TECHNISCHE UNIVERSITÄT WIEN

D I P L O M A R B E I T

Policy Iteration in a Two-Dimensional Ruin Problem

zur Erlangung des akademischen Grades

Diplom-Ingenieur

im Rahmen des Studiums

Finanz- und Versicherungsmathematik

eingereicht von

Stefan Stajic B.Sc.
Matrikelnummer 01525677

ausgeführt am Institut für Stochastik und Wirtschaftsmathematik
der Fakultät für Mathematik und Geoinformation der Technischen Universität Wien

Betreuung
Betreuer: Ao.Univ.Prof. Dipl.-Ing. Dr.techn. Peter Grandits
Mitwirkung: Dr.rer.nat. Maike Klein

Wien, 7. Januar 2022

(Unterschrift Verfasser) (Unterschrift Betreuer)

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub

TECHNISCHE UNIVERSITÄT WIEN

DIPLOMA THESIS

Policy Iteration in a Two-Dimensional Ruin Problem

Submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Financial and Actuarial Mathematics

by

Stefan Stajic B.Sc.
Registration Number 01525677

to the Institute of Statistics and Mathematical Methods in Economics
at the Faculty of Mathematics and Geoinformation TU Wien

Supervision
Advisor: Ao.Univ.Prof. Dipl.-Ing. Dr.techn. Peter Grandits
Assistance: Dr.rer.nat. Maike Klein

Vienna, January 7, 2022

(signature of the author) (signature of the advisor)

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Abstract

This paper is concerned with the application and numerical implementation of the policy iteration method to a two-dimensional stochastic control model with two independent Brownian motions describing the behavior of two firms in a market, as presented in the article by McKean and Shepp [18].

In the article it was shown that the chosen strategy by which the companies are supported by the government in the form of tax breaks varies depending on the optimization criterion. It turns out that if the objective is to maximize the joint survival probability of both firms, the push-bottom strategy, i.e., the strategy that supports the weaker firm at the considered time, is the optimal strategy. If, on the other hand, the criterion is to maximize the expected number of firms that remain solvent, the application of the push-top strategy is partially more suitable than the push-bottom strategy.

Now the question arises whether the optimal strategy for the respective optimization problem can be found even if the companies are subsidized by the government at the beginning according to an arbitrary strategy. Since the underlying partial differential equation for the value function is nonlinear, standard methods for linear partial differential equations cannot be used.

For this reason, we use the policy iteration method, in which an improved policy is derived at each step and the solution of the thereby observed linear partial differential equation is approximated using difference quotients and the successive over-relaxation method.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Kurzfassung

In der vorliegenden Arbeit geht es um die Anwendung und numerische Umsetzung der Policy-Iterationsmethode auf ein zweidimensionales stochastisches Kontrollmodell mit zwei unabhängigen Brownschen Bewegungen, die das Verhalten von zwei Firmen auf einem Markt beschreiben, wie es in dem Artikel von McKean und Shepp [18] vorgestellt wurde.

Im Artikel wurde nachgewiesen, dass die gewählte Strategie, anhand derer die Unternehmen durch den Staat in Form von Steuererleichterungen gefördert werden, je nach Optimierungskriterium variiert. Dabei stellt sich heraus, dass, wenn das Ziel darin besteht, die gemeinsame Überlebenswahrscheinlichkeit beider Unternehmen zu maximieren, die Push-Bottom Strategie, d.h. die Strategie, die das schwächere Unternehmen zum betrachteten Zeitpunkt unterstützt, die optimale Strategie ist. Besteht das Kriterium hingegen in der Maximierung der erwarteten Anzahl von Unternehmen, die zahlungsfähig bleiben, ist die Anwendung der Push-Top Strategie teilweise geeigneter als die Push-Bottom Strategie.

Nun stellt sich die Frage, ob die optimale Strategie für das jeweilige Optimierungsproblem auch dann gefunden werden kann, wenn die Unternehmen zu Beginn entsprechend einer willkürlichen Strategie durch die Regierung gefördert werden. Da die zugrunde liegende partielle Differentialgleichung für die Wertfunktion nichtlinear ist, können Standardmethoden für lineare partielle Differentialgleichungen nicht verwendet werden.

Aus diesem Grund verwenden wir die Policy-Iterationsmethode, bei der in jedem Schritt eine verbesserte Strategie hergeleitet und die Lösung der dabei auftretenden linearen partiellen Differentialgleichung mit Hilfe von Differenzenquotienten und des Überrelaxationsverfahrens, auch SOR-Verfahren genannt, approximiert wird.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Acknowledgement

At this point, I would like to thank all the people who supported me in writing my diploma thesis, in particular my supervisor Ao.Univ.Prof. Dipl.-Ing. Dr.techn. Peter Grandits for his time and patience. He gave me the opportunity to work on an interesting topic and was always being there to answer my questions.

Hereby, I would like to thank Dr.rer.nat. Maike Klein as well, for her support regarding the whole implementation process and her valuable comments on my diploma thesis.

Special thanks to my parents and my older brother for their consistent help and great support during my studies, so that I was not giving up hope for my final degree.

Last but not least I want to mention my friends, fellow students and my work colleagues who always helped me mentally, gave me encouragement in difficult times and happiness in my spare time.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Diplomarbeit selbstständig und ohne fremde Hilfe verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt bzw. die wörtlich oder sinngemäß entnommenen Stellen als solche kenntlich gemacht habe.

Wien, am 7. Januar 2022

Stefan Stajic

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Contents

1 Introduction **1**

2 Stochastic Fundamentals **3**
2.1 Stochastic Processes, Filtrations and Stopping Times . . . 3
2.2 Martingales and Brownian Motion . . . 6
2.3 Stochastic Integration . . . 9
2.4 Ito Process and Ito Formula . . . 12
2.5 Stochastic Differential and Integral Equations . . . 14

3 Risk and Ruin Theory **17**
3.1 Risk Models . . . 17
3.2 Cramer Lundberg Model . . . 20
3.3 Diffusion Approximation . . . 26

4 Stochastic Control Theory **30**
4.1 The Stochastic Control Problem . . . 30
4.2 Dynamic Programming . . . 31
4.2.1 Ito Diffusions and their Generators . . . 31
4.2.2 Bellman Principle & Hamilton-Jacobi-Bellman Equation . . . 35
4.3 Verification Theorems . . . 37
4.3.1 Finite Time Horizon . . . 37
4.3.2 Infinite Time Horizon . . . 41

5 Two-Dimensional Controlled Ruin Problem **43**
5.1 Mathematical Model of the Specific Ruin Problem . . . 43
5.2 Mathematical Difficulties of the Specific Ruin Problem . . . 44
5.3 The Optimal Control Policy for Two Companies . . . 47
5.4 Numerical Approximation & Results . . . 48

6 Policy Iteration Method **51**
6.1 Markov Theory . . . 51
6.2 Reinforcement Learning Problem . . . 55
6.2.1 Policy Evaluation . . . 59
6.2.2 Policy Improvement . . . 60
6.2.3 Policy Iteration . . . 62
6.3 An Illustrative Example . . . 63
6.3.1 Overview . . . 64
6.3.2 State Transition Diagram & Transition Probability Matrix . . . 64
6.3.3 Finding the Optimal Policy . . . 66

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

7 Numerical Analysis **70**
7.1 Heuristic Derivation of the PDE 71
7.2 Finite-Difference Method 73
7.3 Iterative Approaches 78
7.4 Results 80

8 Conclusion **135**

Bibliography **136**

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

1 Introduction

This paper deals with a topic from stochastic control theory, which, given appropriate assumptions, can provide a theoretical model of how a government might design its tax policy and is based on the article of McKean and Shepp [18], in which a simplified model is presented. In their model, an economy is defined over two companies, where the wealth of both firms is each modeled by a Brownian motion with equal volatility. Furthermore, the government payments in order to help the companies are described by the drift of the Brownian motion, where a positive drift means a subsidy that helps only one firm. The joint survival probability of both companies is used as a variable for optimization which implies that a two-dimensional ruin problem is considered. For the sake of simplicity, it is also assumed that there are two choices of control in form of two distinct government tax policies towards companies.

There is a socialist tax system in which the democratic policy is used and therefore only the weaker company for the considered time point is supported by the government but a capitalist tax system is also applicable in which the republican policy is used instead and therefore only the stronger company receives help from the state. By using techniques from stochastic control theory, the authors were able to derive a proof of which tax system should be used, when the optimization criterion was to maximize the joint survival probability of both firms. What can already be heuristically conjectured, in this case the socialist tax system must be used.

For the interested reader, there are several articles dealing with the extension of the obtained results in [18]. For example in an article by Grandits [9], the framework of McKean and Shepp is modified such that the volatility of one of the two firms is slightly increased. In another article, [10], also written by Grandits, it is proved that the push-bottom strategy is not optimal if the optimization criterion is to maximize the expected number of surviving companies. However, there are also more general forms of the model introduced by McKean and Shepp with more than two firms whose wealth processes are modeled by two Brownian motions, which have been treated by authors such as Aldous [1] or Tang and Tsai [27], among others.

The aim of the work presented here is to use an iterative procedure, namely the policy iteration method, and thereby verify the numerical results obtained by the authors in [18]. The question is whether approximately equal solutions can be found using the policy iteration method and whether the results are plausible. Moreover, the question arises as to what the separation curve, i.e., the curve that separates the area under consideration into two surfaces, each of which is supposed to represent the government's support for the respective company, looks like when the volatility of the second company is changed so that $\sigma_x^2 = 1, \sigma_y^2 = 1 + \epsilon$ with $\epsilon > 0$ holds.

This paper is structured as follows.

Chapter 2 summarizes the basic concepts of stochastic analysis that will be used in

the following chapters. Thus, the reader is familiarized with the definition and properties of a stochastic process as well as filtrations. Thereafter, examples of such processes are discussed, namely martingales and Brownian motions. In addition, the Ito formula is introduced in both the one-dimensional and multidimensional cases.

Chapter 3 is about basic principles of risk and ruin theory. It provides results for the individual and the collective risk model but also for survival probabilities, especially on insurance companies. Furthermore, this chapter focuses on the introduction of the Cramer-Lundberg model and the diffusion approximation, which is an approximation to the aforementioned model.

Chapter 4 covers the area of stochastic control theory. The basic concepts of a stochastic optimization problem are explained in more detail to familiarize the reader with stochastic optimization. The principle of dynamic programming is presented and applied to obtain a formal derivation of the Hamilton-Jacobi-Bellman equation. In order to determine whether a solution to this equation is also a value function, the so-called verification theorem is formulated.

Chapter 5 is concerned with the article by McKean and Shepp and its results. This includes a detailed description of the underlying two-dimensional model, which in turn is based on a more general model formulated by David Aldous [1]. Depending on the optimization criterion, that is, maximizing the joint probability of survival or maximizing the expected number of survivors, the model by McKean and Shepp can determine which of the two distinct government tax policies towards companies, i.e., the republican policy or the democratic policy, is more beneficial.

Chapter 6 provides an introduction to the policy iteration method, which is used in reinforcement learning problems to compute an optimal strategy for a given Markov decision process. Therefore, a brief overview of Markov theory is first given before turning to the algorithm of the method, which generally consists of three main steps, namely the initialization, evaluation and improvement of the policy. The last two steps of the algorithm are performed iteratively until the optimal policy for the considered problem is found. At the end of this chapter, an example is presented to illustrate the policy iteration method.

In Chapter 7, first a heuristic derivation of the considered partial differential equation is presented in order to present methods for its numerical solution afterwards. The finite-difference method is applied to approximate the partial differential equation, whose resulting linear system of equations is usually solved by iterative methods such as Jacobi iteration or successive over-relaxation. Last but not least, the results are presented by means of graphs and explained thereafter.

In Chapter 8 a concise summary of the obtained results is presented.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

2 Stochastic Fundamentals

First of all, we want to repeat important mathematical definitions and theorems without proofs from stochastic analysis. This should only be a summarizing repetition. If some of the basics principles are not yet known to the reader we recommend to use the corresponding references[1], which are also used throughout the Sections 2.1–2.5 below.

2.1 Stochastic Processes, Filtrations and Stopping Times

In the following sections we will consider the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, which is composed of

- the event space Ω, which summarizes all possible elementary events $\omega \in \Omega$,
- the σ-Algebra $\mathcal{F}$, which describes the possibles events and
- the probability measure $\mathbb{P}$, which describes the relative frequency with which the events can occur.

A basic concept in probability theory is the so-called stochastic process or more generally called random process, which describes a set of random variables that represents the evolution of a system over time.

Definition 2.1.1 (Stochastic Process)**.**

A stochastic process is a family of random variables $X = (X_t)_{t\in\mathbb{T}}$ on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathbb{T} \neq \emptyset$ denotes the non-empty indexset, which describes the time.

If $\mathbb{T} = \{0, 1, 2, \ldots\}$, then X is a sequence of random variables and is called stochastic process in discrete time. In this paper the index set can be interpret as a time interval, where we will distinguish between a finite interval $\mathbb{T} = [0, T]$, for $0 < T < \infty$ and an infinite time interval $\mathbb{T} = [0, \infty)$. In these cases X is called stochastic process in continuous time.

Definition 2.1.2 (Paths of a Process)**.**

For every $\omega \in \Omega$ the mapping $X(., \omega) : \mathbb{T} \ni t \mapsto X_t(\omega)$ is called realization of the process X, which is also known as trajectory or path of the process.

[1] see [8], [21], [24] and [29]

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Definition 2.1.3 (Càdlàg, Càglàd and Continuous Process)**.**

A stochastic process $X = (X_t)_{t\in\mathbb{T}}$ is called càdlàg if almost all paths of X are right continuous for which the left-side limit exists.
A stochastic process $X = (X_t)_{t\in\mathbb{T}}$ is called càglàd if almost all paths of X are left continuous for which the right-side limit exists.
If all paths of the stochastic process X are continuous, then the process is called continuous.

With increasing time, our knowledge and thus the available information about past processes grows. This modeling of time is achieved in stochastic analysis by the so-called filtration.

Definition 2.1.4 (Filtration)**.**

A filtration $\mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{T}}$ on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is an increasing family of sub-σ-algebras of the σ-Algebra $\mathcal{F}$, which means, that

$$\mathcal{F}_s, \mathcal{F}_t \subset \mathcal{F} \quad \text{and} \quad \mathcal{F}_s \subset \mathcal{F}_t, \quad s, t \in \mathbb{T}, s \leq t.$$

Therefore, $\mathcal{F}_t$ contains the information up to time t, which in financial mathematics is of huge importance, because trading strategies, for instance, are not allowed to use information from the future. The stochastic triple $(\Omega, \mathcal{F}, \mathbb{F})$ is called filtered measurable space.

Definition 2.1.5 (Filtered Probability Space)**.**

If $\mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{T}}$ is a filtration, then the quadruple $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ is called a filtered probability space.

Furthermore, we want to define two important properties for filtrations, which in further analysis will be always assumed:

Definition 2.1.6 (Right-Continuous)**.**

A filtration $\mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{T}}$ is called right-continuous, if $\mathcal{F}_t = \bigcap_{r>t} \mathcal{F}_r \quad \forall t \in \mathbb{T}$.

Definition 2.1.7 (Complete)**.**

A filtration $\mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{T}}$ is called complete, if $\mathcal{F}_0 = \{\emptyset, \Omega\}$ contains all $\mathbb{P}$-nullsets of the σ-algebra $\mathcal{F}$.

If a filtration has the properties 2.1.6 and 2.1.7, then a filtered probability space is said to satisfy the usual hypotheses.
The following definition expresses that when observing the process, no more information can be gained than the filtration is already containing.

Definition 2.1.8 (Adapted Process)**.**

A stochastic process $X = (X_t)_{t\in\mathbb{T}}$ is called adapted with respect to the filtration $\mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{T}}$, if for every $t \in \mathbb{T}$ it holds, that X_t is $\mathcal{F}_t$-measurable.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Definition 2.1.9 (Progressively Measurable Process).

A stochastic process $X = (X_t)_{t\in\mathbb{T}}$ is called $\mathbb{F}$-progressively measurable or simply $\mathbb{F}$-progressive, if for every $t \in \mathbb{T}$ the mapping $(s, \omega) \to X_s(\omega)$ is $\mathcal{B}([0, t]) \otimes \mathcal{F}_t$-measurable on $[0, t] \times \Omega$.

Definition 2.1.10 (Induced or Generated Filtration).

The smallest right-continuous and complete filtration, for which the stochastic process X is adapted, is called the induced or generated filtration by X, also known as natural filtration, and is denoted by $\mathbb{F}^X$.

The concept of stopping times has an important role in the theory of stochastics. Stopping times are often used to determine optimal strategies but also for the optimal exercise time for American options, for instance.

Definition 2.1.11 (Stopping Time).

A random variable $\tau : \Omega \mapsto \mathbb{T} \cup \{\infty\}$ is called stopping time with respect to the filtration $\mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{T}}$, if

$$\{\tau \leq t\} := \{\omega \in \Omega : \tau(\omega) \leq t\} \in \mathcal{F}_t \quad \forall t \in \mathbb{T}.$$

A map $\tau : \Omega \mapsto \mathbb{T} \cup \{\infty\}$ is called random time, if it is $\mathcal{F}$-measurable.

For a stopping time τ we know at every time $t \in \mathbb{T}$ whether we have stopped already, which reflects the idea of "not to look into the future".

Definition 2.1.12 (Stopping Time σ-algebra $\mathcal{F}_\tau$).

Let τ be a stopping time on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. With every τ we associate the σ-algebra

$$\mathcal{F}_\tau = \{A \in \mathcal{F} : A \cap \{\tau \leq t\} \in \mathcal{F}_t, \text{for all } t \in \mathbb{T}\}.$$

This set represents all observed events up to the stopping time τ. In other words, it models the information contained in $\mathbb{F}$, which is revealed up to the stopping time τ.

Definition 2.1.13 (Stopped Process).

For a stochastic process X and a stopping time τ with respect to the filtration $\mathbb{F}$ we define the process $X_t^\tau := X_{\tau\wedge t}$[2] and call it the stopped process with respect to the stopping time τ.

In the following we will list some important theorems without proofs and further definitions regarding stopping times.

Definition 2.1.14 (Enlargement).

A filtration $\mathbb{G} = (\mathcal{G}_t)_{t\in\mathbb{T}}$ of $\mathcal{F}$ is called an enlargement of the filtration $\mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{T}}$, if $\mathcal{F}_t \subset \mathcal{G}_t$ for all $t \in \mathbb{T}$.

[2] $\tau \wedge t := min(\tau, t)$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Theorem 2.1.15 (Properties of Stopping Times and their σ-algebras)**.**
For all $t \in \mathbb{T}$ and stopping times σ and τ, we have

- *the constant time $\Omega \ni \omega \mapsto \tilde{\tau}(\omega) := t$ is a stopping time and $\mathcal{F}_{\tilde{\tau}} = \mathcal{F}_t$,*
- *$\sigma \vee \tau$ (point-wise maximum) and $\sigma \wedge \tau$ (point-wise minimum) are stopping times,*
- *τ is $\mathcal{F}_\tau$-measurable,*
- *if $\sigma \leq \tau$ (pointwise), then $\mathcal{F}_\sigma \subset \mathcal{F}_\tau$,*
- *$\mathcal{F}_{\sigma\wedge\tau} = \mathcal{F}_\sigma \cap \mathcal{F}_\tau$,*
- *$\mathbb{F}^\tau := (\mathcal{F}_{\tau\wedge t})_{t\in\mathbb{T}}$ is a filtration,*
- *$F \cap \{\sigma \leq \tau\} \in \mathcal{F}_{\sigma\wedge\tau}$ for all $F \in \mathcal{F}_\sigma$,*
- *$F \cap \{\sigma = \tau\} \in \mathcal{F}_{\sigma\wedge\tau}$ for all $F \in \mathcal{F}_\sigma$,*
- *if $(\tau_n)_{n\in\mathbb{N}}$ is a sequence of stopping times and $\tau := \sup_{n\in\mathbb{N}} \tau_n$ takes values in $\mathbb{T} \cup \{\infty\}$, then τ is a stopping time, too.*

Definition 2.1.16 (First Entrance Times)**.**
For an $\mathbb{F}$-adapted stochastic process X and an $\mathbb{F}$-stopping time σ we define the first hitting time of a Borel-set A after σ (also called first entrance time) as follows

$$\tau := inf\{t \in \mathbb{T} : \sigma \leq t, X_t \in A\}$$

with the convention $inf\ \emptyset = sup\ \mathbb{T}$.
If the set A is closed and the process X is continuous, then τ is an $\mathbb{F}$-stopping time, too.

2.2 Martingales and Brownian Motion

Another basic concept in probability theory are martingales, which are a special class of stochastic processes. Because of their useful and important properties, martingales are widely used in stochastic calculations.

Definition 2.2.1 (Martingale)**.**
A stochastic process $M = (M_t)_{t\in\mathbb{T}}$ with respect to the filtration $\mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{T}}$ is called martingale (or $\mathbb{F}$-martingale), if

1. *M is integrable, i.e. $\mathbb{E}[|M_t|] < \infty$ for all $t \in \mathbb{T}$,*
2. *M it is adapted to $\mathbb{F}$ and*
3. *M fulfills the martingale condition, i.e. $\mathbb{E}[M_t \mid \mathcal{F}_s] = M_s$ almost surely, for $s \leq t \in \mathbb{T}$.*

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

If in Definition 2.2.1 the martingale condition is not fulfilled, but it holds that

$$\mathbb{E}[M_t \mid \mathcal{F}_s] \leq M_s \text{ almost surely, for } s \leq t \in \mathbb{T},$$

then we call the process M a supermartingale.
If in Definition 2.2.1 the martingale condition is not fulfilled, but it holds that

$$\mathbb{E}[M_t \mid \mathcal{F}_s] \geq M_s \text{ almost surely, for } s \leq t \in \mathbb{T},$$

then we call the process M a submartingale.

Theorem 2.2.2.

M is a submartingale $\Leftrightarrow$ $-M$ is a supermartingale.

Remark 2.2.3.

For an integrable $\mathbb{F}$-adapted process M, the martingale condition is equivalent to

$$\mathbb{E}[M_t - M_s \mid \mathcal{F}_s] = 0 \text{ almost surely, for } s \leq t \in \mathbb{T}.$$

Remark 2.2.4.

For a martingale M the map $\mathbb{T} \ni t \mapsto \mathbb{E}[M_t]$ is constant, because by a property for conditional expectations and Remark 2.2.3 it holds that

$$\mathbb{E}[M_t] - \mathbb{E}[M_s] = \mathbb{E}\left[\mathbb{E}[M_t - M_s \mid \mathcal{F}_s]\right] = 0 \text{ for } s \leq t \in \mathbb{T}.$$

Theorem 2.2.5 (Stopped Martingale)**.**

Let M be a martingale and τ a finite stopping time. Then the stopped martingale M^τ is a martingale, too.

Theorem 2.2.6 (Martingale Convergence Theorem)**.**

Let $(M_t)_{t\in[0,\infty)}$ be a L^1-bounded supermartingale, i.e. it holds that $\sup_{t\in\mathbb{R}^+} \mathbb{E}[|M_t|] < \infty$. Then $\lim_{t\to\infty} M_t$ exists almost surely and it is almost surely finite with $\lim_{t\to\infty} M_t = M_\infty$.

Theorem 2.2.7.

Let M be a non-negative supermartingale. Then $\lim_{t\to\infty} M_t$ exists almost surely.

Definition 2.2.8 (Local Martingale)**.**

A stochastic process M is called local martingale, if there exists a sequence of stopping times $(\tau_n)_{n\in\mathbb{N}}$ such that $\lim_{n\to\infty} \tau_n = \infty$ almost surely and the stopped process $(M_t^{\tau_n})$ is a martingale for every $n \in \mathbb{N}$.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Theorem 2.2.9 (Stopping Theorem)**.**

Let M be a martingale and τ_1, τ_2 be two stopping times. Then it holds:

$$\mathbb{E}[M_t^{\tau_1} \mid \mathcal{F}_{\tau_2}] = M_t^{\tau_1 \wedge \tau_2},$$

which means, that the stopped martingale is a martingale, too.
If we have bounded stopping times, $\tau_2 \leq \tau_1 < \infty$, then it holds that

$$\mathbb{E}[M_t^{\tau_1} \mid \mathcal{F}_{\tau_2}] = M_t^{\tau_2}.$$

A well known example for a martingale is the Brownian motion, also called the Wiener process, which is probably the most important stochastic process in continuous-time in financial and actuarial mathematics. Brownian motions have a wide range of application areas, for instance, in financial mathematics, they are used for modeling equity returns or for risk modeling. Furthermore, many other processes can be modeled by Brownian motions, such as diffusion processes.

Definition 2.2.10 ((Standard) Brownian Motion)**.**

Consider $\mathbb{T} = [0, \infty)$. An $\mathbb{R}^d$-valued stochastic process $W = (W_t)_{t \in \mathbb{T}}$ on the filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ is called d-dimensional Brownian motion or Wiener process with respect to the filtration $\mathbb{F}$ and the probability measure $\mathbb{P}$, if:

- *W_0 is $\mathcal{F}_0$-measurable,*
- *Measurability of past increments: $W_s - W_r$ is $\mathcal{F}_s$-measurable for all $r < s \in \mathbb{T}$,*
- *Independence of future increments: $W_t - W_s$ is independent of $\mathcal{F}_s$ for all $s < t \in \mathbb{T}$,*
- *Stationarity of the distribution of the increments: $\mathcal{L}(W_t - W_s) = \mathcal{L}(W_{t-s} - W_0)$ for all $s < t \in \mathbb{T}$,*
- *Normality of increments: $\mathcal{L}(W_t - W_0) = \mathcal{N}(0, tI_d)$ for all $t \in \mathbb{R}^+$, where $I_d \in \mathbb{R}^{d \times d}$ denotes the identity matrix and*
- *W has continuous paths, meaning that the map $\mathbb{T} \ni t \mapsto W_t(\omega)$ is continuous for every $\omega \in \Omega$.*

The Brownian motion is called standard, if in addition

- *Start at the origin: $\mathbb{P}[W_0 = 0] = 1$.*

Definition 2.2.11 (Levy Process)**.**

If we drop the distributional requirement in Definition 2.2.10 and weaken the continuity requirement to

- *Every path of the process is right-continuous and has left-hand limits at every t which is not the left endpoint of $\mathbb{T}$ (càdlàg paths),*

then we obtain the larger class of (standard) d-dimensional Levy processes.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Proposition 2.2.12.

Let $(W_t)_{t\in\mathbb{T}}$ be a Brownian motion with respect to the filtration $(\mathcal{F}_t)_{t\in\mathbb{T}}$.

1. *Symmetry: $(-W_t)_{t\in\mathbb{T}}$ is also a Brownian motion,*
2. *Scaling: for all $\lambda > 0$, the process $\left(\left(\frac{1}{\lambda}\right) W_{\lambda^2 t}\right)_{t\in\mathbb{T}}$ is also a Brownian motion,*
3. *Invariance by translation: for all $s > 0$, the process $(W_{t+s} - W_s)_{t\in\mathbb{T}}$ is a standard Brownian motion independent of $\mathcal{F}_s$.*

Definition 2.2.13 (Brownian Motion with Drift)**.**

A stochastic process (X_t) is called linear Brownian motion with drift, if it can be written in the form

$$X_t = \mu t + \sigma W_t,$$

where (W_t) is a Brownian motion, $\mu > 0$ is the drift, σ is the volatility and μ, σ are constant.

Definition 2.2.14 (Geometric Brownian Motion)**.**

A stochastic process (M_t) is called geometric Brownian motion, if it can be written in the form

$$M_t = \exp\left(\frac{-2\mu}{\sigma^2} X_t\right),$$

where (X_t) is a Brownian motion with drift (see Definition 2.2.13).
Furthermore, it is easy to show, that (M_t) is a martingale.

Definition 2.2.15 (First Hitting Time of a Level for a Brownian Motion)**.**

The first hitting time of a level $a \in \mathbb{R}$ for a one-dimensional Brownian motion $(W_t)_{t\geq 0}$ is defined as

$$\tau_a = \inf\{t \geq 0 : W_t = a\},$$

which is a stopping time, too.

2.3 Stochastic Integration

In order to perform an integration with respect to a stochastic process, in our case with respect to the Brownian motion, the concept of integration has to be extended, because the Brownian motion is of infinite variation on any interval $\mathbb{T} \subseteq [0, \infty)$. Thus a path-wise definition of the stochastic integral with respect to the Brownian motion is not possible. However, we can take advantage of the fact, that a Brownian motion is a stochastic process and therefore we can use weaker forms of limits. Therefore, we will define the stochastic integral for the Brownian motion according to Ito.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Step 1: Simple integrands

For a partition $\mathcal{T} = \{0 = t_0 < \ldots < t_n = t\}$ we define the space of simple integrands as

$$S_0 := \left\{ H_t : H_t = \phi_1 \mathbb{1}_{\{0\}}(t) + \sum_{i=1}^{\infty} \phi_i \mathbb{1}_{[t_{i-1},t_i]}(t) \right\},$$

where $\phi_i \in L^\infty(\Omega, \mathcal{F}_{t_{i-1}}, \mathbb{P})$.

Now we can define the Ito integral with respect to simple integrands:

Definition 2.3.1 (Ito Integral with respect to Simple Integrands)**.**
With the above notation we define the Ito integral for simple integrands as

$$\int_0^t H_s dW_s := \sum_{i=1}^{n} \phi_i (W_{t_i \wedge t} - W_{t_{i-1} \wedge t}) \text{ for } H \in S_0, \tag{2.1}$$

for which the left hand side of (2.1) is a L^2-bounded martingale.
Furthermore, the Ito-Isometry holds:

$$\mathbb{E}\left[\left(\int_0^t H_s dW_s\right)^2\right] = \mathbb{E}\left[\int_0^t H_s^2 ds\right].$$

Step 2: L^2-adapted integrands

We define the space of L^2-adapted integrands as

$$\mathcal{H} := \left\{ H : H \textit{ adapted process} : \mathbb{E}\left[\int_0^t H_s^2 ds\right] < \infty \right\}.$$

Now we are able to define the Ito integral with respect to L^2-adapted integrands from $\mathcal{H}$:

Definition 2.3.2 (Ito Integral with respect to L^2-adapted integrands from $\mathcal{H}$)**.**
With the above notation we define the Ito integral for L^2-adapted integrands from $\mathcal{H}$ as

$$\int_0^t H_s dW_s := \lim_{n\to\infty} \int_0^t H_s^{(n)} dW_s \tag{2.2}$$

for a sequence $H^{(n)} \in S_0$ with $H^{(n)} \overset{n\to\infty}{\longrightarrow} H$ in L^2.

Remark 2.3.3.
Note that we have used the following assumptions in (2.2):

1. *Assume that for $H \in \mathcal{H}$ there exists $H^{(n)} \in S_0$, such that*

$$\mathbb{E}\left[\int_0^{\mathrm{t}} (\mathrm{H}^{(\mathrm{n})} - \mathrm{H})^2 \mathrm{ds}\right] \overset{n\to\infty}{\longrightarrow} 0.$$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

2. *Assume that $X^{(n)}$ is a sequence of continuous processes, such that $X_t^{(n)}$ is a Cauchy-sequence in $L^2(\mathbb{P})$. Then there exists a unique continuous process X_s, such that*

$$X_s^{(n)} \to X_s$$

in $L^2(\mathbb{P})$ for all $s \in [0,t]$.

According to the Ito-Isometry and 1., it follows that $\int_0^t H^{(n)} dW_s$ is a Cauchy-sequence in $L^2(\mathbb{P})$ and according to 2., it follows that for $H \in \mathcal{H}$ (2.2) is fulfilled.

Proposition 2.3.4 (Properties of the Ito Integral in (2.2))**.**

- *The left hand side of (2.2) is an L^2-bounded martingale*
- *Linearity: $\forall \alpha_1, \alpha_2 \in \mathbb{R}, H^{(1)}, H^{(2)} \in \mathcal{H}$:*

$$\int_0^t \left(\alpha_1 H^{(1)} + \alpha_2 H^{(2)}\right) dW_s = \int_0^t \alpha_1 H^{(1)} dW_s + \int_0^t \alpha_2 H^{(2)} dW_s$$

- *Ito-Isometry:*

$$\mathbb{E}\left[\int_0^t H_s^2 ds\right] = \mathbb{E}\left[\left(\int_0^t H_s dW_s\right)^2\right]$$

- *For a stopping time τ it holds, that*

$$\int_0^{t\wedge\tau} H_s dW_s = \int_0^t H_s \mathbb{1}_{[0,\tau]}(s) dW_s$$

Step 3: Localisation

We define the space of the almost surely L^2-adapted integrands as

$$\mathcal{J} := \left\{H : H \text{ adapted process} : \mathbb{P}\left[\int_0^t H_s^2 ds < \infty\right] = 1\right\}$$

and a stopping time as follows

$$\tau_n = \inf\left\{t > 0 : \int_0^t H_s^2 ds = n\right\}$$

with convention $\inf \emptyset = \mathbb{T}$.

Now we are able to define the Ito integral with respect to the almost surely L^2-adapted integrands from $\mathcal{J}$:

Definition 2.3.5 (Ito Integral with respect to the almost surely L^2-adapted integrands from $\mathcal{J}$)**.**

Let $H_t^{(n)} = H_t \mathbb{1}_{[0,\tau_n]}(t)$, such that $H_t^{(n)} \in \mathcal{H}$ and $H_t \in \mathcal{J}$. Then the Ito integral for almost surely L^2-adapted integrands from $\mathcal{J}$ is defined as

$$\int_0^t H_s dW_s := \lim_{n\to\infty} \int_0^t H_s^{(n)} dW_s \text{ for } t \leq \tau_n \text{ in the intervall } [0,T]. \tag{2.3}$$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Remark 2.3.6.

Note that (2.3) *is not necessarily a martingale, but a local martingale.*

Thus, the stochastic integral with respect to a Brownian motion with almost surely L^2-adapted integrands from $\mathcal{J}$ is defined. It is possible to extend our notation of the stochastic integral by replacing the Brownian motion W_t by a L^2-bounded continuous martingale M_t. Furthermore, the integrands H_s are restricted to the space of the progressively measurable processes (see Definition 2.1.9), for which we need the following definition.

Definition 2.3.7 (Quadratic Variation Process)**.**

$[M]_t = \lim\limits_{n\to\infty} \sum_{i=1}^{n}(M_{t_i} - M_{t_{i-1}})^2$ *with* $\limsup\limits_{n\to\infty} \sup\limits_{i} |t_{i+1} - t_i| = 0$.

So the Ito-Isometry is now written as follows:

$$\mathbb{E}\left[\left(\int_0^t H_s dM_s\right)^2\right] = \mathbb{E}\left[\int_0^t H_s^2 d[M]_s\right].$$

2.4 Ito Process and Ito Formula

In stochastic analysis it is common to work with so-called Ito processes, which are an essential basic concept. Another core statement in stochastics, which in its simplest form can be seen as an integral representation of stochastic processes with respect to a Brownian motion, is the so-called Ito formula.

Definition 2.4.1 (One-dimensional Ito Process)**.**

Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ *be a filtered probability space and* $(W_t)_{t\geq 0}$ *be a one-dimensional Brownian motion with respect to the filtration* $(\mathcal{F}_t)_{t\geq 0}$*. Then a process* $(X_t)_{t\geq 0}$ *of the form*

$$X_t = X_0 + \int_0^t K_s ds + \int_0^t H_s dW_s, \tag{2.4}$$

is called one-dimensional Ito process, where the properties

- X_0 *is* $\mathcal{F}_0$*-measurable*
- K_s *is adapted and* $\mathbb{P}\left[\int_0^t |K_s| ds < \infty\right] = 1$
- H_s *is adapted and* $\mathbb{P}\left[\int_0^t H_s^2 ds < \infty\right] = 1$

are fulfilled.

Remark 2.4.2.

The representation in (2.4) *is unique in the following way:*

$$X_0 + \int_0^t K_s ds + \int_0^t H_s dW_s = X_0' + \int_0^t K_s' ds + \int_0^t H_s' dW_s,$$

where the following holds:

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

- $X_0 = X_0'$ $\mathbb{P}$-*almost surely*
- $K_s = K_s'$ $(ds \otimes d\mathbb{P})$-*almost surely*
- $H_s = H_s'$ $(ds \otimes d\mathbb{P})$-*almost surely.*

Remark 2.4.3.

An example for an Ito process is the Brownian motion with $X_0 = 0$, $K_s = 0$ and $H_s = 1$.

With (2.4) we are able to state the well-known one-dimensional Ito formula, the generalized form for the chain rule:

Definition 2.4.4 (One-dimensional Ito Formula)**.**

Let X_t be a one-dimensional Ito process according to Definition 2.4.1 and $f \in C^2(\mathbb{R})$ be a function from the space of the two times continuously differentiable functions. Then $f(X_t)$ is again a one-dimensional Ito process and it holds that

$$f(X_t) = f(X_0) + \int_0^t f'(X_s)K_s ds + \int_0^t f'(X_s)H_s dW_s + \frac{1}{2}\int_0^t f''(X_s)H_s^2 ds. \tag{2.5}$$

If we adapt our notation

- $dX_s = K_s ds + H_s dW_s$
- $[X]_s = \int_0^t H_s^2 ds$
- $d[X]_s = H_s^2 ds$

then we are able to write (2.5) in short form as

$$f(X_t) = f(X_0) + \int_0^t f'(X_s)dX_s + \frac{1}{2}\int_0^t f''(X_s)d[X]_s.$$

Remark 2.4.5.

If we assume that H is an L^2-bounded continuous martingale, then $H_t^2 - [H]_t$ is a martingale. If $H_t = W_t$ then it implies that $W_t^2 - t$ is a martingale.

Now we are able to extend Definition 2.4.1 and Definition 2.4.4 to the multi-dimensional case:

Definition 2.4.6 (Multi-dimensional Ito Process)**.**

Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a filtered probability space and $W_t = (W_t^1, \dots, W_t^m)^T$ be a m-dimensional Brownian motion, where W^i, $i = \{1, \dots, m\}$, are independent Brownian motions and adapted to the filtration $\mathcal{F}_t$. Then the process

$$X_t = X_0 + \int_0^t b_s\, ds + \int_0^t \sigma_s\, dW_s, \tag{2.6}$$

is called $\mathbb{R}^n$-valued Ito process, where the integrability conditions of the one-dimensional case should be fulfilled component-wise and

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

- $b_s \in \mathbb{R}^n$ *and* $\sigma_s \in \mathbb{R}^{n\times m}$ *are adapted processes*
- $\mathbb{P}\left[\int_0^t |b_s^i| ds < \infty\right] = 1 \ \forall i \in \{1, \dots, n\}$
- $\mathbb{P}\left[\int_0^t (\sigma_s^{ij})^2 ds < \infty\right] = 1 \ \forall i, j \in \{1, \dots, n\}$.

In differential notation[3] (2.6) changes to

$$dX_t = b_t dt + \sigma_t dW_t.$$

Remark 2.4.7.

Let A be a $\mathbb{R}^{n\times m}$-valued matrix, consisting of adapted processes, then the following integral is component-wise defined as

$$\int_0^t A_s dW_s = \begin{pmatrix} \sum_{k=1}^m \int_0^t A_{1k} dW_s^k \\ \sum_{k=1}^m \int_0^t A_{2k} dW_s^k \\ \vdots \\ \sum_{k=1}^m \int_0^t A_{mk} dW_s^k \end{pmatrix}.$$

According to (2.6) we are able to state the multi-dimensional Ito formula.

Definition 2.4.8 (Multi-dimensional Ito Formula)**.**

Let X_t be a $\mathbb{R}^n$-valued Ito process according to Definition 2.4.6 and $f : \mathbb{R}^+ \times \mathbb{R}^n \to \mathbb{R} \in C^{1,2}$ be a continuous function with continuous partial derivatives $\frac{\partial f}{\partial t}$, $\frac{\partial f}{\partial i}$ and $\frac{\partial^2 f}{\partial i \partial j}$ for $i, j \in \{1, \dots, n\}$. Then the process $f(t, X_t)$ is again a $\mathbb{R}^n$-valued Ito process and it holds that

$$\begin{aligned} f(t, X_t) = f(0, X_0) &+ \int_0^t \frac{\partial f}{\partial s}(s, X_s) ds + \sum_{i=1}^n \int_0^t \frac{\partial f}{\partial i}(s, X_s) H_s dX_s^i \\ &+ \frac{1}{2} \sum_{i,j=1}^n \int_0^t \frac{\partial^2 f}{\partial i \partial j}(s, X_s) d[X^i, X^j]_s, \end{aligned} \tag{2.7}$$

where $[X^i, X^j]_s = \int_0^t (\sigma\sigma^T)_{ij} ds$ and $(\sigma\sigma^T)_{ij} = \sum_{k=1}^m \sigma_{ik}\sigma_{jk}$.

2.5 Stochastic Differential and Integral Equations

It is possible to consider an Ito process, as we have seen in Section 2.4, as a stochastic differential or integral equation, which we are going to introduce now.

[3] see Section 2.5

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Definition 2.5.1 (Stochastic Differential and Integral Equation)**.**
Let $(W_t)_{t\geq 0}$ be a $\mathbb{R}^n$-valued Brownian motion, $b : \mathbb{R}^+ \times \mathbb{R}^n \to \mathbb{R}^n$ and $\sigma : \mathbb{R}^+ \times \mathbb{R}^n \to \mathbb{R}^{n\times m}$ measurable functions and $X_0 = \xi$. An equation of the form

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t \tag{2.8}$$

is called stochastic differential equation and

$$X_t = \xi + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s$$

is called stochastic integral equation.

In the next step we want to define a solution for (2.8):

Definition 2.5.2 (Solution of the Stochastic Differential and Integral Equation)**.**
We call a process X_t solution of (2.8), if

1. *X_t fulfills* (2.7)
2. *X_t is adapted with respect to the filtration $\mathcal{F}_t$*
3. $\mathbb{P}\left[\int_0^t |b^i(s, X_s)|ds < \infty\right] = 1 \ \forall i \in \{1, \dots, n\}$
4. $\mathbb{P}\left[\int_0^t (\sigma^{ij})^2(s, X_s)ds < \infty\right] = 1 \ \forall i, j \in \{1, \dots, n\}$

Remark 2.5.3.
In Definition 2.5.2 we are distinguishing between two different types of solutions:

- *X_t is called a strong solution, if it is adapted to the natural filtration of the Brownian motion*[4]
- *X_t is called a weak solution, if the filtration is part of the solution with X_t*

Remark 2.5.4.
An Ito process X_t fulfills a stochastic differential equation with drift coefficient b_t and diffusion coefficient σ_t. For example, the geometric Brownian motion S_t, also known as exponential Brownian motion, is an Ito process and fulfills

$$dS_t = \mu S_t dt + \sigma S_t dW_t$$

with $b_t = \mu$.

[4] see Definition 2.1.10

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Assuming that $\forall t \in \mathbb{R}^+, K > 0$ the following holds

$$\|b(t,x) - b(t,y)\| + \|\sigma(t,x) - \sigma(t,y)\| \leq K\|x - y\| \text{ uniformly} \tag{2.9}$$

and that both X_t and $\hat{X}_t$ are solutions of (2.8), implies that

$$\mathbb{P}\left[X_t = \hat{X}_t \ \forall t \in \mathbb{R}^+\right] = 1,$$

which means, that X_t and $\hat{X}_t$ are indistinguishable. Inequality (2.9) is also known as Lipschitz condition.

Definition 2.5.5 (Modification)**.**
The process X_t is called a modification of $\hat{X}_t$, if the following holds

$$\mathbb{P}\left[X_t = \hat{X}_t\right] = 1 \ \forall t \in \mathbb{R}^+$$

Remark 2.5.6.
If X_t and $\hat{X}_t$ are right-continuous and X_t is a modification of $\hat{X}_t$, then it implies that X_t is indistinguishable of $\hat{X}_t$. The reversal is always valid.

Lemma 2.5.7 (Existence and Uniqueness)**.**
Let X_t be a solution of (2.8). If the coefficients b, σ fulfill

- *Lipschitz condition: $\forall t \in \mathbb{R}^+, \forall x, y \in \mathbb{R}, K > 0$:*

$$\|b(t,x) - b(t,y)\| + \|\sigma(t,x) - \sigma(t,y)\| \leq K\|x - y\| \text{ uniformly}$$

- *Linear boundedness: $\forall t \in \mathbb{R}^+, \forall x \in \mathbb{R}, K > 0$*

$$\|b(t,x)\| + \|\sigma(t,x)\| \leq K(1 + \|x\|)$$

- *X_0 has second moment:*

$$\mathbb{E}\left[X_0^2\right] < \infty \text{ with } X_0 = \xi$$

then there exists a unique strong solution of (2.8) on $\mathbb{R}^+$ with $\mathbb{E}\left[\sup_{t \in \mathbb{R}^+} |X_t|^2\right] < \infty$.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

3 Risk and Ruin Theory

This section deals with the basic principles of risk and ruin theory and is based on [13] and [15]. For a more detailed and advanced illustration, the interested reader is referred to [2] and [7], which are also used as references throughout Sections 3.1–3.3.

3.1 Risk Models

In this section we are going to consider two possible approaches how to model a sum of damage. On the one hand, the so-called individual model by considering the individual risks and, on the other hand, the so-called collective model by considering the risks as a collective depending on the total loss amount and the number of losses.

The individual risk model
In the individual model, a single risk R_i can be interpreted as damage caused by the i-th insurance policy in the period under consideration. Consequently, when considering a portfolio of n policies, the total loss of an insurance company is the sum S of the individual losses $R_1, \ldots, R_n$.

Definition 3.1.1 (Sum of damage in the individual risk model)**.**
Let $R_1, \ldots, R_n$ be the single risks of a portfolio. The total claim amount, which is also called sum of damage is then defined as

$$S := \sum_{i=1}^{n} R_i. \tag{3.1}$$

In (3.1) *the single risks R_i are often assumed to be independent and identically distributed. Thus we are considering a homogeneous portfolio and therefore we can calculate the expected value and variance of S as*

$$\mathbb{E}[S] = \sum_{i=1}^{n} \mathbb{E}[R_i] = n\mathbb{E}[R_1]$$
$$\mathbb{V}[S] = \sum_{i=1}^{n} \mathbb{V}[R_i] = n\mathbb{V}[R_1].$$

Now we are going to describe the collective risk model.

The collective risk model
In contrast to the individual risk model, the collective risk model does not focus on the individual risk, but rather the entire portfolio of risks is regarded as the generator of a random number of losses in an observation period.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Definition 3.1.2 (Sum of damage in the collective risk model)**.**

For independent and identically distributed random variables $Y_i \in \mathbb{R}^+$ describing the amount of losses and a random variable $N \in \mathbb{Z}^+$ describing the incurred number of losses the amount of total damage is defined as a random sum

$$S := \sum_{i=1}^{N} Y_i, \tag{3.2}$$

where $(Y_i)_{i\in\mathbb{N}}$ and N are assumed to be independent. With (3.2) we can calculate the expected value of S by using properties of the conditional expectation and total probability as

$$\begin{aligned}
\mathbb{E}[S] &= \mathbb{E}\big[\mathbb{E}[S \mid N]\big] \\
&= \sum_{n\geq 0} \mathbb{E}[S \mid N = n]\, \mathbb{P}[N = n] \\
&= \sum_{n\geq 0} \mathbb{E}\left[\sum_{i=1}^{N} Y_i \;\middle|\; N = n\right] \mathbb{P}[N = n] \\
&= \sum_{n\geq 0} \mathbb{E}\left[\sum_{i=1}^{n} Y_i \;\middle|\; N = n\right] \mathbb{P}[N = n] \\
&= \sum_{n\geq 0} \mathbb{E}\left[\sum_{i=1}^{n} Y_i\right] \mathbb{P}[N = n] \\
&= \sum_{n\geq 0} n\, \mathbb{E}[Y_1]\, \mathbb{P}[N = n] \\
&= \mathbb{E}[Y_1]\, \mathbb{E}[N]
\end{aligned} \tag{3.3}$$

and for calculating the variance of (3.2) we are using the properties of the conditional variance and total probability respectively

$$\begin{aligned}
\mathbb{V}[S] &= \mathbb{E}[S^2] - (\mathbb{E}[S])^2 \\
&= \mathbb{E}\big[\mathbb{E}[S^2 \mid N]\big] - \big(\mathbb{E}\big[\mathbb{E}[S \mid N]\big]\big)^2 \\
&= \mathbb{E}\left[\mathbb{V}[S \mid N] + (\mathbb{E}[S \mid N])^2\right] - \big(\mathbb{E}\big[\mathbb{E}[S \mid N]\big]\big)^2 \\
&= \mathbb{E}\left[\mathbb{V}[S \mid N]\right] + \mathbb{E}\left[(\mathbb{E}[S \mid N])^2\right] - \big(\mathbb{E}\big[\mathbb{E}[S \mid N]\big]\big)^2 \\
&= \mathbb{E}\left[\mathbb{V}[S \mid N]\right] + \mathbb{V}\left[\mathbb{E}[S \mid N]\right]
\end{aligned}$$

Note that $\mathbb{V}[S]$ can also be calculated as

$$\begin{aligned}
\mathbb{V}[S] &= \mathbb{E}\left[\mathbb{V}[S \mid N]\right] + \mathbb{V}\left[\mathbb{E}[S \mid N]\right] \\
&= \mathbb{E}\big[N\, \mathbb{V}[Y_1]\big] + \mathbb{V}\big[N\, \mathbb{E}[Y_1]\big] \\
&= \mathbb{E}[N]\, \mathbb{V}[Y_1] + (\mathbb{E}[Y_1])^2\, \mathbb{V}[N]
\end{aligned} \tag{3.4}$$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Remark 3.1.3.

If no damage occurs so the total damage is trivially also zero, which means that $N = 0$ implies $S = 0$.

