent gains on long bonds that cannot continue indefinitely, and indeed will reverse in periods of rising interest rates. There is a significant possibility that the massive gains experienced in the last 10 years could be replaced by equally large losses in the next 10 years using the equity/bond investment approach. If the purpose of modern pension risk management is not to generate windfall gains, but to minimize losses and hence minimize contribution variability as much as possible, then the hedging strategy appears far more suitable.

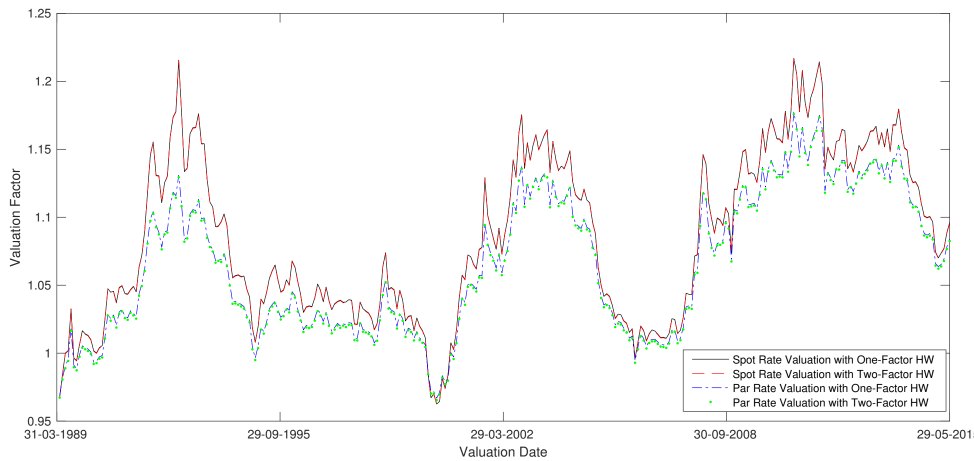


Figure 1.12: Compare Valuation Factors using 30-year spot and YTM crediting rates, 5-year horizon, initial fund \$1000, 10,000 simulations.

1.9 Incorporating pre-retirement exits

In practice, employees may terminate their pension plan prior to the normal retirement date for various reasons: change of employment, disability, death or early retirement. These termination events are often assumed to be independent from the market performance. Here we construct a simple example to illustrate that delta hedging remains effective if the termination events are assumed to be independent from interest rates, and diversifiable. Let $U(t,T)$ denote the valuation factor incorporating pre-retirement exits, and assuming (for simplicity) that the individual will receive their account value at the end of the year of exit.

$$U(t,T) = \sum_{u=\lfloor t+2 \rfloor}^{T-1} V(t,u)\,{}_{\lfloor t+u-1 \rfloor - t}p_x\, q_{x+\lfloor t+u-1 \rfloor - t} \\ + {}_{\lfloor t+1 \rfloor - t}q_x V(t, \lfloor t+1 \rfloor) + {}_{\lfloor t+T-1 \rfloor - t}p_x V(t,T)$$

where ${}_tp_x$ is the probability that a plan member age x is still in the plan at age $x+t$, and q_x is the probability that a plan member age x exits the plan before age $x+1$ (Dickson *et al.* (2013)). We see that $U(t,T)$ is simply a weighted average of $V(t,S)$ where $t \le S \le T$. In Figure 1.16 we show the surface of $V(0,T)$ for a range of horizon periods (T) and start dates. Although $V(0,T)$ is not a strictly increasing function in T it is generally increasing

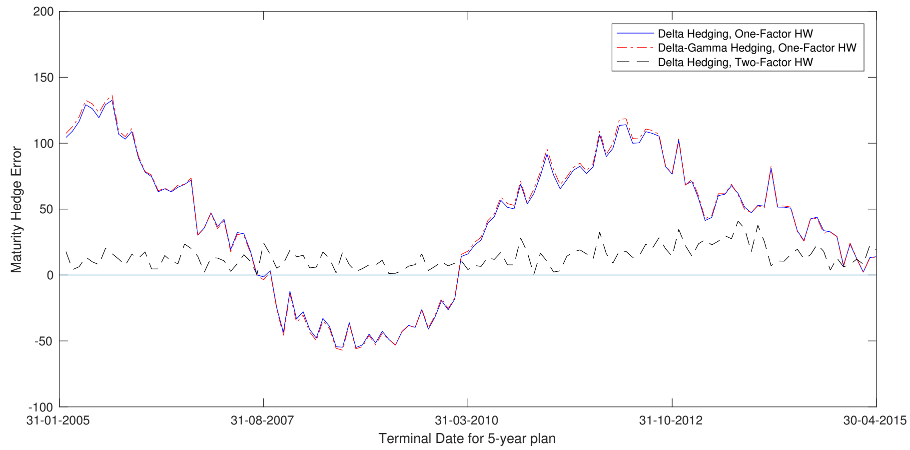


Figure 1.13: MHE for a CB plan with 5-year horizon, maturing at 2005-2015, initial amount \$1000, 30-year Treasury Bond YTM crediting rate, constant parameters.

for $T \in [0, 20]$. Using $V(t, T)$ as an approximation for $U(t, T)$ will tend to over-price and over-hedge the risk.

When the exit events are diversifiable the plan can be hedged by combining the exit probabilities with the appropriate hedge portfolios from Section 1.7.2. For a single cohort, this could be constructed as follows.

- **Step 1**: At time 0, evaluate the valuation factor $V(0, T)$ and construct the hedging portfolio based on the Greeks.
- **Step 2**: At the end of month 1, $100 \times {}_{1/12}p_x\%$ of people remain in the plan (requires $V(1/12, T)$ in liability), and ${}_{1/12}q_x$ percent of people leave the plan and will receive the full account balance at the year end. We measure the hedge loss for the period and re-construct the portfolio.
- **Repeat Step 2** for $t = [1/12, \cdots, T]$.

We illustrate this approach using demographic assumptions from the 2016 Actuarial Report for University of Toronto Pension Plan (see Hewitt (2016) for the report and Pension Experience Subcommittee (2014) for the mortality table), with a maximum horizon of 15 years. We set ${}_{\frac{1}{12}}q_x$ as the aggregate 1-month exit probability for a member age x. The probability that an employee will remain in the plan at the end of 15th year is approximately

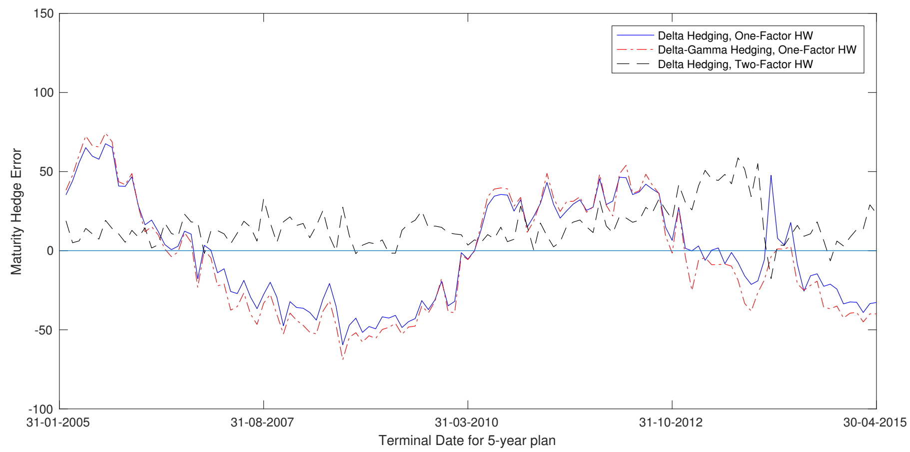


Figure 1.14: MHE for a CB plan with 5-year horizon, maturing at 2005-2015, initial amount \$1000, 30-year Treasury Bond YTM crediting rate, parameters recalibrated monthly.

47%. For a cohort starting on January 1st, 2000 and ending on December 31st 2014, the maturity hedge error is 10.7% of the initial account value. To show how this hedge error arises, in Figure 1.17 we show the monthly hedge errors that accumulate to the maturity hedge error of 10.7% of F_0 (which is set at \$1000), with and without early exits. Notice that the plan allowing for early exits has a slightly more stable hedging performance, but the hedge errors at each re-balancing date are close. Therefore, introducing early exits that are diversifiable will not affect the overall hedging results materially, and diversifiable exits can easily be incorporated into the hedging strategy.

1.10 Conclusion

In this chapter, we have studied dynamic hedging strategies for Cash Balance Pension plans. We have demonstrated that dynamic hedging strategies are relatively simple to construct, and offer effective mitigation of the significant interest rate risk inherent in CB plans with long Treasury bond YTM crediting rates.

We have restricted the examples, for example by excluding consideration of floors, by limiting the horizon to only five years, and by limiting the hedging portfolio to simple Treasury Bonds/STRIPs. Nevertheless, the results when applied to real world scenarios do indicate that the simple delta strategy is able to provide a practical, affordable hedge against interest rate risk, even through the turmoil of the past decade.

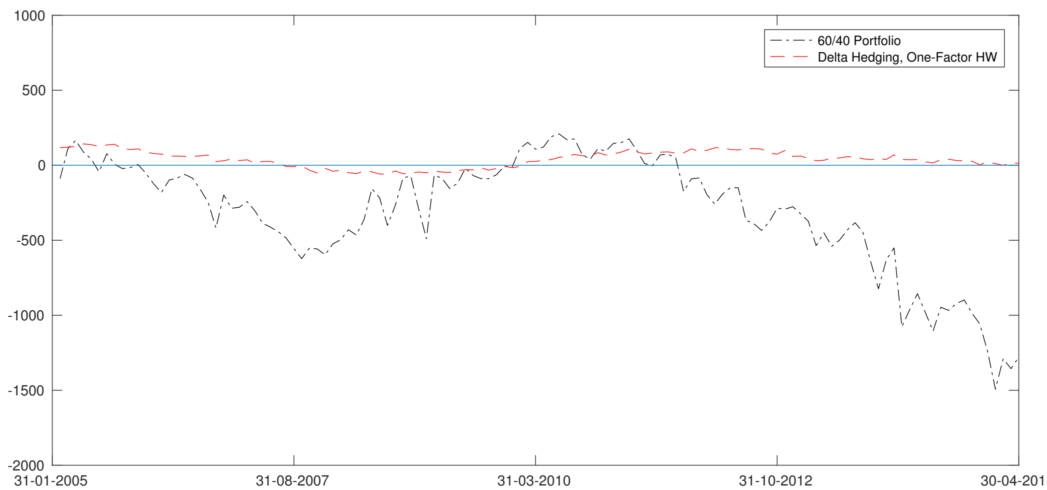


Figure 1.15: Maturity Hedge Loss for 5-year CB plan maturing at 2005-2015, initial account value $F_0 = 1000$, 30 year Treasury spot crediting rate, traditional equity/bond investment strategy and delta hedge strategy. Constant parameters.

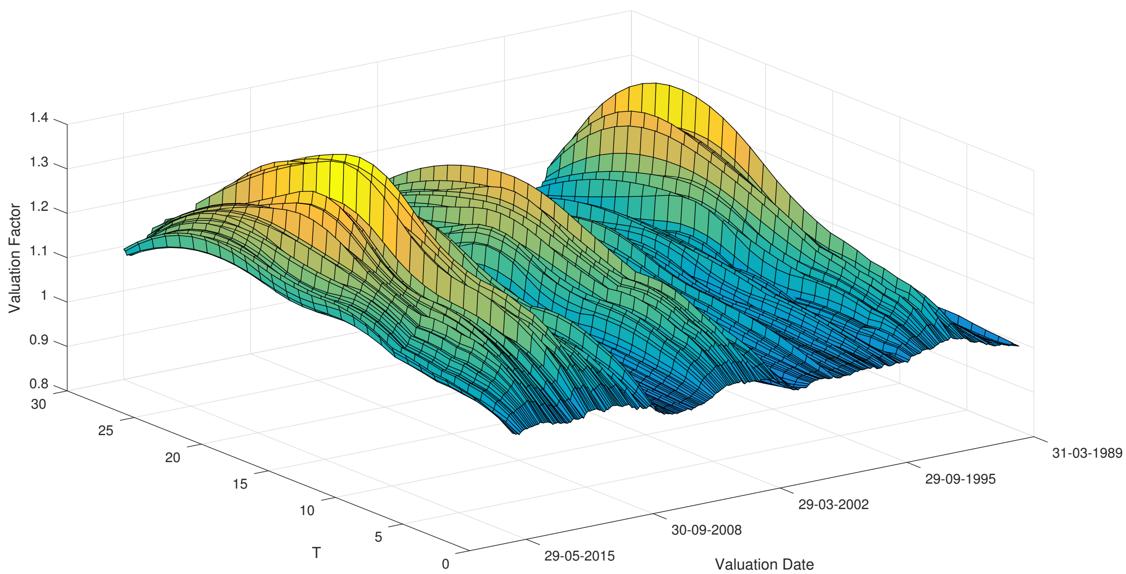


Figure 1.16: $V(0,T)$ using 30 year Treasury Bond YTM crediting rate, with different horizons and start dates, One-Factor HW model.

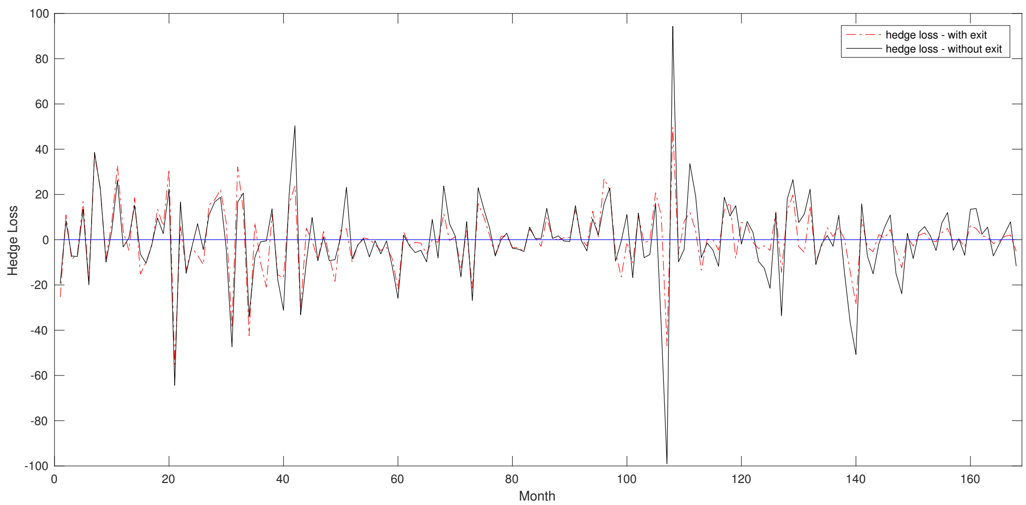


Figure 1.17: Monthly Hedging Errors for a 15-year CB cohort starting on 1/1/2000, with and without allowance for exits. One-Factor HW model, delta hedge, 30 year Treasury Bond YTM crediting rates.

Chapter 2

Early Exercise Defined Benefit Underpin Plan

2.1 Introduction

The rising interest in hybrid pension plans suggests a potential for designing new risk-sharing schemes. In this chapter, a new hybrid pension design is introduced. The intuition behind the new plan is based on two existing embedded options: the Florida second-election option, and the DB underpin.

The Florida second-election option is provided to public employees in Florida. The state sponsors two different pension plans, a DC and a DB. Employees choose either plan at enrollment. Subsequently, they have a one-time opportunity to transfer to the other plan (from DB to DC, or vice versa), prior to their retirement. For an employee switching from the DC to the DB plan, a "buy-in" cost would be calculated, equal to the accrued benefit obligation of the DB plan, based on the employee's current salary and service. The buy-in cost is paid from the DC account balance; if this is insufficient, the employee must cover any difference using his or her own financial resources.

The DB underpin plan is a DC plan that grants a DB type minimum guarantee. At retirement, the retiree's pension is the higher of the guaranteed DB benefit and the income from the annuitized DC funds.

The new hybrid design introduced in this chapter provides natural links to both of these plans. It is an early exercise DB underpin, under which a DC plan member may elect to lock into the DB plan benefit at some point prior to retirement. This is equivalent to the

Florida second election plan but with the transfer of funds at the switch from DC to DB being limited to the DC account balance.

The new hybrid plan's liability is evaluated using market consistent valuation methods and we compare it with the Florida second election and the DB underpin options. This chapter illustrates how the flexible nature of hybrid pension plans allows sponsors to design new risk-sharing schemes based on their risk appetite and objectives. Moreover, we will demonstrate that the creation of the new plan helps in conducting comprehensive and direct comparisons between different existing plans.

The remainder of the chapter is structured as follows. Section 2.2 introduces the notation and assumptions. Sections 2.3 and 2.4 provide detailed background information for the second-election option and the DB underpin option. Section 2.5 develops the new pension plan, as well as presenting some theoretical results in discrete time. Section 2.6 extends the work into the continuous time case and incorporates stochastic salaries. Section 2.7 displays the numerical results. Section 2.8 illustrates an alternative new pension design. Section 2.9 concludes.

2.2 Notation and Assumptions

In this section we introduce the actuarial and economical notation and assumptions adopted throughout this chapter. As is usual in the financial literature, we assume a complete and arbitrage-free market, with all rational agents sharing the same information generated by a filtration. This chapter focuses on the investment risks prior to the retirement date, thus we ignore mortality and other demographic considerations.

- c denotes the annual contribution rate (as a constant proportion of salary) into the DC plan. We assume that contributions are paid annually. We also assume that all contributions are paid by the plan sponsor/employer, although this is easily relaxed to allow for employee contributions.

- b denotes the accrual rate in the DB plan, which is assumed to be a constant.

- T denotes the time of retirement of the employee, which is assumed to be known and non-random.

- r denotes the risk-free rate of interest, compounded continuously, which is assumed to be a constant.

$\ddot{a}(T)$ denotes the actuarial value at retirement of a pension of 1 per year. Since we ignore longevity changes, and we assume a constant interest rate, we implicitly assume $\ddot{a}(T)$ is a constant.

S_t denotes the stochastic price index process of the funds in the DC account. We assume S_t follows a Geometric Brownian Motion, so that

$$\frac{dS_t}{S_t} = rdt + \sigma_S dZ_S^Q(t), \quad S_0 > 0$$

where $Z_S^Q(t)$ is a standard Brownian Motion under the risk neutral measure Q.

L_t denotes the salary from t to $t+1$ for $t = 0, 1, ..., T-1$, where t denotes years of service. We first assume that salaries increase deterministically at a rate μ_L per year, continuously compounded, so that

$$L_t = L_0 e^{t\mu_L}$$

We will also consider a stochastic salary process in the continuous setting, which we assume is hedgeable through the financial market (similar to Bacinello (2000)).

$$\frac{dL_t}{L_t} = rdt + \sigma_L dZ_L^Q(t), \quad L_0 > 0$$
$$dZ_L^Q(t) dZ_S^Q(t) = \rho dt$$

where $Z_L^Q(t)$ is a standard Brownian Motion under the risk neutral measure Q and ρ is the correlation coefficient between Z_L^Q and Z_S^Q (correlation between the stock market and the salary increase).

Let $\mathcal{F} = \{\mathcal{F}_t, t \geq 0\}$ be the filtration generated by $\left\{Z_S^Q(t), Z_L^Q(t), t \geq 0\right\}$ when the salary is stochastic, or $\left\{Z_S^Q(t), t \geq 0\right\}$ when the salary is deterministic. For conditional expectations with respect to the filtration, we simplify the notation such that $E_t^Q[\cdot] = E^Q[\cdot|\mathcal{F}_t]$.

2.3 DB Underpin Plan

The DB Underpin pension plan, also known as the floor-offset plan in the USA, provides a guaranteed defined benefit minimum which underpins a DC plan. Plan sponsors make regular specified contributions into the member's DC account, and separately contribute to an additional fund which covers the extra cost of the guarantee. Employees usually have limited investment options to make the guarantee value more predictable. At the retirement date, after T years of service, if the member's DC balance is higher than the value of the guaranteed minimum defined benefit pension, the plan sponsor will not incur any extra cost. However, if the DC benefit is smaller, the plan sponsor will cover the difference.

Using arbitrage-free pricing, we can calculate the present value of the cumulative cost of the DB underpin plan for the pension sponsor at time $t = 0$ as follows. We assume (more for clarity than necessity) that DB benefits are based on the final 1-year's salary, and we ignore all exits before retirement.

$$C^U = \underbrace{E^Q\left[\sum_{t=0}^{T-1} e^{-rt} c L_t\right]}_{\text{PV of DC Contributions}} + \underbrace{E^Q\left[e^{-rT}\left(bT\ddot{a}(T)L_{T-1} - \sum_{t=0}^{T-1}\frac{S_T}{S_t}cL_t\right)_+\right]}_{\text{Value of DB underpin option}}$$

The option value does not have an explicit solution, but can be determined using Monte Carlo simulation. See Chen and Hardy (2009) for details on the valuation and funding of the DB underpin option.

Notice that we have formulated the problem from the sponsor's perspective. In fact, under a risk neutral valuation, the cost function is equal to the expected discounted benefit received by the employee. For other pension plans studied in this chapter, we will stay with this approach, to be consistent with the pension literature (including, for example, Hardy *et al.* (2014), Bacinello (2000), Chen and Hardy (2009), Guillén *et al.* (2006), etc.).

2.4 Florida Second Election (FSE) Option

Public employees of the State of Florida are given the choice to participate in either the DC or the DB plan at the beginning of their employment. In addition, employees are granted a one-time option to switch to the other plan anytime before their retirement date. This option is called the second election by the Florida Retirement System. In this chapter, we focus on the option to exchange the DC plan for the DB plan, based on the life cycle hypothesis that individuals prefer portability at younger ages (provided by the DC plan) and stable retirement income at older ages (provided by the DB plan). To exercise the second election option, there is a "buy in" cost equal to the accrued benefit obligation (ABO), which is calculated as the present value of the accrued benefit, based on the current service and the current pensionable salary (ABO = accrual rate × current salary × number of years in service × annuity factor × discount factor).

We denote the ABO of the DB benefit for an employee with t years of service by K_t, so that

$$K_t = b\, L_{t-1}\, t\, \ddot{a}(T)\, e^{-r(T-t)}$$

Let τ denote a stopping time with respect to $\mathcal{F}$, which represents the time that an employee chooses to switch from the DC plan to the DB plan. If the ABO at τ is more than the DC account balance, the employee needs to fund the difference from her own financial resources. If the DC account is more than the ABO at transition, then the employee retains the difference in a separate DC top-up account which can be withdrawn at retirement.

Mathematically, assuming that the employee exercises the option at the time which provides most financial benefit, the present value of the total DC and DB benefit cost at inception forms an optimal stopping problem:

$$C^{se} = \sup_{\tau \in [0,1,\cdots,T]} E^Q \left[\sum_{t=0}^{\tau-1} e^{-rt} c L_t + e^{-r\tau} \left(K_T e^{-r(T-\tau)} - K_\tau \right) \right]$$

where the first term, as in the previous section, is the present value of the DC contributions, and the second term is the additional funding required for the DB benefit, offset by the ABO at transition, which is funded from the DC contributions. The 'sup' indicates that the valuation assumes the switch from DC to DB is made at the time to maximize the cost to the employer, which is indeed equivalent to maximizing the expected value of the employee payoff (by the optional sampling theorem):

$$C^{se} = \sup_{\tau \in [0,1,\cdots,T]} E^Q \left[e^{-rT} \left(\left(\sum_{t=0}^{\tau-1} c L_t \frac{S_\tau}{S_t} - K_\tau \right) \frac{S_T}{S_\tau} + K_T \right) \right]$$

Milevsky and Promislow (2004) studied the price and optimal switching time of the Florida option with deterministic assumptions.

2.5 Early Exercise DB underpin plan - Discrete Case

The FSE design has the advantage that it provides employees with the flexibility to choose their plan type based on their changing risk appetite. However, employees retain the investment risk through the DC period of membership, and also have the additional risk of a suboptimal choice of switching time. Moreover, when the DC investment falls below the ABO of the DB plan, the employee may be unable to switch, if they do not have sufficient assets to make up the difference.

Inspired by the idea of combining the DB underpin plan with the Florida option, we investigate a new hybrid design, which we call the Early Exercise DB underpin, that adds a guarantee at the time of the switch from DC to DB. If the employee's DC account is below the ABO when s/he elects to switch, then the plan sponsor will cover the difference.

2.5.1 Problem Formulation

We assume that the contributions are made annually into the DC account until the employee switches to DB, and that we are valuing the benefits at t. We assume also that the employee has not switched from DC to DB before time t, that the DC account is w_t at t, and that future account values (up to the switching time) follow the process

$$W_{t+\tau}^{t,w} = w_t \frac{S_{t+\tau}}{S_t} + \sum_{u=t}^{t+\tau-1} \frac{S_{t+\tau}}{S_u} cL_u \qquad \tau = 1, 2, ..., T-t.$$

We assume that the salary is deterministic, though we will relax this assumption when we consider the continuous case in the next section.

