Elasticity as a Means of Evaluating Bonus-Malus Systems in Automobile Insurance

Maria Virri

Master's Thesis

Spring 2021

Faculty of Science

Department of Mathematics and Statistics

University of Helsinki

HELSINGIN YLIOPISTO — HELSINGFORS UNIVERSITET — UNIVERSITY OF HELSINKI

Tiedekunta/Osasto — Fakultet/Sektion — Faculty	Koulutusohjelma — Utbildningsprogram — Degree programme
Faculty of Science	Master's Programme in Mathematics and Statistics

Tekijä — Författare — Author

Maria Virri

Työn nimi — Arbetets titel — Title

Elasticity as a Means of Evaluating Bonus-Malus Systems in Automobile Insurance

Oppiaine — Läroämne — Subject

Mathematics

Työn laji — Arbetets art — Level	Aika — Datum — Month and year	Sivumäärä — Sidoantal — Number of pages
Master's Thesis	March 2021	47

Tiivistelmä — Referat — Abstract

Bonus-malus systems are used globally to determine insurance premiums of motor liability policy-holders by observing past accident behavior. In these systems, policy-holders move between classes that represent different premiums. The number of accidents is used as an indicator of driving skills or risk. The aim of bonus-malus systems is to assign premiums that correspond to risks by increasing premiums of policy-holders that have reported accidents and awarding discounts to those who have not.

Many types of bonus-malus systems are used and there is no consensus about what the optimal system looks like. Different tools can be utilized to measure the optimality, which is defined differently according to each tool. The purpose of this thesis is to examine one of these tools, elasticity. Elasticity aims to evaluate how well a given bonus-malus system achieves its goal of assigning premiums fairly according to the policy-holders' risks by measuring the response of the premiums to changes in the number of accidents.

Bonus-malus systems can be mathematically modeled using stochastic processes called Markov chains, and accident behavior can be modeled using Poisson distributions. These two concepts of probability theory and their properties are introduced and applied to bonus-malus systems in the beginning of this thesis. Two types of elasticities are then discussed. Asymptotic elasticity is defined using Markov chain properties, while transient elasticity is based on a concept called the discounted expectation of payments. It is shown how elasticity can be interpreted as a measure of optimality.

We will observe that it is typically impossible to have an optimal bonus-malus system for all policy-holders when optimality is measured using elasticity. Some policy-holders will inevitably subsidize other policy-holders by paying premiums that are unfairly large. More specifically, it will be shown that, for bonus-malus systems with certain elasticity values, lower-risk policy-holders will subsidize the higher-risk ones. Lastly, a method is devised to calculate the elasticity of a given bonus-malus system using programming language R. This method is then used to find the elasticities of five Finnish bonus-malus systems in order to evaluate and compare them.

Avainsanat — Nyckelord — Keywords

Bonus-malus systems, automobile insurance, Markov chains, elasticity

Säilytyspaikka — Förvaringsställe — Where deposited

Kumpula Campus Library

Muita tietoja — Övriga uppgifter — Additional information

Notations

K_n	number of claims in year n	p. 14
λ	claim frequency	p. 14
s	number of bonus classes	p. 15
b_i	insurance premium of class i	p. 15
T_k	transition rules	p. 15
$p_k(\lambda)$	probability of k accidents occurring	p. 17
C_n	bonus class of year n	p. 18
$M(\lambda)$	transition matrix of a bonus-malus system	p. 18
$\mu(\lambda)$	stationary distribution of a bonus-malus system	p. 19
$P(\lambda)$	mean stationary premium	p. 20
$\eta(\lambda)$	asymptotic elasticity	p. 24
λ_0	central value	p. 27
B_j	premium paid during year j	p. 30
β	discount factor	p. 30
$\nu_i(\lambda)$	discounted expectation of payments of class i	p. 30
$\gamma_i(\lambda)$	transient elasticity of class i	p. 34

Contents

1 Introduction

A bonus-malus system is a merit rating system used globally to determine insurance premiums of motor liability policy-holders by observing past claims behavior. It penalizes policy-holders that have submitted claims by increasing their premiums and awards discounts to claimless policy-holders. Due to the nature of liability insurance, claims are submitted by the party responsible for the accident, making bonus-malus systems a means to increase premiums of drivers that are responsible for accidents. In practice, policy-holders in such systems move between classes that correspond to different premiums. It is worth mentioning that this type of system may incentivize policy-holders to pay small claims themselves instead of reporting them. This phenomenon is called bonus hunger.

Classification variables (such as age, location of residence or car model) are commonly used to determine initial insurance premiums, but it has been shown that "the best predictor of the future number of claims is not the driver's age, sex, or occupation, but his past claims behavior" [10]. This is because drivers have varying driving skills which affect the probability of future accidents more than other variables. Driving skills may be hard for an insurance company to measure but claims history is a simple indicator of these abilities.

Bonus-malus systems are therefore implemented so that companies can take into account the individual risk that each driver poses based on their personal driving skills. The main goal is for each policy-holder to pay a premium corresponding to their risk. There is, however, no consensus about what an ideal bonus-malus system looks like. Each country uses a different type of bonus-malus system, often determined by legislation, and each company within a country may implement a distinct system as well. It is therefore worthwhile to examine tools that can be used to evaluate the optimality of a given bonus-malus system. These tools measure optimality in different ways, but their common objective is to evaluate how well a given system achieves its goal of assigning premiums fairly according to the drivers' risks.

The purpose of this thesis is to examine one of these tools, elasticity. Elasticity measures the response of the insurance premiums to changes in the number of claims. Two types of elasticities will be discussed, along with their properties. It will be shown that, for bonus-malus systems with certain elasticity values, drivers with lower claim frequencies will inevitably subsidize those with higher claim frequencies. Lastly, the bonus-malus systems of five Finnish insurance companies will be examined and compared using elasticity.

Due to their memoryless nature, bonus-malus systems can be modeled using Markov chains. Section 2 goes over the theory needed to understand Markov chains and their properties, which are then be applied to bonus-malus systems in Section 3. It will also be shown in Section 2 why the Poisson model is often used to model claim behavior.

Two types of elasticity are introduced in Section 4: asymptotic elasticity and transient elasticity. The asymptotic concept of elasticity was introduced in 1972 as "efficiency" by Kari Loimaranta and measures elasticity using an asymptotic property of Markov chains called the stationary distribution. In 1976, Jean Lemaire presented an alternative measure, a transient concept of elasticity, which depends on the policy-holder's starting class. This section also introduces the central value and its implications in relation to elasticity.

In Section 5, a method is devised to calculate elasticity for a given bonus-malus system in R, which is then used to find and compare the elasticities of five Finnish bonus-malus systems.

Jean Lemaire's *Bonus-Malus Systems in Automobile Insurance* is used as the main reference in this thesis. To avoid repetition, the words *policy-holder*, *driver* and *insured* are used interchangeably. The same applies to the words *accident* and *claim*, which implies the assumption that bonus hunger does not occur.

2 Background

2.1 Probability Theory

In order to discuss Markov chains in the following section, some basic concepts of measure theory and probability theory are introduced in this section. It is assumed the reader is familiar with the basics of set notation.

Definition 2.1. Suppose Ω is a set, and $\mathcal{F}$ is a collection of subsets of Ω. The collection $\mathcal{F}$ is called *σ-algebra* if the following conditions are satisfied:

1. $\emptyset \in \mathcal{F}$.
2. If $A \in \mathcal{F}$, then $A^c = \Omega \backslash A \in \mathcal{F}$.
3. If $A_1, A_2, \ldots$ is a sequence of subsets of Ω such that $A_i \in \mathcal{F}$ for all i, then $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$.

A set equipped with a σ-algebra is called a *measurable space.*

A probability measure is a special type of measure such that the measure of the entire set Ω is equal to 1.

Definition 2.2. Given a measurable space $(\Omega, \mathcal{F})$, a function $\mathbb{P} : \mathcal{F} \to [0, 1]$ is called a *probability measure* if the following conditions are satisfied:

1. $\mathbb{P}(\emptyset) = 0$.
2. If $A_1, A_2, \ldots$ is a sequence of sets such that $A_i \cap A_j = \emptyset$ for $i \neq j$ and $A_i \in \mathcal{F}$ for all i, then $\mathbb{P}\left(\cup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}\left(A_i\right)$.
3. $\mathbb{P}(\Omega) = 1$.

We can now define probability spaces using these two concepts.

Definition 2.3. A *probability space* is a triple $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω is a set, $\mathcal{F}$ is a σ-algebra on Ω, and $\mathbb{P}$ is a probability measure on $\mathbb{P}$.

We introduce one additional concept before defining random variables.

Definition 2.4. A *measurable* map is a map $X : \Omega_1 \to \Omega_2$ between two measurable spaces $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ such that for every $A \in \mathcal{F}_2$, $X^{-1}(A) = \{\omega \in \Omega_1 \mid X(\omega) \in A\} \in \mathcal{F}_1$.

Definition 2.5. Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a *random variable* is a measurable map $X : \Omega \to \Omega'$, where $(\Omega', \mathcal{F}')$ is a measurable space.

Lastly, the following definition allows us to discuss the probabilities of events.

Definition 2.6. Suppose X is a random variable with values in Ω'. The measure on $\mathcal{F}'$ given by $P_X = \mathbb{P}(X^{-1}(A)) = \mathbb{P}\{\omega \in \Omega \mid X(\omega) \in A\}$ is called the *distribution* of the random variable X.

Remark 2.1. It is customary to write $\mathbb{P}(X^{-1}(A))$ as $\mathbb{P}(X \in A)$.

2.2 Markov Chains

Markov chains provide a means to examine processes in which the distribution of the chain in the future depends only on its present state and not on any of its previous states. A simple example of this is a game of snakes and ladders. In this classic board game, the player moves through squares on the board from start to finish. From their current square, the player moves forward according to the number thrown on the die. Landing on a square with a ladder or a snake moves the player to a different square. The position of a player after each round is therefore dependent only on their current position and the number they roll on the die, not on any previous positions.

This section is based on Izyurov [6, Chapter 3] with some changes in notation.

Definition 2.7. A *stochastic process* in discrete time is a sequence of random variables $\{X_n |\ n \in \{0, 1, 2, \dots\}\}$ defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$, taking values in a measurable state space S such that $\mathbb{P}(X_n \in S) = 1$.

A Markov chain is a type of stochastic process. While stochastic processes and Markov chains may be continuous, only discrete cases will be discussed. We will also focus only on Markov chains with finite state spaces.

Definition 2.8. A *Markov chain* is a stochastic process on set X_n where

$$\mathbb{P}(X_{n+1} = x_{n+1} |\ X_n = x_n, \dots, X_0 = x_0) = \mathbb{P}(X_{n+1} = x_{n+1} |\ X_n = x_n)$$

for every $x_0, \dots, x_{n+1} \in S$ such that $\mathbb{P}(X_n = x_n, \dots, X_0 = x_0) > 0$. This equality is called the Markov property.

This property shows that Markov chains are "memoryless" in the sense that the history of the chain (states at $0, \dots, n$) does not affect its future (state at $n+1$).

Definition 2.9. The *transition matrix* $M^{(n)}$ of a Markov chain at time $n \in \{1, 2, \dots\}$ is defined by the *transition probabilities*

$$M_{xy}^{(n)} = \mathbb{P}(X_n = y \mid X_{n-1} = x)$$

for $x, y \in S$ such that $M^{(n)} = \left(M_{xy}^{(n)}\right)$.

Let the state space be $S = \{x_1, x_2, \dots\}$ and let $\mu^{(n)} = \begin{bmatrix} \mu_{x_1}^{(n)} & \mu_{x_2}^{(n)} & \dots \end{bmatrix}$ denote the distribution of X_n such that

$$\mu_{x_i}^{(n)} = \mathbb{P}(X_n = x_i).$$

where $n \in \{0, 1, 2, \dots\}$. This is the probability that the chain is at state x_i at time n. As shown in the following lemma, the transition matrix, together with the initial distribution $\mu^{(0)}$, can be used to determine the distribution of X_n.

