Tilburg University

Multi-period risk sharing under financial fairness

Bao, Hailong

Publication date:
2016

Document Version
Publisher's PDF, also known as Version of record

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):
Bao, H. (2016). *Multi-period risk sharing under financial fairness*. CentER, Center for Economic Research.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

MULTI-PERIOD RISK SHARING UNDER FINANCIAL FAIRNESS

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan Tilburg University op gezag van de rector magnificus, prof. dr. E.H.L. Aarts, in het openbaar te verdedigen ten overstaan van een door het college voor promoties aangewezen commissie in de aula van de Universiteit op dinsdag 20 december 2016 om 16.00 uur door

Hailong BAO

geboren op 8 november 1988 te Zibo, Shandong province, China.

PROMOTIECOMMISSIE:

PROMOTORES: Prof. dr. J.M. SCHUMACHER
 Prof. dr. E.H.M. PONDS

OVERIGE LEDEN: Prof. dr. A.M.B. DE WAEGENAERE
 Prof. dr. T.E. NIJMAN
 Prof. dr. ir. M.H. VELLEKOOP
 Prof. dr. B.J.M. WERKER

Acknowledgements

At the beginning of 2013, I was still a master student at the University of Amsterdam. I then started an internship, and every day I biked from my home in Amsterdam-Noord to the company in Amsterdam-Zuid, along the charming canals and beautiful streets in the city. One day in April, I received an invitation email for a PhD study from Hans Schumacher, who later became my supervisor. I didn't realize at that time that the next three years in my life have been so colorful and unforgettable.

My first thanks go to Peter Spreij. Peter was a lecturer for my master program, and was also my supervisor for my master thesis. Without his recommendation, it would not have been possible for me to ever get the chance for this PhD project. Besides, Peter has co-organized the Winter School on Mathematical Finance with my current PhD supervisor Hans Schumacher. The winter school has been a nice platform for me to broaden my vision and get inspired from peers' work.

It has been an honor for me to have Hans Schumacher and Eduard Ponds as my supervisors. Hans is a mathematician with vast expertise in many branches of mathematics, while Eduard is an economist and knows a lot of the pension industry. Their guidance has been extremely important. Hans has always been kind, patient and tolerant; I will never forget the meetings we have had in his office, and the grammatical issues he has pointed out in my writing. I feel privileged to be among his last PhD students at Tilburg University. Eduard always speaks to me with smile; he teaches me how to find economic intuitions behind the formulas and how to link theory to practice. He never hesitates to offer help when asked.

My gratitude goes to Tilburg University, which has offered all the relevant facilities for this PhD project. I would like to thank Netspar, Network for Studies on Pensions, Aging and Retirement, for partly funding my PhD project. Netspar also arranges various seminars and workshops, which have been helpful for me to know more about the pension research in the Netherlands. I thank APG, Cardano and Robeco for being the industry partners of this project; they have given us many valuable comments from the perspective of practitioners.

I am grateful to Anja De Waegenaere, Theo Nijman, Michel Vellekoop and Bas Werker for being the members of the committee. They have carefully read my thesis draft, and pointed out numerous issues that need attention. Their comments are indispensable for me to finalize the thesis.

I am indebted to many of my colleagues who have made it pleasant to work on the 6th floor of the Koopmans building. I would like to thank Anja Manders, Lenie Laurijssen and Heidi Ket for being our secretaries and taking care of all kinds of miscellaneous things for me. My gratitude also goes to Yi He, Hong Li, Mario Rothfelder, Samuel Sender, Lei Shu, Yifan Yu, Trevor Zhen, Kun Zheng, Ye Qiu Zheng, Bo Zhou, and many other colleagues around the floor; they have offered me much help and I enjoy the time with them.

It has been a pleasure for me to live in the city of Tilburg. My footprints have spread over many streets, and I can recall almost every single day when I was there – some are with happiness, some with misery, and together they make my life complete. I also appreciate the services offered by Nederlandse Spoorwegen; they have managed to provide us with high-quality train services despite of many unexpected “storingen” and unavoidable “werkzaamheden”.

Finally, I would like to thank my parents Qiuming Bao and Guizhen Deng, and my girlfriend Yifan Zhang, for their unconditional love and support. My parents are ordinary people from a small town in the east coast of China, and they can hardly understand any words in this thesis except my name. It has always been agonizing for them when I, their only child, say goodbye to them and start the journey of studying on the other side of the continent. They have always been supporting me, and I feel so lucky to be their son. The first time I met my girlfriend Yifan was in a lecture room in the A-building. Ever since our relationship, she has been a lovely companion and always by my side, sharing my laughs and tears. My days have been brighter because of her.

Hailong Bao,
September 28, 2016 in Amsterdam.

*"Hundreds and thousands of times, in vain I looked for her in the crowd;
suddenly, I turned around, to where the lights were waning, and there she stood."*

Xin Qiji, Chinese Poet (1140 - 1207)

Contents

Acknowledgements	iii
1 Introduction	1
2 Multi-Period Risk Sharing under Financial Fairness	7
2.1 Introduction	7
2.2 Model Framework	9
2.3 Pareto Efficiency in the Multi-Period Setting	15
2.4 Financial Fairness	23
2.5 Existence and Uniqueness of the PEFF Risk Sharing Rule	25
2.6 A General Algorithm for Finding PEFF Solution	26
2.7 Examples	30
2.7.1 Implementing the Algorithm: the Case of Two-Valued Random Variables	30
2.7.2 Explicit PEFF Solution: the Case of Exponential Utility Function	35
2.8 Concluding Remarks	36
3 Intertemporal Allocation of Investment Risk in the Decumulation Phase of a Collective DC Scheme	39
3.1 Introduction	39
3.2 The Collective DC Pension System	43
3.2.1 General Framework	43
3.2.2 Annuity-Target Profile	43
3.2.3 Aggregate Benefits	44
3.2.4 Variables	45
3.3 The Notion of PEFF and the Mohopeff Approach	47
3.3.1 The Design Model	47
3.3.2 Pareto Efficiency in a Multi-Period Setting	49
3.3.3 Financial Fairness	49
3.3.4 The Mohopeff Approach	54
3.4 The Mohopeff Approach in a Collective DC System: an ALM Study	55

3.4.1	Calibration	56
3.4.2	Mohopeff under Power Utility: an Overview	58
3.4.3	Comparison to an Indexation Strategy	60
3.4.4	Allocation Rules: Using Different Utility Functions	62
3.4.5	Allocation Rules: the Impact of Risk-Aversion Parameter	63
3.4.6	The Impact of the Asset Mix	65
3.4.7	The Impact of the Horizon Length	66
3.5	Tolerance Band and Sustainability	67
3.6	Concluding Remarks	74
4	Multi-Period Investment Return Allocation within a Heterogeneous Col- lective, with Applications to Collective Defined-Contribution Plan De- sign	77
4.1	Introduction	77
4.2	Personal Pension Account and Individual DC Contracts	81
4.3	Personal Pension Account in a Collective System	82
4.3.1	General Framework	83
4.3.2	Two Stages of Allocation	86
4.3.3	Variables and Notations	87
4.4	The PEFF Approach: Allocating Stochastic Investment Returns	89
4.4.1	Single-Period PEFF: Intra-Group Tranching	89
4.4.2	Mohopeff: Intertemporal Risk Allocation	92
4.5	PEFF Approach in a Mini-Pension Fund: an Example	98
4.5.1	Assumptions	98
4.5.2	Calibration	100
4.5.3	Case 1: the Effect of Different Annuity-Target Profiles	100
4.5.4	Comparison with IDC Schemes	102
4.5.5	Case 2: the Effect of Different Risk-Taking Profiles	105
4.5.6	Financial Fairness on a Generational Level: Analysis in $\mathbb{Q}$ World	107
4.6	Concluding Remarks	109
A	Appendices for Chapter 2	113
A.1	Proofs for Section 2.3	113
A.2	Proofs for Section 2.5	125
A.3	Proofs for Section 2.6	132
A.4	Proofs for Section 2.7.2	132
A.5	Tables for Section 2.7.1	135

Bibliography**137**

List of Abbreviations

AIR	Assumed Interest Rate
ALM	Asset Liability Management
BFR	Benchmark Funding Ratio
CARA	Constant Absolute Risk Aversion
CDC	Collective Defined Contribution
CEA	Certainty Equivalent Annuity
DB	Defined Benefit
DC	Defined Contribution
IBE	Intertemporal Balance Equation
IRS	Intergenerational Risk Sharing
Mohopeff	Moving Horizon PEFF
PEFF	Pareto Efficiency and Financial Fairness
PPA	Personal Pension Account

List of Symbols

Key Symbols in Chapter 2

A_n	the total asset at time t_n
b_n	the lower bound of the domain of the utility function u_n
b_p	the lower bound of the domain of the utility function u_p
C_n	the contingent payment paid out at time t_n
F_n	the buffer size at time t_n
h_n	the implied marginal utility function for F_n in the scope of Theorem 2.3.5
I_n	the inverse of the marginal utility function u'_n
I_p	the inverse of the marginal utility function u'_p
R_n	the gross investment return for the buffer between time t_{n-1} to t_n
u_n	the utility function assigned to C_n
u_p	the utility function assigned to F_N in the OEB case
v_n	the ex-ante market value of C_n
v_p	the ex-ante market value of F_p in the OEB case
X_n	the aggregate financial risk into the system between time t_{n-1} to t_n
θ_n	the weight assigned to C_n in the scope of Theorem 2.3.2
ρ	the generic notation for a risk sharing rule

Key Symbols in Chapter 3 and 4

A_τ	the total asset of the pension fund at time τ
$A_{\tau \tau+s}$	the total asset of the pension fund at time $\tau + s$; a local variable within the D-model at time τ
$AT(\cdot)$	the annuity target of a benefit payment of interest
B_τ	the aggregate benefit paid out at time τ
$B_{t,i;\tau}$	the actual annuity payment at time τ for the pensioner who enters at time t and is indexed by i
$B_{\tau \tau+s}$	the aggregate benefit paid out at time $\tau + s$; a local variable within the D-model at time τ
C_τ	the aggregate contribution paid into the pension system at time τ

$C_{t,i}$	the lump-sum contribution paid by the generation entering at time t with the index i
F_τ	the fund size at time τ
$F_{\tau \tau+s}$	the fund size at time $\tau + s$; a local variable in the D-model at time τ
$R_{\tau t,s}$	the gross nominal return implied by the forward rates from time t to s
$S_{\tau t,s}$	the survival probability that, seen at time τ , the generation entering at time t will survive by the time s
t	general notation mostly used to represent the entry time of a generation
X_τ	the gross asset return from time $\tau - 1$ to τ
$u_{\tau \tau+s}$	the utility function assigned to $B_{\tau \tau+s}$
$u_{\tau p}$	the utility function assigned to $F_{\tau \tau+N}$
τ	a general notation to indicate the time of interest

To my parents.

Chapter 1

Introduction

The main topic of this thesis is risk sharing in a multi-period setting under the notion of Pareto efficiency and financial fairness (PEFF). From a utility perspective, Pareto efficiency is fundamental in systems where participants gather to reallocate their risk exposures; from a value perspective, the value of the risk exposures from each participant is required to stay the same before and after risk sharing, as long as the risks are monetary and can be priced.

Pension System Design: a Financial Engineering Perspective

An important motivation for the multi-period PEFF problem is the design of a defined-contribution (DC) pension system. The pension reform in the Netherlands has shown a tendency of transforming the traditional defined-benefit (DB) schemes into defined-contribution ones while collectivity is preserved. A collective DC (CDC) scheme has two important properties. On the one hand, the pension contract is a financial contract. Participants pay contributions in the early stage of life in exchange for benefits when getting old. Though not an necessity in a collective system, it can be an attractive feature that, for each participant, the market values of benefits and contributions are equal as measured at a given point in time. On the other hand, the pension contract is a social contract. Collectivity makes it possible to share risks among both the current and future participants.

From the perspective of a financial contract, the benefit payments are essentially contingent claims on the assets of the pension fund. In this sense, measured at some given point in time, the market value of the benefits under the risk-neutral valuation should be equal to the value of the paid-in contribution. An important aspect of the system design is to determine the payoffs of the benefits as a function of the pension assets. Pareto efficiency, from the perspective of a social contract, is one condition that can be utilized in determining the payoffs of

the benefits. The problem, therefore, is to find a solution to the multi-period risk sharing problem that satisfies both the PE constraint and the FF constraint.

From Single Period to Multiple Periods

The multi-period PEFF problem is a continuation of the PEFF risk sharing problem in a single-period setting, as discussed by Bühlmann and Jewell [14], Gale [19] and Pazdera et al. [33] amongst others. The main difference is that, in the multi-period problem there is a *buffer* that allows risks to be shared with participants from subsequent periods, while in the single-period situation the risk that realizes within a period can only be divided among the participants in that period.

The intertemporal allocation of risks complicates the problem in the sense that there needs to be a balance between the present and future, both in utility terms and in value terms. According to Borch [10], in a single-period problem, the allocation should be done in such a way that the weighted marginal utility of each participant after allocation should be equal. Advancing to the multi-period situations, such equality evolves to the form of the Euler equation where marginal utilities from future participants are in the form of conditional expectations [7] [23]. The fairness constraint requires that the valuation should be done across multiple periods. It will be investigated whether a PEFF solution still exists in such a setting, and whether the PEFF solution is unique if it exists.

From PEFF to Moving-Horizon PEFF

A primary concern before applying the PEFF risk allocation approach to a pension system is the notion of financial fairness. In the case of a pension fund that continues indefinitely into the future, the fairness constraint is not straightforward to formulate in the sense that the market value of a future benefit is time-dependent, and it needs to be specified at which time point the pension contract is seen as financially fair. Two versions of fairness have been discussed. Bovenberg and Mehlkopf [11] consider *ex-ante* fairness where all the generations are treated financially fairly as seen at the start of the system. Teulings and de Vries [43], on the other hand, consider *ex-interim* fairness, where the market values of benefits and contributions are equal for each generation as seen at the time when the generation enters the system. The *ex-interim* fairness is usually a stronger constraint and implies the *ex-ante* fairness.

The two versions of fairness also differ from each other regarding intergenerational risk sharing (IRS). Ex-ante fairness allows much room for risk sharing across generations. However, as it only specifies a value constraint with market values seen at time zero, the ex-ante fairness gives few restrictions on the buffer. The discontinuity problem can happen: the fund may accumulate a significant surplus or deficit in the buffer, and either the existing or the incoming generations will have the incentive to terminate the system. Under the strict ex-interim fairness constraint, participants are treated fairly in value at the time of entry. However, it also squeezes the space of IRS to a large extent, as it indicates that the market value of the intergenerational transfer regarding any generation can only be zero seen at the time of entry of that generation. Therefore, a version of financial fairness that lies between the ex-ante and ex-interim fairness criteria may be appropriate in the concrete situation of pension funds. A moving-horizon version of PEFF allocation, the *Mohopeff*, is one possibility to balance the fairness and IRS.

From Collective DC to System of Personal Pension Accounts

The personal pension account (PPA) system, as proposed by Bovenberg and Nijman [13], is a pension system where personal accounts can be established with some certain degree of collective risk sharing. In the case where investment risks are also shared among participants, the PPA system needs also to consider the heterogeneity of the participants when allocating the risk. In a PPA system, the principle of PEFF becomes more explicit: setting up personal accounts implies a financial fairness constraint, while the allocation of investment risk should be in line with the risk-taking preferences of the participants.

One possibility to solve the allocation problem is to combine the Mohopeff approach with the single-period PEFF allocation in Pazdera et al. [33]. We then artificially split the whole allocation problem into two sub-problems: allocation across periods (intertemporal allocation), and allocation among participants within a period (intra-group allocation). In such two-stage allocation hierarchy, the Mohopeff is first adopted to give an allocation rule on an aggregate level. Next, the aggregate benefit is distributed among the current pensioners according to their pre-specified risk preferences.

Limitations

The proposed PEFF approach has some limitations that should not be overlooked. First, we investigate the multi-period PEFF allocation with given amounts of risks. In the situation where the risks represent the stochastic investment returns, it may be desirable to include investment decisions which essentially take the distributions of the risks as decision variables as well. Pazdera et al. [32] have investigated the possibility of including investment decisions in a single-period environment. The same problem in a multi-period setting is not addressed in this thesis and may be a future topic of interest.

The drawbacks of the Mohopeff approach include that a moving-horizon structure is an *ad hoc* element and requires extra inputs, such as the horizon length, that may be decided at the discretion of the pension fund. Although aggregating the benefit payments in the same period is a convenient way to simplify the problem, it is a nonstandard way of aggregation in the scope of the common modeling approaches, where benefits are aggregated in terms of utility for each generation (see e.g. [11] [15] [23]).

Structure of the Thesis

The thesis is a compilation of three papers [5], [6] and [4]. The papers are slightly modified to fit as a whole. The first paper [5], as in Chapter 2, establishes from a theoretical perspective the existence and uniqueness of the PEFF allocation solution in a multi-period setting. The second paper [6], as in Chapter 3, adapts the PEFF allocation approach in Chapter 2 into the moving-horizon PEFF approach in a CDC pension system with a more realistic setting. The last paper [4], as in Chapter 4, continues the research in Chapter 3 by taking into accounts the heterogeneity of the individual participants and discussing the intra-group allocation.

In the first paper *Multi-Period Risk Sharing under Financial Fairness*, we start with a general multi-period system where a finite number of agents need to exchange their financial risk exposures. We look for the solutions that are both Pareto efficient regarding the risk preferences of the agents, and financially fair regarding the market value of the positions before and after the risk sharing. A buffer exists and enables the intertemporal capital transfer. Utility functions are used to determine the risk-taking preferences, and a risk-neutral measure is important for determining the allocation. We prove that under general conditions,

there exists a unique allocation solution that is both Pareto efficient and financially fair. An iterative algorithm is then introduced to calculate this rule numerically.

In the second paper *Intertemporal Allocation of Investment Risk in the Decumulation Phase of a Collective DC Scheme*, we adapt the PEFF algorithm in the first paper to a moving-horizon approach and apply it to design the optimal intertemporal allocation of investment risks in a CDC scheme. To incorporate realistic situations, the allocation rule is calculated on a moving-horizon basis in a design model which reflects the current information set and the best estimates. Utility functions specify intertemporal risk preferences, and a moving-horizon version of financial fairness is discussed.

In the last paper *Multi-Period Investment Return Allocation within a Heterogeneous Collective, with Applications to Collective Defined-Contribution Plan Design*, the main focus is to design a method to allocate investment risk within a group of personal pension accounts by further extending the moving-horizon PEFF approach to an individual level. A two-stage allocation structure is adopted: in the first stage, the Mohopeff approach constructs an allocation rule to determine the aggregate benefit; in the second stage, the aggregate benefit is allocated to individual pensioners in line with their risk preferences. We also show an application of the methodology in an example and provide numerical results.

Chapter 2

Multi-Period Risk Sharing under Financial Fairness¹

2.1 Introduction

This chapter explores the intertemporal risk sharing in a multi-period setting under the notion of *Pareto efficiency and financial fairness* (PEFF). Pareto efficiency means that the utility of nobody can be improved without hurting the utility of some others, while financial fairness dictates that the market values of the risk positions before and after risk sharing should be equal. A risk-sharing system with respect to monetary uncertainties – the stochastic returns from the financial market, for instance – can be viewed as a financial contract. On the one hand, Pareto efficiency is fundamental in risk-sharing systems, while on the other hand financial fairness is important in the design of financial contracts.

The model is motivated and abstracted from systems that allow for intertemporal risk sharing. One example is the collective defined-contribution pension systems which can be viewed as a financial contract among both current and future cohorts. The possibility of intertemporal risk sharing with respect to investment risk is due to the incompleteness of the market, i.e. the inability of generations to be exposed to risks outside their own (mature) lifespan. A risk-sharing system tries to partly fix this problem by allowing later generations to take risks before they become participants. Risk sharing can result in welfare gains to the generations; meanwhile, the pension contract should also be fair from a valuation perspective. Another example is the reinsurance market, in which insurance companies reallocate the risks by way of reinsurance contracts among themselves. A multi-period contract is appropriate for dealing with long-term risks, or simply when companies agree to make multi-period arrangements. A similar example

¹For the original paper, see Bao et al. [5]. Some modifications are made to make the paper fit in the entire thesis.

is the design of structured derivatives, for instance, the practice of tranching. In these examples, Pareto efficiency is pertinent for designing the optimal allocation of risks, while financial fairness guarantees that the contract is fairly priced.

The characterization of Pareto efficient solutions in a single-period setting is well studied in quite a lot of papers, which date back to the 1960s with the focus mainly on the field of insurance. For instance, Borch [10] gives a characterization of the Pareto efficient solutions under the situation where expected utility is used to describe the agents' risk preferences, and later DuMouchel [17] gives proof to these results. Similar work also includes Raviv [36] which takes into consideration the existence of market frictions. The fairness criterion is first considered alongside the Pareto efficiency by, amongst others, Gale [19], Bühlmann and Jewell [14] and Balasko [2] in different settings. In these literature, the risk sharing is built over both a utility basis and a valuation basis.

The risk-sharing problem in a multi-period setting is investigated by Barrieu and Scandolo [7] in a general setting; they talk about risk exchanges between two agents over more than one period without taking into consideration any fairness conditions. Other work has been mainly focused on the design of pension systems and the space of intergenerational risk sharing, where risk redistribution can be organized among both the existing and future cohorts. Pareto-efficient risk sharing can be achieved by maximizing the aggregate expected utility of all generations in the situation where a social planner is present (e.g. Gordon and Varian [21], Gollier [20], Bovenberg and Mehlkopf [11]) or by looking for an equilibrium (see Ball and Mankiw [3], Krueger and Kubler [28]). Financial fairness has been considered by Cui et al. [15]; however, the valuation approach is only used to check afterwards whether the distribution rule is fair for the participants. Kleinow and Schumacher [26] analyze the pension system with conditional indexation from the perspective of market value; they investigate whether the pension contract is financially fair for existing and incoming cohorts as well as the sponsor. Risk-neutral valuation becomes essential in Bovenberg and Mehlkopf [11] to determine a unique risk sharing solution by setting the ex-ante market values of the intergenerational transfers to zero.

This chapter explores the Pareto efficient and financially fair risk sharing in a multi-period environment. Expected utility is adopted to evaluate the welfare, and a risk-neutral measure works for the valuation purpose. We shall show the existence and uniqueness of the PEFF solution, and give a numerical algorithm to find it. This can be seen as a direct generalization of the research by Pazdera et al. [33], which explores the Pareto efficient and financially fair risk-sharing rule

in a single-period case. Compared to Barrieu and Scandolo [7], we restrict ourselves to the case of expected utility as the preference functional, and risk-neutral valuation is used to determine a unique solution. Different from Bovenberg and Mehlkopf [11], no parameterization on the risk-sharing rules is needed here; the rules are determined totally under the notion of PEFF. Mathematically, our results resemble the famous consumption-savings model for intertemporal substitution to some extent. The intertemporal balance equation, as we call it, has a close relationship with the Euler equation in the intertemporal substitution theory; see Hall [23]. The main difference is that the model here introduces no subjective discount factor for impatience. The characterization of Pareto efficiency leads to a weighted optimization problem where the weights are unknowns to be determined uniquely by the financial fairness constraints, making use of a risk-neutral measure.

The rest of the chapter is structured as follows. The model setting is set up in Section 2.2 and we formulate the problem of finding PEFF solutions mathematically. Next we establish the existence and uniqueness of the solution in Section 2.5. Explicit solution exists when we assume exponential utility functions to all the agents and deterministic asset returns; other than that, there appears to be no hope for an explicit solution. We then develop an iterative algorithm to numerically find the solution. The case of the explicit solution is dealt with in Section 2.7; besides, we also give a simple example where the numerical algorithm is implemented. Some remarks conclude the chapter in Section 2.8.

2.2 Model Framework

We assume a finite discrete-time system in which a finite number of agents gather to share their risks. As a result of the risk sharing, the agents expect to receive contingent payments from the system. Each agent is assumed to get one single contingent payment. The term “contingent payment” is general and can have various interpretations in different circumstances. For instance, it can refer to the risk exposure of an insurance company after risk sharing in the case of a reinsurance contract, or the investment risk in the case of a collective pension fund. Alongside there is also a long-lived *buffer* which makes the intertemporal money transfer possible.

The system starts at time t_0 . Assume that altogether there are N contingent payments happening at time $t_1 \leq t_2 \leq \dots \leq t_N$, where N is some positive integer. C_n will stand for the contingent payment paid out from the system at time

FIGURE 2.1: The risk sharing system

t_n ². Let F_n be the buffer size at time t_n . X_n denotes the financial risk coming into the system from the agents from time t_{n-1} to t_n , that is, it is the sum of all the stochastic cash inflows from the agents from time t_{n-1} to t_n . The risk stream $X = (X_1, \dots, X_N)$ is defined in a financial market in which prices are given exogenously. The buffer is invested in a risky asset R which produces stochastic per-dollar gross return R_n from time t_{n-1} to t_n . Here the C_n 's and F_n 's are decision variables, and the X_n 's and R_n 's are the risks to be shared.

The X_n 's and R_n 's are random variables defined on a finite probability space $(\Omega, \mathcal{F}, \mathbb{P})$ where $\mathbb{P}$ is the objective measure. $\mathbb{F}$ is the filtration generated by the X 's and R 's:

$$\mathbb{F} = \{\mathcal{F}_n | n = 1, \dots, N\}, \quad \mathcal{F}_n = \sigma\{(X_1, R_1), \dots, (X_n, R_n)\}.$$

There is also a risk-neutral measure $\mathbb{Q}$ defined on the probability space besides the objective measure $\mathbb{P}$. There is no need to assume the completeness of the market; any given risk-neutral measure $\mathbb{Q}$ will suffice. The only assumption is that the agents have agreed to adopt some probability measures $\mathbb{P}$ and $\mathbb{Q}$, or the measures are simply specified in a situation where a social planner is present. Let

$$\mathbb{E}_n^{\mathbb{P}}[\cdot] = \mathbb{E}^{\mathbb{P}}[\cdot | \mathcal{F}_n].$$

It is assumed that (X_t, R_t) and (X_s, R_s) are independent for $t \neq s$ under $\mathbb{P}$ and $\mathbb{Q}$. For $n = 1, \dots, N$, the random variables X_n and R_n need not be independent, and their joint distribution is known. As we are working on a finite probability space, the total number of outcomes of (X_n, R_n) is finite for all n . Illustrated by Figure 2.2, the risks can be seen as a multinomial tree and every pair (X_n, R_n) can

²The notations are adapted to the level of generality of the chapter; the same notation can have different meanings in different chapters. In this chapter, the variable C is a decision variable; the letter " C " may stand for "contingent payment" or "consumption". In the next two chapters, the letter " C " will stand for "contribution" which is an input variable. We will then use the variable B as the notation for the decision variables. " B " stands for "Benefit" and is more convenient in a pension setting. A list of symbols is given at the beginning of the thesis.

FIGURE 2.2: The first two periods of the multinomial tree for the risks

then be totally characterized by

$$\left\{ ((X_n^j, R_n^j), \mathbb{P}(j), \mathbb{Q}(j)) \mid j = 1, \dots, m_n \right\}$$

where (X_n^j, R_n^j) represents all the possible and distinct values of (X_n, R_n) and $\mathbb{P}(j), \mathbb{Q}(j)$ are the corresponding $\mathbb{P}$ - and $\mathbb{Q}$ -probabilities. A technical requirement is that for any $n = 1, \dots, N$

$$\mathbb{Q}(\{\omega \in \Omega \mid X_n(\omega) = \max X_n, R_n(\omega) = \max R_n\}) > 0, \quad (2.2.1)$$

which means that it is possible for X_n and R_n to attain their maximum under $\mathbb{Q}$ simultaneously. This requirement shall be used in the proofs later. Furthermore we assume that $R_n > 0$ for all n as the R 's have the interpretation as the gross return of the asset R .

We write $J_n = j_1 j_2 \dots j_n$ to represent the trajectory $((X_1^{j_1}, R_1^{j_1}), \dots, (X_n^{j_n}, R_n^{j_n}))$. Let $\mathcal{J}_n$ be the set of all the possible trajectories of (X, R) up to time t_n . $J_n j_{n+1}$ will denote any trajectory whose up-to-time- t_n part is J_n . In such a situation we write $j_{n+1} \in \mathcal{J}_n^{n+1}$ where $\mathcal{J}_n^{n+1}$ denotes the set of all the possible cases of (X_{n+1}, R_{n+1}) .

The risk-neutral measure $\mathbb{Q}$ is used to price the risks X as well as the investment returns R . In this generic setting, write

$$x_n := \mathbb{E}^{\mathbb{Q}} X_n, \quad 1 + r_n := \mathbb{E}^{\mathbb{Q}} R_n, \quad n = 1, \dots, N.$$

The x_n 's are the (ex-ante) market prices of the risks X and the r_n 's are the risk-free returns implied by the pricing measure $\mathbb{Q}$. Please note that now and later we directly work with future values for convenience.

Note that the time points $\{t_0, t_1, \dots, t_N\}$ need not be equidistant. As shown in Figure 2.1, two or more time points can be equal if there are more than one contingent payment paid out at the same time. In that case, say $t_{n-1} = t_n$ for some n , we shall have $X_n \equiv 0$ and $R_n \equiv 1$, because there will be no risks coming in and the buffer will not evolve with respect to asset return.

The utilities of the agents depend solely on the contingent payments they receive. The utility function $u_n(\cdot)$ will be used to evaluate the contingent payment C_n . The function $u_n(x)$ is defined on $x \in (b_n, +\infty)$, where b_n is a constant, either a finite real number (e.g. shifted power utility) or $-\infty$ (e.g. exponential utility). These utility functions are *stereotype utility functions* defined as follows:

1. they are continuous and differentiable;
2. they are strictly concave;
3. the marginal utilities satisfy the *Inada conditions*

$$\lim_{x \downarrow b_n} u'_n(x) = +\infty, \quad \lim_{x \rightarrow \infty} u'_n(x) = 0.$$

For any agent, define $I_n = (u'_n)^{-1}$, which is the inverse function of the marginal utility function. Since u'_n satisfies the Inada conditions, we know that I_n is a strictly decreasing function mapping $(0, +\infty)$ into $(b_n, +\infty)$ and is a bijection.

The budget constraints of the system are then straightforward: at each time point, the invested capital will be distributed between the buffer and the current contingent payment, i.e.

$$F_n + C_n = X_n + F_{n-1}R_n \quad n = 1, \dots, N. \quad (2.2.2)$$

The key problem is to determine the decision variables C_n 's and F_n 's along each trajectory. The interpretation is to divide the total amount of available asset in the system between the current payment and the buffer for future payments.

It is assumed without loss of generality that

$$F_0 = 0.$$

The budget constraint is

$$C_1 + F_1 = X_1 + F_0 R_1 := \tilde{X}_1,$$

which suggests that the situation when F_0 is nonzero or even a random variable can always be dealt with by regarding $X_1 + F_0 R_1$ as a new random variable $\tilde{X}_1$.

The buffer size by the end of the system, F_N , will be referred to as the *end buffer*. Depending on whether the end buffer also takes the risks or not, we may have the following two cases:

- *Closed end buffer (CEB) case*: F_N will be a constant, that is, the buffer will only make the intertemporal transfer possible, but it does not take any risks by the end. Without loss of generality we assume

$$F_N = 0.$$

In the situations where F_N is supposed to be a nonzero constant, we can then redefine a new random variable $\tilde{X}_N$ such that

$$C_N = (X_N - F_N) + F_{N-1} R_N := \tilde{X}_N + F_{N-1} R_N.$$

- *Open end buffer (OEB) case*: F_N will be a decision variable just as the C 's. This means that the buffer provider will also participate in the risk sharing besides acting as a vehicle for intertemporal transfer. In this case, a stereotype utility function u_p will be employed to evaluate the utility of F_N . The function $u_p(x)$ is defined on $x \in (b_p, +\infty)$, and b_p can be either a finite constant or equal to $-\infty$.

It is worth mentioning that there is no explicit constraint on the interim status of the buffer F_n , $n = 1, \dots, N-1$, thus in general they can be positive or negative.

It can be argued as follows that any OEB case can always be converted into a CEB case. For any OEB case $(C_1, \dots, C_N, F_N)$ with utility functions $(u_1, \dots, u_N, u_p)$, we define a new time point $t_{N+1} := t_N$ with $X_{N+1} := 0$ and $R_{N+1} := 1$. The OEB setting is thus formulated into a CEB one with an extra contingent payment C_{N+1} with utility u_p

$$C_{N+1} = X_{N+1} + F_N R_{N+1} = F_N.$$

On the other hand, any CEB setting can be turned into an OEB setting in the sense of Pareto efficiency as we shall see later. In this chapter we will proceed mainly with the OEB setting. The utility of the end buffer F_N will be evaluated according to the utility function u_p .

We will try to determine the C 's and the F 's. For any $n = 1, \dots, N$, both F_n and C_n are by nature $\mathcal{F}_n$ -measurable random variables. We then have the following important definition.

DEFINITION 2.2.1 (*Risk-sharing rule.*) A vector of random variables $(C_1, C_2, \dots, C_N, F_N)$ is called a *risk-sharing rule* if it satisfies

- the measurability condition: $C_n \in \mathcal{F}_n$ for $n = 1, \dots, N$ and $F_N \in \mathcal{F}_N$,
- the budget constraints (2.2.2), and
- the domain requirements of the utility functions, i.e. $C_n > b_n$ for all n and $F_N > b_p$ along any trajectory.