We assume that the exercise dates are at the beginning of each year (or at the end of year T), before the contribution is made into the DC account, so the admissible exercise dates are $\tau = 0, 1, \cdots T - t$.

The present value at time t of the cost of future benefits (past and future service, DC and DB), which is denoted by $C_1(t, w_t)$, can be expressed as the sum of three terms:

1. The present value of the future contributions into the DC account before the member switches to the DB plan.

2. The present value of the cost of the DB benefit, offset by the ABO at the time of the switch.

3. The difference between the ABO and the DC balance at the switching time, if positive.

We take the maximum value of the sum of these three parts, maximizing over all the possible switching dates, as follows.

$$C_1(t,w) = \sup_{\tau\in\{0,1,\cdots,T-t\}} E_t^Q\left[\sum_{u=0}^{\tau-1} e^{-ru}cL_{t+u} + e^{-r\tau}\left(K_Te^{-r(T-t-\tau)} - K_{t+\tau}\right) + e^{-r\tau}\left(K_{t+\tau} - W_{t+\tau}^{t,w}\right)^+\right]$$

Notice that the switching time τ is involved in all three parts, which makes the analysis more complex. We may rewrite the problem using optional sampling theorem, such that for any stopping time τ_1:

$$E_t^Q\left[e^{-r\tau_1}W_{t+\tau}\right] = E_t^Q\left[\sum_{u=0}^{\tau_1-1} e^{-ru}cL_{u+t}\right] + W_t$$

which means the sponsor's cost to fund a DC plan up to a future time τ_1, is equivalent to the risk neutral expectation of the present value of the DC account balance at τ_1. Therefore, our cost function becomes:

$$\begin{aligned}
C_1(t,w) &= \sup_{\tau\in[0,1,\cdots,T-t]} E_t^Q\left[\sum_{u=0}^{\tau-1} e^{-ru}cL_{t+u} + e^{-r\tau}\left(K_Te^{-r(T-\tau-t)} - K_{t+\tau}\right) + e^{-\tau r}\left(K_{t+\tau} - W_{\tau+t}\right)^+\right] \\
&= \sup_{\tau\in[0,1,\cdots,T-t]} E_t^Q\left[\sum_{u=0}^{\tau-1} e^{-ru}cL_{t+u} + K_Te^{-r(T-t)} - e^{-r\tau}W_{\tau+t} + e^{-r\tau}\left(W_{t+\tau} - K_{t+\tau}\right)^+\right] \\
&= \sup_{\tau\in[0,1,\cdots,T-t]} \left\{E_t^Q\left[\sum_{u=0}^{\tau-1} e^{-ru}cL_{t+u}\right] - E_t^Q\left[e^{-r\tau}W_{\tau+t}\right] + E_t^Q\left[K_Te^{-r(T-t)}\right]\right. \\
&\qquad \left. + E_t^Q\left[e^{-r\tau}\left(W_{t+\tau} - K_{t+\tau}\right)^+\right]\right\} \\
&= \underbrace{E_t^Q\left[K_Te^{-r(T-t)}\right]}_{\text{The ABO of DB plan at t}} + \underbrace{\sup_{\tau\in[0,1,\cdots,T-t]} E_t^Q\left[e^{-r\tau}\left(W_{\tau+t} - K_{t+\tau}\right)^+\right]}_{\text{Price of the Option}} - w \qquad (2.1)
\end{aligned}$$

This new formulation can be interpreted as the expected value of the future discounted benefit, which also consists of three terms:

- The present value of the DB plan benefits at time t.

- A Bermudan type call option, with underlying process W_t and strike value K_t.
- The negative of the funds available, w from the existing DC balance at t.

The first and third terms do not affect the switching time. To study the optimal exercise strategy, we may omit the first and third terms and define our value function as

$$\begin{aligned} v_1(t,w) &= \sup_{\tau\in[0,1,\cdots,T-t]} E_t^Q\left[e^{-r\tau}\left(W_{t+\tau}^{t,w} - K_{t+\tau}\right)^+\right] \quad (2.2) \\ &= \sup_{\tau\in[0,1,\cdots,T-t]} E_t^Q\left[e^{-r\tau}\left(\sum_{u=0}^{\tau-1}\frac{S_{t+\tau}}{S_{t+u}}cL_{t+u} + w\frac{S_{t+\tau}}{S_t} - b(t+\tau)\ddot{a}(T)L_{\tau+t-1}e^{-r(T-\tau-t)}\right)^+\right] \end{aligned}$$

At time t, given that the DC account balance is w, we define the *intrinsic value* of the option, denoted $v_1^e(t,w)$, as the value if the member decides to switch at that date, and the *continuation value*, denoted $v_1^h(t,w)$ as the value if the member decides not to switch. Then

$$v_1^e(t,w) = \left(w - tbL_{t-1}\ddot{a}(T)e^{-r(T-t)}\right)^+ \quad (2.3)$$

$$v_1^h(t,w) = E^Q\left[e^{-r}v_1\left(t+1,(w+cL_t)\frac{S_{t+1}}{S_t}\right)\middle|\mathcal{F}_t\right] \quad (2.4)$$

The option is not analytically tractable. In the next section, we present some of the properties of the value function, which will enable us to use numerical solution methods.

2.5.2 The Exercise Frontier

The following proposition gives the general convexity and monotonicity of the value function

Proposition 1. *At each observation date $0 \le t \le T$, the value function $v_1(t,w)$ is a strictly positive, non-decreasing, continuous and convex function of w.*

Proof. It is trivial to observe that $v_1(t,w)$ is strictly positive since $(W_{t+\tau}^{t,w} - K_{t+\tau})^+ \ge 0$ with the inequality sign being strict with a non-zero probability. We only need to show that the value function is non-decreasing, continuous and convex in w.

- **Non-decreasing in** w For $\epsilon > 0$, we have $(w-k)^+ - (w+\epsilon-k)^+ \leq 0$, therefore,

$$\begin{aligned}
&v_1(t,w) - v_1(t,w+\epsilon) \\
&\qquad \leq \sup_{\tau\in[0,1,\cdots,T-t]} E_t^Q\left[e^{-r\tau}\left\{\left(W_{t+\tau}^{t,w} - K_{t+\tau}\right)^+ - \left(W_{t+\tau}^{t,w} + \epsilon\frac{S_{t+\tau}}{S_t} - K_{t+\tau}\right)^+\right\}\right] \\
&\qquad \leq 0
\end{aligned}$$

where the last inequality is based on the fact that $\frac{S_{t+\tau}}{S_t}$ is strictly positive a.s.. Similarly we can show that v_1^h is non-decreasing in w:

$$\begin{aligned}
v_1^h(t,w) - v_1^h(t,w+\epsilon) &= e^{-r}E_t^Q\left[v_1\left(t+1,(w+cL_t)\frac{S_{t+1}}{S_t}\right) - v_1\left(t+1,(w+\epsilon+cL_t)\frac{S_{t+1}}{S_t}\right)\right] \\
&\leq 0
\end{aligned}$$

and immediately we see that v_1^e is non-decreasing in w as well.

- **Continuity in** w : For any $\epsilon > 0$, using the fact that $(w+\epsilon-k)^+ - (w-k)^+ \leq \epsilon$, we have

$$\begin{aligned}
&|v_1(t,w+\epsilon) - v_1(t,w)| \\
&\leq \left|\sup_{\tau\in[0,1,\cdots,T-t]} E_t^Q\left[e^{-r\tau}\left((w+\epsilon)\frac{S_{t+\tau}}{S_t} + \sum_{u=0}^{\tau}\frac{cS_{t+\tau}L_{u+t}}{S_{u+t}} - K_{t+\tau}\right)^+\right.\right. \\
&\qquad\qquad \left. -e^{-r\tau}\left(w\frac{S_{t+\tau}}{S_t} + \sum_{u=0}^{\tau}\frac{cS_{t+\tau}L_{u+t}}{S_{u+t}} - K_{t+\tau}\right)^+\right] \\
&\leq \sup_{\tau\in[0,1,\cdots,T-t]} E_t^Q\left[e^{-r\tau}\left(\epsilon\frac{S_{t+\tau}}{S_t}\right)\right] \\
&\leq \epsilon
\end{aligned}$$

Thus, v_1 is Lipschitz continuous in w and by non-decreasing property, we have $v_1(t,w+\epsilon) \leq v_1(t,w) + \epsilon$. The similar property for v_1^h follows immediately:

$$\begin{aligned}
v_1^h(t,w+\epsilon) &= E_t^Q\left[e^{-r}v\left(t+1,(w+\epsilon+cL_t)\frac{S_{t+1}}{S_t}\right)\right] \\
&\leq E_t^Q\left[e^{-r}v\left(t+1,(w+cL_t)\frac{S_{t+1}}{S_t}\right) + e^{-r}\epsilon\frac{S_{t+1}}{S_t}\right] \\
&= v_1^h(t,w) + \epsilon
\end{aligned}$$

In addition, we observe that w_1^e is Lipschitz continuous in w.

- **Convexity in w:** We follow Ben-Ameur *et al.* (2002), and prove the convexity by induction. For any $w_1 > 0$ and $w_2 > 0$, and $0 \le \lambda \le 1$

$$\begin{aligned} v_1^h(T-1, \lambda w_1 + (1-\lambda) w_2) &= E_{T-1}^Q \left[e^{-r} \left((\lambda w_1 + (1-\lambda) w_2 + cL_{T-1}) \frac{S_T}{S_{T-1}} - bL_{T-1} T \ddot{a}(T) \right)^+ \right] \\ &\le \lambda v_1^h(T-1, w_1) + (1-\lambda) v_1^h(T-1, w_2) \end{aligned}$$

Thus, v_1^h is convex at time T-1, and similarly v_1^e is convex at time $T-1$. Since

$$v_1(T-1, w) = \max\left(v_1^e(T-1, w), v_1^h(T-1, w)\right)$$

v_1 is also convex at time $T-1$.

We now assume that result holds for time $t+1$, where $0 \le t \le T-2$, then the continuation value at time t is

$$\begin{aligned} &v_1^h(t, \lambda w_1 + (1-\lambda) w_2) \\ &\quad = E_t^Q \left[e^{-r} v_1 \left(t+1, (\lambda w_1 + (1-\lambda) w_2 + cL_t) \frac{S_{t+1}}{S_t} \right) \right] \\ &\quad \le E_t^Q \left[e^{-r} \left(\lambda v_1 \left(t+1, (w_1 + cL_t) \frac{S_{t+1}}{S_t} \right) + (1-\lambda) v_1 \left(t+1, (w_2 + cL_t) \frac{S_{t+1}}{S_t} \right) \right) \right] \\ &\quad = \lambda E_t^Q \left[e^{-r} v_1 \left(t+1, (w_1 + cL_t) \frac{S_{t+1}}{S_t} \right) \right] + (1-\lambda) E_t^Q \left[e^{-r} v_1 \left(t+1, (w_2 + cL_t) \frac{S_{t+1}}{S_t} \right) \right] \\ &\quad = \lambda v_1^h(t, w_1) + (1-\lambda) v_1^h(t, w_2) \end{aligned}$$

Thus, v_1^h is a convex function at time t. Since v_1^e is convex for all t, and $v_1(t, w) = \max\left(v_1^e(t, w), v_1^h(t, w)\right)$ holds for all t, then $v_1(t, w)$ is convex function at time t. Then, by induction, we have proved the convexity of $v_1(t, w)$.

□

Like other Bermudan and American-type options, there exists a continuation region **C** and a stopping region (or exercising region) **D**. When the time and DC account value pair, (t, w), is in the continuation region, it is optimal for the member to stay in the DC plan. When (t, w) is in the stopping region, it is optimal for the member to switch to the DB plan. The option to switch is exercised when (t, w) moves into the stopping region. Mathematically, the continuation and stopping regions are defined as

- $v_1^h(t,w) > v_1^e(t,w) \Leftrightarrow (t,w) \in \mathbf{C}$
- $v_1^h(t,w) \leq v_1^e(t,w) \Leftrightarrow (t,w) \in \mathbf{D}$

Proposition 2. *There exists a function $\varphi(t)$ such that*

$$v_1(t,w) = \begin{cases} v_1^e(t,w) & \text{if } w \geq \varphi(t) \\ v_1^h(t,w) & \text{if } w < \varphi(t) \end{cases}$$

Proof. Consider first the case when $w < K_t$ at time t. We have $v_1^e(t,w) = (w - K_t)^+$, so

$$w < K_t \Rightarrow v_1^e(t,w) = 0$$

Then immediately we have $v_1^h(t,w) > 0 = v_1^e(t,w)$. Thus if $w \in [0, K_t]$, it cannot be optimal to exercise. Therefore, in order to explore the exercise frontier, we need only consider cases when $w > K_t$.

When $w > K_t$ we have $v_1^e(t,w) = w - K_t$ and

$$(t,w) \in \mathbf{D} \implies v_1(t,w) = v_1^e(t,w) \implies v_1(t,w) = w - K_t$$

First, assume that $(t,w) \in \mathbf{D}$, then by the continuity of v_1^h and v_1^e in w, we have

$$\begin{aligned} (t,w) \in D &\implies v^e(t,w) \geq v^h(t,w) \\ &\implies \forall \epsilon > 0 \quad v^e(t,w+\epsilon) = v^e(t,w) + \epsilon \geq v^h(t,w) + \epsilon \geq v^h(t,w+\epsilon) \\ &\implies (t, w+h) \in \mathbf{D} \end{aligned}$$

Next, assume that $(t,w) \in \mathbf{C}$. Note that $v_1^h(t,w) + \epsilon \geq v_1^h(t,w+\epsilon)$ for all $\epsilon > 0$ implies that $v_1^h(t,w-\epsilon) \geq v_1^h(t,w) - \epsilon$

$$\begin{aligned} (t,w) \in \mathbf{C} &\implies v_1^h(t,w) > v_1^e(t,w) \\ &\implies \forall \epsilon > 0 \quad v_1^h(t,w-\epsilon) \geq v_1^h(t,w) - \epsilon > v_1^e(t,w) - \epsilon = v_1^e(t,w-\epsilon) \\ &\implies (t,w-\epsilon) \in \mathbf{C} \end{aligned}$$

These results show that it is optimal for the employee to switch to the DB plan only when his/her DC account balance is above a certain threshold ($\varphi(t)$) at each possible switching time. □

The function $\varphi(t)$ is often known as the exercise boundary that separates the continuation and exercise regions. Notice that it is possible, under certain parameters, that $\varphi(t) = \infty$ for some $t < T$, which means that it is not optimal to switch regardless of the DC account value. The next proposition will specify the situations in which $\varphi(t) < \infty$.

Proposition 3. *The behavior of the exercise boundary $\varphi(t)$ depends on the ratio $\dfrac{c}{b\,\ddot{a}(T)\,e^{-rT}}$ as follows.*

(i). If $\dfrac{c}{b\,\ddot{a}(T)e^{-rT}} < 1$, then $\varphi(t) < \infty, \forall t \in [0, T]$.

(ii). If $\dfrac{c}{b\,\ddot{a}(T)e^{-rT}} \geq 1$, there exists a $t_* \in [0, 1, \cdots, T-1]$, such that $\varphi(t) = \infty, \forall t \leq t_*$, and $\varphi(t) < \infty, \forall t > t_*$.

(iii). The value of t_* is $\lfloor t' \rfloor$, where $\lfloor \cdot \rfloor$ represents the floor function, and t' satisfies

$$(t'+1)bL_{t'}\ddot{a}(T)e^{-r(T-t')} - t'bL_{t'-1}\ddot{a}(T)e^{-r(T-t')} - cL_{t'} = 0.$$

(iv). If $c > b\,\ddot{a}(T)\left(\left(1-e^{-\mu_L}\right)T + e^{-\mu_L}\right)e^{-r}$ then $t_* = T-1$, which means $\varphi(t) = \infty$, $\forall t < T$.

(v). If $\dfrac{c}{b\,\ddot{a}(T)e^{-rT}} = 1$, then $t_* = 0$.

Proof. First, notice that if $\varphi(t) < \infty$, then for sufficiently large w

$$\begin{aligned}
v_1(t,w) = v_1^e(t,w) \geq v_1^h(t,w) &= E_t^Q\left[e^{-r}v_1\left(t+1, (w+cL_t)\frac{S_{t+1}}{S_t}\right)\right] \\
&\geq E_t^Q\left[e^{-r}v_1^e\left(t+1, (w+cL_t)\frac{S_{t+1}}{S_t}\right)\right] \\
&= (w+cL_t)\,N\left(d_{1,t,w}\right) - (t+1)bL_t\ddot{a}(T)e^{-r(T-t-1)}e^{-r}N(d_{2,t,w})
\end{aligned}$$

where

$$\begin{aligned}
d_{1,t,w} &= \frac{1}{\sigma_S}\left[\ln\left(\frac{w+cL_t}{(t+1)L_t b\ddot{a}(T)e^{-(T-t-1)r}}\right) + \left(r + \frac{\sigma_S^2}{2}\right)\right] \\
d_{2,t,w} &= d_{1,t,w} - \sigma_S
\end{aligned}$$

which is the Black-Scholes Formula, with initial stock price $w + cL_t$ and strike value

$(t+1)bL_t\ddot{a}(T)e^{-r(T-t-1)}$. Here we define

$$\begin{aligned}
f(t) &= \lim_{w\to\infty} v_1^e(t,w) - E_t^Q\left[e^{-r}v^e\left(t+1,(w+cL_t)\frac{S_{t+1}}{S_t}\right)\right] \\
&= \lim_{w\to\infty} v_1^e(t,w) - \big((w+cL_t)\,N\left(d_{1,t,w}\right) - (t+1)bL_t\ddot{a}(T)e^{-r(T-t-1)}e^{-r}N(d_{2,t,w})\big) \\
&= \lim_{w\to\infty} w - \big((w+cL_t)\,N\left(d_{1,t,w}\right) - (t+1)bL_t\ddot{a}(T)e^{-r(T-t)}N(d_{2,t,w})\big) - tbL_{t-1}\ddot{a}(T)e^{-r(T-t)} \\
&= \lim_{w\to\infty} - \big((t+1)bL_t\ddot{a}(T)e^{-r(T-t)}N(-d_{2,t,w}) - (w+cL_t)\,N\left(-d_{1,t,w}\right)\big) \\
&\qquad\qquad - cL_t + (t+1)bL_t\ddot{a}(T)e^{-r(T-t)} - tbL_{t-1}\ddot{a}(T)e^{-r(T-t)} \\
&= (t+1)bL_t\ddot{a}(T)e^{-r(T-t)} - tbL_{t-1}\ddot{a}(T)e^{-r(T-t)} - cL_t
\end{aligned}$$

Clearly, whenever $\varphi(t) < \infty$, we have $f(t) \geq 0$.

If $\dfrac{c}{b\ddot{a}(T)e^{-rT}} < 1$, we prove $\varphi(t) < \infty, \forall t \in [0,T]$ by induction.

At time T, $\varphi(T) = TbL_{T-1}\ddot{a}(T) < \infty$.

At time t, assume $\varphi(t+1) < \infty$. We observe that for sufficiently large $w < \infty$, we have $v_1(t+1,w) = v_1^e(t+1,w)$. Next, we can show

$$\begin{aligned}
&\lim_{w\to\infty} v_1^h(t,w) - E_t^Q\left[e^{-r}v_1^e\left(t+1,(w+cL_t)\frac{S_{t+1}}{S_t}\right)\right] \\
=&\lim_{w\to\infty} E_t^Q\left[e^{-r}\left(v_1\left(t+1,(w+cL_t)\frac{S_{t+1}}{S_t}\right) - v_1^e\left(t+1,(w+cL_t)\frac{S_{t+1}}{S_t}\right)\right)\right] \\
=&0
\end{aligned}$$

The last line is due to the fact that if $w \geq \varphi(t+1)$

$$v_1\left(t+1,w\right) - v_1^e\left(t+1,w\right) = 0 < \varphi(t+1) < \infty$$

and for $w < \varphi(t+1)$

$$v_1\left(t+1,w\right) - v_1^e\left(t+1,w\right) \leq v_1^e\left(t+1,\varphi(t+1)\right) \leq \varphi(t+1) < \infty$$

since $v_1(t+1,w)$ is an increasing function of w. Thus, the difference is bounded by

$\varphi(t+1) < \infty$ and we are able to apply the Dominated Convergence Theorem. Next,

$$\begin{aligned}\lim_{w\to\infty} v_1^e(t,w) - v_1^h(t,w) &= \lim_{w\to\infty} v_1^e(t,w) - E_t^Q\left[e^{-r} v_1^e\left((w + cL_t)\frac{S_{t+1}}{S_t}\right)\right] \\ &= (t+1)bL_t\ddot{a}(T)e^{-r(T-t)} - tb\ddot{a}(T)e^{-r(T-t)}L_{t-1} - cL_t \\ &> (t+1)bL_t\ddot{a}(T)e^{-r(T-t)} - tb\ddot{a}(T)e^{-r(T-t)}L_{t-1} - L_t b\ddot{a}(T)e^{-rT} \\ &> (t+1)bL_t\ddot{a}(T)e^{-r(T-t)} - tb\ddot{a}(T)e^{-r(T-t)}L_t - L_t b\ddot{a}(T)e^{-r(T-t)} \\ &= 0\end{aligned}$$

Which implies $\varphi(t) < \infty$ (otherwise if $\varphi(t) = \infty$, $\lim_{w\to\infty} v_1^e(t,w) - v_1^h(t,w) \le 0$).
Thus $\varphi(t) < \infty, \forall t \in [0,T]$.

For $\dfrac{c}{b\ddot{a}(T)e^{-rT}} \ge 1$, we split the proof into three parts.

(1) If $\frac{c}{b\ddot{a}(T)e^{-rT}} > 1$, we first prove that there exists a t_* such that $\varphi(t) = \infty, \forall t \le t_*$, then prove that $\varphi(t) < \infty, \forall t > t_*$ by induction from time T to $t_* + 1$ as above.

We have $f(0) < 0$ so that

$$\lim_{w\to\infty} v_1^e(0,w) - v_1^h(0,w) \le f(0) < 0$$

and thus $\varphi(0) = \infty$. Also, notice we can write $f(t)$ in the form

$$f(t) = e^{\mu_L t}\left(te^{-r(T-t)}b\ddot{a}(T)(1 - e^{-\mu_L}) + b\ddot{a}(T)e^{-r(T-t)} - c\right) = e^{\mu_L t}h(t)$$

where $h(t)$ is a strictly increasing function of time t, if both r and μ_L are non-negative, with at least one of them being strictly positive. We have $f(0) < 0$ and

$$f\left(\frac{1}{r}\log\left(\frac{c}{b\ddot{a}(T)e^{-rT}}\right)\right) > 0$$

$$\text{where } \frac{1}{r}\log\left(\frac{c}{b\ddot{a}(T)e^{-rT}}\right) > 0 \quad \text{by assumption.}$$

Then, we can find t' such that

$$\begin{aligned} f(t) &< 0 \text{ for } t < t' \\ f(t) &= 0 \text{ for } t = t' \\ f(t) &> 0 \text{ for } t > t' \end{aligned}$$

Here we set $t_* = \lfloor t' \rfloor$, and we have

$$\lim_{w\to\infty} v_1^e(t,w) - v_1^h(t,w) \le f(t) < 0, \forall t < t_*$$

and for $t = t_*$, first notice that

$$f(t_*) \le 0 \implies c \ge b\ddot{a}(T)e^{-rT}\left(t_* + 1 - t_* e^{-\mu_L}\right)e^{rt_*}$$

Next, for any finite $w > t_* L_{t_*-1} b\ddot{a}(T)e^{-r(T-t_*)}$, we have

$$\begin{aligned}
v_1^h(t_*,w) &= E_t^Q\left[e^{-r}v_1\left(t_*+1,(w+cL_{t_*})\frac{S_{t_*+1}}{S_{t_*}}\right)\right] \\
&> E_t^Q\left[e^{-r}v_1^e\left(t_*+1,(w+cL_{t_*})\frac{S_{t_*+1}}{S_{t_*}}\right)\right] \\
&= E_t^Q\left[e^{-r}\left((w+cL_{t_*})\frac{S_{t_*+1}}{S_{t_*}} - (t_*+1)L_{t_*}b\ddot{a}(T)e^{-r(T-t_*-1)}\right)^+\right] \\
&\ge \max\left(0, w + cL_{t_*} - (t_*+1)L_{t_*}b\ddot{a}(T)e^{-r(T-t_*)}\right) \\
&\ge \max\left(0, w - t_* L_{t_*-1} b\ddot{a}(T)e^{-rT}e^{rt_*}\right) \\
&= v_1^e(t_*,w)
\end{aligned}$$

The third to fourth line follows from Jensen's Inequality:

$$\begin{aligned}
&E_t^Q\left[e^{-r}\left((w+cL_{t_*})\frac{S_{t_*+1}}{S_{t_*}} - (t_*+1)L_t b\ddot{a}(T)e^{-r(T-t_*-1)}\right)^+\right] \\
&\ge \max\left(0, E_t^Q\left[e^{-r}\left((w+cL_{t_*})\frac{S_{t_*+1}}{S_{t_*}} - (t_*+1)L_t b\ddot{a}(T)e^{-r(T-t_*-1)}\right)\right]\right) \\
&\quad = \max\left(0, w + cL_{t_*} - (t_*+1)L_t b\ddot{a}(T)e^{-r(T-t_*)}\right)
\end{aligned}$$

Thus, we have $v_1^h(t_*,w) > v_1^e(t_*,w), \forall w < \infty$, and

$$\lim_{t\to\infty} v_1^e(t_*,w) - v_1^h(t_*,w) \le f(t_*) \le 0$$

which implies $\varphi(t_*) = \infty$.