Lemma 2.1. *We have*

$$\mu_y^{(n)} = \sum_{x \in S} \mu_x^{(n-1)} M_{xy}^{(n)}$$

for all $n \geq 1$ and $y \in S$. In matrix form this is

$$\mu^{(n)} = \mu^{(n-1)} M^{(n)}. \tag{2.1}$$

Proof. We assume that $\mathbb{P}(X_{n-1} = x) > 0$ for all $x \in S$. Using the definition of conditional probability we have

$$\begin{aligned} \mu_y^{(n)} = \mathbb{P}(X_n = y) = \sum_{x \in S} \mathbb{P}\left(X_n = y \cap X_{n-1} = x\right) = \sum_{x \in S} \mathbb{P}\left(X_{n-1} = x\right) \mathbb{P}\left(X_n = y \mid X_{n-1} = x\right) \\ = \sum_{x \in S} \mu_x^{(n-1)} M_{xy}^{(n)}. \end{aligned}$$

□

The Markov chains discussed in this thesis have transition probabilities that are not dependent on time.

Definition 2.10. A Markov chain is called *homogeneous* if the transition matrix $M^{(n)}$ does not depend on n.

We will denote homogeneous transition matrices by M and later by $M(\lambda)$ when the matrix is dependent on a parameter λ. All Markov chains assumed to be homogeneous henceforth. With this assumption, Equation (2.1) will naturally be

$$\mu^{(n)} = \mu^{(n-1)} M. \tag{2.2}$$

We will also assume that all Markov chains in the following sections are irreducible.

Definition 2.11. A Markov chain is called *irreducible* if, for every $x, y \in S$, there exists a $t \in \{0, 1, 2, \dots\}$ such that $\mathbb{P}\left(X_{n+t} = y \mid X_n = x\right) > 0$ where $n \in \{0, 1, 2, \dots\}$.

In other words, a Markov chain is irreducible if and only if it has positive probability to pass through any state.

Example 2.1. Suppose there are two insurance companies, Company X and Company Y. A driver becomes a client at either company and after each year has the option of switching to the other company. It has been noted that an insured at Company X has a 90% probability of staying at Company X and a 10% chance of switching to Company Y on any given year. Likewise, an insured at Company Y has a 70% probability of staying at Company Y and a 30% chance of switching to Company X.

This process can be presented as an irreducible, homogeneous Markov chain with two states x and y, which represent the two companies X and Y respectively, so that the state space is $S = \{x, y\}$. Figure 2.1 shows a state transition diagram of the Markov chain. The transition matrix of the chain is defined by the equations

$$\begin{aligned}
M_{xx} &= \mathbb{P}(X_n = x \mid X_{n-1} = x) = 0.9, \\
M_{xy} &= \mathbb{P}(X_n = y \mid X_{n-1} = x) = 0.1, \\
M_{yx} &= \mathbb{P}(X_n = x \mid X_{n-1} = y) = 0.3, \\
M_{yy} &= \mathbb{P}(X_n = y \mid X_{n-1} = y) = 0.7
\end{aligned}$$

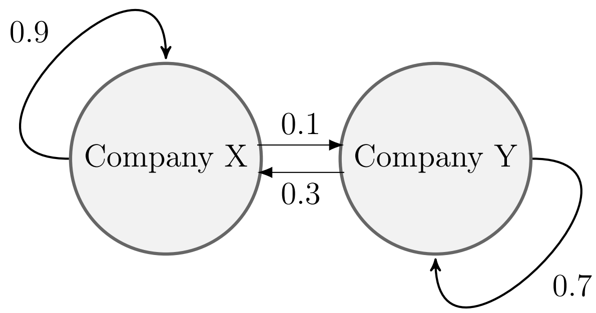

Figure 2.1: Transition diagram of the situation described in Example 2.1. The transition probabilities from Company X to Company Y and from Y to X are 0.1 and 0.3 respectively, while the probabilities of staying at Company X or Company Y are 0.9 and 0.7 respectively.

and hence

$$M = \begin{bmatrix} 0.9 & 0.1 \\ 0.3 & 0.7 \end{bmatrix}.$$

Suppose that the driver is initially at Company X. Then $\mu_x^{(0)} = \mathbb{P}(X_0 = x) = 1$ and $\mu_x^{(0)} = \mathbb{P}(X_0 = y) = 0$ and consequently $\mu^{(0)} = \begin{bmatrix} \mu_x^{(0)} & \mu_y^{(0)} \end{bmatrix} = \begin{bmatrix} 1 & 0 \end{bmatrix}$. We can then use Equation (2.2) to calculate further probabilities. To see which company the driver will be at after the first year we calculate

$$\mu^{(1)} = \mu^{(0)} M = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 0.9 & 0.1 \\ 0.3 & 0.7 \end{bmatrix} = \begin{bmatrix} 0.9 & 0.1 \end{bmatrix}.$$

This tells us that is a 90% chance that the driver will still be at Company X and a 10% chance that they will instead be at Company Y, as can be expected from a client of

Company X. The odds after the second, fifth, 10^{th} and 15^{th} year are

$$\mu^{(2)} = \mu^{(1)}M = \begin{bmatrix} 0.84 & 0.16 \end{bmatrix},$$
$$\mu^{(5)} = \mu^{(4)}M = \mu^{(3)}M^2 = \cdots = \mu^{(0)}M^5 \approx \begin{bmatrix} 0.769 & 0.231 \end{bmatrix},$$
$$\mu^{(10)} = \mu^{(0)}M^{10} \approx \begin{bmatrix} 0.752 & 0.248 \end{bmatrix}$$

and

$$\mu^{(15)} = \mu^{(0)}M^{15} \approx \begin{bmatrix} 0.750 & 0.250 \end{bmatrix},$$

respectively. The probabilities seem to approach the values 0.75 and 0.25. This phenomenon is explained in Section 2.2.1.

2.2.1 Stationary Distributions

Stationary distributions are a central concept when looking into the asymptotic behavior of Markov chains. They describe the behavior of the chain in the long run. If the limit $\mu = \lim_{n\to\infty} \mu^{(n)}$ exists, we can take the limit of both sides of Equation (2.2) so that we get

$$\lim_{n\to\infty} \mu^{(n)} = \lim_{n\to\infty} \mu^{(n-1)}M$$

and therefore

$$\mu = \mu M.$$

Definition 2.12. A probability measure μ that satisfies the equation $\mu = \mu M$ is called a *stationary distribution* when its components satisfy the equation

$$\sum_{x\in S} \mu_x = 1.$$

The components μ_x must also be non-negative for all $x \in S$ because they are probabilities. As Izyurov states, we can think of μ_x as the "proportion of time that the chain, eventually, spends at the state x" [6].

If we assume that a homogeneous Markov chain has a finite state space and is irreducible, its stationary distribution μ exists and is unique [6, Theorem 3.3.3.]. The stationary distribution is then independent of the initial state of the chain. The following example demonstrates how stationary distributions can be found.

Example 2.2 (Continuation of Example 2.1)**.** Suppose the situation is the same as in Example 2.1. We know that the stationary distribution exists because the chain is irreducible.

Method 1: In order to find the stationary distribution we can solve the equation

$$\mu = \mu M$$
$$\begin{bmatrix} \mu_x & \mu_y \end{bmatrix} = \begin{bmatrix} \mu_x & \mu_y \end{bmatrix} \begin{bmatrix} 0.9 & 0.1 \\ 0.3 & 0.7 \end{bmatrix}$$

with the constraint $\mu_x + \mu_y = 1$. We get a system of equations

$$\begin{cases} \mu_x = 0.9\mu_x + 0.3\mu_y \\ \mu_y = 0.3\mu_x + 0.7\mu_y \\ \mu_x + \mu_y = 1 \end{cases}$$

and the stationary distribution μ is the solution of this system $\begin{bmatrix} 0.75 & 0.25 \end{bmatrix}$, the same value that the chain seems to converge to in Example 2.1. This result shows that in the long run (and regardless of which company the driver initially chooses), the driver is expected to choose Company X three quarters of the time Company Y a quarter of the time. In other words, the driver will spend three quarters of his driving career as a client of Company X and the rest as a client of Company Y.

This calculation is simple enough, but for Markov chains with more than two states we get large systems of equations that can be difficult to solve. In these cases it may be better to use a numerical method based on convergence.

Method 2: An easier (but not an exact) way to find the stationary distribution is by raising the transition matrix to a high power. Eventually, the rows of the matrix will become equal if the stationary distribution exists. In the previous example we calculated

$$M^{15} \approx \begin{bmatrix} 0.750 & 0.250 \\ 0.750 & 0.250 \end{bmatrix}$$

to get

$$\begin{bmatrix}1 & 0\end{bmatrix} M^{15} \approx \begin{bmatrix}0.750 & 0.250\end{bmatrix}$$

for a driver who is initially at Company X. Clearly also

$$\begin{bmatrix}0 & 1\end{bmatrix} M^{15} \approx \begin{bmatrix}0.750 & 0.250\end{bmatrix},$$

so the result is the same for a driver who is initially at Company Y. This is expected, because the stationary distribution should not depend on the initial state. The stationary distribution vector can therefore be obtained by raising the transition matrix to a high power and choosing any row. Using this method, however, requires an additional assumption.

The Markov chain of Example 2.2 as well as the Markov chains in the following sections have a property called aperiodicity.

Definition 2.13. Let $\mathcal{T}(x) = \{n > 0 : \mathbb{P}(X_n = x) > 0\}$ represent the set of possible times that a chain starting from state x returns to state x. We call a state x *aperiodic* if the greatest common denominator of the values in $\mathcal{T}(x)$ is 1. A chain is called aperiodic if all its states are aperiodic.

When a Markov chain is aperiodic, irreducible and has a finite state space (also called an ergodic Markov chain),

$$\left\|\mu^{(n)} - \mu\right\| = \left\|\mu^{(0)} M^n - \mu\right\| \to 0 \text{ as } n \to \infty$$

for any initial state $\mu^{(0)}$, where the norm is the l_1 norm defined as $\|v\| = \sum_{x \in S} |v_x|$ [6, Theorem 3.4.3.]. Intuitively this means that the chain will converge to the stationary distribution μ when the transition matrix M is raised to a high power, and we can apply Method 2 of Example 2.2. This also confirms that the result is independent of the initial state.

Markov chains (with finite state spaces that satisfy the previously mentioned assumptions of aperiodicity and irreducibility) converge exponentially fast [14, Fact 3], which is why a power of 15 is sufficient to reach the stationary distribution in Example 2.2. There are different ways to estimate how large the power n should be in order to reach the desired level of accuracy. An overview of these methods of bounding convergence rates can be found in Rosenthal [14].

2.3 The Poisson Model

As we will see in Section 3, Markov chains model bonus-malus systems, but we are yet to show what will determine the transition probabilities of the chain. In this section we justify why the Poisson distribution is often chosen to model claim number distributions. This distribution, along with the rules of the bonus-malus system, will then be used to determine the transition probabilities. This section is based on Lemaire [10, Chapter 3].

When we think about the behavior of drivers, it is natural to assume that the number of accidents in a certain time period does not affect the probability of accidents in another time period, i.e., the claim numbers in different time periods are independent. We may also assume that the probability of a single accident during a small time interval is proportional to the length of the interval. It is five times more likely that an accident will happen during a 5-minute drive compared to a 1-minute drive. Now consider the probability of two or more accidents happening within the same minute, or within the same second. As the time period shortens, two or more accidents happening becomes practically impossible. The last assumption is then that the probability of more than one accident is negligible when the time interval is sufficiently small.