Now we are going to calculate the moment generating function $M_S(\theta)$ of the total claim amount S:

$$\begin{aligned}
M_S(\theta) &= \mathbb{E}[\exp(\theta\, S)] = \mathbb{E}\left[\mathbb{E}\left[\exp(\theta\, S) \,\middle|\, N\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\exp\left(\theta \sum_{i=1}^{N} Y_i\right) \,\middle|\, N\right]\right] \\
&= \sum_{k\geq 0} \mathbb{E}\left[\exp\left(\theta \sum_{i=1}^{N} Y_i\right) \,\middle|\, N = n\right] \mathbb{P}[N = k] \\
&= \sum_{k\geq 0} \mathbb{E}\left[\prod_{i=1}^{k} \exp(\theta Y_i)\right] \mathbb{P}[N = k] \\
&= \sum_{k\geq 0} \prod_{i=1}^{k} \mathbb{E}[\exp(\theta Y_i)]\, \mathbb{P}[N = k] \\
&= \sum_{k\geq 0} \left(\mathbb{E}[\exp(\theta Y)]\right)^k \mathbb{P}[N = k] \\
&= \sum_{k\geq 0} M_Y(\theta)^k\, \mathbb{P}[N = k] \\
&= \sum_{k\geq 0} \exp(\ln(M_Y(\theta)^k))\, \mathbb{P}[N = k] \\
&= M_N(\ln(M_Y(\theta)))
\end{aligned} \tag{3.5}$$

Remark 3.1.4 (Distribution of N)**.**

The number of damages N can be modeled by using, for instance, the following types of distributions:

- *Poisson:* $N \sim Poi(\lambda) \;\Rightarrow \mathbb{E}[N] = \lambda,\; \mathbb{V}[N] = \lambda$
- *Binomial:* $N \sim Bin(m, p) \;\Rightarrow \mathbb{E}[N] = mp,\; \mathbb{V}[N] = mp(1-p)$
- *Negative Binomial:* $N \sim NBin(\alpha, p) \;\Rightarrow \mathbb{E}[N] = \frac{\alpha\,(1-p)}{p},\; \mathbb{V}[N] = \frac{\alpha\,(1-p)}{p^2}$

A Poisson distributed loss number, $N \sim Poi(\lambda)$ is often used in collective models. The distribution of the total loss S is then given by the compound Poisson distribution:

Definition 3.1.5 (Compound Poisson Distribution)**.**

Let N be Poisson distributed with parameter λ, meaning that $N \sim Poi(\lambda)$ and $F_i(\cdot)$ be the distribution function of Y_i, the i-th level of damage. Then the distribution of the total claim amount S is denoted by

$$S \sim CP(\lambda, F(\cdot))$$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

and is called compound Poisson distribution, for which the expected value, variance and moment generating function are calculated from (3.3), (3.4) and (3.5) *as*

$$\mathbb{E}[S] = \lambda\,\mathbb{E}[Y] \quad \mathbb{V}[S] = \lambda\,\mathbb{E}[Y^2] \quad M_S(\theta) = \exp(\lambda\,(M_Y(\theta) - 1)),$$

where $M_Y(\theta) = \exp(\lambda(\exp(\theta) - 1))$ *is the moment generating function of* Y.

3.2 Cramer Lundberg Model

In Section 3.1 we considered the total claim amount S for a given time interval, for instance the amount of damages of a portfolio in one year. Based on the collective risk model, in this section we will follow a more dynamic approach by considering a longer-term horizon and looking at the development of the surplus (profit) of an insurance company. In analogy to the calculation of the total loss in a fixed period of time, where the number of losses was modeled by a Poisson distributed random variable, a so-called Poisson process is now considered, taking into account the time aspect.

Definition 3.2.1 (Total Claims Process)**.**
Let $(N_t)_{t\geq 0}$ *be the number of claims process and* X_n *the amount of damage of the* n*-th claim. Then the total claims process* $(S_t)_{t\geq 0}$ *is defined as*

$$S_t = \sum_{n=1}^{N_t} X_n \quad t \geq 0,$$

where the processes $(N_t)_{t\geq 0}$ *and* $(S_t)_{t\geq 0}$ *are stochastic processes in continuous time.*

Now we are assuming that $(N_t)_{t\geq 0}$ is a homogeneous Poisson process with parameter $\lambda > 0$ and therefore $(S_t)_{t\geq 0}$ is a compound Poisson process according to Definition 3.1.5. The homogeneous Poisson process with parameter $\lambda > 0$ fulfills

- $N_0 = 0$
- $(N_t)_{t\geq 0}$ has right-continuous paths, namely step functions with values in $\mathbb{Z}^+$ and jumps of height 1
- independent increments, namely for $0 \leq t_0 < t_1 < \ldots < t_m$ and $m \in \mathbb{N}$ the increments $N_{t_1} - N_{t_0}, \ldots, N_{t_m} - N_{t_{m-1}}$ are independent
- stationary increments, $N_{t+h} - N_t \sim Poi(h\lambda) \quad \forall h > 0, t \geq 0$ which means that the distribution only depends on the length of the interval and not on the position

According to (3.3) and (3.4) we can denote the moments of $(S_t)_{t\geq 0}$ as

$$\mathbb{E}[S_t] = \lambda t\,\mathbb{E}[X] \quad \mathbb{V}[S_t] = \lambda t\,\mathbb{E}[X^2] \quad \forall t \geq 0 \tag{3.6}$$

The Cramer Lundberg model
For the Cramer Lundberg model we first have to model the premium income c_t. Lundberg assumed that the premium income is continuous over time with a constant premium rate

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

$c > 0$, and for each time interval is proportional to the interval length, which means that up to time t we have ct. Therefore the premium income for the interval $[0, t]$ is defined as

$$c_t = ct \quad t \geq 0.$$

This leads to the following model, known as Cramer Lundberg model or classical risk model, where the risk reserve process U_t of an insurance portfolio describing the amount of money of the company at time t is defined as

$$U_t = u + ct - S_t = u + ct - \sum_{k=1}^{N_t} X_k, \tag{3.7}$$

where

- u is the original capital at time $t = 0$
- c is the premium rate over the interval $(0, t]$
- S_t is the total claims process
- N_t is a homogeneous Poisson process with parameter $\lambda > 0$
- $(X_k)_{k\geq 1}$ is a sequence of independent and identically distributed positive random variables, which are independent of N_t.

We are speaking of the (technical) ruin of an insurance company as soon as (3.7) becomes negative, which means that damages have occurred and that the total claims process exceeds the premium income and therefore the reserve is no longer sufficient. Accordingly, the company is interested in keeping the probability of this event occurring as low as possible. A very frequently used measure of risk is the so-called probability of ruin, for which we need the definition of the ruin time.

Definition 3.2.2 (Ruin Time)**.**
The ruin time T is defined as

$$T := \inf\{t \geq 0 : U_t < 0\},$$

with convention $\inf \emptyset = \infty$.

Definition 3.2.3 (Ruin and Survival Probability)**.**
The ruin probability $\Psi(u)$ over an infinite time horizon is defined as

$$\Psi(u) := \mathbb{P}[T < \infty] = \mathbb{P}\left[\inf_{t\geq 0} U_t < 0 \mid U_0 = u\right]$$

and therefore the survival probability $\Phi(u)$ is defined as

$$\Phi(u) := \mathbb{P}[T = \infty] = 1 - \Psi(u).$$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Remark 3.2.4 (Finite Time Horizon)**.**
The ruin probability $\Psi(u,t)$ over a finite time horizon is defined as

$$\Psi(u,t) := \mathbb{P}[T < t].$$

The calculation and estimation of the ruin probability, which serves as an indicator for the solvency of the insurance company, is one of the central problems of actuarial mathematics. There are basically three methods for calculating $\Psi(u)$:

- Integral equation
- Renewal theory and
- Martingales

In case of renewal theory, we have to reconsider our model in (3.7) but now N_t is a renewal process defined as

$$N_t = \sum_{n\geq 1} \mathbb{1}_{\{T_n \leq t\}} = \sup\{n \geq 0 : T_n \leq t\} = |\{T_n : T_n \leq t\}|, \tag{3.8}$$

where

- $T_n = \sum\limits_{k=1}^{n} W_k$ are the arrival times with $T_0 = 0$ and
- $(W_k)_{k\geq 0}$ is a sequence of independent and identically exponentially distributed random variables, $W_i \sim Exp(\lambda)$, which are called waiting times.

Therefore we can interpret (3.8) as number of jumps that have occurred by time t. Instability is implied with a high $\Psi(u)$, which in this case can be reduced by, for example, adjusting the premium level c or the initial capital u. So that ruin does not occur with certainty, we demand in the following that the premium rate c contains a safety margin $\theta \in \mathbb{R}$ of the form

$$c = (1+\theta)\,\lambda\mu \quad \theta > 0 \tag{3.9}$$

which is equivalent to

$$\theta = \frac{c}{\lambda\mu} - 1 > 0 \tag{3.10}$$

and thus always satisfies

$$c > \lambda\mu. \tag{3.11}$$

Remark 3.2.5.
Note that (3.9) *is motivated by the expected value principle for a premium over the interval* $[0,t]$*:*

$$ct = (1+\theta)\mathbb{E}[S_t] = (1+\theta)\mathbb{E}[X]\mathbb{E}[N_t] = (1+\theta)\mu\lambda t$$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

The parameter in (3.10) is called relative safety margin or safety loading and the inequality in (3.11) is known as net profit condition.

Remark 3.2.6.

If the net profit condition in not fulfilled, then ruin occurs with certainty for both cases:

$$\begin{aligned} c < \lambda\mu \quad &\Rightarrow \quad \Psi(x) = 1 \\ c = \lambda\mu \quad &\Rightarrow \quad \Psi(x) = 1, \end{aligned}$$

where the last equality can be shown by the theorem of the iterated logarithm.

Note, that we are from now on assuming that the net profit condition in (3.11) is always fulfilled.

Definition 3.2.7 (Cramer Lundberg Coefficient)**.**

If there exists $R > 0$, when considering the Cramer Lundberg Model as in (3.7)*, such that the moment generating function of X fulfills $M_X(R) = \mathbb{E}[\exp(RX)] < \infty$ and*

$$\lambda + cR = \lambda M_X(R), \tag{3.12}$$

then this R is called Cramer Lundberg coefficient or adjustment coefficient and the following estimation, known as Cramer Lundberg inequality holds

$$\Psi(x) \leq \exp(-Rx) \quad \forall x \geq 0. \tag{3.13}$$

Note that (3.12) is called Cramer Lundberg Equality, which is not always solvable, and is motivated by the exponential principle for a premium over the interval $[0, t]$:

$$ct = \frac{1}{R} \ln M_{S_t}(R) = \frac{1}{R}\lambda t(M_X(R) - 1),$$

which is equivalent to

$$c = \frac{\lambda}{R}(M_X(R) - 1) \Leftrightarrow \frac{c}{\lambda}R = M_X(R) - 1 \Leftrightarrow M_X(R) = 1 + \frac{c}{\lambda}R$$

where we used that

$$M_{S_t}(z) = M_{N_t}(z)(\ln M_X(z)) = \exp(\lambda t(M_X(t) - 1)).$$

Now we want to prove the Cramer Lundberg inequality in (3.13). Therefore, let

$$\Psi_n(x) = \mathbb{P}[\text{Ruin at } n\text{-th claim or earlier}] = \mathbb{P}\left[\min_{k\in\{1,\dots,n\}} \{x + cT_k - Y_k\} < 0\right],$$

where

- $U_t = x + ct - \sum_{k=1}^{N_t} X_k$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

- $Y_k = \sum_{j=1}^{k} X_j$
- $T_k = \sum_{j=1}^{k} W_j$
- $(W_j)_{j\geq 1} \sim Exp(\lambda)$ independent and identically exponentially distributed
- $(X_j)_{j\geq 1} \sim F(\cdot)$ independent and identically distributed
- $(W_j)_{j\geq 1}$ and $(X_j)_{j\geq 1}$ are independent.

For the proof of the Cramer Lundberg inequality we are considering $x \in \mathbb{R}$ and do not need the condition $x \geq 0$. Moreover, we have

$$\Psi_0(x) = \begin{cases} 0 & x \geq 0 \\ 1 & x < 0. \end{cases} \tag{3.14}$$

We will show this by an induction proof and the idea behind it is that we condition on the first time $T_1 = t$ and on the amount of the first damage $X_1 = y$. So for the case that $n = 0$ we get

$$\Psi_0(x) \leq \exp(-Rx) \quad x \in \mathbb{R}, \tag{3.15}$$

which is true because of (3.14). Now we consider the case $n \geq 1$ and therefore we get

$$\begin{aligned} \Psi_n(x) &= \mathbb{P}\left[\min_{k\in\{1,\dots,n\}} \{x + cT_k - Y_k\} < 0\right] \\ &= \int_{t=0}^{\infty} \left[\int_{y=-\infty}^{x+ct} A\, dF(y) + \int_{y=x+ct}^{\infty} A\, dF(y)\right] \lambda \exp(-\lambda t)\, dt \end{aligned}, \tag{3.16}$$

where

$$A := \mathbb{P}\left[\min_{k\in\{1,\dots,n\}} \{x + cT_k - Y_k\} < 0 \mid T_1 = t,\, X_1 = y\right], \tag{3.17}$$

where we have split the integral into two areas, the first area where no ruin occurs in the first damage, and the second area where ruin occurs. If we use the following

$$x + cT_k - Y_k = x + cT_1 - Y_1 + c(T_k - T_1) - (Y_k - Y_1),$$

with

$$T'_{k-1} := T_k - T_1 = \sum_{j=2}^{k} W'_j, \quad W'_j = W_{j+1}$$

$$Y'_{k-1} := Y_k - Y_1 = \sum_{j=1}^{k-1} X'_j, \quad X'_j = X_{j+1},$$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

which have the same properties as $(X_j)_{j\geq 1}$ and $(W_j)_{j\geq 1}$, we get under the assumption that $y \leq x + ct$ that (3.17) simplifies to

$$\begin{aligned} A &= \mathbb{P}\left[\min_{k\in\{1,\dots,n\}}\{x + cT_k - Y_k\} < 0 \mid T_1 = t,\, X_1 = y\right] \\ &= \mathbb{P}\left[\min_{k\in\{1,\dots,n\}}\{x + ct - y + cT'_k - Y'_k\} < 0 \mid T_1 = t,\, X_1 = y\right] \\ &= \Psi_{n-1}(x + ct - y) \end{aligned} \tag{3.18}$$

because of the fact, that we can omit the conditions in the conditional probability due to independence. Under the assumption that $y > x + ct$ we get

$$\begin{aligned} A &= \mathbb{P}\left[\min_{k\in\{1,\dots,n\}}\{x + cT_k - Y_k\} < 0 \mid T_1 = t,\, X_1 = y\right] \\ &= 1 \\ &= \Psi_{n-1}(x + ct - y). \end{aligned} \tag{3.19}$$

Therefore we get in (3.16) by using (3.18) and (3.19)

$$\begin{aligned} \Psi_n(x) &= \int_{t=0}^{\infty}\left[\int_{y=-\infty}^{\infty}\Psi_{n-1}(x + ct - y)\lambda\exp(-\lambda t)\,dF(y)\right]dt \\ &\overset{(3.15)}{\leq} \int_{t=0}^{\infty}\left[\int_{y=-\infty}^{\infty}\exp(-R(x + ct - y))\,\lambda\exp(-\lambda t)\,dF(y)\right]dt \\ &= \exp(-Rx)\int_{t=0}^{\infty}\lambda\exp(-(\lambda + Rc)t)\,dt\int_{y=-\infty}^{\infty}\exp(Ry)\,dF(y) \\ &= \exp(-Rx)\cdot\frac{\lambda}{\lambda + Rc}\cdot M_X(R) \\ &= \exp(-Rx)\cdot\frac{\lambda}{\lambda + Rc}\cdot\frac{\lambda + Rc}{\lambda} \\ &= \exp(-Rx). \end{aligned}$$

So we have shown that the Cramer Lundberg inequality is valid for all $x \in \mathbb{R}$, which implies that it is valid for all $x \geq 0$ and thus (3.13) holds. Note that we have used the strong Markov property for T_1 and the fact that the ruin process renews itself with the first damage.

Remark 3.2.8.

In conclusion, we obtain the following for the probability of ruin with an initial capital x at time t

$$\Psi(x) = \begin{cases} \int_{t=0}^{\infty}\left[\int_{y=-\infty}^{\infty}\Psi(x + ct - y)\lambda\exp(-\lambda t)\,dF(y)\right]dt & x \geq 0 \\ 1 & x < 0 \end{cases}$$

and for the survival probability $\Phi = 1 - \Psi$

$$\Phi(x) = \begin{cases} \int_{t=0}^{\infty}\left[\int_{y=-\infty}^{x+ct}\Phi(x + ct - y)\lambda\exp(-\lambda t)\,dF(y)\right]dt & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

We get a trivial solution of the Cramer-Lundberg equality for $R = 0$. The left hand side of (3.12) is an affine function, while the right hand side is convex. Moreover, the net profit condition implies that the derivative of the left hand side of (3.12) exceeds the derivative of the right hand side at $R = 0$, and thus the Cramer Lundberg coefficient exists if $M_X(R)$ exists for all $R > 0$.

Theorem 3.2.9 (Upper and Lower Bound for the Adjustment Coefficient)**.**

The adjustment coefficient fulfills:

1. *If $R > 0$ exists, then it holds that*

$$0 < R < 2\theta \frac{\mathbb{E}[X]}{\mathbb{E}[X^2]}.$$

2. *If X is bounded with $\mathbb{P}[X \leq b] = 1$ for $b > 0$, then it holds that:*

$$\exists R > 0 : R > \frac{1}{b} \ln(1 + \theta).$$

Theorem 3.2.10.

If an adjustment coefficient $R > 0$ exists, the it holds that

$$\Psi(x) = \frac{\exp(-Rx)}{\mathbb{E}[\exp(-RU_T) \mid T < \infty]} \quad \forall x \geq 0.$$

Remark 3.2.11.

If we consider our model in (3.7) *without starting capital, then the probability of ruin can be calculated as follows*

$$\Psi(0) = \frac{1}{1 + \theta}$$

where θ is denoting the relative safety margin.

3.3 Diffusion Approximation

This section deals with an approximation of the Cramer Lundberg model. When we are considering a large portfolio, we can assume that the amount of damages are relatively small and reach a high frequency. In this case the risk reserve process in (3.7) can be approximated by a Brownian motion with drift, see Definition 2.2.13.

A possible approximation method for this case is the diffusion approximation and the idea behind it is to match the two first moments of the Brownian motion with drift, and then to find that such an approximation implies in particular that the probabilities of the first pass are close together [2]. We take the content of this section from [2].

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Definition 3.3.1 (Random walk).

A random walk $(S_n^\star)_{n\geq 0}$ is a stochastic process in discrete time, in particular a family of random variables with finite index set.

The basis for the approximation is Donsker's theorem, which is a functional extension of the classical central limit theorem, and is formulated as follows:

Theorem 3.3.2 (Donsker's Theorem).

Let $(S_n^\star)_{n\geq 0}$ be a random walk in discrete time with drift $\mu = \mathbb{E}[S_1^\star]$ and variance $\sigma^2 = \mathbb{V}[S_1^\star]$, then it holds that

$$\left(\frac{1}{\sigma\sqrt{k}}\left(S_{\lfloor tk\rfloor}^\star - tk\mu\right)\right)_{t\geq 0} \xrightarrow{\mathcal{D}} (B_0(t))_{t\geq 0}, \quad k\to\infty, \tag{3.20}$$

where

- $S_n^\star = \sum_{i=1}^{n} X_i$ *for independent and identically distributed random variables X_i with mean 0 and variance 1*
- $(B_\zeta(t))_{t\geq 0}$ *is the Brownian motion with drift ζ and variance 1*
- $\xrightarrow{\mathcal{D}}$ *denotes the weak convergence.*

Definition 3.3.3 (Weak Convergence).

A sequence $(X_n)_{n\geq 1}$ of real-valued random variables converges weakly, denoted by $X_n \xrightarrow{\mathcal{D}} X$, if

$$\lim_{n\to\infty} F_n(x) = F(x) \quad \forall x\in\mathbb{R} \text{ at which } F \text{ is continuous},$$

where F_n and F are denoting the cumulative distribution functions of X_n and X, respectively.

We want to transfer the result in (3.20) to continuous random walks to get an approximation for our risk reserve process in (3.7). For this purpose we want to assume that the relative safety margin in (3.10) is small and positive and that the term $tk\mu$ in (3.20) is inconvenient. To satisfy the condition we consider a family of claim surplus processes $(S_t^{(c)})_{t\geq 0}$, where c represents the fixed premium rate, such that the claim size distribution $F_X(\cdot)$ and frequency intensity λ are the same for all c, meaning that

$$S_t^{(c)} = \sum_{i=1}^{N_t} X_i - ct.$$

Furthermore, we consider the limit $c\to\rho$ where ρ is defined as the critical premium rate $\lambda\mu_{F_X}$.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Theorem 3.3.4.

As $c \to \rho$, we have

$$\left(\frac{|\mu|}{\sigma^2} S^{(c)}_{t\sigma^2/\mu^2}\right)_{t\geq 0} \xrightarrow{\mathcal{D}} (B_{-1}(t))_{t\geq 0},$$

where

- $\mu = \mu_c = \rho - c$
- $\sigma^2 = \mathbb{E}[X^2]$.

Proof.

First of all, it should be noted that

$$\left(\frac{1}{\sigma\sqrt{k}}\left(S^{(c)}_{\lfloor tk\rfloor} - tk\mu_c\right)\right)_{t\geq 0} = \left(\frac{1}{\sigma\sqrt{k}}\left(S^{(\rho)}_{\lfloor tk\rfloor} - tk\rho\right)\right)_{t\geq 0} \xrightarrow{\mathcal{D}} (B_0(t))_{t\geq 0}, \tag{3.21}$$

whenever $k = k_c \to \infty$ as $c \to \rho$. This is a consequence of (3.20) with $S^\star_n = S^{(c)}_n$ and the inequalities

$$S^{(\rho)}_{nk} - \rho k \;\leq\; S^{(\rho)}_t \;\leq\; S^{(\rho)}_{(n+1)k} + \rho k, \quad nk \leq t \leq (n+1)k \tag{3.22}$$

To show these inequalities, let $u, v \geq 0$, such that

$$S_{u+v} \geq S_u - \rho v$$

When there are no arrivals in $(u, u+\rho v]$, then $S_{u+v} - S_u$ attains its minimal value $-\rho v$. In particular, if $t = nk + v$ with $0 \leq v \leq k$, then

$$S_t = S_{nk+v} \geq S_{nk} - \rho v \geq S_{nk} - \rho k$$

Then right inequality in (3.22) is proved similarly.
Letting $k = \frac{\sigma^2}{\mu_c^2} = \frac{\sigma^2}{(\rho-p)^2}$, then we get for (3.21)

$$\left(\frac{|\mu|}{\sigma^2} S^{(c)}_{t\sigma^2/\mu^2} + t\right) \xrightarrow{\mathcal{D}} (B_0(t)),$$
$$\left(\frac{|\mu|}{\sigma^2} S^{(c)}_{t\sigma^2/\mu^2}\right) \xrightarrow{\mathcal{D}} (B_0(t) - t) = (B_{-1}(t)),$$

where

$$\left(\frac{|\mu|}{\sigma^2} S^{(c)}_{t\sigma^2/\mu^2} + t\right) = \left(\frac{|\mu|}{\sigma^2}\left(S^{(c)}_{t\sigma^2/\mu^2} + \frac{\sigma^2}{|\mu|}t\right)\right) = \left(\frac{1}{\sigma\sqrt{\frac{\sigma^2}{|\mu|}}}\left(S^{(c)}_{t\sigma^2/\mu^2} - t\frac{\sigma^2}{\mu^2}(\rho - p)\right)\right).$$

□

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

In conclusion, we have shown the relationship of the excess process in the compound Poisson model and a Brownian motion with drift. Therefore, let the excess process be modeled by a compound Poisson process

$$U_t = u - S_t,$$

with

$$S_t = \sum_{n=1}^{N_t} X_n - ct$$

In addition, we now want to look at another risk process

$$\hat{U}_t = u + \mu t + \sigma W_t,$$

where W_t is a standard Brownian motion, according to Definition 2.2.10. In the next step we want to calculate the first two moments μ and σ of the process $\hat{U}_t$, such that its moments match the moments of the process U_t, whose moments we already know according to (3.6):

$$\begin{aligned}
\mathbb{E}[U_t] &= \mathbb{E}[u - S_t] = \mathbb{E}\left[u - \sum_{n=1}^{N_t} X_n + ct\right] = u + ct - \lambda t\,\mathbb{E}[X] \\
\mathbb{E}[\hat{U}_t] &= \mathbb{E}[u + \mu t + \sigma W_t] = u + \mu t
\end{aligned}$$

$$\begin{aligned}
\mathbb{V}[U_t] &= \lambda t\,\mathbb{E}[X^2] \\
\mathbb{V}[\hat{U}_t] &= \sigma^2 t
\end{aligned}$$

We derive the desired parameters by setting the two processes equal:

$$\begin{aligned}
\mathbb{E}[U_t] = \mathbb{E}[\hat{U}_t] \;&\Leftrightarrow\; u + ct - \lambda t\,\mathbb{E}[X] = u + \mu t \;\Leftrightarrow\; \mu = c - \lambda\mathbb{E}[X] \\
\mathbb{V}[U_t] = \mathbb{V}[\hat{U}_t] \;&\Leftrightarrow\; \lambda t\,\mathbb{E}[X^2] = \sigma^2 t \;\Leftrightarrow\; \sigma^2 = \lambda\mathbb{E}[X^2].
\end{aligned}$$

In summary, we can now approximate a process in the classical Cramer Lundberg model using a Brownian motion with drift, which will be used in the following chapters in this paper. For this representation, control methods are introduced in Section 4.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

4 Stochastic Control Theory

Stochastic control theory deals with the optimization of systems represented by stochastic processes. In this chapter we are going to define basic principles that define a continuous-time stochastic optimization problem based on [8] and [23]. For a more detailed and advanced illustration, the interested reader is referred to [6], [21] and [29], which are also used as references throughout Sections 4.1–4.3.

4.1 The Stochastic Control Problem

Let $(\Omega, \mathbb{F}, \mathbb{P})$ be a stochastic basis, where the filtration $\mathbb{F}$ is assumed to be right-continuous and complete, meaning that it satisfy the usual hypotheses (see Section 2.1). Furthermore, let $W_t = (W_t^1, \dots, W_t^m)^T$ be a m-dimensional Brownian motion with respect to $\mathbb{F}$ and $T > 0$ the maturity.

Now we are going to introduce some general terms from control theory, which are indispensable when we are considering a stochastic optimization problem:

- **Control process** $(u_t)_{t\in[0,T]}$:
  The control process is a progressively-measurable stochastic process with values in $U \subset \mathbb{R}^p, p \in \mathbb{N}$. The number of shares in a portfolio, for example, represents such a control process.

- **State process** X_t:
  The state process is described by the following stochastic differential equation, also known as diffusion equation,

  $$dX_t = b(t, X_t, u_t)\, dt + \sigma(t, X_t, u_t)\, dW_t \quad t \in [0,T] \quad \text{with } X_0 = x, \tag{4.1}$$

  where

  $$b(\cdot,\cdot,\cdot) : [0,T] \times \mathbb{R}^n \times U \to \mathbb{R}^n \text{ is the drift coefficient and}$$
  $$\sigma(\cdot,\cdot,\cdot) : [0,T] \times \mathbb{R}^n \times U \to \mathbb{R}^{n\times m} \text{ is the diffusion coefficient}$$

  of the diffusion process and both are Borel-measurable. An example of a state process is the investor's wealth using the strategy u_t.

- **Objective functional** $J(t, x, u_t)$:
  The objective functional is given by

  $$J(t, x, u_t) = \mathbb{E}\left[\int_t^T \phi(s, X_s, u_s)ds + \psi(T, X_T) \mid X_t = x\right], \tag{4.2}$$

  where the function ϕ describes the running costs or the running benefit and $\psi(T, X_T)$ the final costs or the final benefit at time T. The goal is to maximize $J(t, x, u_t)$.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

- **Admissible Control Processes/Strategies** $\mathcal{A}(t,x)$:
  The set of admissible strategies should fulfill certain conditions:
  1. The stochastic differential equation of the form (4.1), which describes the state process, should have a unique strong solution.
  2. The objective functional $J(t,x,u_t)$ shall be well defined.
- **Value function** $V(t,x)$:
  The value function is given by

$$V(t,x) = \sup_{u \in \mathcal{A}(t,x)} J(t,x,u) \tag{4.3}$$

The aim is now to determine the value function $V(0,x)$ for a giving starting value $X_0 = x$ and to determine an optimal control process $u^\star \in \mathcal{A}(t,x)$, the so-called optimal strategy, in such a way that the objective functional in (4.2) are maximized, therefore the following equations apply:

$$J(t,x,u^\star) = V(t,x)$$
$$J(0,x,u^\star) = V(0,x)$$

Nevertheless for some problems it is common to optimize by minimizing the objective functional. Such a problem would be for example the minimization of the probability of ruin. In order to determine the optimal strategy and the objective functional, it is useful to reformulate the problem.

In control theory we try to control a problem or a state, described by the state process (X_t), by using a control process $(u_t)_{t\in[0,T]}$.

4.2 Dynamic Programming

Now definitions and theorems will be presented, which are needed to derive the so-called Hamilton-Jacobi-Bellman equation. With this equation the value function $V(t,x)$ and the optimal strategy $u^\star \in \mathcal{A}(t,x)$ can be determined. The aim of this chapter is to find a partial differential equation for $V(t,x)$. Therefore, the dynamic programming principle has to be introduced. For this purpose we first have to summarize properties of Ito diffusions and their generators.

4.2.1 Ito Diffusions and their Generators

We are again considering a probability space $(\Omega, \mathbb{F}, \mathbb{P})$ with a m-dimensional Brownian motion $W_t = (W_t^1, \dots, W_t^m)^T$ with respect to the filtration $\mathbb{F}$, which is right-continuous and complete. To go further in detail we need the stochastic differential equation, defined as

$$dX_t = b(t,X_t)\,dt + \sigma(t,X_t)\,dW_t \quad t \in [0,T],$$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

with Borel-measurable functions $b(\cdot,\cdot) : \mathbb{R}^+ \times \mathbb{R}^n \to \mathbb{R}^n$ and $\sigma(\cdot,\cdot) : \mathbb{R}^+ \times \mathbb{R}^n \to \mathbb{R}^{n\times m}$ for $t \in [0,T]$. Furthermore, the drift coefficient and the diffusion coefficient satisfy a uniform Lipschitz condition: $\exists K > 0\, \forall X, Y \in \mathbb{R}^n$:

$$\|b(t,X) - b(t,Y)\| + \|\sigma(t,X) - \sigma(t,Y)\| \leq K\|X - Y\|$$

and the linear growth condition: $\exists K > 0\, \forall X \in \mathbb{R}^n$:

$$\|b(t,X)\| + \|\sigma(t,X)\| \leq K(1 + \|X\|).$$

These conditions imply the existence of a unique strong solution X_t, which is called Ito diffusion. The corresponding matrix defined as

$$a(t,x) = \sigma(t,x)\sigma(t,x)^T$$

is called diffusion matrix of X_t. We will use the following notation for a random variable Y:

$$\begin{aligned} \mathbb{E}_{t,x}[Y] &:= \mathbb{E}[Y \mid X_t = x] \\ \mathbb{E}_x[Y] &:= \mathbb{E}_{0,x}[Y] = \mathbb{E}[Y \mid X_0 = x] \end{aligned}$$

Definition 4.2.1 (Time-homogeneous Ito Diffusion)**.**

A time-homogeneous Ito diffusion is a stochastic process $X_t(\omega) = X(t,\omega) : \mathbb{R}^+ \times \Omega \to \mathbb{R}^n$, which is a solution of the stochastic differential equation

$$dX_t = b(X_t)\, dt + \sigma(X_t)\, dW_t \quad t \in [0,T],$$

where the coefficients $b(x), \sigma(x)$ fulfill the Lipschitz condition and the linear growth condition, in which case they are simplified to

$$\|b(X) - b(Y)\| + \|\sigma(X) - \sigma(Y)\| \leq K\|X - Y\| \quad X, Y \in \mathbb{R}^n.$$

Theorem 4.2.2 (Markov Property for Time-homogeneous Ito Diffusions)**.**

Let X_t be a time-homogeneous Ito diffusion, $f : \mathbb{R}^n \to \mathbb{R}$ be a bounded, Borel-measurable function, $\tau < \infty$ a stopping time and $t, s \geq 0$. Then it holds that

1. ***Markov property**: $\mathbb{E}[f(X_{t+s}) \mid \mathcal{F}_t] = \mathbb{E}_{X_t}[f(X_s)]$*
2. ***Strong Markov property**: $\mathbb{E}[f(X_{\tau+s}) \mid \mathcal{F}_\tau] = \mathbb{E}_{X_\tau}[f(X_s)]$*

This means that the behavior of the process in the future depends only on the present and not on the past. The strong Markov property is often used for hitting times. When the level is hit, the process is restarted, for instance $\tau = \inf\{t : X_t = 1\}$.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Definition 4.2.3 (Generator of X).
The infinitesimal generator L of X is defined as

$$Lf(s,x) := \lim_{t \downarrow s} \frac{1}{t-s} \mathbb{E}_{s,x}[f(t,X_t) - f(s,x)] \quad \forall s \geq 0, x \in \mathbb{R}^n \tag{4.4}$$

and f in the domain $\mathcal{D}_L$ of L which is the class of functions $f : \mathbb{R}^+ \times \mathbb{R}^n \to \mathbb{R}$ for which the limit exists for all s, x, i.e.

$$\mathcal{D}_L := \left\{ f : \mathbb{R}^+ \times \mathbb{R}^n \to \mathbb{R} : (4.4)\ \exists \right\}$$

Definition 4.2.4 (Partial Differential Operator).
The partial differential operator $\mathcal{L}$ is defined as

$$\mathcal{L} := \frac{\partial}{\partial t} + \sum_{i=1}^{n} b_i(t,x) \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^{n} a_{ij}(t,x) \frac{\partial^2}{\partial x_i \partial x_j} \tag{4.5}$$

The partial differential operator can by applied to functions f in

$$\mathcal{D}\mathcal{L} = C^{1,2} := \left\{ f : \mathbb{R}^+ \times \mathbb{R}^n \to \mathbb{R} : \frac{\partial f}{\partial t}, \frac{\partial^2}{\partial x_i \partial x_j} \text{ are continuous } \forall i,j \in \{1,\dots,n\} \right\},$$

which leads to

$$\begin{aligned} \mathcal{L}f(t,x) &= f_t(t,x) + \sum_{i=1}^{n} f_{x_i}(t,x) b_i(t,x) + \frac{1}{2} \sum_{i,j=1}^{n} a_{ij}(t,x) f_{x_i,x_j}(t,x) \\ &= f_t(t,x) + (D_x f(t,x)) b(t,x) + \frac{1}{2} tr((D_{xx} f(t,x)) a(t,x)), \end{aligned} \tag{4.6}$$

where

- $D_x f$ is the gradient of f
- $D_{xx} f$ is the Hessian matrix of f and
- $tr(A)$ is the trace of the matrix A, which is defined as the sum of the diagonal elements of A.

Remark 4.2.5.

- *If f is smooth enough, then Ito's formula in* (2.7) *can be written by using the notation of* (4.5) *as*

$$df(t,X_t) = \mathcal{L}f(t,X_t)\,dt + D_x f(t,X_t)\sigma(t,X_t)\,dW_t$$

- *In the simplest case: For f time-independent and $n = 1$:*

$$\mathcal{L}f = b(x)f'(x) + \frac{1}{2}\sigma^2 f''(x)$$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Now there is the question, under which conditions the generator L is identical to the operator $\mathcal{L}$.

Theorem 4.2.6.

Suppose that $f \in C^{1,2}$, $0 \leq t \leq T$ and that

$$\mathbb{E}_{t,x}\left[\int_t^T |\mathcal{L}f(s, X_s)|ds\right] < \infty$$
$$\mathbb{E}_{t,x}\left[\int_t^T \|D_x f(s, X_s)\sigma(s, X_s)\|^2 ds\right] < \infty,$$

then it holds that $f \in \mathcal{D}_L$ and $Lf(t, x) = \mathcal{L}f(t, x)$.

Proof.

Apply Ito's formula to $f(s, X_s)$ in Definition 4.2.3. With the assumptions in Theorem 4.2.6 the existence of all integrals as well as the interchangeability of expected value and limits is guaranteed. The assertion then follows immediately. □

Remark 4.2.7.

If $f \in C^{1,2}$ with compact support, then the conditions of Theorem 4.2.6 are met automatically, because in this case the appearing derivatives are bounded. This motivates the following theorem.

Theorem 4.2.8 (Dynkin's Formula)**.**

Let $f \in C^{1,2}$ be a function with compact support and τ be a stopping time with $\mathbb{E}_x[\tau] < \infty$. Then it holds that

$$\mathbb{E}_x[f(\tau, X_\tau)] = f(0, x) + \mathbb{E}_x\left[\int_0^\tau \mathcal{L}f(s, X_s)\, ds\right]$$

Proof.

Apply Ito's formula to $f(\tau \wedge n, X_{\tau\wedge n})$ for a fixed $n \in \mathbb{N}$, thus

$$f(\tau \wedge n, X_{\tau\wedge n}) = f(0, x) + \int_0^{\tau\wedge n} \mathcal{L}f(s, X_s)\, ds + \int_0^{\tau\wedge n} \sigma(s, X_s)D_x f(s, X_s)\, dW_s \tag{4.7}$$

If we apply the expected value, then the right summand in (4.7) is omitted, because $f \in C^{1,2}$ with compact support and we get

$$\mathbb{E}_x[f(\tau \wedge n, X_{\tau\wedge n})] = f(0, x) + \mathbb{E}_x\left[\int_0^{\tau\wedge n} \mathcal{L}f(s, X_s)\, ds\right]. \tag{4.8}$$

By estimating the right expected value in (4.8) and because of the dominated convergence theorem the assertion follows for $n \to \infty$. □

Remark 4.2.9.

If τ is an exit time of a bounded set A, then Theorem 4.2.8 is valid $\forall f \in C^{1,2}$.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

4.2.2 Bellman Principle & Hamilton-Jacobi-Bellman Equation

The idea of the dynamic programming principle, which is also known as Bellman principle, is to split the global optimization problem on the interval $[t, T]$ into two local optimization problems on the intervals $[t, t_1]$ and $[t_1, T]$. Using this principle, the Hamilton-Jacobi-Bellman equation can then be derived, that describes the behavior of the value function $V(t, x)$ when t_1 runs against t for $t_1 > t$.
With the setting from Section 4.1 the Bellman principle is formulated as follows:

$$V(t,x) := \sup_{u \in \mathcal{A}(t,x)} \mathbb{E}_{t,x}\left[\int_t^{t_1} \phi(s, X_s, u_s)\, ds + V(t_1, X_1)\right] \quad \text{for } t_1 > t. \tag{4.9}$$

Thus, behaving optimally in the interval $[t, t_1]$ and also acting optimally on $[t_1, T]$ afterwards, this will lead to a global optimum. If we assume that the value function is smooth enough, meaning that $V(t_1, X_{t_1}) \in C^{1,2}$, then by applying Ito's formula we get

$$\begin{aligned} V(t_1, X_{t_1}) = V(t,x) &+ \int_t^{t_1} \left(V_t(s, X_s) + b(s, X_s, u_s) D_x V(s, X_s)\right)\, ds \\ &+ \int_t^{t_1} \frac{1}{2} tr\left(a(s, X_s, u_s) D_{xx} V(s, X_s)\right)\, ds \\ &+ \int_t^{t_1} D_x V(s, X_s) \sigma(s, X_s, u_s)\, dW_s \end{aligned} \tag{4.10}$$

where we define

$$\begin{aligned} g(t,x) := &\int_t^{t_1} \left(V_t(s, X_s) + b(s, X_s, u_s) D_x V(s, X_s)\right)\, ds \\ &+ \int_t^{t_1} \frac{1}{2} tr\left(a(s, X_s, u_s) D_{xx} V(s, X_s)\right)\, ds \\ &+ \int_t^{t_1} D_x V(s, X_s) \sigma(s, X_s, u_s)\, dW_s \end{aligned} \tag{4.11}$$

By assuming that the value function V is so, that the stochastic integral is a martingale (e.g. integrand in $L^2(\omega)$ or bounded), so that it disappears in the expected value, meaning that

$$\mathbb{E}\left[\int_t^{t_1} D_x V(s, X_s) \sigma(s, X_s, u_s)\, dW_s\right] = 0, \tag{4.12}$$

we obtain for (4.9) using (4.10), (4.11) and (4.12) that

$$V(t,x) := \sup_{u \in \mathcal{A}(t,x)} \mathbb{E}_{t,x}\left[\int_t^{t_1} \phi(s, X_s, u_s)\, ds + V(t,x) + g(t,x)\right] \tag{4.13}$$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Because of the conditional expectation we can set $X_t = x$. Subtracting $V(t,x)$ on both sides, dividing by $t_1 - t$ and taking the limit $t_1 \to t$ yields

$$0 = \lim_{t_1 \to t} \sup_{u \in \mathcal{A}(t,x)} \mathbb{E}_{t,x} \left[\frac{\int_t^{t_1} \phi(s, X_s, u_s)\, ds + g(t,x)}{t_1 - t} \right] \tag{4.14}$$

Under the assumption that all interchanges are allowed, meaning that the supremum and the limit but also the expectation and the limit can be interchanged, we can write (4.14) as

$$0 = \sup_{u \in \mathcal{A}(t,x)} \mathbb{E}_{t,x} \left[\lim_{t_1 \to t} \frac{\int_t^{t_1} \phi(s, X_s, u_s)\, ds + g(t,x)}{t_1 - t} \right] \tag{4.15}$$

We are interested in finding the optimal strategy on the interval $[t, t_1]$, but due to the limit $t_1 \to t$ this finding of the optimal control reduces at time t, which means that $u_t = u$, so we get in (4.15)

$$0 = \sup_{u \in U} \left\{ \phi(t,x,u) + V_t(t,x) + (D_x V(t,x)) b(t,x,u) + \frac{1}{2} tr((D_{xx} V(t,x)) a(t,x,u)) \right\},$$

for which we will write by using (4.6)

$$0 = \sup_{u \in U} \{ \phi(t,x,u) + \mathcal{L}^u V(t,x) \} \tag{4.16}$$

The equation in (4.16) is called Hamilton-Jacobi-Bellman equation or HJB equation.

In conclusion it can be seen that under certain conditions the fulfillment of the HJB equation is a necessary condition to the value function to achieve the supremum of the objective functional J. If the value function is not yet known, a candidate for the value function V and the optimal strategy u can usually be determined from the HJB equation using methods of the partial differential equation. This candidate must be confirmed as a value function with a verification theorem, which will be introduced in Section 4.3.

Remark 4.2.10 (Algorithm for solving the control problem)**.**

The following scheme is usually used to calculate the searched value function $V(t,x)$ and optimal strategy $u_t^\star$:

1. *First the maximizer of the Hamilton-Jacobi-Bellman equation is determined as $\hat{u}(t,x) = \hat{u}(t,x,D_xV,D_{xx}V)$ if it exists.*

2. *Then $\hat{u}(t,x)$ is inserted into the HJB equation and a partial differential equation of second order for $V(t,x)$ is obtained*

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

3. *In the next step, an attempt is made to solve this partial differential equation with boundary conditions or terminal conditions $V(T,x) = \psi(T,x)$ and we receive a representation of the value function from which the optimal strategy can be derived. This can be achieved by*
   - *certain approaches*
   - *numerical methods*
   - *Fourier series (in linear differential equations)*
4. *Assuming that the boundary value problem from 3. is solvable, it implies that V is only a possible candidate for the value function. Therefore, we have to consider the following stochastic differential equation:*

$$dX_t^\star = b(t, X_t^\star, \hat{u}(t, X_t^\star))\, dt + \sigma(t, X_t^\star, \hat{u}(t, X_t^\star))\, dW_t$$

5. *If there exists a unique strong solution of the stochastic differential equation in 4., set $u_t^\star = \hat{u}(t, X_t^\star)$*
6. *If $u_t^\star \in \mathcal{A}(t,x)$, then $u_t^\star$ is the optimal strategy and $V(t,x)$ is the searched value function.*

4.3 Verification Theorems

Now the question arises whether a solution of the HJB equation is automatically also the searched value function. This consideration leads to the verification theorems. First we are going to present a verification theorem with finite time horizon followed by the verification theorem with infinite time horizon. These two verification theorems are based on strong assumptions. For many concrete problems own verification theorems are necessary.

4.3.1 Finite Time Horizon

We are considering the setting of Section 4.1, which we summarize again for the purpose of clarity:

$$\begin{aligned} dX_t &= b(t, X_t, u_t)\, dt + \sigma(t, X_t, u_t)\, dW_t \quad t \in [0,T] \quad \text{with } X_0 = x \\ J(t,x,u_t) &= \mathbb{E}\left[\int_t^T \phi(s, X_s, u_s) ds + \psi(T, X_T) \mid X_t = x\right] \\ V(t,x) &= \sup_{u \in \mathcal{A}(t,x)} J(t,x,u) \end{aligned}$$

with

$$\begin{aligned} b(\cdot,\cdot,\cdot) &: \mathbb{R}^+ \times \mathbb{R}^n \times U \to \mathbb{R}^n \\ \sigma(\cdot,\cdot,\cdot) &: \mathbb{R}^+ \times \mathbb{R}^n \times U \to \mathbb{R}^{n \times m} \end{aligned}$$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Remark 4.3.1 (Conditions at $\mathcal{A}(t,x)$)**.**
For strategies $u \in \mathcal{A}(t,x)$ we assume, that

1. $(u_s)_{s\in[t,T]}$ *is progressively measurable, such that* $\mathbb{E}\left[\int_t^T \|u_s\|^2\,ds\right] < \infty$.
2. *The stochastic differential equation in* (4.1) *has a unique strong solution* $X_t = x$, *such that*
$$\mathbb{E}\left[\sup_{t\le s\le T} \|X_s\|^2\right] < \infty$$
for all $u \in \mathcal{A}(t,x)$.
3. *The objective functional* $J(t,x,u)$ *is well-defined for all* $u \in \mathcal{A}(t,x)$.

Theorem 4.3.2 (Verification Theorem with Finite Time Horizon)**.**
Suppose that ϕ is continuous, such that the linear growth conditions for ϕ and σ are fulfilled
$$\|\phi(t,x,u)\|^2 \le C_\phi(1+\|x\|^2+\|u\|^2)$$
$$\|\sigma(t,x,u)\|^2 \le C_\sigma(1+\|x\|^2+\|u\|^2),$$
for some $C_\phi, C_\sigma > 0$ and all $t \ge 0$, $x \in \mathbb{R}^n, u \in U$.

1. *Furthermore, suppose that $\Phi \in C^{1,2}([0,T]\times\mathbb{R}^n)$ with $\|\Phi(t,x)\| \le C_\Phi(1+\|x\|^2)$ fulfills the HJB equation*
$$0 = \sup_{u\in U}\{\phi(t,x,u) + \mathcal{L}^u\Phi(t,x)\} \quad \forall t\in[0,T],\ x\in\mathbb{R}^n, \tag{4.17}$$
and the boundary condition
$$\Phi(T,x) = \Psi(T,x) \quad \forall x \in \mathbb{R}^n$$
Then it holds that
$$\Phi(t,x) \ge V(t,x) \quad \forall t\in[0,T],\ x\in\mathbb{R}^n.$$
2. *If in addition a maximizer $\hat{u}(t,x)$ of* (4.17) *exists, such that $u_t^\star = \hat{u}(t,X_t^\star) \in \mathcal{A}(t,x)$, then it holds that*
$$\Phi(t,x) = V(t,x) \quad \forall t\in[0,T],\ x\in\mathbb{R}^n,$$
where $u_t^\star$ is the optimal strategy, meaning that $V(t,x) = J(t,x,u_t^\star)$, and $X_t^\star$ is the solution of the stochastic differential equation in (4.1)*, meaning that*
$$dX_t^\star = b(t,X_t^\star,\hat{u}(t,X_t^\star))\,dt + \sigma(t,X_t^\star,\hat{u}(t,X_t^\star))\,dW_t.$$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Proof.

Let $t \in [0,T]$, $x \in \mathbb{R}^n$ be fixed. We define a stopping time τ_n as follows

$$\tau_n = T \wedge \inf\{s > t : \|X_s - X_t\| \geq n\} \quad n \in \mathbb{N}$$

with convention $\inf \emptyset = T$. Furthermore let $u \in \mathcal{A}(t,x)$ be an arbitrary admissible strategy. With Ito's formula for $X_t = x$ and $\Phi \in C^{1,2}$ we obtain

$$\Phi(\tau_n, X_{\tau_n}) = \Phi(t,x) + \int_t^{\tau_n} \mathcal{L}^{u_s}\Phi(s,X_s)\,ds + \int_t^{\tau_n} \Phi_x(s,X_s)\sigma(s,X_s,u_s)\,dW_s.$$

From the admissibility of u, the boundedness of X on $[t,\tau_n]$ and the continuity of Φ, we get

$$\mathbb{E}_{t,x}\left[\int_t^{\tau_n} \|\Phi_x(s,X_s)\sigma(s,X_s,u_s)\|^2 ds\right] < \infty$$

and thus

$$\mathbb{E}_{t,x}\left[\int_t^{\tau_n} \Phi_x(s,X_s)\sigma(s,X_s,u_s)\,dW_s\right] = 0.$$

Therefore we receive

$$\begin{aligned}
&\mathbb{E}_{t,x}\left[\int_t^{\tau_n} \phi(s,X_s,u_s)\,ds + \Phi(\tau_n,X_{\tau_n})\right] \\
&= \mathbb{E}_{t,x}\left[\int_t^{\tau_n} \phi(s,X_s,u_s)\,ds + \Phi(t,x) + \int_t^{\tau_n} \mathcal{L}^{u_s}\Phi(s,X_s)\,ds\right], \\
&= \Phi(t,x) + \mathbb{E}_{t,x}\left[\int_t^{\tau_n} (\phi(s,X_s,u_s) + \mathcal{L}^{u_s}\Phi(s,X_s))\,ds\right] \\
&\leq \Phi(t,x)
\end{aligned} \tag{4.18}$$

since Φ satisfies the HJB equation, $u_s \in U$ for $s \in [t,T]$ and the fact, that

$$\phi(s,X_s,u_s) + \mathcal{L}^{u_s}\Phi(s,X_s) \leq 0. \tag{4.19}$$

Now for $n \to \infty$ we get $\lim_{n\to\infty} \tau_n = T$ almost surely. Due to the quadratic growth conditions and the admissibility of u it follows, by defining

$$C := C_\phi(1 + \|X_s\|^2 + \|u_s\|^2)\,ds,$$

that

$$\left|\int_t^{\tau_n} \phi(s,X_s,u_s)\,ds + \Phi(\tau_n,X_{\tau_n})\right| \leq \int_t^{\tau_n} C + C_\Phi(1 + \|X_{\tau_n}\|^2),$$

and by taking the supremum of $\|X_{\tau_n}\|^2$ we get a larger upper bound

$$\left|\int_t^{\tau_n} \phi(s,X_s,u_s)\,ds + \Phi(\tau_n,X_{\tau_n})\right| \leq \int_t^{\tau_n} C + C_\Phi(1 + \sup_{t\leq s\leq \tau_n} \|X_{\tau_n}\|^2) \in L^1(\mathbb{P}),$$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

so there exists an integrable majorante and therefore we can interchange the limit and the expectation. Due to the dominated convergence theorem and the continuity of Φ and $\Phi(T,x) = \Psi(T,x)$ we receive

$$\mathbb{E}_{t,x}\left[\int_t^T \phi(s, X_s, u_s)\, ds + \Phi(T, X_T)\right] \leq \Phi(t,x),$$

which implies, that

$$J(t,x,u) \leq \Phi(t,x) \quad \forall u \in \mathcal{A}(t,x).$$

By taking the supremum it follows, that

$$V(t,x) = \sup_{u \in \mathcal{A}(t,x)} J(t,x,u) \leq \Phi(t,x),$$

which finally proves the first part of Theorem 4.3.2.
For the second part we have equality in (4.18) if we can find a maximizer $\hat{u}(t,x)$ with strategy $u_t^\star = \hat{u}(t, X_t^\star) \in \mathcal{A}(t,x)$, which implies that we have in (4.19)

$$\phi(s, X_s, u_s) + \mathcal{L}^{u_s}\Phi(s, X_s) = 0.$$

The rest of the evidence follows analogously to the first part, for which we get

$$\mathbb{E}_{t,x}\left[\int_t^T \phi(s, X_s, u_s^\star)\, ds + \Phi(T, X_T)\right] = J(t,x,u^\star) = \Phi(t,x)$$

and finally with part one

$$V(t,x) = \Phi(t,x).$$

□

Remark 4.3.3.

