Essays on Defined Benefit Pension Insurance and Participating Life Insurance

Inaugural-Dissertation
zur Erlangung des Grades eines Doktors
der Wirtschafts- und Gesellschaftswissenschaften
durch die
Rechts- und Staatswissenschaftliche Fakultät
der Rheinischen Friedrich-Wilhelms-Universität
Bonn

vorgelegt von

Chunli Cheng

aus Lanzhou (Provinz Gansu, Volksrepublik China)

Bonn 2015

Dekan:	Prof. Dr. Rainer Hüttemann
Erstreferent:	Prof. Dr. An Chen
Zweitreferent:	Prof. Dr. Stefan Ankirchner

Tag der mündlichen Prüfung: 11.12.2015

Acknowledgments

First I would like to thank my first supervisor Prof. Dr. An Chen for her continuous support during my entire PhD studies at the Bonn Graduate School of Economics. She inspired my interest in the research fields of pension insurance and life insurance. During the period when I was working on my first paper which Chapter 2 is now based on, she was so patient to discuss about the research question with me, and teach me how to develop a model for solving the question and write a scientific paper. She shared a lot of her research experience with me, which helped me quickly build up my research skills. Moreover, she always gave me very helpful remarks and comments to improve my papers. I am very appreciated for all of her help.

I want to thank my second supervisor Prof. Dr. Stefan Ankirchner. I have acquired a lot of very useful financial mathematics and technical knowledge in both his lecture and topics course at Bonn University. Besides, we discussed interesting research questions in the mathematical finance field a lot, which helped expand my research view in this field. Moreover, he always carefully read my papers and gave me very helpful comments and suggestions. I am very much appreciated for his support during my PhD studies.

I would also like to thank Prof. Dr. Peter Zweifel, whom I got to know in the 40th Seminar of the European Group of Risk and Insurance Economics. I am appreciated for the insightful conversations and valuable suggestions that he gave to me while I was revising my pension insurance paper which Chapter 2 is based on. Moreover, he patiently answered my questions on health shock and health insurance, which speeded me up in completing a paper that Chapter 4 is based on. He also encouraged me to do refereeing work and gave me instructions on how to do it. I am very grateful for all of what he has taught to me.

I want to thank Prof. Dr. Klaus Sandmann for his help and encouragement. He supported me to participate in many very good conferences during my entire PhD studies. I would also like to thank Prof. Dr. Lorens Imhof. He found time to discuss about my pension insurance paper and provided a proof of Proposition 1 in this paper. I am very grateful to him for very helpful suggestions and contirbuting the proof. I also thank Prof. Dr. Christian Bayer, Prof. Dr. Hendrik Hakenes, Prof. Dr. Harris Schlesinger, Prof. Dr. Hato Schmeiser and Prof. Dr. Alexander Szimayer for insightful conversations and very valuable remarks. Moreover, I would like to thank all my collegues at the Bonn

Graduate School of Economics. In particular, I want to thank Jing Li, my coauthor of a paper, which Chapter 3 is based on, for insightful discussions and great co-working experience; Filip Uzelac, with whom I worked on the pension insurance paper that Chapter 2 is based on; Matthias Meier, with whom I shared my office in the first year and who helped me a lot in and outside the university; and Bo Chen, Adrian Hillenbrand, Ariel Mecikovsky, Nazarii Salish, Alexander Scheer, Benjamin Schickner, Fabian Walders, and Mirjam Wuggenig, with whom I had great both studying and playing time during my entire PhD studies and who all the time gave me encouragement and help whenever I needed. All of them have become very good friends of mine and I feel so lucky to have them. I also want to thank my friend Greg Loutsenko for his support and interesting discussions.

I also thank Prof. Dr. Urs Schweizer, Prof. Dr. Benny Moldovanu and Dr. Silke Kinzig for their great organizational support at the Bonn Graduate School of Economics. Financial support from the German Research Foundation (DFG) through the Bonn Graduate School of Economics is gratefully acknowledged.

I would like to deeply thank my families, in particular, my parents for their moral support, endless encouragement and unconditional love to me during my entire PhD studies; and Christian Hilpert for moral support, great encouragement, and being with me in both happy and sad times.

Contents

List of Figures

List of Tables

1 Introduction

This dissertation analyzes two important types of insurance: defined benefit pension insurance in Chapter 2 and participating life insurance in Chapters 3 and 4.

Private defined benefit (DB) pension plans are insured in many countries by a pension guarantee fund (PGF), which is created by the government to protect pension beneficiaries (employees) from losing their promised pension benefits on retirement.[1] Sponsors (employers) are obliged to participate in the pension-protection scheme by paying a legally specified annual premium to the PGF. As a plan becomes underfunded and the employer is not able to salvage the plan without going out of business, a distress termination of the plan can be initiated, then the PGF steps in to cover its deficit.

However, the DB pension insurance has been abused by sponsors. It has been found that the PGF creates additional incentives for sponsors to underfund their DB pension plans and incur excessive investment risk in their pension portfolios, see e.g., Bodie et al. (1987) and Cooper and Ross (2003), thus taking advantage of the PGF. Together with some other reasons, for example the decline of stock market prices in 2000, the decrease of interest rates, and unfavourable demographic and employment trends, see e.g., Armstrong (2004), Wilcox (2006), and Brown (2008), the PGFs have been accumulating deficits over time and experiencing difficulty in performing the pension-protection function.[2] However, in view of its financial difficulty, the PGF has an involuntary termination of underfunded DB pension plans as its only possibility of intervention due to legislative rules that are placed on the DB pension insurance, see Kalra and Jain (1997). Rather than waiting for a DB plan to become severely underfunded, the PGF can terminate the contract, take over the plan, and cover the deficit.

Chapter 2 proposes a termination rule based on a critical funding ratio for a PGF

[1] Examples include pension guarantee funds in Sweden (created in 1960), Finland (1962), the United States (1974), Germany (1974), Canada (1980), Chile (1981), Switzerland (1985) and Japan (1989).

[2] Data on aggregate surplus and deficits of two pension guarantee funds, i.e., the Pension Benefit Guaranty Corporation in the United States, and the Pension Benefits Guarantee Fund in Canada, during a recent decade, are presented in Chapter 2.

who considers closing an underfunded pension plan.[3] This ratio is determined by solving an expected utility maximization problem on behalf of plan beneficiaries subject to two constraints designed to preserve the PGF's viability. The first constraint is an upper bound on the PGF's annual intervention probability; the second one is a restriction on the expected shortfall of an underfunded pension plan which is not closed.

A participating life insurance contract, being one of the most important life insurance products, employs a profit-sharing mechanism. In addition to a minimum interest guarantee, it promises a bonus payment which is linked to the insurance company's financial performance to policyholders upon survival and upon death. The contract is usually embedded with a surrender option by exercising which policyholders can sell the contract back to the company and receive surrender benefits.

Similar as beneficiaries in a DB pension plan, policyholders face a risk of receiving too little benefit as the insurance company declares bankruptcy at maturity. In order to protect policyholders, regulatory authorities impose early default mechanisms (protection schemes) to monitor insurance companies' financial status and close them before it is too late so.[4] As an early regulatory intervention is imposed by a regulator, an interesting and important question is how the policyholders who now bear early default risk of the insurance company will react on it, for example, by adjusting their surrender behaviors.

Since the early regulatory intervention can reframe the payment structure for the policyholders, as a response they may change their surrender decision-making behaviors, which in turn affects their contracts' payments and value. However, this has not been taken into account in the current literature. In Chapter 3, we analyze the impacts of the early default risk of the insurance company on the pricing of participating policies by endogenously modelling the policyholders' surrender behaviors and uncovering the impacts of the early default risk on the policyholders' surrender decision making.[5] Since policyholders may surrender their contracts for cash in emergency, corresponding to the Emergency Fund Hypothesis, see Kim (2013); and also due to limited financial information and pricing knowledge, not all the policyholders can evaluate their contracts correctly so as to surrender them optimally, see Li and Szimayer (2014), it is more reasonable to consider policyholders as partly rational from a purely financial point of view.

[3]This Chapter is based on Cheng and Uzelac (2015), an earlier version of which has become my Master's thesis.

[4]Examples of early default mechanisms include Solvency II and the Swiss Solvency Test implemented by the Swiss Financial Market Supervisory Authority, see Chapter 3 for more discussions.

[5]This Chapter is based on Cheng and Li (2015).

Policyholders' partial rationality is further linked to their mortality risk by taking into account a systemic health shock[6] in Chapter 4.[7] As the policyholders' normal health state switches to the impaired state due to the systemic health shock, two surrender change scenarios are proposed in this chapter: the first, policyholders hurry to access surrender values due to their threatening financial liquidity (surrender more likely in emergency)[8]; the second, policyholders become more careful in evaluating and surrendering their contracts (surrender more likely at an optimal time point). Based on the two surrender change scenarios, Chapter 4 analyzes the effects of the systemic health shock on the contract valuation, and further discusses the influence of the early default regulatory intervention on the contract value in the two scenarios.

[6]A systemic health shock harms the public health and lasts for a relatively long-time period. One example of the systemic health shock is the smog, which is a severe air pollution event and currently prevalent in Asia, see Chapter 4 for more discussions.

[7]This Chapter is based on Cheng (2015).

[8]As a policyholder's health gets damaged, a high demand of cash spent on necessary medical treatments and improving his living conditions threatens the policyholder's financial liquidity, see Chapter 4 for more discussions.

2 A Termination Rule for Pension Guarantee Funds

2.1 Introduction

A defined benefit (DB) pension plan promises pension beneficiaries (employees) a specified amount of benefits on retirement, which is determined by a formula that usually takes into account the beneficiaries' years of service with the sponsor (the employer), their salary, and their age. Contributions from the sponsor and future investment returns constitute the plan's assets. When assets fall short of accrued pension liabilities, the plan is underfunded, which exposes beneficiaries to the risk of failure to receive the full promised pension benefits on their retirement. Underfunding of DB pension plans in the U.S. and in other countries has been documented in the empirical literature, see e.g., Armstrong (2004), Franzoni and Marin (2006), Selody (2007), and Severinson (2008). Moreover, underfunding has been shown to be an equilibrium outcome in an imperfect financial market, see e.g., Ippolito (1985) and Cooper and Ross (2002). In an attempt to protect pension beneficiaries against loss of promised benefits, many developed countries have created a pension guarantee fund (PGF), also known as pension benefit guarantee insurance.[1] Examples are the Pension Benefit Guaranty Corporation (PBGC) in the United States (created in 1974), the Pension Benefits Guarantee Fund (PBGF) in Ontario, Canada (1980), as well as the guarantee funds in Sweden (1960), Finland (1962), Germany (1974), Chile (1981), Switzerland (1985) and Japan (1989).

The PGF charges the sponsor an annual premium, which is specified by law, for insuring pension benefits promised to beneficiaries. It can become active in two cases. In the first case, when the insured DB plan is underfunded while the sponsor is unable to salvage the plan without going out of business, the sponsor submits a distress termination application and waits for an approval by the PGF. In the second case, the PGF initiates a termination

[1] According to Jametti (2008), the introduction of pension benefit guarantee insurance was motivated by the underfunding problem of DB pension plans. The pension guarantee fund is one of three pension benefit security mechanisms; the other two are solvency requirements and sponsor support, see Broeders and Chen (2013).

of the underfunded pension fund for protecting pension beneficiaries, which is known as an involuntary termination in practice. In each type of termination, the PGF takes over the underfunded DB plan, using its own assets to pay the benefits of current and future retirees, up to a limit called a maximum guarantee. The maximum guarantee is set by law and updated periodically, e.g., annually in the U.S. However, due to a variety of reasons, which include the decline of stock market prices in 2000, the decrease of interest rates, excessive risk-taking by plan sponsors, and unfavourable demographic and employment trends, the PGFs have been experiencing difficulty in performing this function, see e.g., Armstrong (2004), Wilcox (2006), and Brown (2008). Table 2.1 shows the financial status of the PBGC in the U.S. and the PBGF in Ontario, Canada during a recent decade.[2]

	2003	2004	2005	2006	2007
PBGC(USD$)	$-11,499$	$-23,541$	$-23,111$	$-18,881$	$-14,066$
PBGF(CAD$)	-137.5	-107.2	-237.4	-274.2	-112.8
	2008	**2009**	**2010**	**2011**	**2012**
PBGC(USD$)	$-11,151$	$-21,946$	$-23,030$	$-26,036$	$-34,379$
PBGF(CAD$)	-102.2	-47.4	$+103.3$	-6.2	$+76.2$

Table 2.1: *Aggregate surplus* $(+)$ *and deficits* $(-)$ *of PBGC and PBGF from 2003 to 2012 (million).*

In view of the financial status of the PGF, a debate revolving around flaws in its design has sprung up. Thus far most of the academic literature has focused on the lacking incentive compatibility of premium calculation,[3] proposing risk-based premiums instead, see e.g., Lewis and Pennachi (1994), and Chen (2011). However, Kalra and Jain (1997) argue that by law pension benefit guarantors[4] are prevented from controlling the riskiness of investments by pension plans and from adjusting their premiums accordingly. This leaves the involuntary termination of underfunded DB pension plans as their only possibility of intervention.[5] Rather than waiting for a DB plan to become severely underfunded, a PGF can terminate the contract, take over the plan, and cover the deficits.

From the economic point of view, two main arguments support this type of intervention. First, there is a moral hazard problem between the sponsor and the pension benefit

[2]See the PBGC and PBGF's annual financial reports, `http://www.pbgc.gov` and `http://www.fsco.gov.on.ca/EN/PENSIONS/Pbgf/Pages/default.aspx`.

[3]Nevertheless, a variable premium component that takes the degree of underfunding of pension funds into account has been introduced in the U.S., the major part of PBGC's premium are flat. In particular, 70 percent of PBGC's premium were flat in 2010, see `www.pbgc.gov`.

[4]The terms "pension benefit guarantors" and "pension guarantee funds" are equivalent in this paper.

[5]The PBGC states that initiating a termination helps to protect the interests of plan beneficiaries as well as its insurance program. However, it does not specify a termination rule, see `www.pbgc.gov`.

guarantor. Cooper and Ross (2003) find that a PGF creates additional incentives for sponsors to underfund their DB pension plans and to incur excessive investment risk in their pension portfolios. The threat of termination can mitigate these adverse incentives in two cases in particular: sponsors might be tempted to lower current cash payments to their employees, promising them higher future pensions in return thanks to (risky) investments; and they may fail to account for taxes due. Second, financially troubled sponsors are likely to understate their liabilities (Bodie et al. 1987)[6], exposing the PGF to a considerable insolvency risk. Therefore, an involuntary termination can substantially reduce the amount of liability falling on the PGF.

These considerations raise the issue of the proper timing of intervention. In particular, how strongly underfunded should a pension fund be in order to be terminated by the PGF? On the one hand, terminating too early may deprive the pension fund of the opportunity to recover and the beneficiaries of the opportunity to receive full promised pension benefits (recall that they are capped if the PGF takes over). On the other hand, terminating too late may result in higher liabilities for the PGF, jeopardizing its financial equilibrium.

This paper proposes a termination rule based on a critical funding ratio to limit the PGF's own risk of insolvency. It uses an expected utility maximization framework for determining a termination threshold, taking into account both risk aversion of pension beneficiaries and two constraints on the PGF's intervention. First, the intervention probability constraint (IPC) puts a limit on the frequency with which the PGF can step in to cover the pension plan's deficits. Second, the expected shortfall constraint (ESC) limits the value of expected deficits of the underfunded pension plan. This specification has advantages of both identifying pension funds with the highest cost of insolvency, see Doff (2008), and protecting the PGF's future viability. It also reflects U.S. law, which states that the PBGC may terminate a pension plan to avoid being burdened by excessive losses.

The critical funding ratio is derived from a model with a one-year planning horizon by maximizing a power utility function designed to capture the interests of pension beneficiaries. This utility function becomes the objective function of the PGF on the grounds that its mission is to protect pension beneficiaries. The use of the DB plan's funding ratio in the utility function can be motivated in two ways. First, it is a quantity commonly

[6]A popular subterfuge has been to use excessive interest rates for discounting pension liabilities, see Bodie et al. (1987).

used by the industry and in regulatory practice.[7] Second, it reflects the fact that what really matters in pension fund management is not the value of assets *per se*, but how their value relative to the fund's liabilities, see Martellini and Milchau (2008). By determining an optimal critical funding ratio one solves an optimal stopping problem, which has been addressed in the insurance literature. For example, Grosen and Jørgensen (2000), Bacinello (2003), Bauer et al. (2006), Nordahl (2008), and Schmeiser and Wagner (2011) evaluate life insurance contracts with embedded surrender options, which are exercised by policyholders at an optimal point in time.

Work closely related to the present paper includes Acharya and Dreyfus (1989) and Kalra and Jain (1997). Acharya and Dreyfus determine jointly deposit insurance premiums and optimal termination policies applied to financial institutions. They derive an optimal termination threshold in terms of an assets-to-deposits ratio, below which an ailing financial institution should be closed, by minimizing the deposit insurer's net liability. Kalra and Jain adopt the PBGC's point of view; they propose an optimal termination policy that takes into account restrictions on the DB pension insurer's intervention, in particular regarding a premium adjustment. They view the pension benefit guarantee insurance as a down-and-out put option, with which the PBGC takes over a plan if the losses from terminating it are smaller than the losses from letting the plan continue. Hence, termination is governed by profit maximization. By contrast, this paper recognizes the fact that a PGF, being a government agency, has a welfare motivation, aiming at preserving pensions in the interest of beneficiaries.[8] Accordingly, its objective differs from that of a value-maximizing insurance company. This calls for evaluating the PGF's intervention in terms of pension beneficiaries' utility. Additionally, two novel restrictions importantly governing the PGF's financial viability are introduced and a different mechanism of solving the constrained utility maximization problem is used for determining the optimal termination policy.

The remainder of this paper is organized as follows. In Section 2.2, the termination rule is stated and the one-year expected utility objective and the two constraints are introduced. Next, the utility maximization problem is simplified by deriving closed-form expressions for both the expected utility and the two constraints, which are shown to have monotonicity properties. Section 2.3 provides a numerical illustration of expected utility for beneficiaries with different degrees of relative risk aversion over a reasonably

[7]For example, Dutch DB pension plans are obliged to keep their one-year ahead funding ratio at 1.05 or higher.

[8]Salisbury (1996) even views the PGF as a social insurance program that intentionally subsidizes the DB system. The welfare motivation is reflected in the PBGC's statutory mandate, see `www.pbgc.gov`.

wide range of parameter values for the drift, the volatility, and the initial funding ratio determining beneficiaries' one-year expected utility. Preferences regarding intervention by the PGF turn out to consistently depend on relative risk aversion. Section 2.4 focuses on the critical value of risk aversion at which one-year expected utility curve inverts and relates this value to the parameters characterizing the funding ratio process. Also, a functional form for the critical level of risk aversion is conjectured and proved analytically. Solutions to the optimization problem first subject to the IPC only and then, subject to both the IPC and ESC are presented in Section 2.5. Section 2.6 concludes the paper.

2.2 Model framework

2.2.1 A termination rule

Consider a DB pension plan for a homogeneous group of employees which can be represented by a typical beneficiary. This DB pension plan is insured at time $t_0 = 0$ by the PGF. The sponsoring company is charged by an annual premium, which is specified by law on an annual basis. The funding ratio of the plan at time t is denoted by R_t, which is assumed to follow a geometric Brownian motion under the market probability measure $\mathbb{P}$. Therefore,

$$\mathrm{d}R_t = \mu R_t \,\mathrm{d}t + \sigma R_t \,\mathrm{d}W_t^{\mathbb{P}}, \qquad R_0 > 0\,. \tag{2.1}$$

Here $W^{\mathbb{P}}$ is a standard Brownian motion under $\mathbb{P}$, while μ and $\sigma > 0$ denote the constant drift and volatility coefficients, respectively.

The termination rule that the PGF uses to intervene specifies a termination-triggering funding ratio (termination ratio for short) η at time t_0. Once the pension plan's funding ratio falls below η, the plan is closed and taken over by the PGF. In this event, the PGF uses its own assets, which consist of collected premiums and investment returns[9], to cover the deficits of the pension plan and to pay benefits, up to a guarantee limit set by law and adjusted periodically. The time of intervention τ is the first time the pension plan's funding ratio drops below the termination ratio,

$$\tau := \inf\{t \geqslant 0 \,|\, R_t \leqslant \eta\}. \tag{2.2}$$

The higher the termination ratio, the more likely the DB pension plan is closed by the PGF. The PGF is assumed to assess its solvency status[10] and its termination ratio

[9] We abstract from the investment strategy of the PGF, which influences its ability to pay benefits.

[10] While PGF's solvency status influences the feasible set of termination ratios under two constraints introduced in Section 2.3 below, maximizing solvency is not the objective of the PGF in this paper.

annually[11], taking into account its annual premium income. We abstract from any influence the distress termination application initiated by the sponsoring company might have on the PGF's intervention rule, i.e., its termination ratio. The PGF's termination is therefore triggered by the funding ratio of the pension plan only. In order to exclude the case $\tau = +\infty$, i.e., termination never happens, $\eta > 0$ is assumed. Conversely, $\eta < R_0$ is to make sure that the pension plan is not terminated at t_0, the time of issue of the contract (or reset of the termination ratio, respectively). Finally, $\eta < 1$ because it does not make sense for the PGF to close a plan that is fully funded, i.e., $R_t \geqslant 1$. To summarize, the PGF chooses the termination ratio η from the set $\mathbf{H} \equiv (0, \min\{R_0, 1\})$.

2.2.2 One-year expected utility

As stated in the introduction, the PGF pursues a welfare motive, which is reflected in its mission of preserving pension plans and protecting employees against loss of their pensions. Accordingly, the PGF is seen as a perfect agent of the pension beneficiary, implying that it adopts the beneficiary's utility function as its objective function. The higher the funding ratio, the more confident beneficiaries can be to get the promised benefits on retirement. They are characterized by a power utility function with the funding ratio of the DB plan and their degree of relative risk aversion (symbolized by δ, see below) involved.

Given that the PGF sets the termination ratio annually, one-year expected utility is the appropriate criterion. If termination is triggered within the year, it depends on the funding ratio at intervention time R_τ. Since the funding ratio is modelled as a continuous stochastic process, R_τ coincides with the termination ratio η in this case, i.e., if $\tau \leqslant 1$, expected utility is given by $E[U(R_\tau)] = E\left[\frac{\eta^{(1-\delta)}}{1-\delta}\right]$. For $\tau > 1$, the funding ratio at year one is the independent variable, yielding expected utility $E[U(R_1)] = E\left[\frac{R_1^{(1-\delta)}}{1-\delta}\right]$. One-year expected utility thus consists of two parts, conditional on whether or not termination occurs within the year. It can be written compactly as

$$E[U(R_{\tau\wedge 1})] = E\left[\frac{\mathbb{1}_{\{\tau\leqslant 1\}}\eta^{(1-\delta)}}{1-\delta}\right] + E\left[\frac{\mathbb{1}_{\{\tau>1\}}R_1^{(1-\delta)}}{1-\delta}\right], \qquad \delta \geqslant 0,\ \delta \neq 1 \qquad (2.3)$$

where $\tau \wedge 1$ denotes the minimum of τ and 1, and $\mathbb{1}_X$ is the indicator function for a set X.

[11] The time of contract issue $t_0 = 0$ can actually be considered as the time of when the PGF sets the termination ratio.

2.2.3 Two constraints on the termination ratio

In this section, we incorporate two constraints on the termination ratio designed to limit the PGF's insolvency risk. The first is an intervention probability constraint (IPC), which imposes an upper bound on the PGF's annual probability of intervention. Lacking such a constraint, the PGF would easily run out of funds. Define the minimum value of the funding ratio R_t between $t_0 = 0$ and $t_1 = 1$ by X_R,

$$X_R := \inf_{t \in [0,1]} \{R_t\}. \tag{2.4}$$

Then, the IPC states,

$$P(\tau \leqslant 1) = P(X_R \leqslant \eta) \leqslant \epsilon, \tag{2.5}$$

where $\epsilon \in (0, 1]$ that is set by the PGF taking into account its current financial situation and expected pension payment upon intervention, e.g., the maximum guarantee payment. If the PGF is currently not in good financial shape, lowering ϵ works as a solvency cushion because it causes a postponement of intervention.

However, a pension plan, that is not terminated within one year but has a funding ratio below one, is in deficit. Note that the amount of deficits has no upper limit since R_t measures the ratio of the pension plan's assets to its total liabilities. Terminating a plan in deficit can worsen the future financial position of the PGF. To mitigate this drawback of the IPC, consider a second constraint, the expected shortfall constraint (ESC). The ESC seeks to limit the 'long-run' insolvency risk of the PGF by putting a restriction on the size of expected deficits of an ongoing but underfunded pension plan at year one. For simplicity, assume the DB plan's liabilities to be constant within one year. Then, the expected deficit must not exceed a certain percentage $q > 0$ of the plan's liabilities, with q set by the PGF,

$$E\left[(1 - R_1)\mathbb{1}_{\{\tau > 1\}}\mathbb{1}_{\{R_1 \leqslant 1\}}\right] \leqslant q. \tag{2.6}$$

The ESC is also considered by Shi and Werker (2012).[12] It protects the PGF's future financial status by limiting expected pension coverage for an ongoing underfunded pension plan. This is of particular importance in the case of a large pension fund, where a given value $R_t = \eta$ implies large potential liabilities for the PGF.

[12]However, they consider the restriction on the induced expected shortfall subject to a VaR constraint.

2.2.4 Simplifying the utility maximization problem

In this section we formally state the closed-form expressions for the one-year expected utility and the two constraints, resulting in the two constrained utility maximization problems which are numerically analyzed in Section 2.3. At this point, the one-year expected utility is defined,

$$\begin{aligned} &E\left[\frac{\mathbb{1}_{\{\tau\leqslant 1\}}\eta^{(1-\delta)}}{1-\delta}\right]+E\left[\frac{\mathbb{1}_{\{\tau>1\}}R_1^{(1-\delta)}}{1-\delta}\right] \\ &=\frac{\eta^{(1-\delta)}}{1-\delta}P(X_R\leqslant\eta)+\frac{1}{1-\delta}\int_\eta^{R_0}\int_\eta^\infty r^{(1-\delta)}f_{(R_1,X_R)}(r,x)drdx, \end{aligned} \tag{2.7}$$

where $f_{(R_1,X_R)}$ denotes the joint density function of (R_1, X_R). By applying the closed-form expressions of $P(X_R \leqslant \eta)$ and $f_{(R_1,X_R)}$, summarized in Appendix 2.7.1, to the function (2.7), one obtains

$$\begin{aligned} &E\left[\frac{\mathbb{1}_{\{\tau\leqslant 1\}}\eta^{(1-\delta)}}{1-\delta}\right]+E\left[\frac{\mathbb{1}_{\{\tau>1\}}R_1^{(1-\delta)}}{1-\delta}\right] \\ &=\frac{1}{1-\delta}\left(\eta^{(1-\delta)}\left[\Phi\left(\frac{1}{\sigma}\ln\frac{\eta}{R_0}-\frac{\mu}{\sigma}+\frac{1}{2}\sigma\right)+\left(\frac{\eta}{R_0}\right)^{\left(\frac{2\mu}{\sigma^2}-1\right)}\Phi\left(\frac{1}{\sigma}\ln\frac{\eta}{R_0}+\frac{\mu}{\sigma}-\frac{1}{2}\sigma\right)\right]\right. \\ &\quad+R_0^{(1-\delta)}e^{(1-\delta)\left(\mu-\frac{1}{2}\delta\sigma^2\right)}\left[\Phi\left(-\frac{1}{\sigma}\ln\frac{\eta}{R_0}+\frac{\mu}{\sigma}+(\frac{1}{2}-\delta)\sigma\right)\right. \\ &\quad\left.\left.-\left(\frac{\eta}{R_0}\right)^{\left(\frac{2\mu}{\sigma^2}+1-2\delta\right)}\Phi\left(\frac{1}{\sigma}\ln\frac{\eta}{R_0}+\frac{\mu}{\sigma}+(\frac{1}{2}-\delta)\sigma\right)\right]\right), \\ &=: U(\eta,\delta,\mu,\sigma,R_0), \qquad \delta\geqslant 0,\ \delta\neq 1 \end{aligned} \tag{2.8}$$

where $\Phi(.)$ and $\phi(.)$ are the distribution function and the density function of a standard normal variable, respectively. The one-year expected utility is denoted by U, which is a continuously differentiable function of the termination ratio η for given values of the relative risk aversion parameter δ, the drift parameter μ, the volatility σ of the funding ratio process, and the plan's initial funding ratio R_0.

Next, the one-year intervention probability and the expected shortfall are expressed in closed form and their monotonicity properties stated.

Lemma 1 *The one-year intervention probability has the following closed-form expression*

$$\begin{aligned}
&P(\tau \leqslant 1) = P(X_R \leqslant \eta) \\
&= \Phi\left(\frac{1}{\sigma}\ln\frac{\eta}{R_0} - \frac{\mu}{\sigma} + \frac{1}{2}\sigma\right) + \left(\frac{\eta}{R_0}\right)^{\left(\frac{2\mu}{\sigma^2}-1\right)} \Phi\left(\frac{1}{\sigma}\ln\frac{\eta}{R_0} + \frac{\mu}{\sigma} - \frac{1}{2}\sigma\right) \\
&=: h_{IP}(\eta, \mu, \sigma, R_0),
\end{aligned} \tag{2.9}$$

where the one-year intervention probability, denoted by h_{IP}, *is continuous and monotonically increasing from* 0 *to* 1 *as* η *increases from* 0 *to* R_0 *for given values of* (μ, σ, R_0).

If only the IPC ($h_{IP} \leqslant \epsilon$) is imposed, giving rise to the constraint set

$$C_{IPC}(\mu, \sigma, R_0) \equiv \mathbf{H} \cap \{\eta : h_{IP}(\eta, \mu, \sigma, R_0) \leqslant \epsilon\}, \tag{2.10}$$

then the one-year expected utility maximization problem can be compactly written as

$$\max_{\eta \in C_{IPC}(\mu,\sigma,R_0)} U(\eta, \delta, \mu, \sigma, R_0). \tag{2.11}$$

Lemma 2 *The expected shortfall of an ongoing underfunded DB pension plan at time* $t = 1$ *has the following closed-form expression,*

$$\begin{aligned}
&E\left[(1 - R_1)\mathbb{1}_{\{\tau>1\}}\mathbb{1}_{\{R_1\leqslant 1\}}\right] = \Phi\left(-\frac{1}{\sigma}\ln R_0 - \frac{\mu}{\sigma} + \frac{1}{2}\sigma\right) - \Phi\left(\frac{1}{\sigma}\ln\frac{\eta}{R_0} - \frac{\mu}{\sigma} + \frac{1}{2}\sigma\right) \\
&- \left(\frac{\eta}{R_0}\right)^{\left(\frac{2\mu}{\sigma^2}-1\right)} \left[\Phi\left(\frac{1}{\sigma}\ln\frac{\eta}{R_0} + \frac{\mu}{\sigma} - \frac{1}{2}\sigma\right) - \Phi\left(\frac{1}{\sigma}\ln\frac{\eta^2}{R_0} + \frac{\mu}{\sigma} - \frac{1}{2}\sigma\right)\right] \\
&- R_0 e^{\mu}\left[\Phi\left(-\frac{1}{\sigma}\ln R_0 - \frac{\mu}{\sigma} - \frac{1}{2}\sigma\right) - \Phi\left(\frac{1}{\sigma}\ln\frac{\eta}{R_0} - \frac{\mu}{\sigma} - \frac{1}{2}\sigma\right)\right] \\
&+ R_0 e^{\mu}\left(\frac{\eta}{R_0}\right)^{\left(\frac{2\mu}{\sigma^2}+1\right)} \left[\Phi\left(\frac{1}{\sigma}\ln\frac{\eta}{R_0} + \frac{\mu}{\sigma} + \frac{1}{2}\sigma\right) - \Phi\left(\frac{1}{\sigma}\ln\frac{\eta^2}{R_0} + \frac{\mu}{\sigma} + \frac{1}{2}\sigma\right)\right] \\
&=: h_{ES}(\eta, \mu, \sigma, R_0),
\end{aligned} \tag{2.12}$$

where the expected shortfall, denoted by h_{ES}, *is continuous and monotonically decreasing as the termination ratio* η *increases on the interval* $\mathbf{H} \equiv (0, \min\{R_0, 1\})$, *with* $\lim_{\eta\to 0} h_{ES}(\eta, \mu, \sigma, R_0)$ $= \Phi\left(-\frac{1}{\sigma}\ln R_0 - \frac{\mu}{\sigma} + \frac{1}{2}\sigma\right) - R_0 e^{\mu}\Phi\left(-\frac{1}{\sigma}\ln R_0 - \frac{\mu}{\sigma} - \frac{1}{2}\sigma\right)$ *and* $\lim_{\eta\to\min\{R_0,1\}} h_{ES}(\eta, \mu, \sigma, R_0) = 0$ *for given values of* (μ, σ, R_0).

For the derivation of $h_{ES}(\eta, \mu, \sigma, R_0)$ and proof of its monotonicity property, see Appendix 2.7.2.

Note that according to Lemma 1 and Lemma 2, adjustments of the termination ratio η have opposite impacts on the two constraints given fixed values of (μ, σ, R_0). Setting η at a lower value decreases the one-year intervention probability while increasing the expected shortfall. This implies that the sets of termination ratios compatible with the two constraints may not overlap. From a policy point of view, this raises the issue of finding a second-best solution in case the two constraints are not satisfied simultaneously. Hence, the PGF faces a trade-off between controlling its current and future insolvencies. As a second-best solution, we propose a termination ratio which satisfies the IPC but violates the ESC to the least extent possible so that the PGF's financial capacity to protect pension beneficiaries in the long run is preserved. Thus, we give priority to the IPC while minimizing the risk of a future insolvency of the PGF. Since the expected shortfall is monotonically decreasing as the termination ratio increases (see Lemma 2), the second-best solution amounts to the highest termination ratio in the constraint set (2.10). At this termination ratio, IPC binds due to monotonically increasing property of the one-year intervention probability function (see Lemma 1). Therefore, this termination ratio is $\eta_\epsilon = h_{IP}^{-1}(\epsilon)$.

Hence, by defining the constraint set when both the IPC and ESC are incorporated

$$C_{IPC\&ESC}(\mu, \sigma, R_0) \equiv \{\eta_\epsilon\} \cup \left(\{\eta : h_{IP}(\eta, \mu, \sigma, R_0) \leqslant \epsilon\} \cap \{\eta : h_{ES}(\eta, \mu, \sigma, R_0) \leqslant q\} \right), \tag{2.13}$$

the one-year expected utility maximization problem can be compactly written as

$$\max_{\eta \in C_{IPC\&ESC}(\mu,\sigma,R_0)} U(\eta, \delta, \mu, \sigma, R_0). \tag{2.14}$$

Before numerically solving the utility maximization problems presented above, it is appropriate to point out the limitations of this model. First, the expected utility function refers to a group of beneficiaries. This raises problems from the point of view of decision theory since there is no mechanism to aggregate individual utility functions. Second, as is known in practice, different groups of beneficiaries are covered by a DB plan. Therefore, this model neglects the following two complications: there may be conflicting interests of beneficiaries linked to heterogeneous portfolios, see Zweifel and Auckenthalter (2008) for the consequences of such conflicts of interest; the PGF takes over at one point in time rather than a set of times, which would reflect heterogeneous preferences of beneficiaries. Third, since beneficiaries may die and surrender their contracts during the pension fund's accumulation period, the solution to their utility optimization problem depends not only on state preferences but also time preferences. However, when calculating the expected utility given by equation (2.3), this complication is neglected.[13]

2.3 Numerical illustration

For solving the constrained maximization problem (2.11) and (2.14), the properties of the objective function are crucial. However, the expected utility function (2.8) is too complex to be analyzed in closed form. Therefore, an illustration of numerically solving the problem over a reasonably wide range of parameter values is performed in this section. Initial funding ratios for a DB pension fund suggested in Siegmann (2011) are between 0.87 and 1.2. Due to a lack of data for estimating the drift and volatility of the funding ratio process, values ranging from -0.04 to 0.04 and 0.05 to 0.5, respectively are used. The relationship between the optimal termination ratio and the degree of risk aversion characterizing pension beneficiaries is of particular interest. Therefore, relative risk aversion varies between 0 to 5, in keeping with empirically estimated values, see e.g., Szpiro (1986), Jackwerth (2000), Chetty (2006), and Fajardo, Ornelas and De Farias (2012). The parameter values employed are summarized in Table 2.2.