Repeating the induction:

At time T, again we have $\varphi(T) < \infty$.
At time $t > t_*$, assume $\varphi(t+1) < \infty$.

$$\lim_{w\to\infty} v_1^e(t,w) - v_1^h(t,w) = f(t) > 0, t > t_*$$

Thus $\varphi(t) < \infty, \forall t > t_*$.

(2) If $c > b\ddot{a}(T)\left(\left(1-e^{-\mu_L}\right)T + e^{-\mu_L}\right)e^{-r}$

$$\begin{aligned} c &> b\ddot{a}(T)e^{-r}\left(T-(T-1)e^{-\mu_L}\right) \\ &> b\ddot{a}(T)e^{-r} \geq b\ddot{a}(T)e^{-rT} \end{aligned}$$

Thus, we know there exists t_* as defined previously, and for time $T-1$,

$$\begin{aligned} &\lim_{w\to\infty} v_1^e(T-1,w) - v_1^h(T-1,w) \\ &\quad = TbL_{T-1}\ddot{a}(T)e^{-r} - (T-1)bL_{T-2}\ddot{a}(T)e^{-r} - cL_{T-1} \\ &\quad < TbL_{T-1}\ddot{a}(T)e^{-r} - (T-1)bL_{T-2}\ddot{a}(T)e^{-r} - b\ddot{a}(T)\left(\left(1-e^{-\mu_L}\right)T + e^{-\mu_L}\right)e^{-r}L_{T-1} \\ &\quad = 0 \end{aligned}$$

Thus, $\varphi(t) = \infty, \forall t \leq T-1$, and the option simplifies to the DB underpin option.

(3) When $\dfrac{c}{b\ddot{a}(T)e^{-rT}} = 1$, we have $f(0) = 0$. Thus $t_* = t' = 0$, immediately we have $\varphi(0) = \infty$.

□

This proposition demonstrates that the ratio between the DC contribution rate and the accrual rate of the DB benefit will determine the shape of the exercise boundary. In the extreme case, when the DC contribution rate is much higher than the DB accrual rate, it is optimal for the employee to wait until the retirement date, and the Early Exercise DB underpin option simplifies to the DB underpin option.

2.6 Early Exercise DB underpin plan - Continuous Case

By extending the current model to incorporate a stochastic salary process, solving the value of the option becomes a three dimensional problem. The optimal exercise boundary will depend on both the salary and the DC account balance. In this section, we reconstruct the problem in a continuous setting, with a stochastic salary.

The mathematical formulation replaces the summation sign with the integral sign, and extends the admissible stopping time set to $[t, T]$. By the Optional Sampling Theorem, we can define the value function similarly to the discrete case.

$$v_2(t,w,l) = \sup_{0\le\tau\le T-t} E_t^Q\left[e^{-r\tau}\left(W_{t+\tau}-K_{t+\tau}\right)^+ \middle| W_t = w, L_t = l\right]$$

where

$$\begin{aligned}
dW_t =& (rW_t + cL_t)dt + \sigma_S W_t dZ_S^Q(t), \quad W_0 = 0\\
dL_t =& rL_t dt + \sigma_L L_t dZ_L^Q(t), \quad L_0 > 0\\
K_t =& b\, t\, \ddot{a}(T)\, L_t\, e^{-r(T-t)}
\end{aligned}$$

Notice now that the value function also depends on the salary at time t. In the following sections we show that the dimensionality of the problem can be reduced to two, in the sense that the value function only depends on the wealth-salary ratio process.

2.6.1 Definition of the Value Function

We define a new variable $Y_t = \frac{W_t}{L_t}$, which represents the DC wealth-salary ratio. Then we can rewrite our value function:

$$\begin{aligned}
v_2(t,w,l) &= \sup_{0\le\tau\le T-t} E_t^Q\left[e^{-r\tau}\left(W_{\tau+t}^{t,w}-K_{t+\tau}\right)^+\right]\\
&= \sup_{0\le\tau\le T-t} E_t^Q\left[e^{-r\tau}L_{\tau+t}\left(\frac{w}{L_{\tau+t}}\frac{S_{t+\tau}}{S_t} + c\int_0^\tau \frac{S_{\tau+t}L_u}{S_u L_{t+\tau}} - b(t+\tau)\ddot{a}(T)e^{-r(T-t-\tau)}\right)^+\right]\\
&= \sup_{0\le\tau\le T-t} E_t^Q\left[e^{-r\tau}L_{\tau+t}\left(Y_{t+\tau}^{t,w/l} - b(t+\tau)\ddot{a}(T)e^{-r(T-t-\tau)}\right)^+\right]\\
&= \sup_{0\le\tau\le T-t} E_t^Q\left[e^{-r\tau}le^{\left(r-\frac{\sigma_L^2}{2}\right)\tau+\sigma_L(Z_L^Q(t+\tau)-Z_L^Q(t))}\left(Y_{t+\tau}^{t,w/l} - b(t+\tau)\ddot{a}(T)e^{-r(T-t-\tau)}\right)^+\right]\\
&= \sup_{0\le\tau\le T-t} E_t^{\tilde{P}}\left[\left(Y_{\tau+t}^{t,w/l} - b(\tau+t)\ddot{a}(T)e^{-r(T-t-\tau)}\right)^+\right]\times l
\end{aligned}$$

where $\tilde{P}$ is a new probability measure such that $d\tilde{P} = \exp\Big(\sigma_L Z_L^Q(t) - (\sigma_L^2/2)t\Big)\, dQ$. By Girsanov's Theorem, we have

$$\begin{pmatrix} Z_S^{\tilde{P}}(t) \\ Z_L^{\tilde{P}}(t) \end{pmatrix} = \begin{pmatrix} Z_S^Q \\ Z_L^Q \end{pmatrix} - \begin{pmatrix} 0 \\ \sigma_L \end{pmatrix} t$$

as a two-dimensional standard Brownian motion under $\tilde{P}$. It is straightforward to show that under the new probability measure $\tilde{P}$, the wealth-salary ratio has the following SDE:

$$dY_t = cdt + Y_t \sigma_Y dZ_Y^{\tilde{P}}(t), \quad Y_0 = 0$$

where $\sigma_Y = \sqrt{\sigma_S^2 + \sigma_L^2 - 2\sigma_S\sigma_L\rho}$ and $Z_Y^{\tilde{P}}(t)$ is a standard Brownian Motion under the measure $\tilde{P}$. We can define a new function (2-dimensional):

$$v_2(t,y) = \sup_{0 \le \tau \le T-t} E_t^{\tilde{P}} \left[\left(Y_{t+\tau}^{t,y} - b(t+\tau)\ddot{a}(T)e^{-(T-t-\tau)r} \right)^+ \right]$$

Thus, $v_2(t, w/l) = v_2(t, w, l)/l$, which clearly suggests that the exercise rule depends on the wealth-to-salary ratio.

2.6.2 Properties of the Value Function and Exercise Frontier

Proposition 4. *The value function $v_2(t, y)$ is a strictly positive, non-decreasing, continuous and convex function of y, and a continuous and non-increasing function of time t.*

Proof. The strict positive property is trivial to see, and we refer Touzi (2013) for the proofs of other properties:

- **Non-decreasing in y:** By writing the value function explicitly

$$v_2(t,y) = \sup_{0 \le \tau \le T-t} E_t^{\tilde{P}} \left[\left(y \frac{S(t+\tau)}{L(t+\tau)S(t)} + c \int_t^{t+\tau} \frac{S(t+\tau)L(u)}{S(u)L(t+\tau)} du - b(t+\tau)\ddot{a}(T)e^{-(T-t-\tau)r} \right)^+ \right]$$

we immediately see that $y \to v_2(t, y)$ is an increasing and convex function on $[0, \infty)$.

- **Continuity in y:** For $x > y$

$$
\begin{aligned}
&|v_2(t,x) - v_2(t,y)| \\
\le & \left| \sup_{0\le\tau\le T-t} E_t^{\tilde{P}} \left[\left(Y_{t+\tau}^{t,x} - b(t+\tau)\ddot{a}(T)e^{-(T-t-\tau)r} \right)^+ - \left(Y_{t+\tau}^{t,y} - b(t+\tau)\ddot{a}(T)e^{-(T-t-\tau)r} \right)^+ \right] \right| \\
\le & \left| \sup_{0\le\tau\le T-t} E_t^{\tilde{P}} \left[Y_{t+\tau}^{t,x} - Y_{t+\tau}^{t,y} \right] \right| \\
\le & \left| \sup_{0\le\tau\le T-t} E_t^{\tilde{P}} \left[x\frac{L_t S_{t+\tau}}{L_{t+\tau}S_t} - y\frac{L_t S_{t+\tau}}{L_{t+\tau}S_t} \right] \right| \\
\le & \left| (x-y) \sup_{0\le\tau\le T-t} E_t^{\tilde{P}} \left[\frac{L_t S_{t+\tau}}{L_{t+\tau}S_t} \right] \right| \\
\le & (x-y)
\end{aligned}
$$

- **Non-increasing in t:** Since $Y_{t_1+t}^{t_1,y} \overset{law}{=} Y_{t_2+t}^{t_2,y}$, and the strike function $G_t = tbe^{-(T-t)r}\ddot{a}(T)$ is an increasing function of time t, immediately we have $v_2(t,y)$ is non-increasing in t.

- **Continuity in t:** For $t_2 > t_1$,

$$
\begin{aligned}
0 \le & v_2(t_1,y) - E_{t_1}^{\tilde{P}}\left[v(t_2, Y_{t_2}^{t_1,y}) \right] \\
= & \sup_{\tau_1\le T-t_1} E_{t_1}^{\tilde{P}} \left[\mathbf{1}_{\tau_1<t_2-t_1} v_2^e\left(t_1+\tau_1, Y_{t_1+\tau_1}^{t_1,y}\right) + \mathbf{1}_{\tau_1\ge t_2-t_1} v_2\left(t_2, Y_{t_2}^{t_1,y}\right) \right] - E_{t_1}^{\tilde{P}}\left[v_2(t_2, Y_{t_2}^{t_1,y}) \right] \\
\le & \sup_{\tau_1\le T-t_1} E_{t_1}^{\tilde{P}} \left[\mathbf{1}_{\tau_1<t_2-t_1} v_2^e\left(t_1+\tau_1, Y_{t_1+\tau_1}^{t_1,y}\right) + \mathbf{1}_{\tau_1\ge t_2-t_1} v_2\left(t_2, Y_{t_2}^{t_1,y}\right) - v_2(t_2, Y_{t_2}^{t_1,y}) \right] \\
= & \sup_{\tau_1\le T-t_1} E_{t_1}^{\tilde{P}} \left[\mathbf{1}_{\tau_1<t_2-t_1} v_2^e(t_1+\tau_1, Y_{t_1\tau_1}^{t_1,y}) - v_2(t_2, Y_{t_2}^{t_1,y}) \right] \\
\le & \sup_{\tau_1\le T-t_1} E_{t_1}^{\tilde{P}} \left[\mathbf{1}_{\tau_1<t_2-t_1} v_2^e(t_1+\tau_1, Y_{t_1\tau_1}^{t_1,y}) - v_2^e(t_2, Y_{t_2}^{t_1,y}) \right] \\
\le & \sup_{\tau_1\le T-t_1} E_{t_1}^{\tilde{P}} \left[\mathbf{1}_{\tau_1<t_2-t_1} \left(Y_{\tau_1+t_1}^{t_1,y} - (t_1+\tau_1)b\ddot{a}(T)e^{-(T-t_1-\tau_1)r} \right)^+ - \left(Y_{t_2}^{t_1,y} - t_2 b\ddot{a}(T)e^{-(T-t_2)r} \right)^+ \right] \\
\le & E_{t_1}^{\tilde{P}} \left[\sup_{\tau_1\in[t_1,t_2]} \left| Y_{t_1+\tau_1}^{t_1,y} - Y_{t_2}^{t_1,y} \right| + \sup_{\tau_1\in[t_1,t_2]} \left| t_2 b\ddot{a}(T)e^{-(T-t_2)r} - (t_1+\tau_1)b\ddot{a}(T)e^{-(T-t_1-\tau_1)r} \right| \right] \\
\le & E_{t_1}^{\tilde{P}} \left[\sup_{\tau_1\in[t_1,t_2]} \left| Y_{t_1+\tau_1}^{t_1,y} - Y_{t_2}^{t_1,y} \right| \right] + \left(t_2 b\ddot{a}(T)e^{-(T-t_2)r} - t_1 b\ddot{a}(T)e^{-(T-t_1)r} \right) \\
\le & C_1\sqrt{t_2 - t_1}
\end{aligned}
$$

where the last line comes from Touzi (2012), Theorem 2.4, that there exists a constant C such that

$$E_{t_1}^{\tilde{P}}\left[\sup_{\tau_1\in[t_1,t_2]}\left|Y_{t_1+\tau_1}^{t_1,y}-Y_{t_2}^{t_1,y}\right|\right]\leq C(1+|y|)\sqrt{t_2-t_1}$$

Next, we have

$$\begin{aligned}|v_2(t_1,y)-v_2(t_2,y)| &\leq \left|v_2(t_1,y)-E_{t_1}^{\tilde{P}}\left[v_2(t_2,Y_{t_2}^{t_1,y})\right]\right|+\left|E_{t_1}^{\tilde{P}}\left[v_2(t_2,Y_{t_2}^{t_1,y})\right]-v_2(t_2,y)\right| \\ &\leq C_1\sqrt{t_2-t_1}+E_{t_1}^{\tilde{P}}\left[\left|v_2(t_2,Y_{t_2}^{t_1,y})-v_2(t_2,y)\right|\right] \\ &\leq C_1\sqrt{t_2-t_1}+C_2E_{t_1}^{\tilde{P}}\left[\left|Y_{t_2}^{t_1,y}-y\right|\right] \\ &\leq C\sqrt{t_2-t_1}\end{aligned}$$

where the third line follows from the Lipschitz continuity of the value function in y. Thus, we have proved that the value function is Hölder-Continuous in t.

□

As above, we denote the continuation region $\mathbf{C}=\{(t,y)\in(0,T)\times[0,\infty):v_2(t,y)>v_2^e(t,y)\}$ and stopping region $\mathbf{D}=\{(t,y)\in(0,T)\times[0,\infty):v_2(t,y)=v_2^e(t,y)\}$, and let $\varphi(t)$ be the exercise frontier which separates the two regions.

Proposition 5. *There exists a function $\varphi(t)$ such that*

$$v_2(t,y)=\begin{cases}v_2^e(t,y) & \text{if } y\geq\varphi(t)\\ v_2^h(t,y) & \text{if } y<\varphi(t)\end{cases}$$

Proof. Denote the continuation region $\mathbf{C}=\{(t,y)\in(0,T)\times[0,\infty):v_2(t,y)>v_2^e(t,y)\}$ and stopping regions $\mathbf{C}=\{(t,y)\in(0,T)\times[0,\infty):v_2(t,y)=v_2^e(t,y)\}$. Then, by standard arguments based on the strong Markov Property (see Peskir and Shiryaev (2006) Corollary 2.9), the first hitting time $\tau_D=\inf\{0\leq s\leq T-t:(t+s,Y_{t+s})\in D\}$ is optimal, and the value function is $C^{1,2}$ on $\mathbf{C}$ and satisfies:

$$\frac{\partial v_2}{\partial t}+\mathbf{L}_Y v_2=0 \quad in \quad \mathbf{C}$$

where $\mathbf{L}_Y$ is the infinitesimal generator.

If $(t, y) \in \mathbf{D}$,

$$v_2(t, y+h) \leq v_2(t, y) + h = v_2^e(t, y) + h = v_2^e(t, y+h)$$

thus $(t, y+h) \in \mathbf{D}$.

If $(t, y) \in \mathbf{C}$,
if $y - h > G_t$, $v_2(t, y-h) \geq v_2(t, y) - h > v_2^e(t, y) - h = v_2^e(t, y-h)$
if $y - h < G_t$, $v_2(t, y-h) > 0 = v_2^e(t, y-h)$, thus $(t, y-h) \in \mathbf{C}$

□

For numerical solution, we formulate the problem as a free-boundary PDE problem.

$$\begin{aligned}
&\frac{\partial v_2}{\partial t} + c\frac{\partial v_2}{\partial y} + \frac{\sigma_Y^2 y^2}{2}\frac{\partial^2 v_2}{\partial y^2} = 0 \quad in \quad \mathbf{C} \\
&v_2(t, y) = v_2^e(t, y) \quad in \quad \mathbf{D} \\
&v_2(t, y) > v_2^e(t, y) \quad in \quad \mathbf{C} \\
&v_2(t, y) = v_2^e(t, y) \quad for \quad y = \varphi(t) \quad or \quad t = T
\end{aligned}$$

Remark. *Define and substitute $g(t, y) = (y - G_t)$ for $y > G_t$ into the variational inequality, and define*

$$f(t) = -\frac{\partial g}{\partial t} - c\frac{\partial g}{\partial y} - \frac{\sigma_Y^2 y^2}{2}\frac{\partial^2 g}{\partial y^2} = b\ddot{a}(T)e^{-r(T-t)} + rbt\ddot{a}(T)e^{-r(T-t)} - c$$

If $f(t) < 0$, then there is a contradiction with the variational inequality, which means $v_2(t, y) \neq g(t, y)$ for all y, and thus $(t, y) \in \mathbf{C}, \forall y$ ($\varphi(t) = \infty$). Since $f(t)$ is an increasing function of t, there exists a t^* satisfying $f(t^*) = 0$. If $t^* \in [0, T]$, then $(t, y) \in \mathbf{C}, \forall (t, y) \in [0, t^*] \times \mathbf{R}$. In particular, if $c > b\ddot{a}(T)(1 + rT)$ (when the DC contribution rate is extremely high or the horizon is short), the option is equivalent to a European Option.

Figure 2.1 shows an example of the optimal exercise boundary. As mentioned before, the member should switch to the DB plan only when they have already saved a certain extra amount over the cost of ABO, and in this graph, this excess amount is relatively stable. When the time comes very close to the retirement date, the plan member is facing less future opportunity to obtain a higher DC account balance from the equity market, and thus is willing to make the transfer to the DB plan even with lower DC balance, in order to secure their benefit. In this example, the decrease of the optimal exercise boundary starts only at the last few month, and eventually converges to the value of the DB plan.

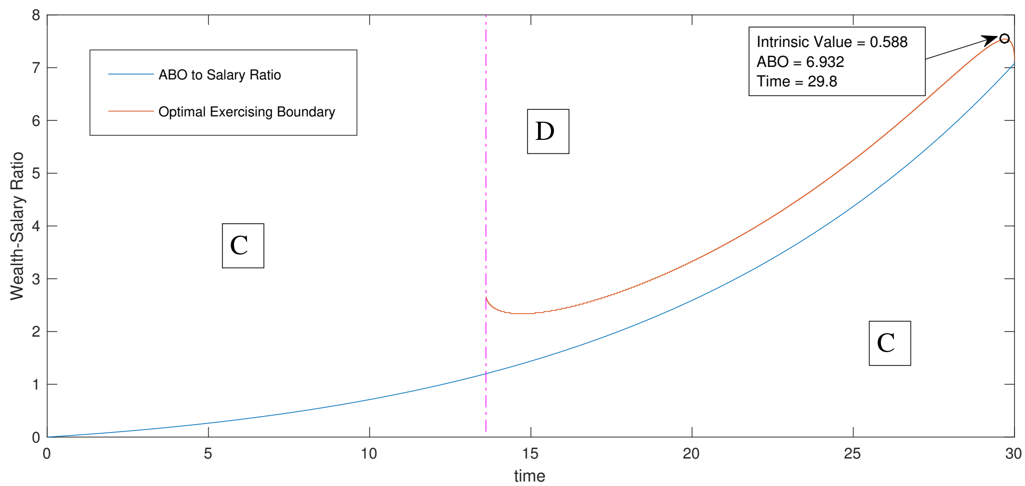


Figure 2.1: Example of Optimal Exercising Boundary, $r = 0.06$, $\sigma_S = 0.15$, $\sigma_L = 0.04$, $c = 0.16$, $b = 0.016$ and $\ddot{a}(T) = 14.75$

2.6.3 Different Discount Rates for ABO Calculation

In practice, actuaries often use a discount rate higher than the observed market risk-free rate to determine the ABO. Here we denote by γ the discount rate for the ABO calculation, so the ABO has the following form:

$$K_t = tbL_t\ddot{a}(T)e^{-\gamma(T-t)}$$

It is not difficult to show that the new cost function can be expressed in the same form as equation (2.1) through the Optional Sampling Theorem. Under our stochastic salary assumption, however, the risk-free rate r only appears in the ABO calculations, which means that γ and r are mathematically indistinguishable. Thus, we only include γ in our study of deterministic salary assumptions. The new value function preserves similar properties (with slight modifications) as those stated in Propositions (1), (2), (3) for the discrete case, and (4) and (5) in the continuous case. Notice that using different discount rates for the ABO does not violate our market-consistent valuation principle. Here we use

v_4 to denote the price with γ as discount rate for the ABO:

$$v_4(t, W_t = w) = \sup_{0 \le \tau \le T-t} E_t^Q \left[e^{-r\tau} \left(W_{t+\tau} - b(t+\tau) L_{t+\tau} \ddot{a}(T) e^{-\gamma(T-t-\tau)} \right)^+ \right]$$

2.7 Numerical Examples

In this section we present numerical results for the values of the Early Exercise DB Underpin plan. For the continuous setting, we use the penalty method (see Forsyth and Vetzal (2002)) and for the discrete case, we use the Least Square Method (see Longstaff and Schwartz (2001)). We compare the Early Exercise option with the DB underpin and with the Florida Second Election option. For the DB underpin option, to be consistent with the valuation method for the Early Exercise DB underpin option, we use Monte Carlo simulation in the discrete setting and the Crank Nicolson finite difference method in the continuous setting. For the second election option, we are able to derive explicit solutions under both the discrete and continuous settings (see Appendix B.1). Furthermore, we include the Early Exercise option with deterministic salary in the continuous setting (denoted by v_3) as an intermediate comparison with the discrete case.

2.7.1 Benchmark Scenario

The initial benchmark parameter set is as follows. Later we perform some sensitivity tests on each of the parameters.

- $\mu_L = r = 0.04$. We set μ_L to be the same as the risk-free rate, to make a consistent comparison with the stochastic salary assumption (when the salary is assumed to be hedgeable).
- $\rho = 0$, assumes no correlation between salary and investment return.
- $\sigma_S = 0.15$, $\sigma_L = 0.04$, $b = 0.016$, $c = 0.125$ and $\ddot{a}(T) = 14.75$.
- $L_0 = 1$, $t = 0$, so that all values are given per unit of starting salary.

2.7.2 Cost

Table 2.1 shows the value of each pension plan in the continuous setting. The prices for the Early Exercise DB-Underpin (EEDBU), the Florida Second Election (FSE) and the DB underpin plans are expressed as an additional cost on top of the base DB plan. We show values for the Early Exercise DB with stochastic salaries v_2, and with deterministic salaries, v_3. Table 2.2 shows the results in the discrete setting.

Time to Retirement T	DB	DC	FSE	DBU	EEDBU(s) v_2	EEDBU(d) v_3
10yr	2.36	1.25	0	0.0023	0.0070	0.0062
15yr	3.54	1.87	0	0.0126	0.0354	0.0315
20yr	4.72	2.50	0.0203	0.0348	0.1010	0.0936
30yr	7.08	3.75	0.2179	0.1199	0.3492	0.3355
40yr	9.44	5.00	0.5837	0.2594	0.7380	0.7194

Table 2.1: Cost of each pension plan per unit of starting salary, continuous setting. Hybrid costs are additional to the basic DB cost. FSE is the Florida second election option; DBU is the DB Underpin, EEDBU(s) v_2 is the Early Exercise DB underpin with stochastic salaries, and EEDBU(d) v_3 is the Early Exercise DB underpin with deterministic salaries.