We will denote the number of claims occurring in time interval (t, u) by $K(t, u)$ where t and u are positive numbers and $t-u$ is the length of the interval. From the independence assumption we have that

$$\mathbb{P}[K(t_1, u_1) = k_1 \text{ and } K(t_2, u_2) = k_2] = \mathbb{P}[K(t_1, u_1) = k_1]\mathbb{P}[K(t_2, u_2) = k_2]$$

when (t_1, u_1) and (t_2, u_2) are disjoint time intervals. Now we consider a time interval $(t, t+h)$ where h is the length of the interval. The second assumption implies that when h is a small, positive number,

$$\mathbb{P}[K(t, t+h) = 1] = \lambda h,$$

where $\lambda \geq 0$ is a constant. Note that this also implies that this probability is the same for all intervals of length h regardless of the starting time t. The last assumption is equivalent with

$$\mathbb{P}[K(t, t+h) > 1] = 0$$

when h is sufficiently small. We will now show that the Poisson distribution is the only distribution that satisfies these three conditions.

If we assume that h is sufficiently small and that the assumptions above hold,

$$\begin{aligned}\mathbb{P}[K(0,t+h)=k] =&\mathbb{P}[K(0,t)=k]\mathbb{P}[K(t,t+h)=0]\\ &+\mathbb{P}[K(0,t)=k-1]\mathbb{P}[K(t,t+h)=1]\\ &+\sum_{i=2}^{k}\mathbb{P}[K(0,t)=k-i]\mathbb{P}[K(t,t+h)=i]\end{aligned}$$

simplifies to

$$\mathbb{P}[K(0,t)=k](1-\lambda h)+\mathbb{P}[K(0,t)=k-1](\lambda h)+0.$$

Rearranging the terms gives

$$\frac{\mathbb{P}[K(0,t+h)=k]-\mathbb{P}[K(0,t)=k]}{h}=-\lambda\mathbb{P}[K(0,t)=k]+\lambda P[K(0,t)=k-1].$$

Now, if $h\to 0$, we have

$$\begin{cases}\dfrac{d\mathbb{P}[K(0,t)=k]}{dt}=-\lambda\mathbb{P}[K(0,t)=k]+\lambda\mathbb{P}[K(0,t)=k-1], & \text{when } k=1,2,\dots\\ \dfrac{d\mathbb{P}(K(0,t)=0]}{dt}=-\lambda P[K(0,t)=0], & \text{when } k=0.\end{cases}$$

Solving this set of equations yields

$$\mathbb{P}[K(0,t)=k]=e^{-\lambda t}\frac{(\lambda t)^k}{k!},\quad k=0,1,2,\dots. \tag{2.3}$$

For more detailed calculations see Cowan [1]. We will now show that, with the assumptions made, the number of claims must have a Poisson distribution.

Definition 2.14. A random variable K is said to have a *Poisson distribution* with parameter $\lambda\geq 0$ if

$$\mathbb{P}(K=k)=e^{-\lambda}\frac{\lambda^k}{k!},\quad k=0,1,2,\dots.$$

When $\lambda=0$ we set $\mathbb{P}(K=0)=1$.

In Section 3, we will see that insurance companies measure the number of claims typically after one policy period. We may therefore assume discrete, unit time periods such that $t=1$ holds always. As previously mentioned, the starting time of the period is not of importance, so the probability will be equal for all policy periods of the same length. We may also assume that each policy period is one year. If K_n denotes the number of accidents in year n, we can see from Equation (2.3) that

$$\mathbb{P}(K_n = k) = \mathbb{P}[K(n, n+1) = k] = \mathbb{P}[K(0, 1) = k] = e^{-\lambda}\frac{\lambda^k}{k!}$$

and K_n has a Poisson distribution for all $n = 1, 2, \ldots$. In conclusion, the yearly claim numbers are identically distributed and have a Poisson distribution.

The use of the parameter λ arises from the fact that drivers have different accident probabilities (and hence different transition probabilities) due to differences in driving skills. The parameter λ represents the yearly *claim frequency* such that a claim frequency $\lambda = 1$ means an average of 1 accident per year and $\lambda = 0.1$ means an average of 1 accident every 10 years. Drivers with higher claim frequencies are considered "worse drivers" and drivers with lower claim frequencies are considered "better drivers". It is worth mentioning that this model is based on the assumption that the parameter λ is constant for every policy period. In reality, an insured's driving skills may very well improve (or worsen) with time, which would result in different parameter values for different years.

While bonus-malus systems and their elasticities will be defined in a more general manner without the explicit use of the Poisson model, we may assume that the assumptions made in this section about the numbers of accidents hold. The Poisson distribution will be implemented in examples and calculations.

3 Bonus-Malus Systems

This section is based on Lemaire [10, Chapter 1] and Lehtomaa [7, Section 5.6], mainly following the notation of the former.

3.1 Definition

In a bonus-malus system, drivers are placed in one of a finite number of classes which determine the drivers' insurance premiums. Each insured's class is updated at the end of each insurance period (year) according to the number of claims made during that period. The idea is that drivers who submit one or more claims will move to classes with higher premiums and drivers with no claims will move to classes with lower premiums. Because there are a finite number of classes, it is possible that the driver stays in their current class.

Definition 3.1. A *bonus-malus system* is made up of the following elements:

1. Bonus classes $1, \dots, s$, where $s \in \mathbb{N}$.
2. Premium scale $(b_1, \dots, b_s)$, where b_i is the insurance premium of class i.
3. Transition rules T_k, $k = 0, 1, 2, \dots$. We denote $T_k(i) = j$ if the insured is transferred from class i to class j when k accidents occur.

We will assume that the lower bonus classes will correspond to lower premiums and the higher bonus classes correspond to higher premiums. Hence,

$$b_1 \leq b_2 \leq \dots \leq b_s.$$

The transition rules can be shown in matrix form $T_k = \left(t_{ij}^{(k)}\right)$, where $i, j \in \{1, \dots, s\}$ and

$$t_{ij}^{(k)} = \begin{cases} 1, & \text{if } T_k(i) = j \\ 0, & \text{otherwise.} \end{cases}$$

Intuitively, this matrix shows whether a transfer happens between two classes when a certain number of accidents happens during the insurance period.

Class i	Premium b_i	T_0	T_1	$T_k(k \geq 2)$
3	150%	2	3	3
2	100%	1	3	3
1	50%	1	2	3

Table 3.1: A bonus-malus system with three classes where each claim per year results in the driver moving up one class and each claimless year results in the driver moving down one class. An insured in class 3 with one or more accidents during the policy period will stay in class 3. Similarly, an insured in class 1 will stay in this class if no accidents occur.

Example 3.1. Suppose that we have a three-class bonus-malus system as shown in Table 3.1. The insured moves between classes 1 and 3 depending on how many accidents occur during the policy period. The transition rules are such that

$$\begin{array}{lll} T_0(1) = 1 & T_0(2) = 1 & T_0(3) = 2 \\ T_1(1) = 2 & T_1(2) = 3 & T_1(3) = 3 \\ T_k(1) = 3 & T_k(2) = 3 & T_k(3) = 3 \end{array}$$

where $k = 2, 3, \ldots$. In matrix form, the rules are

$$T_0 = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \; T_1 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \text{ and } T_k = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \text{ for } k \geq 2.$$

For simplicity, we can choose the premium of class 2 to be $b_2 = 1$ and hence $b_1 = 0.5$ and $b_3 = 1.5$.

3.2 Bonus-Malus Systems as Markov Chains

A policy-holder's bonus class developing over time can be presented as a homogeneous Markov chain. This is because the class of the following year is determined only by the policy-holder's current class, the transition rules and the number of accidents in the policy period. The current class is the current state of the Markov chain and the transition

Class	T_0	T_1	T_2	$T_k(k \geq 3)$
4	3.1	4	4	4
3.0	2	4	4	4
3.1	1	4	4	4
2	1	3.0	4	4
1	1	2	3.0	4

Table 3.2: A non-Markovian bonus-malus system with four classes transformed into a Markovian system. The transition rules are such that each claim per year results in the driver moving up one class while each claimless year results in the driver moving down one class. Additionally, no driver can be above class 1 after two consecutive claim free years. This rule is acknowledged in the table by adding an index to class 3 that counts the number of consecutive claim free years.

rules and accident number distributions determine the transition probabilities. Markov chains formed by bonus-malus systems are irreducible and aperiodic, so the assumptions of Section 2.2 hold.

Remark 3.1. Some bonus-malus systems are not Markovian; they take into account claims of multiple previous years. For example, a system might rule that a policy-holder cannot be in a high class after a certain number of claimless years even though the driver would otherwise be placed there. These systems can, however, be modified and become Markovian. Table 3.2 shows an example of a bonus-malus system with four classes where drivers are always placed in the lowest class after two consecutive claimless years regardless of their current class. The four-class system essentially becomes a five-class system. While the addition of only one class was necessary in this case, more sophisticated systems can grow quite large as seen in Lemaire [10, pp. 9-10].

This section examines the concepts of Section 2.2 in the context of bonus-malus systems, with the addition of the parameter λ. The claim frequency λ represents the driving skills of an insured and allows us to find the transition matrix depending on these skills. We will denote the probability of k accidents happening to a driver with claim frequency λ in any given year by

$$p_k(\lambda) = \mathbb{P}(K_n = k), \quad k = 0, 1, \ldots.$$

We can use a notation independent of year n because, as explained in Section 2.3, the yearly numbers of accidents are assumed to be identically distributed. We also denote the bonus class of year n by C_n. The initial bonus class is $C_1 = i_1$. The next result proves that the process C_n is a Markov chain. The proof follows Lehtomaa [7, Theorem 5.6.].

Theorem 3.1. *The stochastic process* $\{C_n \mid n = 1, 2, \ldots\}$ *is a homogeneous Markov chain. The transition matrix* $M(\lambda)$ *of an insured with claim frequency* $\lambda \geq 0$ *is determined by the transition probabilities*

$$M_{ij}(\lambda) = \sum_{k=0}^{\infty} p_k(\lambda) t_{ij}^{(k)}, \quad i, j \in \{1, \ldots, s\}.$$

In matrix form this is

$$M(\lambda) = \sum_{k=0}^{\infty} p_k(\lambda) T_k. \tag{3.1}$$

Proof. We will show that the Markov property holds for this process. Let $n \in \{0, 1, \ldots\}$ and $i, j \in \{1, \ldots, s\}$. First we note that the number of claims during a year n is naturally independent of the insured's bonus class during that year or any previous years, so

$$\mathbb{P}\left(K_n = k \mid C_n = i\right) = \mathbb{P}\left(K_n = k\right) \tag{3.2}$$

and

$$\mathbb{P}\left(K_n = k \mid C_n = i_n, \ldots, C_1 = i_1\right) = \mathbb{P}\left(K_n = k\right). \tag{3.3}$$

Also, we will be able to focus on the values of k for which $T_k(i) = j$ because

$$\mathbb{P}\left(C_{n+1} = j, K_n = k \mid C_n = i\right) = \begin{cases} \mathbb{P}\left(K_n = k \mid C_n = i\right), & \text{if } T_k(i) = j \\ 0, & \text{otherwise.} \end{cases} \tag{3.4}$$

Now

$$
\begin{aligned}
& \mathbb{P}\left(C_{n+1}=j \mid C_n=i\right) \\
& =\mathbb{P}\left(C_{n+1}=j, K_n=0 \mid C_n=i\right) \cup \mathbb{P}\left(C_{n+1}=j, K_n=1 \mid C_n=i\right) \cup \ldots \\
& =\sum_{k=0}^{\infty} \mathbb{P}\left(C_{n+1}=j, K_n=k \mid C_n=i\right) \\
\text{(from (3.4))} \quad & =\sum_{k: T_k(i)=j}^{\infty} \mathbb{P}\left(K_n=k \mid C_n=i\right) \\
\text{(from (3.2))} \quad & =\sum_{k: T_k(i)=j}^{\infty} \mathbb{P}\left(K_n=k\right) \\
& =\sum_{k: T_k(i)=j}^{\infty} p_k(\lambda)=\sum_{k=0}^{\infty} p_k(\lambda) t_{ij}^{(k)}
\end{aligned}
$$

From (3.3) and similar calculations and we get

$$
\begin{aligned}
& \mathbb{P}\left(C_{n+1}=j \mid C_n=i_n, \ldots, C_1=i_1\right) \\
& =\sum_{k: T_k(i)=j}^{\infty} \mathbb{P}\left(K_n=k \mid C_n=i_n, \ldots, C_1=i_1\right) \\
& =\sum_{k: T_k(i)=j}^{\infty} p_k(\lambda)=\sum_{k=0}^{\infty} p_k(\lambda) t_{ij}^{(k)}
\end{aligned}
$$

We have now shown that $\mathbb{P}\left(C_{n+1}=j \mid C_n=i_n, \ldots, C_1=i_1\right)=\mathbb{P}\left(C_{n+1}=j \mid C_n=i\right)$ and the process is a Markov chain. □

The Markov chain is homogeneous because we assume that neither the bonus rules nor the accident probabilities depend on the year and hence the transition matrix is the same for all n.