1. *Under the conditions of Theorem 4.3.2 the Bellman principle for any stopping time $\tau \in [t,T]$ is formulated as follows:*

$$V(t,x) = \sup_{u \in \mathcal{A}(t,x)} \mathbb{E}_{t,x}\left[\int_t^\tau \phi(s, X_s, u_s)\, ds + V(\tau, X_\tau)\right].$$

2. *The value function is unique by definition. Therefore Theorem 4.3.2 shows that a solution of the HJB equation is unique in the class of $C^{1,2}$-functions which fulfill the quadratic growth condition.*

3. *In general, existence is difficult to show, but sufficient conditions are for instance*
   - *$\Psi(T,x) \in C^3$ with respect to x and bounded*
   - *$\phi(t,x,u), b(t,x,u), \sigma(t,x,u) \in C^{1,2}$ uniformly in u and bounded*
   - *U is compact*
   - *$a(t,x,u)$ is uniformly postiv definit*

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

4.3.2 Infinite Time Horizon

We will now consider the infinite time horizon, where we additionally assume that the running costs $\phi(X_t, u_t)$ but also the coefficients $b(X_t, u_t), \sigma(X_t, u_t)$ do not explicitly depend on time. Therefore, the stochastic differential equation in this section is formulated as follows

$$dX_t = b(X_t, u_t)\, dt + \sigma(X_t, u_t)\, dW_t \quad \text{with } X_0 = x. \tag{4.20}$$

In this case we can simplify the partial differential operator from (4.5) to

$$\mathcal{L}^u f(x) = (D_x f(x))b(x, u) + \frac{1}{2} tr((D_{xx} f(x))a(x, u))$$

So the objective functional from (4.2) can also be rewritten as

$$J(x, u) = \mathbb{E}\left[\int_0^\infty \exp(-\beta t)\phi(X_t, u_t)\, dt\right],$$

where $\beta > 0$ describes the interest intensity. Furthermore, let $A(x)$ be the set of admissible strategies, which leads us to the definition of the value process

$$V(x) = \sup_{u \in \mathcal{A}(x)} J(x, u).$$

Before we state the verification theorem with infinite time horizon, we are denoting the criterions, when a strategy $u \in \mathcal{A}(x)$ is admissible.

Remark 4.3.4 (Conditions at $\mathcal{A}(x)$)**.**
For strategies $u \in \mathcal{A}(x)$ we assume, that

1. *u is progressively measurable, such that $\mathbb{E}\left[\int_0^\infty \|u_t\|^2\, dt\right] < \infty$*
2. *The stochastic differential equation in (4.20) has a unique strong solution X_t, such that $\mathbb{E}\left[\exp(-\beta s)\|X_s\|^2\right] < \infty$*
3. *The objective functional $J(x, u)$ is well-defined*

Theorem 4.3.5 (Verification Theorem with Infinite Time Horizon)**.**
Suppose that ϕ is continuous, such that the quadratic growth conditions for ϕ and σ are fulfilled

$$\|\phi(x, u)\|^2 \leq C_\phi(1 + \|x\|^2 + \|u\|^2)$$
$$\|\sigma(x, u)\|^2 \leq C_\sigma(1 + \|x\|^2 + \|u\|^2),$$

for some $C_\phi, C_\sigma > 0$ and all $x \in \mathbb{R}^n, u \in U$.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

1. *Furthermore, suppose that* $\Phi \in C^2(\mathbb{R}^n)$ *with* $\|\Phi(x)\| \leq C_\Phi(1 + \|x\|^2)$ *fulfills the HJB equation*

$$0 = \sup_{u \in U}\{\phi(x,u) + \mathcal{L}^u\Phi(x) - \beta\Phi(x)\} \quad x \in \mathbb{R}^n, \tag{4.21}$$

and

$$\lim_{T\to\infty} \mathbb{E}[\exp(-\beta T)\Phi(X_T)] = 0$$

Then it holds that

$$\Phi(x) \geq V(x) \quad \forall x \in \mathbb{R}^n.$$

2. *If in addition a maximizer* $\hat{u}(x)$ *of* (4.21) *exists, such that* $u_t^\star = \hat{u}(X_t^\star) \in \mathcal{A}(x)$*, then it holds that*

$$\Phi(x) = V(x) \quad \forall x \in \mathbb{R}^n,$$

where $u_t^\star$ *is the optimal strategy, meaning that* $V(x) = J(x, u_t^\star)$*, and* $X_t^\star$ *is the solution of the stochastic differential equation in* (4.20)*, meaning that*

$$dX_t^\star = b(X_t^\star, \hat{u}(X_t^\star))\,dt + \sigma(X_t^\star, \hat{u}(X_t^\star))\,dW_t.$$

Proof.

The basis for this proof is the proof of the verification theorem with finite time horizon, see Proof 4.3.1. We will take $\tilde{\Phi}(t,x) = \exp(-\beta t)\Phi(x)$ as candidate function for Theorem 4.3.2. By taking $\tilde{\phi}(t,x) = \exp(-\beta t)\phi(x)$, we insert $\tilde{\Phi}(t,x)$ and $\tilde{\phi}(t,x)$ into (4.17) of Theorem 4.3.2 and get with $\Phi(x)$ equivalence to (4.21) in this theorem. If $\Phi(x)$ satisfies (4.17), then $\tilde{\Phi}(t,x)$ satisfies (4.21) and it applies that

$$\tilde{\Phi}(t,x) \geq V(t,x) \text{ with } \Phi(x) = \tilde{\Phi}(0,x) \geq V(0,x) = V(x),$$

for the first part and

$$\tilde{\Phi}(t,x) = V(t,x) \text{ with } \Phi(x) = \tilde{\Phi}(0,x) = V(0,x) = V(x),$$

for the second part of Theorem 4.3.2.
The rest of the proof follows analogously the proof for Theorem 4.3.2. □

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

5 Two-Dimensional Controlled Ruin Problem

This section deals with the summary of the article by H. P. McKean and L. A. Shepp, see [18], in which the authors were considering a two-dimensional controlled ruin problem.

Politics today has a huge influence on the economy of every country in the world. From the perspective of a leading position such as that of the government, for example, we can influence the development and course of the economy on the one hand positively, but on the other hand also negatively through the chosen policy.

Now let us assume that the government of any country intends to grant tax breaks to companies. Moreover we are supposing that the government separates two policies.

1. On one side there is the republican policy, which supports the richer companies by giving them tax breaks in hopes to help them to expand and become even richer.

2. On the other side there is the democratic policy, which supports the weaker companies by giving them tax breaks in hopes to help them not getting ruined and keep them alive and therefore reduce unemployment.

With regard to the country's overall economy, the authors have discussed which policy is better, whereby better can be interpreted in different ways due to the different policies. When looking at the corresponding mathematical model for two companies, which we will discuss in more detail in Section 5.1, the authors have shown that the advantage of capitalism versus socialism depends on the optimization criterion.

5.1 Mathematical Model of the Specific Ruin Problem

The original mathematical model was formulated by David Aldous, see [1]. Let us assume that two independent standard Brownian motions $W_i(t)$ $i \in \{1, 2\}$ can be controlled by adding processes of non-negative drifts $\mu_i(t)$ with sum unit to each process in a non-anticipatory way. Therefore the controlled process is satisfying

$$\begin{aligned} dX_i(t) = \mu_i(t)dt + dW_i(t) \quad \text{with } X_i(0) = x_i \geq 0 \\ \mu_i(t) \geq -\delta,\ \mu_1 + \mu_2 \leq 1,\ -\frac{1}{2} \leq \delta < \infty \end{aligned} \tag{5.1}$$

for $i \in \{1, 2\}$ and $t \geq 0$, where x_i and δ are given. For each $t \geq 0$ we may immediately change the drifts $\mu_i(t)$, but without prior knowledge of the future increments of the independent standard Brownian motion $W_i(t)$. If $X_i(t)$ reaches zero for a finite value of t, it is supposed that the company i goes bankrupt and disappears from the market.

We denote with $\nu \in N := \{0, 1, 2\}$ the number of companies which never go bankrupt, meaning that they never hit zero and therefore survive forever. Considering the above

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

mentioned processes in (5.1), the question is how the risk controller, in our case the government, should support these two companies in the form of tax benefits. This depends on the optimization criteria, which can be applied as follows:

1. To maximize the probability that both companies survive forever, which can be written as

$$\mathbb{P}[\nu = 2] \to \max \tag{5.2}$$

2. To maximize $\mathbb{E}[\nu]$, the expected number of companies that survive forever, which can be written as

$$\mathbb{E}[\nu] = \mathbb{P}[\nu = 1] \cdot 1 + \mathbb{P}[\nu = 2] \cdot 2 \to \max \tag{5.3}$$

This means that these two companies should be supported in such a way that a certain criterion is optimized. Since we are interested in both cases, it is easy to merge these two optimization criteria into one common criterion by inserting a parameter, indicated by α, so we have

$$\alpha \cdot \mathbb{P}[\nu = 1] + (1 - \alpha) \cdot \mathbb{P}[\nu = 2] \to \max \tag{5.4}$$

The optimization problem in (5.4) reduces for $\alpha = 0$ to the first optimization criterion in (5.2), while for $\alpha = \frac{1}{3}$, we get an equivalent optimization problem of the second optimization criterion in (5.3). However, for maximizing the above problem in (5.4), the decision maker must select between two different strategies:

1. The so-called push-bottom strategy or democratic policy, where the drift $\mu = 1$ is assigned to the bottom company.

2. The so-called push-top strategy or republican policy, where the drift $\mu = 1$ is assigned to the upper company.

Obviously, knowing the future can give better results, and for this reason there is no getting around martingale methods, as martingales are the only option to make use of the limitation to non-anticipative strategies. If we consider more than two companies our problem would be much harder, but also if the diffusion constants of the two independent standard Brownian motions are unequal.

5.2 Mathematical Difficulties of the Specific Ruin Problem

In this section we are going to explain some mathematical difficulties when considering the two-dimensional ruin problem in (5.1). We will start with an overview of the theory and the corresponding problems of stochastic optimization in finance and economics. Due to the fact that these problems have a long history, we can say that it perhaps started by John von Neumann and Oskar Morgenstern's "Theory of Games and Economic Behavior" (see [28]). This fundamental motivated and initiated later the breakthrough of the theory

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

of linear programming and, as a consequence, added optimal control and optimal stopping problems.

An example of an early one-dimensional problem is the pricing of the perpetual American option postulated by Paul Samuelson and solved by Henry McKean (see [19]). Maximizing the probability of getting to a specific wealth in a sub-fair casino, set and solved by Dubins and Savage, would be another example for it.

In these examples, martingales were used and therefore seemed to be a new direction of the problem, but Dynkin clearly understood that it is limited to linear programming. Although the dimension of the considered problems was infinite, the method of smooth fitting sometimes made it possible to guess optimal solutions. Ito-calculation played an essential part in the determination of a simple criterion for the estimated optimal process, which can be a martingale or a submartingale. The technique has frequently been used to find solutions to financial and economic problems instead of having to solve the nonlinear differential equations that occur when using the dynamic programming approach.

Now, after a short introduction to the development of stochastic optimization theory and some supposed problem solving techniques, we would like to explain the mathematical difficulties the authors encounter with respect to our model above.

1. **Some problems are too discrete**
   The first difficulty that the authors discovered with the model under consideration is that there is too much discretion and no Itô calculus, which makes it rather difficult to find an optimal solution. For example, suppose we have to pay back a debt and therefore make a size 1 out of an asset $0 \leq f_0 < 1$, where 1 stands for one million rubles in a casino where there is only a sub-fair "double or nothing" bet. You can bet any amount $s < f_0$ and reach the fortune $f_1 = f_0 + s$ with the probability $p < \frac{1}{2}$ or $f_1 = f_0 - s$ with the probability $q = 1 - p > \frac{1}{2}$. It has been proven that a bold play, where a person bets $s = \min(f_0, 1 - f_0)$ on each bet until this person either go broke or reach the target, is optimal.

   If alternatively an interest payment is obtained after each bet, so that $f_1 = \frac{f_0+s}{1+a}$ is reached with probability $p < \frac{1}{2}$, or $f_1 = \frac{f_0-s}{1+a}$ with probability $q = 1 - p > \frac{1}{2}$, by dividing by $1 + a > 1$, then it is even more obvious that bold play is optimal, but this is actually wrong! The optimal strategy seems to be quite tricky to find.

2. **Some problems are higher dimensional**
   The second difficulty in our model is that, in contrast to the previously solved problems in the financial and economic industry, we consider not one but two independent standard Brownian movements. Involving only one Brownian motion is an effective way to reduce the assumption of the optimal control boundary to the assumption of a single threshold, possibly one for each value of a parameter. Moreover the smooth fit method allows the assumption of the threshold because of the one-dimensional Ito equation and the fact that for applying the technique of smooth fit it is only necessary to solve an ordinary differential equation.

   An example for an early one-dimensional problem, which was formulated by Samuel-

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

son and solved by McKean is the pricing of the perpetual American option:

$$V(x) = \sup_{\tau} \mathbb{E}_x \left[\exp(-r\tau)(X_\tau - k)^+ \right],$$

where

$$X = X(t) = x_0 \exp\left(\sigma W(t) + \left(\mu - \frac{\sigma^2}{2} \right) t \right),$$

or in differential notation

$$dX(t) = X(t)(\mu dt + \sigma dW(t)), \quad X(0) = x_0.$$

The value of the option was determined on the basis of the Ito calculus and the principle of smooth fitting and it turned out that the optimal $\tau = \tau_a$ is the first passage time up to a certain level $a > k$. Xin Guo published a generalization of this result in her dissertation in 2000, in which she treated models with arbitrage and insider information (see [12]). When problems arise in the economy rather than in the financial sector, people make business decisions, and to investigate this, Radner introduced a model to study optimal directional decisions and profit taking. In this model the net value of the company at time t follows the stochastic differential equation

$$dX(t) = \mu_{I(t)} dt + \sigma_{I(t)} dW(t) - dZ(t), \quad X(0) = x_0,$$

where $I(t) \in \{1, \ldots, n\}$ and $Z(t)$ is the total dividend or profit taken up to time t. The set $\mathcal{S} = \{(\mu_i, \sigma_i),\ i = 1, \ldots, n\}$ are the eligible business directions and the goal is to maximize the total discounted expected profit

$$V(x_0, \mathcal{S}) = \sup_{Z,I} \mathbb{E}_{x_0} \left[\int_0^{\tau_0} \exp(-rt) dZ(t) \right]$$

until it is declared bankrupt at time $\tau_0 \leq \infty$ by choosing a non anticipatory business direction $I(t)$ and a non anticipatory profit-taking policy $Z(t)$. The value $V(x_0)$ is the true value of the company, and the overall solution has been found. The company in the Radner model bankrupts in a finite time with probability 1. This can be seen as a consequence of unrestrained profit maximization.

3. **Need an explicit solution to make a proof**
   If we succeed in guessing the optimal control strategy in the above model and if we can find the value of the optimal payout amount under this strategy, then the proof that the optimum is the guessed answer is quite simple. Using the explicit formula, it is easy to prove that process $Z(t) = \hat{V}(X_t, Y_t)$ is a supermartingale for every policy. The supermartingal inequality then implies that no policy can do better than to push the bottom company, and since pushing the bottom company is a legitimate policy, optimality follows and thus the solution of the problem. Nevertheless, we can only try to find an explicit solution in our model, when we are considering the parameters $\alpha = 0$ and $\delta = 0$.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

5.3 The Optimal Control Policy for Two Companies

The authors of [18] have successfully and with great effort guessed the optimal payout function V for the case $\delta = 0$ and $\alpha = 0$ with democratic policy as optimal control strategy. In this section we are going to show that this democratic policy, which is also known as "Robin Hood" policy, is optimal. Therefore we are considering the model defined in (5.1) and we will show, but only if $\delta = 0$, that the Robin Hood policy is optimal by specifically calculating the probability that both companies will survive forever under this democratic policy.

Robin Pemantle used a basic coupling argument to show that this policy is indeed optimal for all delta, but without an explicit expression for the value function. In reality, what Robin's argument shows is that for n companies with the given same diffusion constant, pushing the bottom company is ideal to maximize the possibility that all companies will survive. Note that this speculation is quite analogous to the optimal policy of guessing the special extreme point of a convex set on which, as mentioned, a linear programming problem is solved.

In this case, we have a look at the function

$$V(x_1, x_2) = 1 - \exp(-2\min(x_1, x_2)) - 2\min(x_1, x_2)\exp(-x_1 - x_2),$$

which describes the probability that both companies never hit zero and therefore survive forever. Analogously, $V^\pi(x_1, x_2)$ denotes the probability that under the policy π both companies will survive forever, where π is a non-anticipatory policy that impacts either one company or the other. We intend to illustrate that $V^\pi \leq V$, where V denotes the probability that results when the bottom company is driven.

The process $Z(t) = V(X_1^\pi(t), X_2^\pi(t))$ is a supermartingale when and only when its Ito differential locally decreases the expectation, meaning that

$$\mathbb{E}[dZ(t)] = \frac{1}{2}(V_{x_1x_1} + V_{x_2x_2}) + \max(V_{x_1}, V_{x_2}) \leq 0,$$

for all (x_1, x_2) in the first quadrant and where V_{x_1}, V_{x_2} are denoting the partial derivatives first order and $V_{x_1x_1}$, $V_{x_2x_2}$ are denoting the partial derivatives second order. Moreover, we will need that $V \in C^2$ on the diagonal $x_1 = x_2$.

The drift of the process Y is

$$-2\exp(-2x_1) + 2^{-x_1-x_2} - 2x_1\exp(-x_1 - x_2) \leq 0.$$

So we get

$$\mathbb{E}[Y(\infty)] \leq Y(0) = V(x_1, x_2).$$

The following cases can occur at time $t = \infty$:

1. On the one hand none of the two companies has gone bankrupt and therefore both have survived, meaning that:

$$V(\infty, \infty) = 1 \quad \Rightarrow \quad Y(\infty) = 1$$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

2. On the other hand only one company has survived, which means that the other company has hit zero and therefore went bankrupt:

$$V(0,\cdot) = 0 \quad \Rightarrow \quad Y(\infty) = 0$$

Therefore, the supermartingale inequality implies that for each π the probability that both firms will survive forever is at most $V(x_1, x_2)$. This opportunity will be strictly below $V(x_1, x_2)$, unless the maximum company is never pushed. Due to the fact that $V(x_1, x_2)$ is bounded, equality is achieved throughout the evidence for the policy π_0, where the bottom company is always pushed. Thus we have finally showed that $V(x_1, x_2)$ indicates the value of the optimal payoff when the optimal control strategy is actually chosen. As already mentioned above, the authors of [18] were able to solve the Bellman equation, which had the form of a partial differential equation, and were lucky that the solution was guessable. They thought that they could guess a solution with the help of the Kontorovich-Lebedev transformation, because due to the following transformation

$$H(x_1, x_2) = \exp(x_1)V(x_1, x_2)$$

the Helmholtz equation

$$H_{x_1x_1} + H_{x_2x_2} = H$$

was fulfilled in the region $x_1 \leq x_2$ by H.
However, using the following representation with respect to the polar coordinates (r, θ) for (x_1, x_2)

$$V = \exp(-x_1) \int_0^\infty K_{i\nu}(r)[A(\nu)\cosh(\nu\theta) + B(\nu)\sinh(\nu\theta)]\, d\nu$$

no solution can be found, even for the case $\delta = 0$, for which the solution is already known.

The question arises, if it is possible to refer to an existence theorem without actually solving the partial differential equation and therefore if the above martingale/submartingale proof can be used if only a proof of existence and not, as above, an explicitly guessed solution is available.

5.4 Numerical Approximation & Results

Now the question of probability maximization arises, by choosing the criterion in such a way that all companies remain solvent and thus people remain employed, which would mean that socialism brings some advantages and this is certainly better for society, or whether it would be better for the economy if, on the other hand, we chose the criterion in such a way that the expected number of solvent companies is maximized.

The authors of [18] propose a numerical approximation with regard to the solution of the model presented in the previous section. Therefore, they generalized the considered model by looking at the criterion

$$\max_{\pi} \alpha \cdot \mathbb{P}[\nu = 1] + (1 - \alpha) \cdot \mathbb{P}[\nu = 2], \tag{5.5}$$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

where $\nu = \nu^\pi$ is the number of companies which survive forever and therefore never face bankruptcy under a non-anticipatory policy π. For $\alpha = 0$ they got in (5.5)

$$\max_\pi \mathbb{P}[\nu = 2],$$

which reduced the optimization problem to maximizing the probability that both companies will remain solvent. In case $\alpha = \frac{1}{3}$ they got in (5.5)

$$\max_\pi \frac{1}{3} \cdot \mathbb{P}[\nu = 1] + \frac{2}{3} \cdot \mathbb{P}[\nu = 2],$$

which means, that the optimization problem is reduced to maximizing the expected number of companies to remain solvent, where the problem is equivalent up to a multiplication factor of 3 to our origin model.

In order to find an approximate optimal payoff function the authors solved the non-linear partial differential equation

$$\frac{1}{2}(V_{x_1x_1} + V_{x_2x_2}) + \max(V_{x_1}(x_1, x_2), V_{x_2}(x_1, x_2)) = 0,$$

numerically with the following boundary conditions along the axes

$$V(x_1, 0) = \alpha(1 - \exp(-2x_1))$$
$$V(0, x_2) = \alpha(1 - \exp(-2x_2))$$

and the following boundary conditions at infinity

$$V(x_1, \infty) = (1 - \alpha)(1 - \exp(-2x_1)) + \alpha \exp(-2x_1)$$
$$V(\infty, x_2) = (1 - \alpha)(1 - \exp(-2x_2)) + \alpha \exp(-2x_2)$$

where the terms $1-\exp(-2x_1)$ and $1-\exp(-2x_2)$ are the result for the probability of never hitting zero for a Brownian motion with unit drift, $\mu = 1$ and diffusion, $\sigma = 1$ starting at x_1 and x_2, respectively.

Using a finite value instead of infinity and a discrete matrix instead of a continuous square, the authors were able to calculate an approximate optimal V of the partial differential equation in a 500×500 pixel array.

They visualized their results by using black and white dots for each pixel, where a black dot is indicating those values x_1, x_2 of the companies at which the maximum in the non-linear partial differential equation occurs at the smaller of x_1, x_2, namely in the bottom point. Otherwise a white dot is used.

In case of $\alpha = \frac{1}{3}$, meaning that the expected number of companies which are going to stay solvent forever is maximized, the authors used black dots representing those values of the companies where the bottom company is pushed optimally. For the diagonal $x_1 = x_2$ the government can push both companies as it makes no difference which one gets pushed. Due to the fact that for $\alpha = \frac{1}{3}$ the republican policy is optimal, the top companies should be pushed near the origin, where both companies are in danger of getting insolvent.

In case of $\alpha = \frac{1}{2}$, the authors show in their derived plot that the same results as for the case $\alpha = \frac{1}{3}$ are still valid. Remarkable, however, is that now the region near the origin,

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

where both companies are in danger to face bankruptcy and the top company under the republican policy is pushed according to the case $\alpha = \frac{1}{3}$, visualized through white dots, gets larger. This claim could not be verified using our own method, where the corresponding plots can be seen in Chapter 7.

The authors verified their results by implementing a linear program as an alternative numerical verification of the direct solution of the non-linear partial differential equation. They also pointed out that it seems very probable that for any number of companies that all have the same diffusion coefficient and whose drift $\mu_i \geq -\delta$ adds up to one, the government might want to push the bottom company to maximize the probability that all companies will survive forever. However, in the case $\delta = 0$ the authors were not able to solve the partial differential equation for more than two companies.

For the interested reader, there are several articles dealing with the extension of these results. For example [9], written by Grandits, is about the derivation of the following value function for $\delta \in (-\min\{\mu_1, \mu_2\}, 0)$

$$\begin{aligned} V(x,y) &= 1 - \exp(-2(1+\delta)\min\{x,y\}) \\ &\quad + \frac{1+\delta}{\delta}\exp(-(x+y))(1-\exp(-2\delta\min\{x,y\})), \end{aligned}$$

or [10], also written by Grandits, in which it is proved that the push-bottom strategy is not optimal if the expected number of surviving companies is maximized.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

6 Policy Iteration Method

In this chapter we will explain the so-called policy iteration method based on [26] which we are going to use in our numerical analysis in Chapter 7. The interested reader is referred to [5], [16] and [22], where a deeper insight into this approach can be obtained. This method can be used in reinforcement learning problems for calculating an optimal policy for a given Markov decision process. Therefore, we will first explain the most important definitions and theorems with respect to Markov chains based on [14] and [20] in Section 6.1 and then deal with the key components for such a problem in Section 6.2.

6.1 Markov Theory

Markov chains are processes with finite memory, which means that the future development of the process depends only on the present and not on the whole past. In this section we are considering the following setup

- I is the state space, a non-empty finite or countable infinite set
- $i \in I$ are denoting the states
- The family $(\lambda_i)_{i\in I}$ with $0 \leq \lambda_i < \infty \ \forall i \in I$ is considered a measure $\lambda = (\lambda_i)_{i\ inI}$ on $(I, \mathcal{P}(I))$, where $\mathcal{P}(I)$ is denoting the power set
- $\lambda(J) = \sum\limits_{j\in J} \lambda_j, \ J \in \mathcal{P}(I)$
- If $\sum\limits_{i\in I} \lambda_i = 1 \quad \Rightarrow \quad \lambda$ is a distribution on I

We are supposing that for a random variable $X : \Omega \to I$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ the following

$$\lambda_i = \mathbb{P}[X = i] = \mathbb{P}[\{\omega : X(\omega) = i\}], \ i \in I,$$

is describing the distribution of X under $\mathbb{P}$.

Definition 6.1.1 (Stochastic Matrix)**.**
A matrix $P = (p_{ij})_{i,j\in I}$ is called stochastic, if each row $(p_{ij})_{j\in I}$ is a distribution on I, which means, that

$$p_{ij} \geq 0 \quad \forall i, j \in I \quad \text{with} \sum_{j\in I} p_{ij} = 1 \quad \forall i \in I$$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Definition 6.1.2 (Markov Chain).

An I-valued stochastic process in discrete time is a sequence of I-valued random variables $(X_n)_{n\geq 0}$. We say that such a process $(X_n)_{n\geq 0}$ is a Markov chain or also known as Markov process with initial distribution λ and transition matrix P, $(X_n)_{n\geq 0} \sim Markov(\lambda, P)$, if

1. *X_0 has the distribution λ*
2. *For all $n \geq 0$, $i \in I$, conditional on $X_n = i$, the distribution of X_{n+1} is $(p_{ij} : j \in I)$ and is independent of $X_0, \ldots, X_{n-1}$,*

which can be written for $n \geq 0$ and $i_0, \ldots, i_{n+1} \in I$ more precisely as

1. $\mathbb{P}[X_0 = i_0] = \lambda_{i_0}$
2. $\mathbb{P}[X_{n+1} = i_{n+1} \mid X_0 = i_0, \ldots, X_n = i_n] = p_{i_n i_{n+1}}$.

A Markov chain is therefore a random process, which retains no memory of where it has been in the past and only the current state can influence where it goes next:

1. The future X_{n+1} depends only on the present X_n
2. $\mathbb{P}[X_{n+1} = i_{n+1} \mid X_0 = i_0, \ldots, X_n = i_n] = \frac{\mathbb{P}[X_0=i_0,\ldots,X_n=i_n,X_{n+1}=i_{n+1}]}{\mathbb{P}[X_0=i_0,\ldots,X_n=i_n]} = p_{i_n i_{n+1}}$, which is equivalent to

$$\mathbb{P}[X_0 = i_0, \ldots, X_n = i_n, X_{n+1} = i_{n+1}] = \mathbb{P}[X_0 = i_0, \ldots, X_n = i_n] \cdot p_{i_n i_{n+1}}.$$

So if we denote the past by $A = \{X_0 = i_0, \ldots, X_n = i_n\}$, the present by $B = \{X_n = i_n\}$ and the future by $C = \{X_{n+1} = i_{n+1}\}$, it holds that

$$\mathbb{P}[C \mid A \cap B] = \mathbb{P}[C \mid B]$$
$$\mathbb{P}[A \cap C \mid B] = \mathbb{P}[A \mid B] \cdot \mathbb{P}[C \mid B].$$

Theorem 6.1.3 (Construction of Markov chains).

For a distribution λ and a stochastic matrix P on the state space I, there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and an I-valued process $(X_n)_{n\geq 0}$, such that $(X_n)_{n\geq 0} \sim Markov(\lambda, P)$.

Theorem 6.1.4 (Finite-dimensional distributions of a Markov chain).

$(X_n)_{0\leq n\leq N} \sim Markov(\lambda, P) \quad \Leftrightarrow \quad \forall i_0, \ldots, i_N \in I$:

$$\mathbb{P}[X_0 = i_0, \ldots, X_N = i_N] = \lambda_{i_0} \cdot p_{i_0 i_1} \cdot \ldots \cdot p_{i_{N-1} i_N}$$

Definition 6.1.5 (Kronecker-Delta).

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & otherwise \end{cases}$$

For the Dirac-distribution in the state i, which is denoted by δ_i, it holds that

$$X \sim Markov(\delta_i, P) \quad \Rightarrow \quad \mathbb{P}[X_0 = j] = \delta_{ij},$$

which means that the process X is starting in the state i with probability 1.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Theorem 6.1.6 (Markov Property for Markov Chains)**.**
Let $(X_n)_{n\geq 0} \sim Markov(\lambda, P)$. Then conditionally on $\{X_m = i\}$ it holds that the process is $(X_{m+n})_{n\geq 0} \sim Markov(\delta_i, P)$ and is independent of the random variables $X_0, \dots, X_m$.

For distributions λ and stochastic matrices P, Q over I we define for $i, j \in I$ the distribution λP as

$$(\lambda P)_i = \sum_{k\in I} \lambda_k P_{ki} \quad \text{with} \quad \sum (\lambda P)_i = 1$$

and the stochastic matrix PQ as

$$(PQ)_{ij} = \sum_{k\in I} P_{ik} Q_{kj} \quad \text{with row sum 1,}$$

where

$$\begin{aligned} P^0 &= (\delta_{ij})_{i,j\in I} \text{ denotes the identity matrix and} \\ P^n &= P \cdot P^{n-1},\ n \geq 1 \text{ (matrix multiplication).} \end{aligned}$$

In general it holds that

$$(P^n)_{ij} \neq (P_{ij})^n$$

and we are defining

$$\begin{aligned} p_{ij}^{(n)} &:= (P^n)_{ij} \quad \text{the } n\text{-step transition probability from state } i \text{ to state } j \\ \lambda_i^{(n)} &:= (\lambda P^n)_i \end{aligned}$$

Theorem 6.1.7.
Let $(X_n)_{n\geq 0} \sim Markov(\lambda, P)$. Then it holds for all $m, n \geq 0$ that

$$\begin{aligned} \mathbb{P}[X_n = j] &= (\lambda P^n)_j = \lambda_j^{(n)} \\ \mathbb{P}_i[X_n = j] &= \mathbb{P}[X_{m+n} = j \mid X_m = i] = (P^n)_{ij} = p_{ij}^{(n)} \end{aligned}$$

Theorem 6.1.8 (Chapman–Kolmogorov Equations)**.**

$$p_{ij}^{(n+m)} = \sum_{k\in I} p_{ik}^{(n)} p_{kj}^{(m)} \quad \forall i, j \in I$$

Definition 6.1.9 (Communication Classes)**.**
We say a state i leads to state j respectively j is reachable from i, denoted by $i \to j$, if $\mathbb{P}_i[X_n = j \text{ for some } n \geq 0] > 0$. We say the states i and j communicate, denoted by $i \leftrightarrow j$, if $i \to j$ and $j \to i$.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Theorem 6.1.10.

For distinct states i, j the following are equivalent

(i) $i \to j$

(ii) There exists a sequence of states $i_0, i_1, \ldots, i_n$ with $i_0 = i, i_n = j$ such that

$$p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-1} i_n} > 0$$

(iii) $p_{ij}^{(n)} > 0$ *for some* $n \geq 0$.

Communication of states is an equivalence relation, which means that the follwoing properties are fulfilled

(i) reflexive: $i \leftrightarrow i$

(ii) symmetric: $i \leftrightarrow j \Leftrightarrow j \leftrightarrow i$

(iii) transitive: $i \leftrightarrow j,\ j \leftrightarrow k \Rightarrow i \leftrightarrow k$

Equivalence relation partitions the state space into equivalence classes, called communication classes.

Definition 6.1.11.

A class C is called closed, if for $i \in C$ and $i \to j$ it follows that $j \in C$, meaning that we do not escape and therefore stay in our class C. A state i is called absorbing, if $\{i\}$ is a closed class. A chain or transition matrix is called irreducible, if it has only one class.

Definition 6.1.12 (Hitting Times & Absorption Probabilities)**.**

Let $(X_n)_{n\geq 0}$ be a Markov chain with transition matrix P. The mapping $H^A : \Omega \to \{0, 1, \ldots, \} \cup \{\infty\}$ with

$$H^A(\omega) = \inf\{n \geq 0 : X_n(\omega) \in A\}$$

is called the hitting time of a subset A of I with the convention $\inf \emptyset = \infty$. The probability starting from state i that $(X_n)_{n\geq 0}$ ever hits A is defined as

$$h_i^A = \mathbb{P}_i[H^A < \infty]$$

and if A is a closed class, then we call h_i^A absorption probability. The mean time taken for $(X_n)_{n\geq 0}$ to reach A is given by

$$k_i^A = \mathbb{E}_i[H^A] = \sum_{n<\infty} n\mathbb{P}[H^A = n] + \infty\mathbb{P}[H^A = \infty].$$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Theorem 6.1.13.

The vector of hitting probabilities $h^A = (h_i^A : i \in I)$ is the minimal non-negative solution to the system of linear equations

$$h_i^A = \begin{cases} 1 & \forall i \in A \\ \sum_{j \in I} p_{ij} h_j^A & \forall i \notin A \end{cases}$$

Theorem 6.1.14.

The vector of expected hitting times $k^A = (k_i^A : i \in I)$ is the minimal non-negative solution in $\mathbb{R}^+ \cup \{\infty\}$ to the system of linear equations

$$k_i^A = \begin{cases} 0 & \forall i \in A \\ 1 + \sum_{j \notin A} p_{ij} k_j^A & \forall i \notin A \end{cases}$$

Definition 6.1.15 (Stopping Time for Markov Chains)**.**

We say that a random variable $T : \Omega \to \mathbb{Z}^+ \cup \{\infty\}$ is a stopping time for $(X_n)_{n \geq 0}$, if the event $\{T = n\}$ only depends on $X_0, \ldots, X_n$ $\forall n \geq 0$, meaning that the event is measurable with respect to a σ-algebra, denoting by

$$\{T = n\} \in \sigma(X_0, \ldots, X_n) \quad \forall n \geq 0.$$

Theorem 6.1.16 (Strong Markov Property for Markov Chains)**.**

Let $(X_n)_{n \geq 0} \sim Markov(\lambda, P)$ and T be a stopping time for $(X_n)_{n \geq 0}$. Then conditionally on $T < \infty$ and $\{X_T = i\}$ it holds that the process is $(X_{T+n})_{n \geq 0} \sim Markov(\delta_i, P)$ and is conditionally independent of the random variables $X_0, \ldots, X_T$.

6.2 Reinforcement Learning Problem

The interaction between an agent also known as decision maker or controller and the environment also known as controlled system is the key concept in a reinforcement learning problem. We assume that on the one side the agent completely knows and controls everything inside him but on the other side everything that is outside the agent, that is the environment, is not completely controllable but can be known by the agent.

The goal of the agent is to select the best actions in order to maximize the total expected cumulative reward which is a fundamental quantity in reinforcement learning. These rewards R_t that in fact are numbers are indicating how well the agent is doing at time t meaning that they are the basis for evaluating the choices made by the agent.

So more specifically at each time step $t \in \{0, 1, \ldots, T\}$ the agent is responsible for executing actions $A_t \in \mathcal{A}(S_t)$ based on some representation of the environment's state $S_t \in \mathcal{S}$, where $\mathcal{S}$ is denoting the set of possible states and $\mathcal{A}(S_t)$ is the set of actions which are available in state S_t. The agent is then receiving a reward $R_{t+1} \in \mathcal{R} \subset \mathbb{R}$ at time $t+1$ as a consequence of the chosen action at time t and finds himself in a new state S_{t+1}.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Furthermore a policy is describing the agent's behavior for every time step and is denoted by π_t. It describes a mapping from states, which are the basis for making the choices, to probabilities of selecting each possible action. We are distinguishing between two types of policies:

(i) deterministic policy: $\pi_t(s) = a$

(ii) stochastic policy: $\pi_t(a \mid s) = \mathbb{P}[A_t = a \mid S_t = s]$

The next central quantity in our reinforcement learning problem is the value function which is used to predict the future expected reward. Thus, we can say that the value function is a tool for selecting between actions as it tells us how good each state and action is. Therefore the goodness but also the badness of the states can be evaluated with the value function and we get, depending on the current policy π, that

$$V_\pi(s) = \mathbb{E}_\pi \left[\sum_{k=0}^{\infty} \gamma^k R_{t+1+k} \mid S_t = s \right] \tag{6.1}$$

We call (6.1) the state-value function for a policy π and a discount rate $0 \leq \gamma \leq 1$ which gives us the expected discounted return when starting in a state s and following the policy π thereafter.

Similarly, we get the expected discounted return when starting in a state s by taking the action a and thereafter following the policy π as

$$Q_\pi(s, a) = \mathbb{E}_\pi \left[\sum_{k=0}^{\infty} \gamma^k R_{t+1+k} \mid S_t = s, A_t = a \right] \tag{6.2}$$

We call (6.2) the action-value function for a policy π. Monte Carlo methods can be used to estimate the value functions in (6.1) and (6.2) for example.

A Markov decision process is a task in reinforcement learning that satisfies the Markov property. If the set of possible states $\mathcal{S}$ and the set of actions $\mathcal{A}(S_t)$ is finite then we say it is a finite Markov decision process, which are more common in the reinforcement learning theory.

In dynamic programming but also in reinforcement learning it is common to use recursive representations of the value function. This fundamental property holds for any policy π and state s as follows

$$\begin{aligned}
V_\pi(s) &= \mathbb{E}_\pi \left[\sum_{k=0}^{\infty} \gamma^k R_{t+1+k} \mid S_t = s \right] \\
&= \mathbb{E}_\pi \left[R_{t+1} + \gamma \sum_{k=0}^{\infty} \gamma^k R_{t+2+k} \mid S_t = s \right] \\
&= \sum_a \pi(a \mid s) \sum_{s'} \sum_r p(s', r \mid s, a) \left[r + \gamma \mathbb{E}_\pi \left[\sum_{k=0}^{\infty} \gamma^k R_{t+2+k} \mid S_{t+1} = s' \right] \right] \\
&= \sum_a \pi(a \mid s) \sum_{s', r} p(s', r \mid s, a) \left[r + \gamma V_\pi(s') \right]
\end{aligned} \tag{6.3}$$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

where (6.3) is the Bellman equation describing the relationship between the value of the state s and the value of its successor states. For any given state s and action a

$$p(s', r \mid s, a) = \mathbb{P}[R_{t+1} = r, S_{t+1} = s' \mid S_t, A_t] \tag{6.4}$$

is denoting the probability of each possible pair of the next state s' and reward r for which the dynamics of a finite Markov decision process are completely specified. Using (6.4) we can compute the expected rewards for state-action pairs defined as

$$r(s, a) := \mathbb{E}[R_{t+1} \mid S_t = s, A_t = a] = \sum_{r \in \mathcal{R}} r \sum_{s' \in \mathcal{S}} p(s', r \mid s, a),$$

the state-transition probabilities defined as

$$p(s' \mid s, a) := \mathbb{P}[S_{t+1} = s' \mid S_t = s, A_t = a] = \sum_{r \in \mathcal{R}} p(s', r \mid s, a),$$

and the expected rewards for state-action-next-state triples defined as

$$r(s, a, s') := \mathbb{E}[R_{t+1} \mid S_t = s, A_t = a, S_{t+1} = s'] = \frac{\sum_{r \in \mathcal{R}} r \cdot p(s', r \mid s, a)}{p(s' \mid s, a)}.$$

Now we have to define when a policy is better than another policy. When considering a finite Markov decision process we say that a policy π is going to be better than or equal to a policy π' if its expected return is greater than or equal to that of π' for all states, meaning that

$$\pi \geq \pi' \quad \Leftrightarrow \quad V_\pi(s) \geq V_{\pi'}(s) \quad \forall s \in \mathcal{S}. \tag{6.5}$$

The optimal policy $\pi^\star$ is the policy that is better than or equal to all other policies. The optimal state-value function is therefore defined as

$$V^\star(s) = \max_\pi V_\pi(s) \quad \forall s \in \mathcal{S} \tag{6.6}$$

and the optimal action-value function is defined as

$$Q^\star(s, a) = \max_\pi Q_\pi(s, a) \quad \forall s \in \mathcal{S}, a \in \mathcal{A}(s). \tag{6.7}$$

Note that there can be more than one optimal policy whereas for a given Markov decision process the optimal value function for states and state–action pairs are unique. Any policy that is greedy with respect to the optimal value functions must be an optimal policy.

The next equation delivers for the state-action pair (s, a) the expected discounted return when starting in state s by taking the action a and thereafter following the optimal policy $\pi^\star$

$$Q^\star(s, a) = \mathbb{E}_{\pi^\star}[R_{t+1} + \gamma V^\star(S_{t+1}) \mid S_t = s, A_t = a].$$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

The value function in (6.6) must fulfill (6.3) but due to the fact that it is the optimal value function we can rewrite (6.3) as follows

$$\begin{aligned}
V^\star(s) &= \max_{a\in\mathcal{A}(s)} Q_{\pi^\star}(s,a) \\
&= \max_a \mathbb{E}_{\pi^\star}\left[\sum_{k=0}^{\infty}\gamma^k R_{t+1+k} \mid S_t = s, A_t = a\right] \\
&= \max_a \mathbb{E}_{\pi^\star}\left[R_{t+1} + \gamma\sum_{k=0}^{\infty}\gamma^k R_{t+2+k} \mid S_t = s, A_t = a\right] \\
&= \max_a \mathbb{E}_{\pi^\star}\left[R_{t+1} + \gamma V^\star(S_{t+1}) \mid S_t = s, A_t = a\right] \\
&= \max_{a\in\mathcal{A}(s)} \sum_{s',r} p(s',r \mid s,a)\left[r + \gamma V^\star(s')\right] \qquad (6.8)
\end{aligned}$$

which is known as the Bellman optimality equation expressing that the value of a state under an optimal policy must equal the expected return for the best action from that state. Similarly, the Bellman optimality equation for $Q^\star$ is

$$\begin{aligned}
Q^\star(s,a) &= \mathbb{E}_{\pi^\star}\left[R_{t+1} + \gamma\max_{a'} Q^\star(S_{t+1},a') \mid S_t = s, A_t = a\right] \\
&= \sum_{s',r} p(s',r \mid s,a)\left[r + \gamma\max_{a'} Q^\star(s',a'))\right]. \qquad (6.9)
\end{aligned}$$

In order to be able to discover the optimal policy we will show how dynamic programming can be used to compute the value functions in (6.6) and (6.7) which satisfy the Bellman optimality equations in (6.8) and (6.9) respectively. Therefore, we assume that the environment is a finite Markov decision process for which its state, action and reward sets $\mathcal{S}, \mathcal{A}(s)$ and $\mathcal{R}$ for $s \in \mathcal{S}$ are finite.

As we are going to use the policy iteration method to compute the optimal value function for the problem in Chapter 5 and therefore to find the optimal policy, we start with an arbitrary initial policy π_0. After each of the k-iterations of the algorithm we determine a better policy in the sense of (6.5) such that a sequence of monotonically improving policies is created. This will be repeated until we find the optimal policy.

In general the algorithm can be divided into three steps:

1. Initialization

2. Policy Evaluation:
   In this step we want to evaluate the given policy by calculating the value function under the considered policy. So it tells us how good the current policy actually is and how much reward we will collect with this policy. Therefore we have to solve a prediction problem with

   - Input: Markov decision process and policy π
   - Output: Value function V_π

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

3. Policy Improvement:
   In this step we want to improve the considered policy π by acting greedily with respect to the evaluated policy V_π from the previous step, meaning that we are interested in picking actions which give us the maximum action-value function for each state by following the policy π

$$\pi'(s) = greedy(V_\pi) = \underset{a \in \mathcal{A}}{\operatorname{argmax}}\, Q_\pi(s, a)$$

   This will be repeated until we find the best policy for which we get the most reward amongst all policies. Therefore we have to use the Bellman optimality equations to solve a control problem with

   - Input: Markov decision process
   - Output: Optimal value function $V^\star$ and optimal policy $\pi^\star$

Now we are going to consider the second and third step of the policy iteration algorithm in more detail.

6.2.1 Policy Evaluation

After the choice of an arbitrary policy π we will calculate the state-value function $V_\pi(s)$ for all states $s \in \mathcal{S}$ under the current policy π with the aim to evaluate the given policy as follows

$$\begin{aligned} V_\pi(s) &= \mathbb{E}_\pi[R_{t+1} + \gamma V_\pi(S_{t+1}) \mid S_t = s] \\ &= \sum_a \pi(a \mid s) \sum_{s',r} p(s', r \mid s, a) \left[r + \gamma V_\pi(s')\right], \end{aligned} \tag{6.10}$$

where $\pi(a, s)$ is denoting the probability when the agent is taking action a in state s under the policy π. The expected values are indexed here with the policy to show the dependence of π being followed. The existence and uniqueness of the value function V_π can be assured, on the one hand, by the fact that $\gamma < 1$ or, on the other hand, by the fact that a possible termination is guaranteed by all states under the policy.

When assuming that the environment's dynamics are fully known then (6.10) is a system of $|\mathcal{S}|$ linear equations in $|\mathcal{S}|$ unknowns, that are $V_\pi(s), s \in \mathcal{S}$. For these calculations iterative solution methods are suited best. Therefore we are taking the Bellman expectation equation for V_π in (6.3) and turn it into an iterative update which will give us a mechanism for evaluating policies. So we get

$$\begin{aligned} V_{k+1}(s) &= \mathbb{E}_\pi\left[R_{t+1} + \gamma V_k(S_{t+1}) \mid S_t = s\right] \\ &= \sum_a \pi(a \mid s) \sum_{s',r} p(s', r \mid s, a) \left[r + \gamma V_k(s')\right] \end{aligned}$$

where V_0 is arbitrary initial value function, for example $V_0 = 0$ for all $s \in \mathcal{S}$. This iteration will give us a sequence of value functions $\{V_k\}$, meaning that

$$V_0 \to V_1 \to \ldots \to V_\pi$$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

where $V_k = V_\pi$ is a fixed point of this iterative update and in general it can be shown that for $k \to \infty$ the sequence converges to V_π. This algorithm is called iterative policy evaluation.

In order to generate approximations $V_k + 1$ from V_k by using iterative policy evaluation the same operations are done to each state $s \in \mathcal{S}$. Thereby the old value of the state s is replaced by a new one, which in turn is obtained by old values of the successor states s' of s and the expected immediate rewards possible along all one-step transitions under the policy being evaluated, meaning that

- Update $V_{k+1}(s)$ from $V_k(s')$ at each iteration $k+1$ for all states $s \in \mathcal{S}$

An algorithm for iterative policy evaluation with a stopping criterion is stated as follows:

Iterative Policy Evaluation

Input: policy π to be evaluated
Initializiation of an arbitrary value function $\hat{V}(s) = 0$ for all states
Repeat
  $\Delta \leftarrow 0$
  For each $s \in \mathcal{S}$:
    $V \leftarrow \hat{V}(s)$
    $\hat{V}(s) \leftarrow \sum_a \pi(a \mid s) \sum_{s',r} p(s', r \mid s, a) \left[r + \gamma \hat{V}(s')\right]$
    $\Delta \leftarrow \max(\Delta, |V - \hat{V}(s)|)$
until $\Delta < \theta$ which is a small positive number
Output: $\hat{V} \approx V_\pi$

6.2.2 Policy Improvement

After the computation of the value function V_π with respect to the arbitrary policy π from Subsection 6.2.1 we are interested in finding better policies in the sense of (6.5). Assume that we have determined the value function V_π for an arbitrary deterministic policy π. Our interest now lies in deciding if it would be better or worse when taking an action $a \neq \pi(s)$ in state s and thereafter following the old policy π.