One last thing to mention in this section is that the budget constraints (2.2.2) imply a single *global budget constraint* by eliminating the F 's:

$$\sum_{n=1}^{N-1} \left[C_n \left(\prod_{i=n+1}^N R_i \right) \right] + C_N + F_N = \sum_{n=1}^{N-1} \left[X_n \left(\prod_{i=n+1}^N R_i \right) \right] + X_N. \quad (2.2.3)$$

This implies that in order to make the problem well-posed, one needs to have that

$$\sum_{n=1}^{N-1} \left[b_n \left(\prod_{i=n+1}^N R_i \right) \right] + b_N + b_p < \sum_{n=1}^{N-1} \left[X_n \left(\prod_{i=n+1}^N R_i \right) \right] + X_N.$$

This should hold for any realizations of X and R as we now have a finite probability space. Otherwise there will be no possible risk-sharing rules as the domain requirements of the utility functions can never be satisfied.

EXAMPLE 2.2.2 (*Possible variations of the model.*) The budget constraint (2.2.2) shows that the model is very general and can handle different risk sharing systems. Special cases include

- if we let

$$\begin{aligned} t_1 &= t_2 = \dots = t_N \\ X_2 &= \dots = X_N \equiv 0 \\ R_2 &= \dots = R_N \equiv 1 \end{aligned}$$

then the system degenerates to a single-period problem as in Pazdera et al. [33] and the budget constraint becomes

$$\sum_{n=1}^N C_n + F_N = X_1$$

where X_1 represents the aggregate risk to be shared.

- If we only let

$$X_2 = \dots = X_N \equiv 0$$

then this represents a *decumulation system* where the only cash inflow X_1 will be distributed into several contingent payments in the future. The budget constraint can be written as

$$X_1 = C_1 + \frac{C_2}{R_1} + \frac{C_3}{R_1 R_2} + \dots + \frac{C_N + F_N}{\prod_{i=1}^{N-1} R_i} \quad (2.2.4)$$

This case will be discussed later in the chapter with the connection to the assumed interest rate.

- A *defined-contribution pension fund* in the form of a non-overlapping generations model can be modeled by modifying the budget constraint to

$$F_n + C_n = (Y_{n-1} + F_{n-1})R_n \quad n = 1, \dots, N,$$

where the Y 's are the contributions paid into the system by the beginning of each period, the C 's are the benefits paid out from the system by the end of each period and the R 's now represent the returns from a fixed asset mix where the fund invest its capital.

2.3 Pareto Efficiency in the Multi-Period Setting

This section deals with the concept of Pareto efficiency in this multi-period setting, which is the first step to look for a PEFF risk-sharing rule. We shall characterize parametrically all the PE solutions, among which we look for the one that is also financially fair in the following sections.

It may be convenient to introduce first some notations. Let $\mathbb{R}_+^{N+1}$ be the non-negative cone in $\mathbb{R}^{N+1}$, i.e. $\mathbb{R}_+^{N+1} := \{\theta \in \mathbb{R}^{N+1} | \theta_i \geq 0\}$, and define $\mathbb{R}_{++}^{N+1} := \{\theta \in \mathbb{R}^{N+1} | \theta_i > 0\}$ as the strictly positive cone. For two vectors $a = (a_1, \dots, a_N)$,

$b = (b_1, \dots, b_N)$, $a, b \in \mathbb{R}^N$, we write $a \succeq b$ if $a_n \geq b_n$ for all $n = 1, \dots, N$ and there exists some $m = 1, \dots, N$ such that $a_m > b_m$.

For simplicity we write $X := (X_1, \dots, X_N)$ and $R := (R_2, \dots, R_N)$ which are vectors of random variables on Ω . Write $\rho := (C_1, C_2, \dots, C_N, F_N)$ as the generic notation for a risk-sharing rule and the set of all the possible ρ 's is denoted by $\mathcal{RS}$. We will be particularly interested in the subset $\mathcal{P} \subset \mathcal{RS}$ which is the set of all Pareto-efficient risk-sharing rules. First we need the following definition.

DEFINITION 2.3.1 (*Multi-period Pareto efficiency.*) A risk-sharing rule $(C_1, C_2, \dots, C_N, F_N)$ is called *Pareto efficient*, or *Pareto optimal*, if there does not exist another risk-sharing rule $(\tilde{C}_1, \tilde{C}_2, \dots, \tilde{C}_N, \tilde{F}_N)$ such that

$$\left(\mathbb{E}^{\mathbb{P}} u_1(\tilde{C}_1), \dots, \mathbb{E}^{\mathbb{P}} u_N(\tilde{C}_N), \mathbb{E}^{\mathbb{P}} u_p(\tilde{F}_N) \right) \succeq \left(\mathbb{E}^{\mathbb{P}} u_1(C_1), \dots, \mathbb{E}^{\mathbb{P}} u_N(C_N), \mathbb{E}^{\mathbb{P}} u_p(F_N) \right).$$

We then have the following important theorem in this discrete probability space, which can be seen as a generalization of the Borch-type characterization of the Pareto efficiency: every Pareto-efficient risk-sharing rule can be totally characterized by optimizing a weighted time-additive aggregate utility.

THEOREM 2.3.2 (*Characterization of Pareto efficiency.*) For a risk-sharing rule $(C_1, C_2, \dots, C_N, F_N)$, the following statements are equivalent.

1. The risk-sharing rule is Pareto efficient.
2. The risk-sharing rule maximizes

$$\mathbb{E}^{\mathbb{P}} \left[\sum_{n=1}^N \theta_n u_n(C_n) + \theta_p u_p(F_N) \right] \quad (2.3.1)$$

for some strictly positive constants $\theta_1, \dots, \theta_N, \theta_p$.

3. The risk-sharing rule satisfies the following which are hereafter called the intertemporal balance equations (IBEs) for some strictly positive constants $\theta_1, \dots, \theta_N, \theta_p$:

$$\begin{aligned} \theta_n u'_n(C_n) &= \theta_{n+1} \mathbb{E}_n^{\mathbb{P}} [u'_{n+1}(C_{n+1}) R_{n+1}] \quad \text{for } n = 1, \dots, N-1, \\ \theta_N u'_N(C_N) &= \theta_p u'_p(F_N). \end{aligned}$$

PROOF See appendix. □

REMARK 2.3.3 (*Link to Borch [10].*) Consider the case when $t_n = t_{n+1}$ for some n . We must have that $X_{n+1} \equiv 0$ and $R_{n+1} \equiv 1$. Thus $\mathcal{F}_n = \mathcal{F}_{n+1}$ and the IBE becomes

$$\theta_n u'_n(C_n) = \theta_{n+1} \mathbb{E}_n^{\mathbb{P}} [u'_{n+1}(C_{n+1}) R_{n+1}] = \theta_{n+1} u'_{n+1}(C_{n+1}).$$

This means that in a single period setting, the IBEs will coincide with the characterization of PE risk-sharing rules by Borch [10].

REMARK 2.3.4 (*Comparison to the Euler equation.*) The IBEs are very similar to the famous Euler equation derived amongst others by Hall [23] for solving the consumption-savings model. In fact, the model setting in this chapter can be used as a life-cycle model. If we let the time points $\{t_n\}$ be equispaced and set $R_n = 1 + r$ and $u_n = u$ for all n , then the model setting is also similar to Hall's: every period there is a stochastic earning and a consumption, which correspond to the incoming "risk" and the "contingent payment" in this setting.

The optimization targets are different regarding weighing intertemporally the utilities: Hall assumed a single rate of subjective time preference δ while the IBEs are parameterized by weight vector $\theta := (\theta_1, \dots, \theta_N, \theta_p)$.

Formula-wise, Hall gave

$$\mathbb{E}_n u'(C_{n+1}) = \left(\frac{1 + \delta}{1 + r} \right) u'(C_n),$$

while the IBE gives

$$\mathbb{E}_n u'(C_{n+1}) = \left(\frac{\frac{\theta_n}{\theta_{n+1}}}{1 + r} \right) u'(C_n).$$

It is obvious that Hall adopts a specific set of weights in the scope of Theorem 2.3.2. As we shall see later, the weights θ can be seen as unknowns within the framework here and will be determined endogenously by the financial fairness constraint. The interpretation is that, regarding the intertemporal substitution, Hall adopts a single subjective discount factor while in the PEFF framework the discount curve is determined by the market values of the consumption.

The theorem shows that it is equivalent to solve the optimization problem (2.3.1) subject to the budget constraints when one wants to find the corresponding PE risk-sharing rule given any $\theta \in \mathbb{R}_{++}^{N+1}$. We can then construct a mapping to compute the PE solution given any $\theta \in \mathbb{R}_{++}^{N+1}$, which we will call $\Phi : \mathbb{R}_{++}^{N+1} \rightarrow \mathcal{P}$. This can be done by solving the corresponding parameterized optimization

problem of time-additive utility functions:

$$\begin{aligned} & \max_{C_1, \dots, C_N} \mathbb{E}^{\mathbb{P}} \left[\sum_{n=1}^N \theta_n u_n(C_n) + \theta_p u_p(F_N) \right] \\ \text{such that } & F_n + C_n = X_n + F_{n-1} R_n \quad n = 1, \dots, N, \\ & F_0 = 0. \end{aligned}$$

This optimization problem can be solved by dynamic programming. Add in a new time point $t_{N+1} = t_N$, and

$$X_{N+1} \equiv 0, \quad R_{N+1} \equiv 1.$$

Define

$$A_n := X_n + F_{n-1} R_n \quad n = 1, \dots, N+1,$$

which has the interpretation as the total available asset at time t_n to be divided into the current payment and the buffer for later use. Note that by definition $A_{N+1} = F_N$. The A 's are the state variables, the C 's are the decision variables and the X 's and R 's are the risks. Then we shall have the optimization problem formulated as

$$\begin{aligned} & \max_{C_1, \dots, C_N} \mathbb{E}^{\mathbb{P}} \left[\sum_{n=1}^N \theta_n u_n(C_n) + \theta_p u_p(A_{N+1}) \right] \\ \text{such that } & A_{n+1} = X_{n+1} + (A_n - C_n) R_{n+1}, \quad n = 1, \dots, N, \\ & A_1 = X_1. \end{aligned}$$

Proposition 1.3.1 in [8] shows that in order to solve the problem one needs to define the value functions (indirect utility): first for the last period

$$V_{N+1}(A_{N+1}) = \theta_p u_p(A_{N+1}),$$

and then define backwards, for $n = 1, \dots, N$

$$V_n(A_n) = \max_{C_n} \mathbb{E}_n^{\mathbb{P}} [\theta_n u_n(C_n) + V_{n+1}(X_{n+1} + (A_n - C_n) R_{n+1})]. \quad (2.3.2)$$

The final result is presented below. This mapping Φ gives an explicit expression of the risk-sharing rule ρ as a function of the weights θ , which makes it possible to express the financial fairness condition in terms of the weights later in the chapter.

THEOREM 2.3.5 (*The construction of Φ .*) For any given $\theta = (\theta_1, \dots, \theta_N, \theta_p) \in \mathbb{R}_{++}^{N+1}$, the corresponding PE solution $\rho = (C_1, \dots, C_N, F_N)$ is given by

$$A_n = X_n + F_{n-1}R_n \quad n = 1, \dots, N, \quad (2.3.3)$$

$$C_n = I_n \left(\frac{g_n(A_n)}{\theta_n} \right) \quad n = 1, \dots, N, \quad (2.3.4)$$

$$F_n = H_n \left(\frac{g_n(A_n)}{\theta_{n+1}} \right) \quad n = 1, \dots, N-1, \quad (2.3.5)$$

$$F_N = I_p \left(\frac{g_N(A_N)}{\theta_p} \right), \quad (2.3.6)$$

where the functions are defined recursively by

$$\begin{aligned} G_N(x) &:= I_N \left(\frac{x}{\theta_N} \right) + I_p \left(\frac{x}{\theta_p} \right), \\ g_N(x) &:= G_N^{-1}(x), \end{aligned}$$

and for $n = 1, \dots, N-1$

$$h_n(x) = \mathbb{E}_n^{\mathbb{P}} \left[\frac{1}{\theta_{n+1}} g_{n+1}(X_{n+1} + xR_{n+1})R_{n+1} \right] \quad (2.3.7)$$

$$= \mathbb{E}^{\mathbb{P}} \left[\frac{1}{\theta_{n+1}} g_{n+1}(X_{n+1} + xR_{n+1})R_{n+1} \right], \quad (2.3.8)$$

$$H_n = h_n^{-1},$$

$$G_n(x) := I_n \left(\frac{x}{\theta_n} \right) + H_n \left(\frac{x}{\theta_{n+1}} \right),$$

$$g_n(x) := G_n^{-1}.$$

The mapping (2.3.3) - (2.3.6) is denoted as $\Phi : \theta \mapsto \rho, \mathbb{R}_{++}^{N+1} \rightarrow \mathcal{P}$.

PROOF See appendix. Please note that from expression (2.3.7) to (2.3.8) we utilized the assumption that the processes X and R are sequentially independent. $\square$

The functions above have the following interpretation. While u'_n is the marginal utility function of the contingent payment C_n , the function h_n is the implied marginal utility of the buffer F_n and g_n the implied marginal utility of the total available asset A_n . The capital-letter functions I, H, G are the corresponding

inverse functions. The following relationships hold:

$$\begin{aligned} g_n(A_n) &= \theta_n u'_n(C_n) = \theta_{n+1} h_n(F_n), \quad n = 1, \dots, N-1, \\ g_N(A_N) &= \theta_N u'_N(C_N) = \theta_p u'_p(F_N). \end{aligned}$$

The function g 's are also the derivatives of the value functions. The proof in the appendix shows that for any n

$$V'_n(A_n) = g_n(A_n).$$

Write

$$L_n := g_n(A_n),$$

which is interpreted as the weighted marginal utility of the contingent payments. Furthermore, the IBE will be translated into

$$L_n = \mathbb{E}_n^{\mathbb{P}}[L_{n+1} R_{n+1}].$$

The idea of dynamic programming indicates that in each period, the system has to ponder how to distribute the risks between the current contingent payment and all the future contingent payments: for any $n < N$, it compares the marginal utilities of paying out the money now (i.e. C_n) or saving it for the future (i.e. F_n):

$$\theta_n u'_n(C_n) \quad \text{v.s.} \quad \theta_{n+1} h_n(F_n).$$

The h_n function is calculated by “summarizing” the expectations over the future. This property allows us to convert an n -period problem into an *induced* $(n-1)$ -period one, by regarding the time t_{n-1} as the new end of the system and F_{n-1} as the new end buffer with utility h_{n-1} .

This perspective is essential for the proofs later. As a first application, it can help us link the settings of CEB and OEB to each other. First, as we have discussed, any OEB problem can be converted into a CEB problem by regarding F_N as an extra contingent payment C_{N+1} at $t_{N+1} = t_N$. The following result shows that in the sense of Pareto efficiency, the OEB and CEB are equivalent, thus we can work with the two environments interchangeably.

PROPOSITION 2.3.6 (*Equivalence between CEB and OEB problems.*) *The CEB and the OEB are equivalent in the sense that they can always be converted into the form of the other which can produce the identical PE risk-sharing rule.*

PROOF We only need to consider the direction from CEB to OEB. Given a CEB case with PE risk-sharing rule $(C_1, \dots, C_N)$, utility functions $(u_1, \dots, u_N)$ and weights $(\theta_1, \dots, \theta_N)$, we can create a corresponding OEB problem that replicates the original setting for $n = 1, \dots, N - 1$ and truncate the system at time t_{N-1} by defining

$$h(x) := \mathbb{E}_{N-1}^{\mathbb{P}} [u'_N(X_N + xR_N)R_N]$$

as the marginal utility function for the new end buffer F_{N-1} together with weight θ_N . Then according to the IBE for the CEB problem we have

$$\begin{aligned} \theta_{N-1} u'_{N-1}(C_{N-1}) &= \theta_N \mathbb{E}_{N-1}^{\mathbb{P}} [u'_N(C_N)R_N] \\ &= \theta_N \mathbb{E}_{N-1}^{\mathbb{P}} [u'_N(X_N + F_{N-1}R_N)R_N] \\ &= \theta_N h(F_{N-1}) \end{aligned}$$

which matches the final-period IBE in Theorem 2.3.2. Thus according to the theorem the two settings should produce the same PE risk-sharing rules. The only thing left is to verify that the function $h(x)$ defined in this way is indeed a (stereotype) marginal utility function; this has been done in the proof of Theorem 2.3.5. $\square$

There is one degree of freedom extra in determining θ , as for any $c \in \mathbb{R}_{++}$, θ and $c \cdot \theta$ will produce essentially the same optimization target. But if we choose a way of normalizing the θ 's, e.g. restrict the θ 's to the open unit simplex in $\mathbb{R}_{++}^{N+1}$, then we will have the following theorem which indicates that every PE risk-sharing rule $\rho \in \mathcal{P}$ can be uniquely characterized by the weights θ , and the function Φ is a meaningful bijection between all the PE risk-sharing rules ρ 's and the weights θ 's.

THEOREM 2.3.7 *Φ is a one-to-one mapping between the set of all the Pareto efficient risk-sharing rules $\mathcal{P}$ and the open unit simplex in $\mathbb{R}_{++}^{N+1}$, i.e. the set $\mathcal{U} := \{c \in \mathbb{R}_{++}^{N+1} | c_1 + \dots + c_{N+1} = 1\}$.*

PROOF This can be seen as a corollary of Theorem 2.3.2. We discuss the two directions.

1. $\mathcal{U} \rightarrow \mathcal{P}$: the mapping Φ maps any $\theta \in \mathbb{R}_{++}^{N+1}$ into $\mathcal{P}$. This mapping is not injective. Consider some θ and θ' such that $\Phi(\theta) = \Phi(\theta')$. Then we show that there will exist some $c \in \mathbb{R}_{++}$ such that $\theta = c\theta'$ thus Φ is injective if restricted on $\mathcal{U}$.

By the IBEs we know that

$$\frac{\theta_n}{\theta_{n+1}} = \frac{\mathbb{E}_n u'_{n+1}(C_{n+1}) R_{n+1}}{u'_n(C_n)} \quad \text{for } n = 1, \dots, N-1$$

and

$$\frac{\theta_N}{\theta_p} = \frac{u'_p(F_N)}{u'_N(C_N)}.$$

This indicates

$$\frac{\theta_n}{\theta_{n+1}} = \frac{\theta'_n}{\theta'_{n+1}} \quad \text{for } n = 1, \dots, N-1$$

and

$$\frac{\theta_N}{\theta_p} = \frac{\theta'_N}{\theta'_p}.$$

We then have

$$\theta = \frac{\theta_1}{\theta'_1} \theta'.$$

Φ will be an injective mapping if restricted on $\mathcal{U}$.

2. $\mathcal{P} \rightarrow \mathcal{U}$: Theorem 2.3.2 shows that for any element $\rho \in \mathcal{P}$, there exists some $\theta \in \mathbb{R}_{++}^{N+1}$ such that $\Phi(\theta) = \rho$.

We conclude from the above discussion that Φ is both injective and surjective. It must be bijective. $\square$

We conclude this section by some useful properties of the PE risk sharing system. First, we give the following result which seems quite intuitive: every agent will be better off when the realization of the risks is (strictly) better. We call this the monotonicity property of the system with respect to the risks.

LEMMA 2.3.8 (*Monotonicity property of the system with respect to the risks.*) For any $\theta \in \mathbb{R}_{++}^{N+1}$, consider two trajectories $J, J^* \in \mathcal{J}_N$ such that $(X^J, R^J) \succeq (X^{J^*}, R^{J^*})$. Then we have $\rho^J \succeq \rho^{J^*}$.

PROOF See appendix. $\square$

The following result illustrates the impact of the weight θ on the contingent payments: if some weight increases while the others stay the same, then along any trajectory, the corresponding contingent payment will increase while the other contingent payments will decrease.

LEMMA 2.3.9 (*Monotonicity property of the system with respect to the weights.*) Consider two weights $\theta = (\theta_1, \dots, \theta_N, \theta_p), \theta' = (\theta'_1, \dots, \theta'_N, \theta'_p) \in \mathbb{R}_{++}^{N+1}$ such that there

exists some $n = 1, \dots, N, p$ that

$$\theta_n > \theta'_n, \quad \theta_i = \theta'_i \quad \forall i \neq n.$$

Then we have that for any trajectory $J \in \mathcal{J}_N$, the corresponding PE risk-sharing rules satisfy

$$C_n^J > C_n'^J, \quad C_i^J < C_i'^J \quad \forall i \neq n.$$

Here for convenience we let $C_p = F_N$.

PROOF See appendix. □

2.4 Financial Fairness

As we have discussed, the PE risk-sharing rules can be totally characterized by the points on the open unit simplex in $\mathbb{R}_{++}^{N+1}$ and thus there will be infinitely many such PE rules. We will see in the following that the concept of financial fairness will help us narrow down our scope – finally we will arrive at a unique risk-sharing rule that is both PE and FF.

The concept of financial fairness means that when the system starts, for each agent involved, the market value of the risks he contributes into the system should be equal to that of the contingent payments he gets after risk sharing. This is equivalent to say that under the risk-neutral measure $\mathbb{Q}$, for a risk-sharing rule $\rho = (C_1, \dots, C_N, F_N) \in \mathcal{RS}$, the vector

$$\mathbb{E}^{\mathbb{Q}}\rho = (\mathbb{E}^{\mathbb{Q}}C_1, \mathbb{E}^{\mathbb{Q}}C_2, \dots, \mathbb{E}^{\mathbb{Q}}C_N, \mathbb{E}^{\mathbb{Q}}F_N) \in \mathbb{R}^{N+1}$$

is determined and does not change after risk sharing. This vector is called the *value profile*. As before we consider no discounting and we simply use the $\mathbb{Q}$ -expectation as market values.

It is useful to generalize the definition of financial fairness by specifying directly the value profile, i.e. let

$$v = (v_1, v_2, \dots, v_N, v_p) = (\mathbb{E}^{\mathbb{Q}}C_1, \mathbb{E}^{\mathbb{Q}}C_2, \dots, \mathbb{E}^{\mathbb{Q}}C_N, \mathbb{E}^{\mathbb{Q}}F_N), \quad (2.4.1)$$

where the vector v is a vector of constants that we would like the value profile to equal. This definition is more general in the sense that the vector v may not be equal to the corresponding market values of contributions.

The set of all the possible values the value profile can take, $\mathcal{V}$, can only be a restricted subset of $\mathbb{R}^{N+1}$. First note it is trivial that

$$v_n > b_n \quad \text{for } n = 1, \dots, N; \quad v_p > b_p$$

according to the domain requirements of the utility functions. Next, according to the global budget constraint (2.2.3) we shall have, by taking the expectation under $\mathbb{Q}$ to both sides

$$\sum_{n=1}^{N-1} \left[v_n \left(\prod_{i=n+1}^N (1 + r_i) \right) \right] + v_N + v_p = \sum_{n=1}^{N-1} \left[x_n \left(\prod_{i=n+1}^N (1 + r_i) \right) \right] + x_N. \quad (2.4.2)$$

We can then write

$$\mathcal{V} = \left\{ v \in \mathbb{R}^N \mid \text{Eq (2.4.2) holds; } v_n > b_n \text{ for } n = 1, \dots, N; \quad v_{N+1} > b_p \right\} \quad (2.4.3)$$

as the set of all possible value profiles. Note that the set $\mathcal{V}$ is totally determined by the market values of risks and the utility functions.

REMARK 2.4.1 The global budget constraint suggests that for any given value profile vector $v := (v_1, \dots, v_N, v_p)$, we only have to consider any N coefficients. For instance, if the following hold

$$\mathbb{E}^{\mathbb{Q}} C_n = v_n \quad n = 1, \dots, N$$

then

$$\mathbb{E}^{\mathbb{Q}} F_N = v_p$$

will automatically be satisfied.

REMARK 2.4.2 (*Connection to the assumed interest rate.*) The role of assumed interest rate (AIR) in classical life-cycle models has been discussed by Grebentchikova et al. [22] amongst others. The AIR serves to determine how an initial wealth is divided into subsequent periods.

Consider the decumulation case in Example 2.2.2, where the initial wealth X_1 will be divided into a finite number of future payments. In this PEFF framework, taking the expectation under $\mathbb{Q}$ on both sides of the budget constraint (2.2.4), we get

$$x_1 = \mathbb{E}^{\mathbb{Q}} C_1 + \frac{\mathbb{E}^{\mathbb{Q}} C_2}{1 + r_1} + \frac{\mathbb{E}^{\mathbb{Q}} C_3}{(1 + r_1)(1 + r_2)} + \dots + \frac{\mathbb{E}^{\mathbb{Q}}(C_N + F_N)}{\prod_{i=1}^{N-1} (1 + r_i)}.$$

This suggests that in the case where the measure $\mathbb{Q}$ is unique and the rates $\{1+r_i\}$ represent the risk-free rates, one takes the risk-free rates as the AIR to determine how the initial wealth X_1 shall be divided and paid out in several future periods. The difference between the classical models and the PEFF framework is that, in the PEFF framework the initial wealth is divided into future payments in terms of market value.

2.5 Existence and Uniqueness of the PEFF Risk Sharing Rule

The theorems in this section will show that the solution exists and is actually unique if we combine the Pareto efficiency with financial fairness. We continue to work with the general situation when there are N contingent payments alongside the buffer, $N \geq 1$. For any given value profile $v := (v_1, \dots, v_N, v_p) \in \mathcal{V}$, the corresponding PEFF risk-sharing rule is the solution to the following equation system:

1. budget constraints (BCs):

$$F_n + C_n = X_n + F_{n-1}R_n \quad n = 1, \dots, N; \quad (2.5.1)$$

2. intertemporal balance equations (IBEs):

$$\begin{aligned} \theta_n u'_n(C_n) &= \theta_{n+1} \mathbb{E}_n^{\mathbb{P}} [u'_{n+1}(C_{n+1})R_{n+1}] \quad n = 1, \dots, N-1, \\ \theta_N u'_N(C_N) &= \theta_p u'_p(F_N); \end{aligned} \quad (2.5.2)$$

3. financial fairness constraints (FFs):

$$\mathbb{E}^{\mathbb{Q}} C_n = v_n \quad n = 1, \dots, N. \quad (2.5.3)$$

Please note that the BC and IBE equations above are actually equations between functions. The equations should hold for all possible trajectories.

The following theorem is one of the key results of this chapter. It indicates that for the equation system above, the solution always exists and is unique, thus it establishes the existence and uniqueness of the PEFF risk-sharing rule.

THEOREM 2.5.1 (*The existence and uniqueness of the PEFF risk-sharing rule.*) For any given value profile vector $v \in \mathcal{V}$, the PEFF risk-sharing rule exists and is unique. The corresponding θ is unique up to normalization.

PROOF See appendix. □

According to Theorem 2.3.2, the function sets BC and IBE characterize all the possible PE risk-sharing rules by way of weights $\theta \in \mathbb{R}_{++}^{N+1}$. The theorem above then shows that the value profile determines a unique θ .

Recall that in Theorem 2.3.5 Φ defines a bijective mapping from $\mathcal{U}$ to the set of all PE risk-sharing rules $\mathcal{P}$. The mapping Φ then induces a natural mapping Ψ from $\mathcal{U}$ to $\mathcal{V}$: $\Psi(\theta) = \mathbb{E}^Q \Phi(\theta)$. This Ψ links the set of all the possible weights θ and the set of all the possible value profiles.

THEOREM 2.5.2 Ψ is a one-to-one mapping between the set of all possible value profiles $\mathcal{V}$ and the open unit simplex $\mathcal{U}$ in $\mathbb{R}_{++}^{N+1}$.

PROOF Theorem 2.5.1 shows that Ψ is surjective: for any given $v \in \mathcal{V}$ there exists a $\theta \in \mathbb{R}_{++}^{N+1}$ such that $\Psi(\theta) = \mathbb{E}^Q \Phi(\theta) = v$.

This Ψ is also injective restricted on the open unit simplex $\mathcal{U}$ because of the uniqueness of θ up to normalization. Suppose there are $\theta_1, \theta_2 \in \mathcal{U}$ such that $\Psi(\theta_1) = \Psi(\theta_2)$. Theorem 2.5.1 indicates that $\Phi(\theta_1) = \Phi(\theta_2)$, as for each value profile, there will exist exactly one PE risk-sharing rule such that the FF condition is satisfied. According to Theorem 2.3.7, it must be that $\theta_1 = \theta_2$ as they both belong to the open unit simplex $\mathcal{U}$. □

We can then say that the θ uniquely determines the value profile of any PE risk-sharing rule, and also *vice versa*. Instead of talking about the weights θ we can now talk about the value profiles which seem more tangible. However, we cannot say more of the mapping Ψ ; the structure of it can be very complicated depending on the utility functions one uses.

2.6 A General Algorithm for Finding PEFF Solution

Looking for the PEFF risk-sharing rule will come down to solving a system of both linear and non-linear equations. In most cases there's no hope for explicit solutions; fortunately, we have a numerical algorithm that helps to find the PEFF solution.

Recall that

$$L_n = \theta_n u'_n(C_n) \quad n = 1, \dots, N$$

are the weighted marginal utilities of the contingent payments as determined by the risk-sharing rule at time t_n . According to the IBEs

$$L_n = \mathbb{E}_n^{\mathbb{P}}[L_{n+1}R_{n+1}] \quad n = 1, \dots, N-1,$$

thus the whole sequence $\{L_n\}$ is known once L_N is known.

In Theorem 2.3.5 we constructed a mapping $\Phi : \mathbb{R}_{++}^{N+1} \rightarrow \mathcal{P}$ from the sets of functions BC and IBE. Given the mapping Φ , we can deduce another mapping φ_1 by

$$\varphi_1(\theta) = L_N = \theta_N u'_N(C_N) = \theta_N u'_N(\Phi_{(N)}(\theta)),$$

where $\Phi_{(N)}(\cdot)$ stands for the N -th coordinate of this vector-valued function. φ_1 maps any $\theta \in \mathbb{R}_{++}^{N+1}$ into some L_N . For any L_N , another mapping $\varphi_2 : L_N \mapsto \theta$ can be constructed based on the FF constraints: note that according to the mapping Φ we have

$$\begin{aligned} C_n &= I_n \left(\frac{L_n}{\theta_n} \right) \quad n = 1, \dots, N; \\ F_N &= I_p \left(\frac{L_N}{\theta_p} \right), \end{aligned}$$

and

$$L_n = \mathbb{E}_n[L_{n+1}R_{n+1}].$$

This allows us to find a θ' such that the FF conditions are satisfied for the given L_N :

$$\mathbb{E}^{\mathbb{Q}} C_n = \mathbb{E}^{\mathbb{Q}} I_n \left(\frac{L_n}{\theta'_n} \right) = v_n \quad \text{for } n = 1, \dots, N; \quad (2.6.1)$$

$$\mathbb{E}^{\mathbb{Q}} F_N = \mathbb{E}^{\mathbb{Q}} I_p \left(\frac{L_N}{\theta'_p} \right) = v_p. \quad (2.6.2)$$

The function φ_2 is well defined since

$$\mathbb{E}^{\mathbb{Q}} C_n = \mathbb{E}^{\mathbb{Q}} I_n \left(\frac{L_n}{\theta_n} \right) = \sum_{J \in \mathcal{J}_n} \mathbb{Q}(J) I_n \left(\frac{L_n^J}{\theta_n} \right)$$

is a strictly increasing and continuous function in θ_n with $\theta_n \in \mathbb{R}_{++}$. Thus

$$\varphi_{2(n)}(L_N) := \left[\mathbb{E}^{\mathbb{Q}} I_n \left(\frac{L_n}{\cdot} \right) \right]^{-1} (v_n)$$

is well defined. This holds for all $n = 1, \dots, N$ and also for F_N , thus φ_2 is well-defined. Please note that one and only one coordinate of the weight vector θ is solved in every single equation (2.6.1) and (2.6.2).

Consider the composition of the two functions $\varphi = \varphi_2 \circ \varphi_1$: it is a mapping from $\mathbb{R}_{++}^{N+1}$ into itself. Theorem 2.5.1 indicates that there always exists a unique fixed point of this mapping φ , which corresponds to the PEFF risk-sharing rule. The next theorem shows that φ suggests an iterative algorithm for finding the PEFF solution.

THEOREM 2.6.1 (*Feasibility of an iterative algorithm by φ .*) For any given starting point $\theta \in \mathbb{R}_{++}^{N+1}$ with any proper normalization, the sequence of iterates $\{\varphi^{(n)}(\theta) | n \in \mathbb{N}_+\}$ will converge to the fixed point of φ .

PROOF See Appendix. □

Theorem 2.6.1 suggests that starting with any given θ , one first finds the corresponding L_N by φ_1 and then updates the value of θ by φ_2 . It is more convenient, in fact, to use function Φ instead of φ_1 , i.e. we map θ to ρ directly and in the second step we update the θ accordingly. In the first step, we need to calculate numerically the functions g 's and h 's backwards in time, and once all the functions are ready, we then go forwards in time and calculate all the C 's and F 's from the starting distribution X_1 .