The process also has stationary distribution as defined in Section 2.2.1. We will denote the stationary distribution by

$$
\mu(\lambda)=\begin{bmatrix}\mu_1(\lambda) & \ldots & \mu_s(\lambda)\end{bmatrix}.
$$

As shown in Section 2.2.1, the stationary distribution must satisfy the equations

$$\mu(\lambda) = \mu(\lambda)M(\lambda) \tag{3.5}$$

and

$$\sum_{i=1}^{s} \mu_i(\lambda) = 1. \tag{3.6}$$

Because $\mu_i(\lambda)$ can be thought of as the portion of time that the insured eventually spends in class i, we multiply it with the premium of the class to get the amount that the policyholder will pay for the class in the long run. The sum of these amounts for all classes $1, \dots, s$ is called the mean stationary premium, which represents the average premium that a driver with claim frequency λ will pay yearly in the long run.

Definition 3.2. The yearly *mean stationary premium* $P(\lambda)$ (also called the *long-term average yearly premium*) of a bonus-malus system for an insured with a claim frequency λ is defined as

$$P(\lambda) = \sum_{i=1}^{s} \mu_i(\lambda) b_i,$$

where s is the number of classes.

We will now calculate the transition matrix, stationary distribution and mean stationary premium of the bonus-malus system in the previous example.

Example 3.2 (Continuation of Example 3.1)**.** Suppose the number of claims follows a Poisson distribution with parameter λ. The claim probabilities are hence

$$\begin{aligned} p_0(\lambda) = \mathbb{P}(K = 0) =& e^{-\lambda}, \\ p_1(\lambda) = \mathbb{P}(K = 1) =& \lambda e^{-\lambda} \end{aligned}$$

and

$$p_{\geq 2}(\lambda) = p_2(\lambda) + p_3(\lambda) + \ldots = \mathbb{P}(K \geqslant 2) = 1 - \mathbb{P}(k \leq 1) = 1 - \lambda e^{-\lambda} - e^{-\lambda}.$$

Now since $T_2 = T_3 = \dots$, the transition matrix is

$$
\begin{aligned}
M(\lambda) &= \sum_{k=0}^{\infty} p_k(\lambda) T_k \\
&= p_0(\lambda)T_0 + p_1(\lambda)T_1 + p_2(\lambda)T_2 + p_3(\lambda)T_3 + \dots \\
&= p_0(\lambda)T_0 + p_1(\lambda)T_1 + (p_2(\lambda) + p_3(\lambda) + \dots)T_2 \\
&= e^{-\lambda} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} + \left(\lambda e^{-\lambda}\right) \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} + \left(1 - \lambda e^{-\lambda} - e^{-\lambda}\right) \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} e^{-\lambda} & \lambda e^{-\lambda} & 1 - \lambda e^{-\lambda} - e^{-\lambda} \\ e^{-\lambda} & 0 & 1 - e^{-\lambda} \\ 0 & e^{-\lambda} & 1 - e^{-\lambda} \end{bmatrix}.
\end{aligned}
$$

The stationary distribution is then acquired by solving $\mu(\lambda) = \mu(\lambda)M(\lambda)$ such that $\sum_{i=1}^{s} \mu_i(\lambda) = 1$. Calculations as in Example 2.2 show that the stationary distribution is

$$
\begin{cases}
\mu_1(\lambda) = & \dfrac{e^{-2\lambda}}{1 - \lambda e^{-2\lambda}} \\
\mu_2(\lambda) = & \dfrac{e^{-\lambda} - e^{-2\lambda}}{1 - \lambda e^{-2\lambda}} \\
\mu_3(\lambda) = & \dfrac{1 - e^{-\lambda} - \lambda e^{-2\lambda}}{1 - \lambda e^{-2\lambda}}.
\end{cases}
$$

The premium scale is defined as in Example 3.1. The mean stationary premium is then

$$
\begin{aligned}
P(\lambda) &= \sum_{i=1}^{3} \mu_i(\lambda) b_i \\
&= \frac{e^{-2\lambda}}{1 - \lambda e^{-2\lambda}} \cdot 0.5 + \frac{e^{-\lambda} - e^{-2\lambda}}{1 - \lambda e^{-2\lambda}} \cdot 1 + \frac{1 - e^{-\lambda} - \lambda e^{-2\lambda}}{1 - \lambda e^{-2\lambda}} \cdot 1.5 \\
&= \frac{-(0.5 + 1.5\lambda)e^{-2\lambda} - 0.5e^{-\lambda} + 1.5}{1 - \lambda e^{-2\lambda}}.
\end{aligned}
$$

The graph of $P(\lambda)$ can be seen in Figure 3.1. We can see from the graph that $P(\lambda)$ seems to be bounded between $0.5 = b_1$ and $1.5 = b_3$, the minimum and maximum premiums.

The mean stationary premium is, in fact, always bounded by the minumum and maximum premium values. To show this, we can investigate the asymptotic properties of

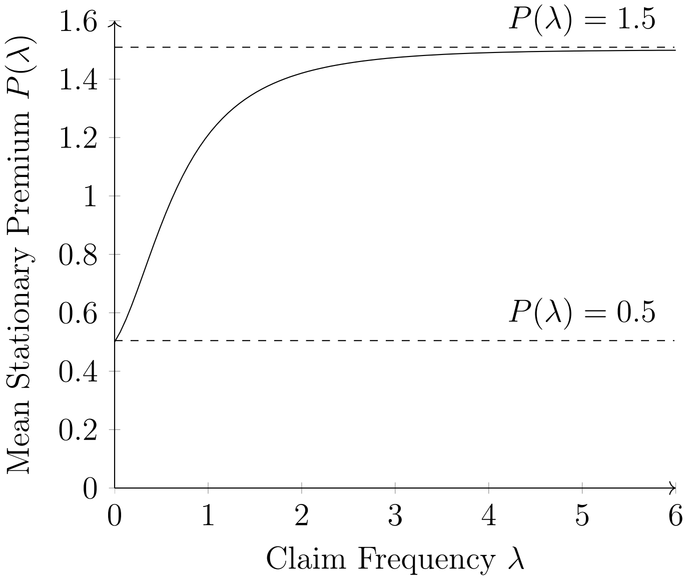

Figure 3.1: Graph of the mean stationary premium of the bonus-malus system in Example 3.2, showing that $P(\lambda)$ is an increasing function of λ bounded by the minimum and maximum premium values of the bonus-malus system. The function eventually converges to the maximum premium value, causing it to resemble a concave function. At values close to 0, $P(\lambda)$ is instead often convex. Although hardly noticeable in this graph, this becomes more apparent with more sophisticated bonus-malus systems, as can be seen in [10, Figure 6-1].

$P(\lambda)$. When $\lambda \to \infty$, $\mu_i(\lambda) \to 0$ for all values of i less than s and $\mu_s(\lambda) \to 1$. The mean stationary premium therefore converges to b_s. Likewise, $P(\lambda) \to b_1$ when $\lambda \to 0$ assuming that it is defined, i.e. that $P(0) > 0$. This is realistic because it is impossible for a policyholder to be paying yearly premiums greater or smaller than the maximum and minimum premiums of the bonus-malus system. A mean stationary premium proportional to λ may be considered ideal theoretically, but if the function $P(\lambda)$ were linear the number of bonus classes would be infinite and the premiums of drivers with high claim frequencies would be unreasonably large.

4 Elasticity of Bonus-Malus Systems

The main advantage of using a bonus-malus system in automobile insurance is that it increases premiums of high-risk “bad drivers” and decreases premiums of low-risk “good drivers”. *But how well do bonus-malus systems achieve the task of assigning premiums that correspond to risks, giving fair punishments to drivers who have accidents and rewarding claim free years reasonably?* This question can be answered by examining a quantity called elasticity.

Definition 4.1. The *elasticity* e of a positive, differentiable function $f(\lambda)$ at a positive point λ is defined as

$$e[f(\lambda)] = f'(\lambda)\frac{\lambda}{f(\lambda)} = \frac{d\log f(\lambda)}{d\log\lambda}.$$

The elasticity of a function is the ratio of a relative change in a function $f(\lambda)$ with respect to a relative change in its variable λ. As Lemaire describes, the elasticity of a bonus-malus system “measures the response of the system to a change in the claim frequency” [10]. More specifically, we want to measure how a change in the claim frequency affects the premium of a driver in a bonus-malus system. Claim frequency is denoted by λ, as before, and $f(\lambda)$ is some function that represents the insurance premium. This section presents two ways of measuring elasticity using two different presentations of the premium.

4.1 Asymptotic Elasticity

Asymptotic elasticity uses the mean stationary premium, introduced in Section 3.2, as the premium of a given bonus-malus system. Consequently, it uses the asymptotic property of ergodic Markov chains, the stationary distribution, which was introduced in Section 2.2.1. This type of elasticity was introduced in 1972 by Kari Loimaranta in his article [11, part 4], which is used as the source of the content in this section unless otherwise mentioned. Asymptotic elasticity measures the relative change in the claim frequency compared to a relative change in the mean stationary premium $P(\lambda)$, as defined in Definition 3.2. It is therefore the elasticity of the mean stationary premium.

Definition 4.2. *Asymptotic elasticity* $\eta(\lambda)$ is defined as

$$\eta(\lambda) = e[P(\lambda)] = P'(\lambda)\frac{\lambda}{P(\lambda)} = \frac{d\log P(\lambda)}{d\log\lambda}.$$

Calculation of $\eta(\lambda)$ requires the derivative of $P(\lambda)$, which is

$$P'(\lambda) = \sum_{i=1}^{s} \mu_i'(\lambda) b_i.$$

Taking the derivatives of Equations (3.5) and (3.6) yields a system of equations

$$\begin{cases} \mu'(\lambda) = \mu'(\lambda)M(\lambda) + \mu(\lambda)M'(\lambda) \\ \sum_{i=1}^{s} \mu_i'(\lambda) = 0, \end{cases}$$

which can be used to determine $d\mu_i(\lambda)/d\lambda$. In the case of Poisson distribution, we can differentiate (3.1) to get

$$M'(\lambda) = \sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!}\left(T_{k+1} - T_k\right)$$

for easier calculation. We will now calculate the asymptotic elasticity of the bonus-malus system in the previous example.

Example 4.1 (Continuation of Example 3.2)**.** We already calculated the mean stationary premium $P(\lambda)$ in the previous example, so we simply differentiate $P(\lambda)$ to get

$$P'(\lambda) = \frac{0.5e^{-3\lambda}\lambda + 0.5e^{-\lambda} + e^{-2\lambda} - 0.5e^{-4\lambda} - 0.5e^{-3\lambda}}{(1 - e^{-2\lambda}\lambda)^2}.$$

Now

$$\begin{aligned}\eta(\lambda) &= P'(\lambda)\frac{\lambda}{P(\lambda)} \\ &= \frac{0.5e^{-3\lambda}\lambda + 0.5e^{-\lambda} + e^{-2\lambda} - 0.5e^{-4\lambda} - 0.5e^{-3\lambda}}{(1 - e^{-2\lambda}\lambda)^2} \frac{\lambda(1 - \lambda e^{-2\lambda})}{-(0.5 + 1.5\lambda)e^{-2\lambda} - 0.5e^{-\lambda} + 1.5} \\ &= \frac{\lambda e^{-4\lambda}\left(0.5\lambda e^{\lambda} - 0.5e^{\lambda} + 0.5e^{3\lambda} + e^{2\lambda} - 0.5\right)}{1.5\lambda^2 e^{-4\lambda} + 0.5\lambda e^{-4\lambda} + 0.5\lambda e^{-3\lambda} - 3\lambda e^{-2\lambda} - 0.5e^{-2\lambda} - 0.5e^{-\lambda} + 1.5}.\end{aligned}$$

The graph of this function is shown in Figure 4.1.