Time to Retirement T	DB	DC	FSE	DBU	EEDBU(d) v_1
10yr	2.2675	1.2500	0	0.0039 (0.0011)	0.0099 (0.0001)
15yr	3.4012	1.8750	0	0.0210 (0.0020)	0.0456 (0.0003)
20yr	4.5349	2.5000	0.0304	0.0458 (0.0029)	0.1190 (0.0006)
30yr	6.8024	3.7500	0.2476	0.1455 (0.0048)	0.3752 (0.0014)
40yr	9.0699	5.0000	0.6280	0.3115 (0.0069)	0.7726 (0.0025)

Table 2.2: Cost of each pension plan, discrete setting. Hybrid costs are additional to the basic DB cost. FSE is the Florida second election option; DBU is the DB Underpin, and EEDBU(d) v_1 is the Early Exercise DB underpin (with deterministic salaries).

Some observations can be made:

- EEDBU v_2 is greater than EEDBU v_3 which reflects the additional costs from stochastic salaries, but the results are fairly close. We would expect to observe a similar result in discrete-time case, such that a stochastic salary would lead to a higher but close cost as in the deterministic salary assumption.

- When the expected retirement date is near, under the benchmark parameters, participants in the Second Election plan should always be in the DB plan ($t^* = 0$), so there will be no extra cost required to fund the second election option.

- Although the Early Exercise DB underpin option is greater than both the DB underpin and the Second Election, none of the values in the benchmark scenario exceed 10% of the cost of a DB plan. For horizons of 30 years or less, the extra cost from the EEDBU option costs around 5% more than the basic DB plan.

- As T increases, the cost is increasing, at an increasing rate, for all three options. The underpin and election options become significantly more costly over 40 years compared with the cost for 30 years.

- In the discrete-time case, the values of the options are generally greater than in the continuous case, and we would expect a larger difference if stochastic salary is incorporated. However, the observations made in the continuous case also apply in the discrete case.

2.7.3 Sensitivity Tests

In this section, we present the sensitivity tests over all parameters in the continuous setting. We consider c, μ_L, r, σ_S, γ and b under the deterministic salary assumption and σ_L and ρ under the stochastic salary assumption. We also fix the time horizon to be 30 years. Details are displayed in Table 2.3.

We note the following points.

- All three plans share the same directional trends as the parameters change, but with some very different sensitivities.

- The additional costs of the hybrid plans react in the opposite direction to the underlying DB costs, for parameters which influence the DB cost. For example, increasing the risk-free rate r decreases the DB cost, but increases the additional hybrid costs. Increasing the accrual rate b increases the DB costs, but decreases the additional hybrid costs. Hence, the sensitivity of the total costs to the changing parameters is rather more muted than the sensitivity of the additional costs shown in the table.

Factor	Sensitivity Tests								
r	0	0.01	0.02	0.03	**0.04**	0.05	0.06	0.07	0.08
DB	23.5064	17.414	12.9006	9.557	7.08	5.245	3.8856	2.8785	2.1325
v^{se}	0	0	0	0.0598	0.2179	0.4045	0.5786	0.7213	0.8276
v_3	0.0174	0.0437	0.1018	0.2023	0.3355	0.451	0.6202	0.7398	0.8327
v_U	0.0093	0.0198	0.039	0.0711	0.1199	0.1878	0.2741	0.374	0.4797
μ_L	0	0.01	0.02	0.03	**0.04**	0.05	0.06	0.07	0.08
DB	2.1325	2.8785	3.8856	5.245	7.08	9.557	12.9006	17.414	23.5064
v^{se}	0.3448	0.2987	0.265	0.2389	0.2179	0.2005	0.1858	0.1732	0.1623
v_3	0.5299	0.473	0.4217	0.3758	0.3355	0.3004	0.2703	0.2446	0.2229
v_U	0.4797	0.374	0.2741	0.1878	0.1199	0.0711	0.039	0.0198	0.0093
c	0.085	0.095	0.105	0.115	**0.125**	0.135	0.145	0.155	0.165
v^{se}	0.0163	0.0466	0.0909	0.1484	0.2179	0.2987	0.3902	0.4917	0.6026
v_3	0.0759	0.1235	0.1831	0.2539	0.3355	0.4271	0.5282	0.6383	0.757
v_U	0.0183	0.0325	0.0533	0.0821	0.1199	0.1679	0.2269	0.2975	0.3801
σ_S	0.07	0.09	0.11	0.13	**0.15**	0.17	0.19	0.21	0.23
v_3	0.2205	0.2314	0.2542	0.2893	0.3355	0.391	0.4538	0.5225	0.5954
v_U	0.0012	0.0094	0.0311	0.0685	0.1199	0.183	0.2552	0.3341	0.418
b	0.012	0.013	0.014	0.015	**0.016**	0.017	0.018	0.019	0.02
DB	5.31	5.7525	6.195	6.6375	7.08	7.5225	7.965	8.4075	8.85
v^{se}	0.4665	0.3896	0.3235	0.2667	0.2179	0.17	0.1401	0.1096	0.0838
v_3	0.5826	0.5075	0.4422	0.3853	0.3355	0.2918	0.2535	0.2199	0.1904
v_U	0.2958	0.2342	0.1864	0.1491	0.1199	0.0969	0.0788	0.0643	0.0527
γ	0	0.01	0.02	0.03	**0.04**	0.05	0.06	0.07	0.08
v^{se}	0	0	0	0.067	0.2179	0.3792	0.5281	0.6591	0.7728
v_3	0.1315	0.1492	0.1835	0.2442	0.3355	0.4513	0.5811	0.715	0.8463

Table 2.3: Sensitivity tests over (c, μ_L, r, σ_S, γ and b), using deterministic salaries. v^{se} is the additional FSE cost, v_3 is the additional EEDBU cost, and v_U is the additional DB underpin cost.

Value	Sensitivity Tests								
σ_L	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
v_3	0.3363	0.3389	0.3432	0.3492	0.357	0.3665	0.3778	0.391	0.4058
v_U	0.1209	0.1238	0.1286	0.1354	0.1443	0.1551	0.1681	0.183	0.2001
ρ	-1	-0.9	-0.5	-0.1	0	0.1	0.5	0.9	1
v_3	0.4538	0.4434	0.4015	0.3596	0.3492	0.3389	0.298	0.2623	0.2542
v_U	0.2552	0.2432	0.195	0.1472	0.1354	0.1238	0.079	0.0396	0.0311

Table 2.4: Sensitivity tests over (σ_L and ρ), stochastic salaries.

- The risk-free rate has a significant impact on the Early Exercise DB underpin option value, especially on the relative cost with respect to the DB plan. However, as shown by the sensitivity test on γ, the impact mostly comes from the value of the ABO. Also, it is interesting to note that when r is high, the cost of the FSE option is quite close to the Early Exercise DB underpin option.

- Decreases in the accrual rate b for the DB plan increase the relative value of the Early Exercise DB option. The extra cost reflects the fact that the funding of the DC plan is higher than the DB plan. When b is high, the fast accumulation of the DB benefit would discourage employees from entering into the DC account, and the option value will be reduced.

- In our model, the cost of the second election option is independent of the market volatility σ_S.

Although the structure of the Early Exercise DB underpin plan provides more flexibility and protection to the employee than both the second election option and the DB underpin plan, the additional cost does not appear as large as one may expect. In most scenarios, the cost is less than 5% of the DB plan. The relative cost is high when the risk-free rate increases, however, it is interesting to note that the cost is very close to the second election option. Similarly, when the salary growth rate is low, the cost is quite close to that of the DB underpin option. This demonstrates that a very large cost comes either from the option which allows employees to switch to DC, or from the guarantee.

Table 2.4 displays the sensitivity test for the stochastic salary parameters. The risks involved in the stochastic salary process have less impact overall. Larger volatility in the salary process, as well as negative correlation between the salary and equity market, will increase the volatility of the wealth-salary ratio process and thus increase the option value. These two risks have same effect on the DB underpin option, which has been previously observed by Chen and Hardy (2009). However, under our assumptions, the value of the second election option is immunized to these two risks.

2.8 DB trade-in with DC

One weakness in the design of the Early Exercise type DB underpin plan is that the employer is indeed bearing more risks than a traditional DB plan. Here we illustrate an example of how to modify the Early Exercise DB underpin plan to retain the employees' flexibility in choosing between DB and DC plans, while restraining the sponsor's cost. Instead of allowing the employees to retain any excess amount in the DC plan at the time of switching from DC to DB, the new plan will force the employee to forfeit all his/her DC account balance. Mathematically, the cost function at time t for this new plan at time t is (by ignoring the employee contribution):

$$\sup_{\tau\geq 0} E_t^Q \left[\underbrace{\int_t^{(t+\tau)\wedge T} e^{-r(u-t)} c L_u du}_{\text{PV of Employer Contribution to DC account}} + \underbrace{\left(K_T e^{-r(T-t)} - e^{-r\tau} W_{\tau+t}\right) \mathbf{1}_{\tau\leq T-t}}_{\text{Cost of DB plan less DC balance}} \right]$$

Clearly, if the employee has a DC balance higher than their accrued benefit obligation under the DB plan, there is no incentive for the employee to make the change. Here we define $C^{trade}(t, w, \tau)$ as the cost of the DB trade-in DC option given the stopping time τ:

$$C^{trade}(t, w, \tau) = E_t^Q \left[\int_0^{\tau\wedge(T-t)} e^{-ru} c L_{t+u} du + \left(K_T e^{-r(T-t)} - W_{t+\tau}^{t,w} e^{-r\tau}\right) \mathbf{1}_{\tau<T-t} \right.$$
$$\left. + \left(K_T e^{-r(T-t)} - W_T^{t,w} e^{-r(T-t)}\right)_+ \mathbf{1}_{\tau=T-t} \right]$$

Denote τ_T as the "waiting untill the end" strategy such that the employee only exercises the option on the retirement date if his/her DC account balance is lower than the fair value of the DB plan, otherwise they will retire with their DC plan. Then, for any admissible optimal stopping time τ, we have

$$
\begin{aligned}
& C^{trade}(t,w,\tau_T) - C^{trade}(t,x,\tau) \\
=& E_t^Q\Bigg[\int_0^{T-t} e^{-ru} cL_{t+u} du + \left(K_T e^{-r(T-t)} - W_T^{t,w} e^{-r(T-t)}\right)_+ \\
& \qquad - \int_0^{\tau\wedge(T-t)} e^{-ru} cL_{t+u} du - \left(K_T e^{-r(T-t)} - W_{t+\tau}^{t,w} e^{-r\tau}\right) \mathbf{1}_{\tau<T-t} \\
& \qquad - \left(K_T e^{-r(T-t)} - W_T^{t,w} e^{-r(T-t)}\right)_+ \mathbf{1}_{\tau=T-t}\Bigg] \\
=& E_t^Q\left[\left(\int_\tau^{T-t} e^{-ru} cL_{t+u} du + \left(K_T e^{-r(T-t)} - W_T^{t,w} e^{-r(T-t)}\right)_+ - \left(K_T e^{-r(T-t)} - W_{t+\tau}^{t,w} e^{-r\tau}\right)\right)\mathbf{1}_{\tau<(T-t)}\right] \\
\geq& E_t^Q\left[\left(\int_\tau^{T-t} e^{-ru} cL_{t+u} du + W_{t+\tau}^{t,w} e^{-r\tau} - W_T^{t,w} e^{-r(T-t)}\right)\mathbf{1}_{\tau<(T-t)}\right] \\
=& E_t^Q\left[\left(\int_\tau^{T-t} e^{-ru} cL_{t+u} du + \int_\tau^{T-t} d\left(e^{-ru} W_{t+u}^{t,w} du\right)\right)\mathbf{1}_{\tau<(T-t)}\right] \\
=& E_t^Q\left[\left(\int_\tau^{T-t} e^{-ru} W_{t+u}^{t,w} \sigma_S dZ_S^Q(t+u)\right)\mathbf{1}_{\tau<(T-t)}\right] \\
=& 0
\end{aligned}
$$

Therefore, the cost of this new plan is equivalent to the cost of sponsoring a DB underpin plan (discrete case can be derived similarly). In practice, employees may choose to switch to a DB plan with cost higher than its "fair value"; possible reasons include their risk-averse behavior, the difficulty to obtain a "fairly" priced annuity in the insurance market, asymmetric information, etc. The new plan discussed in this section is another natural extension of the second election option and the DB underpin plan, and we hope our discussion will inspire more hybrid designs for the pension market.

2.9 Conclusion

In this chapter, we combine Florida's second election option and the DB underpin option to form an Early Exercise type DB underpin plan. We summarize some key characteristics of the option, such as convexity and monotonicity. Also, we provide illustrations of the behavior of the early exercise region, and specifically include the situation where the Early Exercise DB underpin simplifies to the DB underpin plan.

Our numerical illustrations demonstrate that, although the Early Exercise DB underpin option may end up costing more than both the DB underpin option and the second election option combined, it does not cost more than 10% of the DB plan in general. In cases when the relative cost of the option compared with the DB plan is very large, for example when the risk-free rate is high or salary growth rate is low, the actual cost of the Early Exercise DB underpin plan is indeed smaller.

The Early Exercise DB underpin plan shifts more risk and cost to the employer compared with a DB plan, so it may not be an attractive option for pension sponsors, but the overall costs can be managed to some extent by varying the DC contribution rate and the DB accrual rate. Furthermore, it offers an attractive portable benefit for younger employees, which should help with recruitment, and offers a substantial retention benefit for older employees. It offers predictable income in retirement, popular with employees and labour unions. More importantly, we illustrate the simplicity in constructing new hybrid designs that are naturally linked to existing pension structures.

Chapter 3

An Intergenerational Risk Sharing Plan

3.1 Introduction

Most existing hybrid pension plans were designed to re-allocate risk between employers and employees, for example, the Cash Balance plan, the DB underpin plan, and the second election option discussed in Chapters 1 and 2. These hybrids utilize financial option-like benefits which create excessive risk and volatility to employers when they go unhedged. For example, the Cash Balance plan carries significant risk to employers during the accumulation phase, and leaves significant decumulation risks to the members, which does not really address the problems of DB or DC plans.

Recently, there has been growing interest in collective pension designs, often known as Intergenerational Risk Sharing (IRS) plans, where risk is distributed across different generations. In comparison to DB and DC plans, IRS plans emphasize sustainability and smoothness of consumption, meaning that both the contribution volatility and the benefit volatility are reduced.

The development of IRS plans remains in the discussion stage, with only a few practical examples, including some types of Target Benefit plan in Canada and the Collective DC example from the Netherlands. However, numerous IRS plans have been proposed in the literature over the past decade. Goecke (2013) and Bams *et al.* (2016) study a plan where each participant has an individual notional account with crediting rates based on the asset/liability ratio of the entire pension fund. The members continue to participate in the plan after the retirement date, with income varying according to the account balance at the beginning of each period. Cui *et al.* (2011), Khorasanee (2012) and **?** consider a pension plan where both the contribution and the retirement benefit are adjusted by the asset level of the pension fund. Gollier (2008) and Wang *et al.* (2018) formulate an optimal control problem that jointly solves for the optimal portfolio strategy and risk-sharing strategy for a target benefit plan. Boes and Siegmann (2018) suggest setting up an additional insurance fund to smooth retirement benefits. Bovenberg and Mehlkopf (2014) consider non-overlapping generations where a linear function of investment risk is transferred to the next generation. Most of the aforementioned papers demonstrate that introducing IRS features enhances the aggregate utility, which is termed the welfare function.

One attractive feature of the IRS plan is its ability to smooth the effect of investment and longevity risks. In a DC plan, an individual account balance is highly volatile, based on investment experience, and the retirement income is extremely sensitive to the interest rate at the annuitization date (we assume annuitization). On the other hand, when a DB plan remains under-funded for years, its employees bear the risk of sudden changes in the plan structure, either through a reduction in their accrued benefit or as a substantial rise in their contribution, depending on the bargaining power of active workers and retirees[1]. In contrast, a well-structured IRS plan can benefit from the collective risk-sharing mechanism to provide members with more stable lifetime income. In our model, we determine the optimal IRS design for all current and future employees by minimizing the average squared and linear deviation between the actual consumption and a pre-set target. Similar objectives have been considered in other pension-related literature. Gerrard *et al.* (2004) study the optimal investment strategies in a DC plan during the decumulation period, assuming that retirees are making constant withdrawals. Gerrard *et al.* (2006) incorporate the bequest motive, and add withdrawn income as an additional control variable. Wang *et al.* (2018) adopt the framework to an IRS plan with contribution fixed. They are able to derive explicit solutions for both the optimal investment strategy as well as the optimal risk sharing structure.

Our main contribution is to address the issue of intergenerational fairness, of which the discussion is scarce in the current IRS literature. Specifically, we illustrate that for the unconstrained optimal control approach in the study of IRS plans (ex. Cui *et al.* (2011) and Wang *et al.* (2018)) may sacrifice the interests of nearby cohorts for the benefit of future generations. We provide explanations for the abnormal investment strategies observed in the previous literature. In addition, we demonstrate how introducing a regulatory constraint would enhance the fairness across generations. Most of our analysis resides on a stylized IRS plan which is similar to Cui *et al.* (2011), where the structure is transparent

[1] Assuming employees pay all the contribution, implicitly or explicitly.

and easy to understand. We illustrate the appropriateness of such a design by showing its connection with the optimal IRS structure derived from stochastic control problems.

The chapter is organized as follows. In Section 3.2 we introduce the notation and models used in the subsequent sections. In Section 3.3 we derive the optimal IRS structure using two different measures and demonstrate the drawbacks of the unconstrained framework. In Section 3.4 we reformulate the problem and provide explicit solutions. In Section 3.5 we incorporate a regulatory requirement in the constrained optimization and discuss its effect on the pension design. Section 3.6 concludes.

3.2 Model Setup

We adopt a multi-period overlapping generations model for the population structure. All new employees are assumed to live for $N = 60$ years, during which they work for $R = 40$ years and retire for $N - R = 20$ years. We denote by Π the fraction of retirees in the population. As a benchmark, we standardize the unit so that there is exactly one person at each age, and the number of living generations is 60 with $\Pi = 1/3$. We do not consider demographic changes, but one may easily extend Π to be a time-dependent process, Π_t in which case there will be different numbers in each cohort.

The IRS plan discussed here can be regarded as a DB plan with risk-sharing features based on the investment performance. Before introducing the IRS structure, we consider a benchmark de-risked DB plan which is fully funded through fixed employee contributions of p per year, paid continuously, invested in risk-free assets. In this simplified set up there is no remaining asset or demographic risk. Given p, the annual retirement benefit for each retiree, denoted b, can be calculated based on the equivalence principle of present values:

$$p \times \bar{a}_{\overline{R}|} = b \times {}_{R|}\bar{a}_{\overline{N-R}|} \implies \int_0^R e^{-rs} p ds = \int_R^N e^{-rs} b ds$$

In other words, an individual's total contribution is actuarially equivalent to his/her retirement benefit. Here we ignore the inflation risk by normalizing the employee's income to be 1 in real terms; the same assumption is made in Cui *et al.* (2011).

The liability value for this de-risked DB plan is

$$
\begin{aligned}
L &= \int_0^R \left(\underbrace{\int_R^N \frac{N(1-\Pi)}{R} e^{-r(s-x)} b ds}_{\text{Future Benefit for Active Worker}} - \underbrace{\int_x^R \frac{N(1-\Pi)}{R} e^{-r(s-x)} p ds}_{\text{Future Contribution for Active Worker}} \right) dx \\
&\quad + \int_R^N \underbrace{\frac{N\Pi}{N-R} \left(\int_x^N e^{-r(s-x)} b ds \right)}_{\text{Benefit for Retiree}} dx \\
&= \frac{b(N-R) - pR}{r}
\end{aligned}
$$

In reality, although plan sponsors are increasing their holdings in fixed-income securities, a complete de-risking strategy is rare, and investment risk remains the main source of income or contribution volatility. The overall allocation of funds in equities, real estate, hedge funds and other equity-type investments typically represents more than 50% of pension plan assets[2]. Intergenerational risk sharing plans allow the investment risks to be smoothed over time, and across generations, by simultaneously adjusting benefits and contributions in response to variable investment returns.

Let p and b denote target values for contributions and benefits respectively. These are used to determine the liability, L, which is constant over time as the population is stationary, and as we are assuming a fixed rate of interest for the valuation. Let X_t denote the value of the plan assets at t, and p_t and b_t denote the contribution and benefit rates at t, respectively, which are adjusted to reflect the risk sharing resulting from the uncertainty in the asset values.

We define the individual consumption level c_t as the amount of income for each individual at t, after deducting contributions for workers (taxes are not considered in our setting). We have normalized the setting such that workers' earnings are 1 per year (in real terms), so the consumable income is $c_t = 1 - p_t$ for active workers and $c_t = b_t$ for retirees.

We assume the assets, X_t are invested in a mixture of a bank account with risk-free return, and an equity following a Geometric Brownian Motion with drift μ and volatility σ. The stochastic differential equation (SDE) for X_t is

$$
\begin{aligned}
dX_t &= \left(X_t \left(r + \omega_t(\mu - r)\right) + N(1-\Pi)p_t - N\Pi b_t\right) dt + \omega_t \sigma X_t dZ_t, \qquad X_0 = x_0 \\
&= \mu_X(X_t, \omega_t, p_t, b_t) dt + \sigma_X(X_t, \omega_t) dZ_t
\end{aligned} \tag{3.1}
$$

[2] Willis Towers Watson, 2017. "2016 asset allocations in Fortune 1000 pension plans". Available at https://www.towerswatson.com/

where ω_t is the weight of the portfolio invested in equity and Z_t is a standard Brownian Motion. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the complete probability space, where $\mathcal{F} = \{\mathcal{F}_t, t \geq 0\}$ is the $\mathbb{P}$-augmented filtration generated by $\{Z_t, t \geq 0\}$.

Notice that the forms of ω_t, p_t, and b_t are unknown at this point. We define the set of admissible strategies $\pi = \{(\omega_u, p_u, b_u)\}_{u \in [t,\infty)}$ as follows:

Definition 1. *(Admissible Strategy) For any fixed $t \in [0, \infty)$, a strategy*

$$\pi = \{(\omega_u, p_u, b_u)\}_{u \in [t,\infty)}$$

is said to be admissible if

1. *π is $\mathcal{F}_t$-adapted.*
2. *$E[\int_0^T \mu_X(x, \omega_u, p_u, b_u)^2 + \sigma_X(x, \omega_u)^2 du] < +\infty$, $\forall T > 0$, $\forall x \in \mathbb{R}$*
3. *(X^π, π) is the unique solution to (3.1).*

and we denote $\mathcal{A}$ as the set of all admissible strategies π.

Later in this chapter, when we prespecify the risk-sharing structure (the form of p_t and b_t), we can easily modify the above definition for the strategy $\pi = \{\omega_u\}_{u \in [t,\infty)}$ and the corresponding admissible set $\mathcal{A}$.

A simple and intuitive example of an IRS plan is given in Cui *et al.* (2011). In this plan, a portion of the surplus or deficit in the plan (represented by $X_t - L$) is distributed to the current employees as an adjustment to the pension contribution, and/or to the retirees, as an adjustment to the benefit. In the notation defined above, this can be described as follows:

$$\begin{aligned} p_t &= p - \alpha \frac{X_t - L}{\#\text{ of Active Workers}} = p - \alpha \frac{X_t - L}{N(1 - \Pi)} \\ b_t &= b + \beta \frac{X_t - L}{\#\text{ of Retirees}} = b + \beta \frac{X_t - L}{N\Pi} \end{aligned} \tag{3.2}$$

This means that a portion $(\alpha + \beta)$ of the total surplus or deficit $(X_t - L)$ is distributed to participants through reducing/increasing contributions for active workers, and increasing/reducing the benefit for retirees. We will refer back to this plan design in the following sections, as we derive optimal forms for p_t and b_t under a range of assumptions and constraints.

3.3 Optimal Pension Design Without Constraints

In this section we formulate the optimal control problems to derive optimal forms of p_t and b_t, together with the optimal investment strategy ω_t. We consider two objective functions. The first targets overall expected utility; the second targets consumption smoothness. We provide the explicit solutions to both cases, and show that neither case gives satisfactory results without further refinement.

3.3.1 Optimal Control - Welfare Function

The ability to trade between generations can be welfare improving, as demonstrated by, for example, Merton (1983), Gordon and Varian (1988), and Diamond (1965). This idea has been implemented through the social security (or government benefit for seniors) in many countries, but has not been widely applied in the occupational pension context. Here, we follow the standard interpretation of welfare improving from the economics literature, that is, maximizing the utility of the consumption paths, through work and retirement phases, using a Constant Relative Risk Aversion (CRRA) utility function, $U(t,c) = e^{-\delta t}\frac{c^{1-\gamma}}{1-\gamma}$ where $\delta \geq 0$ is the discount rate and $\gamma > 1$ is the risk aversion parameter[3]. For mathematical convenience, we do not differentiate the risk averasion level for active workers and for retirees, thus, γ is interpreted as the average risk aversion level throughout an individual's lifetime.