ALGORITHM 1 (*Numerical algorithm for finding the PEFF solution.*) The following gives a description of the numerical algorithm for finding the PEFF solution.

1. Start with some initial $\theta^{(0)} \in \mathbb{R}_{++}^{N+1}$.
2. For any given $\theta^{(m)}$ with $m \in \mathbb{N}$, calculate backwards in time that

$$G_N^{(m)}(x) := I_N \left(\frac{x}{\theta_N^{(m)}} \right) + I_p \left(\frac{x}{\theta_p^{(m)}} \right),$$

$$g_N^{(m)}(x) := \left(G_N^{(m)} \right)^{-1}(x),$$

and for $n = 1, \dots, N - 1$

$$\begin{aligned} h_n^{(m)}(x) &= \mathbb{E}^{\mathbb{P}} \left[\frac{1}{\theta_{n+1}^{(m)}} g_{n+1}^{(m)}(X_{n+1} + xR_{n+1}) R_{n+1} \right], \\ H_n^{(m)} &= (h_n^{(m)})^{-1}, \\ G_n^{(m)}(x) &:= I_n \left(\frac{x}{\theta_n^{(m)}} \right) + H_n^{(m)} \left(\frac{x}{\theta_{n+1}^{(m)}} \right), \\ g_n^{(m)}(x) &:= (G_n^{(m)})^{-1}. \end{aligned}$$

3. Calculate the decision variables forwards in time by

$$\begin{aligned} A_n^{(m)} &= X_n + F_{n-1}^{(m)} R_n \quad n = 1, \dots, N, \\ C_n^{(m)} &= I_n \left(\frac{g_n^{(m)}(A_n^{(m)})}{\theta_n^{(m)}} \right) \quad n = 1, \dots, N, \\ F_n^{(m)} &= H_n^{(m)} \left(\frac{g_n^{(m)}(A_n^{(m)})}{\theta_{n+1}^{(m)}} \right) \quad n = 1, \dots, N - 1, \\ F_N^{(m)} &= I_p \left(\frac{g_N^{(m)}(A_N^{(m)})}{\theta_p^{(m)}} \right). \end{aligned}$$

4. Update the θ from $\theta^{(m)}$ to $\theta^{(m+1)}$ by solving

$$\begin{aligned} \mathbb{E}^{\mathbb{Q}} C_n^{(m)} &= \mathbb{E}^{\mathbb{Q}} I_n \left(\frac{g_n^{(m)}(A_n^{(m)})}{\theta_n^{(m+1)}} \right) = v_n \quad n = 1, \dots, N; \\ \mathbb{E}^{\mathbb{Q}} F_N^{(m)} &= \mathbb{E}^{\mathbb{Q}} I_p \left(\frac{g_N^{(m)}(A_N^{(m)})}{\theta_p^{(m+1)}} \right) = v_p. \end{aligned}$$

5. Normalize $\theta^{(m+1)}$.

6. If, for some pre-specified error tolerance ε

$$\begin{aligned} |\mathbb{E}^{\mathbb{Q}} C_n^{(m)} - v_n| &< \varepsilon \quad n = 1, \dots, N, \\ |\mathbb{E}^{\mathbb{Q}} F_N^{(m)} - v_p| &< \varepsilon, \end{aligned}$$

we conclude that $\rho^{(m)} = (C_1^{(m)}, \dots, C_N^{(m)}, F_N^{(m)})$ is the PEFF risk-sharing rule we are looking for. Otherwise, go to step 2 with $\theta^{(m+1)}$.

REMARK 2.6.2 (*Comparison to the algorithm proposed by Pazdera et al. [33].*) As has been mentioned, the framework introduced here can also deal with the single-period situation, which has been investigated by Pazdera et al. [33]. There is a significant difference between the two numerical algorithms, though. The algorithm here makes use of the induction technique that the number of contingent payments is reduced by one recursively, thus in each iteration the algorithm always calculate the functions backwards and then the distributions of the decision variables forwards. In contrast, the algorithm in [33] need not use such an induction technique; functions and decision variables can be calculated simultaneously in each iteration. The algorithm in [33] offers more efficiency for the single-period problem, while the algorithm here is more versatile and can deal with multi-period problems.

2.7 Examples

In this section we give two examples of the PEFF risk sharing. In the first example we implement Algorithm 1 in a simple case where each of the financial risks has only two possible outcomes. The second example deals with exponential utility functions where we may have explicit PEFF solution.

2.7.1 Implementing the Algorithm: the Case of Two-Valued Random Variables

We start with a 3-period setting where three agents gather to share their risks. As shown in Figure 2.3, there are four time points $t = 0, 1, 2, 3$. For $n = 1, 2, 3$, agent n exists between time points $n - 1$ and n . He receives a stochastic income X_n as the risk and he gets C_n as the contingent payment. The risks $\{X_n\}$ are assumed to be independent and identically distributed (i.i.d.), and the distributions of X_n are given by

$$\begin{aligned}\mathbb{P}(X_n = 1.2) &= 0.6, & \mathbb{P}(X_n = 0.8) &= 0.4; \\ \mathbb{Q}(X_n = 1.2) &= 0.5, & \mathbb{Q}(X_n = 0.8) &= 0.5.\end{aligned}$$

In the *autarky* case where the agents are all on their own, agent n will get $C_n = X_n$ and there is no risk sharing among the agents.

We consider the situation when the three agents gather to share their stochastic incomes in a PEFF way. The capital in the buffer is always invested in asset R

FIGURE 2.3: Timeline for Section 2.7.1

with $R_n \equiv 1$ for $n = 1, 2, 3$ for simplicity. The budget constraints are then

$$C_n + F_n = X_n + F_{n-1} \quad n = 1, 2, 3.$$

The FF constraints are

$$\mathbb{E}^{\mathbb{Q}} C_n = 1 \quad n = 1, 2, 3$$

and

$$\mathbb{E}^{\mathbb{Q}} F_3 = F_0.$$

We assume that a buffer is available for the agents with initial capital $F_0 = 1$. The reason for starting with a positive buffer size is that we will later use power utility for F_3 and F_3 is required to be strictly positive, and so is $\mathbb{E}^{\mathbb{Q}} F_3$.

Power utility functions are used to evaluate the utility. We assume that the agents use

$$u_n(x) = \frac{x^{1-\gamma}}{1-\gamma} \text{ with } \gamma = 3, \quad n = 1, 2, 3.$$

We

We will consider both the OEB case where the end buffer is also a decision variable, and the CEB case where the end buffer will be a constant. In the OEB case, power utility with $\gamma = 3$ is also used to evaluate the end buffer.

The IBEs in the OEB case are

$$\begin{aligned} \theta_1 u'_1(C_1) &= \theta_2 \mathbb{E}_1^{\mathbb{P}} [u'_2(C_2)], \\ \theta_2 u'_2(C_2) &= \theta_3 \mathbb{E}_2^{\mathbb{P}} [u'_3(C_3)], \\ \theta_3 u'_3(C_3) &= \theta_p u'_p(F_3). \end{aligned}$$

In the CEB case we only have the first two sets of IBEs since the end buffer size is a constant. In any case, the IBEs have to hold along all the trajectories.

Figures 2.4 and 2.5 show the distributions of the risks, under both the PEFF case and the autarky case. The details of the distributions are shown in the appendix.

FIGURE 2.4: Payments to agents under $\mathbb{P}$, CEB FIGURE 2.5: Payments to agents under $\mathbb{P}$, OEB

The interpretation of the results. In both the CEB case and the OEB case, agent 1 and 2 have effectively shifted some of the volatilities to the last agent, which can be seen from the fact that C_1 and C_2 from the PEFF solution are less dispersed than the autarky situation. Agent 3 can be better off in the best-outcome scenario and worse off in the worst-outcome scenario compared to the autarky case. As a compensation for higher volatility, he benefits from a higher expected return. See Table 2.1. An important feature of the PEFF solution is that by design, the PEFF solution satisfies the FF constraints.

	Expected return			Standard deviation		
	PEFF, CEB	PEFF, OEB	Autarky	PEFF, CEB	PEFF, OEB	Autarky
Agent 1	1.0141	1.0101	1.0400	0.0689	0.0497	0.1960
Agent 2	1.0349	1.0236	1.0400	0.1235	0.0826	0.1960
Agent 3	1.0711	1.0431	1.0400	0.2247	0.1271	0.1960

TABLE 2.1: Comparison of PEFF and autarky solutions: statistics

The difference between the CEB and the OEB case is that in the OEB case, F_3 can also absorb some risks. The results are lower expected returns under $\mathbb{P}$ and

smaller standard deviations for the payments C_n compared to the CEB case. Note that C_3 and F_3 have identical distributions. This is because they are assigned the same utility functions and the same ex-ante market values.

Individual rationality. The results also shed some light on the issue of *individual rationality*, which says that if the agents are rational, they are willing to take part in the risk sharing system only when the risk sharing gives welfare improvements. This is a different concept from Pareto efficiency, and in general the PEFF solution does not necessarily result in larger expected utility for *every* agent. In this example, it is possible to compare the expected utility for each agent. Table 2.2 shows the comparison of expected utilities in terms of certainty equivalents. For the OEB case, the certainty equivalent for the end buffer is also included. It is clear that in both the CEB and the OEB cases, the agents all experience welfare improvements and should be willing to participate in the risk sharing system.

	PEFF, CEB	PEFF, OEB	Autarky
Agent 1	0.9905	1.0064	0.9798
Agent 2	1.0113	1.0132	0.9798
Agent 3	1.0068	1.0183	0.9798
End Buffer	-	1.0183	1.0000

TABLE 2.2: Certainty equivalents

Market incompleteness and the role of the risk-neutral measure $\mathbb{Q}$. In the PEFF framework, no market completeness condition is required. The risk-neutral measure $\mathbb{Q}$, which assigns a probability to every possible outcome, is in fact also an input in order to compute the PEFF solution.

In this simple example, the risks are i.i.d. and each of them has only 2 possible outcomes. Hence, the probability of the “good outcome” $\mathbb{Q}(X_n = 1.2)$ uniquely characterizes the measure $\mathbb{Q}$. Note also, that there is a simple one-to-one correspondence between the probability $\mathbb{Q}(X_n = 1.2)$ and the (ex-ante) market value of X_n :

$$\mathbb{E}^{\mathbb{Q}} X_n = 1.2 \cdot \mathbb{Q}(X_n = 1.2) + 0.8 \cdot (1 - \mathbb{Q}(X_n = 1.2)).$$

We can then say that the market value of X_n uniquely characterizes the measure $\mathbb{Q}$. For the risk-neutral measure we also assume that $\mathbb{Q}(X_n = 1.2) \leq \mathbb{P}(X_n = 1.2) = 0.6$.

If we assume no market completeness, i.e. there can be more than one risk-neutral measure, it is interesting to see the effect of different risk-neutral measures on the PEFF risk-sharing rule. We will not go into an extensive analysis on this

FIGURE 2.6: Differences in certainty equivalents as a function of $\mathbb{E}^{\mathbb{Q}} X_n$

point; in this chapter we will only investigate the range of the market value of X_n such that the individual rationality condition is satisfied and it is beneficial for the agents to form a collective to do PEFF risk sharing. To do this, for each given risk-neutral measure $\mathbb{Q}$, we compute the minimum of the differences in certainty equivalents for the three agents between the PEFF and autarky solutions: a positive value will indicate that all the agents experience utility improvement and they should be willing to do risk sharing. For the OEB case, the difference in certainty equivalent for the end buffer is also included to compute the minimum. The range of the probability $\mathbb{Q}(X_n = 1.2)$ is set to be $[0.3, 0.6]$, or equivalently, $\mathbb{E}^{\mathbb{Q}} X_n \in [0.92, 1.04]$. Figure 2.6 shows the results for both the OEB and the CEB cases.

As we can observe from the figures, the OEB case in general results in a higher utility improvement for the agents compared to the CEB case. The intuition is that, in the OEB case, the end buffer can also absorb some risks, and there is larger space for risk sharing compared to the CEB case. For the OEB case, the risk sharing is beneficial for all the agents, including the end buffer, as long as the market values of the risks are larger than 0.95 (and smaller than 1.04). In terms of probability $\mathbb{Q}(X_n = 1.2)$, we have $\mathbb{Q}(X_n = 1.2) \in [0.375, 0.6]$. For the CEB case, the agents will only do the risk sharing when the market prices of the X_n 's fall within $[0.96, 1.012]$, corresponding to $\mathbb{Q}(X_n = 1.2) \in [0.40, 0.53]$. Otherwise, at least one of the agents will find the risk sharing not beneficial and the PEFF collective may not be formed.

Illustration of Algorithm 1. Algorithm 1 indicates that to find the PEFF solution, one starts with some initial weights θ , calculates the functions g_n and h_n , gets the distributions of the decision variables, and then updates the weights until they converge to some θ^* . Setting the error tolerance $\varepsilon = 10^{-6}$, the weight usually converges in less than ten iterations and is not sensitive to the starting values.

FIGURE 2.7: Functions h_n and g_n

Figure 2.7 shows the functions h_n and g_n . Recall that for $n = 1, 2$, h_n can be seen as the implied marginal utility function for the buffer size F_n , and for $n = 1, 2, 3$, g_n can be seen as the implied indirect marginal utility function for the total asset A_n .

2.7.2 Explicit PEFF Solution: the Case of Exponential Utility Function

This section discusses a special case when we assume the R_n 's are all constants (thus only the risks X are stochastic) and exponential utility functions (the constant-absolute-risk-aversion (CARA) utility) are used for all the contingent payments

$$u_n(x) = 1 - e^{-\alpha_n x}, \quad \text{for } n = 1, \dots, N,$$

and also for the end buffer

$$u_p(x) = 1 - e^{-\alpha_p x}.$$

Then we will have an explicit PEFF solution: the contingent payments are actually linear functions of the risks.

THEOREM 2.7.1 (*PEFF solution under CARA utilities and deterministic asset returns.*)
The PEFF solution to an N -period problem with exponential utility functions and deterministic asset returns $\{R_n\}$ is of the form

$$\begin{aligned} C_n &= a_n [(X_n + F_{n-1}R_n) - w_n] + v_n = a_n(A_n - w_n) + v_n, \\ F_n &= A_n - C_n = (1 - a_n)A_n - (v_n - a_n w_n), \end{aligned}$$

where

$$w_n := \mathbb{E}^{\mathbb{Q}} A_n$$

which can be calculated recursively from the budget constraints and the a_n 's are defined recursively by

$$a_N = \frac{\alpha_p}{\alpha_p + \alpha_N}, \quad (2.7.1)$$

$$a_n = \frac{a_{n+1}\alpha_{n+1}R_{n+1}}{\alpha_n + a_{n+1}\alpha_{n+1}R_{n+1}} \quad n = 1, \dots, N-1. \quad (2.7.2)$$

PROOF See appendix. $\square$

Theorem 2.7.1 shows that under CARA utility, each contingent payment only takes a proportion a_n of $A_n - w_n$ which is the excess return from total available asset, thus only takes part of the risk. The remaining part $(1 - a_n)$ is shifted into the future. Under the CARA utility assumption, the risk-sharing rules don't depend on the distribution of the random variables.

REMARK 2.7.2 Suppose $R_n \equiv R = 1 + r$ for $n = 1, \dots, N$. Also, let $\alpha_n \equiv \alpha$ for $n = 1, \dots, N$, that is, we assume the same risk aversion level for all the agents except the buffer. The equations (2.7.2) become

$$a_n = \frac{a_{n+1}R}{1 + a_{n+1}R}.$$

If we let $N \rightarrow \infty$, then we shall have

$$a_n \rightarrow \frac{R-1}{R} \approx r,$$

that is, given a sufficiently long horizon, the proportion that each agent takes from the total excess return is approximately equal to the risk-free rate.

2.8 Concluding Remarks

In this chapter we have explored solving a multi-period risk sharing problem under the concept of Pareto efficiency and financial fairness. The important results are:

1. Theorem 2.3.2 characterizes the Pareto efficient risk-sharing rules: every PE risk-sharing rule can be associated uniquely to an optimization problem with the objective function being the weighted aggregate expected utility of the contingent payments, which can be further translated into the intertemporal balance equations. Theorem 2.3.5 shows how to compute the risk-sharing rule given the weights.

2. Theorem 2.5.1 establishes the existence and uniqueness of a PEFF risk sharing rule. Furthermore, Theorem 2.5.2 indicates that the value profile will uniquely determine the weights.
3. Theorem 2.6.1 guarantees the possibility to find unique the PEFF rule numerically by a universal algorithm.

We conclude this chapter with some comments on further research possibilities.

1. **Multiple payments for each agent.** Throughout this chapter we have assumed that each agent can have only one contingent payment. As a result, the optimization target (2.3.1) is time-additive and the value profile is straightforward to determine. If we make the generalization that each agent can have multiple contingent payments in different periods, two issues need to be resolved. Utility-wise, one needs to choose a preference functional for evaluating the welfare, which may not be time-separable. Value-wise, the fairness constraint in such a setting needs to be formulated. Some cases can be quite different from the setting in this chapter, and the existence and uniqueness of the PEFF solution may have to be re-established.
2. **Serial dependence of risks.** It is assumed that the risks in different periods are independent with each other, which simplifies the optimization problem (2.3.1). In the case where some kind of serial dependence is assumed, solving problem (2.3.1) becomes complicated in the sense that the state space is much larger compared to the case where no serial dependence is present.
3. **Portfolio optimization.** The distributions of the risks here are assumed to be given throughout the thesis, which excludes the possibility of portfolio optimization in the case of collective investment. PEFF risk sharing with collective investment decisions in a single-period case has been investigated by Pazdera et al. [32]. In their research, an extra condition is assumed to sufficiently guarantee the uniqueness of the solution. In this multi-period case, such a condition may also be needed in order to derive the unique PEFF solution. We leave this as a future research topic.
4. **Fairness criterion and issues on discontinuity.** In this chapter the financial fairness is defined in an *ex ante* sense, i.e. the market values of the contingent payments will match the given value profile only at the time when the system starts. The FF will generally not hold *ex interim*, as the

contingent payments are by nature contingent claims and their market values will change after the system starts. For a CDC pension system which may include already the unborn cohorts at start, this issue may result in the so-called *discontinuity problem*: some future cohort may find themselves in a very disadvantageous position when they have to face a large deficit in the buffer left by the previous generations because of some preceding bad financial performance. The later cohort may argue that they didn't have a say when the system was initiated, thus they may choose not to step into the system. Strict ex-interim FF is meaningful, but essentially excludes any possibility of intertemporal capital transfer, thus there is no space for inter-generational risk sharing. One may then adopt some fairness condition that lies between the two extremes as a compromise. This will be covered in the next chapter.

Chapter 3

Intertemporal Allocation of Investment Risk in the Decumulation Phase of a Collective DC Scheme¹

3.1 Introduction

The financial crisis in the last decade have witnessed sharp and sudden falls in the funding ratio of many pension funds in the Netherlands. The current low interest rate environment substantially increases the present values of liabilities. As a result, there has been a trend of transforming the traditional defined-benefit schemes towards the defined-contribution schemes in the second pillar part - the funded part - of the pension system in the Netherlands. The sponsor is retreating from their role and the pensioners may have to bear more investment risks. In a collective scheme, the investment risk needs to be allocated among both the current and future participants. Intertemporal allocation with respect to investment risks is the main topic of this chapter: it involves risk smoothing – allocating risk across current and future benefits for the same generation, and intergenerational risk sharing – allocating risk across current and future generations.

Two main design principles for intertemporal risk allocation in a collective DC system are the notion of “PEFF” – Pareto efficiency from a utility perspective, and financial fairness from a value perspective. The notion of PEFF is motivated from the dual properties of the collective DC pension system. On the one hand, it is a multilateral risk sharing system, where the Pareto efficiency is a fundamental principle. On the other hand, it is essentially a financial contract for each participant. Each participant pays contributions in his early life in exchange for pension payments when he becomes old. It can be attractive if the contract is fairly priced

¹For the original paper, see Bao et al. [6]. Some modifications are made to make the paper fit in the entire thesis.

in the sense that, seen at some given point in time, the market values of benefits and contributions are equal, or at least close to each other, for each participant.

In a system where utility functions are adopted, Pareto efficient risk allocation is an optimization problem: optimizing the weighted aggregate utility totally characterizes the Pareto frontier [7]. Regarding the financial fairness principle, a value-based ALM framework requires using risk-neutral valuation [34], and the fairness principle becomes a constraint that tells how value should be distributed across multiple periods. Thus, under the PEFF principle, the optimal intertemporal allocation of investment risk can be formulated as an optimization of utility with fairness constraints in value.

The efficiency principle is frequently adopted in the theoretical modeling of multi-period allocation of investment risk. The optimal lifetime savings and investment decision is derived based on maximizing the aggregate expected utility over the whole life time of a participant [23] [29]. A similar approach is adopted by Gollier [20] to investigate the optimal allocation of investment risk across generations. The fairness principle, on the other hand, is not often adopted as a design principle and is only checked afterwards. Examples include Cui et al. [15] and Hoevenaars and Ponds [25]. The fairness principle based on market values is adopted as a design principle by Bovenberg and Mehlkopf [11]; it is required that the ex-ante market values of intergenerational transfers, as seen at the start of the system, should be zero. The generations are treated fairly in the sense that the ex-ante market value of benefits is equal to the ex-ante market value of contributions for each generation. Pareto efficiency is also a design principle in [11]. However, the model is limited in the sense that the intergenerational transfer is restricted to be a linear function of the risks in the current and the previous period. In practice, conditional indexation systems are adopted by many CDC pension plans. Indexation of benefits to a certain financial market index is also possible. Though one may assess the efficiency and fairness of such schemes, the PEFF principle is not explicitly incorporated into their original design.

The main purpose of this chapter is to combine the efficiency principle in utility with the fairness principle in value to determine how investment risks should be allocated across different periods in a collective DC pension system. In Chapter 2 we found the unique PEFF solution in a multi-period setting given the utility functions, market values of benefits and distributions of investment risks as inputs. Compared to the solution in [11], there is no need to give restrictions on the allocation rules, and the inputs will uniquely determine the allocation rule that is Pareto efficient and financially fair.

The design of the PEFF allocation in this chapter aims at more realistic situations than highly stylized models. Considering realistic elements complicates the problem to a large extent. First, it is debatable how to evaluate the “true” utility of the total pension benefits for a participant across multiple periods, and Pareto efficiency is not straightforward to define. Second, the system can deviate significantly from ex-interim fairness for later generations if it is designed to be ex-ante financially fair, which can lead to discontinuity problems; on the other hand, strict ex-interim financial fairness squeezes the space for intergenerational risk sharing especially among non-overlapping generations [43]. There should be a balance between the ex-ante and ex-interim fairness, thus between sustainability and intergenerational risk sharing. Lastly, other risks like longevity and inflation risks need to be considered, and frequent revisions of capital market parameters are required if the development of financial market is also considered.

Addressing these issues will call for simplifying the problem in the design procedure and adding in *ad hoc* elements. Regarding the first issue, individual pension payments happening at the same time point are aggregated, and the aggregate benefits are considered. The main reason to do so is that there will be only one benefit payment in each period after aggregation, and we can then totally focus on how to achieve balance between the present and the future when allocating the investment risks in multiple periods. Utility functions are assigned directly to aggregate benefits. Different from the usual way of using utility functions to evaluate welfare, we mainly use the utility function to approximately model the risk preference of an aggregate benefit relative to aggregate benefits in other periods, and control how the risks are allocated among the current and future benefits.

The last two issues give motivation to introduce a moving horizon framework, which we will call “Mohopeff”. The Mohopeff approach takes a more realistic view. We calculate the optimal allocation rule based on a *design model* which simplifies the realistic system but still contains the key information. One example of design model is the calculation of benefits in the system of personal pension accounts (PPA); see Bovenberg and Nijman [13] and the technical explanation in Bovenberg et al. [12]. When calculating the benefits for the current year for a pensioner, they set up a design model with the horizon equal to the rest of the pensioner’s maximal lifetime. In that design model, deterministic predictions are used to represent the risks, including the best estimate of the mortality rates, the expected return of investment and the current interest rate term structure. In the case of the Mohopeff, the design model is set up on a rolling basis with a fixed

horizon length, and the investment risk is allowed to be stochastic. First a horizon length is chosen, say 10 years; in each year, the allocation rule is calculated under the PEFF principle. When we move on to the next year, a new design model is set up again over the next 10 years with updated inputs such as the financial market parameters and mortality tables, and we calculate a new allocation rule based on the new projections.

The moving horizon approach has two main advantages. First, it allows us to find one version of financial fairness that lies between the ex-ante and ex-interim fairness criteria, which is more appropriate in realistic situations where intergenerational risk sharing is desired. Second, it allows us to update our estimation on financial market parameters, and changes in mortality tables and realized inflation can be dealt with by an adjustment mechanism.

The major drawback of the Mohopeff approach is that a moving-horizon structure is an *ad hoc* element and requires extra inputs that may be decided at the discretion of the pension fund, for instance, the length of the horizon. Aggregating the benefit payments in the same period is a convenient way to simplify the problem and enables us to fully focus on the intertemporal risk allocation. However, it is a nonstandard way of aggregation in the scope of the common modeling approaches, where benefits are aggregated in terms of utility for each generation.

To summarize, the Mohopeff approach has the following main innovations. Theory-wise, it offers a methodology to calculate the allocation rules based on economic principles in a more realistic setting, taking into account the balance between fairness and risk sharing. Result-wise, as we shall see later in an ALM study, it achieves a good risk-return trade-off for the benefit payments, and finds a balance between the current and future benefit payments when distributing the investment risks.

The discontinuity problem is still present even if we try to find a balance between ex-ante and ex-interim fairness criteria. Incoming generations may be reluctant to join if the market value of benefits seen at entry is significantly lower than their contributions. On the other hand, the existing generations may want to terminate the fund when a large surplus has been accumulated. To identify these possibilities we employ a utility analysis based on expected lifetime utility: by comparing the expected utility of a generation in the case of receiving the variable annuity from the Mohopeff system with the case of receiving an annuity from the open market as an alternative, we are able to identify the acceptable range of the fund size relative to its liabilities.

The rest of the chapter is structured as follows. We will first describe the collective DC system we work with. Then the PEFF principle will be introduced together with its moving horizon version, the *Mohopeff* approach. The Mohopeff approach will be tested in an ALM study, and in particular, we focus on the performance of the Mohopeff approach if we vary the inputs. We finally move on to the discussion of the discontinuity problem and we introduce the concept of tolerance band for the fund. Some remarks will conclude this chapter in the end.

3.2 The Collective DC Pension System

In this section we describe the collective DC system we work with. Section 3.2.1 describes the structure of the CDC system, Section 3.2.3 introduces the aggregate benefit, and Section 3.2.4 formulates the variables as well as the notations.

3.2.1 General Framework

We consider a realistic collective DC pension fund. The fund is open in the sense that there are always new generations stepping into the system and it continues indefinitely into the future. Compulsory participation is first assumed, and later in the chapter a sustainability analysis shall be performed. The generations, both the existing and the incoming ones, form a collective and there are no other risk sponsors.

We only consider the decumulation phase. Each participant makes a lump-sum contribution at the time of entry, and during the rest of his life he gets a life-long variable annuity. During the decumulation phase, the pension fund invests part of its capital in risky assets. The asset mix is assumed to be fixed. The investment decision is an exogenous element and we focus on the risk allocation when the total amount of investment risk is given.

Regarding the micro-longevity risk, we assume the *survivor dividend*: the pension wealth of those who decease early will be distributed among the survivors.

3.2.2 Annuity-Target Profile

An *annuity-target profile* is assigned to each pensioner at the time he enters the system. Each annuity payment is associated with a target level. The annuity target can be interpreted as a soft promise the pension fund makes to the pensioner, and the actually paid annuity fluctuates around the target level, dependent on the funding status and the performance of the financial market. As we shall see

later, the annuity target will be linked to the market values of that benefit when calculating the allocation rules; in this sense, the target profile indicates how the lump-sum contribution shall be distributed into future benefit payments in terms of market value. See Remark 2.4.2 of Chapter 2.

It is assumed in this chapter that the annuity target level is equal to the amount of the fixed nominal annuity which is calculated in an actuarially fair way by using the nominal interest rate and best estimated mortality rates. The target levels stay the same for the whole life of each pensioner. In the scope of assumed interest rate (AIR), one can say that the nominal term structure is adopted as the AIR when determining the target profile.

Using the nominal term structure is just one possible choice; other rates can also be used. The reason we choose the nominal term structure in this thesis is that, in an ideal environment (no longevity risk, no default risk for the bonds, good liquidity for bonds with different maturities, etc.) the returns implied by the nominal term structure can be achieved if the participant directly goes to the bond market and buys the corresponding bonds at the time of retirement. The target profile calculated according to the nominal term structure is in this sense “attainable”. One may also use the expected return of the asset mix which is higher, but is not attainable without taking investment risk and exposing to the price fluctuations.

3.2.3 Aggregate Benefits

The pension fund needs to allocate the risks within the collective. The pension fund can allocate the risks among both the current and future generations as well as within each generation. There can be infinitely many ways of doing this. We narrow down our scope in this chapter by focusing on *intertemporal allocation* of risks, that is, how to allocate risks among the current and future annuity payments. Further simplifying the problem motivates us to consider the *aggregate benefit* at each time point. The aggregate benefit is defined as the sum of all the single annuity payments from all the existing generations that should be paid during the period. Once the aggregate benefit has been determined, it will be distributed among the constituent annuity payments, which is the topic of the next chapter.

The primary reason for aggregating the annuity payments for each time period is to simplify the allocation problem. A pension fund may consist of many participants which adds up greatly to the complexity of the allocation. By considering the benefits over an aggregate level, there will be only one benefit payment

to be considered in each period. In such a way we fully focus on the intertemporal allocation of risks over an aggregate level, that is, to decide how much should be paid as the current benefit, and how much should be left in the fund for future liabilities.

The drawback is that such aggregation is an *ad hoc* element for simplification purposes and may not be intuitive from an economic point of view. As shown in Figure 3.1, directly aggregating the benefits in the same period is a cross-sectional way of aggregation compared to the common temporal-aggregation approach where the benefits are indirectly aggregated for a specific generation in terms of utility; see e.g. [11] [23].

FIGURE 3.1: Cross-sectional aggregation of benefits

3.2.4 Variables

The unit of time in the system is one year and the contributions and benefits happen annually: lump-sum contributions from the incoming generation come into the system at the beginning of the year while the annuity payments are paid out at the end. The system is essentially a discrete-time system. We are only interested at the time points that are both the end of the previous year and the

beginning of the next year. We use τ as a general notation to represent the time point of interest. For convenience, the term “year τ ” refers to the period between time point $\tau - 1$ to τ .

There is a reference interest rate which is regarded as the risk-free rate; the term structure is available in nominal terms. $R_{\tau|s,t}$ denotes the gross nominal return implied by the forward rates from time s to t according to the nominal interest rate term structure at time τ . X_τ denotes the gross asset return during the year τ .²

$B_{t,i;\tau}$ denotes the actual annuity payment (B stands for “Benefit”) at time τ for the pensioner who enters at time t and is indexed by i . This pensioner pays his lump-sum contribution $C_{t,i}$ at entry (i.e. at time t) and he expects variable annuity payments at time $t + 1, t + 2, \dots, t + T_\omega$ where T_ω is the maximal year that a pensioner can stay in the decumulation phase according to the mortality table. C_t represents the total contribution by the whole generation entering at time t . $S_{\tau|t,s}$ denotes the survival probability that, seen at time τ , the generation entering at time t will survive by the time s .

$AT(B_{t,i;\tau})$ denotes the annuity-target level associated to $B_{t,i;\tau}$. Given the assumption that the annuity-target profile is equal to the actuarially fair fixed nominal annuity, $AT(B_{t,i;\tau})$ can be calculated by solving

$$C_{t,i} = b_{t,i} \cdot \left(\sum_{s=1}^{T_\omega} \frac{S_{\tau|t,t+s}}{R_{\tau|t,t+s}} \right), \quad (3.2.1)$$

$$AT(B_{t,i;t+s}) \equiv b_{t,i} \quad s = 1, \dots, T_\omega. \quad (3.2.2)$$

On the aggregate level at time τ , A_τ denotes the total assets of the fund to be divided between the current aggregate benefit, B_τ , and the fund for future liabilities, F_τ :

$$A_\tau = B_\tau + F_\tau. \quad (3.2.3)$$

According to the definition, B_τ is the sum of all the current single annuity payments, i.e.

$$B_\tau = \sum_{t=\tau-T_\omega}^{\tau-1} \sum_{i \in I_{t;\tau}} B_{t,i;\tau} \quad (3.2.4)$$

²In this and next chapter, the notations C , R and X have different meanings compared to those in Chapter 2. For a complete list of notations, please refer to the list of symbols at the beginning of the thesis.

where $I_{t;\tau}$ is the index set corresponding to the pensioners who entered at time t and are still alive at time τ . Those who decease between time $\tau - 1$ and τ will not receive any annuity payments, and their pension wealth will be distributed among the survivors.

Advancing to the next year, the fund will continue to be invested, together with the new lump-sum contributions paid by the new incoming generation. We then have

$$A_{\tau+1} = (F_{\tau} + C_{\tau})X_{\tau+1}$$

where $X_{\tau+1}$ denotes the stochastic investment return in gross terms from time τ to $\tau + 1$ and C_{τ} denotes the total contribution from the generation entering at time τ .