Funding Ratio Process Parameters	Relative Risk Aversion Parameter
$R_0 \in [0.87, 1.2]$	$\delta \in [0, 1) \cup (1, 5]$
$\mu \in [-0.04, 0.04]$	
$\sigma \in [0.05, 0.5]$	

Table 2.2: *Parameter values for the funding ratio process and relative risk aversion.*

It is important to understand the shape of the one-year expected utility function over the constraint set. Note that the constraint set (2.10) with the IPC only is a superset of the constraint set (2.13) with the second constraint ESC additionally imposed. As a first step, we focus on the utility maximization problem (2.11) only subject to the IPC as defined in (2.10). Let the intervention probability ϵ be 0.025 for computing the feasible set $\eta \in C_{IPC}(\mu, \sigma, R_0)$.[14] Then, for each combination of parameter values (μ, σ, R_0) of Table 2.2 values of the partial derivative of $U(\eta, \delta, \mu, \sigma, R_0)$ w.r.t. the termination ratio η

[13] As an alternative to abstracting from death/surrender of the beneficiary, one could assume an average mortality/surrender propensity which is independent of termination, with payout to the heir/beneficiary invested in a portfolio. If the portfolio develops according to the dynamics given in (2.1) and investment returns are valued in the same way, the optimization problem again involves state-dependent preferences only.

[14] The solvency test for Dutch pension funds uses 2.5% as the maximum allowable probability of the funding ratio dropping below 100% over a one-year horizon.

can be computed over the constraint set $\eta \in C_{IPC}(\mu, \sigma, R_0)$ for a given value of δ.[15] This procedure is repeated for $\delta \in [0, 1) \cup (1, 5]$.[16]

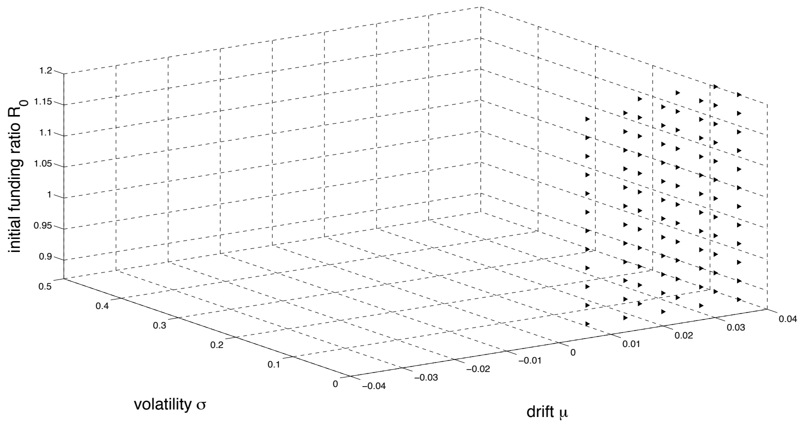
The results of these computations are summarized in Figure 2.1. A "▸" in panel (a) of Figure 2.1 marks the parameter triples (μ, σ, R_0) for which the partial derivative $\frac{\partial U}{\partial \eta}$ is negative over the constraint set for all values of δ. One-year expected utility is decreasing over the constraint set regardless of beneficiaries' risk aversion. In the interest of beneficiaries, the PGF should set the critical funding ratio at a low value, thus avoiding termination most of the time in this case. This occurs when the drift is positive and the volatility is very small ($\sigma < 0.1$). Intuitively, a positive expected change in a DB plan's funding ratio combined with little uncertainty about its future development makes it very likely that the plan will be fully funded again in one year's time. Hence, regardless of their degree of risk aversion, beneficiaries prefer the PGF not to intervene since this would result in a cap on their benefits.

Next, a "◂" in panel (b) of Figure 2.1 marks the parameter triples (μ, σ, R_0) for which $\frac{\partial U}{\partial \eta}$ is positive over the constraint set regardless of beneficiaries' risk aversion. One-year expected utility turns out to be increasing along the feasible termination ratio interval. This outcome obtains once the drift becomes negative, which calls for a high critical funding ratio, making termination by the PGF likely. This is even in the interest of risk-neutral beneficiaries because they have to expect a worsening financial status of the pension fund over time. Through its intervention, the PGF prevents the pension plan's funding level from dropping even further, causing beneficiaries to lose a still greater part of their pension benefits.

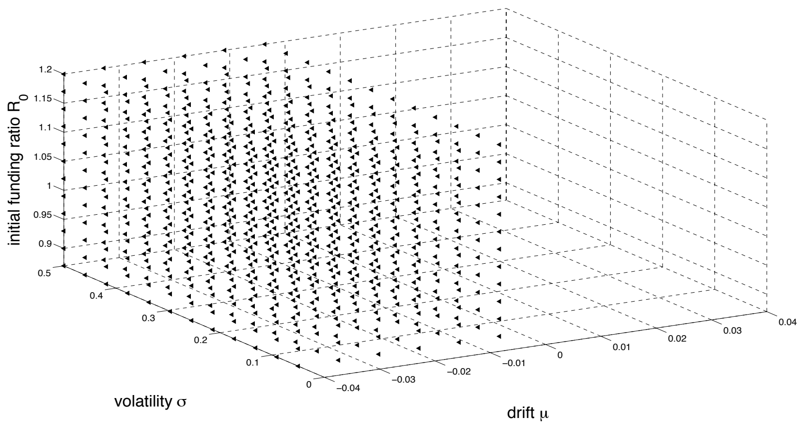
Finally, panel (c) of Figure 2.1 exhibits the parameter triples (μ, σ, R_0), marked with a "•", for which $\frac{\partial U}{\partial \eta}$ changes its sign at a critical value of relative risk aversion, defined as a critical risk aversion level. Below this level, $\frac{\partial U}{\partial \eta}$ is negative over the constraint set; above it, $\frac{\partial U}{\partial \eta}$ turns positive. It means that the one-year expected utility function is decreasing along the feasible termination ratio interval for weakly risk-averse beneficiaries but increasing for strongly risk-averse ones, i.e., the utility curve inverts at the critical risk aversion level. This case obtains when the drift is positive while the volatility is not sufficiently low ($\sigma > 0.1$). Intuitively, less risk-averse beneficiaries are confident that their DB pension plan will be sufficiently funded at the end of the year. They therefore prefer the PGF to have a low termination ratio, causing it to abstain from intervention. However, for beneficiaries with marked risk aversion preferences, uncertainty is not acceptable even

[15]The values $\mu \in [-0.04, 0.04]$, $\sigma \in [0.05, 0.5]$ and $R_0 \in [0.87, 1.2]$ are divided into 12 segments of equal size, resulting in $12^3 = 1,782$ combinations. After calculating the constraint set (2.10), the partial derivative is determined using increments of 0.01 for η.

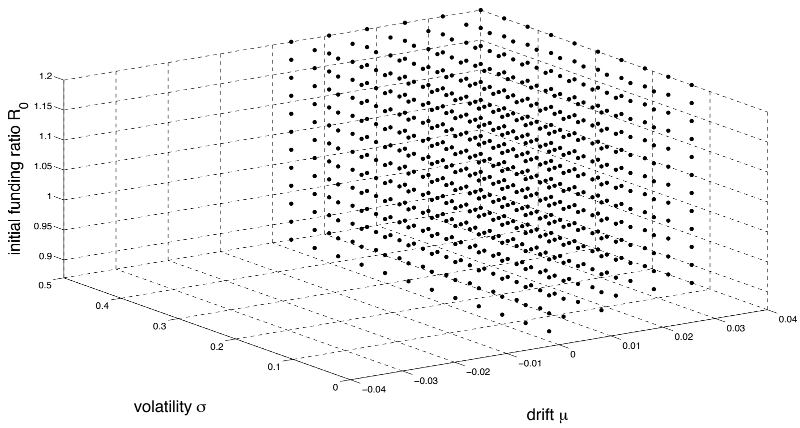
[16]The increment of the coefficient of relative risk aversion is 0.01 as well.



(a)



(b)



(c)

Figure 2.1: *Computation results of partial derivative of beneficiaries' expected utility w.r.t. the termination ratio. Beneficiaries' coefficient of relative risk aversion* $\delta \in [0, 1) \cup (1, 5]$*; funding ratio process specified by drift* $\mu \in [-0.04, 0.04]$*, volatility* $\sigma \in [0.05, 0.5]$ *and initial funding ratio* $R_0 \in [0.87, 1.2]$*.*

though the plan's funding ratio is increasing at a rather high rate. They prefer a high critical funding ratio which is likely to trigger termination by the PGF likely. In sum, demand for intervention by the PGF depends importantly on the risk preferences of pension beneficiaries in this situation.

2.4 Critical risk aversion level

In the previous section, a critical risk aversion level at which the one-year expected utility curve inverts was found for parameter triples (μ, σ, R_0) marked with a "•" sign, see panel (c) of Figure 2.1 again. It is of interest to see how this level, denoted by δ^*, depends on the parameter values (μ, σ, R_0) of the pension plan's funding ratio process. Using the parameter triple $(\mu, \sigma, R_0) = (0.03, 0.2, 1.1)$ which belongs to the set of points marked with "•" in panel (c) of Figure 2.1, $\frac{\partial U}{\partial \eta}$ is plotted over the constraint set $\eta \in C_{IPC}$ in Figure 2.2. Given intervention probability $\epsilon = 0.025$, this constraint set is $C_{IPC} = (0, 0.71]$. The coefficient of relative risk aversion varies in the domain $\delta \in [0,1) \cup (1,5]$. For low values of δ, $\frac{\partial U}{\partial \eta}$ is negative over the entire interval $(0, 0.71]$ of η. However, for values of δ above approximately 1.5, $\frac{\partial U}{\partial \eta}$ turns positive, again over the entire interval $(0, 0.71]$ of η.[17] One-year expected utility curve inverts at $\delta^* = 1.5$, as indicated by the vertical plane of Figure 2.2.

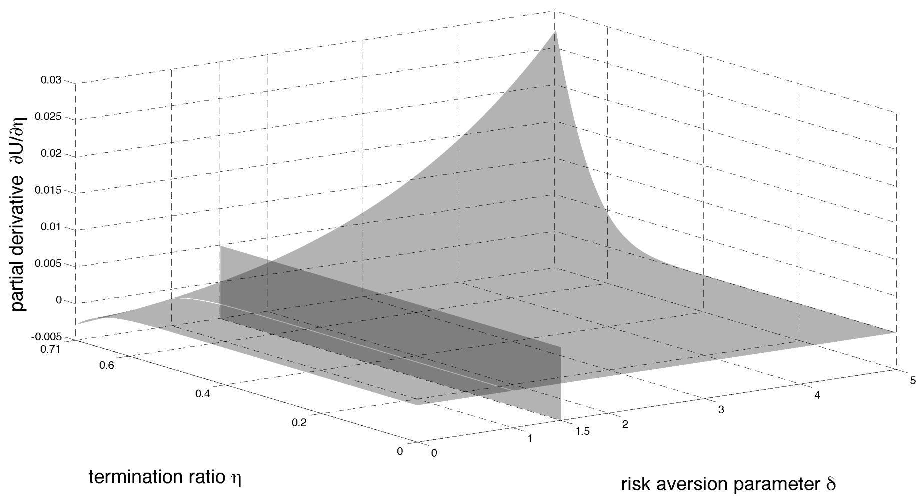


Figure 2.2: *Partial derivative of the one-year expected utility as a function of* $\eta \in (0, 0.71]$ *for values of the coefficient of relative risk aversion* $\delta \in [0,1) \cup (1,5]$ *with* $(\mu, \sigma, R_0) = (0.03, 0.2, 1.1)$. *The vertical plane indicates the critical risk aversion level* $\delta^* = 1.5$.

[17] For relatively small values of η, it becomes difficult to distinguish the sign of $\frac{\partial U}{\partial \eta}$ in Figure 2.2. However, the statement in the text is based on the results of computation.

Next, the relationship between δ^* and the parameter values (μ, σ, R_0) is investigated. First, let the uncertainty concerning the pension plan's funding ratio increase from $\sigma = 0.2$ to $\sigma = 0.4$. This calls for a new computation of the constraint set, which becomes $C_{IPC}(\mu, \sigma, R_0) = (0, 0.43]$. Panel (a) of Figure 2.3 presents $\frac{\partial U}{\partial \eta}$ for termination ratios $\eta \in (0, 0.43]$ and values $\delta \in [0, 1) \cup (1, 5]$, as before. The critical risk aversion value δ^* at which partial derivative over the entire termination ratio interval $(0, 0.43]$ changes sign becomes $\delta^* = 0.38$, down from $\delta^* = 1.5$ in Figure 2.2. Hence, less strongly risk-averse pension beneficiaries with $\delta \in (0.38, 1.5]$ now also exhibit $\frac{\partial U}{\partial \eta} > 0$, causing them to prefer intervention by the PGF. As uncertainty surrounding the funding ratio increases, the probability of the DB plan becoming underfunded increases too; hence, more policyholders prefer intervention by the PGF.

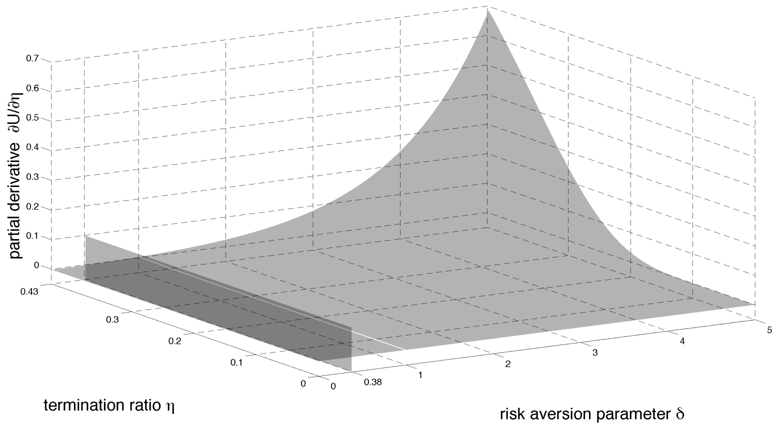
Second, let the drift of the funding ratio process become smaller, with $\mu = 0.015$ rather than $\mu = 0.03$ as in Figure 2.2, resulting in a new constraint set $C_{IPC}(\mu, \sigma, R_0) = (0, 0.7]$. As shown in panel (b) of Figure 2.3, this has a very similar effect on the critical level of risk aversion δ^* as an increase in σ. Now the one-year expected utility curve inverts at $\delta^* = 0.75$ (see the vertical plane in panel (b) of Figure 2.3), with less risk-averse beneficiaries characterized by $\delta \in (0.75, 1.5]$ exhibiting $\frac{\partial U}{\partial \eta} > 0$ and causing them to prefer intervention by the PGF. As the expected growth rate of the funding ratio becomes smaller, the probability of the DB plan becoming underfunded again increases. This causes even less strongly risk-averse beneficiaries to worry about losing their pension benefits and therefore to prefer intervention by the PGF.

Third, consider a lower value of the initial funding ratio $R_0 = 0.9$. With $(\mu, \sigma, R_0) = (0.03, 0.2, 0.9)$, the constraint set changes to $C_{IPC} = (0, 0.58]$. Since in panel (c) of Figure 2.3, the critical level of risk aversion stays the same as in Figure 2.2, i.e., $\delta^* = 1.5$, beneficiaries' preference concerning intervention by the PGF is not influenced by the DB plan's initial funding status. Rather, its future funding status is of importance.

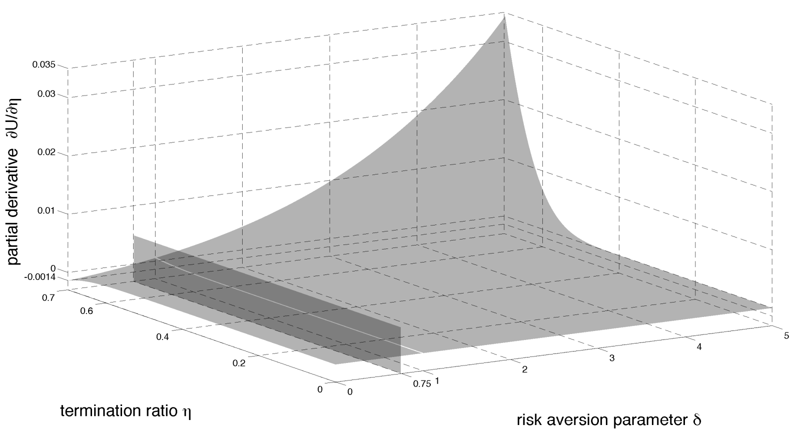
The findings obtained can be summarized in the following five statements.

(1) Variations of the drift parameter μ, the volatility parameter σ, and the initial funding ratio parameter R_0 of the funding ratio process have a substantial impact on the constraint set $C_{IPC}(\mu, \sigma, R_0)$ at a given value of intervention probability ($\epsilon = 0.025$, in the present case).

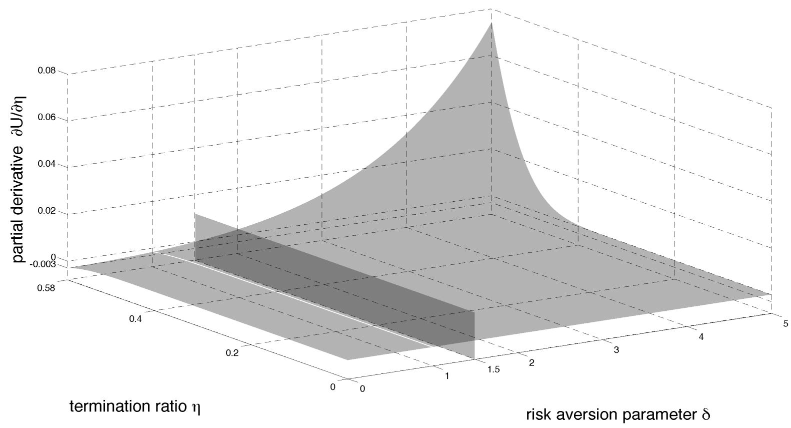
(2) Doubling the volatility σ (from 0.2 to 0.4) causes the critical level of risk aversion where one-year expected utility curve inverts to drop from 1.5 to 0.38, i.e., by one-fourth. Although with $R_0 = 1.1$, the plan is well-funded initially, an increase in



(a) $(\mu, \sigma, R_0) = (0.03, 0.4, 1.1)$; $C_{IPC}(\mu, \sigma, R_0) = (0, 0.43]$, $\delta^* = 0.38$.



(b) $(\mu, \sigma, R_0) = (0.015, 0.2, 1.1)$; $C_{IPC}(\mu, \sigma, R_0) = (0, 0.7]$, $\delta^* = 0.75$.



(c) $(\mu, \sigma, R_0) = (0.03, 0.2, 0.9)$; $C_{IPC}(\mu, \sigma, R_0) = (0, 0.58]$, $\delta^* = 1.5$

Figure 2.3: *Partial derivative of the one-year expected utility as a function of $\eta \in C_{IPC}$ for values of the coefficient of relative risk aversion $\delta \in [0, 1) \cup (1, 5]$ for given values of (μ, σ, R_0). The vertical plane indicates the critical risk aversion level δ^*.*

volatility triggers demand for intervention by the PGF even among almost risk-neutral beneficiaries.

(3) Halving the drift μ (from 0.03 to 0.015) entails a halving of the critical level of risk aversion, from 1.5 to 0.75 in the present case. This suggests that the current near-zero rates of return on the assets typically held by pension funds will be likely to boost beneficiaries' demand for intervention by the PGF.

(4) The critical level of risk aversion drops to one fourth as σ is doubled but only to one half as μ is halved. This difference suggests that demand for intervention by the PGF is more sensitive to the uncertainty of the plan's funding status than its growth rate. This signals to sponsoring companies that an increase in the investment risk of their pension portfolios can more likely induce beneficiaries to call for intervention by the PGF.

(5) A lower initial funding ratio R_0 (from 1.1 to 0.9) does not have an influence on the critical level of risk aversion. This is consistent with the view that pension beneficiaries are worried not about the pension plan's initial funding status, but its future funding status, which crucially depends on the uncertainty and growth rate of its funding ratio.

These findings permit to conjecture a functional form for the critical level of risk aversion, $\delta^* = \frac{2\mu}{\sigma^2}$. This function maps the sets $(\mu, \sigma, R_0) = (0.03, 0.2, 1.1)$, $(\mu, \sigma, R_0) = (0.03, 0.4, 1.1)$ and $(\mu, \sigma, R_0) = (0.015, 0.2, 1.1)$ to $\delta^* = (1.5, 0.38, 0.75)$, as found in Figure 2.3. For this function, Proposition 1 holds, see Appendix 2.7.3 for proof.[18]

Proposition 1 *Let $\delta^* = \frac{2\mu}{\sigma^2}$. Then, the one-year expected utility function $U(\eta, \delta, \mu, \sigma, R_0)$ is strictly decreasing in $\eta > 0$ if $\delta < \delta^*$, strictly increasing in $\eta > 0$ if $\delta > \delta^*$, and constant with respect to $\eta > 0$ if $\delta = \delta^*$ for given values of $\mu \in \mathbb{R}$, $\sigma > 0$ and $R_0 > 0$.*

2.5 The optimal termination ratio

This section is devoted to the solution of the constrained maximization problems (2.11) and (2.14). As in Section 2.4, suppose $(\mu, \sigma, R_0, \epsilon) = (0.03, 0.2, 1.1, 0.025)$, while the coefficient of relative risk aversion assumes the values $\delta = (0, 0.8, 1.6, 2.4)$. By Proposition 1, the critical level of risk aversion can be computed; its value is $\delta^* = 1.5$. Thus, beneficiaries with $\delta = 0$ and $\delta = 0.8$ exhibit $\frac{\partial U}{\partial \eta} < 0$ while those with $\delta = 1.6$ and $\delta = 2.4$ exhibit

[18]We thank Lorens Imhof for contributing the proof.

$\frac{\partial U}{\partial \eta} > 0$. The optimal termination ratio is therefore determined by the boundary of the corresponding constraint sets (2.10) and (2.13).

2.5.1 Expected utility maximization subject to the IPC only

In Section 2.4, the set satisfying the IPC was determined as $\eta \in (0, 0.71]$. Figure 2.4 plots the one-year expected utility as a function of η over this interval for the four values of δ singled out above. It reveals that the utility value is not sensitive to changes in the termination ratio η as long as it is low ($\eta < 0.5$), in keeping with the finding that $\frac{\partial U}{\partial \eta}$ is extremely close to zero regardless of δ (see Figure 2.2 again). For risk-neutral ($\delta = 0$) and weekly risk-averse ($\delta = 0.8$) pension beneficiaries, one-year expected utility decreases progressively as η approaches 0.71 (see panels (a) and (b) of Figure 2.4). For them, a low termination ratio is optimal, rendering intervention by the PGF almost non-applicable.[19] Intuitively, the legal cap on pension benefits paid out by the PGF causes pension beneficiaries of this type to prefer possible recovery of an underfunded DB pension fund, enabling it to pay the full promised benefits, over the prospect of losing them in part when the plan is closed by the PGF.

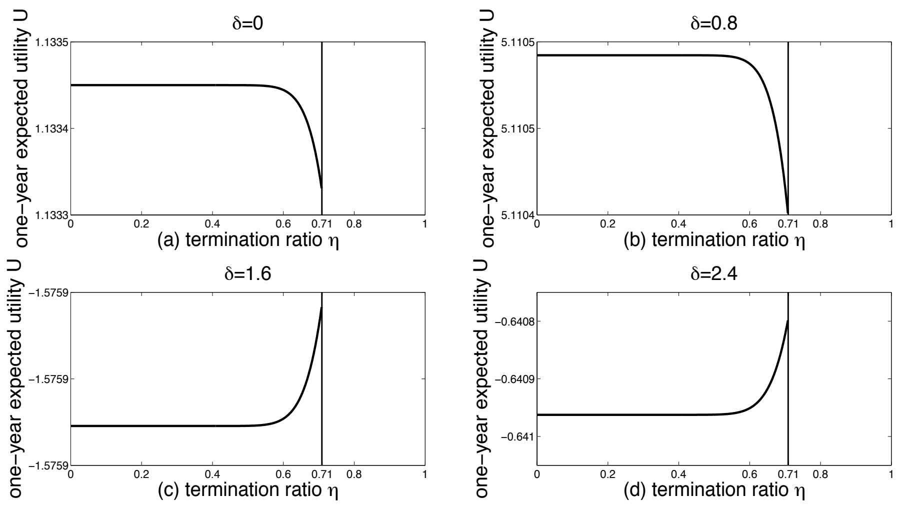


Figure 2.4: *One-year expected utility as a function of the termination ratio $\eta \in (0, 0.71]$ for $\delta = (0, 0.8, 1.6, 2.4)$.*

However, Figure 2.4 also shows that at higher values of relative risk aversion, i.e., $\delta = 1.6$ and $\delta = 2.4$, one-year expected utility starts to increase progressively as the termination ratio approaches the upper bound of $\eta = 0.71$, reaching its maximum at

[19]Note that $\frac{\partial U}{\partial \eta} < 0$ over the interval $\eta \in (0, 0.71]$ for $\delta = (0, 0.8)$, see Proposition 1, causing the optimal termination ratio to be extremely close to 0 with termination hardly ever triggered.

$\eta = 0.71$ where the IPC is binding (see panels (c) and (d) of Figure 2.4). Intuitively, more strongly risk-averse beneficiaries prefer to have the PGF exercise control through termination since they are more concerned about the potential loss of benefits if the pension fund continues operating. The higher the termination ratio, the safer their claims and the higher their expected utility associated with termination.

As in Section 2.4, an analysis of the one-year expected utility can also be performed by halving the drift (to $\mu = 0.015$) and doubling the volatility (to $\sigma = 0.4$). The constraint set shrinks again to become $\eta \in (0, 0.7]$ and $\eta \in (0, 0.43]$, respectively. The critical level of risk aversion updates to be $\delta^* = 0.75$ and $\delta^* = 0.38$, respectively. For risk-neutral beneficiaries, i.e., $\delta = 0$, their one-year expected utility curve has the same pattern as in panel (a) of Figure 2.4, but reaching the minimum at $\eta = 0.7$ and $\eta = 0.43$, respectively. However, for beneficiaries characterized by $\delta = 0.8$ (that is now above the updated critical level of risk aversion $\delta^* = 0.75$ and $\delta^* = 0.38$), their one-year expected utility turns to be increasing as the termination ratio η approaches the upper bound of η rather than decreasing in panel (b) of Figure 2.4. Their optimal termination ratio switches to the upper bounds of $\eta = 0.7$ and $\eta = 0.43$, respectively. At higher coefficients of relative risk aversion ($\delta = 1.6$, $\delta = 2.4$), beneficiaries still exhibit increasing expected utility as in panels (c) and (d) of Figure 2.4, albeit reaching a maximum at a lower value of η. These results are intuitive. Since the DB plan is more likely to end up being underfunded with lower drift or higher volatility, weakly risk-averse beneficiaries ($\delta = 0.8$) now prefer the PGF to intervene, protecting them from a loss of their pension benefits. Lastly, as is known from Proposition 1, the initial funding ratio R_0 does not have an influence on δ^*. This implies that, when a different value of the initial funding ratio is selected, expected utilities for beneficiaries with the four values of δ will have the same pattern as in panels (a)-(d) of Figure 2.4 and their preference concerning the intervention by the PGF is not going to change.

2.5.2 Expected utility maximization subject to both the IPC and ESC

In this section, the ESC is added as a second constraint. It can be shown to be binding at the termination ratio $\eta_q = h_{ES}^{-1}(q) = 0.68$ (expected shortfall is set at $q = 0.03$). The ESC constraint set therefore is $\eta \in [0.68, 1)$, while the IPC constraint set continues to be $\eta \in (0, 0.71]$ as in Section 2.5.1. Therefore, the two constraints are compatible if $\eta \in [0.68, 0.71]$, indicated by the two vertical lines in Figure 2.5. The optimal termination ratio for risk-neutral and weekly risk-averse pension beneficiaries ($\delta = 0, \delta = 0.8$) changes from the close-to-zero value in panels (a) and (b) of Figure 2.4 to the lower bound of $\eta = 0.68$. Thus, while these types of beneficiary prefer to avoid termination, the PGF, seeking to protect its future viability, imposes a higher value of termination threshold η. By way of contrast, the upper bound $\eta = 0.71$ is optimal for more strongly risk-averse beneficiaries ($\delta = 1.6$, $\delta = 2.4$), the same as under the IPC only (see panels (c) and (d) of Figure 2.4 again).

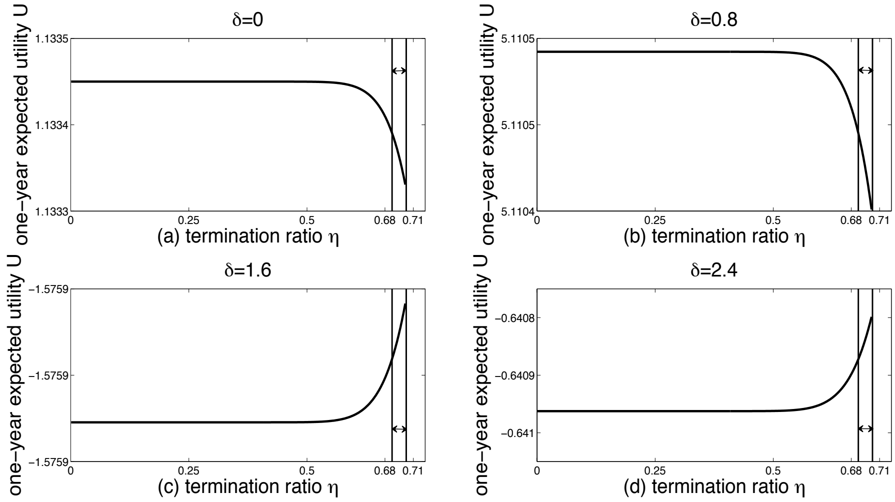


Figure 2.5: *One-year expected utility when the IPC and the ESC are compatible;* $\eta \in (0, 0.71] \cap [0.68, 1)$, $\delta = (0, 0.8, 1.6, 2.4)$, $q = 0.03$.

Clearly, the two constraint sets may fail to overlap. All it takes is to tighten the ESC, e.g., from $q = 0.03$ to $q = 0.015$, meaning that the expected deficit must not exceed 1.5 percent of the plan's liabilities. As shown in Figure 2.6, this causes the termination ratio where the ESC is binding to increase to $\eta = 0.8$. Hence, the two constraint sets now are $\eta \in (0, 0.71]$ and $[0.8, 1)$, respectively. As argued in Section 2.2.4, a second-best solution is the one which satisfies the IPC while violating the ESC to the least extent. By limiting the PGF's annual probability of intervention, this gives priority to its current financial guarantee. While violating the ESC to the least extent, the PGF can still find new funding or adjust its investment strategy. Therefore, the second-best solution is $\eta_\epsilon = h_{IP}^{-1}(\epsilon) = 0.71$. Note from panels (a) to (d) of Figure 2.6 that this solution does not depend on beneficiaries' risk aversion anymore.

To summarize, while the IPC leads to a termination rule that reflects beneficiaries' risk aversion (see Proposition 1 and Figure 2.4 again), this needs not to be true when the ESC is added. The two constraints may not be satisfied simultaneously when the expected deficit relative to the plan's liabilities (q) is set at a sufficiently low value. In that event, satisfying the IPC while violating the ESC to a minimum extent is proposed

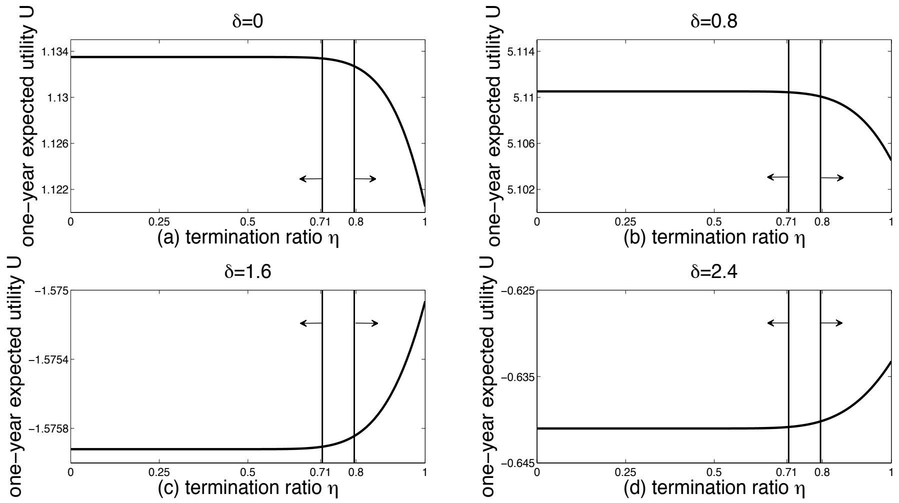


Figure 2.6: *One-year expected utility when the IPC and the ESC are not compatible;* $\eta \in (0, 0.71] \cap [0.8, 1)$*,* $\delta = (0, 0.8, 1.6, 2.4)$*,* $q = 0.015$*.*

as a second-best solution, which however does not reflect beneficiaries' degree of relative risk aversion, anymore.

2.6 Conclusion

The objective of this chapter is to derive a rule for a pension guarantee fund (PGF) when to close an underfunded defined benefit (DB) pension plan. In contrast to designing optimal intervention strategies in the best interest of the PGF, see Kalra and Jain (1997), evaluating the PGF's intervention in terms of pension beneficiaries' utility dispels the main concern of the critics of the PGF's active intervention, being its contradiction to the PGF's social welfare motive. Acting in the interest of pension beneficiaries, the PGF solves a one-year expected utility maximization problem with the termination ratio η, i.e., the funding ratio of the DB pension plan triggering intervention, as the argument. The higher the termination ratio, the more likely the intervention by the PGF is. The two constraints are designed to limit the PGF's insolvency risk, the intervention probability constraint (IPC) which limits the PGF's annual frequency of intervention, and the expected shortfall constraint (ESC) which limits the expected deficit the PGF has to cover when it allows an underfunded pension plan to continue.

In the view of the complexity of the objective function, numerical solutions were presented for combinations of the DB fund's initial funding ratio, the drift of the funding process, and its volatility, generating positive and negative partial derivatives of expected utility w.r.t. the termination ratio. While some of the derivatives kept their sign regardless of beneficiaries' relative risk aversion, quite a few changed sign from negative to positive when the coefficient of relative risk aversion increased above a critical risk aversion level. This finding motivated an analysis of the relationship between the critical risk aversion level, denoted by δ^*, at which one-year expected utility curve inverts, and the parameters characterizing funding ratio process, i.e., the drift, volatility, and initial funding ratio. Following this analysis, a conjecture as to the functional form of δ^* was possible and permitted the formulation of Proposition 1. According to it, the functional form of the critical risk aversion level is $\delta^* = \frac{2\mu}{\sigma^2}$ and the expected utility function exhibits a monotonically decreasing and increasing relationship with the termination ratio η as relative risk aversion δ is below and above δ^*, respectively. This implies that in the interest of beneficiaries, the PGF should set the termination ratio η at a low value, causing intervention to be unlikely, for weakly risk-averse beneficiaries characterized by $\delta < \delta^*$; but at a high value of η, intervention becomes very likely, serving the interests of strongly risk-averse beneficiaries characterized by $\delta > \delta^*$. Therefore, the optimal intervention by the PGF depends crucially on the beneficiaries' risk preferences, which differs from the recommended intervention strategy by Kalra and Jain, i.e., the PBGC intervenes in only extreme cases. Moreover, as volatility σ increases, the critical risk aversion level δ^* drops. This implies that the set of beneficiaries with $\delta > \delta^*$ increases so that more beneficiaries prefer the intervention by the PGF. This finding not only confirms the comparative result in Kalra and Jain (1997) from the PGF's point of view, stating that the PBGC faces a higher risk and intervenes earlier as uncertainty concerning the pension asset process increases, but also confirms from the beneficiaries' point of view that their demand for early intervention by the PGF increases as uncertainty surrounding the funding ratio is larger. In addition to the static analysis on the volatility, a similar effect on the critical risk aversion level as an increase in σ was found by assigning a lower value to the drift μ. However, the initial funding ratio R_0 was found not to affect δ^*, suggesting that pension beneficiaries' demand for earlier intervention by the PGF is driven by the plan's future rather than initial funding status.

Eventually, the optimal termination ratio for beneficiaries with a given value of δ is mainly determined by the boundary of the constraint set after computing δ^* by Proposition 1. As shown in Section 2.5, imposition of only the IPC results in non-applicable intervention by the PGF for risk neutral ($\delta = 0$) and slightly risk-averse ($\delta = 0.8$) beneficiaries. This result may contribute to explaining the passive termination behavior of the PBGC, as noted by Kalra and Jain. When the ESC is added, two outcomes are possible. If the admissible deficit relative to the DB plan's liabilities is not too low, the two constraints overlap, admitting of an optimal termination rule that reflects beneficiaries' relative risk aversion. However, if the relative deficit is set at a very low value, the IPC and ESC fail to result in overlapping constraint sets, calling for a second-best solution.[20] Giving priority to the IPC over the ESC can be justified with the argument that securing the PGF's ability to guarantee pensions now takes precedence over its financial status a year hence, which still can improve in the meantime. However, this second-best solution turns out to have the side effect of neglecting beneficiaries' relative risk aversion.

This analysis could be extended in several ways. First, the PGF is modelled as insuring only one DB pension plan. An important extension for future research is to consider a set of plans, giving rise to portfolio effects. On the one hand, the PGF receives premium income from several sources, permitting it to build its reserves during normal times. On the other hand, it may be confronted with an accumulation of risks during an economic downturn, causing several plans to be underfunded. Evidently, this extension calls for a careful modelling of an aggregate terminating process of the insured DB plans. Second, with regulatory practice specifying annual premiums paid to the PGF over a multi-period horizon, the PGF might anticipate changes in its financial status over the multi-period horizon. This would call for a dynamic termination strategy over the longer horizon taking into account a sequence of expected shortfall values and modelling a series of optimal probabilities of intervention, along the lines of Shi and Werker (2012) in the case of long-term investment strategies. In this context, introducing a utility function with a loss aversion property to characterize pension beneficiaries and studying its effect on the optimal termination rule would be of interest (see e.g., Siegmann (2011)). Finally, one might consider lifting current regulatory norms leading to the restrictions studied in this paper in favor of risk-based premiums paid to the PGF, of the type common in reinsurance. Thus, the PGF would obtain the right to charge DB pension plans with an unfavorable development of their funding ratio an increased premium.[21] But still, this is modelled on the premise of allowing the PGFs' termination rules to reflect pension beneficiaries' risk preferences.

[20]This situation could also obtain if the probability of intervention ϵ in the IPC is set to a very low value, causing the admissible values of the termination ratio η to be low. This case is not pursued here.

[21]The issue of who pays for the deficit of the PGF if it goes bankrupt, is not addressed in this paper.

2.7 Appendix to Chapter 2

2.7.1 Closed-form expressions for pertinent probabilities

We summarize the closed-form expressions for the pertinent probabilities, see, e.g., Jeanblanc, Yor and Chesney (2009), chapter 3, as follows

$$P(R_1 \leqslant r) = \Phi\left(\frac{1}{\sigma}\ln\frac{r}{R_0} - \frac{\mu}{\sigma} + \frac{1}{2}\sigma\right), \quad r > 0; \tag{2.15}$$

$$P(X_R \leqslant x) = \Phi\left(\frac{1}{\sigma}\ln\frac{x}{R_0} - \frac{\mu}{\sigma} + \frac{1}{2}\sigma\right) + \left(\frac{x}{R_0}\right)^{\left(\frac{2\mu}{\sigma^2}-1\right)} \Phi\left(\frac{1}{\sigma}\ln\frac{x}{R_0} + \frac{\mu}{\sigma} - \frac{1}{2}\sigma\right),$$
$$x \in (0, R_0); \tag{2.16}$$

$$P(R_1 \leqslant b, X_R \leqslant a) = P(R_1 \leqslant b) + P(X_R \leqslant a) + P(R_1 > b, X_R > a) - 1$$
$$= \Phi\left(\frac{1}{\sigma}\ln\frac{a}{R_0} - \frac{\mu}{\sigma} + \frac{1}{2}\sigma\right) + \left(\frac{a}{R_0}\right)^{\left(\frac{2\mu}{\sigma^2}-1\right)} \left[\Phi\left(\frac{1}{\sigma}\ln\frac{a}{R_0} + \frac{\mu}{\sigma} - \frac{1}{2}\sigma\right)\right.$$
$$\left. -\Phi\left(\frac{1}{\sigma}\ln\left(\frac{a^2}{bR_0}\right) + \frac{\mu}{\sigma} - \frac{1}{2}\sigma\right)\right], \quad a < b\,,\ a \in (0, R_0); \tag{2.17}$$

$$f_{(R_1,X_R)}(b,a) = \frac{1}{a\sigma}\phi\left(\frac{1}{\sigma}\ln\frac{a}{R_0} - \frac{\mu}{\sigma} + \frac{1}{2}\sigma\right) + \frac{1}{a\sigma}\left(\frac{a}{R_0}\right)^{\left(\frac{2\mu}{\sigma^2}-1\right)} \left[\phi\left(\frac{1}{\sigma}\ln\frac{a}{R_0} + \frac{\mu}{\sigma} - \frac{1}{2}\sigma\right)\right.$$
$$\left. + \left(\frac{2\mu}{\sigma} - \sigma\right)\Phi\left(\frac{1}{\sigma}\ln\frac{a}{R_0} + \frac{\mu}{\sigma} - \frac{1}{2}\sigma\right)\right] + \frac{2}{ab\sigma^3}\ln\left(\frac{bR_0}{a^2}\right)\left(\frac{a}{R_0}\right)^{\left(\frac{2\mu}{\sigma^2}-1\right)}$$
$$\phi\left(\frac{1}{\sigma}\ln\left(\frac{a^2}{bR_0}\right) + \frac{\mu}{\sigma} - \frac{1}{2}\sigma\right), \quad a < b\,,\ a \in (0, R_0) \tag{2.18}$$

where $\Phi(.)$ and $\phi(.)$ denote the CDF and the PDF of a standard normal variable, respectively. (2.15) measures the probability of the funding ratio dropping below x at time $t = 1$, and (2.16), the probability of the minimum value of the funding ratio between $t_0 = 0$ and $t = 1$ falling short of x. The joint probability function and the joint density function of (R_1, X_R) are given by expressions (2.17) and (2.18), respectively.

2.7.2 Derivation of the expected shortfall function

By using the closed-form expressions (2.15) and (2.18) given in Appendix 2.7.1, the expected shortfall of an ongoing underfunded pension plan at $t = 1$ can be calculated as follows,

$$
\begin{aligned}
& E\left[(1-R_1)\mathbb{1}_{\{\tau>1\}}\mathbb{1}_{\{R_1\leqslant 1\}}\right] \\
=& E\left[\mathbb{1}_{\{I_1>\eta\}}\mathbb{1}_{\{R_1\leqslant 1\}}\right] - E\left[R_1\mathbb{1}_{\{I_1>\eta\}}\mathbb{1}_{\{R_1\leqslant 1\}}\right] \\
=& E\left[\mathbb{1}_{\{R_1\leqslant 1\}}\right] - E\left[\mathbb{1}_{\{I_1\leqslant\eta\}}\mathbb{1}_{\{R_1\leqslant 1\}}\right] - E\left[R_1\mathbb{1}_{\{I_1>\eta\}}\mathbb{1}_{\{R_1\leqslant 1\}}\right] \\
=& P(R_1\leqslant 1) - \int_0^\eta\int_0^1 f_{(R_1,X_R)}(r,x)drdx - \int_\eta^{R_0}\int_\eta^1 rf_{(R_1,X_R)}(r,x)drdx \\
=& \Phi\left(-\frac{1}{\sigma}\ln R_0 - \frac{\mu}{\sigma} + \frac{1}{2}\sigma\right) - \Phi\left(\frac{1}{\sigma}\ln\frac{\eta}{R_0} - \frac{\mu}{\sigma} + \frac{1}{2}\sigma\right) \\
& - \left(\frac{\eta}{R_0}\right)^{\left(\frac{2\mu}{\sigma^2}-1\right)}\left[\Phi\left(\frac{1}{\sigma}\ln\frac{\eta}{R_0} + \frac{\mu}{\sigma} - \frac{1}{2}\sigma\right) - \Phi\left(\frac{1}{\sigma}\ln\frac{\eta^2}{R_0} + \frac{\mu}{\sigma} - \frac{1}{2}\sigma\right)\right] \\
& - R_0e^\mu\left[\Phi\left(-\frac{1}{\sigma}\ln R_0 - \frac{\mu}{\sigma} - \frac{1}{2}\sigma\right) - \Phi\left(\frac{1}{\sigma}\ln\frac{\eta}{R_0} - \frac{\mu}{\sigma} - \frac{1}{2}\sigma\right)\right] \\
& + R_0e^\mu\left(\frac{\eta}{R_0}\right)^{\left(\frac{2\mu}{\sigma^2}+1\right)}\left[\Phi\left(\frac{1}{\sigma}\ln\frac{\eta}{R_0} + \frac{\mu}{\sigma} + \frac{1}{2}\sigma\right) - \Phi\left(\frac{1}{\sigma}\ln\frac{\eta^2}{R_0} + \frac{\mu}{\sigma} + \frac{1}{2}\sigma\right)\right] \\
=&: h_{ES}(\eta,\mu,\sigma,R_0), \qquad \eta\in\mathbf{H}\equiv(0,\min\{R_0,1\}).
\end{aligned}
$$

The expected shortfall $h_{ES}(\eta,\mu,\sigma,R_0)$ is continuous on the termination ratio $\eta\in(0,\min\{R_0,1\})$ for given values of (μ,σ,R_0). It is given by,

$$
\begin{aligned}
E\left[(1-R_1)\mathbb{1}_{\{\tau>1\}}\mathbb{1}_{\{R_1\leqslant 1\}}\right] &= E\left[(1-R_1)\mathbb{1}_{\{R_1\leqslant 1\}}\mathbb{1}_{\left\{\inf\limits_{t\in[0,1]}\{R_t\}>\eta\right\}}\right] \\
= E\left[Y\mathbb{1}_{\left\{\inf\limits_{t\in[0,1]}\{R_t\}>\eta\right\}}\right] &= E[Y]E\left[\frac{Y}{E[Y]}\mathbb{1}_{\left\{\inf\limits_{t\in[0,1]}\{R_t\}>\eta\right\}}\right],
\end{aligned}
$$

where $Y=(1-R_1)\mathbb{1}_{\{R_1\leqslant 1\}}$ is non-negative and does not depend on η. Applying a multiplicative factor, this can be expressed under a different probability measure $\mathbb{Q}$,

$$
E\left[(1-R_1)\mathbb{1}_{\{\tau>1\}}\mathbb{1}_{\{R_1\leqslant 1\}}\right] = E[Y]E\left[\frac{Y}{E[Y]}\mathbb{1}_{\left\{\inf\limits_{t\in[0,1]}\{R_t\}>\eta\right\}}\right] = E[Y]E^{\mathbb{Q}}\left[\mathbb{1}_{\left\{\inf\limits_{t\in[0,1]}\{R_t\}>\eta\right\}}\right].
$$

This quantity is decreasing from $E[Y]$ to 0 as the termination ratio η increases from 0 to $\min\{R_0, 1\}$. In addition, when η approaches 0, one has

$$\begin{aligned}
&\lim_{\eta\to 0} E\left[(1-R_1)\mathbb{1}_{\{\tau>1\}}\mathbb{1}_{\{R_1\leqslant 1\}}\right]\\
&= E[Y] = E\left[(1-R_1)\mathbb{1}_{\{R_1\leqslant 1\}}\right]\\
&= \int_0^1 (1-r)dP(R_1\leqslant r) = \Phi\left(-\frac{1}{\sigma}\ln R_0 - \frac{\mu}{\sigma} + \frac{1}{2}\sigma\right) - R_0 e^{\mu}\Phi\left(-\frac{1}{\sigma}\ln R_0 - \frac{\mu}{\sigma} - \frac{1}{2}\sigma\right),
\end{aligned}$$

where $P(R_1 \leqslant r)$ is given in Appendix 2.7.1.

2.7.3 Proof of Proposition 1

In the following proof of Proposition 1, the values of $\mu \in \mathbb{R}$, $\sigma > 0$ and $R_0 > 0$ are fixed. To reduce the number of parameters, introduce

$$\gamma = \frac{\mu}{\sigma} - \frac{\sigma}{2}, \qquad \zeta = 2\gamma + \sigma(1-\delta) = \sigma(\delta^* - \delta),$$

and define for all $x \in \mathbb{R}$,

$$\begin{aligned}
g(x) = &e^{(\zeta-2\gamma)x}\Phi(x-\gamma) + e^{\zeta x}\Phi(x+\gamma)\\
&+ e^{(\frac{1}{2}\zeta^2-\gamma\zeta)}\left[\Phi(-x+\zeta-\gamma) - e^{2(\zeta-\gamma)x}\Phi(x+\zeta-\gamma)\right].
\end{aligned}$$

Then

$$U(\eta,\delta,\mu,\sigma,R_0) = \frac{R_0^{(1-\delta)}}{1-\delta} g\left(\frac{1}{\sigma}\ln\frac{\eta}{R_0}\right).$$

As $\frac{1}{\sigma}\ln\frac{\eta}{R_0}$ is strictly increasing in η, the proposition will follow if $(1-\delta)g'(x)$ can be shown to have the same sign as $\delta - \delta^*$.

Define for every $x \in \mathbb{R}$ a function $\psi_x : \mathbb{R} \to \mathbb{R}$ by

$$\psi_x(z) = e^{\left(\frac{1}{2}z^2 + z(x-\gamma)\right)}\Phi(x+z-\gamma).$$

Then $g'(x)$ can be written as

$$g'(x) = e^{(\zeta-2\gamma)x}[(\zeta-2\gamma)\psi_x(0) + \zeta\psi_x(2\gamma) - 2(\zeta-\gamma)\psi_x(\zeta)], \tag{2.19}$$

where the following equalities are used,

$$e^{(\zeta-2\gamma)x}\phi(x-\gamma) = e^{(\frac{1}{2}\zeta^2-\gamma\zeta)}\phi(-x+\zeta-\gamma), \quad e^{\zeta x}\phi(x+\gamma) = e^{(\frac{1}{2}\zeta^2-\gamma\zeta)}e^{2(\zeta-\gamma)x}\phi(x+\zeta-\gamma).$$

To determine the sign of $g'(x)$, one can use Lemma 3 below. To this end, one has to verify that ψ_x is strictly increasing. The starting point is

$$\psi'_x(z) = e^{\left(\frac{1}{2}z^2+z(x-\gamma)\right)}\left[(x+z-\gamma)\Phi(x+z-\gamma)+\phi(x+z-\gamma)\right].$$

If $x+z-\gamma \geq 0$, then $\psi'_x(z)$ is obviously positive. If $x+z-\gamma < 0$, then the inequality $1-\Phi(a) < \phi(a)/a$ for all $a > 0$, see, e.g., Feller (1968), page 175, implies that

$$\frac{\psi'_x(z)}{e^{\left(\frac{1}{2}z^2+z(x-\gamma)\right)}} = (x+z-\gamma)[1-\Phi(-x-z+\gamma)] + \phi(-x-z+\gamma) > 0.$$

Hence, ψ_x is strictly increasing, and it follows from (2.19) and Lemma 3 that

$$g'(x) < 0 \text{ if } \zeta < \min\{0, 2\gamma\} \text{ or } \zeta > \max\{0, 2\gamma\}, \tag{2.20}$$

$$g'(x) > 0 \text{ if } \min\{0, 2\gamma\} < \zeta < \max\{0, 2\gamma\}. \tag{2.21}$$

Suppose now that $\delta > \delta^*$. Then $\zeta < 0$. If $\delta > 1$, then $\zeta < 2\gamma$, so that, by (2.20), $g'(x) < 0$ and $(1-\delta)g'(x) > 0$. If $\delta < 1$, then $0 > \zeta > 2\gamma$, so that, by (2.21), $g'(x) > 0$ and again $(1-\delta)g'(x) > 0$.

Suppose next that $\delta = \delta^*$. Then $\zeta = 0$ and in view of (2.19), $(1-\delta)g'(x) = 0$.

Suppose finally that $\delta < \delta^*$. Then $\zeta > 0$. If $\delta > 1$, then $0 < \zeta < 2\gamma$, so that, by (2.21), $g'(x) > 0$ and $(1-\delta)g'(x) < 0$. If $\delta < 1$, then $\zeta > 2\gamma$, so that, by (2.20), $g'(x) < 0$ and again $(1-\delta)g'(x) < 0$.

Lemma 3 *Let $\psi : \mathbb{R} \to \mathbb{R}$ be a strictly increasing function. Let $\gamma \in \mathbb{R}$. Then*

$$(z-2\gamma)\psi(0) + z\psi(2\gamma) < 2(z-\gamma)\psi(z) \textit{ if } z < \min\{0, 2\gamma\} \textit{ or } z > \max\{0, 2\gamma\}$$

and

$$(z-2\gamma)\psi(0) + z\psi(2\gamma) > 2(z-\gamma)\psi(z) \textit{ if } \min\{0, 2\gamma\} < z < \max\{0, 2\gamma\}.$$

Proof. If $\gamma = 0$, the claim reduces to $z\psi(0) < z\psi(z)$ for all $z \neq 0$. If $\gamma \neq 0$ and $z = \gamma$, the claim reduces to $-\gamma\psi(0) + \gamma\psi(2\gamma) > 0$. In either case, the claim is an immediate consequence of the assumption that ψ is strictly increasing.

Suppose now that $\gamma \neq 0$. Set $\alpha = \min\{0, 2\gamma\}$, $\beta = \max\{0, 2\gamma\}$, and define for all $z \neq \gamma$,

$$h(z) = \frac{(z-2\gamma)\psi(0) + z\psi(2\gamma)}{z-\gamma}.$$

Then $h(\alpha) = 2\psi(\alpha)$, $h(\beta) = 2\psi(\beta)$, and

$$h'(z) = \frac{\gamma[\psi(0) - \psi(2\gamma)]}{(z-\gamma)^2} < 0 \text{ for all } z \neq \gamma,$$

so that h is strictly decreasing on $(-\infty, \gamma)$ and on (γ, ∞). Hence,

$$\begin{aligned} h(z) > h(\alpha) = 2\psi(\alpha) > 2\psi(z) \quad & \text{if } z < \alpha, \\ h(z) < h(\alpha) = 2\psi(\alpha) < 2\psi(z) \quad & \text{if } \alpha < z < \gamma, \\ h(z) > h(\beta) = 2\psi(\beta) > 2\psi(z) \quad & \text{if } \gamma < z < \beta, \\ h(z) < h(\beta) = 2\psi(\beta) < 2\psi(z) \quad & \text{if } \beta < z. \end{aligned}$$

Multiplying by $(z - \gamma)$ completes the proof.

3 Early Default Risk and Surrender Risk: Impacts on Participating Life Insurance Policies

3.1 Introduction

A typical participating life insurance contract provides policyholders a minimum interest rate guarantee and bonus payments upon death and upon survival which are linked to the performance of the insurance company. Usually, additional options are embedded in the contracts to increase their attraction to the policyholders, among which the most popular one is the surrender option. Surrender option entitles the policyholders the right to terminate their contracts prematurely and to obtain the surrender benefit promised by the insurance companies.

The policyholders may not necessarily receive the payments specified in their contracts even if they hold the contracts till the maturity date. If the insurance company does not have enough reserves to pay back its liabilities at the maturity, the policyholders cannot get more than what remains in the company. To protect the policyholders from collecting too little benefit as the insurance company declares bankruptcy at the maturity, regulatory authorities impose early default mechanisms to monitor insurance companies' financial status and close them before it is too late so. For example, under Solvency II, if a company's capital becomes lower than the minimum capital requirement, its licence is revoked by the supervisory authority, see Gatzert and Wesker (2011). Also, an insurance company supervised by the Swiss Financial Market Supervisory Authority (FINMA) can lose its license when its risk-based capital drops below the lowest threshold specified in the Swiss Solvency Test (SST), see FINMA Circ. 08/44 'SST' (2008). Proceeds from liquidated assets are then paid to stakeholders. Hence, the policyholders also face early default risk of the insurance company accompanied with the early regulatory intervention.

Both surrender decision and early default intervention definitely have an direct impact on the fair value of participating policies since they change the policies' payment streams. In the existing literature, most studies focus on only one of these two aspects. For example, Grosen and Jørgensen (2000), Bacinello (2003) and Andreatta and Corradin (2003) study the fair valuation of participating life insurance contracts with embedded surrender options but have not counted in the early default risk triggered by the bad performance of the insurance company, while Grosen and Jørgensen (2002), Jørgensen (2001), Bernard et al. (2005) and Chen and Suchanecki (2007) take into account regulatory intervention in evaluating participating contracts, but leave out surrender risk. The only work that treats early default risk and surrender risk at the same time is Le Courtois and Nakagawa (2013) who model surrender risk through a Cox process of an intensity which is correlated to the financial market but is independent of the company's liquidation threshold. However, since the early termination of the insurance company imposed by the regulator reframes the payment structure for the policyholders, which we consider as direct impacts, as a response the policyholders may change their surrender decision-making behaviors. This influence of the enforced early bankruptcy on the policyholders' surrender behaviors can be considered as a 'by-product' of the regulatory intervention, which in turn affects the policies' payments and correspondingly, the contracts' value. In this paper, we closely analyze both the direct and indirect ('by-product') impacts of the early default risk on the pricing of the participating policies by endogenously modeling the policyholders' surrender behaviors and uncovering the impacts of the early default regulation on the surrender behaviors. Besides, when regulatory rules change, insurance companies may react differently to the regulation by adopting different investment strategies, which again affects the contract values directly and indirectly through its influence on policyholders' surrender behavior. We hence also study the impacts of the investment strategy on the fair value of the insurance contracts and how the insurance companies choose the investment strategies in face of the regulatory rules.

To describe the early default risk we adopt the regulatory framework in Grosen and Jørgensen (2002), Jørgensen (2001) and Bernard et al. (2005), where liquidation is triggered as the insurance company's asset value drops below a threshold. Concerning surrendering, in most literature, it is assumed that the policyholders are fully rational, which means they can terminate the contract at an optimal time point so that the contract value is maximized, see e.g., Grosen and Jørgensen (1997), Grosen and Jørgensen (2000), Andreatta and Corradin (2003), Bacinello (2003), and Bacinello (2005), to just name a few. However, since there is not an active market to monitor the contract values, the surrender option is hardly exercised at the right time if a policyholder is not capable of evaluating the contract correctly. Also due to the lack of a trading market for the contracts, the policyholders, when in urgent liquidity needs, have to surrender their contracts at the insurance company and collect the surrender guarantees, which are usually lower than the fair contract values. Empirical evidences which confirm the so called emergency hypothesis are found e.g. in Kuo et. al. (2003) and Kiesenbauer (2011). Given the limitations, it is more reasonable to consider policyholders as partly rational from a purely financial point of view. We adopt the approach of modeling policyholders' partial rationality in Li and Szimayer (2014). They consider surrender as a randomized event assuming that arrival of the event follows a Poisson process with an intensity bounded from below and from above. The lower and upper bounds refer to the minimum surrender rate due to exogenous reasons and maximum surrender rate due to limited rationality respectively. The maximum contract values are then derived by choosing surrender intensities within the two bounds in the worst case scenario (from the perspective of the insurers). This approach corresponds to the spirit of Solvency II. CEIOPS[1] has pointed out that it may be necessary to differentiate between different insurance products for the purpose of the mass lapse stress.[2] It is also pointed out by the CEIOPS that the lapse risk should be treated differently for different policyholders. For example, the risk of a mass lapse is substantially greater if the policyholders are institutional investors since they tend to be better informed and react more quickly.[3] This indicates that the rationality level of the policyholders also plays a significant role in the analysis. Therefore, we will consider different bounded values of the surrender intensities and analyze how the influence of regulation rules on surrender behaviors differs with respect to policyholders' rationality. Similar to Li and Szimayer (2014) we derive a partial differential equation (PDE) to characterize the price of a participating policy. However, this PDE is only valid when the liquidation threshold has not been touched yet. Otherwise, the policy takes immediately the liquidation value. In this sense, we are solving a barrier option pricing problem. We apply the finite difference method proposed in Zvan et al. (1996) and Zvan et al. (2000) to solve this problem numerically.

[1]CEIOPS refers to the Committee of European Insurance and Occupational Pensions Supervisors. It was replaced by the European Insurance and Occupational Pensions Authority (EIOPA) since 2011.

[2]CEIOPS advices to take three lapse scenarios into consideration and choose the worst among them to calculate the capital charge for lapse risk, see CEIOPS (2009). The advice of limiting to three scenarios is based on the complexity of situations in reality. The mass lapse event is one of the three scenarios considered during the calculation. In our simplified model, however, we are able to find the worst case scenario dynamically.

[3]See CEIOPS (2009).

The paper is organized as follows. In section 3.2 we model the insurance company and introduce the payoff structure of a participating policy. The early default regulatory framework is specified as well. Besides, both the financial market and the insurance market are modeled with respect to the stochastic processes of the underlying asset, the mortality risk intensity and the surrender risk intensity. In section 3.3 we derive the PDE for the price of the policy. In section 3.4 we analyze the effect of the regulatory framework and the investment strategy on surrender behaviors as well as on contract values. Section 3.5 concludes.

3.2 Model Framework

3.2.1 Company Overview

Inspired by the model framework in Briys and de Varenne (1994), we consider a life insurance company which acquires an asset portfolio with the initial value A_0 at time $t_0 = 0$ financed by two agents, i.e., a policyholder and an equity holder. The policyholder pays a premium to acquire the initial liability $L_0 = \alpha A_0$ with $\alpha \in (0, 1)$. The rest is levied from the equity holder who acquires $E_0 \equiv (1-\alpha)A_0$ with limited liability. The insurance company's balance sheet at time t_0 is shown in Table 3.1. The parameter α is called as a wealth distribution coefficient in Grosen and Jørgensen (2002).

Assets	Liabilities & Equity
A_0	$L_0 \equiv \alpha A_0$
	$E_0 \equiv (1-\alpha)A_0$

Table 3.1: *Insurance company's balance sheet at* t_0

It is assumed that the insurance company operates in a frictionless complete and arbitrage-free financial market over a time interval $[0, T]$, where time T corresponds to the expiration date of the insurance contract. As the insurance contract expires at time T, the insurance company closes and its assets are liquidated and distributed to stakeholders.[4]

[4] For simplicity, we assume the company closes when the contract ends. It is not a strict assumption because it can be considered that assets raised from the policyholder and the equity holder are put in a separate fund, as the contract ends, the fund is closed and assets left in the fund are liquidated and distributed to the stakeholders.

3.2.2 Participating Life Insurance Policy

By investing in the insurance company at time t_0, the policyholder signs a participating insurance contract which promises the policyholder a part of the insurance company's profits in addition to the guaranteed minimum interest rate at maturity date T. If the policyholder dies before time T, the contract pays death benefits. Additionally, the policyholder can exercise the surrender option embedded in the contract before maturity T and collects surrender benefits from the insurance company. To summarize, the contract promises survival benefits, death benefits and surrender benefits, depending on which event happens first. In any event, the policyholder has a priority claim on the company's assets and the equity holder receives what is left.

As the contract expires at maturity T, the policyholder receives a minimum guaranteed benefit, which is given by compounding the initial liability L_0 with a minimum guaranteed interest rate r_g, i.e., $L_T^{r_g} = L_0 e^{r_g T}$, and a bonus conditional on that the asset value generated by the contribution of the policyholder is enough to cover the minimum guaranteed benefit, i.e., $\alpha A_T \geq L_T^{r_g}$. Suppose δ is a participation rate in the asset surplus, the profits shared with the policyholder are $\delta[\alpha A_T - L_T^{r_g}]^+$. However, it may happen that at time T when the company's assets are liquidated, the assets' value is lower than the value of the minimum guaranteed benefit. In this case, based on the assumptions that the policyholder has a priority claim on the company's assets and the equity holder has limited liability, the policyholder collects what is left, i.e., A_T, and the equity holder walks away with nothing in his hands. To sum up, when the contract survives until the maturity T the policyholder receives survival benefits which take the form

$$\Phi(A_T) = L_T^{r_g} + \delta[\alpha A_T - L_T^{r_g}]^+ - [L_T^{r_g} - A_T]^+. \tag{3.1}$$

The policyholder may die before the contract matures. We use τ_d to denote the death time of the policyholder aged x at time t_0. At time $\tau_d < T$, the contract pays death benefits to the policyholder. We assume that death benefits have the same payment structure as survival benefits, but with all the components evaluated at the death time τ_d. We use r_d and δ_d to denote the minimum guaranteed interest rate and the participation rate for calculating the promised minimum guarantee, i.e., $L_{\tau_d}^{r_d} = L_0 e^{r_d \tau_d}$, and the asset surplus, respectively. Then, the death benefits have the following form at time τ_d

$$\Psi(\tau_d, A_{\tau_d}) = L_{\tau_d}^{r_d} + \delta_d[\alpha A_{\tau_d} - L_{\tau_d}^{r_d}]^+ - [L_{\tau_d}^{r_d} - A_{\tau_d}]^+, \tag{3.2}$$

Furthermore, by exercising the surrender option embedded in the contract, the policyholder can terminate the contract before the expiration date T. We use τ_s to denote the surrender time. Once the surrender option is exercised, the company closes and its assets are liquidated and paid to the policyholder as specified in the contract but not more than the liquidated asset value. We consider the following surrender payment form for the policyholder

$$S(\tau_s, A_{\tau_s}) = L_{\tau_s}^{r_s} - [L_{\tau_s}^{r_s} - A_{\tau_s}]^+, \tag{3.3}$$

where $L_{\tau_s}^{r_s} = (1 - \beta_{\tau_s})L_0 e^{r_s \tau_s}$ is the minimum surrender guarantee when the asset value suffices. Here, r_s is the minimum guaranteed interest rate at surrender and β_{τ_s} is a penalty parameter which penalizes the policyholder for early terminating the contract and is assumed to be a deterministic decreasing function of the time. After the policyholder is paid off, the equity holder receives the rest of the asset value.

3.2.3 Early Default Mechanism

Now we introduce early default risk of insurance company into the model. We consider the presence of an external regulator who watches on the insurance company's financial status over its operating time horizon. We abstract away from cumbersome bankruptcy rules and procedures applied to insurance companies in practice and assume the insurance company is on-going until either the external regulator intervenes before T or the insurance contract matures at T. We adopt the regulatory mechanism introduced by Grosen and Jørgensen (2002) and set up a default-triggering barrier based on the minimum survival guarantee $B_t = \theta L_0 e^{r_g t}$, where θ is a default multiplier. Once the company's asset value drops below the barrier before maturity T, the company is closed by the regulator and its assets are liquidated and distributed to the stakeholders. Accordingly, we define the early default time τ_b as the first time that the asset value drops below the barrier,

$$\tau_b = \inf \{t \mid A_t \leq B_t\}. \tag{3.4}$$

At time τ_b, the policyholder receives early default benefits, denoted by $\Upsilon(\tau_b, A_{\tau_b})$, which have the lower value of the liquidated assets and the minimum survival guarantee accrued at the guarantee rate r_g up to the early default date,

$$\Upsilon(\tau_b, A_{\tau_b}) = \min\{A_{\tau_b}, L_{\tau_b}^{r_g}\}, \tag{3.5}$$

where $L_{\tau_b}^{r_g} = L_0 e^{r_g \tau_b}$. Accordingly, if the company has the liquidated assets more than the promised minimum guarantee, the equity holder obtains what is left after paying off the policyholder; otherwise, the equity holder gets nothing.

The default multiplier θ is set by the regulator, which actually reflects how intensively the regulator monitors the insurance company and how strongly the regulator intends to protect the policyholder. If the regulator believes that the insurance company is inclined to take the advantage of the policyholder by running a risky business or is not competent enough to manage its assets, the regulator may set a higher default multiplier to protect the policyholder. This implies that the insurance company must bear a higher early default risk. Otherwise, the regulator will set a lower default multiplier, which allows the insurance company to recover from its temporary bad performance. In our model, we restrict θ to be smaller than $1/\alpha$, which ensures $A_0 > B_0$ so that the insurance company does not default at the initial time t_0 when the contract is just issued to the policyholder. The restriction makes economic sense since in reality there would be no equity holder who would like to invest in a company that will be shut down by the government at the opening day.

3.2.4 Mathematical Formulation

In this section we model the financial market and the insurance market mathematically. We fix a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, where $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ reflects the flow of information available on the financial market and the insurance market. We assume that the company invests its total initial assets in traded (risk-free and/or risky) assets on the financial market, where the risk-free interest rate, denoted by r, is assumed to be deterministic in time. Under the market probability measure $\mathbb{P}$, the company's asset price process A is assumed to be governed by the following stochastic process

$$\mathrm{d}A_t = a(t, A_t)\, A_t \mathrm{d}t + \sigma(t, A_t) A_t \,\mathrm{d}W_t, \quad \forall t \in [0, T]. \tag{3.6}$$

Here W is a standard Brownian motion under $\mathbb{P}$ and generates the filtration $\mathbb{F}^W = \{\mathcal{F}_t\}_{0\leq t\leq T}$. The functions a and $\sigma > 0$ refer to the expected rate of return and the volatility of the asset process respectively, and both are regular enough to guarantee the unique solution of (3.6).