The goal is to maximize the aggregate utility for all generations, taking both the investment strategy and the risk-sharing structure as control variables. That is, we find

$$V^P(X_0) = \sup_{\{\omega_t, p_t, b_t\}\in\mathcal{A}} E\left[\int_0^\infty e^{-\delta t}\left(N(1-\Pi)\frac{(1-p_t)^{1-\gamma}}{1-\gamma} + N\Pi\frac{b_t^{1-\gamma}}{1-\gamma}\right)dt\right] \tag{3.3}$$

Proposition 6. *For the optimal control problem (3.3), the optimal asset allocation policy*

[3] Zhang *et al.* (2018) estimate $e^{-\delta}$ has mean of 0.965 and median of 0.997, and γ to be between 1.74 and 3.74, based on Ontario residents between age 50 and 80.

and risk sharing structure are given by

$$\begin{aligned}
\hat{\omega}_t &= \frac{\mu - r}{\sigma^2 \gamma} \left(\frac{N(1-\Pi)}{r X_t} + 1 \right), \\
\hat{p}_t &= p - \alpha^P \left(\frac{X_t - \psi_w^P L}{N(1-\Pi)} \right), \\
\hat{b}_t &= b + \beta^P \left(\frac{X_t - \psi_r^P L}{N\Pi} \right),
\end{aligned} \tag{3.4}$$

and the corresponding value function is given by

$$V^P(x) = \frac{(A + Bx)^{1-\gamma}}{1-\gamma},$$

where the expressions A, B, ψ_w^P, ψ_r^P, α^P and β^P are given below:

$$\begin{aligned}
A &= \frac{N(1-\Pi)B}{r}, \quad B = \left(\frac{1-\gamma}{N\gamma} \left(\frac{\delta}{1-\gamma} - \frac{(\mu - r)^2}{2\sigma^2\gamma} - r \right) \right)^{\frac{\gamma}{\gamma - 1}}, \\
\psi_w^P &= \frac{N}{L} \left(\frac{1-p}{\xi} - \frac{(1-\Pi)}{r} \right), \qquad \psi_r^P = \frac{N}{L} \left(\frac{b}{\xi} - \frac{(1-\Pi)}{r} \right), \\
\alpha^P &= (1-\Pi)\xi, \qquad \beta^P = \Pi\,\xi, \\
\xi &= \frac{\gamma - 1}{\gamma} \left(\frac{1}{2\gamma} \left(\frac{\mu - r}{\sigma} \right)^2 + r - \frac{\delta}{1-\gamma} \right).
\end{aligned}$$

Proof. The solution can be easily obtained through standard Hamilton-Jacobian-Bellman (HJB) method similar to Merton (1969). First, we are trying to find a function $v(x) \in C^2$ such that

$$\begin{aligned}
\sup_{\{\omega_t, p_t, b_t\} \in \mathcal{A}} \Bigg\{ & [x(r + \omega_t(\mu - r)) + N(1-\Pi)p_t - N\Pi b_t] \frac{\partial v(x)}{\partial x} + \frac{\sigma^2 \omega_t^2 x^2}{2} \frac{\partial^2 v(x)}{\partial x^2} \\
& + N(1-\Pi) \frac{(1-p_t)^{1-\gamma}}{1-\gamma} + N\Pi \frac{b_t^{1-\gamma}}{1-\gamma} \Bigg\} = \delta v(x).
\end{aligned}$$

Then, by the conjecture that $v(x) = \frac{(A+Bx)^{1-\gamma}}{1-\gamma}$ (which implies that $v'' < 0$), we substitute the expression into the HJB equation to obtain the explicit solution. Notice that the expression for $\hat{b}_t$ can be rearranged:

$$\begin{aligned}\hat{b}_t &= B^{-1/\gamma}(A + BX_t) \\ &= \underbrace{B^{-1/\gamma}A + B^{1-1/\gamma}\psi_r^P L}_{\text{set it to be } b \text{ and derive } \psi_r^P} + N\Pi B^{1-1/\gamma}\left(\frac{X_t - \psi_r^P L}{N\Pi}\right).\end{aligned}$$

Then $\hat{p}_t$ can be derived based on the relation between b, L and p. Lastly, by a standard verification argument (for example, see Touzi (2013)), we can conclude that $V^P = v$. □

Remark. *The optimal solution is close to the IRS plan proposed in Cui et al. (2011) (Equation 3.2), with the difference being that the risk sharing is not based on the full surplus $X_t - L$, but on an adjusted surplus, $X_t - \psi_w^P L$ for the workers' contributions and $X_t - \psi_r^P L$ for the retirees' benefits.*

Here we see that ψ_r^P and ψ_w^P are implicitly determined when we fix the target contribution (p). Alternatively, we may choose to specify ψ_w^P and implicitly derive values for p, b and ψ_r^P. The values of α^P and β^P are independent of the choice of p, b, ψ_w^P and ψ_r^P, and are guaranteed to be positive. It is interesting to observe that when the retirees have a lower benchmark consumption level than the active employees ($1 - p > b$), they enjoy a prior claim on the surplus, and have a lower threshold for deficit contributions (since $\psi_w^P > \psi_r^P$). The risk sharing parameters α and β depend on the population structure, the optimal design provide more consumption smoothing to larger population. Moreover, a higher discount rate, which implies more preference of current consumption over the future, results in less risk-sharing incentives.

The mathematical form of the optimal investment strategy also has a clear interpretation. As individuals become more risk-averse, that is, when γ is increased, the proportion in the risky investment will be reduced. In addition, the amount of risky investment decreases as the overall asset level increases; at the extreme case, $\lim_{X_t \to \infty} \hat{\omega}_t = \frac{\mu - r}{\sigma^2 \gamma}$. Also, we notice that when the working population increases ($N(1 - \Pi)$ increases), the amount of equity investment increases. This result aligns with the work by Teulings and De Vries (2006), where the authors consider the optimal investment strategy for a pension fund containing a group of individual Generational Accounts[4].

However, the results from this exercise do not make sense in a practical context. The values for ψ_w^P and ψ_r^P may be negative, for reasonable values of p and other parameters,

[4] In a Generational Account (GA) scheme, investment decisions are made collectively, but there is no risk-sharing feature.

implying that the contributions are always below the benchmark, even if the plan is underfunded, and benefits are always above the benchmark. Alternatively, again for quite reasonable assumed parameters, the values of ψ_w^P and ψ_r^P may be very large, indicating excessive contributions and/or benefits. The SDE of the asset process under the optimal investment strategy and the optimal pension design is:

$$\begin{aligned}
dX_t = &\left\{ X_t \left(r - \alpha^P - \beta^P + \left(\frac{\mu - r}{\sigma}\right)^2 \frac{1}{\gamma} \right) \right. \\
&\left. + \left(\frac{\mu - r}{\sigma}\right)^2 \frac{1}{\gamma} \frac{N(1-\Pi)}{r} + N(1-\Pi)(p+b) - Nb + (\alpha^P \psi_w^P + \beta^P \psi_r^P) L \right\} dt \\
&+ \frac{\mu - r}{\sigma\gamma} \left(\frac{N(1-\Pi)}{r} + X_t \right) dZ_t \\
= &(b_1 X_t + b_2) dt + (c_1 X_t + c_2) dZ_t.
\end{aligned}$$

where the future expected asset value is

$$E[X_t] = (X_0)e^{b_1 t} + \frac{b_2}{b_1}(e^{b_1 t} - 1).$$

Remark. *If $\delta > \left(\frac{\mu-r}{\sigma}\right)^2 \frac{\gamma+1}{2\gamma} + r$, then the long term asset value will converge to $\frac{N(1-\Pi)}{r}$. However, this condition is often violated, and the long term asset value diverges to $+\infty$.*

3.3.2 Optimal Control - Stability of Consumption

The purpose of intergenerational risk sharing is to spread excess profits from the windfall cohorts to support the benefits and contributions of the less fortunate cohorts, so a successful IRS plan should help to limit the variability of both contributions and benefits across generations. In this section we assume that the sponsor will set a target consumption level for all individuals, and the goal is to minimize the squared deviation from the actual consumption and the target consumption. This approach is similar to Wang *et al.* (2018), and is more consistent with the purpose of the IRS than the utility approach. Further discussion on the advantage of this stability measure will be presented in the next section.

For simplicity, we denote c^w and c^r as the target consumption levels of the active workers and the retirees respectively; both are assumed to be constant. Then, we can formulate an ergodic control problem:

$$v^S = \inf_{\{\omega_t, p_t, b_t\} \in \mathcal{A}} \lim_{T\to\infty} \frac{1}{T} E\left[\int_0^T N(1-\Pi)\left\{((1-p_t)-c^w)^2 - \varrho_w((1-p_t)-c^w)\right\}\right.$$
$$\left. + N\Pi\left\{(b_t - c^r)^2 - \varrho_r(b_t - c^r)\right\} dt\right], \tag{3.5}$$

where ϱ_r and ϱ_w are penalty weights for the negative deviation between the actual and the target consumption. Here we want to mention that without taking inflation into the consideration, our objective function also aligns with the economic concept of habit formation preference. Individuals are used to a certain living standard and any deviation from the target consumption would increase his/her disutility. Depending on the individual risk preference (ϱ_r and ϱ_w), one may willing to take some risks on their consumption level for the opportunity of better quality of living. Some discussions on the relationship between quadratic loss function and habit formation in preference can be found in Bruhn and Steffensen (2013).

Proposition 7. *For the ergodic optimal control problem (3.5), if* $2r - \left(\frac{\mu-r}{\sigma}\right)^2 > 0$, *then the optimal asset allocation and risk sharing structure are given by:*

$$\hat{\omega}_t = \frac{\mu - r}{\sigma^2}\left(-1 + \frac{N\Pi(c^r + \frac{\varrho_r}{2}) - N(1-\Pi)(1 - c^w - \frac{\varrho_w}{2})}{rX_t}\right),$$
$$\hat{p}_t = p - \alpha^S\left(\frac{X_t - \psi_w^S L}{N(1-\Pi)}\right), \qquad \hat{b}_t = b + \beta^S\left(\frac{X_t - \psi_r^S L}{N\Pi}\right), \tag{3.6}$$

and the value function is

$$v^S = \left(N(1-\Pi)(1 - c^w - \frac{\varrho_w}{2}) - N\Pi(c^r + \frac{\varrho_r}{2}) - \frac{NB}{4} - \frac{(\mu-r)^2 B}{4\sigma^2 A}\right) B - N(1-\Pi)\frac{\varrho_w^2}{4} - N\Pi\frac{\varrho_r^2}{4},$$

where

$$\beta^S = \Pi\left(2r - \frac{(\mu-r)^2}{\sigma^2}\right), \quad \alpha^S = (1-\Pi)\left(2r - \frac{(\mu-r)^2}{\sigma^2}\right),$$
$$\psi_w^S = \frac{N}{L}\left(\frac{1 - c^w - p - \frac{\varrho_w}{2}}{2r - \left(\frac{\mu-r}{\sigma}\right)^2} - \frac{(1-\Pi)(1 - c^w - \frac{\varrho_w}{2}) - \Pi(c^r + \frac{\varrho_r}{2})}{r}\right),$$
$$\psi_r^S = \frac{N}{L}\left(\frac{b - c^r - \frac{\varrho_r}{2}}{2r - \left(\frac{\mu-r}{\sigma}\right)^2} - \frac{(1-\Pi)(1 - c^w - \frac{\varrho_w}{2}) - \Pi(c^r + \frac{\varrho_r}{2})}{r}\right).$$

Proof. This ergodic control problem can be solved through standard HJB equations, where we are trying to find a pair $(J, v) \in C^{1,2} \times \mathbb{R}$ such that:

$$\inf_{\omega_t, p_t, b_t} \left\{ \left(x(r+\omega_t(\mu-r)) + N(1-\Pi)p_t - N\Pi b_t\right) J_x + \frac{\sigma^2\omega_t^2 x^2}{2} J_{xx} \right.$$
$$+N(1-\Pi)\big(((1-p_t)-c^w)^2 - \varrho_w((1-p_t)-c^w)\big)$$
$$\left. +N\Pi((b_t-c^r)^2 - \varrho_r(b_t-c^r)) \right\} = v.$$

By making the conjecture that $J(x) = Ax^2 + B$, we are able to obtain:

$$A = \frac{2r - \left(\frac{\mu-r}{\sigma}\right)^2}{N},$$
$$B = \frac{2A(N(1-\Pi)(1-c^w-\frac{\varrho_w}{2}) - N\Pi(c^r+\frac{\varrho_r}{2}))}{r}.$$

Notice that $J'' > 0$ if and only if $A > 0 \implies 2r - \left(\frac{\mu-r}{\sigma}\right)^2 > 0$. Lastly, by standard verification theorem (ex.Leizarowitz (1988), Leizarowitz (1990) and Stockbridge (1989)), we have $v^S = v$ whenever $2r > \left(\frac{\mu-r}{\sigma}\right)^2$. □

The optimal design shares exactly the same structure as in the welfare function case and a more detailed comparison will be given in the following section. Here we make a few highlights. First, as $X_t \to \infty$, $\omega_t = -\frac{(\mu-r)}{\sigma^2} < 0$, which means when the asset level is high, the sponsor will take a short position in the risky asset. At first glance, this result is highly counter-intuitive, but soon we will demonstrate that when X_t reaches a certain threshold, all future benefits will be locked and payable through risk-free investment. In addition, any increases in the target consumption level (c^r and c^w) will increase the risk-taking motive; α^S and β^S do not depend on the choice of the target consumption levels c^r and c^w, nor on the penalty parameters on negative returns ϱ_w and ϱ_r. Furthermore, when the penalty terms ϱ_r and ϱ_w increase, more weight is put into the risky investment. Lastly, we are not able to provide the explicit solution when $2r - \left(\frac{\mu-r}{\sigma}\right)^2 \leq 0$, but when $2r - \left(\frac{\mu-r}{\sigma}\right)^2 > 0$, the risk sharing parameters α^S and β^S are guaranteed to be positive.

The asset process under optimal pension design has the following SDE:

$$
\begin{aligned}
dX_t = &\left\{ X_t \left(r - \alpha^S - \beta^S - \left(\frac{\mu - r}{\sigma} \right)^2 \right) \right. \\
&\left. - \left(\frac{\mu - r}{\sigma} \right)^2 \frac{B}{2A} + N(1 - \Pi)(p + b) - Nb + (\alpha^S \psi_w^S + \beta^S \psi_r^S) L \right\} dt \\
&- \left(\frac{\mu - r}{\sigma} \right) \left(X_t + \frac{B}{2A} \right) dZ_t \\
= &(b_1 X_t + b_2) dt + (c_1 X_t + c_2) dZ_t.
\end{aligned}
$$

The future expected asset level is

$$E[X_t] = (X_0) e^{b_1 t} + \frac{b_2}{b_1}(e^{b_1 t} - 1).$$

Notice that $b_1 = -r < 0$ which implies

$$\lim_{t \to \infty} E[X_t] = \frac{-b_2}{b_1} < \infty, \quad \forall r > 0.$$

Remark. *When the asset level X_t reaches $\frac{1}{r}\left[N\Pi(c^r + \frac{\varrho_r}{2}) - N(1 - \Pi)(1 - c^w - \frac{\varrho_w}{2})\right]$, then $\omega_t = 0$ and $dX_t = 0$, which implies that contribution ($\hat{p}_t = 1 - c^w - \frac{\rho_w}{2}$) plus the risk free return of pension asset will be enough to cover the benefit payments ($\hat{b}_t = c^r + \frac{\rho_r}{2}$).*

Unlike the utility maximization approach, the asset level in the consumption stability approach will converge to a constant. However, there remain many practical issues with the result; the thresholds ψ_w^s and ψ_r^s may be less than zero or excessively large; the condition $2r - \frac{\mu - r}{\sigma}^2 > 0$ is too restrictive; and the optimal plan will eventually become a risk-free DB plan. We explore these issues in the next section.

3.3.3 Numerical Illustration

In this subsection, we conduct numerical experiments and compare the pros and cons of the two optimal pension designs derived in Sections 3.3.1 and 3.3.2.

Figures 3.1 and 3.2 present the optimal pension design with different risk-free rate r. For the power utility case, we also compare the optimal design under different levels

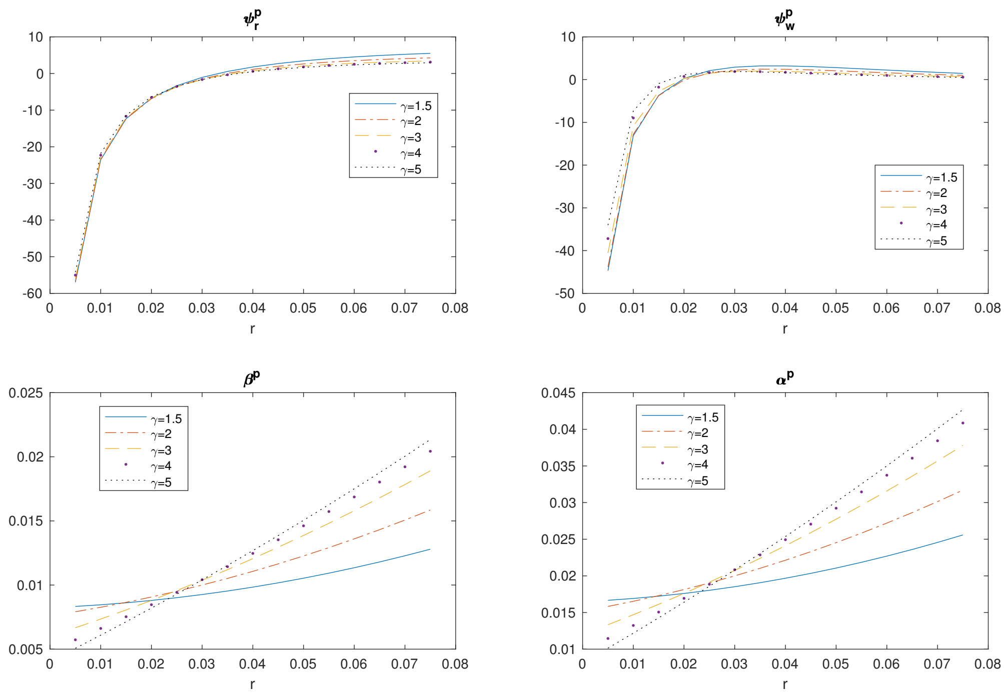


Figure 3.1: The Optimal Pension Design for Improving the Welfare Function

of risk aversion, with $\delta = 0.02$ such that the discount factor is roughly 0.98. For the target consumption case, we examine the effect of the penalty parameter of negative return (assuming $\varrho_r = \varrho_w$, c^r and c^w are fixed at 0.9) and the effect of target consumption (assuming $c^r = c^w$, ϱ_w and ϱ_r are fixed at 0) on ψ_w^S and ψ_r^S. Notice that we did not provide graphs for α^S and β^S as they are independent of ϱ_r, ϱ_r, c^r and c^w. For the benchmark case, we set $\mu = 0.08$, $\sigma = 0.25$ and $p = 0.1$.

Immediately we notice that ψ_r^P loses its financial interpretations when r is small (less than around 0.03), or when r is large. ψ_w^P suffers the same issue but only when r is small. On the other hand, the target consumption case suffers the same interpretation issues when r is close to $\frac{1}{2}\left(\frac{\mu - r}{\sigma}\right)^2$. When r is larger, both ψ_r^S and ψ_w^S appear much more reasonable. Adjusting the values of ϱ_r, ϱ_w, c^w and c^r does not solve this issue, and the optimal pension design is not robust with respect to these input parameters.

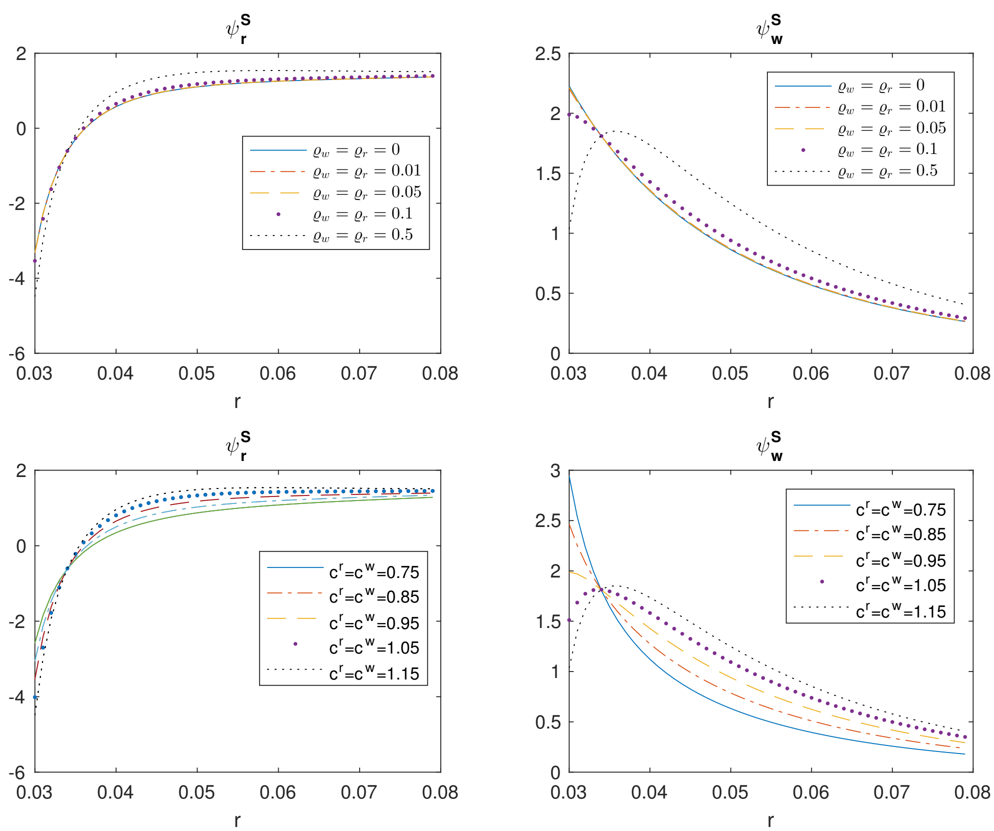


Figure 3.2: The Optimal Pension Design for Improving Consumption Stability

3.4 Optimal Pension Design - Linear Risk Sharing Structure

Despite the critical drawbacks in the unconstrained optimal control problem, it provides theoretical suggestions with respect to the optimality of the linear risk sharing design. Therefore, we now consider a stylized pension design that is similar to Cui *et al.* (2011), where p_t and b_t have the following forms:

$$p_t = p - \alpha \left(\frac{X_t - \psi_w L_t}{N(1-\Pi)} \right)$$
$$b_t = b + \beta \left(\frac{X_t - \psi_r L_t}{N\Pi} \right)$$

and where ψ_w and ψ_r are determined externally, for example, to align with regulatory requirements. Cui *et al.* (2011) have demonstrated that their IRS plan is welfare improving compared to either a DC or DB plan. The authors also calculate the optimal α, β and p, by maximizing the utility of the lifetime consumption of a single generation. Here we re-investigate the problem, and this section outlines the differences between our work and Cui *et al.* (2011).

3.4.1 Consumption Stability of a Single Generation

In this and the following sections we use the second objective function from Section 3.3, where we minimize the variability of individual consumption. The reason is that the utility function approach is unable to reflect the advantages of IRS plans in consumption smoothing. Pure DC or DB plans solely focus on the stability of income either during the pre-retirement or the post-retirement period, but not both (we assume the DB plan is a contribution adjustable plan, where deficit leads to higher contributions from active workers). The quadratic function that measures the squared difference between target and actual consumption would better fit this purpose.

