Rheinland-Pfälzische
Technische Universität
Kaiserslautern
Landau

Equilibrium Premiums and Risk Class Management in Insurance Markets

Marek Oheim

Datum der Disputation:

04. März 2024

Gutachter

Prof. Dr. Jörn Sass
RPTU Kaiserslautern-Landau

Prof. Dr. An Chen
Universität Ulm

Vom Fachbereich Mathematik der Rheinland-Pfälzischen Technischen Universität Kaiserslautern-Landau zur Verleihung des akademischen Grades Doktor der Naturwissenschaften (Doctor rerum naturalium, Dr. rer. nat.) genehmigte Dissertation

DE-386

Abstract

The aim of this thesis is to introduce an equilibrium insurance market model and study its properties and possible applications in risk class management. First, an insurance market model based on an equilibrium approach is developed. Depending on the premium, the insured will choose the amount of coverage they buy in order to maximize their expected utility. The behavior of the insurer in different market regimes is then compared. While the premiums in markets with perfect competition are calculated in order to make no profit at all, insurers try to maximize their margins in a monopolistic market.

In markets modeled in this way several phenomena become evident. Perhaps the most important one is the so-called push-out effect. When customers with different attributes are insured together, insurance might become so expensive for one type of customers that those agents are better off with buying no insurance at all. The push-out effect was already shown for theoretical examples in the literature. We present a comprehensive analysis of the equilibrium insurance market model and the push-out effect for different insurance products such as life, health and disability insurance contracts using real-life data from different sources. In a concluding chapter we formulate indicators when a push-out can be expected and when not.

Machine learning regression approaches such as neural networks have gained vast popularity in recent years. The exponential growth of computing power has enabled larger and more evolved networks that can perform increasingly complex tasks. In our feasibility study about the use of neural networks in the regression of equilibrium insurance premiums it is shown that this regression is quite robust and the risk of overfitting can almost be excluded – as long as the regression is performed on at least a few thousand data points.

Grouping customers of different risk types into contracts is important for the stability and the robustness of an insurance market. This motivates the study of the optimal assignment of risk classes into contracts, also known as rating classes. We provide a theoretical framework that makes use of techniques from different mathematical fields such as non-linear optimization, convex analysis, herding theory, game theory and combinatorics. In addition, we are able to show that the market specifications have a large impact on the optimal allocation of risk classes to contracts by the insurer. However, there does not need to be an optimal risk class assignment for each of these specifications.

To address this issue, we present two different approaches, one more theoretical and another that can easily be implemented in practice. An extension of our model to markets with capacity constraints rounds off the topic and extends the applicability of our approach.

Zusammenfassung

Ziel dieser Arbeit ist die Einführung eines Gleichgewichtsmodells zur Prämienberechnung in Versicherungsmärkten und dessen Eigenschaften sowie mögliche Anwendungen auf das Risikoklassenmanagement zu untersuchen.
Zunächst wird ein Versicherungsmarktmodell entwickelt, das auf einem Gleichgewichtsansatz basiert. Abhängig von der Prämie wählen die Versicherten die Höhe der Versicherung die sie kaufen, um ihren erwarteten Nutzen zu maximieren. Anschließend wird das Verhalten der Versicherer in verschiedenen Marktregimen verglichen. Während die Prämien in Märkten mit vollständigem Wettbewerb so kalkuliert werden, dass sie keinerlei Gewinn machen, versuchen die Versicherer in einem monopolistischen Markt ihre Gewinnmargen zu maximieren.
In den so modellierten Märkten zeigen sich verschiedene Phänomene. Das vielleicht wichtigste ist der so genannte Push-Out-Effekt. Wenn Kunden mit unterschiedlichen Merkmalen gemeinsam versichert werden, kann die Versicherung für eine Art von Kunden so teuer werden, dass es für sie besser ist, überhaupt keine Versicherung abzuschließen. Der Push-Out-Effekt wurde in der Literatur bereits in theoretischen Beispielen gezeigt. Wir präsentieren eine umfassende Analyse von Versicherungsmärkten mit Gleichgewichtsmodells und untersuchen den Push-Out-Effekts für verschiedene Versicherungsprodukte wie Lebens-, Kranken- und Berufsunfähigkeitsversicherungen unter Verwendung realer Daten aus verschiedenen Quellen. Darüber hinaus formulieren wir Indikatoren, wann ein Push-Out zu erwarten ist und wann nicht.
Regressionsansätze des maschinellen Lernens wie neuronale Netze haben in den letzten Jahren stark an Popularität gewonnen. Das exponentielle Wachstum der Rechenleistung ermöglicht größere und besser entwickelte Netzwerke, die immer komplexere Aufgaben erfüllen können. In unserer Machbarkeitsstudie über den Einsatz neuronaler Netze bei der Regression von Versicherungsprämien in unserem Gleichgewichtsmodell zeigt sich, dass diese Regression recht robust ist und das Risiko von Overfitting annähernd ausgeschlossen werden kann – zumindest solange man einige tausend Datenpunkte für die Regression berechnet.
Die Gruppierung von Kunden unterschiedlicher Risikotypen in Verträgen ist wichtig für die Stabilität und Robustheit eines Versicherungsmarktes. Dies motiviert die Untersuchung der optimalen Zuordnung von Risikoklassen zu Verträ- gen, auch bekannt als Ratingklassen. Wir liefern ein theoretisches Rahmenwerk, das sich Techniken aus verschiedenen mathematischen Bereichen wie nicht-linearer Optimierung, konvexer Analysis, Herdenverhalten, Spieltheorie und Kombinatorik zunutze macht. Darüber hinaus können wir zeigen, dass die Marktspezifikationen einen großen Einfluss auf die optimale Zuordnung der Risikoklassen in die Verträge durch den Versicherer haben. Es lässt sich zeigen, dass es nicht für jede dieser Spezifikationen eine optimale Risikoklassenzuordnung geben muss.
Um dieses Problem zu beheben, stellen wir zwei verschiedene Ansätze vor, einen eher theoretischen und einen weiteren, der in der Praxis leicht umgesetzt werden kann. Eine Erweiterung unseres Modells auf Märkte mit Kapazitätsbeschränkungen rundet das Thema ab und erweitern den praktischen Nutzen unseres Ansatzes.

Acknowledgments

First of all, I want to express my sincere gratitude to Prof. Dr. Jörn Sass for giving me the opportunity to conduct my doctoral studies under his supervision. His comments and ideas on my thesis as well as on various talks combined with many inspiring discussions were an invaluable help during my research process. Our cooperation regarding the Financial Engineering distance learning program, where I worked as a teaching assistant, was fruitful and productive in every way. When I had questions about my research and teaching his door was open for me at all times.

Special thanks go to Dr. Jürgen Bierbaum from Alte Leipziger Hallesche Versicherung for introducing me to some present day problems and open questions in real-life insurance markets which greatly inspired the direction which my research took. His comments and subsequent discussions on my thesis were of extraordinary help, especially for the applicability in insurance practice.

I also want to thank the actuarial group of the ARAG Krankenversicherung which provided me with insights of private health insurance during an internship.

Furthermore, I would like to acknowledge my current and former colleagues from my research group at RPTU and ITWM for their advice regarding my thesis and their moral support. Likewise, I want to thank numerous researchers for interesting discussions and valuable suggestions during various conferences.

Finally, I want to express my gratitude to my mother Elga, my cousin Mirjam and my good friends Ajla, Maxi and Ricarda for proofreading my thesis and making numerous suggestions to improve the mathematical and stylistic quality of my thesis.

Marek Oheim

Contents

1 Introduction

Motivation and Research Goals

For thousands of years, people have used insurances or insurance-like contracts to share and pool risks. There is evidence of Greek marine loans and Italian burial insurances, see [Pru15]. The first known insurance contract dates back to 1347 in Genoa. In the 16th and 17th century, the first insurance companies were founded. They operated in the field of fire insurances. The *Hamburger Feuerkasse* (Hamburg fire fund) was founded in 1676 and is the oldest insurance company that still exists today. The *Equitable Life Assurance Society*, founded in 1765, was the first insurance company to use mathematical methods for pricing insurances. Over the years, insurance markets have grown continuously, leading to new methods in pricing and increasingly diverse insurance markets with ever more complex contracts.

Also the legal basis of insurance markets has changed over time. For example, since 2009 the EU Solvency II Directive [Eur09] has been regulating the capital requirements for insurance companies in order to minimize the risk of insolvency.

New regulations on the pricing of (life) insurance products heralded a paradigm shift. On March 1, 2011 the European Court of Justice ruled that insurers could no longer differentiate between the genders of their policyholders by setting up different premiums for men and women [Eur11]. This new law came into force on December 21, 2012, and is a consequence of the 2004 EU Gender Equality Directive [Eur04]. The ruling made the classic pricing for (life) insurance products no longer directly applicable, since life insurance premiums are calculated based on survival and death probabilities which are strongly related to gender. Under the same circumstances, women are generally more likely to survive a given date than men. This led to different premiums for men and women, which is easily recognizable.

Sass and Seifried (2014) [SS14] examine the impact of this regulation on the premiums and the general market behavior and conclude that mandatory unisex tariffs could reduce the market welfare. This reduction arises from the effect that some customers might be priced out of the insurance markets, also known as a market push-out. Rothschild and Stiglitz showed as early as 1976 [RS76] that if the insurer is not able to observe some characteristics of the insured, the low-risk customers might be worse off than if the high-risk customers would be absent. This motivates our first research goal: the question whether these push-out effects can be observed in real-life settings, such as life, health or disability insurance markets. In addition, it is of particular interest to formulate indicators when a push-out can be expected and when not.

To approach this question, we first need to introduce an insurance market model. In this thesis, we make use of an equilibrium market model. The idea behind this type of models is that in almost all market specifications, the price the insurer sets affects the quantity of insurance bought which vice versa has an influence on the price of insurance. The concept of equilibriums in insurance markets gained popularity due to a paper presented by Rothschild and Stiglitz [RS76]. Their work on equilibriums in insurance markets with price-quantity competition became one of the most important papers on actuarial science in the last century. It is shown that under incomplete information this equilibrium might fail to exist. Their concepts have been extended in several papers in the literature. An example is Sass and Seifried [SS14]. They present an equilibrium insurance market model under price and price-quantity competition, compare their results with those in [RS76] and perform some calculations on the welfare of the market. We focus on generalizing this equilibrium market model under price competition so that it is able to handle complex insurance products from different insurance markets, such as life, health or disability insurance. Then, this model can be applied to real-life data to analyze the effects in these insurance markets.

Calculating premiums using this equilibrium approach can become quite time-consuming if the contract duration is long. This raises the question whether machine learning techniques can help us regress premiums from a smaller set of data points. Of course, the parameter specifications have a large impact on the quality of training. Thus, as a second research goal, we are interested in conducting a feasibility study on different machine learning methods along with their parameter specifications.

Insurance products and markets are becoming increasingly complex. In practice it seems reasonable to divide the set of customers into risk classes, where all customers in one class have (almost) the same risk. We have seen in real-life insurance markets that the division of customers into these classes became more distinguished over the years. In addition, grouping customers of different risk classes into rating classes which subsume these customers in contracts gained importance. It is crucial for insurers to do this grouping properly in order to ensure the feasibility of the insurance contracts while deriving policies that are attractive to as many customers as possible. This motivates the third major goal of this thesis: We are interested in finding a way to model this risk class management problem and to decide which assignment of risk classes into contracts is the best. To this end, a characterization of the term “better” must be developed in this setting. Afterwards, we aim for finding an algorithm to perform this optimization task. Note that it is not clear whether an optimal allocation of risk coalitions exists for each market specification.

Tools from different fields of mathematics are needed to formulate and solve this problem. First, our problem formulation is connected to convex optimization problems. This research field in the area of mathematical optimization is well studied. Various books such as the one of Bertsekas

(1999) [Ber99] provide us with a large number of tools to tackle these problems. Since the model depends on the behavior of customers in a market, results from herding theory can find application in our model as well. In 2011, Irle *et al.* [IKLM11] use a Markov chain approach to model behavioral agent-based models. Although they focus on statistical regularities of financial returns, their concepts and results can be applied to our setting. In the 1960s and 1970s, game theory and its various applications gained popularity. Borch has explored the use of game theoretical methods in insurance. Of particular relevance to our market scenario is his 1962 paper [Bor62], in which he presents an application of game theory to several problems in automobile insurance. When it comes to finding the number of possible allocations of customers into contracts, combinatorial arguments such as they can be found in Bell (1934) [Bel34], Dobinski (1877) [Dob77] or Berend and Tassa (2010) [BT10] are useful.

Finally, we are also interested in insurance markets with capacity constraints. These constraints can apply due to market regulations or strategic decisions made by the insurer. One market regulation that finds application here is the EU Solvency II Directive [Eur09] mentioned earlier. This regulation forces insurers to hold risk capital. Even if most insurers hold twice the required amount or even more securities, the amount of risk capital is of course finite. In consequence, the insurance company might want some branches not to grow too large.

It is also interesting to consider markets with capacity constraints because insurance policies are not traded like commodities. Instead, customers must consult an intermediary before signing their contract. These intermediaries may be on the payroll of an insurance company or work for independent agencies. Of course, intermediaries must be paid and there is a limit to the number of contracts they can process in a given amount of time. This provides a second motivation for analyzing constrained markets.

Related Literature and Data Sources

Basic definitions and concepts of insurance mathematics are developed in Olivieri and Pitacco (2011) [OP11]. The insurance market model which we use is based on Sass and Seifried (2014) [SS14] which is in line with Rothschild and Stiglitz (1976) [RS76]. It has already been used in the Master's thesis of Oheim (2020) [Ohe20].

As mentioned before, the paper of Rothschild and Stiglitz can be considered as one of the most important papers on insurance mathematics from the last century. This motivated several authors to extend their model. Wilson (1977) [Wil77] provides a similar analysis as Rothschild and Stiglitz by introducing two new types of equilibriums and analyzing how they behave in different market scenarios. Crocker and Snow (1986) [CS86] extend the analysis of [RS76] and determine the effect of categorization in markets with incomplete information. They conclude that in some market scenarios costless discrimination always increases efficiency. Wambach (2000) [Wam00] augments the model of Rothschild and Stiglitz by adding unobservable wealth and analyzes how this affects the pooling of the different risk and wealth types. Dubey and Geanakoplos (2002) [DG02] modify the analysis of Rothschild and Stiglitz by developing a model of competitive pooling.

In 1970, Akerlof [Ake70] published a paper analyzing so-called markets for "Lemons". Since this paper analyzes car markets under uncertainty, a "Lemon" is not a fruit but a used car of low quality. He shows that asymmetrical information can cause the market to fail.

Hoy and Polborn (2000) [HP00] study the effect of genetic information in the life insurance market.

Schmeiser *et al.* (2014) [SSW14] consider the impact of mandatory unisex tariffs on consumer perceptions as well as their market implications.

In the 19th century Gompertz (1824) [Gom24] and Makeham (1860) [Mak60] began to provide foundations for mortality modeling. These models have evolved over the years. We use the Lee-Carter mortality model (1992) [LC92] in this thesis. Their model along with various model extensions is widely used in insurance practice.

One of the reasons the Lee-Carter model has become so popular is that it is able to deal with the longevity risk, i.e. with a change in the distribution of mortality. Chen *et al.* (2022) [CLS22] approach the longevity risk differently by introducing so-called collective longevity swaps to distribute this risk between the insurers and a reinsurer.

Patton *et al.* (2009) [PCS+09] study global patterns of mortality among young people. Casiglia *et al.* (1993) [CSG+93] describe the mortality for people aged 80 and above, Gavrilov and Gavrilova (2011) [GG11] present results on mortality measurement at advanced ages and the late-life mortality deceleration, a phenomenon that describes the behavior of death probabilities at very old ages.

Chen and Vigna (2017) [CV17] present a new stochastic approach on modeling unisex mortality to comply with the EU Gender Directive [Eur04, Eur11]. To do so, they find a way to mix the gender-specific mortality probabilities.

In some cases, it might be useful to model the interest rates using an interest rate model rather than assuming them to be constant. For our calculations, we use the CIR model named after its developers Cox, Ingersoll and Ross (1985) [CIR85]. A detailed analysis of a broad variety of interest rate models can be found in Brigo and Mercurio (2007) [BM07].

There is quite a number of papers working on the pricing of classic life insurance products such as (pure) endowment and term insurances or pension products, see for example Aase and Persson (1994) [AP94] or Young and Zariphopoulou (2002) [YZ02]. Bacinello *et al.* (2009) [BBM09] price life insurance products which have an early exercise option.

A life insurance product which is about to gain further popularity is the ton-

tine, which can be described as a hybrid between an annuity and a mortality lottery. Several papers can be found on this topic, e.g. Sabin (2010) [Sab10] or more recently Chen and Rach (2023) [CR23].
As mentioned earlier, insurers are required to hold sufficient amounts of risk capital based on the EU Solvency II Directive [Eur09]. This implies that insures must be particularly careful in managing their balance sheet. Boot and Thakor (1991) [BT91] provide research about capital regulations and off-balance sheet liabilities. Berdin and Gründl (2015) [BG15] examine the impact of low interest rate environments on life insurers. Diehl *et al.* (2022) [DHRS22] present an application of an asset-liability management model and analyze its influence on the long-term stability of a life insurer's balance sheet.
Géron (2019) [Gé19] is a well-known introduction to machine learning methods such as neural networks. We refer to this book for detailed explanations of state-of-the-art machine learning tools.
In Hainaut (2018) [Hai18] these methods are applied to predict mortality. Other (less mathematical) papers from clinical research like Simpson *et al.* (2015) [SLC$^+$15] or Lee *et al.* (2018) [LHG$^+$18] also use neural networks to predict mortality probabilities. Another use case of neural networks in the insurance industry can be found in works such as Yeo *et al.* (2001) [YSWB01] and Yu *et al.* (2021) [YGL$^+$21] who use neural networks for forecasting claim sizes or (changes in) insurance premiums.
For basic definitions of health insurance products and related actuarial techniques, Pitacco (2014) [Pit14] is used as a reference. Riedel (2006) [Rie06] shows how unisex tariffs can be calculated in health insurance and examines the effect of premium refund systems on the equilibriums.
Haberman and Pitacco (1999) [HP99] provide foundations about actuarial models and methods in disability insurance. Niemeyer (2015) [Nie15] presents an analysis of disability insurance markets using safety margins.
For our modeling the use of real-life data is of particular interest. The mortality data we use is French Data that originates from the Human Mortality Database (HMD)[1]. For analyzing health insurance policies we use health expenditure data from the German Robert Koch Institute [Rob17a, Rob17b] and the German Federal Supervisory Authority[2]. Finally, the disability insurance data originates from the social security administration of the U.S. government[3]. More details about the data sources can be found in Appendix B at the end of the thesis.

[1] `https://mortality.org/`, visited November 2020
[2] Bundesanstalt für Finanzdienstleitungsaufsicht (BaFin)`https://www.bafin.de/DE/PublikationenDaten/Statistiken/PKV/wahrscheinlichkeitstafeln_node.html`, visited May 2022
[3] `https://www.ssa.gov/oact/NOTES/ran6/`, visited August 2021

Outline

The structure of this thesis is outlined below. First, Chapter 2 establishes a basic insurance market model. As introduced earlier, the core idea of the model is to calculate premiums on an equilibrium approach. Different market settings and regulatory regimes are investigated.
In Chapter 3 the insurance market model is refined to make it suitable for life insurance products. After building a mortality and an interest rate model, the market effects for various life insurance and pension products are examined. The chapter concludes with the introduction of so-called mixing parameters. On one hand these parameters can be used to gain a deeper understanding of the underlying market effects. On the other hand, they help developing an approach for using the type-specific premium model to calculate aggregate premiums. After a conceptualization, we provide different real-life examples. Furthermore, we compare the results in [CV17] with our approaches of finding a mixing parameter for unisex premiums.
Computations can get quite time-consuming for contracts with long durations. In Chapter 4 it is therefore analyzed how various regression techniques, particularly neural networks, can be applied to the life insurance products from Chapter 3. In addition, a feasibility study on the parameter choices of the different models and approaches is presented.
The key part of this thesis is Chapter 5. Instead of limiting to only two types of customers, the model is extended in order to study an arbitrary number of differently risky customers. All customers with the same risk are grouped into what is called a risk class. Insurers are now free to assign risk classes to rating classes, with all customers in one rating class being provided with the same contract. Depending on the risk class allocations chosen by the various insurance companies, customers can decide for the company from which they wish to purchase insurance. The target of an insurer is assumed to be attracting as many customers as possible in its company. After establishing a theoretical framework, an algorithm to determine the optimal risk allocation strategy is presented. We examine markets and their phenomena when the algorithm is applied to a selection of real-life examples. In both, theoretical examples and in practice, an optimal risk allocation might fail to exist. Two different approaches are presented to mitigate this problem. An extension of our model to markets with capacity constraints, i.e. markets where the amount of insurance that can be sold by the insurers is limited, rounds up the topic and the applicability of our approach.
In Chapter 6, we apply our basic insurance market model from Chapter 2 to health insurance markets. The specialty about these markets is that the insured cannot choose the amount of coverage they wish to purchase. Instead, our model is redesigned so that the insurance price now effects the amount of customers buying insurance from a company or heading of to an other one. The newly derived market model is first applied to modeled data

and in a second step to real-life data of German health insurers.
Afterwards, in Chapter 7, different disability insurance products are investigated. After getting disabled, customers can recover and also fall back into disability. The impact of this peculiarity is highlighted in different numerical simulations using the disability insurance data from the social security administration of the U.S. government.
The analysis of the push-out effect is one of our main research objectives and looked at in detail in Chapter 8. Based on the market studies for life, health and disability insurance, we formulate indicators when to expect a push-out and when not. Additional numerical examples are provided to illustrate these rules.
Our work together with our key findings is summarized in a conclusion in Chapter 9.
The notation used in this thesis is summarized in Appendix A. Afterwards, the data sources used in the calculations and numerical examples of this thesis are summed up in Appendix B.

2 Basic Model

We start by defining a mathematical model of an insurance market, which we use in this thesis. The model is a modified version of the model given in [Ohe20], which is in line with Rothschild and Stiglitz (1976) [RS76] and Sass and Seifried (2014) [SS14]. The model is based on an equilibrium principle to find premiums and optimal insurance coverages. We assume knowledge of the basic definitions of life insurance. These can be found in the literature, e.g. in Olivieri and Pitacco (2011) [OP11].
This chapter is based on [Ohe20], especially on Chapters 2 and 4. We start by setting up an insurance market together with an economy in Section 2.1. Before we are able to model insurance products, we need to include the possibility to buy and sell insurance products in our market. This is done in Section 2.2, demand and supply for insurance are discussed in Sections 2.3 and 2.4. We continue by presenting an analysis of our market in Section 2.5. This analysis is substantiated by numerical examples in different market settings. The chapter is concluded by an analysis of the optimal coverage under different utility functions in Section 2.6.

2.1 Insurance Market and the Economy

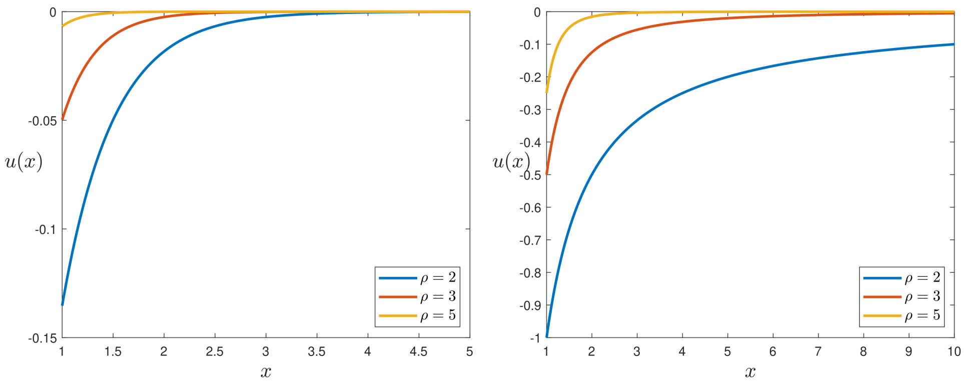
We consider a possibly uncountable set of agents, which we identify as the interval $[0, 1]$. Each agent is represented by one point on the interval. Obviously, in many applications a finite set of agents is sufficient. This can be regarded as a special case of our model. All agents try to maximize their expected utility. To measure the utility, we introduce the concept of a utility function.

Definition 2.1. A real-valued function which is increasing, concave and twice continuously differentiable is called a *utility function*.

The concavity of the utility functions causes that the customers behave risk averse. Contrariwise, insurance companies are modeled risk neutral.
Utility functions can be defined for arbitrary domains. In this thesis we use functions of the form $u\colon]0, \infty[\to \mathbb{R}$ and $u\colon \mathbb{R} \to \mathbb{R}$.

Example 2.2. Prominent examples for utility functions are functions with *constant absolute risk aversion (CARA)*, where the utility function is of the form $u\colon \mathbb{R} \to \mathbb{R}, x \mapsto -e^{-\rho x}$ with a risk aversion parameter $\rho > 0$. To be more precise, a utility function is said to be of constant absolute risk aversion if its *Arrow-Pratt measure of absolute risk aversion* $A(c) = -\frac{u''(c)}{u'(c)}$ is constant in c. One can show that exponential utility (and affine transformations of it) is unique in exhibiting constant absolute risk aversion. One could also regard functions with *constant relative risk aversion (CRRA)*, where the utility function is of the form $u\colon]0, \infty[\to \mathbb{R}, x \mapsto \frac{1}{1-\rho}x^{1-\rho}$ with a

risk aversion parameter $\rho > 0$. Note that the utility function is not defined for $\rho = 1$. In this case we set $u(x) = \ln(x)$ to avoid dividing by zero. In [Ohe20] it is shown that this is a reasonable choice. The class of power utility functions belongs to the class of utility functions which have a constant *Arrow-Pratt measure of relative risk aversion* $R(c) = cA(c) = -\frac{cu''(c)}{u'(c)}$.
For CARA and CRRA functions it is easy to verify that they indeed fulfill the definition of a utility function. The higher the risk aversion parameter ρ is, the faster the absolute value of the utility function is decreasing to zero (for CRRA utility functions we need to assume that $\rho > 1$ to ensure that the function is bounded). This property can be interpreted as follows: A customer with a higher risk aversion parameter is less willing to take risks.



(a) CARA utility (b) CRRA utility

Figure 1: CARA and CRRA utility functions for different risk aversion parameters ρ

In Figure 1, CARA and CRRA functions are plotted for different risk aversion parameters ρ. As mentioned above, a higher value of ρ ensures that the absolute values of the function are smaller and the function converges faster to zero.
In general, the risk aversion parameters of CARA and CRRA utility functions are difficult to compare. As we see later, we are much more interested in the change of utility than in its value itself, therefore this difficulty does not have a big impact on our analysis. Note that CARA functions are specifically useful when we are dealing with losses of unknown or possibly unbounded height, as the function is also able to deal with negative inputs, compare e.g. the modeling of health insurance in Chapter 6.

In our model we equip all agents with a utility function u and an initial wealth $a > 0$. The initial wealth is assumed to be the wealth or amount of money a customer has at his disposal initially, i.e. before damage occurs.
Besides the utility function and the initial wealth, the agents are identical in all but one attribute. Namely, the customers are grouped in two classes

regarding this attribute, the $\oplus$-agents, which constitute a fraction of $w_\oplus$ of the total population and the $\ominus$-agents, which constitute a fraction of $w_\ominus = 1 - w_\oplus$ of the total population.

Remark 2.3. The symbols $\oplus$ and $\ominus$ represent different risk types from the view of the insurer. To be more precise, $\oplus$-agents are supposed to be of lower risk for the insurer than $\ominus$-agents. For example, $\oplus$ could indicate female and $\ominus$ male in the setting of a term insurance, while for a pure endowment insurance $\oplus$ would indicate male and $\ominus$ female, compare Remark 3.6.

Every agent faces an external idiosyncratic risk, where the size and the probability of the damage is determined by the agent's type. Idiosyncratic risks are also called diversifiable or unsystematic risks. With this term we categorize those risks, which can be reduced by the insurer by pooling risks but no cluster risks. To meet this condition we assume that the risk of each customer is independent of the other risks.

Definition 2.4. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Each individual of type $\oplus/\ominus$ faces a *risk* $Z_{\oplus/\ominus}$. The risks are modeled as discrete non-negative random variables, which are independent. The risks of all agents of one risk type are independent and identically distributed with

$$\mathbb{P}(Z_{\oplus/\ominus} = z^k_{\oplus/\ominus}) = p^k_{\oplus/\ominus},$$

with $k \in \mathbb{N}_0$, $z^k_{\oplus/\ominus} \geq 0$ and $p^k_{\oplus/\ominus} \geq 0$ for all k as well as $\sum_{k=0}^{\infty} p^k_{\oplus/\ominus} = 1$. Furthermore we assume $z^k_{\oplus/\ominus} p^k_{\oplus/\ominus} > 0$ for at least one k and that $Z_{\oplus/\ominus}$ is not a.s. constant. The values $z^k_{\oplus/\ominus}$ can be understood as the height of loss in case of damage, which occurs with a probability of $p^k_{\oplus/\ominus}$ for agents of type $\oplus$ and $\ominus$, respectively.

Remark 2.5. Note that $Z_{\oplus/\ominus}$ is not one random variable but a compressed notation for the two random variables $Z_\oplus$ and $Z_\ominus$ which are independent.

Remark 2.6. The risk variables are modeled on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Usually we take discrete sets or $\mathbb{R}$ as the state space and equip it with its standard σ-algebra. We assume that all risk variables in this thesis are (Lebesgue-)integrable.

We assume that the losses given damage $z^k_{\oplus/\ominus}$ and the damage probabilities $p^k_{\oplus/\ominus}$ are known in advance. While [SS14] assumes the risk to be Bernoulli distributed, we extended it to arbitrary discrete distributions.

Remark 2.7. We focus on a one-period setting. At time zero, each agent can decide whether to purchase insurance for a given premium or not. At time one, the losses for each agent are revealed and insurance coverage is paid if the agent has decided to purchase insurance. One could interpret our one-period model also as one with infinitely many periods by revealing the risk probabilities one after another. In this case, whether the agent suffers the loss $z^k_{\oplus/\ominus}$ with probability $p^k_{\oplus/\ominus}$ is revealed at time k.

Remark 2.8. Since $Z_{\oplus/\ominus}$ is used to model the non-deterministic risk an agent faces, respectively the damage an agent suffers, it is reasonable to model it as a non-negative random variable. The advancement $z^k_{\oplus/\ominus} p^k_{\oplus/\ominus} > 0$ for at least one k ensures that the expected loss is non-zero. As mentioned before, $\oplus$ and $\ominus$ denote the risk types, where $\oplus$ indicates good/low and $\ominus$ bad/high risk for the insurer. The easiest approach to decide, whether a type of agent is of good or bad risk, is to compare the expected loss or the net premium of the contracts, which is defined in Definition 2.12. More formally, we assume $\mathbb{E}[Z_\oplus] \leq \mathbb{E}[Z_\ominus]$.

We know that $p_\oplus$ and $p_\ominus$ are stochastic vectors, i.e. their entries are non-negative and sum up to one. Therefore, it is not possible to have $z^k_\oplus \leq z^k_\ominus$ and $p^k_\oplus \leq p^k_\ominus$ for all k, where strict equality holds at least for one k in both equations. This would be closer to the classification in [SS14]. Nonetheless, it is not relevant for the calculations, which group of agents is assigned to which risk class. Instead, we suppose the $\oplus$-agents to have a lower expected loss than the $\ominus$-agents.

Remark 2.9. As mentioned above, in [SS14] it is assumed that the risks are Bernoulli distributed. We revisit two of the examples presented in that paper shortly in Examples 2.29 and 2.31. Most examples in Chapter 5 focus on the case where the risk variables are Bernoulli distributed as well.

2.2 Insurance Contracts

Insurance companies offer *normalized contracts*. This means that the insurance company is willing to cover the whole loss z in case of damage in exchange for a premium $\pi > 0$. Note that π is used as a variable for the insurance premium and should not be confused with the mathematical constant $\pi = 3.14\ldots$. The premium has to be paid in a single payment at the start of the contract. Various papers in the literature such as Chen and Vigna (2017) [CV17] or Aase and Persson (1994) [AP94] analyze markets with periodic premium payments.

An analysis of equilibrium insurance markets under Bernoulli distributed risks is presented in [SS14]. We have already analyzed the effect of periodic premium payments in this market type in [Ohe20].

Customers are allowed to choose the quantity λ of insurance coverage they want to purchase. This means that by paying a premium of $\lambda\pi$, the insurance company is refunding a loss of λz. We usually restrict λ to be non-negative, to rule out short-selling of insurance contracts. The restriction $\lambda \geq 0$ does not rule out overinsurance, i.e. choosing $\lambda > 1$. This is meaningful as the loss z is often chosen as an arbitrary reference point, for example when we analyze life or disability insurance products, see Chapters 3 and 7. Nevertheless, overinsurance is ruled out for many insurance types, like health insurance, which we investigate in Chapter 6. While we restrict λ to be in the interval $[0, 1]$ in these markets, we only require $\lambda \geq 0$ in the general

market setting.
We now consider an insurance contract with premium π and suppose that individuals of type $\oplus/\ominus$ purchase coverage $\lambda_{\oplus/\ominus}$. Be aware that we sometimes omit the agents type if it is not relevant, which type the agent actually has.

Remark 2.10. Note that $\lambda_{\oplus/\ominus}$ is a function depending on π, the agent's initial wealth a and the utility function u. We do not denote the dependency here and write $\lambda_{\oplus/\ominus}$ instead of $\lambda_{\oplus/\ominus}(\pi, a, u)$. Also the equilibrium amount of coverage $\hat{\lambda}_{\oplus/\ominus}$ defined in Definition 2.17 depends on the utility function and in most cases also on the initial wealth, compare Remark 2.36.

Remark 2.11. We could model the initial wealth of the agents as a real-valued random variable and choose the utility function for each agent randomly from a set of possible utility functions. Then, the *average demand* $\tilde{\lambda}(\pi)$ can be defined as

$$\tilde{\lambda}(\pi) := \mathbb{E}[\lambda(\pi, a, u)].$$

The expectation can be understood as the expectation under the joint distribution of the initial wealth and the (risk aversion parameter of) the utility function. When one computes examples for different insurance products one can see that this only complicates the calculations while there are no structural differences in the behavior of the model. All observable effects are qualitatively similar to a model where all agents have the same initial wealth a and utility function u. To ease further computations, we therefore do not use average demands in this thesis but focus on deterministic ones.

Definition 2.12. We define the *net expected value of a contract* as

$$\lambda_{\oplus/\ominus}(\pi) \cdot (\pi - \pi^0_{\oplus/\ominus}),$$

where $\lambda_{\oplus/\ominus}(\pi)$ is the optimal amount of coverage given the premium π and

$$\pi^0_{\oplus/\ominus} := \mathbb{E}[Z_{\oplus/\ominus}] = \sum_{k=0}^{\infty} p^k_{\oplus/\ominus} z^k_{\oplus/\ominus} > 0.$$

The expected value of the risk $\pi^0_{\oplus/\ominus}$ is also called the *net premium.*

Remark 2.13. We restrict the set of possible risks to those with bounded expected loss, i.e. with $\pi^0_{\oplus/\ominus} = \mathbb{E}[Z_{\oplus/\ominus}] < \infty$. If this would not be the case, the expected loss and therefore also the premium would be infinitely large. While the expectation is supposed to be finite, the variance of the risk might be infinite, see Section 6.2 where we use the discretized Pareto distribution with parameters that lead to infinite variance.

Definition 2.14. The insurance contract is said to be *feasible* for the insurer if it does not make expected loss on the aggregate level, i.e. if

$$\mathbb{E}[w_\oplus \lambda_\oplus(\pi - Z_\oplus) + w_\ominus \lambda_\ominus(\pi - Z_\ominus)] = w_\oplus \lambda_\oplus(\pi - \pi^0_\oplus) + w_\ominus \lambda_\ominus(\pi - \pi^0_\ominus) \geq 0.$$

Remark 2.15. Feasibility of an insurance contract does not only depend on the premium charged, but also on the amount of coverage the agents decide to purchase. From $\pi^0_\oplus \leq \pi^0_\ominus$ we know that $\pi - \pi^0_\oplus \geq 0$ is a necessary condition for the feasibility of a contract, as an insurer needs to make profit with at least one type of insured. We might have $\pi - \pi^0_\ominus < 0$ if enough low-risk customers buy the contract and subsidize the high-risk customers with their premium payments.

In our model we assume that an insurance company that sells only feasible contracts will almost never go bankrupt. As we are dealing with idiosyncratic (independent) risks, this can be justified by the strong law of large numbers. Of course, the probability of default is never zero, but can be regarded as so low that it has no practical impact on the choices of the agents or the market behavior and can therefore be ignored for our model. This risk pooling argument is a standard argument in insurance mathematics, see e.g. [OP11]. Later on, in Definition 5.74, we also add a safety loading to the contracts to reduce the default probability even further.

Remark 2.16. Our model describes the setting of *price competition*, where the insurance company is only allowed to specify the price per unit of coverage, but not the amount of coverage an insured is allowed to purchase. Under the market mechanism of *price-quantity competition*, the customer can no longer choose the amount of coverage he wants to purchase. Instead, the insurance company offers contracts with a fixed premium and a fixed coverage. While price competition is common for life and disability insurance products, price-quantity competition can often be found in theft, health or car insurance.
If purchasing multiple insurances is possible or price-quantity competition is ruled out by law, it is reasonable for the insurance company to offer contracts for which the fraction of coverage is linearly scalable. This exact kind of contract was described in this section. We investigate equilibriums under price competition in Section 2.5.

2.3 Demand for Insurance

We recall that each agent is equipped with a utility function u and an initial wealth $a > 0$. Since each $\oplus$- and $\ominus$-agent wants to maximize his utility function, the decision how much insurance to purchase is given by the optimization problem

$$\max_{\lambda_{\oplus/\ominus} \geq 0} \mathbb{E}[u(a - \lambda_{\oplus/\ominus}\pi - (1 - \lambda_{\oplus/\ominus})Z_{\oplus/\ominus})]. \tag{1}$$

Definition 2.17. The maximizer $\hat{\lambda}_{\oplus/\ominus}$ of Equation (1) is called the *equilibrium insurance coverage* of customers of type $\oplus/\ominus$.

Note that the premium π is given by the insurance company, but the insured can choose an arbitrary non-negative amount of coverage they want to purchase. As mentioned in Remark 2.16, this is also known as a market with price competition. If overinsurance is ruled out, we need to restrict our optimization problem so that we maximize over $\lambda_{\oplus/\ominus} \in [0, 1]$.

Remark 2.18. The essential supremum $\text{esssup} Z$ of a random variable Z is defined as the smallest number α such that the set

$$\{\omega : Z(\omega) > \alpha\}$$

has a probability of zero. If Z is discrete, it is the highest value taken by Z with positive probability. Assume that the premium π is greater or equal than this essential supremum of the risk variable Z but $Z < \alpha$ a.s. This implies

$$\arg\max_{\lambda \geq 0} \mathbb{E}[u(a - \lambda\pi - (1 - \lambda)Z)] = 0,$$

as suffering from a loss that is higher than the premium π happens with probability zero. Hence, agents are better of if they just take the risk. This is known as the *no rip-off* property for premium functions in insurance mathematics.

We can reformulate Equation (1) using the discreteness of the random variable $Z_{\oplus/\ominus}$ and obtain the problem

$$\max_{\lambda_{\oplus/\ominus} \geq 0} \sum_{k=0}^{\infty} p^k_{\oplus/\ominus} \cdot u(a - \lambda_{\oplus/\ominus}\pi - (1 - \lambda_{\oplus/\ominus}) z^k_{\oplus/\ominus}). \tag{2}$$

Remark 2.19. It is not clear that this maximizer always needs to exist. We can derive some criterions which ensure the existence of the maximizer. When we have $\pi > \text{essinf} Z$ and the utility function u is bounded from above and satisfies $\lim_{x \to 0} u(x) = -\infty$, a maximizer needs to exist. The conditions on the utility function are satisfied e.g. for CRRA utility with $\rho > 1$ which ensures the existence for most examples in this thesis. In other cases, the maximizer might fail to exist, as it could be the best to buy an infinitely large amount of insurance. Nonetheless, these extreme cases are not relevant in practice, an optimizer exists for all examples in this thesis.

We can then use the first-order condition to find the maximum by solving the equation

$$\sum_{k=0}^{\infty} p^k_{\oplus/\ominus} \cdot u'(a - \hat{\lambda}_{\oplus/\ominus}\pi - (1 - \hat{\lambda}_{\oplus/\ominus}) z^k_{\oplus/\ominus}) \cdot (-\pi + z^k_{\oplus/\ominus}) = 0 \tag{3}$$

in $\hat{\lambda}_{\oplus/\ominus}$. Note that we need to be careful that the solution is non-negative. In the case that the optimal coverage turns out to be negative, we stick to solving the problem in Equation (1) instead.The interchangeability of differentiation and summation can be shown e.g. by Theorem 11.2 in Forster [For12].

Due to the structure of the above equation, we cannot obtain a closed solution for CARA and CRRA utility functions in general. Hence, we have to solve Equation (1) or (3) numerically on the set of non-negative reals.
The expected profit of the insurer is given by the term

$$\mathbb{E}[w_{\oplus}\lambda_{\oplus}(\pi_{\oplus} - Z_{\oplus}) + w_{\ominus}\lambda_{\ominus}(\pi_{\ominus} - Z_{\ominus})],$$

where $w_{\oplus}$ and $w_{\ominus}$ represent the corresponding fractions of $\oplus$- and $\ominus$-agents in the portfolio. As we have already seen in Definition 2.14 we can calculate the expectation explicitly and conclude

$$\mathbb{E}[w_{\oplus}\lambda_{\oplus}(\pi_{\oplus} - Z_{\oplus}) + w_{\ominus}\lambda_{\ominus}(\pi_{\ominus} - Z_{\ominus})] = w_{\oplus}\lambda_{\oplus}(\pi_{\oplus} - \pi^0_{\oplus}) + w_{\ominus}\lambda_{\ominus}(\pi_{\ominus} - \pi^0_{\ominus}). \tag{4}$$

Recall that we set $\pi^0_{\oplus/\ominus} = \mathbb{E}[Z_{\oplus/\ominus}] = \sum_{k=0}^{\infty} p^k_{\oplus/\ominus} z^k_{\oplus/\ominus}$.

Remark 2.20. If $Z_{\oplus/\ominus}$ is Bernoulli distributed, this optimization problem can be solved explicitly using the first-order condition, where we even obtain a closed-form solution for CARA and CRRA utility. Calculating the derivative is in this case very simple, due to the simple distribution of $Z_{\oplus/\ominus}$, see [SS14].

Before continuing our investigation, we show that as long as $\hat{\lambda}$ is strictly positive, it is always differentiable in the premium π. This property eases later computations when it comes to calculating optimal premiums, see Section 2.5. To prove the differentiability, we recap the implicit function theorem.

Proposition 2.21. Let $U, V \subseteq \mathbb{R}$ open and

$$F : U \times V \to \mathbb{R}$$

continuously differentiable. Assume furthermore that $(x_0, y_0) \in U \times V$ satisfies $F(x_0, y_0) = 0$ as well as $\frac{\partial F}{\partial y}(x_0, y_0) \neq 0$. Then there exists an open subset $U_0 \times V_0 \subseteq U \times V$ containing (x_0, y_0) together with a unique continuously differentiable function

$$f : U_0 \to V_0$$

satisfying $f(x_0) = y_0$ such that

$$F(x, y) = 0 \;\Leftrightarrow\; y = f(x)$$

holds for all $(x, y) \in U_0 \times V_0$. Furthermore, the derivative $y' = f'(x)$ is given by

$$y' = f'(x) = -\frac{F_x}{F_y},$$

where F_x and F_y denote the partial derivatives of F satisfying $F_y \neq 0$.

Proof. A proof can be found in Chapter 8 in Forster (2008) [For08]. □

Using this proposition, we can show under very mild conditions that $\hat{\lambda}$ is indeed continuously differentiable in π.

Proposition 2.22. Assume that the utility function u is not only concave but strictly concave. Then $\hat{\lambda}$ given by Equation (1), i.e.

$$\max_{\lambda_{\oplus/\ominus}\geq 0} \mathbb{E}[u(a - \lambda_{\oplus/\ominus}\pi - (1 - \lambda_{\oplus/\ominus})Z_{\oplus/\ominus})]$$

is continuously differentiable in π if $\hat{\lambda} > 0$.

Proof. Recall that using Equation (3) $\hat{\lambda}$ can also be calculated as the solution of

$$\sum_{k=0}^{\infty} p^k \cdot u'(a - \hat{\lambda}\pi - (1 - \hat{\lambda})z^k) \cdot (-\pi + z^k) = 0,$$

as long as $\hat{\lambda} \geq 0$. Here, we dropped the dependence on the agents type to increase the readability. We define $F : \mathbb{R}_{>0} \times \mathbb{R}_{>0} \to \mathbb{R}$ by

$$F(\pi, \hat{\lambda}) = \sum_{k=0}^{\infty} p^k \cdot u'(a - \hat{\lambda}\pi - (1 - \hat{\lambda})z^k) \cdot (-\pi + z^k).$$

In the next step we apply Proposition 2.21 to F.
The function F is a composition of functions that are continuously differentiable in π and $\hat{\lambda}$. As before, summation and differentiation can be interchanged. Hence, F is again continuously differentiable in π and $\hat{\lambda}$.
Furthermore, we calculate

$$\frac{\partial F}{\partial \hat{\lambda}} = \sum_{k=0}^{\infty} p^k \cdot u''(a - \hat{\lambda}\pi - (1 - \hat{\lambda})z^k) \cdot (-\pi + z^k)^2.$$

In order to show that this derivative is non-zero,we make the following observations:

- All p_k are non-negative, where $p_k > 0$ for at least one k.
- The second derivative of the utility function satisfies $u'' < 0$ as u is strictly concave.
- By the non rip-off condition we know that

  $$\pi \leq \sup_k \{z_k : p_k > 0\},$$

  see Remark 2.18. We assume that Z is not a.s. constant, compare Definition 2.4. This implies that the above inequality is strict which yields $(-\pi + z^k)^2 > 0$.

Taking everything together implies $\frac{\partial F}{\partial \hat{\lambda}} \neq 0$ for all $(\pi, \hat{\lambda})$.
Choose a tuple $(\pi_0, \hat{\lambda}_0)$ that satisfies $F(\pi_0, \hat{\lambda}_0) = 0$. By Equation (3) we know that $\hat{\lambda}_0$ is the optimal choice of coverage given a premium π_0. Using Proposition 2.21 we know that there exists a continuously differentiable function $\hat{\lambda} : U_0 \to V_0$ on an open subset $U_0 \subseteq \mathbb{R}_{>0}$ which is continuously differentiable with $\lambda(\pi_0) = \hat{\lambda}_0$ and

$$F(\pi, \hat{\lambda}) = 0 \;\Leftrightarrow\; \hat{\lambda} = \hat{\lambda}(\pi)$$

holds for all $(\pi, \hat{\lambda}) \in U_0 \times V_0$. □

Remark 2.23. We used the strict concavity in Proposition 2.22 to ensure that the derivative $\frac{\partial F}{\partial \hat{\lambda}}$ is non-zero. Both utility functions we use in this thesis, namely the CARA and the CRRA utility function from Example 2.2 are strictly concave. If we would use a utility function that is not strictly concave, the proposition still holds for all points satisfying $\frac{\partial F}{\partial \hat{\lambda}}(\pi, \hat{\lambda}) \neq 0$.

Remark 2.24. In the proof of Proposition 2.22 we assumed that F is defined on $\mathbb{R}_{>0} \times \mathbb{R}_{>0}$. Since $\pi \geq \mathbb{E}[Z] > 0$ the premium has to be positive. We assumed furthermore that there is no short-selling of insurance, i.e. $\hat{\lambda} \geq 0$. As mentioned, we assumed not only that $\hat{\lambda} \geq 0$ but $\hat{\lambda} > 0$ in Proposition 2.22. On one hand, this assumption is of technical nature to ensure that the function is defined on an open set. On the other hand, we can see that the optimal coverage function might indeed not be differentiable in points where it is optimal to buy no coverage. In Example 2.37 we calculate the equilibrium insurance coverage and its derivative as functions of the premium π. It can be seen that the coverage might indeed not be differentiable in zero and the above proposition cannot be applied to this case.

2.4 Supply for Insurance

For the supply side of the insurance market, i.e. the insurance companies, we investigate a market with price competition. We study two different scenarios:

- (M) a monopolistic insurance market;
- (C) a market with perfect competition.

All insurance companies are assumed to be risk neutral. The behavior of the insurer depends on the underlying scenario. While the insurance company maximizes its profit in the monopolistic setting by an optimization problem, this is not possible in a market with perfect competition. In such a market the expected profits on any contract an insurer offers are zero. If a company would make profit, an other company could offer the same products with a smaller profit resulting in a lower price that attracts all customers. This

means that the expected profit as given in Definition 2.14 vanishes.
Besides these two competition scenarios, we consider two different regulatory regimes:

- (E) mandatory equal tariffs (i.e. a ban on discriminative policies);
- (F) free contract design (i.e. no ban on discriminative policies).

In regime (E), insurance companies have to offer the same premium $\pi = \pi_{\odot}$ for all customers. In regime (F), insurance companies are allowed to differ between the different risk classes, in our case denoted by $\oplus$ and $\ominus$, by setting up two different premiums $\pi_{\oplus}$ and $\pi_{\ominus}$, respectively. In both regimes, the insured can choose the amount of coverage they would like to purchase.
The insurance market we currently have in most countries may be considered as almost perfect competitive and is therefore subsumed in scenario (C). Of course, one could argue that firms make profits in real-life insurance markets. Customer choices are not fully rational and the selection of an insurance company is more complex than just looking on prices.
If we interpret the customers of the two risk classes $\oplus$ and $\ominus$ as female and male customers, the ruling of the European Court of Justice [Eur04], [Eur11] changed the market setting from (F) to (E). In this case we speak of unisex premiums in regime (E) and gender-specific premiums in regime (F). As we stay in the general case for the rest of this chapter, we make use of the terms aggregated premium and type-specific premiums instead.

2.5 Equilibrium with Price Competition

In this section we analyze the equilibrium premium and equilibrium insurance per coverage in scenarios (M) and (C). We illustrate our analysis with some examples. The analysis as well as the corresponding examples are based on [SS14].

Monopolistic Insurance Market

We first focus on scenario (M). We know that in this setting the premiums can be determined by a simple optimization problem. As we have seen in Definition 2.14 and Equation (4), the expected payoff for the insurer is given by the term

$$\mathbb{E}[w_{\oplus}\lambda_{\oplus}(\pi_{\oplus} - Z_{\oplus}) + w_{\ominus}\lambda_{\ominus}(\pi_{\ominus} - Z_{\ominus})] = w_{\oplus}\lambda_{\oplus}(\pi_{\oplus} - \pi^0_{\oplus}) + w_{\ominus}\lambda_{\ominus}(\pi_{\ominus} - \pi^0_{\ominus}),$$

which has to be non-negative to satisfy feasibility. Remember that $\pi^0_{\oplus/\ominus}$ denotes the expected value of the loss for the insurer for an agent of type $\oplus/\ominus$, i.e. $\pi^0_{\oplus/\ominus} = \mathbb{E}[Z_{\oplus/\ominus}] = \sum_{k=0}^{\infty} p^k_{\oplus/\ominus} z^k_{\oplus/\ominus}$. In regime (E) the two type-specific premiums $\pi_{\oplus}$ and $\pi_{\ominus}$ are substituted by an aggregated premium $\pi_{\odot}$. Be reminded that the insurer is allowed to determine the premiums

$\pi_{\oplus/\ominus}$ and $\pi_{\odot}$. The amount of coverage $\lambda_{\oplus/\ominus}$ the agents buy are functions of the corresponding premiums, i.e. we have $\lambda_{\oplus/\ominus} = \lambda_{\oplus/\ominus}(\pi_{\oplus/\ominus})$ and $\lambda_{\oplus/\ominus} = \lambda_{\oplus/\ominus}(\pi_{\odot})$, compare Remark 2.10. Remember that $\lambda_{\oplus/\ominus}(\pi_{\odot} - \pi^0_{\oplus/\ominus})$ should be understood as a product of $\lambda_{\oplus/\ominus}(\pi_{\odot})$ and $(\pi_{\odot} - \pi^0_{\oplus/\ominus})$, not as the value of the function $\lambda_{\oplus/\ominus}(\cdot)$ evaluated at $(\pi_{\odot} - \pi^0_{\oplus/\ominus})$.

In regime (E), the *optimal or equilibrium (aggregate) premium* $\hat{\pi}_{\odot}$ is given by

$$\hat{\pi}_{\odot} = \arg\max_{\pi_{\odot}}\{w_{\oplus}\hat{\lambda}_{\oplus}(\pi_{\odot} - \pi^0_{\oplus}) + w_{\ominus}\hat{\lambda}_{\ominus}(\pi_{\odot} - \pi^0_{\ominus})\}, \tag{5}$$

i.e. the insurer tries to maximize its expected payoff. The first-order condition for this equation yields

$$\hat{\pi}_{\odot} = -\frac{w_{\oplus}[-\pi^0_{\oplus}\hat{\lambda}'_{\oplus}(\hat{\pi}_{\odot}) + \hat{\lambda}_{\oplus}(\hat{\pi}_{\odot})] + w_{\ominus}[-\pi^0_{\ominus}\hat{\lambda}'_{\ominus}(\hat{\pi}_{\odot}) + \hat{\lambda}_{\ominus}(\hat{\pi}_{\odot})]}{w_{\oplus}\hat{\lambda}'_{\oplus}(\hat{\pi}_{\odot}) + w_{\ominus}\hat{\lambda}'_{\ominus}(\hat{\pi}_{\odot})}, \tag{6}$$

where we need to have $\hat{\lambda}_{\oplus/\ominus}(\hat{\pi}_{\odot}) \neq 0$ to ensure differentiability, see Proposition 2.22, and $w_{\oplus}\hat{\lambda}'_{\oplus}(\hat{\pi}_{\odot}) + w_{\ominus}\hat{\lambda}'_{\ominus}(\hat{\pi}_{\odot}) \neq 0$ to avoid dividing by zero. For CARA and CRRA utility functions, $\hat{\lambda}_{\oplus/\ominus}(\hat{\pi}_{\odot}) \neq 0$ already implies $\hat{\lambda}'_{\oplus/\ominus}(\hat{\pi}_{\odot}) \neq 0$ and therefore $w_{\oplus}\hat{\lambda}'_{\oplus}(\hat{\pi}_{\odot}) + w_{\ominus}\hat{\lambda}'_{\ominus}(\hat{\pi}_{\odot}) \neq 0$. Analogously in regime (F), the *optimal or equilibrium (type-specific) premiums* $\hat{\pi}_{\oplus}$ and $\hat{\pi}_{\ominus}$ are given by

$$(\hat{\pi}_{\oplus}, \hat{\pi}_{\ominus}) = \arg\max_{(\pi_{\oplus}, \pi_{\ominus})}\{w_{\oplus}\hat{\lambda}_{\oplus}(\pi_{\oplus} - \pi^0_{\oplus}) + w_{\ominus}\hat{\lambda}_{\ominus}(\pi_{\ominus} - \pi^0_{\ominus})\}. \tag{7}$$

In regime (F), the first-order condition delivers a much simpler equation for the optimal premium. The simplification is due to the fact that we can analyze high- and low-risk customers separately, i.e. we have $w_{\ominus} = 1$ and $w_{\oplus} = 0$ or $w_{\ominus} = 0$ and $w_{\oplus} = 1$. We obtain

$$\hat{\pi}_{\oplus/\ominus} = \pi^0_{\oplus/\ominus} - \frac{\hat{\lambda}_{\oplus/\ominus}(\hat{\pi}_{\oplus/\ominus})}{\hat{\lambda}'_{\oplus/\ominus}(\hat{\pi}_{\oplus/\ominus})}. \tag{8}$$

Similarly, we need to have $\hat{\lambda}_{\oplus/\ominus}(\hat{\pi}_{\oplus/\ominus}) \neq 0$ and $\hat{\lambda}'_{\oplus/\ominus}(\hat{\pi}_{\oplus/\ominus}) \neq 0$ to ensure differentiability and avoid dividing by zero.

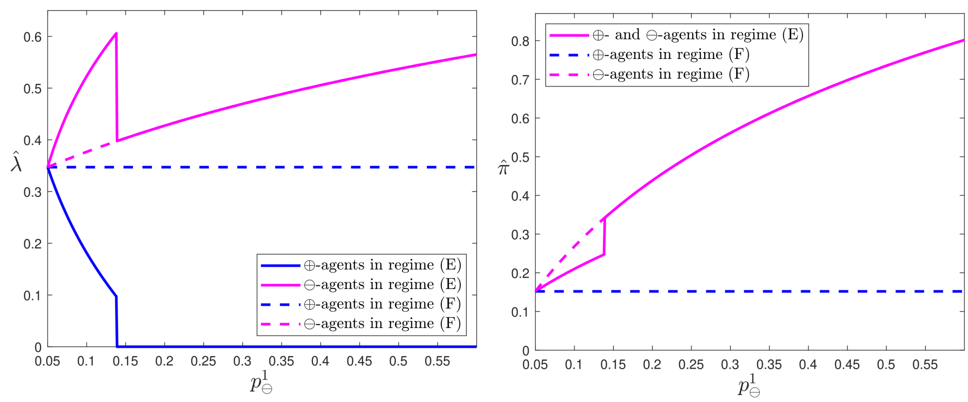
Remark 2.25. The no rip-off condition from Remark 2.18 ensures that the functions we are maximizing over in Equations (5) and (7) have compact support. Using Proposition 2.22 we know that $\hat{\lambda}$ is differentiable as long as it is strictly greater than zero and monotonously decreasing, as its derivative is negative. Furthermore, the intermediate value theorem ensures the for each $\hat{\lambda}$ there exists a π such that $F(\pi, \hat{\lambda}) = 0$, so $\hat{\lambda}(\pi)$ indeed converges continuously to zero. Hence, the functions from Equations (5) and (7) are continuous and bounded on a compact set, so a maximum needs to exist.