As the payoff of the contract depends not only on the asset value itself but also on the occurrence of the death event or the surrender event, we enlarge the filtration $\mathbb{F}^W$ in a minimal way to summarize all the information relevant to the contract valuation. The filtration $\mathbb{F}^W$ is thus enlarged to $\mathbb{G} = \mathbb{F}^W \vee \mathbb{H}$ where $\mathbb{H}$ is jointly generated by the jump processes $H_t = 1_{\{\tau_d \leq t\}}$ and $J_t = 1_{\{\tau_s \leq t\}}$, i.e., the information about whether the policyholder dies before time t and whether he surrenders the contract before time t, respectively. The hazard rate of the random time τ_d, also called mortality intensity, is denoted by μ and is assumed to a deterministic function of time.[5] Similarly, we call the hazard rate of the random time τ_s the surrender intensity and denote it by γ. The surrender intensity γ needs to be modeled here. Based on our arguments in the introduction, we take into account that the policyholder has partial rationality in surrendering the contract. We follow the approach adopted in Li and Szimayer (2014) by assuming the surrender intensity bounded from below by $\underline{\rho}$ as surrendering is not a financially optimal decision for the policyholder, and from above by $\bar{\rho}$ as surrendering becomes financially optimal, with $\bar{\rho} > \underline{\rho}$. Due to personal reasons which urge the policyholder to surrender the contract prematurely, the lower bound of the surrender intensity $\underline{\rho}$ is assumed in any case, i.e., the surrender intensity at least takes the value of $\underline{\rho}$. The size of the increase in the value of surrender intensity as surrendering becomes financial optimal to the policyholder, i.e., $\bar{\rho} - \underline{\rho}$, measures how frequently the policyholder updates his financial market information and makes a surrendering decision when it is optimal to do so. The more frequently he updates and/or analyzes financial information, the larger the increase in the intensity is, accordingly, the more rational he is. In case that $\bar{\rho} = \infty$, we are back in the setting where the policyholder may surrender the contract at any time when it is optimal to do so and a pure American-style contract should be priced by solving an optimal stopping problem. The decision will be made by comparing the continuation value of the contract and the value of surrender benefits, which are denoted by $v(t, A)$ and $S(t, A)$ respectively. Depending on which decision the policyholder makes, the endogenous surrender intensity is thus either 0 or $\bar{\rho} - \underline{\rho}$. To summarize, the surrender intensity takes the form of

$$\gamma_t = \begin{cases} \underline{\rho}, & \text{for } S(t, A) < v(t, A) \\ \bar{\rho}, & \text{for } S(t, A) \geq v(t, A). \end{cases} \tag{3.7}$$

On the current information available on the financial market and the insurance market, the arrival of the death event, the arrival of the surrender event and W are independent.

[5] In the literature there are many discussions on stochastic mortality intensity which is more consistent with the reality, see e.g. Bacinello et al. (2010), Biffis et al. (2010), Dahl (2004), Dahl and Møller (2006). However, the stochastic feature of the mortality intensity does not have too much influence on the contract value, see Li and Szimayer (2011). Hence, we assume a deterministic mortality intensity for simplicity and focus more on the surrender risk and the early default risk.

Thus, we have that W is a $\mathbb{G}$-martingale, and μ and γ are $\mathbb{G}$-intensities of the random death time τ_d and the random surrender time τ_s, respectively.

In the absence of arbitrage, we price the participating life insurance contract by risk-neutral valuation with a risk-neutral measure $\mathbb{Q}$ for the insurer. Under the risk-neutral measure $\mathbb{Q}$ on the filtration generated by the financial market, the company's asset process is then described by

$$\mathrm{d}A_t \quad = \quad r(t)\, A_t \mathrm{d}t + \sigma(t, A_t) A_t \, \mathrm{d}W_t^{\mathbb{Q}}, \;\; \forall t \in [0, T], \tag{3.8}$$

where $W^{\mathbb{Q}}$ is a standard 1-dimensional Brownian motion under $\mathbb{Q}$. Taking the mortality risk and the surrender risk into consideration, pricing the participating life insurance contract under the martingale measure $\mathbb{Q}$ with μ and γ as $\mathbb{G}$-intensities of the random times τ_d and τ_s, respectively, requires additional justification and assumption. Given that the mortality intensity is deterministic, the mortality risk is diversifiable for the insurer if the pool of policyholders is large enough, and thus μ is the $(\mathbb{Q}, \mathbb{G})$-intensity of the arrival of the death. The surrender intensity specified in (3.7) corresponds to the worst-case scenario from the perspective of the insurance company. We therefore assume that the insurance company does not ask for extra risk premium above the worst case surrender intensity any more when changing from the measure $\mathbb{P}$ to the measure $\mathbb{Q}$.[6] The bounds $\underline{\rho}$ and $\bar{\rho}$ are also valid under the measure $\mathbb{Q}$ and the surrender intensity specified in (3.7) corresponds to the $(\mathbb{Q}, \mathbb{G})$-intensity of τ_s on the enlarged market represented by the filtration $\mathbb{G}$.[7] The contract value obtained under the measure $\mathbb{Q}$ with the worst-case surrender intensity γ can be interpreted as the upper price bound of the contract. We address this issue formally in Remark 1. Consequently, the choice of the surrender intensity in (3.7) is not only motivated by the observations of the policyholders' surrender behaviors on the market but also by the worst-case scenario analysis within a reasonable range that is often adopted in practice.

[6] Alternatively, a higher market price for the surrender risk may be charged by lowering the lower bound $\underline{\rho}$ and increasing the upper bound $\bar{\rho}$ under the measure $\mathbb{Q}$. In Proposition 3 we show formally that a lower $\underline{\rho}$ and a higher $\bar{\rho}$ lead to a higher contract value.

[7] Confer Li and Szimayer (2014) for a formal explanation of the surrender intensity after the change of measure.

3.3 Contract Valuation

In this section we evaluate the contract by taking both the (early) default risk and the surrender risk into consideration. We are not able to find a closed-form pricing formula for the contract, since the surrender intensity can only be determined endogenously within our model. However, by applying the PDE approach, we can specify the surrender intensity and the contract value simultaneously. The contract value is thus not represented by a pricing formula but characterized by a PDE equation and solved numerically with the finite difference method. This approach is applied by Li and Szimayer (2014) in a similar setting when they price unit-linked life insurance contracts with surrender guarantees. The crucial point in this chapter is that after introducing the early default mechanism, the contract payoff to the policyholder is linked to the solvency of the company and has a barrier option property. Thus, in order to evaluate the contract, we need to distinguish between the region where $A_t \leq B_t$ and the region where $A_t > B_t$ for $t \in (0, T)$, which is similar to the barrier option pricing. For $A_t \leq B_t$ at time $t \in (0, T)$, the insurance company must be liquidated and the policyholder only obtains $\Upsilon(t, A_t)$. For $A_t > B_t$, we represent the contract value V_t on $\{t \leq \tau_d \wedge \tau_s \wedge T\}$ by

$$V_t = \mathbb{1}_{\{t<\tau_d \wedge \tau_s\}} v(t, A_t) + \mathbb{1}_{\{t=\tau_d, \tau_d<\tau_s, T\}} \Psi(\tau_d, A_{\tau_d}) + \mathbb{1}_{\{t=\tau_s, \tau_s<\tau_d, T\}} S(\tau_s, A_{\tau_s}), \qquad (3.9)$$

where v is a suitably differentiable function $v : [0, T] \times \mathbb{R}^+ \to \mathbb{R}_0^+,\ (t, A) \mapsto v(t, A)$, representing the pre-death/surrender value. Then we apply the balance law based on the no-arbitrage condition on the set $\{t < \tau_d \wedge \tau_s \wedge \tau_b \wedge T\}$

$$r(t)V(t, A_t)\mathrm{d}t = \mathbb{E}_{\mathbb{Q}}[\mathrm{dV_t}|\mathcal{G}_\mathrm{t}]. \qquad (3.10)$$

On the set $\{t < \tau_d \wedge \tau_s \wedge \tau_b \wedge T\}$, we compute the differential of V as[8]

$$\begin{aligned} \mathrm{d}V_t \;&=\; \mathrm{d}v(t, A_t) + (\Psi(t, A_t) - v(t, A_t))\mathrm{d}H_t \\ &\quad +(S(t, A_t) - v(t, A_t))\mathrm{d}J_t, \;\text{ for } 0 \leq t < T, \end{aligned} \qquad (3.11)$$

where H and J refer to the jump processes with the $\mathbb{Q}$-intensities μ and γ respectively. A jump in H or J leads to a change in the payment liability either of the amount $\Psi(t, A_t)-v(t, A_t)$ or $S(t, A_t)-v(t, A_t)$. Plugging (3.11) into (3.10) and using $V_t = v(t, A_t)$

[8]Notice that in the region $A_t > B_t$, there would not be early default after the instantaneous time period $\mathrm{d}t$ since the asset process is assumed to be continuous in our model.

at time $t < \tau_d \wedge \tau_s \wedge \tau_b \wedge T$, we obtain

$$
\begin{aligned}
r(t)v(t,A_t)\mathrm{d}t &= \mathbb{E}_{\mathbb{Q}}[dv(t,A_t)|\mathcal{G}_t] + (\Psi(t,A_t) - v(t,A_t))\mu(t)\mathrm{d}t \\
&\quad +(S(t,A_t) - v(t,A_t))\gamma_t \mathrm{d}t. \qquad (3.12)
\end{aligned}
$$

By applying Ito's Lemma to $\mathrm{d}v(t, A_t)$, we have

$$
\begin{aligned}
\mathbb{E}_{\mathbb{Q}}[\mathrm{d}v(t,A_t)|\mathcal{G}_t] &= \mathbb{E}_{\mathbb{Q}}\left[\mathcal{L}v(t,A_t)\mathrm{d}t + \sigma(t,A_t)A_t\frac{\partial v}{\partial A}(t,A_t)\mathrm{d}W_t^{\mathbb{Q}}\Big|\mathcal{G}_t\right] \\
&= \mathcal{L}v(t,A_t)\mathrm{d}t, \qquad (3.13)
\end{aligned}
$$

where $\mathcal{L}v(t,A) = \frac{\partial v}{\partial t}(t,A) + r(t)A\frac{\partial v}{\partial A}(t,A) + \frac{1}{2}\sigma^2(t,A)A^2\frac{\partial^2 v}{\partial A^2}(t,A)$. Then, on the set $\{t < \tau_d \wedge \tau_s \wedge \tau_b \wedge T\}$ we have

$$
\mathcal{L}v(t,A_t) + \mu(t)\Psi(t,A_t) + \gamma_t S(t,A_t) - (r(t)+\mu(t)+\gamma_t)v(t,A_t) = 0. \qquad (3.14)
$$

We summarize the pricing PDE with the following proposition.

Proposition 2 *For the contract value V described by* (3.9)*, the pre-death/surrender value v for $(t,A) \in [0,T) \times \mathbb{R}^+$ is the solution of the partial differential equation*

$$
\mathcal{L}v(t,A_t) + \mu(t)\Psi(t,A_t) + \gamma_t S(t,A_t) - (r(t)+\mu(t)+\gamma_t)v(t,A_t) = 0\,, \qquad (3.15)
$$

where

$$
\gamma_t = \begin{cases} \underline{\rho}\,, & \text{for } S(t,A_t) < v(t,A_t)\,, \\ \bar{\rho}\,, & \text{for } S(t,A_t) \geq v(t,A_t)\,; \end{cases} \qquad (3.16)
$$

subject to the boundary condition

$$
v(t,A_t) = \Upsilon(t,A_t),\ \text{for } t \in [0,T),\ A_t = B_t = \theta L_0 e^{r_g t}, \qquad (3.17)
$$

and with the termination condition

$$
v(T,A_T) = \Phi(A_T), \quad \text{for } A_T \in \mathbb{R}^+. \qquad (3.18)
$$

The integral representation of the solution to the above pricing PDE is shown in Corollary 1 and proved in Appendix 3.6.1.

Corollary 1 *Suppose the surrender intensity γ is given. The value of the participating* *policy V can be represented on $\{t < \tau_s \wedge \tau_d \wedge \tau_b \wedge T\}$ by*

$$V_t = \mathbb{E}_{\mathbb{Q}}\Bigg[\int_t^{\tau_b \wedge T} e^{-\int_t^m (r(u)+\mu(u)+\gamma(u,A_u))\mathrm{d}u}(\mu(m)\Psi(m,A_m)+\gamma(m,A_m)S(m,A_m))\mathrm{d}m$$
$$+ 1_{\{\tau_b \geq T\}}\Phi(A_T)e^{-\int_t^T (r(u)+\mu(u)+\gamma(u,A_u))\mathrm{d}u} + 1_{\{\tau_b < T\}}\Upsilon(\tau_b, A_{\tau_b})e^{-\int_t^{\tau_b}(r(u)+\mu(u)+\gamma(u,A_u))\mathrm{d}u}\Bigg|\mathcal{G}_t\Bigg]. \tag{3.19}$$

Remark 1 *The pricing problem can be formulated as looking for the worst case of the risk-adjusted surrender intensity γ so that the contract value is maximized under the martingale measure $\mathbb{Q}$ on $\{t < \tau_s \wedge \tau_d \wedge \tau_b \wedge T\}$,*

$$v(t,A) = \sup_{\gamma \in \Gamma(t,A)} \mathbb{E}_{\mathbb{Q}}^{t,A}\Bigg[\int_t^{\tau_b \wedge T} e^{-\int_t^m (r(u)+\mu(u)+\gamma(u,A_u))\mathrm{d}u}(\mu(m)\Psi(m,A_m)+\gamma(m,A_m)S(m,A_m))\mathrm{d}m$$
$$+ 1_{\{\tau_b \geq T\}}\Phi(A_T)e^{-\int_t^T (r(u)+\mu(u)+\gamma(u,A_u))\mathrm{d}u} + 1_{\{\tau_b < T\}}\Upsilon(\tau_b, A_{\tau_b})e^{-\int_t^{\tau_b}(r(u)+\mu(u)+\gamma(u,A_u))\mathrm{d}u}\Bigg] \tag{3.20}$$

where $\Gamma(t,A) = \{\gamma : [t,T] \times \mathbb{R}^+ \mapsto \mathbb{R}^+ : \underline{\rho} \leq \gamma(u,A) \leq \bar{\rho},$ for all $t \leq u \leq T$ and $A \in \mathbb{R}^+\}$ and $\mathbb{E}_{\mathbb{Q}}^{t,A}$ denotes the expectation conditional on $A_t = A$ under measure $\mathbb{Q}$.[9] This is a stochastic control problem, which can be solved, according to the theorem of the Hamilton-Jacobi-Bellman equation (confer Yong (1997) and Yong and Zhou (1999)), by dealing with an equivalent problem

$$0 = \sup_{\gamma \in \Gamma(t,A)} \mathcal{L}v(t,A) + \mu(t)\Psi(t,A) + \gamma(t,A)S(t,A) - (r(t)+\mu(t)+\gamma(t,A))v(t,A), \tag{3.21}$$

subject to $v(t,A) = \Upsilon(t,A)$, for $A = B_t = \theta L_0 e^{r_g t}$, and $v(T,A) = \Phi(A)$, for $A \in \mathbb{R}^+$. γ needs to be optimally controlled: since in the equation above the part that depends on γ is linear in γ, i.e., $\gamma(t,A)(S(t,A) - v(t,A))$, the solution to the problem is exactly the same as is presented in equation (3.7).

Within a given regulatory framework and under a given investment strategy, i.e., for given θ and σ, we can prove formally that a lower value of $\underline{\rho}$ and a higher value of $\bar{\rho}$ lead to an increase of the contract value, see Proposition 3. This is consistent with our

[9] Since the financial market is complete and arbitrage free, the asset process A is unique under measure $\mathbb{Q}$.

intuition, since a lower $\underline{\rho}$ or a higher $\bar{\rho}$ indicates the increase of the rationality level of the policyholder in the monetary sense and thus increases the contract value. The proof is provided in Appendix 3.6.2.

Proposition 3 *Suppose the early default mechanism is characterized by the default multiplier θ and the insurance company's investment strategy by σ. Furthermore, suppose that v is the pre-death/surrender value function of the participating policy with bounds of the surrender intensity being $\underline{\rho}$ and $\bar{\rho}$, and that w is the pre-death/surrender value function of the policy with bounds $\underline{\zeta}$ and $\bar{\zeta}$. Assume that $\underline{\zeta} \leq \underline{\rho}$ and $\bar{\rho} \leq \bar{\zeta}$. Then we have $w(t, A) \geq v(t, A)$, for $(t, A) \in [0, \tau_b \wedge T] \times \mathbb{R}^+$.*

3.4 Numerical Analysis

In this section we adopt the finite difference method proposed by Zvan et al. (1996) and Zvan et al. (2000) to numerically solve the PDE with a continuously applied barrier (3.15) as stated in Proposition 2 and study the effects of the early default risk and the surrender risk on the fair valuation of the contract as well as on the insurance company's investment strategies. The insurance company is set up with initial asset value $A_0 = 100$ and 85% of the asset value is acquired by the policyholder who buys the participating contract at time t_0 as the initial liability, which means $\alpha = 0.85$. The contract matures in $T = 10$ years and promises the same participation rate $\delta = \delta_d = 0.9$ at maturity and at death.[10] The risk-free interest rate is $r = 0.04$. The volatility of the company's asset process is constant, i.e., $\sigma(t, A_t) = 0.2$.[11] The volatility provides information about the riskiness of the insurance company's investment strategy. A higher σ indicates a higher riskiness of the investment strategy while a lower σ implies a more conservative investment strategy.[12] The minimum guaranteed interest rates at survival, at death and at surrender are $r_g = r_d = r_s = 0.02$. As for the mortality intensity, we follow Li and Szimayer (2014) and assume that it follows a deterministic process $\mu(t) = A^\mu + Bc^{x+t}$ for the policyholder aged $x = 40$ at $t_0 = 0$ with $A^\mu = 5.0758 \times 10^{-4}$, $B = 3.9342 \times 10^{-5}$, $c = 1.1029$. Additionally, the penalty

[10] Regulators usually require the participation rate to be kept at least at a certain level. In Germany, e.g., it lies at 90%.

[11] The volatility level is obviously higher than what is adopted in practice. We choose a relatively higher level as a benchmark to see the effect of the regulation on the insurance company's incentive in selecting the riskiness of the investment strategy. We will show that an effective regulation rule leads the insurance company to choose a conservative investment strategy voluntarily.

[12] On a market with only one risk-free and one risky asset, a higher asset volatility is achieved by investing more into the risky asset.

parameter takes the form

$$\beta_t = \begin{cases} 0.05, & \text{for } t \leq 1, \\ 0.04, & \text{for } 1 < t \leq 2, \\ 0.02, & \text{for } 2 < t \leq 3, \\ 0.01, & \text{for } 3 < t \leq 4, \\ 0, & \text{for } t > 4. \end{cases}$$

The parameters are summarized again in Table 3.2.

Market Parameters		Contract Parameters	
A_0	100	α	85%
r	0.04	T	10
σ	0.2	δ, δ_d	0.9
A^{μ}	5.0758×10^{-4}	r_g, r_d, r_s	0.02
B	3.9342×10^{-5}		
c	1.1029		

Table 3.2: *Parameter specifications*

The analysis in the following subsections is conducted for a representative policyholder. Under the assumption that the pool of policyholders is large enough, the surrender intensity of a representative policyholder gives the indication at the portfolio level about the proportion of policyholders who will surrender the contracts. The implications for the large pool of policyholders will be summarized in section 3.5 to conclude the paper.

3.4.1 Effects of Regulatory Frameworks on Contract Valuation

In this section we analyze the effects of the early default risk on fair contract valuation. The magnitude of the early default risk depends on the strictness of the regulatory framework, which in our model is represented by the default multiplier θ that is specified by the regulator. It indicates how the regulator judges the insurance company's ability to manage its assets. If the regulator is very confident about the expertise of the insurance company and about the financial market, it will tolerate the temporary poor performance of the company more and hence choose a lower default multiplier so that the company has the chance to recover. Otherwise, it will set a higher value to protect the policyholders from not being able to obtain the guaranteed benefits promised by the company. Although a lower (higher) default multiplier is less (more) effective to protect the policyholders from the downside development of the company, it gives the company more (less) chance to recover from the temporary bad performance and pay out more (less) to the policyholders when it does recover. Hence, the level of the default multiplier has great influence on the payoff of the contract and thus on the contract value.

Furthermore, the policyholder takes into account the impacts of the protection from the regulator on the payments of his contract and adjusts his surrender behavior accordingly, which indirectly influences the contract value. Intuitively, the policyholder makes his surrender decision not only based on benefits that are promised by the insurance company but also on the ability of the insurance company to meet its promise. The early default mechanism insures the ability of the company to meet its promise by imposing a limit on the asset value of the insurance company. A higher default multiplier indicates that the policyholder has to worry less about the second issue because they are better protected and will surrender the contract only when the surrender benefits are very attractive. On the contrary, if the default multiplier is set lower so that the policyholders would not be protected completely, the policyholders must take the default risk of the insurance company seriously into account when implementing their surrender strategies. In this case, the policyholders may be willing to surrender the contract earlier to avoid losing too much of their initial investment.

In Table 3.3 we present the contract values for different default multipliers θ and different rationality levels represented by $(\underline{\rho}, \bar{\rho})$. In the second column are the contract values in the case when there is no early default mechanism. From the third to the fifth column are the contract values with different levels of regulatory strength which are represented by the different values of the default multiplier θ. For example, $\theta = 0.7$ means that the regulator does not allow the insurance company's asset value to drop below 70% of the minimum guarantee. $\theta = 1.1$ indicates that the regulator is more conservative and requires the company's asset value to lie above 110% of the minimum guarantee. Comparing the contract values in columns 2-4 where the early default regulation is first introduced, then strengthened, the contract value increases gradually for all the types of policyholders. Introducing the early termination rule protects the policyholder from the downside risk and increasing the default multiplier enlarges the protection level. An interesting feature is that the effects of early termination regulation not only depend on the default multiplier θ but also on the rationality level of the policyholder. For example, a policyholder with $(\underline{\rho}, \bar{\rho}) = (0, 0)$ never surrenders her contract. This is equiv-

	no early default	with early default		
		$\theta = 0.7$	$\theta = 0.9$	$\theta = 1.1$
(0, 0)	85.6141	86.8199	90.4937	89.6619
(0, 0.03)	86.0368	87.0668	90.5088	89.6619
(0, 0.3)	88.1531	88.4680	90.6270	89.6619
$(0, \infty)$	92.0546	92.0548	92.0628	89.6619
(0.03, 0.03)	81.8567	82.9119	86.6744	87.9197
(0.03, 0.3)	84.2656	84.5696	86.8343	87.9197
$(0.03, \infty)$	88.5391	88.5392	88.5436	87.9197
(0.3, 0.3)	75.4561	75.7496	78.0482	83.2947
$(0.3, \infty)$	80.7500	80.7500	80.7500	83.2951

Table 3.3: *Contract values for different default multipliers θ and different rationality levels represented by $(\underline{\rho}, \bar{\rho})$.*

alent to a European-type contract which does not allow early termination.[13] We can see that in the case when the policyholder is not competent enough to adopt a rational surrender strategy, an effective early termination regulation may help the policyholder improve his position. However, the benefits from the protection become smaller as the policyholder becomes financially more rational. The default multiplier does not play a significant role when the policyholder is able to exercise the surrender option optimally, i.e. $(\underline{\rho}, \bar{\rho}) = (0, \infty)$.[14] Since the fully rational policyholder can find the optimal surrender strategy anyway, he does not need the protection of the regulator. However, the positive effect of the strengthening of the regulation disappears as an over-regulation rule is carried out. As the default multiplier increases from 0.9 to 1.1, the contract value decreases in some cases, e.g., when $(\underline{\rho}, \bar{\rho}) = (0, \cdot)$, $(\underline{\rho}, \bar{\rho}) = (0.03, \infty)$, among which we can even observe the disadvantage of introducing the early default regulation. For the policyholder with $(\underline{\rho}, \bar{\rho}) = (0, \infty)$ and $(\underline{\rho}, \bar{\rho}) = (0.03, \infty)$, the contract becomes even lower than when there is no early default risk. As we have mentioned in Section 3.2.4 that the policyholder with $\bar{\rho} = \infty$ may surrender the contract at any time when it is optimal to do so, irrespective of exogenous reasons, he is able to protect himself from the downside risk. However, enforcing a termination regulation with a very large default multiplier stops him from obtaining more benefits in the favorable development of the insurance company, which actually lowers the contract's value. Additionally, if we take a look at the column of the contract values for $\underline{\rho} = 0$ and $\theta = 1.1$, the contracts have the same value. Since when

[13]However, the policyholder would be better off if he terminates the contract and collects the surrender guarantee when the value of the contract drops down. Notice that the contract value increases when $\bar{\rho} > 0$.

[14]The contract values are all around 92 for $(\underline{\rho}, \bar{\rho}) = (0, \infty)$ and $\theta = \{0, 0.7, 0.9\}$.

the default multiplier is so high that the benefits that the policyholder obtains at the liquidation of the insurance company are higher than the surrender benefits, surrendering the contract becomes unattractive, which means there would also be no endogenous reasons to surrender the contracts prematurely. Therefore, if the policyholder does not surrender his contract for exogenous reasons ($\underline{\rho} = 0$), the contract value stays at the same level as the European-style contract value 89.6619, no matter how financially rational the policyholder is.

We have discussed in section 3.1 that due to personal reasons, the policyholder surrender his contract even though he knows the value of surrender guarantee offered by the insurance company is lower than the value of his contract. If the policyholder has to liquidate his contract before maturity, which means exogenous surrender does exist, i.e., $\underline{\rho} > 0$, the contract's fair value should be lower than when no exogenous surrender exists, $\underline{\rho} = 0$. The decrease in contract value for a given $\bar{\rho}$ measures the premium that the insurance company should not have charged due to personal non-avoidable liquidity reasons. We call this premium the liquidity premium. In Figure 3.1 we present the liquidity premia under different regulatory rules for different exogenous surrender intensities when $\bar{\rho} = \infty$. We observe the following trends. First, within the same regulatory framework, the liquidity premium becomes larger as the exogenous surrender intensity increases. As

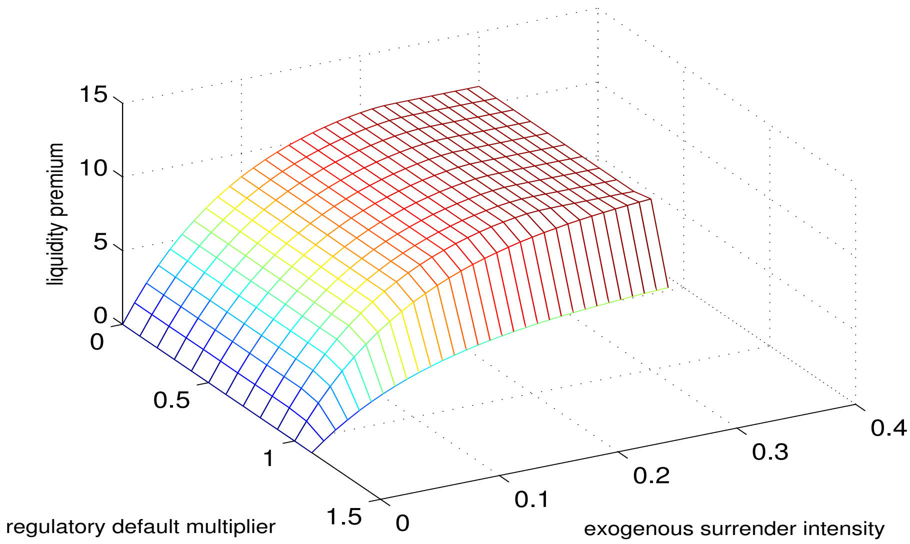


Figure 3.1: *Liquidity premia as a function of the exogenous surrender intensity* $\underline{\rho} \in [0, 0.3]$ *and the default multiplier* $\theta \in [0, 1.1]$ *when* $\bar{\rho} = \infty$.

the non-rational exogenous surrender intensity increases, the insurance company needs to compensate the policyholder more in terms of lowering contract value in order to make the contract more attractive to the policyholder. Second, the liquidity premium increases faster at a lower θ-level while more slowly at a higher θ-level, until the exogenous surrender intensity $\underline{\rho}$ also becomes quite large. This indicates that the value of the liquidity premium is more sensitive to the policyholder's exogenous surrender intensity level $\underline{\rho}$ at a lower θ-level, where the protection from the regulator is low and it is more necessary for the regulator to urge the insurance company to assess the exogenous surrender rates more correctly. On the contrary, as the intervention by the regulator is enhanced, the probability that the insurance company is closed increases. Liquidation may happen before the policyholder exercises the surrender option due to exogenous reasons. Since the policyholder is not penalized at the liquidation, he may receive more than the surrender guarantees he may otherwise obtain from surrendering the contract.

Similar to the above discussion on the impacts of exogenous surrender intensity on the contract value, endogenous surrender intensity also influences the contract value. Since the policyholder has limited information on financial market situation and/or limited knowledge to evaluate the contract on his own, i.e., $\bar{\rho} < \infty$, he may fail to surrender the contract when he should do so. The contract value decreases for given $\underline{\rho}$ as the upper bound surrender intensity $\bar{\rho}$ changes from infinitely large value to a not enough high value. This decrease in the contract value measures the premium that the insurance company should not have charged due to limited information and/or limited evaluation ability, which we name as rationality premium. In Figure 3.2 we plot the rationality premia as a function of the upper bound surrender intensity $\bar{\rho}$ and the default multiplier θ given $\underline{\rho} = 0$. It is natural to observe that the rationality premium decreases with the increase of $\bar{\rho}$ at a given protection level θ settled by the regulator, compare to Proposition 3. It is natural to observe that the rationality premium decreases with the increase of $\bar{\rho}$ at a given protection level θ settled by the regulator, compare to Proposition 3. Furthermore, for given rationality level $(\underline{\rho}, \bar{\rho})$, the rationality premium decreases when the intervention level from the regulator is enhanced until the default multiplier θ reaches 1. Intuitively, after the early default mechanism is introduced and even enhanced, the policyholder has fewer surrender strategies at disposal to carry out which lower his gains from a rational surrender action. Therefore, the insurance company should charge a lower contract price. On the other hand, the intervention by the regulator helps the policyholder close his contract prematurely without bringing penalties to him. Such an intervention protects the policyholder at the moment when he should exercise the surrender option but does not do it for various reasons. Hence, the contract value may also increase. The overall effect depends on the power of these two aspects and is reflected by the final contract value presented e.g. in Table 3.3. We also see that the value of the rationality premium is higher when the early default threshold is lower indicating the importance of assessing the endogenous surrender intensity correctly in an non-strict regulation environment.

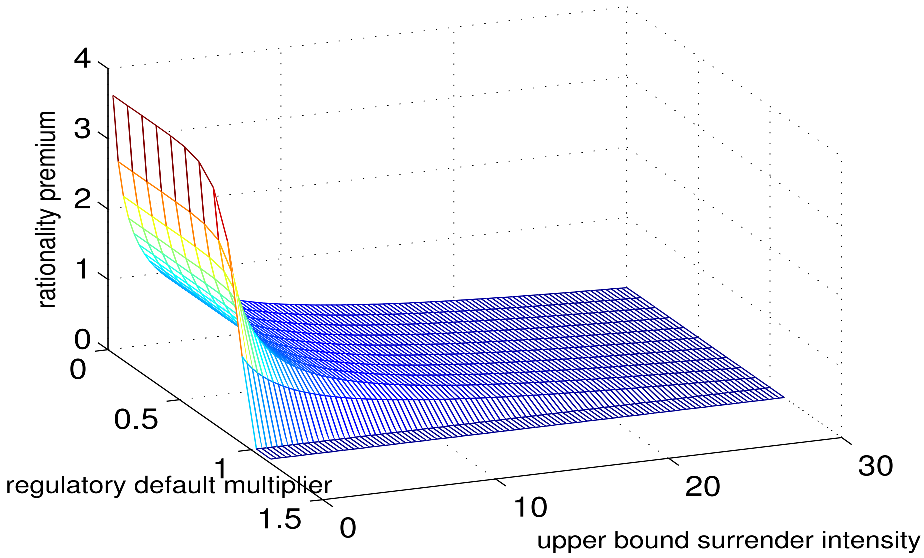


Figure 3.2: *Rationality premia as a function of the upper bound surrender intensity* $\bar{\rho} \in [0.3, 30]$ *and the default multiplier* $\theta \in [0, 1.1]$.

3.4.2 Effects of Insurance Company's Investment Strategies

In this section we focus on the influence of insurance company's investment strategies on contract valuation. Furthermore, we analyze the insurance company's risk-shifting incentives within the two regulatory frameworks. The investment strategy is represented by the volatility σ of the underlying asset A. The higher the volatility σ, the higher the risk that the insurance company has entered into. In Table 3.4 we display the contract values for different values of σ. The default multiplier θ is set to be 0.9.[15]

The influences of the investment strategies are different in different regulatory frameworks. For the no early default case, we observe three tendencies, which depend on the policyholder's rationality. When $(\underline{\rho}, \bar{\rho}) = (0, \infty), (0.03, \infty)$, the policyholder is considered to be very financially rational because the contract will be terminated immediately when

[15]We have also studied the cases with $\theta = 0.7$ and 1.1. However, we have not found any qualitative difference in the effect of the volatility σ and hence do not present all the results here.

	no early default			with early default		
	$\sigma = 0.1$	$\sigma = 0.2$	$\sigma = 0.3$	$\sigma = 0.1$	$\sigma = 0.2$	$\sigma = 0.3$
(0, 0)	85.3380	85.6141	84.7199	86.4204	90.4937	92.4513
(0, 0.03)	85.5737	86.0368	85.2578	86.5152	90.5088	92.4547
(0, 0.3)	86.7156	88.1531	87.9902	87.0566	90.6270	92.4836
$(0, \infty)$	88.3422	92.0546	93.3676	88.3424	92.0628	93.4082
(0.03, 0.03)	82.8209	81.8567	79.7188	83.7463	86.6744	88.1438
(0.03, 0.3)	84.0278	84.2656	83.0419	84.3402	86.8343	88.1934
$(0.03, \infty)$	85.5405	88.5391	89.6150	85.5407	88.5436	89.6439
(0.3, 0.3)	78.2582	75.4561	71.5565	78.4962	78.0482	77.8317
$(0.3, \infty)$	80.7500	80.7500	80.7500	80.7500	80.7500	80.7500

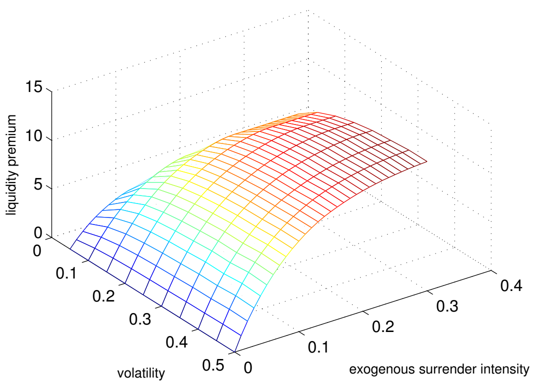
Table 3.4: *Contract values for different investment strategies represented by σ and different rationality levels represented by $(\underline{\rho}, \bar{\rho})$, $\theta = 0.9$*

it is optimal to do so. The increase of the underlying asset risk also implies the potentially higher rate of return, which a rational policyholder can exactly capture. Hence, the contract value increases with the asset risk. For $(\underline{\rho}, \bar{\rho}) = (0, 0)$, $(0, 0.03)$, $(0, 0.3)$ and $(0.03, 0.3)$, which indicates either low exogenous surrender intensity or relatively but not enough high endogenous surrender intensity, we observe firstly the increase and then the decrease in the contract value. When the underlying asset risk increases but still stays at a lower level, the downside risk is still limited and the optimal surrender intensity during the life time of the contract stays anyway at a lower level. However, the chance in the participation of the favorable development of the asset value increases. Hence, overall the contract value increases slightly when σ increases from 0.1 to 0.2. When the asset risk further increases, the downside risk could be so high that it is necessary to check more frequently whether to surrender the contract or not. A lower rational surrender intensity in this case would then lead to a lower contract value. When $(\underline{\rho}, \bar{\rho}) = (0.03, 0.03)$ or $(0.3, 0.3)$, the endogenous surrender intensity is zero and the policyholder surrenders his contract only for exogenous reasons that is not related to the contract value at all. A higher asset risk requires a more rapid and correct response to the changing market situation. When the policyholder is not willing to do so, the contract value for the policyholder will decrease with the increase of the volatility σ.

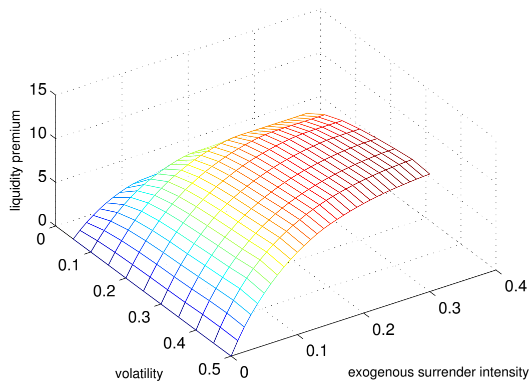
As the early default mechanism is implemented by the regulator, the contract value increases as the volatility σ increases in most cases, except when the probability that the policyholder surrender the contract due to exogenous reasons is relatively large, i.e., $\underline{\rho} = 0.3$. From Table 3.4 we observe that the contract value decreases and stays constant with σ when $(\underline{\rho}, \bar{\rho}) = (0.3, 0.3)$ and $(\underline{\rho}, \bar{\rho}) = (0.3, \infty)$ respectively. In these cases, the policyholder surrenders the contract even when the asset value develops, which deprives him of the chance of participating in the asset appreciation. Since the policyholder is protected by the regulator through the early default barrier, the potential downside risk is limited while the potential participation in the favorable asset performance is still possible. As long as the policyholder is not so urgent in cashing out of his contract, he can benefit more from the regulator's protection as the riskiness of the investment strategy increases and his contract value increases accordingly.

Similar to what we did in section 3.4.1, we present in Figure 3.3 the liquidity premia under different investment strategies (adopted by the insurance company) in both a world without the intervention of the regulator, see Figure 3.3 (a), and a world with the early default mechanism, see Figure 3.3 (b). We see that, for given rationality level $(\underline{\rho}, \bar{\rho})$, liquidity premium increases with the volatility of the underlying asset in both cases. As the investment risk of the insurance company increases with a larger volatility σ, the probability that the policyholder sells back his contract due to exogenous reasons to the insurance company which has been experiencing financial difficulties becomes higher. This indicates that increasing the riskiness of the investment generally does harm to the policyholder who is likely to cash out of his contract when personal difficulties occur. Hence, the contract value decreases more in order to attract policyholders as the insurance company's investment risk increases, which happens in both cases with and without the early default regulation. In addition, the introduction of the early default mechanism lowers the value of the liquidity premium for a given rationality level of $(\underline{\rho}, \bar{\rho})$ and an investment strategy of the insurance company σ. Since the regulatory intervention, as a substitute of an exogenous surrender of the policyholder, helps the policyholder close his contract prematurely without bringing him any penalty, the insurance company accordingly lowers its compensation to the policyholder.

We present the rationality premia depending on investment strategies and endogenous surrender intensities in Figure 3.4. We see that, given the same endogeneous surrender intensity $\bar{\rho}$, the rationality premium increases monotonically with the riskiness of the investment strategy σ when there is no early default mechanism. The rationality premium is much higher when the endogenous surrender intensity $\bar{\rho}$ is low. Unlike a policyholder who can track the performance of the insurance company and act optimally to maximize their benefits, a partially rational policyholder faces the risk of mistakenly holding a contract whose value is lower than the value of the surrender guarantee. The risk increases as the company's asset value becomes more volatile and is reflected by the increasing rationality premium with respect to σ. However, when the early default mechanism exists,

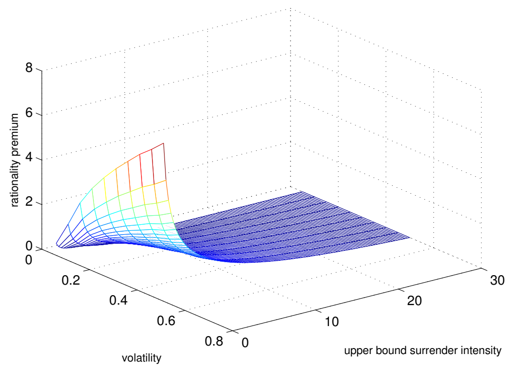


(a) no early default

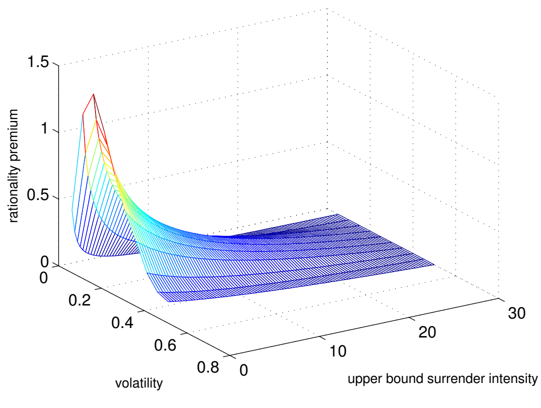


(b) $\theta = 0.9$

Figure 3.3: *Liquidity premia as a function of the exogenous surrender intensity* $\underline{\rho} \in [0, 0.3]$ *and the volatility* $\sigma \in [0.05, 0.5]$ *when* $\bar{\rho} = \infty$.



(a) no early default



(b) $\theta = 0.9$

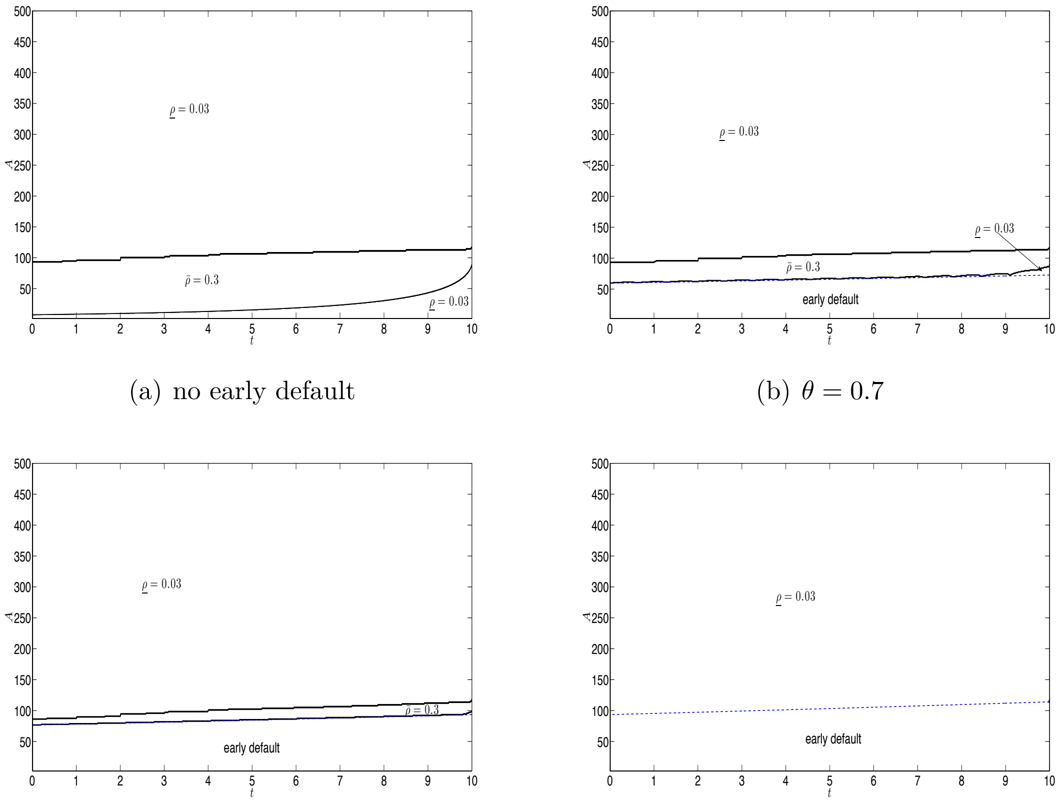
Figure 3.4: *Rationality premia as a function of the upper bound surrender intensity* $\bar{\rho} \in [0.3, 30]$ *and the volatility* $\sigma \in [0.05, 0.5]$ *when* $\underline{\rho} = 0$.

the rationality premium first increases and then decreases with σ. The decreasing effect can be explained by the protection from the early default mechanism, which as a remedy for the policyholder's 'insufficient' surrendering leads the insurance company to lower the rationality premium.

Due to the influence of the policyholder' rationality and the regulatory framework on contract value, as a response, the insurance company may change its investment strategy. We assume that the insurance company performs in the interest of the equity holder. Since the contract value can be regarded as the market value of the insurance company's liabilities when the insurance company is ongoing, the purpose of the company, maximizing the residual value for the equity holder, is thus to minimize the value of the policyholder's policy. From Table 3.4 we can infer which investment strategy the insurance company tends to adopt. If there is no early default rule and the rationality level of the policyholders is very high, the insurance company will prefer to take low risk investment. This gives us two implications. First, if the policyholder is rational enough to surrender their policies, the regulator does not need to interfere into the insurance company in order for the company to avoid too risky investment. Second, looking back into the history, insurance companies have not always taken conservative strategies. Although there are many reasons for them not to do so. For example, the market interest rate was too low in the past and the insurance company has to invest more riskily to achieve higher excess return so as to meet their payment obligations. Another aspect that we can infer from our study is that the insurance company has actually assumed that the policyholders will not always act optimally. Considering this, it is then inappropriate to price surrender option as a pure American-style option as it is often assumed in the literature, since the policy tends to be overpriced under this assumption which is unfair for the policyholders. Hence, if the insurance company chooses parameters for the contract such that the contract's value is exactly equal to the policyholder's payment by assuming a high rationality level and leading us to think that it will adopt an investment strategy with low risk under its assumption, the company actually has the incentive to increase the riskiness of its investment strategy afterwards. This problem will be avoided most likely as the early default regulation is introduced. We can read out from Table 3.4 that the insurance company prefers low risk investments in all cases but one.

3.4.3 Effects of Regulatory Frameworks and Investment Strategies on Surrender Behaviors

To demonstrate the effect of the regulatory framework on the policyholder's surrender behavior, we depict in Figure 3.5 the separating boundaries which illustrate the regions where the policyholder surrenders the contract for exogenous reasons and the regions where the policyholder surrenders the contract for endogenous reasons. When the early default regulation is enforced, part of the surrender region will be replaced by the early default region.

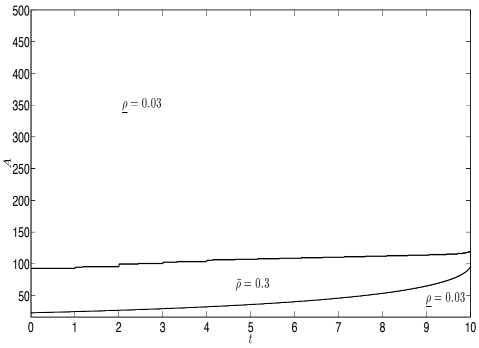


(a) no early default (b) $\theta = 0.7$ (c) $\theta = 0.9$ (d) $\theta = 1.1$

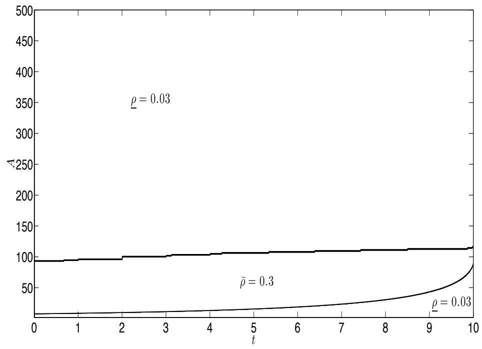
Figure 3.5: *The separating boundary of surrender behaviors for the two regulatory frameworks and for the different default multipliers.*

Based on our consideration about the policyholder's degree of rationality as is illustrated in section 3.4.1 and 3.4.2, we assume $\underline{\rho} = 0.03$ and $\bar{\rho} = 0.3$. We begin with the graph for the case where there is no early termination rule, see Figure 3.5 (a). Here we can observe three regions. When the asset price A is relatively high, the policyholder only surrenders for exogenous reasons, because participation in asset performance is very attractive, which is, according to the contract design, only possible when the policyholder holds the contract until death or until maturity. When asset price A is very low, there would also only be exogenous surrender. This is because in this case participation in asset performance at death or at maturity is hardly possible and early surrender carries penalty on the minimum guarantee, the policyholder would rather to stay in the contract if he does not have other exogenous surrender reasons. The region in the middle of the graph corresponds to the case when the policyholder surrenders his contract for endogenous reasons. In this region the probability that the policyholder surrenders the contract increases mainly due to the reason that the policyholder wants to protect himself from the potential downside risk when it is still not too late to do so. If the regulator intervenes, see Figure 3.5 (b)-(d), we observe that the region with $\bar{\rho} = 0.3$ is more and more replaced by the early default which is triggered by the regulatory rule. When $\theta = 1.1$ the policyholders only surrender the contract for exogenous reasons. This is consistent with the results presented in Figure 3.2, where the rationality premium reduces to 0 when $\theta = 1.1$.

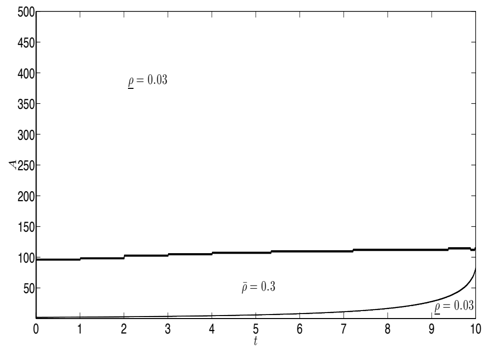
In Figure 3.6 and Figure 3.7 we present the separating boundaries of surrender behaviors as the volatility increases from 0.1 to 0.3 when there is no early default risk and when there is default risk respectively, assuming the rationality level of the representative policyholder is $(\underline{\rho}, \bar{\rho}) = (0.03, 0.3)$. The first interesting observation is that the region $\bar{\rho} = 0.3$ is larger in the case where there is no early default risk than that in the case where there is early default risk for every asset risk σ. Since policyholders feel protected by the early default regulation imposed by the regulator, some closures of policies are carried out by the regulator instead of the policyholders themselves. more surrenders are triggered due to exogenous reasons instead of financially related reasons. The second interesting observation is that as volatility σ increases the region $\bar{\rho} = 0.3$ expands when there is no early default regulation, while it shrinks when regulator specifies an early termination barrier during the life time of the contract. It indicates that policyholders are becoming more sensitive and exercise more surrender options as the financial world becomes more volatile when no early termination mechanism is established by the regulator. However, when the early termination mechanism is introduced, due to limited rationality, more policyholders may decide to rely on the protection from the regulator as the financial world becomes more volatile so that the region $\bar{\rho} = 0.3$ shrinks as σ increases, and at the same time, the regulator's early intervention becomes more intensive.



(a) $\sigma = 0.1$

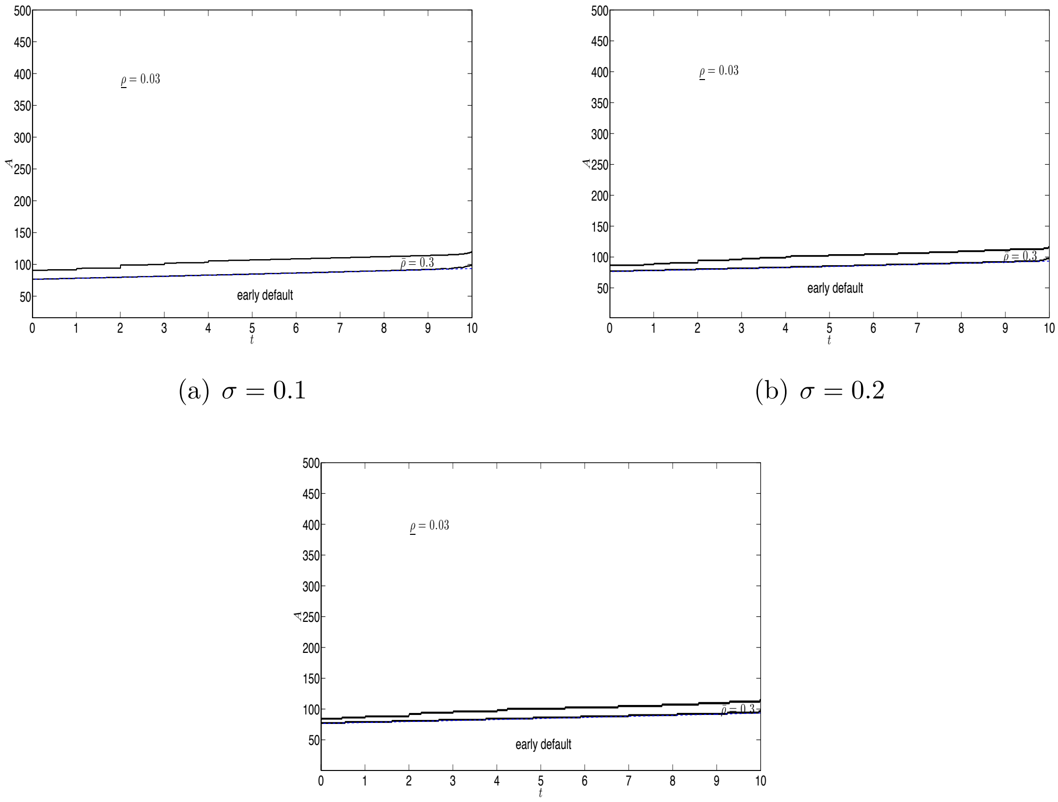


(b) $\sigma = 0.2$



(c) $\sigma = 0.3$

Figure 3.6: *The separating boundary of surrender behaviors when there is no early default risk*



(a) $\sigma = 0.1$

(b) $\sigma = 0.2$

(c) $\sigma = 0.3$

Figure 3.7: *The separating boundary of surrender behaviors when there is early default risk*

3.5 Conclusion

In this chapter we have studied the impacts of early default risk and surrender risk on participating life insurance policies. Early default is triggered once its asset value touches a prespecified threshold. Surrender risk is represented by a surrender intensity which is bounded from below and from above and accounts for the limited rationality of a representative policyholder in making surrender decisions. The lower bound refers to the policyholder's surrender intensity for exogenous reasons while the upper bound is achieved if the surrender value is higher than the active contract value. Since early default risk affects the contract's payment structure, it influences the policyholder's surrender behavior. We have derived the pricing PDE equation which characterizes the contract value and solved it numerically with the finite difference method. Based on the numerical examples, we have analyzed the influence of early default risk, given that the insurance company's investment strategy is known, on the policyholder's surrender behavior and consequently on contract valuation. Furthermore, we have analyzed the influence of insurance company's investment strategy on the policyholder's surrender behaviors as well as on contract values given the regulatory rule prescribed by the regulator.

The analysis for a representative policyholder can be transferred to a large pool of policyholders. Many implications can be drawn from our analysis. First, if policyholders are able to surrender their participating policies optimally, it is not necessary for the regulator to set a regulatory rule to monitor the insurance company. The insurance company is actually monitored by the policyholders themselves. However, since policyholders are mostly not completely rational to make financially optimal decisions, an early default regulation protects these policyholders as long as the regulation is not too strict. Overregulation is disadvantageous to most policyholders. Second, enhancing the regulation exerts a negative effect on the rationality premium for those policyholders who would have more freedom to construct their own optimal stopping strategy if there were fewer regulatory constraints. Since the effects of the regulation are different for the policyholders with different surrender reasons and the contract value that we are looking for is the average contract value for all the policyholders, which is denoted as the contract value for a representative policyholder in our paper, the overall effect of enhancing the regulation on the contract value may be positive. Third, without the introduction of the regulatory framework, we are not clear about the effect of the investment strategy on contract value. We find that the equity holder prefers to adopt a less risky investment strategy if the policyholders are able to surrender the contracts optimally. Since the equity holder knows that policyholders are most of the time not financially rational enough and there are always exogenous reasons for them to surrender the contracts prematurely, he actually tends to invest more riskily. However, when the early default barrier is settled, an increase in the riskiness of the investment strategy will generally have a positive effect on the contract value from the perspective of the policyholders. The equity holder will then have the incentive to reduce the riskiness of their investment, which is independent of the rationality level of the policyholders. This result is consistent with the goal of the regulator.

3.6 Appendix to Chapter 3

3.6.1 Proof of Corollary 1

We follow the proof of Theorem 2.1 in Freidlin (1985) for a similar Dirichlet problem.

Define

$$\begin{aligned} g(t, A_t) &:= \mu(t)\Psi(t, A_t) + \gamma(t, A_t)S(t, A_t), \\ c(t, A_t) &:= r(t) + \mu(t) + \gamma(t, A_t), \\ Y_t^A &:= -\int_0^t c(z, A_z)\mathrm{d}z, \\ U_t^A &:= v(t, A_t)e^{Y_t^A}. \end{aligned}$$

According to the Ito's Lemma, we obtain, for all $t < m < \tau_b \wedge T$ where $t < \tau_d \wedge \tau_s \wedge \tau_b \wedge T$, the stochastic differential equation of U_m^A as

$$\begin{aligned} \mathrm{d}U_m^A &= \frac{\partial U_m^A}{\partial m}\mathrm{d}m + \frac{\partial U_m^A}{\partial A_m}\mathrm{d}A_m + \frac{1}{2}\frac{\partial^2 U_m^A}{\partial A_m^2}\mathrm{d}A_m^2 \\ &= \left[\frac{\partial v}{\partial m}(m, A_m)e^{Y_m^A} - v(m, A_m)e^{Y_m^A}c(m, A_m)\right]\mathrm{d}m \\ &\quad + \frac{\partial v}{\partial A_m}(m, A_m)e^{Y_m^A}\left(r(m)A_m\mathrm{d}m + \sigma(m, A_m)A_m\mathrm{d}W_m\right) \\ &\quad + \frac{1}{2}\frac{\partial^2 v}{\partial A_m^2}(m, A_m)e^{Y_m^A}\sigma^2(m, A_m)A_m^2\mathrm{d}m \\ &= e^{Y_m^A}(\mathcal{L}v(m, A_m) - c(m, A_m)v(m, A_m))\mathrm{d}m + e^{Y_m^A}\frac{\partial v}{\partial A_m}(m, A_m)\sigma(m, A_m)A_m\mathrm{d}W_m \\ &= e^{Y_m^A}(-g(m, A_m))\mathrm{d}m + e^{Y_m^A}\frac{\partial v}{\partial A_m}(m, A_m)\sigma(m, A_m)A_m\mathrm{d}W_m \end{aligned}$$

The last equation follows from equation (3.15) in Proposition 2.

Integrate both sides of the above equation from t to $\tau_b \wedge T$ and take the expectation on both sides. Under the assumption that

$$\mathbb{E}_{\mathbb{Q}}\left[\int_0^{\tau_b \wedge T} ||e^{Y_m^A}\frac{\partial v}{\partial A_m}\sigma(m, A_m)A_m||^2 \mathrm{d}m\right] < \infty$$

which ensures

$$\mathbb{E}_{\mathbb{Q}}\left[\int_t^{\tau_b \wedge T} e^{Y_m^A}\frac{\partial v}{\partial A_m}(m, A_m)\sigma(m, A_m)A_m \mathrm{d}W_m \Big| \mathcal{G}_t\right] = 0,$$

we obtain

$$\mathbb{E}_{\mathbb{Q}}[U^A_{\tau_b \wedge T}|\mathcal{G}_t] \quad = \quad U_t^A + \mathbb{E}_{\mathbb{Q}}\left[\int_t^{\tau_b \wedge T} e^{Y_m^A} g(m, A_m)\mathrm{d}m \Big| \mathcal{G}_t\right],$$

and thus

$$v(t, A_t) \quad = \quad \mathbb{E}_{\mathbb{Q}}\left[\int_t^{\tau_b \wedge T} e^{Y_m^A - Y_t^A} g(m, A_m)\mathrm{d}m + e^{Y^A_{\tau_b \wedge T} - Y_t^A} v(\tau_b \wedge T, A_{\tau_b \wedge T}) \Big| \mathcal{G}_t\right].$$

Since

$$\begin{aligned} e^{Y^A_{\tau_b \wedge T} - Y_t^A} v(\tau_b \wedge T, A_{\tau_b \wedge T}) &= 1_{\{\tau_b < T\}} e^{Y^A_{\tau_b}} v(\tau_b, A_{\tau_b}) + 1_{\{\tau_b \geq T\}} e^{Y_T^A} v(T, A_T) \\ &= 1_{\{\tau_b < T\}} e^{Y^A_{\tau_b}} \Upsilon(\tau_b, A_{\tau_b}) + 1_{\{\tau_b \geq T\}} e^{Y_T^A} \Phi(A_T), \end{aligned}$$

where the last equation results from the boundary conditions in Proposition 2, we obtain, by substituting $g(\cdot, \cdot)$, $c(\cdot, \cdot)$, and Y with their original forms,

$$\begin{aligned} V_t \quad = \quad \mathbb{E}_{\mathbb{Q}}\Bigg[&\int_t^{\tau_b \wedge T} e^{-\int_t^m (r(u)+\mu(u)+\gamma(u,A_u))\mathrm{d}u} (\mu(m)\Psi(m, A_m) + \gamma(m, A_m)S(m, A_m))\mathrm{d}m \\ &+ 1_{\{\tau_b \geq T\}} \Phi(A_T) e^{-\int_t^T (r(u)+\mu(u)+\gamma(u,A_u))\mathrm{d}u} + 1_{\{\tau_b < T\}} \Upsilon(\tau_b, A_{\tau_b}) e^{-\int_t^{\tau_b} (r(u)+\mu(u)+\gamma(u,A_u))\mathrm{d}u} \Big| \mathcal{G}_t \Bigg] \end{aligned}$$

Corollary 1 is therefore proved.

3.6.2 Proof of Proposition 3

The pre-death/surrender value function v is the solution of the PDE (3.15) with terminal condition $v(T, A_T) = \Phi(A_T)$ and boundary condition $v(t, A_t) = \Upsilon(t, A_t)$ for $A_t \leq B_t$, and bounds $\underline{\rho}$ and $\bar{\rho}$. The pre-death/surrender value function w is the solution of the same

PDE (3.15) with identical terminal condition $w(T, A_T) = \Phi(A_T)$ and boundary condition $w(t, A_t) = \Upsilon(t, A_t)$ for $A_t \leq B_t$ but different bounds $\underline{\zeta}$ and $\bar{\zeta}$. Assume that $\underline{\zeta} \leq \underline{\rho}$ and $\bar{\rho} \leq \bar{\zeta}$. Now define $z = w - v$. It follows directly that $z(T, A_T) = w(T, A_T) - v(T, A_T) = \Phi(A_T) - \Phi(A_T) = 0$ and $Z(t, A_t) = \Upsilon(t, A_t) - \Upsilon(t, A_t) = 0$ for $A_t \leq B_t$. To obtain the dynamics of z take the difference of the PDEs describing w and v, i.e.:

$$\begin{aligned} 0 &= \mathcal{L}w(t, A_t) + \mu(t)\Psi(t, A_t) + \gamma^w(t, A_t)S(t, A_t) - (r(t) + \mu(t) + \gamma^w(t, A_t))\, w(t, A_t) \\ &\quad - (\mathcal{L}v(t, A_t) + \mu(t)\Psi(t, A_t) + \gamma^v(t, A_t)S(t, A_t) - (r(t) + \mu(t) + \gamma^v(t, A_t))\, v(t, A_t)) \\ &= \mathcal{L}z(t, A_t) + (\gamma^w(t, A_t) - \gamma^v(t, A_t))\, (S(t, A_t) - w(t, A_t)) - (r(t) + \mu(t) + \gamma^v(t, A_t))\, z(t, A_t)\,, \end{aligned}$$

where γ^v and γ^w, respectively, are given by (3.7) using the appropriate bounds. Similar to the proof of Corollary 1, we obtain the stochastic representation of z as follows

$$\begin{aligned} z(t, A) &= \mathbb{E}_{\mathbb{Q}}^{t,A}\Bigg[\int_t^{\tau_b \wedge T} e^{-\int_t^m (r(u)+\mu(u)+\gamma^v(u,A_u))\,\mathrm{d}u}(\gamma^w(m, A_m) - \gamma^v(m, A_m))\,(S(m, A_m) \\ &\quad -w(m, A_m))\mathrm{d}m\Bigg|\mathcal{G}_t\Bigg], \end{aligned}$$

where $\mathbb{E}_{\mathbb{Q}}^{t,A}$ denotes the expectation conditioned on $A_t = A$. From the definition of γ^w in (3.7) and the assumption $\bar{\zeta} \geq \bar{\rho}$ we see that if $(S - w) \geq 0$ we have $\gamma^w = \bar{\zeta} \geq \bar{\rho} \geq \gamma^v$ and thus $(\gamma^w - \gamma^v) \geq 0$. On the other hand, if $(S - w) < 0$ then $\gamma^w = \underline{\zeta}$. By assumption we have $\underline{\zeta} \leq \underline{\rho}$ and thus $\gamma^w \leq \underline{\rho} \leq \gamma^v$, or, $(\gamma^w - \gamma^v) \leq 0$. Thus, we see that the integrand in the above equation is nonnegative and therefore $z \geq 0$. Since $z = w - v$ we obtain $w \geq v$.

4 Linking Surrender Risk to Mortality Risk: Does the Systemic Health Shock Matter?

4.1 Introduction

The evidence of policyholders' health status being an influencing factor of the surrender of life insurance contracts has been recorded in the current literature. For example, Doherty and Singer (2003a, 2003b) argue that the surrender rate of impaired policyholders is higher than that of normal policyholders on a competitive settlement market. Besides, Sutherland and Drivanos (1999) summarize findings on the financial hardship that families with a seriously ill member experienced in the United States[1] and emphasize that the ill patient sells a life insurance policy to acquire needed cash for dealing with such financial hardship. Giacalone (2001) also points out that the demand of selling a life insurance policy by ill insureds, which is an alternative to surrendering the policy, is a major factor on viatical settlement markets. In addition, many empirical papers on lapse determinants have found that policyholders' characteristics, for example age and gender, are important drivers for lapse rates, see Cerchiara et al. (2008), Milhaud et al. (2011), Eling and Kiesenbauer (2013), Fier and Liebenberg (2013) and Kim (2013), just to name a few. Since we know in actuarial literature, policyholders' characteristics, in particular policyholders' age, have an influence on their survival rates, see for example laws of mortatlity proposed by Gompertz 1825, Makeham 1860, and Oppermann 1870, the health stage of the policyholders do influence the life insurance lapse.

Even though both the mortality risk and surrender risk have been analyzed thoroughly in the current life insurance literature, the relationship between the two types of risk is rarely modelled. A few literatures that we are aware of and model the relationship between the mortality risk and the surrender risk are Jones (1998) and Gatzert et al. (2009).

[1] The original findings on the financial hardship were publish in 1994 on the Journal of the American Medical Association.

The two papers recognize mortality heterogeneity of poliyholders and they adopt a frailty factor as modelled in Vaupel et al. (1979) to measure a risk level of each policyholder in the heterogeneous insured group. By assuming each policyholder's surrender is a function of the risk level, Jones (1998) analyzes the effect on the cohort force of mortality under different assumptions on the relationship between the lapse rate and the mortality risk level. While, Gatzert et al. (2009) focus on asymmetric surrender decisions triggered by the secondary market, where only contracts of impaired policyholders are sold and kept on the market, and analyze the effect of the adverse surrender behavior on the primary insurer's surrender profits. The two papers do consider the relationship between the health status and surrender of policyholders, but rather in a deterministic way. First, the frailty factor is drawn from a given distribution function to specify different health states in a portfolio. However, any changes in the health states of the policyholders in the portfolio are not tracked in the two papers and consequently any changes in the policyholders' surrender decisions due to their health state changes are not taken into account. Second, in the paper of Gatzert et al. (2009), adverse surrender behavior of impaired policyholders implicated by the secondary market is exogenously given, rather than being endogenously linked to the triggering factor.

In this paper, we take into account a systemic health shock which harms the public health and lasts for a relatively long-time period. Hence, the systemic health shock applies to all of policyholders in the pool and damages on the policyholders' health last for a relatively long-time period.[2] One example of the systemic health shock is the smog, which is a severe air pollution event and currently prevalent in Asia.[3] Starting from the beginning of 2013, northern China was hit by multiple severe smog events, which damage rural residents' health and lead to higher mortality, see Zhou et al. (2015). Similar health effects and the positive relationship between the smog and the mortality during smog episodes in industrialized countries have also been recorded in some empirical studies, see e.g., Wichmann et al. (1989) in Germany, and Anderson et al. (1995) in United Kingdom. As the policyholders' normal health state switches to the impaired state due

[2]As long as the pool is large enough, a health shock which applies only to an individual policyholder and consequently influences the policyholder's surrender, similar as the stochastic feature of the policyholder's mortality intensity, does not have much influence on the contract value, see Li and Szimayer (2011). Hence, we focus on the health shock which applies to all policyholders in the pool in this paper.

[3]The smog also happened intensively worldwide in the 19th and 20th centuries, e.g., Belgium, United Kingdom, United States, Germany, Mexico, and Chile. In addition, other examples of the systemic health shock in history include slums in European cities, ca. 1800-1900, potato famine in Ireland beginning in 1845, black population in South Africa during Apartheid ca. 1955-1993, etc.

to the systemic health shock, we expect changes in the policyholders' surrender behavior, which consequently have effects on the contract valuation. However, the connection between the mortality and the surrender triggered by the systemic health shock, and the presumed reactions of the policyholders on the surrender have not attracted attention in the current life insurance literature.

We use a jump process to model the systemic health shock and construct a continuous Markov chain to track a representative policyholder's health state change. We then adopt the intensity-based approach of modelling partially rational surrender of the policyholder in Li and Szimayer (2014), and propose two state-dependent surrender change scenarios. In the first scenario, as the policyholder's health gets damaged because of the systemic health shock, a large demand of cash spent on necessary medical treatments and improving his living conditions threatens the policyholder's financial liquidity.[4] As a consequence, the policyholder is hurry to access surrender values in order to deal with the financial liquidity problem. It corresponds to the Emergency Fund Hypothesis (EFH), which has been investigated and confirmed in many empirical studies on life insurance lapses, see Dar and Dodds (1989), Outreville (1990), Liebenberg et al. (2012), Fier and Liebenberg (2013), and Kim (2013), just to name a few. Hence, in this scenario the policyholder surrenders the policy more likely for exogenous reasons. In the second scenario, suppose the policyholder becomes more sensitive in handling his assets after experiencing the health damage. Since the insurance policy is also a part of the policyholder's assets, see Sutherland and Drivanos (1999), we consider that the policyholder becomes more careful in dealing with his insurance policy. Higher concentration on managing his insurance contract requires the policyholder to collect more financial information and improve evaluating skills, which leads to a higher surrender rationality. Based on the two presumed scenarios, we analyze effects of the systemic health shock on the contract valuation, and further discuss the influence of an early default regulatory rule on the contract value in the two scenarios.

This paper is organized as follows. In Section 4.2 we present the insurance company's financial structure and introduce the participating life insurance policy. Both mortality risk and surrender risk are modelled by a jump process and linked through a systemic health shock, which is also modelled by a jump process. Based on the health shock jump process, we construct a continuous two-state Makov chain and accordingly introduce

[4]We have talked about the financial hardship of ill patients in the beginning of the introduction, see Sutherland and Drivanos (1999). Here, we use a more general term, financial liquidity problem, to include policyholders who are not terminally ill but still need a large amount of cash to pay for necessary medical treatments.

the two surrender change scenarios as the policyholder's health state becomes impaired. Besides, we also specify the early default regulatory rule imposed by a regulator. In Section 4.3 we derive a coupled partial differential equation (PDE) system for the price of the policy. In Section 4.4 we study effects of both the systemic health shock and regulatory frameworks on the contract valuation in the two surrender change scenarios. We conclude in Section 4.5.

4.2 The Model Framework

4.2.1 Insurance company and Participating Life Insurance Policy

We fix a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{Q})$, where the filtration $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ reflects the flow of information available on both the financial market and insurance market, and $\mathbb{Q}$ is an martingale measure in the absence of arbitrage. Following Cheng and Li (2015), we consider an insurance company founded at time $t_0 = 0$ by two agents, a representative policyholder and an equity holder. The representative policyholder pays a single premium for a participating life insurance contract and acquires an initial liability $L_0 \equiv \alpha A_0$ with $a \in (0, 1)$, where A_0 denotes the company's initial assets' value. The equity holder invests the remaining $E_0 \equiv (1-\alpha)A_0$ in the insurance company. The insurance company invests its initial assets A_0 in the traded risky and risk-free assets on the financial market. Under the martingale measure $\mathbb{Q}$, we assume the company's asset price satisfies the following stochastic process

$$dA_t = r(t)A_t\mathrm{d}t + \sigma(t, A_t)A_t dW_t^{\mathbb{Q}}, \ \forall t \in [0, T] \text{ and } A_0 > 0, \tag{4.1}$$

where the short rate r is assumed to be a deterministic function of time, the function $\sigma > 0$ is the volatility of the asset price process, and T refers to the maturity of the participating contract. $\mathbb{W}^{\mathbb{Q}}$ is a standard 1-dimensional Brownian motion under $\mathbb{Q}$, which generates the filtration $\mathbb{F}^W = (\mathcal{F}_t)_{t\geq 0}$. For simplicity, we assume the insurance company announces a closure as the contract matures at time T and distributes its liquidated assets to the stakeholders. The policyholder is assumed to have a priority claim to the insurance company's assets and the equity holder is subject to limited liability. Therefore, as the participating life insurance contract continues and expires at maturity T, the policyholder receives survival benefits, which have the following structure[5]

$$\Phi(A_T) = L_T^{r_g} + \delta[\alpha A_T - L_T^{r_g}]^+ - [L_T^{r_g} - A_T]^+. \tag{4.2}$$

The first term following the equality refers to the minimum guaranteed benefits promised to the policyholder, which are calculated by compounding the initial liability with a minimum guaranteed interest rate r_g, i.e., $L_T^{r_g} = L_0 e^{r_g T}$. The second term describes the share of the company's profits written in the contract, i.e., the bonus payment $\delta(\alpha A_T - L_T^{r_g})$ conditional on $\alpha A_T > L_T^{r_g}$, where δ is called a participation rate. Given that the equity holder has limit liability, when the company has less-than-minimum-guarantee assets at maturity T, the policyholder only receives what is left in the company A_T, which is captured by the last term in the above equation.

4.2.2 Morality Risk and Systemic Health Shock

We model the policyholder's death by a jump process $H = (H_t)_{t\geq 0}$ with $H_t = \mathbb{1}_{\{\tau_d \leq t\}}$ for $0 \leq t \leq T$, where τ_d denotes the death time of the policyholder. We denote the smallest filtration making τ_d as a stopping time by $\mathbb{H} = (\mathcal{H}_t)_{t\geq 0}$ with $\mathcal{H}_t = \sigma(H_s : 0 \leq s \leq t)$. If the policyholder dies before the contract ends naturally, the insurance policy pays out death benefits. We use $\Psi(\tau_d, A_{\tau_d})$ to denote the death benefits paid out at $\tau_d < T$, which are assumed to have the same payment structure as the survival benefits

$$\Psi(\tau_d, A_{\tau_d}) = L_{\tau_d}^{r_d} + \delta_d[\alpha A_{\tau_d} - \L_{\tau}^{r_d}]^+ - [L_{\tau}^{r_d} - A_{\tau_d}]^+,\ L_t^{r_d} = L_0 e^{r_d \tau_d}, \tag{4.3}$$

where r_d and δ_d denote a minimum guaranteed interest rate and a participation rate at death, respectively. Moreover, the jump process has intensity μ_t, named as mortality intensity. As we have introduced in Section 4.1, that the public may encounter a systemic health shock during the insuring period, e.g., the smog, which damages the representative policyholder's health. We model the systemic health shock by a jump process $L = (L_t)_{t\geq 0}$ with $L_t = \mathbb{1}_{\{\tau_l \leq t\}}$ for $t \geq 0$, where τ_l denotes the arrival of the health shock. We assume that as the systemic health shock occurs, the policyholder's health state changes from the 'normal' state to the 'impaired' state. Since the health damage preserves for a relative long period, for example, the smog in the European cities lasted for decades and people lived in highly polluted environment for a very long period, for simplicity we assume that the representative policyholder stays in the impaired state after the health shock shortest till the maturity of the contract. We assume a state-dependent mortality intensity which

[5]We adopt the contract payment forms in Cheng and Li (2015).

has the following form[6]

$$\mu_t = m_0(t) + m_1 \mathbb{1}_{\{\tau_l \le t\}}, \quad m_1 > 0. \tag{4.4}$$

Here $m_0(t)$ is a deterministic intensity function of time t, which represents a realistic prediction of the mortality intensity as long as the policyholder stays in the normal health status. As the health shock impairs the policyholder's health status, the mortality intensity increases by $m_1 > 0$, i.e., a systematic deviation of mortality from the mortality prediction $m_0(t)$, resulting in a deterministic intensity function $m_0(t) + m_1$ for the policyholder in the impaired state.

Now based on the jump process $L_t = \mathbb{1}_{\{\tau_l \le t\}}$, we construct a continuous-time two-state Markov chain

$$X = (X_t)_{t \ge 0} \tag{4.5}$$

to describe the change in policyholder's health state, see chapter 2 of Norris (1997). We describe the state space of the Markov chain X by the set of two unit vectors, denoted by $\mathcal{E} = \{e_1, e_2\}$, where $e_1 = (1\ 0)'$ and $e_2 = (0\ 1)'$ refer to the normal health state and the impaired health state, respectively, see Buffington and Elliot (2002). The two-state Markov chain generates the filtration $\mathbb{F}^X = (\mathcal{F}_t^X)_{t \ge 0}$, which is independent of the deterministic predicted mortality component $m_0(t)$. Under the assumption that the jump process L_t has intensity κ and the policyholder's health stays in the impaired state after the systemic health shock, we have the following generator matrix

$$C = \begin{pmatrix} -\kappa & \kappa \\ 0 & 0 \end{pmatrix}. \tag{4.6}$$

κ describes the intensity that the policyholder's health state jumps from the normal state to the impaired state. Accordingly, we have the transition probability, $P_{ij}(t) = \mathbb{P}(X_t = e_j | X_0 = e_i), i, j \in \{1, 2\}, t \ge 0$, as the solution of the following forward equation

$$\frac{d\mathbf{P}}{dt}(t) = \mathbf{P}(t)C, \quad \mathbf{P}(0) = I_2, \tag{4.7}$$

where $\mathbf{P}(t) = (P_{ij}(t))_{i,j=1,2}$ is the transition matrix of X and I_2 is the identity matrix

[6] In this paper, we focus on the linkage of mortality risk and surrender risk triggered by the systemic health shock. Hence, we model the mortality intensity function form (4.4) with m_1 exclusively capturing the systematic deviation of mortality from expected mortality rates due to the health shock. We refer readers who are interested in dynamics of human mortality taking into account random fluctuations and systematic deviations from the chosen mortality basis to Biffis (2005).

of two dimensions.[7] The generator matrix C yields the decomposition for the stochastic differential of X, $dX_t = C'X_t dt + dM_t$, where M_t is a two-dimensional martingale on the filtration $\mathbb{F}^X$, see Elliot et al. (1994). Since the mortality jump process of H is conditionally independent of the filtration $\mathbb{F}^X$, M_t is also a martingale on the enlarged filtration $\mathbb{Z} = \mathbb{H} \vee \mathbb{F}^X$ under $\mathbb{Q}$,[8] see chapter 6 of Bielecki and Rutkowski (2004). Moreover, given that the mortality risk is not linked to the financial market, we can work on the enlarged filtration $\mathbb{G} = \mathbb{F}^W \vee \mathbb{Z}$, where $W^{\mathbb{Q}}$ is a $(\mathbb{Q}, \mathbb{G})$ standard Brownian, M_t is a $(\mathbb{Q}, \mathbb{G})$ martingale, and μ_t is a $(\mathbb{Q}, \mathbb{G})$ intensity of random death time τ_d.

4.2.3 Surrender Risk and State-dependent Surrender Scenarios

A surrender option is written in the contract, by exercising which the policyholder can terminate his contract before the maturity T. In this section, we describe surrender behavior of the policyholder. By taking into account that the policyholder's surrender decision-making behavior is subject to his health status, we model the policyholder's surrender by the first arrival of a Markov-modulated Poisson process. It generates the filtration $\mathbb{J} = (\mathcal{J}_t)_{t \geq 0}$, with $\mathcal{J}_t = \sigma(J_s = \mathbb{1}_{\{\tau_s \leq s\}} : 0 \leq s \leq t)$, where τ_s denotes the random surrender time. If the contract is terminated at $\tau_s < T$, surrender benefits of the following form

$$S(\tau_s, A_{\tau_s}) = L_{\tau_s}^{r_s} - [L_{\tau_s}^{r_s} - A_{\tau_s}]^+, \; L_{\tau_s}^{r_s} = (1 - \beta_{\tau_s}) L_0 e^{r_s \tau_s} \tag{4.8}$$

are paid to the policyholder. r_s is used to calculate the minimum guaranteed benefits in the event of surrender. By terminating the contract before the contract matures, the policyholder does not participate in the company's profits, but only obtains the minimum surrender guarantee $L_{\tau_s}^{r_s}$ as the company has enough assets, i.e., $A_{\tau_s} > L_{\tau_s}^{r_s}$, otherwise, the policyholder receives only what is left in the company A_{τ_s}. Additionally, a penalty parameter β_t, which is assumed as a decreasing function of time, is applied for calculating the minimum surrender guarantee. The earlier the contract is terminated by the policyholder, the higher punishment that the policyholder bears for doing so.

We denote the hazard rate of the surrender time τ_s by γ_t, called surrender intensity, which depends on the health state of the policyholder X_t, the value of surrender benefits

[7] For explicitly solving the differential equation (4.7), see chapter 2 of Norris (1997).

[8] Even though choosing the martingale measure $\mathbb{Q}$ is not discussed in this paper, it is worth emphasizing that the diversification argument on risk-neutral measure does not apply when systematic deviations of mortality from expected mortality rates are considered. A risk premium needs to be priced by the market, see more discussions on calibration in Biffis (2005).

$S(t, A_t)$ and the contract's continuation value, denoted by $v(t, A_t, X_t)$.[9] In this paper, we consider that the policyholder is partially rational and the surrender intensity is bounded. More in detail, due to exogenous reasons, for example, liquidity demand, the policyholder decides to surrender his contract for cash even though it is not financially optimal to do so, i.e., the value of surrender benefits is lower than the contract's continuation value. Therefore, in this case a lower bound value is assigned to the surrender intensity, and depending on the policyholder's health state, we write the lower bound of the surrender intensity as $\underline{\rho}X_t$ with $\underline{\rho} = (\underline{\rho}_1 \quad \underline{\rho}_2)$ for $X_t \in \{e_1, e_2\}$. $\underline{\rho}_1$ and $\underline{\rho}_2$ are the lower bound surrender intensity value in the normal state and in the impaired state respectively. On the contrary, when the value of surrender benefits is higher than the contract continuation value, it happens that since the policyholder has restricted financial information and/or limited financial competency of evaluating the contract, he does not surrender his contract. Hence, we assign a upper bound value to the surrender intensity. Same as the case with the lower bound surrender value, the upper bound intensity value is also state-dependent, which can be written as $\bar{\rho}X_t$ with $\bar{\rho} = (\bar{\rho}_1 \quad \bar{\rho}_2)$, $\bar{\rho}_1 > \underline{\rho}_1$ and $\bar{\rho}_2 > \underline{\rho}_2$ for $X_t \in \{e_1, e_2\}$. $\bar{\rho}_1$ and $\bar{\rho}_2$ are the upper bound surrender intensity value in the normal health state and in the impaired health state respectively. The difference between the upper bound intensity value and the lower bound intensity value, i.e., $\rho_i^E = \bar{\rho}_i - \underline{\rho}_i$ for $i \in \{1, 2\}$, also called endogenous surrender intensity, describes how sensitive the policyholder is about the financial information and how much efforts the policyholder puts into analyzing financial information and evaluating his insurance contract. Since the surrender process J is also conditionally independent of the filtration generated by the Markov-chain $\mathbb{F}^X$, we can now work on the enlarged filtration $\mathbb{F} = \mathbb{J} \vee \mathbb{G}$, where $W^{\mathbb{Q}}$ is a $(\mathbb{Q}, \mathbb{F})$ standard Brownian, C is the generator of Markov-chain X with M_t as a $(\mathbb{Q}, \mathbb{F})$ martingale, and μ_t and γ_t are $(\mathbb{Q}, \mathbb{F})$ the intensity of the jump process H and the jump process J, respectively.

In the present paper, we consider the following two presumed state-dependent surrender scenarios. First, we have argued in the introduction that the policyholder's health gets damaged after the systemic health shock, which worsens his financial position as the demand of cash raises up consequently. We would expect that the systemic health shock shortens the expected duration of the arrival of policyholder's exogenous surrender. As we have pointed out in the introduction, this consequence is supported under the Emergency Fund Hypothesis (EFH), which describes that policyholders surrender their contracts to

[9] The continuation value of the contract $v(t, A_t, X_t)$ is the pre-death-and-surrender contract value, which is further explained in Section 4.3.

cope with their personal financial distress.[10] Hence, in Scenario 1, as the policyholder's health status switches to the impaired state after the health shock, the policyholder becomes more likely to surrender his contract for exogenous reasons, which implies an increase in the exogenous surrender intensity, i.e., $\underline{\rho}_2 > \underline{\rho}_1$. At the same time, as the policyholder hurries to his contract's surrender value, he overlooks the information on the financial market and reduces the efforts in analyzing the contract continuation value. It results in a lower value of the endogenous surrender intensity after the health shock, i.e., the difference between lower bound intensity value and the upper bound intensity value becomes smaller, $\rho_2^E < \rho_1^E$. For simplicity, we assume the upper bound surrender intensities in both health states have the same value, i.e., $\bar{\rho}_1 = \bar{\rho}_2$ with $\bar{\rho}_1 = \underline{\rho}_1 + \rho_1^E$, and $\bar{\rho}_2 = \underline{\rho}_2 + \rho_2^E$. We summarize the first state-dependent surrender scenario in Remark 2.

Remark 2 *In Scenario 1, we model the representative policyholder's surrender by the Markov-modulated Poisson process with the surrender intensity* γ_t

$$\gamma_t = \begin{cases} (\underline{\rho}_1 \quad \underline{\rho}_2) X_t\,, & \textit{for } S(t, A_t) < v(t, A_t, X_t), \\ (\bar{\rho}_1 \quad \bar{\rho}_1) X_t\,, & \textit{for } S(t, A_t) \geq v(t, A_t, X_t), \end{cases} \tag{4.9}$$

for $(t, A_t) \in (0, T) \times \mathbb{R}^+$ *and* $X_t \in \{e_1, e_2\}$*, where* $\underline{\rho}_1 < \underline{\rho}_2 < \bar{\rho}_1$ *and* $v(t, A_t, X_t)$ *is the contract continuation value.*

In the second state-dependent surrender scenario, we suppose that the systemic health shock does not have significant influence on the policyholder's exogenous surrender decision making, but leads the policyholder to be more cautious and careful in handling his personal investment portfolio, including his participating life insurance contract which yields the insurance company's profit-sharing benefits. So, in this scenario, first, the exogenous surrender intensities share the same value in both the normal health state and the impired health state, i.e., $\underline{\rho}_2 = \underline{\rho}_1$. Second, since the policyholder pays more attention on collecting the financial information and puts more efforts into evaluating the contract after the systemic health shock, he becomes financially more rational, which implies a higher endogenous surrender intensity value in the impaired state than the one in the normal state, i.e., $\rho_2^E > \rho_1^E$. Therefore, given $\bar{\rho}_1 = \underline{\rho}_1 + \rho_1^E$ and $\bar{\rho}_2 = \underline{\rho}_2 + \rho_2^E$, we have a higher upper bound surrender intensity value in the impaired health state than the value in the normal health state. We summarize the second state-dependent surrender scenario in Remark 3.

[10]literatures on confirming the EFH are given in Section 4.1.

Remark 3 *In Scenario 2, we model the representative policyholder's surrender by the Markov-modulated Poisson process with the surrender intensity* γ_t

$$\gamma_t = \begin{cases} (\underline{\rho}_1 \quad \underline{\rho}_1)X_t\,, & \text{for } S(t, A_t) < v(t, A_t, X_t), \\ (\bar{\rho}_1 \quad \bar{\rho}_2)X_t\,, & \text{for } S(t, A_t) \geq v(t, A_t, X_t), \end{cases} \tag{4.10}$$

for $(t, A_t) \in (0, T) \times \mathbb{R}^+$ *and* $X_t \in \{e_1, e_2\}$*, where* $\underline{\rho}_1 < \bar{\rho}_1 < \bar{\rho}_2$ *and* $v(t, A_t, X_t)$ *is the contract continuation value.*

4.2.4 Early Default Mechanism

We assume there is an external regulator who monitors the insurance company's financial structure continuously and adopt the early default mechanism introduced by Grosen and Jørgensen (2002). The regulator sets a default-triggering threshold, which is based on the minimum survival guaranteed benefits promised by the insurance company, $B_t = \theta L_0 e^{r_g t}$ where θ is called the default multiplier. As the value of the insurance company's assets drops below the threshold value before the maturity T, the regulator intervenes and closes the insurance company. The default multiplier θ reflects the strictness of the early default mechanism imposed by the regulator. As the regulator is confident on the insurance company's operation of its assets, she will set a lower default multiplier value so that she will not intervene when the insurance company has only temporarily bad performance. Otherwise, a higher default multiplier is used to protect the policyholder by shutting down the insurance company before its asset value drops too much. In the model, we restrict the default multiplier to be $\theta < 1/\alpha$ so that $A_0 > B_0$ and the regulator does not close the company at the contract-issuing time t_0. As the regulator closes the insurance company at the early default time $\tau_b := \inf\{t \mid A_t < B_t\}$, the company's assets are liquidated and distributed to its stakeholders. We use $\Upsilon(\tau_b, A_{\tau_b})$ to denote the early default benefits paid to the policyholder, which have the lower value of the company's liquidated assets and the minimum survival guaranteed benefits,

$$\Upsilon(\tau_b, A_{\tau_b}) = \min\{A_{\tau_b}, L^{r_g}_{\tau_b}\}, \quad L^{r_g}_{\tau_b} = L_0 e^{r_g \tau_b}. \tag{4.11}$$

If the value of the company's liquidated assets is higher than that of the minimum survival guarantee, i.e., $A_{\tau_b} > L^{r_g}_{\tau_b}$, after paying off the policyholder, the equity holder takes away what is left in the company. Otherwise, the policyholder gets the company's liquidated assets A_{τ_b} and the equity holder receives nothing.

4.3 Contract Valuation

In this section, by using the balance law, see Dai et al. (2007), we derive a system of two coupled partial differential equations following Uzelac and Szimayer (2014) to evaluate the participating life insurance contract. As the contract's payoffs depend on the time t and the company's asset value A_t, we have the contract's value V as a function of t and A_t. Additionally, since the systemic health shock has influence on the mortality intensity of the death process and also the surrender intensity of the surrender process, which accordingly affects the payoffs to the policyholder, the contract's value also depends on the policyholder's health state X_t. Therefore, at time t we have the contract's value $V_t = V(t, A_t, X_t)$. Moreover, as the early default mechanism has influence on the payoffs to the policyholder, same as in Cheng and Li (2015), we also distinguish the region where the early bankruptcy occurs and the region where the company is on-going to evaluate the contract. First, if the company's asset value drops below the default barrier at $t \in (0, T)$, i.e., $A_t \leq B_t$, the regulator closes the company and the policyholder receives the early default benefits. It implies that in the region of $A_t \leq B_t$, the policyholder's contract has the value of the early default benefits $V_t = \Upsilon(t, A_t)$. Second, in the region of $A_t > B_t$ where the insurance company is on-going, i.e., $t < \tau_b$, we suppose the contract's value V_t has the form of $V_t = \mathbb{1}_{\{t<\tau_d,\tau_s\}} v(t, A_t, X_t)$ on the set $\{t < \tau_d, \tau_s\} \cap \{t \leq T\}$, where $v(t, A_t, X_t)$ measures the pre-death-and-surrender value for any $t \in [0, T]$, $A_t \in \mathbb{R}^+$ and $X_t \in \mathcal{E}$. For each state $X_t = e_i$, $i \in \{1, 2\}$, we define $v_i(t, A_t) := v(t, A_t, e_i)$, which is a suitable differential function $v_i : [0, T] \times \mathbb{R}^+ \to \mathbb{R}^+$, $(t, A_t) \to v_i(t, A_t)$. Accordingly, the vector of pre-death-and-surrender contract values for the two states can be defined as $v(t, A_t) = (v_1(t, A_t) \quad v_2(t, A_t))$, with that we have $v(t, A_t, X_t) = v(t, A_t)X_t$. Further, we write the state-dependent mortality intensity specified in (4.4) as $\mu_t = (\mu_1(t) \quad \mu_2(t))X_t$, where $\mu_1(t) = m_0(t)$ and $\mu_2(t) = m_0(t) + m_1$, for any $t \in [0, T]$. As the policyholder is in the normal state, i.e., $X_t = e_1$, he has the predictable mortality intensity $\mu_t = \mu_1(t) = m_0(t)$. Once the health shock occurs, the policyholder's health changes to the impaired state, i.e., $X_t = e_2$, his mortality intensity increases by m_1. Moreover, the surrender intensity γ described in Section 4.2.3 in both scenarios is a function of the time t, the contract's value V_t and the health state X_t, which can be written as $\gamma_t = \gamma(t, V_t, X_t)$. By defining the surrender intensity in each state $X_t = e_i$, $i \in \{1, 2\}$ by $\gamma_i(t, V_t) := \gamma(t, V_t, e_i)$, we have the state-dependent surrender intensity that $\gamma_t = \gamma(t, v_t, X_t) = (\gamma_1(t, v_1) \quad \gamma_2(t, v_2))X_t = (\gamma_1(t, v(t, A_t)e_1) \quad \gamma_2(t, v(t, A_t)e_2))X_t$ on the set $\{t < \tau_d, \tau_s\} \cap \{t < T\}$.

Suppose now $A_t > B_t$, we represent the contract's value on the set $\{t \leq \tau_d \wedge \tau_s \wedge T\}$ by

$$V_t = \mathbb{1}_{\{t<\tau_d,\tau_s\}} v(t, A_t, X_t) + \mathbb{1}_{\{t=\tau_d,\tau_d<\tau_s,T\}} \Psi(t, A_t) + \mathbb{1}_{\{t=\tau_s,\tau_s<\tau_d,T\}} S(t, A_t), \tag{4.12}$$

which implies that as either the death or surrender occurs over $(t, t + dt)$, the payment liability changes by either $\Psi(t, A_t) - v(t, A_t)$ or $S(t, A_t) - v(t, A_t)$. First, we apply the balance law based on the no-arbitrage condition to the contract's value V_t on the set $\{t < \tau_d \wedge \tau_s \wedge T\}$

$$r(t)V_t dt = E_{\mathbb{Q}}[dV_t \,|\, \mathcal{F}_t] \tag{4.13}$$

and obtain

$$\begin{aligned} r(t)v(t, A_t, X_t)dt &= E_{\mathbb{Q}}[dv(t, A_t, X_t) + (\Psi(t, A_t) - v(t, A_t, X_t))dH_t \\ &\quad + (S(t, A_t) - v(t, A_t, X_t))dJ_t \,|\, \mathcal{F}_t] \\ &= E_{\mathbb{Q}}[dv(t, A_t, X_t) \,|\, \mathcal{F}_t] + (\Psi(t, A_t) - v(t, A_t, X_t))\mu(t, X_t)dt \\ &\quad + (S(t, A_t) - v(t, A_t, X_t))\gamma(t, V_t, X_t)dt. \end{aligned} \tag{4.14}$$

Then, similar as in Uzelac and Szimayer (2014), by applying Ito's lemma to the differential of $v(t, A_t, X_t)$ and using the decomposition for the stochastic differential of the Markov chain X, $dX_t = C'X_t dt + dM_t$, where M_t is a $\mathbb{F}$ martingale and C' is the transpose of the generator matrix of X, we have

$$\begin{aligned} dv(t, A_t, X_t) &= \mathcal{L}v(t, A_t, X_t)dt + v(t, A_t)dX_t + \sigma(t, A_t, X_t)A_t \frac{\partial v}{\partial A}(t, A_t, X_t)dW_t^{\mathbb{Q}} \\ &= \mathcal{L}v(t, A_t, X_t)dt + v(t, A_t)C'X_t dt + v(t, A_t)dM_t \\ &\quad + \sigma(t, A_t, X_t)A_t \frac{\partial v}{\partial A}(t, A_t, X_t)dW_t^{\mathbb{Q}}, \end{aligned} \tag{4.15}$$

where $\mathcal{L}$ is define as $\mathcal{L}v(t, A, X) = \frac{\partial v}{\partial t}(t, A, X) + r(t)A\frac{\partial v}{\partial A}(t, A, X) + \frac{1}{2}\sigma^2(t, A)A^2\frac{\partial^2 v}{\partial A^2}(t, A, X)$.[11] We now plug (4.15) to (4.14) and use the following result

$$E_{\mathbb{Q}}[v(t, A_t)dM_t \,|\, \mathcal{F}_t] = E_{\mathbb{Q}}[\sigma(t, A_t, X_t)A_t \frac{\partial v}{\partial A}(t, A_t, X_t)dW_t^{\mathbb{Q}} \,|\, \mathcal{F}_t] = 0 \tag{4.16}$$

[11] For the Ito's formula for general semimartingales, see section II of Protter (2005).

to obtain

$$
\begin{aligned}
r(t)v(t,A_t,X_t) &= \mathcal{L}v(t,A_t,X_t)+v(t,A_t)C'X_t+(\Psi(t,A_t)-v(t,A_t,X_t))\mu(t,X_t)\\
&\quad +(S(t,A_t)-v(t,A_t,X_t))\gamma(t,V_t,X_t). \qquad (4.17)
\end{aligned}
$$

By arranging the terms in the equation (4.17), we have the following partial differential equation (PDE)

$$
\begin{aligned}
0 &= \mathcal{L}v(t,A_t,X_t)+\Psi(t,A_t)\mu(t,X_t)+S(t,A_t)\gamma(t,V_t,X_t)\\
&\quad -(r(t)+\mu(t,X_t)+\gamma(t,V_t,X_t))v(t,A_t,X_t)+v(t,A_t)C'X_t, \qquad (4.18)
\end{aligned}
$$

which holds for any $t\in[0,T)$, $A_t\in\mathbb{R}^+$ and $X_t\in\mathcal{E}$. So for each state $X_t=e_i$, $i=1,2$, we have the following state-dependent PDE

$$
\begin{aligned}
0 &= \mathcal{L}v_i(t,A_t)+\Psi(t,A_t)\mu_i(t)+S(t,A_t)\gamma_i(t,v_i(t,A_t))\\
&\quad -(r(t)+\mu_i(t)+\gamma_i(t,v_i))v_i(t,A_t)+v(t,A_t)C'e_i, \qquad (4.19)
\end{aligned}
$$

where $\mathcal{L}v_i(t,A_t)=\mathcal{L}v(t,A_t,e_i)$. The no-arbitrage condition in state $i=1,2$ is $v_i(T,A)=\Phi(T,A)$, $A\in\mathbb{R}^+$. Given the pre-death-and-surrender value v_i for $i=1,2$, we can write the contract value V_t on the set $\{t\le\tau_d\wedge\tau_s\wedge T\}$ in the region of $A_t>B_t$ as follows

$$
\begin{aligned}
V_t &= \mathbb{1}_{\{t<\tau_d,\tau_s\}}(\mathbb{1}_{\{X_t=e_1\}}v_1(t,A_t)+\mathbb{1}_{\{X_t=e_2\}}v_2(t,A_t))\\
&\quad +\mathbb{1}_{\{t=\tau_d,\tau_d<\tau_s,T\}}\Psi(t,A_t)+\mathbb{1}_{\{t=\tau_s,\tau_s<\tau_d,T\}}S(t,A_t). \qquad (4.20)
\end{aligned}
$$

We summarize the results in the following proposition.

Proposition 4 *Given the contract value V_t described in* (4.20)*, the pre-death-and-surrender value functions $v_i(t,A_t)$, $i=1,2$ for $t\in[0,T)$ and $A_t\in\mathbb{R}^+$ are the solutions of the following system of partial differential equations:*

$$
\begin{aligned}
0 &= \mathcal{L}v_1(t,A_t)+\Psi(t,A_t)\mu_1(t)+S(t,A_t)\gamma_1(t,v_1(t,A_t))\\
&\quad -(r(t)+\mu_1(t)+\gamma_1(t,v_1(t,A_t)))v_1(t,A_t)+\kappa(v_2(t,A_t)-v_1(t,A_t)), \qquad (4.21)\\
0 &= \mathcal{L}v_2(t,A_t)+\Psi(t,A_t)\mu_2(t)+S(t,A_t)\gamma_2(t,v_2(t,A_t))\\
&\quad -(r(t)+\mu_2(t)+\gamma_2(t,v_2(t,A_t)))v_2(t,A_t), \qquad (4.22)
\end{aligned}
$$

where $\mu_1(t) = m_0(t)$ *and* $\mu_2(t) = m_0(t) + m_1$*; subject to the boundary conditions*

$$v_1(t, A_t) = v_2(t, A_t) = \Upsilon(t, A_t) \quad for \quad A_t = B_t = \theta L_0 e^{r_g t} \tag{4.23}$$

and with the termination conditions

$$v_1(T, A_T) = v_2(T, A_T) = \Phi(A_T) \quad for \quad A_T \in \mathbb{R}^+. \tag{4.24}$$

The surrender intensity in the state $i = 1, 2$ *has the form of*

$$\gamma_i(t, v_i) = \begin{cases} (\underline{\rho}_1 \quad \underline{\rho}_2)e_i\,, & for\ S(t, A_t) < v_i(t, A_t), \\ (\bar{\rho}_1 \quad \bar{\rho}_2)e_i\,, & for\ S(t, A_t) \geq v_i(t, A_t), \end{cases} \tag{4.25}$$

where $\underline{\rho}_1 < \underline{\rho}_2 < \bar{\rho}_1 = \bar{\rho}_2$ *in the scenario 1, or* $\underline{\rho}_1 = \underline{\rho}_2 < \bar{\rho}_1 < \bar{\rho}_2$ *in the scenario 2.*

Furthermore, the contract continuation values v_1 and v_2 are influenced by the bounds $(\bar{\rho}_1, \underline{\rho}_1)$ and $(\bar{\rho}_2, \underline{\rho}_2)$ of surrender intensity. It can be proved that v_1 and v_2 increase as a higher value of $\bar{\rho}_1$, and lower values of $\underline{\rho}_1$ and $\underline{\rho}_2$ are imposed in Scenario 1; or higher values of $\bar{\rho}_1$ and $\bar{\rho}_2$, and a lower value of $\underline{\rho}_1$ are imposed in Scenario 2. This has economic intuition. Since the policyholder with a lower value of exogenous surrender intensity and a higher level of monetary rationality in state $i = 1, 2$ can carry out the surrender decision more likely at the optimal time, his contract value then increases in either state. See Appendix 4.6.1 for proof of Proposition 5.[12]

Proposition 5 *Within the given regulatory framework characterised by the default multiplier* θ *and the given investment strategy of the insurance company described by* σ*, suppose pre-death-and-surrender value functions* v_1 *and* v_2 *are solutions to the system of the two PDEs in Proposition 4. Furthermore, suppose that* $\tilde{v}_1$ *and* $\tilde{v}_2$ *are also solutions to the system of the two PDEs in Proposition 4 but with* $\tilde{\bar{\rho}}_1 \geq \bar{\rho}_1$*,* $\tilde{\underline{\rho}}_1 \leq \underline{\rho}_1$ *and* $\tilde{\underline{\rho}}_2 \leq \underline{\rho}_2$ *in Scenario 1, or* $\tilde{\bar{\rho}}_1 \geq \bar{\rho}_1$*,* $\tilde{\bar{\rho}}_2 \geq \bar{\rho}_2$ *and* $\tilde{\underline{\rho}}_1 \leq \underline{\rho}_1$ *in Scenario 2. Then we have* $\tilde{v}_1(t, A) \geq v_1(t, A)$ *and* $\tilde{v}_2(t, A) \geq v_2(t, A)$ *for* $(t, A) \in [0, \tau_b \wedge T] \times \mathbb{R}^+$.

4.4 Numerical Analysis

In this section we numerically solve the coupled PDE system with a continuously applied barrier in Proposition 4 by using the finite difference method proposed by Zvan et al. (1996) and Zvan et al. (2000). First, we study effects of the systemic health shock

[12] For an alternative proof of Proposition 5, confer Uzelac and Szimayer (2014).

together with changes in the surrender behavior due to the health shock on the contract valuation. We analyze how the effects change with respect to the health damage level represented by m_1 and we further compare these effects within different regulatory frameworks θ. Second, we focus on the effects of regulatory frameworks on the contract valuation given that the systemic health shock is taken into account for pricing together with different considerations of changes in the policyholder's surrender behavior after the health damage.

We use the following benchmark values of the relevant parameters for the numerical computation. At time $t_0 = 0$, the insurance company is set up by issuing a participating life insurance contract to the policyholder and levying money from the equity holder. It has initial assets $A_0 = 100$, which are invested into assets traded on the financial market. The risk-free interest rate is $r = 0.04$ and the volatility of the company's asset process is assumed to be constant $\sigma = 0.2$. 85% of the company's initial assets are labelled as initial liabilities, i.e., $\alpha = 0.85$, in the participating policy that matures in 10 years. The contract guarantees a minimum interest rate 0.02 at maturity, death and surrender, which means $r_g = r_d = r_s = 0.02$. On top of the minimum interest rate guarantee, the contract promises a participation rate 0.9 both at maturity and at death, i.e., $\delta = \delta_d = 0.9$. However, if the policyholder terminates his contract before maturity, the following penalty parameter value is applied to calculate the surrender benefits

$$\beta_t = \begin{cases} 0.05, & \text{for } t \leq 1, \\ 0.04, & \text{for } 1 < t \leq 2, \\ 0.02, & \text{for } 2 < t \leq 3, \\ 0.01, & \text{for } 3 < t \leq 4, \\ 0, & \text{for } t > 4. \end{cases}$$

It implies that the earlier the policyholder surrenders his contract, the more punishment he faces by doing so. We assume that the policyholder who buys the contract at t_0 is at age $y = 40$ and his predictable mortality intensity function $m_0(t)$ has the form $m_0(t) = A^\mu + Bc^{y+t}$ with $A^\mu = 5.0758 \times 10^{-4}$, $B = 3.9342 \times 10^{-5}$, $c = 1.1029$, which is used in Cheng and Li (2015) and Li and Szimayer (2014). At the occurrence of the health shock, the policyholder's health state switches from the normal state to the impaired state, and his mortality intensity shifts up by $m_1 = 0.05$. The intensity of the health shock process is assumed to be $\kappa = 0.075$ in the benchmark. Lastly, we assume that the benchmark value of the regulatory threshold θ is equal to 0.9.

4.4.1 Effects of the Systemic Health Shock on the Contract Valuation

In this section we analyze the effects of the systemic health shock on the contract's fair value in both the case where the health shock does not influence the policyholder's surrender decision making and the case where the policyholder adjusts his surrender behavior after the health shock, as described in both Scenario 1 and Scenario 2. In order to further isolate influence of the early default regulation on these effects, we conduct the analysis in both regulation-free and with-regulation environments. For the sake of discussion, in the following subsections we suppose assume that the insurance company overlooks the systemic health shock while pricing the contract, which means the contract value computed with no health shock is charged to the policyholder.

4.4.1.1 When Surrender Risk is not Linked to Mortality Risk

We start in the situation where there is no early fault mechanism and the policyholder's surrender behavior does not change after the health shock. We present contract values for different types of policyholders, whose surrender rationalities in the normal and impaired states are represented by $(\underline{\rho}_1, \bar{\rho}_1)$ and $(\underline{\rho}_2, \bar{\rho}_2)$ respectively, satisfying $(\underline{\rho}_1, \bar{\rho}_1) = (\underline{\rho}_2, \bar{\rho}_2)$, before and after the systemic health shock is taken into account for pricing in Table 4.1. m_1 measures the damage level of the health shock, which takes values of 0.01, 0.03, 0.05 and 0.07 in the computation. In column 2 are contract values which are computed when the systemic health shock is overlooked by the insurance company and charged to the different types of policyholder. We see that the contract value is highest for the fully rational policyholder represented by $(0, \infty)$ and decreases as the exogenous surrender intensity increases and the endogenous surrender intensity decreases. As the health shock with a minor damage level is presented for pricing, see column 3 with $m_1 = 0.01$, we observe that except for the fully rational policyholder, the contract value for the rest of policyholders becomes higher. Furthermore, the contract value keeps increasing with the damage level of the health shock for these policyholders, but decreasing for the fully rational policyholder. We know that the participating life insurance contract attracts policyholders by promising a participation in the company's favorable asset performance. For the fully rational policyholder who is capable of making an optimal surrendering decision, an increase in his mortality due to the systemic health shock lowers his chance of participating in the company's asset surplus at maturity. Hence, the insurance company has to lower the contract price in order to compensate the fully rational policyholder and the compensation increases with the damage level of the health shock. However, for

$(\underline{\rho}_1, \bar{\rho}_1) = (\underline{\rho}_2, \bar{\rho}_2)$	no early default regulation				
	no health shock	with health shock			
		$m_1 = 0.01$	$m_1 = 0.03$	$m_1 = 0.05$	$m_1 = 0.07$
(0, 0)	85.6140	85.6442	85.7003	85.7512	85.7974
(0, 0.03)	86.0368	86.0614	86.1071	86.1486	86.1863
(0, 0.3)	88.1531	88.154	88.1558	88.1574	88.159
(0, ∞)	92.0546	92.0376	92.0055	91.9759	91.9485
(0.03, 0.03)	81.8567	81.9306	82.0676	82.1913	82.3034
(0.03, 0.3)	84.2656	84.3098	84.3918	84.466	84.5332
(0.03, ∞)	88.5391	88.5593	88.5964	88.6293	88.6585
(0.3, 0.3)	75.4561	75.534	75.68	75.8138	75.9367
(0.3, ∞)	80.75	80.75	80.75	80.75	80.75

Table 4.1: *Contract values for different rationality levels represented by $(\underline{\rho}_i, \bar{\rho}_i)$, $i = 1, 2$ with $\underline{\rho}_1 = \underline{\rho}_2$ and $\bar{\rho}_1 = \bar{\rho}_2$ and different health damage levels represented by m_1 in the situation when there is no early default mechanism.*

partially rational policyholders, there are two situations where bearing a higher mortality intensity after the health shock may help improve their positions, leading to a higher contract value. The first situation is that for a partially rational policyholder with $\bar{\rho}_i < \infty$, $i = 1, 2$ who has difficulties in collecting information on the financial market and analyzing the contract's fair value, the increasing death probability after the health shock exempts him from making endogenous surrender decisions, in particular at the time when he should surrender his contract but fails to do so. The second situation refers to the case where the policyholder with $\underline{\rho}_i > 0$, $i = 1, 2$ has to surrender the contract due to personal non-avoidable reasons. If the policyholder dies before he surrenders his contract for exogenous reasons, he receives death benefits without any punishment. And these protecting effects of the systemic health shock on the partially rational policyholders become more significant, i.e., the contract value keeps increasing, as the damage level m_1 increases from 0.01 to 0.07. Therefore, in the regulation-free environment, the contract of the partially rational policyholder is undervalued, while the contract of the fully rational policyholder is overvalued.

Next, we take into consideration the early default mechanism imposed by the regulator. We start with a relatively mild regulatory rule represented by $\theta = 0.7$, which requires the insurance company to hold assets' value not less than 70% of the promised interest rate guarantee. Contract values for the different types of policyholders with the different increases in the mortality intensity after the health shock are presented in Table 4.2. Now

$(\underline{\rho}_1, \bar{\rho}_1) = (\underline{\rho}_2, \bar{\rho}_2)$	with early default regulation, $\theta = 0.7$				
	no health shock	with health shock			
		$m_1 = 0.01$	$m_1 = 0.03$	$m_1 = 0.05$	$m_1 = 0.07$
(0, 0)	86.8199	86.833	86.8574	86.8797	86.9001
(0, 0.03)	87.0668	87.0773	87.0968	87.1146	87.131
(0, 0.3)	88.468	88.4657	88.4615	88.4576	88.4542
(0, ∞)	92.0548	92.0377	92.0057	91.976	91.9486
(0.03, 0.03)	82.9119	82.971	83.0805	83.1796	83.2693
(0.03, 0.3)	84.5696	84.6107	84.687	84.7561	84.8186
(0.03, ∞)	88.5392	88.5594	88.5965	88.6294	88.6586
(0.3, 0.3)	75.7496	75.8243	75.9641	76.0924	76.2103
(0.3, ∞)	80.7500	80.7500	80.7500	80.7500	80.7500

Table 4.2: *Contract values for different rationality levels represented by $(\underline{\rho}_i, \bar{\rho}_i)$, $i = 1, 2$ with $\underline{\rho}_1 = \underline{\rho}_2$ and $\bar{\rho}_1 = \bar{\rho}_2$ and different health damage levels represented by m_1 in the situation when the regulator imposes an early default mechanism with $\theta = 0.7$.*

we see that, with the protection of the regulator, the contract value decreases not only for the fully rational policyholder but also for the policyholder with $(\underline{\rho}_i, \bar{\rho}_i) = (0, 0.3)$, $i = 1, 2$ as the health shock is taken into account for pricing. As we have discussed in Section 4.2.4, the early default regulation helps protect the policyholder from the downside risk of the company's asset process. More in detail, it protects the policyholder who is not competent enough to terminate his contract as the company's asset performance worsens and his contract value drops below the value of surrender benefits by closing the company and paying him early default benefits. Under the protection of the regulator, this policyholder, who surrenders his contract with relatively high competency, i.e., $\bar{\rho}_i = 0.3$, and at the same time does not surrender his contract because of exogenous reasons, actually acts like the fully rational policyholder. Hence, the policyholder with $(0, 0.3)$ faces a similar situation as the fully rational policyholder when the systemic health shock is taken into consideration for pricing, which is a higher death probability after the health shock lowers the probability that the policyholder participates in the company's favourable asset performance at maturity. So the contract becomes less attractive for the policyholder under the protection of the regulator. However, for the rest policyholders who have relatively low financial rationality and/or do surrender contracts for exogenous reasons, since the mild regulatory rule does not fully assist them in carrying out optimal surrender strategies, their contract values increase as the systemic health shock is considered for pricing, the same as in Table 4.1.

Suppose now a stricter regulatory rule, i.e., $\delta = 0.9$, is imposed and we present contract values for different types of policyholders within this regulatory framework in Table 4.3. As a higher company's asset value, i.e., 90% of minimum interest rate guaranteed value, is required by the regulator, policyholders get more intensively protected from the downside risk of the insurance company. It results in that more policyholders (the policyholders who have relatively low financial competency, e.g., $(\underline{\rho}_i, \bar{\rho}_i) = (0, 0.03)$ and $(\underline{\rho}_i, \bar{\rho}_i) = (0, 0)$, are

$(\underline{\rho}_1, \bar{\rho}_1) = (\underline{\rho}_2, \bar{\rho}_2)$	with early default regulation, $\theta = 0.9$				
	no health shock	with health shock			
		$m_1 = 0.01$	$m_1 = 0.03$	$m_1 = 0.05$	$m_1 = 0.07$
(0, 0)	90.4937	90.4817	90.4592	90.4385	90.4194
(0, 0.03)	90.5088	90.4967	90.474	90.4531	90.4339
(0, 0.3)	90.627	90.6142	90.5902	90.5682	90.5487
(0, ∞)	92.0628	92.0457	92.0135	91.9838	91.9563
(0.03, 0.03)	86.6744	86.7049	86.7611	86.8114	86.8566
(0.03, 0.3)	86.8343	86.8634	86.9169	86.9649	87.0079
(0.03, ∞)	88.5436	88.5638	88.6009	88.6338	88.663
(0.3, 0.3)	78.0482	78.1081	78.2204	78.3234	78.4181
(0.3, ∞)	80.75	80.75	80.75	80.75	80.75

Table 4.3: *Contract values for different rationality levels represented by* $(\underline{\rho}_i, \bar{\rho}_i)$, $i = 1, 2$ *with* $\underline{\rho}_1 = \underline{\rho}_2$ *and* $\bar{\rho}_1 = \bar{\rho}_2$ *and different health damage levels represented by* m_1 *in the situation when the regulator imposes an early default mechanism with* $\theta = 0.9$.

$(\underline{\rho}_1, \bar{\rho}_1) = (\underline{\rho}_2, \bar{\rho}_2)$	with early default regulation, $\theta = 1.1$				
	no health shock	with health shock			
		$m_1 = 0.01$	$m_1 = 0.03$	$m_1 = 0.05$	$m_1 = 0.07$
(0, 0)	89.6619	89.659	89.6534	89.6482	89.6433
(0, 0.03)	89.6619	89.659	89.6534	89.6482	89.6433
(0, 0.3)	89.6619	89.659	89.6534	89.6482	89.6433
(0, ∞)	89.6619	89.659	89.6534	89.6482	89.6433
(0.03, 0.03)	87.9197	87.934	87.9602	87.9838	88.005
(0.03, 0.3)	87.9197	87.934	87.9602	87.9838	88.005
(0.03, ∞)	87.9197	87.934	87.9602	87.9838	88.005
(0.3, 0.3)	83.2948	83.3209	83.3698	83.4148	83.4563
(0.3, ∞)	83.2952	83.3212	83.37	83.415	83.4565

Table 4.4: *Contract values for different rationality levels represented by* $(\underline{\rho}_i, \bar{\rho}_i)$, $i = 1, 2$ *with* $\underline{\rho}_1 = \underline{\rho}_2$ *and* $\bar{\rho}_1 = \bar{\rho}_2$ *and different health damage levels represented by* m_1 *in the situation when the regulator imposes an early default mechanism with* $\theta = 1.1$.

included) face the same situation as the fully rational policyholder, where their contract values decrease as the systemic health shock is considered for pricing. Therefore, more contracts are overpriced by the insurance company as the regulation strengthens, which helps improve the company's position. Further, we obtain similar results when an over-regulation rule is imposed, see the contract values in Table 4.4 where $\theta = 1.1$. Moreover, we observe the dominating positive effects of the health shock on the contract valuation for policyholders with non-zero exogenous surrender intensities $\underline{\rho}_i > 0$ in any regulatory environment, i.e., their contract value increases as the health shock is considered in Table 4.1, 4.2, 4.3 and 4.4.

4.4.1.2 When Surrender Risk is linked to Mortality Risk

Now we focus on the situation when the policyholder's surrender behavior changes after the health shock. Table 4.5-4.8 present contract values for policyholders whose surrender behaviors change as described in Scenario 1 after the health shock within different regulatory frameworks. In the third column of Table 4.5-4.8 are contract values which are computed by assuming no systemic health shock so that policyholders' surrender behaviors are consistent during the entire insuring period. For the sake of discussion, suppose that these contract prices are charged to the policyholders. We observe that contract values for all different types of policyholders in columns 4-7 where the systemic health shock is considered for pricing are lower than the charged prices, which holds true within all different regulatory frameworks. As we have described in Section 4.2.3 that in Scenario 1 policyholders bear more financial distress after experiencing the health damage, it is more likely that the policyholders surrender their contracts for exogenous reasons and also become impatient in analyzing the contracts' value, which consequently lowers their contracts' value. By comparing to changes in the contract value in Section 4.4.1.1, we notice that this negative effect of the policyholder's exogenous irrationality after the health shock dominates the positive effects of the systemic health shock on the contract valuation that we have mentioned in Section 4.4.1.1, which holds true within all different regulatory frameworks. The higher the exogenous surrender intensity increases, the more likely the policyholder exercises his contract for exogenous reasons and the more his contract value decreases. The decrease in the contract value measures the premium that the insurance company should have not charged given that the health shock can occur and the policyholder becomes more irrational after experiencing it. Therefore, if the company overlooks the systemic health shock, an overly-priced contract is going to be sold to the policyholder, which is definitely not a wish of the regulator. But sadly, the regulator cannot help change the policyholder's position by imposing the early default mechanism in this situation.

$(\underline{\rho}_1, \bar{\rho}_1)$	$(\underline{\rho}_2, \bar{\rho}_2)$	no early default regulation				
		no health shock	with health shock & scenario 1			
			$m_1 = 0.01$	$m_1 = 0.03$	$m_1 = 0.05$	$m_1 = 0.07$
(0, 0.03)	(0.03, 0.03)	86.0368	84.612	84.7559	84.8861	85.004
(0, 0.3)	(0.03, 0.3)	88.1531	86.8415	86.9303	87.0107	87.0838
	(0.3, 0.3)		82.0224	82.2885	82.5373	82.7686
(0.03, 0.3)	(0.3, 0.3)	84.2656	80.1099	80.3463	80.5667	80.771
(0, ∞)	(0.03, ∞)	92.0546	90.9427	90.983	91.0188	91.0508
	(0.3, ∞)		86.8433	87.0237	87.2001	87.3668
(0.03, ∞)	(0.3, ∞)	88.5391	84.9128	85.0756	85.2352	85.386

Table 4.5: *Contract values for different rationality levels in the normal state represented by $(\underline{\rho}_1, \bar{\rho}_1)$ and in impaired state represented by $(\underline{\rho}_2, \bar{\rho}_2)$ in Scenario 1 and different health damage levels represented by m_1 in the situation when there is no early default mechanism.*

$(\underline{\rho}_1, \bar{\rho}_1)$	$(\underline{\rho}_2, \bar{\rho}_2)$	with early default regulation, $\theta = 0.7$				
		no health shock	with health & scenario 1			
			$m_1 = 0.01$	$m_1 = 0.03$	$m_1 = 0.05$	$m_1 = 0.07$
(0, 0.03)	(0.03, 0.03)	87.0668	85.6502	85.7661	85.8711	85.9663
(0, 0.3)	(0.03, 0.3)	88.468	87.1544	87.2372	87.3123	87.3803
	(0.3, 0.3)		82.3384	82.6003	82.8429	83.0686
(0.03, 0.3)	(0.3, 0.3)	84.5696	80.4093	80.6407	80.8557	81.0571
(0, ∞)	(0.03, ∞)	92.0548	90.9428	90.9831	91.019	91.0509
	(0.3, ∞)		86.8433	87.0238	87.2002	87.3668
(0.03, ∞)	(0.3, ∞)	88.5392	84.9128	85.0757	85.2353	85.3861

Table 4.6: *Contract values for different rationality levels in the normal state represented by $(\underline{\rho}_1, \bar{\rho}_1)$ and in impaired state represented by $(\underline{\rho}_2, \bar{\rho}_2)$ in Scenario 1 and different health damage levels represented by m_1 in the situation when the regulator imposes an early default mechanism with $\theta = 0.7$.*

In addition, we observe that as the damage level of the health shock represented by m_1 increases, the contract value increases for all types of policyholders. Since after the health shock, the policyholder is more likely to terminate his contract for personal non-avoidable liquidity reasons, experiencing a higher level of health damage increases the chance that the policyholder gets death benefits before he exercises the surrender option and gets surrender benefits with punishments. Furthermore, by comparing the decreases in the contract value as the health shock is taken into consideration in Table 4.5-4.8, we find that the decrease in the contract value for any type of the policyholder is significantly smaller in the case when an over-regulation is imposed by the regulator, i.e., $\theta = 1.1$. With the over-regulation rule, it is more likely that the regulator forces the company

$(\underline{\rho}_1, \bar{\rho}_1)$	$(\underline{\rho}_2, \bar{\rho}_2)$	with early default regulation, $\theta = 0.9$				
		no health shock	with health shock & scenario 1			
			$m_1 = 0.01$	$m_1 = 0.03$	$m_1 = 0.05$	$m_1 = 0.07$
(0, 0.03)	(0.03, 0.03)	90.5088	89.2506	89.3111	89.3655	89.4144
(0, 0.3)	(0.03, 0.3)	90.627	89.3855	89.4434	89.4953	89.5421
	(0.3, 0.3)		84.8058	85.0368	85.2511	85.4504
(0.03, 0.3)	(0.3, 0.3)	86.8343	82.8091	83.018	83.2115	83.391
(0, ∞)	(0.03, ∞)	92.0628	90.9493	90.9896	91.0255	91.0575
	(0.3, ∞)		86.8468	87.0273	87.2038	87.3705
(0.03, ∞)	(0.3, ∞)	88.5436	84.9154	85.0783	85.2379	85.3888

Table 4.7: *Contract values for different rationality levels in the normal state represented by* $(\underline{\rho}_1, \bar{\rho}_1)$ *and in impaired state represented by* $(\underline{\rho}_2, \bar{\rho}_2)$ *in Scenario 1 and different health damage levels represented by* m_1 *in the situation when the regulator imposes an early default mechanism with* $\theta = 0.9$.

$(\underline{\rho}_1, \bar{\rho}_1)$	$(\underline{\rho}_2, \bar{\rho}_2)$	with early default regulation, $\theta = 1.1$				
		no health shock	with health shock & scenario 1			
			$m_1 = 0.01$	$m_1 = 0.03$	$m_1 = 0.05$	$m_1 = 0.07$
(0, 0.03)	(0.03, 0.03)	89.6619	89.1331	89.1614	89.1869	89.2098
(0, 0.3)	(0.03, 0.3)	89.6619	89.1331	89.1614	89.1869	89.2098
	(0.3, 0.3)		87.1187	87.2198	87.3136	87.4009
(0.03, 0.3)	(0.3, 0.3)	87.9197	86.1287	86.2213	86.3071	86.3868
(0, ∞)	(0.03, ∞)	89.6619	89.1331	89.1614	89.1869	89.2098
	(0.3, ∞)		87.1188	87.2198	87.3136	87.4009
(0.03, ∞)	(0.3, ∞)	87.9197	86.1288	86.2213	86.3071	86.3868

Table 4.8: *Contract values for different rationality levels in the normal state represented by* $(\underline{\rho}_1, \bar{\rho}_1)$ *and in impaired state represented by* $(\underline{\rho}_2, \bar{\rho}_2)$ *in Scenario 1 and different health damage levels represented by* m_1 *in the situation when the regulator imposes an early default mechanism with* $\theta = 1.1$.

to shut down and pay out its liquidated assets very early, even at the time when the company is operating well. Hence, the effects of the health damage due to the systemic health shock and the accompanied more irrational surrender behavior after the health shock on the contact valuation are less significant than those in the environment with no regulation or with a mild regulatory rule, i.e.,$\theta = 0.7$ and $\theta = 0.9$. If the regulator is aiming at lowering the difference between the contract's selling price and its fair value, an overly-regulated early default rule should be imposed.

So far we have discussed about the Scenario 1 and now we present contract values for policyholders whose surrender behaviors change after the health shock as described in Scenario 2 within different regulatory frameworks in Table 4.9-4.12. Policyholders become more careful and responsible in handling their contracts after experiencing the health damage in Scenario 2, which has positive effects on the contract valuation. Hence, first, we observe that contract values in Table 4.9-4.11 are higher than the ones in Table 4.1-4.3 where the policyholders' surrender behaviors are assumed to be consistent. The larger the increases in the policyholders' financial rationalities, the higher the policyholders' contract values. Second, as the regulation is harshly imposed, i.e., $\theta = 1.1$, we observe that the improvements of policyholders' endogenous surrender behaviors do not have any impacts on the contract valuation and in addition, the contract values for the policyholders who have the same value of the exogenous surrender intensity are the same as the values in the case when the policyholders' surrender behaviors are consistent before and after the health shock, see Table 4.4 and 4.12. Since as the insurance company is liquidated under the overly regulatory rule, the policyholder receives early default benefits which have a higher value than that of surrender benefits, the policyholder is not going to surrender his contract for endogenous reasons. Therefore, whether there is an increase in the policyholder's financial intensity and how large the increase is after the health shock become irrelevant for determining the contracts' values.

Now we discuss about the effects of taking into account the health shock for pricing within the different regulatory frameworks. Same as in the previous subsection, suppose the company overlooks the health shock while pricing the contract. First, Table 4.9 summarizes contract values when there is no early default regulatory rule, where we observe higher contract values for all the levels of the health damage when the health shock is considered for pricing. The more rational the policyholder becomes after experiencing the health shock, i.e., the larger $\bar{\rho}_2 - \bar{\rho}_1$, the larger the increase in his contract value and the more the contract is undervalued. Additionally, we observe that the increase in the contract value for the policyholder who has zero exogenous surrender intensity and an infinitely large value of endogenous surrender intensity after experiencing the health shock becomes smaller as the damage level of the health shock increases. Since the policyholder with $(0, \infty)$ in the impaired health state can exercise his contract optimally but an increase in the mortality rate lowers the chance that the policyholder participates in sharing profits of the company in the long run, the contract value decreases with respect to the health damage level for these types of policyholders. This phenomenon also corresponds to the decreasing contract value with respect to m_1 for the fully rational policyholder in Table 4.1. The rest policyholders with a non-zero exogenous intensity or a finite endogenous surrender intensity in the impaired state, i.e., $\underline{\rho}_2 = \{0.03, 0.3\}$ or

$(\underline{\rho}_1, \bar{\rho}_1)$	$(\underline{\rho}_2, \bar{\rho}_2)$	no early default regulation				
		no health shock	with health & scenario 2			
			$m_1 = 0.01$	$m_1 = 0.03$	$m_1 = 0.05$	$m_1 = 0.07$
(0, 0)	(0, 0.03)	85.614	85.7461	85.7936	85.8368	85.876
	(0, 0.3)		86.3106	86.316	86.3212	86.326
	(0, ∞)		87.5263	87.4844	87.4454	87.4092
(0, 0.03)	(0, 0.3)	86.0368	86.6051	86.6099	86.6145	86.6188
	(0, ∞)		87.7756	87.7341	87.6956	87.6599
(0, 0.3)	(0, ∞)	88.1531	89.0928	89.0541	89.0182	88.9849
(0.03, 0.03)	(0.03, 0.3)	81.8567	82.5029	82.5941	82.6769	82.7522
	(0.03, ∞)		83.692	83.7313	83.766	83.7968
(0.03, 0.3)	(0.03, ∞)	84.2656	85.2254	85.2641	85.2985	85.329
(0.3, 0.3)	(0.3, ∞)	75.4561	76.2958	76.3865	76.4778	76.5642

Table 4.9: *Contract values for different rationality levels in the normal state represented by $(\underline{\rho}_1, \bar{\rho}_1)$ and in impaired state represented by $(\underline{\rho}_2, \bar{\rho}_2)$ in Scenario 2 and different health damage levels represented by m_1 in the situation when there is no early default mechanism.*

$(\underline{\rho}_1, \bar{\rho}_1)$	$(\underline{\rho}_2, \bar{\rho}_2)$	with early default regulation, $\theta = 0.7$				
		no health shock	with health shock & scenario 2			
			$m_1 = 0.01$	$m_1 = 0.03$	$m_1 = 0.05$	$m_1 = 0.07$
(0, 0)	(0, 0.03)	86.8199	86.8905	86.9106	86.929	86.9459
	(0, 0.3)		87.2396	87.2363	87.2334	87.2309
	(0, ∞)		88.2585	88.217	88.1785	88.1427
(0, 0.03)	(0, 0.3)	87.0668	87.4174	87.414	87.411	87.4083
	(0, ∞)		88.4096	88.3685	88.3304	88.2949
(0, 0.3)	(0, ∞)	88.468	89.3082	89.2695	89.2336	89.2003
(0.03, 0.03)	(0.03, 0.3)	82.9119	83.3358	83.4183	83.4931	83.5609
	(0.03, ∞)		84.3451	84.3845	84.4193	84.4502
(0.03, 0.3)	(0.03, ∞)	84.5696	85.4357	85.4745	85.5089	85.5394
(0.3, 0.3)	(0.3, ∞)	75.7496	76.5002	76.5906	76.6816	76.7679

Table 4.10: *Contract values for different rationality levels in the normal state represented by $(\underline{\rho}_1, \bar{\rho}_1)$ and in imparied state represented by $(\underline{\rho}_2, \bar{\rho}_2)$ in Scenrario 2 and different levels represented by m_1 in the situation when the regulator imposes an early default mechanism with $\theta = 0.7$.*

$\bar{\rho}_2 = \{0.03, 0.3\}$, would still benefit from the increasing probability of dying and receiving death benefits, as we have discussed in Section 4.4.1.1. Suppose now an early default regulatory rule with a relatively low threshold $\theta = 0.7$ is imposed by the regulator, see Table 4.10. We observe the same effects of improvements in the policyholders' financial rationality after the health shock on the contract valuation as in the case of no regulation, i.e., higher contract values for all types of policyholders as the health shock is taken into

$(\underline{\rho}_1, \bar{\rho}_1)$	$(\underline{\rho}_2, \bar{\rho}_2)$	with early default regulation, $\theta = 0.9$				
		no health shock	with health shock & scenario 2			
			$m_1 = 0.01$	$m_1 = 0.03$	$m_1 = 0.05$	$m_1 = 0.07$
(0, 0)	(0, 0.03)	90.4937	90.4857	90.463	90.4421	90.4228
	(0, 0.3)		90.5172	90.4929	90.4705	90.45
	(0, ∞)		90.8561	90.8196	90.7859	90.7546
(0, 0.03)	(0, 0.3)	90.5088	90.528	90.5038	90.4815	90.4609
	(0, ∞)		90.8658	90.8295	90.7958	90.7646
(0, 0.3)	(0, ∞)	90.627	90.9437	90.9077	90.8743	90.8435
(0.03, 0.03)	(0.03, 0.3)	86.6744	86.7406	86.7946	86.8429	86.8863
	(0.03, ∞)		87.0891	87.1278	87.1621	87.1925
(0.03, 0.3)	(0.03, ∞)	86.8343	87.1995	87.238	87.2722	87.3025
(0.3, 0.3)	(0.3, ∞)	78.0482	78.3673	78.4534	78.5401	78.6222

Table 4.11: *Contract values for different rationality levels in the normal state represented by $(\underline{\rho}_1, \bar{\rho}_1)$ and in imparied state represented by $(\underline{\rho}_2, \bar{\rho}_2)$ in Scenrario 2 and different levels represented by m_1 in the situation when the regulator imposes an early default mechanism with $\theta = 0.9$.*

$(\underline{\rho}_1, \bar{\rho}_1)$	$(\underline{\rho}_2, \bar{\rho}_2)$	with early default regulation, $\theta = 1.1$				
		no health shock	with health shock & scenario 2			
			$m_1 = 0.01$	$m_1 = 0.03$	$m_1 = 0.05$	$m_1 = 0.07$
(0, 0)	(0, 0.03)	89.6619	89.659	89.6534	89.6482	89.6433
	(0, 0.3)		89.659	89.6534	89.6482	89.6433
	(0, ∞)		89.659	89.6534	89.6482	89.6433
(0, 0.03)	(0, 0.3)	89.6619	89.659	89.6534	89.6482	89.6433
	(0, ∞)		89.659	89.6534	89.6482	89.6433
(0, 0.3)	(0, ∞)	89.6619	89.659	89.6534	89.6482	89.6433
(0.03, 0.03)	(0.03, 0.3)	87.9197	87.934	87.9602	87.9838	88.005
	(0.03, ∞)		87.934	87.9602	87.9838	88.005
(0.03, 0.3)	(0.03, ∞)	87.9197	87.934	87.9602	87.9838	88.005
(0.3, 0.3)	(0.3, ∞)	83.2948	83.3209	83.3698	83.4148	83.4563

Table 4.12: *Contract values for different rationality levels in the normal state represented by $(\underline{\rho}_1, \bar{\rho}_1)$ and in imparied state represented by $(\underline{\rho}_2, \bar{\rho}_2)$ in Scenrario 2 and different levels represented by m_1 in the situation when the regulator imposes an early default mechanism with $\theta = 1.1$.*

account for pricing. Additionally, as we have explained in Section 4.4.1.1 that with the regulator's protection, policyholders with $(\underline{\rho}_2, \bar{\rho}_2) = (0, 0.3)$ in the impaired state do actually behave like the fully rational policyholder, so they also dislike the increasing damage level of the health shock and their contract values are decreasing with the health damage level. Therefore, under the mild protection of the regulator, even though all the different types of policyholders are still in an favorable position, there are more policyholders who are concerned about the damage level of the health shock.

However, as the regulatory rule becomes stricter, see Table 4.11, we observe the disadvantage of overlooking the health shock while pricing the participating contract by the insurance company for some types of policyholders. Fair contracts' values for the policyholder with $(\underline{\rho}_1, \bar{\rho}_1) = (0, 0)$ in the normal state and $(\underline{\rho}_2, \bar{\rho}_2) = (0, 0.03)$ or $(\underline{\rho}_2, \bar{\rho}_2) = (0, 0.3)$ in the impaired state, and the policyholder with $(\underline{\rho}_1, \bar{\rho}_1) = (0, 0.03)$ in the normal state and $(\underline{\rho}_2, \bar{\rho}_2) = (0, 0.3)$ in the impaired state are lower than the prices which are computed with no health shock and paid to the insurance company at some damage levels represented by m_1. As the regulatory threshold θ increases from 0.7 to 0.9, more assets are required to be kept in the company so that the policyholder is more intensively protected from the downside risk of the company's asset process. Under such intensive protection, the improvement of the policyholder's financial rationality after the systemic health shock plays a less significant role in enhancing his contract's value. But still, a higher mortality in the impaired state lowers the policyholder's chance to participate in the company's long-run profits, which consequently lowers his contract value. So under the regulator's intensive protection, the policyholder's contract value becomes very sensitive to the trade-off between the increase in the policyholder's financial rationality given by $\bar{\rho}_2 - \bar{\rho}_1$ and the damage level of the health shock represented by m_1. For example, the contract value for the policyholder who has no exogenous surrender reasons and only a marginal improvement in his endogenous surrender intensity after the health shock, i.e., $(\underline{\rho}_1, \bar{\rho}_1) = (0, 0)$, $(\underline{\rho}_2, \bar{\rho}_2) = (0, 0.03)$, is lower than the contract value without taking into account the health shock. In addition, for policyholders whose surrender intensities switch from $(\underline{\rho}_1, \bar{\rho}_1) = (0, 0)$ or $(\underline{\rho}_1, \bar{\rho}_1) = (0, 0.03)$ in the normal state to $(\underline{\rho}_2, \bar{\rho}_2) = (0, 0.3)$ in the impaired state, their contract values are higher than the values obtained without taking into account the shock if the health damage level is relatively low, i.e., $m_1 = 0.01$, but are lower when $m_1 \geq 0.03$. To sum up, in such the moderately intensive regulatory environment, policyholders' financial positions turn to be sensitive to the magnitude of the health shock and the changes in the policyholders' surrender behaviors. Some policyholders are being overly charged by the insurance company and this overpricing scenario is going to enlarged as the regulatory rule is further strengthened with $\theta = 1.1$, see Table 4.12. We see that the contracts' values for the policyholders who do not surrender their contracts for exogenous reasons are lower than when the health shock is not considered, no matter how financially rational the policyholders become after experiencing the health shock. As we have mentioned in the beginning of this section that with the over protection of the regulator the policyholder would not surrender his contract on his own, the increase in his endogenous surrender intensity does not effectively bring positive effects on the contract valuation. Therefore, as the health shock is taken into account for pricing, the contract value drops consequently. Moreover, same as in Scenario 1, the difference in the contract's value before and after the health shock is considered for pricing is smallest for all the different types of policyholders when the regulation is imposed harshly by the regulator. Lastly, from Table 4.9-4.12 we see that policyholders with non-zero exogenous surrender intensities benefit from the company's pricing strategy which does not take into account the systemic health shock within all the different regulatory frameworks, which will draw the attention of the insurance company to policyholders' financial capability.

4.4.2 Effects of Regulatory Frameworks on Contract Valuation

Suppose now the insurance company is aware of the existence of the systemic health shock and takes into account the shock while pricing the contract, where m_1 has the benchmark value 0.05. So, in this section, we analyze effects of different regulatory frameworks on the contract valuation given different surrender reactions of the policyholder after experiencing the health shock. First, we present contract values when the policyholder's surrender decision making is not influenced by the health shock, i.e., $(\underline{\rho}_1, \bar{\rho}_1) = (\underline{\rho}_2, \bar{\rho}_2)$, in Table 4.13. We observe that as the regulation is imposed and moderately strengthened, see columns 3-5, the contract value increases for all the different types of policyholders. And the increase in the contract value becomes less significant as the policyholder is more financially rational. Since it is more likely that financially more rational policyholders carry out an optimal surrender strategy on their own,benefits of the regulator's protection

$(\underline{\rho}_1, \bar{\rho}_1)$	$(\underline{\rho}_2, \bar{\rho}_2)$	with health shock ($m_1 = 0.05$) & without change in the surrender behavior			
		no early default regulation	with early default regulation		
			$\theta = 0.7$	$\theta = 0.9$	$\theta = 1.1$
(0, 0)	(0, 0)	85.7512	86.8797	90.4385	89.6482
(0, 0.03)	(0, 0.03)	86.1486	87.1146	90.4531	89.6482
(0, 0.3)	(0, 0.3)	88.1574	88.4576	90.5682	89.6482
(0, ∞)	(0, ∞)	91.9759	91.976	91.9838	89.6482
(0.03, 0.03)	(0.03, 0.03)	82.1913	83.1796	86.8114	87.9838
(0.03, 0.3)	(0.03, 0.3)	84.466	84.7561	86.9649	87.9838
(0.03, ∞)	(0.03, ∞)	88.6293	88.6294	88.6338	87.9838
(0.3, 0.3)	(0.3, 0.3)	75.8138	76.0924	78.3234	83.4148
(0.3, ∞)	(0.3, ∞)	80.75	80.75	80.75	83.415

Table 4.13: *Contract values for different multipliers θ and different rationality levels represented by $(\underline{\rho}_i, \bar{\rho}_i)$, $i = 1, 2$.*

become lower. However, as the regulation parameter θ increases from 0.9 to 1.1, we see that the contract value for the policyholder with $(\underline{\rho}_i, \bar{\rho}_i) = (0, \cdot)$, who does not surrender his contract for exogenous reasons, and $(\underline{\rho}_i, \bar{\rho}_i) = (0.03, \infty)$, who has a very low exogenous surrender intensity but an infinitely large endogenous surrender intensity, becomes smaller. Too much intervention from the regulator starts hurting the policyholder by depriving him of the chance to keep his contract and actively participate in the favorable development of the company. Hence, strengthening the early default regulatory rule is not always favorable, which is consistent with Cheng and Li (2015) where participating contracts are evaluated in a model with no systemic health shock. By comparing our results to Cheng and Li (2015), we notice that assuming the health shock but no changes in the policyholder's surrender behavior after the health shock does not bring more complexities to the effects of the regulatory frameworks on the contract valuation.

Suppose now as described in Scenario 1, the policyholder becomes more impatient after experiencing the health shock and surrenders his contract more likely for personal non-avoidable reasons, which means $\underline{\rho}_2 > \underline{\rho}_1$. We present the contract values for all the different types of policyholders within different regulatory frameworks in Table 4.14. Same as in Table 4.13, we observe that introducing the early default regulatory rule and moderately enhancing its protection level do improve the position of the policyholder with all the different rationality levels. However, an interesting feature is that the effect of an over-regulation rule on the contract valuation depends on the policyholder's surrender reaction after experiencing the health shock. Without knowing the change in the policyholder's surrender behavior, the effect of further strengthening the early default regulatory rule becomes unclear. For example, for the policyholder with $(\underline{\rho}_1, \bar{\rho}_1) = (0, 0.3)$ and $(\underline{\rho}_1, \bar{\rho}_1) = (0, \infty)$ in the normal state, his contract value can either decrease or increase as the default multiplier increases from 0.9 to 1.1 depending on how impatient the policyholder becomes after experiencing the health shock. If the policyholder's exogenous surrender intensity increases to 0.3, the over-regulation does further improve his position. However, if there is only a small change in the policyholder's exogenous surrender-decision making, i.e., $\underline{\rho}_2 = 0.03$, there is a decrease in his contract value. We know that even though the over-regulation rule lowers the chance of the policyholder to participate in the company's profits in the long run, it protects the policyholder from getting punished when he terminates the contract due to exogenous reasons by liquidating the company and paying out the early default benefits. Hence, as the policyholder hurries to surrender his contract for liquidity reasons after having the health damage, it is better for the policyholder to have the over-regulation rule, by applying which the

$(\underline{\rho}_1, \bar{\rho}_1)$	$(\underline{\rho}_2, \bar{\rho}_2)$	with health shock ($m_1 = 0.05$) & with the change in the surrender behavior (Scenario 1)			
		no early default regulation	with early default regulation		
			$\theta = 0.7$	$\theta = 0.9$	$\theta = 1.1$
(0, 0.03)	(0.03, 0.03)	84.8861	85.8711	89.3655	89.1869
(0, 0.3)	(0.03, 0.03)	87.0107	87.2123	89.4953	89.1869
	(0.3, 0.3)	82.5373	82.8429	85.2511	87.3136
(0, ∞)	(0.03, ∞)	91.0188	91.019	91.0255	89.1869
	(0.3, ∞)	87.2001	87.2002	87.2038	87.3136
(0.03, 0.3)	(0.3, 0.3)	80.5667	80.8557	83.2115	86.3071
(0.03, ∞)	(0.3, ∞)	85.2352	85.2353	85.2379	86.3071

Table 4.14: *Contract values for different multipliers θ and different rationality levels represented by $(\underline{\rho}_i, \bar{\rho}_i)$, $i = 1, 2$ which are consistent with the description in Scenario 1.*

policyholder receives early default benefits that are higher than the surrender benefits. The policyholder with $(\underline{\rho}_1, \bar{\rho}_1) = (0.03, \infty)$ in the normal state also belongs to this group. Overall, depending on the changes in the policyholders' surrender behaviors after the health shock, the over-regulation rule can be welcomed by more policyholders.

Suppose, being different from Scenario 1, policyholders become financially more rational after experiencing the health shock as described in Scenario 2. We present contract values for different types of policyholders within different regulatory frameworks in Table 4.15. We observe the same positive effects of imposing the early default regulatory rule and moderately strengthening it on the contract valuation as in Table 4.13 and Table 4.14, i.e., the contract value increases as the regulatory rule with $\theta = 0.7$ is introduced and the regulation threshold θ increases to 0.9. However, as θ increases further to 1.1, with which the company is going to be liquidated even when it has enough assets to cover the promised minimum guarantee, such the over-regulation rule is not welcomed by the policyholders who do not surrender the contract for exogenous reasons. It has the same effect of the over regulation on the contract valuation as in Table 4.13. However, since in Scenario 2 the policyholder is becoming financially more rational in handling his contract after the health shock, such the over-protection does block him from participating in the company's profits to a further extent. Hence, we observe that the decrease of the contract value for the policyholder with $(0, \cdot)$ in Table 4.15 is larger than that in Table 4.13 where there is no improvement in the policyholder's financial rationality. It is also the reason for the policyholder, who has a non-zero exogenous surrender intensity, that his contract value in Scenario 2 does not increase as much as in the case where there is no improve-

$(\underline{\rho}_1, \bar{\rho}_1)$	$(\underline{\rho}_2, \bar{\rho}_2)$	with health shock ($m_1 = 0.05$) & with the change in the surrender behavior (Scenario 2)			
		no early default regulation	with early default regulation		
			$\theta = 0.7$	$\theta = 0.9$	$\theta = 1.1$
(0, 0)	(0, 0.03)	85.8368	86.929	90.4421	89.6482
	(0, 0.3)	86.3212	87.2334	90.4705	89.6482
	(0, ∞)	87.4454	88.1785	90.7859	89.6482
(0, 0.03)	(0, 0.3)	86.6145	87.411	90.4815	89.6482
	(0, ∞)	87.6956	88.3304	90.7958	89.6482
(0, 0.3)	(0, ∞)	89.0182	89.2336	90.8743	89.6482
(0.03, 0.03)	(0.03, 0.3)	82.6769	83.4931	86.8429	87.9838
	(0.03, ∞)	83.766	84.4193	87.1621	87.9838
(0.03, 0.3)	(0.03, ∞)	85.2985	85.5089	87.2722	87.9838
(0.3, 0.3)	(0.3, ∞)	76.4778	76.6816	78.5401	83.4148

Table 4.15: *Contract values for different multipliers θ and different rationality levels represented by $(\underline{\rho}_i, \bar{\rho}_2)$, $i = 1, 2$ which are consistent with the description in Scenario 2.*

ment in his surrender decision making, see Table 4.13. The more financially rational the policyholder becomes after experiencing the health shock, the more the policyholder dislikes the over-protection of the regulator.

4.5 Conclusion

In this paper we have modelled the connection between the mortality risk and surrender risk trigged by the systemic health shock. We take into account the partial rationality of a representative policyholder and assume his surrender intensity is bounded from below and from above due to exogenous surrender reasons and restricted evaluating capacity respectively. By modelling the systemic health shock by a jump process, we construct a continuous two-state Markov chain to capture both the state-dependent mortality intensity and the state-dependent surrender intensity. When the systemic health shock occurs, the representative policyholder's mortality intensity increases, and, at the same time, either the policyholder becomes more impatient in accessing surrender values in Scenario 1 so that his lower bound surrender intensity increases, or the policyholder becomes more careful in surrendering the contract in Scenario 2 so that his upper bound surrender intensity increases. The two scenarios are analyzed both in a regulation-free environment and in an environment with an early default mechanism imposed by a regulator. The regulator monitors the company's financial performance and closes the company when its asset value drops below a prespecified threshold according to the early default regulatory rule.

We have derived a coupled PDE system for pricing the contract and solved it numerically with the finite difference method. In order to easily discuss our numerical results, we suppose that the insurer charges the policyholder the price which is determined when the health shock is overlooked. We quantify the disadvantage on the policyholder with all different levels of rationality, who is more likely to surrender his contract for exogenous reasons after experiencing the health shock, i.e., in Scenario 1. The early default regulation lowers the magnitude of the disadvantage, but it does not help eliminate the disadvantage on the policyholder. Differently, if the policyholder's financial rationality increases after he becomes impaired, he takes advantage of the insurer by paying the price which is lower than the fair contract value when the regulatory rule is absent or weakly imposed. However, such the advantage disappears for some policyholders when the early default regulation becomes stricter. All policyholders who do not surrender the contract for exogenous reasons pay a price which is higher than the contract fair value when the regulation is overly imposed. Recognizing the link between the mortality risk and surrender risk, and the policyholder's surrender change is important for both the policyholder and the insurer. It is also important for the regulator to design an protective early default mechanism.

This paper can be extended further. For example, instead of modelling an increase in the policyholder's mortality intensity over the rest of the contract life as the health shock occurs, first, one can consider a mean reversion process. If the policyholder slowly recovers from the necessary medical treatments, the increase in the mortality intensity vanishes slowly and his mortality intensity may converge to the predicted intensity level before the contract ends. Second, since the policyholder's health damage may accumulate as the health shock exists over the rest of the contract term, one could model the increase in the mortality intensity to be a time-dependent function. No matter which modelling of the mortality intensity, the change in the policyholder's surrender behavior can be linked to the deviation to the expected mortality level so that a continuous surrender intensity change is constructed.

4.6 Appendix to Chapter 4

4.6.1 Proof of Proposition 5

Proof of Proposition 5 in Scenario 1. We follow the proof of Proposition 3 in Chapter 3 for a similar comparative study. Suppose pre-death-and-surrender value functions $v_1(t, A_t)$ and $v_2(t, A_t)$ are solutions to the system of the two PDEs in Proposition 4 with bounds of $(\underline{\rho}_1, \bar{\rho}_1)$ at state $i = 1$, $(\underline{\rho}_2, \bar{\rho}_2)$ at state $i = 2$, and $\bar{\rho}_1 = \bar{\rho}_2$ in Scenario 1, $v_1(t, A_t)$ and $v_2(t, A_t)$ satisfy the two PDEs (4.21) and (4.22) with terminal conditions $v_1(T, A_T) = v_2(T, A_T) = \Phi(A_T)$ and boundary conditions $v_1(t, A_t) = v_2(t, A_t) = \Upsilon(t, A_t)$ for $A_t \leq B_t$. Suppose pre-death-and-surrender value functions $\tilde{v}_1(t, A_t)$ and $\tilde{v}_2(t, A_t)$ are solutions to the system of the two PDEs in Proposition 4 but with bounds of $(\tilde{\underline{\rho}}_1, \tilde{\bar{\rho}}_1)$ at state $i = 1$, $(\tilde{\underline{\rho}}_2, \tilde{\bar{\rho}}_2)$ at state $i = 2$, and $\tilde{\bar{\rho}}_1 = \tilde{\bar{\rho}}_2$ in Scenario 1, $\tilde{v}_1(t, A_t)$ and $\tilde{v}_2(t, A_t)$ satisfy the two PDEs (4.21) and (4.22) with terminal conditions $\tilde{v}_1(T, A_T) = \tilde{v}_2(T, A_T) = \Phi(A_T)$ and boundary conditions $\tilde{v}_1(t, A_t) = \tilde{v}_2(t, A_t) = \Upsilon(t, A_t)$ for $A_t \leq B_t$. Assume $\tilde{\underline{\rho}}_1 \leq \underline{\rho}_1$, $\tilde{\underline{\rho}}_2 \leq \underline{\rho}_2$ and $\tilde{\bar{\rho}}_1 \geq \bar{\rho}_1$. Define $z_1(t, A_t) = \tilde{v}_1(t, A_t) - v_1(t, A_t)$ and $z_2(t, A_t) = \tilde{v}_2(t, A_t) - v_2(t, A_t)$. It follows directly that $z_1(T, A_T) = \tilde{v}_1(T, A_T) - v_1(T, A_T) = \Phi(A_T) - \Phi(A_T) = 0$ and $z_2(T, A_T) = \tilde{v}_2(T, A_T) - v_2(T, A_T) = \Phi(A_T) - \Phi(A_T) = 0$. Also, $z_1(t, A_t) = \Upsilon(t, A_t) - \Upsilon(t, A_t) = 0$ and $z_2(t, A_t) = \Upsilon(t, A_t) - \Upsilon(t, A_t) = 0$ for $A_t \leq B_t$. To obtain the dynamics of z_1 and z_2 take the difference of the two PDEs describing $\tilde{v}_1$ and $\tilde{v}_2$ and the two PDEs describing v_1 and v_2, respectively, i.e.,

$$\begin{aligned}
0 &= \mathcal{L}\tilde{v}_1(t, A_t) + \Psi(t, A_t)\mu_1(t) + S(t, A_t)\tilde{\gamma}(t, A_t, e_1) \\
&\quad -(r(t) + \mu_1(t) + \tilde{\gamma}(t, A_t, e_1))\,\tilde{v}_1(t, A_t) + \kappa(\tilde{v}_2(t, A_t) - \tilde{v}_1(t, A_t)) \\
&\quad - (\mathcal{L}v_1(t, A_t) + \Psi(t, A_t)\mu_1(t) + S(t, A_t)\gamma(t, A_t, e_1) \\
&\quad -(r(t) + \mu_1(t) + \gamma(t, A_t, e_1))\,v_1(t, A_t) + \kappa(v_2(t, A_t) - v_1(t, A_t))) \\
&= \mathcal{L}z_1(t, A_t) + (\tilde{\gamma}(t, A_t, e_1) - \gamma(t, A_t, e_1))\,(S(t, A_t) - \tilde{v}_1(t, A_t)) + \kappa z_2(t, A_t) \\
&\quad -(r(t) + \mu_1(t) + \gamma(t, A_t, e_1) + \kappa)\,z_1(t, A_t)\,;
\end{aligned}$$

$$\begin{aligned}
0 &= \mathcal{L}\tilde{v}_2(t, A_t) + \Psi(t, A_t)\mu_2(t) + S(t, A_t)\tilde{\gamma}(t, A_t, e_2) - (r(t) + \mu_2(t) + \tilde{\gamma}(t, A_t, e_2))\,\tilde{v}_2(t, A_t) \\
&\quad - (\mathcal{L}v_2(t, A_t) + \Psi(t, A_t)\mu_2(t) + S(t, A_t)\gamma(t, A_t, e_2) - (r(t) + \mu_2(t) + \gamma(t, A_t, e_2))\,v_2(t, A_t)) \\
&= \mathcal{L}z_2(t, A_t) + (\tilde{\gamma}(t, A_t, e_2) - \gamma(t, A_t, e_2))\,(S(t, A_t) - \tilde{v}_2(t, A_t)) \\
&\quad -(r(t) + \mu_2(t) + \gamma(t, A_t, e_2))\,z_2(t, A_t)\,,
\end{aligned}$$

where $\gamma(t, A_t, e_i)$ and $\tilde{\gamma}(t, A_t, e_i)$, $i = 1, 2$, are given by (4.25) using the appropriate bounds mentioned above. Similar to the proof of Corollary 1 in Chapter 3, we obtain the stochastic representation of z_1 and z_2 as follows

$$\begin{aligned} z_1(t,A) &= \mathbb{E}_{\mathbb{Q}}^{t,A}\left[\int_t^{\tau_b\wedge T} e^{-\int_t^m (r(u)+\mu_1(u)+\gamma(u,A_u,e_1)+\kappa)\,\mathrm{d}u}((\tilde{\gamma}(m,A_m,e_1)-\gamma(m,A_m,e_1))\,(S(m,A_m) \right. \\ & \qquad \left. -\tilde{v}_1(m,A_m)) + \kappa z_2(m,A_m))\mathrm{d}m \middle| \mathcal{F}_t\right]; \qquad (4.26) \\ z_2(t,A) &= \mathbb{E}_{\mathbb{Q}}^{t,A}\left[\int_t^{\tau_b\wedge T} e^{-\int_t^m (r(u)+\mu_2(u)+\gamma(u,A_u,e_2))\,\mathrm{d}u}(\tilde{\gamma}(m,A_m,e_2)-\gamma(m,A_m,e_2))\,(S(m,A_m) \right. \\ & \qquad \left. -\tilde{v}_2(m,A_m))\mathrm{d}m \middle| \mathcal{F}_t\right], \qquad (4.27) \end{aligned}$$

where $\mathbb{E}_{\mathbb{Q}}^{t,A}$ denotes the expectation conditioned on $A_t = A$. From the definition of $\tilde{\gamma}(t, A, e_i)$, $i = 1, 2$ in (4.25) and assumptions $\tilde{\underline{\rho}}_1 \leq \underline{\rho}_1$, $\tilde{\underline{\rho}}_2 \leq \underline{\rho}_2$ and $\tilde{\bar{\rho}}_1 \geq \bar{\rho}_1$, we see that if $S(t, A) - \tilde{v}_2(t, A) \geq 0$ we have $\tilde{\gamma}(t, A, e_2) = \tilde{\bar{\rho}}_2 = \tilde{\bar{\rho}}_1 \geq \bar{\rho}_1 = \bar{\rho}_2 \geq \gamma(t, A, e_2)$ in Scenario 1 and thus $\tilde{\gamma}(t, A, e_2) - \gamma(t, A, e_2) \geq 0$. On the other hand, if $S(t, A) - \tilde{v}_2(t, A) < 0$ we have $\tilde{\gamma}(t, A, e_2) = \tilde{\underline{\rho}}_2 \leq \underline{\rho}_2 \leq \gamma(t, A, e_2)$. Hence, $\tilde{\gamma}(m, A_m, e_2) - \gamma(m, A_m, e_2) \leq 0$. In sum, the integrand in (4.27) is nonnegative and therefore $z_2(t, A) \geq 0$. Since $z_2(t, A) = \tilde{v}_2(t, A) - v_2(t, A)$, we have $\tilde{v}_2(t, A) \geq v_2(t, A)$. Same as above proving $z_2(t, A) \geq 0$ in (4.27), we obtain that the integrand in (4.26), i.e., $(\tilde{\gamma}(t, A, e_1) - \gamma(t, A, e_1))\,(S(t, A - \tilde{v}_1(t, A)) + \kappa z_2(t, A)$, is also nonnegative given $z_2(t, A) \geq 0$ and therefore $z_1(t, A) \geq 0$. Since $z_1(t, A) = \tilde{v}_1(t, A) - v_1(t, A)$, we have $\tilde{v}_1(t, A) \geq v_1(t, A)$. Proposition 5 in Scenario 2 can be proved in a similar way.

Bibliography

Acharya, S., and Dreyfus, J. F.; 1989: Optimal bank reorganization policies and the pricing of federal deposit insurance, *The Journal of Finance*, 44(5): 1313-1333.

Anderson, H.R., Limb, E.S., Bland, J.M., de Leon, A.P., Strachan, D.P., and Bower, J.S.; 1995: Health Effects of an Air Pollution Episode in Long, December 1991, *Thorax*, 50(11): 1188-1193.

Andreatta, G. and Corradin, S.; 2003: Fair Value of Life Liabilities with Embedded Options: an Application to a Portfolio of Italian Insurance Policies, Working Paper.

Armstrong, J.; 2004: What is the Funding Status of Corporate Defined-Benefit Pensions Plans in Canada?, *Financial System Review (June)*: 45-52.

Bacinello, A. R.; 2003: Pricing Guaranteed Life Insurance Participating Policies with Annual Premiums and Surrender Option, *North American Actuarial Journal*, 7(3): 1-17.

Bacinello, A.R.; 2005: Endogeneous Model of Surrender Conditions in Equity-Linked Life Insurance, *Insurance: Mathematics and Economics*, 37: 270-296.

Bacinello, A.R., Biffis, E. and Millossovich, P.; 2010: Regression-based algorithms for life insurance contracts with surrender guarantees, *Quantitative Finance*, 10(9): 1077-1090.

Bauer, D., Kiesel, R., Kling, A., and Ruß, J.; 2006: Risk-neutral Valuation of Participating Life Insurance Contracts, *Insurance:Mathematics and Economics*, 39(2): 171-183.

Bernard, C., Le Courtois, O. and Quittard-Pinon, F.; 2005: Market value of life insurance contracts under stochastic interest rates and default risk, *Insurance: Mathematics and Economics*: 499-516.

Bielecki, T. and Rutkowski, M.; 2004: Credit Risk: Modeling, Valuation and Hedging, Springer, Berlin.

Biffis, E.; 2005: Affine Processes for Dynamic Mortality and Actuarial Valuations, *Insurance: Mathematics and Economics*, 37: 443-468.

Biffis, E., Denuit, M. and Devolder, P.; 2010: Stochastic Mortality Under Measure Changes, *Scandinavian Actuarial Journal*, 2010(4): 284-311.

Bodie, Z., Light, J. O., Morck, R., and Taggart, Jr., R. A.; 1987: Funding and asset allocation in corporate pension plans: An empirical investigation, Issues in Pension Economics, University of Chicago Press, Chicago, III.

Briys, E. and de Varenne, F.; 1994: Life Insurance in a Contingent Claim Framework: Pricing and Regulatory Implications, *The Geneva Papers on Risk and Insurance Theory*, 19(1): 53-72.

Broeders, D., and Chen, A.; 2013: Pension Benefit Security: A Comparison of Solvency Requirements, A Pension Guarantee Fund and Sponsor Support, *Journal of Risk and Insurance*, 80(2): 239-272.

Brown, J. R.; 2008: Guaranteed Trouble: The Economic Effects of the Pension Benefit Guaranty Corporation, *Journal of Economic Perspectives*, 22(1): 177-198.

Buffington, J. and Elliott, R.J.; 2002: American Options with Regime Switching, *International Journal of Theoretical and Applied Finance*, 5(5): 497-514.

CEIOPS; 2009: CEIOPS' Advice for Level 2 Implementing Measures on Solvency II: Standard Formula SCR-Article 109c Life Underwriting Risk, URL: `https://eiopa.europa.eu/fileadmin/tx_dam/files/consultations/consultationpapers/CP49/CEIOPS-L2-Final-Advice-on-Standard-Formula-Life-underwriting-risk.pdf`, visited on 31.12.2013.

Cerchiara, R. R., Edwards, M., and Gambini, A.; 2008: Generalized Linear Models in Life Insurance: Decrements and Risk Factor Analysis Under Solvency II, Working paper.

Chen, A. and Suchanecki, M.; 2007: Default risk, Bankruptcy Procedures and the Market Value of Life Insurance Liabilities, *Insurance: Mathematics and Economics*, 40: 231-255.

Chen, A.; 2011: A Risk-based Model for the Valuation of Pension Insurance, *Insurance: Mathematics and Economics*, 49(3): 401-409.

Cheng, C. and Uzelac, F.; 2015: A Termination Rule for Pension Guarantee Funds, Working paper.

Cheng, C. and Li, J.; 2015: Early Default Risk and Surrender Risk: Impacts on Participating Life Insurance Policies, Working paper.

Cheng, C.; 2015: Linking Surrender Risk to Mortality Risk: Does the Systemic Health Shock Matter?, Working paper.

Chetty, R.; 2006: A New Method of Estimating Risk Aversion, *The American Economic Review*, 96(5): 1821-1834.

Cooper, R. W., and Ross, T. W.; 2002: Pensions: Theories of Underfunding, *Labour Economics*, 8: 667-689.

Cooper, R. W., and Ross, T. W.; 2003: Protecting Underfunded Pensions: The Role of Guarantee Funds, *Journal of Pension Economics and Finance*, 2(3): 247-272.

Le Courtois, O. and Nakagawa, H.; 2013: On Surrender and Default Risks, *Mathematical Finance*, 23(1): 143-168.

Dahl, M.; 2004: Stochastic Mortality in Life Insurance: Market Reserves and Mortality-linked Insurance Contracts, *Insurance: Mathematics and Economics*, 35: 113-136.

Dahl, M. and Møller, T.; 2006: Valuation and Hedging of Life Insurance Liabilities with Systematic Mortality Risk, *Insurance: Mathematics and Economics*, 39: 193-217.

Dai, M., Kwok, Y.K, and You, H.; 2007: Intensity-based Framework and Penalty Formulation of Optimal Stopping Problems, *Journal of Economic Dynamics & Control*, 31: 3860-3880.

Dar, A., and Dodds, C.; 1989: Interest Rates, the Emergency Fund Hypothesis and Saving through Endowment Policies: Some Empirical Evidence for the U.K., *Journal of Risk and Insurance*, 56(3): 415-433.

Doff, R.; 2008: A Critical Analysis of the Solvency II Proposals, *The Geneva Papers*, 33: 193-206.

Doherty, N.A., and Singer, H.J.; 2003a: The Benefits of a Secondary Market for Life Insurance Policies, *Real Property, Probate and Trust Journal*, 38(3): 449-478.

Doherty, N.A., and Singer, H.J.; 2003b: Regulating the Secondary Market for Life Insurance Policies, *Journal of Insurance Regulation*, 21(4): 63-99.

Eling, M. and Kiesenbauer, D.; 2013: What Policy Feature Determine Life Insurance Lapse? An Analysis of the German Market, *Journal of Risk and Insurance*, 81(2): 241-269.

Elliot, R., Aggoun, L. and Morre, J., 1994: Hidden Markov Models-Estimation and Control, Springer: Berlin.

Fajardo, J., Ornelas, J. R. H., and De Farias, A. R.; 2012: Estimating Risk Aversion, Risk-Neutral and Real-World Densities using Brazilian Real Currency Options, *Economia Aplicada* , 16(4): 567-577.

Feller, W.; 1968: An Introduction to Probability Theory and Its Applications, vol. 1, 3rd ed., Wiley.

Fier, S.G. and Liebenberg, A.P.; 2013, Life Insurance Lapse Behavior, *North American Actuarial Journal*, 17(2): 153-167.

Financial Market Supervision Authority (FINMA) Circular 2008/44: Swiss Solvency Test, `http://www.finma.ch/`.

Freidlin, M.; 1985: Functional Integration and Partial Differential Equations, Princeton University Press.

Franzoni, F., and Marin, J. M.; 2006: Pension Plan Funding and Stock Market Efficiency, *The Journal of Finance*, 61(2): 921-956.

Gatzert, N. and Hoermann, G. and Schmeiser H.; 2009: The Impact of the Secondary Market on Life Insurers' Surrender Profits, *The Journal of Risk and Insurance*, 76(4): 887-908.

Gatzert, N. and Wesker, H.; 2011: A Comparative Assessment of Basel II/III and Solvency II, Working Paper.

Giacalone, J.A.; 2001: Analyzing an Emerging Industry: Viatical Transactions and the Secondary Market for Life Insurance Policies, *Southern Business Review*, 27(1): 1-7.

Grosen, A. and Jørgensen, P.L.; 1997: Valuation of Early Exercisable Interest Rate Guarantees, *The Journal of Risk and Insurance*, 64: 481-503.

Grosen, A. and Jørgensen, P.L.; 2000: Fair Valuation of Life Insurance Liabilities: the Impact of Interest Rate Guarantees, Surrender Options, and Bonus Policies, *Insurance: Mathematics and Economics*, 26(1): 37-57.

Grosen, A. and Jørgensen, P.L.; 2002: Life Insurance Liabilities at Market Value: An Analysis of Insolvency Risk, Bonus Policy, and Regulatory Intervention Rules in a Barrier Option Framework, *Journal of Risk and Insurance*, 69(1): 63-91.

Ippolito, R. A.; 1985: The Labor Contract and the True Economic Pension Liabilities, *The American Economic Review*, 75(5): 1031-1043.

Jackwerth, J. C.; 2000: Recovering Risk Aversion from Option Prices and Realized Returns, *The Review of Financial Studies*, 13(2): 433-451.

Jametti, M.; 2008: Underfunding of Defined Benefit Pension Plans and Benefit Guarantee Insurance: An Overview of Theory and Evidence, *Canadian Public Policy/Analyse de Politiques, Special Supplement on Private Pensions and Income Security in Old Age: An Uncertain Future*, 34: 39-45.

Jeanblanc, M., Yor, M., and Chesney, M.; 2009: Mathematical Methods for Financial Markets, 2009th ed., Springer Finance.

Jones, B.L.; 1998: A Model for Analyzing the Impact of Selective Lapsation on Mortality, *North American Actuarial Journal*, 2(1): 79-86.

Jørgensen, P.L.; 2001: Life Insurance Contracts with Embedded Options: Valuation, Risk Management, and Regulation, *Journal of Risk Finance*, 3(1): 19-30.

Kalra, R., and Jain, G.; 1997: A continuous-time model to determine the intervention policy for PBGC, *Journal of Banking and Finance*, 21(8): 1159-1177.

Kiesenbauer, D.; 2011: Main Determinants of Lapse in the German Life Insurance Industry, *North American Actuarial Journal*, 16(1): 52-73.

Kim, C.; 2013: Modeling Surrender and Lapse Rates with Economic Variables, *North American Actuarial Journal*, 9(4): 56-70.

Kuo, W., Tsai, C. and Chen, W.; 2003: An Empirical Study on the Lapse Rate: the Cointegration Approach, *The Journal of Risk and Insurance*, 70(3): 489-508.

Lewis, C. M., and Pennacchi, C. G.; 1994: The value of pension benefit guaranty corporation insurance, *Journal of Money, Credit and Banking*, 26(3): 735-753.

Li, J. and Szimayer, A.; 2011: The Uncertain Mortality Intensity Framework: Pricing and Hedging Unit-Linked Life Insurance Contracts, *Insurance: Mathematics and Economics*, 49(3): 471-486.

Li, J. and Szimayer, A.; 2014: The Effect of Policyholders' Rationality on Unit-linked Life Insurance Contracts with Surrender Guarantees, *Quantitative Finance*, 14(2): 327-342.

Liebenberg, A.P., Carson, J.M., and Dumm, R.E.; 2012: A Dynamic Analysis of the Demand for Life Insurance, *Journal of Risk and Insurance*, 79(3): 619-644.

Martellini, L., and Milhau, V.; 2008: How Costly is Regulatory Short-Termism for Defined Benefit Pension Funds?, EDHEC Working Paper Series 2008.

Milhaud, X., Loisel, S. and Maume-Deschamps, V.; 2011: Surrender Triggers in Life Insurance: What Main Features Affect the Surrender Behavior in a Classical Economic Context, *Bulletin Français d'Actuariat*, 11(22): 5-48.

Nordahl, H. A.; 2008: Valuation of Life Insurance Surrender and Exchange Options, *Insurance: Mathematics and Economics*, 42(3): 909-919.

Norris J.R.; 1997: Markov Chains, Cambridge University Press.

Outreville, J.F.; 1990: Whole-life Insurance Lapse Rates and the Emergency Fund Hypothesis, *Insurance: Mathematics and Economics*, 9(4): 249-255.

Protter, P.E.; 2005: Stochastic Integration and Differential Equations, Springer, Berlin.

Salisbury, D. L.; 1996: ‘Comments’ (on James Pesando’s ’The Government’s Role in Insuring Pensions.’), in: Zvi Bodie, Olivia S. Mitchell, and John A. Turner, eds., *Security Employer-Based Pensions: An International Perspective Philadelphia: Pension Research Council.*

Schmeiser, H., and Wagner, J.; 2011: A Joint Valuation of Premium Payment and Surrender Options in Participating Life Insurance Contracts, Insurance:Mathematics and Economics, 49(3): 580-596.

Selody, J.; 2007: Vulnerabilities in Defined-Benefit Pension Plans, Discussion Paper 2007-3, Bank of Canada, Ottawa.

Severinson, C.; 2008: Accounting for Defined Benefit Plans: An International Comparison of Exchanged-Listed Companies, OECD Working Papers on Insurance and Private Pensions, 23, OECD Publishing.

Shi, Z., and Werker, J. M.; 2012: Short Horizon Regulation for Long-Term Investors, Journal of Banking and Finance, 36(12): 3227-3238.

Siegmann, A.; 2011: Minimal Funding Ratios for Defined-Benefit Pension Funds, Journal of Pension Economics and Finance, 10(3): 417-434.

Sutherland, W.P. and Drivanos, P.O.; 1999: Viatial Settlements: Life Insurance as a Liquid Asset for the Seriously Ill, Journal of Financial Planning 12(5): 74-78.

Szpiro, G. G.; 1986: Measuring Risk Aversion: An Alternative Approach, The Review of Economics and Statistics, 68(1): 156-159.

Uzelac, F. and Szimayer, A.; 2014: Valuation of Equity-Linked Life Insurance Contracts with Surrender Guarantees in a Regime-Switching Rational Expectation Model, Quantitative Finance, 14(2): 357-368.

Vaupel, J.W., Manton, K.G. and Stallard, E.; 1979: The Impact of Heterogeneity in Individual Frailty on The Dynamics of Mortality, Demography, 16(3): 439-454.

Wichmann, H.E., Mueller, W., Allhoff, P., Beckmann, M., Bocter, N., Csicsaky, M.J., Jung, M., Molik, B., and Schoeneberg, G.; 1989: Health Effects During a Smog Episode in West Germany in 1985, Environmental Health Perspectives, 79: 89-99.

Wilcox, D. W.; 2006: Reforming the Defined-Benefit Pension System, Brookings Papers on Economic Activity, 1: 235-304.

Yong, J.; 1997: Relations Among ODEs, PDEs, FSDEs, BSDEs, and FBSDEs, Proceedings of the 36th Conference on Decision & Control, San Diego, California USA.

Yong, J. and Zhou, X.Y.; 1999: Stochastic Controls: Hamiltonian Systems and HJB Equations, Springer, New York.

Zhou, M., He, G., Fan, M., Wang, Z., Liu, Y., Ma, J., Ma, Z., Liu, J., Liu, Y., Wang, L., and Liu, Y.; 2015: Smog Episodes, Fine Particulate Pollution and Mortality in China, Environmental Research, 136: 396-404.

Zvan, R., Forsyth, P.A. and Vetzel K.R.; 1996: Robust Numerical Methods for PDE Models of Asian Options, Technical Report CS-96-28, Department of Computer Science, University of Waterloo.

Zvan, R., Forsyth, P.A. and Vetzal, K.R.; 2000: PDE Methods for Pricing Barrier Options, Journal of Economic Dynamics and Control, 24: 1563-1590.

Zweifel, P. and Auckenthaler, C.; 2008: On the Feasibility of Insurers' Investment Policies, The Journal of Risk and Insurance, 75(1): 193-206.