To begin with, we consider the single generation model, by replacing the utility function in Cui *et al.* (2011) with a quadratic function, and defining the value function as:

$$V(t,x;p,\alpha,\beta) = \inf_{\omega_t \in \mathcal{A}_t} E\left[\int_{\min(t,R)}^{R} (c^w - (1-p_s))^2 - \varrho_w(1-p_s-c^w)ds \right.$$
$$\left. + \int_{\max(t,R)}^{N} (c^w - b_s)^2 - \varrho_r(b_s - c^r)ds \middle| \mathcal{F}_t \right], \tag{3.7}$$
$$\text{where } dX_t = (X_t(r+\omega_t(\mu - r)) + Rp_t - (N-R)b_t)\, dt + \omega_t \sigma X_t dZ_t.$$

The optimal pension design can be easily obtained by static optimization:

$$(\hat{p}, \hat{\alpha}, \hat{\beta}) = \arg\min_{p,\alpha,\beta} V(t,x;p,\alpha,\beta)$$

In this section, we will provide the explicit solution to (3.7). For clarity, we write the value function such that:

$$V(t,x;p,\alpha,\beta) = \inf_{\omega_t \in \mathcal{A}_t} E\left[\int_{\min(t,R)}^{R} (\rho_1 + \lambda_1 X_s)^2 ds + \nu_1 ds + \int_{\max(t,R)}^{N} (\rho_2 + \lambda_2 X_s)^2 ds + \nu_2 ds \middle| \mathcal{F}_t \right],$$
$$\rho_i = \begin{cases} c^w - 1 + p + \frac{\alpha\psi_w}{N(1-\Pi)}L + \frac{\varrho_w}{2}, & i=1 \\ c^r - b + \frac{\beta\psi_r}{N\Pi}L + \frac{\varrho_r}{2}, & i=2, \end{cases}$$
$$\lambda_i = \begin{cases} -\frac{\alpha}{N(1-\Pi)}, & i=1 \\ -\frac{\beta}{N\Pi}, & i=2, \end{cases} \quad \nu_i = \begin{cases} -\frac{\varrho_w^2}{4}, & i=1 \\ -\frac{\varrho_r^2}{4}, & i=2, \end{cases}$$
$$dX_t = [X_t(a_1 + \omega_t a_2) + a_3]dt + \omega_t \sigma X_t dZ_t,$$
$$a_1 = r - \alpha - \beta, \quad a_2 = (\mu - r),$$
$$a_3 = N(1-\Pi)(p+b) - Nb + (\alpha\psi_w + \beta\psi_r)L.$$

Proposition 8. *For the optimal control problem (3.7), the optimal asset allocation is*

$$\hat{\omega}_t^S = \begin{cases} \frac{-a_2(2A(t)X_t + B(t))}{2\sigma^2 X_t A(t)}, & t \in [R,N] \\ \frac{-a_2(2P(t)X_t + Q(t))}{2\sigma^2 X_t P(t)}, & t \in [0,R], \end{cases} \tag{3.8}$$

and the value function is

$$V(t,x) = \begin{cases} A(t)x^2 + B(t)x + C(t), & t \in [R,N] \\ P(t)x^2 + Q(t)x + W(t), & t \in [0,R], \end{cases}$$

where the expressions are given below:

$$A(t) = \begin{cases} A_1 \left(e^{A_2(N-t)} - 1\right), & 2a_1 - \frac{a_2^2}{\sigma^2} \neq 0 \\ \lambda_2^2(N-t), & 2a_1 - \frac{a_2^2}{\sigma^2} = 0, \end{cases}$$

$$B(t) = \begin{cases} B_4 \left(e^{B_2(N-t)} - 1\right) + \frac{-2\lambda_2^2 a_3}{B_2}(N-t), & 2a_1 - \frac{a_2^2}{\sigma^2} = 0 \\ \frac{2a_3 A_1}{A_2} e^{A_2(N-t)} + (2a_3 A_1 - 2\rho_2\lambda_2)(N-t), & a_1 - \frac{a_2^2}{\sigma^2} = 0 \\ B_1 \left(e^{B_2(N-t)} - 1\right) + 2a_3 A_1 (N-t) e^{B_2(N-t)}, & a_1 = 0 \\ B_1 \left(e^{B_2(N-t)} - 1\right) + B_3 \left(e^{B_2(N-t)} - e^{A_2(N-t)}\right), & o/w, \end{cases}$$

$$C(t) = \int_t^N \left(a_3 B(s) - \frac{a_2^2 B(s)^2}{4\sigma^2 A(s)} + \rho_2^2 + \nu_2\right) ds,$$

$$P(t) = \begin{cases} e^{A_2(R-t)} A(R) + \frac{\lambda_1^2}{A_2} \left(e^{A_2(R-t)} - 1\right), & 2a_1 - \frac{a_2^2}{\sigma^2} \neq 0 \\ \lambda_1^2(R-t) + \lambda_2^2(N-R), & 2a_1 - \frac{a_2^2}{\sigma^2} = 0, \end{cases}$$

$$Q(t) = \begin{cases} B(R) e^{B_2(R-t)} + Q_3 \left(e^{B_2(R-t)} - 1\right) - \frac{2a_3\lambda_1^2}{B_2}(R-t) e^{B_2(R-t)}, & 2a_1 - \frac{a_2^2}{\sigma^2} = 0 \\ B(R) + \frac{Q_2}{A_2} \left(1 - e^{A_2(R-t)}\right) + Q_1(R-t), & a_1 - \frac{a_2^2}{\sigma^2} = 0 \\ B(R) e^{B_2(R-t)} + \frac{Q_1}{B_2} \left(e^{B_2(R-t)} - 1\right) + 2a_3 \left(A(R) + \frac{\lambda_1^2}{A_2}\right)(R-t) e^{B_2(R-t)}, & a_1 = 0 \\ B(R) e^{B_2(R-t)} + \frac{Q_1}{B_2} \left(e^{B_2(R-t)} - 1\right) + Q_2 \left(e^{B_2(R-t)} - e^{A_2(R-t)}\right), & o/w, \end{cases}$$

$$W(t) = \int_t^R \left(a_3 Q(s) - \frac{a_2^2 Q(s)^2}{4\sigma^2 P(s)} + \rho_1^2 + \nu_1\right) ds + e^{-\delta(R-t)} C(R),$$

$$A_1 = \frac{\lambda_2^2}{(2a_1 - \frac{a_2^2}{\sigma^2})}, \quad A_2 = \left(2a_1 - \frac{a_2^2}{\sigma^2}\right),$$

$$B_1 = \frac{-2a_3 A_1 + 2\rho_2\lambda_2}{B_2}, \quad B_2 = \left(a_1 - \frac{a_2^2}{\sigma^2}\right),$$

$$B_3 = \frac{-2a_3 A_1}{a_1}, \quad B_4 = 2\lambda_2^2 a_3 + \frac{2\rho_2\lambda_2}{B_2}.$$

$$Q_1 = 2\rho_1\lambda_1 - 2a_3\frac{\lambda_1^2}{A_2}, \quad Q_2 = \frac{-2a_3\left(A(R) + \frac{\lambda_1^2}{A_2}\right)}{a_1},$$

$$Q_3 = \frac{2}{B_2}\left(\frac{a_3\lambda_1^2}{B_2} + a_3 A(R) + \rho_1\lambda_1\right).$$

Proof. The explicit solution can be derived similarly to Gerrard *et al.* (2004) and Wang *et al.* (2018). The objective function changes at time R and we need to iteratively derive the solution. For time $t \in [R, N]$, the associated HJB equation is

$$V_t + \inf_{\omega_t \in \mathcal{A}_t} \left[(x(a_1 + \omega_t a_2) + a_3) V_x + \frac{\sigma^2 \omega_t^2}{2} x^2 V_{xx} \right] + (\rho_2 + \lambda_2 x)^2 + \nu_2 = 0, \quad (t, x) \in ([R, N] \times \mathrm{I\!R}) .$$

By substituting $V(t,x) = A(t)x^2 + B(t)x + C(t)$, we obtained that $A(t) > 0$. Then by standard verification theorem, we have derived the solution for $t \in [R, N]$. Since $V(R, x)$ satisfies the quadratic growth condition, then the results in Touzi (2012) can be directly applied so that the value function at time $t \in [0, R]$ can be expressed as

$$V(t, x; p, \alpha, \beta) = \inf_{\omega_t \in \mathcal{A}_t} E\left[\int_t^R (\rho_1 + \lambda_1 X_s)^2 ds + \nu_1 ds + A(R)x^2 + B(R)x + C(R) \middle| \mathcal{F}_t \right].$$

Following the exact same procedure as before, we derived the solution. □

3.4.2 Multi-generation problem

Since multiple cohorts are participating in the intergenerational risk-sharing plan, designing the optimal pension structure based on any particular generation will inevitably violate the interests of the others. Figure 3.3 provides a fan-chart (5% to 95% quantiles) that displays an example of the optimal investment $\hat{\omega}_t$ under the single generation model via simulation. Notice that there is a significant change in the portfolio weights near the retirement date. Similar patterns can be found in Cui *et al.* (2011) with a power utility function. The observation is consistent with the life-cycle hypothesis for a single generation; however, this approach sacrifices the interests of all other generations, which misses the main objective of the IRS framework. Therefore, a multi-generation approach is necessary.

In the multi-generation approach we consider the financial stability of all generations up to time T, giving an objective function:

$$\begin{aligned} V(t, x; T) = \inf_{\omega_t \in \mathcal{A}_t} E\Bigg[\int_t^T & N(1 - \Pi)\{(c^w - (1 - p_s))^2 - \varrho_w(1 - p_s - c^w)\} \\ & + N\Pi\{(c^r - b_s)^2 - \varrho_r(b_s - c^r)\} ds + \varrho_1(X_T - d)^2 - \varrho_2 X_T \Big| \mathcal{F}_t \Bigg] \end{aligned} \tag{3.9}$$

The choice of finite T can be arbitrary, but it is necessary to include the penalty terms $\varrho_1(X_T - d)^2$ and $\varrho_2 X_T$ to ensure that the interests of generations beyond time T will not

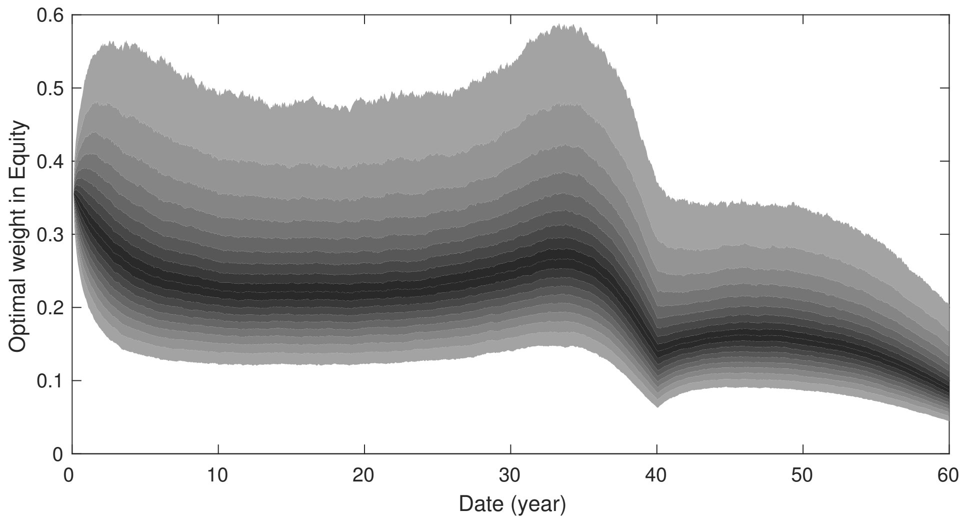


Figure 3.3: The 90% fan-chart of the optimal portfolio weight on equity under the single generation model $\varrho_r = \varrho_w = 0$, $\alpha = \beta = 0.1$ and $p = 0.2$.

be ignored. It would be more satisfying and less arbitrary to consider $T \to \infty$ which will be present in the following section. Similarly to the single-generation case, we simplify the notation such that:

$$\begin{aligned}
V(t,x;T) &= \inf_{\omega_t \in \mathcal{A}_t} E\left[\int_t^T \left(\theta(\eta + X_s)^2 + \nu\right) ds + \varrho_1 (X_T - d)^2 + \varrho_2 X_T \middle| \mathcal{F}_t\right], \\
\theta &= \frac{\alpha^2}{N(1-\Pi)} + \frac{\beta^2}{N\Pi}, \quad \eta = -\frac{\alpha(\rho_1 + \frac{\varrho_w}{2}) + \beta(\rho_2 + \frac{\varrho_r}{2})}{\theta}, \\
\nu &= N(1-\Pi)(\rho_1 - \varrho_w)\rho_1 + N\Pi(\rho_2 - \varrho_r)\rho_2 - \theta\eta^2, \\
\rho_i &= \begin{cases} c^w - 1 + p + \frac{\alpha\psi_w}{N(1-\Pi)} L & i = 1 \\ c^r - b + \frac{\beta\psi_r}{N\Pi} L & i = 2, \end{cases} \\
dX_t &= [X_t(a_1 + \omega_t a_2) + a_3]dt + \omega_t \sigma X_t dZ_t, \\
a_1 &= r - \alpha - \beta, \quad a_2 = (\mu - r), \\
a_3 &= N(1-\Pi)(p+b) - Nb + (\alpha\psi_w + \beta\psi_r)L.
\end{aligned}$$

Proposition 9. *For the optimal control problem (3.9), the optimal asset allocation is*

$$\hat{\omega}(t;T) = \frac{-a_2\left(2A(t;T)X_t + B(t;T)\right)}{2\sigma^2 X_t A(t;T)}, \tag{3.10}$$

and the value function has the following form:

$$V(t,x;T) = A(t;T)x^2 + B(t;T)x + C(t;T),$$

where

$$A(t;T) = \begin{cases} \frac{\theta}{A_1}\left(e^{A_1(T-t)} - 1\right) + \varrho_1 e^{A_1(T-t)}, & 2a_1 - \frac{a_2^2}{\sigma^2} \neq 0 \\ \theta(T-t) + \varrho_1, & 2a_1 - \frac{a_2^2}{\sigma^2} = 0 \end{cases}$$

$$B(t;T) = \begin{cases} (\varrho_2 - 2d\varrho_1 + B_3(T-t))e^{B_1(T-t)} + B_4\left(1 - e^{B_1(T-t)}\right), & 2a_1 - \frac{a_2^2}{\sigma^2} = 0 \\ (\varrho_2 - 2d\varrho_1 + B_5(T-t)) + \frac{B_2}{a_1}\left(1 - e^{A_1(T-t)}\right), & a_1 - \frac{a_2^2}{\sigma^2} = 0 \\ (\varrho_2 - 2d\varrho_1)e^{B_1(T-t)} + \frac{B_5}{B_1}\left(e^{B_1(T-t)} - 1\right) + B_2(T-t)e^{B_1(T-t)}, & a_1 = 0 \\ (\varrho_2 - 2d\varrho_1)e^{B_1(T-t)} + \frac{B_5}{B_1}\left(e^{B_1(T-t)} - 1\right) - \frac{B_2}{a_1}\left(e^{B_1(T-t)} - e^{A_1(T-t)}\right), & o/w \end{cases}$$

$$C(t;T) = \varrho_1 d^2 + \int_t^T \left(a_3 B(s;T) - \frac{a_2^2 B(s;T)^2}{4\sigma^2 A(s;T)} + \theta\eta^2 + \nu\right) ds$$

$$A_1 = 2a_1 - \frac{a_2^2}{\sigma^2}, \quad B_1 = a_1 - \frac{a_2^2}{\sigma^2}, \quad B_2 = \frac{2a_3\theta}{A_1} + 2\varrho_1 a_3,$$

$$B_3 = \frac{-2a_3\theta}{B_1}, \quad B_4 = \frac{B_3 - 2\theta\eta - 2a_3\varrho_1}{B_1}, \quad B_5 = 2\theta\eta - \frac{2a_3\theta}{A_1}.$$

The proof is exactly the same as in the single-generation case.

Remark. *When each individual follows their own optimal asset allocation (3.7), then the the cumulative investment strategy follows a similar structure:*

$$\hat{\omega}^S_{\sum}(X_t) = \frac{1}{T}\left(\int_0^R \frac{1}{R}\hat{\omega}^S_t dt + \int_R^T \frac{1}{T-R}\hat{\omega}^S_t dt\right) = \frac{-(\mu - r)}{\sigma^2}\left(1 + \frac{C}{X_t}\right),$$

where C is a constant. However, it is only in special cases where $\hat{\omega}^S_{\sum}(X_t) = \hat{\omega}(t;T)$. An example would be $\alpha = 0$, $r = \beta$, $c^w = 1 - p$ and $c^r = b$. Otherwise, it is simple to show that $\hat{\omega}^S_{\sum}(X_t) \neq \hat{\omega}(t;T)$. This illustrates that the group optimality cannot be achieved at the individual level.

3.4.3 Infinite Time Horizon ($T \to \infty$)

The terminal condition in the multi-generation problem involves subjective choice of the penalty term ϱ_2. We now focus on ensuring the long-term sustainability of the pension plan by letting $T \to \infty$. To avoid an infinite value function, we calculate the long-term average consumption dispersion from the target, to form an ergodic control problem similar to the one considered in Section 3.3.2:

$$v = \inf_{\omega_t \in \mathcal{A}} \limsup_{T\to\infty} \frac{1}{T} E\left[\int_0^T N(1-\Pi)((c^w - (1-p_s))^2 - \varrho_w(1-p_s-c^w)) \right. \\ \left. + N\Pi((c^r - b_s)^2 - \varrho_r(b_s - c^r))ds \right]. \tag{3.11}$$

Again, the problem can be simplified that

$$v = \inf_{\omega_t \in \mathcal{A}} \limsup_{T\to\infty} \frac{1}{T} E\left[\int_0^T \theta(\eta + X_t)^2 + \nu dt\right],$$

where all notations are kept the same as in Section 3.4.2.

Proposition 10. *For optimal control problem (3.11), if $r - \alpha - \beta - \frac{1}{2}\left(\frac{\mu-r}{\sigma}\right)^2 < 0$, then the optimal asset allocation is*

$$\hat{\omega}_t = \frac{\mu - r}{\sigma^2}\left(-1 - \frac{1}{X_t}\frac{a_3 - \left[\alpha(\rho_1 + \frac{\varrho_w}{2}) + \beta(\rho_2 + \frac{\varrho_r}{2})\right]\left(2\alpha + 2\beta - 2r + \left(\frac{\mu-r}{\sigma}\right)^2\right)}{\alpha + \beta - r + \left(\frac{\mu-r}{\sigma}\right)^2}\right), \tag{3.12}$$

and the value function is

$$v = \nu + \theta\eta^2 + a_3 B - \left(\frac{\mu - r}{\sigma}\right)^2 \frac{B^2}{4A},$$

where

$$A = \frac{\theta}{-2(r-\alpha-\beta) + \left(\frac{\mu-r}{\sigma}\right)^2}, \quad B = \frac{2Aa_3 + 2\theta\eta}{-(r-\alpha-\beta) + \left(\frac{\mu-r}{\sigma}\right)^2}.$$

The proof follows exactly as Section 3.3.2.

Remark. *When $r - \alpha - \beta - \frac{1}{2}\left(\frac{\mu-r}{\sigma}\right)^2 < 0$, we can show the relationship between the optimal control problems (3.9) and (3.11). It is trivial to notice that $\lim_{T\to\infty} B(t;T) = B$*

and $\lim_{T\to\infty} A(t;T) = A$, *and since* $\lim_{T\to\infty} \frac{A(t;T)}{T} = 0$ *and* $\lim_{T\to\infty} \frac{B(t;T)}{T} = 0$, *we have*

$$\begin{aligned}
&\lim_{T\to\infty} \frac{C(t;T)}{T} \\
=&\lim_{T\to\infty} \frac{1}{T}\int_t^T a_3 B(s;T) - \frac{a_2^2 B(s;T)^2}{4\sigma^2 A(s;T)} ds + \theta\eta^2 + \nu \\
=&\lim_{T\to\infty} \left\{ a_3 B(T;T) - \frac{a_2^2 B(T;T)^2}{4\sigma^2 A(T;T)} + \int_t^T \frac{d}{dT}\left(a_3 B(s;T) - \frac{a_2^2 B(s;T)^2}{4\sigma^2 A(s;T)} \right) ds + \theta\eta^2 + \nu \right\} \\
=&\lim_{T\to\infty} \left\{ a_3 B(T;T) - \frac{a_2^2 B(T;T)^2}{4\sigma^2 A(T;T)} + \int_t^T \frac{d}{ds}\left(-a_3 B(s;T) + \frac{a_2^2 B(s;T)^2}{4\sigma^2 A(s;T)} \right) ds + \theta\eta^2 + \nu \right\} \\
=&\lim_{T\to\infty} \left\{ a_3 B(T;T) - \frac{a_2^2 B(T;T)^2}{4\sigma^2 A(T;T)} - a_3 B(s;T)\Big|_t^T + \frac{a_2^2 B(T;T)^2}{4\sigma^2 A(s;T)}\Big|_t^T + \theta\eta^2 + \nu \right\} \\
=&\lim_{T\to\infty} \left\{ a_3 B(t;T) - \frac{a_2^2 B(t;T)^2}{4\sigma^2 A(t;T)} + \theta\eta^2 + \nu \right\} \\
=&v.
\end{aligned}$$

Therefore, $\lim_{T\to\infty} \frac{1}{T} V(t,x;T) = v$ *and* $\lim_{T\to\infty} \hat{\omega}(t;T) = \hat{\omega}_t$. *This continuity feature ensures that the optimal pension design derived from (3.9) with sufficiently long horizon will be close to the optimal pension design in (3.11).*

Remark. *Similarly to Section 3.3.2, the explicit solution is given on a subset of parameters. However, the condition differs in that the risk-sharing parameters* α *and* β *are involved. In the next section, we will demonstrate that when a regulatory constraint is included, the condition* $r - \alpha - \beta - \frac{1}{2}\left(\frac{\mu - r}{\sigma}\right)^2 < 0$ *is automatically satisfied.*

3.5 Optimal Pension Design with Regulatory Constraints

In this section, we will illustrate that the unconstrained optimization problem discussed in Section 3.4 has similar issues as the optimal control problems in Section 3.3.

3.5.1 Regulatory Constraint

When the pension plan is designed for the long-term financial stability of all participants, the interest of any particular generation would not be important. Therefore, the "optimal" plan, in the mathematical sense, may sacrifice the current generation. The "optimal" strategy ensures the asset is able to achieve a certain funding level, where all future payments can be guaranteed through the risk-free investment. An example of the portfolio weight is given in Figure 3.4. Possible ways to alleviate this issue include: incorporating a discount factor, selecting appropriate penalty parameters (ϱ_r and ϱ_w), fixing the maturity time T, and constraining the parameter space. However, most of these are not suitable in our context. The discount factor, penalty terms and fixed maturity date, all suffer from interpretation issues, along with subjective choices of the values. Although it is natural to set constraints on the portfolio weight (ex, $\underline{\omega} \leq \omega_t \leq \bar{\omega}$), it may not solve the problem as the primary issue comes from the unreasonable α and β in the optimal pension structure.

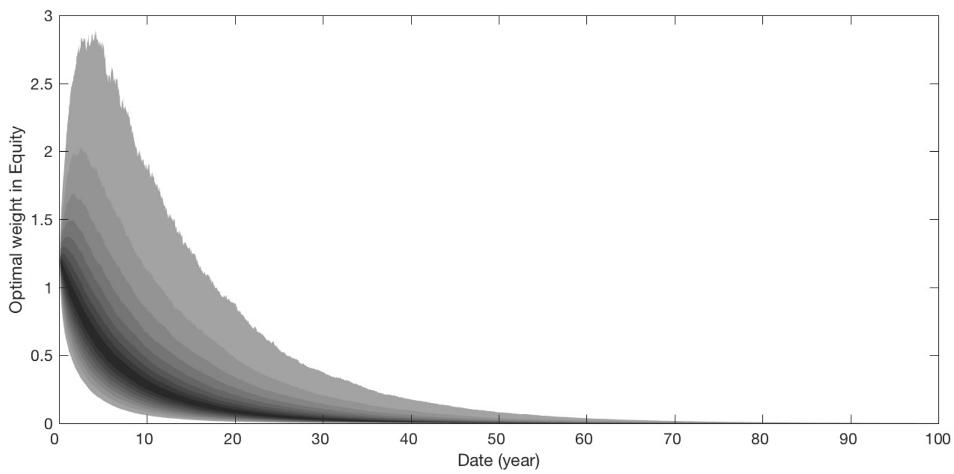


Figure 3.4: The 90% fan-chart of the optimal portfolio weight on equity under the multi-generation model, with $T \to \infty$, using the 'optimal' parameters.

Therefore, the optimization must be done under some restricted domain for the risk-sharing parameters. Setting α and β to be between 0 and 1 would be a natural choice, but more restrictive constraints can be applied to align with regulatory requirements.

When a deficit occurs, sponsors are often required to create a recovery plan that is expected to clear the deficit within a specified period. The risk-sharing feature in the IRS plan will quicken the recovery process, and in particular, the funding level can be recovered through risk-free investment. In our analysis, we define the recovery period as the number of years that the pension fund will need to return to the target level if all assets are invested in the risk-free bond. Specifically, for the target funding ratio of f_r, we define

$$t^* = \inf\left\{ t : t > 0, \frac{X_t}{L} = f_r \middle| \frac{X_0}{L} < f_r \right\}, \tag{3.13}$$
$$\text{where } dX_t = (rX_t - N(1-\Pi)p_t + N\Pi b_t)\, dt.$$

Theoretically, one can use any investment strategy in defining the recovery period; however, it would be illogical to consider the investment strategy before the decision of the pension structure. Therefore, the calculation of the recovery period should be based on objective methods. There are at least three options available: using industry standard practice (for example, using the average portfolio in the pension industry); using the risk-free assets (the conservative choice); and the "optimal" choice, for example, Equation (3.12). We should mention that the "optimal" choice may become subjective if the objective function is utility based (where each pension plan has their own risk aversion parameter). In this section, we use the risk-free investment.