Remark 2.26. Note that if we set $w_{\oplus/\ominus}$ to one for one type of agents, Equation (6) simplifies to Equation (8).

Remark 2.27. Proposition 2.22 ensures that $\hat{\lambda}$ is indeed differentiable in all points where it is strictly positive. Note that the proof of this proposition makes use of the implicit function theorem which gives us an opportunity to calculate the derivative, see Proposition 2.21. We compute the formula for the derivative together with some examples in Section 2.6. Unfortunately, $\lambda_{\oplus/\ominus}$ and its derivative can usually not be given in closed form.

Remark 2.28. Due to the restrictions on $\hat{\lambda}$ and its derivative respectively on the denominators of the fractions, we do not use the above representations for the monopolistic case, but stick to the original maximization problem given in Equations (5) and (7) instead.

Example 2.29. In this example, we illustrate the values of the equilibrium insurance premiums and the corresponding equilibrium insurance coverages for concrete values of the parameters. Therefore we assume the risk variables $Z_{\oplus/\ominus}$ to be Bernoulli distributed. We set the loss probabilities to $p^1_\oplus = 5\%$ and $p^0_\oplus = 1 - p^1_\oplus$, vary $p^1_\ominus$ between 5% and 60% and set $p^0_\ominus = 1 - p^1_\ominus$. The damage in case of loss is set to $z^1_\oplus = z^1_\ominus = 1$, $z^0_\oplus = z^0_\ominus = 0$ and all other $p^i_{\oplus/\ominus}$ and $z^i_{\oplus/\ominus}$ to zero. The initial wealth for all agents is set to $a = 2$ and the share of $\oplus$- and $\ominus$-agents in the portfolio to $w_\oplus = w_\ominus = 50\%$. Furthermore, we equip all agents with CRRA utility functions with a risk aversion parameter of $\rho = 3$, i.e. $u(x) = -\frac{1}{2x^2}$.
Now we calculate the premiums $\pi_\odot$, $\pi_\oplus$ and $\pi_\ominus$ for these parameters. As our choice does not admit a closed form, we have to solve Equations (5) and (7), respectively. Recall that these equations depend on the equilibrium insurance coverages $\hat{\lambda}$ which are functions of the premium. To determine the corresponding insurance demands, we plug the optimal premium into the optimal coverage function. The results are plotted in Figure 2 below.



(a) Eq. insurance coverage $\hat{\lambda}_{\oplus/\ominus}$ (b) Eq. premium per coverage $\hat{\pi}_{\oplus/\ominus}$

Figure 2: Equilibrium insurance coverage $\hat{\lambda}_{\oplus/\ominus}$ and equilibrium premium per coverage $\hat{\pi}_{\oplus/\ominus}$ as functions of the high-risk agents loss probability $p^1_\ominus$ in scenario (M)

In Figure 2, we plot the equilibrium insurance coverage and the equilibrium premium per coverage over the value of the high-risk customers loss probability $p^1_\ominus$, while we keep the low-risk customers loss probability constant at $p^1_\oplus = 5\%$. We are interested in investigating the effects of the high-risk agent's damage probability on the behavior of the agents. As the damage probability of the low-risk agents does not change, the effects are based more on the *(relative and absolute) difference* of the probabilities between the two types of agents rather than its actual *value*.
In Figure 2b we can see that the premium per unit of coverage, which can be understood as the relative price of the insurance, rises with an increasing value of $p^1_\ominus$. This should not be surprising, as the expected loss for the insurer ascends with $p^1_\ominus$, which results in higher premiums.
If the damage probability of the $\ominus$-agents exceeds a certain level (for our specifications if $p^1_\ominus \geq 14\%$, can be noted by the jump in the function at that point), the low-risk customers are pushed out of the insurance market. If so, the low-risk customers choose not to buy any insurance, and the high-risk customers end up paying exactly the same as before. Then, the only difference between regimes (E) and (F) is that the low-risk agents do not purchase any insurance in regime (E), as they are driven out of the insurance market. If $p^1_\ominus$ is below the push-out level, low-risk agents subsidize high-risk agents with their premium payments. A higher value of $p^1_\ominus$ causes the aggregate premium to rise, which results in a higher subsidy. This happens until the push-out level is reached.

Remark 2.30. From Figure 2 we can see that there are two local maximums for the insurer's optimization problem in regime (E) formulated in Equation (5). If the high-risk agent's damage probability $p^1_\ominus$ is low enough, we are in the first local maximum (if $p^1_\ominus < 0.14$). Here, the insurance is bought by both types of agents. Buying insurance in the second local maximum (if $p^1_\ominus \geq 0.14$) is only attractive for high-risk customers, as the premium is too high for the low-risk customers. The jump in the equilibrium insurance coverage and equilibrium premium per coverage denotes the change from the first to the second local maximum, which coincides with the point where the low-risk customers are driven out of the insurance market.

Competitive Insurance Supply

In contrast to the monopolistic insurance scenario (M) we now assume that there is perfect competition in the market, i.e. that we are in market scenario (C). In Equation (4), we calculated the profit of the insurer. As noted in Section 2.4, this profit given by

$$\begin{aligned}&\mathbb{E}[w_\oplus \lambda_\oplus(\pi_{\oplus/\odot} - Z_\oplus) + w_\ominus \lambda_\ominus(\pi_{\ominus/\odot} - Z_\ominus)] \\ &= w_\oplus \lambda_\oplus(\pi_{\oplus/\odot} - \pi^0_\oplus) + w_\ominus \lambda_\ominus(\pi_{\ominus/\odot} - \pi^0_\ominus)\end{aligned}$$

must equal zero. Note that the insurer can charge a type-specific premium $\pi_{\oplus/\ominus}$ or an aggregate premium $\pi_{\odot}$ depending on whether the market allows for free contract design or not, i.e. if we are in regime (F) or (E).
We first look at regime (F). The premiums are required to be non-negative and the contracts for the $\oplus$- and the $\ominus$-agents have to be seen as separate contracts, which both need to be feasible. As we are pricing the two groups of customers separately, the $\oplus$-agents cannot subsidize the $\ominus$-agents as described in Remark 2.15. As a result, both contracts need to be feasible. Furthermore, we assume hat the insurer does not set the premiums so high that no coverage is bought. Therefore, the only non-trivial solution to the above equation is given by

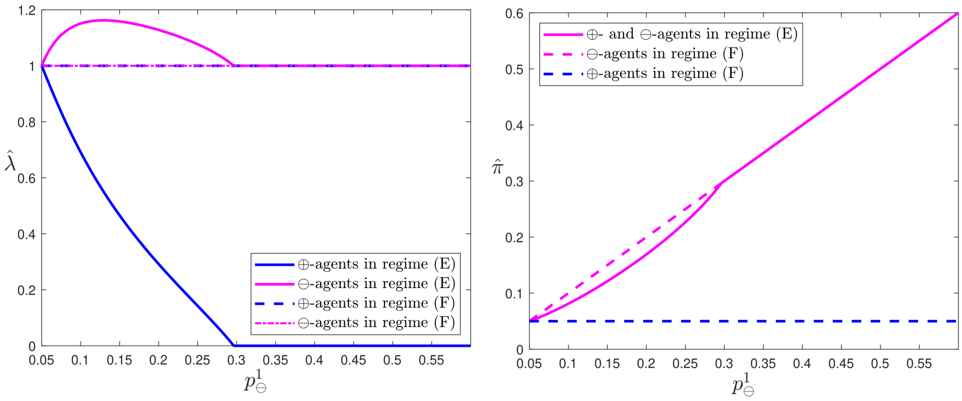
$$\hat{\pi}_{\oplus} = \pi^0_{\oplus}, \quad \hat{\pi}_{\ominus} = \pi^0_{\ominus}. \tag{9}$$

Recall that $\hat{\pi}_{\oplus}$ and $\hat{\pi}_{\ominus}$ denote the optimal type-specific premiums for $\oplus$- and $\ominus$-agents, respectively.
In regime (E), the insurer is not allowed to differ between $\oplus$- and $\ominus$-agents when it sets the premiums. Given the equilibrium insurance demand functions $\hat{\lambda}_{\oplus}$ and $\hat{\lambda}_{\ominus}$, the optimal aggregate premium can be computed as

$$\hat{\pi}_{\odot} = \frac{w_{\oplus}\hat{\lambda}_{\oplus}}{w_{\oplus}\hat{\lambda}_{\oplus} + w_{\ominus}\hat{\lambda}_{\ominus}} \cdot \pi^0_{\oplus} + \frac{w_{\ominus}\hat{\lambda}_{\ominus}}{w_{\oplus}\hat{\lambda}_{\oplus} + w_{\ominus}\hat{\lambda}_{\ominus}} \cdot \pi^0_{\ominus}. \tag{10}$$

Recall that the functions $\hat{\lambda}_{\oplus/\ominus}$ depend on the premium $\hat{\pi}_{\odot}$, compare Remark 2.10. Therefore, the optimal premium $\hat{\pi}_{\odot}$ can be calculated by solving Equation (10) numerically.

Example 2.31. We examine the setting of Example 2.29 again, but this time in scenario (C). We use the same parameters as in Example 2.29: loss probabilities $p^1_{\oplus} = 5\%$, $p^0_{\oplus} = 1 - p^1_{\oplus}$, vary $p^1_{\ominus}$ between 5% and 60%, $p^0_{\ominus} = 1 - p^1_{\ominus}$, damage in case of loss $z^1_{\oplus} = z^1_{\ominus} = 1$, all other $p^i_{\oplus/\ominus}$ and $z^i_{\oplus/\ominus}$ to zero, initial wealth $a = 2$, CRRA utility with $\rho = 3$, equal share of $\oplus$- and $\ominus$-agents.
Based on these parameters, we calculate the equilibrium insurance coverage $\hat{\lambda}_{\oplus/\ominus}$ and the equilibrium premium per coverage $\hat{\pi}_{\oplus/\ominus}$ in the competitive insurance market. The results are given in Figure 3 below.



(a) Eq. insurance coverage $\hat{\lambda}_{\oplus/\ominus}$ (b) Eq. premium per coverage $\hat{\pi}_{\oplus/\ominus}$

Figure 3: Equilibrium insurance coverage $\hat{\lambda}_{\oplus/\ominus}$ and equilibrium premium per coverage $\hat{\pi}_{\oplus/\ominus}$ as functions of the high-risk agents loss probability $p^1_\ominus$ in scenario (C)

Similar to the monopolistic insurance market, the premium per coverage increases with a rising value of $p^1_\ominus$. In competitive insurance supply, the growth is linear for higher values of $p^1_\ominus$ in regime (E) and generally for $\ominus$-agents in regime (F). This is easy to explain as $\hat{\pi}_\ominus$ and $\hat{\pi}_\odot$ in the second local maximum (compare Remark 2.30) are given as $\pi^0_{\oplus/\ominus} = z^1_{\oplus/\ominus} p^1_{\oplus/\ominus}$. Again, low-risk customers are driven out of the insurance market, but in this scenario, the phenomenon turns out to be much weaker. This effect may be explained by the fact that the point where only high-risk customers remain in the portfolio is reached at a higher value of $p^1_\ominus$. In a market with free competition, this critical value is 30% instead of 14% for the monopolistic insurance market. Furthermore, for competitive insurance supply, there is no drop-off, at which the equilibrium insurance coverage for the $\ominus$-agents drops to zero. Instead, the equilibrium insurance coverage decreases continuously to zero in this market. We take a second look on the push-out probabilities in this market in Example 5.79.

If $p^1_\oplus$ and $p^1_\ominus$ are sufficiently close to each other, we can see that $\ominus$-agents buy insurance coverage in excess of their losses. This is due to the fact that overinsurance in this case is a gamble with positive expectation. As the $\oplus$-agents subsidize the $\ominus$-agents with their premium payments, the coverage is affordable. Thus the additional payment in case of loss is higher than the cost due to a higher premium. If overinsurance is prohibited, i.e. the high-risk insurance demand is bounded by $z^1_\ominus$, we can obtain similar results as above, but with the high-risk agents coverage capped to $z^1_\ominus = 1$.

Furthermore, we note that all agents in regime (F) buy full coverage, i.e. we have $\hat{\lambda}_\oplus = \hat{\lambda}_\ominus = 1$. This can be understood quite easily, as the insured solve

$$\max_{\lambda \geq 0} \mathbb{E}[u(a - \lambda\pi - (1-\lambda)Z)]$$

which becomes

$$\max_{\lambda \geq 0} \mathbb{E}[u(a - \lambda \mathbb{E}[Z] - (1 - \lambda)Z)]$$

in our setting. Buying full coverage leaves us with the (fully deterministic) utility $u(a - \mathbb{E}[Z])$. As the premium equals the expected loss, buying more or less coverage just adds variance to the utility. Using Jensen's inequality, the concavity of utility functions shows that with the same expectation this is less favorable than choosing full coverage where we end up in a deterministic setting with zero variance.

Remark 2.32. Again, there are two equilibriums. One, where both types of agents purchase insurance and one, where only high-risk customers choose to purchase insurance. We switch from the first to the second equilibrium at $p^1_{\ominus} = 30\%$.

Remark 2.33. Note that the difference between the equilibrium premium per coverage in scenarios (M) and (C) is very high. For example, $\oplus$-agents in regime (F) pay up to three times more if there is only one monopolistic insurer as they would pay in a market with perfect competition.

Remark 2.34. The plots in Figures 2 and 3 are based on 551 data points for each function. For the figures in the other chapters of this thesis corresponding remarks and explanations can be found in the according chapters, compare Remarks 3.9, 5.11, 6.4, 7.1 and 8.1.

Remark 2.35. The calculations in Examples 2.29, 2.31 and all other examples in Chapters 2, 3, 5 and 8 were implemented in MATLAB using the `fminsearch` optimizer. For the aggregate premium in the competitive scenario this optimizer appears to perform better than MATLAB's equation solving algorithm `solve`. It is easy to see that solving $x = f(x)$ is equivalent to minimizing $|x - f(x)|$ or $(x - f(x))^2$ in x, as long as we make sure that the solution we found indeed fulfills $x = f(x)$.
For the monopolistic market setting, we decided to solve the original maximization problem given by Equations (5) and (7) instead of solving Equations (6) and (8) derived by the first order condition. This is due to the restrictions on the differentiability of $\hat{\lambda}$.

2.6 Optimal Coverage under CARA and CRRA Utility

When it comes to analyzing phenomena in our insurance market model, it is crucial to understand the behavior of the optimal coverage problem defined in Equation (1).
Before providing an analysis of the equilibrium insurance coverage, we illustrate that the equilibrium insurance coverage (and therefore also the premiums) in a market with CARA utility does not depend on the height of the initial wealth a of the agents.

Remark 2.36. In Examples 2.29 and 2.31 we use the CRRA utility function (power utility, $u(x) = \frac{1}{1-\rho}x^{1-\rho}$) defined in Example 2.2. Performing the analysis of the two above examples with an other value for the initial wealth leads to (slightly) different results. This is due to the fact that insurance becomes less important for people with higher wealth in this market setting, compare Figure 65. For CARA utility (exponential utility, $u(x) = -e^{-\rho x}$), this is not the case. As one can already guess from the abbreviation CARA (constant absolute risk aversion, compare the explanation for the CRRA utility function in Example 2.2), the absolute risk aversion given by

$$A(c) = -\frac{u''(c)}{u'(c)} = -\frac{\rho^2 u(c)}{-\rho u(c)} = \rho$$

is constant in x. Using exponential law we get

$$\begin{aligned} &\max_{\lambda_{\oplus/\ominus}\geq 0} \mathbb{E}[u(a' - \lambda_{\oplus/\ominus}\pi - (1-\lambda_{\oplus/\ominus})Z_{\oplus/\ominus})] \\ = u(a'-a) &\max_{\lambda_{\oplus/\ominus}\geq 0} \mathbb{E}[u(a - \lambda_{\oplus/\ominus}\pi - (1-\lambda_{\oplus/\ominus})Z_{\oplus/\ominus})], \end{aligned}$$

where $a, a' \in \mathbb{R}_{\geq 0}$ denote two different values for the initial wealth. Note that $u(a'-a)$ can be taken out of the expectation and maximization, as it does not depend on Z or λ. Hence, the maximization procedure and therefore also its result are independent of the initial wealth.

In Proposition 2.22 we have shown that the equilibrium insurance coverage $\hat{\lambda}$ is differentiable in the premium π as long as $\hat{\lambda}$ is strictly positive. The implicit function theorem from Proposition 2.21 equips us with an explicit way to calculate the coverage. As in the proof of Proposition 2.22 we define $F : \mathbb{R}_{>0} \times \mathbb{R}_{>0} \to \mathbb{R}$ by

$$F(\pi, \hat{\lambda}) = \sum_{k=0}^{\infty} p^k \cdot u'(a - \hat{\lambda}\pi - (1-\hat{\lambda})z^k) \cdot (-\pi + z^k).$$

Using this definition together with Proposition 2.21, the derivative of $\hat{\lambda}$ in π can be calculated as

$$\hat{\lambda}' = \hat{\lambda}'(\pi) = -\frac{F_\pi}{F_{\hat{\lambda}}},$$

where we need to assume that $\hat{\lambda} > 0$. Later in the proof we calculated

$$F_{\hat{\lambda}} = \frac{\partial F}{\partial \hat{\lambda}} = \sum_{k=0}^{\infty} p^k \cdot u''(a - \hat{\lambda}\pi - (1-\hat{\lambda})z^k) \cdot (-\pi + z^k)^2.$$

We furthermore obtain

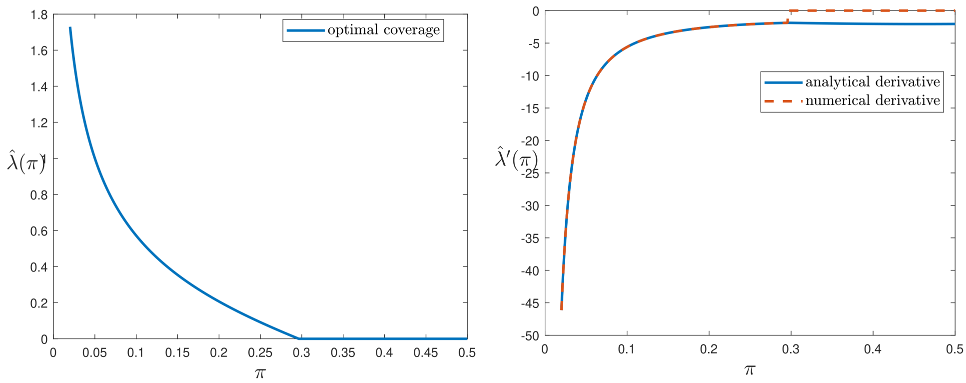
$$F_\pi = \frac{\partial F}{\partial \pi} = \sum_{k=0}^{\infty} p^k \cdot (-u'(a - \hat{\lambda}\pi - (1-\hat{\lambda})z^k) - \hat{\lambda} \cdot u''(a - \hat{\lambda}\pi - (1-\hat{\lambda})z^k) \cdot (-\pi + z^k)).$$

Combining these equations reads

$$\hat{\lambda}'(\pi) = -\frac{\sum_{k=0}^{\infty} p^k \cdot (-u'(a - \hat{\lambda}\pi - (1-\hat{\lambda})z^k) - \hat{\lambda} \cdot u''(a - \hat{\lambda}\pi - (1-\hat{\lambda})z^k) \cdot (-\pi + z^k))}{\sum_{k=0}^{\infty} p^k \cdot u''(a - \hat{\lambda}\pi - (1-\hat{\lambda})z^k) \cdot (-\pi + z^k)^2}. \tag{11}$$

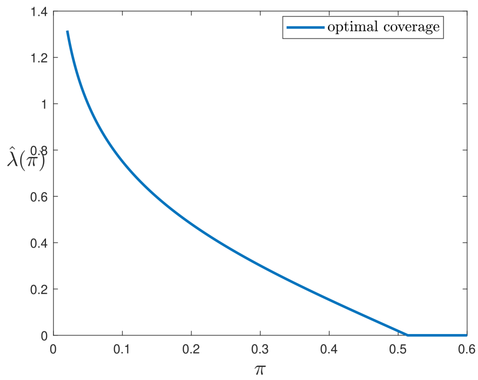
We can use this representation to calculate the equilibrium insurance coverage and its derivative for an example.

Example 2.37. We consider a setting as in Examples 2.29 and 2.31. Different from these examples we assume there is one type of agents with a damage probability of $p = 5\%$ and a loss in case of damage of $z = 1$. All agents are equipped with an initial wealth of $a = 2$. In two different calculations we assume the agents to have CRRA or CARA utility, respectively. In both cases, the risk aversion parameter is set to $\rho = 3$. Based on these parameters, we compute the equilibrium insurance coverage and its derivative using Equation (11). As a comparison we provide the numerical derivative of the coverage using the midpoint rule.

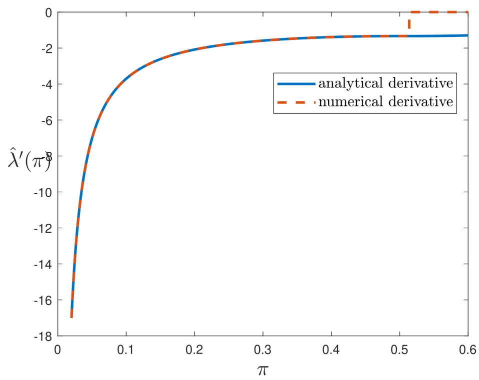


(a) Eq. insurance coverage $\hat{\lambda}_{\oplus/\ominus}$ (b) Derivative of eq. ins. coverage $\hat{\lambda}'_{\oplus/\ominus}$

Figure 4: Equilibrium insurance coverage $\hat{\lambda}$ and its derivative $\hat{\lambda}'$ as functions of the premium π for CRRA utility



(a) Eq. insurance coverage $\hat{\lambda}_{\oplus/\ominus}$



(b) Derivative of eq. ins. coverage $\hat{\lambda}'_{\oplus/\ominus}$

Figure 5: Equilibrium insurance coverage $\hat{\lambda}$ and its derivative $\hat{\lambda}'$ as functions of the premium π for CARA utility

When we compare the values of the analytical derivative from the implicit function theorem with its numerical approximation, we can see that the analytical approach indeed only works if $\hat{\lambda} > 0$. If $\hat{\lambda} = 0$, the derivative is either not defined or zero, but not strictly negative, as the analytical formula would suggest. Hence, we need to pay attention, that the purchased coverage is indeed strictly positive before we differentiate implicitly.

3 Life Insurance

In this chapter, we apply the basic model we have introduced in a theoretical setting in Chapter 2 to real-life data and insurance products. Life insurance policies are more complex than the products from our theoretical examples. This is not an issue as we have defined our basic model for general non-negative discrete risk variables. Nonetheless, some preparatory work is needed before we can price life insurance policies.
We start this chapter by investigating different life insurance products. These policies are defined in Section 3.1. To set up these contracts, we need an appropriate mortality and interest rate model. All these components come together in Section 3.2, where we provide different numerical examples to analyze the parameter dependencies. As stated in Chapter 1, an analysis of the push-out effect is of special interest for our research. We conclude the chapter with Section 3.3, where we define different mixing parameters to be able to compute unisex premiums like gender-specific premiums. The according numerical analysis serves for a deeper understanding of the structure of our model and its application to life insurance policies.
For basic definitions in life insurance mathematics and their related actuarial concepts we refer to Olivieri and Pitacco (2011) [OP11].

3.1 Life Insurance Products

We start this section by introducing four basic life insurance products which we investigate in the following. In this regard, we introduce the Lee-Carter mortality model in Section 3.1.1. The big advantage of using this model is that it is able to deal with the so-called longevity risk, i.e. the risk that the distribution of mortality changes over the years. As the insurance contracts usually need to be paid at the beginning of the contract duration that often lasts several years, an interest rate for discounting the premium payment is necessary. To do so, the CIR interest rate model is introduced in Section 3.1.2.

Example 3.1. A *pure endowment insurance* pays one unit of money after a maturity T if the policyholder survives until then. We set $z^1_{\oplus/\ominus} = B(0,T)$, where $B(0,T)$ is the time-T-discount factor, see Section 3.1.2. As the payment by the insurance is made at the maturity but the premium needs to be paid at the start of the contract, the insurance company can invest the premium in the money market. Also the insured could decide to invest the money in the market instead of buying insurance.
The according damage probability is set to $p^1_{\oplus/\ominus} = {}_Tp_x^{\oplus/\ominus}$, where ${}_Tp_x^{\oplus/\ominus}$ is the time-T-survival probability of an $\oplus/\ominus$-agent with initial age x. All other $z^k_{\oplus/\ominus}$ are set to zero. The net expected value of the contract is given by

$$\pi^0_{\oplus/\ominus} = B(0,T) \cdot {}_Tp_x^{\oplus/\ominus}.$$

Example 3.2. A *term insurance* is an insurance contract that pays out one unit of money at the end of the year if the insured does not survive until maturity T. We therefore set $z^k_{\oplus/\ominus} = B(0, k+1)$ for $k = 0, \dots, T-1$, where $B(0, k+1)$ is the discount factor from zero to $k+1$. We are discounting until $k+1$ instead of k because the insurance is always paid out at the end of the year. Furthermore, $p^k_{\oplus/\ominus} = {}_{k|1}q_x^{\oplus/\ominus}$ holds, where ${}_{k|1}q_x^{\oplus/\ominus}$ is the k-year deferred death probability of an agent of type $\oplus/\ominus$. In other words, ${}_{k|1}q_x^{\oplus/\ominus}$ describes the probability of death during year k (but before year $k+1$). This time period is given by the interval $(k, k+1]$, so ${}_{k|1}q_x^{\oplus/\ominus} = {}_kp_x^{\oplus/\ominus} - {}_{k+1}p_x^{\oplus/\ominus}$. This leads to the net expected loss

$$\pi^0_{\oplus/\ominus} = \sum_{k=0}^{T-1} B(0, k+1) \cdot {}_{k|1}q_x^{\oplus/\ominus}.$$

Example 3.3. An *endowment insurance* is a combination of a term insurance with a pure endowment insurance: it pays out one unit of money at the end of the year if the insured does not survive until maturity T. If the policyholder survives until then, it pays one unit of money after a maturity T. Similar to the term and the pure endowment insurance we set $z^k_{\oplus/\ominus} = B(0, k+1)$ for $k = 0, \dots, T-1$ and $z^T_{\oplus/\ominus} = B(0, T)$. As before, $B(0, k+1)$ is the discount factor from zero to $k+1$. Furthermore, $p^k_{\oplus/\ominus} = {}_{k|1}q_x^{\oplus/\ominus}$ holds and we set $p^T_{\oplus/\ominus} = {}_Tp_x^{\oplus/\ominus}$. This leads to the net expected loss

$$\pi^0_{\oplus/\ominus} = \sum_{k=0}^{T-1} B(0, k+1) \cdot {}_{k|1}q_x^{\oplus/\ominus} + B(0, T) \cdot {}_Tp_x^{\oplus/\ominus}.$$

Remark 3.4. The pure endowment part of the endowment insurance could also be understood as a "money-back guarantee" for the premium of the term insurance for the case that the policy holder survives the contract.

Example 3.5. An *immediate lifetime annuity* is a pension product that pays out one unit of money at the end of each year which the policyholder survives. We have $z^k_{\oplus/\ominus} = \sum_{j=0}^{k-1} B(0, j+1)$ and $p^k_{\oplus/\ominus} = {}_{k|1}q_x^{\oplus/\ominus}$. All other $z^k_{\oplus/\ominus}$ are set to zero. The net expected value of the contract is given by

$$\begin{aligned}\pi^0_{\oplus/\ominus} &= \sum_{k=0}^{\omega-x-1} \left(\sum_{j=0}^{k-1} B(0, j+1)\right) \left({}_kp_x^{\oplus/\ominus} - {}_{k+1}p_x^{\oplus/\ominus}\right) \\ &= \sum_{k=1}^{\omega-x-1} B(0, k) \cdot {}_kp_x^{\oplus/\ominus},\end{aligned}$$

where ω denotes the maximal age (we assume $\omega = 110$).

Remark 3.6. We already mentioned in Remark 2.3 that $\oplus$ and $\ominus$ could indicate males and females. In all of our numerical examples in Section 3.2, we use males and females as the $\oplus$- and $\ominus$-agents. Based on our model, $\oplus$-agents are of lower risk for the insurance company. Given our mortality model from Section 3.1.1 below and positive interest rates (e.g. interest rates modeled by the CIR model from Section 3.1.2), we receive the following correspondences.

	$\oplus$-agents	$\ominus$-agents
pure endowment insurance	males	females
term insurance	females	males
endowment insurance	females	males
immediate lifetime annuity	males	females

Table 1: Correspondence between agent type and gender for four different life insurance products

Remember that we always assumed that the premium is paid in a single payment at the beginning of the contract duration. The endowment insurance is certainly paying out one unit of money, the only question is when the insurance company needs to pay, compare Remark 3.4. Given that our interest rate is positive, it is advantageous for the insurer to delay the payment as long as possible. Hence, customers with lower death probabilities are preferable for the insurer.

As we can see in Section 3.1.1, women are more likely to survive a given time point than men. Hence, the number of years which they survive and therefore also the number of years in which the annuity needs to be paid is higher. This makes females less attractive for the insurer for the immediate lifetime annuity.

If the interest rate is negative the males can be identified as the $\oplus$-agents for the endowment insurance, while the females become the $\ominus$-agents. For the other insurance products it is not clear which gender is assigned to which type of agents. This motivates us to use an interest rate model that ensures positive interest rate, such as the CIR model from Section 3.1.2, compare also Remark 3.8.

In this thesis we assume that there are only two genders. The mortality of transgenders, i.e. how a "change of gender" affects the death probabilities of this person is little explored so far. While the exposure to genetically driven and gender dependent illnesses is not changed, the behavior of the person might be closer to a person of the lived instead of the biological gender.

Remark 3.7. Another interesting life insurance product is the *tontine*, which is kind of a mixture between an (immediate) lifetime annuity and a mortality lottery. When signing the contract, a customer pays a premium that entitles him for an annuity payment. Each time a customer in this contract dies, his payments are split upon the other customers in that contract who are still alive. The contract ends after the last customer had died. We do not investigate tontines further but refer to the literature, e.g. Sabin (2010) [Sab10] or more recently Chen and Rach (2023) [CR23].

3.1.1 Mortality Models

To determine the net expected values and the premiums of our life insurance products, we need to model mortality. By setting up a convenient model, we can calculate the (deferred) death and survival probabilities which are required for later calculations.
We use the Lee-Carter model which was first presented by Lee and Carter in 1992 [LC92]. With this model we can calculate the so-called centralized death probabilities, i.e. the deaths in one year normalized by the average population within that year. In the Lee-Carter model, the centralized death probabilities ${}_tm_x$ are given by

$$\ln({}_tm_x) = a_x + \kappa_t b_x + \varepsilon_{x,t}.$$

Here $(a_x)_x$ and $(b_x)_x$ are age dependent factors, while $(\kappa_t)_t$ models the evolution of mortality over time. With $\varepsilon_{x,t}$ we add an error term. Following [LC92], this error is said to have mean zero and a variance of σ_ε^2 and is introduced to reflect the particular age- and time-specific influences not captured by the model. Other sources speak of $\varepsilon_{x,t}$ being white noise, which is then defined as a vector of statistically independent random variables with a mean of zero and finite variance. This definition coincides with the requirements in [LC92] and should not be mixed up with the definition of white noise in the sense of the (generalized) derivative of a Brownian motion. For the vectors $(b_x)_x$ and $(\kappa_t)_t$ the conditions

$$\sum_x b_x = 1 \text{ and} \sum_t \kappa_t = 0$$

have to hold. Given the centralized death probabilities ${}_tm_x$ we obtain the one-year death probabilities $q_{x,t}$ at time t by

$$q_{x,t} = \frac{{}_tm_x}{1 + 1/2 \cdot {}_tm_x}.$$

Different to the centralized death probabilities, we use death probabilities that depend on the population at the beginning of a year instead of the average population of that year for normalizing the one-year death probabilities, which delivers us ${}_tm_x \geq q_{x,t}$ for all t and x. All other probabilities we need for modeling can be calculated from the one-year death probabilities.
We can calibrate the parameters using life tables. The data we use for

calibration are French life tables that origin from the Human Mortality Database (HMD)[4]. The calibration can be performed using singular value decomposition (SVD), see [LC92], or by various other approaches. We use a MATLAB toolbox[5] for calibration, but apply it on French data instead of U.S. data.
We continue by performing a quadratic polynomial regression on the vector $(\kappa_t)_t$, to be able to forecast the mortality data. As $(a_x)_x$ and $(b_x)_x$ are age dependent and not time dependent factors, we do not need to perform a regression on them.
In Figure 6 below, the three parameters $(a_x)_x$, $(b_x)_x$ and $(\kappa_t)_t$ are plotted. As $(a_x)_x$ and $(b_x)_x$ are age dependent, they are plotted over the age, while $(\kappa_t)_t$ is plotted over the birth year.

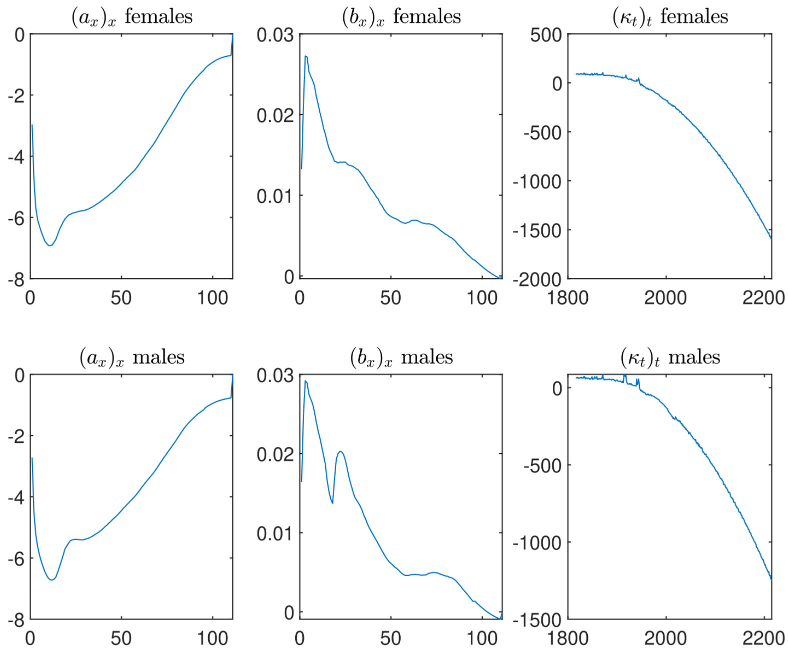
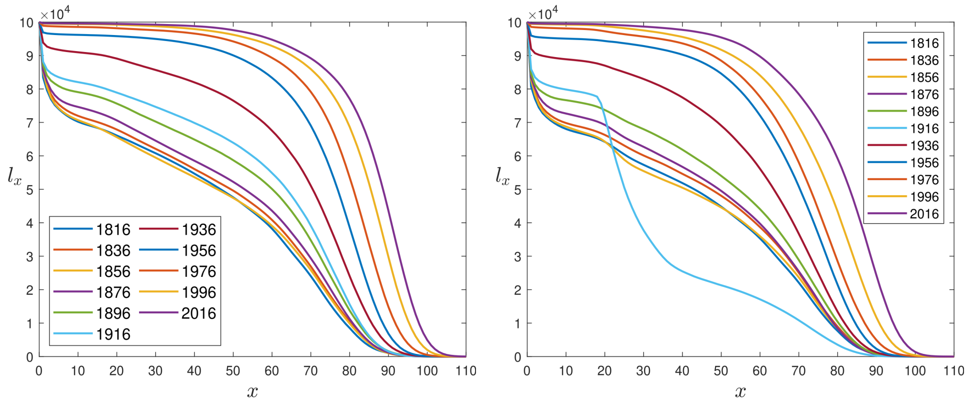


Figure 6: Parameters of the Lee-Carter model for males and females

We can clearly observe that the life expectancy rises over the years as the parameter $(\kappa_t)_t$ decreases over time. Also the mortality shocks due to World War II or pandemics can be seen in the data, compare Figures 7 and 8 below for the according mortality and survival probabilities. Another interesting observation is that the parameter $(b_x)_x$ for males has a local maximum

[4]`https://www.mortality.org/`, visited November 2020
[5]`https://github.com/mrockinger/Matlab-Longevity-Toolbox`, visited February 2021

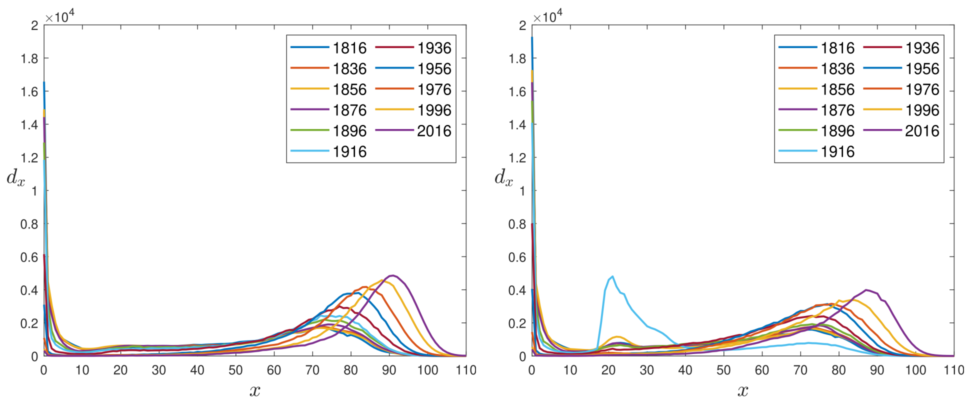
around the age of 28. As already analyzed in different studies like Patton *et al.* (2009) [PCS+09], the reason for this peak is that young males are often risk-loving and adventurous. This results in higher mortality rates for males between 20 and 35, mainly due to (traffic, sports,...) accidents. This effect can also be observed for the females, but is much weaker in this case.
One advantage of the Lee-Carter model is that it uses a time dependent factor $(\kappa_t)_t$ and is able to manage the longevity risk. This risk describes that not only the probabilities, but also the distribution of mortality changes over time, see Figures 7 and 8.



(a) Females (b) Males

Figure 7: Number of females and males l_x that still lived at age x from an initial population of 100,000 for different years

As one can see, the shapes of the curves in Figure 7 change over time. Our model is able to deal with this change in the distribution of mortality.



(a) Females (b) Males

Figure 8: Number of females and males d_x that died at age x from an initial population of 100,000 for different years

Longevity is a serious risk for insurance companies, not taking this risk into account might result in too high or (even worse) too low premiums and in extreme cases to the bankruptcy of the whole company. There are other approaches to mitigate this risk for the insurer in addition to using a mortality model capable of dealing with longevity risk. One possibility is to consider reinsurance and set up collective longevity swaps between the insurers and the reinsurer, as it is done in Chen *et al.* (2022) [CLS22].

3.1.2 Interest Rate Models

We model the time-T-discount factor $B(0,T)$ by

$$B(0,T) = \exp\left(-\int_0^T r(s)\mathrm{d}s\right),$$

where r is given as the (instantaneous) short-rate. To model the short-rate, we use the CIR model, named after their developers Cox, Ingersoll and Ross (1985) [CIR85]. Based on their approach, r is given by $r_0 > 0$ and

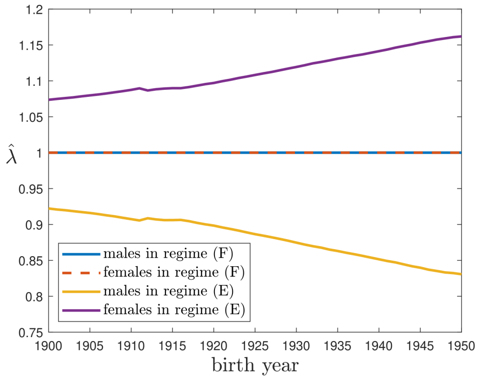
$$\mathrm{d}r_t = a(b - r_t)\mathrm{d}t + \sigma\sqrt{r_t}\mathrm{d}W_t.$$

Here, $(W_t)_t$ is a standard Brownian motion and a, b and σ are (deterministic) parameters. The stochastic differential equation above can be solved by various methods. We make use of the built-in MATLAB CIR solver.

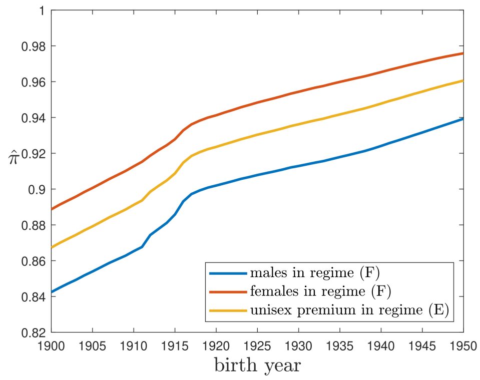
Remark 3.8. The CIR model ensures mean reversion to the long-term mean b, where the parameter a controls the adjustment speed. If a and b are both positive, the CIR model guarantees non-negative interest rates, the condition $2ab \geq \sigma^2$ grants strict positivity, see for example Brigo and Mercurio (2007) [BM07] for more details. This has the advantage that it is clear, which gender can be identified with which group of insured ($\oplus$ or $\ominus$), see Remark 3.6. We furthermore know that the CIR model has an affine term structure.

3.2 Numerical Examples

We continue by providing some numerical examples for the gender-specific and unisex premiums for the four life insurance products which we introduced in Section 3.1. We regard a portfolio consisting of the same number of males and females, all equipped with a CRRA utility function with risk aversion parameter $\rho = 3$, i.e. $u(x) = -\frac{1}{2x^2}$ and an initial wealth $a = 2$ ($a = 45$ for the immediate lifetime annuity). We assume all customers have an initial age of 30 and we choose the parameters of the CIR model such that we have a (long-term) risk-free interest rate of 3%. As the interest rate does not play a relevant role for the pure endowment insurance, there is only one possible time point for payoff, we assume a (constant) risk-free interest rate of 0% for the pure endowment insurance. We assume that the contract duration is 20 years for all insurance products but the immediate lifetime annuity, as there is no such contract duration for this pension product. The premiums and corresponding optimal coverages are then given as functions of the birth year. The corresponding figures are displayed below.



(a) Eq. insurance coverage $\hat{\lambda}$



(b) Eq. premium per coverage $\hat{\pi}$

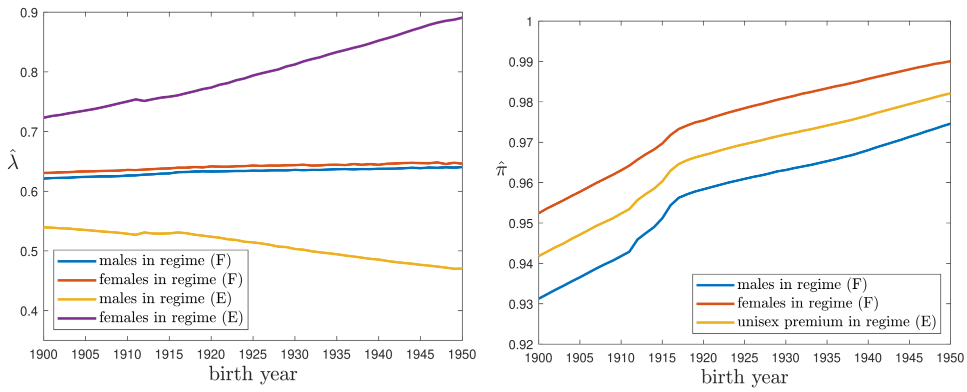
Figure 9: Equilibrium insurance coverage $\hat{\lambda}$ and equilibrium premium per coverage $\hat{\pi}$ of a pure endowment insurance as functions of the birth year in scenario (C)

We can observe the following phenomena:

- First, we compare the premiums in scenario (C) with those in scenario (M), take for example Figures 9 with 10 for the pure endowment insurance. As expected, we see that the monopolistic premiums always lie above the corresponding competitive premiums. Therefore, the equilibrium insurance coverages in a monopolistic market always lie lower than in a corresponding competitive market.

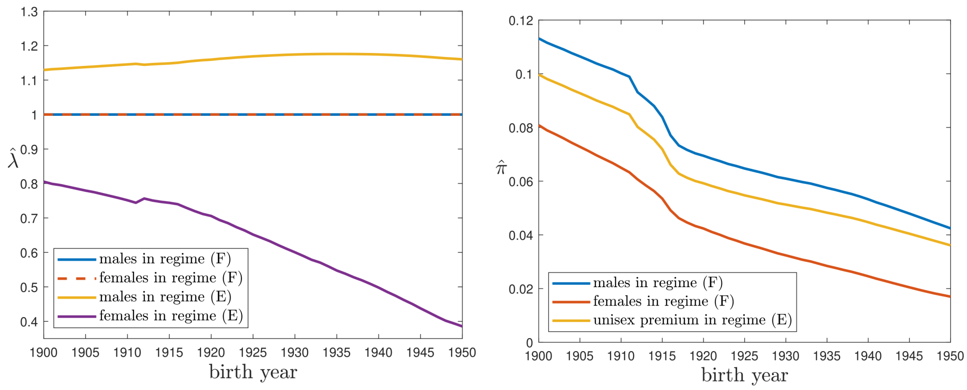
- As a second observation from this comparison, the relative difference between the coverages in markets with free contract design and mandatory unisex tariffs is much higher in scenario (M) than it is in scenario (C). Also the push-out effect is stronger, compare e.g. Figure 13 with 14.

- In competitive markets with free contract design, customers always buy full coverage, see e.g. Figure 9a and compare the argument in Example 2.29.



(a) Eq. insurance coverage $\hat{\lambda}$ (b) Eq. premium per coverage $\hat{\pi}$

Figure 10: Equilibrium insurance coverage $\hat{\lambda}$ and equilibrium premium per coverage $\hat{\pi}$ of a pure endowment insurance as functions of the birth year in scenario (M)



(a) Eq. insurance coverage $\hat{\lambda}$ (b) Eq. premium per coverage $\hat{\pi}$

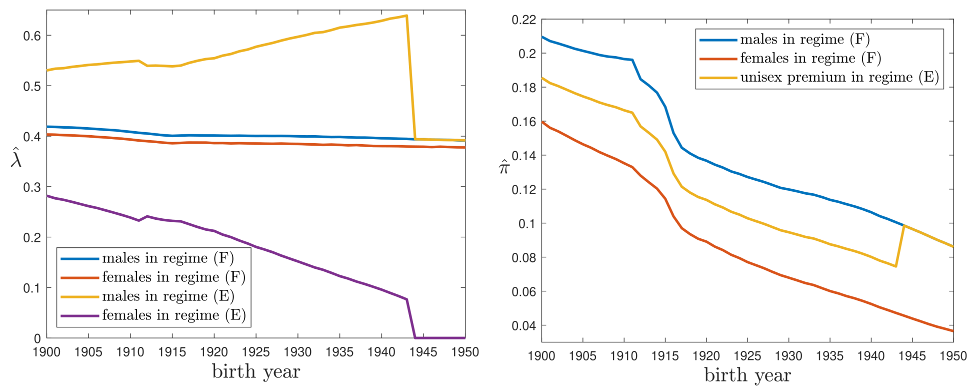
Figure 11: Equilibrium insurance coverage $\hat{\lambda}$ and equilibrium premium per coverage $\hat{\pi}$ of a term insurance as functions of the birth year in scenario (C)

- As already argued in Section 3.1.1, the life expectation has risen over time, compare Figures 7 and 8. Over the (birth) years, this naturally leads to rising premiums for the pure endowment and endowment insurance as well as for the immediate lifetime annuity and to falling premiums for the term insurance.

- For the pure endowment insurance (Figures 9 and 10) and the immediate lifetime annuity (Figures 15 and 16), females pay more than males in regime (F), while it is the other way round for the term insurance (Figures 11 and 12) and the endowment insurance (Figures 13 and 14). This relation was already addressed in Remark 3.6.

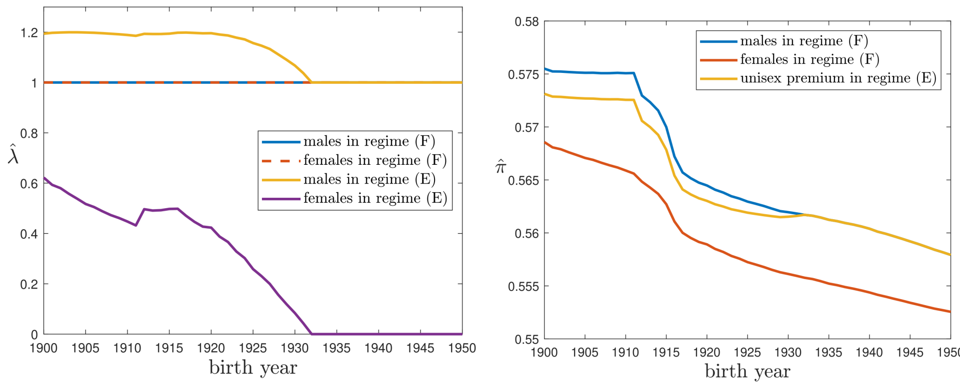
- The unisex premium always lies between the male and the female premium. For the term insurance and the immediate lifetime annuity for monopolistic markets (Figures 12 and 16) as well as for the endowment for both market settings (Figures 13 and 14) we can see a push-out of one gender out of the insurance market. In Examples 2.29 and 2.31 this phenomenon is already obtained in some theoretical examples. These examples originate from [SS14], where it is argued that one can obtain two different equilibriums, compare Remark 2.30. In one equilibrium, agents of both genders purchase insurance. In the other one only the agents with the higher risk purchase insurance. It is concluded that the push-out point corresponds to the point where we switch from the first to the second equilibrium. This argument transfers to our extended model as well.

- We have seen the effect of World War II in the mortality data, compare Figures 7 and 8. Of course, this impacts the premiums, e.g. by adding some roughness to them, especially to the male one. For the endowment insurance in monopolistic markets this even drives the females to return to the market after there was a push-out for some birth years, regard Figure 14. When it comes to regressing the life insurance premiums in Chapter 4, the regressions of the male premiums cause higher training errors for the female and unisex ones. It can be assumed that extra roughness is one of the reasons for this, as taking the according birth years out of the regression leaves us with training errors for the males that are approximately as high as for the females.



(a) Eq. insurance coverage $\hat{\lambda}$ (b) Eq. premium per coverage $\hat{\pi}$

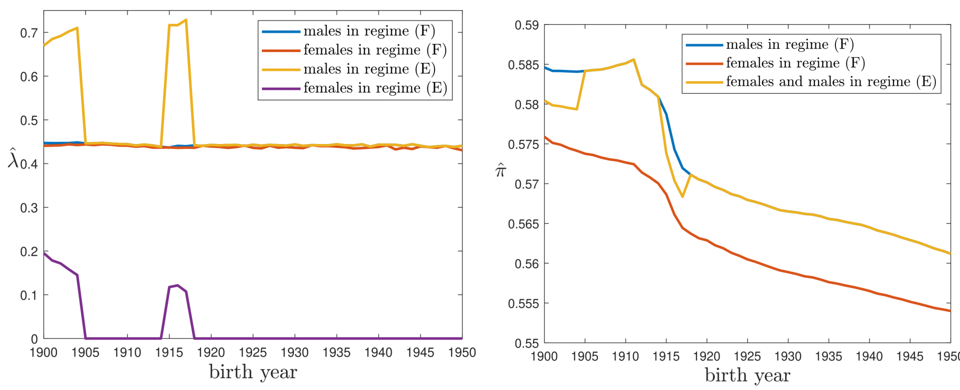
Figure 12: Equilibrium insurance coverage $\hat{\lambda}$ and equilibrium premium per coverage $\hat{\pi}$ of a term insurance as functions of the birth year in scenario (M)



(a) Eq. insurance coverage $\hat{\lambda}$ (b) Eq. premium per coverage $\hat{\pi}$

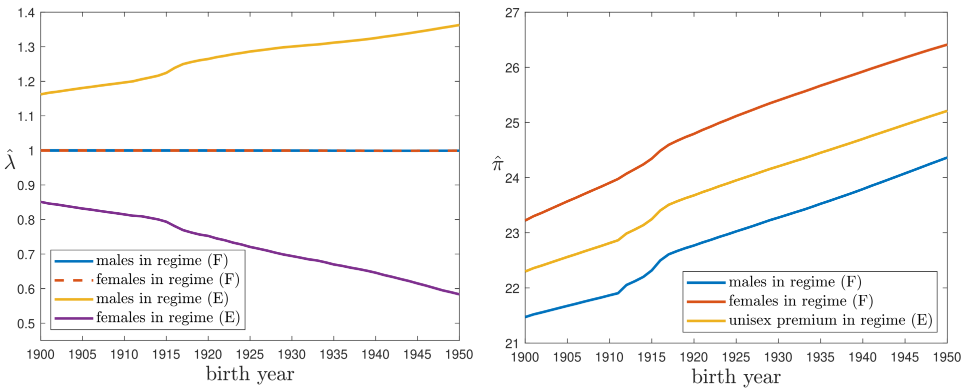
Figure 13: Equilibrium insurance coverage $\hat{\lambda}$ and equilibrium premium per coverage $\hat{\pi}$ of an endowment insurance as functions of the birth year in scenario (C)

- As mentioned before, the strength of the push-out effect depends highly on the underlying product. As one can see in Figures 13 and 14, especially the endowment insurance is very susceptible for the push-out effect. This can be explained by the fact that the time point where the agent is dying only determines when, but not if he is paid out by the insurance company, compare also Remark 3.4. Therefore the risk has a much lower variance and the demand for insurance is reduced as well. One can also see from the diagrams that the gender-specific premiums lie much closer together.



(a) Eq. insurance coverage $\hat{\lambda}$ (b) Eq. premium per coverage $\hat{\pi}$

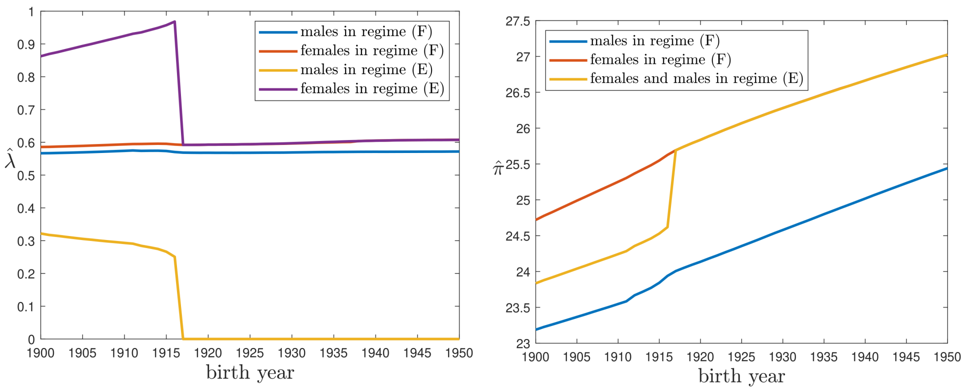
Figure 14: Equilibrium insurance coverage $\hat{\lambda}$ and equilibrium premium per coverage $\hat{\pi}$ of an endowment insurance as functions of the birth year in scenario (M)



(a) Eq. insurance coverage $\hat{\lambda}$ (b) Eq. premium per coverage $\hat{\pi}$

Figure 15: Equilibrium insurance coverage $\hat{\lambda}$ and equilibrium premium per coverage $\hat{\pi}$ of an immediate lifetime annuity as functions of the birth year in scenario (C)

- Interestingly, the unisex premium often lies relatively closer to the lower premium in scenario (M) than it does in scenario (C), see e.g. Figures 18 and 19 from the next section. Sometimes, the unisex premium is lower than the weighted average of the gender-specific premiums, compare also Figure 25 in Section 5.1. An explanation for this could be the structure of the optimization problem in the monopolistic market. By setting the unisex premium comparably low, more insurance is bought in total and the profit of the insurer might rise.



(a) Eq. insurance coverage $\hat{\lambda}$ (b) Eq. premium per coverage $\hat{\pi}$

Figure 16: Equilibrium insurance coverage $\hat{\lambda}$ and equilibrium premium per coverage $\hat{\pi}$ of an immediate lifetime annuity as functions of the birth year in scenario (M)

Remark 3.9. All functions in the figures of this chapter and Section 3.3 are functions of the birth years. Hence, the number of points we use for each function in a plot is just the number of birth years we are plotting over. For all figures besides Figures 6, 7 and 8, we use the birth years between 1900 and 1950, so 51 data points have to be computed.

Remark 3.10. In real-life insurance markets, insurers will charge different types of costs. In life insurance, these costs can usually be grouped into

- α-costs, i.e. acquisition costs (costs for signing the contract like provision for the sales agent, costs of a medical test,...),
- β-costs, i.e. collection expenses (banking fees, investment costs),
- γ-costs, i.e. administration expenses (wages, buildings, costs for data processing,...) and
- κ-costs, i.e. the annual fixed costs of an insurer, often included in the γ-costs.

While α- and γ-costs are charged as a fraction of the insured sum (e.g. $\alpha = 1\%$ and $\gamma = 0.2\%$), β-costs depend on the premiums (e.g. $\beta = 3\%$). When it comes to calculations, we can see that it is easy to integrate these costs into our model, but the observed effects do not change qualitatively. To ease further computations, we therefore do not regard costs in our model.

3.3 Mixing Parameters

Instead of calculating the unisex premiums directly by the approach used in Section 2.5, we could try to find a mixing parameter according to which the gender-specific premiums or model parameters have to be mixed to compute the unisex premium. The actual premium calculation could then be performed by using the gender-specific formulas only. This procedure has the advantage that the gender-specific approach needs much lower computational effort than to compute unisex premiums. Especially when it comes to products with very long contract durations, such as the immediate lifetime annuity, an approach using mixing parameters could save some computation time.

A second target for this section is to gain a better understanding of the behavior of this model. The equilibrium approach we are using is based on two connected optimization problems that rely on each other. Therefore, predicting the exact output for a given specification of parameters is not straightforward.

The conceptualization of the mixing parameter approach is done in Section 3.3.1. By setting up different mixing parameters in Definitions 3.11, 3.12 and 3.13, we try to create a better understanding of the market model itself. In Section 3.3.2, we study numerical examples where our definitions are applied, recognizing some phenomena from Section 3.2 but also making new observations.

3.3.1 Conceptualization

We present three different approaches for defining a mixing parameter.

a) The easiest approach to set up a mixing parameter is to define it by the weight given to the males when one computes the unisex premium as a weighted sum of the gender-specific premiums. Of course, one could also choose the weight given to the females instead. More formal, we receive the following definition.

   Definition 3.11. The *mixing parameter based on premiums* is defined as the parameter ξ_1 that solves the equation

   $$\Pi_u = \xi_1 \cdot \Pi_m + (1 - \xi_1) \cdot \Pi_f.$$

   Here, Π_u, Π_m and Π_f denote the unisex, male and female premium, respectively.

b) All of our insurance products depend – implicitly or explicitly – on survival probabilities. We denote this dependency by writing $\Pi(S)$. Therefore we could also define a mixing parameter based on the survival probabilities.

   Definition 3.12. We can find the *mixing parameter based on probabilities* as the parameter ξ_2 solving the equation

   $$\Pi_u = \Pi(\xi_2 \cdot S_m + (1 - \xi_2) \cdot S_f).$$

   Here, S_m and S_f denote male and female survival probability, respectively. Furthermore, $\Pi(\cdot)$ denotes the gender-specific premium based on a given survival probability.

   It is easy to show that the mixing parameters calculated by the first two approaches coincide if we are in market scenario (C), see Proposition 3.15.

c) As explained in Section 3.1.1, we are using the Lee-Carter mortality model [LC92]. This model uses three parameters, $(a_x)_x$, $(b_x)_x$ and $(\kappa_t)_t$. We can now replace one or more unisex model parameters by the weighted sum of the gender-specific parameters. If we do not replace all parameters accordingly, a push-out of one gender out of the market would not (necessarily) result in the mixing parameter to be zero or one. This is a desirable property of the mixing parameter, which is also fulfilled by the first two approaches.

Definition 3.13. We define the *mixing parameter based on parameters* as the parameter ξ_3 that solves the equation

$$\begin{aligned}\Pi_u =& \Pi(S(\xi_3 \cdot a_x^m + (1-\xi_3) \cdot a_x^f, \\ & \xi_3 \cdot b_x^m + (1-\xi_3) \cdot b_x^f, \xi_3 \cdot \kappa_t^m + (1-\xi_3) \cdot \kappa_t^f)).\end{aligned}$$

Here, a_x^i,b_x^i and κ_t^i denote the male and female parameters of the Lee-Carter model for $i = m, f$ and $S(\cdot)$ denotes the (Lee-Carter) survival probability given the corresponding parameters.

Remark 3.14. The above definition is of course only one approach of defining a mixing parameter based on parameters. Instead of mixing all three parameters of the Lee-Carter model, we could also mix only one or two and let the remaining parameter(s) unchanged. As argued, we want a push-out of one gender to result in the mixing parameter being zero or one. Furthermore, for being able to model both, the age and the time dependent factors which behave differently for males and females in practice, it does not seem meaningful to mix only one or two parameters. Another option would be to mix the parameters b_x and κ_t with $\sqrt{\xi_3}$ and $\sqrt{1-\xi_3}$ instead of ξ_3 and $1-\xi_3$. The reasoning to follow this approach would be that the parameters b_x and κ_t get multiplied in the Lee-Carter model, see Section 3.1.1. By using the root, the product $\kappa_t b_x$ gets mixed by ξ_3 and $1-\xi_3$ instead of their squares. In practice, this leads to a mixing parameter that is highly unstable and not robust. Hence, we decided to model the parameter according to Definition 3.13.

Before continuing, we give a short proof that the mixing parameters based on premiums and probabilities coincide for scenario (C).

Proposition 3.15. Assume a competitive insurance market, where we have an insurance product corresponding to a risk $Z_{m/f}$. As in Definition 2.4, the risks are assumed to be discrete, non-negative random variables with

$$\mathbb{P}(Z_{m/f} = z_{m/f}^k) = p_{m/f}^k,$$

where $k \in \mathbb{N}_0$, $z_{m/f}^k \geq 0$ and $p_{m/f}^k \geq 0$ for all k as well as $\sum_{k=0}^{\infty} p_{m/f}^k = 1$. Moreover we assume that $z_m^k = z_f^k$ for all $k = 0, 1, \ldots$, i.e. the damage amounts of the two risks are equivalent, Z_m and Z_f only differ in the probability that this damage occurs.
In this market setting, the mixing parameters based on premiums ξ_1 and probabilities ξ_2 coincide.

Proof. We need to show that $\xi_1 = \xi_2$ holds. From Section 2.5, namely Equation (9), we know that the gender-specific premiums are given by the net value of the contract, i.e. by

$$\Pi_{m/f} = \mathbb{E}[Z_{m/f}] = \sum_{k=0}^{\infty} p^k_{m/f} z^k_{m/f}.$$

Note that in our setting, $z^k_m = z^k_f$ holds for all $k = 0, 1, \dots$, so we can write z^k instead of z^k_m and z^k_f.
The mixing parameter by premiums is given as the solution of the equation

$$\begin{aligned}\Pi_u &= \xi_1 \cdot \Pi_m + (1 - \xi_1) \cdot \Pi_f \\ &= \xi_1 \cdot \left(\sum_{k=0}^{\infty} p^k_m z^k\right) + (1 - \xi_1) \cdot \left(\sum_{k=0}^{\infty} p^k_f z^k\right).\end{aligned}$$

On the other hand, the mixing parameter by probabilities is given as the solution of

$$\begin{aligned}\Pi_u &= \Pi(\xi_2 \cdot S_m + (1 - \xi_2) \cdot S_f) \\ &= \sum_{k=0}^{\infty} (\xi_2 \cdot p^k_m + (1 - \xi_2) \cdot p^k_f) z^k. \\ &= \xi_2 \cdot \left(\sum_{k=0}^{\infty} p^k_m z^k\right) + (1 - \xi_2) \cdot \left(\sum_{k=0}^{\infty} p^k_f z^k\right)\end{aligned}$$

As the sums in the equations are non-zero and finite by definition (see Definition 2.4 and Remark 2.13), we finally obtain

$$\xi_1 = \xi_2.$$

□

Remark 3.16. In Chen and Vigna (2017) [CV17], the term of the *unisex fairness principle* is introduced. Therefore, we regard a given portfolio that consists of m males and n females with premiums Π_m and Π_f, respectively. The unisex premium Π_u is calculated according to the unisex fairness principle if

$$\Pi_u = w \cdot \Pi_m + (1 - w) \cdot \Pi_f,$$

where

$$w = \frac{m}{m + n}$$

is the *fraction of males in the portfolio*. In other words, a premium is said to fulfill the principle if the unisex premium is given as the weighted average of the two gender-specific premiums. This implies that changing the regime from (F) to (E), the amount of premium, the insurance company earns does not change if the amount of coverage purchased by the insured does not

change. As we have seen and continue to see in various examples, this assumption does not hold in our insurance market model.
If the unisex fairness principle is fulfilled, they call the premium Π_u the *unisex fair premium.* According to Definition 3.11, this is the case if $\xi_1 = w$. As we can see in Section 3.3.2, we almost always have $\xi_1 \neq w$, i.e. the unisex fairness principle from [CV17] is almost never satisfied in our setting.

Remark 3.17. Sticking to the unisex fairness principle from the last remark, [CV17] use this principle to define a so-called *fair unisex mortality intensity.* This mortality intensity is chosen in such a way that the resulting unisex premium fulfills the unisex fairness principle. Note that the mixing parameter based on probabilities ξ_2 from Definition 3.12 of our approach turns out to be the closest to the approach in [CV17]. Nonetheless, there are two differences, where the second one is indeed of fundamental kind:

a) While our modeling deals with survival rates, [CV17] uses the mortality intensities. This difference is more of a technical kind and does not result in a big difference in the computations. If one takes the death probability (i.e. the integrated mortality intensity), the two approaches are the same. This can be seen in the calculation below.
Therefore we consider arbitrary survival probabilities p_m and p_f of males and females, respectively. With $q_m = 1 - p_m$ and $q_f = 1 - p_f$ we denote the corresponding death or death probabilities. If we receive the unisex survival probability by mixing the gender-specific survival probabilities by a mixing parameter ξ, i.e. by setting

$$p_u = \xi p_m + (1 - \xi)p_f,$$

we equivalently obtain

$$\begin{aligned} q_u &= 1 - p_u \\ &= 1 - (\xi p_m + (1 - \xi)p_f) \\ &= 1 - (\xi(1 - q_m) + (1 - \xi)(1 - q_f)) \\ &= \xi q_m + (1 - \xi)q_f \end{aligned}$$

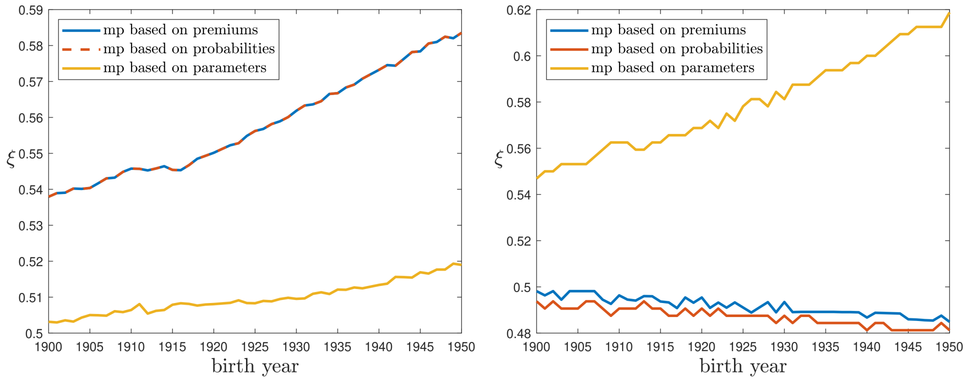
and vice versa.

b) Our approach sets the focus to find the mixing parameter such that the unisex premium using the equilibrium approach from Chapter 2 is the same as calculating it as a gender-specific premium with accordingly mixed survival probability. Contrariwise, [CV17] models it in such a way that the total amount of premium paid in the two insurance markets with regimes (F) and (E) are the same. Note that the amount of coverage purchased by the customers is assumed to be equal in both regimes. Different to the first point, this makes indeed a different approach.

Remark 3.18. The values of the mixing parameters highly depend on the values of the model parameters. To use this approach for the calculation of unisex premiums, we need to perform a regression on the values of the mixing parameters. We can then use them for pricing unisex premiums of parameter specifications we do not know so far. One can assume that the regression works much better if the premiums and therefore the mixing parameters do not have jumps, as most of the classic regression approaches perform better, the smoother the function we want to regress is. As we have seen in Figure 13, a push-out of one gender out of the market does not necessarily result in a jump of the premiums. In Chapter 4 different regression methods for regressing equilibrium insurance premiums are presented.

3.3.2 Numerical Examples

In the next step, we compare the values of the mixing parameters from the three approaches for our four life insurance products. Some observations for the given examples can be found in Remark 3.19. In a further step, we could perform a regression on the data and compare the quality of the regression for the different approaches.
As in Section 3.2, we present examples for the pure endowment insurance defined in Example 3.1, the term insurance introduced in Example 3.2, the endowment insurance from Example 3.3 and the immediate lifetime annuity introduced in Example 3.5. For all examples, we use the standard parameters from Section 3.2. There we assumed a portfolio that consists of equal shares of males and females. The agents are said to have CRRA (power) utility with $\rho = 3$ and an initial wealth of $a = 2$ ($a = 45$ for the immediate lifetime annuity). The initial age of the customers is assumed to be 30 years while the contract duration for all products but the immediate lifetime annuity is 20 years. While we assume the interest rate to be constant $r = 0\%$ for the pure endowment insurance, we use the CIR model such that we have a (long-term) risk-free interest rate of $r = 3\%$ for the other three insurance products. The results are displayed in the figures below.

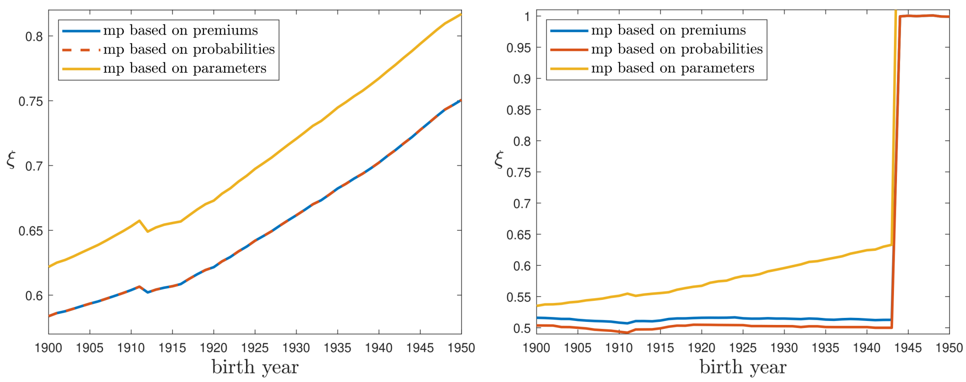


(a) Scenario (C)

(b) Scenario (M)

Figure 17: Mixing parameters (mp) ξ_1, ξ_2 and ξ_3 for the pure endowment insurance in scenarios (C) and (M) as functions of the birth year

In Figure 17 above, the three mixing parameters we defined before are calculated and plotted for the pure endowment insurance. The corresponding premiums and coverages on which the calculations depend implicitly and explicitly can be found in Figures 9 and 10 but are not given here again.



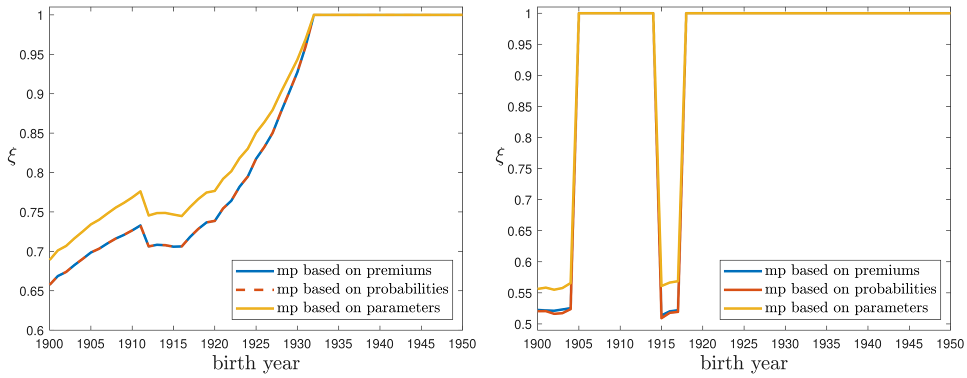
(a) Scenario (C)

(b) Scenario (M)

Figure 18: Mixing parameters (mp) ξ_1, ξ_2 and ξ_3 for the term insurance in scenarios (C) and (M) as functions of the birth year

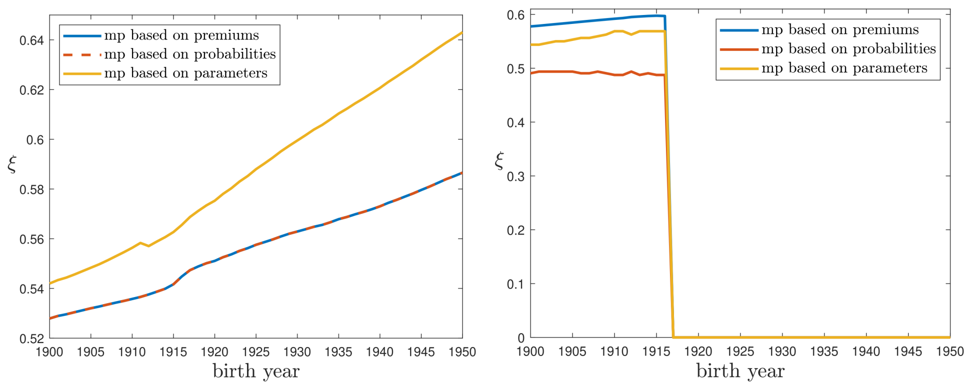
Remark 3.19. We summarize our observations from the last figures in the following:

- If one gender is pushed out of the insurance market, all three mixing parameters are zero or one, respectively, compare e.g. Figure 20b or 18b. This is meaningful, as a market where one gender is pushed completely out equals a market where only customers of one gender sign contracts.



(a) Scenario (C) (b) Scenario (M)

Figure 19: Mixing parameters (mp) ξ_1, ξ_2 and ξ_3 for the endowment insurance in scenarios (C) and (M) as functions of the birth year



(a) Scenario (C) (b) Scenario (M)

Figure 20: Mixing parameters (mp) ξ_1, ξ_2 and ξ_3 for the immediate lifetime annuity in scenarios (C) and (M) as functions of the birth year

- As proven in Proposition 3.15, the mixing parameter based on premiums ξ_1 and the mixing parameter based on probabilities ξ_2 coincide in the competitive case, see for example Figure 17a. For our examples in the monopolistic market, the mixing parameter based on premiums is always bigger than the mixing parameter based on probabilities in the monopolistic case, see e.g. Figure 17b. When one plays around with the market and product specifications, one can realize that this is true for most but not all settings. It seems that the higher the contract duration or the initial ages are, the bigger the difference becomes. Only for specifications with really low initial ages, say 15 or less, ξ_2 sometimes is larger than ξ_1.
- In most examples, the mixing parameter ξ_3 based on parameters lies

clearly above the two other mixing parameters. Nonetheless, there are market settings in which it is opposite. As we have seen for the monopolistic case of the immediate lifetime annuity (Figure 20b), ξ_3 can also lie between the other two parameters. In general, it does not seem to be possible to formulate a proposition about the behavior of ξ_3 in comparison to the other parameters.

- If there is no push-out, the mixing parameters in a competitive market are often higher than in a monopolistic one, compare for example Figure 18a with 18b.
- One would expect that the unisex premium lies closer to the higher gender-specific premium. Therefore, one would expect the mixing parameters when there is no push-out to be smaller than 0.5 for the pure endowment insurance and the immediate lifetime annuity, and to be greater than 0.5 for the term and the endowment insurance. As we can see, these expectations are not always met, compare Figures 17a and 20b.
- In general, the mixing parameters are growing over time. Nonetheless, it is also possible to obtain a reduction of the values. While a negative trend is rare (ξ_1 and ξ_2 in Figure 17b), a temporary reduction due to the effect of World War II can be observed in many examples, especially in Figure 19a. We already noticed this effect in most of the examples in Section 3.2, where we calculated the premiums and equilibrium insurance coverages of the life insurance products.

Remark 3.20. In [CV17], a proposition is presented which assumes that the male force of mortality is greater than the female one at each time point. If this is true, their mixing parameter ξ^* for the pure endowment insurance is always lower or equal the fraction of males in the portfolio.
As long as it is assumed that the mixing parameter is constant over time, it always lies close to the fraction of males in the portfolio. This changes as soon as one starts to analyze their mixing parameter time dependently. Based on the insurance product, the difference between the mixing parameter and the fraction of males in the portfolio can become quite large. Nonetheless, the difference vanishes over the contract time, reaching almost zero at the end of the duration time. For more details we refer to [CV17].

Example 3.21. As a last example of this section, we visualize how the fraction of males in a portfolio changes over the years. We therefore model six different portfolios with initial ages of 0, 20, 40, 60, 80 and 100. All the portfolios consist of equal shares of males and females at the beginning of the contract. The evolution of this fraction over time is shown in Figure 21 below. The data used in this plot is the French data from the HMD for the

2016 cohort.
As expected, the fraction of males never surpasses 0.5. The decline of the fraction accelerates over the years. Interestingly it slows down again at very old ages. While the exact behavior is highly affected by the small portfolio size, this deceleration in the decline can be observed for almost all birth years. There is some research going on for mortality at very high ages which might explain this phenomenon, see for example Gavrilov and Gavrilova [GG11] or Casiglia *et al.* (1993) [CSG+93]. The International Database on Longevity (IDL)[6] collects information about supercentenarians, i.e. people that lived at least until age 105.

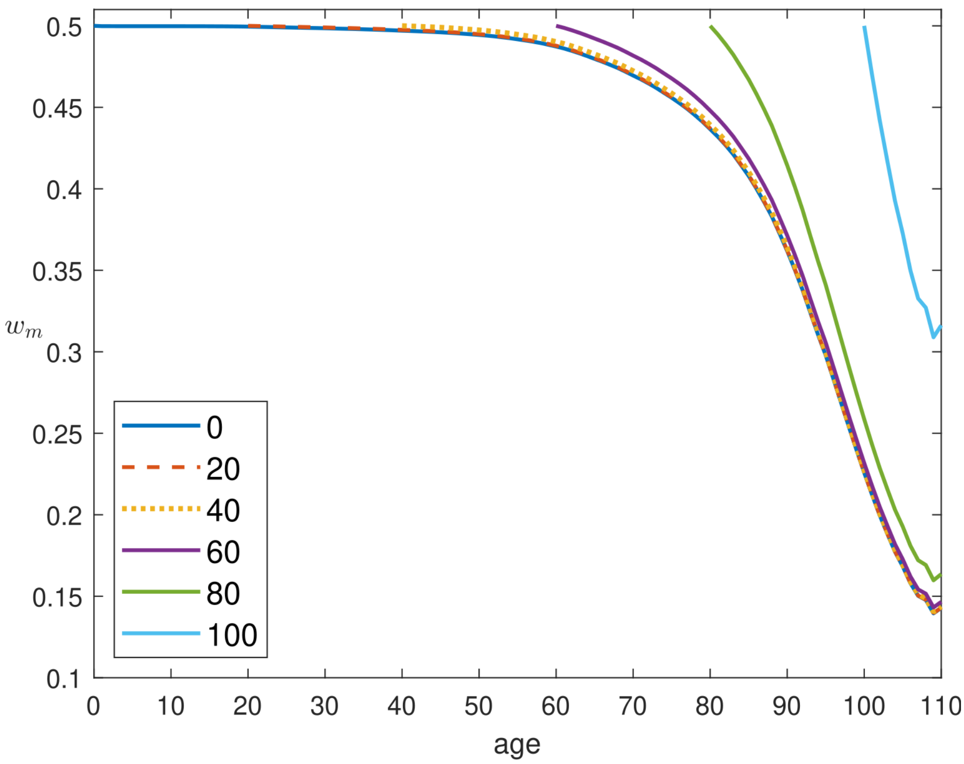


Figure 21: Evolution of the fraction of males w_m for portfolios with different initial ages over the time

[6] https://www.supercentenarians.org/en/

4 Regression of Life Insurance Premiums

The calculation of premiums, especially for more complex products like the endowment insurance or the immediate lifetime annuity can become very time-consuming as two connected optimization problems or equations need to be solved for each parameter specification. Therefore, we want to generate only a small amount of data points and fit a regression through them to gain the premiums for a large amount of parameter specifications. More details about the used methods can be found e.g. in Géron (2019) [Gé19].
In this chapter, we perform some regressions on the life insurance premiums we obtained earlier in Section 3.2. We can see in Section 4.1 that neural networks are indeed a quite promising tool to perform this task. Nonetheless, we are going to analyze and compare other classic regression approaches in Section 4.2.

4.1 Neural Networks

A promising approach for regressing life insurance premiums are neural networks. We implement the neural networks which we use for regression with the `keras` package of Python and focus on deep neural nets with dense layers.

Remark 4.1. An *(artificial) neural network* is a method in machine learning trying to rebuild a neural network which can be found in the brains of animals and humans. A network consists of some input neurons, different layers of neurons, so-called *hidden layers* in between and one or more output neurons. If each neuron of one layer is connected with all the neurons of the next one, we speak of a *network with dense layers.* The term of a *deep neural network* is not clearly defined, usually one refers to a network with more than one hidden layer. From layer to layer, each neuron calculates an output as a weighted sum of the inputs. We then apply a so-called *activation function* to this sum. Examples for different activation functions can be found in Figure 23 below.
Before being able to use a neural network for regression, we need to train it. This is done using a method called *backpropagation.* Therefore, one uses data inputs, one already knows the output for. These inputs are passed during the network performing a so-called *forward pass.* In the next step, the difference between the calculated and the real output is measured using a *loss function*, often the mean squared error function is taken for this purpose. The actual backpropagation step is then performed by traveling backwards through the layers to measure the error contribution of each connection. In a final step, the weights are adjusted by an *optimizer.* The intensity of changing can be controlled by the *learning rate* of the optimizer.
Not only the specifications of the neural network, like its size, used activation function etc. are important for the quality of training. Also the size

of the training set, i.e. the number of points used for training is relevant. Furthermore, the number of *epochs*, i.e. how many training cycles are performed for each point plays a fundamental role. More details together with some algorithms can be found in [Gé19].

In Figure 22 below[7] a visualization of a neural network with three hidden layers is given.

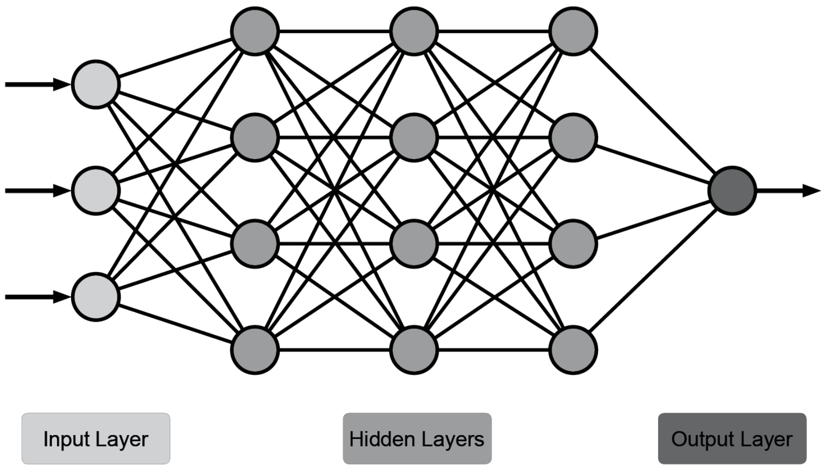


Figure 22: Neural network with three hidden layers

In the following, we investigate the effect of several network specifications on the goodness of our approximation. Of course, the results presented in this section have to be understood only as a feasibility study. As the quality of approximation is strongly related to the underlying data set, one cannot say, which network specification is generally better for the given problem. This is not a specific issue for our setting but a general characteristic of neural network regression.

For our training, we use a set of data points generated using the calculation method from Chapter 3. The insurance product we are focusing on is the pure endowment insurance in scenario (C). For scenario (M) as well as for other insurance products such as the term insurance, the observable phenomena and effects are similar to the pure endowment insurance in a competitive market. Therefore we decided to analyze the product, where the data points are the easiest to generate. The set of data points contains all combinations of birth years between 1900 and 1950, initial ages between 20 and 70 and contract durations between 1 and 50 years. In total, our

[7] Figure is based on `https://www.spotfire.com/glossary/what-is-a-neural-network`

complete data set contains

$$51 \text{ (diff. birth years)} \cdot 51 \text{ (diff. initial ages)} \cdot 50 \text{ (diff. contract durations)}$$
$$= 130,050 \text{ (data points)}$$

from which we usually only use a small share for training, validating and testing our network, as explained below.
In the following, we compare a variety of neural net regression approaches which differ in only one network specification. All other specifications are set to their standard values. With "standard neural network" we address a network with

- three hidden layers consisting of 150, 100 and 30 neurons,
- the Adam optimizer with learning rate 0.001 (standard value for the Adam optimizer) and 1000 epochs,
- the elu activation function and
- the mean-squared error loss function.

The sizes of the train, validation and test sets are 1000, 100 and 10000, respectively. The test set is chosen to be so large to ensure that the loss does not depend too much on the choice of the test set.
The test set is chosen at random, also the network is initialized randomly. To avoid being too dependent on chance, it is state of the art to calculate the average root-mean squared errors over ten networks or data sets. To reduce the effect of the choice of the test set even further, we train ten different neural networks with a new and randomly chosen test set for each network. All ten networks are trained from scratch.

The Network Size

One important question when it comes to neural networks is the number of layers and the number of neurons per layer which the neural network should have. Below, the errors are given for different neural network sizes.

network size	males	females	unisex
200-150-100-30-1	0.00324212	0.00205556	0.00243482
150-100-30-1	0.00344637	0.00284644	0.00240020
100-30-1	0.00372505	0.00230863	0.00337127
30-1	0.00413402	0.00229974	0.00272141
100-100-100-1	0.00377898	0.00275292	0.00259546
200-200-200-1	0.00407800	0.00179374	0.00244923
30-100-150-200-1	0.00369453	0.00216836	0.00256288

Table 2: Average root-mean-squared errors of different neural network sizes for the pure endowment insurance in scenario (C)

Here, e.g. 200-150-100-30-1 means that we have a neural network with four hidden layers consisting of 200, 150, 100 and 30 neurons, respectively. The final "1" represents the output neuron.
We can see that the size of the neural network does not really matter. Also the shape of the network, i.e. if we vary the numbers of neurons for different layers does not seem to have a big impact on the quality of training. As the size of the network did not have a big influence on our training times for our examples, we decided to keep the "standard neural network" at a medium size of 150-100-30-1.

The Optimizer

Next, we take a look at eight of the optimizers which are implemented in `keras`.

optimizer	males	females	unisex
SGD	0.00581211	0.00472995	0.00509333
Adam	0.00344637	0.00284644	0.00240020
Nadam	0.00455761	0.00229244	0.00265699
Adamax	0.00381995	0.00275219	0.00328929
Adadelta	0.01465838	0.01162739	0.01238723
Adagrad	0.00844584	0.00798959	0.00826941
Ftrl	0.01747911	0.01018802	0.01175904
RMSprop	0.00724953	0.00577141	0.00452675

Table 3: Average root-mean-squared errors of different optimizers for the pure endowment insurance in scenario (C)

One can see that Adadelta and Ftrl perform the worst on our data set, while Adam, Nadam and Adamax work the best.
The Adam optimizer got his name from the abbreviation of "adaptive moment estimation" and uses exponentially decaying averages of (the l_1- and the l_2-norm of) past and squared gradients.
Adamax uses the same approach as Adam but with the l_∞-norm instead of the l_2-norm. As the gradients are of finite dimension, we can speak of the maximum norm instead of the l_∞-norm. This variant of the optimizer was developed to ensure a more stable optimization. Indeed, for our examples the variance of the root-mean-squared errors is lower for the Adamax optimizer.
Finally, the Nadam optimizer is another variant of Adam making use of the so-called "Nesterov trick", which might result in a slightly faster convergence. In our examples, the differences are marginal.
Explanations for the other optimizers and further details as well as the exact algorithms can be found in Chapter 11 of [Gé19].

The Size of the Training Set

Of course, the size of the training set enormously affects the goodness of the approximation. We try different sizes of training sets, while keeping the test set fixed to 10,000 data points and the validation set to 10% of the size of the training set.

size of training set	males	females	unisex
10	0.24052890	0.26863978	0.17851720
100	0.05509833	0.04385099	0.04788966
1000	0.00344637	0.00284644	0.00240020
10,000	0.00141310	0.00101410	0.00102928
109,136	0.00130531	0.00107693	0.00112867

Table 4: Average root-mean-squared errors of different test set sizes for the pure endowment insurance in scenario (C)

We can see that the neural network needs approximately 1000 data points for training to work properly. It also seems that there is no increase in approximation quality, if the training set consists of more than 10,000 data points. In some cases, the results turn out to be even worse.
For the last example we used all available data points. By reserving 10,000 of our 130,050 data points for testing we are left with 120,050 points for training and validation. As mentioned above we use 10% of the points for validation leaving us a set of size $120,050 \cdot \frac{10}{11} = 109,136$ for training.

The Number of Epochs and the Learning Rate

The number of epochs and the learning rate have a fundamental effect on the quality of training.

epochs, learning rate	males	females	unisex
100, 0.0001	0.00867770	0.00869722	0.00803047
1000, 0.0001	0.00427497	0.00415512	0.00430157
10000, 0.0001	0.00193950	0.00127069	0.00143566
100, 0.001	0.00910728	0.01346069	0.01082322
1000, 0.001	0.00344637	0.00284644	0.00240020
10000, 0.001	0.00148179	0.00088758	0.00109058
100, 0.01	0.00704989	0.00539653	0.00567253
1000, 0.01	0.02846418	0.01525617	0.01988522
10000, 0.01	0.03003984	0.01552650	0.01904158

Table 5: Average root-mean-squared errors of different learning rates and epochs for the pure endowment insurance in scenario (C)

We can see that if the learning rate is not too high, more epochs result in a lower error. This keeps to be true if one takes even higher number of epochs. Hence it does not seem to be needed to develop a soft stopping rule if the learning rate is low enough. For a learning rate of 0.01, this is not longer the case. Here we are overfitting the training set for larger amounts of epochs which results in a poorer performance of the network on our test set.

The Activation Function

Finally, we look at different activation functions. As the activation function defines the output of a neuron given its inputs, also the choice of the activation functions used plays a crucial role for the performance of the regression. We try the nine different functions which are provided by `keras`, where the learning rate is set to the standard value chosen by `keras`. Using the exponential activation function results in infinite training, validation and test losses most of the time. We therefore do not display the values for this activation function in the table. If the regression terminates, it delivers root-mean-squared errors around 0.01. In Figure 23 we can see that the exponential function takes quite high values compared to the other activation functions, which might explain this problem.

activation function	males	females	unisex
relu	0.00226433	0.00142918	0.00159407
sigmoid	0.00400089	0.00244167	0.00306375
softmax	0.00345298	0.00275257	0.00281909
softplus	0.00444880	0.00240328	0.00392410
softsign	0.00359321	0.00222828	0.00249798
tanh	0.00425666	0.00376827	0.00395937
selu	0.00351502	0.00390210	0.00315989
elu	0.00344637	0.00284644	0.00240020

Table 6: Average root-mean-squared errors of different activation functions for the pure endowment insurance in scenario (C)

We can see that all activation functions perform quite similar. In our calculation, the relu activation function, i.e. $f(x) = \max(0, x)$, provides the best results. In Figure 23 below, the nine different activation functions we used are plotted. For a detailed discussion and definitions for all functions we refer once more to [Gé19]. Note that some of the functions use parameters as their input. We set all these parameters to their default value. For the selu function we take $\lambda = 1.0507$ and $\alpha = 1.6733$, all other parameters are set to one.

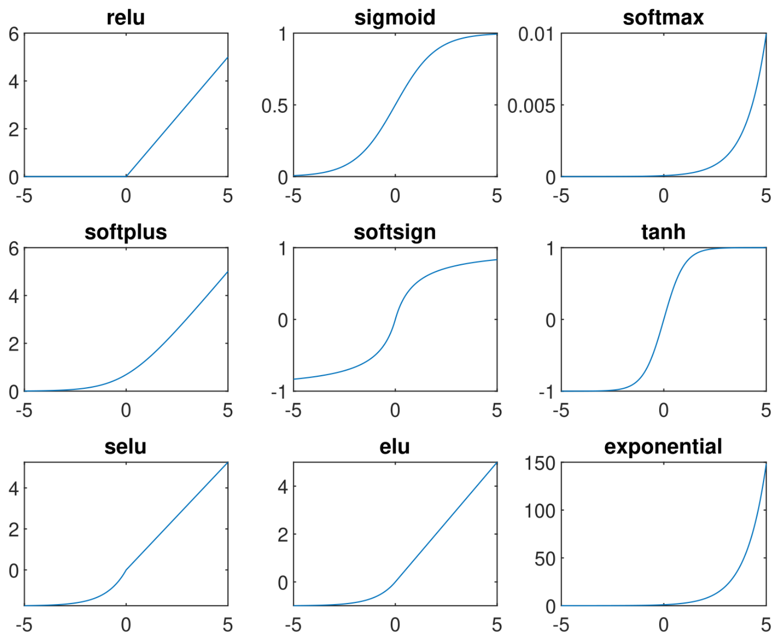


Figure 23: nine different activation functions

General Remarks

The goodness of the unisex approximation can be improved by using the gender-specific premiums as input.

input	premium
standard input	0.00240020
standard input and male premium	0.00175024
standard input and female premium	0.00117153
standard input, male and female premium	0.00101441
male and female premium	0.00078625

Table 7: Average root-mean-squared errors of the unisex premiums for the pure endowment insurance in scenario (C) using different inputs

With "standard input" we mean the input used for regressing the gender-specific and unisex premiums before, i.e. the initial age, birth year and contract duration. Interestingly, we obtain the best approximation if we only use the two gender-specific premiums as the input.
For most examples one can observe that the female error is lower than the

male one. This comes due to the effect of World War II we observed in Section 3.2. By taking out the birth years until 1920, this effect is removed and one can no longer observe a structural difference in the errors.
All neural networks used for this section make use of the mean-squared error loss function. This function performed much more stable and robust than other loss functions, so we decided to stick to the mean-squared error loss for performing the backpropagation.

4.2 Other Regression Approaches

Neural networks clearly outperform all other classic regression approaches in terms of the (average) root-mean-squared error. Even if the training of a neural network takes longer than performing the regression for most of the approaches presented here, we consider them to be the overall preferred regression tool, as they have several other useful properties. We have seen that by choosing the learning rate sufficiently low, we can avoid overfitting, see Table 5. Furthermore, the disadvantage of longer training times are not a big issue, as once the network is trained, it can be used without taking further time for training.
We therefore do not present details for the other regression approaches here but summarize them shortly. To make the results comparable, we calculate the 10-fold cross-validation score based on root-mean-squared errors for several regression methods. The so-called *k-fold cross-validation* is a common technique in machine learning. To make use of it, we randomly shuffle the data and split it up in k shares of equal size. We then perform k independent regressions, where we use $k-1$ of the shares for training and the last one for validation, each share is used for validation once. To obtain the k-fold cross-validation score, we simply average over the according (root-mean-squared) errors.
The data set on which the regressions are performed is again a set of male, female and unisex pure endowment insurance premiums in scenario (C), respectively. Just like before, it contains 130,050 data points, all combinations of birth years between 1900 and 1950, initial ages between 20 and 70 and contract durations between 1 and 50 years are regarded. We use Python's `Scikit-learn` package for implementation. The scores are presented in the following table.

regression method	males	females	unisex
linear regression	0.11918597	0.15409744	0.13613016
squared regression	0.06172708	0.06105916	0.05253646
degree 10 polynomial regression	0.01370156	0.02403878	0.01854339
linear SVR	0.14486327	0.16314460	0.14619527
squared SVR	0.11760384	0.15261114	0.13759061
degree 10 SVR	0.07741539	0.07333816	0.07022085
SGD regression	0.14838842	0.16500266	0.15445413
decision tree regression	0.03952165	0.03618673	0.03661325

Table 8: 10-fold cross-validation scores based on root-mean-squared errors of different regression methods for the pure endowment insurance in scenario (C)

If the regression method has hyperparameters which can be chosen by the user, we used a grid search approach on a suitable parameter grid to determine the optimal parameter choice. SVR stands for support vector regression and SGD regression for regression by stochastic gradient decent. Details about the methods can be found in the literature, e.g. in [Gé19].

5 Risk Classes and Risk Relations

In this chapter we are going to investigate markets with more than two risk classes. This extension opens a window for a deep analysis of risk class and contract management in insurance markets. A possible application for this type of markets lies in life or disability insurances. An analysis of different life insurance products can be found in Chapter 3. More details about disability insurances for two risk classes (males and females) using real-life data of the social security administration of the U.S. government can be found in Chapter 7.

We start by generalizing our basic model from Chapter 2 such that it can deal with more than two risk classes in Section 5.1. In a next step, we assign the risk classes to rating classes, where customers of each rating class share the same policy. This is done in Section 5.2. We derive a concept for managing these risk and rating classes in insurance markets. Different criterions for optimality of rating class vectors are presented and analyzed. The section is rounded up by presenting an algorithm and its application to real-life examples. In Section 5.3 we present an extension of our model by including capacity constraints into the market. Here, we include the option to cap the maximal amount of contracts that can be sold by each insurance company.

5.1 Risk Classes

A natural extension for our basic model is to allow for more than two risk classes. As stated in Equation (4) in Section 2.5, the expected profit of the insurer is given by the term

$$\begin{aligned}&\mathbb{E}[w_\oplus \lambda_\oplus (\pi_\oplus - Z_\oplus) + w_\ominus \lambda_\ominus (\pi_\ominus - Z_\ominus)]\\ =&w_\oplus \lambda_\oplus (\pi_\oplus - \mathbb{E}[Z_\oplus]) + w_\ominus \lambda_\ominus (\pi_\ominus - \mathbb{E}[Z_\ominus]),\end{aligned}$$

which can, after replacing $\oplus$ and $\ominus$ by the numbers 1 and 2, respectively, be rewritten as

$$\sum_{i=1}^{2} w_i \lambda_i (\pi_{i/\odot} - \mathbb{E}[Z_i]).$$

As we had $\mathbb{E}[Z_\oplus] \leq \mathbb{E}[Z_\ominus]$, we have after renaming $\mathbb{E}[Z_1] \leq \mathbb{E}[Z_2]$. In practice we assume that the equality is strict. From the representation above, we can see that the most straightforward and meaningful way of extending the model for more risk classes is to increase the number of summands. We therefore define a new set of risk classes and adapt our model of the insurance market to it.

Definition 5.1. The *set of risk classes* is modeled as a set $\mathcal{C}$, where $\mathcal{C} \subseteq \mathbb{N}$ or an interval $\mathcal{C} \subseteq \mathbb{R}$. Furthermore, $(w_i)_{i \in \mathcal{C}}$ are the corresponding *fractions* or the *density of i-agents in the market*. As before, we assume that $w_i \geq 0$ holds for all i and $\sum_{i \in \mathcal{C}} w_i = 1$ or $(w_i)_{i \in \mathcal{C}}$ is (Lebesgue-)integrable with $\int_{\mathcal{C}} w_i \mathrm{d}i = 1$, respectively. All agents of one risk class face a risk with the same distribution denoted by the random variable Z_i for agents of class $i \in \mathcal{C}$. We assume that the risks of one class are i.i.d. and independent of the risks of the other classes. The expected risk is increasing with the number of the risk class, i.e. $\mathbb{E}[Z_i]$ is increasing in i. All random variables are (Lebesgue-) square-integrable. Finally, the number of customers in the corresponding risk class is given by $\eta_i \in \mathbb{N}_0$.

While we assumed that all random variables are integrable in Chapter 2, compare Remark 2.6, we now assume square-integrability. This is needed, as we are need the variances of the random variables later, in Definition 5.74. Definition 5.1 leads us to a new definition of the profit of the insurer.

Definition 5.2. We define the expected *insurer's profit* as

$$\sum_{i=1}^{n} w_i \lambda_i (\pi_{i/\odot} - \mathbb{E}[Z_i]), \quad n \in \mathbb{N} \cup \{\infty\}$$

for the discrete case and

$$\int_{\mathcal{C}} w_i \lambda_i (\pi_{i/\odot} - \mathbb{E}[Z_i]) \mathrm{d}i$$

for the continuous case.

By setting $n = 2$ the above equation simplifies to Equation (4). With $\mathbb{N} \cup \{\infty\}$ we include the case that we might have infinitely many risk classes. Remember that λ is a function that depends on π, see Remark 2.10, so one actually has $w_i \cdot \lambda_i(\pi_{i/\odot}) \cdot (\pi_{i/\odot} - \mathbb{E}[Z_i])$ as the summands/integrand.

Remark 5.3. Similar to Chapter 2, the equilibrium demand for insurance $\hat{\lambda}$ can be obtained by solving the problem

$$\max_{\lambda \geq 0} \mathbb{E}[u(a - \lambda\pi + (1 - \lambda)Z)],$$

see Definition 2.17. Similar to Chapter 2, it is not clear in general, why this maximizer needs to exist. As mentioned, one can work out criterions for some special cases and the maximizer exists for all of our examples, compare Remark 2.19.
As before, the solution to the problem of the insured can often not be given in closed form. Nonetheless, we are again able to find some closed form representations for the insurer's problem, as we can see in Proposition 5.6.

Similar to Chapter 2, we introduce the optimal premiums for the different market scenarios by the following definition.

Definition 5.4. In a competitive market with equal tariffs (scenario (C) and regime (E)), the *optimal premium* $\hat{\pi}_\odot$ is given as the solution of

$$\sum_{i=1}^{n} w_i \hat{\lambda}_i (\pi_\odot - \mathbb{E}[Z_i]) = 0$$

or

$$\int_{\mathcal{C}} w_i \hat{\lambda}_i (\pi_\odot - \mathbb{E}[Z_i]) \mathrm{d}i = 0.$$

Under free contract design (regime (F)), the *optimal premium for i-agents* is given as the solution of

$$w_i \hat{\lambda}_i (\pi_i - \mathbb{E}[Z_i]) = 0.$$

In a monopolistic market (scenario (M)) the *optimal premium* in regime (E) is given by

$$\hat{\pi}_\odot = \arg\max_{\pi_\odot} \left\{ \sum_{i=1}^{n} w_i \hat{\lambda}_i (\pi_\odot - \pi_i^0) \right\}$$

where $n \in \mathbb{N} \cup \{\infty\}$ and

$$\hat{\pi}_\odot = \arg\max_{\pi_\odot} \left\{ \int_{\mathcal{C}} w_i \hat{\lambda}_i (\pi_\odot - \pi_i^0) \mathrm{d}i \right\}.$$

Furthermore, the *optimal premium for i-agents* is given by

$$\hat{\pi}_i = \arg\max_{\pi_i} \left\{ \hat{\lambda}_i (\pi_i - \pi_i^0) \right\}.$$

Similar to Chapter 2, π_i^0 is defined as $\pi_i^0 := \mathbb{E}[Z_i]$ and the equilibrium insurance demand $\hat{\lambda}_i$ depends on the premium π.
If we are in regime (F), where the optimization in scenario (M) is performed for each risk class individually, we speak of a *risk class specific premium* or *type-specific premium*. The same applies in case of scenario (C) where we solve the equation individually. If we are in regime (E), where this is done for all risk classes at the same time, we speak of a *global premium* or an *aggregate premium*.

Remark 5.5. The no rip-off property from Remark 2.18 ensures the existence of the (argument of the) maximums in Definition 5.4, see also Remark 2.25.

Proposition 5.6. Let $n \in \mathbb{N} \cup \{\infty\}$. In scenario (C), the optimal premiums are given by

$$\hat{\pi}_i = \pi_i^0$$

if we have free contract design and by

$$\pi_\odot = \sum_{i=1}^{n} \frac{w_i \hat{\lambda}_i}{\sum_{i=1}^{n} w_i \hat{\lambda}_i} \cdot \pi_i^0$$

or

$$\pi_{\odot} = \int_{\mathcal{C}} \frac{w_i \hat{\lambda}_i}{\int_{\mathcal{C}} w_i \hat{\lambda}_i \mathrm{d}i} \cdot \pi_i^0 \mathrm{d}i$$

if we have mandatory equal tariffs.
In scenario (M) the optimal premiums are given by

$$\hat{\pi}_i = \pi_i^0 - \frac{\hat{\lambda}_i(\hat{\pi}_i)}{\hat{\lambda}_i'(\hat{\pi}_i)}$$

for a setting with free contract design, where we need to have $\hat{\lambda}_i(\hat{\pi}_i) \neq 0$ and $\hat{\lambda}_i'(\hat{\pi}_i) \neq 0$ for the formula to be well-defined. In a regime with mandatory equal tariffs we calculate

$$\hat{\pi}_{\odot} = -\frac{\sum_{i=1}^{n} w_i[-\pi_i^0 \hat{\lambda}_i'(\hat{\pi}_{\odot}) + \hat{\lambda}_i(\hat{\pi}_{\odot})]}{\sum_{i=1}^{n} w_i \hat{\lambda}_i'(\hat{\pi}_{\odot})}.$$

and

$$\hat{\pi}_{\odot} = -\frac{\int_{\mathcal{C}} w_i[-\pi_i^0 \hat{\lambda}_i'(\hat{\pi}_{\odot}) + \hat{\lambda}_i(\hat{\pi}_{\odot})]\mathrm{d}i}{\int_{\mathcal{C}} w_i \hat{\lambda}_i'(\hat{\pi}_{\odot})\mathrm{d}i}.$$

Similar to the case with free contract design, we need to have $\hat{\lambda}_i(\hat{\pi}_i) \neq 0$ as well as $\sum_{i=1}^{n} w_i \hat{\lambda}_i'(\hat{\pi}_{\odot}) \neq 0$ or $\int_{\mathcal{C}} w_i \hat{\lambda}_i'(\hat{\pi}_{\odot})\mathrm{d}i \neq 0$ to ensure that the formula is well defined.

Proof. We start by investigating the competitive case. If we have free contract design, we have to solve

$$w_i \hat{\lambda}_i(\pi - \mathbb{E}[Z]) = 0.$$

First of all, we assume $w_i \neq 0$. Otherwise there would be no agents of this type in the market and calculating the premium is obsolete. By setting $\pi = \mathbb{E}[Z] = \pi_i^0$ we get that $\hat{\lambda}_i(\pi) = 1$, compare Example 2.31. This solves the equation. As a second option, infinitely many solutions can be obtained by choosing for such a high premium, that no insurance is bought at all, e.g. by choosing $\pi_i > \text{esssup}[Z_i]$. As the insurer is interested in selling contracts, this option can be ignored.
In a market with mandatory equal tariffs and perfect competition, we have to solve

$$\sum_{i=1}^{n} w_i \hat{\lambda}_i(\pi_{\odot} - \mathbb{E}[Z_i]) = 0.$$

Rearrangements of the equation shows

$$
\begin{aligned}
&\sum_{i=1}^{n} w_i \hat{\lambda}_i (\pi_{\odot} - \mathbb{E}[Z_i]) = 0 \\
\Leftrightarrow \quad &\sum_{i=1}^{n} w_i \hat{\lambda}_i \pi_{\odot} = \sum_{i=1}^{n} w_i \hat{\lambda}_i \mathbb{E}[Z_i] \\
\Leftrightarrow \quad &\pi_{\odot} = \sum_{i=1}^{n} \frac{w_i \hat{\lambda}_i}{\sum_{j=1}^{n} w_j \hat{\lambda}_j} \cdot \pi_i^0 .
\end{aligned}
$$

The integral case can be proven analogously.
Let us continue by looking at a monopolistic market. If we have free contract design, we need to solve

$$
\arg\max_{\pi_i} \left\{ w_i \hat{\lambda}_i(\pi_i) \cdot (\pi_i - \pi_i^0) \right\},
$$

where we recall that $\hat{\lambda}$ is a function that depends on the premium. This function is continuously differentiable in all points in which the function is strictly positive, as we have seen in Proposition 2.22. As before, we assume $w_i \neq 0$. We apply the FOC and use the product rule to arrive at

$$
\begin{aligned}
&w_i(\hat{\lambda}_i'(\pi_i)(\pi_i - \pi_i^0) + \hat{\lambda}_i(\pi_i)) = 0 \\
\Leftrightarrow \quad &\pi_i = \pi_i^0 - \frac{\hat{\lambda}_i(\pi_i)}{\hat{\lambda}_i'(\pi_i)} .
\end{aligned}
$$

Here we used $\hat{\lambda}_i(\hat{\pi}_i) \neq 0$ to ensure the differentiability and $\hat{\lambda}_i'(\hat{\pi}_i) \neq 0$ to avoid dividing by zero. Note that for most utility functions, $\hat{\lambda}_i(\hat{\pi}_i) \neq 0$ implies $\hat{\lambda}_i'(\hat{\pi}_i) \neq 0$. Similarly we calculate for the case with mandatory equal tariffs

$$
\begin{aligned}
&\sum_{i=1}^{n} w_i(\hat{\lambda}_i'(\pi_{\odot})(\pi_{\odot} - \pi_i^0) + \hat{\lambda}_i(\pi_{\odot})) = 0 \\
\Leftrightarrow \quad &\pi_{\odot} = -\frac{\sum_{i=1}^{n} w_i[-\pi_i^0 \hat{\lambda}_i'(\pi_{\odot}) + \hat{\lambda}_i(\pi_{\odot})]}{\sum_{i=1}^{n} w_i \hat{\lambda}_i'(\pi_{\odot})} .
\end{aligned}
$$

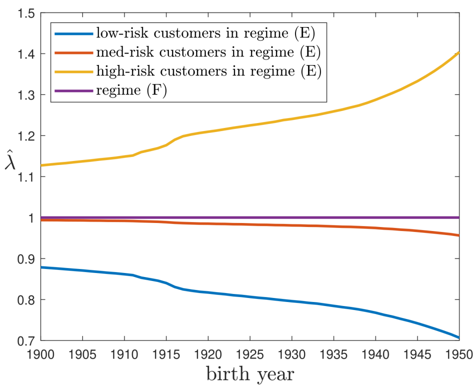
Alike in the case with free contract design $\hat{\lambda}_i(\hat{\pi}_i) \neq 0$ ensures the differentiability while $\sum_{i=1}^{n} w_i \hat{\lambda}_i'(\hat{\pi}_{\odot})$ (or $\int_{\mathcal{C}} w_i \hat{\lambda}_i'(\hat{\pi}_{\odot}) \mathrm{d}i$) ensures that we are not dividing by zero. As before, the proof works similarly for the integral case. The interchangeability of the summation/integration with the differentiation can be shown e.g. by Theorem 11.2 in Forster [For12]. □

Remark 5.7. If one breaks down Definition 5.4 and Proposition 5.6 to two risk classes, one ends up with the same premiums as given in Equations (9) and (10) for a market with perfect competition, where the insurer's profits are zero. For scenario (M), the premiums from above coincide with the premiums given by Equations (5) and (7) which maximize the insurer's profit. As in Chapter 2, the conditions on $\hat{\lambda}$ and its derivative (or the denominator) limit the practical use of the above proposition for the monopolistic case. To avoid problems later on, we do not use this proposition for calculations but stick to solving the maximization problem given in Definition 5.4 numerically.

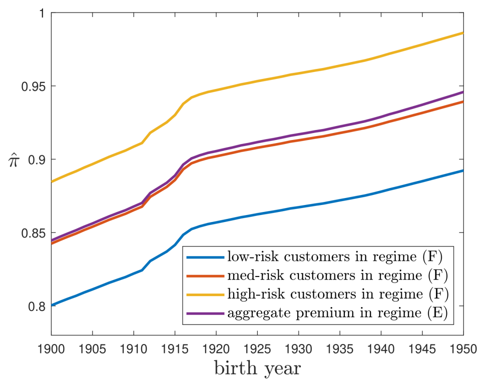
Remark 5.8. As one can see in Section 5.2, it is meaningful to group some risk classes to a so-called *rating class*, where all customers pay the same premium, see Definition 5.14. Then the premiums can be calculated using the formulas for the case with mandatory equal tariffs by taking only the risk classes belonging to the corresponding rating class into account.

We illustrate this extension with an example:

Example 5.9. We reconsider the pure endowment insurance from Example 3.1. We assume that our portfolio consists of males, which are either of medium risk (same risk as before), or their risk is enlarged or reduced by 5% (i.e. the loss in case of damage is multiplied by 0.95 or 1.05, respectively). The customers with unchanged risk might also be called customers with medium risk and are assumed to take a share of 60% of the customers, while the high- and the low-risk customers take a share of 20% each. We then compute the type-specific and the aggregate premiums and their corresponding equilibrium insurance coverages.



(a) Eq. insurance coverage $\hat{\lambda}$

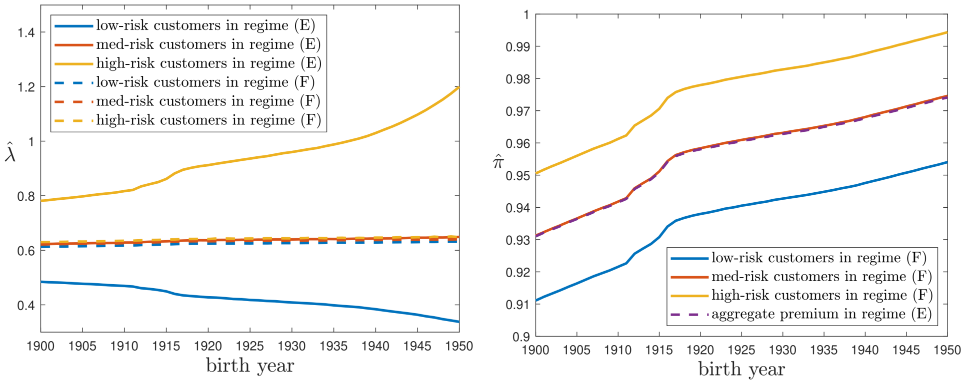


(b) Eq. premium per coverage $\hat{\pi}$

Figure 24: Equilibrium insurance coverage $\hat{\lambda}$ and equilibrium premium per coverage $\hat{\pi}$ of a pure endowment insurance as functions of the birth year in scenario (C) with three risk classes

As one would expect, the aggregate premium lies slightly above the weighted average of the type-specific premiums, which would be exactly the premium of the risk neutral customers. Similar to our previous observations, the high-risk customers are subsidized by the insured of lower risk, in this case the medium- and low-risk customers. As the high-risk customers get subsidized by two risk classes, their equilibrium coverage rises more than the coverages of the other insured decreases.

Example 5.10. We investigate the setting of the last example in a monopolistic market.



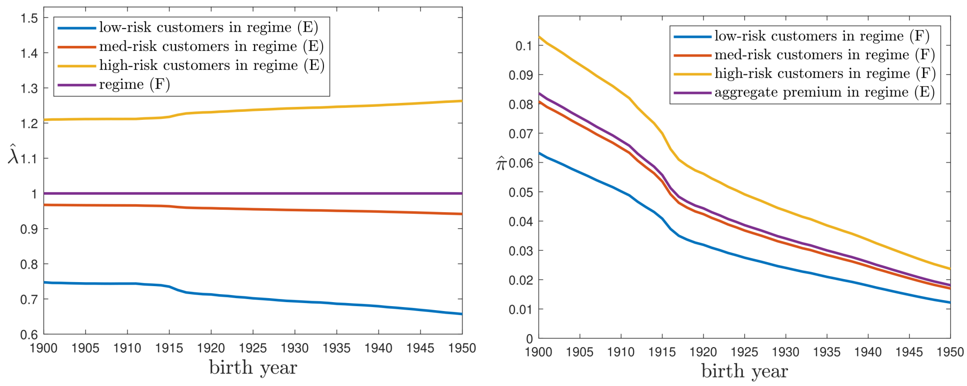
(a) Eq. insurance coverage $\hat{\lambda}$ (b) Eq. premium per coverage $\hat{\pi}$

Figure 25: Equilibrium insurance coverage $\hat{\lambda}$ and equilibrium premium per coverage $\hat{\pi}$ of a pure endowment insurance as functions of the birth year in scenario (M) with three risk classes

Most of the phenomena are similar to those which we have observed in the last example. In monopolistic markets the aggregate premium lies slightly below the average of the premiums in regime (E), compare also the figures in Section 3.2. This is due to the structure of the underlying optimization problem in the monopolistic market, see also Section 3.3.

Remark 5.11. All figures in this section are based on the birth years between 1900 and 1950, therefore all plots are based on 51 data points.

Example 5.12. Next, we analyze a product with more than one possible loss size, namely the term insurance from Example 3.2.



(a) Eq. insurance coverage $\hat{\lambda}$ (b) Eq. premium per coverage $\hat{\pi}$

Figure 26: Equilibrium insurance coverage $\hat{\lambda}$ and equilibrium premium per coverage $\hat{\pi}$ of a term insurance as functions of the birth year in scenario (C) with three risk classes

In Figure 26, we can obtain similar effects as for the pure endowment insurance from Example 5.9. Note that the differences in the equilibrium amounts of coverage are generally bigger but do not change so much over time. Compared to the example with two risk classes from Figure 11, the coverage reduction for the low-risk customers is weaker in our setting with three risk classes.

Neither in Example 5.9, nor in 5.10 or 5.12, a push-out was observable. As we can see in Example 8.2 it is even possible to observe two push-outs in a setting with three risk classes – one where the low-risk customers and a second where the medium-risk customers are pushed out of the market. More general, in a market with n risk classes, up to $n - 1$ push-outs are possible. We now assume that there are multiple insurers operating in the same insurance market. It is then important to keep track what other insurers do and how they set up their contracts. If an other insurance company has set up a new contract, which is more favorable to low-risk customers (e.g. due to lower premiums) but cannot be bought by the medium- and high-risk customers, the amount of low-risk customers in our contract will shrink. If this is not recognized in time, this can lead to the supposed equilibrium aggregate premium no longer being the equilibrium aggregate premium in reality. Instead, we end up with a premium that is too low, as the next example illustrates.

Example 5.13. We use the setting of Example 5.9 but assume that the distribution of the different risk classes in our contract has shifted from 20%, 60%, 20% for low, medium and high risk, respectively, to 10%, 70%, 20%. In the following figure we compare the equilibrium premium of our contract in the old and the new setting.

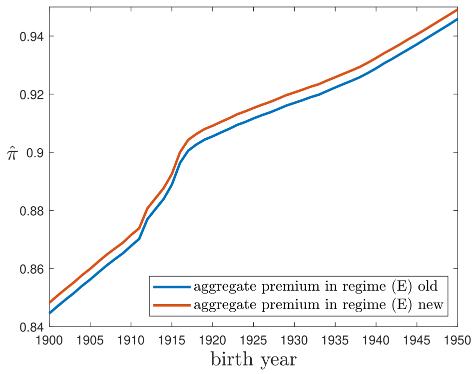


Figure 27: Equilibrium premium per coverage $\hat{\pi}$ of a pure endowment insurance as functions of the birth year in scenario (C) in the old and the new setting

As mentioned before, not recognizing the change in the risk class allocation brings us in a situation where we are still charging the old premium plotted in blue in Figure 27. Due to the change of the distribution of people in the risk classes, we would need to charge an adapted premium plotted in orange. If we continue charging the old (blue) premium instead, our insurance company could fall bankrupt in the long run, as the premiums we are charging lie below the expected losses of the new market scenario.

5.2 Rating Classes

Example 5.13 motivates us to take a closer look at how to choose the rating classes, which group the different risk classes into contracts. Our goal for this section is to provide tools for modeling this together with an analysis of optimal risk class management in insurance markets.
After setting up the basic concept in Section 5.2.1, we provide a motivation of the further modeling using a continuous-time Markov chain approach in Section 5.2.2.
To be able to compare different rating class vectors, we present a preference relation in Section 5.2.3. We define the (local) optimality of a vector with respect to this relation in Section 5.2.4, and present an example in which no optimizer exists.
We introduce an approach for mitigating the problem of non-existing optimizers in Section 5.2.5 by providing an analysis of the properties of optimal rating class vectors together with an approach for finding an optimizer in convex sets. As the sets on which we are optimizing are usually not convex, we need to find a way to apply these results on discrete sets. This is done by defining an appropriate procedure using a metric in Section 5.2.6.
In Section 5.2.7 we concretize the approach of applying the preference rela-

tion to rating class vectors and therewith present the last piece needed for our model.
Before turning to real-life examples, we summarize the strategic decisions that can be made by the insurance companies in our market in Section 5.2.8. All our previous work of this chapter is summed up when we set up an algorithm for finding optimal rating class vectors in Section 5.2.9, which is getting applied in Section 5.2.10.

5.2.1 Basic Concept

Definition 5.14. Let $f : \mathcal{C} \to \mathbb{N}$ be a function taking values in the positive integers. The sets of the form

$$r_f^i = \{c \in \mathcal{C} : f(c) = i\},\ i \in \mathbb{N}$$

are called *rating classes* where the function f is called the corresponding *rating class assignment function.* The set of sets

$$R_f = \{r_f^i : r_f^i \neq \emptyset\}$$

is called a *set of rating classes* or a *rating class set.* This set of sets of course depends on the choice of f. The *number of rating classes given* f is denoted by $m_f = |R_f|$. Finally, we call the set

$$\mathcal{R} = \{R_f : f \text{ is a rating class assignment function}\}$$

the *set of all possible rating class sets.*

If it is clear from the context, which function f we use for the rating class assignment, we might drop the dependence on f. Our interpretation of a rating class is that we assume that all customers in one rating class get assigned to one insurance contract and therefore pay the same premium.

Remark 5.15. We can calculate the (optimal) premium $\hat{\pi}_k$ of the customers in risk class k similarly to how we obtained $\hat{\pi}_\odot$ in Definition 5.4 and Proposition 5.6. We just need to exchange the set of all risk classes $\mathcal{C}$ by the set of the risk classes in rating class k.

Remark 5.16. As the term rating class already indicates, the classes rely on an insurer that rates the risk classes into rating classes and are not coalitions formed by the insured. By definition, the number of rating classes is always lower or equal than the number of risk classes.

Remark 5.17. One could interpret the set $\mathcal{R}$ also as the set of possible partitions of $\mathcal{C}$. As we see in Remark 5.19 and later on, our definition has the advantage that each element of a partition gets a naturally assigned number.

In the remainder of this chapter, $\mathcal{C}$ is going to be finite. W.l.o.g. we can assume that $\mathcal{C} = \{1, \dots, n\}$ holds for $n \in \mathbb{N}$. Instead of dealing with partitions, we work with vectors of the form $\{1, \dots, n\}^n$, as illustrated in the next definition.

Definition 5.18. Let $\mathcal{C} = \{1, \dots, n\}$, $n \in \mathbb{N}$ be finite. Let furthermore $f : \mathcal{C} \to \{1, \dots, n\}$ be a function taking positive integer values up to n. A vector of the form $(f(1), \dots, f(n))$ is called a *vector of rating classes* or a *rating class vector*.

Remark 5.19. Definition 5.18 introduces a notation that complements Definition 5.14. Each element $R_f \in \mathcal{R}$ with a rating class assignment function f that maps to the set $\{1, \dots, n\}$ gets assigned a vector of the form $(f(1), \dots, f(n)) \in \{1, \dots, n\}^n$. Given a number i, all elements of this vector that contain the same number $f(c) = i$ form a rating class. Following our interpretation, we assume that all risk classes with the same number pay the same premium.

As the actual number a rating class gets assigned is not relevant, we define an equivalence relation to characterize all relations that differ only in their numeration.

Definition 5.20. The *set of all possible rating class vectors modulo numeration* $\mathcal{R}'$ is defined as the quotient set $\mathcal{R}' := \mathcal{R}/\sim$. The *counting relation* $\sim$ is an equivalence relation given on $\mathcal{R} \times \mathcal{R}$ by

$$R_f \sim R_g :\Leftrightarrow \forall (i, j) \in \mathcal{C}^2 : f(i) = f(j) \Leftrightarrow g(i) = g(j).$$

Here f and g are functions taking values in $\{1, \dots, n\}$, see Definition 5.18 and Remark 5.19.

Remark 5.21. It is not difficult to show that $\sim$ indeed defines an equivalence relation. Knowing about the slight notation abuse, we are going to use $\mathcal{R}'$ as a set of risk relations instead of as a set of equivalence classes of risk relations in the following.

Let us illustrate the above definitions with an easy example.

Example 5.22. In the setting of the previous chapters, where we only had two risk classes (i.e. $n = 2$, all $\oplus$- and $\ominus$-agents face the same risks $Z_\oplus$ and $Z_\ominus$), the set of all possible rating class vectors $\mathcal{R}$ for the set $\mathcal{C} = \{1, 2\}$ is given by

$$\mathcal{R} = \{(1, 1), (1, 2), (2, 1), (2, 2)\}.$$

As already argued it is only interesting which risk classes are grouped up in one contract by assigning the same number to them but not which number is actually assigned to that contract. Note that for the rating class vectors in $\mathcal{R}$ it holds that $(1,1) \sim (2,2)$ as well as $(1,2) \sim (2,1)$ for our equivalence relation $\sim$. We therefore obtain

$$\mathcal{R}' = \{(1,1),(1,2)\},$$

where one could also have chosen other representatives for the rating classes, e.g. $(2,2)$ instead of $(1,1)$. Given our setting from Chapters 2 and 3, the rating class $(1,1)$ indicates that we are in regime (E), i.e. in a regime where all customers need to buy the same contract and the number of rating classes is $m = 1$. Contrariwise, $(1,2)$ indicates that we are in regime (F), where customers of each risk class are equipped with an individual contract and we obtain $m = 2$.

Proposition 5.23. Suppose there are n risk classes. The number $|\mathcal{R}'|$ of possible rating class vectors modulo numeration is given by the Bell number B_n. These numbers can be calculated recursively by

$$B_{n+1} = \sum_{k=0}^{n} \binom{n}{k} B_k, \quad B_0 = 1.$$

Using Dobinski's formula, we receive the closed form representation

$$B_n = \frac{1}{e} \sum_{k=0}^{\infty} \frac{k^n}{k!}.$$

Proof. With the n-th Bell number, we can describe the number of partitions of a set with n elements, see Bell (1934) [Bel34]. As mentioned in Remark 5.17, grouping risk classes in rating classes can also be interpreted as a partitioning problem. Since we are regarding $\mathcal{R}'$ instead of $\mathcal{R}$ the actual number a rating class gets assigned is not relevant at all, see also the explanations in Example 5.22.
Over 60 years before Bell's paper, Dobinski found a way for calculating the series $\sum_{k=0}^{\infty} \frac{n^m}{n!}$ in 1871. Without knowing, he derived an explicit formula for calculating the Bell numbers. Using his results we get

$$B_n = \frac{1}{e} \sum_{k=0}^{\infty} \frac{k^n}{k!}.$$

His findings were published in a book [Dob77] that can now be found online, available only in German. □

Example 5.24. Let us assume that we have two insurers which compete for the largest share of customers in a market. We consider an example, where we assume that the market consists of 4000 customers in 20 risk classes where each customer faces a Bernoulli distributed risk. The damage probability p_k for risk class $k \in \{1, \dots, 20\}$ is given by

$$p_k = k \cdot 0.1\%$$

and all risks are independent. Each risk class $k \in \{1, \dots, 20\}$ consists of $\eta_k = 200$ customers, i.e. all risk classes are of the same size. The two insurance companies can set up rating classes to group the customers of the different risk classes such that each rating class is assigned its own premium. In a next step we compare the equilibrium insurance coverages $\hat{\lambda}_k^1$ and $\hat{\lambda}_k^2$ the customers of risk class k would buy from Company 1 and 2 for the given premiums of their corresponding rating classes. Recall that both companies independently set up their allocation of risk classes into rating classes. The 200 customers from each risk class are split proportional according to the corresponding coverages. If, for example, the equilibrium insurance coverage for one insurer is 0.5 and for the other insurer 1.5, we assume that 50 of the 200 customers buy insurance from the first insurer, 150 from the second insurer. These numbers can be calculated by (rounding)

$$\eta_k \cdot \frac{\hat{\lambda}_k^1}{\hat{\lambda}_k^1 + \hat{\lambda}_k^2} \quad \text{and} \quad \eta_k \cdot \frac{\hat{\lambda}_k^2}{\hat{\lambda}_k^1 + \hat{\lambda}_k^2}.$$

Indeed we calculate that $50 = 200 \cdot \frac{0.5}{0.5+1.5}$ customers decide to purchase insurance from Company 1, while $150 = 200 \cdot \frac{1.5}{0.5+1.5}$ customers choose for Company 2. If both premiums for one risk class are equal, then also the equilibrium insurance coverage is equal and the customers of that risk class split up equally. We justify and formalize this concept in the next section.

5.2.2 Modeling the Customer Flow using a Markov Chain Approach

Let us consider each risk class individually. Each risk class has a population of N agents, $N \in \mathbb{N}$, who need to decide from which of the two companies on the market they wish to buy insurance. Following Irle *et al.* (2011) [IKLM11], we can model the decision process of customers by a continuous-time Markov chain. In our setting, the two types of customers are not optimists and pessimists or chartists and fundamentalists in a financial market as assumed in [IKLM11], but customers buying insurance from Company 1 or 2.

The number of customers purchasing from Company 1 can be modeled as a continuous-time homogeneous Markov chain $(Z_t^N)_t$ with state space $\{0, \dots, N\}$. Agents can freely choose from which insurer they want to buy. The switching process is modeled by a birth rate and a death rate. The birth rate ν models the customer flow from Company 2 to Company 1 while the death rate μ models the flow in the opposite direction. In other words, viewed from the perspective of insurance Company 1, the birth rate models the "new born" customers in the insurance company, while the death rate models the customers who "die" and leave the insurance company by switching to the other. [IKLM11] denotes the birth rate as λ instead of ν. We changed this notation to avoid confusion with the insurance coverage. Both rates depend on the number of customers currently buying insurance from Company 1. First, we use an extensive, N-dependent approach, i.e. Model 1 from [IKLM11]. This model looks at birth and death rates

$$\nu_i = (N-i)\left(a_1 + b\frac{i}{N}\right), \quad \mu_i = i\left(a_2 + b\frac{N-i}{N}\right). \tag{12}$$

In the above formulas, ν_i and μ_i are the birth and death rates if i of the N customers currently buy insurance from Company 1. Following [IKLM11], the (positive) parameters a_1 and a_2 describe the overall tendency to switch, which is often conceptualized as the impact of new arriving customers. The parameter b is positive and models the actual herding propensity among the agents.

By setting $X_t^N := \frac{Z_t^N}{N}$, we define a standardized process with values in $[0,1]$, which coincide with the fraction of customers in the market buying insurance from Company 1 at time t.

It is shown in [IKLM11] that the standardized processes $(X_t^N)_t$ converge to a non-random limiting process $(X(t))_t$ given by

$$X(t) = \frac{a_1}{a_1+a_2} - \left(x_0 - \frac{a_1}{a_1+a_2}\right)e^{-(a_1+a_2)t}$$

if the number of customers N in the market tends to infinity. Letting also the time go to infinity, we see that $X(t)$ converges to the deterministic fraction $\frac{a_1}{a_1+a_2}$ as $t \to \infty$. We adapt the above formula to our setting by choosing $a_1 = \hat{\lambda}_k^1$ and $a_2 = \hat{\lambda}_k^2$ as the equilibrium insurance coverages for customers of risk class k in Company 1 and 2, respectively. We then obtain, similar to Example 5.24 above, the fraction $\frac{\hat{\lambda}_k^1}{\hat{\lambda}_k^1+\hat{\lambda}_k^2}$ as the limit of the limiting process $X(t)$. This limit can be interpreted as the fraction of customers in the market buying insurance from Company 1. We formalize this in Definition 5.78.

This is done for all possible risk classes k. Hence, the total share of customers insurance Company 1 and 2 get, is given by

$$\frac{1}{n}\sum_{k=1}^{n}\frac{\hat{\lambda}_k^1}{\hat{\lambda}_k^1+\hat{\lambda}_k^2} \quad \text{and} \quad \frac{1}{n}\sum_{k=1}^{n}\frac{\hat{\lambda}_k^2}{\hat{\lambda}_k^1+\hat{\lambda}_k^2},$$

where $\hat{\lambda}^1$ and $\hat{\lambda}^2$ are the vectors containing the equilibrium coverages for all risk classes. Note that the above formula applies to the simplified case that all risk classes are of equal size, which we assume for the remainder of this chapter, unless stated otherwise.

Remark 5.25. In the general case, one needs to multiply the fraction $\frac{\hat{\lambda}_k^1}{\hat{\lambda}_k^1+\hat{\lambda}_k^2}$ by the number of customers η_k and divide the sum by the total amount of customers in the market instead of the number of risk classes, see also Remark 5.80. More formally we write

$$\frac{1}{\sum_{i=1}^n \eta_i} \sum_{k=1}^n \eta_k \cdot \frac{\hat{\lambda}_k^1}{\hat{\lambda}_k^1 + \hat{\lambda}_k^2} \quad \text{and} \quad \frac{1}{\sum_{i=1}^n \eta_i} \sum_{k=1}^n \eta_k \cdot \frac{\hat{\lambda}_k^2}{\hat{\lambda}_k^1 + \hat{\lambda}_k^2}$$

as the total share of customers buying insurance from Company 1 or 2, respectively. This is indeed a generalization, as the numbers η_k simply cancel if one has $\eta_1 = \cdots = \eta_n$.

Remark 5.26. As a possible model extension, we can take a second look at [IKLM11]. Besides the extensive, N-dependent approach, there is a second one – the non-extensive, N-independent case. In this model, the birth and death rates are given by

$$\nu_i = (N-i)\,(a_1 + bi)\,, \quad \mu_i = i\,(a_2 + b(N-i))\,. \tag{13}$$

We can reformulate Equation (13) by using $y = \frac{i}{N}$ to obtain

$$\nu_i = N^2(1-y)\left(\frac{a_1}{N} + by\right), \quad \mu_i = N^2 y\left(\frac{a_2}{N} + b(1-y)\right). \tag{14}$$

Following the argumentation in [IKLM11], we obtain two fundamental changes compared to the first model given by Equation (12). By using the representation $y = \frac{i}{N}$ again, the birth and death rate in this model can be written as

$$\nu_i = N(1-y)\,(a_1 + by)\,, \quad \mu_i = Ny\,(a_2 + b(1-y))\,.$$

From this representation, the differences can be spotted quite easy. First of all, the total number of switching per time period is now of order N^2 instead of order N. Secondly, the overall tendency of switching are now given by $\frac{a_1}{N}$ and $\frac{a_2}{N}$, instead of a_1 and a_2 decreasing the order to $\frac{1}{N}$. These two differences dramatically change the behavior of the model. The convergence of the standardized process $(X_t^N)_t$ to a non-random limiting process as before does not hold any longer. Instead, $(X_t^N)_t$ converges to a diffusion on $[0,1]$, where the drift is given by

$$\mu(y) = (1-y)a_1 - ya_2$$

while the diffusion term reads as

$$\sigma^2(y) = 2by(1-y),$$

see [IKLM11] for more details. An introduction about diffusions can be found in Karatzas and Shreve (1991) [KS91].

While the first model provides us with a deterministic fraction for our customer sharing problem, this model brings in a random factor. This is interesting, as this additional randomness could be understood as the influence of further factors to our model that effect the customers choices. As already argued above, the variable b describes the herding propensity among the agents. It opens the possibility to include factors that are not reflected by the price of the insurance into the model such as how easy it is to get your money from the insurer, whether the insurer is known to be reliable and so on. It is not clear how to model this parameter and one needs to model this factor carefully in order to have good control over its influence on the model.

5.2.3 Preference Relations

To be able to compare two rating class vectors and decide which one is preferable, we need to define a preference relation. The exact application of the relation is explained in Definition 5.77 and Section 5.2.8.

Remark 5.27. To be more precise, we are going to use the preference relation from the following definition to compare equilibrium coverage vectors, not the rating class vectors themselves. More details about this can be found in the later part of this section, see for example Definitions 5.74 and 5.77.

Definition 5.28. We consider a set of vectors of the same length n. All entries of the vectors are non-negative. We then define a relation $\sqsubseteq$ and say that vector a is *smaller than b in terms of* $\sqsubseteq$ if

$$a \sqsubseteq b :\Leftrightarrow \sum_{k=1}^{n} \frac{a_k}{a_k + b_k} \leq \sum_{k=1}^{n} \frac{b_k}{a_k + b_k},$$

where we set $\frac{0}{0} = 0$. In other words, $a \sqsubseteq b$ means that the sum of the (component wise) normalized vector a is smaller than the sum of the normalized vector b. If $a_i + b_i > 0$ for all $i \in \{1, \ldots, n\}$, this is equivalent to

$$a \sqsubseteq b \Leftrightarrow \sum_{k=1}^{n} \frac{a_k}{a_k + b_k} \leq \frac{n}{2}.$$

Remark 5.29. In practice we might want to use the equivalent definition of the relation $\sqsubseteq$ to compare two vectors which have a zero entry in the same component, compare Example 5.87. In this case, we replace the $\leq \frac{n}{2}$ by $\leq \frac{|\{i \in \{1,\ldots,n\} : a_i + b_i > 0\}|}{2}$ in the original definition. Another option would be to set the fraction $\frac{0}{0}$ to $1/2$ instead of to 0.

Remark 5.30. For $n = 1$ the relation $\sqsubseteq$ is equivalent to the relation $\leq$. This is easy to see as $a \sqsubseteq b \Leftrightarrow \frac{a}{a+b} \leq \frac{b}{a+b} \Leftrightarrow a \leq b$ if at least a or b are non-zero.

Remark 5.31. For $n = 2$ we conclude using Proposition 5.32 that $a \sqsubseteq b \Leftrightarrow a_1 a_2 \leq b_1 b_2$, compare Remark 5.34. This delivers us a characterization and structure of $\sqsubseteq$ in the two dimensional case, which corresponds to our basic model from Chapter 2 extended by a safety loading, as we can see below, e.g. in Example 5.79 and Proposition 5.56. In Proposition 5.45 we are going to present an optimality criterion for sets for which all vectors have the same component sum. Assume for example that we have the constraint $b_1 + b_2 = u$ for some non-negative real-valued constant u. We then calculate that it is optimal to set $b_1 = b_2 = \frac{u}{2}$ if one wants to have $a \sqsubseteq b$ for all vectors $a \in \mathbb{R}^n$ with $a_1 + a_2 = u$. More details and a proof can be found in Proposition 5.45.

Proposition 5.32. Using the relation from Definition 5.28, we have that for $a, b \in \mathbb{R}^n$ with $a = (a_1, \ldots, a_n)$ and $b = (b_1, \ldots, b_n)$ it holds that $a \sqsubseteq b$ is equivalent to

$$\begin{aligned} &\sum_{\{(k_1,\ldots,k_n):k_i\in\{0,1\}\}} |\{i : k_i = 0\}| \prod_{i=1}^{n} ((1-k_i)a_i + k_i b_i) \\ \leq &\sum_{\{(k_1,\ldots,k_n):k_i\in\{0,1\}\}} \frac{n}{2} \prod_{i=1}^{n} ((1-k_i)a_i + k_i b_i). \end{aligned} \tag{15}$$

The weights of the form $|\{i : k_i = 0\}|$ in the left sum as well as the weights of the form $\frac{n}{2}$ in the right sum add up to $n \cdot 2^{n-1}$.

Proof. The first part follows from rearranging the terms from the definition of

$$a \sqsubseteq b \;\Leftrightarrow\; \sum_{k=1}^{n} \frac{a_k}{a_k + b_k} \leq \sum_{k=1}^{n} \frac{b_k}{a_k + b_k}.$$

Therefore, we multiply both sides with $\prod_{i=1}^{n} a_i + b_i$ and multiply all summands out. The weights can then be obtained by counting the number of appearances of each summand. The notation with the Boolean vectors is only needed to ensure that the counting is done correctly. If $a_i \neq b_i$, we can use the alternative formulation from Remark 5.33.
For the second part, we analyze both sides separately. Firstly, we note that $|\{(k_1, \ldots, k_n) : k_i \in \{0,1\}\}| = 2^n$. From this we conclude that the sum of the weights on the right-hand side is given by $\frac{n}{2} \cdot 2^n = n \cdot 2^{n-1}$.
Secondly, we obtain by a combinatorial argument that

$$|\{(k_1, \ldots, k_n) : |\{i : k_i = 0\}| = l\}| = \binom{n}{l}.$$

Using the symmetry of the binomial coefficient, i.e. that $\binom{n}{l} = \binom{n}{n-l}$, it follows that

$$\begin{aligned}\sum_{\{(k_1,\dots,k_n):k_i\in\{0,1\}\}} |\{i : k_i = 0\}| &= \sum_{l=0}^{n} l \cdot \binom{n}{l} \\ &= \sum_{l=0}^{n/2} l \cdot \binom{n}{l} + \sum_{l=n/2+1}^{n} l \cdot \binom{n}{n-l} \\ &= \sum_{l=0}^{n/2} n \cdot \binom{n}{l} \\ &= n \cdot \frac{1}{2} \cdot 2^n = n \cdot 2^{n-1}\end{aligned}$$

if n is even, where we used the symmetry again. If n is odd, the computation works similarly by taking $\lfloor n/2 \rfloor$ and $\lceil n/2 \rceil$ as the end and starting point of the splitted sum. □

Remark 5.33. If $a_i \neq b_i$ holds for all i, we can rewrite Equation (15) as

$$\sum_{\{(k_1,\dots,k_n):k_i\in\{a_i,b_i\}\}} |\{i : k_i = a_i\}| \prod_{i=1}^{n} k_i \leq \sum_{\{(k_1,\dots,k_n):k_i\in\{a_i,b_i\}\}} \frac{n}{2} \prod_{i=1}^{n} k_i.$$

Remark 5.34. Proposition 5.32 derives an alternative definition for the relation $\sqsubseteq$. It furthermore provides us with the insight that one can also understand the relation as a weighted sum of the "mixed" vectors, where $a \sqsubseteq b$ holds if the inequality from Equation (15) with shifted weights in the sum is fulfilled. The weights are shifted in such a way that the weight is larger if the proportion of a is higher.
If n is even, some of the terms on both sides of the inequality cancel. As the left-hand side solely allows for integer weights, this is only possible for the case that n is even.
Let us focus on the case that $n = 2$. We therefore consider two vectors, $a = (a_1, a_2),\ b = (b_1, b_2) \in \mathbb{R}^2_{\geq 0}$. According to Proposition 5.32 we then have that

$$a \sqsubseteq b \ \Leftrightarrow\ 0 \cdot b_1b_2 + 1 \cdot a_1b_2 + 1 \cdot a_2b_1 + 2 \cdot a_1a_2 \leq b_1b_2 + a_1b_2 + b_1a_2 + a_1a_2.$$

As already stated in Remark 5.31, elimination provides

$$a \sqsubseteq b \ \Leftrightarrow\ a_1a_2 \leq b_1b_2.$$

In order to improve our understanding of the relation, we calculate all vectors which are equivalent in the sense of $\sqsubseteq$.

Definition 5.35. Two vectors $a, b \in \mathbb{R}^n$ are *equivalent in the sense of* $\sqsubseteq$ if $a \sqsubseteq b$ and $b \sqsubseteq a$ holds. We denote this equivalence by $a \equiv b$.

The set of vectors $\mathcal{E}(a)$ which are equivalent to the vector a is called the *indifference set (of a w.r.t. $\sqsubseteq$)* and is defined by

$$\mathcal{E}(a) := \left\{\tilde{a} \in \mathbb{R}^n_{\geq 0} : \tilde{a} \equiv a\right\}.$$

Remark 5.36. Our definition of indifference sets is equivalent to defining it by

$$\mathcal{E}(a) = \{\tilde{a} : f_{\tilde{a}}(a) = f_a(\tilde{a})\} = \left\{\tilde{a} : f_a(\tilde{a}) = \frac{n}{2}\right\},$$

where

$$f_b(a) = \sum_{i=1}^{n} \frac{a_i}{a_i + b_i}. \tag{16}$$

This can easily be seen by the definition of the risk relation $\sqsubseteq$ from Definition 5.28. Furthermore, $a_i + b_i > 0$ needs to hold for all i or we need to set $\frac{0}{0}$ to $1/2$, if we use the set relying on the alternative definition of $\sqsubseteq$, compare Remark 5.29.

Remark 5.37. By maximizing f_b in a, we find the *optimal respond* to the vector b in a set $\tilde{\mathcal{R}}$:

$$f^{\text{or}}(b) = \max_a \{f_b(a) : a \in \tilde{\mathcal{R}}\}.$$

Of course, the maximizer does not need to exist in general. If $\tilde{\mathcal{R}}$ is finite, the maximum exists. The function f_b is concave in a, compare Proposition 5.50. This provides us with a lot of tools from convex optimization to find the maximizer. Furthermore, if the maximizer exists and the vector b has only strictly positive entries, f_b is strictly concave and the maximizer is unique.

Example 5.38. In the case of $n = 2$, calculating the indifference sets is an easy task: We can compute all vectors $\tilde{a} = (\tilde{a}_1, \tilde{a}_2)$ that are equivalent to a by letting $\tilde{a}_1$ be arbitrary and setting $\tilde{a}_2 = \frac{a_1 a_2}{\tilde{a}_1}$. Therefore we obtain the indifference set

$$\mathcal{E}(a) = \left\{(\tilde{a}_1, \tilde{a}_2) : \tilde{a}_1 \in \mathbb{R}_{>0},\ \tilde{a}_2 = \frac{a_1 a_2}{\tilde{a}_1}\right\}.$$

In Figure 28 the indifference sets $\mathcal{E}((0.3, 0.1))$, $\mathcal{E}((0.2, 0.2))$, $\mathcal{E}((0.3, 0.3))$, $\mathcal{E}((0.3, 0.5))$ and $\mathcal{E}((0.5, 0.5))$ are shown.

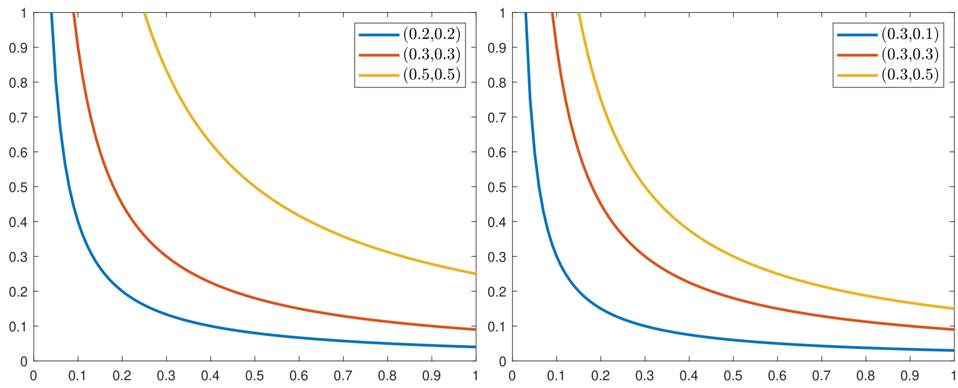


Figure 28: Indifference sets $\mathcal{E}(a)$ for five different vectors a

5.2.4 (Non-)existence of an Optimizer

Returning to the original problem, we aim to find a vector b, which is the maximizer of the set A of all vectors in the sense that $a \sqsubseteq b$ for all $a \in A$. Looking at Figure 28 above one could assume that in the case of $n = 2$ one can find a unique optimizer if the set A is convex, see Proposition 5.56 below. In the proof of this proposition we use that the intersection of two convex sets in $\mathbb{R}^n$ needs to have zero, one or infinitely many points. Before continuing we formalize the concept of a maximizer of a set A.

Definition 5.39. A vector b is called *(globally) optimal* or a *(global) maximizer* of A if it holds that $a \sqsubseteq b$ for all $a \in A$. Furthermore, a set $\mathcal{O} \subseteq A$ is called a set of *locally optimal* vectors if for all $a \in A$ there exists a vector $b \in \mathcal{O}$ such that $a \sqsubseteq b$. The smallest such set in terms of cardinality is called a *minimal set of locally optimal vectors.*

Remark 5.40. If a vector b is globally optimal, the set $\{b\}$ is also a one-elemental set of locally optimal vectors and therefore also the minimal set of locally optimal vectors. Of course, we are not interested in finding an arbitrary set of locally optimal vectors, but the smallest set of optimal vectors. Indeed, finding a set of locally optimal vectors is trivial, as A itself is by definition always a set of locally optimal vectors.

A definition for a coverage relation $\sqsubseteq_{\text{cov}}$ to compare the coverage vectors given the corresponding risk class vectors is given in Definition 5.77, see also Definition 5.74. After defining the coverage relation we introduce a concept for optimality that is similar to the concept derived in Definition 5.39.
It is easy to show that a global maximizer does not need to exist in general.

Example 5.41. Consider the set

$$A_1 := \{(1, 2, 0), (0, 1, 2), (2, 0, 1)\}.$$

Then

$$(0,1,2) \sqsubseteq (1,2,0),$$
$$(1,2,0) \sqsubseteq (2,0,1),$$
$$(2,0,1) \sqsubseteq (0,1,2),$$

but not vice versa and therefore a global maximizer cannot exist.

As we can see in Example 5.87, also in a real-life example it is possible that an optimal vector does not exist.
In a next step, we are interested in conditions which are sufficient and necessary for the existence of a maximizer. Additionally, in case of existence, we aim for finding a way to find this maximizer.

5.2.5 Optimizing on Convex Sets and Hulls

We try to mitigate the problem of a possibly non-existing (global) maximizer by optimizing on the convex hull conv$A$ instead of A. This is one step of our procedure to find optimal points of non-convex sets presented in the beginning of Section 5.2.6. A second approach to overcome the issue of non-existing optimizers is given in Section 5.3, compare Remark 5.104.

Definition 5.42. The *convex hull* conv$X$ of a set X which is the subset of a real or complex vector space V is given as the (unique) smallest convex set containing X.

Remark 5.43. Numerous equivalent definitions for the convex hull can be found in the literature, e.g. by defining it as the intersection of all convex sets containing X or as the set of all convex combinations of points in X.

Example 5.44. The set A_1 from Example 5.41 is of course not convex. The convex hull of A_1, the set convA_1 is given by the triangle spanned by the points $(1,2,0), (0,1,2), (2,0,1)$. More formally we write

$$\text{conv}A_1 = \{\delta_1(1,2,0) + \delta_2(0,1,2) + \delta_3(2,0,1) : \sum_{i=1}^{3} \delta_i = 1,\ \delta_1, \delta_2, \delta_3 \geq 0\}.$$

To ease further computations we now consider

$$A_2 := \{(1,2,0), (2,1,0), (0,1,2), (0,2,1), (2,0,1), (1,0,2)\}$$

instead of A_1. The set A_2 contains all possible permutations of 0, 1 and 2 in one vector. It is then easy to show that the equality

$$\text{conv}A_2 = \{(x,y,z) : \ x+y+z = 3,\ 0 \leq x,y,z \leq 2\}$$

holds.

Before continuing with the example, we state a proposition which helps us to find the maximizer of the set convA_2 from the example above.

Proposition 5.45. For $u \in \mathbb{R}_{\geq 0}$ we consider the set

$$A = \{x \in \mathbb{R}^n : x_i \geq 0 \text{ for all } i,\ \|x\|_1 = u\},$$

i.e. the set of all vectors with the same component sum u and non-negative entries. In this set, the optimal vector problem under $\sqsubseteq$ is uniquely solved by the vector $(u/n, \dots, u/n)$. In other words, $a \sqsubseteq (u/n, \dots u/n)$ holds for all $a \in A$.

Proof. Firstly, we notice similar to Remark 5.36 that for $b \in A$ the condition $a \sqsubseteq b$ for all $a \in A$ is equivalent to the condition that b maximizes the function f_b given by

$$f_b(a) = \sum_{i=1}^{n} \frac{a_i}{a_i + b_i}$$

in A with $b \in \mathbb{R}^n_{\geq 0}$. Note that the parameter of f_b and its optimizer are indeed the same vector. This follows directly by the definition of $a \sqsubseteq b$ as

$$a \sqsubseteq b \;\Leftrightarrow\; \sum_{i=1}^{n} \frac{a_i}{a_i + b_i} \leq \frac{n}{2} = \sum_{i=1}^{n} \frac{b_i}{b_i + b_i}.$$

In a second step we need to show that the vector $(u/n, \dots, u/n)$ is the unique optimizer of

$$f_{(u/n,\dots,u/n)}(a) = \sum_{i=1}^{n} \frac{a_i}{a_i + \frac{u}{n}}$$

in A. To do so, we use the condition $\|a\|_1 = u$ from the definition of A. This condition is equivalent to $\sum_{i=1}^{n} a_i = u$, as all a_i are real-valued and non-negative. We reformulate this condition to obtain $a_n = u - \sum_{i=1}^{n-1} a_i$. By plugging this into the definition of $f_{(u/n,\dots,u/n)}$ we conclude that maximizing $f_{(u/n,\dots,u/n)}$ in A is equivalent to maximizing the function

$$\tilde{f}(a_1, \dots, a_{n-1}) = \sum_{i=1}^{n-1} \frac{a_i}{a_i + \frac{u}{n}} + \frac{u - \sum_{i=1}^{n-1} a_i}{u - \sum_{i=1}^{n-1} a_i + \frac{u}{n}}$$

in the set $\mathbb{R}^{n-1}_{\geq 0}$, where also $u - \sum_{j=1}^{n-1} a_j \geq 0$.
This maximization is performed by a classic first-order-condition approach. For the i-th component we receive the equation

$$\frac{\partial}{\partial a_i} \tilde{f}(a_1, \dots, a_{n-1}) = \frac{\frac{u}{n}}{\left(a_i + \frac{u}{n}\right)^2} - \frac{\frac{u}{n}}{\left(u - \sum_{j=1}^{n-1} a_j + \frac{u}{n}\right)^2} = 0.$$

This equation is equivalent to

$$\left(a_i + \frac{u}{n}\right)^2 = \left(u - \sum_{j=1}^{n-1} a_j + \frac{u}{n}\right)^2. \tag{17}$$

We now remove the squares from both sides in Equation (17). By taking the root, we need to make sure that the expressions that are squared on either sides are non-negative. We therefore conclude that

a) The expression $a_i + \frac{u}{n}$ is non-negative as all a_i are non-negative.

b) Also the expression $u - \sum_{j=1}^{n-1} a_j + \frac{u}{n}$ is non-negative, as $\sum_{j=1}^{n-1} a_j \leq u$.

Hence, Equation (17) is equivalent to

$$\begin{aligned} a_i + \frac{u}{n} &= u - \sum_{j=1}^{n-1} a_j + \frac{u}{n} \\ \Leftrightarrow \quad \sum_{j=1}^{n-1} a_j + a_i &= u. \end{aligned}$$

From the system of equations $\sum_{j=1}^{n-1} a_j + a_i = u$ for $i = 1, \ldots n-1$ we obtain $a_1 = \cdots = a_{n-1} = \frac{u}{n}$. Checking the second derivatives delivers that the $(n-1)$-dimensional vector $(u/n, \ldots, u/n)$ is indeed a maximizer of $\tilde{f}$. From the definition of a_n above we can conclude that the n-dimensional vector $(u/n, \ldots, u/n)$ is a maximizer of $f_{(u/n,\ldots,u/n)}$.
Finally, we prove that the vector $(u/n, \ldots, u/n)$ is a global optimizer. Therefore we need to check the edge cases, i.e. the points $(a_1, \ldots, a_n)$ where at least one a_i equals zero. Due to symmetry, we can choose w.l.o.g. a_1 to be zero. Note that $f_{(u/n,\ldots,u/n)}(u/n, \ldots, u/n) = \frac{n}{2}$. If we have that $a_1 = 0$, it is optimal to choose $a_2 = \cdots = a_n = \frac{u}{n-1}$ with the same argument as above. We then compute

$$\begin{aligned} f_{(u/n,\ldots,u/n)}(0, u/(n-1), \ldots, u/(n-1)) =& (n-1) \cdot \frac{\frac{u}{n-1}}{\frac{u}{n-1} + \frac{u}{n}} \\ =& (n-1) \cdot \frac{u}{\frac{un+u(n-1)}{n}} \\ =& (n-1) \cdot \frac{un}{un + u(n-1)} \\ <& (n-1) \cdot \frac{un}{u(n-1) + u(n-1)} \\ =& \frac{n}{2} = f_{(u/n,\ldots,u/n)}(u/n, \ldots, u/n), \end{aligned}$$

so $(0, u/(n-1), \ldots, u/(n-1))$ cannot be optimal as

$$f_{(u/n,\ldots,u/n)}(0, u/(n-1), \ldots, u/(n-1)) < f_{(u/n,\ldots,u/n)}(u/n, \ldots, u/n)$$

holds. With a similar argument we can also show that if two or more a_i are zero, one can only achieve smaller function values than $n/2$. Hence, $(u/n, \ldots, u/n)$ is the unique optimizer of $f_{(u/n,\ldots,u/n)}$ in the set A, so the optimal vector problem under $\sqsubseteq$ is solved by $(u/n, \ldots, u/n)$. □

Remark 5.46. As an alternative proof, one could use a Lagrange multiplier approach instead of reformulating the constraint and plugging it into the function.

Remark 5.47. The result from Proposition 5.45 is in line with our characterization for the two-dimensional case in Remark 5.31.

Example 5.48. According to Proposition 5.45, we find that the point $(1, 1, 1)$ is the unique maximizer of the set $\{x \in \mathbb{R}^3_{\geq 0} : \|x\|_1 = 3\}$. Indeed, the point $(1, 1, 1)$ also fulfills the condition that each entry of the vector lies in the interval $[0, 2]$ which implies $(1, 1, 1) \in \text{conv}A_2$, so $(1, 1, 1)$ is also the unique optimizer of the set $\text{conv}A_2$ from Example 5.44. Note that we can represent $(1, 1, 1)$ in $\text{conv}A_2$ (and also in $\text{conv}A_1$) by taking

$$(1, 1, 1) = \frac{1}{3}(1, 2, 0) + \frac{1}{3}(0, 1, 2) + \frac{1}{3}(2, 0, 1).$$

Furthermore, we can show that for a point $(x, y, z) \in \text{conv}A_2$ the smaller the distance in the sense of Definition 5.66 between (x, y, z) and $(1, 1, 1)$ is, the greater the value $f_{(1,1,1)}(x, y, z)$ becomes. Recall that we defined f_b in Equation (16) by

$$f_b(a) = \sum_{i=1}^{n} \frac{a_i}{a_i + b_i}.$$

All points with the same distance to $(1, 1, 1)$ have the same function value. In other words, the intersection between the indifference sets of different points and the set $\text{conv}A_2$ are the circles around $(1, 1, 1)$ (intersected with $\text{conv}A_2$). Note that the circles do not look like the circles induced by the Euclidean norm but more like a "crumbled" version of them. In Figures 29 and 30 below, two circles of points satisfying $f_{(1,1,1)}(x, y, z) = \frac{8}{6}$ or $f_{(1,1,1)}(x, y, z) = \frac{7}{6}$ are given. The hexagonal surface in the two figures is the set $\text{conv}A_2$.

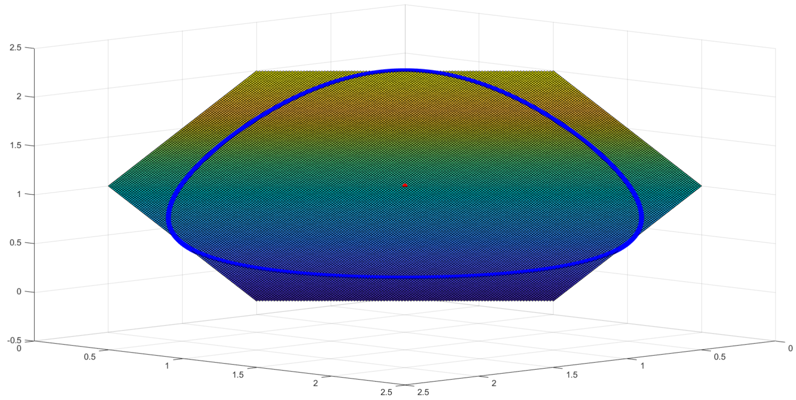


Figure 29: Indifference set (blue curve) around $(1,1,1)$ (red) for function value $\frac{8}{6}$ under the constraint $x+y+z=3$

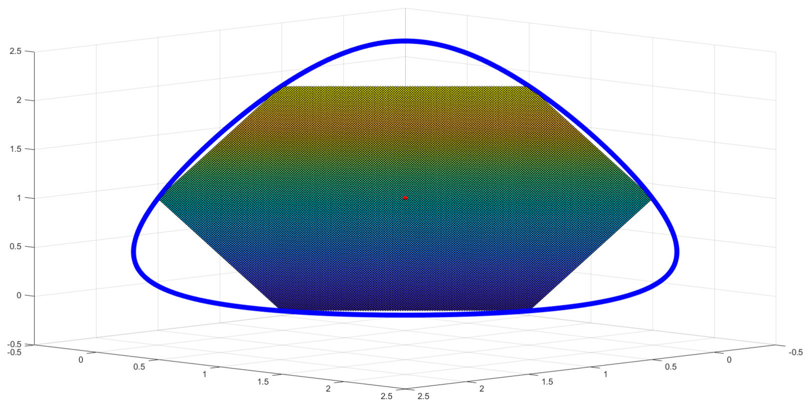


Figure 30: Indifference set (blue curve) around $(1,1,1)$ (red) for function value $\frac{7}{6}$ under the constraint $x+y+z=3$

It can be seen that some points of the indifference sets (blue curves) lie outside the set convA_2 (rainbow shaded hexagon). Therefore, one needs to intersect convA_2 with the indifference circles to obtain the points in convA_2 that are indifferent for a given function value. We decided to plot the whole indifference sets in Figures 29 and 30 and not only the intersections with convA_2 so that one gets a feeling how the curves actually look. If we optimize in $\{(x,y,z) \in \mathbb{R}^3_{\geq 0} : x+y+z=3\}$ instead of in convA_2, taking the intersection is not needed.

Remark 5.49. One can analyze the setting presented here also from a game theoretical point of view. The set A_1 from the Examples 5.41 and 5.44 can be seen as three possible strategies of a game. We compare the setting of these examples with the analysis of the game "Rock, Paper, Scissors". Two players choose their strategy from the set {Rock, Paper, Scissors}. Each strategy wins and loses against one other strategy: Paper beats Rock, Rock beats Scissors and Scissors beats Paper. A strategy ties if both players decide to pick the same symbol. By using $\mathrm{conv} A_1$ instead of A_1 we allow for mixed strategies instead of only pure ones. Various sources from the game theory literature show that it is a Nash equilibrium to play each symbol randomly, where each symbol is played on average every three moves. This strategy coincides with our point $(1, 1, 1)$, where every "pure strategy" of the original set A_1 receives a weight of $\delta_{1,2,3} = \frac{1}{3}$.

Proposition 5.50. The function $f_b : \mathbb{R}^n_{\geq 0} \to \mathbb{R}$ defined in Equation (16) by

$$f_b(a) = \sum_{i=1}^{n} \frac{a_i}{a_i + b_i},$$

$b \in \mathbb{R}^n_{\geq 0}$ is concave in a. If one restricts b to only have strictly positive entries, f_b is strictly concave.

Proof. Similar as in the proof of Proposition 5.45 we can show that f_b is twice continuously differentiable with

$$\frac{\partial}{\partial a_i} f_b(a) = \frac{b_i}{(a_i + b_i)^2}$$

for each $i = 1, \ldots, n$. Furthermore,

$$\frac{\partial^2}{\partial a_i^2} f_b(a) = -\frac{2b_i}{(a_i + b_i)^3}$$

and

$$\frac{\partial^2}{\partial a_i a_j} f_b(a) = 0$$

for $i, j = 1, \ldots, n$ with $i \neq j$. Therefore the Hessian matrix of f_b is given by

$$H_{f_b}(a) = \operatorname{diag}\left(-\frac{2b_1}{(a_1 + b_1)^3}, \ldots, -\frac{2b_n}{(a_n + b_n)^3}\right).$$

Due to the non-negativity of a and b, H_{f_b} is therefore negative semi-definite and hence f_b is concave. If b is strictly positive, H_{f_b} is even negative definite and therefore f_b is strictly concave in this case. □

Remark 5.51. Similar to Proposition 5.50 one can show that $f_b(a)$ is convex in b. Alternatively one can argue that $f_b(a)$ has to be convex in b, as $f_a(b)$ is concave in b and

$$f_b(a) = |\{i \in \{1, \ldots, n\} : a_i + b_i > 0\}| - f_a(b)$$

holds, see also Remark 5.36.

Remark 5.52. The (strict) concavity of f_b (i.e. convexity of $-f_b$) in a ensures that finding a rating class vector which is the optimal response to the choice of a competing insurer is a maximization problem on a concave function (minimization problem on a convex function), see also Remark 5.37 Section 5.2.8. This type of problems has some useful properties, e.g. we know that a local optimizer is always a global optimizer which is unique if f_b is strictly concave. There is a broad literature about convex optimization providing us with various tools to tackle this problem.

The convexity of $-f_b$ can be used to apply the following theorem known from the literature, e.g. Proposition 2.1.2 in Bertsekas (1999) [Ber99].

Proposition 5.53. Let X be a non-empty and convex set and $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable over X. If x^* is a local minimum of f over X, then

$$\nabla f(x^*)'(x - x^*) \geq 0$$

for all $x \in X$.
If furthermore f is convex on X, the above condition is also sufficient for x^* to be a local minimum of f over X.

Remark 5.54. Using the convexity of $-f_b$ and the fact that the set

$$A = \{x \in \mathbb{R}^n : x_i \geq 0 \text{ for all } i,\ \|x\|_1 = u\},\ u \in \mathbb{R}_{\geq 0},$$

from Proposition 5.45 is convex, we can apply Proposition 5.53 on $-f_b$ to obtain an alternative proof for Proposition 5.45.

For dimension two, i.e. a market where there are only two risk classes, say males or females, there exists a closed-form solution for the problem of finding a global optimizer for our optimal vector problem in convex sets. To find it, recall a proposition where we use the convexity of a function f to show that the epigraph of f, i.e. the set of all points lying on or above the graph, is convex as well. More formally, we state the following proposition which can be found in standard analysis literature, e.g. in Section 3.1.7 in Boyd and Vandenberghe (2004) [BV04].

Proposition 5.55. Let the *epigraph* epi$f$ *of* $f : X \to \mathbb{R}$ be defined by

$$\text{epi}f := \{(x, \mu) \in X \times \mathbb{R} : f(x) \leq \mu\} \subseteq X \times \mathbb{R}.$$

Then a function is (strictly) convex if and only if epi$f$ is a (strictly) convex set.

This leads us to the following proposition.

Proposition 5.56. Set $n = 2$ and let $A \subseteq \mathbb{R}^2_{\geq 0}$ be compact and convex. Then, the optimal vector problem has a unique solution in A which is given by the vector that has a maximal product of components, i.e. the solution is given by

$$\arg\max\{a_1 a_2 : a = (a_1, a_2) \in A\}.$$

Proof. We already know from Example 5.38 that the indifference sets for $n = 2$ are given by the sets

$$\mathcal{E}(a) = \left\{(\tilde{a}_1, \tilde{a}_2) : \tilde{a}_1 \in \mathbb{R}_{\geq 0},\ \tilde{a}_2 = \frac{a_1 a_2}{\tilde{a}_1}\right\},$$

i.e. by the graphs of the function family $g_z(a) = \frac{z}{a}$, where the function parameter z is given by $z = a_1 a_2$. The function g_z is strictly convex for all z, hence the epigraph epig_z is also strictly convex.
We assumed that the set of vectors $A \subseteq \mathbb{R}^2_{\geq 0}$ we are optimizing over is compact and convex. It is easy to show that the intersection of two (closed) convex sets is either empty, one-elemental or has uncountably many elements. As A is compact, we know that the function $h(x, y) = xy$ has a maximizer $(\tilde{x}, \tilde{y})$ on A. Choose $z = \tilde{x}\tilde{y}$. Then $(\tilde{x}, \tilde{y}) \in \text{epi}g_z$ and therefore $(\tilde{x}, \tilde{y}) \in \text{epi}g_z \cap A$.
All the points on the boundary of epig_z have the same product z of their two components. Assume now, there are two (or more) points a_1 and a_2 in A which maximize h, i.e. with component product z. Due to the structure of A and h, the points need to lie on the boundary of A. Then, a_1 and a_2 both lie in epi$g_z \cap A$, which is convex. Due to the convexity, also $\frac{a_1+a_2}{2}$ lies in epi$g_z \cap A$. The strict convexity of epig_z ensures that $\frac{a_1+a_2}{2}$ does not lie on the boundary, but in the interior of epig_z. Hence, the product of the components of $\frac{a_1+a_2}{2}$ is greater than z. This implies that a_1 and a_2 do not maximize the product function h, which is a contradiction. □

Remark 5.57. The results from Propositions 5.45 and 5.56 coincide, if we optimize on the set $A = \{b \in \mathbb{R}^2_{\geq 0} : b_1 + b_2 = u\}$, where $u \geq 0$. While Proposition 5.45 guarantees that $(u/2, u/2)$ is optimal in A, Proposition 5.56 ensures optimality for the vector in A which maximizes its component product. Indeed, the vector $(u/2, u/2)$ is the vector with the maximum component product in A.

For spaces with higher dimension, we could not find a closed form solution as Proposition 5.56. Nonetheless, we can provide some analysis to overcome the restriction from Proposition 5.45, i.e. that all vectors in our set need to have the same component sum.
Suppose we optimize over a set A which contains a vector of the form $(u/n, \ldots, u/n)$. By Proposition 5.45 we know that for this vector it holds

that $(u/n, \dots, u/n) \sqsupseteq x$ for all vectors x with $\|x\|_1 \leq u$. Nonetheless, there might be a vector $y \in A$ with $\|y\|_1 = w > u$ which satisfies $x \sqsubseteq y$ or equivalently $f_x(y) \geq \frac{n}{2}$. We are interested in characterizing these vectors.

Proposition 5.58. Suppose there are two vectors, $x, y \in \mathbb{R}^n_{\geq 0}$ with $\|x\|_1 = u$ and $\|y\|_1 = w$, where $w > u$. Let us furthermore assume that x is given by $x = (u/n, \dots, u/n)$. Then $x \sqsubseteq y$ can only hold if

$$y_i \in [a, b],$$

where

$$a = \frac{-\left(-\frac{nw}{2} + \frac{nu}{2} - u\right) - \sqrt{\left(-\frac{nw}{2} + \frac{nu}{2} - u\right)^2 - (wu(2-n) + (n+1)u^2)}}{n}$$

and

$$b = \frac{-\left(-\frac{nw}{2} + \frac{nu}{2} - u\right) + \sqrt{\left(-\frac{nw}{2} + \frac{nu}{2} - u\right)^2 - (wu(2-n) + (n+1)u^2)}}{n}.$$

Proof. We prove the condition for $i = 1$, the proof for all other y_i can be done equivalently. Let us fix y_1 and set

$$y_2 = \dots = y_n = \frac{w - y_1}{n - 1}.$$

By construction it holds that $\|y\|_1 = w$. Furthermore, y is defined in such a way that $(y_2, \dots, y_n)$ maximizes $f_{(u/n,\dots,u/n)}$ in $\mathbb{R}^{n-1}_{\geq 0}$. This can be argued by following an equivalent argument as in the proof of Proposition 5.45.
It is of interest how big or small we can choose the components of y such that

$$f_{(u/n,\dots,u/n)}(y) \geq \frac{n}{2}$$

still holds. Pay attention that the vector $(u/n, \dots, u/n)$ we use as a reference point for $f_{(u/n,\dots,u/n)}$ can have $n-1$ or n elements in this proof. Of course, the length is chosen such that the function is well defined and fits to the length of the given input. Based on our previous argumentation, we need to solve $f_{(u/n,\dots,u/n)}(y) \geq \frac{n}{2}$. By construction of y we obtain

$$\begin{aligned} f_{(u/n,\dots,u/n)}(y) &= \frac{y_1}{y_1 + \frac{u}{n}} + \sum_{i=2}^{n} \frac{\frac{w-y_1}{n-1}}{\frac{w-y_1}{n-1} + \frac{u}{n}} \\ &= \frac{y_1}{y_1 + \frac{u}{n}} + (n-1)\frac{\frac{w-y_1}{n-1}}{\frac{w-y_1}{n-1} + \frac{u}{n}}. \end{aligned}$$

Hence, we need to solve

$$\frac{y_1}{y_1 + \frac{u}{n}} + (n-1)\frac{\frac{w-y_1}{n-1}}{\frac{w-y_1}{n-1} + \frac{u}{n}} = \frac{n}{2}$$

in y_1. The above equation is equivalent to finding the roots of a quadratic polynomial, i.e. to solving

$$\frac{n}{2}y_1^2 + \left(-\frac{nw}{2} + \frac{nu}{2} - u\right) y_1 + \left(\frac{wu(2-n)}{2n} + \frac{u^2}{2} + \frac{u^2}{2n}\right) = 0$$

in y_1. We can use the quadratic formula that provides us with the solutions

$$a = \frac{-\left(-\frac{nw}{2} + \frac{nu}{2} - u\right) - \sqrt{\left(-\frac{nw}{2} + \frac{nu}{2} - u\right)^2 - (wu(2-n) + (n+1)u^2)}}{n}$$

and

$$b = \frac{-\left(-\frac{nw}{2} + \frac{nu}{2} - u\right) + \sqrt{\left(-\frac{nw}{2} + \frac{nu}{2} - u\right)^2 - (wu(2-n) + (n+1)u^2)}}{n}.$$

If $y_1 \notin [a, b]$, we even receive $f_{(u/n,\dots,u/n)}(y) < \frac{n}{2}$ when we decide for the optimal choice $y_2 = \cdots = y_n = \frac{w-y_1}{n-1}$, for all other choices the function value of $f_{(u/n,\dots,u/n)}$ is even smaller. With the same argument we can show that

$$y_i \in [a, b]$$

needs to hold for all i. □

Remark 5.59. Proposition 5.58 above states that the entries of a vector y with $\|y\|_1 > \|(u/n, \dots, u/n)\|_1$ need to be "sufficiently equal" such that $y \sqsupseteq (u/n, \dots, u/n)$ can hold. We can see in Example 5.61 that this condition is only necessary but not sufficient for having $y \sqsupseteq (u/n, \dots, u/n)$.

Remark 5.60. Depending on the parameter choice, a as defined in Proposition 5.58 might be negative. As y_i needs to be non-negative for each i by definition, one could shrink the interval by requiring $y_i \in [\min\{0, a\}, b]$ instead of $y_i \in [a, b]$.

Let us illustrate Proposition 5.58 with an easy example.

Example 5.61. Assume we are trying to find the optimizer in a set A with $(1, 1, 1) \in A$. A vector y with $\|y\|_1 = 4$ needs to fulfill

$$y_i \in \left[\frac{9-\sqrt{57}}{6}, \frac{9+\sqrt{57}}{6}\right] \approx [0.24, 2.76]$$

for each i. For example $y^1 = (0.24, 1.88, 1.88)$ and $y^2 = (2.76, 0.62, 0.62)$ fulfill

$$f_{(1,1,1)}(y^1) = f_{(1,1,1)}(y^2) = \frac{3}{2} = f_{(1,1,1)}(1, 1, 1)$$

while all components of y^1 and y^2 lie in $[0.24, 2, 76]$. Keep in mind that the above condition is only necessary but not sufficient for $(1,1,1) \sqsubseteq y$. For example, for $y^3 = (2.75, 0.25, 1)$ we obtain

$$\frac{2.75}{3.75} + \frac{0.25}{1.25} + \frac{1}{2} = \frac{43}{30} < \frac{3}{2},$$

i.e. $(1,1,1) \sqsupseteq y^3$.

From a practical standpoint, Proposition 5.58 is more useful to detect whether a rating class vector is *not* optimal than to ensure that a rating class vector is optimal. In this regard, we develop a second proposition in the style of Proposition 5.58 which ensures that a given rating class vector is at least better than all possible rating class vectors with a 1-norm up to u. Of course, the boundaries for the proposition with the sufficient condition need to be more narrow than in the one with the necessary condition.

Proposition 5.62. Suppose again there are two vectors, $x, y \in \mathbb{R}^n_{\geq 0}$ with $\|x\|_1 = u$ and $\|y\|_1 = w$, where $w > u$ and $x = (u/n, \ldots, u/n)$. Then

$$y_i \in \left[\frac{w-b}{n-1}, \frac{w-a}{n-1}\right]$$

for at least all but one $i \in \{1, \ldots, n\}$ implies that $x \sqsubseteq y$. As before,

$$a = \frac{-\left(-\frac{nw}{2} + \frac{nu}{2} - u\right) - \sqrt{\left(-\frac{nw}{2} + \frac{nu}{2} - u\right)^2 - (wu(2-n) + (n+1)u^2)}}{n}$$

and

$$b = \frac{-\left(-\frac{nw}{2} + \frac{nu}{2} - u\right) + \sqrt{\left(-\frac{nw}{2} + \frac{nu}{2} - u\right)^2 - (wu(2-n) + (n+1)u^2)}}{n}.$$

Proof. In Proposition 5.58 we have shown that choosing $y_1 = a$ or $y_1 = b$ and $y_i = \frac{w-y_1}{n-1}$ ensures $f_{(u/n,\ldots,u/n)}(y) = \frac{n}{2}$. Based on the choice of y_1 this implies $y_i = \frac{w-a}{n-1}$ or $y_i = \frac{w-b}{n-1}$, where $a < b$ implies $\frac{w-b}{n-1} < \frac{w-a}{n-1}$.
Note that for most examples it holds that

$$a, b \notin \left[\frac{w-b}{n-1}, \frac{w-a}{n-1}\right],$$

see e.g. Example 5.64 below. Since we assumed

$$y_i \in \left[\frac{w-b}{n-1}, \frac{w-a}{n-1}\right]$$

for at least all but one $i \in \{1, \ldots, n\}$ this is not a problem.
As we already know, the function $g(y) = \frac{y}{y+b}$ monotonously increases in y while its derivative monotonously decreases. Hence, reducing $y_1 = b$ while enlarging some other y_i or enlarging $y_1 = a$ while reducing some other y_i preserves the inequality $f_{(u/n,\ldots,u/n)}(y) \geq \frac{n}{2}$. □

Remark 5.63. Instead of using the result of Proposition 5.58, one could also follow the exact same argument as in the proof of this proposition but fix $y_2 = \cdots = y_n$ which implies

$$y_1 = w - (n-1)y_2.$$

Recall that we fixed y_1 and set

$$y_2 = \cdots = y_n = \frac{w - y_1}{n-1}$$

when we proved Proposition 5.62. Of course, y_2 has to be chosen such that $y_1 \geq 0$, which is equivalent to $y_2 \leq \frac{w}{n-1}$.

Example 5.64. Let us continue Example 5.61, where we calculated

$$a = \frac{9 - \sqrt{57}}{6} \text{ and } b = \frac{9 + \sqrt{57}}{6}.$$

Remember that we are in a setting where we want to find a vector y with $\|y\|_1 = 4$ satisfying $y \sqsupseteq (1,1,1)$. Using Proposition 5.62 we calculate further that all vectors with

$$y_i \in \left[\frac{15 - \sqrt{57}}{12}, \frac{15 + \sqrt{57}}{12}\right] \approx [0.62, 1.88]$$

for all but one i satisfy $y \sqsupseteq (1,1,1)$.

Remark 5.65. Let us summarize the results of Propositions 5.58 and 5.62. These two propositions deliver some criterions to compare vectors with different component sum. While Proposition 5.58 is useful to prove that a vector following the optimal allocation given by Proposition 5.45 is indeed the optimum, Proposition 5.62 might be used as a tool to disprove it. The proofs of the two propositions show that the given bounds are indeed sharp.

Note that it might not always be possible to find a unique optimizer on conv$A$. Nonetheless, it is often easier to work on the convex hull.

5.2.6 Returning to Non-Convex Sets Using a Metric

After some analysis for the case that the set A we are optimizing on is convex, we return to the non-convex case. If the set A is not convex, we proceed as follows:

1. Calculate the convex hull conv$A$ of our set A.
2. Perform the optimization on the (convex) set conv$A$.
3. Return to the original set by finding a point of A lying close to the calculated optimum.

To be able to perform the third step, we need to derive a concept what "close" means in our setting. Therefore we introduce a metric d_b on the set $\mathbb{R}^n_{\geq 0}$. As we can see in Proposition 5.69, the metric should furthermore satisfy some stylized facts, compare also Remark 5.71.

Definition 5.66. Let b be a real-valued vector of length n with non-negative entries and choose $\varepsilon > 0$. We define the *metric* $d_b(\cdot,\cdot)$ *induced by* f_b by

$$d_b(x, y) = |f_b(x) - f_b(y)| + \varepsilon \cdot \mathbb{1}_{\{x \neq y\}},$$

where $x, y \in \mathbb{R}^n_{\geq 0}$ and f_b is defined as in Equation (16) by

$$f_b(a) = \sum_{i=1}^{n} \frac{a_i}{a_i + b_i}.$$

Remark 5.67. The vector b serves as the reference point of our optimization problem. If the optimizer is known, one should choose the reference point b to be the optimizer, see Proposition 5.45 or 5.56 for finding it. If it is clear from the context, we might drop the dependency on b and write $d(\cdot,\cdot)$ instead of $d_b(\cdot,\cdot)$.
Furthermore, we need to decide how to choose ε. We added this term in the definition above to ensure that $d_b(\cdot,\cdot)$ is a metric, see Proposition 5.68 for more details. In practice we often use $d_b(\cdot,\cdot)$ to measure the distance between the optimum and other points. In these cases, the positive definiteness property is needed to avoid dividing by zero later on. Unless stated otherwise, we fix ε to be a very small but positive value.

To be able to use $d_b(\cdot,\cdot)$ for measuring distances we show that it is a metric on $\mathbb{R}^n_{\geq 0}$.

Proposition 5.68. The function $d_b(\cdot,\cdot)$ defined in Definition 5.66 is a metric on $\mathbb{R}^n_{\geq 0}$.

Proof. We need to check the three properties of a metric:

Positive definiteness: By definition we have $d_b(x, x) = 0$ for all $x \in \mathbb{R}^n_{\geq 0}$. Let us assume that $x \neq y$. We then have $d_b(x, y) > 0$ since the absolute value is non-negative and $\varepsilon \cdot \mathbb{1}_{\{x \neq y\}}$ is strictly positive.

Symmetry: The symmetry follows directly from the properties of the absolute value $|\cdot|$ and the symmetry of the inequality in the indicator function.

Δ-inequality: Let $x, y, z \in \mathbb{R}^n_{>0}$. Then we have

$$\begin{aligned} d_b(x, z) &= |f_b(x) - f_b(z)| + \varepsilon \cdot \mathbb{1}_{\{x \neq z\}} \\ &\leq |f_b(x) - f_b(y)| + |f_b(y) - f_b(z)| + \varepsilon \cdot (\mathbb{1}_{\{x \neq y\}} + \mathbb{1}_{\{y \neq z\}}) \\ &= d_b(x, y) + d_b(y, z). \end{aligned}$$

Hence, all the properties are fulfilled and $d_b(\cdot,\cdot)$ defines a metric on $\mathbb{R}^n_{\geq 0}$. $\square$

There are a few stylized facts that our metric $d_b(\cdot,\cdot)$ needs to satisfy. They are summarized in the next proposition.

Proposition 5.69. Choose $b \in \mathbb{R}^n_{>0}$ as the optimizer of the optimal vector problem. The metric $d_b(\cdot,\cdot)$ from Definition 5.66 satisfies the following stylized facts:

1. Equality: For $x, y \in \mathbb{R}^n_{\geq 0}$ we have

$$f_b(x) = f_b(y) \;\Leftrightarrow\; d_b(x,b) = d_b(y,b),$$

i.e. all points with the same function value have the same distance to b.

2. Inverse proportionality: For $x, y \in \mathbb{R}^n_{\geq 0}$ the equality

$$f_b(x) < f_b(y) \;\Leftrightarrow\; d_b(x,b) > d_b(y,b)$$

is fulfilled. In other words, all points with a lower function value have a larger distance to b.

3. Positivity: For all $x \in \mathbb{R}^n_{\geq 0}$ with $x \neq b$ it holds that $d_b(x,b) > 0$.

Proof. The positivity is just one of the properties of the metric, see Proposition 5.68. For the two other facts we reconsider the definition of $d_b(\cdot,\cdot)$, which was given by

$$d_b(x,y) = |f_b(x) - f_b(y)| + \varepsilon \cdot \mathbb{1}_{\{x \neq y\}},$$

see Definition 5.66 for more details. The optimality of b ensures $f_b(b) \geq f_b(x)$ for all $x \in \mathbb{R}^n_{\geq 0}$. Therefore, we can drop the absolute value and receive

$$d_b(x,b) = f_b(b) - f_b(x) + \varepsilon,$$

if $b \neq x$. From this equation we directly obtain

$$f_b(x) = f_b(y) \;\Leftrightarrow\; d_b(x,b) = d_b(y,b)$$

and

$$f_b(x) < f_b(y) \;\Leftrightarrow\; d_b(x,b) > d_b(y,b).$$

$\square$

Remark 5.70. Note that by the first stylized fact we expect all points on the indifference circles in Example 5.48 to have equal distance to the point $(1,1,1)$. As desired, we derive by construction that $d_{(1,1,1)}(x,(1,1,1)) = \frac{1}{6}+\varepsilon$ or $d_{(1,1,1)}(x,(1,1,1)) = \frac{1}{3} + \varepsilon$ for all points on the circles, respectively.

Remark 5.71. There are more approaches to define the metric d_b which are different to the one we presented in Definition 5.66. One way would be to define a metric related to the one induced by the 1-norm and set

$$d_b(x,y) = \sum_{i=1}^{n} \left| \frac{x_i}{x_i + b_i} - \frac{y_i}{y_i + b_i} \right|,$$

for $x, y \in \mathbb{R}^n_{\geq 0}$. This approach indeed sets up a metric, as one can easily show. Nonetheless, the desired equality property

$$d_b(x,b) = d_b(y,b) \;\Leftrightarrow\; f_b(x) = f_b(y)$$

from Proposition 5.69 as well as the inverse proportionality property does not hold for this approach. Take for example the points $x = (0.6218, 0.4, 1.9782)$ and $y = (0.8158, 0.3, 1.8842)$ on the indifference set around $b = (1,1,1)$ for function value $\frac{8}{6}$, see Figure 29. For these points it holds that $f_b(x) = f_b(y)$ but

$$d_b(b,x) = 0.4951 \neq 0.4732 = d_b(y,b).$$

In a similar way we can construct a counterexample for the other implication and for the inverse proportionality. These flaws make the above definition practically useless for the next step of our procedure.
Another approach is to follow the idea of the SNCF-metric by defining

$$d_b(x,y) = \begin{cases} |f_b(x) - f_b(y)|, & \text{if } x \text{ and } y \text{ lie on a half-line with } b, \\ f_b(x) + f_b(y), & \text{else.} \end{cases}$$

With a half-line we mean a (straight) line going from b first through x and then to y or vice versa. In other words, we want x or y to be able to be displayed by a convex combination of y and b or x and b, respectively. Also this defines a metric. As we are only calculating distances of the form $d_b(b,x)$ in the next step, the case distinction does not limit the practical use of this approach. Nonetheless, we prefer to use the metric from Definition 5.66.

After setting up a metric to measure distances, we present different approaches to return from conv$A$ to A if the optimizer we found on conv$A$ does not lie in A:

a) Simply choose the point that lies closest to the optimum b, i.e. solve

$$\min_{x \in A} d_b(b,x).$$

If this point is not uniquely defined, choose one of the points that minimize this problem at random. The stylized facts from Proposition 5.69 ensure that we choose a point where f_b is maximal.

b) Choose a point out of the set A at random, where the probabilities are inversely proportional to their distance to the optimizer. For example one could find the probabilities by normalizing the terms

$$\frac{1}{d_b(b, x)},$$

where b is the optimizer found in conv$A$. The normalization is needed to ensure that we end up with a probability density. Note that the third stylized fact, i.e. the positivity from Proposition 5.69 holds. Together with the assumption that the optimizer does not lie in A this ensures that we are not dividing by zero. Furthermore, the first two facts, equality and proportionality, guarantee that points with the same value of f_b are picked with the same probability. In addition, the greater the function value of a point is, the higher its probability to be chosen becomes.

c) Instead of taking $\frac{1}{d_b(b,x)}$, one could also use the (normalization of the) terms

$$\frac{1}{g(d_b(b, x))},$$

where $g : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ is monotonously increasing. Examples for a reasonable choice of g would be $g(x) = x^n$ for $n \in \mathbb{N}$ ($n = 1$ yields the second approach) or $g(x) = \exp(x)$. Compared to the previous approach, this increases the preference of points with a small distance to the optimizer. As for the second approach, the desired properties are ensured by our stylized facts.

Remark 5.72. Let us come back to our application, where two insurance companies compete for the biggest share of customers in their company. Following a probabilistic instead of a deterministic strategy might be advantageous for an insurer, as the competing company does not know in advance which strategy is picked. As we have seen before, knowing the chosen strategy of your opponent in advance leaves you with an advantage, as finding the best response is much easier than finding the best strategy in general.

Example 5.73. Let us continue with Example 5.48 from before. All six points from our set A_2 have the same distance from the optimum $(1, 1, 1)$, namely

$$\begin{aligned} & \left|f_{(1,1,1)}(1,1,1) - f_{(1,1,1)}(x)\right| + \varepsilon \cdot \mathbb{1}_{\{(1,1,1)\neq x\}} \\ = & \quad \frac{9}{6} \quad - \quad \frac{7}{6} \quad + \quad \varepsilon \\ = & \frac{1}{3} + \varepsilon. \end{aligned}$$

Therefore it does not make a difference which of our approaches from above we apply. All three correspond to choosing one of the six points in A_2 with equal probability. This is indeed a desired result, compare also Remark 5.49.

5.2.7 Preference Relations for Rating Class Vectors

We now apply the strategy to determine an optimal vector in terms of $\sqsubseteq$ to the optimal coverage vectors $\hat{\lambda}$ from Definition 5.74. By doing so, we see that it is optimal to insure each risk class individually. The reason for this is that grouping up different risk classes results in a lower total equilibrium coverage, compare Section 3.2. Recall that those customers which subsidize other customers with higher risk reduce their coverage compared to a market where they are equipped with an own contract. This reduction is larger than the additional coverage purchase of the subsidized customers. The effect shows up due to the concavity of the utility function u, we refer to Example 5.79 as well as Examples 2.29 and 2.31 for more details. This motivates us to apply an extension to our model, which also helps to make our model more realistic.
Similar as in Borch (1962) [Bor62] we therefore assume that we need to add a safety loading to all insurance contracts which are issued by an insurer. We assume the safety loading to be the standard deviation of the (compound) risk. Note that [Bor62] assumes the safety loading to be three times the standard deviation. The safety loading in a contract is split proportionally to the purchased coverage among all customers in this contract. We formalize this approach by the following definition.

Definition 5.74. Given a rating class vector b, the *loaded premium vector* γ is given as the sum of the premium vector π plus the safety loading s, i.e. $\gamma = \pi + s$. To be able to define these two quantities we assume that the rating class vector b contains m_b rating classes. Rating class k consists of e_k risk classes, where $1 \leq e_k \leq n$ and $\sum_{k=1}^{m_b} e_k = n$. In order to define the safety loading, we need to introduce the concept of premium and optimal coverage vectors first.

- The *premium vector* $\pi \in \mathbb{R}^n$ is a vector that contains the premiums for each risk class according to the vector of rating classes b, see Definition 5.14. The premium π_i of the customers in risk class i is given as the premium of its according rating class which can be calculated as explained in Remark 5.15 with the help of Definition 5.4 and Proposition 5.6.

- Given a loaded premium vector γ, the *optimal coverage vector* $\hat{\lambda} \in \mathbb{R}^n$ is given by $\hat{\lambda} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_n)$ with

$$\hat{\lambda}_i = \hat{\lambda}_i(\gamma_i),$$

for each $i \in \{1, \dots, n\}$, where $\hat{\lambda}_i(\cdot)$ is the equilibrium insurance demand/coverage of the customers in risk class i.

- The *safety loading* $s \in \mathbb{R}^n$ is a vector that contains the amount of safety loading per customer for each risk class. The safety loading s_i of each customer of risk class i from rating class k is given by solving the following equation in s_i.

$$s_i = \hat{\lambda}_i(\pi_i + s_i) \cdot \frac{\sqrt{\sum_{l=1}^{e_k} \mathbb{V}(X_{k_l}) \cdot \eta_{k_l} \cdot \hat{\lambda}_{k_l}(\pi_{k_l} + s_{k_l})}}{\sum_{l=1}^{e_k} \eta_{k_l} \cdot \hat{\lambda}_{k_l}(\pi_{k_l} + s_{k_l})}. \tag{18}$$

  Here we denote the risk of the customers of risk class k_l by X_{k_l} and the corresponding number of customers in that risk class by η_{k_l}.

Remark 5.75. Let us provide some additional details regarding the definition of the safety loading in Equation (18). As mentioned before, the safety loading of a contract is supposed to be the standard deviation of the compound risk. This risk is given by

$$\sum_{l=1}^{e_k} X_{k_l} \cdot \eta_{k_l} \cdot \hat{\lambda}_{k_l}(\gamma_{k_l}),$$

where we used the same notation as in Definition 5.74 above. In Equation (18) we used $\pi + s$ instead of γ to illustrate the dependency of s of the right hand-side. Note that the compound risk depends on the amounts of coverage bought by the customers of each risk class. Keeping in mind that the risk variables are all independent, we can use Bienaymé's identity to calculate the safety loading of the contract as the standard deviation of the compound risk by

$$\sqrt{\sum_{l=1}^{e_k} \mathbb{V}(X_{k_l}) \cdot \eta_{k_l} \cdot \hat{\lambda}_{k_l}(\gamma_{k_l})}.$$

This safety loading is supposed to be split among all customers in the contract proportional to their purchased coverage. This finally provides us with Equation (18) to compute the safety loading per head.

Remark 5.76. The loaded premium vector is given as the premium plus a safety loading. All other costs such as signing fees, bank costs etc. are assumed to be zero, compare e.g. Remark 3.10 or 6.9. Introducing these costs and including them into the loaded premium vector would not structurally change the phenomena we can observe in our model.

We calculate the safety loading for each contract and not for the whole insurance company to make sure that each contract the insurance company sells is not going to make loss, compare the term of feasibility in Definition 2.14.

The amount of safety loading each customer needs to pay decreases with convergence rate 1/2 for a rising number of customers in the contract and vanishes if the number of customers tends to infinity. Hence, a larger amount of risk classes in a contract often results in a lower amount of safety loading for each customer, especially when the risks of the risk classes do not differ too much. At the same time, larger rating classes lead to less attractive premiums for some customers and therefore to a lower total equilibrium insurance purchase, as we have argued before. Therefore, it is not trivial to decide which of the two effects is larger. In Examples 5.79 and 5.85 we are going to analyze the strength of the effects in a real-life setting.

Definition 5.77. We define the *coverage relation* $\sqsubseteq_{\text{cov}}$ on the set of possible rating class vectors $\mathcal{R}'$ modulo numeration such that $a \sqsubseteq_{\text{cov}} b$ if $\hat{\lambda}_a \sqsubseteq \hat{\lambda}_b$ for the corresponding optimal coverage vectors. We call the rating class vector b *(globally) optimal rating class vector* if $a \sqsubseteq_{\text{cov}} b$ for all possible vectors of rating classes (modulo numeration) $a \in \mathcal{R}'$, i.e. if it is the global maximizer in the set of all coverage vectors corresponding to a risk coalition vector in $\mathcal{R}'$, see Definition 5.39. Furthermore, a set of rating class vectors $\mathcal{O} \subseteq \mathcal{R}'$ is called a *set of locally optimal rating class vectors*, if for all rating class vectors $a \in \mathcal{R}'$ there exists a vector $b \in \mathcal{O}$ such that $a \sqsubseteq_{\text{cov}} b$ holds. The smallest such set in terms of cardinality is called a *minimal set of locally optimal rating class vectors*, see also Remark 5.40.

Generally, we never compare two rating class vectors by $\sqsubseteq$ but only by $\sqsubseteq_{\text{cov}}$ in this thesis, see also Remark 5.27.

Definition 5.78. Given two insurance companies 1 and 2 with optimal coverage vectors $\hat{\lambda}_a$ and $\hat{\lambda}_b$, we call ψ the *fraction vector of customers in Company 1* given by

$$\psi = \frac{\hat{\lambda}_a}{\hat{\lambda}_a + \hat{\lambda}_b},$$

where the division is performed element wise.

In the next step, we provide an example to illustrate the effect of the safety loading to the optimal choice of a rating class vector. We therefore reconsider the very easy setting of Examples 2.29 and 2.31. In these examples we constituted a market consisting of equal shares of two different types of customers and varied the damage probability of the high-risk agents. We combine this setting with our thoughts from Example 5.22, where we concluded that the two possible rating class vectors are $(1,1)$ (one contract for all customers) and $(1,2)$ (two different contracts). Which one is preferable does not only depend on (the difference between) the damage probabilities but also on the number of customers in the market.

Example 5.79. We look (again) at an insurance market which consists of equal shares of two types of customers. These customers are facing Bernoulli distributed risk variables $Z_{1/2}$ with loss probability $p_1 = 5\%$ and p_2 between 5% and 60%. Furthermore, the loss in case of damage is considered to be $z_1 = z_2 = 1$, the initial wealth of the customers is $a = 2$ and the utility function is a CRRA utility function with risk aversion parameter $\rho = 3$, i.e. $u(x) = -\frac{1}{2x^2}$. Based on the number of customers in the market we can then calculate the optimal contract setup.

The closer the two risk probabilities lie together, the more favorable it is to group the two risk classes in one contract. Assume that the total number of customers in the market is low and therefore the safety loading per customer is high, compare Definition 5.74. This implies that the point at which it is optimal to split up the customers into two different contracts is reached for higher values of the high-risk customers damage probability p_2. In the following table, the splitting probabilities are given: If the high-risk agent's damage probability is higher than this value, it is better for the insurer to group the insured in two contracts based on the number of customers in the market. Furthermore, the push-out probability quantifies at which value of p_2 the low-risk customers are completely pushed out of the market in case there is only one contract given. Both calculations are done in scenarios (M) and (C).

	splitting probability		push-out probability	
number of customers	(M)	(C)	(M)	(C)
2	–	11.07%	–	14.73%
20	8.52%	9.75%	10.82%	24.10%
200	7.60%	7.84%	13.83%	27.87%
2000	6.52%	6.59%	13.83%	29.08%
20000	5.84%	5.89%	13.83%	29.46%
∞	5.00%	5.00%	13.83%	29.63%

Table 9: Splitting and push-out probabilities based on the number of customers in the market in scenarios (M) and (C)

Note that the results from the last row correspond to the results from Figures 2 and 3. In the case of a market with an unbounded amount of customers in the market, the per customer amount of safety loading vanishes and we are back in the exact setting of Examples 2.29 and 2.31.

With a rising number of customers, the per customer safety loading is reduced which leads to lower premiums and therefore to higher push-out probabilities (i.e. the push-out occurs later). Interestingly, the push-out probability in scenario (M) is only affected if the number of customers is very low. Reducing the amount of safety loading per person lowers the benefit of grouping both risk classes into one contract, as the net reduction of safety loading per person becomes smaller. Therefore, the splitting probability de-

creases if the number of customers in the market increases. As we can see, there are cases in which it is better to have only one contract in a competitive scenario, while one should insure each type of customers individually in a corresponding monopolistic market.
In a monopolistic market with only two customers, one of high and one of low risk, no insurance is bought at all, so no probabilities are given for this market setting.

5.2.8 Strategic Decisions and the Insurer's Game

Before continuing, we summarize our model approach from the sections before. The procedure of choosing the rating classes and calculating the customer flows is described below. From a game theoretical standpoint we are modeling a game between two insurers which can choose for allocations of the customers into contracts called rating class vectors in order to maximize the number of customers in their company.

0. First of all, the market specifications need to be fixed:
   - if we are in scenario (C) or (M);
   - the number of risk classes n together with their sizes η_k and the risk Z_k for each $k \in \{1, \dots, n\}$;
   - the specifications of the agents, i.e. their utility function u and their initial wealth a.

   Note that scenario (M) is actually duopolistic. We assume that the two insurers agree on charging the monopolistic premiums to maximize their profits.

1. At the same time, insurance Company 1 fixes its rating class vector a, Company 2 its rating class vector b from the set of possible rating class vectors modulo numeration $\mathcal{R}'$. As mentioned, this decision is made by the insurers, not the agents, and forms the (only) strategic decision the insurance companies can make.

2. Based on these rating class vectors, the premium vectors π_a and π_b are calculated according to Definitions 5.4 and 5.74 as well as Proposition 5.6. We then obtain the loaded premium vectors γ_a and γ_b by solving Equation (18) in Definition 5.74. Following this definition, we can use the loaded premium vectors to calculate the optimal coverage vectors $\hat{\lambda}_a$ and $\hat{\lambda}_b$ in a final step.

3. After all these steps we can finally decide which rating class vector is better in the sense of $\sqsubseteq_{\text{cov}}$. According to Definitions 5.78, we can calculate the fraction vector ψ of customers in Company 1. Each entry ψ_k of ψ describes the fraction of customers from risk class k

buying insurance from Company 1. As we have seen in Definitions 5.28 and 5.77, $a \sqsubseteq b$ holds if and only if $\sum_{i=1}^{n} \psi_i \leq \frac{n}{2}$, where we assumed for simplification that all risk classes are of equal size.

Remark 5.80. In the more general case that there are n risk classes with sizes $\eta_1, \dots, \eta_n$, $\hat{\lambda}_a \sqsubseteq \hat{\lambda}_b$ holds if and only if $\sum_{i=1}^{n} \eta_i \psi_i \leq \frac{\sum_{i=1}^{n} \eta_i}{2}$, compare Remark 5.25. If $\eta_1 = \cdots = \eta_n$, this condition simplifies to $\sum_{i=1}^{n} \psi_i \leq \frac{n}{2}$. As the observable phenomena in markets with different class sizes do not differ structurally from those in markets with equal class sizes, we stick to risk classes of equal size to ease the notation and the computations.

The insurers try to find a rating class vector such that the agents buying insurance from their company is maximized. Viewed from a game theoretical standpoint, we are seeking for a Nash equilibrium, i.e. a state of the game, where no insurance company can enlarge the amount of customers in its insurance company by deciding for a different rating class vector in the first step. As the game is symmetrical, in a Nash equilibrium it always needs to hold that $\sum_{i=1}^{n} \psi_i = \frac{n}{2}$ if an equilibrium exists. This can be understood, as the company gaining less customers due to the choice of its rating class vector can always choose for the vector of the other company to raise its customer share to be half of the total amount of customers.

Remark 5.81. If we restrict our problem to the choice of only one insurer, where the rating class vector of the other insurer is known in advance, we end up in a convex optimization problem that is therefore easy to solve, compare Remark 5.52. Depending on the assignment the other insurer made, this problem might even have a unique solution.

Remark 5.82. We could easily extend our model and the corresponding risk class management problem to a setting with more than two insurance companies. All extensions are straightforward and most of the results carry over to this extended model.

5.2.9 The Optimal Rating Class Algorithm

We continue by investigating the findings and phenomena described in Examples 5.22 and 5.79 for a setting with more than two risk classes. Therefore, we make some observations.
Firstly, we can note that in practice it is not optimal to set up rating classes which are not connected. For all of our examples, it was preferable to choose connected rating classes. For some easy examples we can derive an argument for this. Assume for example we are in a market with three sufficiently large risk classes, where we assume as usual $\mathbb{E}[Z_1] < \mathbb{E}[Z_2] < \mathbb{E}[Z_3]$. The rating class vector $(1, 2, 1)$ with a non-connected rating class cannot be optimal, as it holds that $(1, 2, 1) \sqsubseteq_{\text{cov}} (1, 1, 2)$. This can be seen as the push-out

effect and the aggregated coverage bought is lower, the higher the difference between the damage probabilities is. In Remark 5.83 we see that choosing for connected rating classes drastically reduces the computational effort.
Secondly, it is easy to see that the preference relation is not transitive, see Example 5.41. This makes further computations more complicated, as it is not sufficient to go through the set of all possible rating class vectors (modulo numeration) and compare all vectors with the currently largest in terms of $\sqsubseteq_{\text{cov}}$. Nonetheless, comparing each vector only with the current maximizer and not with all other vectors is a strategy that often works in practice, see Example 5.85.
Thirdly, one can see from numerical examples that it is not optimal to assume the rating classes to be of the same size. The lower the relative difference between the risks are, the larger the rating classes are going to be.
Next, we present an idea which ensures that each rating class vector with connected classes is considered but the rating classes are not necessarily of the same size. Our idea is the following: We travel systematically through the set

$$\mathcal{R}^* = \{(a_1, \dots, a_n) \in \mathbb{N}^n : 1 = a_1 \leq \dots \leq a_n \text{ and } a_i - a_{i-1} \leq 1 \text{ for all } i = 2, \dots, n\},$$

which is according to our restrictions the set of all possible rating class vectors. As we are considering monotonously increasing integer vectors with a one as the first value and a maximum increase of one in each element, every vector of rating classes can also be represented by a $\{0,1\}^{n-1}$ vector, which describes if there is an increase in entry $2, \dots, n$ or not. Hence, traveling through $\mathcal{R}^*$ is equivalent to traveling through

$$\{(b_2, \dots, b_n) \in \{0,1\}^{n-1}\},$$

where we set $b_i = a_i - a_{i-1}$ to make the connection between the two sets clear. Our algorithm makes use of the set $\mathcal{R}^*$, which can be traveled much easier, as we can make use of the binary representation of numbers, see Remark 5.84.

Remark 5.83. Our algorithm allows us to restrict to $|\mathcal{R}^*| = 2^{n-1}$ rating class vectors instead of $|\mathcal{R}| = n^n$ ones. Proposition 5.23 already showed that it is sufficient to regard $|\mathcal{R}'| = B_n$ instead of $|\mathcal{R}| = n^n$ rating class vectors, where B_n is the n-th Bell number. Below, we compare these three numbers for some examples. The sequence of Bell numbers together with many interesting literature sources and comments can be found as sequence A000110 in the OEIS.[8]

[8]Online Encyclopedia of Integer Sequences, `https://oeis.org/A000110`

n	2^{n-1}	B_n	n^n
2	2	2	4
3	4	5	27
4	8	15	256
5	16	52	3125
10	512	115,975	10,000,000,000
20	$\approx 5 \cdot 10^5$	$\approx 5 \cdot 10^{13}$	$\approx 10^{26}$

Table 10: Number of rating class vectors to check for a market with n risk classes using different approaches

This table provides us with a first intuition how the Bell numbers evolve. To analyze the asymptotics and therefore the computing time of an algorithm comparing the vectors in $\mathcal{R}$, $\mathcal{R}'$ and $\mathcal{R}^*$ we refer to Berend and Tassa (2010) [BT10]. They show different bounds for the Bell numbers, e.g.

$$B_n < \left(\frac{0.792n}{\ln(n+1)}\right)^n \quad \forall n \in \mathbb{N}.$$

The constant 0.792 can be lowered to $\exp(-0.6+\varepsilon)$ which is for small choices of ε even smaller than 0.55. Lowering this constant implies that the bound for B_n only holds if n lies beyond a certain threshold, see [BT10] for details. All together, we can see that the number of rating class vectors we need to check can be lowered by the factor $\left(\frac{\ln(n+1)}{0.792}\right)^n$ if we regard the set $\mathcal{R}'$ instead of $\mathcal{R}$. As already mentioned, the constant in the factor can be lowered to $\exp(-0.6+\varepsilon)$.
Considering $\mathcal{R}^*$ instead of $\mathcal{R}$ we can lower the number of rating class vectors we need to check even further, compare Table 10. In this case the factor of cardinalities between the sets is given by $n \cdot \left(\frac{n}{2}\right)^n$.

Let us formalize our considerations in the following algorithm.

Algorithm 1 Optimal Rating Class Algorithm

Input: Set of risk classes with the corresponding risk variables and class sizes, parameters to compute the premiums and coverages according to Section 5.1 (Definition 5.4 and Proposition 5.6) and Definition 5.74

Output: (possibly one elemental) locally optimal set of rating class vectors $\mathcal{O}$ for the given risk classes

$\mathcal{O} = \emptyset$
$b_{\text{opt}} = (1, \ldots, 1)$
calculate the loaded premium vector γ_{opt} according to Definition 5.74 for b_{opt}
$\hat{\lambda}_{\text{opt}} = (\hat{\lambda}_1(\gamma_{\text{opt}_1}), \ldots, \hat{\lambda}_n(\gamma_{\text{opt}_n}))$
for $(b_2, \ldots, b_n)$ in $\{0,1\}^{n-1}$ **do**
  $b_{\text{sum}} = (1, 1 + b_2, 1 + b_2 + b_3, \ldots, 1 + \sum_{k=2}^{n} b_k)$
  calculate the loaded premium vector γ_{sum} according to Definition 5.74 for b_{sum}
  $\hat{\lambda}_{\text{sum}} = (\hat{\lambda}_1(\gamma_{\text{sum}_1}), \ldots, \hat{\lambda}_n(\gamma_{\text{sum}_n}))$
  if $\hat{\lambda}_{\text{opt}} \sqsubseteq \hat{\lambda}_{\text{sum}}$ **then**
    $b_{\text{opt}} = b_{\text{sum}}$
    $\hat{\lambda}_{\text{opt}} = \hat{\lambda}_{\text{sum}}$
    $\mathcal{O} = \mathcal{O} \cup \{b_{\text{opt}}\}$
  end if
end for
if $\mathcal{O} = \emptyset$ **then**
  $\mathcal{O} = \{b_{\text{opt}}\}$
end if
return $\mathcal{O}$

Remark 5.84. The **for loop** needs us to somehow travel through the set $\{0,1\}^{n-1}$. This can be done by using a **for loop** that goes from 1 to 2^{n-1} and transforming this number into its binary representation written in a vector, where the k-th entry of the vector corresponds to the k-th digit of the binary number.

5.2.10 Numerical Analysis

Now, we apply our newly developed algorithm to the example we threw up in the beginning of this section.

Example 5.85. We consider a similar setting as in Example 5.24 and assume there is an economy in a competitive market with 4000 customers which consists of 20 risk classes with equal size. All agents face Bernoulli distributed risks. We compare four different scenarios which differ in the risk probabilities of the agents:

(a) risk probabilities of 0.1%, 0.2%, …, 2%;

(b) risk probabilities of 1.1%, 1.2%, ..., 3%;

(c) risk probabilities of 2.1%, 2.2%, ..., 4%;

(d) risk probabilities of 0.2%, 0.4%, ..., 4%.

All agents of all risk classes in all examples are assumed to have a loss in case of damage of $z = 1$. Furthermore, all agents are equipped with an initial wealth of $a = 2$ and CRRA utility with $\rho = 3$.
In Table 11 below, the optimal rating class vectors for these four examples are given. It is easy to check that all of our given rating class vectors are global optimizers for the corresponding example.

risk class	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Example (a)	1	1	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3
Example (b)	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2
Example (c)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Example (d)	1	1	2	2	2	2	3	3	3	3	3	3	4	4	4	4	4	4	4	4

Table 11: Optimal rating class vectors for customers of 20 different risk classes for four examples in scenario (C)

For Example (a), we end up with three rating classes, one for the two classes with the lowest risk, one for the next six and a final one for the rest.
Shifting the risk probabilities up results in a lower relative difference between the risks and therefore (as expected) in a reduced number of rating classes. For example if the risks lie between 1.1% and 3% we only receive two rating classes, one for the first eight risk classes and one for the rest (Example (b)). Shifting even more, e.g. to 2.1% til 4% results in only one big rating class (Example (c)).
Contrariwise, multiplying the original risks by a factor such that the risk probabilities are spread over a larger interval, e.g. 0.2%, 0.4%,...,4% (Example (d)) or even by larger factors results in a higher number of rating classes.
When it comes to the number and the size of the rating classes, there are two opposing effects working against each other, which we already explained before. To emphasize the importance of these two effects, we investigate them again for this real-life example.
First of all there is the push-out effect, we already observed in theoretical and real-life examples in the last chapters, see e.g. Example 2.31 or the examples in Section 3.2. This effect wants us to choose the number of risk classes m to be large and the classes itself to be small. This is due to the reason that the average amount of coverage bought reduces, the more risk classes there are in one rating class.
On the other hand, we have added a variance based safety loading to the

premium of each contract, which reduces its cost per head with a rising number of customers in a rating class. This effect pushes towards a small number of large rating classes. Which of these effects is stronger highly depends on the market specifications.
As a thumb rule one can say that the bigger the (absolute and relative) difference between the damage probabilities of the different risk classes is, the higher the number of risk classes m becomes. Note that we were already observing similar effects as these described here in Example 5.79.

Example 5.86. In a monopolistic market the rating class get more evolved. Generally, the optimal number of rating classes m in a monopolistic market is higher than in the corresponding competitive market. Furthermore, it is no longer the case that the number of risk classes in each rating class needs to get larger. In the following table we can find the four optimal rating class vectors for the Examples (a) to (d) from the last example, Example 5.85, in this case for the monopolistic market.

risk class	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Example (a)	1	1	1	1	1	1	1	2	2	2	2	2	3	3	3	3	3	3	3	3
Example (b)	1	1	1	1	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3
Example (c)	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2
Example (d)	1	1	1	1	2	2	3	3	3	3	4	4	4	4	5	5	5	5	5	5

Table 12: (Element of a set of locally) Optimal rating class vectors for customers of 20 different risk classes for four examples in scenario (M)

Recall that we assume the two insurance companies in scenario (M) to decide for a price fixing on the monopolistic premiums which impedes competition but ensures that their profits are maximized.
The effects we have observed for the competitive case in Example 5.85 show up for monopolistic markets as well. We have already seen that the push-out effect in monopolistic markets is stronger than in competitive ones, compare e.g. Example 2.29 with 2.31. This leads to a larger number of smaller rating classes compared to competitive markets.
As we can see in the next example, Example 5.87, there is no globally optimal rating class vector for Example (d). In this case we give one of the locally optimal rating class vectors. For all other examples, the given optimizer is a global optimizer.

Example 5.87. Let us reconsider Example 5.86. We take a second look at part (d), i.e. risk classes with risk probabilities of 0.2%,...,4% in a monopolistic market. Therefore, we analyze the rating class vectors listed in the table below.

risk class	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
rating class vector (1)	1	1	1	1	2	2	3	3	3	3	4	4	4	4	5	5	5	5	5	5
rating class vector (2)	1	1	1	1	2	2	3	3	3	4	4	4	4	4	5	5	5	5	5	5
rating class vector (3)	1	1	1	1	2	2	3	3	3	4	4	4	4	5	5	5	5	5	5	5
rating class vector (4)	1	1	1	1	2	2	2	3	3	3	3	4	4	4	4	4	5	5	5	5
rating class vector (5)	1	1	1	1	2	2	2	3	3	3	3	4	4	4	4	5	5	5	5	5
rating class vector (6)	1	1	1	1	2	2	2	3	3	3	4	4	4	4	4	5	5	5	5	5
rating class vector (7)	1	1	1	1	2	2	3	3	3	3	4	4	4	4	4	5	5	5	5	5
rating class vector (8)	1	1	1	1	2	2	2	3	3	3	4	4	4	4	5	5	5	5	5	5

Table 13: Rating class vectors for customers of 20 different risk classes in scenario (M)

We can then show that for the rating class vectors it holds (1) $\sqsubseteq_{\text{cov}}$ (2), (2) $\sqsubseteq_{\text{cov}}$ (3), ..., (8) $\sqsubseteq_{\text{cov}}$ (1) but not vice versa. For example we obtain that $\hat{\lambda}_1 = \frac{1}{100}(0, 12, 28, 41, 23, 31, \mathbf{21}, \mathbf{27}, \mathbf{32}, \mathbf{37}, \mathbf{25}, \mathbf{29}, \mathbf{32}, \mathbf{36}, 25, 28, 31, 33, 36, 38)$, $\hat{\lambda}_2 = \frac{1}{100}(0, 12, 28, 41, 23, 31, \mathbf{23}, \mathbf{29}, \mathbf{34}, \mathbf{23}, \mathbf{27}, \mathbf{31}, \mathbf{34}, \mathbf{38}, 25, 28, 31, 33, 36, 38)$, where the coverages are rounded to full percent. The differences in two vectors above are displayed bold. It is easy to calculate that $\hat{\lambda}_1 \sqsubseteq \hat{\lambda}_2$ and therefore (1) $\sqsubseteq_{\text{cov}}$ (2). This can be seen as we obtain that $9.4982 < 9.5018$, where we omitted the calculation for the first entry of the vector, as we would get a division by zero. These calculations imply that there cannot be an optimal rating class vector in terms of $\sqsubseteq_{\text{cov}}$ for this example. Nonetheless, the rating class vectors from the above table form a set $\mathcal{O} = \{(1), \dots (8)\}$ of locally optimal rating class vectors.
By defining

$$\hat{\lambda} := \frac{\hat{\lambda}_1 + \cdots + \hat{\lambda}_8}{8},$$

we find a vector in $\operatorname{conv}\{\hat{\lambda}_1, \dots, \hat{\lambda}_8\}$ which satisfies

$$\hat{\lambda}_i \sqsubseteq \hat{\lambda}$$

for all $i = 1, \dots, 8$.
As we can show later, using Algorithm 3, a refined version of Algorithm 1, Vector (3) seems to be the "best" rating class vector out of the eight locally optimal ones. The reason for this is that it is considered to be in the locally optimal set with a probability of over 90% if we travel through the set of all rating class vectors at random instead of in order. The second best rating class vector, Vector (2), only appears with a probability of about 80%. We refer to Remark 5.104 and also Section 5.3.2 for more details about the used algorithm.
All eight locally optimal rating class vectors share that the first four risk classes are grouped in one rating class. Note that the first risk class does not purchase any insurance. Hence, one could ask why one does not choose for one rating class with the first risk class and another one with the second to the fourth risk class. Nonetheless, the safety loading per customer for the

first risk class is so high that the agents do not purchase any insurance at all, even if they are insured alone. Hence, we end up lying in the exact same situation as by choosing only one risk coalition for the first four risk classes.

5.3 Capacity Constraints

Until now, we have assumed insurance markets to have an infinite capacity. In practice, it might be possible that the market has only a limited capacity c. The conceptualization of capacity constrained markets is done in Section 5.3.1. After evolving the optimal rating class algorithm we apply it to different numerical examples in Section 5.3.2.
Capacity constraints could apply due to market regulations or strategic decisions made by the insurer. The reason for the latter could be that an insurance company wants to ensure all branches of its insurance portfolio to be more or less equally occupied. Note that the insurer has only a certain amount of risk capital, so some branches cannot become bigger than a certain threshold. Furthermore, a lot of insurances are sold via insurance intermediaries. Because the number (and of course also the working speed) of these intermediaries is bounded, this might result in a limitation of the market capacity due to a limited insurance distribution.
The capacity constraints could either apply to the whole market, single risk classes or contracts (i.e. rating classes). As each risk class consists of a fixed amount of customers, the most meaningful way would be to set a limit for the amount of contracts that can be sold, i.e. to cap the proportion of the coverages respectively the fraction vector ψ to a capacity constraint, compare Remark 5.98. To cap the coverages themselves instead does not really make sense. We are going to present an analysis of our approach shortly.

5.3.1 Conceptualization

We start by presenting a three-part definition of constrained vectors.

Definition 5.88. A vector $a \in \mathbb{R}^n_{\geq 0}$ is called *element wise constrained* by a capacity constraint $c \in \mathbb{R}_{\geq 0}$ if it fulfills

$$a_i \leq c \text{ for all } i \in \{1, \ldots, n\}.$$

The vector a is said to be *contract wise constrained* by the capacity constraint c and contract vector $d \in \{1, \ldots, n\}^n$ if it fulfills

$$\sum_{i=1}^{n} a_i \cdot \mathbb{1}_{\{d_i = j\}} \leq c \text{ for all } j \in \{1, \ldots, n\}.$$

Finally, the vector is called *globally constrained* by c if

$$\sum_{i=1}^{n} a_i \leq c.$$

Remark 5.89. From a practical standpoint a globally constrained market is the most important type of constrained markets. To include the argument with the risk capital in the market, one could add weights to the summands leaving us with the condition

$$\sum_{i=1}^{n} w_i \cdot a_i \leq c,$$

where $(w_1, \ldots, w_n)$ is a vector taking real non-negative values. We can see later on, in Example 5.100, that adding a global constraint on an insurance market does not have a big influence on the market behavior.

Remark 5.90. If it is clear from the context, in which kind a vector a is constrained by a capacity constraint c, we simply write that a is constrained by c.

It is also of interest what happens to the customers who are not able to buy an insurance contract from one company due to a capacity constraint. The behavior of the affected customers depends on the fact, whether the insurance is compulsory (like health insurance in Germany) or not. We model the decisions of the customers by a changing parameter $h \in [0, 1]$. It describes how many customers, who are not able to buy insurance from one company due to the lack of capacity, switch to the other insurance company. Hence, $1 - h$ describes the amount of customers vanishing from the market.

Remark 5.91. As mentioned before, in case of a compulsory insurance we need to set the changing parameter to $h = 1$.

The impact of the value of the changing parameter on the optimal risk class allocation is not to be underestimated, an analysis of this parameter is presented in Example 5.108.
In the case that the capacity constraint applies not only to one risk class, but to a contract or even the whole market, we need to model in which order the possible customers are not served. One way would be to reduce the amount of customers equally, i.e. by proportionally reducing the customers of each risk class until the constraint is met. Another approach would be to reduce the classes with the highest or lowest volumes first. As the equilibrium amount of coverage is lower, the higher the premium is the customers need to pay, it might be meaningful to reduce the risk classes with the high coverages first, as they make less profit for the insurer. Of course, one needs to pay attention that changing the ratio between the risk classes in one contract could imply further changes in the contract, as it might become more or less attractive for some types of customers. We therefore reduce the coverages proportional to their purchased coverage. Let us formalize our previous thoughts.

Definition 5.92. Let $a,\ b$ be vectors in $\mathbb{R}^n$ and c be a real-valued capacity constraint such that a is not capacity constrained by c in the sense of Definition 5.88. Let furthermore $h \in [0,1]$. We then call the vectors $\tilde{a}$ and $\tilde{b}$ defined in the following as *the capped vectors a and b.*
If we are in the setting of element wise constrained vectors we define

$$\tilde{a},\ \tilde{b} \in \mathbb{R}^n \text{ by } \tilde{a}_i = \min(a_i, c) \text{ and } \tilde{b} = b + h \cdot (a - \tilde{a}).$$

Furthermore, if we are in the setting of contract wise constrained vectors with a contract vector d, we define

$$\tilde{a},\ \tilde{b} \in \mathbb{R}^n \text{ by}$$

$$\tilde{a} = \sum_{j=1}^{n} \frac{c}{\max\left(\sum_{i=1}^{n} a_i \cdot \mathbb{1}_{\{d_i=j\}}, c\right)} \cdot \left(a_1 \cdot \mathbb{1}_{\{d_1=j\}}, \ldots, a_n \cdot \mathbb{1}_{\{d_n=j\}}\right)$$

and

$$\tilde{b} = b + h \cdot (a - \tilde{a}).$$

Finally, if we are in the setting of globally constrained vectors we define

$$\tilde{a},\ \tilde{b} \in \mathbb{R}^n \text{ by } \tilde{a} = \frac{c}{\max\left(\sum_{i=1}^{n} a_i, c\right)} \cdot a \text{ and } \tilde{b} = b + h \cdot (a - \tilde{a}).$$

Remark 5.93. The vector $\tilde{a}$ is defined in such a way that it is now constrained by c (see Proposition 5.95), where the constraint is exactly met. The vector $\tilde{b}$ is defined such that it contains the amounts of customers of each risk class which were insured before we applied the constraint to vector a. Furthermore, we add the amount of customers that changes from the other insurance company, which are not able to be insured there, according to the changing parameter h.

Remark 5.94. Of course, high changing parameters in combination with a too small capacity constraint could result in an error, as customers would be sent forth and back infinitely often. Hence, one needs to choose the capacity constraint and the changing parameter accordingly to prevent this problem. Nonetheless, if $h < 1$ and not all customers can be insured, the amount of agents bouncing between the two insurance companies reduces in each step. After a significantly large time interval, all uninsurable customers have left the market, so we end up with the capacity constraints being fulfilled.
As an alternative, we could mitigate this problem by extending our model like we did it in Remark 2.11. There we assumed that the risk aversion varies among the customers in each risk class, leading to different optimal amounts of coverage between the agents in one risk class. Hence, some customers might be willing to leave the insurance market, as there benefit of insurance is comparably small. The implementation of this extension is not straightforward and therefore left for future research.

Proposition 5.95. Following the procedure in Definition 5.92, the vector $\tilde{a}$ is chosen such that it is capacity constrained by the capacity constraint c and the vector $\tilde{b}$ is minimal with respect to the 1-norm.

Proof. We need to show the proposition for the three possible types of constraints. Before we start, we note that $\tilde{b} = b + h \cdot (a - \tilde{a})$ is minimal w.r.t. the 1-norm if and only if $a - \tilde{a}$ is minimal w.r.t. the 1-norm.

a) element wise constrained vectors: It is clear from the definition that the capacity constraint is met and $a - \tilde{a}$ is minimal.

b) globally constrained vectors: Let $\tilde{c}$ be the sum of the vector a. If $\tilde{c} \leq c$, then a is already capacity constrained by c and we obtain $\tilde{a} = a$ and b is obviously minimal. If $\tilde{c} > c$, then $\frac{c}{\max(\tilde{c},c)} = \frac{c}{\tilde{c}}$ and therefore the sum of $\tilde{a}$ equals $\frac{c}{\tilde{c}} \cdot \tilde{c} = c$. Hence, $\tilde{a}$ is capacity constrained by c. Furthermore, as the sum of $\tilde{a}$ equals exactly c, $a - \tilde{a}$ is minimal w.r.t. the 1-norm.

c) contract wise constrained vectors: Note that by definition $\tilde{a}$ is a vector, which is a sum of n vectors. For each entry of the vector $\tilde{a}$ exactly one of the entries of the n vectors which are summed up is non-zero. Now we can easily apply the argument from the last paragraph to show that also in this case the proposition is fulfilled.

□

5.3.2 Evolution of the Optimal Rating Class Algorithm and Numerical Examples

After building the theoretical background, we illustrate our model with some examples. Before we do so, we are going to present an adapted version of Algorithm 1, our optimal rating class algorithm. The adaption introduces the opportunity to deal with capacity constraints and changing parameters. Keep in mind that Algorithm 2 is indeed a generalization of Algorithm 1 from the last section, since setting the capacity constraint high enough reflects an unconstrained market.

Algorithm 2 Optimal Rating Class Algorithm under Capacity Constraints

Input: Set of risk classes with the corresponding risk probabilities and class sizes, parameters to compute the premiums and coverages according according to Section 5.1 (Definition 5.4 and Proposition 5.6) and Definition 5.74, capacity constraint c, changing parameter h

Output: (possibly one elemental) locally optimal set $\mathcal{O}$ of rating class vectors for the given risk classes

$\mathcal{O} = \emptyset$
$b_{\text{opt}} = (1, \dots, 1)$
calculate the loaded premium vector γ_{opt} according to Definition 5.74 for b_{opt}
$\hat{\lambda}_{\text{opt}} = (\hat{\lambda}_1(\gamma_{\text{opt}_1}), \dots, \hat{\lambda}_n(\gamma_{\text{opt}_n}))$
for $(b_2, \dots, b_n)$ in $\{0,1\}^{n-1}$ **do**
    $b_{\text{sum}} = (1, 1 + b_2, 1 + b_2 + b_3, \dots, 1 + \sum_{k=2}^{n} b_k)$
    calculate the loaded premium vector γ_{sum} according to Definition 5.74 for b_{sum}
    $\hat{\lambda}_{\text{sum}} = (\hat{\lambda}_1(\gamma_{\text{sum}_1}), \dots, \hat{\lambda}_n(\gamma_{\text{sum}_n}))$
    Calculate the capped fraction vectors ψ^{optcc} and ψ^{sumcc} according to Definition 5.92 using c and h and with regard to the desired type of capacity constraint
    if $\sum_{i=1}^{n} \psi_i^{\text{optcc}} \leq \sum_{i=1}^{n} \psi_i^{\text{sumcc}}$ **then**
        $b_{\text{opt}} = b_{\text{sum}}$
        $\hat{\lambda}_{\text{opt}} = \hat{\lambda}_{\text{sum}}$
        $\mathcal{O} = \mathcal{O} \cup \{b_{\text{opt}}\}$
    end if
end for
if $\mathcal{O} = \emptyset$ **then**
    $\mathcal{O} = \{b_{\text{opt}}\}$
end if
return $\mathcal{O}$

Remark 5.96. As the capacity constraint is applied to the fraction vectors and not the coverage vectors itself, we now compare the sums of the capped fraction vectors by $\leq$ instead of the coverage vectors by $\sqsubseteq$. If the capacity constraint is so high that it does not get applied, it makes no difference whether we compare the sums of the capped fraction vectors by $\leq$ or the coverage vectors by $\sqsubseteq$, see Definitions 5.78 and 5.28. Therefore, Algorithm 2 is indeed a generalization of Algorithm 1.

As we can see in the following examples, the problem discussed in Example 5.87 that a maximizer does not need to exist is very relevant for capacity constrained markets.