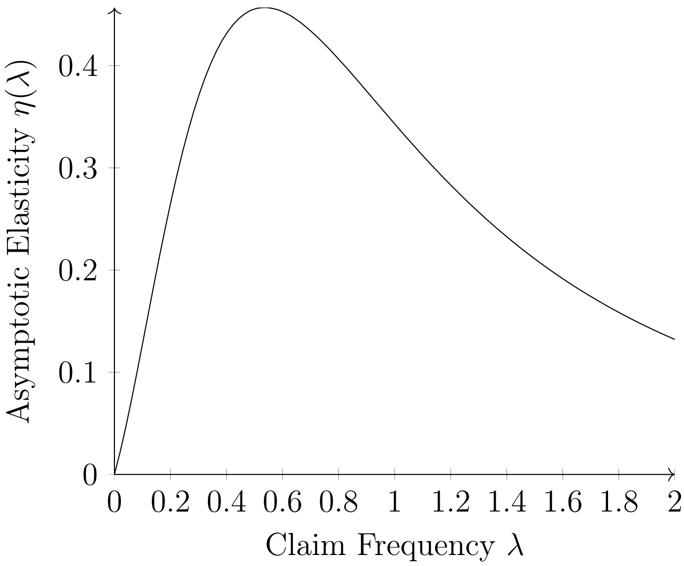

Figure 4.1: Asymptotic elasticity of the bonus-malus system in Example 4.1. The system only reaches elasticity values below 0.5 and is therefore quite inelastic.

The properties of the mean stationary premium determine the shape of the elasticity function. In Section 3.2 we saw that the mean stationary premium converges to b_s, a constant, when $\lambda \to \infty$. Consequently, the elasticity, its logarithmic derivative, tends to zero. Likewise, $P(\lambda) \to b_1$ when $\lambda \to 0$ and the elasticity goes to zero when the claim frequency is zero. For all other positive values of λ, the asymptotic elasticity is positive because $P(\lambda)$ is increasing. As expected, the elasticity function in Figure 4.1 corroborates these findings.

4.1.1 Perfect Elasticity and Optimality

We will now discuss how elasticity can be interpreted as a measure of how optimal a bonus-malus system is. Although perfect elasticity will be introduced in the context of asymptotic elasticity, it can be applied to any elasticity function.

Definition 4.3. A bonus-malus system is called *perfectly elastic* for the values of λ where $\eta(\lambda) = 1$.

From the definition of elasticity we have that

$$e[P(\lambda)] = P'(\lambda)\frac{\lambda}{P(\lambda)} = \lim_{x\to\lambda} \frac{P(x) - P(\lambda)}{x - \lambda}\frac{\lambda}{P(\lambda)} = \lim_{x\to\lambda} \frac{P(x) - P(\lambda)}{P(\lambda)}\frac{\lambda}{x - \lambda}.$$

Now we can see that perfect elasticity occurs when

$$\frac{P(x) - P(\lambda)}{P(\lambda)} = \frac{x - \lambda}{\lambda}.$$

This equation shows the percentage change of the claim frequency being equal to the percentage change of the mean stationary premium. This gives us a more intuitive interpretation of perfect elasticity. If the system is perfectly elastic at $\lambda = 0.10$, a driver with claim frequency 0.11 will pay 10% more premiums in the long run than a driver with claim frequency 0.10 [10]. This is because

$$\frac{0.11 - 0.10}{0.10} = 10\%.$$

Perfect elasticity is considered to be the optimal situation according to this measure because it implies that a relative change in claim frequency does not result in an unfairly large or small relative change in the premium and the system achieves its goal of assigning premiums that represent the risks: a 1% increase in the claim frequency causes a 1% increase in the premium, a 3% decrease in the claim frequency causes a 3% decrease in the premium and so on. The closer the system is to perfect elasticity, the better it achieves this goal.

Nevertheless, the mean stationary premium is a function bounded by b_1 and b_s and it is impossible to have perfect elasticity for all values of λ. This is evident from the previous discussion regarding the shape of the elasticity curve. Consequently, we cannot deem a system optimal for all claim frequencies. Then again, we are not very interested in the extreme cases of very large claim frequencies or claim frequencies equal to zero because they are highly unlikely. It is more important that elasticity close to 1 is reached for more likely values of λ such as the range 0.05 to 0.5 [10].

Elasticity measures the fairness of premiums with respect to the premiums of policyholders with different claim frequencies, but does not address fairness of premiums in general. Suppose a bonus-malus system has an elasticity of 0.1 at $\lambda = 0.10$. Then the two drivers, with claim frequencies 0.10 and 0.11, pay premiums such that the second driver pays only 1% more in the long run. With 10% being the ideal price difference, we can see that either the first driver pays too much or the second driver pays too little. Of course it is also possible that the premiums of both drivers are overpriced or underpriced. How do we know who pays the correct price? The answer to this question is discussed in Section 4.1.2.

4.1.2 The Central Value of Bonus-Malus Systems

The central value of a bonus-malus system is the parameter value for which the premium of the system equals the risk premium and can therefore be considered the optimal premium of a policy-holder. In the case of asymptotic elasticity, we use the mean stationary premium as the premium of the system. The risk premium is the expectation of the total claim costs i.e. the size of the claim multiplied by the expected number of claims. To simplify calculations, we can choose a monetary unit such that the average claim size is 1. Now the risk premium is equal to λ, the expected number of claims. The information and calculations of this section are based on [10, Chapter 6] and [2].

Definition 4.4. The *central value* (in the context of asymptotic elasticity) is a fixed point of function $P(\lambda)$ i.e. a value λ_0 such that $P(\lambda_0) = \lambda_0$.

Suppose again that a bonus-malus system has an elasticity of 0.1 at $\lambda = 0.10$ and suppose that $P(0.10) = 0.10$, so that 0.10 is the central value of the system. Then, to answer the question in the end of Section 4.1.1, the first driver ($\lambda = 0.10$) pays the optimal amount and the second driver ($\lambda = 0.11$) pays too little. If instead 0.11 is the central value, the second driver pays the correct amount and the first driver pays too much. If we were to set all of the premium values such that $P(\lambda) = \lambda$ would hold for all claim frequencies, then $P(\lambda)$ would be linear and the system would have an elasticity of 1 for all values of λ, along with an infinite number of central values. Note that this last case is actually impossible when $P(\lambda)$ is the mean stationary premium, as we saw in the previous section. Nevertheless, when $\eta(\lambda) \geq 1$ for some λ there may be more than one central value. We can show, however, that when the elasticity is less than 1 for all λ, only one value of λ satisfies the equation $P(\lambda) = \lambda$.

Theorem 4.1. *The central value λ_0 is unique when $\eta(\lambda) < 1$ for all $\lambda > 0$.*

Proof. Let $\lambda < \lambda_0$ and assume $P(\lambda_0) = \lambda_0$. We have

$$\begin{aligned}
\int_{\lambda}^{\lambda_0} \eta(\lambda) d\log\lambda &= \int_{\lambda}^{\lambda_0} \frac{d\log P(\lambda)}{d\log\lambda} d\log\lambda \\
&= \int_{\lambda}^{\lambda_0} d\log P(\lambda) \\
&= \log P(\lambda_0) - \log P(\lambda) \\
&= \log\lambda_0 - \log P(\lambda) \\
&= \log\lambda_0 - \log\lambda + \log\lambda - \log P(\lambda) \\
&= \int_{\lambda}^{\lambda_0} d\log\lambda + \log\lambda - \log P(\lambda).
\end{aligned}$$

Rearranging the terms we get

$$\begin{aligned}
\log P(\lambda) &= \int_{\lambda}^{\lambda_0} d\log\lambda + \log\lambda - \int_{\lambda}^{\lambda_0} \eta(\lambda) d\log\lambda \\
&= \int_{\lambda}^{\lambda_0} (1 - \eta(\lambda)) d\log\lambda + \log\lambda.
\end{aligned}$$

By taking the exponential of both sides, we get

$$\exp[\log P(\lambda)] = \exp\left[\int_{\lambda}^{\lambda_0} (1 - \eta(\lambda)) d\log\lambda + \log\lambda\right]$$

so that

$$\begin{aligned}
P(\lambda) &= \exp\left[\int_{\lambda}^{\lambda_0} (1 - \eta(\lambda)) d\log\lambda\right] \exp(\log\lambda) \\
&= \lambda \exp\left[\int_{\lambda}^{\lambda_0} (1 - \eta(\lambda)) d\log\lambda\right].
\end{aligned}$$

If $\eta(\lambda) < 1$ for all values of λ, the integral is positive. Then

$$\exp\left[\int_{\lambda}^{\lambda_0} (1 - \eta(\lambda)) d\log\lambda\right] > 1$$

and $P(\lambda) > \lambda$ when $\lambda < \lambda_0$. Similarly for the case $\lambda > \lambda_0$, we get

$$P(\lambda) = \lambda \exp\left[-\int_{\lambda_0}^{\lambda} (1 - \eta(\lambda)) d\log\lambda\right]$$

and

$$\exp\left[\int_{\lambda}^{\lambda_0}(1-\eta(\lambda))d\log\lambda\right] < 1$$

and hence $P(\lambda) < \lambda$ when $\lambda > \lambda_0$. Therefore $P(\lambda) \neq \lambda$ whenever $\lambda \neq \lambda_0$. □

These results show that the solution to $P(\lambda_0) = \lambda_0$ is unique whenever the asymptotic elasticity is less than 1. This also shows that for insureds with higher (lower) risk parameters, the mean stationary premium is lower (higher) than their risk. Intuitively λ_0 is the parameter for which the policy-holder pays the correct price in relation to his or her risk. This means that the better drivers subsidize the worse driver or as Lemaire writes, "The better risks ($\lambda < \lambda_0$) pay too much; the worse risks ($\lambda > \lambda_0$) do not pay enough" [10]. We know now that the bonus-malus system in the previous examples has a unique central value, which we will calculate in the following example.

Example 4.2 (Continuation of Example 3.2)**.** Solving the equation

$$P(\lambda_0) = \frac{-(0.5+1.5\lambda_0)e^{-2\lambda_0} - 0.5e^{-\lambda_0} + 1.5}{1-\lambda_0 e^{-2\lambda_0}} = \lambda_0$$

gives $\lambda_0 \approx 1.31$. A claim frequency this large is not common. Most policy-holders in this system will have a lower claim frequency and pay "too much" to subsidize those few with a claim frequency higher than 1.31. To make this system fairer, premiums should be lowered.

More than one central value may exist for bonus-malus systems with elasticity greater or equal to 1, but the conclusion that drivers with certain claim frequencies subsidize the rest holds. This is easily observed by drawing a straight line $P(\lambda) = \lambda$ onto a mean stationary premium graph. The central values are the points where the line intersects the mean stationary premium, while the parts of the mean stationary premium that remain on the left side of the line represent the overpriced policies. In any case, drivers with claim frequencies close to zero will inevitably pay too much and drivers with very high claim frequencies will pay too little due to the shape of the mean stationary premium curve.

4.2 Transient Elasticity

In 1976, Jean Lemaire defined a new way to calculate the elasticity of bonus-malus systems. This section follows Lemaire [10] unless otherwise stated. The transient concept of elasticity does not involve stationary distributions and is instead dependent on the policy-holder's starting class. It is common for policy-holders to enter a bonus-malus system starting from different classes based on classification variables such as age, location of residence or car model. Another factor that can influence the starting class is the driver's class in another insurance company, if applicable.

Transient elasticity is defined in the same way as asymptotic elasticity with the difference that the mean stationary premium is replaced by a function called the discounted expectation of payments, which depends on the driver's starting class i. In order to define this function, we introduce a random variable B_j, which denotes the premium paid during year $j \in \{1, \dots, n\}$, and a *discount factor* β, which determines the present value of a payment. The value n denotes the length of the policy-holder's driving lifetime. Note that B_j is dependent on the starting class i. The expectation of B_1 is b_i, a known value. The expectations of the premiums of the following years are determined by the possible outcomes (the class of the driver given the previous class) and the probabilities of these outcomes. As before, $p_k(\lambda)$ denotes the probability of k accidents occurring to a driver with claim frequency λ.

Definition 4.5. Suppose $0 < \beta < 1$ is a discount factor. The *discounted expectation of payments* $\nu_i^{(n)}(\lambda)$ of class i is defined by

$$\begin{aligned}\nu_i^{(n)}(\lambda) &= \mathbb{E}(B_1 + \beta B_2 + \beta^2 B_3 + \cdots + \beta^{n-1} B_n) \\ &= b_i + \beta \sum_{k_1=0}^{\infty} p_{k_1}(\lambda) b_{T_{k_1}(i)} + \beta^2 \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} p_{k_1}(\lambda) p_{k_2}(\lambda) b_{T_{k_2}\left(T_{k_1}(i)\right)} \\ &\quad + \ldots + \beta^{n-1} \sum_{k_1=0}^{\infty} \cdots \sum_{k_{n-1}=0}^{\infty} p_{k_1}(\lambda) \ldots p_{k_{n-1}}(\lambda) b_{T_{k_{n-1}} \cdots \left(T_{k_2}\left(T_{k_1}(i)\right)\right)},\end{aligned}$$

where $i = 1, \dots, s$ and k_j is the number of accidents during year j.

Function $\nu_i^{(n)}(\lambda)$ is the discounted expectation of all the premiums paid by a policy-holder with claim frequency λ, who begins their n-year driving lifetime in class i. This can also be expressed using recursive formulas. The idea is that the insured's total premium is the premium of the first year b_i plus the present value of the expectation of the premiums payed from the second year until the end of the n-year driving lifetime. The premium from the second year until year n is then dependent on the claim number of the first year

k and calculated using the same formula. This process is repeated until we reach $n = 1$. The recursive formulas are

$$\nu_i^{(n)}(\lambda) = b_i + \beta \sum_{k=0}^{\infty} p_k(\lambda) \nu_{T_k(i)}^{(n-1)}(\lambda)$$

for $i = 1, \ldots, s$. To simplify calculations, we introduce the infinite-horizon discounted expectation by taking the limit of $\nu_i^{(n)}(\lambda)$.

Definition 4.6. The *infinite-horizon discounted expectation of payments* $\nu_i(\lambda)$ is defined by

$$\nu_i(\lambda) = \lim_{n\to\infty} \nu_i^{(n)}(\lambda)$$

and hence

$$\nu_i(\lambda) = b_i + \beta \sum_{k=0}^{\infty} p_k(\lambda) \nu_{T_k(i)}(\lambda) \tag{4.1}$$

for $i = 1, \ldots, s$.

Like the mean stationary premium, the functions $\nu_i(\lambda)$ are increasing and bounded. As before, we assume that the yearly claim numbers $p_k(\lambda)$ follow a Poisson distribution. When $\lambda \to 0$, $p_0(\lambda) \to 1$ and $p_j(\lambda) \to 0$ for $j \geq 1$, so

$$\lim_{\lambda\to 0} \nu_i(\lambda) = b_i + \beta \lim_{\lambda\to 0} \nu_{T_0(i)}(\lambda).$$

This is the lower bound of $\nu_i(\lambda)$. Let's consider what happens when $\lambda \to \infty$. Transition rules are specified up to a certain accident number $m \in \{0, 1, 2, \ldots\}$, after which the transition rules for accident numbers $k > m$ are the same as for m accidents. In Example 3.1, $m = 2$. We can therefore write Equation (4.1) as

$$\begin{aligned} \nu_i(\lambda) &= b_i + \beta \left(p_0(\lambda)\nu_{T_0(i)}(\lambda) + \cdots + p_m(\lambda)\nu_{T_m(i)}(\lambda) + p_{m+1}(\lambda)\nu_{T_m(i)}(\lambda) + \ldots \right) \\ &= b_i + \beta \left(p_0(\lambda)\nu_{T_0(i)}(\lambda) + \cdots + p_{\geq m}(\lambda)\nu_{T_m(i)}(\lambda) \right), \end{aligned}$$

where $p_{\geq m}(\lambda) = \mathbb{P}(K \geq m)$. Now we can see that when $\lambda \to \infty$, $p_j(\lambda) \to 0$ for $j < m$ and $p_{\geq m}(\lambda) \to 1$, and so the upper bound is

$$\lim_{\lambda\to\infty} \nu_i(\lambda) = b_i + \beta \lim_{\lambda\to\infty} \nu_{T_m(i)}(\lambda),$$

where m is the largest accident number specified by the transition rules. Often a very large number of accidents results in a transition to the highest class so that $T_m(i) = s$ and

$$\lim_{\lambda\to\infty} \nu_i(\lambda) = b_i + \beta \lim_{\lambda\to\infty} \nu_s(\lambda),$$

as is the case in Example 3.1.

The use of the infinite-horizon discounted expectations of payments to define transient elasticity requires the solution of the set of equations $\nu_i(\lambda)$ to be unique.

Theorem 4.2. *The set of equations*

$$\begin{cases} \nu_1(\lambda) = b_1 + \beta \sum_{k=0}^{\infty} p_k(\lambda)\nu_{T_k(1)}(\lambda) \\ \vdots \\ \nu_s(\lambda) = b_s + \beta \sum_{k=0}^{\infty} p_k(\lambda)\nu_{T_k(s)}(\lambda) \end{cases}$$

has a unique solution.

In order to prove Theorem 4.2, we first introduce contraction mappings and the contraction mapping theorem. The theory behind these concepts and the proof of the contraction mapping theorem are omitted in this thesis. They can be found in Peeler [12].

Definition 4.7. Let (K, d) be a metric space. The map $f : K \longrightarrow K$ is a *contraction mapping* of (K, d) if for some real number $0 \leq \beta < 1$ we have

$$d(f(x), f(y)) \leq \beta d(x, y)$$

for all $x, y \in K$.

Lemma 4.1 (Contraction mapping theorem)**.** *Let (K, d) be a non-empty, complete metric space and let $f : K \longrightarrow K$ be a contraction mapping of the space. Then f has a unique fixed point in K i.e. there exists an x_0 such that $f(x_0) = x_0$.*

Proof. See Peeler [12] for the proof. □

Now we can show that the set of equations defined by Equation (4.1) has a unique solution. The following proof is based on [9, p. 169].

Proof of Theorem 4.2. Let $f(x) = y$ and $f(y) = x$, where x and y are vectors such that

$$y_i(\lambda) = b_i + \beta \sum_{k=0}^{\infty} p_k(\lambda) x_{T_k(i)}(\lambda)$$

and

$$x_i(\lambda) = b_i + \beta \sum_{k=0}^{\infty} p_k(\lambda) y_{T_k(i)}(\lambda).$$

We set $d(x, y) = \max_i |x_i - y_i|$. Now we have

$$\begin{aligned}
d(f(y), f(x)) =& d(x, y) \\
=& \max_i |x_i - y_i| \\
=& \max_i \left| b_i + \beta \sum_{k=0}^{\infty} p_k(\lambda) y_{T_k(i)}(\lambda) - b_i - \beta \sum_{k=0}^{\infty} p_k(\lambda) x_{T_k(i)}(\lambda) \right| \\
=& \max_i \left| \beta \sum_{k=0}^{\infty} p_k(\lambda) \left(y_{T_k(i)}(\lambda) - x_{T_k(i)}(\lambda) \right) \right| \\
=& \beta \max_i \left| \sum_{k=0}^{\infty} p_k(\lambda) \left(y_{T_k(i)}(\lambda) - x_{T_k(i)}(\lambda) \right) \right| \\
\text{(Triangle inequality)} \qquad \leq& \beta \sum_{k=0}^{\infty} \max_i \left| p_k(\lambda) \left(y_{T_k(i)}(\lambda) - x_{T_k(i)}(\lambda) \right) \right|.
\end{aligned}$$

Now, since $T_k(i) = j$ and $\sum_{k=0}^{\infty} p_k(\lambda) = 1$, we have

$$\begin{aligned}
& \beta \sum_{k=0}^{\infty} \max_i \left| p_k(\lambda) \left(y_{T_k(i)}(\lambda) - x_{T_k(i)}(\lambda) \right) \right| \\
=& \beta \sum_{k=0}^{\infty} p_k(\lambda) \max_j |(y_j(\lambda) - x_j(\lambda))| \\
=& \beta \max_j |y_j(\lambda) - x_j(\lambda)| \\
=& \beta d(y, x),
\end{aligned}$$

and so $d(f(y), f(x)) \leq \beta d(y, x)$. Therefore f is a contraction mapping and from Theorem 4.1 we can see that there exists a unique vector ν such that $f(\nu) = \nu$ or

$$f(\nu_i) = b_i + \beta \sum_{k=0}^{\infty} p_k(\lambda) \nu_{T_k(i)}(\lambda) = \nu_i(\lambda).$$

for all $i = 1, \ldots, s$. $\square$

We can now define transient elasticity as the elasticity of the infinite-horizon discounted expectation of payments $\nu_i(\lambda)$.

Definition 4.8. *Transient elasticity* $\gamma_i(\lambda)$ of class i is defined as

$$\gamma_i(\lambda) = e[\nu_i(\lambda)] = \nu_i'(\lambda)\frac{\lambda}{\nu_i(\lambda)} = \frac{d\log\nu_i(\lambda)}{d\log\lambda} \tag{4.2}$$

for $i = 1, 2, ..., s$.

The derivative of $\nu_i(\lambda)$ is

$$\nu_i'(\lambda) = \beta\sum_{k=0}^{\infty}\left[p_k'(\lambda)\nu_{T_k(i)}(\lambda) + p_k(\lambda)\nu_{T_k(i)}'(\lambda)\right],$$

which is

$$\nu_i'(\lambda) = \beta\sum_{k=0}^{\infty}\frac{e^{-\lambda}\lambda^k}{k!}\left[\left(\frac{k}{\lambda} - 1\right)\nu_{T_k(i)}(\lambda) + \nu_{T_k(i)}'(\lambda)\right]$$

when the number of claims follows the Poisson distribution.

We can expect the transient elasticity curves to have a similar shape to those of the asymptotic elasticity because the functions $\nu_i(\lambda)$ are bounded, increasing functions just like the mean stationary premium $P(\lambda)$. The discussions on perfect elasticity and the central value of Sections 4.1.1 and 4.1.2 apply also to transient elasticity. It can be shown by calculations similar to those in the proof of Theorem 4.1 that the central value is unique provided that $\gamma_i(\lambda) < 1$. The same conclusion as in Section 4.1.2 holds: the better risks subsidize the worse risks. Next we will calculate the transient elasticity of the bonus-malus system in the previous examples to see how similar the two elasticities are.

Example 4.3 (Continuation of Example 3.1)**.** Suppose the discount factor $\beta = 0.9$. The premium scale (b_1, b_2, b_3) is chosen to be $(0.5, 1, 1.5)$ as in Example 3.1. The number of claims follows a Poisson distribution with parameter λ, so the probabilities $p_k(\lambda)$ are as in Example 3.2. We begin by inserting the values of the probabilities, the discount factor, the premiums and the transition rules into Equation (4.1) to get

$$\begin{cases} \nu_1(\lambda) = 0.5 + 0.9\left(e^{-\lambda}\nu_1(\lambda) + \lambda e^{-\lambda}\nu_2(\lambda) + \left(1 - \lambda e^{-\lambda} - e^{-\lambda}\right)\nu_3(\lambda)\right) \\ \nu_2(\lambda) = 1 + 0.9\left(e^{-\lambda}\nu_1(\lambda) + \left(1 - e^{-\lambda}\right)\nu_3(\lambda)\right) \\ \nu_3(\lambda) = 1.5 + 0.9\left(e^{-\lambda}\nu_2(\lambda) + \left(1 - e^{-\lambda}\right)\nu_3(\lambda)\right). \end{cases}$$

Using Wolfram Mathematica [5] we get

$$\begin{cases} \nu_1(\lambda) = \dfrac{5\left(-9e^{\lambda}\lambda - 243\lambda - 99e^{\lambda} + 280e^{2\lambda} - 81\right)}{100e^{2\lambda} - 81\lambda} \\ \nu_2(\lambda) = \dfrac{5\left(-243\lambda - 99e^{\lambda} + 290e^{2\lambda} - 81\right)}{100e^{2\lambda} - 81\lambda} \\ \nu_3(\lambda) = \dfrac{15\left(-81\lambda - 30e^{\lambda} + 100e^{2\lambda} - 27\right)}{100e^{2\lambda} - 81\lambda} \end{cases}$$

as the solution to the set of equations. Once we take the derivatives of the equations above and insert the values into Equation (4.2), we have that the elasticities are

$$\gamma_1(\lambda) = \frac{9\lambda\left(81e^{\lambda}\left(\lambda^2 + 11\lambda - 11\right) + 100e^{3\lambda}(\lambda+10) + 180e^{2\lambda}(2\lambda+9) - 729\right)}{\left(100e^{2\lambda} - 81\lambda\right)\left(-9e^{\lambda}(\lambda+11) + 280e^{2\lambda} - 81(3\lambda+1)\right)},$$

$$\gamma_2(\lambda) = \frac{9\lambda\left(891e^{\lambda}(\lambda-1) + 1100e^{3\lambda} + 90e^{2\lambda}(2\lambda+19) - 729\right)}{\left(100e^{2\lambda} - 81\lambda\right)\left(-81(3\lambda+1) - 99e^{\lambda} + 290e^{2\lambda}\right)}$$

and

$$\gamma_3(\lambda) = \frac{3\left(810e^{\lambda}(\lambda-1) + 1800e^{2\lambda} + 1000e^{3\lambda} - 729\right)\lambda}{\left(100e^{2\lambda} - 81\lambda\right)\left(-81\lambda - 30e^{\lambda} + 100e^{2\lambda} - 27\right)}$$

for classes 1, 2 and 3 respectively. Figure 4.2 shows the graphs of these functions, along with the asymptotic elasticity function of the same system shown in Figure 4.1.

Transient elasticity poses advantages when compared to asymptotic elasticity. Firstly, it allows us to evaluate bonus-malus systems based on the starting class of the insured, and can consequently be thought of as a more accurate measure. This is especially useful when an insurance company places new policy-holders in different classes depending on criteria chosen by the company. Secondly, asymptotic elasticity is valid only when the stationary state has been reached, which may never be the case due to changing economic conditions and the flow of customers [8]. We can also apply transient elasticity easily to non-Markovian bonus-malus systems, because transient elasticity is not based on the Markov properties.

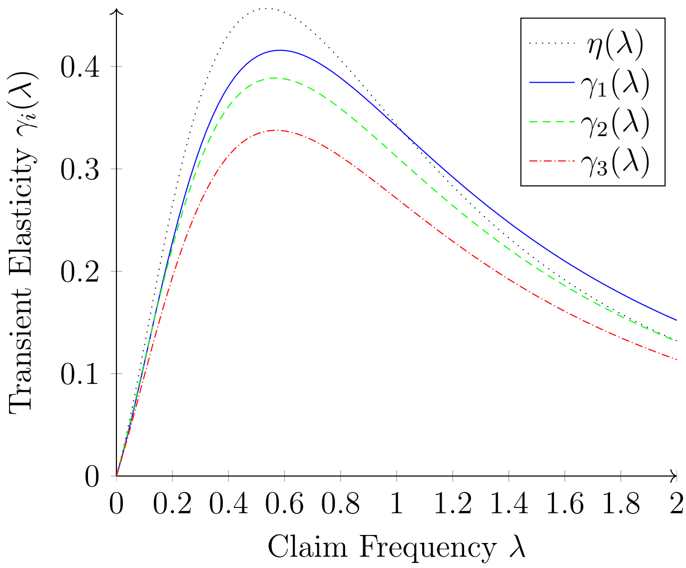

Figure 4.2: Transient elasticities of the bonus-malus system in Example 4.3. Overall, the largest elasticity values seem to be reached by those who start in class 1, although the differences between the elasticities are negligible for very small claim frequencies. The asymptotic elasticity function $\eta(\lambda)$ has a higher peak than the transient elasticity functions. Nonetheless, the shapes of the elasticities are similar and the same conclusions can be drawn from both types: the most optimal values of elasticity are reached somewhere between λ =0.5 and $\lambda = 0.6$, but because these values are lower than 0.5, this bonus-malus system is quite inelastic.

5 Application to Real-Life Bonus-Malus Systems

The purpose of this section is to find, assess and compare the elasticities of real-life bonus-malus systems, namely, bonus-malus systems of Finnish insurance companies. While transient elasticity is better for evaluation of bonus-malus systems depending on a driver's starting class, asymptotic elasticity is simpler when comparing different systems, which is why only asymptotic elasticity is considered in this section. As before, we use the Poisson distribution to model yearly claim quantities.

The bonus-malus systems evaluated in this application are those of five well known Finnish insurance companies that offer automobile insurance. Although the companies offer both compulsory liability insurance and voluntary motor vehicle insurance, only the bonus-malus systems of the former are considered. Many companies also specify different transition rules and bonuses for commercial and private vehicle drivers or for car and motorcycle drivers, but only the bonus-malus systems of drivers of non-commercial cars are considered. The insurance premiums, expressed as bonus percentages, and transition rules of the systems have been collected from publicly accessible sources in October 2020.

5.1 Finding Elasticities Using R

Real-life bonus-malus systems are notably more sophisticated than the three class example explored throughout this thesis. Finnish insurance companies' bonus-malus systems tend to have approximately 20 classes. As we saw in Method 1 of Example 2.2, finding the exact value of the stationary distribution (and consequently the asymptotic elasticity) consists of solving a system of $s+1$ equations, where s is the number of classes. It is apparent that solving large systems of equations can be complicated, or tedious at the very least. For the purpose of this application, it is thus favorable to devise a more effective method than the one used to find the elasticity of the three class system in the previous examples.

The aim is to develop a more efficient way to find the elasticity

$$\eta(\lambda) = P'(\lambda)\frac{\lambda}{P(\lambda)} = \sum_{i=1}^{s} \mu_i'(\lambda) b_i \frac{\lambda}{\sum_{i=1}^{s} \mu_i(\lambda) b_i} \tag{5.1}$$

of a bonus-malus system, given its transition rules and premium scale. This can be done by creating a function in R [13] that takes the transition rules, premium scale and claim frequency λ as variables and returns the value of the asymptotic elasticity at that claim frequency. The function can then be visualized by plotting its values for various claim frequencies. The R code of the function used can be found in Appendix A.

The first step is to find the transition matrix $M(\lambda)$. Transition rules can be expressed in matrix form T_k as in Example 3.1 and the transition matrix is calculated from the formula $M(\lambda) = \sum_{k=0}^{\infty} p_k(\lambda) T_k$ as in Example 3.2. Notice that only matrices $T_0 \dots T_k$ are needed where k is the largest accident number specified by the bonus-malus system. For example, a system specifying different rules for "0", "1" and "2 or more" accidents will require the input of matrices T_0, T_1 and T_2 into R, along with the Poisson probabilities $\mathbb{P}(K = 0)$, $\mathbb{P}(K = 1)$ and $\mathbb{P}(K \geq 2)$, which can be found as functions in R. The code in Appendix A admits transition rules of up to six accidents (because it is the highest number specified by Finnish insurance companies), but can be easily modified to accommodate rules of higher accident numbers.

The stationary distribution vector $\mu(\lambda) = \left[\mu_1(\lambda) \dots \mu_s(\lambda)\right]$ can then be calculated as in Method 2 of Example 2.2, raising the transition matrix to a high power using the matrix power function from package "expm" [4] and choosing the first row of the resulting matrix. The next step is to find the derivative of each class's stationary distribution. Since we have the stationary distribution of each class as a vector of values instead of functions (and differentiating a function involving matrices would be complicated in R), the derivative is most easily found through estimation. We can obtain an estimate of the derivative vector at λ by choosing values λ_1 and λ_2 closely on either side of λ and calculating the gradient

$$\frac{\mu(\lambda_2) - \mu(\lambda_1)}{\lambda_2 - \lambda_1} \approx \left[\mu_1'(\lambda) \dots \mu_s'(\lambda)\right].$$

This, of course, entails finding the stationary distributions at λ_1 and λ_2 beforehand. After finding the stationary distribution and its derivative as vectors, the elasticity can easily be calculated from Equation (5.1) using the sum function in R.

5.2 Evaluation and Comparison of Finnish Bonus-Malus Systems

Figure 5.1 shows the elasticities of the five Finnish bonus-malus systems calcuated in R. The curves have clear peaks and tend towards zero on both ends of the bottom axis, as

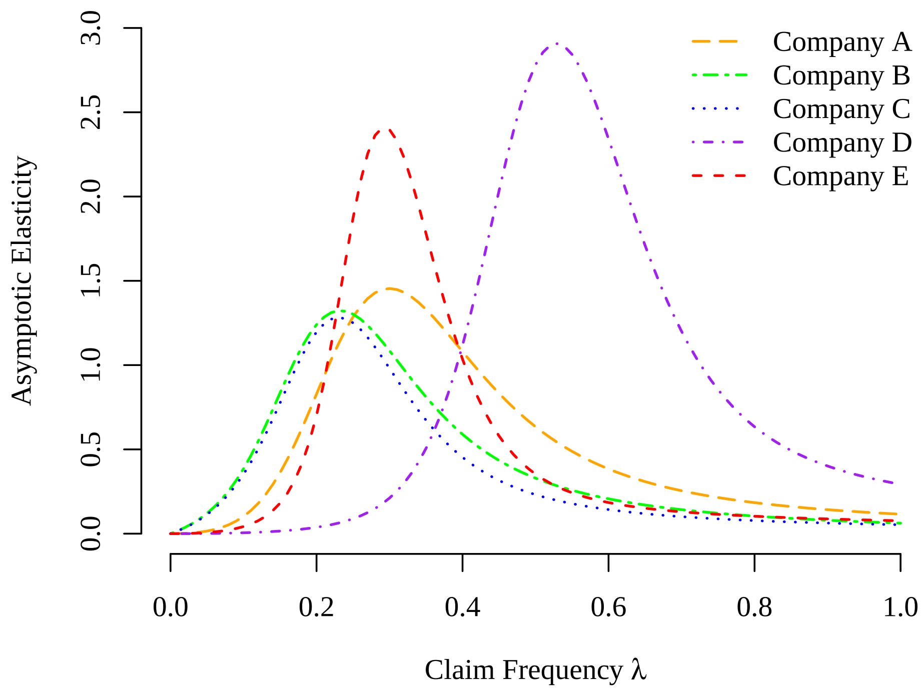

Figure 5.1: Asymptotic elasticity functions of five Finnish bonus-malus systems for $\lambda \in [0, 1]$. All of the maximum elasticities are above 1, implying that the systems are highly elastic.

expected. What is perhaps most notable, is how high the peaks are and how exaggerated the elasticities of companies D and E are.

Definition 5.1. If $\eta(\lambda) > 1$ for some λ we call the system *overelastic* for that value of λ.

All of the bonus-malus systems are overelastic for some values of λ. This is somewhat surprising because in his 1995 book, Lemaire states that elasticity remains below 1 in most cases and that overelasticity is uncommon [10]. This claim is corroborated by his comparison of the Swiss, Japanese, Taiwanese and Belgian systems, of which only the Swiss system reaches and goes above an elasticity of 1. Other sources also support

Lemaire's claim. In 1972, a comparison by Vepsäläinen [15] between the Norwegian, West German, Danish, Finnish, Swedish and Swiss bonus-malus systems showed that only the Norwegian and Swiss systems had elasticities over 1, the Norwegian system having maximum elasticity just above 1 and the Swiss having a maximum elasticity approaching 2. In 2011 Duplan et al. [3] found in a comparison between the bonus-malus systems of Brazil, Switzerland, Japan and Hong Kong that the Japanese system had a maximum elasticity slightly above 1 and the Swiss system had a maximum elasticity of almost 1.6, while the rest remained below 1.

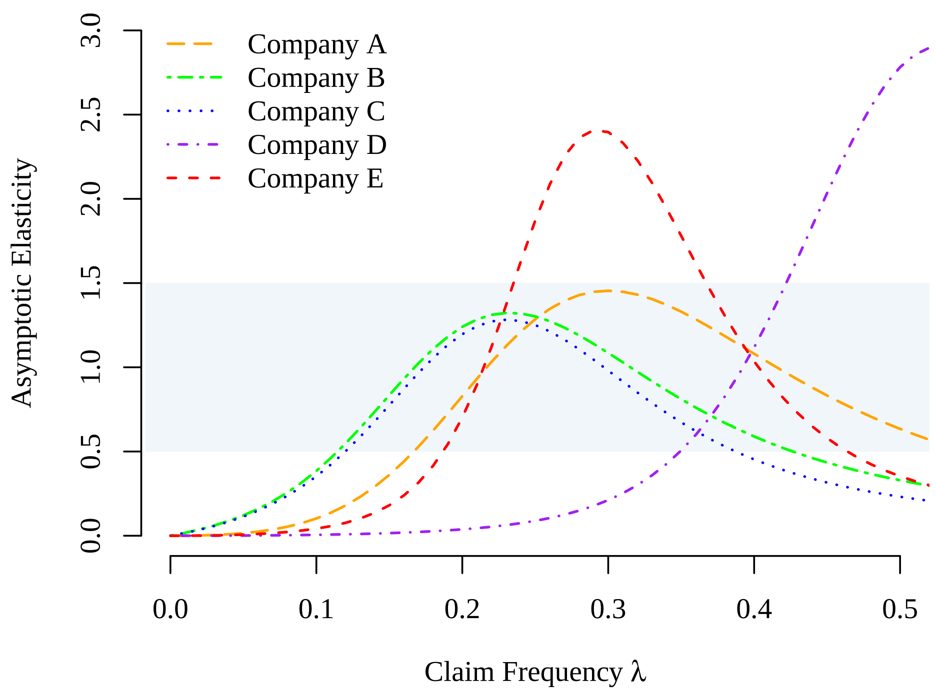

Figure 5.2: Asymptotic elasticity functions of five Finnish bonus-malus systems for $\lambda \in [0, 0.5]$. The shaded region helps to visualize where the elasticities are in the vicinity of 1 (perfect elasticity).

In Section 4 we established that an elasticity of 1 is the optimum, but it is not evident from the graph which bonus-system is the most ideal seeing that all of the systems have elasticities equal or close to 1 at various claim frequencies. If we focus on the more realistic claim frequencies, we can draw better conclusions. Figure 5.2 shows the same graph for claim frequencies up to 0.5, which amounts to one accident every two years. In this range, the elasticities of companies A, B and C remain below 1.5 for all claim frequencies and above 0.5 more often than not, as shown by the shaded region. The elasticity of Company D, on the other hand, goes quickly from one extreme to another, spending a relatively small amount of time in the vicinity of 1. It is between 0.5 and 1.5 only for relatively high claim frequencies. The elasticity of Company E also remains beyond the shaded region for most claim frequencies.

For even more exact results we can examine specific values of λ. While each bonus-malus system seems to reach perfect elasticity approximately between the claim frequencies 0.15 and 0.40, Lemaire states that the most common claim frequecies are less than or equal to 0.15 [10]. Table 5.1 shows the elasticities of each system for claim frequencies 0.05, 0.10 and 0.15. The order of the systems from closest to perfect elasticity to furthest from perfect elasticity is the same for all three claim frequencies and is shown in the last column of the table.

Insurance Company	Asymptotic Elasticity at			Rank
	$\lambda = 0.05$	$\lambda = 0.10$	$\lambda = 0.15$	
A	0.02	0.10	0.36	3
B	0.12	0.38	0.83	1
C	0.11	0.35	0.78	2
D	0.00	0.01	0.02	5
E	0.01	0.04	0.18	4

Table 5.1: Asymptotic elasticities of five Finnish bonus-malus systems for three realistic claim frequencies. The systems are ranked according to closeness to perfect elasticity.

Overelasticity is no longer present for these values of λ. Only the systems of Companies B and C reach values close to 1, doing so only at the highest chosen claim frequency. On the other end we have Company D with extremely low elasticities. To interpret the results, we recall that a driver with claim frequency 0.11 should, in the long run, pay 10% more premiums than a driver with claim frequency 0.10. At Company D, however, the elasticity at $\lambda = 0.10$ being only 0.01, the first driver will only pay 0.1% more premiums, an insignificant punishment. The same driver at Company A will pay 1% more, a more appropriate amount, but still very little.

Regardless of which of the three claim frequencies we choose, we can conclude that the best elasticity values are observed for Company B, closely followed by Company C. Even for higher frequencies the elasticities of these two systems remain within a reasonable range. The same can be said for Company A, but for lower values of λ its elasticity is less than half of the elasticities of B and C. Similarly, the values of Company E are approximately half of those of Company A for the claim frequencies shown in Table 5.1. While Company D's system may have initially seemed highly elastic from Figure 5.1, its elasticity for these lower values is extremely low, making it the unfairest system of the five. Through this discussion, we may conclude that the ranking in Table 5.1 corresponds to the ranking of the systems from "most optimal" to "least optimal" according to this measure.

6 Conclusion

The purpose of this thesis is to give insight to one of the tools used to measure the performance of bonus-malus systems, elasticity, and to compare Finnish bonus-malus systems using this tool. We have shown how both the mean stationary premium and the discounted expectation of payments can be used as representations of the insurance premium of a given policy-holder, resulting in two types of elasticity: asymptotic elasticity and transient elasticity. While the former is a single function useful for comparison between bonus-malus systems, the latter provides additional information based on the insured's starting class. In Section 5, asymptotic elasticity was used to rank the bonus-malus systems of five Finnish insurance companies. Future work could include inspecting the transient elasticity of these systems. Other ways of defining elasticity exist as well and are worth examining. For instance, Nelson De Pril generalizes the elasticities proposed by Loimaranta and Lemaire in his 1978 article [2].

Inspecting the central value in the context of elasticity gave us an interesting result in the studied model: when the elasticity is less than 1 for all claim frequencies, the low-risk drivers subsidize the high-risk drivers. If the elasticity is higher, the central value concept can still be used to determine fair premiums. However, both perfect elasticity for all claim frequencies and infinite central values are impossible to achieve in this model due to the characteristics of bonus-malus systems, which means that no system is completely optimal. It would be interesting to also calculate the central values of the five Finnish bonus-malus systems, but this would require knowledge of claim sizes and actual premium values (instead of bonus percentages).

Although we explore elasticity as a measure of optimality, this approach is very theoretical and in practice there are many aspects to be considered when evaluating bonus-malus systems. For instance, a highly efficient system often gives rise to more bonus hunger, which is when policy-holders choose not to report small claims in order to maintain low insurance costs [8]. This subsequently leads to inaccurate perceptions of the drivers' risks, undermining the purpose of the bonus-malus system. In fact, Jean Lemaire prefers to talk about *toughness* instead of optimality in his book *Bonus-Malus Systems in Automobile Insurance* [10]. His book includes three other tools that measure toughness which, together with elasticity, provide a more comprehensive evaluation than elasticity alone does.

References

[1] G. Cowan. Derivation of the Poisson distribution. *Royal Holloway, University of London. Dec*, 2009.

[2] N. De Pril. The efficiency of a bonus-malus system. *ASTIN Bulletin: The Journal of the IAA*, 10(1):59–72, 1978.

[3] N. Duplan, C. Hall, J. Nguyen, C. Willis, and K. P. Das. Evaluating the elasticity of a bonus-malus system. *Pi Mu Epsilon Journal*, 13(5):269–276, 2011.

[4] V. Goulet, C. Dutang, M. Maechler, D. Firth, M. Shapira, and M. Stadelmann. *expm: Matrix Exponential, Log, 'etc'*, 2020. R package version 0.999-5.

[5] W. R. Inc. *Mathematica, Version 12.1.1.0.* Champaign, IL, 2020.

[6] K. Izyurov. *Probability Theory.* University of Helsinki, 2019.

[7] J. Lehtomaa. *Tariff Theory.* University of Helsinki, 2020.

[8] J. Lemaire. Driver versus company. *Scandinavian Actuarial Journal*, 1976(4):209–219, 1976.

[9] J. Lemaire. *Automobile insurance: actuarial models*, volume 4. Springer Science & Business Media, 1985.

[10] J. Lemaire. *Bonus-malus systems in automobile insurance*, volume 19. Springer science & business media, 1995.

[11] K. Loimaranta. Some asymptotic properties of bonus systems. *ASTIN Bulletin: The Journal of the IAA*, 6(3):233–245, 1972.

[12] L. Peeler. Metric spaces and the contraction mapping principle. *Math. Unchicago. edu*, 2011.

[13] R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2020.

[14] J. S. Rosenthal. Convergence rates for Markov chains. *Siam Review*, 37(3):387–405, 1995.

[15] S. Vepsäläinen. Applications to a theory of bonus systems. *ASTIN Bulletin: The Journal of the IAA*, 6(3):212–221, 1972.

A R Code

```
#The following function gives the elasticity of a
#bonus-malus system at point x.

#The function takes variables x, b, T0, T1, T2, T3, T4, T5 and T6.

#Value x is the claim frequency. Vector b is the bonus scale.
#T0, T1, T2, T3, T4, T5 and T6 are the transition rules in
#matrix form for 0, 1, 2, 3, 4, 5 and 6 (or more) accidents.

#The function dpois(k,x) gives the probability of k accidents
#occurring when the claim frequency is k.
#The function ppois(k,x,lower.tail = F) gives the probability
#of more than k accidents occurring when the claim frequency is k.

#Package used to calculate matrix powers
library(expm)

elasticity <- function(x,b,T0,T1,T2,T3,T4,T5,T6) {

    #Number of classes
    s <- length(b)

    #Classes 0 to s
    i <- 0:s

    #A value approaching x from the left
    x1 <- x-0.0001

    #A value approaching x from the right
    x2 <- x+0.0001

    #Transition matrix at point x
```

```
    M <- dpois(0,x)*T0+dpois(1,x)*T1+dpois(2,x)*T2
         +dpois(3,x)*T3+dpois(4,x)*T4+dpois(5,x)*T5
         +ppois(5,x,lower.tail = F)*T6

    #Transition matrices at points x1 and x2
    M1 <- dpois(0,x1)*T0+dpois(1,x1)*T1+dpois(2,x1)*T2
          +dpois(3,x1)*T3+dpois(4,x1)*T4+dpois(5,x1)*T5
          +ppois(5,x1,lower.tail = F)*T6
    M2 <- dpois(0,x2)*T0+dpois(1,x2)*T1+dpois(2,x2)*T2
          +dpois(3,x2)*T3+dpois(4,x2)*T4+dpois(5,x2)*T5
          +ppois(5,x2,lower.tail = F)*T6

    #Stationary distribution vector at point x, the first row
    #of the 200th power of the transition matrix
    sd <- (M%^%200)[1,]

    #Stationary distribution vectors at points x1 and x2
    sd1 <- (M1%^%200)[1,]
    sd2 <- (M2%^%200)[1,]

    #Derivative of the stationary distribution at point x,
    #estimated from the gradient between x1 and x2
    d <- (sd2-sd1)/(x2-x1)

    #Part 1 of the elasticity equation
    p1 <- sum(d[i]*b[i])

    #Part 2 of the elasticity equation
    p2 <- x/sum(sd[i]*b[i])

    #The function returns the product of the two parts,
    #the elasticity at point x
    return(p1*p2)
}
```