It is easy to recognize the technical drawback of setting $f_r = \psi_r = \psi_w = 1$ which leads to a recovery time of infinity. Fortunately, from both the theoretical (Section 3.3) and practical perspective, we can safely modify this assumption. For example, from the Financial Toezichtskader (FTK), established by the Dutch National Bank, the pension sponsor must implement a recovery plan when the funding level is below 123.9%, and cut the pension benefit if the funding level is below 104.2% for five consecutive years; in addition, no indexation is allowed if the funding level is below 110%.

For illustration purposes, we implement a simplified constraint in the following way. Assume that the surplus/deficit is distributed at 110% funding level, $\psi_r = \psi_w = 110\%$, and that regulation requires that, if the plan is below 90% funded, it must schedule payments to achieve 105% funding within 10 years. Of course, in practice, constraints are more complicated, but the general idea is captured. When $\psi_r = \psi_w = \psi$, the constraint on the recovery period can be transformed into a constraint on the risk-sharing parameter $\xi = \alpha + \beta$. Equation (3.13) becomes

$$t^*(\xi) = \frac{1}{r-\xi} \log\left(\frac{(f_r - 1)r + (\psi - f_r)\xi}{\left(\frac{x_0}{L} - 1\right) r + \left(\psi - \frac{x_0}{L}\right)\xi} \right) \tag{3.14}$$

where $\frac{x_0}{L}$ is the initial funding level. Figure 3.5 presents the recovery period for each deficit level back to 105%, using the largest risk-sharing parameters studied by Cui *et al.* (2011) ($\alpha = 0.06$, $\beta = 0.02$). Notice that when the funding level is 90%, the recovery period is about 17 years, which is above most regulatory requirements, which are generally in the range of 5 to 10 years. The recovery period remains high for any reasonable choice of the risk-free rate.

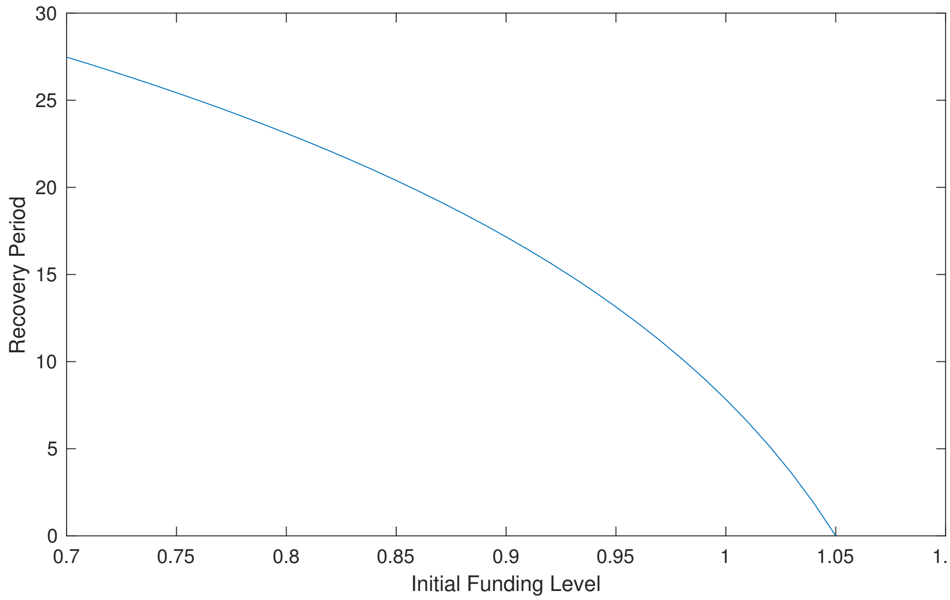


Figure 3.5: The recovery period when $\alpha = 0.06$ and $\beta = 0.02$.

To apply the regulatory constraint, we denote the mandatory recovery period to be t_r, which is set to be 10 in our benchmark examples. We are trying to find ξ that solves the equation:

$$t^*(\xi) - t_r = 0$$

However, the solution may not be unique, and to avoid any ambiguity, we define:

$$\xi^* = \max\left\{\xi' \middle| t^*(\xi') = t_r, \left(\frac{dt^*(\xi)}{d\xi} < 0, \quad \forall \xi > \xi'\right)\right\}, \tag{3.15}$$

such that ξ^* is the largest parameter that satisfies the regulatory constraint. Then the constrained parameter domain is simply $\{(\alpha, \beta) | \xi^* \leq \alpha + \beta \leq 1, \alpha \geq 0, \beta \geq 0\}$.

Proposition 11. *If* $\frac{f_r - f_0}{(\psi-1)r} > t_r$*, then* ξ^* *exists and* $\xi^* > r$*.*

Proof. By taking the derivative of $t^*(\xi)$, we have

$$\frac{dt^*}{d\xi} = \frac{1}{(r-\xi)^2}\log\left[\frac{(f_r-1)r+(\psi-f_r)\xi}{(f_0-1)r+(\psi-f_0)\xi}\right] + \frac{1}{r-\xi}\frac{\frac{(\psi-f_r)}{(f_0-1)r+(\psi-f_0)\xi} + \frac{-(\psi-f_0)((f_r-1)r+(\psi-f_r)\xi)}{((f_0-1)r+(\psi-f_0)\xi)^2}}{\frac{(f_r-1)r+(\psi-f_r)\xi}{(f_0-1)r+(\psi-f_0)\xi}}$$
$$= \frac{1}{r-\xi}\left\{\frac{1}{r-\xi}\log\left[\frac{(f_r-1)r+(\psi-f_r)\xi}{(f_0-1)r+(\psi-f_0)\xi}\right] + \frac{(f_r-x_0)(1-\psi)r}{((f_r-1)r+(\psi-f_r)\xi)((f_0-1)r+(\psi-f_0)\xi)}\right\}.$$

Then, using the Taylor expansion of the log function:

$$\frac{1}{r-\xi}\log\left[\frac{(f_r-1)r+(\psi-f_r)\xi}{(f_0-1)r+(\psi-f_0)\xi}\right]$$
$$= \frac{1}{r-\xi}\log\left[1-\frac{(f_0-f_r)(r-\xi)}{(f_0-1)r+(\psi-f_0)\xi}\right]$$
$$= \left(\frac{-(f_0-f_r)}{(f_0-1)r+(\psi-f_0)\xi} - \frac{1}{2}\left(\frac{(f_0-f_r)}{(f_0-1)r+(\psi-f_0)\xi}\right)^2(r-\xi) - \frac{1}{3}()^3(r-\xi)^2\cdots\right)$$

If $\xi > r$ then $\left|\frac{(f_0-f_r)(r-\xi)}{(f_0-1)r+(\psi-f_0)\xi}\right| < 1$, which ensures the convergence. By combining the first part of the Taylor expansion and the second part of $\frac{dt_r^*}{d\xi}$, we have

$$\frac{(f_r-f_0)((f_r-1)r+(\psi-f_r)\xi)+(f_r-f_0)(1-\psi)r}{((f_r-1)r+(\psi-f_r)\xi)((f_0-1)r+(\psi-f_0)\xi)}$$
$$=\frac{(f_r-\psi)(f_r-f_0)(r-\xi)}{((f_r-1)r+(\psi-f_r)\xi)((f_0-1)r+(\psi-f_0)\xi)} > 0$$

All other terms in the Taylor expansion are positive, therefore $\frac{dt^*}{d\xi} < 0$, when $\xi > r$. Furthermore, by taking the limit, we have

$$\lim_{\xi\to r^+} t_r^*(\xi) = \lim_{\xi\to r^+}\frac{\frac{d}{d\xi}\log\left[\frac{(f_r-1)r+(\psi-f_r)\xi}{(f_0-1)r+(\psi-x_0)\xi}\right]}{-1}$$
$$= \lim_{\xi\to r^+}\frac{(f_r-f_0)(\psi-1)r}{((f_r-1)r+(\psi-f_r)\xi)((f_0-1)r+(\psi-f_0)\xi)}$$
$$=\frac{f_r-f_0}{(\psi-1)r}$$

Therefore, if $\frac{f_r-f_0}{(\psi-1)r} > t_r$, then $\xi^* > r$ exists. ∎

Using our benchmark parameter set, the constraint is transformed to $\{0.1375 < \alpha+\beta < 1, \alpha \geq 0, \beta \geq 0\}$. Figure 3.6 displays the optimal weight in equity under the optimal pension design given the regulatory constraints. Apparently the constrained optimal pension design avoids the short-term over-investment in risky assets. Therefore, the nearby generations will have a relatively stable income, and the issue of intergenerational unfairness is lessened.

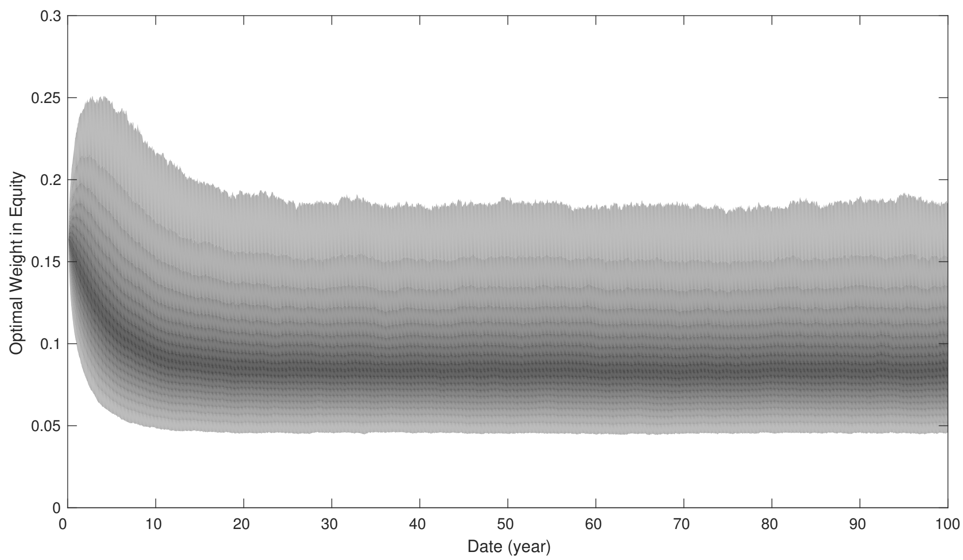


Figure 3.6: The 90% fan-chart of optimal portfolio weight on equity under the multi-generation model with $T \to \infty$, using optimal parameters constrained by regulation. The optimal parameters are $\alpha = 0.1084$, $\beta = 0.0292$ and $p = 0.2241$, for a target consumption of 0.9 throughout one's life. The long term funding level is 117.26%.

3.5.2 Sensitivity Tests

This section performs sensitivity tests over different parameters, to check the robustness of the optimal pension design.

Figure 3.7 presents the optimal pension design for different target consumption levels. We have set $c^w, c^r \in [0.9, 1.15]$ with $c^r < c^w$ (target consumption level for retirees smaller than for active workers). We kept $\mu = 0.08$, $r = 0.02$, $\sigma = 0.25$, and $\psi_r = \psi_w = 1.1$, and made the following observations:

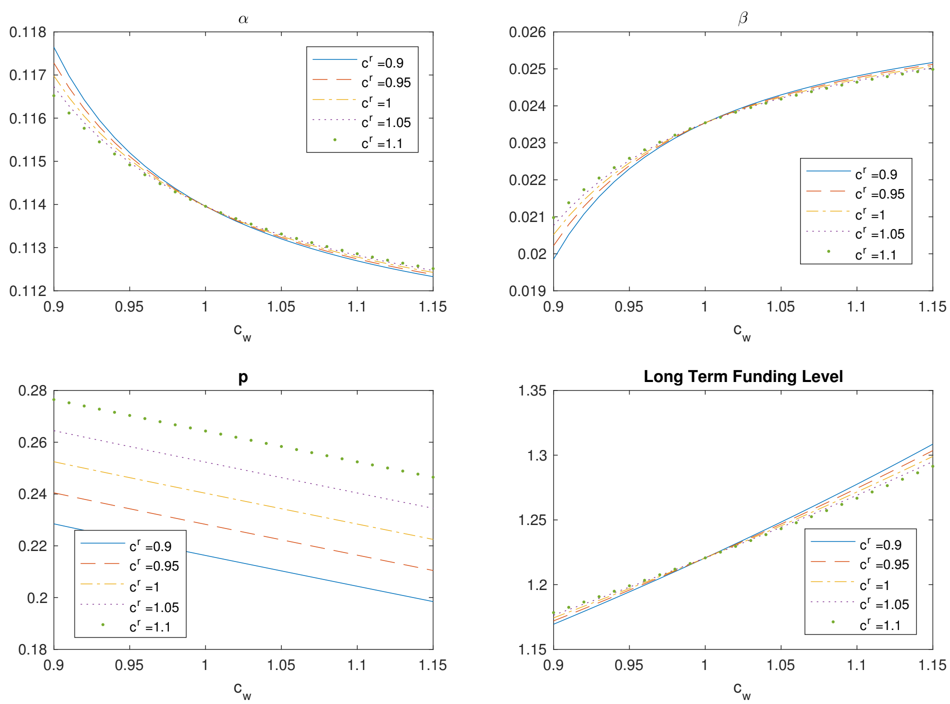


Figure 3.7: Sensitivity of optimal design over different c^w and c^r.

- α and β are extremely robust against changes in c^w and c^r. Although there is a negative relation between the target consumption level and the pension parameters, the scale is relatively small.

- When we increase c^w, p is decreasing (less contribution) which also leads to a decrease in benefit b and liability value, which increases the long term funding level. Exactly the opposite observation can be made for c^r. However, p remains relatively stable in $[0.2, 0.28]$.

- As the target consumption levels increase, it is intuitive to expect that the required funding level also increases. Therefore, it is not surprising to observe a strong positive relationship between long-term funding level and the target consumption. Again, the long-term funding level is quite robust and lays in the range of $[1.15, 1.3]$.

The last observation is interesting since it suggests the necessity of reasonable over-funding, and coincides with the current regulations on DB plans in the Netherlands (where full indexation of DB plans is permissible only when the funding level is above 130%). It also indicates that strict tax limits on surplus should be reconsidered to achieve funding and benefit stability.

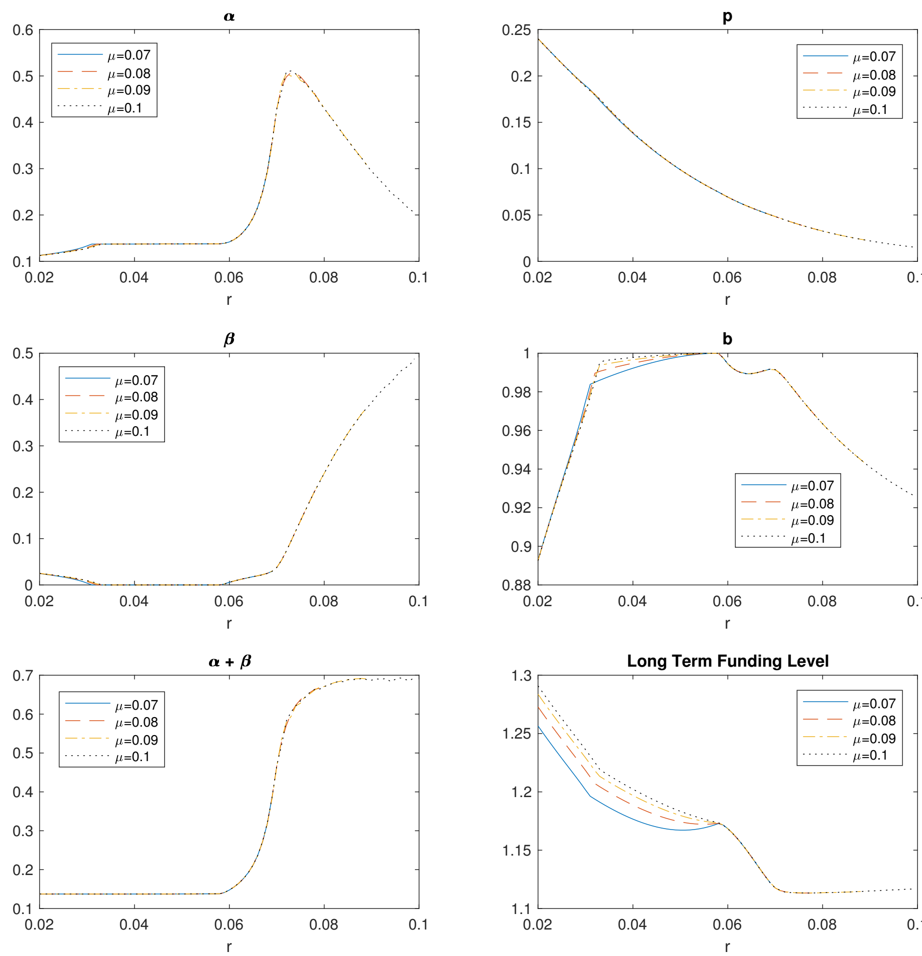


Figure 3.8: Sensitivity of optimal design over different μ and r, when $c^w = c^r = 1$.

Figure 3.8 presents the optimal pension design for different equity returns μ and risk-free rates r. We have set $c^w = c^r = 1$; all other assumptions are unchanged. We make the following observations:

- The optimal pension design is robust with respect to μ, which is desirable as there is significant uncertainty in estimating the expected equity return.
- The optimal pension design is sensitive to the risk-free rate, r. Part of the reason for this depends on the relationship between b and p, such that a small increase in r may substantially increase the value of the retirement benefit b. The hump observed in α will be discussed in more detail in the next section.
- Increases in the risk-free rate allow more risk-sharing in the pension design.
- Increases in r mean less contribution is required for the same amount of benefits, and thus lead to a decrease in p.
- The long term funding level is between 1.1 and 1.3. In a low interest rate environment the optimal pension design will achieve a long term funding level in the range $[1.15, 1.3]$.

3.5.3 Optimality of Pure DC or DB Plan

In this section we comment further on designing the optimal intergenerational risk-sharing pension, and on selecting the appropriate target consumption level. Figure 3.9 presents the same sensitivity test as Figure 3.8, except the target consumption is $c^w = c^r = 0.9$. The hump of α in Figure 3.8 has been replaced by a cut-off when the interest rate is between 0.04 and 0.05. This is caused by the situation where the optimal benchmark consumption level $(1-p)$ for active employees is larger than their target consumption level $(1-p > c^w)$, in which case the optimal IRS plan collapses to a DC plan. In fact, when the interest rate is high, the target consumption level is achievable through risk-free investment, and the ultimate long term weight in equity becomes negative.

This issue can be resolved in two ways: either by taking $\varrho_w > 0$ and $\varrho_r > 0$, or by raising the target consumption level. Since the equity return is above the risk-free rate, increasing the risky investment should lead to a higher expected future benefit than b (and lower contribution than p). Therefore, it is reasonable to set a consumption goal that is higher than 1 for employees to participate in market growth.

Figure 3.10 demonstrates the increasing effectiveness of the intergenerational risk-sharing structure as we increase the target consumption level. We plot the value function v for different α and β under the optimal p. The grid surface represents the value function, and the vertical plane represents the regulatory requirement on $\alpha + \beta$. The optimal

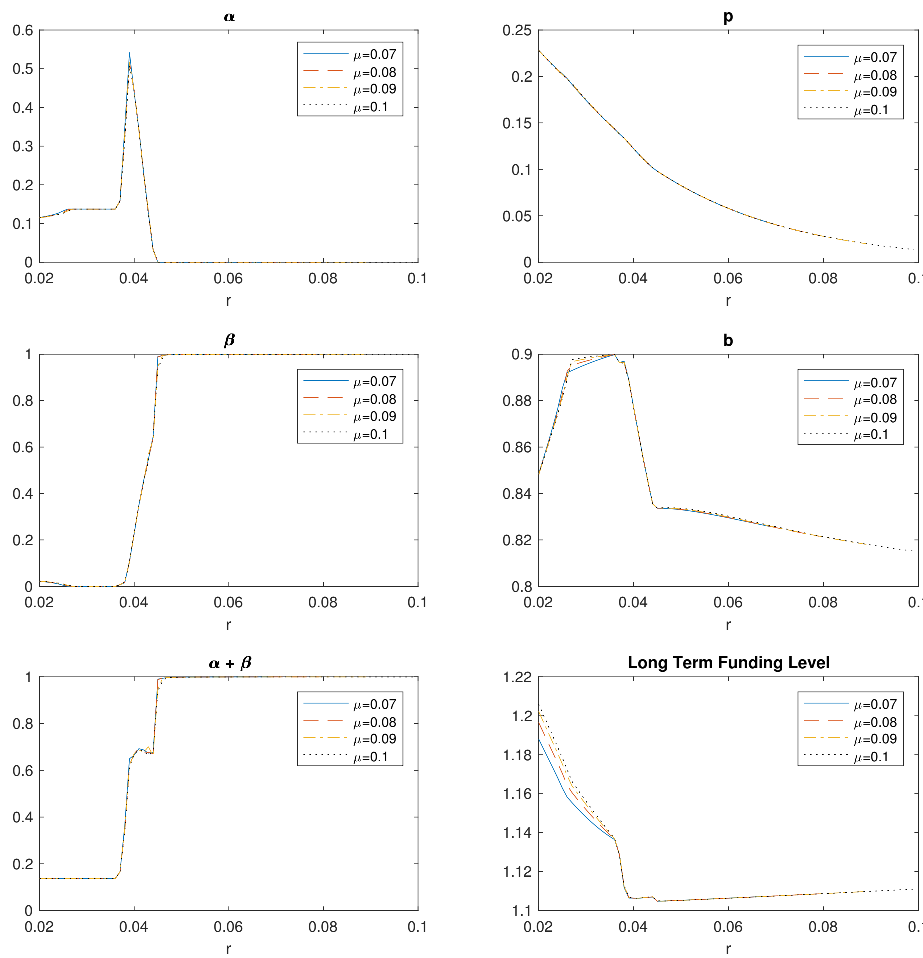
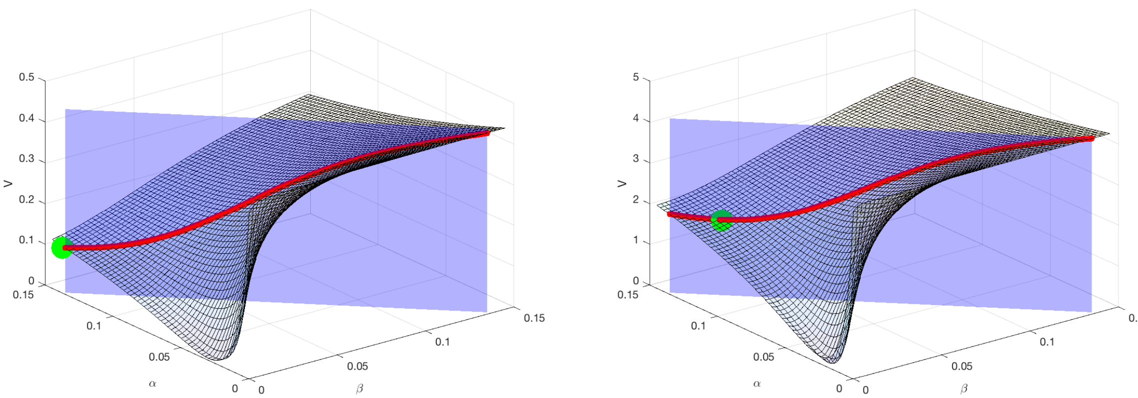


Figure 3.9: Sensitivity of optimal design over different μ and r, when $c^w = c^r = 0.9$.

pension design is the minimum point on the intersection between the regulatory plane and the value function grid surface (the red line). The optimal design is close to a traditional DB plan when $c^w = c^r = 0.9$, except that the market risk borne by the active employees is smoothed across time ($\alpha \neq 1$). As c^w increases from 0.9 to 1.1, the optimal design moves away from the DB plan.