The action-value function in this case can be determined as

$$\begin{aligned} Q_\pi(s,a) &= \mathbb{E}_\pi[R_{t+1} + \gamma V_\pi(S_{t+1}) \mid S_t = s, A_t = a] \\ &= \sum_{s',r} p(s', r \mid s, a) \left[r + \gamma V_\pi(s')\right] \end{aligned}$$

If this value is greater than the old value V_π, meaning that selecting a once in s and then following π is better than following π all the time, then it seems to be logical to execute this action a every time we are in state s and therefore to replace the old policy π by the new one as a whole. This special case of the policy improvement theorem is in fact true.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

We assume now that for two deterministic policies π and π' it holds for all states $s \in \mathcal{S}$ that

$$Q_\pi(s, \pi'(s)) \geq V_\pi(s) \tag{6.11}$$

which implies, that the policy π' is as good as or better than the policy π. Thus, it holds that

$$V_{\pi'}(s) \geq V_\pi(s) \tag{6.12}$$

meaning that we obtain greater or equal expected return from all states under the policy π'. If (6.11) holds with strict inequality at any state then (6.12) holds with strict inequality at least one state. This is true for the considered policies π and π' which is identical to the original policy except that $\pi'(s) = a \neq \pi(s)$. The new policy is of course better than the old if the following holds

$$Q_\pi(s, a) > V_\pi(s)$$

To show (6.12) we use (6.11) several times until we fet the desired inequality.

$$\begin{aligned}
V_\pi(s) &\leq Q_\pi(s, \pi'(s)) \\
&= \mathbb{E}_{\pi'}[R_{t+1} + \gamma V_\pi(S_{t+1}) \mid S_t = s] \\
&\leq \mathbb{E}_{\pi'}[R_{t+1} + \gamma Q_\pi(S_{t+1}, \pi'(S_{t+1})) \mid S_t = s] \\
&= \mathbb{E}_{\pi'}[R_{t+1} + \gamma \mathbb{E}_{\pi'}[R_{t+2} + \gamma V_\pi(S_{t+2})] \mid S_t = s] \\
&= \mathbb{E}_{\pi'}[R_{t+1} + \gamma R_{t+2} + \gamma^2 V_\pi(S_{t+2}) \mid S_t = s] \\
&\leq \mathbb{E}_{\pi'}[R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \gamma^3 V_\pi(S_{t+3}) \mid S_t = s] \\
&\vdots \\
&\leq \mathbb{E}_{\pi'}[R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \gamma^3 R_{t+4} + \cdots \mid S_t = s] \\
&= V'_\pi(s)
\end{aligned}$$

As we have shown for a given policy and its value function that we can evaluate a change in the policy at a single state to a particular action, it is also possible to extend this behavior to all states and all actions. Therefore we are interested in picking at each state the action which appears best according to the action-value function which leads to a new greedy policy π' given by

$$\begin{aligned}
\pi'(s) &= \operatorname*{argmax}_a Q_\pi(s, a) \\
&= \operatorname*{argmax}_a \mathbb{E}[R_{t+1} + \gamma V_\pi(S_{t+1}) \mid S_t = s, A_t = a] \\
&= \operatorname*{argmax}_a \sum_{s',r} p(s', r \mid s, a) \left[r + \gamma V_\pi(s')\right]
\end{aligned} \tag{6.13}$$

where the value of action a at which the expression that follows is maximized is denoted by $\operatorname*{argmax}_a$. The greedy policy π' fulfills (6.11) and is therefore as good as or better than the

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

original policy. This process of finding a new policy improving on an old policy by making it greedy with respect to V_π is called policy improvement.

Assume now that the new greedy policy π' is not better than but as good as the old policy π which implies that $V_\pi = V_{\pi'}$. We obtain from (6.13) for all states $s \in \mathcal{S}$ that

$$\begin{aligned} V_{\pi'}(s) &= \max_a \mathbb{E}[R_{t+1} + \gamma V_{\pi'}(S_{t+1}) \mid S_t = s, A_t = a] \\ &= \max_a \sum_{s',r} p(s', r \mid s, a) \left[r + \gamma V_{\pi'}(s')\right] \end{aligned}$$

which is equal to the Bellman optimality equation in (6.8) which implies that $V_{\pi'} = V^\star$ and both policies π and π' are optimal policies.

If we consider stochastic policies $\pi(a \mid s)$ instead, the ideas of this section can be easily extended. We get for the policy improvement theorem in the stochastic case that

$$Q_\pi(s, \pi'(s)) = \sum_a \pi'(a \mid s) Q_\pi(s, a)$$

Moreover, if in the stochastic case the maximum is achieved for several actions then we are not forced to choose a single action amongst them. We can give these actions a portion of the probability of being selected in the new greedy policy, instead.

6.2.3 Policy Iteration

We can create a sequence of monotonically improving policies such that

$$\pi_0 \xrightarrow{E} V_{\pi_0} \xrightarrow{I} \pi_1 \xrightarrow{E} V_{\pi_1} \xrightarrow{I} \pi_2 \xrightarrow{E} \cdots \xrightarrow{I} \pi^\star \xrightarrow{E} V^\star$$

where $\xrightarrow{E}$ denotes a policy evaluation and $\xrightarrow{I}$ a policy improvement by using the value function V_{π_0} to get a better policy π_1 than the original policy π_0 in the first step, followed by using the value function V_{π_1} to get a better policy π_2 than the previous policy π_1 and so on. This will be repeated until we find the optimal policy. Because of our assumption that we are considering a finite Markov decision process this process must converge to an optimal policy and optimal value function in a finite number of iterations. This algorithm is called policy iteration and is summarized below.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Policy Iteration Algorithm

1. Initialization
   $\hat{V}(s) \in \mathbb{R}$ and $\pi(s) \in \mathcal{A}(s)$ for all $s \in \mathcal{S}$

2. Policy Evaluation
   Repeat
   $\Delta \leftarrow 0$
   For each $s \in \mathcal{S}$:
   $V \leftarrow \hat{V}(s)$
   $\hat{V}(s) \leftarrow \sum_{s',r} p(s', r \mid s, \pi(s)) \left[r + \gamma \hat{V}(s')\right]$
   $\Delta \leftarrow \max(\Delta, |V - \hat{V}(s)|)$
   until $\Delta < \theta$ which is a small positive number

3. Policy Improvement
   policy-stable $\leftarrow$ *true*
   For each $s \in \mathcal{S}$:
   $a \leftarrow \pi(s)$
   $\pi(s) \leftarrow \operatorname{argmax}_a \sum_{s',r} p(s', r \mid s, a) \left[r + \gamma \hat{V}(s')\right]$
   If $a \neq \pi(s)$, then *policy-stable* $\leftarrow$ *false*
   If *policy-stable* $\leftarrow$ *true*, then stop and return $\hat{V}$ and π, else go to 2.

In conclusion, we have assumed, as it is usual for dynamic programming methods, that we have a perfect model of the environment's Markov decision process. In the policy evaluation step we are calculating the state-value function $V(s)$ for a given policy π. At each state s, we look ahead one step at each possible action a and next state s', which can only be done because of our assumption to have a perfect model of the environment. In the next step the so-called policy improvement step we can act greedily with respect to a given and correct state-value function for a policy meaning that we are picking the best action at each state. Then we are guaranteed to improve the policy or keep it fixed if it's already optimal. In the last step we iteratively perform the previous steps until we reach the optimal policy which is known as policy iteration.

6.3 An Illustrative Example

In this section we are going to illustrate the policy iteration algorithm in hopes for better understanding through an example. Let us assume that our goal is to find the optimal policy that maximizes the long-term discounted reward.

Furthermore, we assume that a person is in possession of an old treasure map. Since the person encounters several forks in the path over time and every option has its own rewards and risks, he or she has several ways to reach the goal. Based on the route choices, the person faces different places that are marked with a red cross on the map. The person must therefore decide which way is the best in the middle of risky and advantageous situations.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

6.3.1 Overview

Suppose that a person has already set out on a treasure hunt and is about to reach the first fork in the path. By taking the route to the North, the person will arrive to a marked place on the map, where he or she can find a flashlight. By then taking the route to the South, the person will get to a dark cave where the treasure can be found with the flashlight. However, if the person decides to go north after collecting the flashlight then he or she will get into a swamp where no rescue is possible and it will end with the death of the person.

If the person takes the route to the South instead, there will be a matchbox on the marked spot on the map which can also be used to find the treasure when taking the route to the North afterwards. However, moving South after collecting the matchbox will lead to a trap where the person gets injured.

As we are interested in stochasticity, we say, every time, the person decides to make a move towards north, he or she takes the route to the North with a probability of 0.7. However, the person might not always know where north and south lies and therefore can also take the wrong route and reach south instead with a probability of 0.3.

Similarly, if the person is interested to take the route to the South, he or she makes a move towards south with a probability of 0.7 and gets north with a probability of 0.3.

In case of reaching the marked spots on the treasure map we want to assign some rewards to each spot:

- Reaching the cave and finding the treasure: +1.5 points
- Reaching the spot where the flashlight can be found: +1 point
- Reaching the spot where the matchbox can be found: +0.5 points
- If the person gets into a swamp and therefore dies: -2 points
- If the person gets trapped and therefore injured: -1 point

So our aim is to reach the cave and find the treasure with maximum rewards with the considered reward set $\mathcal{R} = \{-2, -1, \frac{1}{2}, 1, \frac{3}{2}\}$.

6.3.2 State Transition Diagram & Transition Probability Matrix

Now we are going to define the state transition diagram and the transition probability matrix. To be able to build our state transition diagram we have to define our states and actions. In our example we have two actions, north and south, and six states which we are going to label as follows

(i) Start state S_1

(ii) State of reaching the marked spot where the flashlight can be found S_2

(iii) State of reaching the marked spot where the matchbox can be found S_3

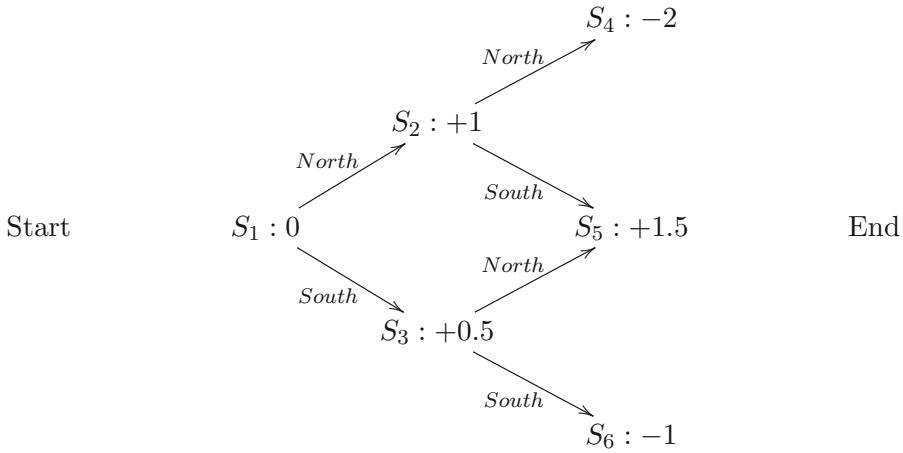
(iv) State of reaching the marked spot where the person gets into a swamp and dies S_4

(v) State of reaching the cave and finding the treasure S_5

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

(vi) State of reaching the marked spot where the person gets trapped and injured S_6

where the reward for the start state S_1 is zero. Therefore we get our state transition diagram



With the state transition diagram and the probabilities in Subsection 6.3.1 we can now calculate the transition probabilities for all states. As there are two actions in our example we are going to compute the transition probability matrix for both actions:

1. Transition Probability Matrix for Action North:

$$P[A(North)] \begin{bmatrix} & S_1 & S_2 & S_3 & S_4 & S_5 & S_6 \\ S_1 & 0 & 0.7 & 0.3 & 0 & 0 & 0 \\ S_2 & 0 & 0 & 0 & 0.7 & 0.3 & 0 \\ S_3 & 0 & 0 & 0 & 0 & 0.7 & 0.3 \\ S_4 & 0 & 0 & 0 & 1 & 0 & 0 \\ S_5 & 0 & 0 & 0 & 0 & 1 & 0 \\ S_6 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

2. Transition Probability Matrix for Action South:

$$P[A(South)] \begin{bmatrix} & S_1 & S_2 & S_3 & S_4 & S_5 & S_6 \\ S_1 & 0 & 0.3 & 0.7 & 0 & 0 & 0 \\ S_2 & 0 & 0 & 0 & 0.3 & 0.7 & 0 \\ S_3 & 0 & 0 & 0 & 0 & 0.3 & 0.7 \\ S_4 & 0 & 0 & 0 & 1 & 0 & 0 \\ S_5 & 0 & 0 & 0 & 0 & 1 & 0 \\ S_6 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Here $p(S, a, S')$ denotes the transition probabiity when moving from state S to state S' after executing an action a when Bellman's equation in (6.14) is applied.

In summary, we are considering the finite state set $\mathcal{S}^+ = \{S_1, \ldots, S_6\}$ but due to the fact that the states S_4, S_5 and S_6 are terminal states with $V(S_t) = 0$ for $t \in \{4, 5, 6\}$ we restrict

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

ourselves to the state set $\mathcal{S} = \{S_1, S_2, S_3\}$. Furthermore, the associated set of actions for every time step t is given as $\mathcal{A}(S_t) = \{South, North\} = \{S, N\}$.

Now we are going to give a short review about the policy iteration algorithm which can be divided into three steps:

1. **Initialization of a random policy**:
   $V_k(s) = 0 \ \forall s \in \mathcal{S}$ for iteration step $k = 0$

2. **Policy Evaluation**:
   Based on Bellman's equation

$$V_{k+1}(s) = \max_{a \in \{S,N\}} \sum_{s',r} p(s', r \mid s, a) \left[r + \gamma V_k(s') \right] \tag{6.14}$$

   we get an action a for every state s in the policy and evaluate the value function V, where p is denoting the transition probability.

3. **Policy Improvement**:
   Using

$$\pi(s) = \operatorname*{argmax}_a \sum_{s',r} p(s', r \mid s, a) \cdot V(s') \tag{6.15}$$

   we get the best action from the value function for every state and if it is better than the present policy action, then we are replacing the current action by the best action.

We are iterating step 2 and step 3 until the policy did not change after an iteration. If this is the case then we can assume that the policy iteration algorithm has converged.

6.3.3 Finding the Optimal Policy

Now we are going to apply the policy iteration algorithm in Subsection 6.3.1 step by step by setting $\gamma = 1$. According to the policy iteration algorithm, we start with an arbitrarly policy, for example

$$\pi_0 = \{North, North, North, North, North, North\}$$

meaning that we are going north for all states, for which $V_{\pi_0}(s) = 0 \ \forall s \in \mathcal{S}$. According to the state transition diagram we can see that the states S_4, S_5 and S_6 are end states and therefore no actions will be taken. So we are going to restrict our policy and focus only to the first three states where we can perform an action:

$$\pi_0 = \{North, North, North\}.$$

According to the policy evaluation step we have

$$V_{k+1}(s) = \max_{a \in \{S,N\}} \sum_{s' \in \mathcal{S}} \sum_{r \in \mathcal{R}} p(s', r \mid s, a) \left[r + \gamma V_k(s') \right],$$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

implying the following three equations when applying the above equation for our considered state and action set $\mathcal{S}$ and $\mathcal{A}(S_t)$ for $k = 0$

$$
\begin{aligned}
V_{k+1}(S_1) &= \max \left\{ \begin{matrix} p(S_2,1|S_1,N)[1+V_k(S_2)]+p(S_3,\frac{1}{2}|S_1,N)[\frac{1}{2}+V_k(S_3)]; \\ p(S_2,1|S_1,S)[1+V_k(S_2)]+p(S_3,\frac{1}{2}|S_1,S)[\frac{1}{2}+V_k(S_3)] \end{matrix} \right\} \\
V_{k+1}(S_2) &= \max \left\{ \begin{matrix} p(S_4,-2|S_2,N)[-2+V_k(S_4)]+p(S_5,\frac{3}{2}|S_2,N)[\frac{3}{2}+V_k(S_5)]; \\ p(S_4,-2|S_2,S)[-2+V_k(S_4)]+p(S_5,\frac{3}{2}|S_2,S)[\frac{3}{2}+V_k(S_5)] \end{matrix} \right\} \\
V_{k+1}(S_3) &= \max \left\{ \begin{matrix} p(S_5,\frac{3}{2}|S_3,N)[\frac{3}{2}+V_k(S_5)]+p(S_6,-1|S_3,N)[-1+V_k(S_6)]; \\ p(S_5,\frac{3}{2}|S_3,S)[\frac{3}{2}+V_k(S_5)]+p(S_6,-1|S_3,S)[-1+V_k(S_6)] \end{matrix} \right\}.
\end{aligned}
$$

When applying the transition probabilities and the fact that $V_k(s) = 0 \ \forall s \in \mathcal{S}, k = 0$ we derive the following

$$
\begin{aligned}
V_1(S_1) &= \max \left\{ 0.7 \cdot 1 + 0.3 \cdot \frac{1}{2}; 0.3 \cdot 1 + 0.7 \cdot \frac{1}{2} \right\} \\
&= \max\{0.85; 0.65\} \\
&= 0.85 \quad \Rightarrow N \\
V_1(S_2) &= \max \left\{ 0.7 \cdot (-2) + 0.3 \cdot \frac{3}{2}; 0.3 \cdot (-2) + 0.7 \cdot \frac{3}{2} \right\} \\
&= \max\{-0.95; 0.45\} \\
&= 0.45 \quad \Rightarrow S \\
V_1(S_3) &= \max \left\{ 0.7 \cdot \frac{3}{2} + 0.3 \cdot (-1); 0.3 \cdot \frac{3}{2} + 0.7 \cdot (-1) \right\} \\
&= \max\{0.75; -0.25\} \\
&= 0.75 \quad \Rightarrow N.
\end{aligned}
$$

In summary, we get for the first iteration in the policy evaluation part of the algorithm

$$
\begin{matrix} V_1[S_1] = 0.85 & V_1[S_2] = 0.45 & V_1[S_3] = 0.75 \\ V_1[S_4] = 0 & V_1[S_5] = 0 & V_1[S_6] = 0 \end{matrix} \tag{6.16}
$$

and for the policy improvement part we get in (6.15)

State	Action	Value Function
S_1	*North*	$V[S_1] = 0.7 \cdot 1 + 0.3 \cdot 0.5 = 0.85$
	South	$V[S_1] = 0.3 \cdot 1 + 0.7 \cdot 0.5 = 0.65$
S_2	*North*	$V[S_2] = 0.7 \cdot (-2) + 0.3 \cdot 1.5 = -0.95$
	South	$V[S_2] = 0.3 \cdot (-2) + 0.7 \cdot 1.5 = 0.45$
S_3	*North*	$V[S_3] = 0.7 \cdot 1.5 + 0.3 \cdot (-1) = 0.75$
	South	$V[S_3] = 0.3 \cdot 1.5 + 0.7 \cdot (-1) = -0.25$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

So we get for state S_1 that the route to the north, for state S_2 that the route to the south and for state S_3 that the route to the north is the best, which implies our new policy according to the best actions

$$\pi_1 = \{North, South, North\}.$$

For the second iteration of the policy evaluation part of the algorithm the value function V for each state is again calculated according to (6.14). Therefore we get for $k = 1$

$$\begin{aligned}
V_{k+1}(S_1) &= \max \begin{Bmatrix} p(S_2,1|S_1,N)[1+V_k(S_2)]+p(S_3,\frac{1}{2}|S_1,N)[\frac{1}{2}+V_k(S_3)]; \\ p(S_2,1|S_1,S)[1+V_k(S_2)]+p(S_3,\frac{1}{2}|S_1,S)[\frac{1}{2}+V_k(S_3)] \end{Bmatrix} \\
V_{k+1}(S_2) &= \max \begin{Bmatrix} p(S_4,-2|S_2,N)[-2+V_k(S_4)]+p(S_5,\frac{3}{2}|S_2,N)[\frac{3}{2}+V_k(S_5)]; \\ p(S_4,-2|S_2,S)[-2+V_k(S_4)]+p(S_5,\frac{3}{2}|S_2,S)[\frac{3}{2}+V_k(S_5)] \end{Bmatrix} \\
V_{k+1}(S_3) &= \max \begin{Bmatrix} p(S_5,\frac{3}{2}|S_3,N)[\frac{3}{2}+V_k(S_5)]+p(S_6,-1|S_3,N)[-1+V_k(S_6)]; \\ p(S_5,\frac{3}{2}|S_3,S)[\frac{3}{2}+V_k(S_5)]+p(S_6,-1|S_3,S)[-1+V_k(S_6)] \end{Bmatrix}.
\end{aligned}$$

When applying the transition probabilities and the calculated V_k in (6.16) for $k = 1$ we derive the following

$$\begin{aligned}
V_2(S_1) &= \max\left\{0.7\cdot(1+0.45)+0.3\cdot\left(\frac{1}{2}+\frac{3}{4}\right);0.3\cdot(1+0.45)+0.7\cdot\left(\frac{1}{2}+\frac{3}{4}\right)\right\} \\
&= \max\{1.39;1.31\} \\
&= 1.39 \quad \Rightarrow N \\
V_2(S_2) &= \max\left\{0.7\cdot(-2+0)+0.3\cdot\left(\frac{3}{2}+0\right);0.3\cdot(-2+0)+0.7\cdot\left(\frac{3}{2}+0\right)\right\} \\
&= \max\{-0.95;0.45\} \\
&= 0.45 \quad \Rightarrow S \\
V_2(S_3) &= \max\left\{0.7\cdot\left(\frac{3}{2}+0\right)+0.3\cdot(-1+0);0.3\cdot\left(\frac{3}{2}+0\right)+0.7\cdot(-1+0)\right\} \\
&= \max\{0.75;-0.25\} \\
&= 0.75 \quad \Rightarrow N.
\end{aligned}$$

In summary, we get for the second iteration in the policy evaluation part of the algorithm

$$\begin{aligned}
&V_2[S_1] = 1.39 \qquad V_2[S_2] = 0.45 \qquad V_2[S_3] = 0.75 \\
&V_2[S_4] = 0 \qquad V_2[S_5] = 0 \qquad V_2[S_6] = 0
\end{aligned}$$

and for the policy improvement part we get in (6.15)

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

State	Action	Value Function
S_1	$North$	$V[S_1] = 0.7 \cdot 1.45 + 0.3 \cdot 1.25 = 1.39$
	$South$	$V[S_1] = 0.3 \cdot 1.45 + 0.7 \cdot 1.25 = 1.31$
S_2	$North$	$V[S_2] = 0.7 \cdot (-2) + 0.3 \cdot 1.5 = -0.95$
	$South$	$V[S_2] = 0.3 \cdot (-2) + 0.7 \cdot 1.5 = 0.45$
S_3	$North$	$V[S_3] = 0.7 \cdot 1.5 + 0.3 \cdot (-1) = 0.75$
	$South$	$V[S_3] = 0.3 \cdot 1.5 + 0.7 \cdot (-1) = -0.25$

So we get for the second iteration that for state S_1 again the route to the north, for state S_2 again the route to the south and for state S_3 again the route to the north is the best, which implies our new policy according to the best actions

$$\pi_2 = \{North, South, North\}.$$

By comparing the policy π_2 with the obtained policy in the first iteration π_1, we can see that the policies did not change implying that our algorithm has converged because afterwards neither $V_k(S_j)$ for $k \geq 2$, $j \in 1, \ldots, 6$ nor the policies would change and therefore we have found the optimal policy for our example. In summary, we obtain from the algorithm that for the considered example the optimal policy is

$$\pi^\star = \{North, South, North\}$$

and the associated payoff is 1.39.

However, looking at the example, on the one side it might be tempting to go south first, because going north and finding the flashlight brings more points, but due to the fact that we could lose also two points when getting stuck into a swamp and die there the route to the north involves higher risks on the other side. Nevertheless, it is better to choose the northern route in order to maximize our reward in the long term which was our aim in this example.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

7 Numerical Analysis

In this chapter, we will try to verify the results obtained by the authors in Chapter 5 using the method explained in Chapter 6. In [18], McKean and Shepp generalized the underlying model by inserting a parameter α such that the following optimization problem was considered.

$$\max_{\pi} \alpha \cdot \mathbb{P}[\nu = 1] + (1 - \alpha) \cdot \mathbb{P}[\nu = 2], \tag{7.1}$$

where $\nu = \nu^{\pi} \in \{0, 1, 2\}$ is denoting how many companies survive forever under a non-anticipatory policy π.

The boundary conditions for the underlying partial differential equation for the value function V, which is stated as

$$\frac{1}{2}(V_{xx} + V_{yy}) + \max(V_x(x, y), V_y(x, y)) = 0, \tag{7.2}$$

are given as follows. If $\alpha = 0$, corresponding to the criterion to maximize the probability that both companies remain solvent, we get

$$\begin{aligned} V(x, 0) &= 0 \\ V(0, y) &= 0 \\ V(x, \infty) &= 1 - \exp(-2x) \\ V(\infty, y) &= 1 - \exp(-2y) \end{aligned}$$

and if $\alpha = \frac{1}{3}$, corresponding to the criterion to maximize the expected number of companies that remain solvent, the boundary conditions are given as

$$\begin{aligned} V(x, 0) &= \alpha(1 - \exp(-2x)) \\ V(0, y) &= \alpha(1 - \exp(-2y)) \\ V(x, \infty) &= (1 - \alpha)(1 - \exp(-2x)) + \alpha \exp(-2x) \\ V(\infty, y) &= (1 - \alpha)(1 - \exp(-2y)) + \alpha \exp(-2y). \end{aligned}$$

In order to find an approximate optimal payoff function we are not able to use standard methods for solving linear partial differential equations, as (7.2) is a non-linear partial differential equation. Therefore, we use the policy iteration method and solve (7.2) numerically by following the next steps.

1. First, we take an arbitrary strategy, for example

$$u_0 = u_0(x, y) = \mathbb{1}_{\{x \leq \frac{1}{2} y\}}, \tag{7.3}$$

   according to which the companies are supported on the basis of the addition of a drift $\mu = \mu_x + \mu_y = 1$, where x and y are denoting the wealth of the respective company. We distinguish the following two cases when assigning the drift μ to the companies:

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

- If $x \leq \frac{1}{2}y$, then support company x with $\mu_x = 1$ and $\mu_y = 0$.
- If $x > \frac{1}{2}y$, then support company y with $\mu_x = 0$ and $\mu_y = 1$.

With this fixed strategy u_0, we can then determine a linear partial differential equation for the survival probability $J = J(x, y; u_0)$ starting in x and y. The latter is

$$\frac{1}{2}\sigma_x^2 J_{xx} + \frac{1}{2}\sigma_y^2 J_{yy} + u_0(x,y)J_x + (\mu - u_0(x,y))J_y = 0, \tag{7.4}$$

where σ_x^2 and σ_y^2 are representing the diffusion coefficients of the respective companies. A heuristic derivation of (7.4) is given in Section 7.1.

2. Subsequently, the partial derivatives occurring in (7.4) are approximated by means of finite-differences. This procedure is introduced in Section 7.2.

3. Now we can try to determine a solution $J(x, y; u_0)$ of (7.4) by the method of successive over-relaxation (SOR). This iterative procedure is presented and explained in more detail in Section 7.3.

4. Finally, as the HJB equation can be written in the form as in (7.2), we get the first improved strategy u_1, by the policy iteration method, as

$$u_1 = u_1(x, y; u_0) = \mathbb{1}_{\{J_x(x,y;u_0) \geq J_y(x,y;u_0)\}}.$$

Analogously, we get $u_2, u_3, \ldots$, and

$$V(x, y) = \lim_{n\to\infty} J(x, y; u_n).$$

7.1 Heuristic Derivation of the PDE

For our two-dimensional ruin-problem we are considering the following partial differential equation

$$\frac{1}{2}\sigma_x^2 J_{xx} + \frac{1}{2}\sigma_y^2 J_{yy} + uJ_x + (1-u)J_y = 0,$$

where $u = u(x, y)$ is a strategy taking values in $[0, 1]$ and $\sigma_x^2 = \sigma_y^2 = 1$ is assumed for the sake of simplicity. The dynamics of our problem can be rewritten as

$$X_t^{x,u} = x + B_t + \int_0^t u(X_s^{x,u}, Y_s^{y,u})ds$$
$$Y_t^{y,u} = y + W_t + \int_0^t (1 - u(X_s^{x,u}, Y_s^{y,u}))ds,$$

where (B_t) and (W_t) are independent Brownian motion processes. The ruin time is defined as

$$\tau_{Ruin} = \tau_{Ruin}^{x,y;u} = \inf\{t \in [0, \infty) : X_t^{x,u} \leq 0 \vee Y_t^{y,u} \leq 0\}.$$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

The objective function $J(x,y)$ is the common survival function of X and Y depending on the strategy u, defined as

$$\begin{aligned} J(x,y) = J^u(x,y) = J(x,y;u) &= \mathbb{P}[\tau_{Ruin}^{x,y;u} = \infty] \\ &= \mathbb{P}[X_t^{x,u} > 0, Y_t^{y,u} > 0 \quad \forall t \in [0,\infty)]. \end{aligned}$$

Let τ be an arbitrary stopping time with $\tau < \infty, \mathbb{P}$-almost surely. Then it holds for $A := \{\tau_{Ruin}^{x,y;u} = \infty\}$ that

$$\begin{aligned} J(x,y) &= \mathbb{P}[X_t^{x,u} > 0, Y_t^{y,u} > 0 \, \forall t] = \mathbb{E}[\mathbb{1}_A] = \mathbb{E}[\mathbb{E}[\mathbb{1}_A \mid \mathcal{F}_\tau]] \\ &= \mathbb{E}[\mathbb{E}[\mathbb{1}_{\{\tau \geq \tau_{Ruin}\}}\mathbb{1}_A + \mathbb{1}_{\{\tau < \tau_{Ruin}\}}\mathbb{1}_A \mid \mathcal{F}_\tau]] \\ &= \mathbb{E}[\mathbb{1}_{\{\tau < \tau_{Ruin}\}}\mathbb{E}[\mathbb{1}_A \mid \mathcal{F}_\tau]] \\ &= \mathbb{E}[\mathbb{1}_{\{\tau < \tau_{Ruin}\}}\mathbb{E}[\mathbb{1}_{\left\{X_s^{X_\tau^{x,u}} > 0, Y_s^{Y_\tau^{y,u}} > 0 \, \forall s\right\}} \mid \mathcal{F}_\tau]] \\ &= \mathbb{E}[\mathbb{1}_{\{\tau < \tau_{Ruin}\}} J(X_\tau^{x,u}, Y_\tau^{y,u}; u)]. \end{aligned} \tag{7.5}$$

The fifth equality in (7.5) holds because of the fact that $\mathbb{1}_{\{\tau \geq \tau_{Ruin}\}}\mathbb{1}_A$ is zero as on $\{\tau \geq \tau_{Ruin}\}$ it holds that $\tau_{Ruin} < \infty$. Furthermore we used that $\{\tau < \tau_{Ruin}\} \in \mathcal{F}_\tau$ and on $\{\tau < \tau_{Ruin}\}$ ruin has not yet occurred until τ, in particular $X_t^{x,u} > 0, Y_t^{y,u} > 0$ hold for $t \leq \tau$. The sixth equality in (7.5) is the result when applying the strong Markov property.

Let us now assume that $J \in C^2((0,\infty) \times (0,\infty))$, $x, y > 0$ be fixed and $T > 0$. We define

$$\Theta = \inf\{t \in [0,\infty) : (X_t^{x,u}, Y_t^{y,u}) \notin \mathcal{B}_r(x,y)\},$$

with $r = \min\{\frac{x}{2}, \frac{y}{2}\}$. Note that for any other point y in a metric space (X,d) where the metric d defines the distance between x and y, $d(x,y)$, the open ball with radius r about x is defined as $\mathcal{B}_r(x) = \{y \in X : d(x,y) < r\}$. Then it holds that $T \wedge \Theta < \tau_{Ruin}$ and $T \wedge \Theta \leq T < \infty, \mathbb{P}$-almost surely. Therefore we have, when subtracting the left term in (7.5) on both sides,

$$\begin{aligned} 0 &= \mathbb{E}\left[\mathbb{1}_{\left\{T\wedge\Theta < \tau_{Ruin}\right\}} J(X_{T\wedge\Theta}^{x,u}, Y_{T\wedge\Theta}^{y,u}) - J(x,y)\right] \\ &= \mathbb{E}\left[\int_0^{T\wedge\Theta} J_x(X_s^{x,u}, Y_s^{y,u}) dB_s + \int_0^{T\wedge\Theta} J_y(X_s^{x,u}, Y_s^{y,u}) dW_s\right] \\ &+ \mathbb{E}\left[\int_0^{T\wedge\Theta} \left(\frac{1}{2}J_{xx} + \frac{1}{2}J_{yy} + uJ_x + (1-u)J_y\right)(X_s^{x,u}, Y_s^{y,u}) ds\right]. \end{aligned} \tag{7.6}$$

The second equality in (7.6) is the result when applying the multi-dimensional Ito formula (see in (2.7)). The first term of the second equation in (7.6) is zero, since after choosing Θ it holds that $(X_s^{x,u}, Y_s^{y,u}) \in \mathcal{B}_r(x,y) \quad \forall s \leq T \wedge \Theta$, so the integrands are bounded and as we already know as stochastic integrals are martingales the expected value is therefore zero. This implies the following

$$0 = \mathbb{E}\left[\frac{1}{T}\int_0^{T\wedge\Theta} \left(\frac{1}{2}J_{xx} + \frac{1}{2}J_{yy} + uJ_x + (1-u)J_y\right)(X_s^{x,u}, Y_s^{y,u}) ds\right]. \tag{7.7}$$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

If the integrand is continuous, then the right side of (7.7) converges for $T \downarrow 0$ to

$$\left(\frac{1}{2}J_{xx} + \frac{1}{2}J_{yy} + uJ_x + (1-u)J_y\right)(x,y) = 0.$$

For the value function $V(x,y)$ we get therefore the HJB equation

$$\sup_{u\in[0,1]}\left\{\frac{1}{2}V_{xx} + \frac{1}{2}V_{yy} + uV_x + (1-u)V_y\right\}$$

or

$$\frac{1}{2}V_{xx} + \frac{1}{2}V_{yy} + \max(V_x, V_y) = 0.$$

At this point it is not clear if the integrand is continuous for arbitrary strategies u and if $J^u \in C^2$ holds. If we assume that $J^u \in C^2$ holds and that (7.7) converges to $\frac{1}{2}J_{xx} + \frac{1}{2}J_{yy} + uJ_x + (1-u)J_y$, we can solve the partial differential equation with the method of successive over-relaxation, which we are going to introduce in Section 7.3.

Now we are going to explain the finite-difference method in more detail, which is an important tool in our analysis.

7.2 Finite-Difference Method

This section gives a brief overview of the finite-difference method (FDM) for solving partial differential equations (PDEs) and is based on [4]. The interested reader is referred to [17] and [25], where a deeper insight into this approach can be obtained. Due to the fact that an analytical or in other words exact solution of PDEs is not always existing, numerical methods, which are a convenient alternative to analytical methods, are used instead in order to provide an approximate solution.

The finite-difference method is a method that relies on approximations in place of the partial derivatives that appear in the PDE. Upon some modifications, a finite-difference scheme is constructed, from which the approximate solution is derived. A simple and at the same time important numerical tool on which the method is fundamentally based is the so-called Taylor's theorem, which is described below.

Theorem 7.2.1 (Taylor's Theorem)**.**
Let $J : [a,b] \to \mathbb{R}$ be a continuous function and $J', J'', \dots, J^{(n-1)}$ be continuous derivatives of J on the interval $[a,b]$. Suppose that the n-th derivative of J exists on the interval (a,b). Then it holds for $a < x_0 < x_0 + h < b$ that

$$J(x_0+h) = J(x_0) + hJ'(x_0) + h^2\frac{J''(x_0)}{2!} + \cdots + h^{n-1}\frac{J^{(n-1)}(x_0)}{(n-1)!} + \mathcal{O}(h^n), \tag{7.8}$$

where

- $J' = \frac{dJ}{dx}, J'' = \frac{d^2J}{dx^2}, \dots, J^{(n-1)} = \frac{d^{n-1}J}{dx^{n-1}}$.
- *$J'(x_0)$ is the derivative of J with respect to x evaluated at $x = x_0$.*

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

- *$\mathcal{O}(h^n)$ is an unknown error term with $\mathcal{O}(h^n) \leq Ch^n$.*

If we drop the error term $\mathcal{O}(h^n)$ in (7.8), we obtain, by knowing the value of the point x_0 and also the values of its derivatives, an approximation to $J(x_0 + h)$, which yields to

$$J(x_0 + h) \approx J(x_0) + hJ'(x_0) + h^2 \frac{J''(x_0)}{2!} + \cdots + h^{n-1} \frac{J^{(n-1)}(x_0)}{(n-1)!}.$$

In the FDM, approximations are accomplished at discrete values of the independent variables, more precisely, by taking the definition domain of the independent variable and replacing it by a finite grid, also called mesh, of points at which the function with its dynamics given by the PDE is familiar. This is where the Taylor theorem comes into play, as we approximate the partial derivatives in the PDE at each grid point from the nearby values, similar to the approximation in (7.8), but interpreted in a different way. Due to the fact that both x_0 and $x_0 + h$ are grid points in FDM but also the function values at the grid points $J(x_0)$ and $J(x_0 + h)$ are known, we are able to rearrange (7.8) to obtain finite-difference approximations to derivatives which have $O(h^n)$ error terms.

Since we consider a PDE in two independent variables with given initial and/or boundary conditions, we want to derive some finite-difference approximations for partial derivatives occurring in our PDE. To do this, we first consider a finite grid of discrete x values, $x_i, i \in \{1, \ldots, N\}$ and denote by $y_j, j \in \{1, \ldots, M\}$ the finite grid of discrete y values. Furthermore, we assume that $N = M$ and that the grid spacing is constant, meaning that $\Delta x := x_{i+1} - x_i = y_{j+1} - y_j =: \Delta y$, which yields to

$$\begin{aligned} x_{i+1} &= x_i + \Delta x \quad \forall i \in \{1, \ldots, N\} \\ y_{j+1} &= y_j + \Delta x \quad \forall j \in \{1, \ldots, N\}. \end{aligned}$$

We will approximate the partial derivatives of J with respect to x, while keeping y constant. The approximation of the partial derivatives of J with respect to y works analogously if x is kept constant. Accordingly, equation (7.8) changes to

$$\begin{aligned} J(x_0 + \Delta x, y) = J(x_0, y) + \Delta x J_x(x_0, y) + \frac{(\Delta x)^2}{2!} J_{xx}(x_0, y) + \cdots \\ + \frac{(\Delta x)^{n-1}}{(n-1)!} J_{(n-1)}(x_0, y) + \mathcal{O}((\Delta x)^n), \end{aligned} \tag{7.9}$$

where

$$J_x = \frac{\partial J}{\partial x}, J_{xx} = \frac{\partial^2 J}{\partial x^2}, \ldots, J_{(n-1)} = \frac{\partial^{n-1} J}{\partial x^{n-1}}$$

are denoting the partial derivatives of J with respect to x.

As we are only interested in the approximation of partial derivatives up to second order, we will write (7.9) as follows

$$J(x_0 + \Delta x, y) = J(x_0, y) + \Delta x J_x(x_0, y) + \mathcal{O}((\Delta x)^2), \tag{7.10}$$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

which yields to, by rearranging (7.10) and by division through Δx,

$$J_x(x_0, y) = \frac{J(x_0 + \Delta x, y) - J(x_0, y)}{\Delta x} - \mathcal{O}((\Delta x)). \tag{7.11}$$

Dropping the error term $\mathcal{O}((\Delta x))$ in (7.11) yields to the first order forward difference approximation to $J_x(x_0, y)$, which in fact has its name because of the first power of Δx in the error term $\mathcal{O}((\Delta x))$ and the fact that starting in point x_0 and moving forward to the point $x_0 + \Delta x$. It can be written as

$$J_x(x_0, y) \approx \frac{J(x_0 + \Delta x, y) - J(x_0, y)}{\Delta x}. \tag{7.12}$$

The first order backward difference approximation to $J_x(x_0, y)$ can be derived in the same way when starting at the same point x_0 and moving backwards instead to the point $x_0 - \Delta x$

$$J_x(x_0, y) \approx \frac{J(x_0, y) - J(x_0 - \Delta x, y)}{\Delta x}. \tag{7.13}$$

So far we have derived two finite-difference approximations which are first order in x but when increasing the order we can obtain a higher accuracy of the approximation. Therefore, we have to consider more terms in the Taylor series which yields for example to the second order central difference approximation to $J_x(x_0, y)$

$$J_x(x_0, y) \approx \frac{J(x_0 + \Delta x, y) - J(x_0 - \Delta x, y)}{2\Delta x}. \tag{7.14}$$

It is also possible to construct higher order finite-difference approximations to $J_x(x_0, y)$ by only taking more terms in the Taylor series, but for our problem we are considering first and second order approximations to first order partial derivatives. However, what we do need are approximations for second order partial derivatives as most of the considered PDEs will contain them and even higher order partial derivatives. Therefore, we will derive some approximations for second order unmixed partial derivatives in the following.

Truncating (7.9) to $\mathcal{O}((\Delta x)^4)$ gives

$$\begin{aligned} J(x_0 + \Delta x, y) = J(x_0, y) + \Delta x J_x(x_0, y) &+ \frac{(\Delta x)^2}{2!} J_{xx}(x_0, y) \\ &+ \frac{(\Delta x)^3}{3!} J_{xxx}(x_0, y) \\ &+ \mathcal{O}((\Delta x)^4), \end{aligned} \tag{7.15}$$

and by replacing Δx by $-\Delta x$ we obtain

$$\begin{aligned} J(x_0 - \Delta x, y) = J(x_0, y) - \Delta x J_x(x_0, y) &+ \frac{(\Delta x)^2}{2!} J_{xx}(x_0, y) \\ &- \frac{(\Delta x)^3}{3!} J_{xxx}(x_0, y) \\ &+ \mathcal{O}((\Delta x)^4). \end{aligned} \tag{7.16}$$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Addition of equation (7.15) and (7.16) yields to

$$J(x_0 + \Delta x, y) + J(x_0 - \Delta x, y) = 2J(x_0, y) + (\Delta x)^2 J_{xx}(x_0, y) + \mathcal{O}((\Delta x)^4), \tag{7.17}$$

which can be rewritten after rearrangement, division through $(\Delta x)^2$ and by dropping the error term $\mathcal{O}((\Delta x)^2)$ as

$$J_{xx}(x_0, y) \approx \frac{J(x_0 + \Delta x, y) - 2J(x_0, y) + J(x_0 - \Delta x, y)}{(\Delta x)^2}. \tag{7.18}$$

Approximation (7.18) is called second order symmetric difference approximation to $J_{xx}(x_0, y)$.

To sum it up, we will put the possible approximations for all partial derivatives up to the order of 2 in the following table, where $J_{i,j}$ is denoting the value of the function J at the grid points $x_i, i \in \{1, \ldots, N\}$ and $y_j, j \in \{1, \ldots, N\}$.

Partial Derivative	**FD-Approximation**	**Type**	**Order**
J_x	$\frac{J_{i+1,j}-J_{i,j}}{\Delta x}$	forward	first in x
	$\frac{J_{i+1,j}-J_{i-1,j}}{2\Delta x}$	central	second in x
	$\frac{J_{i,j}-J_{i-1,j}}{\Delta x}$	backward	first in x
J_{xx}	$\frac{J_{i+1,j}-2J_{i,j}+J_{i-1,j}}{(\Delta x)^2}$	symmetric	second in x
J_y	$\frac{J_{i,j+1}-J_{i,j}}{\Delta y}$	forward	first in y
	$\frac{J_{i,j+1}-J_{i,j-1}}{2\Delta y}$	central	second in y
	$\frac{J_{i,j}-J_{i,j-1}}{\Delta y}$	backward	first in y
J_{yy}	$\frac{J_{i,j+1}-2J_{i,j}+J_{i,j-1}}{(\Delta y)^2}$	symmetric	second in y

Table 7.1: Finite-Difference Approximations for Partial Derivatives

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Now we are able to create a so-called finite-difference scheme by using the corresponding finite-difference approximations from Table 7.1 instead of the partial derivatives occurring in our PDE in (7.4), where $\sigma_x^2 = \sigma_y^2 = 1$ and $\mu = \mu_x + \mu_y = 1$. With those appropriate FD-approximations, (7.3), the constant grid spacing Δx and the notation $J(x_i, y_j) = J_{i,j}$ for $i, j \in \{2, \dots, N-1\}$, our PDE in (7.4) can be rewritten as follows

$$0 = \frac{1}{2}(J_{xx} + J_{yy}) + u_0(x,y)J_x + (1 - u_0(x,y))J_y \quad \Leftrightarrow$$

$$\begin{aligned} 0 = {} & \frac{1}{2}\left(\frac{J_{i+1,j} - 2J_{i,j} + J_{i-1,j}}{(\Delta x)^2} + \frac{J_{i,j+1} - 2J_{i,j} + J_{i,j-1}}{(\Delta x)^2}\right) \\ & + u_0(x,y)\left(\frac{J_{i+1,j} - J_{i-1,j}}{2\Delta x}\right) + (1 - u_0(x,y))\left(\frac{J_{i,j+1} - J_{i,j-1}}{2\Delta x}\right) \quad \Leftrightarrow \end{aligned}$$

$$\begin{aligned} 0 = {} & \frac{J_{i+1,j} - 2J_{i,j} + J_{i-1,j} + J_{i,j+1} - 2J_{i,j} + J_{i,j-1}}{2(\Delta x)^2} \\ & + (\Delta x)u_0(x,y)\left(\frac{J_{i+1,j} - J_{i-1,j}}{2(\Delta x)^2}\right) + (\Delta x)(1 - u_0(x,y))\left(\frac{J_{i,j+1} - J_{i,j-1}}{2(\Delta x)^2}\right) \quad \Leftrightarrow \end{aligned}$$

$$\begin{aligned} 0 = {} & J_{i+1,j} - 2J_{i,j} + J_{i-1,j} + J_{i,j+1} - 2J_{i,j} + J_{i,j-1} \\ & + (\Delta x)u_0(x,y)(J_{i+1,j} - J_{i-1,j}) + (\Delta x)(1 - u_0(x,y))(J_{i,j+1} - J_{i,j-1}) \quad \Leftrightarrow \end{aligned}$$

$$\begin{aligned} 0 = {} & (1 + (\Delta x)u_0(x,y))J_{i+1,j} + (1 - (\Delta x)u_0(x,y))J_{i-1,j} \\ & + (1 + (\Delta x)(1 - u_0(x,y)))J_{i,j+1} + (1 - (\Delta x)(1 - u_0(x,y)))J_{i,j-1} - 4J_{i,j} \quad \Leftrightarrow \end{aligned}$$

$$\begin{aligned} J_{i,j} = {} & \frac{(1 + (\Delta x)u_0(x,y))J_{i+1,j} + (1 - (\Delta x)u_0(x,y))J_{i-1,j}}{4} \\ & + \frac{(1 + (\Delta x)(1 - u_0(x,y)))J_{i,j+1} + (1 - (\Delta x)(1 - u_0(x,y)))J_{i,j-1}}{4}. \end{aligned}$$

By defining four constants,

$$\begin{aligned} c_1 &:= \frac{1 + (\Delta x)u_0(x,y)}{4}, \\ c_2 &:= \frac{1 - (\Delta x)u_0(x,y)}{4}, \\ c_3 &:= \frac{1 + (\Delta x)(1 - u_0(x,y))}{4}, \\ c_4 &:= \frac{1 - (\Delta x)(1 - u_0(x,y))}{4}, \end{aligned}$$

we can write the above equation as

$$J_{i,j} = c_1 J_{i+1,j} + c_2 J_{i-1,j} + c_3 J_{i,j+1} + c_4 J_{i,j-1}, \tag{7.19}$$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

which can be reduced to a system of linear equations of the well-known form

$$Ax = b, \tag{7.20}$$

where

A is the $(N-2)(N-2) \times (N-2)(N-2)$ coefficient matrix

x is the $(N-2)(N-2)$ column vector of unknowns of $J_{i,j}$ and

b is the $(N-2)(N-2)$ column vector of constant terms,

from which an approximate solution can be obtained in several ways. For example, we can use the direct way and get such a solution through Gaussian elimination, but due to the fact that we are considering very large systems, we do not prefer this way because of the efficiency. However, we can use so-called iterative methods instead, for example Jacobi iteration, Gauss-Seidel iteration or the successive over-relaxation method. The last one is a direct application of the Gauss-Seidel iteration and we will use this method for our problem. A deeper insight in these iterative approaches can be obtained in the following Section 7.3.

7.3 Iterative Approaches

In practice, it is common to encounter more complex problems where A in (7.20) can be quite a large coefficient matrix. In this case, it is recommended to exclude the direct solution approach by using the standard Gaussian elimination in order to achieve a solution of (7.20) since it is computationally inefficient, and to choose more efficient methods with corresponding iterative approaches.

In this section, which is based on [4], we present three methods using iterative approaches with constant grid spacing, $\Delta x = \Delta y$, where an original estimate for x is updated in (7.20) to obtain an even better estimate. By assuming that the iterative process converges to the solution, this procedure is continued until the distance between successive estimates is smaller than a given tolerance.

The interested reader is referred to [3] and [11], where iterative methods for solving linear equations are explained in more detail.

Since the outcome of these iterations is a vector consisting of x-values, the distance between two vectors $x = (x_1, \ldots, x_N)^T, y = (y_1, \ldots, y_N)^T \in \mathbb{R}^N$ is denoted by $d(x, y)$ and also known as the infinity-norm, which is given as follows

$$d(x, y) := \max_i(|x_i - y_i|) = \|x - y\|_\infty. \tag{7.21}$$

In order to be able to display the current iteration step for the methods below, we need a new variable k, which is introduced as a superscript and thus represents the iteration index.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Jacobi Iteration

By using the iteration index k we can write (7.19) as the Jacobi formula, which is also known as the point-Jacobi formula,

$$J_{i,j}^{k+1} = c_1 J_{i+1,j}^{k} + c_2 J_{i-1,j}^{k} + c_3 J_{i,j+1}^{k} + c_4 J_{i,j-1}^{k}. \tag{7.22}$$

Since we are working on a mesh, we can obtain $J_{i,j}^{k+1}$ at the next iteration step $k+1$ from (7.22) for every interior grid point (i,j). Prior to starting the iteration, that is, at index $k=0$, the function values for the interior points of $J_{i,j}$ must be determined. However, on the one side we can use the given boundary conditions of the function of our considered problem to determine them. Therefore we know the values of $J_{i,j}$ at its boundaries and we can use them to interpolate the unknown inner values of $J_{i,j}$. On the other side, we could also simply set these inner grid points to 0. After each iteration step k we are obtaining a vector x containing all values for the interior grid points. We are then computing the distance between our new obtained vector x^{k+1} and the old vector x^k. If

$$\|x^{k+1} - x^k\|_\infty < TOL, \tag{7.23}$$

where TOL is a pre-defined tolerance, the algorithm is finished and the solution vector of (7.20) is x^{k+1}, otherwise the vector x^{k+1} is not close enough to the vector x^k and therefore the iteration algorithm continues.