How to split the total asset between the current aggregate benefit and the remainder for future liabilities in (3.2.3) is the main problem in this chapter.

3.3 The Notion of Pareto Efficiency and Financial Fairness and the Mohopeff Approach

This section deals with the principle of Pareto efficiency and financial fairness and the moving-horizon PEF (Mohopeff) approach. We first introduce the design model, then go on to the notion of PEF in a multi-period setting, and finally discuss in detail how the Mohopeff approach will work.

3.3.1 The Design Model

The optimal allocation of risk should indicate a good balance between the current and future benefits, which depends on our expectations on the future pension system. One possibility is to calculate the allocation rule in a *design model*. The design model, or D-model, is a simple model for the pension system over a horizon into the future. It needs to be decided what risks should be included in the D-model; best estimates shall be adopted to represent other risks. We calculate the optimal allocation rules according to the setting in the D-model. In this chapter we assume that only the investment returns are stochastic in the D-model; for the inflation and mortality table we use the best estimates and exclude any randomness.

Suppose the current time is τ , and a D-model is set up. Our purpose is to calculate an allocation rule for $B_{\tau+1}$ as a function of $A_{\tau+1}$, that is, at time τ we already determine how to divide the total asset $A_{\tau+1}$ by the end of the year between the current payments and the remainder for future payments.

Consider the D-model at time τ with a fixed horizon length N , e.g. 10 years. This means that we consider the projected pension system from time τ to $\tau + N$. The current fund size is F_τ . It ends at time $\tau + N$ with the undistributed capital in the fund, $F_{\tau|N}$, for future liabilities. $B_{\tau|s}$ denotes the aggregate benefit at time $\tau + s$, $s = 1, \dots, N$. The notation " $\tau|$ " indicates that the variables are local variables in the D-model at time τ .

We then need the predictions on cash flows into the fund within the horizon, i.e. the contributions into the system by future generations. Under the assumption of fixed contribution rate, the stream of contributions $\tilde{C}_{\tau|1}, \dots, \tilde{C}_{\tau|N}$ can be determined jointly by the projected number of future participants, predicted inflation and rate of return on the investments of the corresponding generations. The tilde indicates the nature of prediction, and $\tilde{C}_{\tau|0}$ is C_τ which has already realized at time τ .

To include investment risks the D-model is equipped with a probability space, including an objective measure $\mathbb{P}$ and a risk-neutral measure $\mathbb{Q}$. $X_{\tau|s}$ denotes the stochastic annual gross per-unit-of-money return from the fixed asset mix investment during the year $\tau + s$. They are seen as random variables. It is assumed that the risk stream $\{X_{\tau|s} | s = 1, \dots, N\}$ is sequentially independent under both $\mathbb{P}$ and $\mathbb{Q}$, i.e. $X_{\tau|s_1}$ and $X_{\tau|s_2}$ are independent for $s_1 \neq s_2$. We need to give an estimated probability distribution to each $X_{\tau|s}$, both under the given objective probability measure $\mathbb{P}$ and risk-neutral measure $\mathbb{Q}$. We further assume the absence of stochastic interest rates in order to keep the D-model simple.

The budget constraints within the horizon is as follows. At the end of each period, it has to be determined how to distribute the total asset between the current aggregate benefit and the remainder in the fund, i.e.

$$B_{\tau|s} + F_{\tau|s} = A_{\tau|s} = (\tilde{C}_{\tau|s-1} + F_{\tau|s-1})X_{\tau|s}, \quad \forall s = 1, \dots, N.$$

There can be many ways of splitting $A_{\tau|s}$ into $B_{\tau|s}$ and $F_{\tau|s}$. The following part of this chapter is dedicated to explore the allocation rules under the principle of PEFF. It is also worth mentioning that, as we will apply the PEFF methodology in Chapter 2, the D-model in this chapter will be established in such a way that all the technical conditions of the PEFF framework in Chapter 2 will be satisfied.

3.3.2 Pareto Efficiency in a Multi-Period Setting

Pareto efficiency is an important concept in academic studies of risk allocation. To talk Pareto efficiency we employ utility functions to represent the risk preferences. The utility functions here satisfy the usual conditions including continuity, differentiability, concavity and Inada conditions. Utility function $u_{\tau|\tau+s}$ is assigned to aggregate benefit $B_{\tau|\tau+s}$. It is a basic and important assumption that the utility function can give an adequate representation of average risk aversion among the group of all beneficiaries at a given point in time. Allocation to individual beneficiaries will be discussed in the next chapter.

In line with Chapter 2, we will also give a utility function $u_{\tau|p}$ to the fund size by the end of the horizon, $F_{\tau|\tau+N}$. Similar to the definition of end buffer in Chapter 2, for convenience $F_{\tau|\tau+N}$ will be called the *end fund*.

Consider the D-model with horizon length N . An allocation rule $(B_{\tau|\tau+1}, B_{\tau|\tau+2}, \dots, B_{\tau|\tau+N}, F_{\tau|\tau+N})$ is *Pareto efficient* if there does not exist another allocation $(B'_{\tau|\tau+1}, B'_{\tau|\tau+2}, \dots, B'_{\tau|\tau+N}, F'_{\tau|\tau+N})$ such that

$$\begin{aligned} &(\mathbb{E}_{\tau}^{\mathbb{P}} u_{\tau|\tau+1}(B'_{\tau|\tau+1}), \dots, \mathbb{E}_{\tau}^{\mathbb{P}} u_{\tau|\tau+N}(B'_{\tau|\tau+N}), \mathbb{E}_{\tau}^{\mathbb{P}} u_{\tau|p}(F'_{\tau|\tau+N})) \geq \\ &(\mathbb{E}_{\tau}^{\mathbb{P}} u_{\tau|\tau+1}(B_{\tau|\tau+1}), \dots, \mathbb{E}_{\tau}^{\mathbb{P}} u_{\tau|\tau+N}(B_{\tau|\tau+N}), \mathbb{E}_{\tau}^{\mathbb{P}} u_{\tau|p}(F_{\tau|\tau+N})) \end{aligned}$$

for the given utility functions $u_{\tau|\tau+1}, \dots, u_{\tau|p}$. All the expectations here are taken at time τ , i.e. the beginning of the D-model.

It becomes clear from the definition that, in the scope of Pareto efficiency, the utility function can be seen to specify the risk aversion with regard to an aggregate benefit relative to other aggregate payments, thus it will have an impact on the risk allocation. According to the result in Chapter 2, there will be infinite allocation rules that satisfy the PE condition. Specifying the market values of the aggregate benefits and the end fund can be one way of finding a unique allocation.

3.3.3 Financial Fairness

Financial Fairness stipulates that for each participant, the market value of the benefits should equal the market value of the paid-in contributions. Depending on whether the equality is seen at the beginning of the system or at the entry time of the participant, two versions of fairness are generally recognized:

- ex-ante financial fairness, as discussed by Bovenberg and Mehlkopf [11]. It requires that the market values of benefits and contributions are equal only

as seen when the system starts; that is, the market values of benefits and contributions of each generation are evaluated at time zero. The advantage is that it allows much room for risk sharing across generations. The disadvantage is that the market value of benefits at entry can deviate significantly from the value of contributions for the generations who enter long after the system starts, which may incur the discontinuity problem: either the existing or the incoming generations may have an incentive to terminate the system if such deviation is too large.

- ex-interim financial fairness, as discussed by Teulings and de Vries [43] in the form of generational accounts. This version of fairness requires that the market values of benefits and contributions should be equal for each generation as seen at the entry time of that generation; that is, the market values of benefits and contributions of a generation are evaluated at the time when that generation enters the system. It is a more strict constraint and usually implies ex-ante fairness. As a result, the discontinuity problem may be less likely to happen. The disadvantage however is that intergenerational risk sharing between non-overlapping generations is difficult to arrange.

It is more appropriate in a realistic situation to combine the advantages of the two versions of fairness. On the one hand, we would like the intergenerational risk sharing to persist; on the other hand, the deviation from ex-interim fairness should be limited to an acceptable range. This motivates us to find a new version of financial fairness based on a moving-horizon framework, that lies between the ex-ante and ex-interim criteria; see Figure 3.2. The idea is to link the target levels to the market values of benefits, adjusted according to the funding status.

Connecting to the earlier discussion in Chapter 2, the notion of financial fairness is linked to the *value profile* which is the vector of market values of the aggregate benefits as seen at the beginning of the D-model:

$$V_\tau = (\mathbb{E}_\tau^\mathbb{Q} B_{\tau|\tau+1}, \mathbb{E}_\tau^\mathbb{Q} B_{\tau|\tau+2}, \dots, \mathbb{E}_\tau^\mathbb{Q} B_{\tau|\tau+N}, \mathbb{E}_\tau^\mathbb{Q} F_{\tau|\tau+N}).$$

According to the definition in Chapter 2, financial fairness with respect to the risk allocation means that the market values of the cash outflows shall match the given value vector. In other words, the risk allocation should not change the given market values of the cash outflows. Regarding the benefits, we directly work with the future values for convenience.

FIGURE 3.2: Versions of financial fairness

In this chapter we propose a version of financial fairness that is based on the moving horizon framework. The basic idea is that in a D-model, the market values of benefits, seen at the beginning of the D-model, are equal to the corresponding annuity-target levels adjusted to a version of funding ratio which will be introduced shortly.

We start with the annuity-target levels of the aggregate benefits. For a D-model at time τ , any generation which has entered at or before time τ has already been assigned the annuity-target profile. Assigning annuity-target profiles can also be done for all generations which will enter within the horizon of the D-model. However, as those generations have not yet stepped into the system, predictions on contributions, economic variables as well as the mortality rates will be used. The annuity-target levels calculated in such a way are the best prediction on the future annuity targets based on current information.

We can then calculate the annuity-target level for $B_{\tau|\tau+s}$ by summing up the target levels from the constituent benefits, adjusted by the survival probabilities:

$$AT(B_{\tau|\tau+s}) = \sum_{t=\tau+s-T_\omega}^{\tau} \sum_{i \in I_{t;\tau}} (AT(B_{t,i;\tau}) \cdot S_{\tau|t,\tau+s}) + \sum_{t=\tau+1}^{\tau+s-1} \sum_{i \in I_{t;\tau}} (\tilde{AT}(B_{t,i;\tau}) \cdot S_{\tau|t,\tau+s}). \quad (3.3.1)$$

Equation (3.3.1) indicates that the annuity target for the aggregate benefit $B_{\tau|\tau+s}$ is calculated based on the expected number of participants that are still alive at time $\tau + s$. Again, the tilde indicates the nature of an estimation.

The main reason to use annuity target in terms of expected number of participants rather than the actual number of participants is that, in scope of the PEFF theory in Chapter 2, the value profile needs to be a vector of constants rather than stochastic variables. As we shall see shortly, in the Mohopeff approach we need to link the annuity targets to the value profile in the D-model. If we use the actual number of participants and turn the annuity targets into random variables, some valuation operator needs to be used to arrive at a deterministic value profile, which goes beyond the scope of this thesis.

Analogously, we can determine an “annuity-target” for the fund F_τ . Unlike DB schemes, there are no defined liabilities in this CDC system. The *benchmark liability* is intended to measure the pension liability in terms of the annuity-target levels. Denoted as L_τ , it is defined as the present value of all the annuity-target levels of the unpaid benefits for the existing generations, adjusted to the mortality rates and discounted at the nominal interest rates suggested by the term structure at time τ . It is calculated right after the benefit B_τ is paid, but before new contributions C_τ comes into the system. L_τ can be explicitly defined by

$$L_\tau = \sum_{j=1}^{T_\omega-1} \frac{\left(\sum_{t=\tau}^{\tau-(T_\omega-1-j)} \sum_{i \in I_{t,\tau}} (AT(B_{t,i;j}) \cdot S_{\tau|t,\tau+j}) \right)}{R_{\tau|\tau,\tau+j}}.$$

The *benchmark funding ratio* (BFR) measures the ability of the fund to meet its future liability in terms of annuity target levels:

$$\kappa_\tau = \frac{F_\tau}{L_\tau}.$$

To link the annuity-target to the actual market values of benefits we proceed as follows. Note that we have a budget constraint for the value profile, that is, the present market values of cash inflows and outflows should be equal³:

$$\sum_{s=1}^N \frac{\mathbb{E}_\tau^\mathbb{Q} B_{\tau|\tau+s}}{R_{\tau|\tau,\tau+s}} + \frac{\mathbb{E}_\tau^\mathbb{Q} F_{\tau|\tau+N}}{R_{\tau|\tau,\tau+N}} = F_\tau + C_\tau + \sum_{s=1}^{N-1} \frac{\tilde{C}_{\tau|\tau+s}}{R_{\tau|\tau,\tau+s}}. \quad (3.3.2)$$

³Equation (3.3.2) makes use of the assumption that the dynamics of the money market account is deterministic according to the observed term structure.

We then let

$$\mathbb{E}_\tau^\mathbb{Q} F_{\tau|\tau+N} = \tilde{L}_{\tau|\tau+N}, \quad (3.3.3)$$

that is, seen from now, the market value of the fund by the end of the horizon, $F_{\tau|\tau+N}$, is equal to the projected benchmark liability at that time. Both are in future values for convenience. The tilde over L indicates that the projected benchmark liability by the end of the horizon is an estimated variable.

If we also let $\mathbb{E}_\tau^\mathbb{Q} B_{\tau|\tau+s} = AT(B_{\tau|\tau+s})$ for all the aggregate benefits, the global budget constraint (3.3.2) may not hold together with Equation (3.3.3). To resolve this problem we look for an *adjustment ratio* δ_τ s.t.

$$\mathbb{E}_\tau^\mathbb{Q} B_{\tau|\tau+s} = \delta_\tau AT(B_{\tau|\tau+s}) \quad \text{for } s = 1, \dots, N.$$

This number δ_τ will be uniquely determined by

$$\delta_\tau \left(\sum_{s=1}^N \frac{AT(B_{\tau|\tau+s})}{R_{\tau|\tau,\tau+s}} \right) + \frac{\mathbb{E}_\tau^\mathbb{Q} F_{\tau|\tau+N}}{R_{\tau|\tau,\tau+N}} = F_\tau + \sum_{s=0}^{N-1} \frac{\tilde{C}_{\tau|\tau+s}}{R_{\tau|\tau,\tau+s}}. \quad (3.3.4)$$

where, as we have specified

$$\mathbb{E}_\tau^\mathbb{Q} F_{\tau|\tau+N} = \tilde{L}_{\tau|\tau+N}.$$

The philosophy behind this is that if the starting fund size is too high or too low, the surplus or deficit relative to the projected benchmark liability will be spread among the aggregate benefits within the N -year horizon in the form of market value. In such a way the fund size is anchored to the liability in terms of target levels, and any Ponzi scheme that shifts positive or negative values indefinitely into the future is prohibited.

The adjustment ratio δ_τ is jointly determined by the current benchmark funding ratio and the projected cash flows, and it measures the deviation of the actual market value of the pension payment from its annuity-target level. It can be larger than 1 in case the BFR is larger than 1, meaning that the benefit is worth more than its target level, and vice versa when BFR is smaller than 1.

There can be other ways of determining the value profile, i.e. there are other ways to define a moving-horizon version of financial fairness. The definition of the moving-horizon fairness is an ad hoc element and is subject to discretionary concerns. For instance, one may want to discriminate between the incoming and

existing generations by setting the adjustment ratio equal to 1 for the new generation during their first several years as a means of extra protection.

3.3.4 The Mohopeff Approach

The main result in Chapter 2 indicates that in a D-model, given the utility functions, value profiles and the distributions of investment risks as inputs, the PEFF allocation rule will always exist and is unique. According to Theorem 2.3.5 in Chapter 2, the PEFF allocation rules are increasing functions $\{f_{\tau|\tau+s}\}$, which tells how to distribute the total assets between the current benefit and the remainder in the fund at each time point. The benefits are calculated as, for $s = 1, \dots, N$

$$B_{\tau|\tau+s} = f_{\tau|\tau+s}(A_{\tau|\tau+s}).$$

In most cases, the PEFF will not lead to analytical solutions. An iterative numerical algorithm is proposed in [5] to compute the allocation rule numerically.

To apply the PEFF algorithm the needed direct inputs are

- The deterministic stream of projected future contributions;
- The utility functions assigned to the aggregate benefits and the end fund;
- The distributions of the investment risks to be allocated, both under $\mathbb{P}$ and under $\mathbb{Q}$;
- The value profile.

The Mohopeff approach is a moving horizon version of the PEFF approach. In the Mohopeff approach, the allocation rule is calculated in the D-model each year. When we move on to the next year, a new D-model will be established and a new allocation rule is calculated based on the new D-model. Regarding the D-model at time τ , only $f_{\tau|\tau+1}$ will be actually implemented to determine $B_{\tau+1}$ at time $\tau + 1$; for $s > 1$, the $f_{\tau,\tau+s}$'s are only auxiliary and will not be executed. When we move on to the next year $\tau + 1$ and after $B_{\tau+1}$ have been paid, another D-model will be set up and the allocation rule $f_{\tau+1|\tau+2}$ shall be applied by the year $\tau + 2$ to determine $B_{\tau+2}$ at time $\tau + 2$.

The advantages of the Mohopeff are twofold. First, it allows us to find a version of financial fairness that lies between the ex-ante and ex-interim fairness criteria. Strict ex-interim fairness constraint is relaxed so that intergenerational risk sharing can be achieved. Second, it allows us to update our estimation on financial market parameters, mortality tables and inflation, etc. This is done by updating the design model in each year.

According to the Mohopeff approach, the actually-paid benefit consists of three components:

- the annuity-target component. It is the amount equal to the specified annuity-target level.
- The adjustment component. Adjustment is done to the target level to get the market value of the annuity payment. The adjustment can be either positive or negative based on the current funding status.
- The contingent claim component. After the market value of the annuity payment is determined, the actually paid annuity then depends on the performance of the financial market.

3.4 The Mohopeff Approach in a Collective DC System: an ALM Study

This section tests the Mohopeff model by means of an ALM study. A number of economic scenarios will be simulated by a risk model to represent the reality, and the pension arrangements will be calculated by the Mohopeff approach along these scenarios.

It is worth noting that the Mohopeff approach does not depend on the risk models in the sense that it can be tested over scenario sets from different risk models. However, the D-models should be calibrated in such a way that the risk dynamics in the D-model mimics the risk model as closely as possible, and one needs to adopt the best estimation on future cash inflows and inflation. In the following ALM study, since the D-model will be set up at every time point along every scenarios, the calibration of the D-models has to be done systematically in order to make the entire computation procedure possible, and some assumptions therefore are purely for computational purposes.

Two ratios will appear as result amongst others:

- benchmark funding ratio:

$$\kappa_\tau = \frac{F_\tau}{L_\tau},$$

which measures the ability of the fund to satisfy the future targets;

- *benefit ratio*:

$$\pi_\tau = \frac{B_\tau}{AT(B_\tau)},$$

which measures to what extent the annuity target level is achieved. Note that since there is no change in demography, the annuity target for B_τ is simply the sum of the target levels of all the constituent benefits.

We will also show the allocation rules under various circumstances. The allocation rule is a function from the gross asset return to the actual benefit ratio; it tells how to distribute the total asset given the investment results.

3.4.1 Calibration

In this chapter, we will not consider further any changes in mortality rates; we will proceed with each generation having a fixed population size normalized to one all the time for simplification purpose. There is no need to simulate any demography scenarios; we can then totally focus on how the financial risks can be allocated. Each generation will stay in the system for 20 years during their retirement. Also, for simplicity, it is assumed that the incoming aggregate contribution grows at the rate of realized inflation, thus the contribution flow can be totally normalized and represented by the price index.

The risk model. For the economic scenarios we employ the financial market model proposed by Draper [16] from CPB (the Netherlands bureau for economic policy analysis). We use the parameter value set that is consistent with the updated calibration from CPB; for a complete description of the risk model, see Muns [30]. 1000 scenarios will be generated, each consisting of economic data over 100 years. The Mohopeff calculation starts from the beginning of the 20th year; the first 19 years are used to generate the annuity targets for the older generations as well as the benchmark liability L_0 when Mohopeff starts. The starting benchmark funding ratio F_0/L_0 is set to some κ_0 at the beginning.

Along each scenario, the model generates a nominal term structure in gross term at time τ : $\{R_{\tau|\tau,\tau+s}|s = 0, 1, 5, 10, 30, 100\}$, where $R_{\tau|\tau,\tau+100}$ is derived from the ultimate forward rate. Other economic variables include term structure of future gross inflation $\tilde{\eta}_{\tau|\tau,\tau+s}$ which is implied by the difference between the nominal and real term structure, the realized gross inflation η_τ and the realized return from the fixed asset mix X_τ .

The fund invests in a risky portfolio: 40% in a stock index and 60% in roll-over bond positions equally distributed in quantity over bonds with maturities from 1 to 20 years. The composition of the portfolio is rebalanced each year.

For each generation, the variable annuity takes the actuarially fair fixed nominal annuity as the annuity-target profile, which is calculated according to Equation (3.3.1). In this setting, there is no need to consider survival probabilities.

The D-model. Each D-model will be calibrated as follows. We first consider a moving horizon of 10 years. At each time point along each scenario, a D-model will be established. The nominal and real term structure from the risk model will be directly taken into the D-model as input for risk-free rates and expected inflation. Within the D-model, the lump-sum contributions from the generations are assumed to grow at the rate of expected inflation, and the asset returns are assumed to be independent from year to year. Returns from stock market follow finitely discretized log-normal distributions with the same expected excess return 4% and the same standard deviation 20%.

There are many assumptions made in this section; some of them are purely for computational purposes while some are essential from a theoretical perspective. The assumption of fixed generation sizes is one example that facilitates the ALM study; it may be removed at the cost of more complicated or time-consuming computation procedures. The assumption of sequentially independent investment returns and deterministic interest rate within the D-model is critical since they are required by the PEFF theory in Chapter 2; removing these assumptions will call for innovations in the theoretical framework of the PEFF allocation.

Utility functions and the benefit ratio. Several utility functions will be experimented, including power utility, exponential utility and kinked utility functions which will be introduced shortly. For each utility function, we need to determine its risk-aversion parameter. We then proceed by assuming that for the aggregate benefits, the utility function is specified for the benefit ratio rather than the benefit itself. The same is for the end fund: we directly specify the utility function for the end fund size in terms of benchmark funding ratio.

The main motivation is that the magnitudes of the aggregate benefits can differ significantly from each other as well as from the end fund size. Considering benefit ratio is one way to normalize the benefits, and the utility functions then serve to characterize the relative risk aversion with regard to the aggregate benefit, since we will consider the percentage of change rather than the absolute change.

Assigning utility functions to benefit ratios makes it easier to specify the parameters for *kinked utility function* which assumes that risk aversion rises to another level when the benefit is too low. Figure 3.3 shows an example in the form of marginal utility. The x -axis stands for the benefit ratio, i.e. the aggregate benefit normalized to its annuity target. At the level of $x = 0.95$ there is a kink; below and above this level we assume power utility functions with $\gamma = 7$ and 3 respectively. The motivation is that people become more risk averse when they face benefit cuts; it helps to characterize the bottomline of the amount of annuity that pensioners don't want to cross. As can be seen, we can then specify the location of the kink in terms of benefit ratio rather than in terms of the absolute benefit which can differ significantly in different cases.

FIGURE 3.3: Construction of kinked utility function: example

3.4.2 Mohopeff under Power Utility: an Overview

In this section we implement the Mohopeff approach using 1000 economic scenarios from the CPB model. The starting BFR is set to be 100%. We proceed first with the power utility

$$u(x) = \frac{x^{1-\gamma}}{1-\gamma}.$$

Within each D-model, the risk aversion parameter γ is set to be 3 for both the aggregate benefits and the end fund.

Figure 3.4 shows the distributions of the benchmark funding ratio and the benefit ratio. Two observations can be made. First, the distributions will stabilize to a stationary distribution after 10 to 20 years. After arriving at the stationary

distribution, the volatility from the investment risks is allocated between the current benefits and future benefits in a balanced manner. A stationary distribution also means that no Ponzi solution that shifts volatility indefinitely into the future is possible under the Mohopeff.

FIGURE 3.4: Power utility with $\gamma = 3$, 100% starting BFR

The second observation is that the benchmark funding ratio and the benefit ratio stay above 0.85 and below 1.6 within a 90% confidence interval. There is a large probability that the BFR is larger than 1. The reason is that part of the excess return from the risky investment is deposited in the fund. Such an “excess return effect” is also true for the benefit ratio.

Figure 3.5 gives another perspective with respect to the excess return effect: it shows the distributions of the BFR and the benefit ratio using the simulated scenarios under $\mathbb{Q}$ by the CPB model. The excess return effect no longer exists in a risk-neutral world. The medians of the two ratios lie around 1, which means that the probabilities of upward or downward fluctuations relative to the target level are almost the same.

Figure 3.6 shows the result when we start with the BFR equal to 85%, a low funding status. The BFR and the benefit ratio also stabilize after approximately 20 years to a stationary distribution as in the case in Figure 3.4. The cost is then borne mostly by the first several annuity payments.

As the pension fund has investments in risky assets and is thus exposed to investment risk, the pensioners shall expect to get a variable annuity. Generally, the actually paid benefit can range from 90% to 160% of the annuity-target level. One question is how volatile it can be. Figure 3.7 shows the distributions of the yearly changes in BFR and benefit ratio. In extreme cases, the pensioner can face a $\pm 15\%$ yearly change in benefit payments.

FIGURE 3.5: Distributions under $\mathbb{Q}$, power utility with $\gamma = 3$, 100% starting BFR

FIGURE 3.6: Power utility with $\gamma = 3$, 85% starting BFR

The benefit payments seem much more volatile compared to what can be perceived in reality. There are several reasons for this. As we work with a decumulation system and there is no other risk sponsor, the volatility must be totally absorbed by the current and future benefit payments. As the result of the Mhohpeff approach, the volatility in the system is allocated between the current and future benefits in a balanced manner. The volatility in benefit ratio is at the implied proper level in order to keep the distributions of the benchmark funding ratio stable.

3.4.3 Comparison to an Indexation Strategy

To compare the Mhohpeff strategy with an alternative that mimics conditional indexation, we introduce the *simplified indexation* strategy which mimics the policy

FIGURE 3.7: Yearly changes in BFR and benefit ratio

of conditional indexation in a very simple way. The simplified indexation indicates that the actual benefit is indexed according to the funding status, and the indexation rate is determined by the difference between the current funding ratio and the target funding ratio. In this chapter, the benchmark funding ratio will be used, and the actual benefit payment is calculated by:

$$B_{\tau} = AT(B_{\tau}) \cdot \left(1 + \frac{\kappa_{\tau-1} - \kappa^*}{N} \right),$$

where κ^* is the target BFR level which we set to be 100%.

FIGURE 3.8: Simplified indexation, 100% starting BFR

As is shown in Figure 3.8, under the simplified indexation strategy the volatility is not effectively shared among the current and future benefit payments, which in fact leads to a Ponzi solution that shifts the volatility indefinitely into the future. This is part of the reason why in reality a set of rules has to be introduced, rather than merely a simple linear rule like the simplified indexation. In contrast, Mohopeff shows a better performance: it gives situation-dependent allocation

rules and results in a balanced distribution of volatility between the current and future cohorts.

3.4.4 Allocation Rules: Using Different Utility Functions

The utility functions assigned to the aggregate benefits and the end fund are important inputs for the D-model. In the previous sections we have used the same power utility function for the aggregate benefits and the end fund: a power utility function with $\gamma = 3$. In this section and next section we vary the input utility functions and see the impact to the allocation rule.

Seen at the beginning of the D-model, the PEF allocation will not change the market values of the benefits. It is the volatility from the investment risks that needs to be allocated, and the utility function plays an important role in determining how the volatility is allocated. The results in previous sections are based on the power utility function; we shall also experiment with other kinds of utility functions, for instance, the exponential (or CARA, constant absolute risk aversion) utility function, and kinked utility function.

Figure 3.9 shows the distributions of BFR and benefit ratio under exponential utility functions with risk aversion parameter equal to 1, i.e.

$$u(x) = 1 - e^{-\alpha \cdot x} \quad \text{with} \quad \alpha = 1.$$

Compared to the results in Figure 3.4, one can conclude that it makes very little difference whether CARA or power utility functions are used. This may be due to the fact that in most situations only a very limited domain of the utility function is used, and the difference between CARA and power utility is small.

FIGURE 3.9: CARA utility function, 100% starting BFR

Figure 3.10 shows the allocation rules under 3 different utility specifications for the first year along a sample path – benefit ratio as a function of the current gross asset return. The utility functions include power utility with $\gamma = 3$, CARA utility with $\alpha = 1$, and kinked utility with parameters specified in Section 3.4.1. The starting BFR is set to be 1. As can be seen, the differences among the 3 allocation rules are small, especially between the power utility case and the CARA utility case. Adopting power or CARA utility functions results in almost linear allocation rules, while in the kinked utility case the allocation rule is close to piece-wise linear: the line becomes steeper when the return turns out to be high, and flatter if the return is below 1.

FIGURE 3.10: Comparison of allocation rules under different utility functions

3.4.5 Allocation Rules: the Impact of Risk-Aversion Parameter

The risk aversion parameters can also be used to control how volatility is allocated. One possibility is to treat the aggregate benefits and the end fund differently, by assigning to them different risk aversion parameters. Examples include

1. “Front protection”: the aggregate benefits are assumed to be more risk averse than the end fund, thus more volatility is shifted into the future payments.
2. “Back protection”: the aggregate benefits are assumed to be less risk averse than the end fund, thus more volatility is absorbed by the current benefit payments.
3. “Equal protection”: assume equal risk aversion levels to both the aggregate benefits and the end fund, as we have done in previous sections.

Figure 3.11 shows the resulted allocation rules based for the first year of a randomly chosen scenario under

1. Front protection: power utility with $\gamma = 5$ for the aggregate benefits and power utility with $\gamma = 3$ for the end fund.
2. Back protection: power utility with $\gamma = 3$ for the aggregate benefits and power utility with $\gamma = 5$ for the end fund.
3. Equal protection: power utility with $\gamma = 3$ for both the aggregate benefits and the end fund.

FIGURE 3.11: Comparison of allocation rules under different risk aversion parameters

As expected, front protection leads to a flatter line while back protection leads to a steeper one, for the aggregate benefits. Compared to the equal protection, the current pensioners get protected when the BFR is low in the case of front protection, and in return they get less when the BFR is high.

FIGURE 3.12: Comparison of allocation rules under different utility policies, with kinked utility

It makes some difference when kinked utility is involved. We consider combining the power utility with the kinked utility in specifying the risk aversion levels:

1. Front protection: kinked utility with $\gamma = 7, 3$ for the aggregate benefits, and power utility with $\gamma = 3$ for the end fund.
2. Back protection: power utility with $\gamma = 3$ for the aggregate benefits and kinked utility with $\gamma = 7, 3$ for the end fund.
3. Equal protection: power utility with $\gamma = 3$ for both the aggregate benefits and the end fund.

The lines in Figure 3.12 vary from Figure 3.11 where only power utilities are employed. In the case when the current aggregate benefits are assigned the kinked utility, the current pensioners are protected by a flatter curve when the investment return decreases to a dangerous level. In contrast, in the back protection case, the current pensioners need to sacrifice in order to help to pull the low BFR back. We shall note that the difference is fairly smaller than in Figure 3.11: in most cases the difference is no larger than 2% of the target. The reason may be that only varying risk aversion for part of the utility function is not enough to result in significant front/back loading effect in volatility.

3.4.6 The Impact of the Asset Mix

The results so far are based on the assumption that the pension fund has a fixed investment strategy, that is, the asset mix of the fund is fixed. Optimizing the investment decisions is an important part for the pension system design, but is not addressed in the PEFF theory in Chapter 2. In this section, we investigate the impact of the investment decisions by changing the composition of the asset mix. Two asset mixes will be considered: the first consists of 30% in the stock index and 70% in the bond portfolio; the second consists of 50% in the stock index and 50% in the bond portfolio. The bond portfolio consists of equal amounts of bonds with maturities from 1 to 20 years.

Figure 3.13 and 3.14 show the distributions of the BFR and the benefit ratio when we vary the proportion invested in stock. As can be seen from the results, when only 30% of the total asset is invested in stock, the distributions of the two ratios are more compact compared to the case when half of the asset is invested in stock. Also, the medians of the ratios are lower. Varying the asset mix is thus related to the balance between risk and return from the investment.

FIGURE 3.13: Changing the asset mix: 30% in stock

FIGURE 3.14: Changing the asset mix: 50% in stock

3.4.7 The Impact of the Horizon Length

Although in the results above we have made it as default that the horizon length is 10 years, the length of the horizon is also a parameter which can be tuned. Figure 3.15 to 3.17 show the quantiles of the BFR and benefit ratio under a 5-year, 10-year and 15-year horizon. The starting benchmark funding ratio is set to be 85 %, that is, the fund is suffering from a low funding status. The utility functions used are power utility functions with $\gamma = 3$. The fund always invests 40% of its asset in the stock index.

The choice of the horizon length has the following two main effects:

1. A longer horizon means that the quantiles of the BFR have a wider range compared to the case when a shorter horizon is chosen. The reason is that when a longer horizon is chosen, the current beneficiaries are exposed to less adjustment in the scope of Equation (3.3.4).

FIGURE 3.15: The impact of horizon length: $N = 5$, $BFR = 85\%$ FIGURE 3.16: The impact of horizon length: $N = 10$, $BFR = 85\%$

2. A longer horizon means that the fund can arrive at a stationary distribution more slowly than when a shorter horizon is chosen. As the figures show, it takes only roughly 15 years for the fund to get to a stationary distribution in a 5-year horizon setting, while in a 15-year horizon setting the benefit payments in up to 40 years can also be affected. A stationary distribution still exists, but comes later: the effects of the initially low BFR are felt for a longer time.

3.5 Tolerance Band and Sustainability

The discussion above has implicitly assumed a high degree of compulsory participation and has taken into no consideration the *discontinuity problem*, that is, if the funding ratio of the fund is too low which leads to a low level of benefit payment, the prospective cohorts may find the scheme unattractive and they may

FIGURE 3.17: The impact of horizon length: $N = 15$, $BFR = 85\%$

stay out of the system by looking for alternatives like individual annuity product from insurance companies. It is then natural to ask what level of the fund should be called “too low”. Symmetrically, when the funding ratio is too high, the existing cohorts may want to terminate the system immediately and distribute all the wealth accumulated in the fund. In this section we propose a way to find this break-up point based on a utility analysis.

Individuals participate in the system to share risks within the collective and this results in welfare improvement for them compared to the stand-alone situation. Thus the individual is still willing to accept a relatively low benefit payment to some certain extent as long as he still finds it attractive to participate in the system in the coming years. Conversely, he is also willing to stay in the system instead of terminating the fund when the funding status is high, as long as he still expects to be better off in the system. This idea motivates the concept of the *tolerance band* of the fund: as long as the fund size stays within a certain range, both the existing and the prospective cohorts will still like to participate and the pension system will not break up.

The concept of tolerance band is thus built on an evaluation of welfare from the life-long annuity. One possibility to describe the total utility for a participant who entered in year τ is to assume that the participant calculates the expected aggregate utility at time τ :

$$\mathbb{E}_{\tau}^{\mathbb{P}} \sum_{s=1}^{T_{\omega}} d^{t-1} u_I(B_{\tau,i;\tau+s}), \quad (3.5.1)$$

where d is a subjective discount factor and u_I a utility function. Following a conventional way we let $d = 0.97$ and $u_I(x) = \frac{x^{1-\gamma}}{1-\gamma}$ with $\gamma = 3$.

While cross-sectional aggregation is used in the design process of the allocation rules where we consider aggregating the benefit payments in each period, in Expression (3.5.1) we use the temporal aggregation to do a sustainability analysis, where the benefit payments belonging to the same participant are aggregated in terms of utility.

To compute the tolerance band we need to compare the expected aggregate utilities from two pension plans for a participant. It is then more convenient to make the comparison in terms of certainty equivalent. This motivates us to raise the idea of *certainty equivalent annuity* (CEA). For an annuity stream $\{B_{\tau,i;\tau+s} | s = 1, \dots, T_\omega\}$, the CEA is defined as the constant annuity payment B^C such that the following holds:

$$\sum_{s=1}^{T_\omega} d^{t-1} u_I(B^C) = \mathbb{E}_\tau^\mathbb{P} \sum_{s=1}^{T_\omega} d^{t-1} u_I(B_{\tau,i;\tau+s}).$$

For the purpose of comparison we need alternative annuity products from the open market. Two possible choices are

1. a fixed nominal annuity. The annuity is calculated in the actuarially fair way at the time of entry, thus it equals the annuity-target profile of the corresponding variable annuity from the Mohopeff system.
2. the *doorbeleggen*, *smart* variant of the IDC scheme mentioned in Steenkamp [40]. For a pensioner with the index i who retires at year τ with lump-sum contribution $C_{\tau,i}$, the actual annuity paid at year $\tau + s$ is determined by

$$B_{\tau,i;\tau+s} = \frac{IA_{\tau+s;i}}{AF_{\tau+s}}$$

where $IA_{t;i}$ is the size of the individual account at time t , and AF_t is the annuity factor calculated based on the *expected return* from the investment over the rest of the life. The investment strategy follows a life-cycle pattern; for simplicity, here we assume that the proportion in stock decreases linearly from 40% to zero.⁴

The lower bound of the tolerance band. Consider the generation as a whole which enters at time zero. To calculate the break-up point, we consider that this generation needs to decide whether to step into this Mohopeff pension system by

⁴Comparison to this type of variable annuity is also done in Section 4.5.4 from a different perspective.

comparing the CEA between taking part in the Mohopeff system and purchasing an annuity from the two alternatives above. To determine the actual benefits this generation can get in the Mohopeff system, we need to disaggregate the aggregate benefits calculated in the previous sections. To this point we assume that the aggregate benefit shall be distributed among its constituent benefits proportional to the corresponding annuity-target levels; a detailed discussion on how to distribute the aggregate benefit will be the topic of the next chapter.

Figure 3.18 shows the difference in CEA, as a function of starting benchmark funding ratio, between the Mohopeff system and the two alternative annuity arrangements. The CEA is normalized by dividing the CEA of the fixed nominal annuity. The figure suggests that the critical break-up point of the BFR for the Mohopeff system is roughly 0.85 when the fixed nominal annuity is taken as an alternative – that is, if the BFR goes below 0.85, the about-to-enter generation may choose not to step in and they may go for the fixed nominal annuity. The collective system then faces the danger of collapse.

Compared to the fixed nominal annuity, the *doorbeleggen smart* annuity leads to a higher aggregate utility – roughly 8% more in terms of CEA. When it is taken as an alternative, the break-up BFR rises to 1.01, which means that the Mohopeff system is attractive compared to the variable annuity only if its asset is more than its benchmark liability - that is, there has been some excess return accumulated in the pension fund.

FIGURE 3.18: Difference in CEA relative to fixed nominal annuity,
 $\gamma = 3$

The risk-aversion parameter γ in Expression (3.5.1) also plays an important

role. Figure 3.19 shows the result regarding the lower bound when we set $\gamma = 7$. When a higher risk aversion is assumed, the fixed annuity becomes more attractive to the pensioners compared to the variable annuities. The break-up BFR of the Mohopeff system with the fixed nominal annuity as an alternative rises up to 0.92. When the doorbeleggen smart annuity is taken as the alternative, the break-up BFR is still around 1.01 even if a higher risk aversion is assumed for the pensioners.

FIGURE 3.19: Difference in CEA relative to fixed nominal annuity,
 $\gamma = 7$

The upper bound of the tolerance band. Regarding the upside of the tolerance band, we employ the same approach as for the downside of the tolerance band, i.e. we consider the existing generations at time zero who compare the aggregate utility between staying in the system or terminating the system and immediately converting what they get into an individual annuity product for the rest of their lives. We further assume that upon termination, what each generation shall get equals the current BFR times the discounted sum of their corresponding unpaid annuity-target levels, that is, the generations dismiss the fund by taking the capital in the fund proportionally to the present value of the unpaid annuity-target levels, discounted at the current term structure.

Figure 3.20 shows the number of existing generations who will stay in the Mohopeff system as a function of the starting BFR. When taking the fixed nominal annuity as an alternative, more than half of the generations will be in favor of terminating the Mohopeff system and turning to individual products when the

BFR reaches 1.15. We then conclude that when one takes the fixed nominal annuity as an alternative, the tolerance band of the Mohopeff system in terms of benchmark funding ratio is (0.85, 1.15). The results in Section 3.4 show that the probability that the BFR goes above 1.15 is significant. Thus the Mohopeff system is relatively stable with respect to the downside of the tolerance band, but not the upside: there can be a significant probability that the existing generations may decide to terminate the system when the asset in the fund is worth 15% more than the corresponding benchmark liability.

If the doorbeleggen smart annuity is taken as an alternative, the corresponding break-up BFR becomes 1.01 when more than half of the generations will vote for terminating the system. This means that as long as there is some excess return accumulated in the system, the existing generations will have the motivation to terminate the system and take the excess return. Note that 1.01 is also the lower bound of the tolerance band; this suggests that the Mohopeff system is not stable when taking the doorbeleggen smart annuity as the alternative: the incoming generation wants to join only when the existing generations want to close the system.

FIGURE 3.20: Number of generations in favor of staying in the Mohopeff system, $\gamma = 3$

When we increase the risk-aversion parameter from 3 to 7, the fixed annuity becomes more attractive and more than one half of the generations will want to terminate the system when the BFR reaches 1.08 and the tolerance band becomes (0.92, 1.08). Raising the risk-aversion parameter leads to a narrower tolerance

band when the fixed nominal annuity is taken as an alternative. When the doorbeleggen smart annuity is taken as the alternative, the corresponding BFR rises to around 1.04 and the corresponding tolerance band stays close to the result when $\gamma = 3$ is assumed. This means that also in this case there is essentially no tolerance band when the variable annuity is taken as an alternative.

FIGURE 3.21: Number of generations in favor of staying in the Mohopeff system, $\gamma = 7$

Comments regarding the results. First, the results in this section have been derived under various assumptions. One of these is that particular investment strategies have been adopted both in the Mohopeff system and in the variable annuity. Another assumption is that the time-separable aggregate utility defined in Expression (3.5.1) can measure the welfare of the participant. A future research topic may be to use other kinds of preference functionals to characterize the total welfare of the pensioner.

Second, the calculation in this section is simple and does not take into consideration any transaction costs of getting an individual annuity product from the open market. In the case where insurance companies charge a higher premium, the individual pension products can be less attractive. The result is that the tolerance band of the Mohopeff system becomes wider when transaction costs are considered.

Lastly, the calculation of the upper bound makes use of the assumption that it is possible for the existing generations to terminate the system whenever they like to do so. In realistic situations, they may face legal and regulatory issues, and

some deduction on the pension wealth may be applicable in the case of immediate termination. These factors will have an impact on the tolerance band, and the modeling of them can be an interesting future research topic.

3.6 Concluding Remarks

In this chapter we adapt the PEFF algorithm in the previous chapter to a moving-horizon approach and apply it to design the optimal intertemporal allocation of investment risks in a CDC scheme. To incorporate realistic situations, the allocation rule is calculated on a moving-horizon basis in a design model which reflects the current information set and the best estimates over the future. A moving-horizon version of financial fairness is discussed to find a balance between ex-ante and ex-interim fairness criteria.

An ALM study shows that the Mohopeff approach results in a balanced distribution of investment risks between the current and future benefit payments in terms of value and volatility. Depending on the inputs, linear or piece-wise linear allocation rules can be optimal.

Finally, we analyze the sustainability of the pension scheme with the help of the tolerance band, a utility-based concept. The tolerance band is applicable to all kinds of collective pension schemes in principle; it can be used to find out the acceptable range of the funding ratio such that the fund remains attractive to both the incoming generations and the existing generations.

The major drawbacks of the Mohopeff approach include that it is an *ad hoc* element and requires extra inputs, such as the horizon length, that may be decided at the discretion of the pension fund. Although aggregating the benefit payments in the same period is a convenient way to simplify the problem, it is an cross-sectional aggregation in the scope of the common modeling approach, where benefits are aggregated in terms of utility for each generation. Furthermore, it may not be easy to specify a proper utility function for the aggregate benefit in order to describe the average risk preferences of the current beneficiaries.

There are many directions that can be further explored.

1. While we have assumed a simple fixed asset mix, the investment policy is an important part of the decumulation phase. Different investment strategy can give different risk-return profile and have an influence over the distributions of the benefits. The Mohopeff approach, as in its current form, is not capable of optimizing asset allocation. It will be an interesting topic to include portfolio optimization in the Mohopeff approach.

2. We have assumed in this chapter that the accumulation and decumulation are separated and the fund mainly performs as the decumulation platform. While this can be one possibility in reality and computations are simplified under such an assumption, it is interesting whether it can be generalized to a classic CDC situation where the two phases are integrated. The answer shall be “yes”; one solution is to convert the contributions into deferred annuity targets at the time they are paid into the system. In any case, the Mohopeff approach will work as long as the annuity-target can be determined.
3. The D-models can be more complicated. In this chapter we calibrate the D-model in a rather simple way, and only investment risk is included. It remains to be explored whether other risks, like the longevity risks and stochastic inflation rates, can be included in the design model.
4. So far it has not been discussed how aggregate benefits should be distributed among current beneficiaries. The efficient and fair allocation within the groups of current beneficiaries will be the topic of the next chapter.

Chapter 4

Multi-Period Investment Return Allocation within a Heterogeneous Collective, with Applications to Collective Defined-Contribution Plan Design¹

4.1 Introduction

The Dutch government has expressed its intention to arrive at a major revision of the current pension system in the Netherlands by 2020 [42]. In anticipation of the results of the ongoing debate, and under pressure of current low interest rates, a law has already been passed by Dutch parliament which expands the possibilities for retiring individuals to opt for variable annuities, rather than fixed annuities [46]. At the same time, following the Dutch tradition of collective funds based on compulsory participation, collective elements are likely to play an important role in the new system. In fact, it has been argued that a system with compulsory participation may raise legal issues under European competition law unless strong collective elements are present [24]. Similar discussions also take place in Canada, where target-benefit plans have been proposed to provide pensioners with variable annuities in a collective scheme [40].

If the collective system is to survive, it is therefore crucial to design risk sharing mechanisms that bring clear benefits to participants. The main risks involved are micro and macro longevity risk, and investment risk. Sharing of these risks within a collective may have a disadvantage that individual property rights are

¹For the original paper, see Bao et al. [4]. Some modifications are made to make the paper fit in the entire thesis.

not clearly defined. To address this issue, a system of *personal pension accounts* (PPA) has recently been proposed by Bovenberg and Nijman [13]. In this system, participants have individual pension accounts, but collective risk sharing can still be realized through a system of agreements which allow amounts to be added to or subtracted from individual accounts according to pre-specified rules. For instance, sharing of micro-longevity risk can be implemented through a system of survival dividends. The possibility of sharing investment risk in a similar way is mentioned in [13], but not worked out. Both for the benefit of participants and for legal reasons as mentioned above, it is important to develop risk allocation mechanisms with respect to the investment risk.

A number of guiding principles for sharing of investment risk are mentioned in the memorandum that has been published by the Dutch government in conjunction with the proposal for modification of the Pension Act [45]. The document states first of all that individual property rights should be clearly defined; this is in line with the PPA proposal of Bovenberg and Nijman. The memorandum furthermore stipulates that an allocation mechanism for investment risk should be constructed in such a way that no ex-ante redistribution of wealth is generated [45, p. 13]. In addition to this, the system must be such that it fits with the risk profile of participants, in line with the prudent person principle [45, pp. 15–18].

A key question is, therefore, how to design an allocation mechanism for sharing of investment risk that satisfies the dual criteria of no ex-ante redistribution of wealth and matching with the risk profiles of participants. In this chapter we propose a design methodology that has these two principles built in.

The existing literature that deals with multi-period allocation of investment risk in a collective DC setting includes Gollier [20] and Bovenberg and Mehlkopf [11] amongst others. They mainly focus on modeling the investment risk sharing across generations. Regarding the two principles mentioned above, they both model the risk preferences of the generations by using utility functions. Gollier [20] does not consider fairness criteria in value terms. Bovenberg and Mehlkopf [11] does incorporate the principle of no ex-ante redistribution in the design of the allocation system; however, the optimization is in a limited context in the sense that a mean-variance optimization is used as an approximation. In practice, conditional indexation systems are generally adopted in Dutch pension schemes. In a recent report from the Social and Economic Council of the Netherlands (SER) [38], a PPA system is discussed which enables sharing investment risks within the collective on the basis of a set of explicit rules. Such rules cannot by design guarantee the absence of ex-ante redistribution or matching with the risk profiles

of participants.

Our starting point is Chapter 2 which gives a procedure to design a multi-period allocation system that takes into account the preferences of agents as well as fairness in terms of market value. It combines *Pareto efficiency* from a utility perspective and *financial fairness* from a value perspective, and is referred to as the *PEFF* approach. It is a good candidate for the core part of the methodology since it is in line with the two principles mentioned in the memorandum: Pareto efficiency is fundamental in risk sharing systems and takes the risk profiles of the participants into account, while financial fairness controls how values are distributed in the risk sharing by making use of risk-neutral valuation, thus deals with the issue of no ex-ante value redistribution.

The modeling approach in Chapter 2 offers the possibility to design a CDC pension system on a long horizon that allocates the investment risk in a PEFF way. For Pareto efficiency, the pensioners' welfare of the stochastic annuity payments can be expressed by aggregated time-additive utility functions which are specified according to the risk profiles of the pensioners, and for financial fairness, the ex-ante market value of the benefits must equal the ex-ante value of contribution for each pensioner. However, some problems will stand in the way if we move closer to realistic situations:

1. Financial fairness is specified in an ex-ante sense, i.e. the participants are treated fairly only as seen from the beginning of the system. Financial fairness is difficult to justify for later generations if the ex-interim fairness deviates too much from the ex-ante fairness – in the case of a significant surplus or deficit in the fund, for instance – and a discontinuity problem will occur.
2. Pareto efficiency should in principle be formulated on the individual level. This adds up greatly to the computational complexity if the pension funds consists of thousands even millions of pensioners over a long horizon.
3. Planning over a long horizon may not be practical in reality. The optimal allocation rule is calculated at time 0 when the system starts, that is, the allocation rules are already determined for all the involved future participants at time 0. Revision of inputs and parameters, like the mortality tables and the investment return distributions, is likely to be called for after some time.

To solve these problems and make the PEFF methodology feasible in a more realistic setting calls for more elaborations, possibly with *ad hoc* elements, of the methodology. We propose in this chapter a *moving-horizon* framework which

turns the PEF methodology into a dynamic approach. A moving-horizon version of the scheme in Chapter 2 has been proposed already in Chapter 3 as a more practical alternative. In this setting, a risk-allocation plan is made within a design model for a horizon of, say, ten years, but only the first year is actually implemented. After the first year has passed, a new plan is made for the same horizon length; again, only the first year of this is implemented, and so on. The main advantages of a moving-horizon framework are the following:

- fairness in value terms is more acceptable and feasible on a shorter horizon for perpetual entities like pension funds, which corresponds to problem 1 mentioned above;
- the computational problem of designing a Pareto efficient and financially fair risk allocation scheme is more manageable on a relatively short horizon, which corresponds to problem 2;
- as part of the computation of a new plan, the model on which the design is based may be updated to adapt to revisions of for instance capital market parameters or mortality tables, which corresponds to problem 3.

The primary disadvantage of the moving-horizon approach is that it is only an approximation of the complete solution that considers all the participants of an infinite horizon, which is difficult to formulate and solve in realistic situations. The moving horizon approach also brings in an *ad hoc* element, that is, one needs to choose the horizon length as the input.

Towards further alleviation of problem 2, a *two-stage allocation* structure is proposed alongside with the moving-horizon framework. In Chapter 3, attention was limited to what might be called *intertemporal* risk sharing, which concerns the balance between current and future benefit payments. Further refinement is needed if we come down to the individual level. The risk allocation problem is still too complex to solve even within the moving-horizon framework, unless substantial aggregation is applied. We implement this via a two-stage approach. In the first stage, the method in Chapter 3 is used to construct an intertemporal allocation rule. Single annuity payments in each year are first aggregated and the first-stage allocation concerns the question how much can be spent on benefit payments in a given year, and how much should be placed in or extracted from the fund for future pension liabilities. In the second stage, the aggregate benefits that are determined in this way are allocated to participants according to their preferences. In this stage we use the single-period theory of Pareto efficient and financially fair allocation as discussed in [33]. However, the aggregation of

individual benefits at each time point also means that it provides a suboptimal solution in regard to the whole allocation problem.

This chapter can be seen as a continuation of the research in Chapter 3. The two chapters share the same philosophy that the risk allocation in the collective pension system is calculated based on design principles. The main difference is that, a two-stage allocation structure is proposed in this chapter as a further step of the methodology in Chapter 3. The two-stage allocation structure enables the pension fund to allocate the investment risks among the individual pensioners by taking into consideration the heterogeneity in risk-taking preferences, and we are then able to set up personal pension accounts in a collective system and design the system down to an individual level. Compared to Chapter 3, numerical results show that pensioners can choose their own target and risk-taking profiles to customize their own annuity.

The rest of the chapter is structured as follows. We first briefly review the individual DC schemes and the PPA proposal in [13] in Section 4.2. Section 4.3 then gives an introduction to the DC system we work with. Section 4.4 describes in detail the PEFF approach and how it can be applied to calculate the variable annuity. In Section 4.5 we implement the PEFF approach in the case of a simplified mini-pension fund to see how the investment risk can be allocated within the collective and how the PEFF approach differs from the conventional ones. Some remarks in Section 4.6 conclude the chapter.

4.2 Personal Pension Account and Individual DC Contracts

We start with a brief overview of the individual DC contracts in the Netherlands. As in Chapter 3, we mainly focus on the decumulation phase in this chapter; we simply assume that the pensioners have accumulated their pension capital when they are still active and they convert all the pension capital into an annuity at retirement. The accumulation phase is beyond the scope of this chapter. Throughout the chapter we call the pension capital at retirement *lump-sum contribution*.

The most common type of annuity in a DC contract is life-long fixed annuity, which is available from insurance companies in the open market. This kind of annuity provides fixed and guaranteed annuity cash flows until death, where no investment risk is present for the pensioner and he is insured against micro-longevity risk. There can be variations in patterns; for instance, choices include the nominal annuity in which the annuity remains fixed in nominal terms, and

the escalating annuity in which the payments increase at a pre-determined rate over time.

The advantage of a fixed annuity stream is that the pensioner faces few uncertainties during the years of retirement. The disadvantages are also obvious: first, insurance companies may charge a higher premium fee compared to collective schemes. Second, the guaranteed fixed payments rule out the possibility of benefiting from potential excess returns which may be available if the pension capital could be invested in a more aggressive and diversified way. This has been an important issue in the Dutch pension reform and it is desired that pensioners may be enabled to benefit from such excess returns by investing in the financial market. In return, there will be no guaranteed payments and the pensioners will have to bear the investment risks. Advantages of variable annuity over fixed annuity have been discussed by Koijen et al. [27] amongst others. Furthermore, new trends are that risks like the macro-longevity risk are also no longer insured and are then borne by the pensioners.

There can be many ways of determining a variable annuity. Bovenberg and Nijman [13] have recently proposed a personal pension account scheme where the pensioner affords the investment risk in the decumulation phase and longevity risk can be shared collectively. It is also possible to share investment risks within the collective. There can be many variations; for instance, the variable annuity can be calculated by incorporating the principle of shock smoothing.

Longevity risk – both macro- and micro-longevity risks – can be shared within a collective. This is done by prescribing that any remaining pension capital of any pensioner who dies early is transferred to the collective and distributed among the survivors. This is called *survivor dividend*, which gradually becomes important for those pensioners who live to an older age.

The most important feature of such IDC contracts is that the investment account is clearly defined for each pensioner. The key steering instrument is the investment policy; each pensioner can realize a certain degree of customization by modifying his investment policy and optimizing his own risk-return objective over time.

4.3 Personal Pension Account in a Collective System

In this section we describe the basic setting of the collective DC system we work with. The system is referred to as a *collective PPA* scheme, as it is a collective DC plan where personal accounts are set up for each participant. The fund is open to

new generations, and compulsory participation is assumed. There are no external risk sponsors, and the participants form a collective to share all the risks. The risks that they face include investment risk, longevity risk, and inflation risk. In this chapter we mainly focus on the design of allocation rules regarding investment risk.

FIGURE 4.1: Collective PPA and other plans

4.3.1 General Framework

This section depicts the general structure of the collective PPA system. The annuity-target profile and risk-taking profile are new concepts; other than that, it is just a description of the framework that is standard in practice.

In this system, the accumulation phase is separated from the decumulation phase. In the accumulation phase, the active cohorts accumulate their pension wealth through defined-contribution patterns. When the participants retire, the accumulated capital in his PPA will be converted into a variable annuity. In the decumulation phase, the pension fund invests part of its capital in risky assets; annuities are paid to the pensioners based on certain calculations and the risks are allocated. Figure 4.2 shows the structure of this collective PPA system.

No *hard promises* are made by the fund to the pensioners; instead, the pensioners will receive life-long variable annuity payments. In this framework, each pensioner chooses an *annuity-target profile*, which may be interpreted a kind of soft promise that the fund strives to achieve, and a *risk-taking profile* which specifies the relative risk preferences of the pensioner throughout his life. The actual annuity payments will fluctuate around the target level, dependent on the funding status and investment returns; see Figure 4.3.

The accumulation phase. Participants pay contributions into the PPA system during the accumulation phase. The accumulation phase serves as a way to accumulate capital; when the participants retire and are no longer active, the total amount in the PPA will be converted into variable annuities. This chapter is mainly focused on the decumulation phase; for the accumulation phase we assume that pensioners adopt defined-contribution types, and direct lump-sum contributions are possible.

FIGURE 4.2: Framework of collective PPA pension system

Conversion into annuity. After reaching the retirement age, the participants step into the decumulation phase by converting the total amount of accumulated contribution into a variable annuity. The conversion is made by assigning an annuity-target profile and a risk-taking profile to each pensioner, which are assumed to be fixed during the rest of his life. The annuity-target profile decides the basic pattern of the annuity stream, and the risk-taking profile describes how much risk the pensioner would like to take in his annuity payment each year. For each pensioner, each single annuity payment is assigned an annuity target and a risk-taking level at inception.

Several types of annuities can be chosen by the pensioners, for instance:

- nominal variable annuity. The annuity-target levels remain constant for the whole life time of the pensioner.
- Escalating variable annuity. The annuity target level grows at a fixed rate each year (e.g. 2%) over the life time of the pensioner.
- Other types of variable annuity, like the high-low or low-high patterns, as long as the annuity targets can be determined at the time of conversion.

For each participant, the target profile is determined based on the nominal interest rate term structure and the mortality table at the time of entry. The target profile can be realized if the participant directly turns to an insurance company

FIGURE 4.3: Annuity-target and risk-taking profiles

from the open market at retirement, that is, the participant can get a fixed annuity equal to the target profile. In contrast, the collective PPA system offers the participant a variable annuity which fluctuates around the target profile.

The risk-taking profile works in the form of utility functions that are specified directly to each single annuity payment, e.g. power utility functions. Possibilities of the risk-taking profiles include

- the risk taking levels stay the same for all the annuity payments;
- assume a higher risk taking appetite at the early stage of retirement, and a lower level when the pensioner gets older.

Decumulation phase. In the decumulation phase, the pensioners will receive a stream of annuity payments until death, i.e. a life-long variable annuity. Regarding the investment risk, the actual annuity payment is calculated based on the principle of Pareto efficiency and financial fairness, and is directly linked to the target and risk-taking profiles for each pensioner. Survivor dividend is present for dealing with micro-longevity risk: the deceased will receive no more payments and those payments will be distributed among the survivors.

4.3.2 Two Stages of Allocation

The pension fund during the decumulation phase is a dynamic system: cash outflows are constantly paid from the fund to the existing pensioners, while new lump-sum contributions are paid into the fund as cash inflows which are actually liabilities of the fund to the incoming participants. The fund faces various sources of uncertainties, including the risk from investment returns on risky asset and changes in demographic profiles as well as the inflation changes.

Regarding the investment risk, the pension fund needs to allocate the stochastic investment returns along two dimensions – across time and among individual pensioners at each time point. There can be many ways of allocating these risks. In this chapter we adopt the *two stages of allocation* perspective for calculation: first consider allocation on an aggregate level, then consider allocation with respect to each individual pensioner; see Figure 4.4. The first stage corresponds to *intertemporal risk allocation* along the dimension of time while the second stage is *intra-group tranching* along the dimension of current participants at each time point.

The primary reason for introducing the two-stage allocation structure is that it is computationally difficult to compute the solution under the principle of Pareto efficiency and financial fairness if we consider all the annuity payments of all the

FIGURE 4.4: Aggregate benefit and two stages of allocation

participants in multiple periods. As a compromise, the two-stage allocation approach allows us to decompose the problem into two sub-problems. The main disadvantage is that the solution is only an approximation and one needs to aggregate the benefit payments at each time point. It is also required to specify risk preferences for aggregate benefits as inputs, which is described in the next section. To specify risk preferences for aggregate benefits is an *ad hoc* element and the PEFF principle gives no instruction on this.

4.3.3 Variables and Notations

In this section we formulate the variables and the associated notation for later use.

For simplification purposes we assume in this chapter that the cash flows happen only annually: lump-sum contributions are paid into the system at the beginning of the year while the annuity payments are paid out at the end of each year. τ is a general notation representing the time of interest.

There is a reference interest rate which is regarded as the risk-free rate; the term structure is available, both in nominal and real terms. $R_{\tau|s,t}$ denotes the gross nominal return implied by the forward rates from time s to t according to

the nominal interest rate term structure at time τ . X_τ denotes the gross asset return from time $\tau - 1$ to time τ . In this chapter we always work with some given investment policy, thus there are no investment decisions involved: we focus on the risk allocation when the total amount of investment risk is fixed.

$B_{t,i;\tau}$ denotes the actual annuity payment (B stands for “Benefit”) at time τ for the pensioner who enters at time t and is indexed by i . This pensioner pays his lump-sum contribution $C_{t,i}$ at entry (i.e. at time t) and he expects variable annuity payments at time $t + 1, t + 2, \dots, t + T_\omega$ where T_ω is the maximal year that a pensioner can stay in the decumulation phase according to the mortality table (here we keep the same T_ω for all generations). In this pension system, each pensioner also chooses his own annuity-target profile, which determines the basic pattern of his annuity stream, and risk-taking profile, which determines the volatility of his annuity stream.

On the aggregate level at time τ , A_τ denotes the total assets to be divided between the current aggregate benefit, B_τ , and the fund for future liabilities, F_τ . That is,

$$A_\tau = B_\tau + F_\tau. \quad (4.3.1)$$

Advancing to the next year, the fund will continue to be invested, together with the new lump-sum contributions paid by the new incoming generation. We then have

$$A_{\tau+1} = (F_\tau + C_\tau)X_{\tau+1}$$

where $X_{\tau+1}$ denotes the stochastic investment return in gross terms from time τ to $\tau + 1$ and C_τ denotes the total contribution from the generation entering at time τ .

According to the definition, B_τ is the sum of all the current single annuity payments, i.e.

$$B_\tau = \sum_{t=\tau-T_\omega}^{\tau-1} \sum_{i \in I_{t;\tau}} B_{t,i;\tau} \quad (4.3.2)$$

where $I_{t;\tau}$ is the index set corresponding to the pensioners who entered at time t and are still alive at time τ .

Equation (4.3.1) and (4.3.2) characterize the problem that we will solve under the PEFf approach in the next section. Equation (4.3.1) requires us to split the total asset between the current aggregate benefit and the remainder for future liabilities, while Equation (4.3.2) is the second-stage intra-group distribution among

all the current pensioners.

4.4 The PEFF Approach: Allocating Stochastic Investment Returns

To allocate the investment risk and calculate the variable annuity within the collective PPA system the PEFF approach will be applied extensively. The acronym “PEFF” stands for “Pareto efficiency and financial fairness”. This suggests that the approach works under the notion of Pareto efficiency from a utility perspective and financial fairness from a value perspective. Utility functions are used to specify risk preferences and a risk-neutral measure is adopted to measure the market values of the annuity payments. Given the prespecified market values of the payments, the PEFF approach can optimize the risk allocation while keeping the market values equal to the prespecified values. The feasibility of finding a unique PEFF solution has been investigated theoretically by Pazdera et al. [33] in a single-period setting and by Bao et al. [5] (see Chapter 2) in a multi-period setting.

The PEFF approach is appropriate to solve the two-stage allocation problem. Pareto efficiency can be linked to risk-taking profiles and financial fairness can be linked to annuity-target profiles. The multi-period PEFF is suitable for the first-stage intertemporal risk allocation while the single-period PEFF can work in the second-stage allocation for intra-group tranching according to the heterogeneity in target and risk-taking levels.

We start with the single-period PEFF which is simpler. Then we advance to the introduction of multi-period PEFF and the moving-horizon version of it, which we call the *Mohopeff* approach. For proofs and details on computational algorithms we refer interested readers to Chapter 2, 3 and [33].

4.4.1 Single-Period PEFF: Intra-Group Tranching

Consider a single-period environment where a stochastic amount of money X is going to be divided among n agents by the end of the period. X will realize during the period, and the distributions of it under $\mathbb{P}$ and $\mathbb{Q}$ are known at the beginning. Agent i expects to get cash flow y_i ; the y_i 's are by nature contingent

claims over X . The budget constraint (BC) is then formulated as

$$\sum_{i=1}^n y_i = X.$$

The allocation $\{y_i\}$ is Pareto efficient if there does not exist another distribution $\{\tilde{y}_i\}$ such that

$$(\mathbb{E}^{\mathbb{P}} u_1(\tilde{y}_1), \dots, \mathbb{E}^{\mathbb{P}} u_N(\tilde{y}_N)) \succeq (\mathbb{E}^{\mathbb{P}} u_1(y_1), \dots, \mathbb{E}^{\mathbb{P}} u_N(y_N))$$

where the u_i 's are utility functions for evaluating the stochastic payoffs y_i . These functions are assumed to be strictly increasing and concave, and the Inada conditions must hold. Examples include power utility and exponential utility function. The utility functions are the main steering tool for determining how volatility shall be allocated.

Financial fairness requires that the ex-ante market values of cash flows must be equal to a given vector $v = (v_1, \dots, v_n)$, i.e.

$$\mathbb{E}^{\mathbb{Q}} y_i = v_i \quad \forall i = 1, \dots, n.$$

We shall call the vector v the *value profile*. It is given exogenously.

It can then be shown that it is possible to find such a solution that is PE while the FF constraints are satisfied at the same time; moreover, the PEFF solution is unique; see Bühlman and Jewell [14], Gale [19] and Pazdera et al. [33]. The y 's are increasing functions of X which can be solved numerically.

We give a simple example where 3 agents with different risk preferences share the risk from investment in a stock. The 3 agents employ power utility with risk-aversion parameters $\gamma = 6, 3, 2$ respectively. The risk is assumed to be log-normally distributed under objective probability measure $\mathbb{P}$ with expected return 5%, risk-free rate 2%, and standard deviation 20%. Under the risk-neutral measure $\mathbb{Q}$ the risk is also log-normally distributed, and the expected return is then the risk-free rate. The market values of the cash outflows for each agent are all equal to 1, i.e. total market value is 3. Seen from Figure 4.5, compared to the autarky situation where each agent gets exactly 1/3 from the total return, the most risk-averse agent is better protected when the return is low and he gives up part of the upside return. The least risk-averse agent stands at the opposite side: he takes up a large proportion of the upside return and has to suffer in bad scenarios.

Figure 4.6 shows the allocation rules as a function of the gross investment return. The rules are almost linear though in fact they are not exactly linear;

FIGURE 4.5: Intra-group risk allocation among 3 heterogeneous agents: percentages

the rule for the less risk-averse agent is steeper, which is in accordance with our expectation: the payoff for this agent is more sensitive to the investment result.

FIGURE 4.6: Intra-group risk allocation among 3 heterogeneous agents: allocation rules

The single-period PEFF optimizes the risk allocation while keeping the market values unchanged; the result is that the stochastic amount X is divided into tranches that go to different agents with different risk aversion and market value specifications. Table 4.1 summarizes the key elements regarding the single-period PEFF. It works as a processor: given the distributions of the stochastic cash inflow, the utility functions and the ex-ante market values associated to the cash

outflows, the single-period PEFF determines the allocation rules as an increasing function of the cash inflow.

Time	0	1	1	...	1
Risks / cash inflow	X			...	
Decision v'ble: cash out-flow		y_1	y_2	...	y_n
Input: utility function		u_1	u_2	...	u_n
Input: ex-ante market value		v_1	v_2	...	v_n
Output: PEFF allocation rule		$f_1(X)$	$f_2(X)$	...	$f_n(X)$

TABLE 4.1: Single-period PEFF: summary

Single-period PEFF is applied in the second-stage allocation as will be discussed later. Aggregate benefit is the stochastic amount that needs to be trached; the agents are then the single annuity payments of the existing pensioners.

4.4.2 Mohopeff: Intertemporal Risk Allocation

It makes some difference if we go from the single-period setting to a multi-period one. For intertemporal risk allocation one needs a *buffer* so that intertemporal transfer is possible. We first introduce the multi-period PEFF briefly, and then go on to its moving-horizon variant.²

Multi-Period PEFF

Suppose we have a collective pension fund in a discrete-time environment. Time t goes from 0 to N . At time t there is a new contribution C_t coming into the system after a benefit B_t is paid out from the system. Let F_t be the fund size at time t . The fund F_t and the new contribution C_t are invested in an asset mix whose gross return from time t to $t + 1$ is denoted by X_{t+1} . The X_t 's are stochastic and their distributions under $\mathbb{P}$ and $\mathbb{Q}$ are known at time 0. In this context, the C_t 's are deterministic, the X_t 's are risks to be shared and the B_t 's and F_t 's are decision variables. The budget constraints are

$$B_t + F_t = (C_{t-1} + F_{t-1})X_t \quad t = 1, \dots, N.$$

²The contents in this section have already been discussed in Chapter 3. We still cover them here in order to keep the story as a whole. Readers are referred to Chapter 3 for technical details.

In this multi-period setting, an allocation $(B_1, B_2, \dots, B_N, F_N)$ is Pareto efficient if, for the given utility functions $u_1, \dots, u_N, u_F$, there does not exist any other allocation $(\tilde{B}_1, \tilde{B}_2, \dots, \tilde{B}_N, \tilde{F}_N)$ such that

$$\left(\mathbb{E}^{\mathbb{P}} u_1(\tilde{B}_1), \dots, \mathbb{E}^{\mathbb{P}} u_N(\tilde{B}_N), \mathbb{E}^{\mathbb{P}} u_F(\tilde{F}_N) \right) \succeq \left(\mathbb{E}^{\mathbb{P}} u_1(B_1), \dots, \mathbb{E}^{\mathbb{P}} u_N(B_N), \mathbb{E}^{\mathbb{P}} u_F(F_N) \right).$$

The fund size by the end of the system, F_N , is referred to as the *end-phase fund*, or *end fund*. It is also evaluated by a utility function u_F . This helps prevent any Ponzi solutions which may shift all the volatility into the buffer. Also note that the expectation is taken as viewed from time 0.

Financial fairness is a constraint based on market values. Recall that in the single-period situation the FF can be defined in the sense that the ex-ante market values of the cash outflows will equal a given value profile. In the multi-period setting we define the FF in the same manner, that is, the market values of the benefits should match the given value profile $v = (v_1, \dots, v_N, v_F)$:

$$(\mathbb{E}^{\mathbb{Q}} B_1, \dots, \mathbb{E}^{\mathbb{Q}} B_N, \mathbb{E}^{\mathbb{Q}} F_N) = (v_1, \dots, v_N, v_F).$$

The expectation under $\mathbb{Q}$ is unconditional and we directly work with future values for convenience.

The main result in Chapter 2 is that, under the conditions given above, there exists a unique allocation solution that satisfies both the Pareto efficiency and the financial fairness conditions. The solution has the following form

$$B_t = f_t(A_t), \quad t = 1, \dots, N,$$

where

$$A_t = (C_{t-1} + F_{t-1})X_t$$

is the total asset at time t to be distributed.

The functions f 's are increasing functions and the exact forms are determined by the model inputs. They can be calculated numerically by an iterative algorithm given in Section 2.6 in Chapter 2. A summary of the multi-period PEFF is shown in Table 4.2. Multi-period PEFF requires inputs during the next N years, including the cash inflows, distributions of risks, and the market values at time 0 of the cash outflows.

Time	0	1	2	...	N	N
Input: cash inflow	C_0, F_0	C_1	C_2	...		
Input: risks		X_1	X_2	...	X_N	
Decision v'ble: cash out-flow		B_1	B_2	...	B_N	F_N
Input: utility function		u_1	u_2	...	u_N	u_F
Input: ex-ante market value		v_1	v_2	...	v_N	v_F
Output: PEFF allocation rule		$f_1(A_1)$	$f_2(A_2)$	...	$f_N(A_N)$	$f_F(A_N)$

TABLE 4.2: Multi-period PEFF: Summary

The Mohopeff Approach: the Moving-Horizon PEFF

The Mohopeff approach is a dynamic approach that adopts the aforementioned multi-period PEFF approach on a moving horizon basis. First we choose some certain N to be our planning horizon, e.g. $N = 10$ years. At the beginning of each year we set up a *design model* (D-model) within which we execute the multi-period PEFF approach. It needs to be decided what risks should be included in the D-model; best estimates shall be adopted to represent other risks. As in Chapter 3, we assume that only the investment returns are stochastic in the D-model; for the inflation and mortality table we use the best estimates and exclude any randomness. The PEFF solution can be calculated based on these information. Mohopeff requires that only the allocation rule for the next year is really implemented; when we advance to the next year, another D-model is set up based on the most updated information and we do the same within that D-model to see what is the optimal allocation strategy.

The moving horizon version of PEFF is motivated by several concerns. The primary reason is that the moving horizon approach allows us to find a compromise between ex-ante financial fairness and ex-interim financial fairness in terms of market value. Ex-ante fairness means that the total market values of benefits and contributions from each pensioner are equal only as seen at the beginning of the system, while ex-interim fairness means that the values are equal as seen at the time the pensioner retires and enters the system. The problem with ex-ante fairness is that for participants who enter the system long after the system starts, the total market value of benefits can deviate significantly from the value of the

contributions, thus the system may run into a discontinuity problem. The problem with ex-interim fairness is that it squeezes the space for risk sharing among non-overlapping generations. Taking into accounts the realistic situations, it is desirable to find a version of financial fairness that lies between the ex-ante and ex-interim versions.

Second, it may not be practical to formulate an infinite-horizon problem in realistic situations, especially when we consider risk allocation down to an individual level. A moving horizon approach makes it possible to focus on the information in the near future. Combined with the two-stage allocation framework, the moving horizon approach simplifies the problem and gives an approximate solution.

Lastly, the moving horizon approach allows revision and updating of the input parameters. The investment returns are allowed to be stochastic in the D-model. Other risks are excluded; we always use the best estimates regarding the inflation and mortality rates. This means that the PEFF solution is optimal only if the demography and inflation exactly follow the projections. Such an issue is on the one hand due to the limitation of the PEFF algorithm, as it requires deterministic contribution stream and market values of benefits as direct inputs; on the other hand, it is not easy to give explicit distributions of demography changes under a risk-neutral measure. A moving horizon approach allows us to update these best estimates once the risks have realized. At the same time, the parameters of the quantitative model that quantifies the investment risk need to be updated constantly, which should stay in line with the developments of the financial market over time.

The main disadvantage of the moving horizon framework is that it comes as an approximation of the complete solution that takes into account an infinite horizon. The horizon length N is a necessary input variable that needs to be specified exogenous to the PEFF principle, that is, it is an *ad hoc* element brought in by the moving horizon framework.

We briefly discuss how the moving horizon version of financial fairness is formulated, and put the technical details in the appendix. First, note that for each participant, the annuity target profile is determined in an actuarially fair way by using the nominal interest rates and mortality table at the time of entry. The target profile can thus be seen as the ex-ante market values of annuity payments as seen at entry. A system that aims at ex-interim financial fairness will make the market values of annuity payments equal to the corresponding annuity targets. In this moving horizon framework, this ex-interim fairness constraint is relaxed

to allow risk sharing across generations. The market values of annuity payments are linked to the corresponding annuity targets, but are dynamically adjusted based on the funding status. We call the adjustment process *value adjustment* in this chapter.

We adopt a way of value adjustment by prioritizing, with respect to the annuity payments within the horizon, the capability of the fund to meet its future liabilities beyond the horizon in terms of annuity targets. It is then natural to introduce a version of funding ratio which measures the ability of the fund to meet its future liabilities. Different from the defined-benefit plans, the collective PPA system does not have defined liabilities; instead we introduce the *benchmark liability*, denoted by L , which is the present value of all the unpaid future annuity payments of the existing generations in terms of their target levels. The benchmark liability L_τ in year τ is calculated just after the aggregate benefit B_τ has been paid out. We then introduce the *benchmark funding ratio* (BFR) which measures the ability of the fund to meet its future liability in terms of annuity target levels:

$$\kappa_\tau = \frac{F_\tau}{L_\tau}.$$

In each D-model, we always require that the market value of the fund at the end of the horizon equals the projected benchmark liability at that time, and the difference between the current funding status and that projected funding status is spread proportionally among the aggregate benefits within the horizon in terms of market value. The purpose is to anchor the fund size to the liability in terms of target levels, and prevent any Ponzi scheme that shifts positive or negative values indefinitely into the future.

Suppose the current time (year) is τ . The current fund size is denoted by F_τ , and our aim is to determine the aggregate benefit for the next year, $B_{\tau+1}$, as a function of the total asset at time $\tau + 1$, $X_{\tau+1}$. To do this, a D-model at time τ needs to be established as our design model for calculation. In the D-model we consider the time point $\tau + 1, \dots, \tau + N$, that is, consider a moving horizon of length N . The estimated aggregate contribution at time t is denoted by $\tilde{C}_{\tau|t+s}$ for $s = 1, \dots, (N - 1)$, where the tilde indicates the nature of an estimate, and τ and “|” express that it is a local variable in the D-model at time τ . We also need the decision variables $B_{\tau|\tau+1}, \dots, B_{\tau|\tau+N}$, and the end-phase fund size $F_{\tau|\tau+N}$. A utility function $u_{\tau|t}(\cdot)$ is used to express the risk preference of $B_{\tau|t}$, and $u_{\tau|F}(\cdot)$ is used for $F_{\tau|\tau+N}$. The investment risks considered are gross asset returns $X_{\tau|\tau+s}$ with $s = 1, \dots, N$.

Like every single annuity payment from each participant, each aggregate benefit is also associated with an annuity target, which is the sum of the corresponding target levels of all its constituent single annuity payments times the corresponding survival probability. Note that this entails giving certain projections on the annuity target levels for future participants who will take part within N years.

We let the value adjustment be determined in the first place by the requirement

$$\mathbb{E}_\tau^\mathbb{Q} F_{\tau|\tau+N} = \tilde{L}_{\tau|\tau+N},$$

that is, the market value of the end-phase fund should equal the projected benchmark liability at that time. Any surpluses or deficits of the fund are then spread proportionally among all the aggregate benefits in terms of market value. That leads to an adjustment ratio δ_τ such that

$$\mathbb{E}_\tau^\mathbb{Q} B_{\tau|\tau+s} = \delta_\tau AT(B_{\tau|\tau+s}), \quad s = 1, \dots, N$$

where $AT(B_{\tau|\tau+s})$ is the annuity target level of $B_{\tau|\tau+s}$. The adjustment ratio δ_τ indicates the proper level of adjustment to the current aggregate benefit in terms of market value if we want the funding status to converge to the desired level within the next N years and any surpluses or deficits shall be smoothed proportionally among the current benefits. It can be larger than 1 in case the BFR is larger than 1, meaning that the benefit is worth more than its target level, and vice versa when BFR is smaller than 1.

Multi-period PEFF then computes the allocation rules $\{f_{\tau|\tau+s} | s = 1, \dots, N\}$ as output:

$$B_{\tau|\tau+s} = f_{\tau|\tau+s}(A_{\tau|\tau+s}),$$

i.e. the benefit is a contingent claim with the total asset as underlying.

As we are working on a moving-horizon basis, only $f_{\tau|\tau+1}$ is actually implemented. When we move on to the next year, a new D-model is established and one performs the same procedure. To be exact, we can write

$$B_{\tau+1} = f_{\tau|\tau+1}(A_{\tau+1}) \tag{4.4.1}$$

which is actually an increasing function of the gross asset return $X_{\tau+1}$ in the next year. The function $f_{\tau|\tau+1}$ includes our expectation on the economic environment under the notion of PEFF.

Once the aggregate benefit is determined, the single-period PEFF then indicates how to distribute it among the current pensioners. The adjustment ratio is applied to get the market values of the individual annuity payments. The final result is that the annuity payment of each individual pensioner is an increasing function of the total asset, and the function summarizes the information including the funding status, investment return distributions, relative risk-taking levels, etc.

Micro-longevity risk is not taken into account in the PEFF approach, though. The total size of the survivor dividend can be determined, but one still needs exogenous mechanism to allocate it to each individual. This is because in the second-stage allocation it has to be assumed that each pensioner will not pass away till the end of the year, and as the mortality rates have already been incorporated in the first-stage allocation, the fund just needs to re-distribute among the survivors the part of aggregate benefit that belongs to those who have actually passed away during the year. The wealth from those who die early becomes the survival dividend for those who live longer. How to determine this survivor dividend is beyond the scope of this thesis.

4.5 PEFF Approach in a Mini-Pension Fund: an Example

In this section we carry out the PEFF approach in the case of a mini-pension fund. The purpose of this section is to visualize what the pensioners can get under the PEFF approach when different profiles are chosen, and to investigate numerically the differences between the PEFF system and the IDC plans regarding the allocation of investment risk in the decumulation phase.

4.5.1 Assumptions

The mini-pension fund exists in a world where the interest rates and the inflation rate are stochastic. There are no longevity risks: the mini-pension fund consists of pensioners who will stay in the system for exactly 20 years, and we totally focus on dealing with investment risk. As we have discussed, each pensioner is associated with an annuity-target profile and a risk-taking profile at the time of conversion. Therefore we assume that in this mini-pension fund, each generation consists of only two participants who differ either in annuity-target profile or in risk-taking profile, so that we can see the effect of that factor. We further assume

that the lump-sum contributions from the generations grow at the rate of realized inflation and the contributions from the two pensioners are always the same, so as to make things comparable. Pensioners have the following choices regarding the profiles:

- Annuity-target profile:
 1. actuarially-fair nominal annuity, computed based on the nominal interest rate term structure at the time of conversion. The target levels remain flat over the pensioner's lifetime;³
 2. actuarially-fair escalating annuity, with the target level increasing by 2% each year, computed based on the risk-free interest rate term structure at the time of conversion.
- Risk-taking profile:
 1. Power utility, with γ (the risk-aversion parameter) equal to 3 all the time: the pensioner assumes a flat risk-taking profile;
 2. power utility, with γ equal to 2 for the first 7 years, then 3 for the next 7 years, and 6 for the last 6 years: the pensioner assumes a stepwise-decreasing risk-taking level over time.

In the following sections we consider 2 cases where the pensioners have the specific choices shown in Table 4.3.

	Pensioner A	Pensioner B
Case 1	AT1, RT1	AT2, RT1
Case 2	AT1, RT1	AT1, RT2

TABLE 4.3: Choices of pensioners

³There is one comment regarding the conversion risk. Here it is assumed that the annuity-target profile is determined in an actuarially fair way, that is, it heavily depends on the term structure and mortality rates at the time of conversion. While the assumption here is mainly for simplification concern, one possibility to resolve this issue is to include in the investment during the accumulation phase certain financial instruments to help the pensioners hedge the interest rate risk, which goes beyond the scope of this chapter. Another possibility is to determine the annuity-target profile in a more comprehensive way. For instance, when the participant is still active, the paid-in contribution in each year can be converted into a deferred annuity target based on the term structure and mortality table in that year. When the participant gets retired, the final annuity-target profile is the sum of all the deferred annuity targets from the previous years. In such a way a natural smoothing procedure is incorporated and one actually uses multiple rates instead of one.

4.5.2 Calibration

We employ the CPB model with updated calibration based on Dutch data (see Draper [16], Muns [30]) to simulate the economic environment and the financial market, including the term structure, realized and expected inflation and stock returns. One thousand scenarios are simulated, each consisting of 100 time points (in years). Along each scenario, the data from the first 19 time points are used to generate the starting benchmark liability of the fund, thus the pension system starts at the 20th time point. To avoid any ambiguity we shall call the 20th time point “year 0” which is the actual starting point of the system.

The fund manages its assets collectively and there are no investment decisions involved. For simplification purpose, the fund invests its capital in a fixed asset mix which consists of 40% in a stock index with no dividends and 60% in roll-over zero-coupon bond positions equally distributed among bonds with maturities 1 to 20 years. The bond prices correspond to the nominal term structure and are thus stochastic. The composition of the asset mix is rebalanced each year. We assume that the starting benchmark funding ratio is 100%, that is, at the start the fund size equals the present value of all the future annuity target levels, discounted at the nominal term structure.

The D-model is calibrated by assuming simpler dynamics. The expected inflation is used to predict the future lump-sum contributions, and the current interest rate term structure is used to compute the target levels for future generations who will come into the system within the N -year horizon. Regarding the investment risks, we assume that the stock prices are always log-normal distributed and the excess return is 4%. Doing as such makes it feasible to perform the PEFF calculations along all the scenario paths, and the difference between the distribution of investment risk according to the D-model and according to the CPB model helps to mimic the fact that we cannot totally model the risks in reality. This provides a test of the approach under such misspecification. Regarding the utility functions for the aggregate benefits as well as for the end fund, we use power utility with $\gamma = 3$. The benefit normalized by its target level is evaluated by the utility function rather than the benefit itself: in such a way the utility function measures the percentage of change.

4.5.3 Case 1: the Effect of Different Annuity-Target Profiles

We first consider the case when the two pensioners within each generation both choose the same risk-taking profile, while one chooses a nominal annuity target,

and the other chooses an escalating annuity target. Figure 4.7 shows the distribution of the benchmark funding ratio. It is noteworthy that the BFR converges to a stationary distribution after around 20 years. The value adjustment process serves as an automatic balance mechanism: the BFR will be “pulled back” if it gets too high or too low. Ponzi solutions are essentially excluded. Another observation is that there is a large probability that the BFR reaches levels above 1, which means that the total wealth in the fund exceeds the present value of the future targets. This is caused by the fact that part of the excess return from the risky investment is deposited in the fund, which can be viewed as a “safety reserve” for potential financial catastrophes in the future.

FIGURE 4.7: Benchmark funding ratio, Case 1

We then focus on the generation which enters at year 20 when the BFR has converged to a stationary distribution. We further assume that both participants in this generation start with a lump-sum contribution of 200 000 Euros so as to make things comparable.⁴ Figure 4.8 exhibits the distribution of the actually paid annuities of the two pensioners who have chosen different annuity-target profiles. The dashed line is the mean of the target levels. One can easily see that the volatilities of the annuity payments are very stable over the lifetime of the pensioners: for Pensioner A who chooses nominal annuity target, the quantiles stay almost the same during the life time, while for Pensioner B who chooses escalating annuity target, the quantiles increase steadily as time passes.

Figure 4.8 also shows statistically the relationship between the actual annuities and their target levels. As is desired, the actual annuity lies around the mean

⁴One may argue that the lump-sum contribution should be different along different economic scenarios since, for instance, the investment policy may be designed to mitigate the conversion risk by using fixed income instruments. We will not go that far in this chapter; we let the lump-sum contribution be fixed for simplification purposes.

FIGURE 4.8: Actual annuity in nominal terms

of the target level, and one can see that the probability that the pensioner gets more than the target is larger than 50%. This is the excess return effect due to the use of the risk-neutral valuation in the PEFF calculation.

Figure 4.9 shows the distributions of the annuities in real terms by discounting the annuity payments using the realized inflation along each scenario. The variable escalating annuity provides protection against inflation from a statistical point of view, as the quantiles remain flat if the effect of inflation is stripped.

FIGURE 4.9: Actual annuity in real terms

4.5.4 Comparison with IDC Schemes

It is interesting to see numerically how the PEFF approach differs from the IDC contracts when dealing with investment risk during the decumulation phase. A good and simple candidate for comparison is the *doorbeleggen, basis* variant of the IDC scheme mentioned in Steenkamp [40]. Under absence of longevity risk, for a pensioner who retires at time τ with lump-sum contribution C_τ^I , the actual

annuity paid at time $t > \tau$ is determined by

$$B_t^I = \frac{PPA_t^I}{IAF_t}$$

where PPA_t^I is the size of the individual account at time t , and IAF_t is the annuity factor calculated based on the *expected return* from the investment over the rest of the life. We then make a comparison between the Pensioner A in case 1 and another pensioner C who participates in the IDC doorbeleggen basis variant. We also have to assume the following:

- the IDC follows the same investment policy as the PEFF system. In such a way it is clear how the two methodologies differ with respect to allocating investment risk.
- Pensioner A and C are identical in other aspects: they retire at the same time (year 20), stay alive for exactly 20 years and have accumulated the same amount of lump-sum contribution (200 000 Euros);
- the actual annuities of the two pensioners are calculated based on the same economic scenarios.

Figure 4.10 gives the comparison in distributions of actual annuity payments in nominal terms between pensioners A and C. A striking fact is that while Pensioner A can expect an variable annuity stream with very stable volatility, Pensioner C has to endure more volatility at a later stage of his life. This can also be seen from Figure 4.11, where the distributions of year-to-year changes in actual annuity are displayed.

FIGURE 4.10: PEFF system vs. an IDC scheme: comparison of annuity distributions

FIGURE 4.11: PEFF system vs. an IDC scheme: comparison of yearly changes in annuity

This phenomenon results directly from the different underlying philosophies of the two schemes. The IDC variant always calculates the benefit payment by re-annuitizing the total amount of wealth in the account every year based on the most recent estimates of the expected returns. When the pensioner gets to almost the end of his life, he has to confront the accumulated volatility. This leads to the observed end-of-the-world effect. The PEFF approach is different in contrast, as it is based on an open-fund assumption i.e. there are always new generations coming into the system. For each pensioner, the annuity payments he receives when he is young or old are treated indifferently in the PEFF approach from an aggregate level, and the end-of-the-world effect vanishes due to intergenerational risk sharing.

Similar to the PEFF approach which calculates the annuity within the D-model, the IDC *doorbeleggen basis* variant in a sense adopts a “D-model” as well. This D-model is much simpler in that it uses deterministic best estimates for all risks including investment risk. Also, while the PEFF approach adopts a moving but fixed-length horizon for the D-model, the IDC variant adopts the remaining lifetime of the pensioner as the horizon, which means that the horizon becomes shorter and shorter and the ability of risk smoothing is weakening. Amending this problem will typically lead to a life-cycle pattern of investment, that is, the amount of risk taken in the investment gradually decreases as one gets old. Concrete examples are the *doorbeleggen* variants in [40] which adopt life-cycle investment strategies. Figure 4.12 shows the quantiles of actual annuity payments of the IDC plan when the percentage of pension capital invested in stock decreases

linearly from 40% to zero over the pensioner's lifetime.⁵ Compared to Figure 4.11, the distribution becomes more compact since the pensioner is exposed less to the risk from the stock market; the other side of the coin is that he then benefits less from the possible excess returns.

Within the PEFF system, the investment is managed collectively and the proceeds as well as the associated volatility are allocated dynamically – both intertemporally and intra-group – under the PEFF principle. In contrast, pension funds that adopt IDC plans set up separate investment accounts for different age groups. This leads to our discussion of case 2 below where we can see how PEFF can help determine the intra-group risk allocations.

FIGURE 4.12: PEFF system vs. an IDC scheme: comparison of annuity distributions, decreasing equity proportion

4.5.5 Case 2: the Effect of Different Risk-Taking Profiles

In case 2 we still consider the situation where each generation consists of only two pensioners. Different from case 1, now the pensioners choose different risk-taking profiles, while keeping the annuity-target profiles identical. One pensioner chooses steady risk taking levels for the whole lifetime, while the other wants to decrease his risk taking levels when he grows old. Based on the same scenario set, Figure 4.13 shows the distribution of the BFR along time in case 2. It is very similar to the BFR in case 1 in Figure 4.7: the figure still shows a stationary distribution as well as the excess return effect.

Consider the generation which enters the system at year 20 with per-capita lump-sum contribution 200 000 Euros. Pensioner A chooses RT type 1 and B

⁵This is intended to mimic the life-cycle pattern of investment in a simple way. We limit our discussions mainly on the proportion of stock in the asset mix, and we avoid going into an extensive discussion on duration management regarding the investment in bonds.

FIGURE 4.13: Benchmark funding ratio, case 2

chooses RT type 2, i.e. the risk aversion parameter increases from 2 to 6 stepwise. A simple illustration of the PEFF intra-group risk sharing under these RT levels is already provided in Figure 4.5 and 4.6. As can be seen from Figure 4.14, the difference between the quantiles of the two annuity streams is small. However, if we look at the volatility in those annuity streams in Figure 4.15 we observe the expected effects: while the quantiles of the yearly changes under RT type 1 remain stable over time, Pensioner B takes more volatility when he is (relatively) young and less when old.

FIGURE 4.14: Distribution of actual annuity payments, case 2

Figure 4.16 gives another perspective of the intra-group risk sharing effect: it shows the difference between the payoffs of the two pensioners in cases of very high or very low returns. The left figure shows that in cases of high returns (annual return larger than 10%), Pensioner B gets more when he is young under a higher risk appetite, and gets less when he is old under lower risk tolerance. Vice versa, when the returns are low, Pensioner A gets protected when young

FIGURE 4.15: Distributions of yearly changes in annuity payments

and suffers to a limited extent when old. The RT profile thus plays an important role in the intra-group distribution of investment risks.

FIGURE 4.16: Differences in actual annuity payments in cases of high or low returns, $A - B$

4.5.6 Financial Fairness on a Generational Level: Analysis in $\mathbb{Q}$ World

As we have mentioned, the PEFF system allows much room for intergenerational risk sharing, since by design the annuity payments of each individual at different ages are treated indifferently on the aggregate level, and the natural end-of-life boundary is thus eliminated. However, allowing intergenerational risk sharing will definitely indicate that *ex-interim* financial fairness cannot be achieved. Fairness from a value perspective is not a problem for IDC contracts if one considers no longevity risks and thus no survivor dividends, as for each generation the annuity stream is totally generated from its paid-in contribution. For a collective

pension fund this is different: from an *ex interim* perspective, some generations may benefit from the intergenerational transfers while some other generations may be negatively affected. One natural question would be to what extent the ex-interim financial fairness can be violated. In this subsection we explore financial fairness at a generational level. To do this we implement the PEFF approach within 1000 $\mathbb{Q}$ -scenarios generated by the CPB model.

Figure 4.17 exhibits the benchmark funding ratio in the $\mathbb{Q}$ -world. Compared to its $\mathbb{P}$ -world counterpart in Figure 4.7, the quantiles here lie more symmetrically around 1, meaning that the $\mathbb{Q}$ -probabilities of over- and under-funding is approximately equal.

FIGURE 4.17: Benchmark funding ratio in case 1, risk-neutral world

Simulation under $\mathbb{Q}$ enables us to evaluate the financial fairness by comparing the values of cash inflows and outflows of the generations. We consider the *benefit-contribution ratio* (B-C ratio) which is defined as the ratio between the present market value of all the annuity payments and the contribution for a given generation, calculated at the time of entry. We compute the former by

$$\text{B-C Ratio} = \frac{1}{1000} \sum_{\text{path}} \frac{\sum_t \text{Annuity}_t / \text{SDF}_t}{\text{Lump-sum con.}}.$$

We consider the generation entering at year 0 with different starting funding status. Figure 4.18 shows the relationship between the starting BFR and the B-C ratio. It is almost linear and the slope is approximately 0.443, meaning that if the BFR increases/decreases by 1% then the market value of the annuity stream for the incoming generation will increase/decrease by roughly 0.44%. The ex-interim financial fairness is not violated too much as long as the BFR stays close to one;

it can be considered as the price that the participants pay in order to have the collective risk sharing at hand.

FIGURE 4.18: Benefit-contribution ratio

A too low BFR may render the pension system unattractive to new participants, and a too high BFR may give the existing generations incentives to terminate the system and distribute the capital in the fund. Based on a very similar setting, in Chapter 3 we have carried out a sustainability analysis from a utility perspective. The concept of tolerance band was introduced by comparing the aggregate expected utility for the generations when staying in the system to the utility when going for individual contract (a fixed nominal annuity or the door-beleggen smart annuity) from the open market. The result is that when the fixed nominal annuity is taken as an alternative, the system becomes unstable if the BFR takes values outside the interval (0.85, 1.15): if the BFR gets too low, then the incoming generation becomes reluctant to join, and if it gets too high then the existing generations may want to terminate the system. One comment is that as there is a large probability that the BFR can reach levels above 1.15 because of the excess return effect, the pressure from the existing participants to terminate the system can be high. The pension fund should be careful when offering participants the option of *early exit*, that is, to allow participants leave the system with lump-sum benefits.

4.6 Concluding Remarks

Guided by the trend of the Dutch pension reform, our main focus in this chapter is to design a methodology such that the investment return can be efficiently distributed among the individual pensioners in a multi-period setting, in line with

their annuity-target profiles and risk-taking preferences.

In the designed PPA system, each pensioner receives a variable annuity which is anchored to an annuity-target profile, and the risk taking is directly linked to a risk-taking profile. The pensioners can realize some degree of customization within the collectively managed pension fund by choosing their own target and risk-taking profiles. On an aggregate level, numerical results show that volatility is allocated in a balanced manner between the current and future benefits, and in particular, intergenerational risk sharing helps mitigate the volatility in annuity payments for older pensioners.

There is much room for collective risk sharing in this PEFF-featured collective system. In contrast to IDC systems with supplementary risk sharing, dynamic risk allocation is a built-in feature under the PEFF approach, and shock smoothing is a natural consequence of the system design. The investment risk has already been taken into account in each design model, and is reflected in the allocation rules. Because we use best estimates of the mortality table, the inflation and the interest rates, the macro-longevity risk and inflation risk are allocated through an adjustment mechanism. By tying the market value of the fund size by the end of the horizon to the projected benchmark liability in each D-model, the risk sharing is restricted within the near future to a large extent so that Ponzi schemes are ruled out. Intergenerational risk sharing makes it possible to share risks efficiently among both the current and future generations, and it can be avoided that pensioners of an old age have to suffer from too much volatility in annuity payments. Intra-group risk sharing utilizes the intra-group heterogeneity of risk preferences. Rather than directly manipulating the investment policy for each individual, the PEFF approach allows the fund to manage the investments collectively and then allocate the returns according to the pre-specified risk-taking levels, which resembles the practice of tranching in designing derivative products. In this way the risk sharing is tailored down to an individual level.

The PEFF-featured collective PPA system offers pensioners many possibilities to customize their pension products. Besides the flexibility during the accumulation phase, each pensioner can choose their preferred annuity-target and risk-taking profiles. The annuity-target profile is the benchmark of the market values of the annuity payments and thus determines the basic pattern of the annuity, while the risk-taking profile indicates how much risk one would like to take during different stages of life. The PEFF system enables the fund to pool pensioners with various annuity and risk-taking profiles together and allocate accordingly

the returns from the collective investment, which is different from the IDC patterns where one directly adjusts the investment policy to realize the risk-return target over time.

The PEFF methodology discussed in this chapter is closely connected to the current Dutch pension reform. The value adjustment procedure in Section 4.4.2 can be seen as a risk smoothing mechanism, and a version of funding ratio is introduced and plays an important role in risk allocation. There are also differences if we compare the PEFF methodology to the collective pension schemes in practice, for example, the IV-B variant in the SER report [38]. One key difference is the definition of the funding ratio. Unlike the conventional definition, In this PEFF framework, the funding ratio is defined based on the concept of annuity target and is thus given the name “benchmark funding ratio” for distinction. Regarding the smoothing mechanism, the IV-B variant uses pre-specified rules on funding ratio, while in the PEFF approach smoothing is done in terms of market value and the allocation rules are calculated based on design principles. In this sense, the PEFF approach is complementary to the existing methodologies and contributes to the current pension reform by offering a different perspective.

The PEFF approach has several limitations and can be refined and extended.

1. It is assumed throughout the chapter that the amounts of risks are given; no investment decisions are involved. PEFF focuses on how the investment risk can be allocated given the amount of risk, rather than how to optimize the asset allocation at the same time. Working with a fixed asset mix might be acceptable if the composition of the pensioners is relatively stable over time; however, in cases where the overall risk-taking willingness of the group changes over time, it should be appropriate to adjust the asset mix accordingly. The PEFF approach as discussed in this chapter gives no hint on how to adjust the asset mix. Pazdera et al. [32] have discussed allocation of financial risk under the PEFF principle when investment decision is endogenous in a single period setting; it may serve as the starting point for multi-period cases.
2. Only investment risk is included in the design model; for other risks we use the best estimates which are updated annually on a moving horizon basis. It can be a further research topic whether one can include other risks, like the macro-longevity risk, in the design model.
3. The two-stage allocation structure is an approximation which mainly serves to make the numerical calculation feasible.

4. The PEFF approach uses utility functions to model risk preferences. There have been other more elaborate ways of evaluating the welfare of multiple cash flows proposed in academic research from a theoretical point of view, for instance, habit formation; to incorporate them into the allocation system needs more effort, both in the relevant theory and numerical procedures.

Appendix A

Appendices for Chapter 2

A.1 Proofs for Section 2.3

For any risk sharing rule $\rho = (C_1, \dots, C_N, F_N) \in \mathcal{RS}$, let

$$u(\rho) := (u_1(C_1), \dots, u_N(C_N), u_p(F_N))$$

and

$$\phi := \mathbb{E}^{\mathbb{P}} u(\rho) = (\mathbb{E}^{\mathbb{P}} u_1(C_1), \dots, \mathbb{E}^{\mathbb{P}} u_N(C_N), \mathbb{E}^{\mathbb{P}} u_p(F_N)) \in \mathbb{R}^{N+1}.$$

First note that ϕ is a strictly concave and increasing function of ρ with co-domain $\mathbb{R}^{N+1}$. The PE optimization target then becomes

$$\mathbb{E}^{\mathbb{P}} \left[\sum_{n=1}^N \theta_n u_n(C_n) + \theta_p u_p(F_N) \right] = \langle \theta, \phi \rangle$$

where $\theta = (\theta_1, \dots, \theta_N, \theta_p) \in \mathbb{R}_{++}^{N+1}$.

We need the following definitions and results in preparation for the proof of Theorem 2.3.2.

LEMMA A.1.1 *Consider n concave functions $\{f_i | i = 1, \dots, n\}$ from a common domain K to $\mathbb{R} \cup \{-\infty\}$. Then $F(K) - \mathbb{R}_+^n := \{x - y | \forall x \in F(K), y \in \mathbb{R}_+^n\}$ is convex where $F := (f_1, f_2, \dots, f_n)$.*

PROOF See the proof of Proposition 2.6 from Aubin [1]. □

We will use a separation theorem in the proof of Theorem 2.3.2. We then need to introduce the following definitions.¹

DEFINITION A.1.2 (*Affine sets in $\mathbb{R}^n$.*) A subset $M \in \mathbb{R}^n$ is called an affine set if $(1 - \lambda)x + \lambda y \in M$ for any $x, y \in M$ and $\lambda \in \mathbb{R}$.

¹Interested readers are referred to Rockafeller [37] for more details.

DEFINITION A.1.3 (*Affine hull.*) The affine hull of any subset $M \in \mathbb{R}^n$, which is denoted as $\text{aff}(M)$, is the smallest affine set that contains M .

DEFINITION A.1.4 (*Relative interior and boundary.*) The relative interior of a convex set $C \subset \mathbb{R}$, which is denoted as $\text{ri}(C)$, is defined as the interior of C when it is regarded as a subset of $\text{aff}(C)$. The relative boundary of C is the difference of the closure of C and the relative interior of C .

The following lemma is crucial in proving Theorem 2.3.2.

LEMMA A.1.5 *Let C be a convex set. A point $x \in C$ is a relative boundary point of C if and only if there exists a linear function not constant on C such that it achieves its maximum over C at x .*

PROOF See Corollary 11.6.2 by Rockafellar [37]. □

PROOF OF THEOREM 2.3.2.

$1 \Rightarrow 2$: Let $\rho = (C_1, C_2, \dots, C_N, F_N)$ be PE. Then we have that $\phi(\mathcal{RS}) - \mathbb{R}_+^{N+1}$ is convex by Lemma A.1.1. Note that an element ρ^* is PE if and only if

$$\{\phi(\rho^*)\} \cap (\phi(\mathcal{RS}) - \mathbb{R}_+^{N+1}) = \{\phi(\rho^*)\}$$

and

$$\{\phi(\rho^*)\} \cap (\phi(\mathcal{RS}) - \mathbb{R}_+^{N+1})^\circ = \emptyset.$$

Otherwise, if $\{\phi(\rho^*)\} \in (\phi(\mathcal{RS}) - \mathbb{R}_+^{N+1})^\circ$, then there exist $\rho' \in \mathcal{RS}$ and $c \in \mathbb{R}_+^{N+1}$ with $c \neq 0$ such that $\phi(\rho^*) = \phi(\rho') - c$, which means ρ' results in a Pareto improvement. This is in contradiction with the assumption that ρ^* is PE.

$\phi(\mathcal{RS}) - \mathbb{R}_+^{N+1}$ is a full-dimensional set thus its relative interior is the same as its interior. Write $\phi^* = \phi(\rho^*)$. Then ϕ^* is a relative boundary point of $\phi(\mathcal{RS}) - \mathbb{R}_+^{N+1}$, as it belongs to $\phi(\mathcal{RS}) - \mathbb{R}_+^{N+1}$, thus to its closure, but not its relative interior. According to Lemma A.1.5, for this ϕ^* , there exists a $\theta^* \neq 0$ such that

$$\sup_{\phi \in \phi(\mathcal{RS}) - \mathbb{R}_+^{N+1}} \langle \theta^*, \phi \rangle \leq \langle \theta^*, \phi^* \rangle.$$

First note that any coordinates of θ^* cannot be negative as then

$$\sup_{\phi \in \phi(\mathcal{RS}) - \mathbb{R}_+^{N+1}} \langle \theta^*, \phi \rangle = +\infty.$$

No coordinates of θ^* can be zero. If this would be the case, suppose $\theta_1^* = 0$ while $\theta_2^* > 0$ without loss of generality. Then any $\rho = (C_1, C_2, \dots, C_N, F_N)$ cannot

be optimal since for any small $\epsilon > 0$ such that $C_1^{j_1} - \epsilon > b_1$ for all $j_1 \in \mathcal{J}_1$, $\rho_\epsilon = (C_1 - \epsilon, C_2 + R_2\epsilon, \dots, C_N, F_N)$ will result in a larger optimization target because u_2 is strictly increasing.

$2 \Rightarrow 1$: consider a risk-sharing rule ρ that maximizes $\langle \theta, \phi \rangle$ for some $\theta \in \mathbb{R}_+^{N+1}$. If ρ is not PE, then there exists another $\tilde{\rho}$ such that $\tilde{\phi} \not\geq \phi$ and hence

$$\langle \theta, \tilde{\phi} \rangle > \langle \theta, \phi \rangle$$

which results in a contradiction.

$2 \Leftrightarrow 3$: as we are working with a finite probability space, we may use the Lagrangian multiplier method to solve the maximization problem.

For $n = 1, \dots, N$, reorganize the budget constraint and we have

$$F_n^{J_{n-1}j_n} + C_n^{J_{n-1}j_n} - X_n^{j_n} - F_{n-1}^{J_{n-1}} R_n^{j_n} = 0.$$

Define

$$F_n^{J_{n-1}j_n} + C_n^{J_{n-1}j_n} - X_n^{j_n} - F_{n-1}^{J_{n-1}} R_n^{j_n}$$

as $BC^{J_{n-1}j_n}$ or BC^{J_n} .

We then maximize

$$\sum_{n=1}^N \left\{ \theta_n \sum_{J_n \in \mathcal{J}_n} \mathbb{P}(J_n) u_n(C_n^{J_n}) + \sum_{J_n \in \mathcal{J}_n} \lambda^{J_n} BC^{J_n} \right\} + \theta_p \sum_{J_N \in \mathcal{J}_N} \mathbb{P}(J_N) u_p(F_N^{J_N})$$

where the λ 's are the Lagrangian multipliers.

For any $n < N$, setting the first-order partial derivative with respect to $C_n^{J_n}$ to zero will help us find a stationary point of the optimization problem. It gives

$$\mathbb{P}(J_n) \theta_n u'_n(C_n^{J_n}) + \lambda^{J_n} = 0 \quad \forall J_n \in \mathcal{J}_n.$$

For $n + 1$ similarly we have, along the trajectory J_n

$$\mathbb{P}(J_n j_{n+1}) \theta_{n+1} u'_{n+1}(C_{n+1}^{J_n j_{n+1}}) + \lambda^{J_n j_{n+1}} = 0 \quad \forall j_{n+1} \in \mathcal{J}_n^{n+1}.$$

Now take the partial derivative with respect to F_n and set to zero

$$\lambda^{J_n} = \sum_{j_{n+1} \in \mathcal{J}_n^{n+1}} \lambda^{J_n j_{n+1}} R_{n+1}^{j_{n+1}}.$$

This will lead to

$$\theta_n u'_n(C_n^{J_n}) = \theta_n \sum_{j_{n+1} \in \mathcal{J}_n^{n+1}} u'_{n+1}(C_{n+1}^{J_n j_{n+1}}) R_{n+1}^{j_{n+1}} \frac{\mathbb{P}(J_n j_{n+1})}{\mathbb{P}(J_n)}.$$

By the assumption of sequential independence we have

$$\frac{\mathbb{P}(J_n j_{n+1})}{\mathbb{P}(J_n)} = \mathbb{P}(j_{n+1}).$$

Then the equation can be further rewritten as

$$\theta_n u'_n(C_n) = \theta_{n+1} \mathbb{E}_n^\mathbb{P} [u'_{n+1}(C_{n+1}) R_{n+1}] \quad \text{for } n = 1, \dots, N-1.$$

C_N and F_N are both $\mathcal{F}_N$ -measurable and we have

$$\theta_N u'_N(C_N^{J_N}) = \theta_p u'_p(F_N^{J_N}) = -\lambda^{J_N}$$

by taking partial derivatives with respect to C_N and F_N and setting them to be zero.

We have arrived at a stationary point thanks to the Lagrangian multiplier method; this stationary point is the unique global optimum once we note that the optimization target is a concave function with respect to the decision variables and the feasible set is convex. $\square$

PROOF OF THEOREM 2.3.5. The optimization target (2.3.1) is a parameterized optimization problem of time-additive utility functions:

$$\begin{aligned} & \max_{C_1, \dots, C_N} \mathbb{E}^\mathbb{P} \left[\sum_{n=1}^N \theta_n u_n(C_n) + \theta_p u_p(F_N) \right] \\ & \text{such that } F_n + C_n = X_n + F_{n-1} R_n \quad n = 1, \dots, N, \\ & F_0 = 0. \end{aligned}$$

This optimization problem can be solved by dynamic programming. Add in a new time point $t_{N+1} = t_N$, and

$$X_{N+1} \equiv 0, \quad R_{N+1} \equiv 1.$$

Define

$$A_n := X_n + F_{n-1} R_n \quad n = 1, \dots, N+1,$$

which has the interpretation as the total available asset at time t_n to be divided into the current cash flow and the buffer for later use. Note that by definition $A_{N+1} = F_N$. The A 's are the state variables, the C 's are the decision variables and the X 's and R 's are the risks. Then we shall have the optimization problem formulated as

$$\begin{aligned} & \max_{C_1, \dots, C_N} \mathbb{E}^{\mathbb{P}} \left[\sum_{n=1}^N \theta_n u_n(C_n) + \theta_p u_p(A_{N+1}) \right] \\ \text{such that } & A_{n+1} = X_{n+1} + (A_n - C_n)R_{n+1}, \quad n = 1, \dots, N, \\ & A_1 = X_1. \end{aligned}$$

Proposition 1.3.1 in [8] tells that in order to solve the problem one needs to define first

$$V_{N+1}(A_{N+1}) = \theta_p u_p(A_{N+1}),$$

and then define backwards, for $n = 1, \dots, N$

$$V_n(A_n) = \max_{C_n} \mathbb{E}_n^{\mathbb{P}} [\theta_n u_n(C_n) + V_{n+1}(X_{n+1} + (A_n - C_n)R_{n+1})]. \quad (\text{A.1.1})$$

This can be solved by taking the derivative of

$$\mathbb{E}_n^{\mathbb{P}} [\theta_n u_n(C_n) + V_{n+1}(X_{n+1} + (A_n - C_n)R_{n+1})] \quad (\text{A.1.2})$$

with respect to C_n and setting it to zero. We will start from period N and go backwards in time in order to verify the differentiability of the V_n 's. For period N , note that the target (A.1.2) becomes

$$\theta_N u_N(C_N) + \theta_p u_p(F_N) = \theta_N u_N(C_N) + \theta_p u_p(A_N - C_N).$$

The conditional expectation drops out because of the measurability of C_N and F_N . It is continuous and differentiable with respect to C_N . Take the derivative and set it to zero; we get

$$\theta_N u'_N(C_N^*) = \theta_p u'_p(A_N^* - C_N^*) := L_N^*.$$

Here the star indicates that it is the optimal solution. Next, define

$$G_N(x) := I_N \left(\frac{x}{\theta_N} \right) + I_p \left(\frac{x}{\theta_p} \right),$$

$$g_N(x) := G_N^{-1}.$$

Both G_N and g_N are well-defined. G_N is the sum of two strictly decreasing bijective functions thus it is strictly decreasing and bijective from $\mathbb{R}_{++}$ to $(\max\{b_N, b_p\}, +\infty)$, and it follows that g_N is also strictly decreasing and bijective from $(\max\{b_N, b_p\}, +\infty)$ to $\mathbb{R}_{++}$. The Inada conditions tell

$$\lim_{x \rightarrow 0} G_N(x) = +\infty, \quad \lim_{x \rightarrow +\infty} G_N(x) = \max\{b_N, b_p\}$$

and thus

$$\lim_{x \rightarrow \max\{b_N, b_p\}} g_N(x) = +\infty, \quad \lim_{x \rightarrow +\infty} g_N(x) = 0.$$

L_N^* can then be calculated as

$$L_N^* = g_N(A_N^*)$$

and

$$C_N^* = I_N \left(\frac{L_N^*}{\theta_N} \right), \quad F_N^* = I_p \left(\frac{L_N^*}{\theta_p} \right).$$

The value function is

$$V_N(A_N^*) = \theta_N u_N(C_N^*) + \theta_p u_p(A_N^* - C_N^*),$$

which is a differentiable function of A_N^* when we regard A_N^* as its argument:

$$V_N'(A_N^*) = \theta_p u_p'(A_N^* - C_N^*) = \theta_p u_p'(F_N^*) = g_N(A_N^*).$$

Going one period backwards, we have the value function

$$V_{N-1}(A_{N-1}) = \max_{C_{N-1}} \mathbb{E}_{N-1}^{\mathbb{P}} [\theta_{N-1} u_{N-1}(C_{N-1}) + V_N(X_N + (A_{N-1} - C_{N-1})R_N)].$$

The part

$$\begin{aligned} & \mathbb{E}_{N-1}^{\mathbb{P}} [\theta_{N-1} u_{N-1}(C_{N-1}) + V_N(X_N + (A_{N-1} - C_{N-1})R_N)] \\ &= \sum_{j \in \mathcal{J}_{N-1}^N} \mathbb{P}(j) [\theta_{N-1} u_{N-1}(C_{N-1}) + V_N(X_N^j + (A_{N-1} - C_{N-1})R_N^j)] \end{aligned}$$

is a differentiable function of C_{N-1} when we regard C_{N-1} as its argument. We then take the derivative with respect to C_{N-1} and set it to zero. Differentiation and conditional expectation can be interchanged, since we are working on a finite probability space. We have

$$\begin{aligned} & \sum_{j \in \mathcal{J}_{N-1}^N} \mathbb{P}(j) [\theta_{N-1} u'_{N-1}(C_{N-1}) + V'_N(A_N^j)(-R_N^j)] \\ &= \theta_{N-1} u'_{N-1}(C_{N-1}) - \mathbb{E}_{N-1}^{\mathbb{P}} V'_N(A_N) R_N = 0, \end{aligned}$$

which leads us to

$$L_{N-1}^* = \theta_{N-1} u'_{N-1}(C_{N-1}^*) = \mathbb{E}_{N-1}^{\mathbb{P}} [g_N(A_N^*) R_N] = \mathbb{E}_{N-1}^{\mathbb{P}} [L_N^* R_N].$$

We then define

$$h_{N-1}(x) := \frac{1}{\theta_N} \mathbb{E}_{N-1}^{\mathbb{P}} [g_N(X_N + x R_N) R_N].$$

Due to the assumption of sequential independence, $h_{N-1}(x)$ can further be written in the form of an unconditional expectation

$$h_{N-1}(x) = \frac{1}{\theta_N} \mathbb{E}^{\mathbb{P}} [g_N(X_N + x R_N) R_N]$$

since both X_N and R_N are independent from $\mathcal{F}_{N-1}$. Note that h_{N-1} is invertible since by definition it is a weighted sum of strictly decreasing functions; thus h_{N-1} is also a strictly decreasing function with domain $(d_{N-1}, +\infty)$, where d_{N-1} is defined as

$$d_{N-1} = \inf \left\{ d \in \mathbb{R} \mid X_N^j + d R_N^j \geq \max\{b_N, b_p\} \quad \forall j \in \mathcal{J}_{N-1}^N \right\}.$$

Furthermore, h_{N-1} can be viewed as the marginal utility of a stereotype utility function since

- it is continuous and strictly decreasing,
- it satisfies

$$\lim_{x \rightarrow d_{N-1}} h_{N-1}(x) = +\infty, \quad \lim_{x \rightarrow +\infty} h_{N-1}(x) = 0.$$

Write

$$H_{N-1} := h_{N-1}^{-1}.$$

Then once we combine

$$C_{N-1}^* + F_{N-1}^* = X_{N-1} + F_{N-2}^* R_{N-1} = A_{N-1}^*$$

with

$$L_{N-1}^* = \theta_{N-1} u'_{N-1}(C_{N-1}^*) = \theta_N h_{N-1}(F_{N-1}^*),$$

we have

$$I_{N-1} \left(\frac{L_{N-1}^*}{\theta_{N-1}} \right) + H_{N-1} \left(\frac{L_{N-1}^*}{\theta_N} \right) = A_{N-1}^*.$$

Next, define

$$\begin{aligned} G_{N-1}(x) &:= I_{N-1} \left(\frac{x}{\theta_{N-1}} \right) + H_{N-1} \left(\frac{x}{\theta_N} \right), \\ g_{N-1}(x) &:= G_{N-1}^{-1}. \end{aligned}$$

G_{N-1} and g_{N-1} are well-defined just as G_N and g_N . L_{N-1} can then be calculated as

$$L_{N-1}^* = g_{N-1}(A_{N-1}^*)$$

and

$$C_{N-1}^* = I_{N-1} \left(\frac{L_{N-1}^*}{\theta_{N-1}} \right), \quad F_{N-1}^* = H_{N-1} \left(\frac{L_{N-1}^*}{\theta_N} \right).$$

For the value function

$$V_{N-1}(A_{N-1}^*) = \theta_{N-1} u_{N-1}(C_{N-1}^*) + \mathbb{E}_{N-1}^{\mathbb{P}} V_N [X_N + (A_{N-1}^* - C_{N-1}^*) R_N],$$

it follows that V_{N-1} is differentiable when we regard A_{N-1}^* as the argument and one can calculate $V'_{N-1}(A_{N-1}^*)$ as follows:

$$V'_{N-1}(A_{N-1}^*) = \mathbb{E}_{N-1}^{\mathbb{P}} [V'_N(A_N^*) \cdot R_N] = \mathbb{E}_{N-1}^{\mathbb{P}} [L_N^* R_N] = L_{N-1}^* = g_{N-1}(A_{N-1}^*).$$

Proceeding one period backwards, we then have the corresponding value function

$$V_{N-2}(A_{N-2}) = \max_{C_{N-2}} \mathbb{E}_{N-2}^{\mathbb{P}} [\theta_{N-2} u_{N-2}(C_{N-2}) + V_{N-1}(X_{N-1} + (A_{N-2} - C_{N-2}) R_{N-1})].$$

To solve the right hand side, note that the expression

$$\theta_{N-2} u_{N-2}(C_{N-2}) + V_{N-1}(X_{N-1} + (A_{N-2} - C_{N-2}) R_{N-1})$$

is a differentiable function of C_{N-2} . We then take the derivative with regard to C_{N-2} and set it to zero:

$$\begin{aligned} & \theta_{N-2} u'_{N-2}(C_{N-2}) + \mathbb{E}_{N-2}^{\mathbb{P}} [V'_{N-1}(A_{N-1}) \cdot (-R_{N-1})] \\ &= \theta_{N-2} u'_{N-2}(C_{N-2}) - \mathbb{E}_{N-2}^{\mathbb{P}} [g_{N-1}(A_{N-1}) R_{N-1}] \\ &= \theta_{N-2} u'_{N-2}(C_{N-2}) - \mathbb{E}_{N-2}^{\mathbb{P}} [g_{N-1}(X_{N-1} + (A_{N-2} - C_{N-2}) R_{N-1}) R_{N-1}] = 0. \end{aligned}$$

We can then repeat what has been done in period $N - 1$. This recursive procedure can be continued backwards in time until we arrive at the first period. That is, we can always define recursively for $n = 1, \dots, N - 2$

$$\begin{aligned} h_n(x) &= \mathbb{E}_n^{\mathbb{P}} \left[\frac{1}{\theta_{n+1}} g_{n+1}(X_{n+1} + x R_{n+1}) R_{n+1} \right] \\ &= \mathbb{E}^{\mathbb{P}} \left[\frac{1}{\theta_{n+1}} g_{n+1}(X_{n+1} + x R_{n+1}) R_{n+1} \right], \\ H_n &= h_n^{-1}, \\ G_n(x) &:= I_n \left(\frac{x}{\theta_n} \right) + H_n \left(\frac{x}{\theta_{n+1}} \right), \\ g_n(x) &:= G_n^{-1}, \end{aligned}$$

and the decision variables are given by

$$\begin{aligned} C_n^* &= I_n \left(\frac{g_n(A_n^*)}{\theta_n} \right) \quad n = 1, \dots, N, \\ F_n^* &= H_n \left(\frac{g_n(A_n^*)}{\theta_{n+1}} \right) \quad n = 1, \dots, N - 1, \\ F_N^* &= I_p \left(\frac{g_N(A_N^*)}{\theta_p} \right). \end{aligned}$$

This will be the unique solution of the optimization problem, as the optimization target is concave with respect to the decision variables and the feasible set is convex. $\square$

PROOF OF LEMMA 2.3.8. By definition the function g 's are all strictly decreasing. We have

$$\begin{aligned} C_n &= I_n \left(\frac{g_n(X_n + F_{n-1}R_n)}{\theta_n} \right) \quad \text{for } n = 1, \dots, N, \\ F_n &= H_n \left(\frac{g_n(X_n + F_{n-1}R_n)}{\theta_{n+1}} \right) \quad \text{for } n = 1, \dots, N-1, \\ F_N &= I_p \left(\frac{g_N(X_N + F_{N-1}R_N)}{\theta_p} \right), \end{aligned}$$

thus both C_n and F_n are increasing functions of $A_n = X_n + F_{n-1}R_n$.

We only have to consider the case when only one coordinate of $(X, R) = (X_1, \dots, X_N, R_1, \dots, R_N)$ increases. First consider two trajectories J, J^* such that there is a time point $\tau = 1, \dots, N$ such that $X_\tau^J > X_\tau^{J^*}$ and other random variables from (X, R) are equal. Since

$$F_n = H_n \left(\frac{g_n(X_n + F_{n-1}R_n)}{\theta_{n+1}} \right) \quad \text{for } n = 1, \dots, N-1$$

then $F_1^J = F_1^{J^*}$, and this will lead to $F_2^J = F_2^{J^*}$. Doing this recursively we conclude that $F_n^J = F_n^{J^*}$ for any $n < \tau$. Then as

$$X_\tau^J + F_{\tau-1}^J R_\tau^J > X_\tau^{J^*} + F_{\tau-1}^{J^*} R_\tau^{J^*}$$

we have

$$C_\tau^J > C_\tau^{J^*}, \quad F_\tau^J > F_\tau^{J^*},$$

and the latter will tell that $C_n^J > C_n^{J^*}$ for all $n > \tau$. Also $F_N^J > F_N^{J^*}$. Then $\rho^J \not\geq \rho^{J^*}$.

The cases when only $R_\tau^J > R_\tau^{J^*}$ follows analogously. $\square$

It is convenient to have the following definition before we continue to the proof of Lemma 2.3.9.

DEFINITION A.1.6 (*N-PE Problem.*) An N -PE problem refers to the 4-tuple $((X, R), \rho, u', \theta)$ and the corresponding equation systems BC (2.5.1) and IBE (2.5.2), where (X, R) is a vector of random variables, ρ a vector of decision variables, u' an $(N+1)$ -tuple of stereotype marginal utility functions and θ a constant vector,

i.e.

$$\begin{aligned}(X, R) &= (X_1, \dots, X_N, R_2, \dots, R_N) \in \mathcal{L}^{2N+1}, \\ \rho &= (C_1, \dots, C_N, F_N) \in \mathcal{L}^{N+1}, \\ u' &= (u'_1, \dots, u'_N, u'_p), \\ \theta &= (\theta_1, \dots, \theta_N, \theta_p) \in \mathbb{R}_{++}^{N+1},\end{aligned}$$

where $\mathcal{L} := \mathbb{R}^\Omega$ is the space of random variables over the underlying probability space.

PROOF OF LEMMA 2.3.9. The key point of the proof is that otherwise, the IBE and the BC cannot hold simultaneously.

We use mathematical induction to show this. First consider $N = 1$. For a 1-PE problem this is true; we only have two agents including the buffer and there will be only one family of IBE:

$$\theta_1 u'_1(C_1) = \theta_p u'_p(F_1),$$

and the budget constraints are

$$C_1 + F_1 = X_1.$$

For any trajectory $J \in \mathcal{J}_1$, if θ_1 increases, then we argue that C_1^J cannot decrease. Otherwise (i.e. C_1^J decreases), by the budget constraint F_1^J will increase, but according to the IBE it will decrease, which is a contradiction. For the same reason C_1^J cannot stay the same. Thus C_1^J will increase and F_1^J has to decrease. As there is a symmetry between C_1 and F_1 , we conclude that the argument is true for single-period problems.

Assume the statement holds true for an N -PE problem, $N > 1$. Then consider the case of an $(N + 1)$ -PE problem with the conventional notations

$$\begin{aligned}(X, R) &= (X_1, \dots, X_{N+1}, R_2, \dots, R_{N+1}), \\ \rho &= (C_1, \dots, C_{N+1}, F_{N+1}), \\ u' &= (u'_1, \dots, u'_{N+1}, u'_p), \\ \theta &= (\theta_1, \dots, \theta_{N+1}, \theta_p).\end{aligned}$$

First consider if some θ_n increases, $n < N + 1$. Then as we have discussed, this $(N + 1)$ -PE problem can be converted into an induced N -PE problem by truncation

at time point t_N and define h_N as has been defined in Theorem 2.3.5, that is, the 4-tuple

$$\begin{aligned}(X, R)_{[N]} &= (X_1, \dots, X_N, R_2, \dots, R_N), \\ \rho_{[N]} &= (C_1, \dots, C_N, F_N), \\ u'_{[N]} &= (u'_1, \dots, u'_N, h_N), \\ \theta_{[N]} &= (\theta_1, \dots, \theta_N, \theta_{N+1}).\end{aligned}$$

Consider this N -PE problem. According to the induction assumption, we will have that for any $J \in \mathcal{J}_{N+1}$, C_n^J will increase if θ_n increases, while other cash outflows will decrease. So F_N^J will decrease and so is A_{N+1}^J . Note that by definition the function g_{N+1} will stay the same if θ_n increases. Thus C_{N+1}^J and F_{N+1}^J will both decrease as they are increasing functions of A_{N+1}^J .

Now consider the situation if θ_{N+1} increases. We will show that $F_{N+1}^{J_{N+1}}$ will decrease. Otherwise (i.e. $F_{N+1}^{J_{N+1}}$ either increases or stays the same), by the final period IBE

$$\theta_{N+1} u'_{N+1}(C_{N+1}^{J_{N+1}}) = \theta_p u'_p(F_{N+1}^{J_{N+1}}) = L_{N+1}^{J_{N+1}}$$

we have that $C_{N+1}^{J_{N+1}}$ has to increase because of the monotonicity of u'_{N+1} and u'_p . Then by the budget constraint for that period

$$C_{N+1}^{J_{N+1}} + F_{N+1}^{J_{N+1}} = X_{N+1}^{j_{N+1}} + F_N^{J_N} R_{N+1}^{j_{N+1}}$$

$F_N^{J_N}$ also has to increase. This will lead to the fact that $L_{N+1}^{J_N j_{N+1}^*}$ will not increase for any $j_{N+1}^* \in \mathcal{J}_N^{N+1}$. This is because we have

$$L_{N+1}^{J_N j_{N+1}} = g_{N+1}(X_{N+1}^{j_{N+1}} + F_N^{J_N} R_{N+1}^{j_{N+1}})$$

and

$$L_{N+1}^{J_N j_{N+1}^*} = g_{N+1}(X_{N+1}^{j_{N+1}^*} + F_N^{J_N} R_{N+1}^{j_{N+1}^*})$$

which shows that $L_{N+1}^{J_N j_{N+1}}$ and $L_{N+1}^{J_N j_{N+1}^*}$ should have the same monotonicity property with respect to θ_{N+1} . The result is that $\mathbb{E}_N(L_{N+1}^{J_N} R_{N+1})$ will not increase.

According to the global budget constraint along that trajectory, there has to be at least one n such that $C_n^{J_n}$ will decrease. Let the set of such n 's be denoted by $\mathcal{T}$. Consider first the situation that $\max\{\mathcal{T}\} = N$. Then $L_N^{J_N} = \theta_N u'_N(C_N^{J_N})$ will increase. On the other hand, $\mathbb{E}_N(L_{N+1}^{J_N} R_{N+1})$ will not increase. We then arrive at

a contradiction by noting that by IBE we should have

$$L_N^{J_N} = \mathbb{E}_N(L_{N+1}^{J_N} R_{N+1}).$$

Then consider more generally that $\tau = \max\{\mathcal{T}\} < N$. Then as $F_N^{J_N}$ will increase and $C_N^{J_N}$ will not decrease, by budget constraint we know $F_{N-1}^{J_{N-1}}$ will increase. Repeat this reasoning until we get that $F_\tau^{J_\tau}$ will have to increase. Then by analogy as above we will have that $\mathbb{E}_\tau(L_{\tau+1}^{J_\tau} R_{\tau+1})$ will not increase. However, $L_\tau = \theta_\tau u'_\tau(C_\tau^{J_\tau})$ will increase as $C_\tau^{J_\tau}$ decreases. The IBE will then not hold. We conclude that $F_{N+1}^{J_{N+1}}$ will decrease and $L_{N+1}^{J_{N+1}}$ will increase. According to

$$L_n = \mathbb{E}^\mathbb{P}[L_{n+1} R_{n+1}]$$

we know that for any $n < N+1$, along the trajectory J_n which is the up-to-time- t_n part of J_{N+1} , $L_n^{J_n}$ will increase. Then $C_n^{J_n}$ will decrease since

$$L_n^{J_n} = \theta_n u'_n(C_n^{J_n}).$$

Finally, consider the global budget constraint (2.2.3) along the trajectory J_{N+1} . It must be that $C_{N+1}^{J_{N+1}}$ will have to increase since all the other C 's and F_{N+1} will decrease.

The case when only θ_p increases follows analogously as there is symmetry between C_{N+1} and F_{N+1} . This completes the proof. $\square$

A.2 Proofs for Section 2.5

Please note that some of the proofs in this section make use of the mapping φ defined in Section 2.6.

DEFINITION A.2.1 (*N*-PEFF Problem.) An *N*-PEFF problem refers to the 4-tuple $((X, R), \rho, u', v)$ and the corresponding equation systems (2.5.1), (2.5.2) and (2.5.3), where (X, R) is a vector of random variables, ρ a vector of decision variables, u' an $(N + 1)$ -tuple of stereotype marginal utility functions and v a value profile

vector, i.e.

$$\begin{aligned}(X, R) &= (X_1, \dots, X_N, R_2, \dots, R_N) \in \mathcal{L}^{2N+1}, \\ \rho &= (C_1, \dots, C_N, F_N) \in \mathcal{L}^{N+1}, \\ u' &= (u'_1, \dots, u'_N, u'_p), \\ v &= (v_1, \dots, v_N, v_p) \in \mathcal{V}.\end{aligned}$$

The set $\mathcal{V}$ is totally determined by (X, R) and u' according to Expression (2.4.3).

DEFINITION A.2.2 (*Hilbert metric on $\mathbb{R}_{++}^n$* .) The Hilbert metric defines a distance as

$$d(x, y) = \log \frac{\max_i \{x_i/y_i\}}{\min_i \{x_i/y_i\}}$$

for any $x, y \in \mathbb{R}_{++}^n$. It is not a real metric as

$$d(x, y) = 0 \quad \Leftrightarrow \quad \exists c \in \mathbb{R}^+ \quad \text{such that} \quad y = cx.$$

It will become a true metric if restricted on e.g. the open unit simplex in $\mathbb{R}_{++}^n$.

LEMMA A.2.3 *If $\phi : \mathbb{R}_{++}^n \rightarrow \mathbb{R}_{++}^n$ is homogeneous and strongly monotone, then ϕ is contractive with respect to the Hilbert metric.*

PROOF See for instance Lemma 4.5 in Pazdera et al [33]. □

Any contractive mapping ϕ can only have one fixed point. Suppose there are two, namely x and y with $d(x, y) > 0$. Then by contractiveness we have

$$d(x, y) = d(\phi(x), \phi(y)) < d(x, y)$$

which is contradictory. Then $d(x, y) = 0$. Note that the uniqueness is in the sense of Hilbert metric.

The following lemma is the key part of proving the uniqueness of the PEFF solution.

LEMMA A.2.4 *The mapping φ_1 defined in Section 2.6 is strictly increasing, i.e. for any trajectory $J \in \mathcal{J}_N$, we have that*

$$L_N^J(\theta') \geq L_N^J(\theta'') \quad \forall \theta' \geq \theta''.$$

PROOF To show this we only need to show that L_N^J is strictly increasing with respect to any one of the coordinates of θ . We can utilize Lemma 2.3.9.

Consider first that only θ_n increases while the other θ 's stay the same, $n = 1, \dots, N$. Then according to Lemma 2.3.9, F_N^J will decrease thus

$$L_N^J = \theta_p u'_p(F_N^J)$$

will increase. The case when only θ_p increases follows analogously as there is symmetry between C_N and F_N . $\square$

LEMMA A.2.5 (*The uniqueness of the PEFF rule.*) *For any given value profile $v = (v_1, \dots, v_N, v_p) \in \mathcal{V}$, the corresponding PEFF risk-sharing rule will be unique if it exists.*

PROOF The main point of this proof is to show that φ defined in Section 2.6 is homogeneous and strictly monotone thus by Lemma A.2.3 it can only have one fixed point (up to normalization) if it has.

The mapping is homogeneous by definition thus we only have to consider monotonicity. First, according to Lemma A.2.4 φ_1 is strictly increasing with respect to θ along all possible trajectories. Then L_n^J is also increasing since

$$L_n = \mathbb{E}_n^{\mathbb{P}}[L_{n+1}R_{n+1}].$$

Now consider $\theta' \geq \theta''$. Then for any $J \in \mathcal{J}_N$ we have that $L_N^J(\theta') > L_N^J(\theta'')$. For any possible n , the n -th coordinate of φ_2 : $\varphi_{2(n)}(L_N) = [\mathbb{E}^{\mathbb{Q}} I_n(\frac{L_n}{\cdot})]^{-1}(v_n)$ will lead to that $\varphi_{2(n)}(L_N(\theta')) > \varphi_{2(n)}(L_N(\theta''))$. This is because φ_2 will always require that

$$\mathbb{E}^{\mathbb{Q}} C_n = \mathbb{E}^{\mathbb{Q}} I_n\left(\frac{L_n}{\theta_n}\right) = \sum_{J \in \mathcal{J}_n} \mathbb{Q}(J) I_n\left(\frac{L_n^J}{\theta_n}\right) = v_n.$$

If L_n^J increases for all $J \in \mathcal{J}_n$, then θ_n also will increase according to this φ_2 . The result is that

$$\varphi(\theta') = \varphi_2(L_N(\theta')) > \varphi_2(L_N(\theta'')) = \varphi(\theta''),$$

i.e. φ is strictly increasing with respect to θ . $\square$

PROOF OF THEOREM 2.5.1. The proof uses mathematical induction. Note that we can always fix $\theta_p = 1$ as a normalization to the θ 's unless specified otherwise.

For any 1-PEFF problem, there is only one random variable X_1 to be shared. One needs to solve

$$\begin{aligned} C_1 + F_1 &= X_1, \\ \theta_1 u'_1(C_1) &= \theta_p u'_p(F_1), \\ \mathbb{E}^\mathbb{Q} C_1 &= v_1. \end{aligned}$$

For any given θ_1 , the equations of BC and IBE will jointly produce a certain risk sharing rule according to the mapping Φ in Theorem 2.3.5. However, the third FF equation may not hold. We need to show that there will exist some θ_1 such that the FF equation will hold. We define

$$w(\theta_1) = \mathbb{E}^\mathbb{Q} C_1 = \sum_{J \in \mathcal{J}_1} \mathbb{Q}(J) C_1^J.$$

It is a continuous function of θ_1 which follows as a property of the mapping Φ . Next we will show that the value of the function w can be both above and below v_1 , so that there exists some θ_1^* such that $w(\theta_1^*) = v_1$ since w is continuous. This will be done by taking θ_1 to the limits.

First consider $\lim_{\theta_1 \rightarrow 0} w(\theta_1)$. Then along any trajectory $J \in \mathcal{J}_1$ it must be that $C_1^J \rightarrow b_1 < v_1$. Otherwise, suppose there exists some sequence of θ_1 , say $\{\hat{\theta}^{[m]}\}$ with $\hat{\theta}^{[m]} \rightarrow 0$ as $m \rightarrow \infty$, such that

$$\lim_{m \rightarrow \infty} C_1^J(\hat{\theta}^{[m]}) \geq b_1 + \varepsilon$$

for some trajectory J and some $\varepsilon \in \mathbb{R}_{++}$. If $b_1 = -\infty$ then this is interpreted as bounded from below. Then according to the IBE

$$\theta_1 u'_1(C_1^J) = \theta_p u'_p(F_1^J)$$

the left hand side will go to zero as $u'_1(C_1^J)$ will be bounded. As a result, F_1^J will have to go to $+\infty$ which is not possible if we take into consideration the budget constraint. We conclude that $C_1^J \rightarrow b_1 < v_1$ along all the J 's if we let $\theta_1 \rightarrow 0$.

Next consider $\lim_{\theta_1 \rightarrow \infty} w(\theta_1)$. Now we drop the normalization constraint $\theta_p = 1$. Taking into consideration the freedom of choosing a way of normalization, it follows that the following two statements are equivalent:

- fix θ_p and let $\theta_1 \rightarrow +\infty$;
- fix θ_1 and let $\theta_p \rightarrow 0$.

Then following the analogy above we have $F_1^J \rightarrow b_p$ for all $J \in \mathcal{J}_1$ as $\theta_p \rightarrow 0$. Thus $\lim_{\theta_1 \rightarrow \infty} \mathbb{E}^{\mathbb{Q}} F_1 = b_p$ and according to the budget constraint

$$\lim_{\theta_1 \rightarrow \infty} w(\theta_1) = \lim_{\theta_1 \rightarrow \infty} \mathbb{E}^{\mathbb{Q}} C_1 = v_1 + v_p - b_p.$$

Then since

$$v_p - b_p > 0$$

must hold, we have

$$v_1 < v_1 + v_p - b_p.$$

By a simple intermediate value theorem we know that there will exist some θ_1^* such that $w(\theta_1^*) = v_1$. Then we have found a weight vector θ (i.e. $(\theta_1^*, \theta_p = 1)$) that leads to a PEFF solution to the system. This indicates that the fixed points of the mapping φ will exist; the fixed point must be unique according to Lemma A.2.5, i.e. the vector θ is unique. The uniqueness is up to normalization.

Let's assume that there always exists a unique solution for an N -PEFF problem, $N > 1$. Consider an $(N + 1)$ -PEFF problem using our conventional notations

$$\begin{aligned} (X, R) &= (X_1, \dots, X_{N+1}, R_2, \dots, R_{N+1}), \\ \rho &= (C_1, \dots, C_{N+1}, F_{N+1}), \\ u' &= (u'_1, \dots, u'_{N+1}, u'_p), \\ v &= (v_1, \dots, v_N, v_{N+1}, v_p) \end{aligned}$$

Consider the corresponding $(N + 1)$ -PE problem with some given weight θ . Use $\theta_p = 1$ as a normalization. As we have discussed, the whole system will degrade to an induced N -PEFF problem with F_N now being the “final” buffer whose risk aversion is characterized by h_N given by Theorem 2.3.5. That is,

$$\begin{aligned} (X, R)_{[N]} &= (X_1, \dots, X_N, R_2, \dots, R_N), \\ \rho_{[N]} &= (C_1, \dots, C_N, F_N), \\ u'_{[N]} &= (u'_1, \dots, u'_N, h_N), \\ v_{[N]} &= (v_1, \dots, v_N, \mathbb{E}^{\mathbb{Q}} F_N) \end{aligned}$$

where $\mathbb{E}^{\mathbb{Q}} F_N$ can be calculated according to the global budget constraint of the induced N -PEFF problem.

For any given θ_{N+1} , according to the assumption, the degraded system has a unique PEFF solution with coefficients $(\theta_1, \dots, \theta_N)$. This solution, together with

the θ_{N+1} and $\theta_p = 1$, satisfies all the equations except the following one²

$$\mathbb{E}^{\mathbb{Q}} C_{N+1} = \sum_{J \in \mathcal{J}_{N+1}} \mathbb{Q}(J) C_{N+1}^J = v_{N+1}.$$

Next we will show that there exists θ_{N+1} such that the equation above will hold. Then by Theorem A.2.5 the solution θ_{N+1} will be unique.

Define

$$w(\theta_{N+1}) = \mathbb{E}^{\mathbb{Q}} C_{N+1} = \sum_{J_{N+1} \in \mathcal{J}_{N+1}} \mathbb{Q}(J_{N+1}) C_{N+1}^{J_{N+1}}.$$

Note that $C_{N+1}^{J_{N+1}}$ is a continuous function of θ_{N+1} for any J which follows from Theorem 2.3.5 and so is w itself. Next we will show that the value of the function w can be both above and below v_{N+1} , so that there exists θ_{N+1}^* such that $w(\theta_{N+1}^*) = v_{N+1}$ since w is continuous. This will be done by taking θ_{N+1} to the limits.

First consider $\lim_{\theta_{N+1} \rightarrow 0} w(\theta_{N+1})$. We will distinguish between the following two cases.

A. The lower bounds of the utility functions b_n are all finite. We will then have

$$\lim_{\theta_{N+1} \rightarrow 0} C_{N+1}^{J_{N+1}} = b_{N+1} \quad \forall J_{N+1} \in \mathcal{J}_{N+1}.$$

Otherwise, suppose there exists a sequence of θ_{N+1} , say $\{\hat{\theta}^{[m]}\}$ with $\hat{\theta}^{[m]} \rightarrow 0$ as $m \rightarrow \infty$, such that

$$\lim_{m \rightarrow \infty} C_{N+1}^{J_{N+1}}(\hat{\theta}^{[m]}) \geq b_{N+1} + \varepsilon$$

for some trajectory J_{N+1} and some $\varepsilon \in \mathbb{R}_{++}$. Then according to the final period IBE

$$\theta_{N+1} u'_{N+1}(C_{N+1}^{J_{N+1}}) = \theta_p u'_p(F_{N+1}^{J_{N+1}})$$

the left hand side will go to zero as $u'_{N+1}(C_{N+1}^{J_{N+1}})$ will be bounded. As a result, $F_{N+1}^{J_{N+1}}$ will have to go to $+\infty$ which is not possible when all the C 's can only be finite.

B. Consider when $b_n = -\infty$ for some n . $\mathcal{T}$ denotes the set of all such n 's. We will show that still

$$\lim_{\theta_{N+1} \rightarrow 0} C_{N+1}^{J_{N+1}} = b_{N+1} \quad \forall J_{N+1} \in \mathcal{J}_{N+1}.$$

We then only have to show this for a special J' which satisfies that for any n , $X_n^{J'} = \max_{J \in \mathcal{J}_{N+1}} X_n^J$ and $R_n^{J'} = \max_{J \in \mathcal{J}_{N+1}} R_n^J$, i.e. $(X^{J'}, R^{J'})$ is the attainable

²The equation $\mathbb{E}^{\mathbb{Q}} F_{N+1} = v_p$ will also not hold. However, as we have discussed, we don't have to consider this equation, since it will be automatically satisfied if other FF conditions hold.

“upper bound” of all trajectories. This is possible because the number of trajectories is finite, the condition (2.2.1) holds and the risk stream is sequentially independent. Once we show that $\lim_{\theta_{N+1} \rightarrow 0} C_{N+1}^{J'} = b_{N+1}$, by Lemma 2.3.8, the limit of C_{N+1} of all other trajectories cannot be larger than b_{N+1} , and also cannot be smaller than b_{N+1} .

Otherwise, suppose there exist a sequence of θ_{N+1} , say $\{\hat{\theta}^{[m]}\}$ with $\hat{\theta}^{[m]} \rightarrow 0$ as $m \rightarrow \infty$, and $\varepsilon > 0$, such that

$$\lim_{m \rightarrow \infty} C_{N+1}^{J'}(\hat{\theta}^{[m]}) \geq b_{N+1} + \varepsilon.$$

If $b_{N+1} = -\infty$ then the equation above is interpreted as that the sequence $\{C_{N+1}^{J'}(\hat{\theta}^{[m]})\}$ is bounded from below.

Then by final period IBE

$$\theta_{N+1} u'_{N+1}(C_{N+1}^{J'}) = \theta_p u'_p(F_{N+1}^{J'})$$

we have that $F_{N+1}^{J'}$ will have to go to $+\infty$ since $u'_{N+1}(C_{N+1}^{J'})$ will be bounded. Consider the global budget constraint: now since $C_{N+1}^{J'} + F_{N+1}^{J'} \rightarrow +\infty$, there will exist $\tau \in \mathcal{T}$ such that $C_\tau^{J'} \rightarrow -\infty$. By the definition of J' , we have that $C_\tau^J \rightarrow -\infty$ for any other possible $J \in \mathcal{J}_{N+1}$, thus the value profile condition for C_τ will not hold. This is a contradiction.

To conclude: we have shown that

$$\lim_{\theta_{N+1} \rightarrow 0} C_{N+1}^J = b_{N+1} \quad \forall J \in \mathcal{J}_{N+1}$$

whatever the value of b_{N+1} is. Thus

$$w(\theta_{N+1}) = \mathbb{E}^\mathbb{Q} C_{N+1} = \sum_{J \in \mathcal{J}_{N+1}} \mathbb{Q}(J) C_{N+1}^J \rightarrow b_{N+1}.$$

Next consider $\lim_{\theta_{N+1} \rightarrow \infty} w(\theta_{N+1})$. Now we drop the normalization constraint $\theta_p = 1$. Taking into consideration the freedom of choosing a way of normalization, we can conclude that the following two statements are equivalent:

- fix θ_p and let $\theta_{N+1} \rightarrow +\infty$;
- fix θ_{N+1} and let $\theta_p \rightarrow 0$.

Then following the analogy we have $F_{N+1}^J \rightarrow b_p$ for all $J \in \mathcal{J}_{N+1}$ as $\theta_p \rightarrow 0$. Thus according to the budget constraint for the last period, we conclude that

$$\lim_{\theta_{N+1} \rightarrow \infty} w(\theta_{N+1}) = \lim_{\theta_{N+1} \rightarrow \infty} \sum_{J \in \mathcal{J}_{N+1}} \mathbb{Q}(J) C_{N+1}^J = v_{N+1} + v_p - b_p.$$

Then since

$$v_p - b_p > 0$$

must hold, we have

$$v_{N+1} < v_{N+1} + v_p - b_p.$$

By a simple intermediate value theorem we know that there will exist some θ_{N+1}^* such that $w(\theta_{N+1}^*) = v_{N+1}$. Then we have found a weight vector θ that leads to a PEFF solution to the system. This indicates that the fixed points of the mapping φ will exist; the fixed point must be unique according to Theorem A.2.5, i.e. the solution θ is unique. The uniqueness is up to normalization. This finishes the proof. $\square$

A.3 Proofs for Section 2.6

LEMMA A.3.1 *When $(\mathcal{X}, d)$ is a locally compact and connected metric space, and $f : \mathcal{X} \rightarrow \mathcal{X}$ is a contractive mapping with fixed point $x^* \in \mathcal{X}$, then for every $x \in \mathcal{X}$ the sequence of iterates $\{f^{(n)}(x) | n \in \mathbb{N}_+\}$ converges to x^* .*

PROOF See Thm. 1 by Nadler [31]. $\square$

PROOF OF THEOREM 2.6.1. Lemma A.2.5 has shown that the mapping φ is contractive with respect to the Hilbert metric. The theorem then is a direct result of Lemma A.3.1. $\square$

A.4 Proofs for Section 2.7.2

PROOF OF THEOREM 2.7.1. The proof of the theorem is actually a process of calculation. First we need the following preparations. By Theorem 2.3.5, for any given θ we can define $f_n(\cdot)$ such that

$$C_n = f_n(X_n + F_{n-1}R_n).$$

By the IBE for the last period we have

$$\theta_p u'_p(X_N + F_{N-1}R_N - C_N) = \theta_N u'_N(C_N),$$

which will translate into

$$\theta_p \alpha_p \exp[-\alpha_p(X_N + F_{N-1}R_N - C_N)] = \theta_N \alpha_N \exp[-\alpha_N C_N].$$

Take the logarithm on both sides and after rearranging the items we get

$$C_N = \frac{\alpha_p}{\alpha_p + \alpha_N}(X_N + F_{N-1}R_N) + \frac{1}{\alpha_p + \alpha_N} \ln \frac{\theta_N \alpha_N}{\theta_p \alpha_p}.$$

Take the $\mathbb{Q}$ -expectation and we shall have

$$\mathbb{E}^{\mathbb{Q}} C_N = v_N$$

which gives us

$$C_N = \frac{\alpha_p}{\alpha_p + \alpha_N} [X_N + F_{N-1}R_N] + (v_N - \frac{\alpha_p}{\alpha_p + \alpha_N} w_N),$$

where

$$w_N := \mathbb{E}^{\mathbb{Q}} A_N = \mathbb{E}^{\mathbb{Q}}(C_N + F_N) = v_N + v_p.$$

Next we show that for any possible n , f_n should be linear. We will show this by first showing that if f_{n+1} is linear with positive slope, then so is f_n .

By IBE

$$\theta_n u'_n(C_n) = \theta_{n+1} \mathbb{E}_n^{\mathbb{P}} [u'_{n+1}(C_{n+1}) R_{n+1}]$$

we have

$$\frac{\theta_n}{R_{n+1} \theta_{n+1}} u'_n(f_n(x)) = \sum_{j \in \mathcal{J}_n^{n+1}} \mathbb{P}(j) u'_{n+1} [f_{n+1}(X_{n+1}^j + (x - f_n(x)) R_{n+1})],$$

where x is the variable standing for the available assets.

Assume that $f_{n+1}(x) = a_{n+1}x + e_{n+1}$ with $a_{n+1} > 0$. We have

$$\begin{aligned} & \frac{\theta_n}{R_{n+1}\theta_{n+1}} \exp(-\alpha_n f_n(x)) \\ &= \sum_{j \in \mathcal{J}_n^{n+1}} \mathbb{P}(j) \exp \left\{ -\alpha_{n+1} [a_{n+1}(X_{n+1}^j + (x - f_n(x))R_{n+1}) + e_{n+1}] \right\} \\ &= \exp \{ -\alpha_{n+1} a_{n+1} R_{n+1} (x - f_n) \} \cdot \left\{ \sum_{j \in \mathcal{J}_n^{n+1}} \mathbb{P}(j) \exp [-\alpha_{n+1} (a_{n+1} X_{n+1}^j + e_{n+1})] \right\} \end{aligned}$$

Take the logarithm on both sides:

$$\ln \left(\frac{\theta_n}{R_{n+1}\theta_{n+1}} \right) - \alpha_n \cdot f_n = \ln \kappa_{n+1} - \alpha_{n+1} a_{n+1} R_{n+1} (x - f_n),$$

finally

$$f_n(x) = \frac{a_{n+1}\alpha_{n+1}R_{n+1}}{\alpha_n + a_{n+1}\alpha_{n+1}R_{n+1}}x + \frac{1}{\alpha_n + a_{n+1}\alpha_{n+1}R_{n+1}} \ln \left(\frac{\theta_n}{R_{n+1}\theta_{n+1}} \frac{1}{\kappa_{n+1}} \right)$$

where

$$\kappa_{n+1} = \frac{\mathbb{E}u'_{n+1}(f_{n+1}(X_{n+1}))}{\alpha_{n+1}}.$$

It follows that all f_n should be linear with positive slope since f_N is. The slope satisfies

$$a_n = \frac{a_{n+1}\alpha_{n+1}R_{n+1}}{\alpha_n + \alpha_{n+1}a_{n+1}R_{n+1}}.$$

By recursion we know that if we start with $a_N = \frac{\alpha_p}{\alpha_p + \alpha_N}$, then all the a_n 's can be calculated. Hence

$$C_n = f_n(X_n + F_{n-1}R_n) = a_n(X_n + F_{n-1}R_n) + \text{constant}.$$

Taking the expectation under $\mathbb{Q}$ immediately gives the constant part and finally we have

$$C_n = a_n(X_n + F_{n-1}R_n) + (v_n - a_n w_n),$$

where $w_n = \mathbb{E}^{\mathbb{Q}} A_n$ can be recursively calculated according to the relationship

$$A_{n+1} = X_{n+1} + (A_n - C_n)R_{n+1}.$$

□

A.5 Tables for Section 2.7.1

Tables A.1 – A.4 shows the distributions of the decision variables for Section 2.7.1.

X_1	$\mathbb{P}$ -prob.	$\mathbb{Q}$ -prob.	C_1 : OEB	C_1 : CEB	C_1 : au- tarky
1.2	0.6	0.5	1.0507	1.0704	1.2
0.8	0.4	0.5	0.9493	0.9296	0.8

TABLE A.1: Distributions of payments for agent 1

X_1	X_2	$\mathbb{P}$ -prob.	$\mathbb{Q}$ -prob.	C_2 : OEB	C_2 : CEB	C_1 : au- tarky
1.2	1.2	0.36	0.25	1.1197	1.1741	1.2
0.8	1.2	0.24	0.25	1.0181	1.0376	1.2
1.2	0.8	0.24	0.25	0.9818	0.9632	0.8
0.8	0.8	0.16	0.25	0.8801	0.8251	0.8

TABLE A.2: Distributions of payments for agent 2

X_1	X_2	X_3	$\mathbb{P}$ -prob.	$\mathbb{Q}$ -prob.	C_3 : OEB	C_3 : CEB	C_1 : au- tarky
1.2	1.2	1.2	0.216	0.125	1.2157	1.3556	1.2
0.8	1.2	1.2	0.144	0.125	1.1167	1.2327	1.2
1.2	0.8	1.2	0.144	0.125	1.0832	1.1665	1.2
0.8	0.8	1.2	0.096	0.125	0.9844	1.0452	1.2
1.2	1.2	0.8	0.144	0.125	1.0157	0.9556	0.8
0.8	1.2	0.8	0.096	0.125	0.9167	0.8327	0.8
1.2	0.8	0.8	0.096	0.125	0.8832	0.7665	0.8
0.8	0.8	0.8	0.064	0.125	0.7844	0.6452	0.8

TABLE A.3: Distributions of payments for agent 3

X_1	X_2	X_3	$\mathbb{P}$ -prob.	$\mathbb{Q}$ -prob.	F_3 : OEB	F_3 : CEB
1.2	1.2	1.2	0.216	0.125	1.2157	1
0.8	1.2	1.2	0.144	0.125	1.1167	1
1.2	0.8	1.2	0.144	0.125	1.0832	1
0.8	0.8	1.2	0.096	0.125	0.9844	1
1.2	1.2	0.8	0.144	0.125	1.0157	1
0.8	1.2	0.8	0.096	0.125	0.9167	1
1.2	0.8	0.8	0.096	0.125	0.8832	1
0.8	0.8	0.8	0.064	0.125	0.7844	1

TABLE A.4: Distributions of end buffer

Bibliography

- [1] Jean-Pierre Aubin. *Optima and equilibria: an introduction to nonlinear analysis*. Berlin: Springer, 2013.
- [2] Yves Balasko. "Budget-constrained Pareto-efficient allocations". *Journal of Economic Theory* 21.3 (1979), pp. 359–379.
- [3] Laurence Ball and N Gregory Mankiw. "Intergenerational risk sharing in the spirit of Arrow, Debreu, and Rawls, with applications to social security design". *Journal of Political Economy* 115.4 (2007), pp. 523–547.
- [4] Hailong Bao, Eduard Ponds, and Hans Schumacher. "Multi-period investment return allocation within a heterogeneous collective, with applications to collective defined-contribution plan design". *Working paper* (2016).
- [5] Hailong Bao, Eduard Ponds, and Hans Schumacher. "Multi-period risk sharing under financial fairness". *SSRN* (2015). URL: <http://ssrn.com/abstract=2690270>.
- [6] Hailong Bao, Roderick Molenaar, Eduard Ponds, and Hans Schumacher. "Intertemporal allocation of investment risk in the decumulation phase of a collective DC scheme". *Working paper* (2016).
- [7] Pauline Barrieu and Giacomo Scandolo. "General Pareto optimal allocations and applications to multi-period risks". *ASTIN Bulletin* 38.1 (2008), p. 105.
- [8] Dimitri P Bertsekas. *Dynamic programming and optimal control*. Vol. 1. 2. Belmont, MA: Athena Scientific, 1995.
- [9] Jan Bonenkamp, Lex Meijdam, Eduard Ponds, and Ed Westerhout. "Reinventing intergenerational risk sharing". *Netspar Panel Paper* 40 (2014).
- [10] Karl Borch. "Equilibrium in a Reinsurance Market". *Econometrica* 30.3 (1962), pp. 424–444.
- [11] Lans Bovenberg and Roel Mehlkopf. "Optimal design of funded pension schemes". *Annual Review of Economics* 6.1 (2014), pp. 445–474.

- [12] Lans Bovenberg, Roel Mehlkopf, and Theo Nijman. "Techniek achter persoonlijke pensioenrekeningen in de uitkeringsfase". *Netspar Occasional Paper* (2014).
- [13] Lans Bovenberg and Theo Nijman. "Persoonlijke pensioenrekeningen met risicodeling". *Netspar NEA Paper* 56 (2014).
- [14] Hans Bühlmann and William S Jewell. "Optimal risk exchanges". *ASTIN Bulletin* 10.03 (1979), pp. 243–262.
- [15] Jiajia Cui, Frank de Jong, and Eduard Ponds. "Intergenerational risk sharing within funded pension schemes". *Journal of Pension Economics & Finance* 10.1 (2011), p. 1.
- [16] Nick Draper. *A financial market model for the Netherlands*. CPB Background Document. 2014.
- [17] William H DuMouchel. "The Pareto optimality of an n-company reinsurance treaty". *Scandinavian Actuarial Journal* 1968.3-4 (1968), pp. 165–170.
- [18] Damir Filipovic. *Term-structure models: a graduate course*. Springer Finance, Berlin, 2009.
- [19] David Gale. "Fair division of a random harvest." *DTIC Document* (1977).
- [20] Christian Gollier. "Intergenerational risk-sharing and risk-taking of a pension fund". *Journal of Public Economics* 92.5 (2008), pp. 1463–1485.
- [21] Roger H Gordon and Hal R Varian. "Intergenerational risk sharing". *Journal of Public Economics* 37.2 (1988), pp. 185–202.
- [22] Anna Grebentchikova, Roderick Molenaar, Peter Schotman, and Bas Werker. "Default Life-Cycles for Retirement Savings". *Netspar Working paper* (2016).
- [23] Robert E Hall. "Stochastic implications of the life cycle-permanent income hypothesis: theory and evidence". *The Journal of Political Economy* 86.6 (1978), pp. 971–987.
- [24] Mark Heemskerk, René Maatman, and Bas Werker. "Heldere en harde pensioenrechten onder een PPR". *Netspar Design Paper* 46 (2016).
- [25] Roy Hoevenaars and Eduard Ponds. "Valuation of intergenerational transfers in funded collective pension schemes". *Insurance: Mathematics and Economics* 42.2 (2008), pp. 578–593.
- [26] Torsten Kleinow and Johannes M Schumacher. "Financial fairness and conditional indexation". *Scandinavian Actuarial Journal* (To appear).

- [27] Ralph Koijen, Theo Nijman, and Bas Werker. "Optimal annuity risk management". *Review of Finance* 15 (2011), pp. 799–833.
- [28] Dirk Krueger and Felix Kubler. "Pareto-improving social security reform when financial markets are incomplete!?" *The American Economic Review* 96.3 (2006), p. 737.
- [29] Robert C Merton. "Optimum consumption and portfolio rules in a continuous-time model". *Journal of economic theory* 3.4 (1971), pp. 373–413.
- [30] Sander Muns. "A financial market model for the Netherlands: A Methodological Refinement". *Netspar Discussion Paper* (2015).
- [31] Sam B Nadler. "A Note on an iterative test of Edelstein". *Canadian Mathematical Bulletin* 15.3 (1972), pp. 381–386.
- [32] Jaroslav Pazdera, Hans Schumacher, and Bas Werker. "Cooperative investment in incomplete markets under financial fairness". *Netspar Discussion Paper* (2014).
- [33] Jaroslav Pazdera, Johannes M Schumacher, and Bas Werker. "The composite iteration algorithm for finding efficient and financially fair risk-sharing rules". SSRN (2014). URL: <http://ssrn.com/abstract=2393288>.
- [34] Eduard Ponds. "Pension funds and value-based generational accounting". *Journal of Pension Economics and Finance* 2.03 (2003), pp. 295–325.
- [35] Monika Queisser and Edward R Whitehouse. *Neutral or fair?: actuarial concepts and pension-system design*. OECD Publishing. 2006.
- [36] Artur Raviv. "The design of an optimal insurance policy". *American Economic Review* 69.1 (1979), pp. 84–96.
- [37] R Tyrrell Rockafellar. *Convex analysis*. Princeton university press, 2015.
- [38] Sociaal-Economische Raad (SER). *Verkenning persoonlijk pensioenvermogen met collectieve risicodeling*. May, 2016. URL: https://www.ser.nl/~media/db_adviezen/2010_2019/2016/persoonlijk-pensioenvermogen.ashx.
- [39] Jana R. Steele, Angela Mazerolle, and Mel Bartlett. *Target-benefit plans in Canada – an innovation worth expanding*. July, 2016. URL: https://www.cdhowe.org/pdf/Commentary_411.pdf.
- [40] Tom Steenkamp. "Individueel defined contribution in de uitkeringsfase". *Netspar Design Paper* 50 (2016).

- [41] Ministerie SZW. *Concept wetsvoorstel variabele pensioenuitkering*. July 2015. URL: <https://www.rijksoverheid.nl/documenten/kamerstukken/2015/07/13/concept-wetsvoorstel-variabele-pensioenuitkering>.
- [42] Ministerie SZW. *Hoofdpijnennota optimalisering wettelijk kader voor premieovereenkomsten*. December 2014. URL: <https://www.rijksoverheid.nl/documenten/rapporten/2014/12/19/hoofdpijnennota-optimalisering-wettelijk-kader-voor-premieovereenkomsten>.
- [43] Coen N Teulings and Casper G De Vries. "Generational accounting, solidarity and pension losses". *De Economist* 154.1 (2006), pp. 63–83.
- [44] Bas Werker. *Tilburg Finance Tool (v 1.5.106)*. 2014. URL: https://www.tilburguniversity.edu/nl/webwijs/show/b.j.m.werker_nl.htm.
- [45] *Wetsvoorstel variabele pensioenuitkering, memorie van toelichting*. November 23, 2015. URL: <https://zoek.officielebekendmakingen.nl/kst-34344-3.pdf>.
- [46] *Wetsvoorstel verbeterde premieregeling*. March 10, 2016. URL: <https://zoek.officielebekendmakingen.nl/dossier/34255/kst-34255-A>.