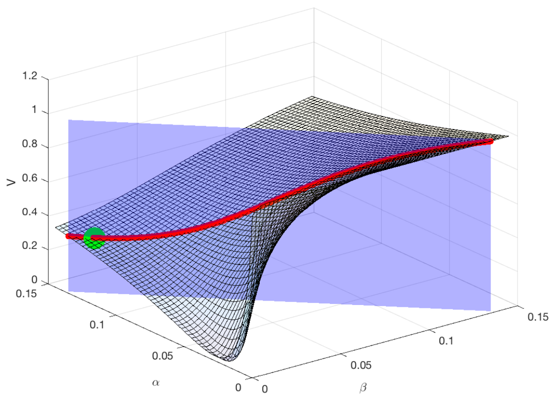


(a) $c^w = 0.9$, $c^r = 0.9$, $r = 0.025$

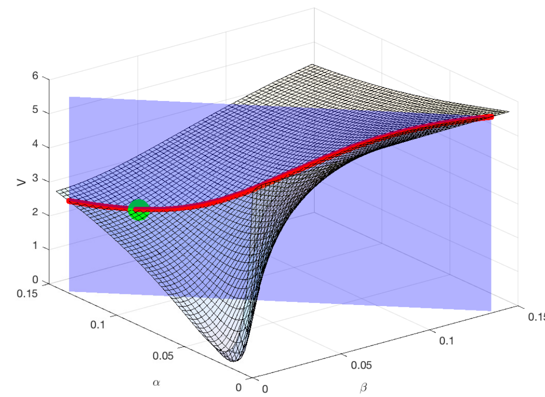
(b) $c^w = 1.1$, $c^r = 1.1$, $r = 0.025$

Figure 3.10: The value function under different α and β with optimal p, the red line is the intersection between the regulatory constraint (blue plane) and the value function surface.

As a final note, we have not discussed the risk from demographic changes, which is an important reason why the IRS plan should be more effective than the DB plan. A well-designed IRS should have the ability to smooth both market risk and the risk associated with population structure. Figure 3.11 duplicates Figure 3.10 with $R = 36$ ($\Pi = 0.4$ instead of $1/3$). Clearly, as the number of retirees increases, more risk is to be shared between generations.



(a) $c^w = 0.9$, $c^r = 0.9$, $r = 0.025$



(b) $c^w = 1.1$, $c^r = 1.1$, $r = 0.025$

Figure 3.11: The value function under different α and β with optimal p, when the percentage of retirees is $\Pi = 0.4$, the red line is the intersection between the regulatory constraint (blue plane) and the value function surface.

3.6 Conclusion

In this chapter, we have studied a stylized intergenerational risk-sharing structure that is similar to that in Cui *et al.* (2011). We provided some theoretical justification for our pension structure and illustrate some shortcomings in the existing research. We incorporated a regulatory requirement as a natural constraint to our optimization process, which enhances the fairness across generations. In addition, we demonstrated that the impact of regulation may force the design to a pure DB or a pure DC pension.

Chapter 4

Future Work

This thesis explores three topics in hybrid pension plans, namely, the hedging strategy, the liability valuation and the optimal hybrid design. This chapter discusses future work on each topic.

Chapter 1 investigates the performance of dynamic hedging strategies on Cash Balance Pension Plans, using the popular one- and two-factor Hull White model. Although the liability valuation is quite robust with respect to the choice of the models, hedging results are sensitive to the underlying assumption and thus more sophisticated models should be considered. In addition, since the pension funds are unlikely to rebalance their portfolio in short horizon, static or semi-static hedging strategies should be investigated.

Chapter 2 illustrates how new hybrid pension designs can be created to flexibly fulfill the needs for both sponsors and employees. However, our assumptions are oversimplified; for example, salaries can only be partially hedged in the market, interest rates and annuity factors should not be constant over the time, and longevity risk in DB plans cannot be ignored. Without addressing the aforementioned issues, our valuation is likely to underestimate the true risks embedded in the plan. Moreover, our valuation is based on a sponsor's conservative views; quantitative analysis from an employee's perspective should also be performed. A more comprehensive comparison between the sponsors' cost and employees' utility gain should be studied to assess the feasibility of the new design.

Chapter 3 studies the fairness issue in the intergenerational risk sharing pension plan. Both our underlying model and plan structure are stylized. The fixed population structure, the exclusion of sponsor's responsibility, and the simple Geometric Brownian Motion for the financial market have all limited the practicality of our results. In addition, other than the optimal design of the IRS plan, another practical issue is on the optimal transition from an existing pension plan to the new IRS plan. The goal of achieving long-term sustainability of a pension fund may conflict with the interests of the current retirees. How to reform the current pension system without violating the equality and equity among different generations remains an unsolved question.

Essentially, pension reform is a collective bargaining agreement, that involves policy-makers, pension sponsors, pension trusts, labour unions, retirees, current workers and all future generations. Most current literature resides on the interests of one party only, with limited discussions that involve more than two parties. Considerably more effort should be made on strategic balancing of interests between different stakeholders.

References

Bacinello AR (2000). "Valuation of contingent-claims characterising particular pension schemes." *Insurance: Mathematics and Economics*, **27**(2), 177–188.

Bams D, Schotman PC, Tyagi M (2016). "Optimal risk sharing in a collective defined contribution pension system." *Discussion paper*, Network for Studies on Pensions, Aging and Retirement, Tilburg University.

Ben-Ameur H, Breton M, L'Ecuyer P (2002). "A Dynamic Programming Procedure for Pricing American-Style Asian Option." *Management Science*, **48**(5), 625–643.

Biggs AG (2010). "An options pricing method for calculating the market price of public sector pension liabilities." *AEI (American Enterprise Institute) Working Paper*, (164).

Biggs AG, Smetters K (2013). "Understanding the argument for market valuation of public pension liabilities." *American Enterprise Institute*.

Björk T (2009). *Arbitrage Theory in Continuous Time*. Oxford University Press.

Boes MJ, Siegmann A (2018). "Intergenerational risk sharing under loss averse preferences." *Journal of Banking and Finance*, **92**, 269–279.

Bovenberg L, Mehlkopf R (2014). "Optimal design of funded pension schemes." *Annu. Rev. Econ.*, **6**(1), 445–474.

Boyle P, Hardy M (2003). "Guaranteed annuity options." *ASTIN Bulletin: The Journal of the IAA*, **33**(2), 125–152.

Brigo D, Mercurio F (2007). *Interest rate models-theory and practice: with smile, inflation and credit*. Springer Science & Business Media.

Brown DT, Dybvig PH, Marshall WJ (2001). "The cost and duration of cash-balance pension plans." *Financial Analysts Journal*, **57**(6), 50–62.

Bruhn K, Steffensen M (2013). "Optimal smooth consumption and annuity design." *Journal of Banking & Finance*, **37**(8), 2693–2701.

Chen K, Hardy MR (2009). "The DB underpin hybrid pension plan: fair valuation and funding." *North American Actuarial Journal*, **13**(4), 407–424.

Coronado JL, Copeland PC (2004). "Cash balance pension plan conversions and the new economy." *Journal of Pension Economics & Finance*, **3**(3), 297–314.

Cui J, De Jong F, Ponds E (2011). "Intergenerational risk sharing within funded pension schemes." *Journal of Pension Economics & Finance*, **10**(1), 1–29.

Diamond PA (1965). "National debt in a neoclassical growth model." *The American Economic Review*, **55**(5), 1126–1150.

Dickson DC, Hardy MR, Waters HR (2013). *Actuarial Models for Life Contingent Risks (2nd Edition)*. Cambridge University Press.

Driessen J, Klaassen P, Melenberg B (2003). "The performance of multi-factor term structure models for pricing and hedging caps and swaptions." *Journal of Financial and Quantitative Analysis*, **38**(3), 635–672.

Enev E (2011). "Gaussian affine term structure models within the Duffie-Kan framework and their use for forecasting." *Master's thesis, University of Amsterdam.*

Fan R, Gupta A, Ritchken P (2001). "On Pricing and Hedging in the Swaption Market: How Many Factors, Really?" *Working Paper.*

Forsyth PA, Vetzal KR (2002). "Quadratic convergence for valuing American options using a penalty method." *SIAM Journal on Scientific Computing*, **23**(6), 2095–2122.

Gerrard R, Haberman S, Vigna E (2004). "Optimal investment choices post-retirement in a defined contribution pension scheme." *Insurance: Mathematics and Economics*, **35**(2), 321–342.

Gerrard R, Haberman S, Vigna E (2006). "The management of decumulation risks in a defined contribution pension plan." *North American Actuarial Journal*, **10**(1), 84–110.

Glasserman P (2013). *Monte Carlo methods in financial engineering*, volume 53. Springer Science & Business Media.

Goecke O (2013). "Pension saving schemes with return smoothing mechanism." *Insurance: Mathematics and Economics*, **53**(3), 678–689.

Gold J (2001). "Shareholder-optimal design of cash balance pension plans. Pension Research Council." *Technical report*, Working Paper 2001-7.

Gollier C (2008). "Intergenerational risk-sharing and risk-taking of a pension fund." *Journal of Public Economics*, **92**(5-6), 1463–1485.

Gordon RH, Varian HR (1988). "Intergenerational risk sharing." *Journal of Public Economics*, **37**(2), 185–202.

Guillén M, Jørgensen PL, Nielsen JP (2006). "Return smoothing mechanisms in life and pension insurance: path-dependent contingent claims." *Insurance: Mathematics and Economics*, **38**, 229–252.

Gupta A, Subrahmanyam MG (2005). "Pricing and hedging interest rate options: Evidence from cap–floor markets." *Journal of Banking & Finance*, **29**(3), 701–733.

Gurrieri S, Nakabayashi M, Wong T (2009). "Calibration Methods of Hull-White Model." *SSRN Electronic Journal*.

Hardy M (2003). *Investment guarantees: modeling and risk management for equity-linked life insurance*, volume 215. John Wiley & Sons.

Hardy MR, Saunders D, Zhu X (2014). "Market-consistent valuation and funding of cash balance pensions." *North American Actuarial Journal*, **18**(2), 294–314.

Harvey J (2012). "Strategies for hedging interest rate risk in a cash balance plan." *Russell Investments*.

Hewitt A (2016). "Actuarial Valuation as at July 1, 2016 for University of Toronto Pension Plan." *Technical report*.

Khorasanee ZM (2012). "Risk-sharing and benefit smoothing in a hybrid pension plan." *North American Actuarial Journal*, **16**(4), 449–461.

Kravitz (2018). "2018 National CASH BALANCE Research Report." *Technical report*, Kravitz, Inc.

Leizarowitz A (1988). "Controlled diffusion processes on infinite horizon with the overtaking criterion." *Applied Mathematics and Optimization*, **17**(1), 61–78.

Leizarowitz A (1990). "Optimal controls for diffusion in RdA min-max max-min formula for the minimal cost growth rate." *Journal of mathematical analysis and applications*, **149**(1), 180–209.

Longstaff FA, Schwartz ES (2001). "Valuing American options by simulation: a simple least-squares approach." *Review of Financial Studies*, pp. 113–147.

Marshall C (2011). *Financial Risk Management of Guaranteed Minimum Income Benefits Embedded in Variable Annuities*. Ph.D. thesis.

Merton RC (1969). "Lifetime portfolio selection under uncertainty: the continuous-time case." *The Review of Economics and Statistics*, **51**(3), 247–257.

Merton RC (1983). "On the role of social security as a means for efficient risk sharing in an economy where human capital is not tradable." In *Financial aspects of the United States pension system*, pp. 325–358. University of Chicago Press.

Milevsky MA, Promislow SD (2004). "Florida's pension election: from DB to DC and back." *Journal of Risk and Insurance*, **71**(3), 381–404.

Murphy RJ (2001). "The Cash Balance Funding Method." In *Cash Balance Symposium Monograph, Society of Actuaries*.

Niehaus G, Yu T (2005). "Cash-Balance plan conversions: evidence on excise taxes and implicit contracts." *Journal of Risk & Insurance*, **72**(2), 321–352.

Novy-Marx R, Rauh JD (2009). "The liabilities and risks of state-sponsored pension plans." *Journal of Economic Perspectives*, **23**(4), 191–210.

Pension Experience Subcommittee (2014). "Canadian Pensioners' Mortality." *Technical report*, Canadian Institute of Actuaries.

Peskir G, Shiryaev A (2006). *Optimal stopping and free-boundary problems*. Springer.

PricewaterhouseCoopers (2014). "Embedded Options in Pension Plans - Valuation of Guarantees in Cash Balance Plans." *Technical report*, Society of Actuaries.

Rogerscasey S (2015). "Examining Cash-Balance Plans' Investment Risk." *Technical report*, Segal Rogerscasey Investment Brief.

Schrager DF, Pelsser AA (2006). "Pricing swaptions and coupon bond options in affine term structure models." *Mathematical Finance*, **16**(4), 673–694.

Stockbridge RH (1989). "Time-average control of martingale problems: the Hamilton-Jacobi-Bellman equation." *Stochastics: An International Journal of Probability and Stochastic Processes*, **27**(4), 249–260.

Teulings CN, De Vries CG (2006). "Generational accounting, solidarity and pension losses." *de Economist*, **154**(1), 63–83.

Touzi N (2012). *Optimal stochastic control, stochastic target problems, and backward SDE*, volume 29. Springer Science & Business Media.

Touzi N (2013). *Optimal Stochastic Control, Stochastic Target Problems, and Backward SDE*. Springer-Verlag New York.

Wang S, Lu Y, Sanders B (2018). "Optimal investment strategies and intergenerational risk sharing for target benefit pension plans." *Insurance: Mathematics and Economics*, **80**, 1–14.

Zhang S, Hardy M, Saunders D (2018). "Retirement Consumption, Risk Perception and Planning Objectives." *Technical report*, Canadian Institute of Actuaries.

Zhu X (2015). *Performance of Dynamic Hedging Strategies for Cash Balance Pension Plans*. Master's thesis, University of Waterloo.

Zhu X, Hardy M, Saunders D (2018a). "Valuation of an early exercise defined benefit underpin hybrid pension." *Scandinavian Actuarial Journal*, **2018**(9), 823–844.

Zhu X, Hardy MR, Saunders D (2018b). "Dynamic Hedging Strategies For Cash Balance Pension Plans." *ASTIN Bulletin: The Journal of the IAA*, **48**(3), 1245–1275.

Appendix A

A.1 Continuous Crediting Based on the k-year spot rate: one-factor HW Model

This section and the following three sections provide formulae for valuation factors. Detailed proofs can be found in Hardy *et al.* (2014) and Zhu (2015). For CB plans with the k-year spot rate as the continuous crediting rate, the valuation factor using the one-factor HW model is

$$\begin{aligned} V(t,T) &= E_t^Q\left[e^{\int_0^T r_k(s)ds - \int_t^T r(s)ds + m(k)T}\right] \\ &= e^{\int_0^t r_k(s)ds} e^{-\int_t^T \frac{A(s,s+k)}{k} ds} P_\gamma(t,T) e^{m(k)T} \end{aligned}$$

where

$$
\begin{aligned}
P_\gamma(t,T) &= \exp\{A_\gamma(t,T) - B_\gamma(a,T-t)r(t)\} \\
A_\gamma(t,T) &= \gamma \log\left\{\frac{P^M(0,T)}{P^M(0,t)}\right\} + \left\{\frac{\sigma^2\gamma}{4a}\{\frac{2}{a}(\gamma-1)(T-t) + (e^{-2at}-\gamma)B^2(a,T-t) \right. \\
&\qquad \left. + \frac{1}{a}(2-2\gamma)B(a,T-t)\} + \gamma f^M(0,t)B(a,T-t)\right\} \\
B_\gamma(a,T-t) &= \gamma B(a,T-t) \\
B(a,k) &= \frac{1-e^{-ak}}{a} \\
A(t,t+k) &= \log\frac{p(0,t+k)}{p(0,t)} + f(0,t)B(a,k) - \frac{\sigma^2}{4a}B(a,k)^2(1-e^{-2at}) \\
\gamma &= 1 - \frac{B(a,k)}{k}
\end{aligned}
$$

The price of zero coupon bond $P(t,T)$ can be regarded as a special case such that $P(t,T) = P_{\gamma=1}(t,T)$.

A.2 Discrete Crediting Based on the k-year spot rate: one-factor HW Model

A CB plan with the k-year spot rate as the crediting rate and crediting frequency n times per year, has the valuation factor using the one-factor HW model

$$V(t,T) = e^{\sum_{s=1}^{tn} \frac{r_k(\frac{s}{n})}{n}} e^{A(t,T)-B(a,T-t)r(t)} e^{EH(t,T,n)+\frac{1}{2}VH(t,T,n)} \tag{A.1}$$

where

$$
\begin{aligned}
EH(t,T,n) = &\sum_{i=tn+1}^{Tn} \frac{-A\left(\frac{i}{n}, \frac{i}{n}+k\right)}{nk} \\
&+ \frac{B(a,k)}{nk} \sum_{i=tn+1}^{Tn} \left[r(t)e^{-a\left(\frac{i}{n}-t\right)} + \varphi\left(\frac{i}{n}\right) - \varphi(t)e^{-a\left(\frac{i}{n}-t\right)} \right. \\
&\left. - \frac{\sigma^2}{a^2}\left(1 - \frac{1}{2}e^{-a\left(T-\frac{i}{n}\right)} - e^{-a\left(\frac{i}{n}-t\right)} + \frac{1}{2}e^{-a\left(T+\frac{i}{n}-2t\right)}\right)\right] \\
VH(t,T,n) = &\left(\frac{\sigma B(a,k)}{nk}\right)^2 \sum_{j=tn+1}^{Tn} \left(\frac{e^{-\frac{a}{n}(j-1)} - e^{-aT}}{e^{\frac{a}{n}} - 1}\right)^2 \frac{e^{2a\frac{j}{n}} - e^{2a\frac{j-1}{n}}}{2a} \\
\varphi(t) = &f^M(0,t) + \frac{\sigma^2}{2a^2}(1 - e^{-at})^2
\end{aligned}
$$

$f^M(0,t)$ is the observed instantaneous forward rate at time t, and $A(t,T)$, $B(a,k)$ are defined in the previous section.

A.3 Continuous Crediting Based on the k-year spot rate: two-factor HW Model

For a CB plan with the k-year spot rate as the continuous crediting rate, the valuation factor using the two-factor HW model is

$$
\begin{aligned}
V(t,T) = &\exp\left\{\int_0^t r_k(s)ds\right\} \exp\{m(k)T\} \exp\left\{-\int_t^T \frac{A^G(s,s+k)}{k}ds\right\} \\
&\exp\left(A^*(t,T) - \gamma(a_1,k)B(a_1,T-t)x(t) - \gamma(a_2,k)B(a_2,T-t)y(t)\right)
\end{aligned}
$$

where

$$
\begin{aligned}
A^G(t,T) &= \log \frac{p(0,T)}{p(0,t)} + \frac{1}{2}(\nu(T-t) + \nu(t) - \nu(T)) \\
\nu(k) &= \frac{\sigma_1^2}{a_1^2}(k - 2B(a_1,k) + B(2a_1,k)) + \frac{\sigma_2^2}{a_2^2}(k - 2B(a_2,k) + B(2a_2,k)) \\
&\quad + \frac{2\rho\sigma_1\sigma_2}{a_2 a_2}(k - B(a_1,k) - B(2a_2,k) + B(a_1+a_2,k)) \\
B(a,k) &= \frac{1}{a}(1 - e^{-ak}) \\
A^*(t,T) &= \log \frac{p(0,T)}{p(0,t)} + \frac{1}{2}(\nu^*(T-t) + \nu(t) - \nu(T)) \\
\nu^*(k) &= \frac{\gamma(a_1,k)^2\sigma_1^2}{a_1^2}(k - 2B(a_1,k) + B(2a_1,k)) + \frac{\gamma^2(a_2,k)\sigma_2^2}{a_2^2}(k - 2B(a_2,k) + B(2a_2,k)) \\
&\quad + \frac{2\rho\gamma(a_1,k)\gamma(a_2,k)\sigma_1\sigma_2}{a_1 a_2}(k - B(a_1,k) - B(a_2,k) + B(a_1+a_2,k)) \\
\gamma(a_j,k) &= 1 - \frac{B(a_j,k)}{k}, \qquad j = 1,2
\end{aligned}
$$

A.4 Discrete Crediting Based on the k-year spot rate: two-factor HW Model

The CB plan with the k-year spot rate as the crediting rate and crediting frequency n times per year, has the valuation factor using the two-factor HW model

$$
V(t,T) = e^{\sum_{s=1}^{tn} \frac{r_k(\frac{s}{n})}{n}} e^{A^G(t,T) - B(a_1,T-t)x(t) - B(a_2,T-t)y(t)} e^{EG(t,T,n) + \frac{1}{2}VG(t,T,n)}
$$

where

$$
\begin{aligned}
EG(t,T,n) = \sum_{i=tn+1}^{Tn} & \frac{-A^G\left(\frac{i}{n}, \frac{i}{n}+k\right)}{nk} + \frac{B(a_1,k)}{nk}\left[x(t)e^{-a_1(\frac{i}{n}-t)} - M_x^T(t,\frac{i}{n})\right] \\
& + \frac{B(a_2,k)}{nk}\left[y(t)e^{-a_2(\frac{i}{n}-t)} - M_y^T(t,\frac{i}{n})\right] \\
VG(t,T,n) = & \left(\frac{\sigma_1 B(a_1,k)}{nk}\right)^2 \sum_{j=tn+1}^{Tn}\left(\frac{e^{-\frac{a_1}{n}(j-1)} - e^{-a_1 T}}{e^{\frac{a_1}{n}}-1}\right)^2 \frac{e^{2a_1\frac{j}{n}} - e^{2a_1\frac{j-1}{n}}}{2a_1} \\
& + \left(\frac{\sigma_2 B(a_2,k)}{nk}\right)^2 \sum_{j=tn+1}^{Tn}\left(\frac{e^{-\frac{a_2}{n}(j-1)} - e^{-a_2 T}}{e^{\frac{a_2}{n}}-1}\right)^2 \frac{e^{2a_2\frac{j}{n}} - e^{2a_2\frac{j-1}{n}}}{2a_2} \\
& 2\rho\left(\frac{\sigma_1 B(a_1,k)}{nk}\right)\left(\frac{\sigma_2 B(a_2,k)}{nk}\right) \\
& \sum_{j=tn+1}^{Tn}\left(\frac{e^{-\frac{a_1}{n}(j-1)} - e^{-a_1 T}}{e^{\frac{a_1}{n}}-1}\right)\left(\frac{e^{-\frac{a_2}{n}(j-1)} - e^{-a_2 T}}{e^{\frac{a_2}{n}}-1}\right)\frac{e^{(a_1+a_2)\frac{j}{n}} - e^{(a_1+a_2)\frac{j-1}{n}}}{a_1+a_2} \\
M_x^T(s,t) = & \left(\frac{\sigma_1^2}{a_1} + \rho\frac{\sigma_1\sigma_2}{a_2}\right)B(a_1,t-s) - \frac{\sigma_1^2}{a_1}e^{-a_1(T-t)}B(2a_1,t-s) - \frac{\rho\sigma_1\sigma_2}{a_2}e^{-a_2(T-t)}B(a_1+a_2,t-s) \\
M_y^T(s,t) = & \left(\frac{\sigma_2^2}{a_2} + \rho\frac{\sigma_1\sigma_2}{a_1}\right)B(a_2,t-s) - \frac{\sigma_2^2}{a_2}e^{-a_2(T-t)}B(2a_2,t-s) - \frac{\rho\sigma_1\sigma_2}{a_1}e^{-a_1(T-t)}B(a_1+a_2,t-s)
\end{aligned}
$$

Appendix B

B.1 Price of Second Election

Here we provide the pricing formulae for second election option under three scenarios. Milevsky and Promislow (2004) has provide more analytic insights relating to the cost function.

- Stochastic Salary in Continuous Setting:

$$C^s(t, L_t) = \left(L_t c(t^* - t) - t^* L_t b\ddot{a}(T) e^{-r(T-t^*)}\right) + Tb\ddot{a}(T)L_t$$

for which we can find a $t^* = \max(\min(T, t^{'}), t)$, such that $t^{'}$ satisfies

$$c - b\ddot{a}(T)e^{-r(T-t^{'})} - rt^{'}b\ddot{a}(T)e^{-r(T-t^{'})} = 0$$

- Deterministic Salary in Continuous Setting:

$$C^s(t, L_t) = \frac{cL_t}{\mu_L - r}\left(e^{(\mu-r)(t^*-t)} - 1\right) - t^* b\ddot{a}(T) L_t e^{(\mu_L - r)(t^*-t)} e^{-\gamma(T-t^*)} + TL_t e^{(\mu_L-r)(T-t)} b\ddot{a}(T)$$

where $t^* = \min(\max(t, t^{'}), T)$ and $t^{'}$ satisfies

$$c - b\ddot{a}(T)e^{-\gamma(T-t^{'})} - t^{'}(\mu_L - r + \gamma)b\ddot{a}(T)e^{-\gamma(T-t^{'})} = 0$$

- Deterministic Salary in Discrete Setting:

$$C^s(t, L_t)$$

$$= \max_{t^* \in t, t+1, \cdots T} L_t \left(c\frac{1 - e^{(\mu_L - r)(t^*-t)}}{1 - e^{\mu_L - r}} - t^* b\ddot{a}(T) e^{\mu_L(t^*-1-t)} e^{-rt^*} e^{-\gamma(T-t^*)} + Tb\ddot{a}(T)e^{\mu_L(T-1)}e^{-r(T-t)}\right)$$