Gauss-Seidel Iteration

From an efficiency point of view, the so-called Gauss-Seidel iteration is potentially more recommendable because, in contrast to the Jacobi iteration, we do not wait until the next iteration step to use the values $J_{i,j}$ we obtained in the current iteration, but instead use these determined values already in the current iteration step. If we traverse the considered grid from left to right, starting from the top corner, once we are at position (i, j) we have previously updated $J_{i-1,j}$ and $J_{i,j+1}$, thus we can apply them and the formula (7.22) turns into

$$J_{i,j}^{k+1} = c_1 J_{i+1,j}^{k} + c_2 J_{i-1,j}^{k+1} + c_3 J_{i,j+1}^{k+1} + c_4 J_{i,j-1}^{k}, \tag{7.24}$$

which is known as the Gauss-Seidel formula or point Gauss-Seidel. The implementation of this method is performed analogously to the Jacobi iteration.

Successive Over-Relaxation

This method is based on the fact that in an iterative formula, the value of $J_{i,j}$ for each grid point (i,j) at the new iteration $k+1$ depends on the value of $J_{i,j}$ plus an error (or residual) term at the previous iteration k at these grid points,

$$J_{i,j}^{k+1} = J_{i,j}^{k} + R_{i,j}^{k}, \tag{7.25}$$

where $R_{i,j}^{k}$ is called the residual or error term. Thus, the idea behind this methodology is to use the values of $J_{i,j}$ computed in the current iteration step to approximate other values

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

of $J_{i,j}$ within the same iteration step. By introducing a so-called relaxation parameter $\omega \in (0,2)$ accordingly, we can potentially accelerate the convergence of the iteration procedure. In doing so, equation (7.25) is transformed by adding a weighting to the residual with the relaxation parameter ω, which finally leads to the following equation

$$J_{i,j}^{k+1} = J_{i,j}^{k} + \omega R_{i,j}^{k}. \tag{7.26}$$

If ω takes values in $(0,1)$, equation (7.26) corresponds to under-relaxation and if ω takes values in $(1,2)$, equation (7.26) corresponds to over-relaxation. This approach is implemented to modify and improve the point-Gauss-Seidel method, which can be re-written as

$$J_{i,j}^{k+1} = J_{i,j}^{k} + c_1 J_{i+1,j}^{k} + c_2 J_{i-1,j}^{k+1} - J_{i,j}^{k} + c_3 J_{i,j+1}^{k+1} + c_4 J_{i,j-1}^{k},$$

which equals formula (7.25) with

$$R_{i,j}^{k} = c_1 J_{i+1,j}^{k} + c_2 J_{i-1,j}^{k+1} - J_{i,j}^{k} + c_3 J_{i,j+1}^{k+1} + c_4 J_{i,j-1}^{k}.$$

Finally, by further inserting the relaxation-parameter ω, we obtain

$$J_{i,j}^{k+1} = (1-\omega) J_{i,j}^{k} + \omega (c_1 J_{i+1,j}^{k} + c_2 J_{i-1,j}^{k+1} + c_3 J_{i,j+1}^{k+1} + c_4 J_{i,j-1}^{k}), \tag{7.27}$$

which is called the point-successive over-relaxation method. For $\omega = 1$, equation (7.27) reduces to the Gauss-Seidel formula, see (7.24). This method is implemented as the previous iterative methods. For further improvements regarding the efficiency of the SOR method, the interested reader is referred to [4], where the Line SOR method is presented.

As we have now introduced all required methods, we are going to show our numerical results and corresponding plots in the following Section 7.4.

7.4 Results

First, we refer to the article by McKean and Shepp and review the results obtained there. The authors consider a two-dimensional model representing the economy of a state and examine which tax policy is optimal for the government based on the optimization criterion. Here, a distinction is made between maximizing the joint probability of survival and maximizing the expected number of survivors. Mathematically, the model can be formulated as follows: The economy consists of two firms whose dynamics can be written as

$$\begin{aligned} dX_1(t) &= \mu_1(t)dt + \sigma_1 dW_1(t) \quad \text{with } X_1(0) = x_1 \geq 0, \sigma_1^2 = 1 \\ dX_2(t) &= \mu_2(t)dt + \sigma_2 dW_2(t) \quad \text{with } X_2(0) = x_2 \geq 0, \sigma_2^2 = 1, \end{aligned}$$

where $X_i(t), i \in \{1,2\}$ represents the wealth of the respective firm and $W_i(t), i \in \{1,2\}$ are two independent standard Brownian motions, which can be controlled by adding processes of non-negative drifts $\mu_i(t), i \in \{1,2\}$ with $\mu_1 + \mu_2 = 1$ to each process in a non-anticipatory way. If, for example, $\mu_1 \geq \mu_2$, then the first company is receiving more government support than the second company. Here, the aim is to maximize the joint survival probability of

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

both companies. We will distinguish between two arbitrary initial strategies s, which are reflected in the drifts as follows

$$\mu_1(t) = s(t)$$
$$\mu_2(t) = 1 - s(t)$$

and present the obtained results using graphs.

In the first case, the government is allocating the drifts to the companies by using the following strategy at each time point t

$$s(t) = \begin{cases} 1, & \frac{1}{2}x_1(t) \leq x_2(t) \\ 0, & \frac{1}{2}x_1(t) > x_2(t). \end{cases} \tag{7.28}$$

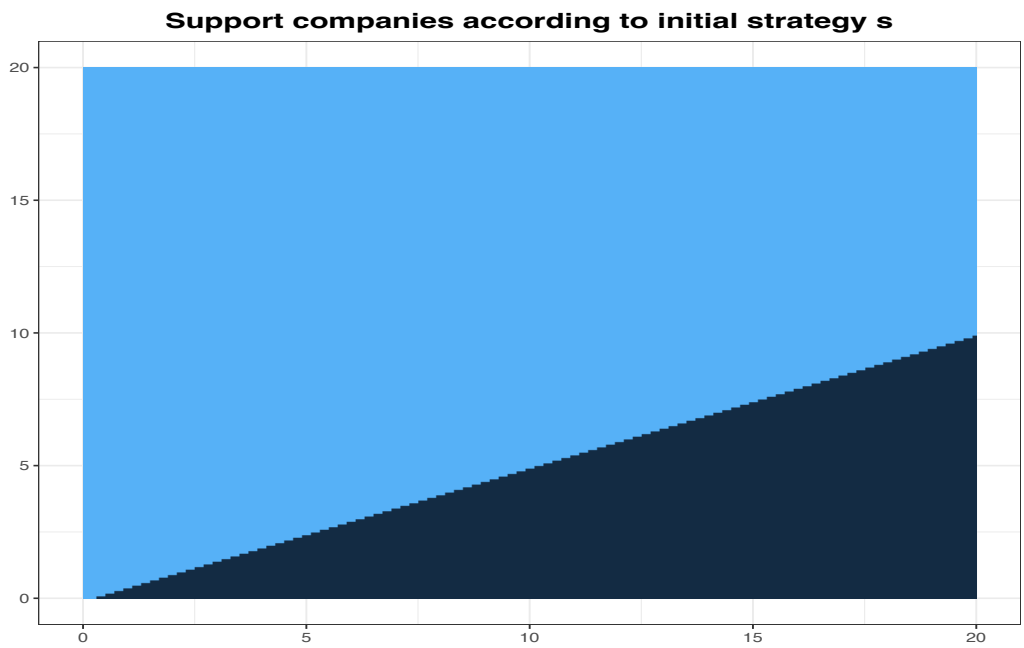


Figure 7.1: Surface plot of initial strategy as defined in (7.28)

In Figure 7.1, the light blue colored area represents the support by the government for the first company while the dark blue colored area represents the support for the second company according to (7.28). By using the policy iteration method we obtain a sequence of monotonically improving policies, $u01 - u20$, according to which the firms are supported by the state, where the colored area for each new strategy is again shown in light blue and dark blue, respectively.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

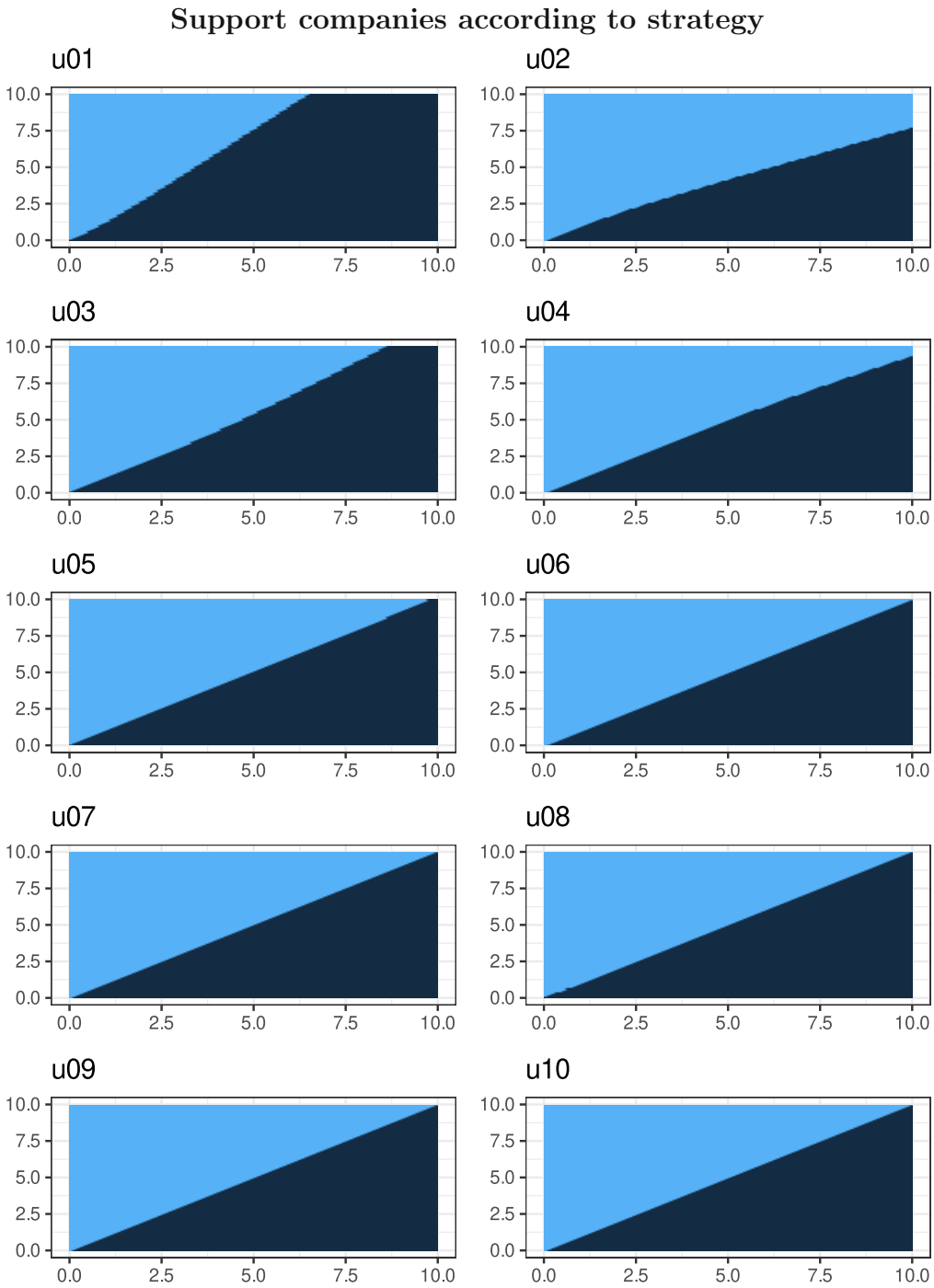


Figure 7.2: $J_1 - J_2$ for $\sigma_1^2 = \sigma_2^2 = 1, \mu = 1$ and $\delta = 0$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

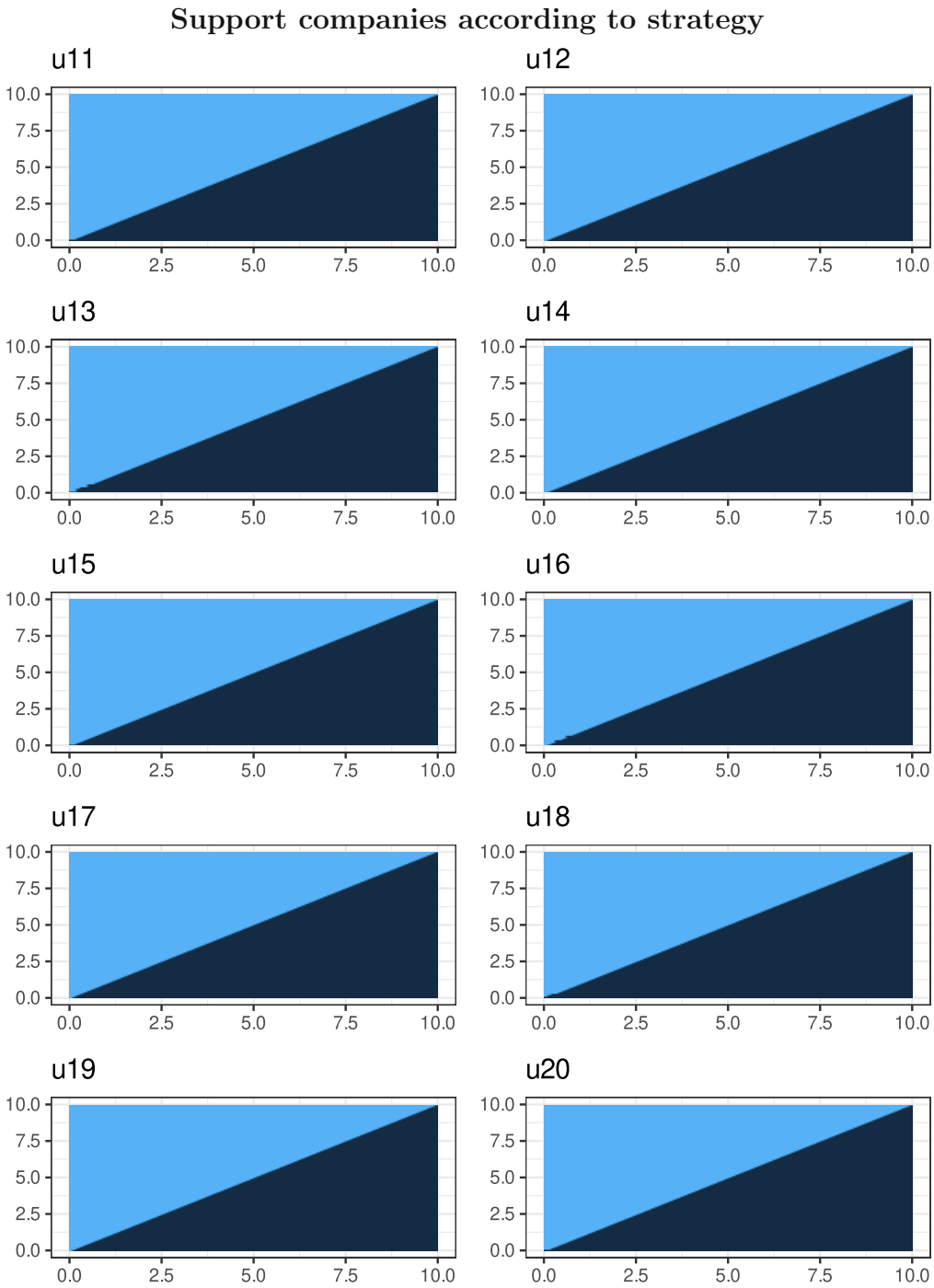


Figure 7.3: $J_1 - J_2$ for $\sigma_1^2 = \sigma_2^2 = 1, \mu = 1$ and $\delta = 0$

Since we have only approximated the boundary condition at infinity, we cannot expect anything reasonable to come out of there. Therefore, the surface plots in Figure 7.2 and in Figure 7.3 do not show the entire range, but only the support by the government in the

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

square $[1, 10] \times [1, 10]$. Each determined policy is guaranteed to be a strict improvement over the previous one, unless there is no difference between the previous policy and the current one, which implies that the latter is already optimal. Moreover, the policy improvement step of the policy iteration algorithm, see Section 6.2.3, ensures us that these policies are better than the original random policy as defined in (7.28). In Figure 7.2, however, we can see that the fifth strategy $u05$ is hardly different from the strategies derived in Figure 7.3, since the error rates are already so small that we get the next policy after just one iteration. These strategies are not only better, but even optimal, because the firms are supported in the form of tax breaks in a way that simultaneously maximizes our objective function, which is the joint survival probability of both firms.

Therefore, we create another plot below which is showing the separation curves, i.e. the curves separating the light blue colored area from the dark blue colored area, for the respective strategies.

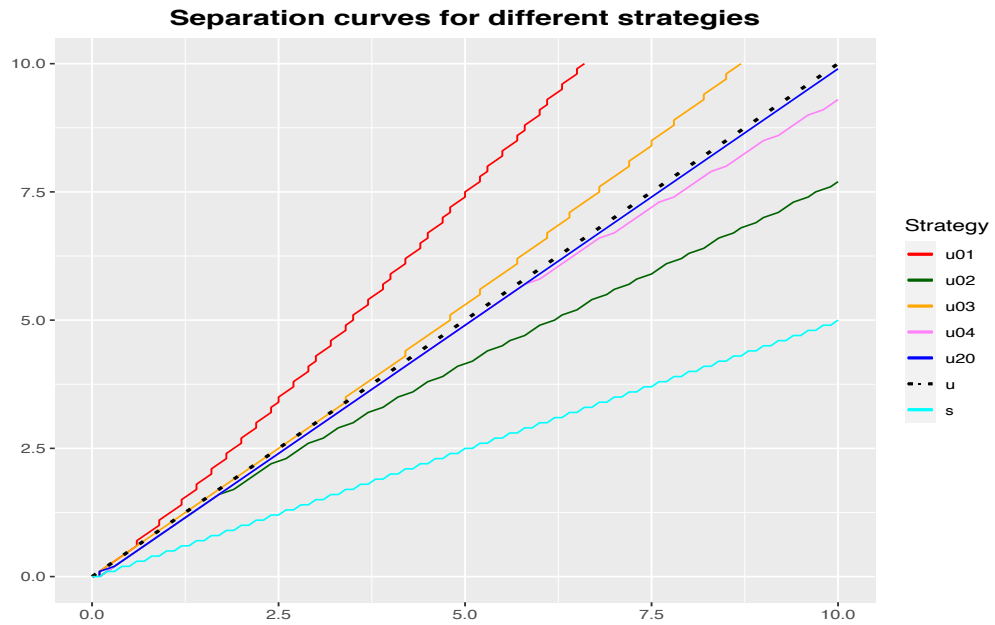


Figure 7.4: Plotting the separation curves of $J_1 - J_2$ with $\alpha = 0$

As we already know from the article by McKean and Shepp, the optimal strategy is the one that changes the support of the state based on the first median when the optimization criterion is to maximize the joint probability of survival. The separation curve of this optimal strategy u has been drawn as the black colored dashed line in Figure 7.4. The turquoise colored line is representing the separation curve for our initial strategy s as defined in (7.28). Furthermore, it is clearly visible that the determined separation curves of the new strategies are getting closer and closer to the separation curve of the optimal strategy, which was also the goal of this task.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

In the second case, the government is allocating the drifts to the companies by using the following strategy at each time point t

$$s(t) = \begin{cases} 1, & \frac{7}{10}x_1(t) \leq x_2(t) \\ 0, & \frac{7}{10}x_1(t) > x_2(t). \end{cases} \tag{7.29}$$

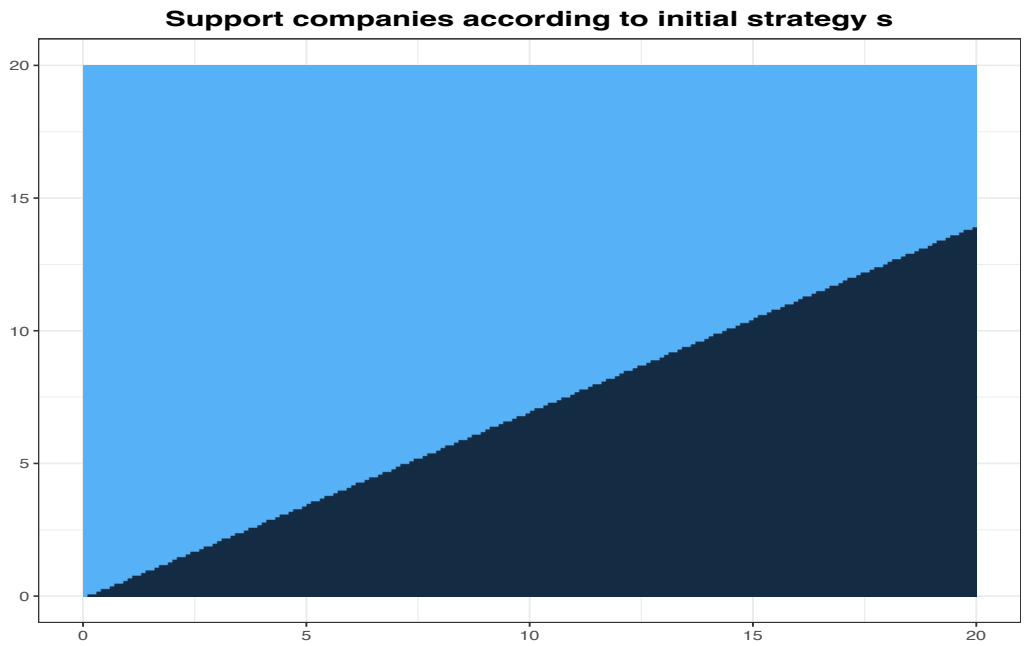


Figure 7.5: Surface plot of initial strategy as defined in (7.29)

In Figure 7.5 the light blue colored area represents the area, where the first company is receiving the drift from the government, whereas in the dark blue colored area the second company is supported according to the policy as defined in (7.29). Compared to the plot derived from the (7.28), the area in which X_2 receives support is now slightly larger. By using the policy iteration method we obtain a sequence of monotonically improving policies, $u01 - u20$, according to which the firms are supported by the state, where the colored area for each new strategy is again shown in light blue and dark blue, respectively.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

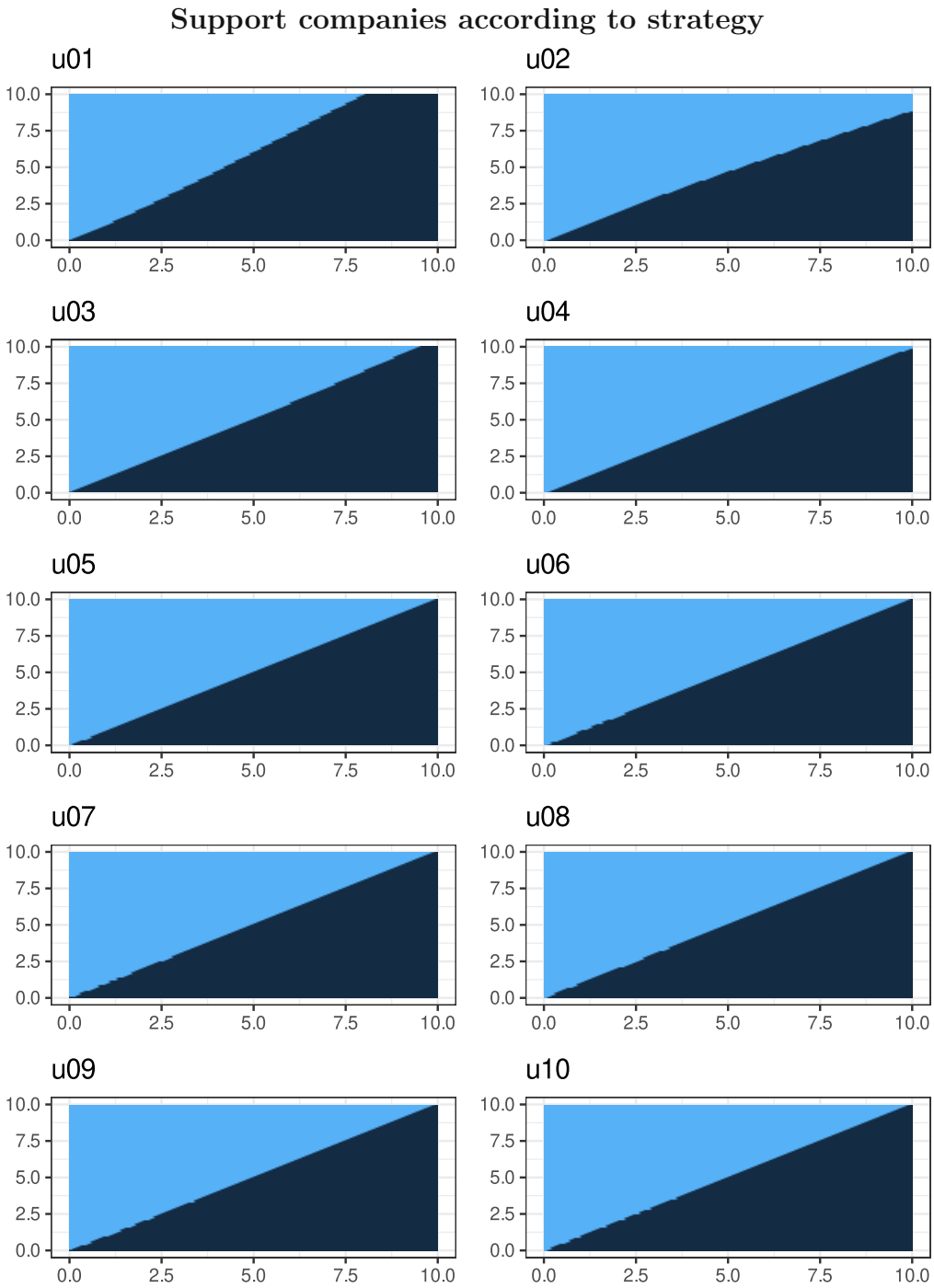


Figure 7.6: $J_1 - J_2$ for $\sigma_1^2 = \sigma_2^2 = 1, \mu = 1$ and $\delta = 0$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

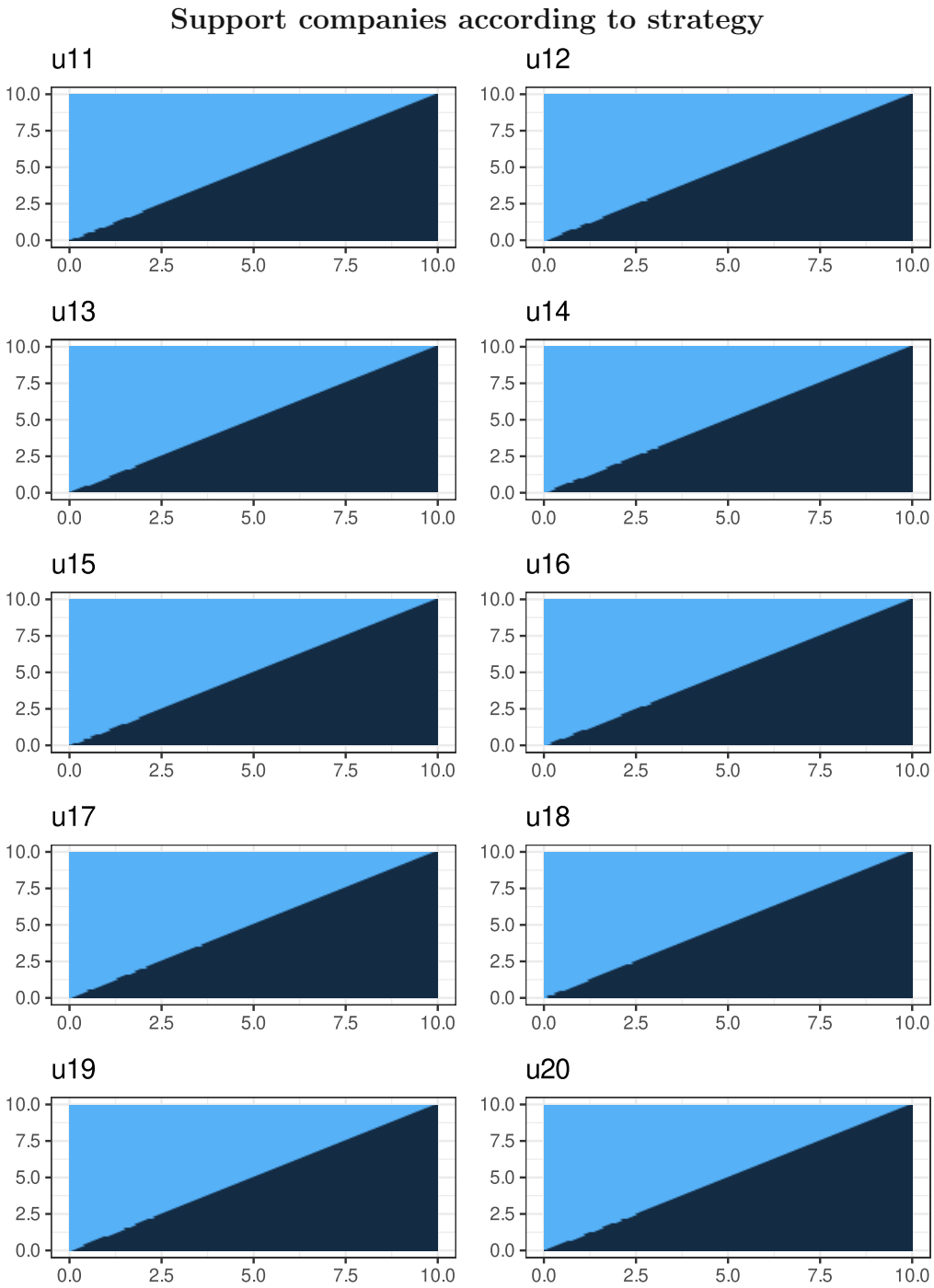


Figure 7.7: $J_1 - J_2$ for $\sigma_1^2 = \sigma_2^2 = 1, \mu = 1$ and $\delta = 0$

Since we have only approximated the boundary condition at infinity, we cannot expect anything reasonable to come out of there. Therefore, the surface plots in Figure 7.6 and in Figure 7.7 do not show the entire range, but only the support by the government in the

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

square $[1, 10] \times [1, 10]$. Each determined policy is guaranteed to be a strict improvement over the previous one, unless there is no difference between the previous policy and the current one, which implies that the latter is already optimal. Moreover, the policy improvement step of the policy iteration algorithm, see Section 6.2.3, ensures us that these policies are better than the original random policy as defined in (7.29). In Figure 7.6, however, we can see that the fourth strategy $u04$ is hardly different from the strategies derived in Figure 7.7, since the error rates are already so small that we get the next policy after just one iteration. These strategies are not only better, but even optimal, because the firms are supported in the form of tax breaks in a way that simultaneously maximizes our objective function, which is the joint survival probability of both firms. Therefore, we create another plot below which is showing the separation curves for the respective strategies.

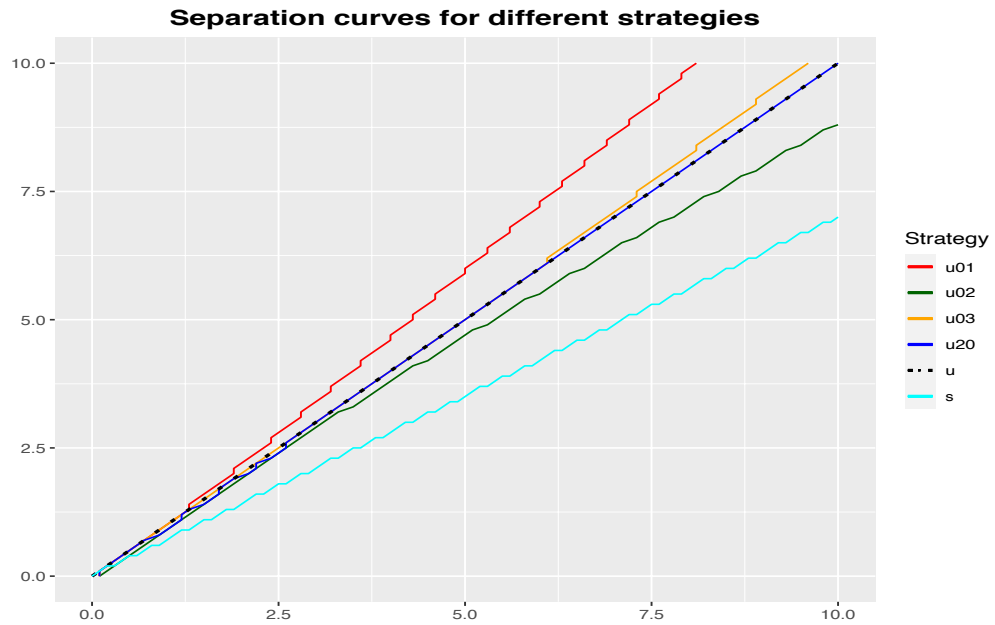


Figure 7.8: Plotting the separation curves of $J_1 - J_2$ with $\alpha = 0$

The separation curve of the optimal strategy u is again drawn as the black colored dashed line whereas the turquoise colored line is representing the separation curve for our initial strategy s as defined in (7.29) in Figure 7.8. Here it can be seen that the separation curve of strategy $u20$ is equal to the separation curve of the optimal strategy u from a certain point.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Now we would like to modify the initial model presented by the authors minimally by adjusting the volatility of the second company using an additional parameter ϵ and thus consider different volatilities for the companies. The dynamics of the customized model can be described as follows:

$$\begin{aligned} dX_1(t) &= \mu_1(t)dt + \sigma_1 dW_1(t) \quad \text{with } X_1(0) = x_1 \geq 0, \sigma_1^2 = 1 \\ dX_2(t) &= \mu_2(t)dt + \sigma_2 dW_2(t) \quad \text{with } X_2(0) = x_2 \geq 0, \sigma_2^2 = \sigma_1^2 + \epsilon = 1 + \epsilon, \end{aligned}$$

where $X_i(t), i \in \{1, 2\}$ represents the wealth of the respective firm and $W_i(t), i \in \{1, 2\}$ are two independent standard Brownian motions, which can be controlled by adding processes of non-negative drifts $\mu_i(t), i \in \{1, 2\}$ with $\mu_1 + \mu_2 = 1$ to each process in a non-anticipatory way. Moreover, it holds that $\sigma_2^2 = \sigma_1^2 + \epsilon \geq \sigma_1^2 = 1, \epsilon > 0$ and if the volatility of a company increases, then its probability of survival decreases according to the fact that the survival probability of a company that follows a Brownian motion can be written as $1 - \exp(-\frac{2\mu x}{\sigma^2})$. Here, the aim is again to maximize the joint survival probability of both companies. Therefore, the company with a lower probability of survival receives the support from the government, which means, that in this case, the firm with a higher volatility should be granted a higher financing margin than assumed in the previous simplified model. Thus, the company with the higher volatility is now offered subsidies until its wealth exceed the wealth of the other company plus an asset-based threshold. We will assign three different values to the parameter ϵ and present the obtained results using graphs. The drifts are allocated to the companies at each time point t by using the optimal strategy for the optimization problem of maximizing the joint survival probability of both companies as initial strategy, which is defined as

$$u(t) = \begin{cases} 1, & x_1(t) \leq x_2(t) \\ 0, & x_1(t) > x_2(t). \end{cases} \tag{7.30}$$

and this strategy is reflected in the drifts as follows

$$\begin{aligned} \mu_1(t) &= u(t) \\ \mu_2(t) &= 1 - u(t). \end{aligned}$$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

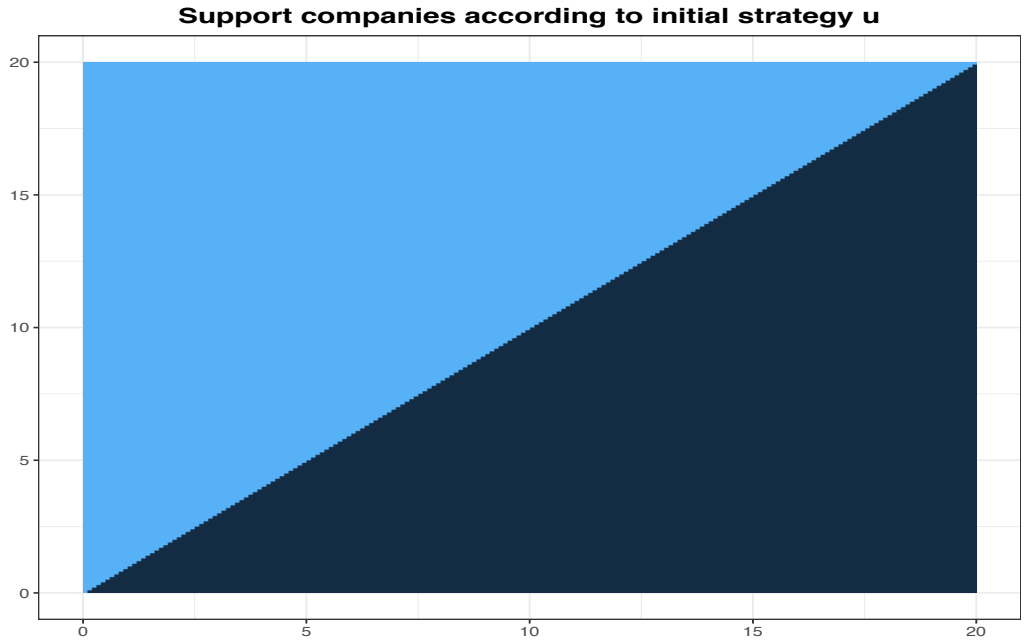


Figure 7.9: Surface plot of initial strategy as defined in (7.30)

In Figure 7.9, company X_1 is receiving the drift from the government in the light blue colored area, whereas in the dark blue colored area company X_2 is supported by the state according to the initial strategy as defined in (7.30). By using the policy iteration method we obtain a sequence of monotonically improving policies, $u01 - u20$, according to which the government is allocating the drifts to the respective firm. The volatility for X_1 remains the same, $\sigma_1^2 = 1$, whereas the volatility of X_2 is now higher, $\sigma_2^2 = 1 + \epsilon$, which is due to the additional parameter $\epsilon = 0.2$.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

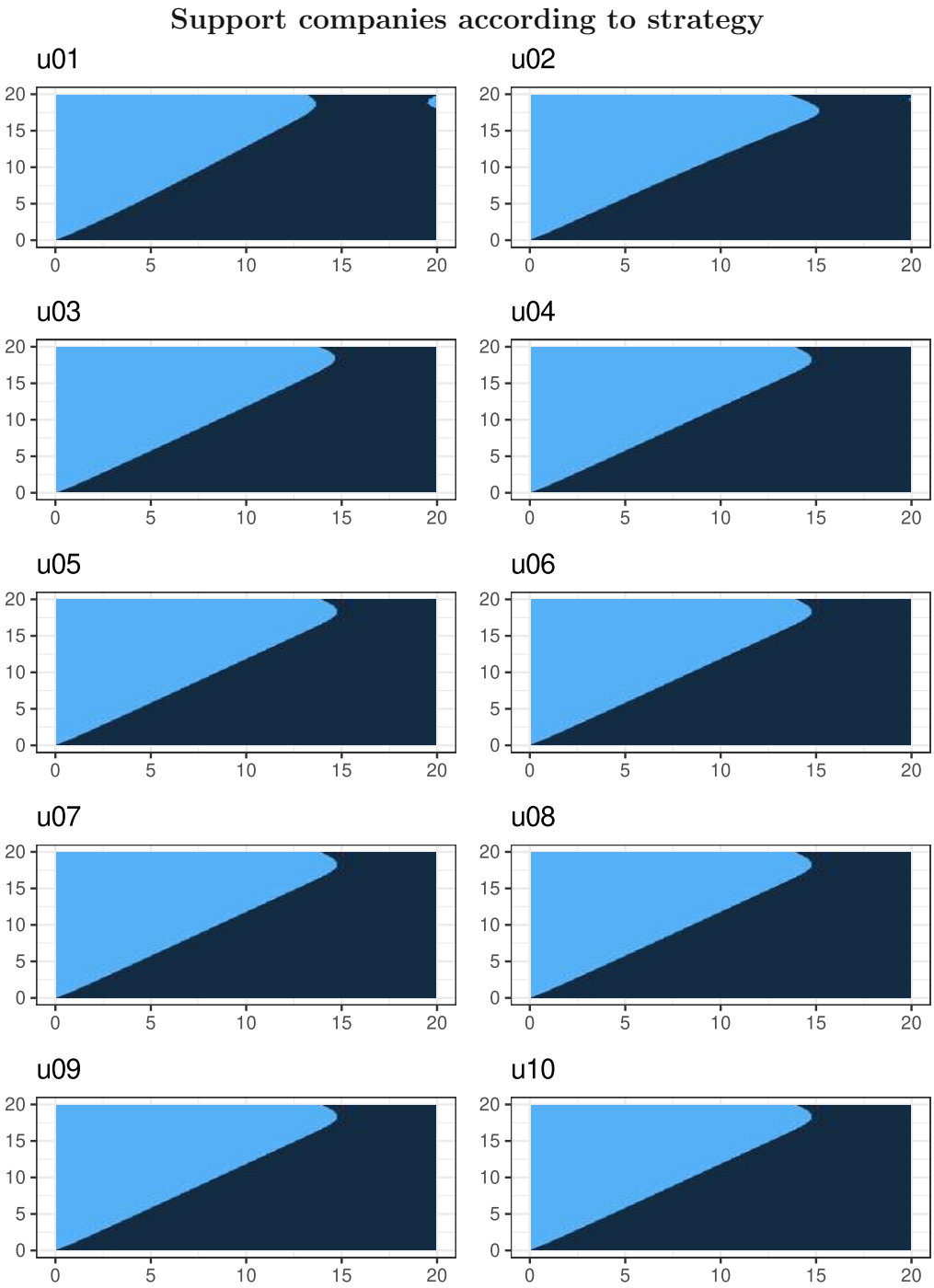


Figure 7.10: $J_1 - J_2$ for $\sigma_1^2 = 1, \sigma_2^2 = 1 + \epsilon = 1.2, \mu = 1$ and $\delta = 0$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

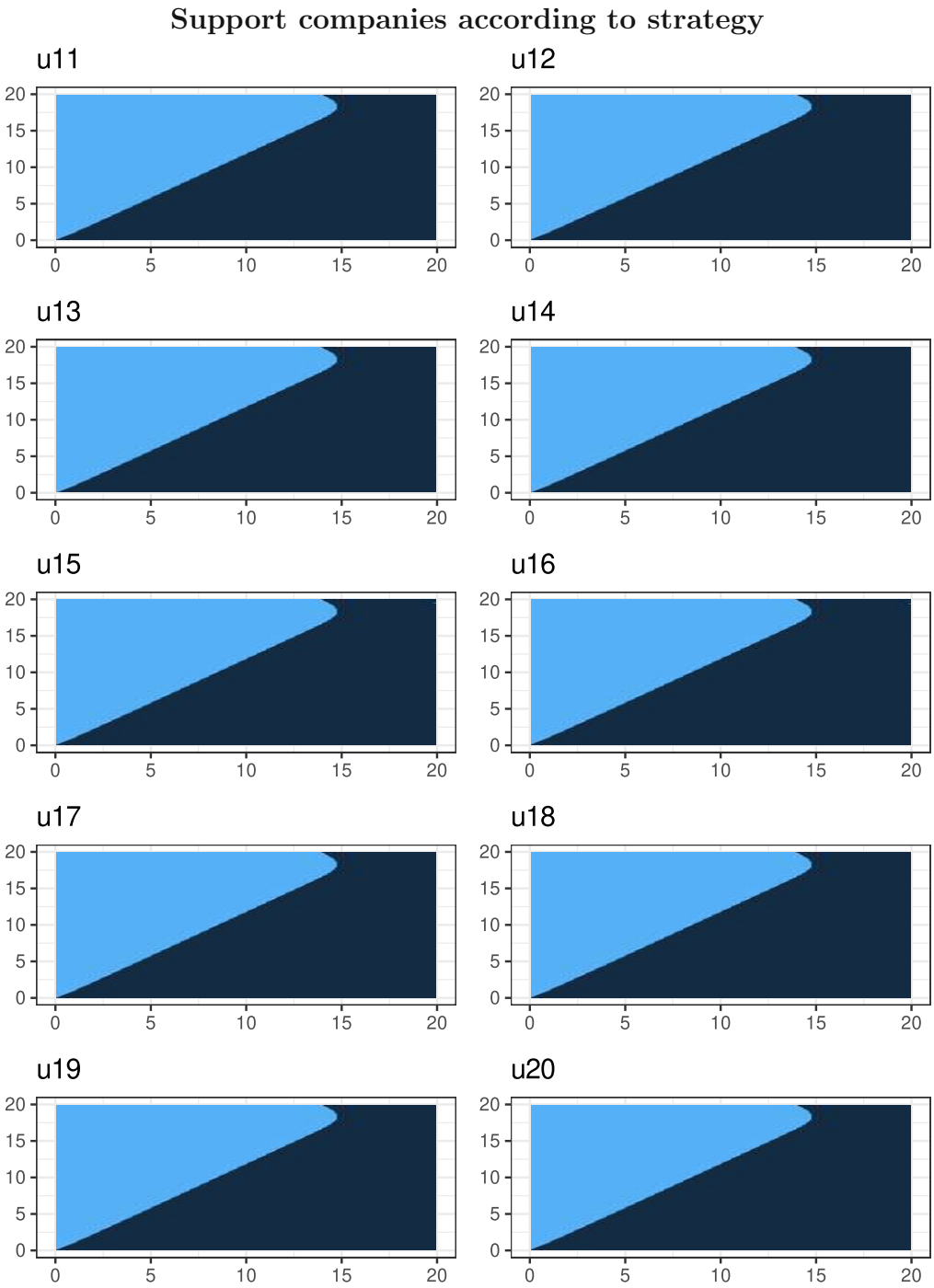


Figure 7.11: $J_1 - J_2$ for $\sigma_1^2 = 1, \sigma_2^2 = 1 + \epsilon = 1.2, \mu = 1$ and $\delta = 0$

In Figure 7.10 and Figure 7.11 we can now see that the dark blue colored area representing support for company X_2 is significantly larger than the area representing support for company X_1. This is based on the fact that the second company has a lower probability of

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

survival than the first and, accordingly, is more often supported by the state in the form of tax relief. Each determined policy is guaranteed to be a strict improvement over the previous one, unless there is no difference between the previous policy and the current one, which implies that the latter is already optimal. Moreover, the policy improvement step of the policy iteration algorithm, see Section 6.2.3, ensures us that these policies are better than the original random policy as defined in (7.30).

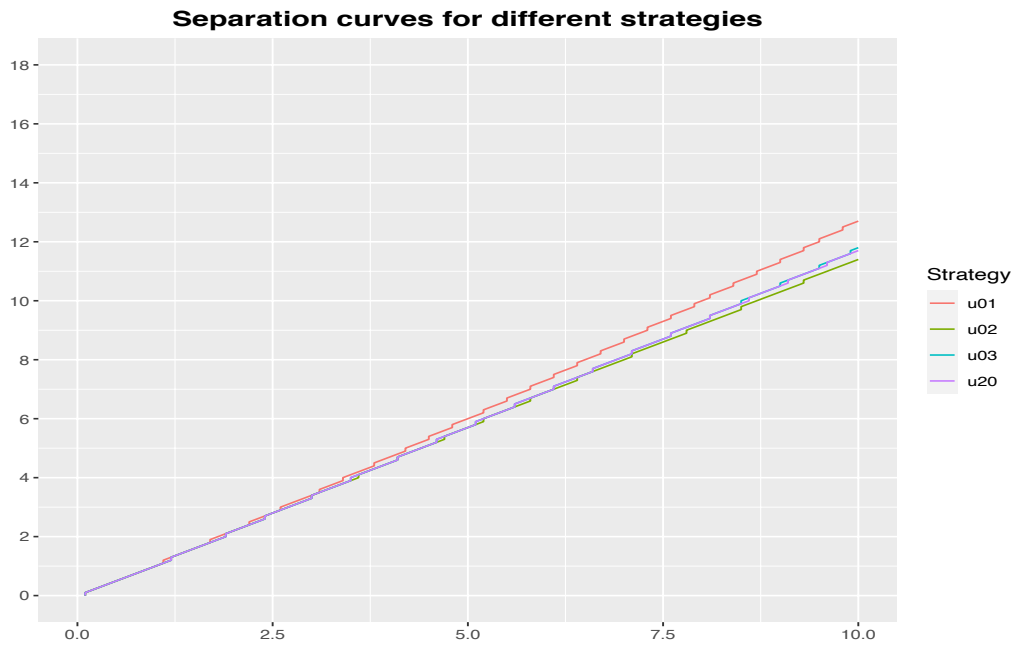


Figure 7.12: Plotting the separation curves of $J_1 - J_2$ with $\alpha = 0$

In Figure 7.12, the separation curves for different strategies are visible for the case $\sigma_1^2 = 1, \sigma_2^2 = 1 + \epsilon = 1.2, \mu = 1$ and $\delta = 0$, where the optimization criterium was to maximize the probability that both companies survive forever.