Example 5.97. We begin by examining the market for element wise constrained vectors. We fix the changing parameter for all calculations in this example to $h = \frac{2}{3}$. Similar to part (a) of Example 5.85, we consider 20 risk classes of equal size and Bernoulli distributed risk with damage probabilities of 0.1%, 0.2%,...,2%. Furthermore, the agents of all risk classes are assumed to have a loss in case of damage of $z = 1$. All agents are equipped with an initial wealth of $a = 2$ and CRRA utility with $\rho = 3$. The rating class vectors for different capacity constraints are presented in the following table.

risk class	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$c = 0.4$	1	1	1	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3
$c = 0.45$	1	1	1	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3
$c = 0.5$	1	1	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3	3
$c = 0.55$	1	1	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3
$c = 0.6$	1	1	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3	3
$c = 0.65$	1	1	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3

Table 14: (Element of a set of locally) Optimal rating class vectors for customers of 20 different risk classes for varied element wise capacity constraints in scenario (C)

For all values of c greater than 0.65 we receive the same results as for $c = 0.65$. It is not surprising that this rating class vector coincides with the rating class vector for the corresponding unconstrained case. As we have already argued, the constrained case with $c = 1$ is equivalent to the unconstrained case. Note that we have set the changing parameter to $h = \frac{2}{3}$ to avoid that there are customers that cannot be insured at all, compare Remark 5.94.

As we can see, the capacity constraint does not have a big influence on the optimal rating class vectors. For small capacity constraints, the rating classes seem to get a bit bigger. Be reminded that bigger rating classes result in lower coverages for the customers with lower risk. Hence the fraction of agents signing contracts at this company is lowered. Nonetheless this is not such a big deal, as the (small) capacity constraint would cap the fraction anyway.

Especially when one takes ridiculously small values such as $c = 0.1$ or even smaller, one ends up with a market where it is optimal to have only two rating classes. In this case, one rating class consists of the first two or three risk classes, while the other contains the rest. Note that such small constraints are not meaningful in practice and are therefore not analyzed further.

Remark 5.98. Keep in mind that if a fraction $\psi_k = 1$ of customers buy insurance from Company 1 would imply that all 200 customers of that risk class k sign their contract at Company 1, see Definition 5.78. A smaller fraction ψ'_k implies that $200 \cdot \psi'_k$ (rounded) customers sign the contract at the corresponding company. Hence, a capacity constraint of c'_k caps the number of contracts that can be signed at one company to $200 \cdot c'$ and vice versa.

We continue with an example, where we investigate contract wise constrained markets.

Example 5.99. We look at a market with the same settings as in Example 5.97 which is now contract wise and not element wise constrained. This means that the maximum amount of policies that can be sold for one rating class respectively contract is capped. The capping is applied to the sum of the fractions, i.e. the total amount of customers, in one contract. For element wise capacity constraint markets, a fraction of $\psi_k = 1$ for risk class k implies that all 200 customers of risk class k sign the contract at one company. Also in this setting, one can easily transform the capacity constraint to the number of customers that are allowed to sign the contract by multiplying it with 200. The optimal rating class vectors for different capacity constraints can be found in the table below.

risk class	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$c = 1$	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8	8	9	9	10	10
$c = 2$	1	1	2	2	2	3	3	3	3	4	4	4	4	5	5	5	5	6	6	6
$c = 3$	1	1	2	2	2	2	2	2	3	3	3	3	3	3	4	4	4	4	4	4
$c = 4$	1	1	2	2	2	2	3	3	3	3	3	3	4	4	4	4	4	4	4	4
$c = 5$	1	1	1	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3
$c = 6$	1	1	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3

Table 15: (Element of a set of locally) Optimal rating class vectors for customers of 20 different risk classes for varied contract wise capacity constraints in scenario (C)

Again, for capacity constraints larger than $c = 6$ we end up in a market that is de facto unconstrained. In these cases the rating class vector is the one we obtain for $c = 6$ or from part (a) of Example 5.85.
It should not be surprising that small capacity constraints lead to small risk classes, as deciding for bigger ones would end in large amounts of customers getting capped out of the insurance company. The smaller the constraint is, the stronger this effect becomes.

Before continuing with the presentation of a refined version of Algorithm 2, we investigate a third example, this time for a globally constrained market. Actually, it was quite hard to find an example, where the (global) capacity constraint has an influence on the optimal rating class vector. For most examples, especially all other examples we have investigated so far, there is no difference between a constrained and an unconstrained market. This is not very surprising, as this type of constraining affects the whole market instead of single risk classes or contracts. As an insurance company can never perform better than the capacity constraint of the market, it might be favorable to choose for a risk allocation that is not optimal to avoid losing customers to the other company. Nonetheless, the insurance company also harms itself which makes this strategy not optimal for most market specifications, so we often get the same results as for an unconstrained market.

Example 5.100. Let us consider a market similar to those from the last examples, but this time with customers facing damage probabilities of 0.25%, 0.5%, ..., 5% in a monopolistic market. As before, each risk class consists of 200 people. Using different capacity constraints together with a changing parameter of $h = 1$ we obtain:

risk class	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$c = 10$	1	1	1	2	2	3	3	4	4	4	4	5	5	5	5	6	6	6	6	6
$c = 10.2$	1	1	1	2	2	3	3	3	4	4	4	4	4	5	5	5	5	5	5	5
$c = 10.4$	1	2	2	3	3	4	4	4	5	5	5	5	5	6	6	6	6	6	6	6
$c = 10.6$	1	1	1	2	2	3	3	3	4	4	4	4	4	5	5	5	5	5	5	5

Table 16: (Element of a set of locally) Optimal rating class vectors for customers of 20 different risk classes for varied global capacity constraints in scenario (M)

For all capacity constraints larger than $c = 10.6$, we end up for the same optimal rating class vectors as for this value. Interestingly, the strength of the effect of the capacity constraint does not seem to be monotone. This comes due to that the algorithm we are using is not very robust when we are dealing with capacity constraints. Hence, it seems meaningful to derive a new, improved algorithm which is more robust.

As mentioned in Example 5.100, the issue we addressed in Example 5.87 is getting even more relevant in the case of capacity constrained insurance markets. Let us therefore focus on the capacity constraint $c = 0.55$ from Example 5.97 above. Depending on the value of the changing parameter h algorithm provides us with a set of up to 900 locally optimal vectors. Hence, we need to refine our optimal rating class algorithm under capacity constraints, Algorithm 2, to receive better results.
In Remark 5.84 we mentioned that we systematically travel through the set $\{0,1\}^{n-1}$ by using the binary representations of the set of integers from 1 to 2^{n-1}. This leads to the fact that we often compare rating class vectors that differ in few entries only. Instead of going from 1 to 2^{n-1} in order, we now shuffle the integers randomly and use this random shuffled numbers. This already leads to a significantly smaller set of locally optimal vectors, often the size shrinks by factor two or more, as the rating class vectors that get compared are more distinct. To shrink this set further, we apply our algorithm k times to receive the locally optimal sets $\mathcal{O}_1, \dots, \mathcal{O}_k$. We can then count how often each rating class vector shows up in the sets. These ideas are formalized in the definition below.

Definition 5.101. Consider a (finite) set of rating class vectors $\tilde{\mathcal{R}}$, which is ordered in the sense that each element r_i of $\tilde{\mathcal{R}}$ has a uniquely assigned number $i \in \{1, \dots, |\tilde{\mathcal{R}}|\}$. Furthermore, we consider k (locally optimal) subsets $\mathcal{O}_1, \dots, \mathcal{O}_k$ of $\tilde{\mathcal{R}}$. By freq we define the *frequency vector* given $\mathcal{O}_1, \dots, \mathcal{O}_k$ as

$$\text{freq}_i = |\{j : r_i \in \mathcal{O}_j\}|,$$

i.e. the number of sets $\mathcal{O}_j$ in which the i-th element of $\tilde{\mathcal{R}}$ occurs.

The ideas are summarized in a new algorithm, Algorithm 3, below.

Algorithm 3 Counting Algorithm under Capacity Constraints

Input: Set of risk classes with the corresponding risk probabilities and class sizes, parameters to compute the premiums and coverages according to Section 5.1 (Definition 5.4 and Proposition 5.6) and Definition 5.74, capacity constraint c, changing parameter h, repetition value k

Output: (absolute) frequency vector freq of all rating class vectors

freq $= (0, \dots, 0)$
for i in $\{1, \dots, k\}$ **do**
  calculate the (locally optimal set of rating class vectors) $\mathcal{O}$ according to Algorithm 2, where the set of all possible rating class vectors is traveled at random instead of in order
  update the freq by setting $\text{freq}_j = \text{freq}_j + 1$ if rating class vector j appears into the set $\mathcal{O}$
end for
return freq

Remark 5.102. Algorithm 3 requires to travel the set of all possible rating class vectors at random. As mentioned in Remark 5.84, the original approach is to set up a for loop from 1 to 2^{n-1} and transform this number into binary representation. Instead of doing this one can shuffle the set $\{1, \dots, 2^{n-1}\}$ before traveling through it.

Remark 5.103. Algorithm 3 is a generalization of Algorithm 2 (which is a generalization of Algorithm 1, see Remark 5.96). This can easily be seen by setting the repetition value in Algorithm 3 to $k = 1$ and browse the set in order instead of at random.

Remark 5.104. By setting the capacity constraint large enough, we can use this algorithm for unconstrained markets as well. Since we might not able to find a unique optimizer in some examples, we can use Algorithm 3 to derive which of the rating class vectors appears in locally optimal sets $\mathcal{O}$ most often. This equips us with an easy approach for mitigating the non-existence of an optimizer, see Example 5.87.

We illustrate the procedure of Algorithm 3 with an example where we consider an element wise capacity constrained market with $c = 0.55$ and $h = 1$.

Example 5.105. We consider a competitive market with the specifications from Example 5.97 together with a capacity constraint of $c = 0.55$ and a changing parameter of $h = 1$. The agents face Bernoulli distributed risks with damage probabilities between 0.1% and 2% with a loss in case of damage of $z = 1$, initial wealth of $a = 2$ and CRRA utility with $\rho = 3$. For this setting, we calculate ten locally optimal sets of rating class vectors $\mathcal{O}_1, \ldots, \mathcal{O}_{10}$. Interestingly, all these sets are almost of same size and contain between 400 and 450 vectors (compared to about 900 when we use Algorithm 2). The intersection of all sets contain just one vector, the rating class vector $(1, 1, 1, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3)$.

The capacity constraints $c = 0.4$ and $c = 0.45$ in Example 5.97 lead to significantly large sets of locally optimal vectors if we set the capacity constraint to $h = 1$. In Example 5.106 we use our newly derived Algorithm 3 from above to calculate the rating class vectors, which are locally optimal for the most sets.
As we already stated in Remark 5.94, these parameter choices actually would not work in reality, as some customers are just unable to be insured, while the insurance is compulsory. Here, we just assume that all customers who cannot be insured at one company insure at the other. The capacity constraint is not applied there. Example 5.106 does not really have a practical relevance and is just included to show how big the optimal sets can become if one does not choose the parameters carefully. Choosing the capacity constraint and the changing parameter careless can cause tremendously big locally optimal sets containing almost half of all rating class vectors.

Example 5.106. We consider a competitive market with the specifications from Example 5.97 with a capacity constraint of $c = 0.4$ and set the changing parameter to $h = 1$. For this setting, we calculate 25 locally optimal sets of rating class vectors $\mathcal{O}_1, \ldots, \mathcal{O}_{25}$. Interestingly, all these sets are almost of same size and contain between 109,000 and 111,000 vectors. The intersection of all sets is empty, while two vectors, the rating class vectors $(1, 1, 1, 2, 2, 2, 2, 3, 3, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4)$ and $(1, 1, 1, 2, 2, 2, 2, 2, 3, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4)$ show up in all but one set. In Figure 31 below, the number of vectors for each frequency from 0 to 25 is plotted. As mentioned above, there is no vector included in all 25 sets.
It is easy to see that the number of rating class vectors which appear in exactly k sets of locally optimal vectors is decreasing in k. The total amount of possible rating class vectors is $2^{19} = 524,288$. Hence, about every seventh rating class vector is not contained in any local optimal set, less than 0.1% of the rating class vectors is contained in 20 or more of the 25 sets.
If we keep the changing parameter at $h = 1$, enlarge the capacity constraint to $c = 0.45$ and calculate 50 optimal sets (containing about 200,000 vectors each), the vector $(1, 1, 1, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 4, 4, 5, 5, 5, 5, 5)$ is contained in all sets for our example.

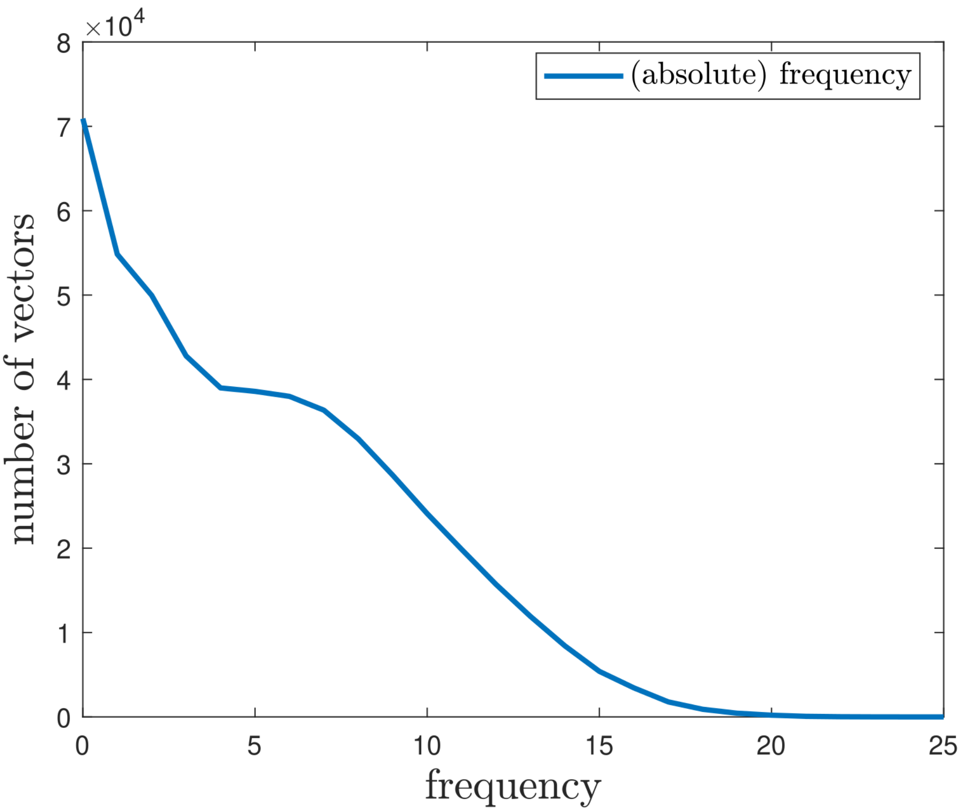


Figure 31: Absolute frequencies of appearance of different rating class vectors for each possible frequency from 0 to 25 in an element wise capacity constrained insurance market with capacity constraint $c = 0.4$ in scenario (C)

Remark 5.107. The number of locally optimal sets we calculate for an example is based on the size of the underlying sets of locally optimal rating class vectors. As mentioned before, the sizes of the different locally optimal sets for each capacity constraint are quite similar, in our examples between about 400 to 450 for $c = 0.55$ in Example 5.105 and about 200,000 for $c = 0.45$ in the above example. The smaller the average set size is, the less sets we need to calculate to reach the point where there are only few vectors which are contained in all or all but one set. The examples also show that using small capacity constraints together with large changing parameters cause for markets that are not robust at all.

After analyzing how the market behaves under different capacity constraints c, we finally examine how the value of the changing parameter h affects the market.

Example 5.108. We look at a competitive market with element wise constrained vectors. In this example, we are reconsidering part (d) from Example 5.85. This means, we have a market with 4000 customers in 20 risk classes of equal size, suffering from damage probabilities of 0.2%, 0.4%,. . . ,4%. While our capacity constraint is fixed to be $c = 0.5$, the changing parameter h is varied in order to analyze its effect.

risk class	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$h = 0$	1	1	2	2	2	2	3	3	3	3	3	3	4	4	4	4	4	4	4	4
$h = 0.2$	1	1	2	2	2	2	3	3	3	3	3	3	4	4	4	4	4	4	4	4
$h = 0.4$	1	2	2	2	3	3	3	3	3	3	4	4	4	4	4	4	4	4	4	4
$h = 0.6$	1	2	2	2	3	3	3	3	3	3	4	4	4	4	4	4	4	4	4	4
$h = 0.8$	1	2	2	2	2	3	3	3	3	3	4	4	4	4	4	4	4	4	4	4
$h = 1$	1	2	2	2	3	3	3	3	3	4	4	4	4	4	4	4	4	4	4	4

Table 17: (Element of a set of locally) Optimal rating class vectors for customers of 20 different risk classes for varied changing parameters in scenario (C)

We use our counting algorithm under capacity constraints, Algorithm 3, for this example as well. Compared to Examples 5.105 and 5.106, our locally optimal sets are quite small. All sets contain between 30 and 50 rating class vectors, resulting in a low number of only five repetitions to choose an optimizer.
We can see that a larger changing parameter tends to result in smaller risk classes, especially for the low-risk customers. This can be explained by the fact that larger rating classes lead to more cases, where the capacity constraint is reached. The higher the changing parameter is, the worse this becomes for the insurer, as more and more customers change to the concurring insurance company.

6 Health Insurance

In this chapter we apply our basic model from Chapter 2 to health insurance markets. In Section 6.1 we provide some adjustments in order to make our model applicable to health insurance data. We continue by applying our model to simulated data after modeling the risk in Section 6.2. In a further step, in Section 6.3, we look at real-life data. The provided data and the adjusted model are based on expected losses and do not take the loss distributions into account. In Section 6.4, we model different possible risk distributions and adjust our model to be able to deal with them. We illustrate these extensions with some numerical examples. Health insurance products are usually life-long contracts, where the customer pays a so-called flat premium, i.e. a premium which is constant over lifetime. We therefore need to model reserves, which are first built up and then consumed in later ages. This modeling is done in Section 6.5. For basic definitions and concepts of health insurance together with related actuarial techniques we refer to Pitacco (2014) [Pit14].

6.1 Adjustments of the Basic Model

In contrast to life insurance contracts, health insurance contracts do not allow the customer to choose the amount of coverage he wants to purchase. Speaking in terms of our basic model from Chapter 2, this means that the amount of coverage, the agent purchases, is set to $\hat{\lambda} = 1$. Of course, the demand of insurance in the market still depends on the price of it. It affects the number of customers which are willing to purchase insurance for a given price. In Chapter 2 we have seen that customers in life insurance markets purchase full coverage if the price of the insurance is equal to the net expected loss of the insured. Therefore the point where the market premium corresponds to the net expected loss is used as a reference point.
We model the relative amount of customers purchasing health insurance of price π by

$$e^{-\rho^{-1}\frac{\pi-\pi^0_{\oplus/\ominus}}{\pi^0_{\oplus/\ominus}}},$$

where $\pi^0_{\oplus/\ominus}$ denotes the net expected loss of an insured of type $\pi^0_{\oplus/\ominus}$ and ρ is the risk aversion parameter of our demand function.
From the representation

$$e^{-\rho^{-1}\frac{\pi-\pi^0_{\oplus/\ominus}}{\pi^0_{\oplus/\ominus}}} = \frac{-e^{-\rho^{-1}\frac{\pi-\pi^0_{\oplus/\ominus}}{\pi^0_{\oplus/\ominus}}}}{-e^{-\rho^{-1}\frac{\pi^0_{\oplus/\ominus}-\pi^0_{\oplus/\ominus}}{\pi^0_{\oplus/\ominus}}}} \tag{19}$$

we can see that our new demand function originates from a CARA utility function. As desired, the function decreases with an increasing value of π.

Note that taking the inverse of the risk aversion parameter ensures that a higher risk aversion parameter models a higher risk aversion. If we would take the parameter itself instead of its inverse, a higher risk aversion parameter would result in a lower risk aversion and vice versa.
As we already know, the value of the equilibrium insurance coverage function $\hat{\lambda}$ does not depend on the initial wealth of the customers a if one equips the agents with CARA utility, see Remark 2.36. As adding or subtracting an initial wealth in the exponential functions in the nominator and the denominator of Equation (19) does not change the value of the fraction, also the values of the new demand function are independent of the initial wealth of the customers.
Note that this adjustment takes the distribution of the risk completely out of the computation, as the value of the demand function solely depends on the expected value of the risk but not on the distribution of the risk. We return to that problem later, in Section 6.4.
Using

$$\hat{\lambda}_{\oplus/\ominus}(\pi) = e^{-\rho^{-1}\frac{\pi-\pi^0_{\oplus/\ominus}}{\pi^0_{\oplus/\ominus}}}$$

as an optimal demand function, we can perform a similar analysis as in Chapter 2. Let us summarize this procedure shortly. To do so, we switch the notation from $\oplus$- and $\ominus$-agents to males and females. This comes due to the fact that we do not know which gender is generally more favorable to the insurance company without knowing the exact market specifications but always deal with data that relies on genders.

1. We start by setting up all parameters that characterize the insurance market. This is done by choosing the fraction of males w_m in the insurance market as well as the risk aversion parameter ρ for our demand function. As mentioned above, this function is independent of the agents initial wealth a. For the risk variables of the males and females $Z_{m/f}$ we calculate the net premiums $\pi^0_{m/f} = \mathbb{E}[Z_{m/f}]$. Note that the risks of the agents of each gender are i.i.d. and independent of the other gender.

2. We set $\hat{\lambda}_{m/f}(\pi) = e^{-\rho\frac{\pi-\pi^0_{m/f}}{\pi^0_{m/f}}}$ as the (optimal) demand function $\hat{\lambda}$ that depends on the premium π.

3. The optimization problems are now the same as in Chapter 2. If we are in a monopolistic market, we solve

$$\hat{\pi}_\odot = \arg\max_{\pi_\odot}\{w_m\hat{\lambda}_m(\pi_\odot - \pi^0_m) + w_f\hat{\lambda}_f(\pi_\odot - \pi^0_f)\}$$

in regime (E) and

$$(\hat{\pi}_m, \hat{\pi}_f) = \arg\max_{(\pi_m,\pi_f)}\{w_m\hat{\lambda}_m(\pi_m - \pi^0_m) + w_f\hat{\lambda}_f(\pi_f - \pi^0_f)\}$$

in regime (F). The exponential structure of the equations ensures the existence of the maximizers. In a market with perfect competition we solve

$$\pi_{\odot} = \frac{w_m \hat{\lambda}_m}{w_m \hat{\lambda}_m + w_f \hat{\lambda}_f} \cdot \pi_m^0 + \frac{w_f \hat{\lambda}_f}{w_m \hat{\lambda}_m + w_f \hat{\lambda}_f} \cdot \pi_f^0$$

in regime (E) and set

$$\hat{\pi}_m = \pi_m^0, \quad \hat{\pi}_f = \pi_f^0$$

in regime (F). For more details we refer to Section 2.5. From the definition of $\hat{\lambda}$ we can easily see that this demand function is differentiable in the premium π. Hence, no involved argument with the implicit function theorem as in Proposition 2.22 is needed.

Remark 6.1. Instead of replacing the equilibrium coverage by an optimal demand function, we could model the market as a market with price-quantity competition. Therefore we would need to set markets with different coverages and corresponding premiums and perform a similar analysis such as in [SS14].

Remark 6.2. There are not many papers which try to investigate health insurance markets in a regime with mandatory unisex tariffs. The closest paper to this topic might be Riedel (2006) [Rie06]. Especially the effect of premium refund systems are investigated there. Unfortunately, the formulas used lack some details and it is not clear how the parameters were chosen. Hence, the numerical results cannot be recalculated and therefore also not be used as a comparison in this thesis.

6.2 Modeling the Health Insurance Risk

In this section we model an ambulant and stationary health insurance. We then use our simulated data to calculate premiums according to our adjusted model.
The risk for the insurance company is supposed to consist of two components: A normally distributed component to model medical consultations and a Pareto distributed component to model hospitalization expenses. To be more precise, we use

$$X_{\text{doctor}} \sim \mathcal{N}(100, 25) \text{ and } X_{\text{hospital}} \sim \text{Par}\left(1000, \frac{4}{3}\right).$$

Here, the first parameter of the Pareto distribution is its scale parameter, i.e. the minimal value the distribution can take, and the second parameter is its shape parameter. Hence $\mathbb{E}(X_{\text{hospital}}) = 4000$ and $\mathbb{V}(X_{\text{hospital}}) = \infty$. Like already mentioned in Remark 2.13, the expectation of the risk is bounded as desired, but not so the variance. It is not uncommon in practice to use heavy tailed distributions like the Cauchy or Pareto distribution for modeling, e.g. for fire insurances. Because the expenses for medical treatments in a hospital might be unbounded as well, we chose to use a Pareto distribution here. Note that our model from Section 6.1 is not taking the distribution or variance into account. We will return to that problem later, in Section 6.4. Depending on age and gender, we model the frequency of the claim occurrence for hospital stays and doctor's consultations. This is done in two steps. First we include the probability that an agent needs medical treatment during one year. The probabilities we use for this part of our model are based on data sets of the German Robert Koch Institute (RKI) [Rob17a, Rob17b] from 2017 (only available in German) and are summarized in the following tables.

age	doctor's visit	hospital stay
18-29	0.784	0.087
30-44	0.776	0.095
45-64	0.850	0.159
older than 65	0.937	0.258

Table 18: Probabilities of males to have at least one doctor's visit or hospital stay during one year given the age

age	doctor's visit	hospital stay
18-29	0.904	0.151
30-44	0.878	0.114
45-64	0.908	0.144
older than 65	0.940	0.259

Table 19: Probabilities of females to have at least one doctor's visit or hospital stay during one year given the age

We assume that a 20-year old person is having 1.5 doctor's visits on average each year, given that he has seen a doctor at least once during the year. This number is modeled to increase exponentially with the age, doubling each 20 years.

Furthermore, we assume that a 20-year old person is having one hospital stay per year if this person has to stay in the hospital during the year. Again we assume that this number is increasing exponentially with the age. As the probability for staying in a hospital at least once per year is already increasing with the age (especially for men), we assume the number to go up by 60% each 20 years. Remark 6.5 we can see that these choices reflect the reality quite well.

To be exact, the formula for expected number of doctors visits for a person aged x is given as

$$1.5 \cdot 2^{\frac{x-20}{20}} \cdot (0.784 \cdot \mathbb{1}_{x\in[18,29]} + 0.776 \cdot \mathbb{1}_{x\in[30,44]} + 0.850 \cdot \mathbb{1}_{x\in[45,64]} + 0.937 \cdot \mathbb{1}_{x\in[64,\infty)})$$

for males and

$$1.5 \cdot 2^{\frac{x-20}{20}} \cdot (0.904 \cdot \mathbb{1}_{x\in[18,29]} + 0.878 \cdot \mathbb{1}_{x\in[30,44]} + 0.908 \cdot \mathbb{1}_{x\in[45,64]} + 0.940 \cdot \mathbb{1}_{x\in[64,\infty)})$$

for females. The expected number of hospital stays is given by

$$1.6^{\frac{x-20}{20}} \cdot (0.087 \cdot \mathbb{1}_{x\in[18,29]} + 0.095 \cdot \mathbb{1}_{x\in[30,44]} + 0.159 \cdot \mathbb{1}_{x\in[45,64]} + 0.258 \cdot \mathbb{1}_{x\in[64,\infty)})$$

for males and

$$1.6^{\frac{x-20}{20}} \cdot (0.151 \cdot \mathbb{1}_{x\in[18,29]} + 0.114 \cdot \mathbb{1}_{x\in[30,44]} + 0.144 \cdot \mathbb{1}_{x\in[45,64]} + 0.259 \cdot \mathbb{1}_{x\in[64,\infty)})$$

for females. Note that the above formulas give the expected number of doctors visits and hospital stays, not the expected costs. In order to receive the (variables for the) costs one needs to multiply the expected numbers by X_{doctor} and X_{hospital}, respectively.

Example 6.3. Given all this information, it is easy to calculate the net expected losses of the insured and therewith also the monopolistic and unisex premiums and the corresponding insurance demands.

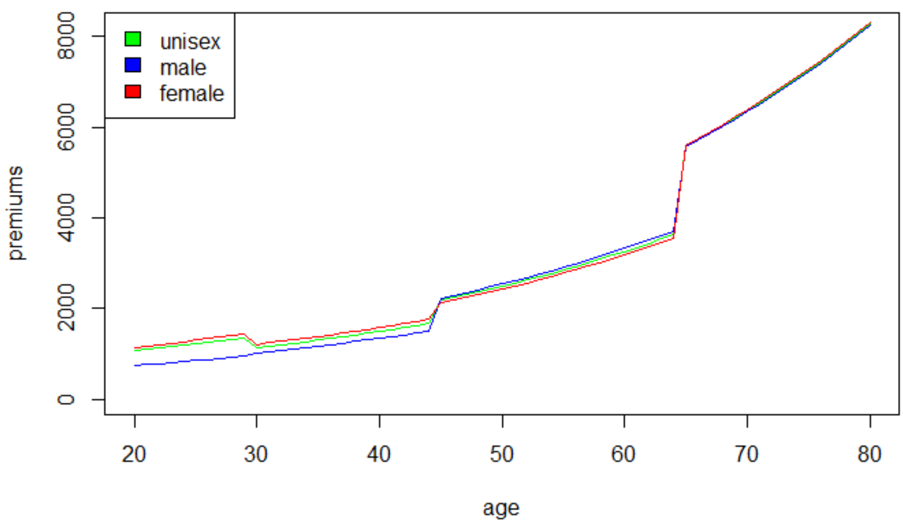


Figure 32: One-year health premiums as functions of the initial age in scenario (C)

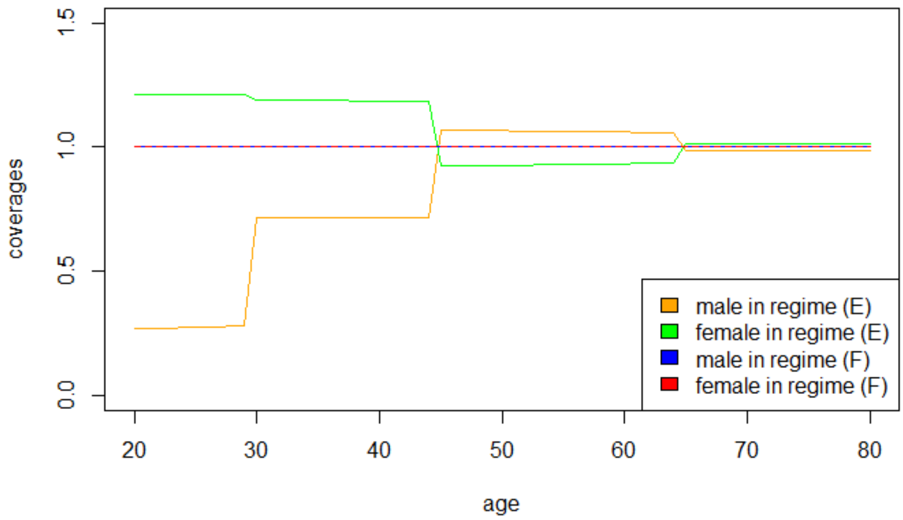


Figure 33: Equilibrium insurance demands of one-year health premiums as functions of the initial age in scenario (C)

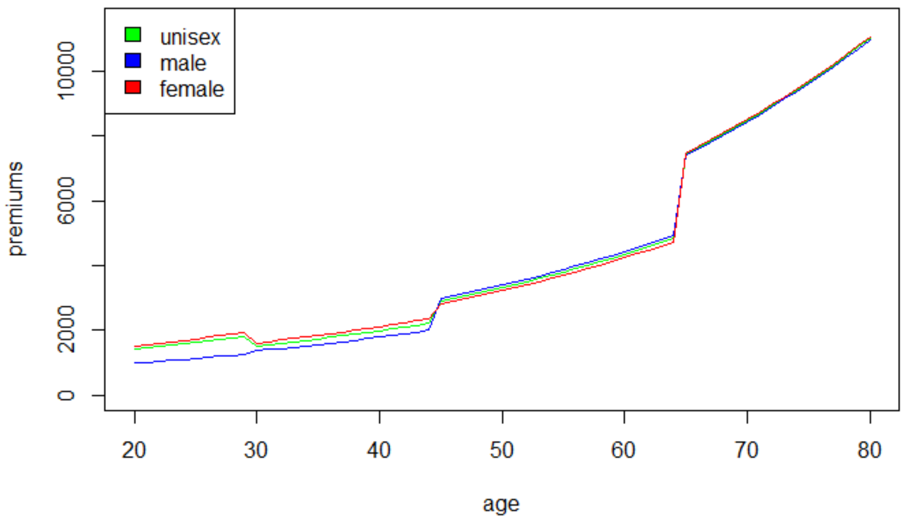


Figure 34: One-year health premiums as functions of the initial age in scenario (M)

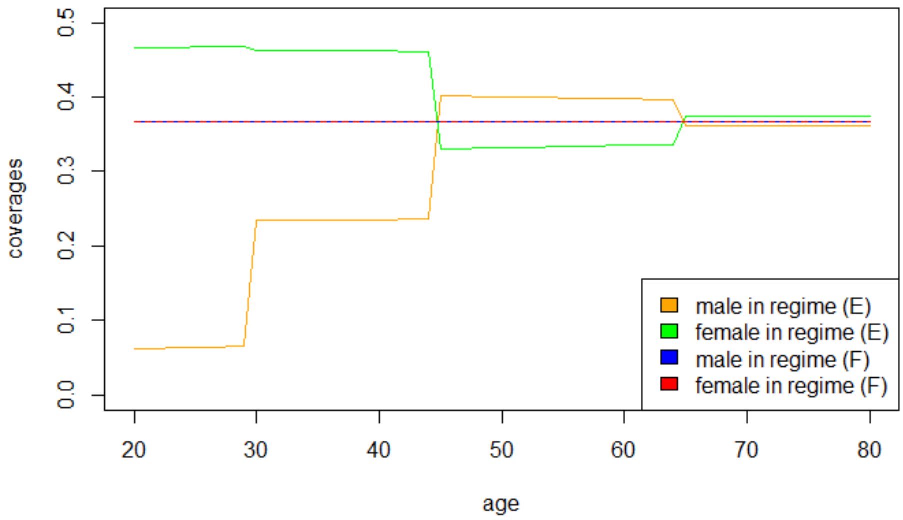


Figure 35: Equilibrium insurance demands of one-year health premiums as functions of the initial age in scenario (M)

In Figures 32 and 34 these premiums are plotted for different ages in the competitive and the monopolistic market. The corresponding insurance demands are given in Figures 33 and 35. We assume the risk aversion parameter of our demand function to be $\rho = 3$ and the portfolio to consist of equal shares of males and females, i.e. $w_m = 50\%$.
Of course, the piece wise modeling of the probabilities can be seen in the plots. As expected, the unisex premium lies always between the male and the female premium. If the (relative) difference of the two gender-specific premiums is high, the unisex premium stays close to the higher gender-specific premium. As usual, the monopolistic premiums are higher than the competitive premiums, while the difference of the premiums in the two scenarios increases with an increasing value of the risk aversion parameter ρ.
The comparison of Table 18 and 19 reveals that the gender has a big influence on the number of doctor's visits and hospital stays. Which gender produces higher medical expenses depends on the age of the customers. Particularly, this can be seen when comparing the insurance demands. For more details we refer to Remark 6.5 and Section 6.3.

Remark 6.4. Different to the computations in the previous chapters, the calculations in this chapter and Chapter 7 are implemented in R. The reason for this is that R is much more popular in the health insurance sector than MATLAB, as it can be used for free. In R we are using the built in `optimize` function and proceed similar as for the previous chapters, see Remark 2.35 for more details. We made use of the `actuar` package of R. This package

contains actuarial functions and heavy tailed distributions for R and can be found on GitLab[9]. We use the version of November 2021.
In this chapter, all plots are based on functions over the initial age. Hence, each line in a plot is based on a number of points equal to the number of initial ages investigated. In this section we always investigate initial ages between 20 and 80, resulting in the calculation of 61 data points for each line. In later sections, we are interested in initial ages between 21 and 100, so we need to calculate 80 data points per line.

Remark 6.5. As we see in Section 6.3, the premium heights we are ending up with are indeed quite realistic. The ambulant health costs we are dealing with in Examples 6.6 and 6.7 later on make up a bit more than half of our total medication costs, also the premiums in these cases are about half as high. The total medication costs are in this setting given by the sum of the ambulant and the stationary costs. Other expenses like seeing the dentist or sickness day allowances are not included here. Nonetheless, we can observe two differences:

1. Because of shrinking portfolio sizes in higher ages, the real-life data source we are going to use starts to average over the costs for customers over 85, therefore the costs and also the premiums turn out to be constant there.

2. We can see that the male and female costs differ more from each other for the real-life data. Especially the pregnancy costs are much more visible and pronounced in real-life data. While in our modeled data there is only a small gap, we are able to observe a so-called "pregnancy hill" known as "Schwangerschaftshügel" in German health care companies.

6.3 Real-Life Data

As mentioned before, it is our goal to apply our model to real-life health insurance data. Before doing so, we need to collect appropriate data sets.
Each year, the German Federal Financial Supervisory Authority (Bundesanstalt für Finanzdienstleistungsaufsicht, BaFin) publishes probability tables for private health insurance [10]. These tables consist of five different parts, where the structure and exact names of the tables have changed over time.

- "Profile": standardized loss profiles, standardized and unstandardized losses per head and portfolio sizes for each health insurance product,

[9] `https://gitlab.com/vigou3/actuar`
[10] `https://www.bafin.de/DE/PublikationenDaten/Statistiken/PKV/wahrscheinlichkeitstafeln_node.html`, also called loss per head statistics, only available in German; visited May 2022

- “Grundkopfschäden”: basic losses per head for each health insurance product,
- “Fiktive Selbstbehalte”: fictitious deductibles for each health insurance product that contains ambulant insurance benefits,
- “Erklärungen”: explanations about the data sets, used abbreviations, etc.,
- “Grafiken”: visualizations for the data from the data sets.

Let us give some further explanations about the contents of the “Profile” table. This table lists separated data for males, females and pregnancy costs and contains data for stationary and ambulant costs, costs resulting from dental treatments and the (compulsory) long-term care insurance as well as daily sickness allowances. Furthermore, different lapse rate tables are included. For more details about the tables and the exact columns used for our computations, we refer to Appendix B.
The probability tables of the BaFin is not the only possible data source for real-life health insurance data. Nonetheless, many other possible sources, like the RePortal data from the umbrella organization of the German private health insurers (Dachverband der privaten Krankenversicherer in Deutschland) are not open source.
Unfortunately, it is quite difficult to find data which contains the distributions for the losses per head and not only expected values. The distribution of the losses highly effects the demand for insurance, as we have seen in Section 3.2, compare also Chapter 8. Therefore, we need to generate a distribution for the losses and include them in our model. We return to this problem later, in Section 6.4.

Example 6.6. We first investigate the (one-year) health insurance premiums and demands for males and females as well as the unisex premium in a competitive scenario. Therefore we analyze an ambulant health care with “Beihilfe”. The so-called “Beihilfe” is a system in Germany that provides financial aid for medical costs in the private health insurance sector. Civil servants such as teachers, police officers or fire fighters as well as soldiers and judges and under specific circumstances the relatives of all these people are eligible to receive this or a similar kind of financial aid. We model a market without deductible, where the losses per head are coming from the BaFin data from 2020. We assume that the agents are equipped with a risk aversion parameter of $\rho = 3$ and the fraction of males w_m in the portfolio to be 50%.

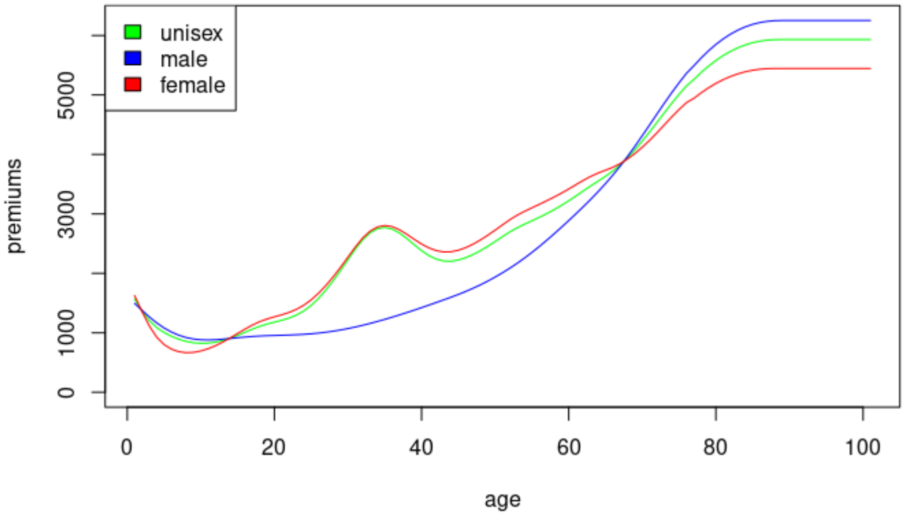


Figure 36: One-year health premiums as functions of the initial age including pregnancy costs in scenario (C)

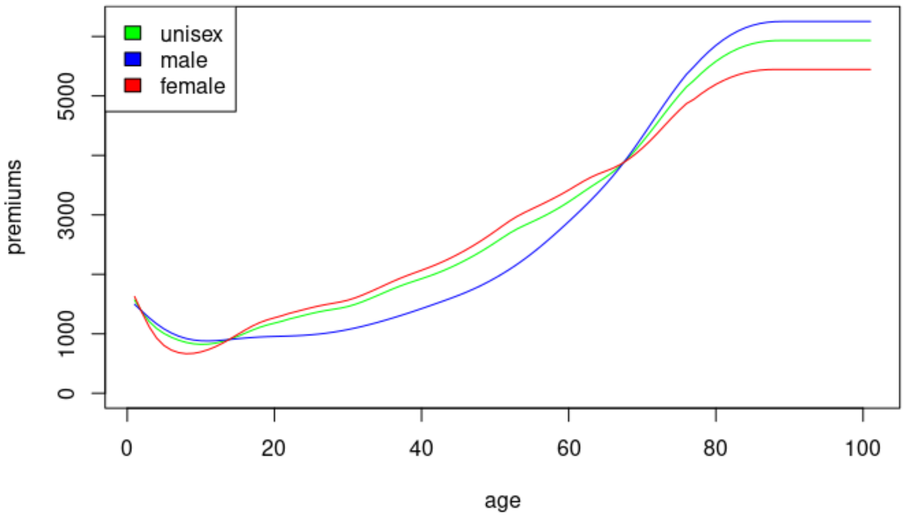


Figure 37: One-year health premiums as functions of the initial age excluding pregnancy costs in scenario (C)

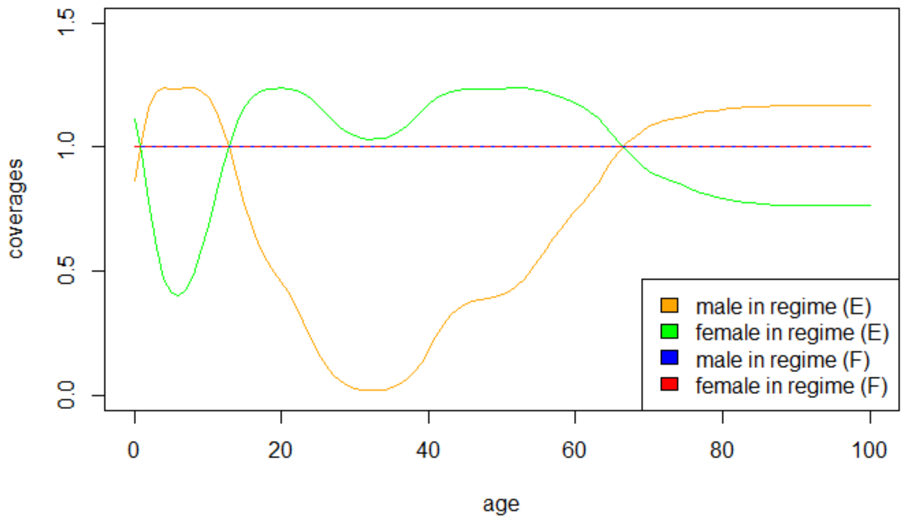


Figure 38: Equilibrium insurance demands of one-year health premiums as functions of the initial age including pregnancy costs in scenario (C)

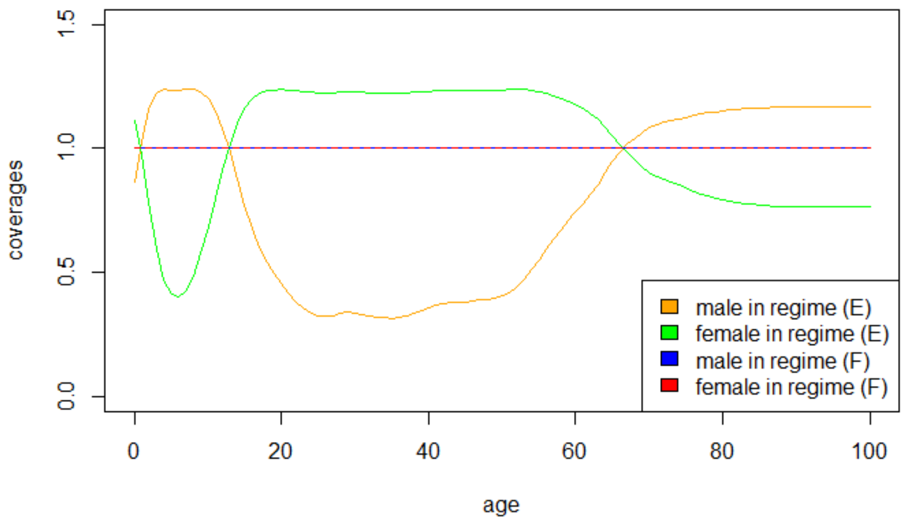


Figure 39: Equilibrium insurance demands of one-year health premiums as functions of the initial age excluding pregnancy costs in scenario (C)

As we can see, it depends on the age of the customer whether men or women have higher average medical costs. Young men up to an age of 16 tend to need more medical aid than females of the same age. The reason for this

phenomenon might be that boys hurt themselves more often while playing than girls. For an age between 16 and 70, women are producing higher medical costs not only due to pregnancy (note the difference between Figure 36 and 37) but also due to the fact that they make use of medical check-ups and preventive examinations more often. The lack of these kind of treatments might also be a factor for the higher losses per head of the males over 70.
Note that the medical costs are not constant for very high ages, as shown in Figures 36 and 37, but the numbers of customers in these ages are so small that the costs are averaged over all these ages.
Due to the structure of the renewed demand function, the demand can never be zero, compare Section 6.1. Nonetheless, there are situations, where almost no customer of one gender buys insurance. As constructed, the insurance demand is one when there is free contract design in the market. In regime (E), one gender subsidizes the other. While this could already be seen for life insurance policies in Chapter 3, the subsidizing and subsidized genders change depending on the age. Especially when we include pregnancy costs, males between 30 and 35 are almost driven out of the insurance market.

Example 6.7. Of course, we are also interested in the monopolistic premiums and insurance demands for the health insurance product from the last example.

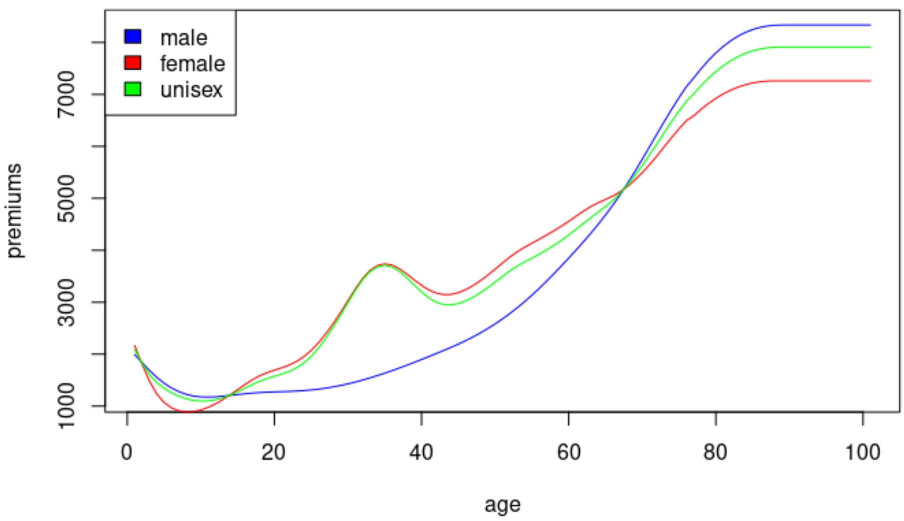


Figure 40: One-year health premiums as functions of the initial age including pregnancy costs in scenario (M)

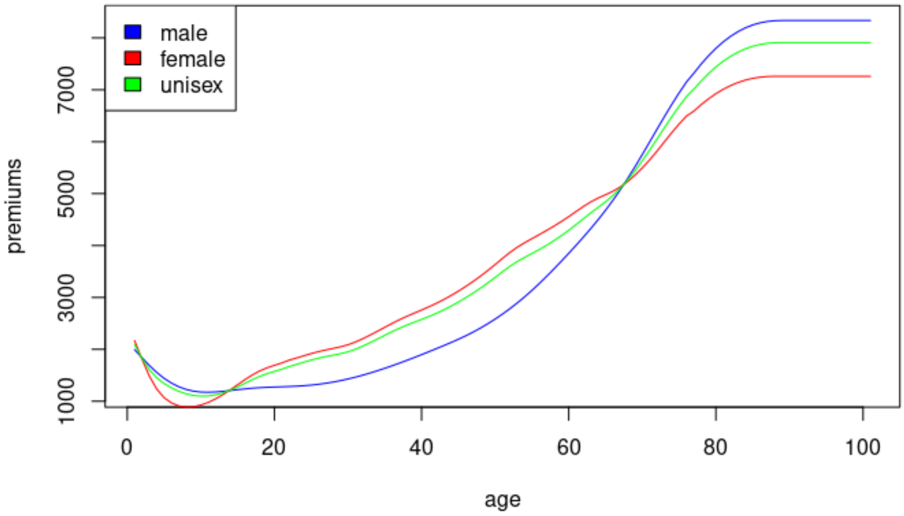


Figure 41: One-year health premiums as functions of the initial age excluding pregnancy costs in scenario (M)

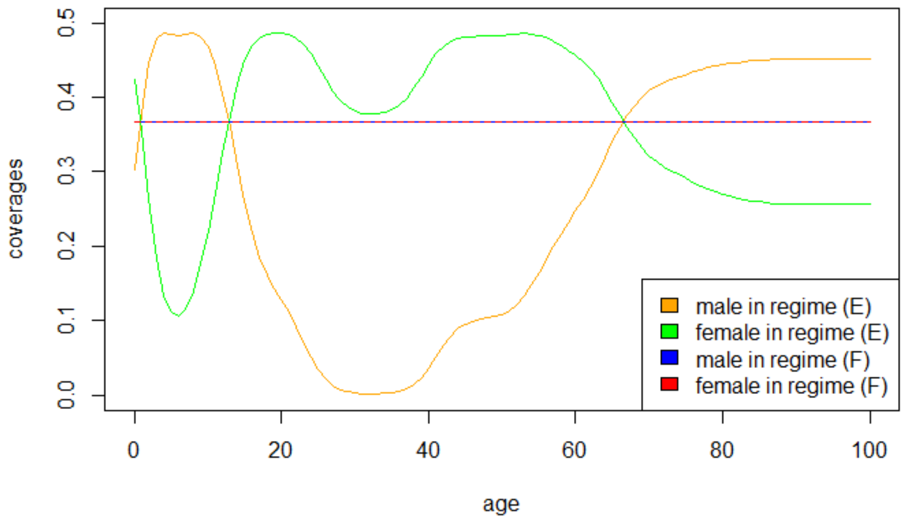


Figure 42: Equilibrium insurance demands for one-year health premiums as functions of the initial age including pregnancy costs in scenario (M)

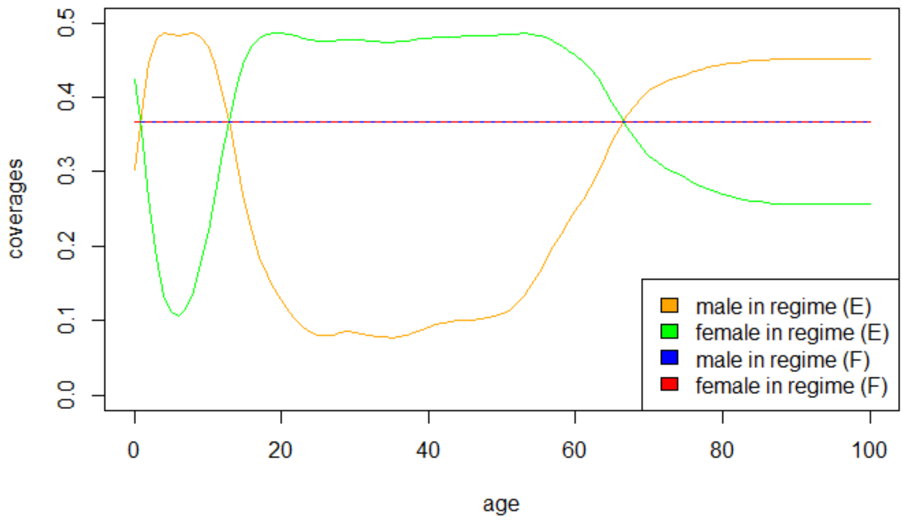


Figure 43: Equilibrium insurance demands for one-year health premiums as functions of the initial age excluding pregnancy costs in scenario (M)

Comparing Example 6.6, the shapes of the premium curves remain more or less unchanged while the price of the insurance increases roughly by 33%. This rise in the premiums gets reflected in the demands. As for the premiums, the shape of the demand curves are almost unchanged, while the insurance demand, respectively the number of customers deciding to purchase insurance from a company, is drastically lowered.

Remark 6.8. As the phenomena that can be observed for the demand functions in later examples do not qualitatively differ from these ones, we decided to omit plotting the demands and focus on the premiums instead.

Remark 6.9. Usually, an insurer would charge costs in order to meet its expenses, operating costs etc. The types of costs together with their allowed heights are fixed for health insurance products in Germany. In detail, the following costs can be charged:

- acquisition costs of 6.4% of the expected average basic losses per head,
- loss adjustment expenses of 4% of the expected average basic losses per head,
- administrative expenses of 2.3% of the expected average basic losses per head,
- safety loading of 4% of the basic loss per head of the corresponding age,

- 0.7% of the basic loss per head of the corresponding age as a surcharge for the standard and basic tariff; these tariffs cover the (basic) medical expenses of customers that are not able to pay the normal premiums anymore.

Similar to life insurance products, including costs in health insurance contracts is straightforward, see Remark 3.10. As before, the observed effects do not change qualitatively. Hence, all costs are omitted to keep the calculations simple.

6.4 Further Adjustments of the Model

As mentioned before, the adjusted optimal demand function

$$\hat{\lambda}_{m/f}(\pi) = e^{-\rho^{-1}\frac{\pi-\pi^0_{m/f}}{\pi^0_{m/f}}}$$

from Section 6.1 depends solely on the expected value of the risk and not on the distribution. As we have seen in Section 3.2, the distribution, especially the variance of the risk, has a high effect on the demand of insurance. As we have seen when comparing Figures 11 and 13, the demand for insurance in case of the term insurance is much higher than in the case of the endowment insurance. This makes sense, as an insurance is needed especially for risks, which are hard to predict and have a high worst-case loss.
In the following, we present three different possible adjustments for the demand function. Again, $\pi^0_{m/f}$ denotes the net expected loss of an insured of type $\pi^0_{m/f}$ with risk $X_{m/f}$ and ρ is the risk aversion parameter of our demand function. With $\mathbb{V}(u(X))$ and $\sigma(u(X))$, we denote the variance and standard deviation of $u(X)$, where u is a utility function.

a) $e^{-\rho^{-1}\frac{\pi-\pi^0_{m/f}}{\pi^0_{m/f}+\mathbb{V}(u(X))}}$: One option is to add the variance of the utility of the risk to the denominator. A higher variance results in a lower value of the fraction and therefore to a value of the demand function that lies closer to one.
Unfortunately, this approach only enlarges the demand for insurance if the variance is high but does not reduce it if it is low. To solve this problem, we present a second approach.

b) $e^{-\frac{\rho^{-1}}{\mathbb{V}(u(X))}\frac{\pi-\pi^0_{m/f}}{\pi^0_{m/f}}}$: By manipulating the (inverse) risk aversion parameter ρ by dividing it through the variance of the utility of the risk, we receive a demand function which matches our criteria: It lies closer to one, the bigger the variance is and equals one whenever the premium is equal to $\pi^0_{m/f}$. Note that a higher variance results in a higher risk aversion parameter and therefore in a higher risk aversion. The problem with this approach is that even by normalizing the risk X by its expectation, our model is not robust and reacts very sensitive to changes in the risk of the customers. These changes might affect the variance and therefore also the demand. By replacing the variance by the standard deviation, this effect can be weakened.

c) $e^{-\frac{\rho^{-1}}{\sigma(u(X))}\frac{\pi-\pi^0_{m/f}}{\pi^0_{m/f}}}$: Here, the (inverse) risk aversion parameter ρ is manipulated by dividing it through the standard deviation (square root of the variance) of the utility of the (normalized) risk. This seems like a good choice to depict the spread of the distribution in our demand function. Unless otherwise stated, we are going to use this demand function in the following.

The classic pricing of health insurance products only uses the expected values of the losses, not their distributions. Also, all public data sources we found only contain expected losses but no loss distributions. Therefore, we need to derive a distribution on our (real-life) data.
We already know, what the expected values of the distributions have to be, namely the corresponding (expected) value of our real-life data. It is our goal, to model different distributions, where we choose the parameters of the distribution in such a way that the desired expectation is reached, while we are often able to choose our variance by determining the values of the parameters accordingly. Of course, this is impossible for distributions which only depend on one parameter. Possible candidates for distributions are:

- the exponential distribution (variance cannot be chosen, as we have only one parameter, which is fixed due to the expectation constraint),
- the normal distribution (variance can be easily chosen with the variance parameter σ^2),
- the uniform distribution (variance can be chosen by setting the size of the interval on which the density of the distribution is non-zero accordingly),
- the Pareto distribution (variance can be chosen by setting suitable values for the shape and the scale parameter, it is also possible to choose the parameters such that the distribution has infinite variance and is heavy tailed, which might be useful for stationary costs, compare Section 6.1).

Of course, one could extend this list by adding more distributions.

Remark 6.10. Assume, we are choosing a Pareto distribution with a parameter setting that ensures infinite variance, i.e. we set the shape parameter a

to be in $(1, 2]$. In this case, not only X but also $u(X)$ would have infinite variance. This would lead to a division by ∞, which cannot be calculated directly. If we calculate the limit instead, we see that we always end up with an insurance demand of one. Especially in a monopolistic market setting, this leads to exploding premiums, as customers always buy insurance, no matter how high the premium is. In this case it might be better not to use the Pareto distribution itself but to cut it off at a certain value, e.g. after 50 or 100 times the expected damage.
On the other hand, if the variance is zero, i.e. if the risk is a.s. constant, we also receive a division that is not defined. Taking the limit in approach b) and c) leaves us with a setting where no insurance is bought at all. Indeed, if the premium is at least as high as the expected loss an agent is not interested in buying any insurance.

Example 6.11. As in Example 6.6, we consider an ambulant health care with "Beihilfe" in a competitive scenario without deductible, where the losses per head are coming from the BaFin data from 2020. We assume that the risk aversion parameter is given as $\rho = 3$. In Figure 44, we assume the risk to be uniformly distributed on the interval $[0, 2 \cdot \pi^0_{\oplus/\ominus}]$, i.e. between zero and twice the expected loss. In Figure 45, we assume the loss to be exponentially distributed where the parameter is chosen to be the inverse of the expected loss. All the examples are calculated with the third approach for the demand function, i.e. by taking $\hat{\lambda}_{m/f}(\pi) = e^{-\frac{\rho-1}{\sigma(u(X))}\frac{\pi-\pi^0_{m/f}}{\pi^0_{m/f}}}$. The utility function for calculating $\sigma(u(X))$ is chosen to be CARA utility with a risk aversion parameter of three.

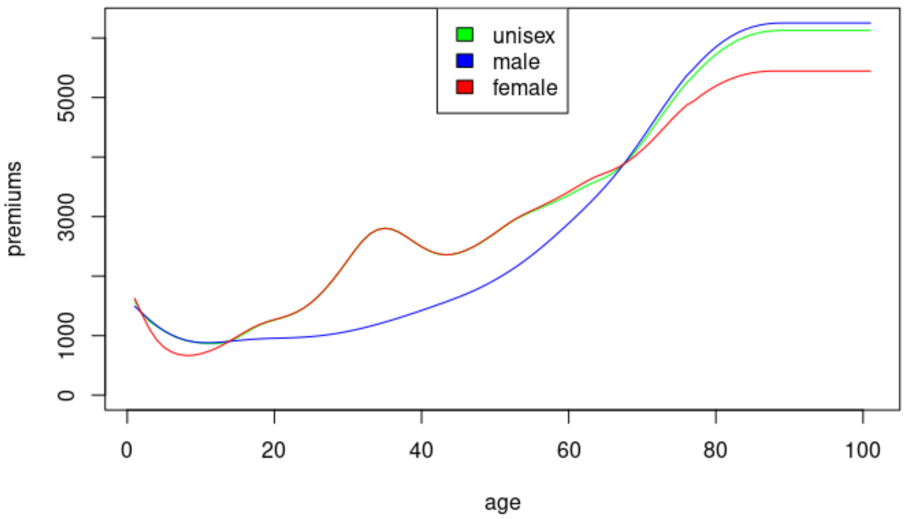


Figure 44: One-year health premiums as functions of the initial age with uniform distribution in scenario (C)

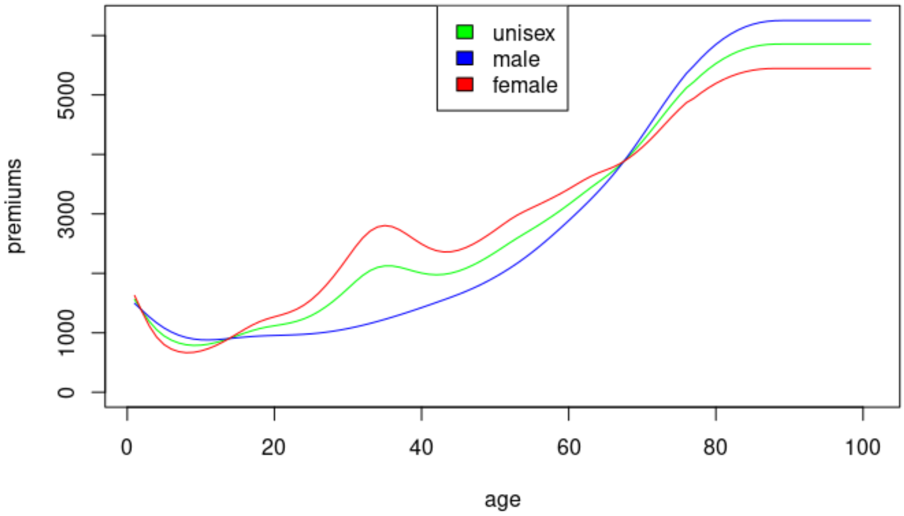


Figure 45: One-year health premiums as functions of the initial age with exponential distribution in scenario (C)

As we can see, the demand for insurance and therefore the unisex premium highly depends on the underling distribution. If the risk is uniformly distributed and therefore bounded, the demand for insurance is much lower than if it is unbounded, like in the exponential case. As mentioned above, the demand does not depend on the initial wealth of the agents, see the explanations at the beginning of Section 6.1 and compare Remark 2.36. Replacing the risk aversion parameter by a modified one like we did it in this section does not affect this feature.

As there is no structural difference for the competitive and the monopolistic scenario here, we do not present examples for the scenario (M) for this market setting.

6.5 Reserves and Bruttopremiums

Health insurance products are usually life-long contracts with a (constant) flat premium, also called the bruttopremium, over the years. Of course, insurance companies can raise the annual contributions under certain conditions. As stated in Remark 6.14, including medical inflation in the model would be straightforward. To ease the notation, we do not regard medical inflation here.
To calculate the bruttopremium, we need to calculate a pension series ps in order to determine the future cash flow. We calculate the series by backward induction. We drop the dependency on the gender of the agent in all formulas of this section to increase the readability.

Definition 6.12. Assuming that the interest rate r is constant, we define the *pension series* ps by

$$ps(\omega+1) = 0 \text{ and } ps(k) = \frac{1}{1+r} ps(k+1) + \tilde{p}_k,$$

where ω is the maximum age of the population (here $\omega = 100$) and $k = 0, \dots, \omega - 1$. Furthermore, $\tilde{p}_k$ is the probability that a customer "survives" the year, i.e. he is still a customer of the insurance company at the end of the year, so he has neither died nor canceled the contract.

With the pension series and the annual premiums we can calculate the bruttopremium. This is done by dividing the (discounted) sum of the remaining one-year premiums π_k by the corresponding value of the pension series. We summarize this in a formula in the following definition.

Definition 6.13. The *bruttopremium bp* of an x-year old is defined as

$$bp(x) = \frac{\sum_{k=x}^{\omega} \left(\frac{1}{1+r}\right)^{k-x} \pi_k}{ps(x)},$$

where the value π_k denotes the premium of a customer aged k.

Remark 6.14. If we are in a competitive market with free contract design, the premium π_k equals expected loss π_k^0 of a k-year old person. Note that one could model the premium π_k to depend on the initial age/birth year of the customer, as the medical inflation affects the losses per head over time. Including this dependency in our model is straightforward. To ease the notation and the computations, we omit this dependency here.

Finally we can calculate the reserve res_x of an initially x-year old person by subtracting the remaining sum of expected losses per head from the remaining discounted expected (brutto)premium payments. More formally, we receive the following definition.

Definition 6.15. The *reserve* res_x of an initially x-year old person now aged l is given by

$$res_x(l) = bp(x) \cdot ps(l) - \sum_{k=l}^{\omega} \left(\frac{1}{1+r}\right)^{k-x} \cdot \pi_k^0,$$

where l is an integer greater or equal x. Here, the value π_k^0 denotes the expected medical expenses of a k-year old agent.

Note that the value of the reserve depends on the initial age as well, because the expected costs rely on the survival probabilities. By construction, the reserve has a value of zero at the end of the lifespan, i.e. at terminal age plus one $\omega + 1$. In scenario (C) with free contract design, i.e. a market, where premium π_k and expected costs π_0^k are equal, also the reserve at the beginning of the contract is zero. This can be easily seen by calculating

$$bp(x) \cdot ps(x) = \sum_{k=x}^{\omega} \left(\frac{1}{1+r}\right)^{k-x} \pi_k$$
$$= \sum_{k=x}^{\omega} \left(\frac{1}{1+r}\right)^{k-x} \pi_k^0$$

which implies $res_x(x) = 0$. As we can see in Examples 6.16 and 6.17 below, this is not true for the monopolistic case or unisex premiums.

Example 6.16. We continue the investigation about the scenario from Example 6.6. As before, we consider an ambulant health care with "Beihilfe" in a competitive scenario without deductible, where the losses per head are coming from the BaFin data from 2020. Different to Section 6.4, we use the model that does not depend on the standard deviation/distribution together with a risk aversion parameter $\rho = 3$. The fraction of males in the portfolio is given by $w_m = 50\%$. Including the distributions of the risks does not change the results qualitatively. We analyze the reserves of male customers with different initial ages, in Figure 46 in a competitive market, in Figure 47 in a monopolistic one. The reserves for the females including pregnancy costs are given in Figure 50.

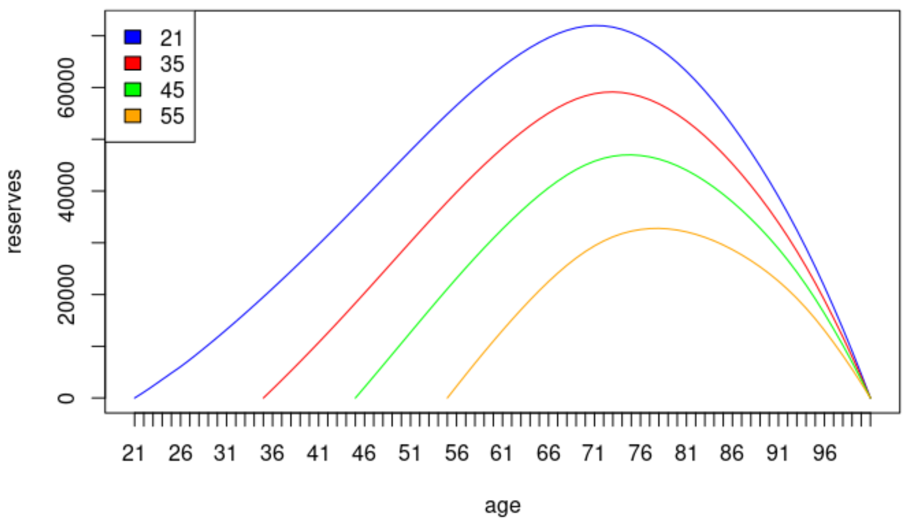


Figure 46: Male reserves as functions of the age with different initial ages in scenario (C)

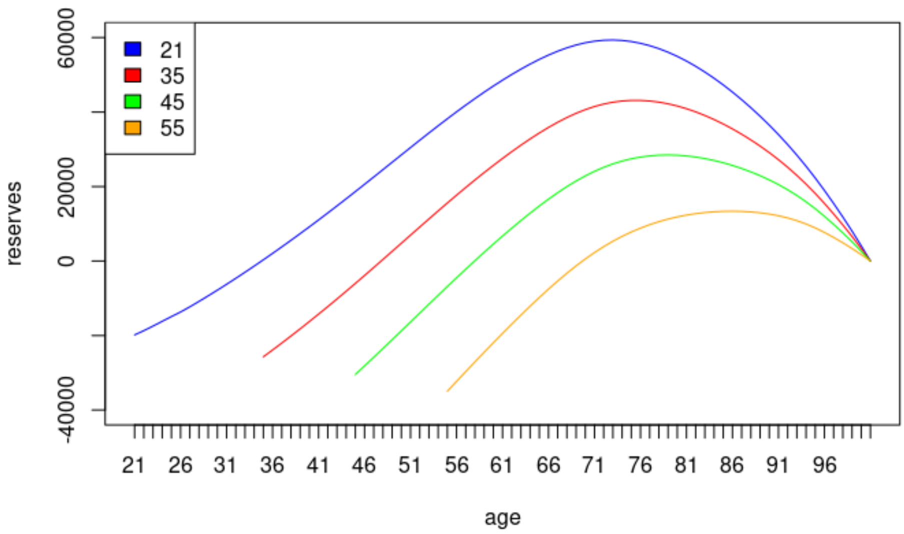


Figure 47: Male reserves as functions of the age with different initial ages in scenario (M)

The lower the initial age is, the higher the reserves become. This is not surprising, as the expected medical costs highly depend on the age of the customer and becomes larger, the older the customer becomes. Hence, a lower initial age results in a longer time, where the expected losses lie below the bruttopremium and therefore let the reserve grow. After the time point, where the expected losses exceed the bruttopremium, the reserve starts to shrink. As mentioned above, it has to be zero at terminal time by construction.
As the premiums in the monopolistic case lie clearly above the expected losses, our reserves become negative in young ages. Different to other cases, for example if the losses are falling for higher ages, this is no issue for the insurer, as this results from charging very high premiums. Later, we discuss how one could evolve the model to avoid negative reserves if we regard the unisex case.

The final problem which we need to deal with in this chapter is how to transform the concept of reserves to our unisex scenario. The naive idea would be, to mix all the gender-related input variables by a constant factor, which is equivalent to the fraction w_m of males in the portfolio. Note that we assume that our portfolio consists of 50% males and females, respectively.

Example 6.17. In the setting of the last example, we redeem the reserves in a competitive market including pregnancy costs, shown below in Figure 48.

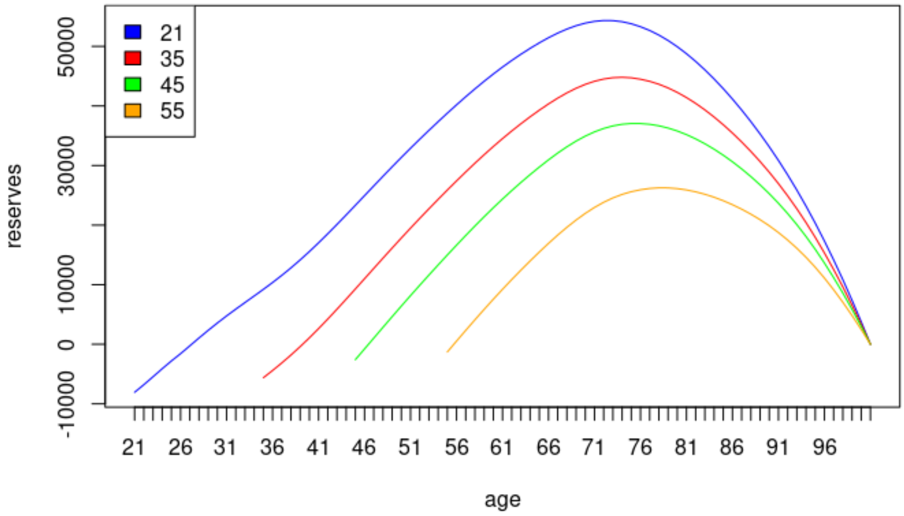


Figure 48: Unisex reserves as functions of the age with different initial ages and constant mixing in scenario (C)

As we can see, it does not make sense to mix the gender-related parameters with a constant mixing parameter, as it leads to negative reserves, even in the competitive market. The more natural approach is to mix the parameters according to a dynamic mixing parameter, which relies on the age dependent fraction of males in the portfolio. This mixing dynamic parameter ξ_n at the beginning of year n is given by

$$\xi_n = \frac{\eta_n^m}{\eta_n^m + \eta_n^f},$$

where η_n^m and η_n^f describe the number of males and females in the portfolio at the beginning of year n. Of course, these two numbers depend on the amount of customers dying or canceling their contracts but in practice also on the number of newly arriving customers. We assume that no new customers arrive once the contract has started, so the number of males and females in the portfolio can be calculated by (rounding)

$$\eta_n^i = \eta_0^i \cdot \prod_{i=1}^{n} \hat{p}_k^i$$

for $i = m, f$ Here, $\hat{p}_k^i$ are the probabilities that a customer of gender i neither cancels, nor dies within year k, compare Definition 6.12.
For more details about using mixing parameters to mix (gender-specific life insurance) premiums we refer to Section 3.3. The fraction of males in a

life insurance portfolio where the lapse rates are zero was already given in Figure 21.

Example 6.18. In Figure 49 we can see that our new approach indeed prevents negative reserves. For more discussions regarding mixing parameters, we refer to Section 3.3.

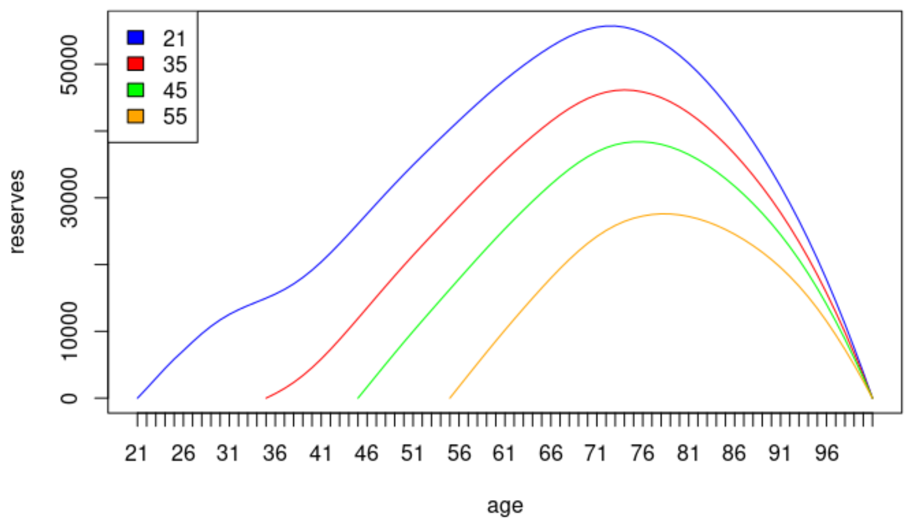


Figure 49: Unisex reserves as functions of the age with different initial ages and dynamic mixing in scenario (C)

The flattening of the curve for ages in the upper 20s and the 30s can be explained by the higher fraction of females in these ages. As we have seen in Example 6.6, health insurance is of special interest for females in these ages, the market consists almost solely of them. Hence an effect which is observable for the female reserves can also be observed here: Due to rising costs related to pregnancies, the growth of the reserves slows down. At an age of around 40, where pregnancy costs play almost no role anymore, the growth of the reserve accelerates again. In these older ages, the fraction of males in the portfolio rises again. This also explains why this change of growth cannot be observed for the reserves with an initial age of 45 and 55. In Figure 50, the reserves for females including pregnancy costs are plotted. The effect of the pregnancy costs can also be observed quite clearly in this plot.

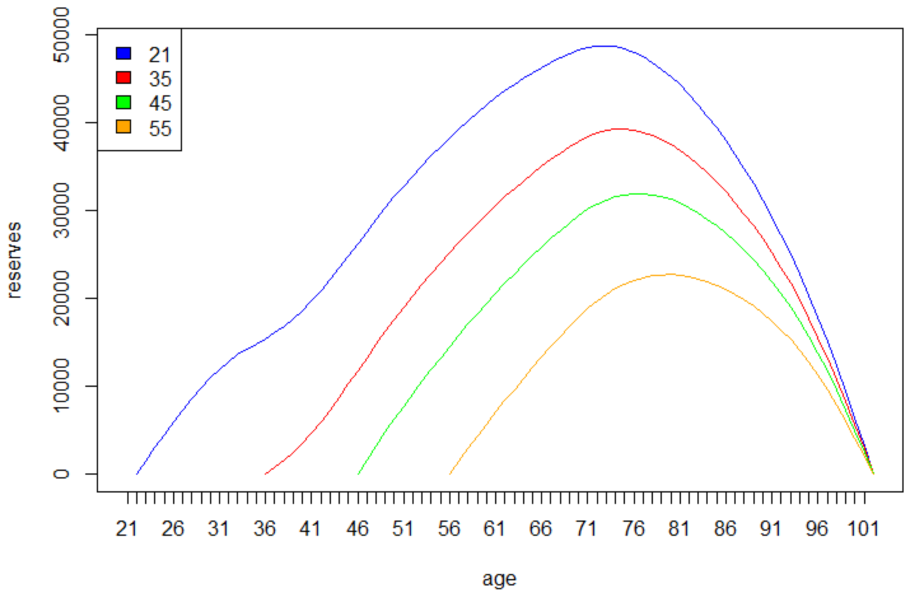


Figure 50: Female reserves as functions of the age with different initial ages in scenario (C)

7 Disability Insurance

In this chapter we model disability insurance products. These types of policies pay out some money if the insured person falls seriously ill or becomes disabled during his working life. As the disability insurance market is usually a market with price competition, we can directly make use of the model of Chapter 2. Of course, we need to replace the mortality model by an appropriate disability model. The U.S. Social Security Administration has published actuarial notes, including disability and mortality tables for insured workers. More details are presented in Section 7.1.
We continue by analyzing different insurance contracts in Section 7.2. The products which we are applying our model to are a disability annuity, contracts which pay out a lump sum in case of disability and a combination of the two contract types.
Different to classic life insurance products, we deal with four states when it comes to disability insurances. More details about actuarial models and methods in disability insurance can be found in Haberman and Pitacco (1999) [HP99]. The four possible states in our model are:

- (a) active: agents which are alive, healthy and have never been disabled,
- (r) recovered: agents which recovered from a disability and are (alive and) active again,
- (i) disabled: agents which are alive but disabled,
- (d) dead: agents which are dead.

Active and recovered agents can become disabled (again) or die with certain probabilities. Disabled agents can recover from their disability, putting them in the recovered state to mark that they have been disabled in the past. Of course, disabled agents can die as well, we assume furthermore that the customers cannot rise from death.
Sometimes, the active and recovered agents, i.e. the agents that are not disabled and alive are subsumed in a non-disabled status (nd). For computations it is not only relevant which status an agent is currently in, but also his history, see the formulas in Section 7.2. The real-life data source we are using distinguishes between active and recovered agents. Especially when we compare the probabilities of becoming disabled (again), agents with a case of disability in their history are more likely to fall into disability again. Hence, we need to make sure that we are carefully distinguishing between these cases when it comes to computing, compare Remark 7.3.
The possible transitions between the different status are illustrated in Figure 51 below.

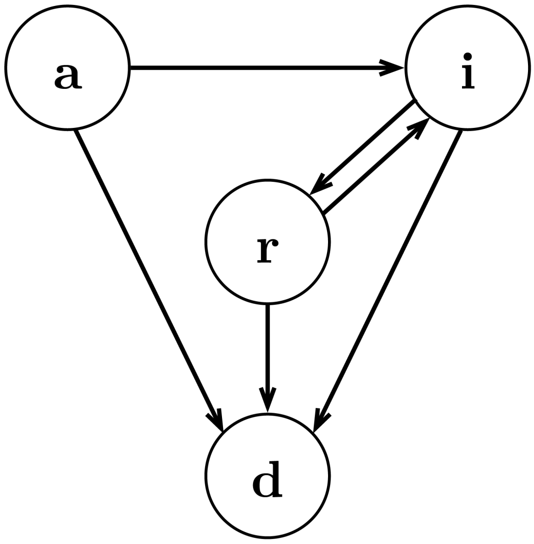


Figure 51: Possible transitions between the four possible states of an agent

7.1 Data Basis and Adjustments of the Model

Before continuing, we present details how the payment of the disability insurance can come to an end. Complementary to the immediate lifetime annuities, which we have observed Chapter 3, e.g. in Example 3.5, there are three options how the payment of a disability insurance can terminate:

a) The agent can leave the system by dying. This is the only way the payment of an immediate lifetime annuity introduced in Definition 3.5 can end. As we are considering a disability insurance, the part of our customers which are still alive can be partitioned in three parts. One part consist of the agents which have never been disabled (state (a)), a second one contains the agents that are recovered (state (r)). The third part is formed by those customers which are disabled (state (i)). As the disability status highly affects the survival probabilities, we need to distinguish between disabled and non-disabled people when it comes to survival and mortality rates, see also Remark 7.3.

b) Different to pension products like the immediate lifetime annuity, the payment does not continue until the end of the life (modeled by a maximum age ω, in our model $\omega = 110$). Contrariwise, it ends when an agent reaches a predetermined retirement age. In Germany, most contracts end when the customer has reached the age of 67. At this age, he is eligible to receive payments from the federal pension fund.

c) Some disabilities are not permanent. Once the disability is gone and the insured is able to work again, also the payment of the disability annuity ends. This is called a *recovery*, the corresponding probabilities are called *recovering probabilities* r_k, respectively ${}_l r_k$ if we look at the cumulative probabilities over l years, see Example 7.2. When a customer gets recovered, he is set back to the non-disabled but recovered status, from where he can become disabled again, retire or die. Note that one could model the probabilities not only age dependent, but also depending on a parameter that models how often the insurance company checks whether the insured is still disabled. The more money an insurance company spends for trying to detect insurance fraud by customers that do not report their recovery, the higher the recovering probabilities becomes. Vice versa, spending no money on investigating lowers the detection cost but therefore also the recovering probability, which leads to higher annuity payments. We leave this task of finding the optimal checking procedure as an open problem for future research.

Considering these thoughts, our basic model does not need to be changed too much. We just need to carefully choose the damage probabilities and heights and apply our basic model from Chapter 2 to it, see Examples 7.2, 7.6 and 7.8.

As mentioned before, our analysis is based on data from the social security administration of the U.S. government[11]. Each year, a death and disability life table for insured workers born 20 years before the current year is published by the office of chief actuary in Baltimore, Maryland. The publications are PDF files containing five (before 2012) or six (since 2012) pages each. In these pages, an introduction and an analysis of the assumptions and methods used is presented. From the four (before 2014) to five (since 2014) tables (named A, B, C, D and E) we are using tables C[12] and D[13]. The tables mentioned above consist of four sections:

a) In the columns “Living at Beginning of Year” one can find the numbers, how many people of a fictive population of 1,000,000 individuals which were active and alive at the age of 20 are alive, active, got disabled or recovered.

b) Under the heading “Deaths”, the number of deaths is analyzed, in total as well as disaggregated to active, disabled and recovered people.

c) Under the heading “Newly Disabled”, the number of newly disabled persons is summarized, again as a total number and split up in active and recovered persons.

[11] `https://www.ssa.gov/oact/NOTES/ran6/`, visited August 2021

[12] “Illustrations of Survival and Disability Status for Insured Males Attaining Age 20”; name has changed over time; before 2014 denoted as Table B

[13] “Illustrations of Survival and Disability Status for Insured Females Attaining Age 20”; name has changed over time; before 2014 denoted as Table C

d) Finally, the column “Newly Recovered” counts the number of newly recovered persons. As people can only recovered from a disability status and naturally not from the active or dead status, no further disaggregation is needed.

Unfortunately, at the time of working the tables were only available for a limited amount of years. We are using the data as of August 2021, i.e. data for the birth years 1985 and 1991 till 2000. To deal with this lack of data, we regress the given data by using ensemble learning and train a set of 20 independent neural networks for males and females, respectively. We are regressing nine different numbers: The number of active, disabled and recovered persons living at the beginning of a year, the newly died active, disabled and recovered customers, the newly disabled active and recovered agents and the newly recovered agents.
The networks have the following specifications, which seem to work good for our setting:

- The network has three layers with 150, 100 and 30 neurons and nine output neurons.
- We make use of the mean-squared error for the loss and the elu activation function.
- The optimizer is chosen to be the Nadam optimizer with a learning rate of $3 \cdot 10^{-4}$.
- The test set is supposed to contain 10% of the original 7248 points in our data set, while the validation set contains 10% of the remaining points.

The final regression is then performed by taking the mean of the regressions done by the neural networks, see Chapter 4 for more applications of neural networks. Again, it seems like the effect of the network specifications is not very large if one is not using extreme specifications like a very high learning rate. The alternative regression methods we tried in Section 4.2 are also not very promising in this setting. There is some research going on in the area of forecasting mortality probabilities using neural networks. We refer to Hainaut (2018) [Hai18] or other (less mathematical) papers from clinical research like Simpson *et al.* (2015) [SLC+15] or Lee *et al.* (2018) [LHG+18].

7.2 Numerical Examples

After these theoretical considerations we are interested how our model works in practice. Therefore we are going to investigate three insurance products. We start by analyzing the disability annuity in Section 7.2.1, a product that ensures a constant payment stream to substitute income loss due to disability. We continue by investigating a lump sum insurance in Section 7.2.2.

This type of insurance pays out one unit of money when an agent becomes disabled, so the medical and private expenses like remodeling the house to meet the needs of a disabled person can be made. Finally, in Section 7.2.3, we are looking at full disability insurance, a combination of the two above products.

Remark 7.1. As in the last chapter, Chapter 6 which was about health insurance, all computations are done in R and all costs are omitted. Again, the number of data points we are calculating for each line in a figure is determined by the number of initial ages or birth years we are plotting over.

7.2.1 Disability Annuities

First, we take a look at disability annuities. A disability annuity pays out one unit of money at the beginning of each year, starting the year after the disability occurred. The contract can only be signed by those customers, who are active, not by disabled customers. The payment ends either when the agent is retired (in Germany at the age of 67), or if the agent dies or recovers before getting retired, compare Section 7.1. As for the life insurance products in Chapter 3, setting the payout to one unit of money is only done for the sake of creating an arbitrary reference point and has no further meaning.

Example 7.2. To ease the quite complicated notation which is upcoming next, we do not denote the gender of the agent in the formulas. Nonetheless, one actually needs two versions, one for males and one for females. We calculate all probabilities for a person that is x years old at the beginning of the contract. As it is common practice in many insurance companies, we assume that it is only possible to sign an insurance contract when one has reached a certain age, say 20 for our model. This age is chosen as it fits the minimal age of our data set.
Following the notation of Definition 2.4, the risk variables Z are given by

$$\mathbb{P}(Z = z^k) = {}^{nd}_{k}p_x \cdot d_{x+k} \cdot \mathbb{1}_{\{k+x \leq 67\}},$$

where $k = 0, 1, \ldots$. In the formula above, ${}^{nd}_{k}p_x$ denotes the k-year survival probability of a non-disabled x-year old agent. We can calculate this probability as

$$ {}^{nd}_{k}p_x = \frac{{}^{a}_{k}l_x}{{}^{nd}_{k}l_x} \cdot {}^{a}_{k}p_x + \frac{{}^{r}_{k}l_x}{{}^{nd}_{k}l_x} \cdot {}^{r}_{k}p_x. \tag{20}$$

Here, ${}^{r}_{k}p_x$ and ${}^{a}_{k}p_x$ denote the according survival probabilities of recovered and active customers. With $\frac{{}^{r}_{k}l_x}{{}^{nd}_{k}l_x}$ and $\frac{{}^{a}_{k}l_x}{{}^{nd}_{k}l_x}$ we describe the fractions of recovered and active non-disabled customers. As a non-disabled person can be either recovered or active, these fractions add up to one. Customers can only switch from the active to the recovered status but not vice versa, so

$\frac{{}^r_k l_x}{{}^{nd}_k l_x}$ increases in k. Finally, by d_{x+k} we describe the probability that an $x+k$-year old customer is falling into disability before the age of $x+k+1$. Actually one needs to mix this probability from the recovered and active probabilities as in Equation (20):

$$d_{x+k} = \frac{{}^a_k l_x}{{}^{nd}_k l_x} \cdot {}^a d_{x+k} + \frac{{}^r_k l_x}{{}^{nd}_k l_x} \cdot {}^r d_{x+k}. \tag{21}$$

Here, ${}^a d_{x+k}$ and ${}^r d_{x+k}$ denote the probabilities of an $x+k$-year old active or recovered agent falls into disability, respectively. The damage heights z^k can be calculated by

$$z^k = da(x,k).$$

With $da(x,k)$ we denote the value of a disability annuity for an initially x-year old person that falls into disability at age $x+k$. For $k = 0, 1, \dots, 67-x$, $da(x,k)$ can be calculated by

$$da(x,k) = \sum_{l=k+1}^{67-x} B(0,l)({}_{l-k}^{\;\;i}p_{x+k} - {}_{l-k}r_{x+k}).$$

In the formula above, $B(0,l)$ denotes the discount factor from age 0 to l, see Section 3.1.2 for more details. Note that if the interest rate is not assumed to be constant but time dependent, the discount factor depends on the exact year in which the agent falls into disability. To ease the notation, we do not regard this dependence here.

By ${}^i_l p_{x+k}$ the l-year survival probability of an inactive (disabled) $x+k$-year old person is given. In other words, this term describes the probability that a person with initial age $x+k$ survives until he is $x+k+l$ years old. Note that the index shift from ${}^i_l p_{x+k}$ to ${}_{l-k}^{\;\;i}p_{x+k}$ is needed as the sum starts at $k+1$. Finally, as the agent could recover, ${}_l r_{x+k}$ represents the probability that a person which got disabled at the age of $x+k$ recovers within the next l years. To be able to recover, the person especially needs to survive, so ${}_{l-k}^{\;\;i}p_{x+k} - {}_{l-k}r_{x+k}$ is always non-negative.

Remark 7.3. In Example 7.2 we distinguished between active, recovered and disabled survival probabilities. There is indeed a crucial difference between the active, recovered and disabled persons. In most cases, the disabled death probabilities are five to ten times higher than for the active ones. For example the male one-year death probability for a 50-year old active agent is about 0.25%, while it is approximately 0.4% for the recovered and about 2% for the disabled agents.

In Figures 52, 56 and 58 and 54, 57 and 59, the premiums are plotted for different birth years, namely 1997, 1950 and 2040, in scenarios (C) and (M), respectively. Figure 60 shows the premiums for customers with initial age 40 for different birth years in scenario (C). In Figures 53 and 55 the coverages for the customers born in 1997 are given for scenarios (C) and (M). As for the health insurance, the observed phenomena for the coverages do not seem to depend too much on the underlying product and the birth year, so we set our focus on analyzing the premiums in the following.

Remark 7.4. The premiums are calculated based on U.S. data, see Section 7.1 for more details. Unless otherwise stated, we assume that the agents are equipped with a constant interest rate of $r = 3\%$ and CARA utility (exponential utility) with a risk aversion parameter of $\rho = 0.03$. As mentioned in Section 2.1, it is difficult to compare the risk aversion parameters of a CARA and a CRRA utility function. By choosing $\rho = 3$ as for the CRRA utility function when we investigated life insurance products in Chapter 3, the monopolistic premiums are about as high as the worst-case loss. Recall that the initial wealth is not relevant, as the equilibrium insurance demand is independent of the initial wealth under CARA utility, compare Remark 2.36.

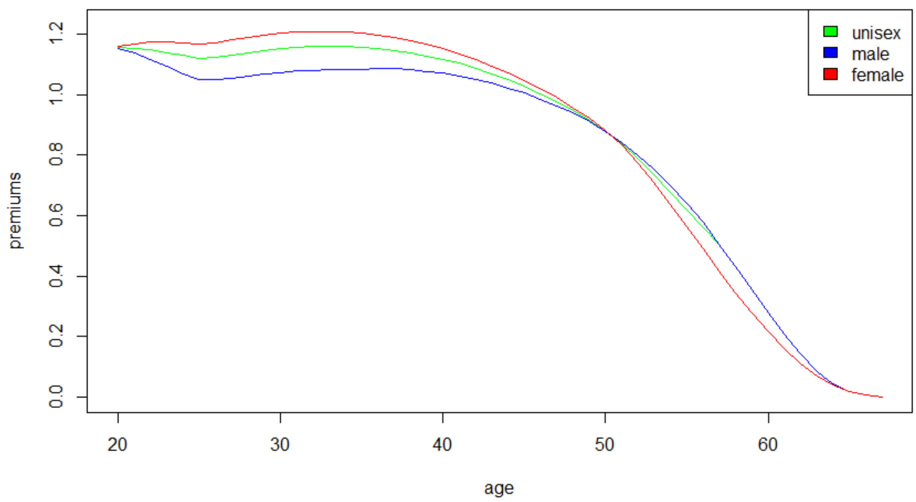


Figure 52: Premiums of a disability annuity for customers born in 1997 as functions of the initial age in scenario (C)

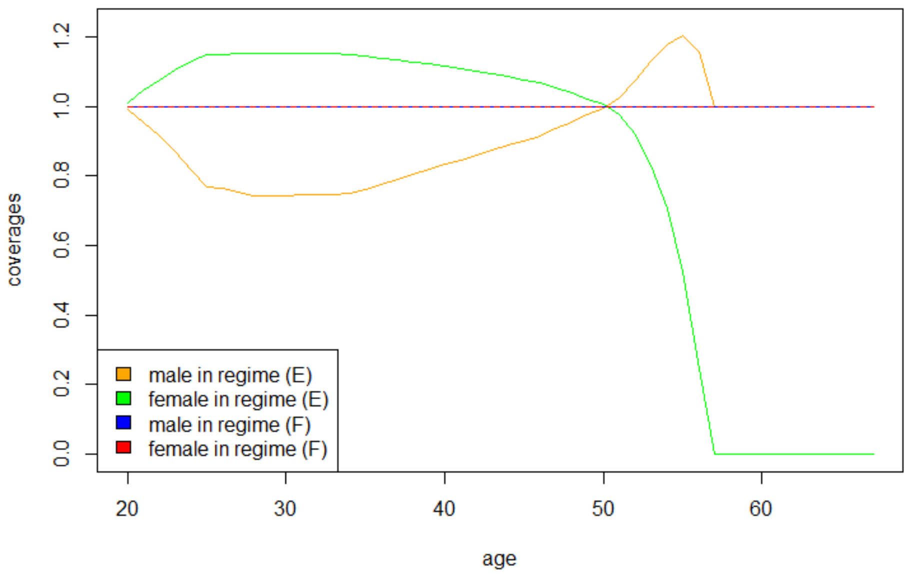


Figure 53: Equilibrium insurance coverages of a disability annuity for customers born in 1997 as functions of the initial age in scenario (C)

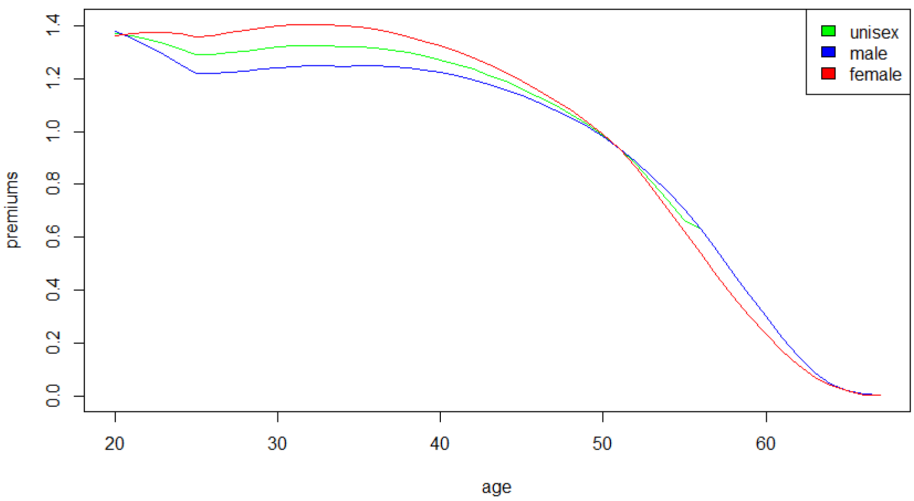


Figure 54: Premiums of a disability annuity for customers born in 1997 as functions of the initial age in scenario (M)

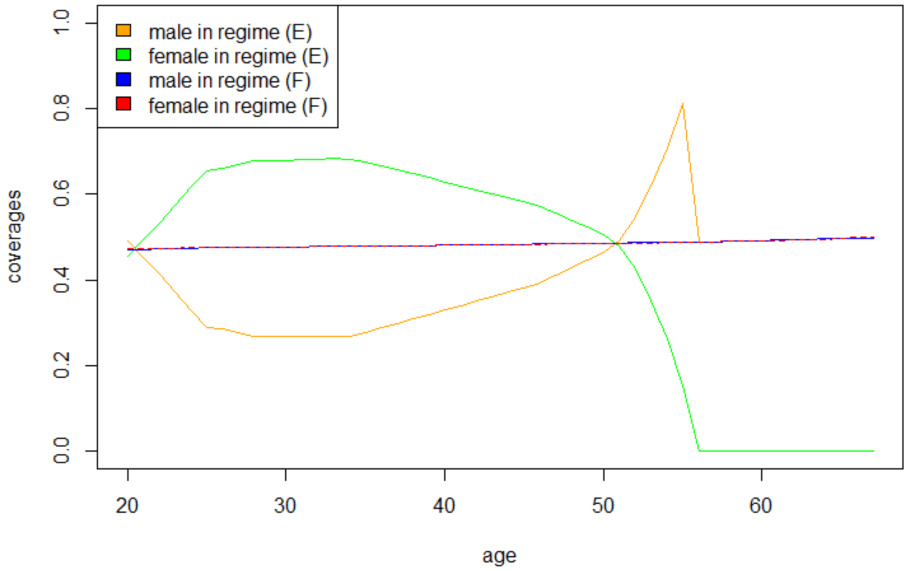


Figure 55: Equilibrium insurance coverages of a disability annuity for customers born in 1997 as functions of the initial age in scenario (M)

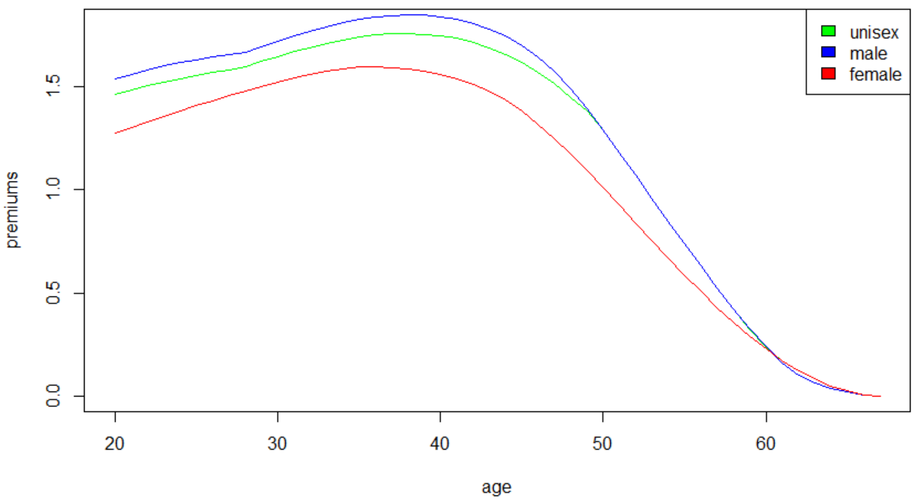


Figure 56: Premiums of a disability annuity for customers born in 1950 as functions of the initial age in scenario (C)

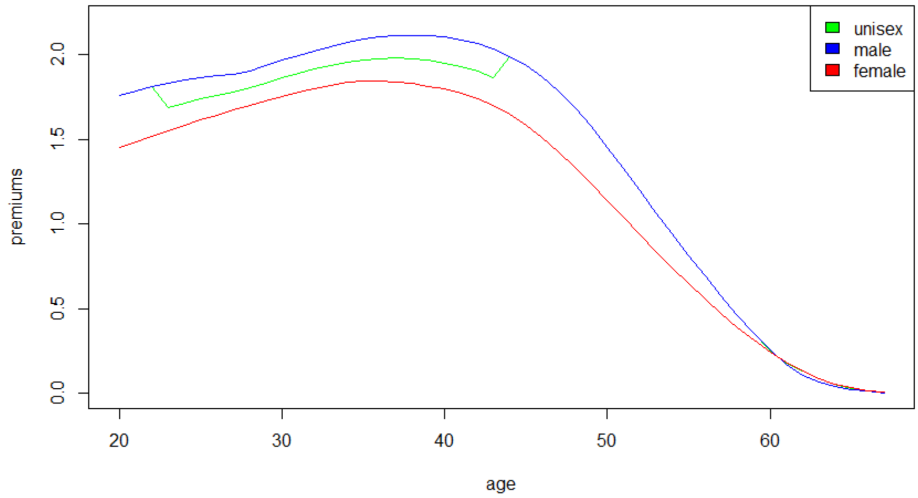


Figure 57: Premiums of a disability annuity for customers born in 1950 as functions of the initial age in scenario (M)

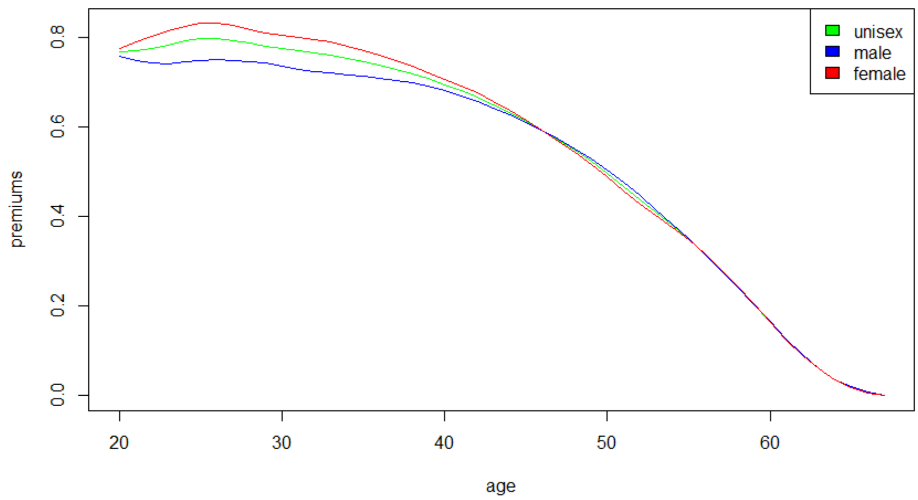


Figure 58: Premiums of a disability annuity for customers born in 2040 as functions of the initial age in scenario (C)

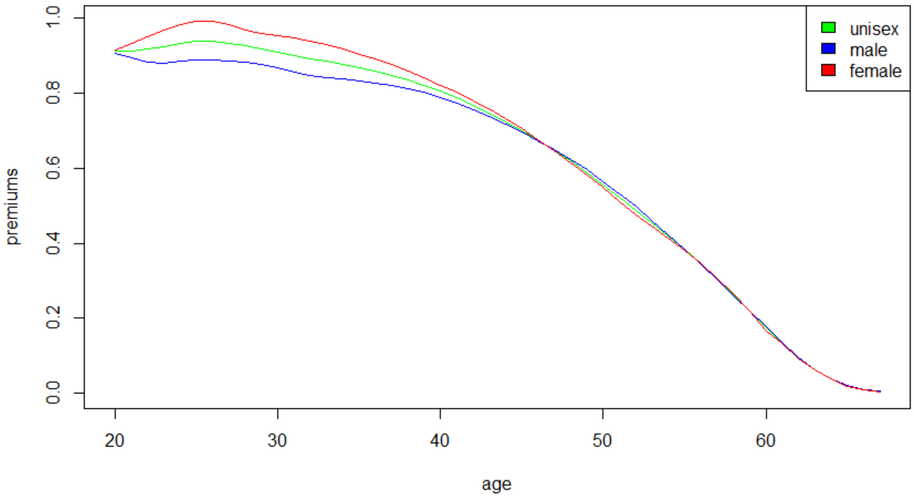


Figure 59: Premiums of a disability annuity for customers born in 2040 as functions of the initial age in scenario (M)

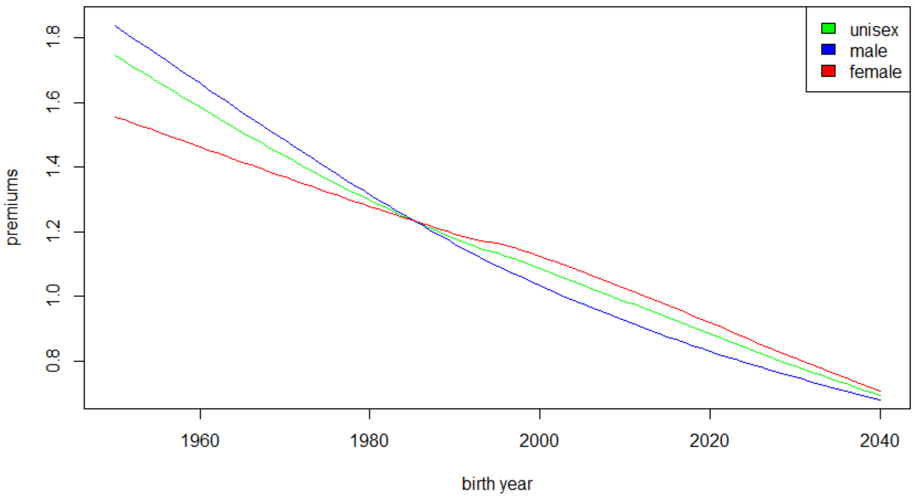


Figure 60: Premiums of a disability annuity for customers with initial age 40 as functions of the birth year in scenario (C)

Remark 7.5. In the nine figures above, we can observe the following phenomena:

- As we can see, the probability of falling into disability and therewith the price of insurance is getting lower, the later the customers are born, see Figure 60.

- Interestingly, the premiums are not decreasing monotonously with rising age, especially for early birth years, see e.g. Figure 52. The slight increase for younger ages can be explained by the fact that the decrease of the contract duration also results in a shorter time in which the premium payment can be invested. Furthermore, the probabilities of becoming disabled in young ages is quite small. Be reminded that a customer needs to be active at the beginning of the contract. This effect weakens over time, compare e.g. the customers born in 1950 (Figure 56) with those born in 2040 (Figure 58).

- The annuity ends when the customer retires, therefore the insurance company will not make any payments after the customer turns 67. Because of this, the premiums are tending to zero when the age of the customer approaches the retirement age of 67, see Figure 52 as an example.

- For customers born in 1950, males pay more than females for almost all ages. In both scenarios this difference is so high that the females are pushed out of the market for many initial ages, see Figures 56 and 57. The push-out effect in scenario (M) is stronger than in scenario (C). This is in line with our results from Chapters 2 and 3, where we already could observe the effect. Also for customers born in 1997 a push-out can be observed but occurs at higher ages compared to the customers in 1950. The push-out gets particularly visible when observing the coverages, see Figures 53 and 55.

- While males paid more for almost all initial ages if they were born in 1950, this is no longer the case for later birth years. For example for customers born in 1997 it is more expensive for females up to an age of 53, afterwards males become more expensive again. Interestingly, the premium difference between males and females vanishes over the recent and upcoming birth years. Given our interpolated data, the premiums are going to be almost the same for many initial ages in the future, see Figures 58 and 59.

7.2.2 Lump Sum Payments

It is of interest to provide an option to include lump sum payments in our disability insurance contracts. If the insured has bought such a contract, the insurance company pays out a predefined amount of money in case of disability.

Example 7.6. The risk variables Z for the *lump sum insurance* are quite similar to the variables of the disability annuity from Example 7.2. The two products only differ in the damage heights but not in the corresponding probabilities. As before, these probabilities are given by

$$\mathbb{P}(Z = z^k) = {}_k^{nd}p_x \cdot d_{x+k} \cdot \mathbb{1}_{\{x+k \leq 67\}},$$

where $k = 0, 1, \ldots$. Again, ${}_k^{nd}p_x$ denotes the (mixed) k-year survival probability of a non-disabled x-year old agent, see Equation (20). Furthermore, by d_{x+k}, we describe the (mixed) probability that an $x+k$-year old customer falling into disability before the age of $x + k + 1$, see Equation (21). The damage heights z^k for the lump sum insurance can be calculated by

$$z^k = B(0, k + 1),$$

the discount factor from year 0 to $k + 1$.

Remark 7.7. We assume in our model that agents can recover and become disabled again. In this case, the lump sum is going to be paid again. The agents do not need to pay the lump sum back in case they recover.

The effects observed for the lump sum insurance are quite similar to those observed for the disability annuity in Section 7.2.1, see Remark 7.5. Therefore, we focus on comparing different risk aversion parameters for the unisex premium in a competitive scenario. When there is perfect competition, the gender-specific premiums are just given by the expectation of the corresponding risk variable. One can see in Figure 61 that the unisex premium for the lump sum insurance highly depends on the value of the risk aversion parameter ρ. It should not be surprising that the higher the risk aversion parameter is, i.e. the more risk averse a person is, the closer the unisex premium lies to the average of the gender-specific premiums. Lower risk aversion parameters such as $\rho = 0.2$ or $\rho = 0.1$ lead to push-outs, in our example at ages 57 and 51.
For comparison, some calculations for the monopolistic case is given in Figure 62. The risk aversion parameter for the gender-specific premiums as well as for the unisex premium is given by $\rho = 0.5$.

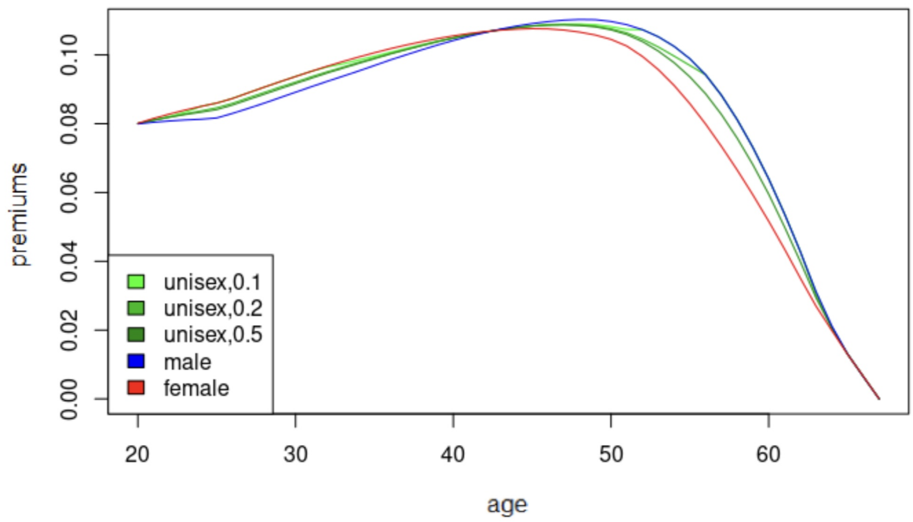


Figure 61: Premiums of a lump sum insurance for customers born in 1997 as functions of the initial age in scenario (C) for different risk aversion parameters ρ for the unisex premium

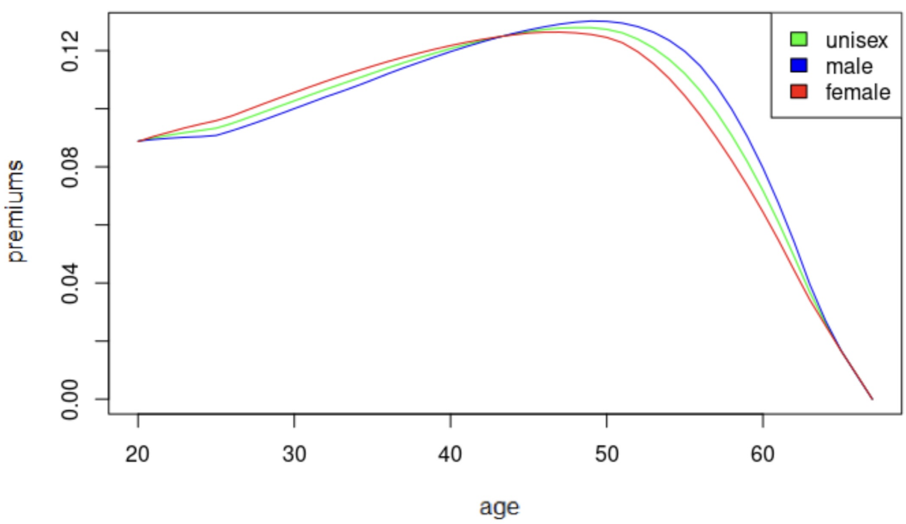


Figure 62: Premiums of a lump sum insurance for customers born in 1997 as functions of the initial age in scenario (M)

7.2.3 Full Disability Insurance

Finally, we take a look at full disability insurances, i.e. insurances, where agents receive periodical payments beginning in the year after disability and in addition a single lump sum payment in the beginning of the year after disability occurred. The lump sum payment is fixed to ten units of money, while the annuity payment is still one amount of money per year in which the agent is disabled at the beginning.
As in Section 7.2.1 we suppose that the annuity can end with the death or recovery of the agent, but always ends with the retirement of the insured. Similar to Section 7.2.2, a customer that has recovered and falls into disability again is eligible to receive a second lump sum payment. Additionally, the first payment does not need to be returned in case of a recovery of the agent.

Example 7.8. Let us now investigate the full disability insurance. We have seen in Examples 7.2 and 7.6 that the damage probabilities of the disability annuity and a lump sum insurance are given by

$$\mathbb{P}(Z = z^k) = {}^{nd}_{k}p_x \cdot d_{x+k} \cdot \mathbb{1}_{\{k+x \leq 67\}},$$

Furthermore, the damage heights can be calculated as

$$z^k = 10 \cdot B(0, k+1) + da(x, k),$$

since it is given by combining a disability annuity with a lump sum insurance that pays out ten units of money in the end of the year, in which the customer becomes disabled. More details can be found in Examples 7.2 and 7.6.

Figures 63 and 64 show numerical values for the premiums in a competitive and monopolistic scenario. Similar to the disability annuities in Section 7.2.1, the risk aversion parameter is set to $\rho = 0.03$.

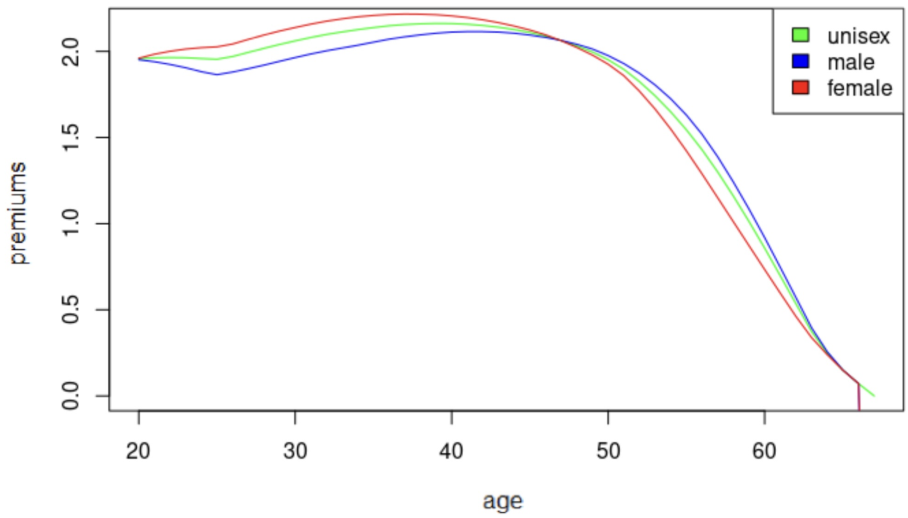


Figure 63: Premiums of a full disability insurance for customers born in 1997 as functions of the initial age in scenario (C)

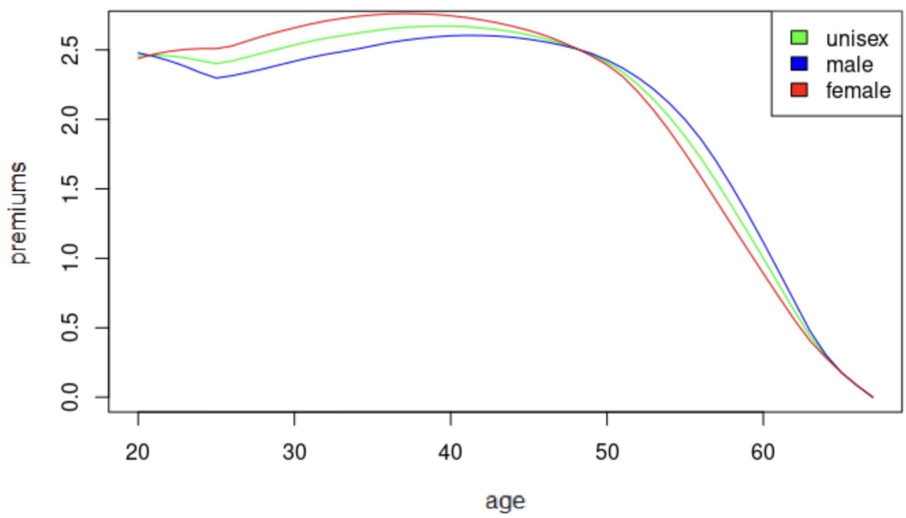


Figure 64: Premiums of a full disability insurance for customers born in 1997 as functions of the initial age in scenario (M)

8 Analysis of the Push-Out Effect

When analyzing insurance products in our model, we were able to observe the push-out effect, a phenomenon where customers of one type/gender were pushed out of the insurance market. The reason for the appearance of push-outs is that insurance is too expensive for one type of agents, so they are better off if they do not purchase insurance at all.
In [SS14] a market under price competition with two types of customers is analyzed in a theoretical setting. It is argued that there are two equilibriums in the market, one where both types of agents buy insurance and one where only the type with the higher risk decides to buy coverage. The push-out point coincides with the point where we switch from the first to the second equilibrium. This argument is also applicable for our setting. In a market with n risk classes one can observe n equilibriums, one where all customers are buying insurance, one where all customers but those with the lowest risk buy insurance and so on. As for markets with two types of customers, each push-out coincides with a point where we switch from one equilibrium to an other, so $n-1$ push-outs can be observed. In Example 8.2, a market with three risk classes, three equilibriums and two push-outs is analyzed.
We have first seen this phenomenon in theoretical examples in Chapter 2 which originate from [SS14]. But also when discussing real-life data, we could observe this effect for life insurance products in Chapter 3 as well as for disability insurance products in Chapter 7. Due to the adjustments of the model for health insurance products, push-outs are (even theoretically) not possible in these markets, see Chapter 6. Nonetheless, unisex premiums might reduce the share of customers of one gender in the insurance market to almost zero.
In this chapter we present a short analysis on the parameter-dependency on the appearance of these push-out effects and show some more numerical examples in order to illustrate our argumentation. We restrict ourselves to present examples only for some meaningful parameters. For a deep analysis for each parameter we refer to [Ohe20].

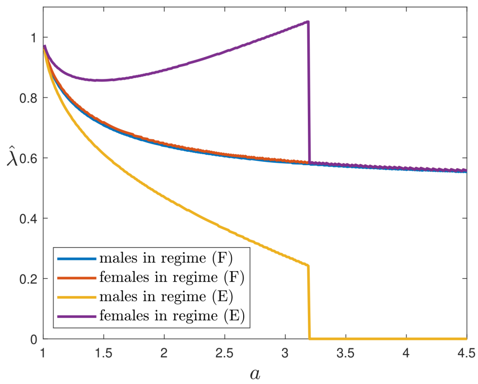
Remark 8.1. All figures in this chapter rely on an implementation in MATLAB, compare Remark 2.35. We calculated between 101 and 1000 data points for each curve.

The probability for a push-out

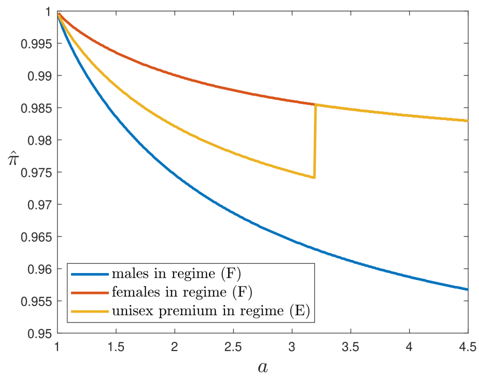
- is bigger in scenario (M) than in scenario (C): As we have obtained in various examples, market settings where the insurer tries to maximize its profits strengthen the push-out effect. This can be seen for example when comparing Figure 15 with 16 or Figure 56 with 57.
- is larger, the smaller the variance of the risk is: In Chapter 3 we have seen that the push-out effect for the endowment insurance is much

stronger than for other insurance products, see Figures 13 and 14. This is reasonable as the customers choose to buy insurance in order to swap the unpredictable risk against a predictable premium payment. If the variance of the risk is very low, this exchange is less attractive.

- is larger, the higher the initial wealth of the customers is: When the initial wealth of a customer is much higher than maximum loss that can occur, it gets less attractive for the customers to purchase insurance. If we are not using CARA utility, the equilibrium insurance coverage $\hat{\lambda}$ is a function that depends on the initial wealth, compare also Remark 2.10.
  Figure 65 illustrates the equilibrium premiums and the equilibrium insurance coverages as functions of the initial wealth a of the agents. All other parameters are set to their standard values: The insurance portfolio consists of 50% males and females born in 1950 with an initial age of 30. The contract duration is 20 years, and we use CRRA utility (power utility) with a risk aversion parameter of $\rho = 3$, i.e. $u(x) = -\frac{1}{2x^2}$. We do not charge any costs or safety loading/risk premium.



(a) Eq. insurance coverage $\hat{\lambda}$

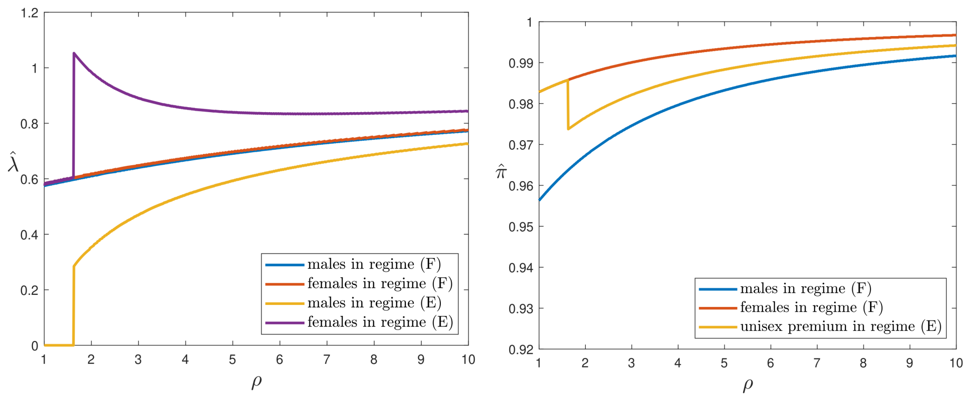


(b) Eq. premium per coverage $\hat{\pi}$

Figure 65: Equilibrium insurance coverage $\hat{\lambda}$ and equilibrium premium per coverage $\hat{\pi}$ of a pure endowment insurance as functions of the initial wealth a of the customers in scenario (M)

- is lower, the larger the risk aversion parameter ρ is: The customer's willingness to take risk is modeled by a risk aversion parameter ρ. The larger this parameter is, the more risk averse the agents are. Hence, customers that are very risk averse are willing to pay higher premiums in order to see their risks covered. In Figure 61 the unisex premiums for a disability lump sum insurance are compared for different risk aversion parameters ρ. It can be seen that the higher the risk aversion

parameter is, the later the push-out occurs.
We give a second example to illustrate the influence of the risk aversion parameter on the push-out effect. In Figure 66, the equilibrium premiums and equilibrium coverages are given as functions of the risk aversion parameter ρ. Again, all other values are set to their standard values. We assume furthermore that the initial wealth is set to $a = 2$ and that the underlying utility function is a CRRA utility function, i.e. $u(x) = \frac{1}{1-\rho}x^{1-\rho}$ for $\rho > 0$ and $u(x) = \ln(x)$ if we have $\rho = 1$.

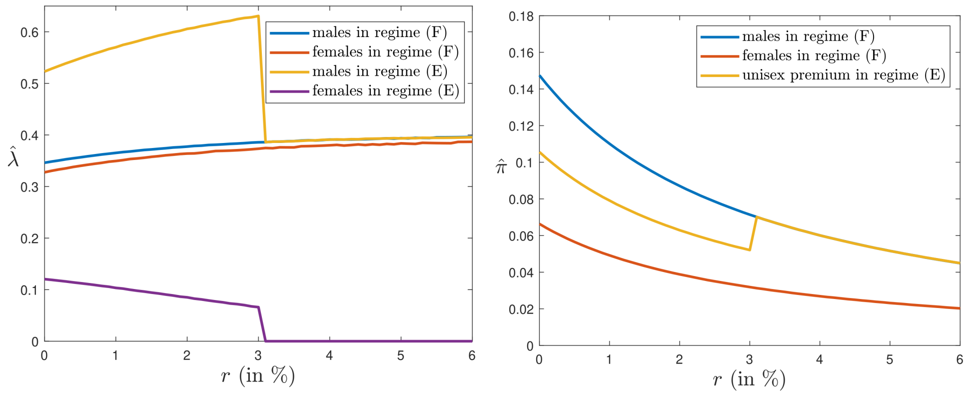


(a) Eq. insurance coverage $\hat{\lambda}$ (b) Eq. premium per coverage $\hat{\pi}$

Figure 66: Equilibrium insurance coverage $\hat{\lambda}$ and equilibrium premium per coverage $\hat{\pi}$ of a pure endowment insurance as functions of the risk aversion parameter ρ in scenario (M)

- is higher, the more the damage probabilities differ, especially regarding the relative difference: For high ages, the survival and death probabilities differ more than for lower ages, for long contract durations more than for short ones. The longevity effect causes the death probabilities to fall over the birth years, which leads to a bigger relative difference between the two genders. As we can see in Figure 12, this can result in a push-out.
- is larger, the higher the (annual) interest rate r is: We model the interest rate either to be constant or using a CIR model, see Section 3.1.2. If the interest rate is high enough, it gets more and more attractive to take the risk and invest the premium in the money market instead of paying it to the insurance company. As argued before, the damage height compared to the initial wealth of the customers might get to low, which can result in a push-out, compare Figure 65. We calculated an additional example for the term insurance in scenario (M), where we analyzed the dependency of the annual interest rate the equilibrium

premium and insurance coverage. The results are given in Figure 67. Note that for an interest rate of 0% we cannot use the CIR model but set the rate to be constant at 0%.



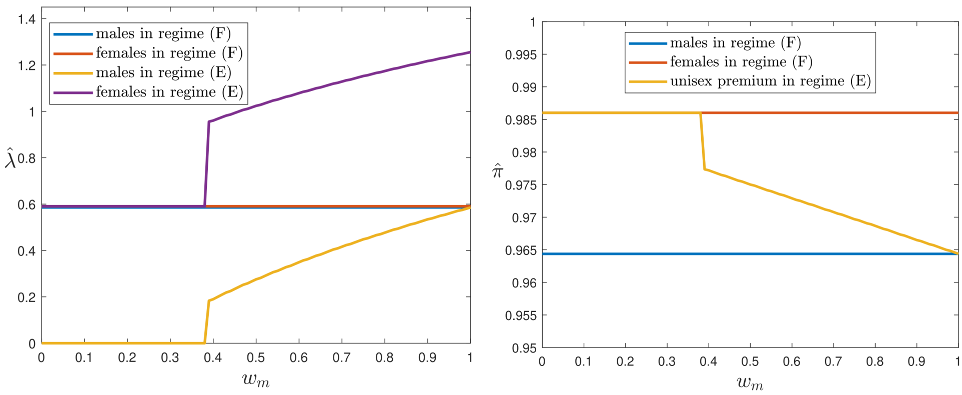
(a) Eq. insurance coverage $\hat{\lambda}$

(b) Eq. premium per coverage $\hat{\pi}$

Figure 67: Equilibrium insurance coverage $\hat{\lambda}$ and equilibrium premium per coverage $\hat{\pi}$ of a term insurance as functions of the (annual) interest rate r (in %) in scenario (M)

- is higher, the less customers there are in the portfolio (as long as we charge a safety loading in addition to the net premium): We have seen in Example 5.79 that the number of customers in the market affects the point where the push-out appears. The less customers there are, the higher is the influence of the (variance based) safety loading on the premium. When comparing scenarios (C) and (M), customers with lower risk are more sensitive to an increase of the premium. Hence, the effect is stronger, the less customers there are in the market.

- is higher, the less customers of low risk there are in the portfolio: The fewer low-risk customers there are in a portfolio, the closer the aggregate premium lies to the premium of the high-risk customers. This enhances the chance that insurance gets too expensive for the low-risk customers and therefore leave the insurance market. To illustrate this effect, we calculated an example where we vary the fraction of males w_m for the pure endowment insurance. Recall from Remark 3.6 that the males are the low risk customers when we regard the pure endowment insurance. The example is given in Figure 68. Note that we needed to set the initial wealth of the agents to $a = 3$ in order to cause a push-out.

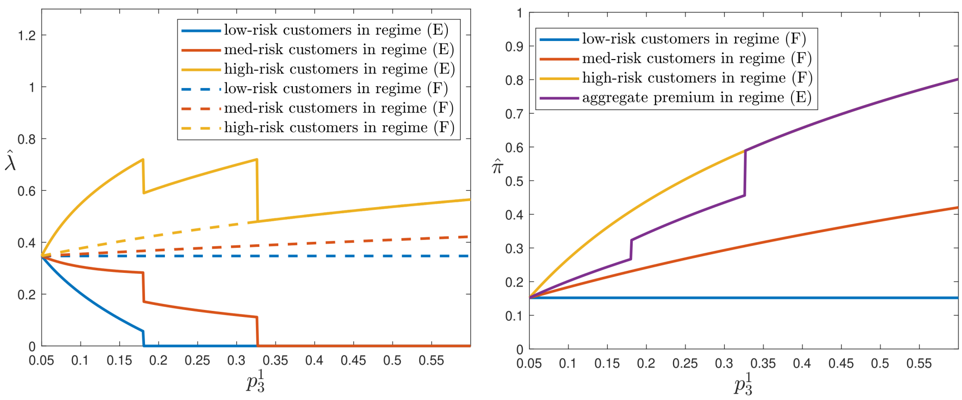


(a) Eq. insurance coverage $\hat{\lambda}$ (b) Eq. premium per coverage $\hat{\pi}$

Figure 68: Equilibrium insurance coverage $\hat{\lambda}$ and equilibrium premium per coverage $\hat{\pi}$ of a pure endowment insurance as functions of the fraction of males w_m in the portfolio in scenario (M)

To conclude this chapter, we provide an example in order to show that for markets with more than two risk classes, multiple push-outs are possible.

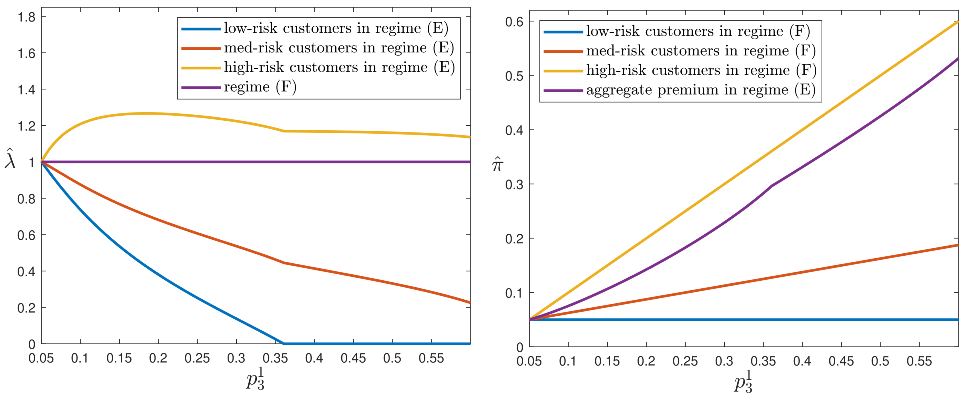
Example 8.2. We extend the setting of Example 2.29. Recall that we assumed there are agents equipped with CRRA utility with a risk aversion parameter of $\rho = 3$ and an initial wealth of $a = 2$. In contrast to the example from Chapter 2, we now have three types of agents where each risk type shows up equally likely, i.e. $w_1 = w_2 = w_3 = \frac{1}{3}$. All agents face a loss in case of damage of $z_1^1 = z_2^1 = z_3^1 = 1$. The damage probability of the low-risk customers is set to $p_1^1 = 5\%$. The damage probability of the high-risk customers p_3^1 is as before varied between 5% and 60%. For the customers of medium risk we assume that the damage probability p_2^1 rises by 1% for each 4% rise of p_3^1, in formulas $p_2^1 = \frac{p_3^1 - 5\%}{4} + 5\%$. For example, if we have $p_3^1 = 9\%$, we set $p_2^1 = 6\%$ and so on. We furthermore assume that $p_i^0 = 1 - p_i^1$ for $i = 1, 2, 3$ and that all other damage probabilities and losses in case of damage are zero.



(a) Eq. insurance coverage $\hat{\lambda}$ (b) Eq. premium per coverage $\hat{\pi}$

Figure 69: Equilibrium insurance coverage $\hat{\lambda}$ and equilibrium premium per coverage $\hat{\pi}$ as functions of the (medium-) and high-risk agents loss probability p_3^1 in scenario (M)

In scenario (M), we can see that there are two push-outs. In the first one, the low-risk customers leave the market and only the medium- and high-risk agents stay in the market. This happens at $p_3^1 = 18\%$, so the existence of medium-risk agents in the market defers the push-out of low-risk customers compared to a market with only two types of agents, compare Figure 2. At $p_3^1 = 33\%$ there is a second push-out, where also the customers of medium risk are pushed out of the market. We then end up in a market consisting of high-risk customers only.
For the monopolistic case there are three local maximums for the insurer's optimization problem in regime (E), compare Remark 2.30.



(a) Eq. insurance coverage $\hat{\lambda}$ (b) Eq. premium per coverage $\hat{\pi}$

Figure 70: Equilibrium insurance coverage $\hat{\lambda}$ and equilibrium premium per coverage $\hat{\pi}$ as functions of the (medium-) and high-risk agents loss probability p_3^1 in scenario (C)

In the competitive market, there is only one push-out for our example. As for the monopolistic market setting, the existence of medium-risk customers defers the point where the push-out happens, compare Example 2.31. Nonetheless, in a market setting where the medium-risk agents damage probability rises slow enough, we are observing a second push-out in the competitive market as well.

9 Conclusion

In this chapter we conclude the thesis by recapturing our main findings and how they relate to our research aims and questions. We discuss the value and contribution, but also the limitations of our studies and suggest future research opportunities.

Generalization of the Insurance Market Model and Observability of Push-Out Effects

We have generalized the insurance market model presented by Sass and Seifried in 2014 [SS14] in such a way that our model is able to handle arbitrary discrete risks. It is based on the computation of equilibriums in insurance markets under imperfect information which was introduced by Rothschild and Stiglitz in 1976 [RS76]. While a similar insurance market model was already presented in [Ohe20], we could develop an even more general model in this thesis. We used the Lee-Carter mortality model to deal with longevity, modeled the interest rate through the CIR model and introduced the possibility for different values of the initial wealth and utility functions among the agents. In Chapter 5 the market model was generalized further so that it is able to deal with an possibly even (uncountably) infinite amount of agent types.

After providing some theoretical foundations, our model was applied by introducing different insurance products and calculating the equilibrium premiums and equilibrium insurance coverages using real-life data. In doing so, we found that the push-out effect, already observed in theoretical examples in Chapter 2, also occurs in real-life examples.

Three major groups of insurance contracts were analyzed: In Chapter 3, we examined life insurance and annuity products. The analysis was conducted using French mortality data from the Human Mortality Database (HMD). In Chapter 6 our focus was on the analysis of health insurance products. Before doing so, we needed to provide some adjustments to the model. The adjusted model was then applied to simulated data that was based on [Rob17a, Rob17b] and real-life data that originated from the German Federal Financial Supervisory Authority (Bundesanstalt für Finanzdienstleistungsaufsicht, BaFin). Finally, Chapter 7 examined three disability insurance products. The data source for this chapter was the Social Security Administration of the U.S. government. We regressed their data in order to predict the past and future probability tables using a neural net approach.

Some of the phenomena that can be identified are evident in all types of insurance we investigated. One phenomenon that was of particular interest in our considerations is the push-out effect. To investigate its manifestations, we have undertaken a general analysis of this effect in Chapter 8. Since equilibrium premiums and coverages are calculated as numerical solutions of two connected optimization problems, it is difficult to formulate analytical criterions for the occurrence of a push-out. Nonetheless, we were able to devise indicators what increases or decreases the probability that a push-out shows up.

In Section 3.3 we calculated different types of mixing parameters and compared our approach with [CV17]. On one hand, these parameters can be used to calculate the aggregate/unisex premiums using the type-/gender-specific formulas. This drastically lowers the computation times of the aggregate/unisex premiums. On the other hand, this analysis helped us to gain a deeper understanding of the behavior of the equilibrium premiums (and coverages). For future research it might be interesting to generalize the concept of mixing parameters further and to combine it with our (neural network) regression approach from Chapter 4 or other insurance products like the health and disability contracts from Chapters 6 and 7.

In practice, one can see that push-outs show up less frequently than in our model calculations. There are two reasons that can explain this difference between the reality and our model. First, customers are not fully rational. Even if they would be economically better off if they did not purchase insurance, some customers choose to buy it to satisfy their need for security. Second, insurance companies can design policies that are specifically attractive to men or women. To do so, one would need to extend the model in a way that it is able to deal with more complex contracts, e.g. by implementing a bonus malus system, premium refunds, deductibles or others. Another option is to develop contracts that can only be bought by people who meet certain criterions, such as belonging to some occupational group. For example, by selling a contract that can only be bought by roofers, one can create a portfolio that primarily consists of males without excluding females from buying the contract.

In general, we can conclude that our equilibrium insurance market model is capable of pricing even complex insurance products. The equilibrium approach, unless precluded by market regulations, could be used as an alternative technique for pricing in insurance markets.

Regression via Machine Learning Approaches

Machine learning regression approaches such as neural networks have gained vast popularity in recent years. The exponential growth in computing power enabled larger and more sophisticated networks that can perform increasingly complex tasks.

As shown in Chapter 4, these methods can also be used to regress equilibrium premiums of life insurance products. For our considerations the phenomena for the different products and the two scenarios (M) and (C) were structurally the same, so we decided to focus on the product where the generation of data points made the lowest effort, i.e. the pure endowment

insurance in competitive markets.
The neural net approach provides us with a new finding of the insurance premiums that is quite robust, as the specifications of the network do not have a large impact on the quality of training. If the learning rate is chosen sufficiently small, say e.g. 0.001 or 0.0001, the risk of overfitting can be almost eliminated.
Furthermore, we have observed that the training quality in the unisex premium regression can be increased by adding the male and female premiums to the input. Nevertheless, we have seen that one needs to calculate several thousand data points to run a reasonable regression. In addition, regulations in most countries require insurers to calculate premiums directly rather than approximating them. These issues limit the practical use of neural networks to regress premiums.
While we have restricted ourselves to regress the insurance premiums, one could develop a new pricing approach by regressing the equilibrium insurance coverage function and perform a similar analysis like we did but with the regressed demand function instead of solving it each time. It is worth considering the approach to regress the mixing parameters from Section 3.3 and calculate the aggregate/unisex premiums based on the type-/gender-specific premiums and the regressed mixing parameters. Also the regression of other insurance types and an analysis how further input such as the death or survival probabilities or the discount factors improve the quality of training are of interest. Finally, one could try to use neural networks for detecting push-outs.
Overall, it can be concluded that neural networks seem to be a powerful tool when it comes to regressing equilibrium insurance premiums, but are limited in practical use.

Risk Class Management

The core of our work is Chapter 5, where we introduced the concepts of risk classes and risk relations to solve the problem of optimal risk class management in insurance markets. The goodness of the risk class allocation chosen by an insurer given the choice(s) of its competitor(s) is characterized by the number of customers who decide to buy insurance from that company. In order to analyze the customer count of an insurer, we needed a model to predict the customer flow in our insurance market. This is done using a Markov chain approach, first presented by Irle *et al.* in 2011 [IKLM11]. Based on this approach, we were able to introduce a preference relation to compare different allocations of customers into contracts. This allowed us to derive the concept of global and local optimizers for a set of risk class allocations. As we have seen in theoretical but also real-world examples, a global optimizer might fail to exist.
In practice we found two ways to overcome this issue. The first one is more

of theoretical nature. Instead of solving the problem in a discrete set, we transferred it to its convex hull, where we have proven various propositions about existence and location of optimums. We then defined a procedure for returning to the discrete set using a metric. This approach works well in theory but is of limited use in practice, as computing the convex hull and performing the optimization on it becomes very complex when the underlying discrete set is very large or has a high dimension. According to preliminary considerations, it is sufficient in practice to regard only 2^{n-1} contracts if there are n risk classes in the market. Therefore, the number of potential contracts only doubles for each additional risk class. Nonetheless, this approach is not really applicable in practice.

An alternative idea with broader practical application is developed in the section about capacity constrained markets, Section 5.3. There we presented an adapted version of Algorithm 1, our optimal rating class algorithm. This refined version of our original approach was presented in Algorithm 3 under the name "Counting Algorithm (under Capacity Constraints)". This already describes the core idea behind the procedure. Following the algorithm, the task of finding an optimal rating class vector is performed several times, randomly traversing the set of all possible vectors. Using a so-called frequency vector, the prevalence of each possible risk class allocation in the set of locally optimal vectors is recorded. After a sufficient number of repetitions, we choose for the vector which appeared as a local optimum the most often. What "sufficient" means must be evaluated individually for each example. As a rule of thumb, the more local optimums are found in each iteration, the higher the number of iterations must be. Of course, this procedure increases the computation time but serves as an option to increase the robustness of the method and to find a way to choose for a risk class allocation, if no global optimizer exists.

As mentioned in the introduction, the formulation and management of this problem and the development of an algorithm to solve it is motivated by the trend that the allocation of customers into risk classes became more distinguished over the years. We hence presented a new approach to a problem that dates back several decades but has gained importance in recent years. Game theoretical approaches to address this problem existed as early as the 1960s, see for example the studies of Borch from 1962 [Bor62]. Our approach developed in this thesis provides a new perspective to the problem and satisfies some desirable properties. This is particular evident when it comes to finding the optimal response to a given risk class allocation of an other insurance company. Given the structure of our model this can be formulated as a convex optimization problem, which is therefore easy to solve. We have seen that the fraction of customers buying insurance from one company converges towards a deterministic fraction. As a possible generalization, one could model the customer flow such that the limiting process is random. For example the standardized process we use to model the customer

share converges to a (random) diffusion if we use the second non-extensive and N-independent model in [IKLM11]. Models in which customers of the different risk classes do not behave independently of customers in other classes are also possible.
The goal for the insurers in our model was to attract a customer share to their company that is as large as possible. This is motivated by our study of herding theory, which serves as a theoretical foundation for why the convergence against the desired fractions makes sense. Nonetheless, other objectives are also conceivable, such as weighting the customers according to their acquired coverage or assuming that the companies want to maximize their total profit in the monopolistic/duopolistic market setting.
Taking everything together, our model has indeed given answers to the underlying research questions, namely how to compare different allocations of risk classes into rating classes and how to model the risk class management problem based on this comparison. All our reasoning got summarized in Algorithm 1 and its refinements. We have seen that an optimal risk class allocation does not need to exist and found ways to mitigate this problem.
The risk class management application of our equilibrium insurance market model forms the main part of this thesis and concludes the analysis how an equilibrium pricing model affects the behavior and the characteristics of insurance markets.

A Notation

We summarize the notation used in this thesis in the following table.

Notation	Explanation
Ch 2: Basic Model	
$u(x)$	(real-valued) utility function
ρ	risk aversion parameter of the (CARA and CRRA) utility function, always $\rho > 0$
a	initial wealth of a customer
$\oplus/\ominus$	agent type (low-risk/high-risk) from the insurers view
$w_\oplus/w_\ominus$	fraction of the low-risk/high-risk customers in the market
$Z_{\oplus/\ominus}$	risk of customer of type $\oplus/\ominus$ (discrete and independent non-negative RV)
$(\Omega, \mathcal{F}, \mathbb{P})$	probability space on which the risk variables Z are defined
$p^k_{\oplus/\ominus}$	damage probability for customers of type $\oplus/\ominus$ (in damage case k, $k \in \mathbb{N}_0$), $p^k_{\oplus/\ominus} \in [0,1]$ for all k
$z^k_{\oplus/\ominus}$	damage amount for customers of type $\oplus/\ominus$ (in damage case k, $k \in \mathbb{N}_0$), $z^k_{\oplus/\ominus} \geq 0$ for all k
$\pi^0_{\oplus/\ominus} = \mathbb{E}[Z_{\oplus/\ominus}]$	insurance contract's net expected value of an agent of type $\oplus/\ominus$
$\pi_{\oplus/\ominus}$	(type-specific) premiums in regime (F) for customers of type $\oplus/\ominus$
$\pi_\odot$	(aggregated) premium in regime (E)
$\lambda_{\oplus/\ominus}$	purchased coverage/insurance demand of customers of type $\oplus/\ominus$
$\hat{\lambda}_{\oplus/\ominus}$	equilibrium insurance coverage of customers of type $\oplus/\ominus$
$\hat{\pi}_{\oplus/\ominus}$	optimal type-specific premiums for customers of type $\oplus/\ominus$
$\hat{\pi}_\odot$	optimal aggregate premium
F	function derived from the optimal coverage equation used for the implicit calculation of the derivative $\hat{\lambda}'_{\oplus/\ominus}$ of the optimal coverage
Ch 3: Life Insurance	
$B(0,T)$	time-T-discount factor
${}_Tp_x^{\oplus/\ominus}$	time-T-survival probability of an $\oplus/\ominus$-agent with initial age x
${}_{k\|1}q_x^{\oplus/\ominus}$	k-year deferred death probability of an $\oplus/\ominus$-agent with initial age x
ω	maximal age an agent can reach (we assume $\omega = 110$)

${}_tm_x$	centralized death probability of an x-year old person at time t, gender/agent type is omitted here
$(a_x)_x$, $(b_x)_x$	age dependent factors of the Lee-Carter mortality model
$(\kappa_t)_t$	time dependent factor of the Lee-Carter mortality model
$\varepsilon_{x,t}$	error term of the Lee-Carter mortality model
$q_{x,t}$	one-year death probability of an x-year old at time t
r_t	short-rate (CIR model) at time t
Π_u, Π_m, Π_f	unisex, male and female premium
ξ_1, ξ_2, ξ_3	mixing parameter based on premiums, probabilities and parameters
Ch 5: Risk Classes and Risk Relations	
$\mathcal{C}$	set of risk classes
η_i	number of customers in risk class $i \in \mathcal{C}$
λ_i	purchased coverage/insurance demand of customer of type $i \in \mathcal{C}$, hat denotes optimality
π_i	(type-specific) premium for customer of type $i \in \mathcal{C}$, hat denotes optimality, $\pi_\odot$ denotes that the premium is global/aggregate
$(w_i)_{i\in\mathcal{C}}$	fraction (or density) of i-agents in the market
Z_i	risk variable of agents of type $i \in \mathcal{C}$
$\pi_i^0 = \mathbb{E}[Z_i]$	insurance contract's net expected value of an agent of type $i \in \mathcal{C}$
n	number of risk classes (if $\mathcal{C} = \{1, \dots, n\}$)
$f : \mathcal{C} \to \mathbb{N}$	rating class assignment function
r_f^i	rating class i given rating class assignment function f
R_f	set of rating classes given f
$m_f = \|R_f\|$	number of rating classes given f
$\mathcal{R}$	set of possible rating class sets/vectors
$\sim$	counting relation, equivalence relation used to find rating class vectors that only differ in the numeration of the rating classes
$\mathcal{R}' = \mathcal{R}/\sim$	set of (equivalence classes) of possible rating class vectors modulo numeration
B_n	n-th Bell number; if there are n risk classes it holds that $\|\mathcal{R}'\| = B_n$
$(Z_t^N)_t$	time homogeneous Markov chain with state space $\{0, \dots, N\}$ that models the number of customers buying insurance from a company
ν_i, μ_i	birth and death rate of the Markov chain $(Z_t^N)_t$
$X_t^N = \frac{Z_t^N}{N}$	standardized version of the Markov chain (values in $[0, 1]$)

$(X(t))_t$	non-random limiting process of the Markov chain
$\sqsubseteq$	preference relation to compare two vectors of the same length n
$\equiv$	equivalence of two vectors in the sense of $\sqsubseteq$ ($a \equiv b :\Leftrightarrow a \sqsubseteq b$ and $b \sqsubseteq a$)
$\mathcal{E}(a)$	indifference curve of vector a w.r.t. $\sqsubseteq$
f_b	function that needs to be maximized in order to find the optimal respond to a vector b
A_1	set used for the discrete example to show the possible non-existence of an optimizer, $A_1 = \{(1,2,0),(0,1,2),(2,0,1)\}$
$\mathcal{O}$	set of locally optimal vectors
conv$X$	convex hull of set X
A_2	extended set for the discrete example, $A_2 = \{(1,2,0),(0,1,2),(2,0,1),(0,2,1),(2,0,1),(1,0,2)\}$
epi$f$	epigraph of a function f
$d_b(x,y)$	distance between x and y induced by f_b, $x,y,b \in \mathbb{R}^n_{\geq 0}$, might write $d(\cdot,\cdot)$ instead of $d_b(\cdot,\cdot)$
$\pi \in \mathbb{R}^n$	premium vector given a rating class vector
$\gamma \in \mathbb{R}^n$	loaded premium vector given a rating class vector
$\hat{\lambda} \in \mathbb{R}^n$	optimal coverage vector given a loaded premium vector γ
$s \in \mathbb{R}^n$	vector which contains the amounts of safety loadings given a rating class vector
$\sqsubseteq_{\text{cov}}$	preference relation to compare the coverage vectors given their rating class vectors
ψ	fraction vector of customers in Company 1
$\mathcal{R}^*$	set of risk coalitions considered by Algorithm 1
c	capacity constraint of an insurance company, can be element wise, contract wise or globally
h	changing parameter which describes how many of the customers which do not meet the capacity constraint of a company switch to the other
$\tilde{a}, \tilde{b}$	capped vectors a and b
freq	frequency vector given $\mathcal{O}_1, \dots, \mathcal{O}_k$ used by Algorithm 3
Ch 6: Health Insurance	
X_{doctor}	random variable of the ambulant costs/medical consultation expenses (normally distributed)
X_{hospital}	random variable of the stationary costs/hospitalization expenses (Pareto distributed)
ps	pension series, vector in $\mathbb{R}^{\omega+1}$

ω	maximal age (here we assume contrary to the life insurance chapter $\omega = 100$)
$\tilde{p}_k$	probability that a customer "survives" the year, i.e. he is still a customer of the insurance company at the end of the year, so he has neither died nor canceled the contract
bp	brutto premium, i.e. constant annual premium that does not rise with the age
π_k	one-year premium of a k-year old person
res_x	(premium) reserve of an x-year old person
π_k^0	expected medical expenses of a k-year old person
$\xi_n = \frac{\eta_n^m}{\eta_n^m + \eta_n^f}$	dynamic mixing parameter for the unisex reserves, i.e. fraction of males in the portfolio at the beginning of year n
η_n^m, η_n^f	number of males and females in the portfolio at the beginning of year n
Ch 7: Disability Insurance	
r_k	probability that a k-year old disabled person recovers within the next year
${}_l r_k$	probability that a k-year old disabled person recovers within the next l years
${}^{nd}_{k}p_x$	k-year survival probability of a non-disabled x-year old agent
${}^{a}_{k}p_x$	k-year survival probability of an active x-year old agent
${}^{r}_{k}p_x$	k-year survival probability of a recovered x-year old agent
$\frac{{}^{a}_{k}l_x}{{}^{nd}_{k}l_x}$	fraction of active among the non-disabled customers of a population of initially x-year old agents after k years
$\frac{{}^{r}_{k}l_x}{{}^{nd}_{k}l_x}$	fraction of recovered among the non-disabled customers of a population of initially x-year old agents after k years
d_k	probability that a k-year old customer falls into disability before the age of $k+1$
${}^{a}d_k$	probability that a k-year old active customer falls into disability before the age of $k+1$
${}^{r}d_k$	probability that a k-year old recovered customer falls into disability before the age of $k+1$
$da(k)$	value of a disability annuity for a person that falls into disability at age k
${}^{i}_{l}p_x$	l-year survival probability of an inactive (disabled) x-year old person

B Data Sources

In this chapter, we summarize the data sources used in this thesis. The author of this thesis greatly acknowledges the availability of the given data sets, which made it possible to apply the presented model to real-life data.

Mortality Data

The mortality data we used in Chapters 3, 4, 5 and 6 originates from the Human Mortality Database (HMD), which can be found under `https://www.mortality.org/`. The database is a project by the University of California, Berkeley, USA and the Max Planck Institute for Demographic Research, Rostock, Germany. It is an open source database providing detailed population and mortality data to researchers and others.
To be able to calculate the premiums for recent birth years, we need to predict the future mortality. To this end, we make use of the Lee-Carter model. More details about this model can be found in Section 3.1.1.
As this thesis is written and supervised at a German university, one would expect the data to be for the German population as well. Nonetheless, there are several problems when using the German data. Due to the division of Germany after World War II, there are two datasets, one for West and one for East Germany. Not only could mixing these datasets cause trouble and produce misleading results, but there was also no data collected before 1956. To make the results more reliable, the data used in this thesis is of French origin, containing data from 1816 to 2018. For our thesis we used the "Deaths 1x1" data as of November 2020. The latest French mortality data can be found under `https://www.mortality.org/Country/Country?cntr=FRATNP`.

Health Data

For our analysis of health insurance products in Section 6.2 we use data that we model from data sets of the German Robert Koch Institute (RKI) [Rob17a, Rob17b]. These data set are called "Inanspruchname ambulanter ärztlicher Versorgung in Deutschland" and "Inanspruchname von Krankenhausbehandlungen in Deutschland" translating to "Utilization of Ambulant Medical Care in Germany" and "Utilization of Hospital Treatments in Germany" and are only available in German. In this data set, the RKI summarizes how often men and women of different ages see the doctor or visit a hospital. We use these frequencies to model the expected losses, see Section 6.2 for more details.
In the rest of Chapter 6 we use real-life health insurance data. Each year, the German Federal Financial Supervisory Authority (Bundesanstalt für Finanzdienstleitungsaufsicht, BaFin) publishes probability tables for private health insurance. The data can be found under `https://www.bafin.de/D`

E/PublikationenDaten/Statistiken/PKV/wahrscheinlichkeitstafe ln_node.html. Each year, the tables from the year before are published. Unfortunately, the data is available only in German. We use the tables as of May 2022.
While we use the data from 2020 for the examples in this thesis, we might also use the mixed data from 2018, 2019 and 2020. In this case, which is also used by insurance companies, we calculate the “mixed” damage heights by simply taking the arithmetic mean of the corresponding damage heights of the three years. The most important part of the data sets given by the BaFin are the “Profile” tables. From these we use the “..._ks” columns, where we replace the dots according to the gender and insurance type we would like to investigate. In this case, “ks” means “Kopfschäden”, a German word translating to loss per head.
As we focus on ambulant health insurances, we need to choose the ambulant loss per head columns which can be found right in the beginning of the table. Note that there are different tables for “Beihilfe” (financial aid for civil servants by the government) and other customers. We also need to decide whether we want to add pregnancy and maternity costs for females. These costs are known as “S-Kosten”, where the S stands for “Schwangerschaft” translating to pregnancy. In the future, the costs for maternity might not be included in “S-Kosten” anymore but given as “M-Kosten”. While these costs are not included at first, we can add them by taking the sum of the female head per loss and pregnancy column. To be precise, we are using the columns “01_KKV_amb_B_M_ks”, “01_KKV_amb_B_W_ks”, “01_KKV_amb_B_S_ks” for males, females and pregnancy cost for “Beihilfe”. If we want analyze the other customers, we can use the columns “01_KKV_amb_N_M_0-100_ks”, “01_KKV_amb_N_W_0-100_ks” and “01_KKV_amb_N_S_0-100_ks”. Note that “0-100” indicates that this is the data for customers with a deductible between 0 and 100€ per year. The tables we are taking the data from are called “KRAWATTE_2018_Tafeln_komplett”,
“PKV_Kopfschadenstatistik_2019_Profile” and
“PKV_Kopfschadenstatistik_2020_TAFELN” for the years 2018, 2019 and 2020, respectively.
When it comes to the calculation of reserves and bruttopremiums, we also need the probabilities that a customer leaves the system. This could happen due to two reasons. First of all, a customer could die. To calculate these probabilities we use our French mortality data. Keep in mind that the health data is of German origin. The two different origins of the probabilities should not cause any trouble, as the French and German mortality probabilities lie close together, especially the observed effects are (almost) the same. Secondly, a customer could cancel his private health insurance and switch back to the public one. These probabilities are given in the “06_STORNO...” columns, where “Storno” can be translated by lapse rate. From 2021, these probabilities are collected in an own table, where more details are provided.

Due to a rule in the German health care system, a customer cannot cancel his contract if he is older than 55. This rule is known as the 55-rule ("55er-Regel" in German) and causes all lapse rates for customers older than 55 to be zero.

Disability Data

In the examples in Chapter 7 we rely our analysis on U.S. data. More precisely, the data originates from the social security administration of the U.S. government and can be found as PDF files under `https://www.ssa.gov/oact/NOTES/ran6/`. We use the tables for customers born in 1985 together with the birth years from 1991 to 2000 as of August 2020. From these eleven documents we are getting the general tables which we use for our computations by regressing the given data with a neural net approach, see Section 7.1 for more details. The tables provide disability, recovery and death probabilities. The death probabilities are disaggregated to active, disabled and recovered customers, while the disability probabilities are split up into active and recovered customers. As the tables already provide death probabilities, also for the active, recovered and disabled customers, we do not need to make use of the mortality data from the HMD. As we have seen in Remark 7.3 it is indeed relevant to distinguish between the different states an insured can be in.

References

[Ake70] George Akerlof. The market for 'lemons': Quality uncertainty and the market mechanism. *quarterly Journal of Economics*, 84(3):488–500, 1970.

[AP94] Knut K. Aase and Svein-Arne Persson. Pricing of unit-linked life insurance policies. *Scandinavian Actuarial Journal*, 1994(1):26–52, 1994.

[BBM09] Anna Rita Bacinello, Enrico Biffis, and Pietro Millossovich. Pricing life insurance contracts with early exercise features. *Journal of computational and applied mathematics*, 233(1):27–35, 2009.

[Bel34] Eric Temple Bell. Exponetial numbers. *The American Monthly*, 41:411–419, 1934.

[Ber99] Dimitri Bertsekas. *Nonlinear Programming.* Athena Scientific, 1999.

[BG15] Elia Berdin and Helmut Gründl. The effects of a low interest rate environment on life insurers. *The Geneva Papers on Risk and Insurance-Issues and Practice*, 40:385–415, 2015.

[BM07] Damiano Brigo and Fabio Mercurio. *Interest Rate Models - Theory and Practice.* Springer-Verlag GmbH, 2007.

[Bor62] Karl Borch. Application of game theory to some problems in automobile insurance. *ASTIN Bulletin*, 2(2):208–221, 1962.

[BT91] Arnoud W.A. Boot and Anjan V. Thakor. Off-balance sheet liabilities, deposit insurance and capital regulation. *Journal of Banking & Finance*, 15(4-5):825–846, 1991.

[BT10] Daniel Berend and Tamir Tassa. Improved bounds on bell numbers and on moments of sums of random variables. *Probability and Mathematical Statistics*, 30(2):185–205, 2010.

[BV04] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization.* Cambridge University Press, March 2004.

[CIR85] John C. Cox, Jonathan E. Ingersoll, and Stephen A. Ross. A theory of the term structure of interest rates. *Econometrica*, 53(2):385, March 1985.

[CLS22] An Chen, Hong Li, and Mark Schultze. Collective longevity swap: A novel longevity risk transfer solution and its economic pricing. *Journal of Economic Behavior & Organization*, 201:227–249, 2022.

[CR23] An Chen and Manuel Rach. The tontine puzzle. *Available at SSRN 4106903*, 2023.

[CS86] Keith J. Crocker and Arthur Snow. The efficiency effects of categorical discrimination in the insurance industry. *Journal of Political Economy*, 94(2):321–344, April 1986.

[CSG+93] Edoardo Casiglia, P. Spolaore, G. Ginocchio, G. Colangeli, G. Di Menza, M. Marchioro, A. Mazza, and Giovanni Battista Ambrosio. Predictors of mortality in very old subjects aged 80 years or over. *European journal of epidemiology*, 9:577–586, 1993.

[CV17] An Chen and Elena Vigna. A unisex stochastic mortality model to comply with EU gender directive. *Insurance: Mathematics and Economics*, 73:124–136, 2017.

[DG02] Pradeep Dubey and John Geanakoplos. Competitive pooling: Rothschild-Stiglitz reconsidered. *The Quarterly Journal of Economics*, 117(4):1529–1570, 2002.

[DHRS22] Maximilian Diehl, Roman Horsky, Susanne Reetz, and Jörn Sass. Long-term stability of a life insurer's balance sheet. *European Actuarial Journal*, 13(1):147–182, 2022.

[Dob77] Gabriel Dobinski. Archiv der Mathematik und Physik, 1877. `https://archive.org/details/archivdermathem88unkngoog/page/332/mode/2up?view=theater`, visited August 2023.

[Eur04] European Union. COUNCIL DIRECTIVE 2004/113/EC of 13 December 2004 implementing the principle of equal treatment between men and women in the access to and supply of goods and services, 2004. `https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2004:373:0037:0043:EN:PDF`, visited August 2023.

[Eur09] European Union. DIRECTIVE 2009/138/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the taking-up and pursuit of the business of Insurance and Reinsurance (Solvency II), 2009. `https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02009L0138-20210630`, visited August 2023.

[Eur11] European Union. Sex Discrimination in Insurance Contracts: Statement by European Commission Vice-President Viviane Reding, the EU's Justice Commissioner, on the European Court of Justice's ruling in the Test-Achats case, 2011. `https://ec.europa.eu/commission/presscorner/detail/en/MEMO_11_123`, visited August 2023.

[For08] Otto Forster. *Analysis 2: Differentialrechnung im IRn, gewöhnliche Differentialgleichungen.* Analysis / Otto Forster. Vieweg+Teubner Verlag, 2008.

[For12] Otto Forster. *Analysis 3: Maß-und Integrationstheorie, Integralsätze im IRn und Anwendungen.* Springer Science & Business Media, 2012.

[GG11] Leonid A. Gavrilov and Natalia S. Gavrilova. Mortality measurement at advanced ages. *North American Actuarial Journal*, 15(3):432–447, Jul. 2011.

[Gom24] Lewis Gompertz. *Moral Inquiries on the Situation of Man and of Brutes.* E. Mellen Press, 1824.

[Gé19] Aurélien Géron. *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow.* O'Reilly UK Ltd., 2019.

[Hai18] Donatien Hainaut. A neural-network analyzer for mortality forecast. *ASTIN Bulletin: The Journal of the IAA*, 48(2):481–508, 2018.

[HP99] Steven Haberman and Ermanno Pitacco. *Actuarial models for disability insurance.* Chapman and Hall, 1999.

[HP00] Michael Hoy and Mattias Polborn. The value of genetic information in the life insurance market. *Journal of Public Economics*, 78(3):235–252, Nov. 2000.

[IKLM11] Albrecht Irle, Jonas Kauschke, Thomas Lux, and Mishael Mlaković. Switching Rates and the asymptotic Behavoiur of Herding Models. *Advances in Complex Systems*, 14(03):359–376, Jun. 2011.

[KS91] Ioannis Karatzas and Steven Shreve. *Brownian motion and stochastic calculus*, volume 113. Springer Science & Business Media, 1991.

[LC92] Ronald D. Lee and Lawrence R. Carter. Modeling and forecasting U. S. mortality. *Journal of the American Statistical Association*, 87(419):659, Sep. 1992.

[LHG+18] Christine K. Lee, Ira Hofer, Eilon Gabel, Pierre Baldi, and Maxime Cannesson. Development and validation of a deep neural network model for prediction of postoperative in-hospital mortality. *Anesthesiology*, 129(4):649–662, 2018.

[Mak60] William Matthew Makeham. On the law of mortality and the construction of annuity tables. *The Assurance Magazine and Journal of the Institute of Actuaries*, 8(6):301–310, Jan. 1860.

[Nie15] Andreas Niemeyer. Safety margins for systematic biometric and financial risk in a semi-markov life insurance framework. *Risks*, 3(1):35–60, 2015.

[Ohe20] Marek Oheim. Unisex Tariffs: Stochastic Mortality Models and Equilibrium Premiums. *Unpublished Master's Thesis*, Sep. 2020.

[OP11] Annamaria Olivieri and Ermanno Pitacco. *Introduction to Insurance Mathematics.* Springer-Verlag GmbH, Jan. 2011.

[PCS+09] George C. Patton, Carolyn Coffey, Susan M. Sawyer, Russell M. Viner, Dagmar M. Haller, Krishna Bose, Theo Vos, Jane Ferguson, and Colin D. Mathers. Global patterns of mortality in young people: a systematic analysis of population health data. *The Lancet*, 374(9693):881–892, Sep. 2009.

[Pit14] Ermanno Pitacco. Health insurance. *Basic Actuarial Models, Cham, Switzerland: Springer Verlag*, 2014.

[Pru15] Prudential Insurance Company of America. The documentary history of insurance, 1000 B.C. - 1875 A.D., 1915. `https://archive.org/details/cu31924030231736/page/n3/mode/2up`, visited August 2023.

[Rie06] Oliver Riedel. Unisex tariffs in health insurance. *The Geneva Papers on Risk and Insurance - Issues and Practice*, 31(2):233–244, Apr. 2006.

[Rob17a] Robert Koch-Institut. Inanspruchnahme ambulanter ärztlicher Versorgung in Deutschland. *Journal of Health Monitoring*, (4):88–94, 2017. `https://www.rki.de/DE/Content/Gesundheitsmonitoring/Gesundheitsberichterstattung/GBEDownloadsJ/FactSheets/JoHM_04_2017_Ambulante_aerztliche_Versorgung.pdf?__blob=publicationFile`, visited August 2023.

[Rob17b] Robert Koch-Institut. Inanspruchnahme von Krankenhausbehandlungen in Deutschland. *Journal of Health Monitoring*, (4):95–100, 2017. `https://www.rki.de/DE/Content/Gesundheitsmonitoring/Gesundheitsberichterstattung/GBEDownloadsJ/FactSheets/JoHM_04_2017_Krankenhausbehandlungen.pdf?__blob=publicationFile`, visited August 2023.

[RS76] Michael Rothschild and Joseph Stiglitz. Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. *The Quarterly Journal of Economics*, 90(4):629, Nov. 1976.

[Sab10] Michael J. Sabin. Fair tontine annuity. *Available at SSRN 1579932*, 2010.

[SLC+15] Scot H. Simpson, Jayson Lee, Sabina Choi, Ben Vandermeer, Ahmed S. Abdelmoneim, and Travis R. Featherstone. Mortality risk among sulfonylureas: a systematic review and network meta-analysis. *The lancet Diabetes & endocrinology*, 3(1):43–51, 2015.

[SS14] Jörn Sass and Frank Thomas Seifried. Insurance markets and unisex tariffs: is the European court of justice improving or destroying welfare? *Scandinavian Actuarial Journal*, 2014(3):228–254, May 2014.

[SSW14] Hato Schmeiser, Tina Störmer, and Joël Wagner. Unisex insurance pricing: Consumers' perception and market implications. *The Geneva Papers on Risk and Insurance-Issues and Practice*, 39:322–350, 2014.

[Wam00] Achim Wambach. Introducing heterogeneity in the Rothschild-Stiglitz model. *Journal of Risk and Insurance*, pages 579–591, 2000.

[Wil77] Charles Wilson. A model of insurance markets with incomplete information. *Journal of Economic Theory*, 16(2):167–207, Dec. 1977.

[YGL+21] Wenguang Yu, Guofeng Guan, Jingchao Li, Qi Wang, Xiaohan Xie, Yu Zhang, Yujuan Huang, Xinliang Yu, and Chaoran Cui. Claim amount forecasting and pricing of automobile insurance based on the bp neural network. *Complexity*, 2021:1–17, 2021.

[YSWB01] Ai Cheo Yeo, Kate A. Smith, Robert J. Willis, and Malcolm Brooks. Modeling the effect of premium changes on motor insurance customer retention rates using neural networks. In *Computational Science-ICCS 2001: International Conference San Francisco, CA, USA, May 28—30, 2001 Proceedings, Part II 1*, pages 390–399. Springer, 2001.

[YZ02] Virginia R. Young and Thaleia Zariphopoulou. Pricing dynamic insurance risks using the principle of equivalent utility. *Scandinavian Actuarial Journal*, 2002(4):246–279, 2002.

Scientific and Professional Career

Education

10/2020 – present	PhD student of Prof. Dr. Jörn Sass, *RPTU Kaiserslautern-Landau*
01/2018 – 09/2020	Master of Science Mathematics International, *TU Kaiserslautern*
01/2018 – 05/2018	Semester abroad in Singapore, *National University of Singapore*
10/2014 – 12/2017	Bachelor of Science Mathematics, *TU Kaiserslautern*
06/2014	Abitur, *Dürer-Gymnasium, Nürnberg*

Experience

04/2021 – present	Researcher in the Financial Mathematics Group, *RPTU Kaiserslautern-Landau*
10/2018 – 03/2021	Research assistant, *TU Kaiserslautern*
04/2016 – 12/2017	Student assistant, *TU Kaiserslautern*

Wissenschaftlicher und Beruflicher Werdegang

Ausbildung

10/2020 – heute	Doktorand bei Prof. Dr. Jörn Sass *RPTU Kaiserslautern-Landau*
01/2018 – 09/2020	Masterstudium Mathematics International, *TU Kaiserslautern*
01/2018 – 05/2018	Auslandssemester in Singapur, *National University of Singapore*
10/2014 – 12/2017	Bachelorstudium Mathematik, *TU Kaiserslautern*
06/2014	Abitur, *Dürer-Gymnasium, Nürnberg*

Erfahrung

04/2021 – heute	Wissenschaftlicher Mitarbeiter in der AG Finanzmathematik, *RPTU Kaiserslautern-Landau*
10/2018 – 03/2021	Wissenschaftliche Hilfskraft, *TU Kaiserslautern*
04/2016 – 12/2017	Studentische Hilfskraft, *TU Kaiserslautern*

Statutory Declaration

I declare that the submitted dissertation was written independently, that the aids used for the work are named and that the results of any collaborators or other authors are clearly marked.
Neither the dissertation nor parts of the dissertation have already been submitted to another department as a separate examination paper. I have not made any publications about the topics of this dissertation.

Marek Oheim, Kaiserslautern, December 2023