If we increase the parameter ϵ to $\epsilon = 0.5$, we expect the range in which firm X_2 is supported by the government to become even larger than in the case of $\epsilon = 0.2$. Therefore, we used the policy iteration method to derive a sequence of monotonically improving policies under which firms are supported by the state.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

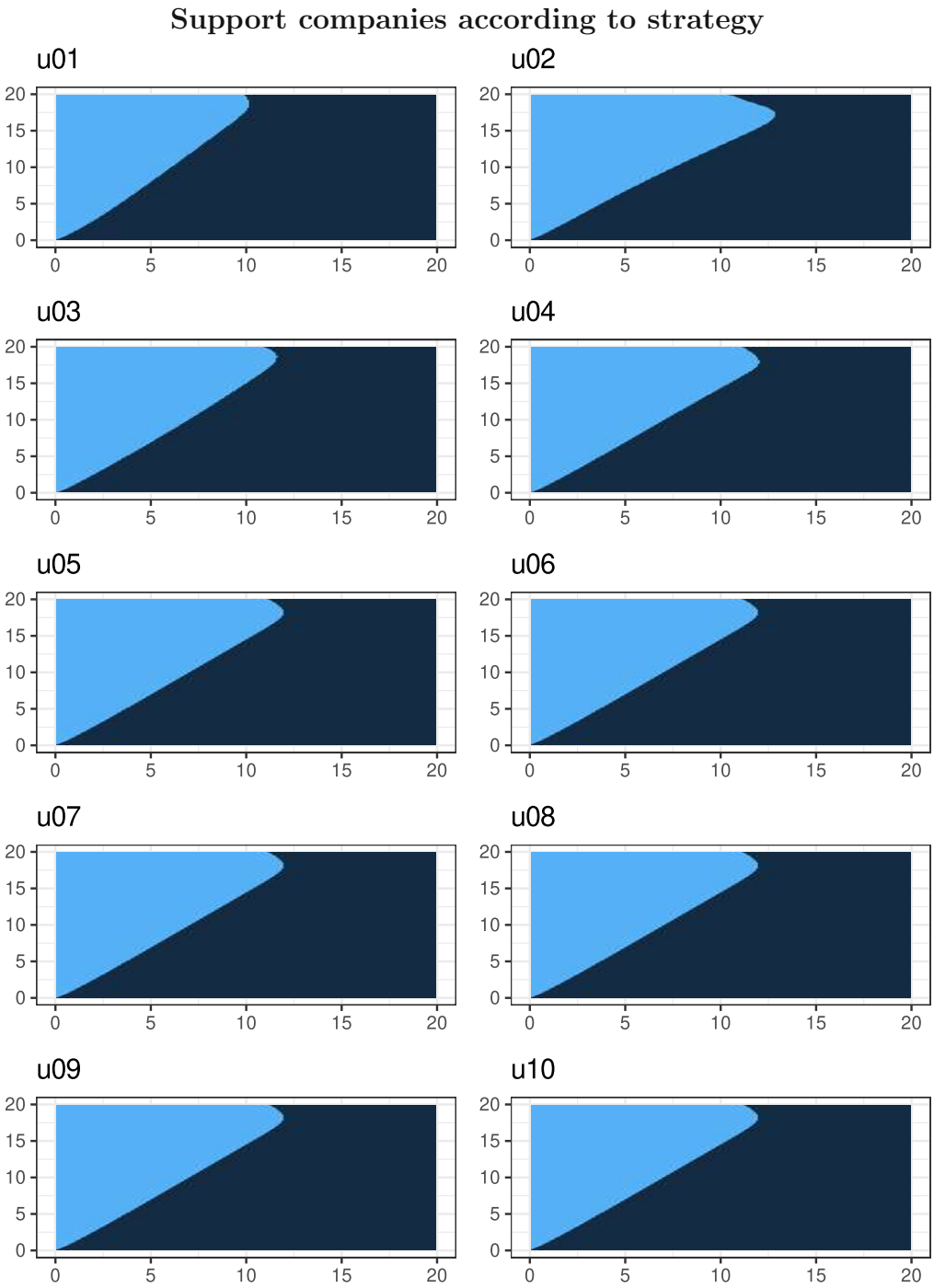


Figure 7.13: $J_1 - J_2$ for $\sigma_1^2 = 1, \sigma_2^2 = 1 + \epsilon = 1.5, \mu = 1$ and $\delta = 0$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

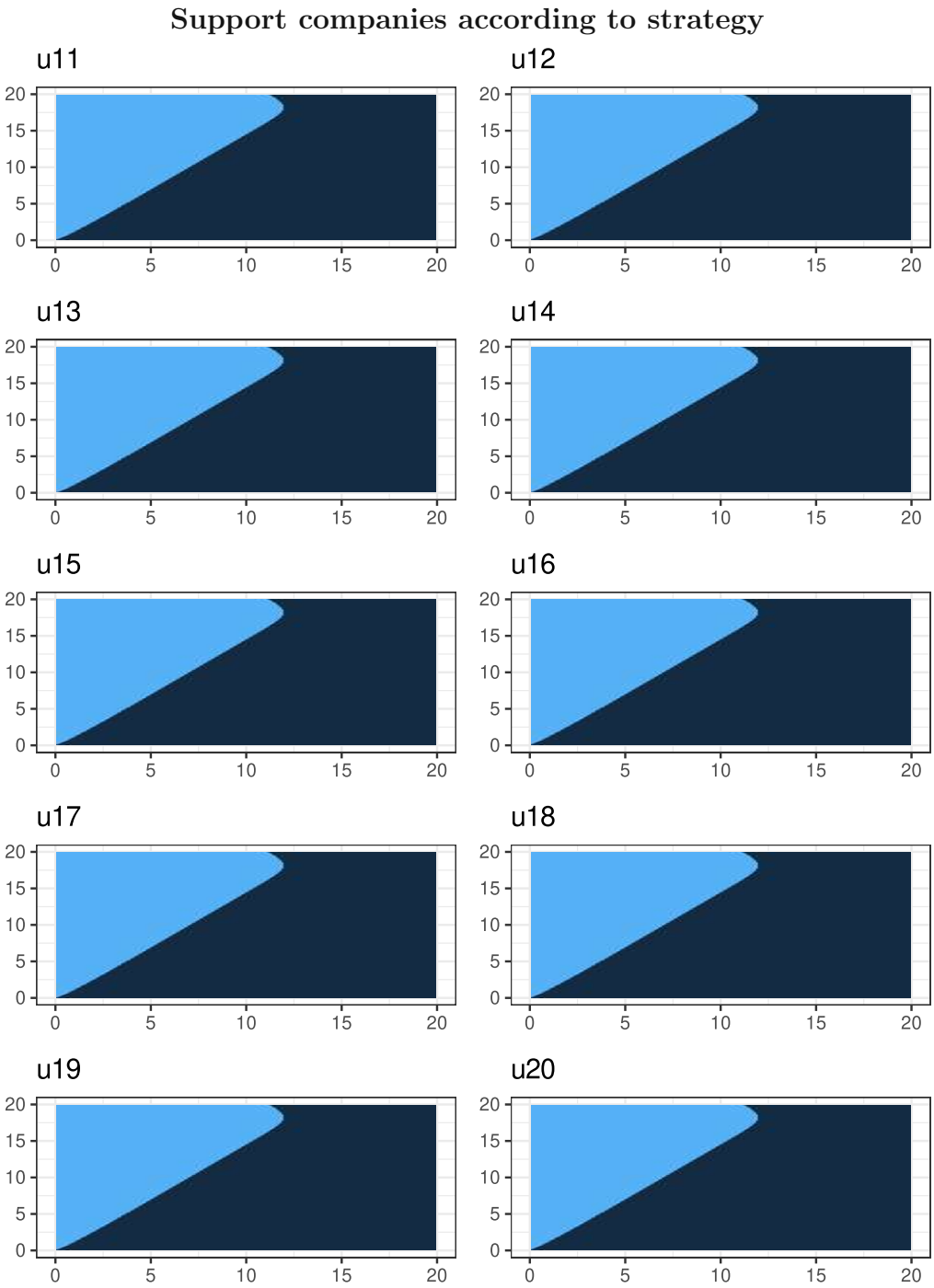


Figure 7.14: $J_1 - J_2$ for $\sigma_1^2 = 1, \sigma_2^2 = 1 + \epsilon = 1.5, \mu = 1$ and $\delta = 0$

In Figure 7.13 and Figure 7.14 we can now see that the region colored in dark blue illustrating government support for company X_2 is noticeably larger than the region representing support for company X_1 in each of the twenty new strategies $u01 - u20$. This is

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

based on the fact that the second company has a lower probability of survival than the first one due to its higher volatility, represented by the parameter ϵ, and is accordingly more often pushed by the government. Each determined policy is guaranteed to be a strict improvement over the previous one, unless there is no difference between the previous policy and the current one, which implies that the latter is already optimal. Moreover, the policy improvement step of the policy iteration algorithm, see Section 6.2.3, ensures us that these policies are better than the original random policy as defined in (7.30).

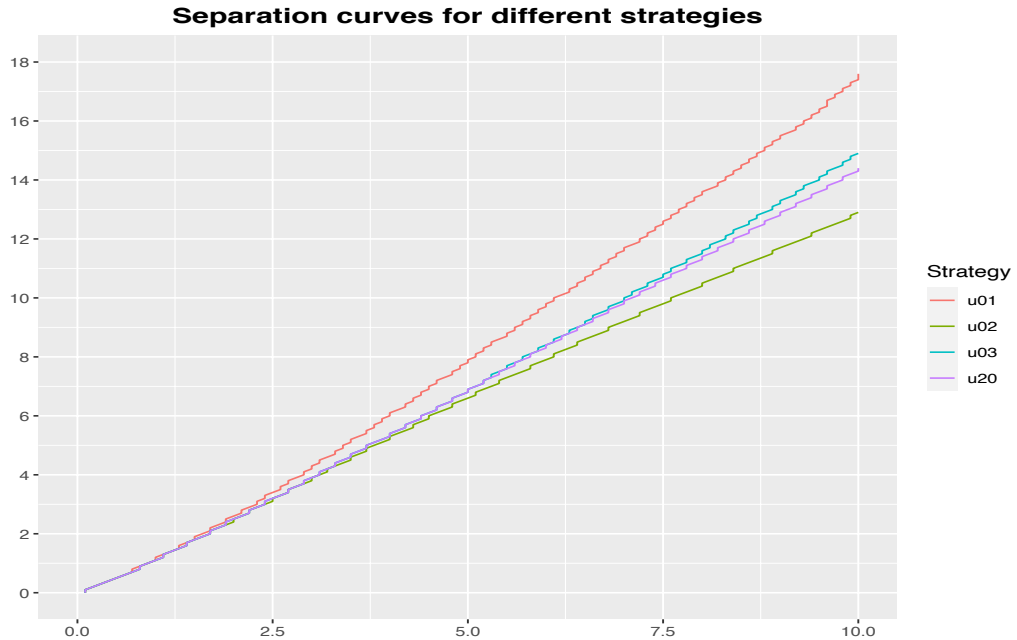


Figure 7.15: Plotting the separation curves of $J_1 - J_2$ with $\alpha = 0$

In Figure 7.15, the separation curves for different strategies are visible for the case $\sigma_1^2 = 1, \sigma_2^2 = 1 + \epsilon = 1.5, \mu = 1$ and $\delta = 0$, where the optimization criterion was to maximize the joint survival probability of both firms.

If we increase the parameter ϵ to $\epsilon = 0.8$, we expect the range in which firm X_2 is supported by the government to become even larger than in the case of $\epsilon = 0.5$. Therefore, we applied the policy iteration method to obtain a sequence of monotonically improving policies under which the government is allocating drifts to the companies.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

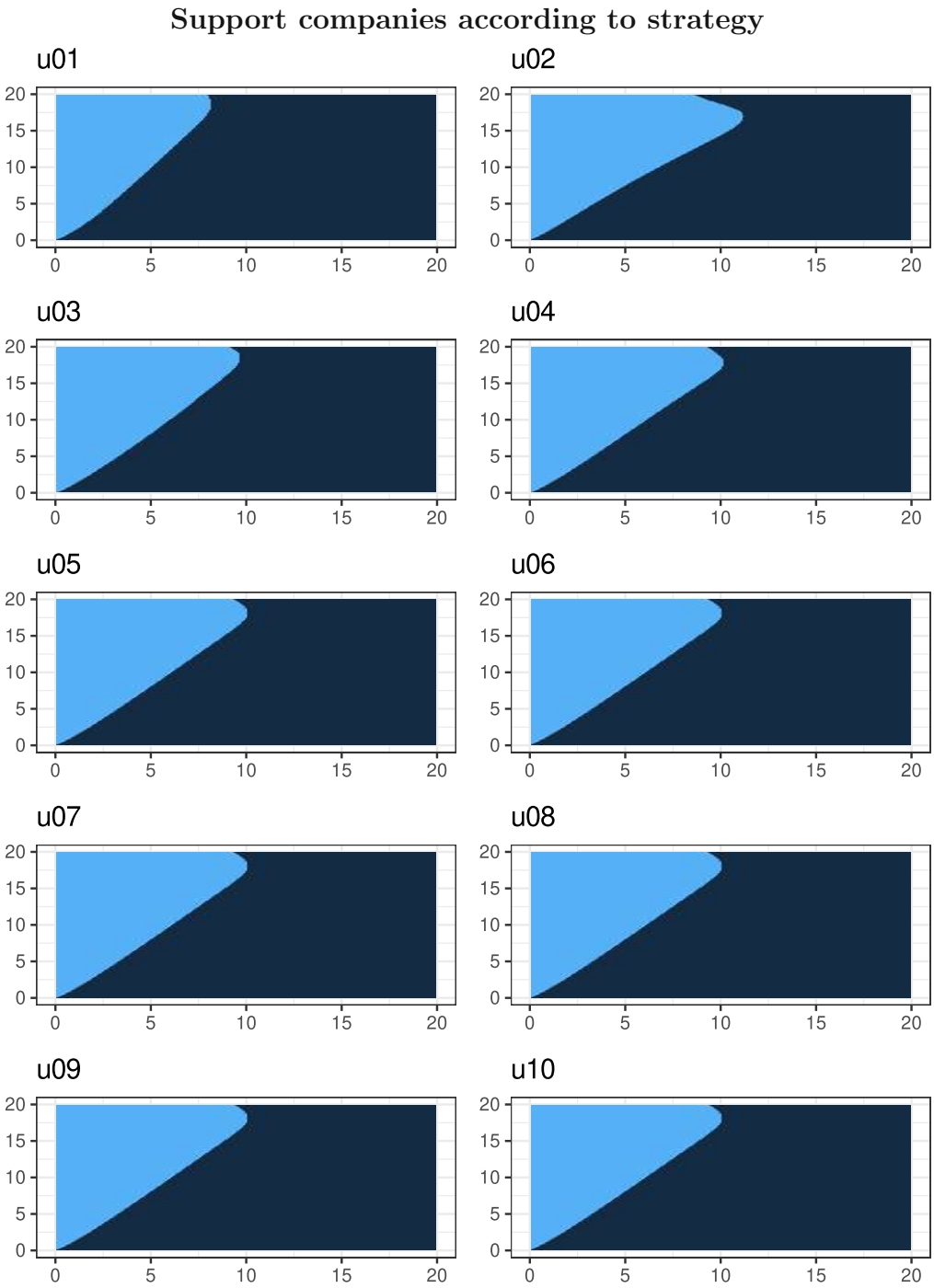


Figure 7.16: $J_1 - J_2$ for $\sigma_1^2 = 1, \sigma_2^2 = 1 + \epsilon = 1.8, \mu = 1$ and $\delta = 0$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

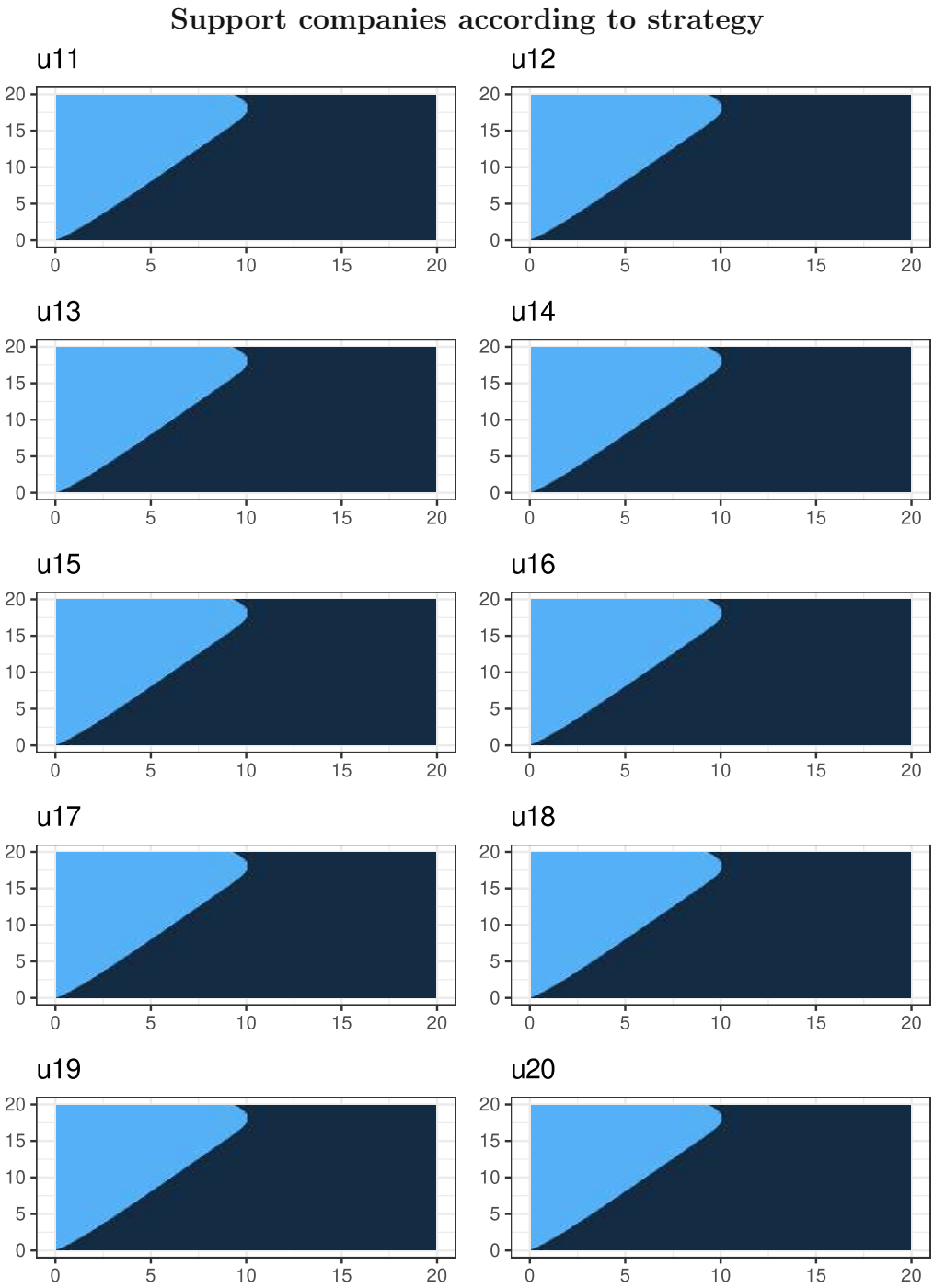


Figure 7.17: $J_1 - J_2$ for $\sigma_1^2 = 1, \sigma_2^2 = 1 + \epsilon = 1.8, \mu = 1$ and $\delta = 0$

In Figure 7.16 and Figure 7.17 we can see that the dark blue colored area depicting the assistance provided by the state to firm X_2 is clearly larger than the area depicting the assistance provided to company X_1 n each of the twenty new strategies $u01 - u20$. This

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

is again based on the fact that the second company has higher volatility, represented by the parameter ϵ, than the other company and, accordingly, is more often backed by the government. Each determined policy is guaranteed to be a strict improvement over the previous one, unless there is no difference between the previous policy and the current one, which implies that the latter is already optimal. Moreover, the policy improvement step of the policy iteration algorithm, see Section 6.2.3, ensures us that these policies are better than the original random policy as defined in (7.30).

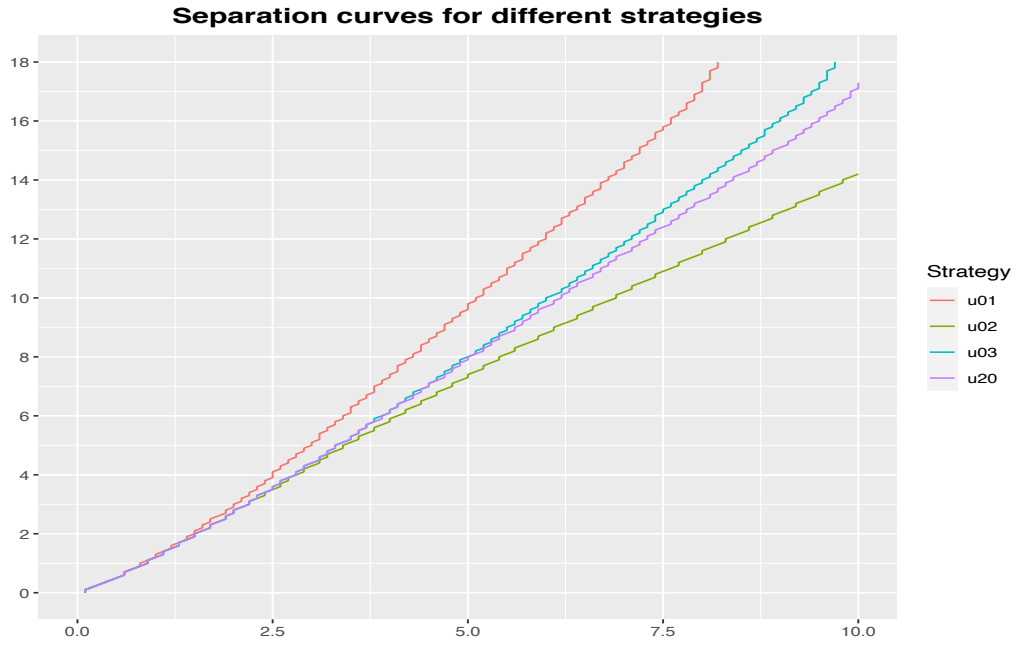


Figure 7.18: Plotting the separation curves of $J_1 - J_2$ with $\alpha = 0$

In Figure 7.18, the separation curves for different strategies are visible for the case $\sigma_1^2 = 1, \sigma_2^2 = 1 + \epsilon = 1.8, \mu = 1$ and $\delta = 0$, where the optimization criterion was to maximize the probability that both companies would survive in the long term.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

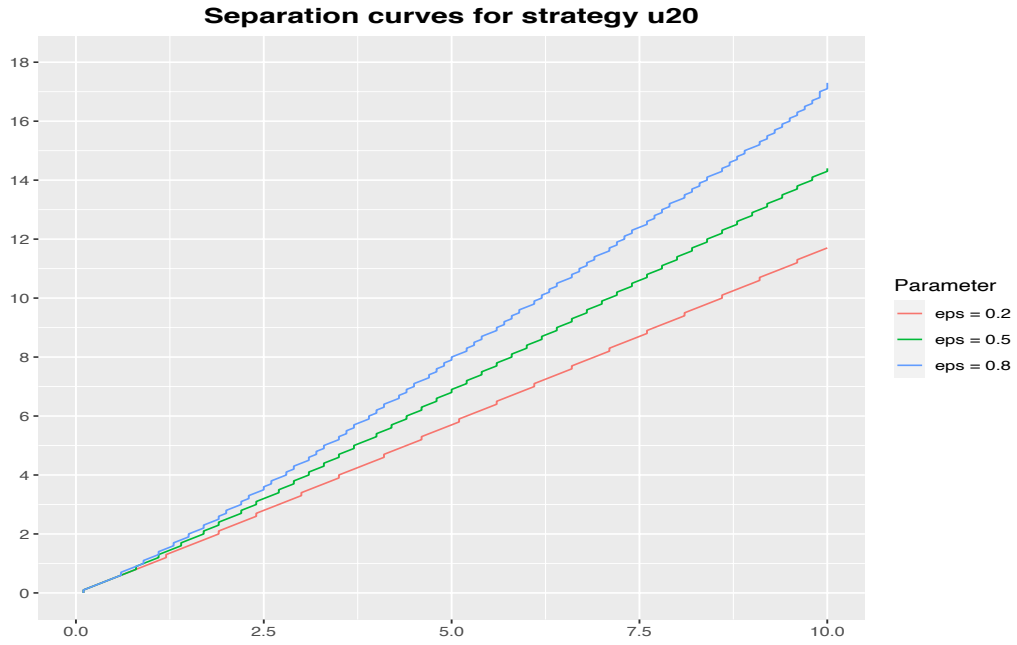


Figure 7.19: Plotting the separation curves of $J_1 - J_2$ with $\alpha = 0$

In Figure 7.19, the separation curves for the strategy $u20$ are plotted again for different volatilities for the case $\sigma_1^2 = 1, \sigma_2^2 = 1 + \epsilon, \mu = 1$ and $\delta = 0$, where $\epsilon =$ eps. Here we see once again that the larger the volatility of X_2, the larger the area of support for X_2, if the optimization criterion is to maximize the joint survival probability of both firms.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

The authors of [18] have also shown that the push-bottom strategy is not the optimal strategy when maximizing the expected numbers of survivors. To prove this, several numerical approximations of the objective function have been made for different values of the occuring parameter α in (7.1).

First, we will consider the case of $\alpha = \frac{1}{2}$, which means that the probability that one company never goes bankrupt and the probability that both companies stay solvent are equally weighted, see (7.1). Note that $\alpha = \frac{1}{2}$ corresponds also to the optimization problem in maximizing the probability that at least one company survives forever. In the following, we will use different initial strategies and show, that each of them is leading to the same result in the sense of the optimal strategy for the considered optimization problem.
The first initial strategy is the push-bottom strategy according to which the drifts are allocated to the weaker company at each time point t, which is defined as

$$u(t) = \begin{cases} 1, & x_1(t) \leq x_2(t) \\ 0, & x_1(t) > x_2(t). \end{cases} \tag{7.31}$$

and this strategy is reflected in the drifts as follows

$$\mu_1(t) = u(t)$$
$$\mu_2(t) = 1 - u(t).$$

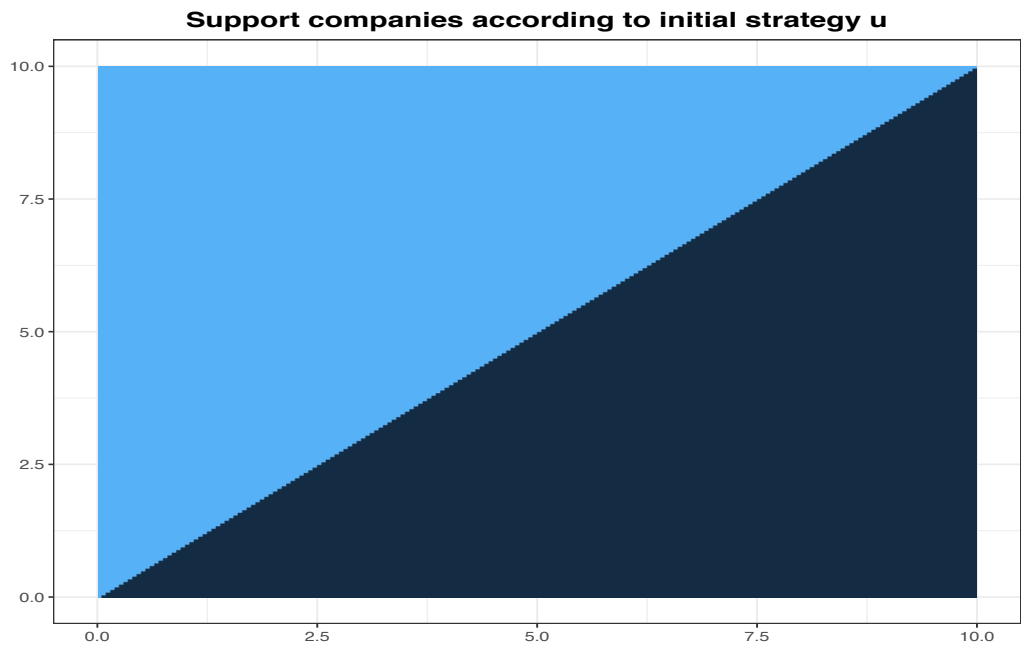


Figure 7.20: Surface plot of initial strategy as defined in (7.31)

In Figure 7.20, the light blue colored area represents the area, where the first company is receiving the drift from the government, whereas in the dark blue colored area the second company is supported according to the policy as defined in (7.31). By using the policy

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

iteration method we obtain a sequence of monotonically improving policies, $u01 - u20$, according to which the firms are supported by the state, where the colored area for each new strategy is again shown in light blue and dark blue, respectively.

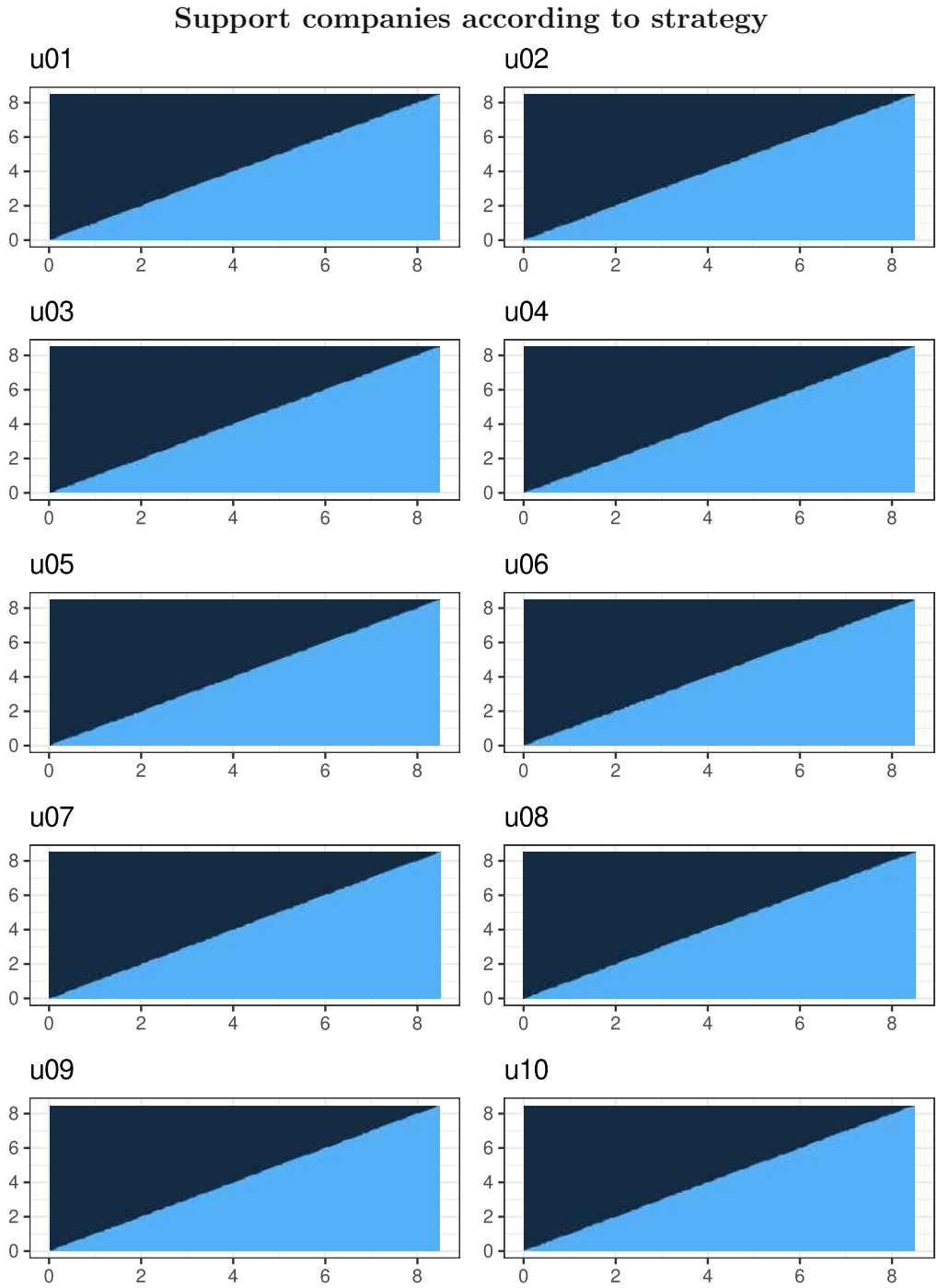


Figure 7.21: $J_1 - J_2$ for $\sigma_1^2 = \sigma_2^2 = 1, \mu = 1$ and $\alpha = \frac{1}{2}$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

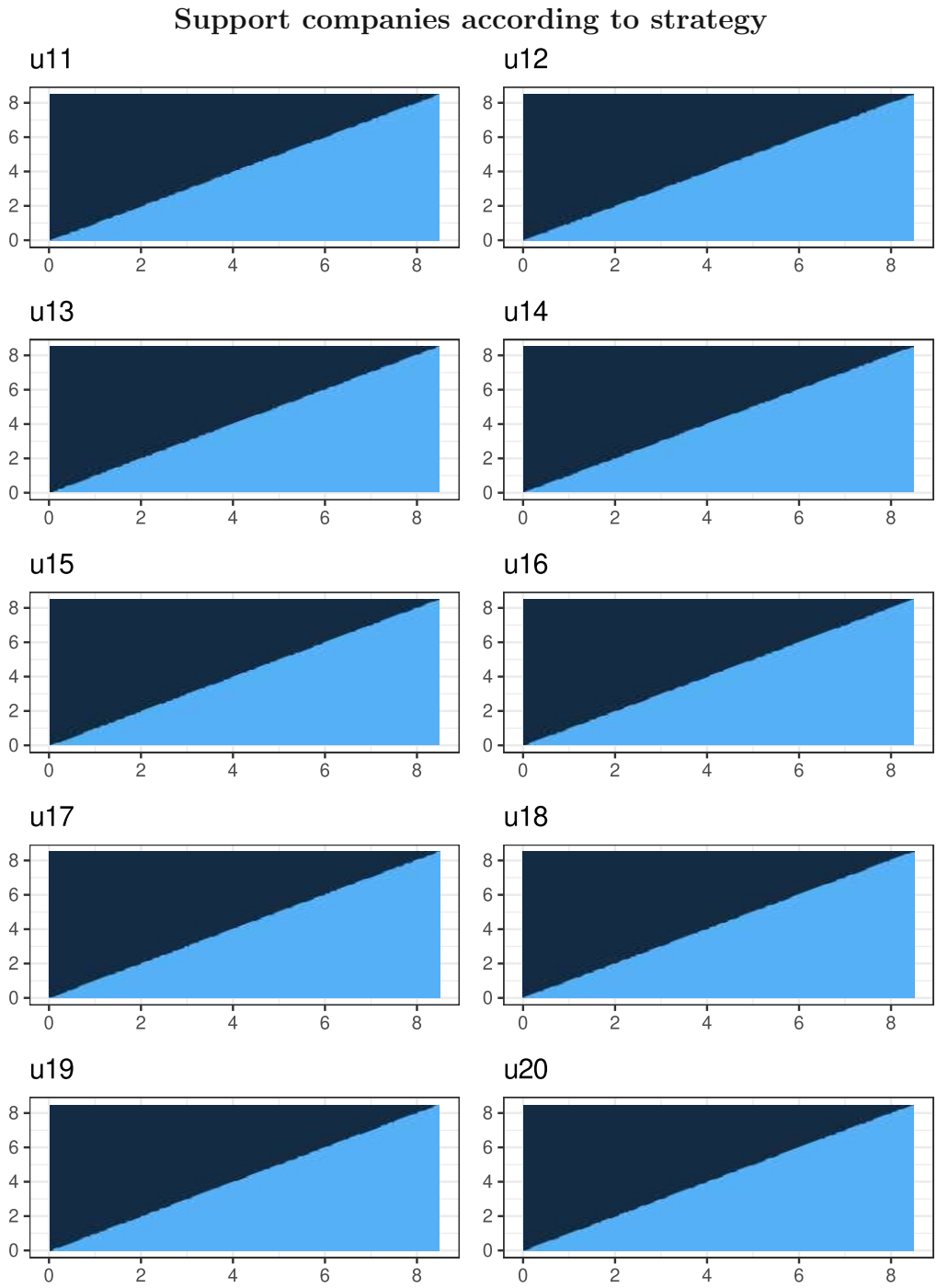


Figure 7.22: $J_1 - J_2$ for $\sigma_1^2 = \sigma_2^2 = 1, \mu = 1$ and $\alpha = \frac{1}{2}$

Each determined policy is guaranteed to be a strict improvement over the previous one, unless there is no difference between the previous policy and the current one, which implies that the latter is already optimal. Here, in Figure 7.21, we can see that after the first

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

determined strategy $u01$, the following strategies, as determined in Figure 7.22 look very similar and therefore we can say that the push-top strategy, which always supports the stronger firm at the considered time, is the optimal strategy for maximizing the expected number of surviving firms with $\alpha = \frac{1}{2}$. Moreover, the policy improvement step of the policy iteration algorithm, see Section 6.2.3, ensures us that these policies are better than the original random policy as defined in (7.31).

For the second initial strategy, let us define four sets, where our initial strategy for the considered optimization problem is supporting the firms respectively.

$$\begin{aligned} A &:= \{(x_1(t), x_2(t)) : (x_1(t) \geq x_2(t)) \wedge (x_1(t) + x_2(t) \leq K)\} \\ B &:= \{(x_1(t), x_2(t)) : (x_1(t) < x_2(t)) \wedge (x_1(t) + x_2(t) \leq K)\} \\ C &:= \{(x_1(t), x_2(t)) : (x_1(t) \leq x_2(t)) \wedge (x_1(t) + x_2(t) > K)\} \\ D &:= \{(x_1(t), x_2(t)) : (x_1(t) > x_2(t)) \wedge (x_1(t) + x_2(t) > K)\} \end{aligned} \tag{7.32}$$

The government is allocating the drifts to the companies by using the following strategy at each time point t

$$u(t) = \begin{cases} 1, & A \vee C \\ 0, & B \vee D. \end{cases} \tag{7.33}$$

and this strategy is in turn reflected in the drifts as follows

$$\begin{aligned} \mu_1(t) &= u(t) \\ \mu_2(t) &= 1 - u(t). \end{aligned}$$

For the sake of better visualization, the strategy defined in (7.33) is illustrated by means of a plot.

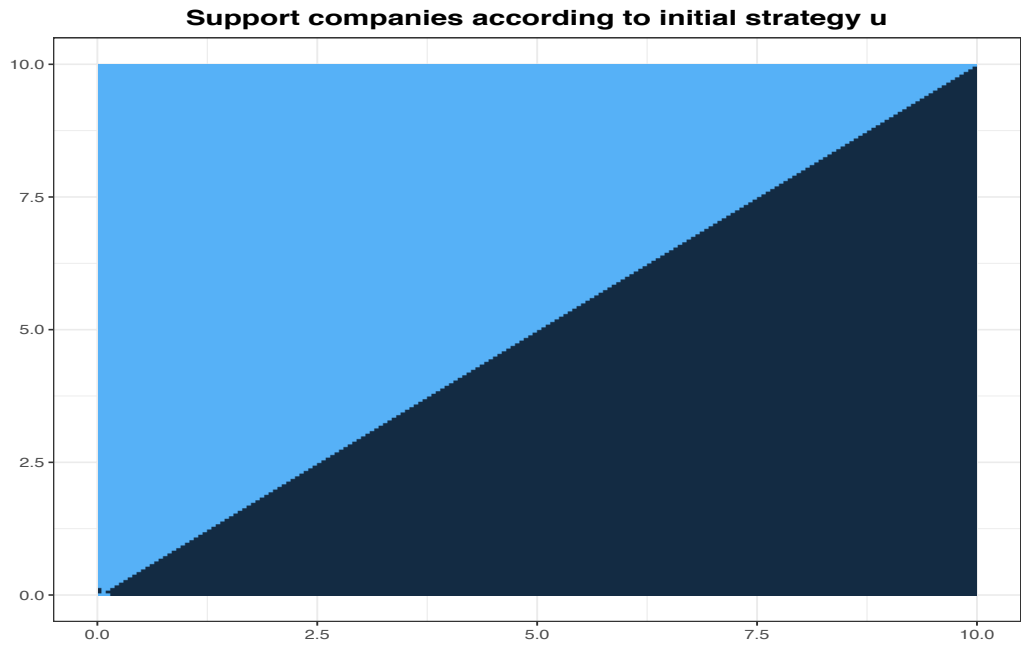


Figure 7.23: Surface plot of initial strategy as defined in (7.33)

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

In Figure 7.23, the light blue colored area represents again the area, where the first company is receiving the drift from the government, whereas in the dark blue colored area the second company is supported according to the policy as defined in (7.33). Here, we can see that near the origin, in $[0, K] \times [0, K]$ with $K \approx 0.15$, the push-top strategy is used. Therefore, we made a second plot to illustrate this area.

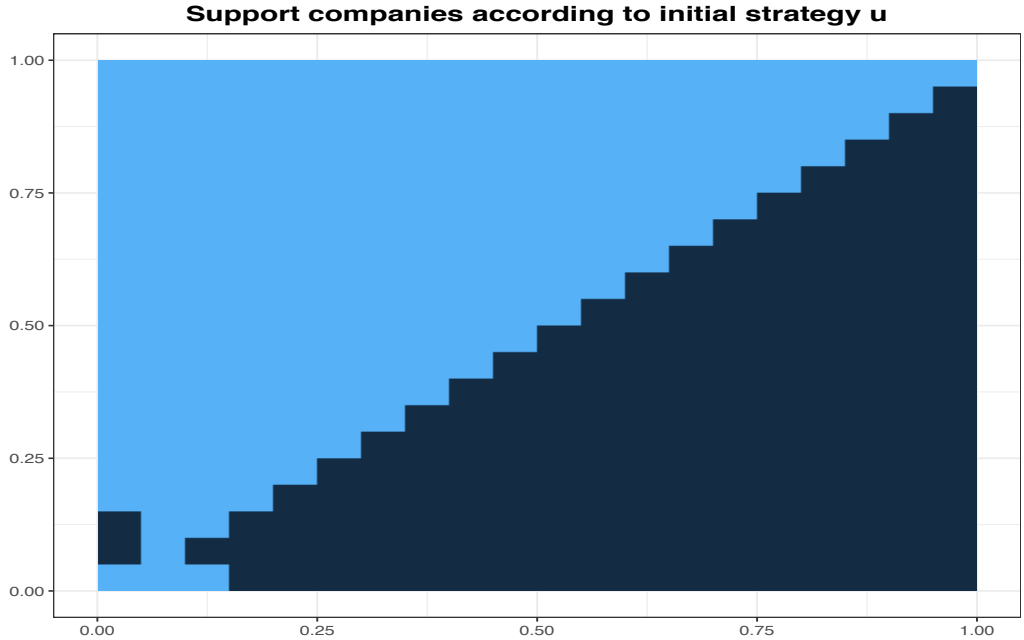


Figure 7.24: Surface plot of initial strategy near origin as defined in (7.33)

By using the policy iteration method we obtain again a sequence of monotonically improving policies, $u01 - u20$, according to which the firms are supported by the state, where the colored area for each new strategy is again shown in light blue and dark blue, respectively.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

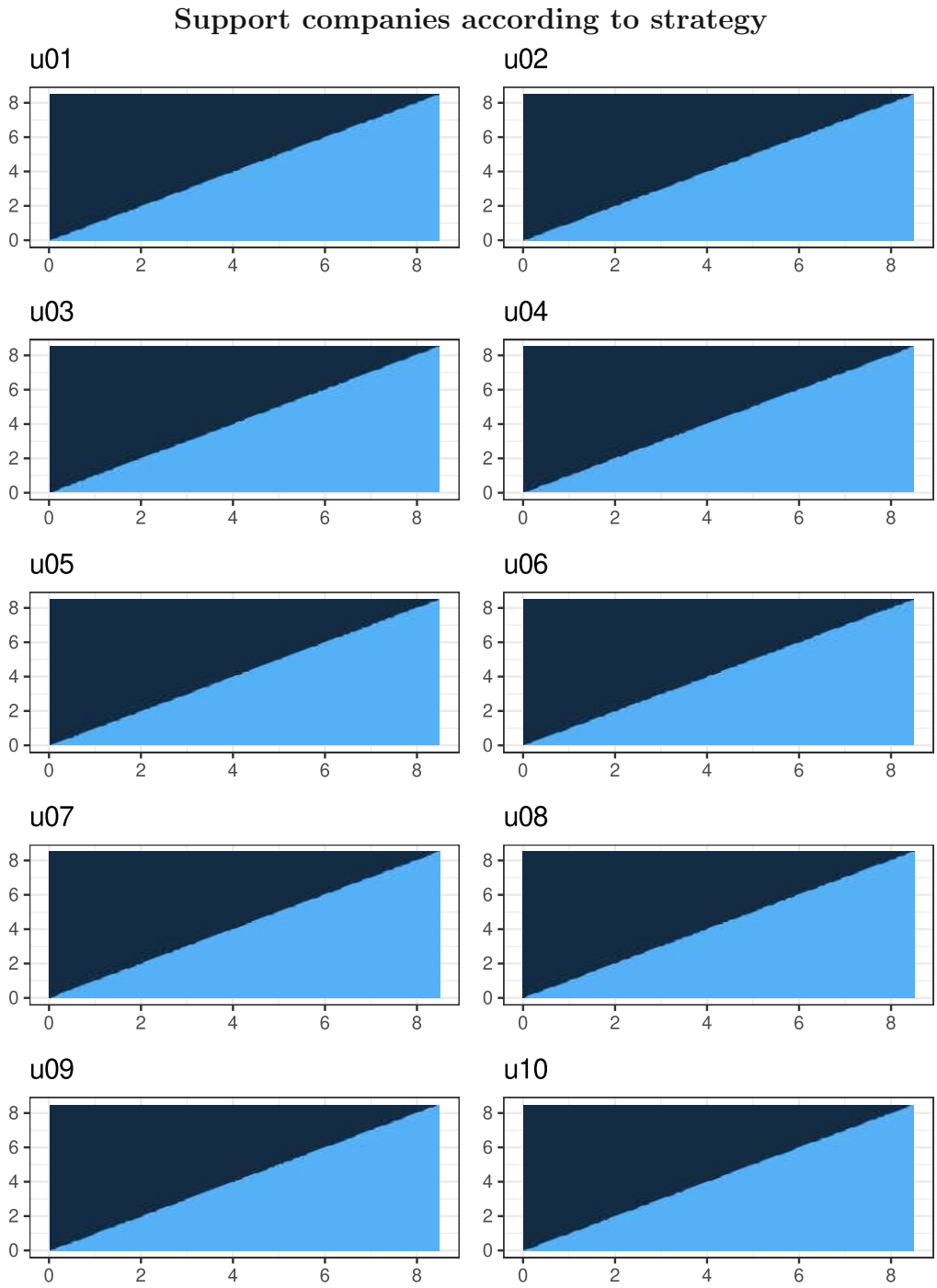


Figure 7.25: $J_1 - J_2$ for $\sigma_1^2 = \sigma_2^2 = 1, \mu = 1$ and $\alpha = \frac{1}{2}$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

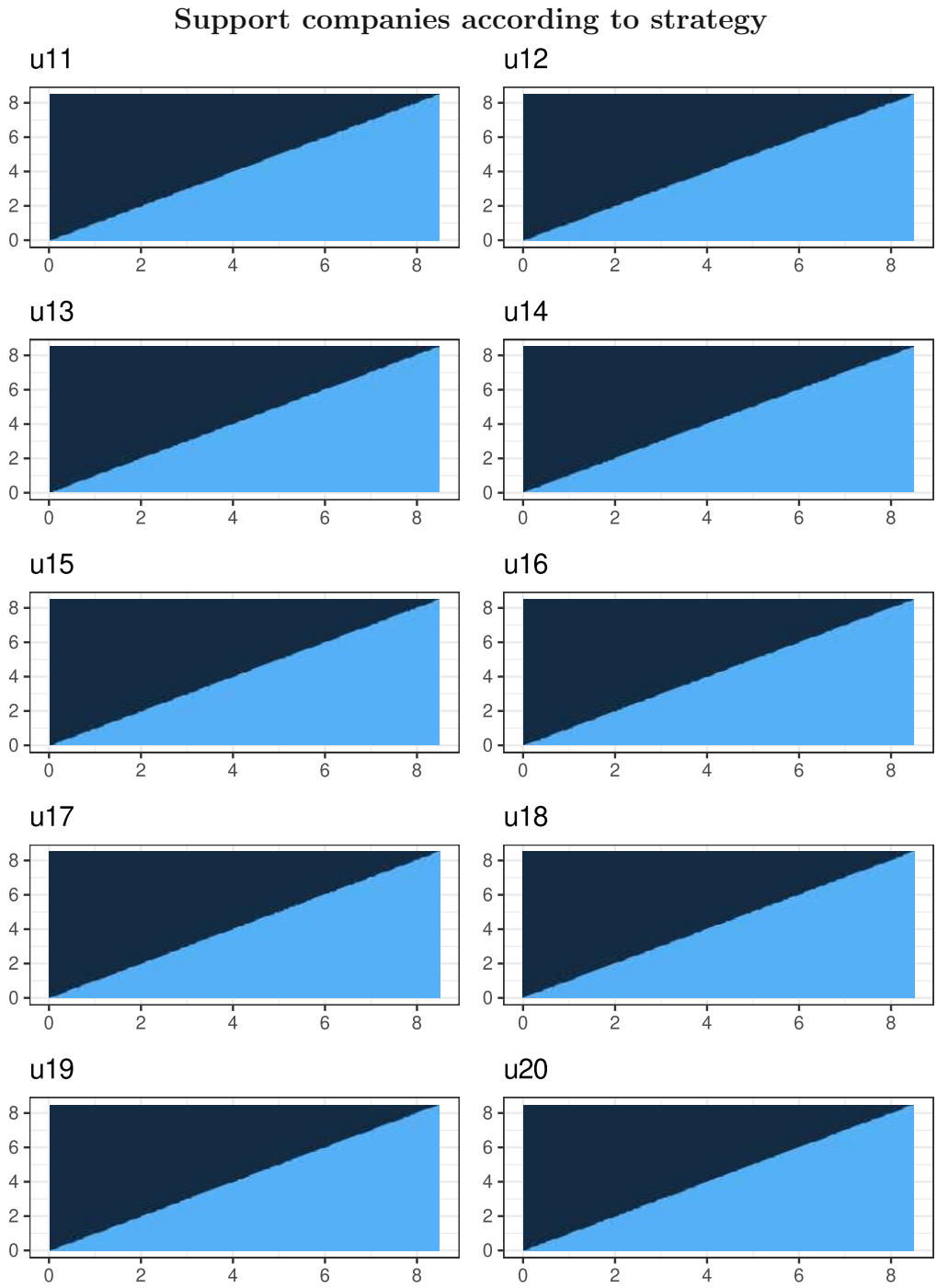


Figure 7.26: $J_1 - J_2$ for $\sigma_1^2 = \sigma_2^2 = 1, \mu = 1$ and $\alpha = \frac{1}{2}$

Each determined policy is guaranteed to be a strict improvement over the previous one, unless there is no difference between the previous policy and the current one, which implies that the latter is already optimal. We can again see that after the first determined strategy

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

$u01$ in Figure 7.25, the following strategies in Figure 7.26 look very similar and therefore we can say that the push-top strategy, which always supports the stronger firm at the considered time, is the optimal strategy for maximizing the expected number of surviving firms with $\alpha = \frac{1}{2}$. Moreover, the policy improvement step of the policy iteration algorithm, see Section 6.2.3, ensures us that these policies are better than the original random policy as defined in (7.31).

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Now, we will consider the case of $\alpha = \frac{1}{3}$. In the following, we will use again different initial strategies and show, that each of them is leading to the same result in the sense of the optimal strategy for the considered optimization problem.

The first initial strategy is the push-top strategy according to which the drifts are allocated to the stronger company at each time point t, which is defined as

$$u(t) = \begin{cases} 0, & x_1(t) \leq x_2(t) \\ 1, & x_1(t) > x_2(t). \end{cases} \tag{7.34}$$

and this strategy is reflected in the drifts as follows

$$\mu_1(t) = u(t)$$
$$\mu_2(t) = 1 - u(t).$$

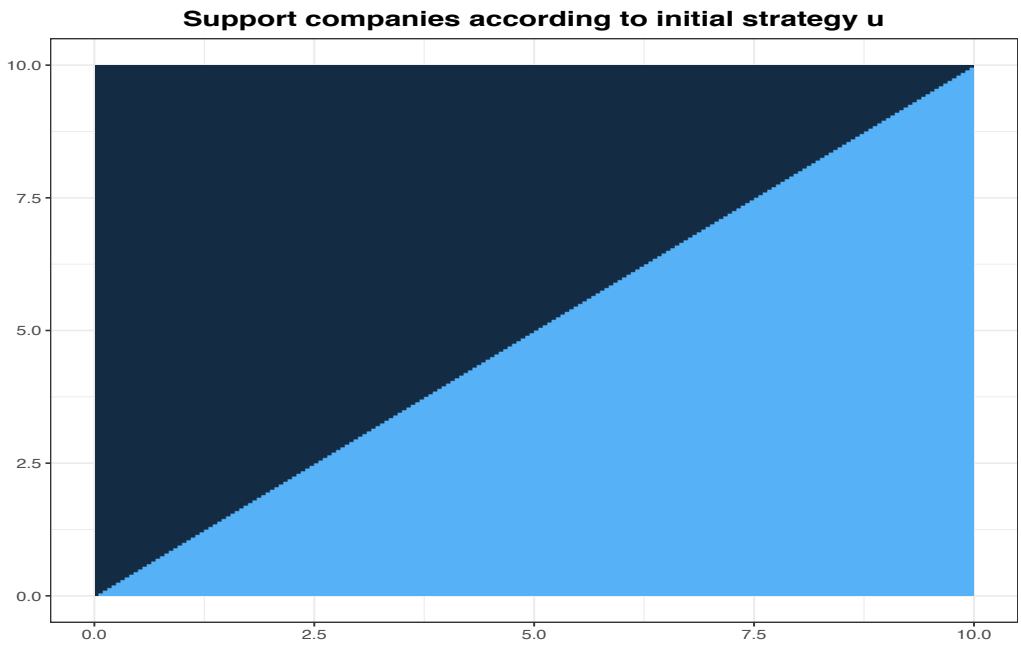


Figure 7.27: Surface plot of initial strategy as defined in (7.34)

In Figure 7.27, the light blue colored area represents the area, where the first company is receiving the drift from the government, whereas in the dark blue colored area the second company is supported according to the policy as defined in (7.34). By using the policy iteration method we obtain a sequence of monotonically improving policies, $u01 - u20$, according to which the firms are supported by the state, where the colored area for each new strategy is again shown in light blue and dark blue, respectively.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

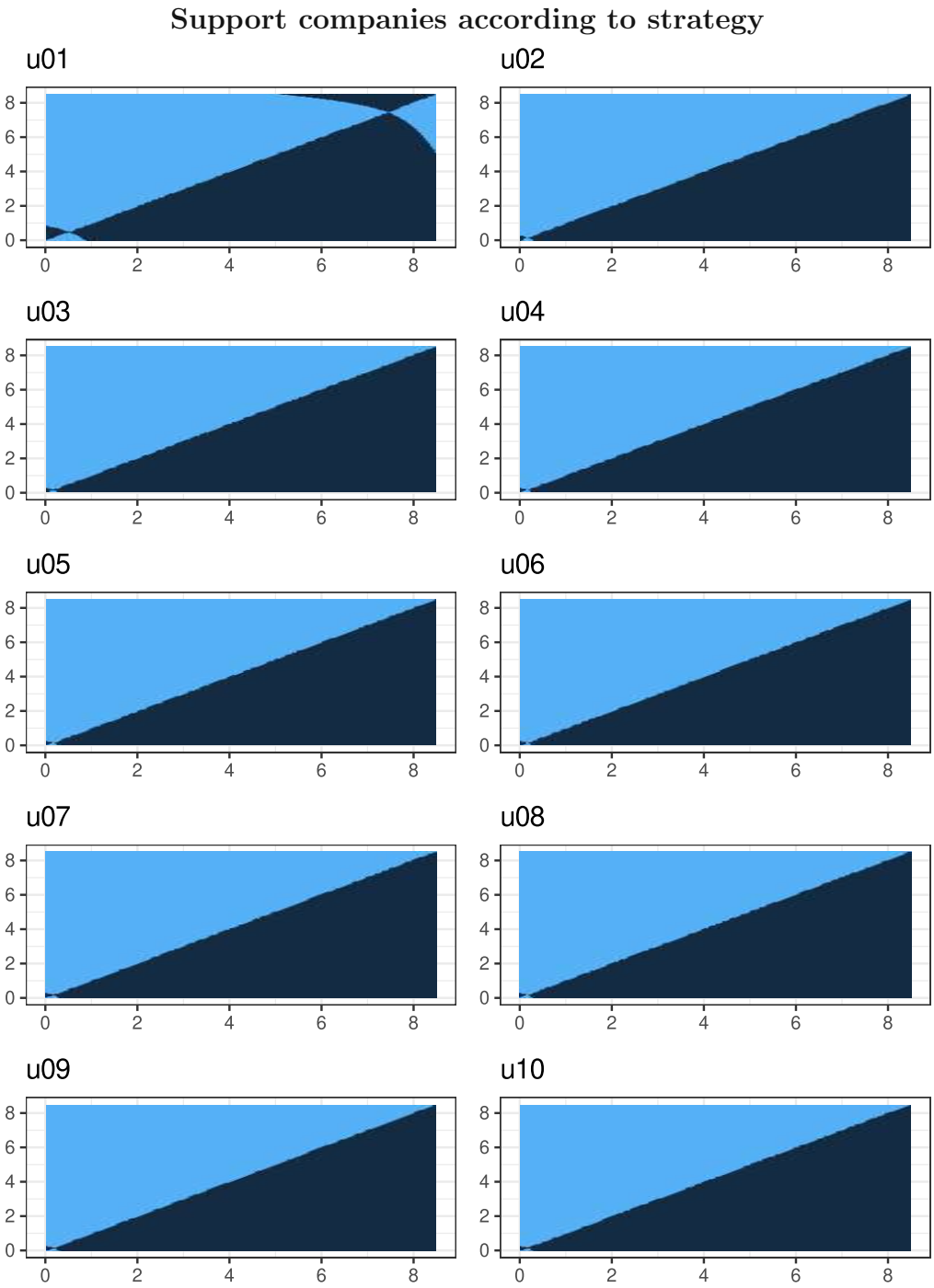


Figure 7.28: $J_1 - J_2$ for $\sigma_1^2 = \sigma_2^2 = 1, \mu = 1$ and $\alpha = \frac{1}{3}$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

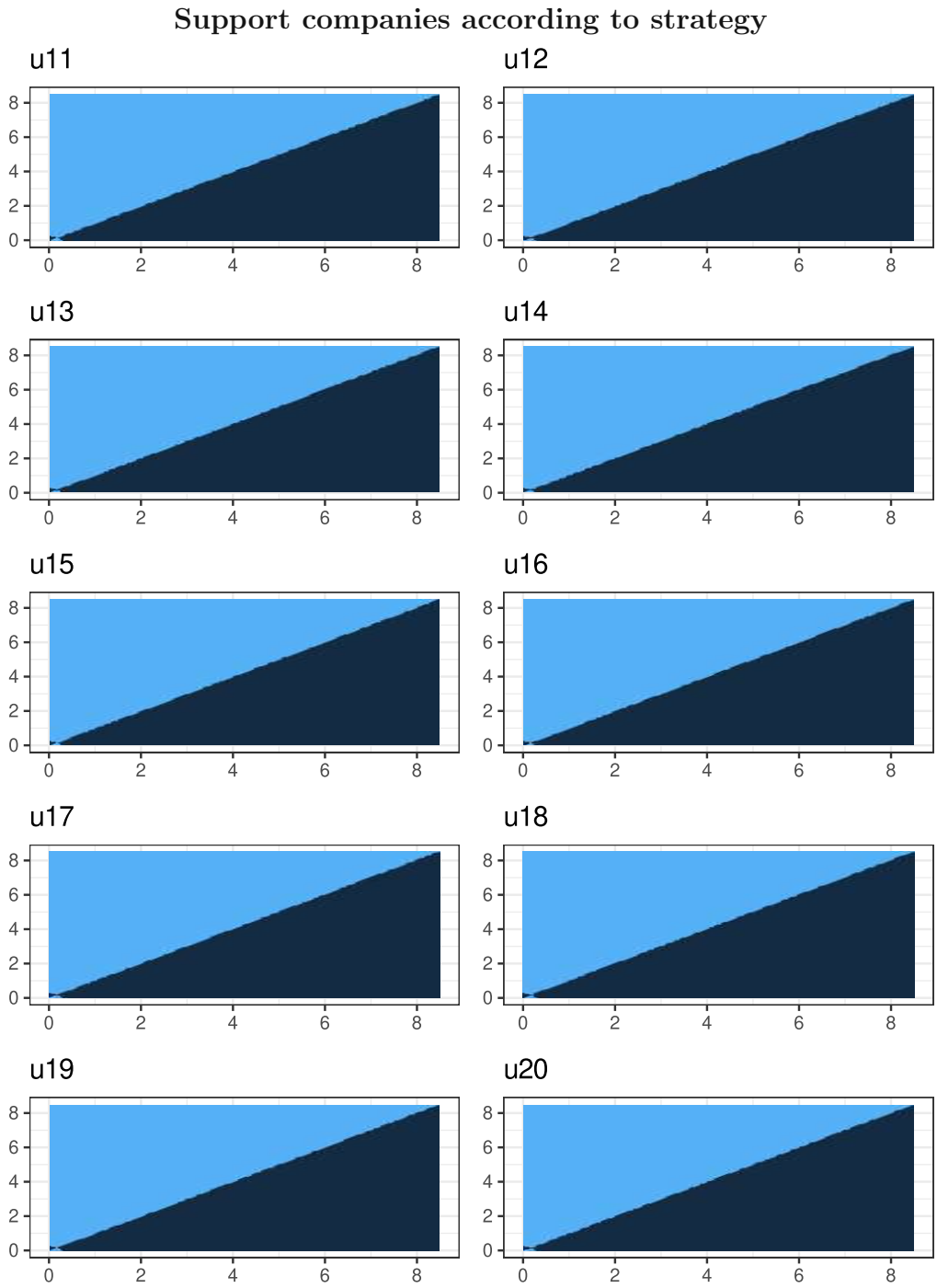


Figure 7.29: $J_1 - J_2$ for $\sigma_1^2 = \sigma_2^2 = 1, \mu = 1$ and $\alpha = \frac{1}{3}$

Each determined policy is guaranteed to be a strict improvement over the previous one, unless there is no difference between the previous policy and the current one, which implies that the latter is already optimal. In Figure 7.28 we can see that after the second determined

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

strategy $u02$, the following strategies in Figure 7.29 look very similar. However, if we take a closer look at the area near the origin, we can see that the push-bottom strategy is not optimal everywhere. Therefore we create some additional plots with focus on the area near the origin in order to say, which policy is the optimal policy for the considered optimization problem.

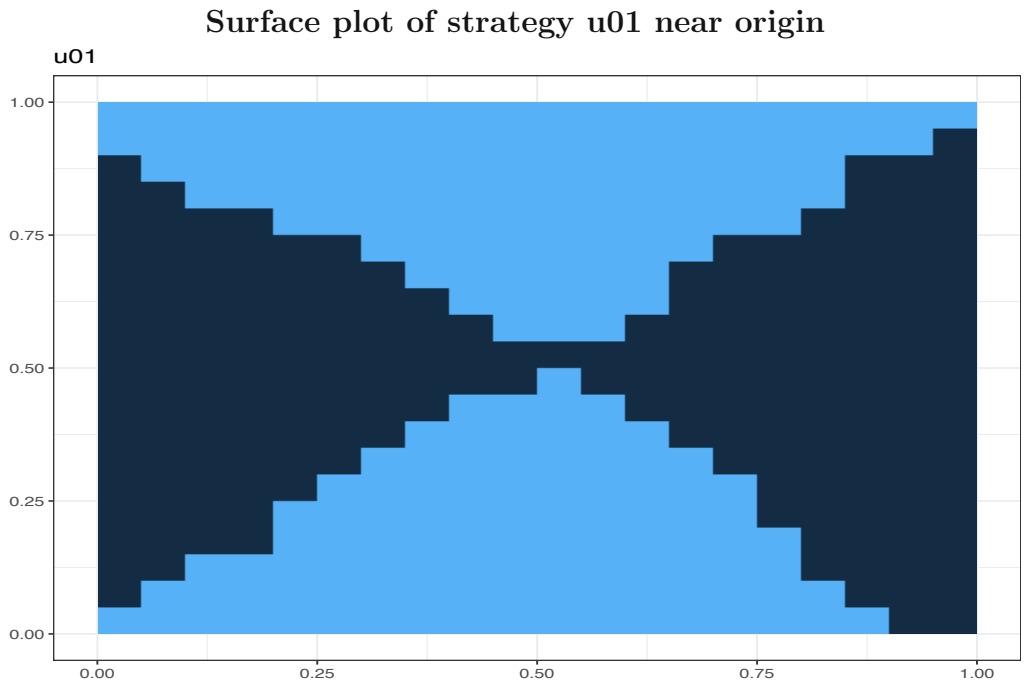


Figure 7.30: $J_1 - J_2$ for $\sigma_1^2 = \sigma_2^2 = 1, \mu = 1$ and $\alpha = \frac{1}{3}$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

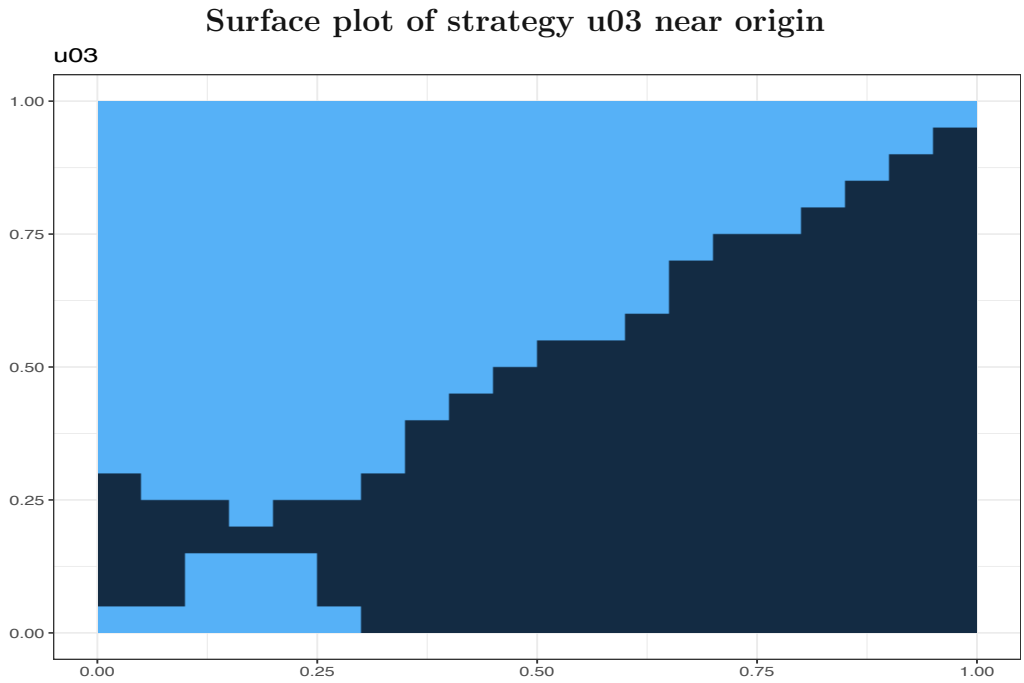


Figure 7.31: $J_1 - J_2$ for $\sigma_1^2 = \sigma_2^2 = 1, \mu = 1$ and $\alpha = \frac{1}{3}$

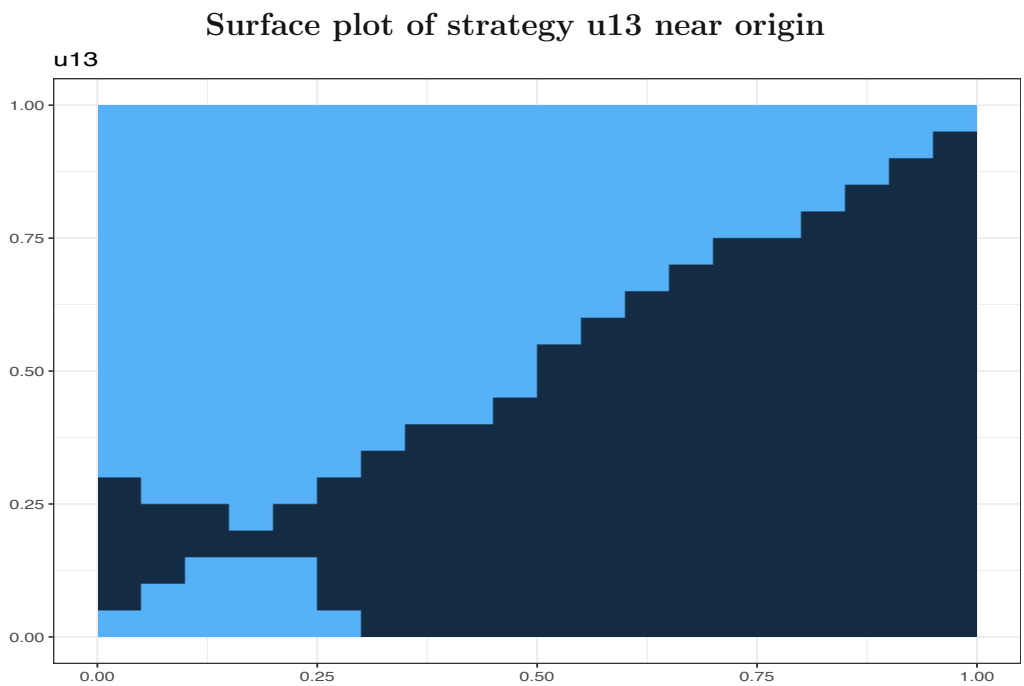


Figure 7.32: $J_1 - J_2$ for $\sigma_1^2 = \sigma_2^2 = 1, \mu = 1$ and $\alpha = \frac{1}{3}$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

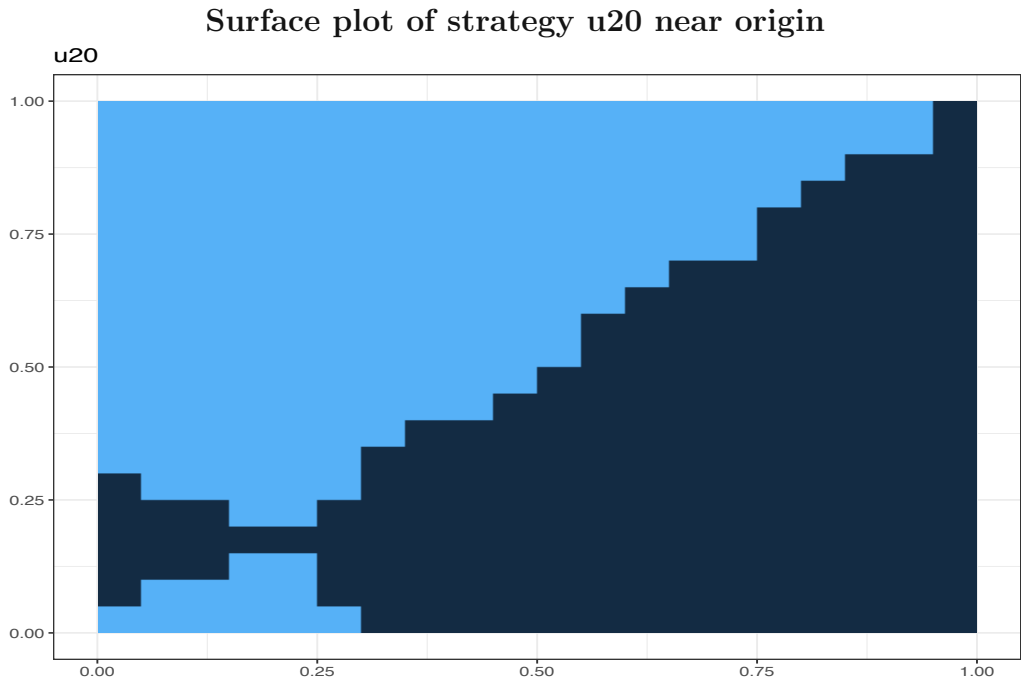


Figure 7.33: $J_1 - J_2$ for $\sigma_1^2 = \sigma_2^2 = 1, \mu = 1$ and $\alpha = \frac{1}{3}$

In Figure 7.30 we can see that up to a certain point, that is approximately at 0.9, the push-top strategy is optimal, meaning that $u[x, y] = 0$, if $x_1(t) \leq x_2(t)$, and otherwise, $u[x, y] = 1$. However, the range in which the push-top strategy is optimal becomes smaller as we can see in Figures 7.31, 7.32 and 7.33. Thus, we can say that a mixture of the push-top and the push-bottom strategy is the optimal strategy for maximizing the expected number of surviving firms with $\alpha = \frac{1}{3}$.

The second initial strategy is the push-bottom strategy as defined in (7.31). By using the policy iteration method we obtain a sequence of monotonically improving policies, $u01 - u20$, according to which the firms are supported by the state, where the colored area for each new strategy is again shown in light blue and dark blue, respectively.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

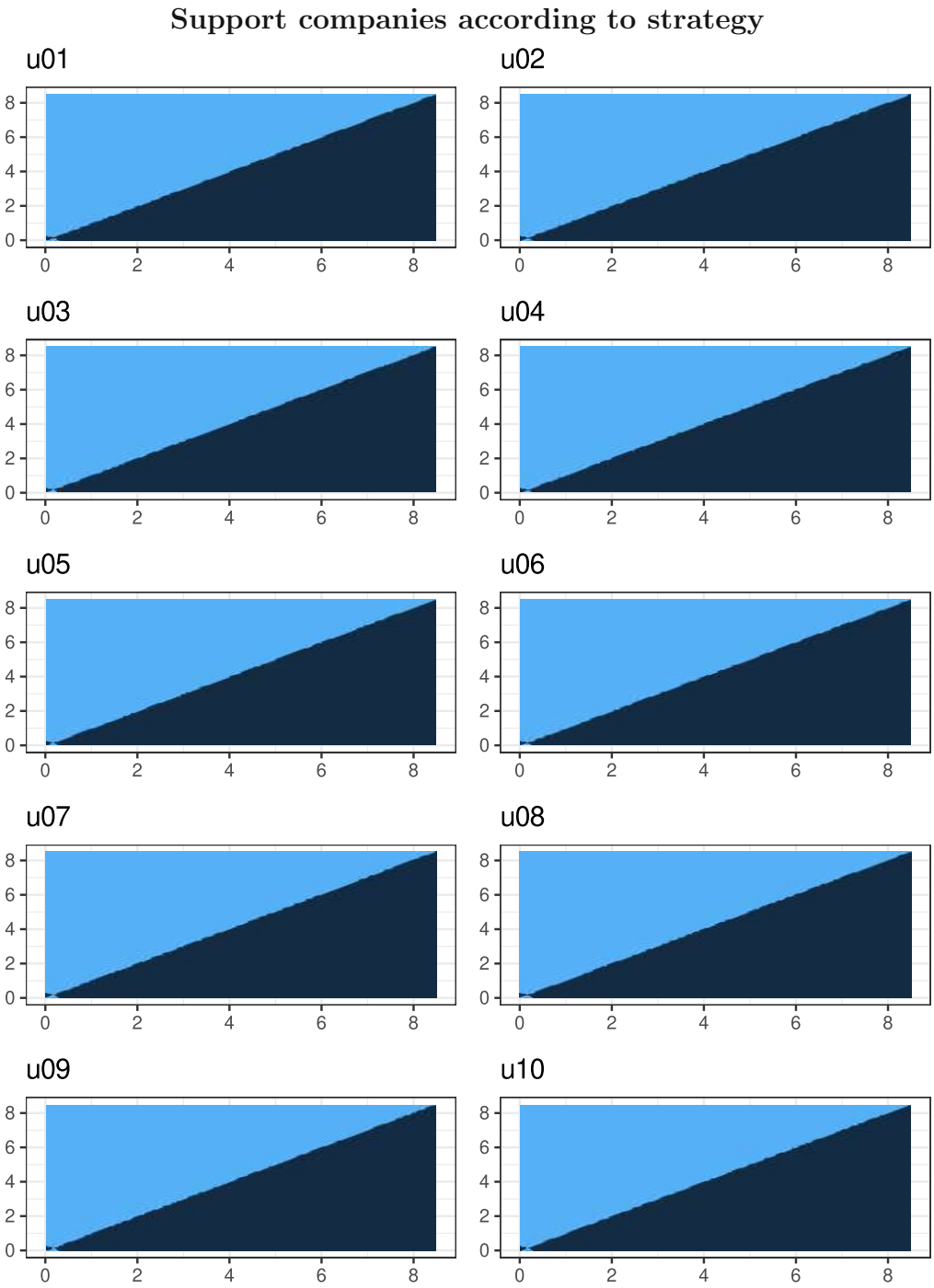


Figure 7.34: $J_1 - J_2$ for $\sigma_1^2 = \sigma_2^2 = 1, \mu = 1$ and $\alpha = \frac{1}{3}$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

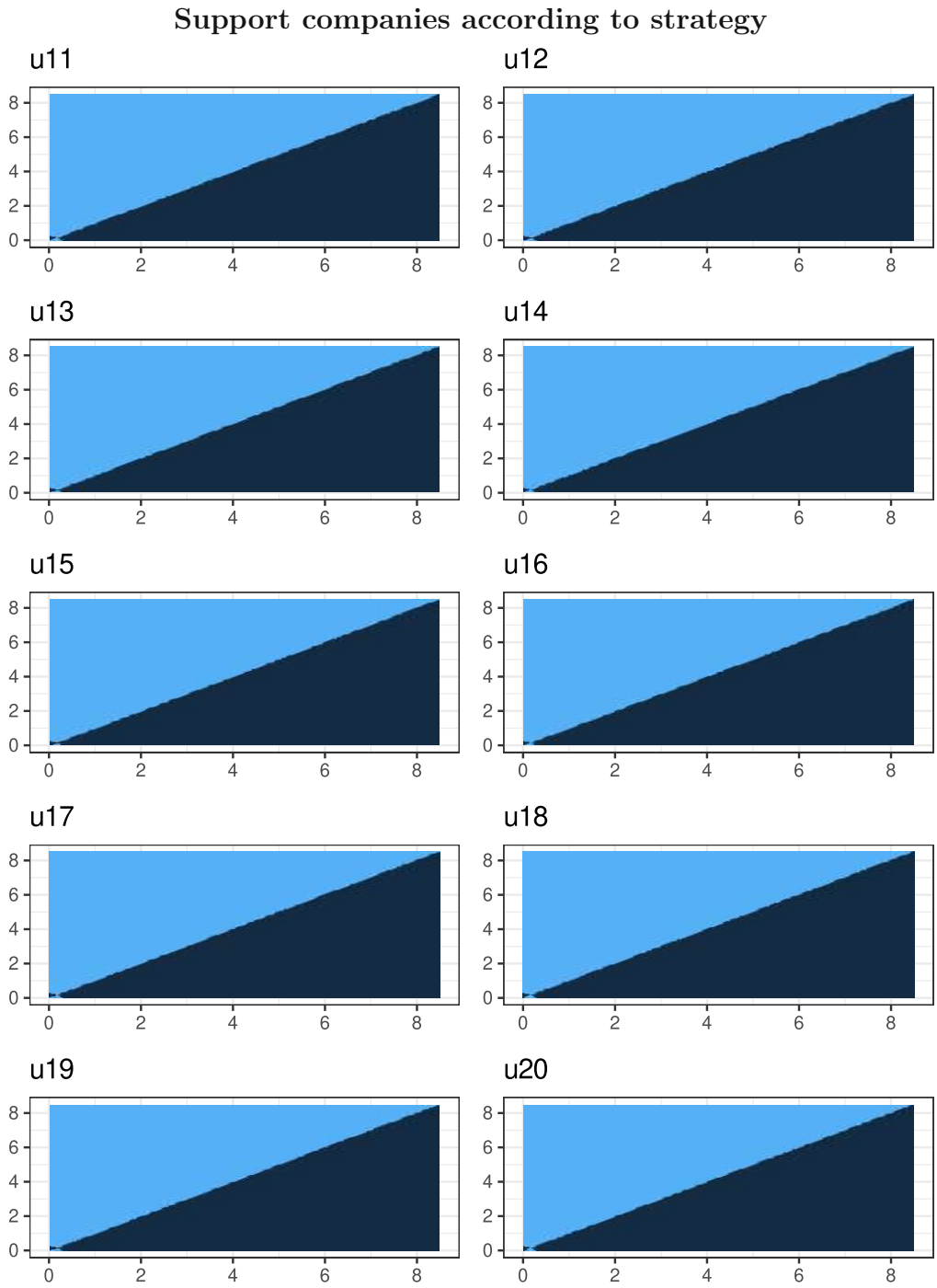


Figure 7.35: $J_1 - J_2$ for $\sigma_1^2 = \sigma_2^2 = 1, \mu = 1$ and $\alpha = \frac{1}{3}$

Each determined policy is guaranteed to be a strict improvement over the previous one, unless there is no difference between the previous policy and the current one, which implies that the latter is already optimal. In Figure 7.34 we can see that after the first determined

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

strategy $u01$, the following strategies in Figure 7.35 look very similar. However, if we take a closer look at the area near the origin, we can see that the push-bottom strategy is not optimal everywhere. Therefore we create some additional plots with focus on the area near the origin in order to say, which policy is the optimal policy for the considered optimization problem.

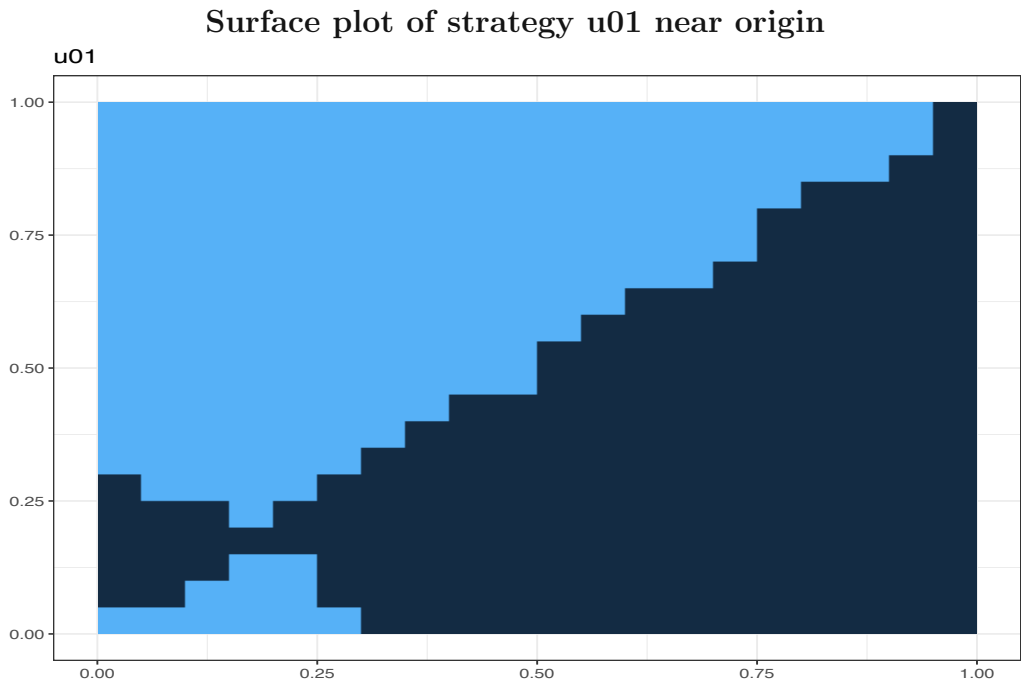


Figure 7.36: $J_1 - J_2$ for $\sigma_1^2 = \sigma_2^2 = 1, \mu = 1$ and $\alpha = \frac{1}{3}$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Surface plot of strategy u03 near origin

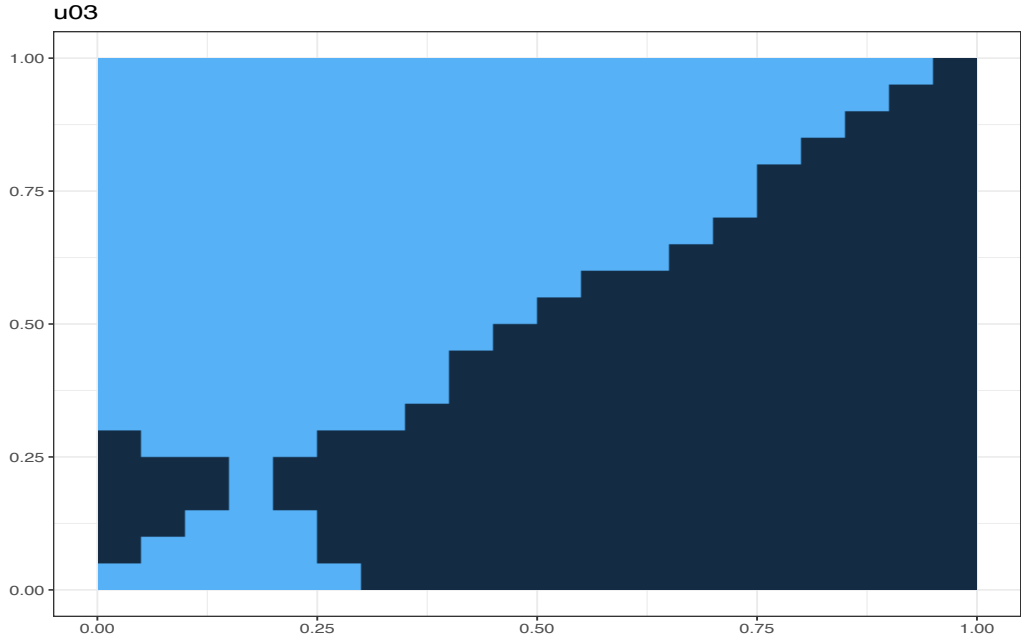


Figure 7.37: $J_1 - J_2$ for $\sigma_1^2 = \sigma_2^2 = 1, \mu = 1$ and $\alpha = \frac{1}{3}$

Surface plot of strategy u18 near origin

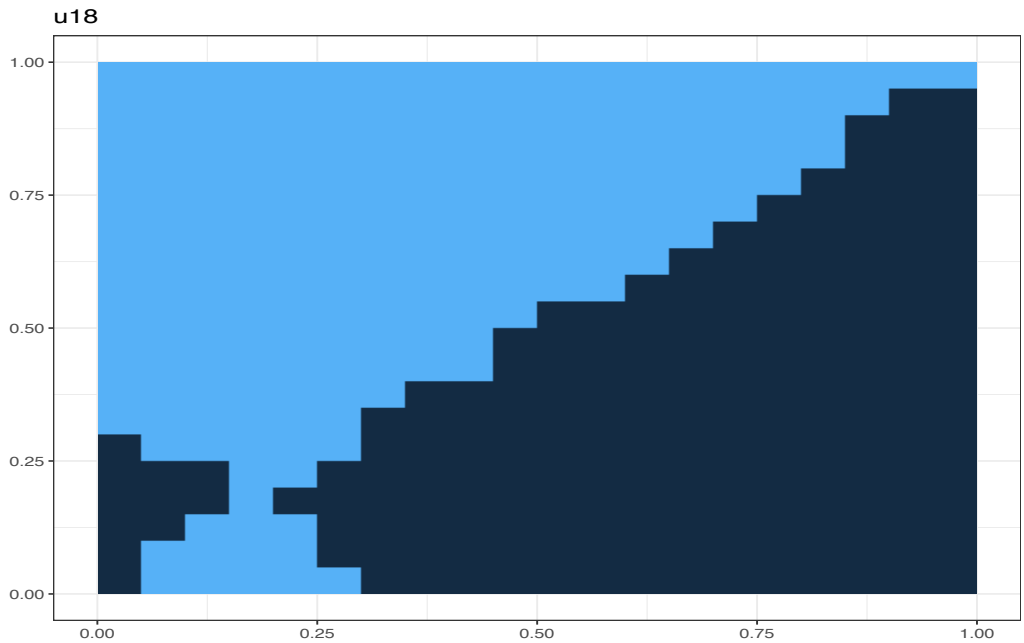


Figure 7.38: $J_1 - J_2$ for $\sigma_1^2 = \sigma_2^2 = 1, \mu = 1$ and $\alpha = \frac{1}{3}$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

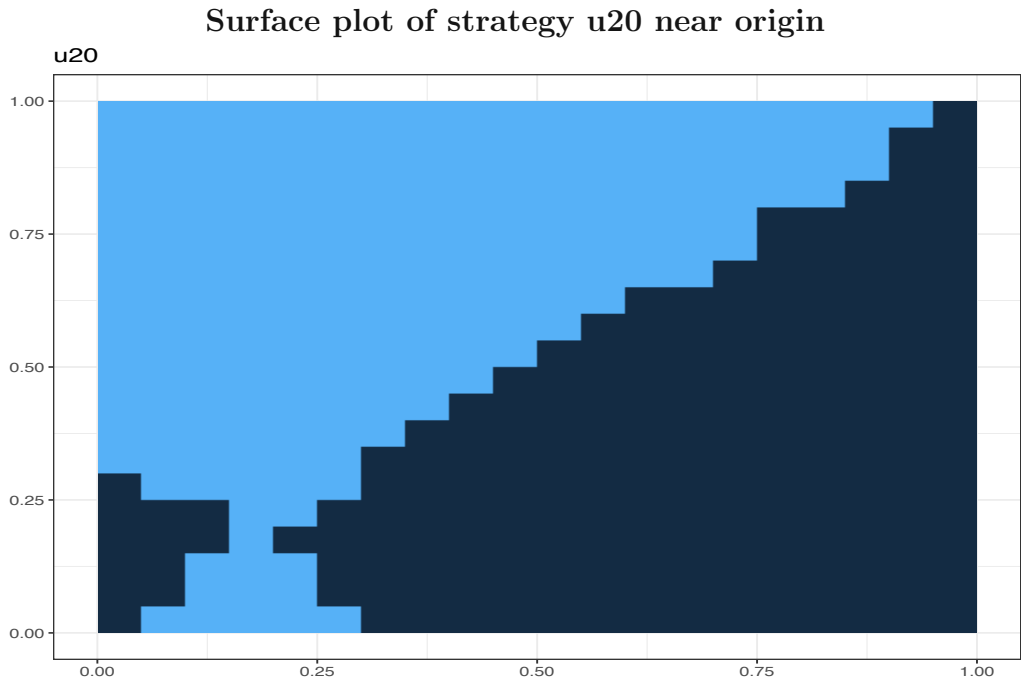


Figure 7.39: $J_1 - J_2$ for $\sigma_1^2 = \sigma_2^2 = 1, \mu = 1$ and $\alpha = \frac{1}{3}$

In Figures 7.36, 7.37, 7.38 and 7.39 we can again see that near the origin, when both companies are in danger of going bankrupt it is optimal to always apply the drift to the stronger company, meaning that $u[x, y] = 0$, if $x_1(t) \leq x_2(t)$, and otherwise, $u[x, y] = 1$. Thus, we can say that a mixture of the push-top and the push-bottom strategy is the optimal strategy for maximizing the expected number of surviving firms with $\alpha = \frac{1}{3}$.

The third initial strategy is the strategy defined in (7.33). By using the policy iteration method we obtain again a sequence of monotonically improving policies, $u01 - u20$, according to which the firms are supported by the state, where the colored area for each new strategy is again shown in light blue and dark blue, respectively.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

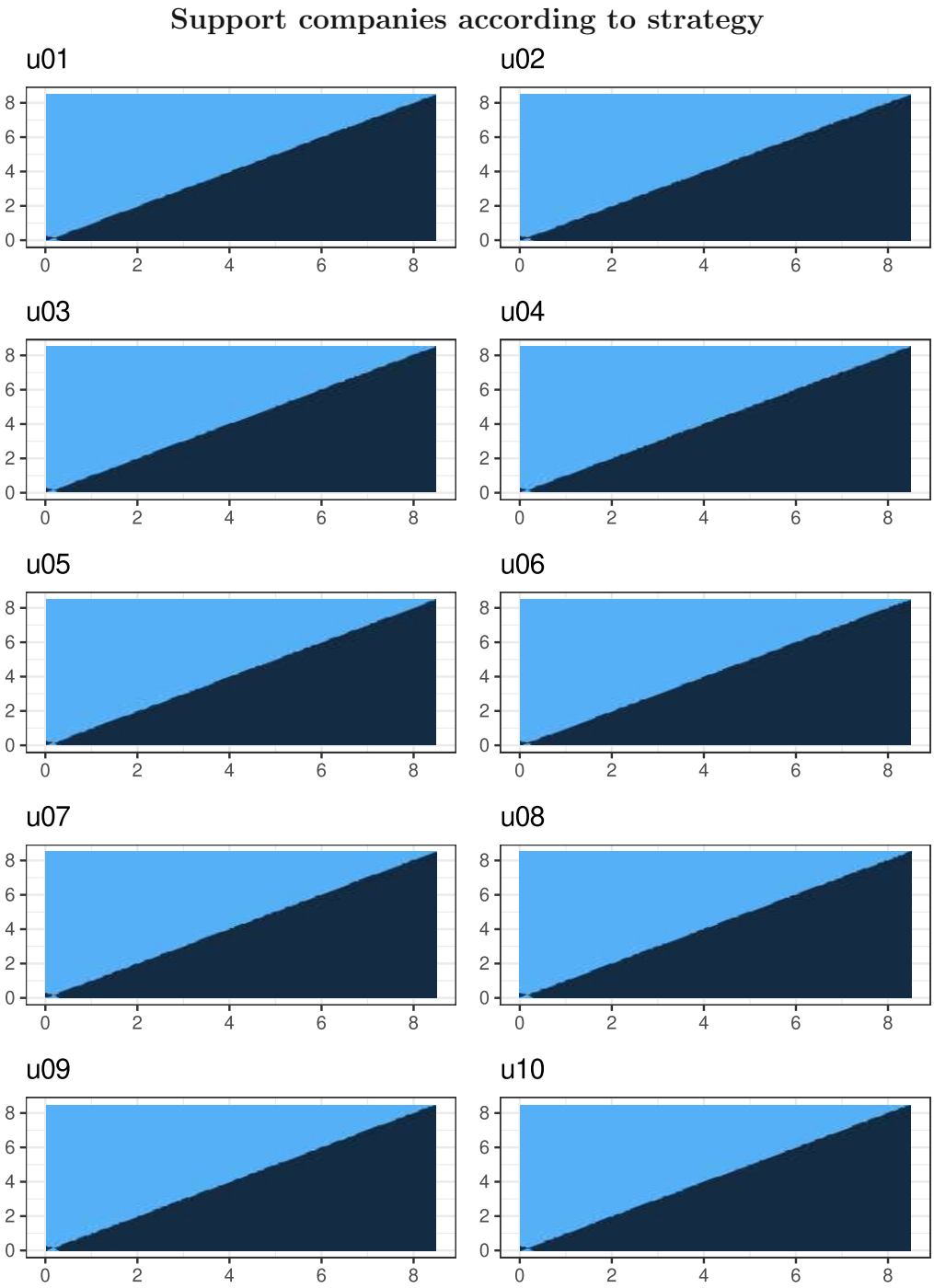


Figure 7.40: $J_1 - J_2$ for $\sigma_1^2 = \sigma_2^2 = 1, \mu = 1$ and $\alpha = \frac{1}{3}$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

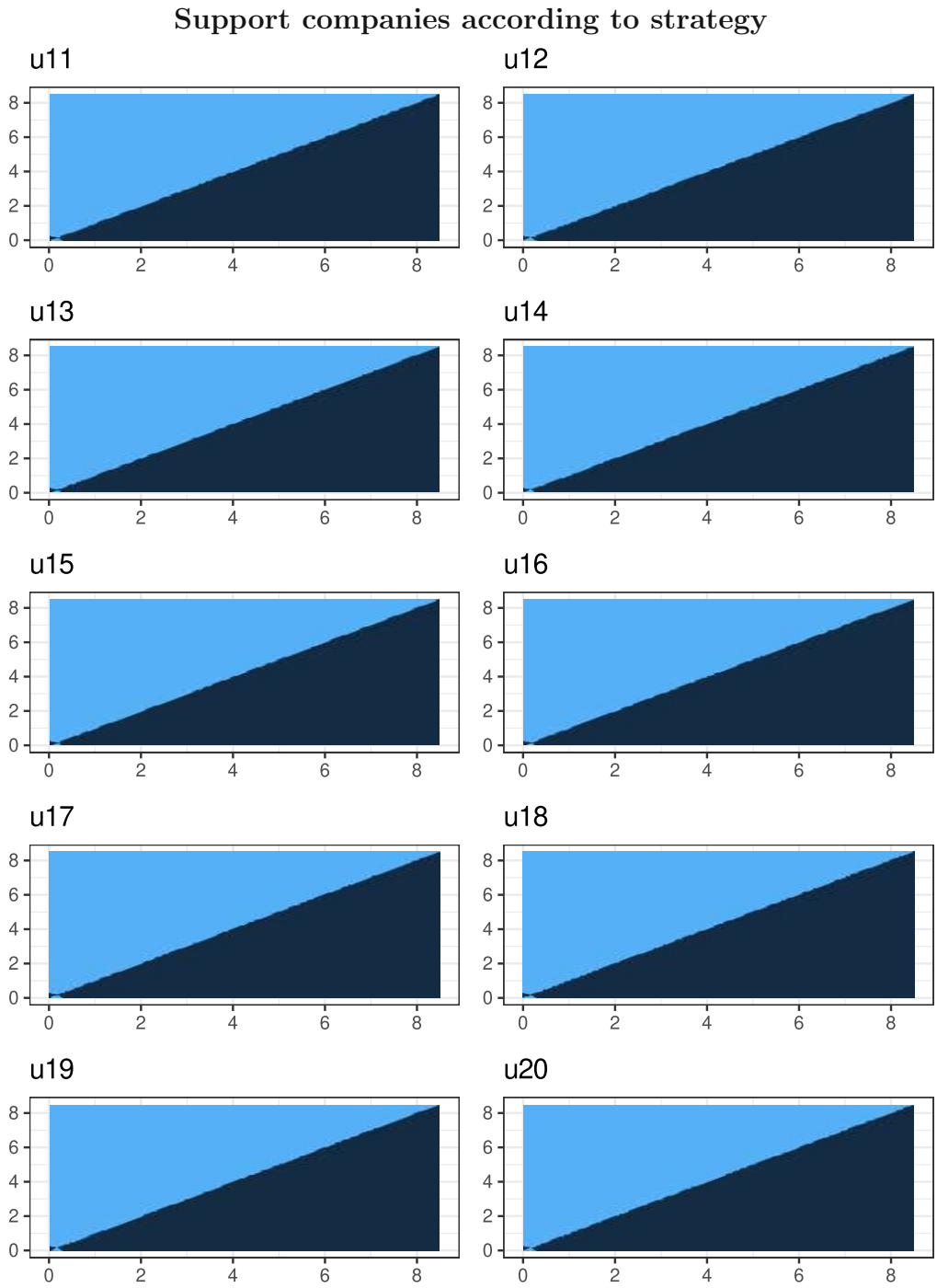


Figure 7.41: $J_1 - J_2$ for $\sigma_1^2 = \sigma_2^2 = 1, \mu = 1$ and $\alpha = \frac{1}{3}$

Each determined policy is guaranteed to be a strict improvement over the previous one, unless there is no difference between the previous policy and the current one, which implies that the latter is already optimal. In Figure 7.40 we can see that after the first determined

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

strategy $u01$, the following strategies in Figure 7.41 look very similar. However, if we take a closer look at the area near the origin, we can see that the push-bottom strategy is not optimal everywhere. Therefore we create some additional plots with focus on the area near the origin in order to say, which policy is the optimal policy for the considered optimization problem.

Surface plot of strategy u01 near origin

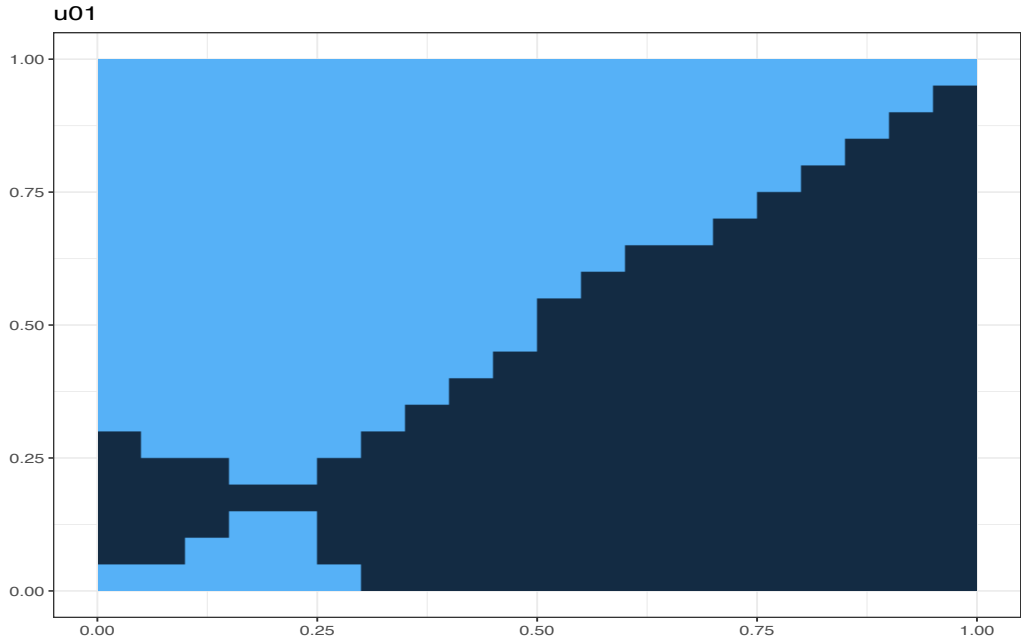


Figure 7.42: $J_1 - J_2$ for $\sigma_1^2 = \sigma_2^2 = 1, \mu = 1$ and $\alpha = \frac{1}{3}$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Surface plot of strategy u03 near origin

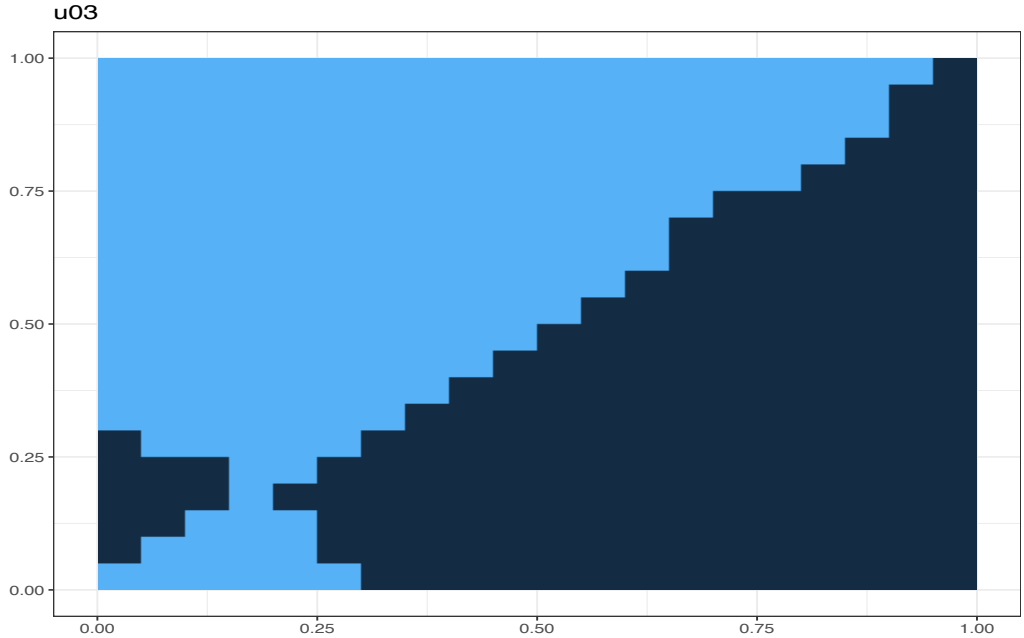


Figure 7.43: $J_1 - J_2$ for $\sigma_1^2 = \sigma_2^2 = 1, \mu = 1$ and $\alpha = \frac{1}{3}$

Surface plot of strategy u14 near origin

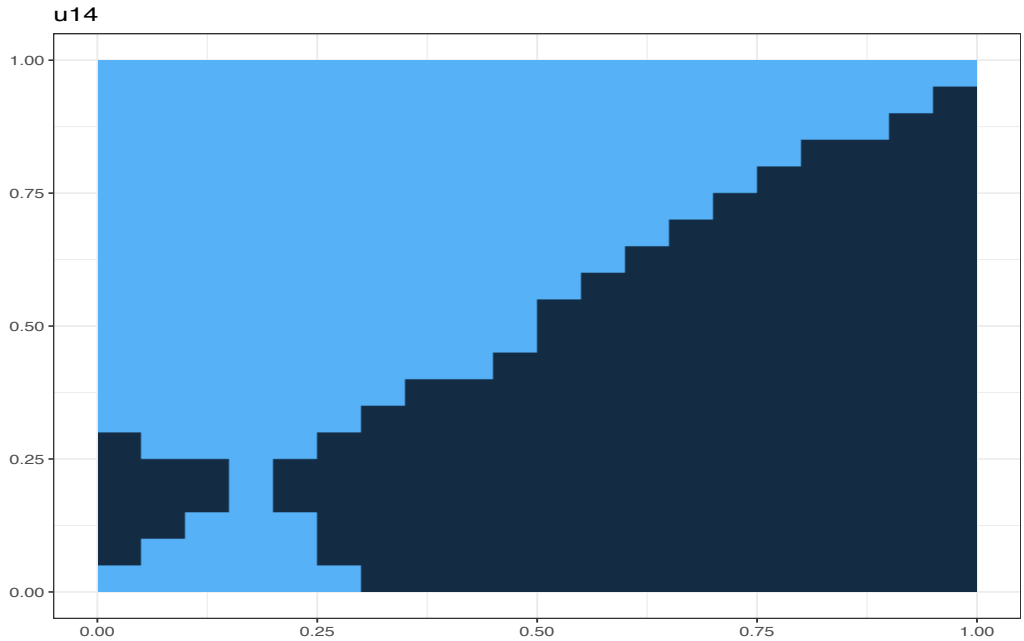


Figure 7.44: $J_1 - J_2$ for $\sigma_1^2 = \sigma_2^2 = 1, \mu = 1$ and $\alpha = \frac{1}{3}$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

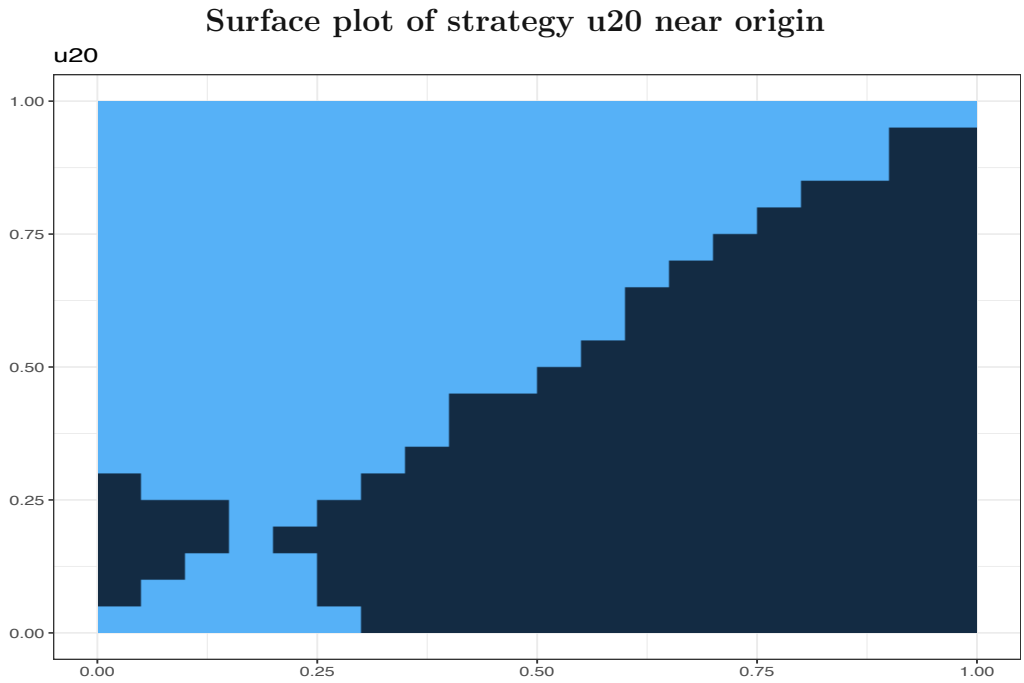


Figure 7.45: $J_1 - J_2$ for $\sigma_1^2 = \sigma_2^2 = 1, \mu = 1$ and $\alpha = \frac{1}{3}$

In Figures 7.42, 7.43, 7.44 and 7.45 we can again see that up to a certain point, that is approximately at 0.3, the push-top strategy is optimal, meaning that $u[x, y] = 0$, if $x_1(t) \leq x_2(t)$, and otherwise, $u[x, y] = 1$. Thus, we can say that a mixture of the push-top and the push-bottom strategy is the optimal strategy for maximizing the expected number of surviving firms with $\alpha = \frac{1}{3}$.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Now, we will consider the case of $\alpha = \frac{2}{5}$. In the following, we will use again different initial strategies and show, that each of them is leading to the same result in the sense of the optimal strategy for the considered optimization problem, that is, as in the case for $\alpha = \frac{1}{3}$, a mixture of the push-top and push-bottom strategy. However, we except that the range, where the push-top strategy is optimal will increase. Therefore, we will create plots that only show the area near the origin, since we already know that the push-bottom strategy is optimal for the area not near the origin.

First, we use the push-bottom strategy as defined in (7.31) as initial strategy.

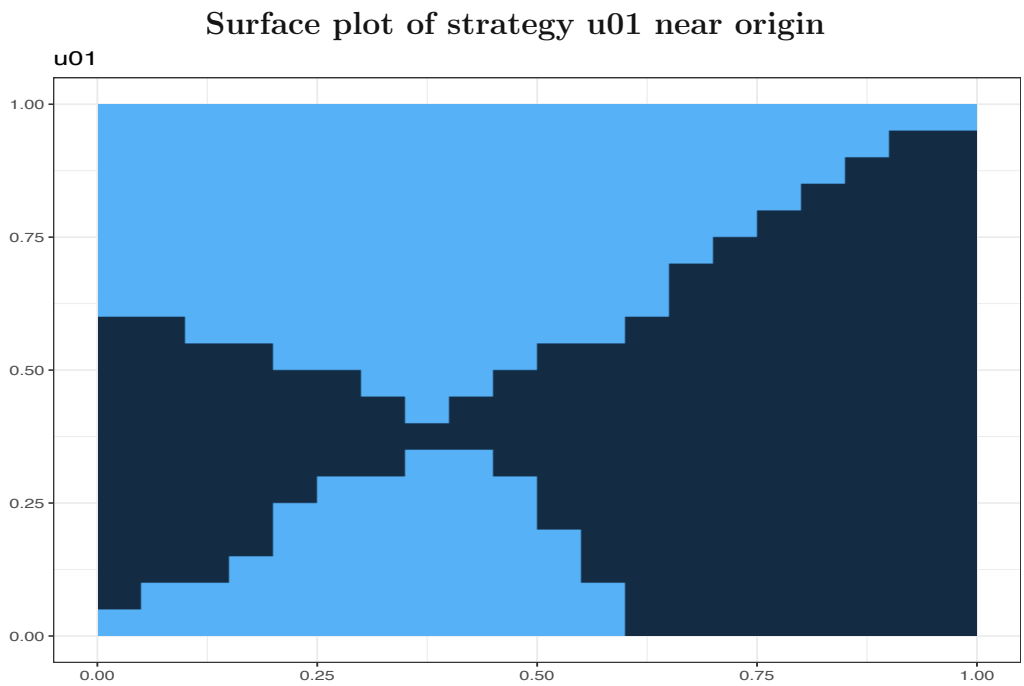


Figure 7.46: $J_1 - J_2$ for $\sigma_1^2 = \sigma_2^2 = 1, \mu = 1$ and $\alpha = \frac{2}{5}$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Surface plot of strategy u03 near origin

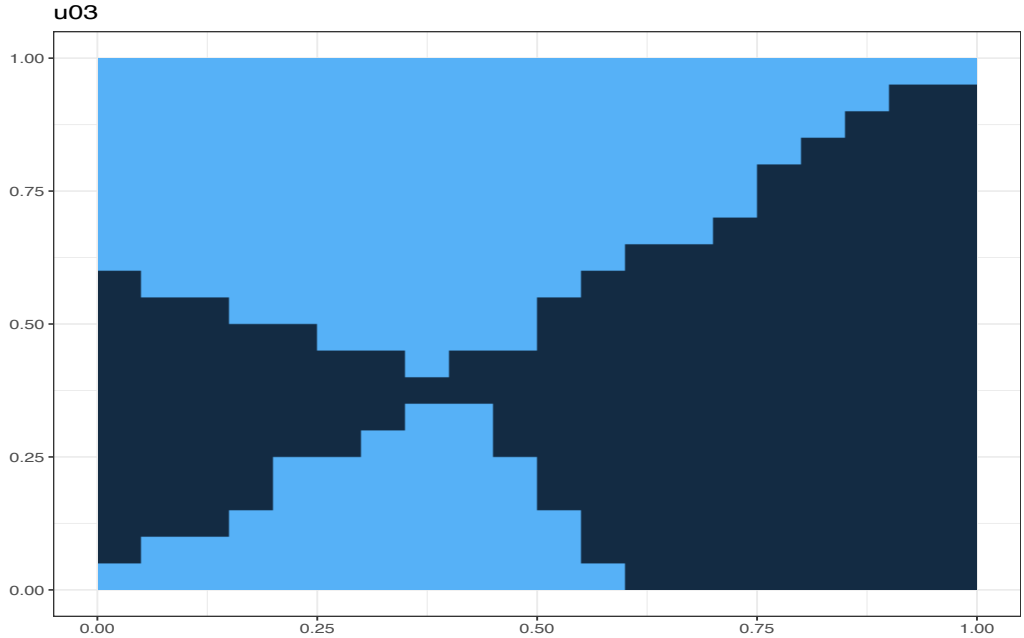


Figure 7.47: $J_1 - J_2$ for $\sigma_1^2 = \sigma_2^2 = 1, \mu = 1$ and $\alpha = \frac{2}{5}$

Surface plot of strategy u04 near origin

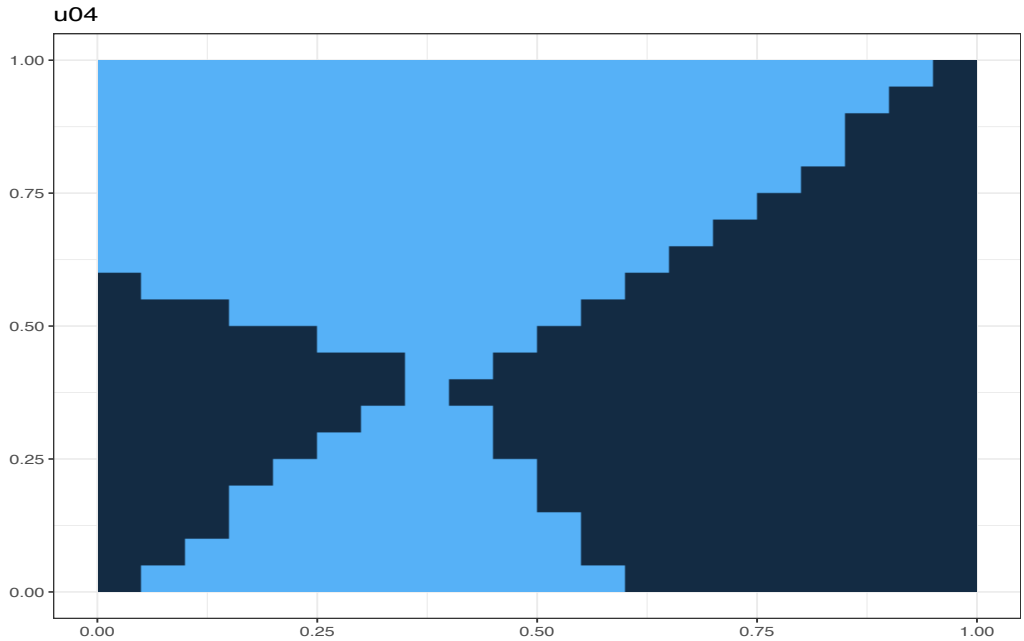


Figure 7.48: $J_1 - J_2$ for $\sigma_1^2 = \sigma_2^2 = 1, \mu = 1$ and $\alpha = \frac{2}{5}$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

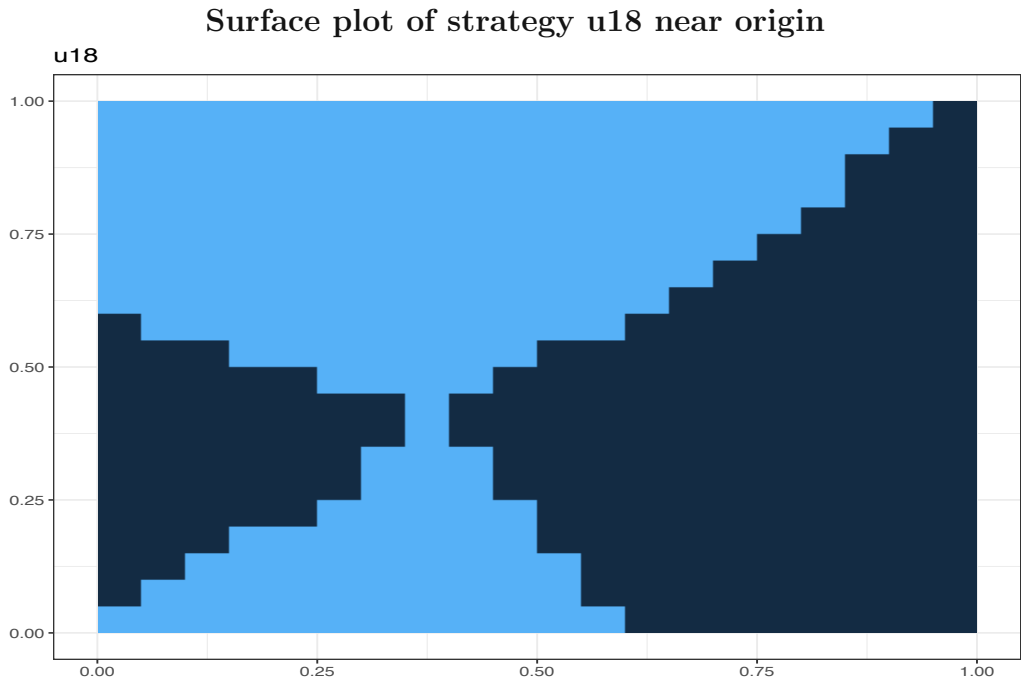


Figure 7.49: $J_1 - J_2$ for $\sigma_1^2 = \sigma_2^2 = 1, \mu = 1$ and $\alpha = \frac{2}{5}$

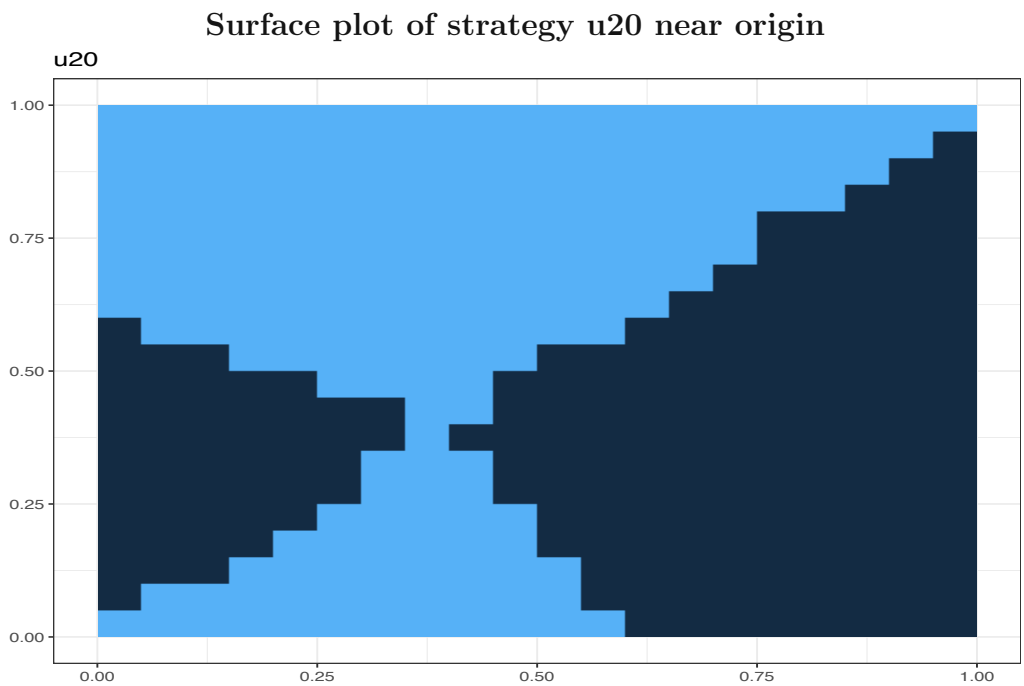


Figure 7.50: $J_1 - J_2$ for $\sigma_1^2 = \sigma_2^2 = 1, \mu = 1$ and $\alpha = \frac{2}{5}$

In Figures 7.46, 7.47, 7.48, 7.49 and 7.50 we can see that, as we expected, in the area near

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

the origin, where both companies are in danger of going bankrupt, it is optimal to push the stronger company, meaning that the push-top strategy is optimal. However, the area now is larger than for the case $\alpha = \frac{1}{3}$. Hence, up to a certain point, that is approximately at 0.6, the stronger company is receiving the drift from the government at the considered time point, meaning that $u[x, y] = 0$, if $x_1(t) \leq x_2(t)$, and otherwise, $u[x, y] = 1$. Thus, we can say that a mixture of the push-top and the push-bottom strategy is again the optimal strategy for maximizing the expected number of surviving firms with $\alpha = \frac{2}{5}$.

Second, we use the push-top strategy as defined in (7.34) as initial strategy.

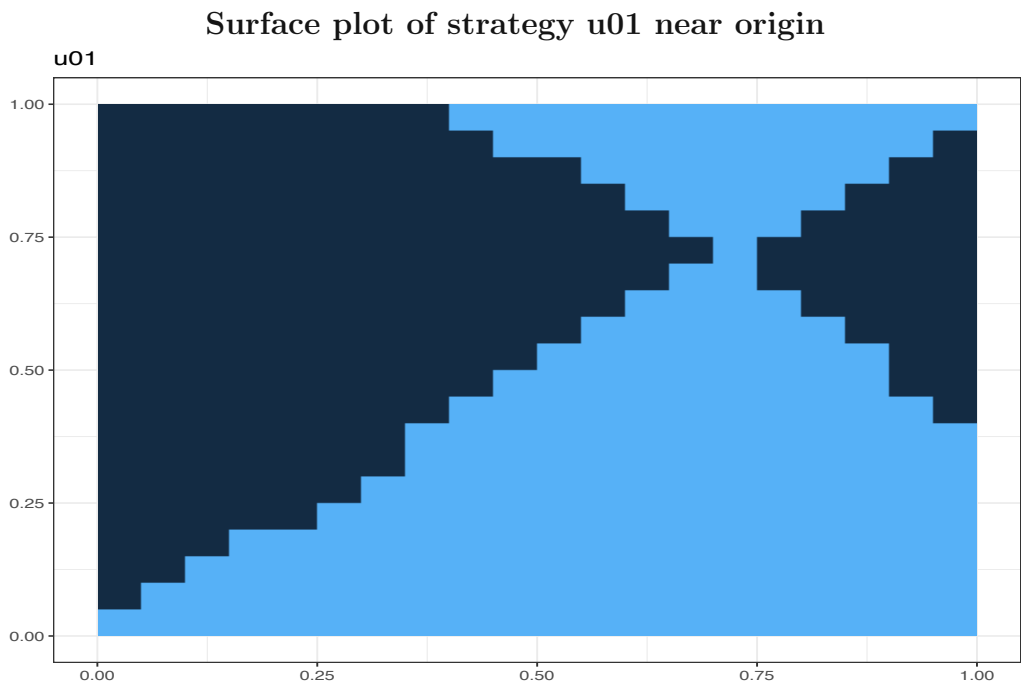


Figure 7.51: $J_1 - J_2$ for $\sigma_1^2 = \sigma_2^2 = 1, \mu = 1$ and $\alpha = \frac{2}{5}$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Surface plot of strategy u05 near origin

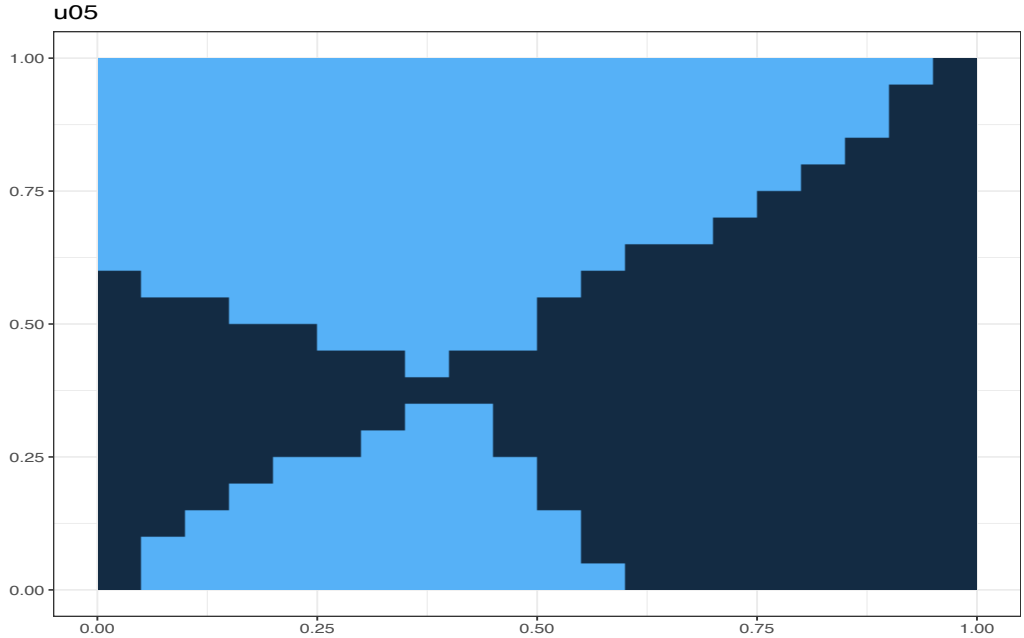


Figure 7.52: $J_1 - J_2$ for $\sigma_1^2 = \sigma_2^2 = 1, \mu = 1$ and $\alpha = \frac{2}{5}$

Surface plot of strategy u06 near origin

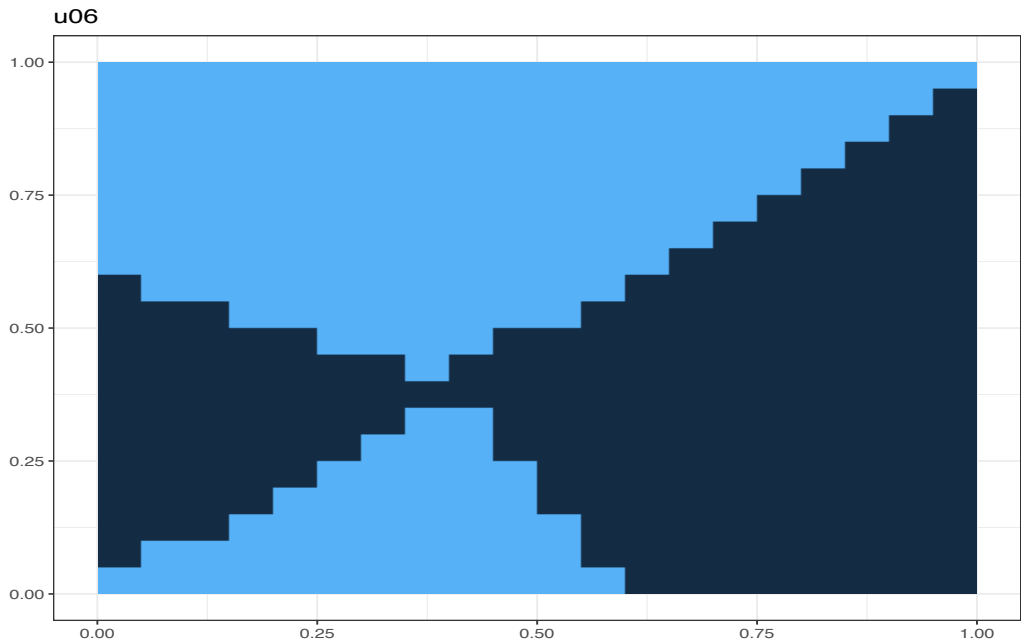


Figure 7.53: $J_1 - J_2$ for $\sigma_1^2 = \sigma_2^2 = 1, \mu = 1$ and $\alpha = \frac{2}{5}$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

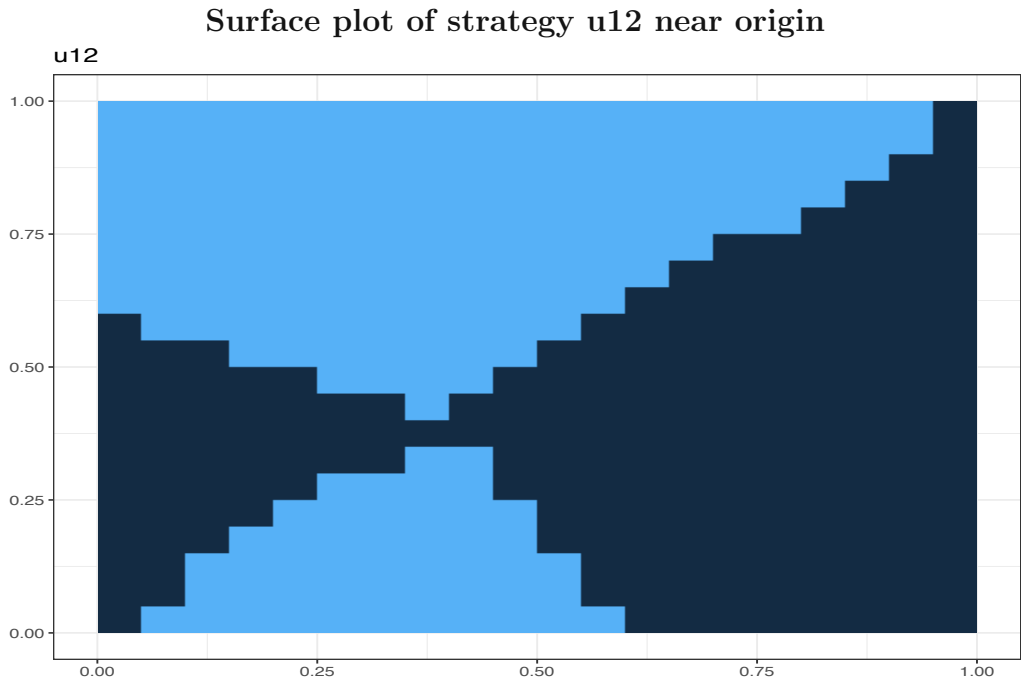


Figure 7.54: $J_1 - J_2$ for $\sigma_1^2 = \sigma_2^2 = 1, \mu = 1$ and $\alpha = \frac{2}{5}$

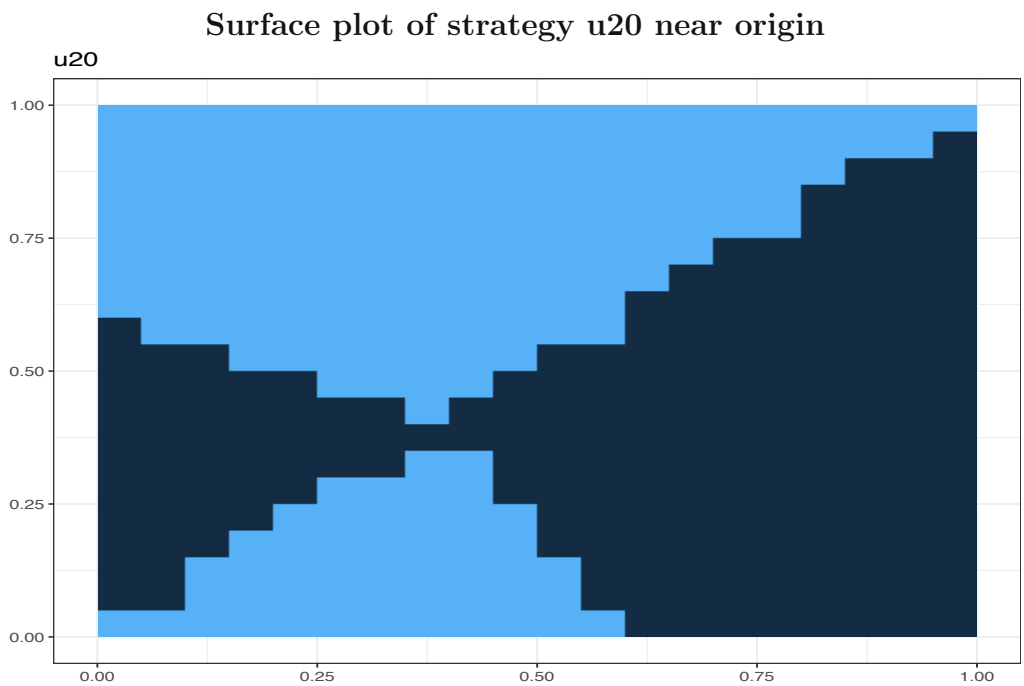


Figure 7.55: $J_1 - J_2$ for $\sigma_1^2 = \sigma_2^2 = 1, \mu = 1$ and $\alpha = \frac{2}{5}$

In Figures 7.52, 7.53, 7.54 and 7.55 we can again see that the push-top strategy is optimal

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

near the origin. Hence, up to a certain point, that is approximately at 0.6, the stronger company is receiving the drift from the government at the considered time point, meaning that $u[x, y] = 0$, if $x_1(t) \leq x_2(t)$, and otherwise, $u[x, y] = 1$. Thus, we can say that a mixture of the push-top and the push-bottom strategy is again the optimal strategy for maximizing the expected number of surviving firms with $\alpha = \frac{2}{5}$.

Third, we use the strategy as defined in (7.33) as initial strategy.

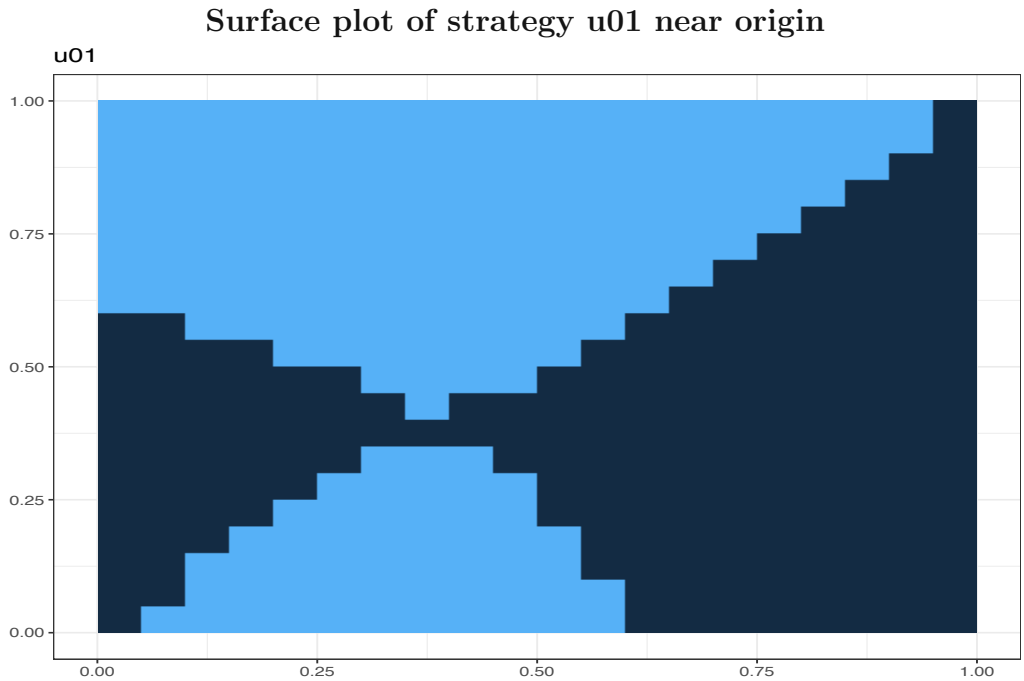


Figure 7.56: $J_1 - J_2$ for $\sigma_1^2 = \sigma_2^2 = 1, \mu = 1$ and $\alpha = \frac{2}{5}$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

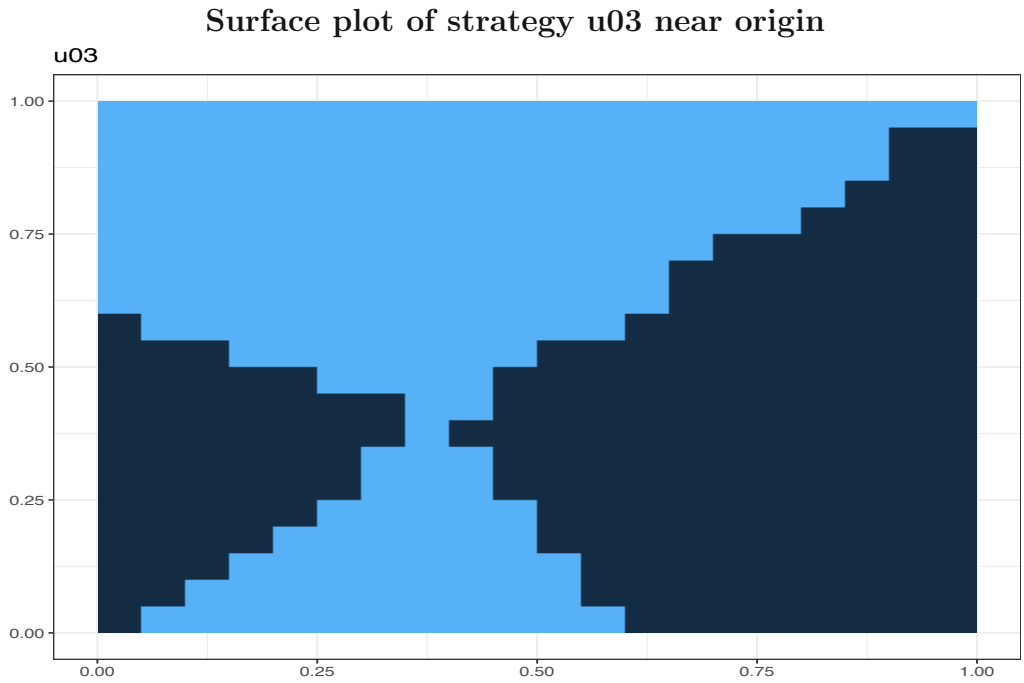


Figure 7.57: $J_1 - J_2$ for $\sigma_1^2 = \sigma_2^2 = 1, \mu = 1$ and $\alpha = \frac{2}{5}$

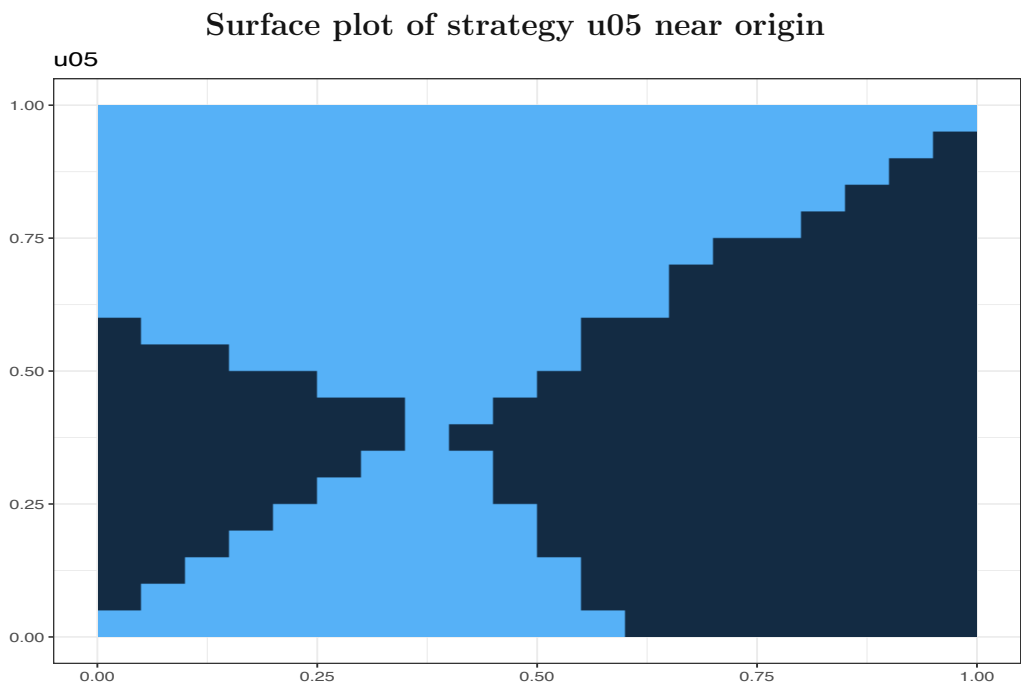


Figure 7.58: $J_1 - J_2$ for $\sigma_1^2 = \sigma_2^2 = 1, \mu = 1$ and $\alpha = \frac{2}{5}$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

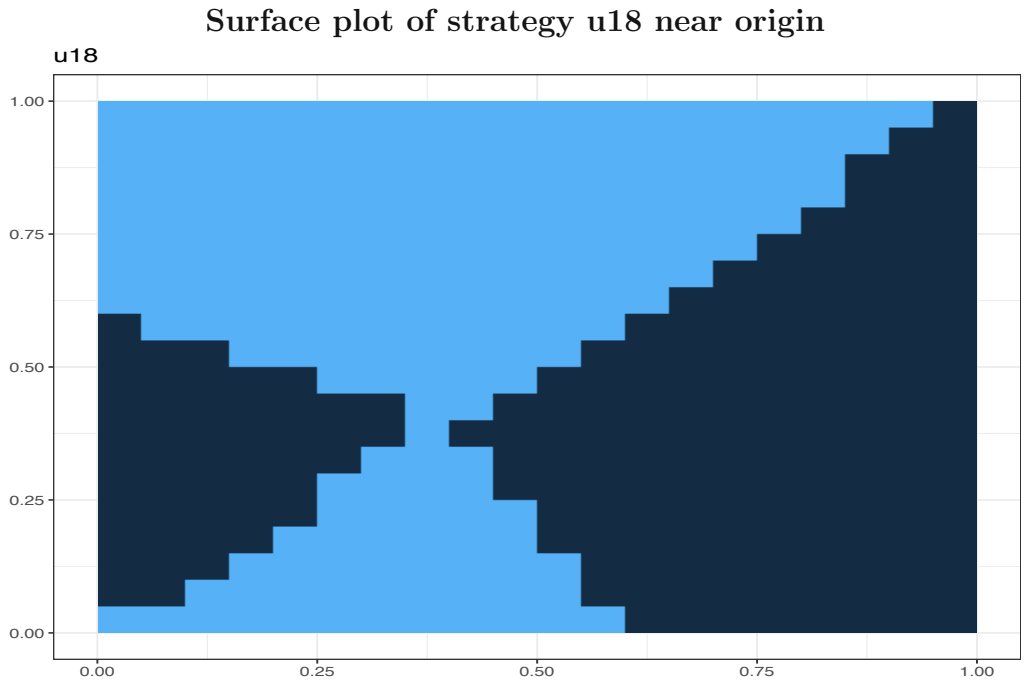


Figure 7.59: $J_1 - J_2$ for $\sigma_1^2 = \sigma_2^2 = 1, \mu = 1$ and $\alpha = \frac{2}{5}$

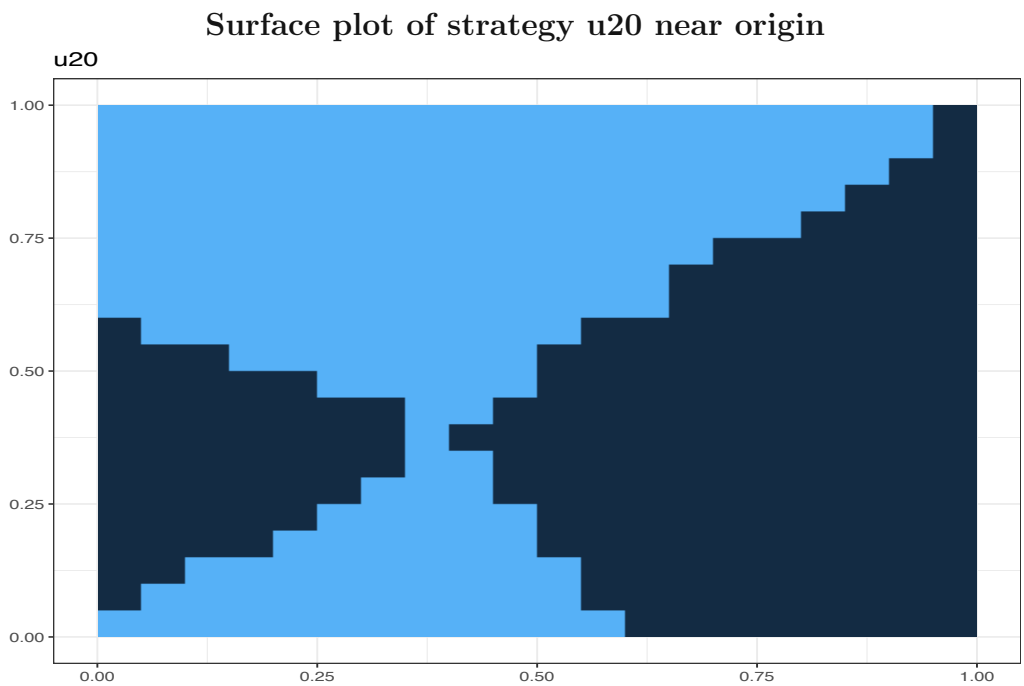


Figure 7.60: $J_1 - J_2$ for $\sigma_1^2 = \sigma_2^2 = 1, \mu = 1$ and $\alpha = \frac{2}{5}$

In Figures 7.56, 7.57, 7.58, 7.59 and 7.60 we can again see that the drifts are always

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

allocated to the stronger company under this republican criterion of optimization. Thus, we can say that a mixture of the push-top and the push-bottom strategy is again the optimal strategy for maximizing the expected number of surviving firms with $\alpha = \frac{2}{5}$.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

8 Conclusion

In conclusion, the choice of the optimal strategy depends on the optimization criterion, which can be either the maximization of the joint probability of survival or the maximization of the expected number of surviving companies. Thus, in our thesis, we were considering the following optimization problem

$$\max_{\pi} \alpha \cdot \mathbb{P}[\nu = 1] + (1 - \alpha) \cdot \mathbb{P}[\nu = 2],$$

where $\nu = \nu^{\pi} \in \{0, 1, 2\}$ is denoting how many companies survive forever under a non-anticipatory policy π.

If the criterion is to maximize the probability that both companies never face bankruptcy, which means that $\alpha = 0$, the authors of [18] have guessed the optimal payoff function "with great luck and also great difficulty" but only for $\delta = 0$ by using the push-bottom strategy as optimal control strategy. Our numerical results, see Section 7.4, also show that the democratic policy is the optimal control strategy for the case $\alpha = 0$.

If the criterion is to maximize the expected number of companies that never go bankrupt, which means that $\alpha = \frac{1}{3}$, the authors of [18] have obtained that near the origin it is optimal to use the push-top strategy, and otherwise the push-bottom strategy as optimal control strategy. Here, the probability that both firms remain solvent is assigned more weight by a larger weighting factor $(1 - \alpha)$. Our numerical approximations indicate, that near the origin, where both companies are in danger to face bankruptcy, the republican policy is optimal. Otherwise, the democratic policy should be applied in order to maximize the expected number of companies that remain solvent. Hence, for the case $\alpha = \frac{1}{3}$, a mixture of the push-top and the push-bottom strategy is optimal.

If the criterion is to maximize the probability that at least one company survives forever, which means that $\alpha = \frac{1}{2}$, the authors have derived the same results as for the case $\alpha = \frac{1}{3}$. Here, the probabilities that one company remains solvent and the probability that both companies survive forever are equally weighted. Our numerical approximations show that in this case the republican policy is the optimal control strategy, where always the top company is pushed.

We were able to obtain similar results, as shown by the authors of [18] in their article, except for the case $\alpha = \frac{1}{2}$. However, since $\alpha = \frac{1}{2}$ also corresponds to the optimization problem of maximizing the probability that at least one firm survives forever, we showed numerically, which could already be surmised heuristically, that the capitalist tax system must be used in this case.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Bibliography

[1] D. Aldous. "up the river" game story. 2000.

[2] S. Asmussen and H. Albrecher. *Ruin Probabilities*. 01 2010.

[3] C. Byrne. *Applied Iterative Methods*. 01 2008.

[4] P. Causon and P. Mingham. *Introductory Finite Difference Methods for PDEs*. Bookboon.

[5] J. Duan, S. E. Li, Z. Liu, M. Bujarbaruah, and B. Cheng. Generalized policy iteration for optimal control in continuous time, 2019.

[6] W. Fleming and H. Soner. *Controlled Markov Processes and Viscosity Solutions*. Applications of mathematics. Springer-Verlag, 1993.

[7] H. U. Gerber. *An introduction to mathematical risk theory*. S.S.Huebner Foundation for Insurance Education, Philadelphia, 1979.

[8] P. Grandits. *Stochastische Kontrolltheorie für FVM - Vorlesungsmitschrift TU Wien*. 2018.

[9] P. Grandits. On the gain of collaboration in a two dimensional ruin problem. *European Actuarial Journal*, 9, 02 2019.

[10] P. Grandits. A ruin problem for a two-dimensional brownian motion with controllable drift in the positive quadrant. *Theory of Probability & Its Applications*, 64(4):646–655, 2020.

[11] A. Greenbaum. *Iterative methods for solving linear systems*. SIAM, 1997.

[12] X. Guo and G. Brunswick. Inside information and stock fluctuations. 10 1999.

[13] F. Hubalek. *Risiko und Ruintheorie - Vorlesungsmitschrift TU Wien*. 2018.

[14] F. Hubalek and W. Scherrer. *Einführung in die Stochastischen Prozesse und Zeitreihen - Vorlesungsmitschrift TU Wien*. 2017.

[15] R. Kaas, M. Goovaerts, J. Dhaene, and M. Denuit. *Modern Actuarial Risk Theory: Using R, Second Edition*. Springer-Verlag Berlin Heidelberg, 2009.

[16] H. Kushner and P. Dupuis. *Numerical Methods for Stochastic Control Problems in Continuous Time*. Applications of Mathematics, Stochastic Modelling and Applied Probability. Springer, 2012.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

[17] R. J. LeVeque. *Finite Difference Methods for Ordinary and Partial Differential Equations*. Society for Industrial and Applied Mathematics, 2007.

[18] H. P. McKean and L. A. Shepp. The advantage of capitalism vs. socialism depends on the criterion. *Journal of Mathematical Sciences*, 328:160–168, 2005.

[19] J. H. P. MCKEAN. Appendix : a free boundary problem for the heat equation arising from a problem of mathermatical economics. *Industrial Management Review*, 6:32–39, 1965.

[20] J. R. Norris. *Markov Chains*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1997.

[21] H. Pham. *Continuous-Time Stochastic Control and Optimization with Financial Applications*. Springer-Verlag, 2009.

[22] M. L. Puterman and S. L. Brumelle. On the convergence of policy iteration in stationary dynamic programming. *Mathematics of Operations Research*, 4(1):60–69, 1979.

[23] J. Sass. *Stochastic Control: With Applications to Financial Mathematics*. Unterlagen zur Vorlesung an der JKU Linz. 2006.

[24] U. Schmock. *Stochastische Analysis für FVM 1 - Vorlesungsskript TU Wien*. 2020.

[25] M. Shashkov and S. Steinberg. *Conservative Finite-Difference Methods on General Grids*. 01 1996.

[26] R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*. The MIT Press, second edition, 2018.

[27] W. Tang and L.-C. Tsai. Optimal surviving strategy for drifted Brownian motions with absorption. *The Annals of Probability*, 46(3):1597 – 1650, 2018.

[28] J. Von Neumann and O. Morgenstern. *"Theory of Games and Economic Behavior (60th Anniversary Commemorative Edition)*. 1944.

[29] B. K. Øksendal. *Stochastic Differential Equations. An Introduction with Applications*. Springer-Verlag, Heidelberg New York, 2000.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN