Stockholms universitet

Optimization of Deductible Levels to Maximize Portfolio Utility in Insurance

Ida Lundmark

Masteruppsats i matematisk statistik
Master Thesis in Mathematical Statistics

Masteruppsats 2025:6
Matematisk statistik
Juni 2025

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Mathematical Statistics
Stockholm University
Master Thesis **2025:6**
http://www.math.su.se

Optimization of Deductible Levels to Maximize Portfolio Utility in Insurance

Ida Lundmark*

June 2025

Abstract

This thesis investigates how an insurance company can determine optimal deductible levels using Borch's theorem to maximize expected utility across a diverse portfolio, given an expected premium. Individual wealth characteristics, a key factor in Borch's framework, are assigned to each policyholder. Assuming a Bernoulli utility function, we compare the impact of Gamma and compound Poisson loss distributions on a representative policyholder, ultimately selecting the compound Poisson for final analysis. Using this framework, optimal deductibles are then numerically calculated for each policyholder and clustered into two- and three-level deductible options. The study also examines how varying wealth levels affect these results and the application of Borch's theorem.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: idalundmark@live.com. Supervisor: Mathias Millberg Lindholm.

Acknowledgment

I would like to thank Länsförsäkringar AB for providing the data used in this thesis. Their contribution has been important for completing this work. I would also like to thank my mentor at Stockholm University, Associate Professor Mathias Millberg Lindholm, at the Department of Mathematics, for his constructive feedback, support, and guidance throughout the process. His help has been crucial in improving the quality of this thesis.

AI Statement

Throughout the process of writing this thesis, I made use of language and AI-based tools, including ChatGPT and Grammarly. Grammarly was primarily used to support grammar and language refinement, ensuring clarity and consistency in the written text. ChatGPT was used to troubleshoot code, and ensure consistency in the formulation.

Contents

1 Introduction

Insurance is an essential arrangement for individuals and businesses to protect themselves financially against uncertain future events. In non-life insurance, this typically involves protection against damages to property, vehicles, or liability for harm caused to others. A policyholder (i.e., the insured) pay a premium to the insurer, who in return assumes the financial risk associated with these events (Ohlsson and Johansson, 2010).

A common feature in many insurance contracts is the deductible, i.e. the amount the policyholder must pay out of pocket before the insurer contributes to a claim. Deductibles serve to reduce the insurer's expected claim costs and, consequently, often result in lower premiums. As the deductible increases, the premium typically decreases, making the choice of deductible level a key consideration for policyholders seeking to balance risk and cost (Ohlsson and Johansson, 2010).

In today's competitive insurance market, policyholders are presented with a variety of products, including options with differing deductible levels. This means they must decide not only which insurer to choose but also which deductible level aligns with their financial capacity and risk preferences. For insurers, strategically offering deductible levels that appeal to a broad range of policyholders can constitute a significant competitive advantage. This raises a central question: How can an insurance company optimally set deductible levels across a heterogeneous portfolio of policyholders?

This thesis develops a method for determining deductible levels that an insurance company can offer across a diverse portfolio of policyholders. The method builds on Borch's classical utility-based framework, which defines the optimal deductible for an individual as the one that maximizes expected utility, accounting for the individual's wealth. This individual-level approach is extended to a portfolio context, with the goal of maximizing total utility across the insurer's entire portfolio.

The structure of the thesis is as follows. In Section 2, we review relevant insurance theory, with a focus on the concepts and models used in this study. Section 3 presents the methodological framework for determining optimal deductibles, including data preparation, wealth modeling, and numerical solutions. Section 4 outlines the results, and Section 5 concludes with a discussion of key findings and their implications.

2 Model Theory

2.1 Basic Insurance Theory

In insurance mathematics, the premium charged to each policyholder is generally based on the expected value of the losses they are likely to generate, along with a loading. This loading is intended to cover additional costs such as administration, capital costs, and profit margins, etc. Since different policyholders present different levels of risk, statistical models are used to determine appropriate premiums. According to Ohlsson and Johansson (2010), factors considered when estimating the expected cost of claims typically fall into three categories: properties of the policyholder, properties of the insured object, and properties of the geographic region.

A key concept in this context is the pure premium, which represents the expected cost of claims per policyholder, before any loading is added. It is calculated as:

$$\text{Pure Premium} = \text{Claim Frequency} \times \text{Claim Severity}$$

Here, claim frequency refers to the expected number of claims per policyholder over a given period, and claim severity refers to the average cost per claim. Multiplying the two gives the expected loss, which forms the basis for setting a fair and risk-adjusted premium. To set premiums, insurers often use tariff models, where the pure premium is modeled as a function of several rating factors. For insurance pricing, multiplicative models are particularly suitable. In the case of three rating factors, the pure premium π_{ijk} for a policyholder in class i for the first factor, class j for the second factor, and class k for the third factor can be expressed as:

$$\pi_{ijk} = \gamma_0 \cdot \gamma_{1i} \cdot \gamma_{2j} \cdot \gamma_{3k}.$$

To ensure that the model parameters are uniquely identified, a reference cell (also called a base cell) is defined, typically the one with the largest exposure. The multiplicative relativities for this base cell are set to 1. Under this constraint, γ_0 is the base value, and the parameters γ_{1i}, γ_{2j}, and γ_{3k} represent multiplicative relativities, that is, the relative effect of belonging to a specific level of a rating factor compared to the base level (Ohlsson and Johansson, 2010).

In practice, insurers may either model the pure premium directly, or separately model the claim frequency and claim severity components. A common choice is to use a Poisson distribution for the claim frequency and a Gamma distribution for the claim severity. Both components can then be modeled using generalized linear models (GLMs), and their product yields the estimated pure premium.

Another important element in insurance pricing and pure premium estimation is the deductible. A deductible is the amount the policyholder agrees to pay before the insurer covers any claim costs. By introducing a deductible, the insurer transfers a portion of the financial responsibility to the policyholder, which typically results in a lower pure premium. This reduction occurs because the insurer's expected cost of claims decreases when small claims are absorbed by the policyholder. From a modeling perspective, deductibles are often handled by analyzing claim amounts net of the deductible when the deductible level is fixed for all policyholders. However, when policyholders can choose among multiple deductible options, the situation becomes more complex. As shown in Ohlsson and Johansson (2010), deductibles do not act as standard multiplicative rating factors. In such cases, layered modeling approaches may be used, where each deductible level is treated as a separate coverage layer and analyzed independently. While this can work in simple cases, it may become impractical in settings with many deductible choices (Ohlsson and Johansson, 2010).

2.2 Utility Theory and Risk Aversion

A central concept in explaining why individuals are willing to pay an insurance premium that exceeds the expected loss (the pure premium) is utility theory. According to this theory, individuals assign a value through a utility function $u(w)$ to their wealth w, which represents the level of satisfaction or utility derived from that wealth. This utility function is typically assumed to be non-decreasing, reflecting the natural assumption that more wealth leads to greater utility. When faced with uncertainty, such as a potential loss X, a decision maker evaluates the desirability of different alternatives by comparing their expected utilities. In particular, the choice between bearing a random loss and paying a fixed insurance premium is modeled by comparing $\mathbb{E}[u(w-X)]$ and $u(w-P)$, where P is the premium. The value of P that satisfies:

$$\mathbb{E}[u(w-X)] = u(w-P) \tag{1}$$

represents the maximum premium the insured is willing to pay. At this point, the insured is indifferent, in terms of utility, between purchasing insurance and retaining the risk.

A key assumption in this framework is that individuals are risk averse, meaning they prefer a certain outcome over a risky one with the same expected value. This behavior is captured mathematically by the concavity of the utility function, which implies a decreasing marginal utility of wealth. In other words, while the utility increases as wealth increases, the rate at which utility increases diminishes as wealth grows. Formally, this means the first derivative of the utility function, $u'(w)$, is positive, but the second derivative, $u''(w)$, is negative. This decreasing marginal utility reflects the intuitive idea that each additional unit of wealth provides less additional satisfaction than the previous one. Jensen's inequality then guarantees that for any concave utility function u, it holds that:

$$\mathbb{E}[u(w - X)] \leq u(w - \mathbb{E}[X]).$$

Consequently, a risk-averse individual prefers to pay a certain amount equal to the expected loss rather than face the uncertainty of the actual loss. This preference explains why the maximum premium a risk-averse individual is willing to pay often exceeds the expected value of the loss (Kaas et al., 2008).

The foundations of utility theory were laid by the St. Petersburg paradox, which revealed that individuals are not willing to pay large amounts for gambles with infinite expected value (see Section 2.2.1) (Aase, 2001). Daniel Bernoulli proposed resolving this paradox by introducing the logarithmic utility function $u(w) = \log(w)$, defined for positive wealth $w > 0$ (Bernoulli, 1954). Throughout this thesis, this simpler form is used, assuming strictly positive wealth. Other commonly used utility functions include the linear utility function $u(w) = w$, and the power utility function $u(w) = w^c$, where $0 < c \leq 1$ (Kaas et al., 2008).

2.2.1 The St. Petersburg Paradox

Since the St. Petersburg game is defined in terms of repeated fair coin tosses, the number of tosses until the first head is a discrete random variable. Accordingly, the original formulation uses a summation rather than an integral to compute expected utility. In this section, we adopt this discrete approach to remain consistent with the historical and probabilistic structure of the paradox.

The St. Petersburg Paradox was originally presented by Nicolaus Bernoulli and later analyzed by Daniel Bernoulli. It is a well-known example in economic theory that illustrates the limitations of relying solely on the expected monetary value when making rational decisions under uncertainty. Initially in probability theory, it was assumed that a rational agent should choose between risky prospects by maximizing the expected monetary value, calculated as:

$$\mathbb{E}[S] = \sum_i s_i p_i$$

where s_i is the monetary outcome and p_i is the probability of that outcome. Under this criterion, a rational agent should be willing to pay any amount up to the expected monetary value of the gamble in order to participate (Aase, 2001; Machina, 1987).

> **St. Petersburg paradox**
> Consider a game in which a fair coin is repeatedly tossed until heads appears. At each toss, the player's potential profit is doubled. If heads appears on the i-th toss, the player's gain is 2^i. The expected profit is given by the following summation:
>
> $$\mathbb{E}[S] = \sum_{i=1}^{\infty} \frac{1}{2^i} \cdot 2^i = \infty.$$
>
> Despite this infinite expected value, no rational person would be willing to pay a very large amount to enter such a game (Kaas et al., 2008).

Daniel Bernoulli, as cited by Aase (2001), criticized the expected value criterion by noting: *"there should be no sensible man who would not be willing to sell his right to this gain for 20 ducats."* To resolve the paradox, Bernoulli proposed that decisions under uncertainty should be based not on expected monetary value, but on what he called *moral expectation*, which we today refer to as expected utility.

Bernoulli addressed the paradox by introducing the concept of diminishing marginal utility of wealth, the idea that the utility derived from an additional amount of money decreases as a person's wealth increases. He suggested using a logarithmic utility function, $u(w) = \log(w)$, where w denotes the individual's wealth. Instead of computing the expected monetary gain, Bernoulli proposed evaluating the expected change in utility from participating in the gamble, i.e. the utility gain or loss from the gamble:

$$\mathbb{E}[\Delta u] = \sum_{i=1}^{\infty} \frac{1}{2^i} \log\left(\frac{w + 2^i}{w}\right).$$

This expression captures the expected utility gain from receiving 2^i, taking into account the individual's initial wealth. The series converges to a finite value for all realistic values of initial wealth w (Bernoulli, 1954; Machina, 1987). This utility-based approach laid the foundation for modern expected utility theory, where the logarithmic function is generalized to any increasing and concave utility function $u(w)$, as later formalized by von Neumann and Morgenstern (1947) (Aase, 2001).

2.3 Borch's Model and the Optimal Deductible

This section is based on Karl Borch's article on optimal insurance arrangements (Borch, 1975), which builds on the foundational work by Arrow (1974). Both approaches analyze insurance contracts under the assumption that individuals maximize expected utility.

Consider a setting where an individual faces a random loss X, described by a cumulative distribution function $F(x) = P(X \leq x)$ for $x \geq 0$. The individual may purchase an insurance contract that provides compensation $y(x) > 0$ in the event of a loss x, and the objective is to find the function $y(x)$ that maximizes expected utility. The individual's preferences toward risk are represented by a Bernoulli utility function $u(x)$, and they possess initial wealth w.

The premium for the insurance is defined by a functional $P(y)$, which is assumed to be proportional to the expected compensation, adjusted for a loading λ:

$$P(y) = (1+\lambda)\int_0^\infty y(x)\,\mathrm{d}F(x) = (1+\lambda)\mathbb{E}[y(X)].$$

The optimization problem is to choose a contract $y(X) \in \mathbf{Y}$, where $\mathbf{Y}$ denotes the set of feasible insurance policies, in order to maximize the expected utility:

$$\int_0^\infty u(w - P(y) - x + y(x))\,\mathrm{d}F(x) = \mathbb{E}[u(w - P(y) - X + y(X))],$$

where $w - P(y) - x + y(x)$ represents the individual's wealth after paying the insurance premium $P(y)$, experiencing the loss x, and receiving the insurance compensation $y(x)$. This expression captures the net wealth available to the individual in each loss scenario, which determines their utility. As shown by Arrow (1974), the optimal contract that maximizes expected utility under this premium structure takes the form:

$$y(x) = \begin{cases} 0, & x < m, \\ x - m, & x \geq m, \end{cases}$$

for some deductible m. This contract structure corresponds to a stop-loss insurance policy, i.e. the insured is responsible for all losses up to m, and is fully compensated for any amount exceeding that threshold. Given this form of $y(x)$, the premium can be written as a function of m:

$$P(m) = (1+\lambda)\int_m^\infty (x-m)\,\mathrm{d}F(x) = (1+\lambda)\mathbb{E}[(X-m)_+]. \tag{2}$$

For notational convenience, we write P instead of $P(m)$ in what follows, implicitly treating P as a function of m. Hence, when differentiating expressions involving P, we account for the fact that $P = P(m)$.

As shown by Borch (1975), the optimization problem thus reduces to determining the optimal deductible m. The expected utility associated with such a contract is given by:

$$U(m) = \int_0^m u(w - P - x)\,\mathrm{d}F(x) + \int_m^\infty u(w - P - m)\,\mathrm{d}F(x).$$

To characterize the optimal deductible, we differentiate $U(m)$ with respect to m, accounting for the fact that P also depends on m. The derivative of the premium is:

$$\frac{\mathrm{d}P}{\mathrm{d}m} = -(1 + \lambda)(1 - F(m)). \tag{3}$$

Applying the chain rule yields:

$$\begin{aligned}\frac{\mathrm{d}U}{\mathrm{d}m} = & -\frac{\mathrm{d}P}{\mathrm{d}m}\int_0^m u'(w - P - x)\,\mathrm{d}F(x) \\ & -\left(1 + \frac{\mathrm{d}P}{\mathrm{d}m}\right) u'(w - P - m)(1 - F(m)).\end{aligned}$$

A full derivation of $\frac{\mathrm{d}P}{\mathrm{d}m}$ and $\frac{\mathrm{d}U}{\mathrm{d}m}$ is provided in the Appendix, see Section 7. Setting $\frac{\mathrm{d}U}{\mathrm{d}m} = 0$ and substituting Equation (3), we obtain the condition characterizing the optimal deductible:

$$(1 + \lambda)\int_0^m u'(w - P - x)\,\mathrm{d}F(x) = [(1 + \lambda)F(m) - \lambda]\, u'(w - P - m). \tag{4}$$

This equation implicitly defines the optimal m. Furthermore, under the standard assumption of strict risk aversion, i.e. $u''(x) < 0$, the optimal deductible increases with the loading factor λ, as shown by Borch (1975).

2.4 Key Distributions in Actuarial Science

2.4.1 The Gamma Distribution

The Gamma distribution is a continuous probability distribution commonly used in insurance mathematics to model non-negative stochastic quantities, such as claim sizes. A random variable $X \sim \text{Gamma}(\alpha, \rho)$ has the probability density function (PDF):

$$f(x) = \frac{\rho^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\rho x}, \quad x > 0,$$

where $\alpha > 0$ is the shape parameter and $\rho > 0$ is the rate parameter. Note that in some formulations, the scale parameter $\theta = 1/\rho$ is used instead (Kaas et al., 2008). The term $\Gamma(\alpha)$ in the denominator refers to the gamma function, which is defined for real $\alpha > 0$ as:

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx.$$

The mean and variance of a Gamma-distributed random variable are given by α/ρ and α/ρ^2, respectively (Ohlsson and Johansson, 2010).

2.4.2 The Poisson Distribution

The Poisson distribution is a discrete probability distribution frequently used in actuarial science to model event frequency, such as the number of insurance claims within a fixed time period. Let $N \sim \text{Poisson}(\kappa)$, where $\kappa > 0$ denotes the rate parameter, representing the expected number of events in the given interval. The probability mass function (PMF) of the Poisson distribution is given by:

$$\mathbb{P}(N = k) = \frac{\kappa^k e^{-\kappa}}{k!}, \quad k = 0, 1, 2, \ldots$$

A key property of the Poisson distribution is that its expected value and variance are both equal to κ (Ohlsson and Johansson, 2010; Kaas et al., 2008).

2.4.3 The Compound Poisson Distribution

The Compound Poisson distribution is a discrete-continuous mixture distribution commonly used in insurance mathematics. It describes the total amount of claims occurring within a fixed period of an insurance contract, where the number of claims is random and each individual claim amount follows its own probability distribution.

Formally, let $N \sim \text{Poisson}(\kappa)$ denote the number of claims during a fixed time interval, and let $X_1, X_2, \ldots$ be independent and identically distributed non-negative random variables representing individual claim amounts. Assume further that the claim sizes X_i are independent of the claim count N. The total claim amount Z is then defined as:

$$Z = \sum_{i=1}^{N} X_i.$$

The distribution of Z is referred to as the Compound Poisson distribution. Under the assumption of independence, the expected value and variance of Z are given by:

$$\mathbb{E}[Z] = \mathbb{E}[N] \cdot \mathbb{E}[X], \quad \mathrm{Var}(Z) = \mathbb{E}[N] \cdot \mathbb{E}[X^2],$$

where $\mathbb{E}[N] = \kappa$. While the gamma distribution is commonly used to model claim severity X, other possible choices include the log-normal and inverse Gaussian distributions (Kaas et al., 2008; Ohlsson and Johansson, 2010).

2.5 Generalized Linear Models

Generalized Linear Models (GLMs) are a class of statistical models that extend the classical linear model by (i) allowing the stochastic variables to follow distributions other than the normal distribution, such as Poisson, Binomial, and Gamma, and (ii) modeling the mean of the response variable as a linear function on a different scale, such as the logarithmic scale. This transformation leads to a multiplicative model, in contrast to the additive model used in classical linear regression (Kaas et al., 2008). For insurance pricing, multiplicative models are more suitable since claim amounts and claim frequencies are often modeled with non-normal distributions, such as Poisson, binomial, or gamma (Ohlsson and Johansson, 2010).

A Generalized Linear Model (GLM) is composed of three main components: a stochastic component, a systematic component, and a link function. The stochastic component defines the probability distribution of the response variable U, which lies within the exponential family. The systematic component represent the linear predictor η. This is a linear function of the regressors $V_1, V_2, \ldots, V_p$, typically written as:

$$\eta = \beta_0 + \beta_1 V_1 + \beta_2 V_2 + \cdots + \beta_p V_p$$

where $\beta_0, \beta_1, \ldots, \beta_p$ are the parameters that are estimated during model fitting. The link function, denoted as $g(\cdot)$, connects the expected value of the response variable $\mathbb{E}(U)$ to the linear predictor η. The relationship is expressed as (Kaas et al., 2008):

$$g(\mathbb{E}(U)) = \eta$$

and must be a monotone and differentiable function. A common link function is the logarithmic link $g(\mathbb{E}(U)) = \log(\mathbb{E}(U))$ which gives us a multiplicative model and, as stated before, is more suitable for insurance pricing. As mentioned in Section 2.1, we have a claim severity and a claim frequency model, for a GLM we assume that the claim frequency follows a poisson distribution and often a gamma distribution for the claim severity, and both a logarithmic link (Ohlsson and Johansson, 2010).

2.6 K-Means Clustering

K-means clustering is a fundamental unsupervised learning method used to partition a dataset $(\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_p)$, where each observation $\mathbf{q}_n \in \mathbb{R}^D$, into K distinct clusters. The absence of labeled responses distinguishes this problem as unsupervised, the algorithm seeks to discover intrinsic groupings within the data based solely on the observed feature vectors.

Formally, the goal of K-means is to determine both an assignment of each data point to one of the K clusters and a corresponding set of cluster centroids $\{\mathbf{c}_1, \ldots, \mathbf{c}_K\} \subset \mathbb{R}^D$ that minimize the within-cluster sum of squared Euclidean distances. Introducing indicator variables $r_{nk} \in \{0, 1\}$ such that:

$$r_{nk} = \begin{cases} 1, & \text{if } \mathbf{q}_n \text{ is assigned to cluster } k, \\ 0, & \text{otherwise,} \end{cases}$$

with the constraint $\sum_{k=1}^{K} r_{nk} = 1$ for all n, ensuring that each data point is assigned to exactly one cluster, the objective function becomes:

$$J = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \|\mathbf{q}_n - \mathbf{c}_k\|^2.$$

The algorithm proceeds iteratively via two alternating steps. First, we initialize the cluster centroids, commonly by randomly selecting K points from the dataset. Then the algorithm iterates between the following steps: Each data point is assigned to the cluster whose centroid is closest in Euclidean distance:

$$r_{nk} = \begin{cases} 1, & \text{if } k = \arg\min_j \|\mathbf{q}_n - \mathbf{c}_j\|^2, \\ 0, & \text{otherwise.} \end{cases}$$

Then, each cluster centroid is updated as the mean of all data points currently assigned to that cluster:

$$\mathbf{c}_k = \frac{\sum_{n=1}^{N} r_{nk} \mathbf{q}_n}{\sum_{n=1}^{N} r_{nk}}.$$

These steps are repeated until convergence. Because each phase reduces the value of the objective function J, convergence of the algorithm is guaranteed. Specifically, the update for the cluster centroids $\mathbf{c}_k$ is obtained by minimizing J with respect to $\mathbf{c}_k$, which leads to a closed-form solution by setting the derivative of J with respect to $\mathbf{c}_k$ equal to zero. However, the procedure may converge to a local rather than a global minimum of J (Bishop, 2006).

3 Methods

3.1 Data Description

The dataset used in this study was provided by Länsförsäkringar AB and contains detailed information on insurance contracts related to collision damage for privately owned passenger cars. The dataset includes 3 586 572 insurance policies and consists of 13 variables, of which seven are explanatory variables describing characteristics of either the policyholder or the insured vehicle. These variables include, for example, the age of the policyholder, the age of the car, annual mileage, and vehicle risk zone, but do not contain information on policyholders' wealth or income, nor the market value of the insured vehicles. The dataset also contains information on claims: whether a claim occurred, the number of claims per contract, and the total cost of reported claims. Out of all contracts, 109 070 policies had at least one claim, and the total number of claims was 119 676, indicating that some policies experienced more than one claim during the contract period.

Quantile	95%	99%	99.5%	99.9%
Total cost	0.00	25 607.36	39 482.84	80 511.04

Table 1: Quantiles of total claim costs (SEK)

In Table 1, we can see selected quantiles of the total claim costs (in SEK) reported in the dataset. The table illustrates the highly skewed nature of the claim cost distribution, showing that while most claims are of relatively low cost (with 99% of claims below 25 607 SEK), there are some very large claims, as indicated by the 99.9th percentile at 80 511 SEK.

It is important to note that originally all policyholders had a deductible threshold of 3000 SEK. However, since the aim of this thesis is to estimate the optimal deductible m, any deductible originally applied to the claims was added back to the observed claim costs before any further analysis. This adjustment ensures that the estimations accurately reflect the insurer's full risk exposure rather than only the portion exceeding the deductible. The dataset also contained some observations with zero claim cost despite an indicated claim, typically because the claim amount falls below the deductible threshold. These observations were excluded from the analysis since the actual loss amount is unknown. This limitation is inherent in insurance data, as claims below the deductible are generally not reported, introducing a systematic bias that cannot be fully avoided even with complete data

Furthermore, despite the presence of extreme claim costs, no outliers were removed nor were the data trimmed at any upper percentiles. This decision is motivated by the fact that large claims represent genuine risks that insurers face and should therefore be included in premium calculations to avoid underestimating the insurer's exposure.

3.2 Modeling Individual Wealth

To model individual wealth for all policyholders we assume that a person's monthly income is directly proportional to their wealth. We base our analysis on income data from Statistics Sweden (SCB) for 2024, where the mean monthly salary is 39 900 SEK and the median monthly salary is 35 600 SEK. Since income distributions are typically right-skewed, with a long right tail, we model the distribution of monthly income using a lognormal distribution. Specifically, we assume:

$$W \sim \text{Lognormal}(\mu, \sigma^2)$$

with an expected value $\mathbb{E}[W] = 39\,900$ SEK and a median value $\text{Median}(W) = 35\,600$ SEK. Here, μ and σ are the location and scale parameters of the lognormal distribution, i.e., the mean and standard deviation of the log-transformed income $\ln(W)$. The standard deviation parameter σ determines the dispersion of incomes: a lower σ results in incomes clustered more closely around the median, while a higher σ results in greater variability, with many low-income individuals and a few very high-income earners.

To account for a reasonable degree of income inequality, we estimate μ and σ using the expected value and the median of the lognormal distribution. For a lognormal distribution, the expected value and median are related to μ and σ as follows (Ross, 2014):

$$\mathbb{E}[W] \;=\; \exp\Big(\mu + \tfrac{\sigma^2}{2}\Big)$$

$$\text{Median}(W) \;=\; \exp(\mu)$$

We compute μ from the median and calculate σ using the relationship between the expected value and the median of the lognormal distribution:

$$\mu = \ln\big(\text{Median}(W)\big)$$

$$\sigma = \sqrt{2\ln\left(\frac{\mathbb{E}[W]}{\text{Median}(W)}\right)}$$

By solving for μ and σ using these equations, we obtain $\mu = 10.480$ and $\sigma = 0.478$. Based on these estimates, we simulate the monthly incomes for each policyholder.

3.3 Numerical Solution of Borch's Equation

While Equation (4) provides an implicit condition for the optimal deductible m, it cannot be solved analytically due to the involvement of the cumulative distribution function $F(x)$, as well as the dependency of the premium $P(m)$ on the deductible itself. In particular, $P(m)$ is a function of the expected claims retained by the insurer, which depends on the distribution of the loss amount in excess of the deductible, i.e., $(x - m)_+$.

To overcome this, we adopt a numerical approach. Specifically, we evaluate both sides of Equation (4) over a grid of candidate deductible values M, and select the value $m \in M$ that minimizes the absolute deviation between the two sides. All computations are carried out in `R`, and the loading factor is set to $\lambda = 0.15$, resembling a realistic practice at Länsförsäkringar AB.

We assume a logarithmic utility function $u(x) = \log(x)$, consistent with the Bernoulli utility principle as advocated by Borch (1975). The marginal utility is then given by $u'(x) = 1/x$, which is used in the evaluation of both sides of the condition. For each candidate value of $m \in M$, the corresponding premium $P(m)$ is computed, and both sides of Equation (4) are then evaluated. The optimal deductible is selected as:

$$m^* = \arg\min_{m \in M} |\text{LHS}(m) - \text{RHS}(m)| .$$

This simulation-based framework provides a flexible and robust method for determining optimal deductibles under varying distributional assumptions. In the next section, we implement this procedure under two specific models for the loss distribution $F(x)$: a Gamma distribution and a Compound Poisson distribution with Gamma-distributed severities. As will be shown, the latter better accommodates the structure of the problem, particularly the possibility of zero-loss outcomes, making it more appropriate for the numerical solution of Borch's equation.

3.4 Parameter Estimation and Model Selection

Our initial thought was to assume a Gamma distribution for $F(x)$, since it should describe a setting where an individual faces a random loss x and in insurance mathematics the severity model is often assumed to be Gamma distributed (Ohlsson and Johansson, 2010). However, this specification implies that a loss always occurs and therefore does not account for the possibility of zero claims. To assess whether this simplification is justified, we performed an exploratory analysis on one individual from the dataset, representing the most typical policyholder characterized by the median values of the explanatory variables and median wealth. Specifically, we solved the numerical version of Borch's equation, as outlined in Section 3.3, under two alternative distributional assumptions: (i) a pure Gamma distribution, and (ii) a continuous compound distribution, where the number of claims follows a Poisson distribution and the claim sizes are Gamma distributed.

To numerically solve Borch's equation under these assumptions, we begin by estimating the relevant parameters for both the Gamma and Poisson distributions. This is done by fitting Generalized Linear Models (GLMs) with a logarithmic link function for both a frequency and a severity component, each with four predictors. The predictors include two continuous variables (e.g., age of the policyholder and the vehicle) and two categorical variable (risk zone and mileage). Through these models, we obtain individual estimates for each policyholder i in the dataset, allowing the parameters to reflect each policyholder's specific risk profile rather than relying on general estimates across the entire portfolio.

For the severity component, we fit a Gamma-distributed GLM using only observations with at least one reported claim. The logarithm of the number of claims is included as an offset in the model, enabling us to estimate the expected claim cost per claim for each policyholder. To ensure that the resulting quantity corresponds to the expected cost of a single claim, we generate predictions using a modified version of the dataset in which the number of claims is set to one for all individuals. Consequently, individuals sharing the same combination of explanatory variable classes receive identical predicted mean claim costs μ_i. These predicted means are then used to derive the parameters of the Gamma(α, ρ_i) distribution for each individual. Specifically, the shape parameter α is obtained from the GLM dispersion parameter ϕ, defined by:

$$\alpha = \frac{1}{\phi},$$

and the rate parameter ρ_i is computed for each individual as:

$$\rho_i = \frac{\alpha}{\mu_i},$$

derived from the expected value of the Gamma Distribution (Ohlsson and Johansson, 2010). For the frequency component, we fit a Poisson-distributed GLM using only observations with strictly positive exposure. The logarithm of the duration is included as an offset to model the number of claims per unit time. To obtain the expected annual claim frequency κ_i, we generate predictions on a modified dataset where the exposure duration is set to one for all individuals. Similarly to the Gamma predictions, individuals with identical combinations of explanatory variables receive the same estimated frequency parameters.

To illustrate the application of the numerical solution of Borch's equation under different distributional assumptions, we use the estimated parameters corresponding to the most typical policyholder as defined earlier. These parameters include the individual's wealth w_i, expected claim frequency κ_i, and the Gamma distribution parameters (α, ρ_i) used to model claim severity.

We begin by solving Borch's equation under the assumption that $F(x)$ follows a pure Gamma distribution, using the estimated parameters for individual i, which corresponds to a single-claim severity model. In this case, the cumulative distribution function $F(x)$ is available in closed form, allowing us to directly apply the numerical solution procedure described in Section 3.3.

Under the Compound Poisson assumption, where the total loss Z is the sum of a random number of Gamma-distributed claims, no closed-form expression exists for the cumulative distribution function $F(x)$. To handle this, we approximate $F(x)$ empirically through Monte Carlo simulations. Specifically, we simulate the total loss by first drawing the number of claims N from a Poisson distribution with parameter κ_i, and then summing N independent Gamma-distributed severities with parameters (α, ρ_i). If $N = 0$, the total loss is set to zero, reflecting a year without any claims (Kaas et al., 2008). This process is repeated K times to generate a sample $\{Z^{(1)}, Z^{(2)}, \ldots, Z^{(K)}\}$ of total losses. Here, we set the search space to values $m \leq 0.9 \cdot w$ to ensure strictly positive post-loss wealth and avoid undefined utility values under log utility. A more detailed explanation is given in the discussion section.

The empirical cumulative distribution function at deductible m is then estimated as the proportion of simulated losses not exceeding m, i.e. $F(m) = P(Z \leq m)$, given by:

$$\hat{F}(m) = \frac{1}{K} \sum_{k=1}^{K} \mathbf{1}\{Z^{(k)} \leq m\}.$$

The insurance premium $P(m)$ for a given deductible m is defined as the expected retained loss for the insurer, including the loading factor λ, as expressed in Equation (2). Hence, in this case, we estimate the premium by:

$$\hat{P}(m) = (1 + \lambda) \times \frac{1}{K} \sum_{k=1}^{K} \max(Z^{(k)} - m, 0).$$

In the optimal deductible condition in Equation (4), the left-hand side represents the expected marginal utility of wealth over all losses up to M, weighted by the distribution of losses. We approximate this integral by the empirical average over the simulated losses:

$$(1+\lambda) \times \frac{1}{K} \sum_{k=1}^{K} u'\big(w - \hat{P}(m) - Z^{(k)}\big) \mathbf{1}\{Z^{(k)} \leq m\}.$$

Here, the indicator function $\mathbf{1}\{Z^{(k)} \leq m\}$ ensures that only losses below the deductible contribute to the expectation, and the average is taken over the entire sample to properly approximate the integral with respect to the loss distribution F. This simulation-based approach enables numerical evaluation of the optimal deductible condition despite the absence of closed-form expressions, allowing us to solve for the deductible m.

After computing the optimal deductible m^* and the corresponding premium $P(m^*)$ under both distributional assumptions, we compare the results to determine which loss model provides a better representation for our analysis.

3.5 Calculation and Clustering of Optimal Deductibles

Based on the exploratory analysis presented in Section 3.4, it will be shown that the compound Poisson distribution provides a more realistic model. Accordingly, we adopt the compound Poisson framework when solving Borch's equation for all individuals in the dataset. Although the detailed methodology in Section 3.2 was illustrated for a single representative individual, the same simulation-based procedure is applied to each policyholder i, using their respective estimated parameters $(w_i, \kappa_i, \alpha_i, \rho_i)$. Specifically, for each individual, Borch's equation is numerically solved under the compound Poisson assumption to determine the optimal deductible m_i^*.

After obtaining the set of optimal deductibles, we apply the k-means clustering algorithm to identify representative contract structures, as outlined in Section 2.6, following the methodology inspired by Bishop (2006). Since the number of clusters t is not known a priori, we use the elbow method to determine an appropriate choice of t. This involves running k-means clustering for various values of t and identifying the point at which the marginal reduction in within-cluster sum of squares begins to level off. Based on this criterion, we select an appropriate number of clusters and perform k-means clustering with multiple random initializations in `R` to ensure stability. The resulting cluster centroids represent distinct levels of deductibles, allowing for a meaningful segmentation of the dataset.

3.6 Modeling Wealth Effects

In the baseline simulations, policyholder's initial wealth levels are randomly assigned using a log-normal distribution, as described in Section 3.2. This approach does not incorporate any characteristics of the policyholder or the insured vehicle, which may result in unrealistic allocations of wealth. In practice, a policyholder's wealth is expected to correlate with both individual-specific attributes and the value of the insured asset. Ignoring this relationship reduces the realism of the model and may distort the resulting insurance design. To investigate the implications of this simplification and analyze how wealth influences the choice of optimal deductible m^*, we conduct a targeted simulation study.

We start by calculating the pure premium for each insurance contract using the previously estimated frequency and severity models. For simplicity, instead of a full spectrum of contracts, we select three representative contracts corresponding to the 25th, 50th, and 75th percentiles of the premium distribution. This approach captures low, medium, and high-risk profiles, allowing us to explore how deductible choices vary across different levels of risk exposure while maintaining a streamlined analysis.

For each of these contracts, the optimal deductible is simulated over a range of wealth levels. The optimization procedure for the deductibles follows the same framework as in earlier simulations, using a logarithmic utility function and solving Borch's condition via Monte Carlo simulation. Premiums are updated dynamically to reflect the deductible level and contract-specific risks. The relationship between wealth and the optimal deductible for these three risk profiles is illustrated using line plots.

4 Results

This section presents results from the preliminary analysis, optimization, and the investigation of wealth effects on deductible choice. Section 4.1 covers parameter estimation and the selection of a loss distribution, with the latter illustrated through a representative individual. Section 4.2 shows the optimal deductibles based on the chosen model, followed by clustering results that identify recommended deductible levels for the insurer. Finally, Section 4.3 explores how varying policyholders' wealth impacts the selection of optimal deductibles across different risk profiles.

4.1 Preliminary Analysis

As outlined in the methodology, the parameters of the Poisson(κ_i) and Gamma(α, ρ_i) distributions were estimated via generalized linear models (GLMs) for each unique risk profile. A risk profile here corresponds to a unique combination of explanatory variables. The Gamma shape parameter α is constant across profiles and estimated from the overall dispersion, yielding $\alpha = 1.16$. Since the predictions correspond to claim counts per unit exposure, all individuals sharing the same risk profile receive identical parameter estimates. This allows us to summarize the results in an aggregated table with one row per unique risk profile. Given the multiple classes within the four explanatory variables, presenting all possible combinations is impractical, therefore, Table 2 shows a selected subset of estimated parameters for illustrative risk profiles. The table is sorted by decreasing κ, with the overall highest claim frequencies at the top and the lowest at the bottom. The starred row highlights the most typical policyholder profile.

Owner Age	Vehicle Age	Mileage (km)	Risk Zone	κ	$\rho/10^{-5}$	
18	0	>2500	3	0.44	3.078	
19	0	>2500	3	0.44	3.094	
36	5	≤1000	1	0.11	4.24	
34	7	≤2500	1	0.11	4.41	
48	16	≤2000	3	0.08	6.41	
52	9	≤1500	1	0.074	5.13	*
60	19	≤1500	1	0.03	5.048	
85	20	≤1000	1	0.02	7.15	

Table 2: Estimated parameters κ and ρ from a generalized linear model (GLM) for selected risk profiles, sorted in descending order by κ. The star marks the most typical policyholder profile.

From Table 2, we observe that younger policyholders with new vehicles and high mileage exhibit the highest claim frequencies κ, while older individuals with older vehicles exhibit the lowest. Although variation in the severity parameter ρ is less pronounced, it still reflects meaningful differences in expected claim size across profiles.

The result of evaluating the loss distributions on a single representative individual, referred to as the typical policyholder (marked with $*$ in Table 2) with a median wealth of 35 600, is as follows. Using the estimated parameters for this profile, the Gamma model suggests an optimal deductible of 11 220 and a premium of 15 505, but it fails to provide meaningful results at lower wealth levels. The reason for this will be discussed later. In contrast, the compound Poisson model estimates a deductible of 4 800 and a premium of 1 740, which align better with practical expectations. Therefore, we proceed with the compound Poisson model for the subsequent analyses.

4.2 Optimization Results

Under the compound Poisson assumption, we numerically solved Borch's equation for each policyholder to determine their individual optimal deductible and corresponding premium. Note that even policyholders with identical risk profiles may receive different deductible values due to variations in individual wealth w_i.

To summarize these individualized results into representative contract structures, we applied k-means clustering to all computed optimal deductible values across policyholders. The clustering was based on minimizing the total within-cluster sum of squares, and the optimal number of clusters t was determined using the elbow method, as illustrated in Figure 1. The plot suggests that $t = 3$ is a reasonable choice. However, since it is not uncommon for insurers to offer only two deductible levels, we also report results for $t = 2$.

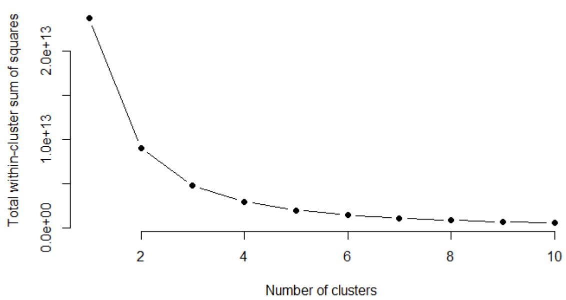

Figure 1: Total within-cluster sum of squares for different number of clusters t.

After constructing contract structures with two and three deductible levels using $k-$means, we obtained the cluster centers shown in Table 3, which are interpreted as representative deductible levels. Exact values are reported without rounding.

Deductible Level	Three-Level Grouping	Two-Level Grouping
Level 1	3,375	4,054
Level 2	6,558	9,180
Level 3	11,691	–

Table 3: Deductible levels for $t = 3$ and $t = 2$.

4.3 Wealth Effects on Deductibles

Figure 2 presents the optimal deductible as a function of individual wealth across three levels of pure premium (25th, 50th, and 75th percentiles). A clear positive relationship is observed, with the optimal deductible increasing nearly linearly with wealth. This suggests that wealthier individuals are generally willing to assume a greater portion of the risk themselves, which aligns with economic theory predicting higher risk tolerance with increased capital.

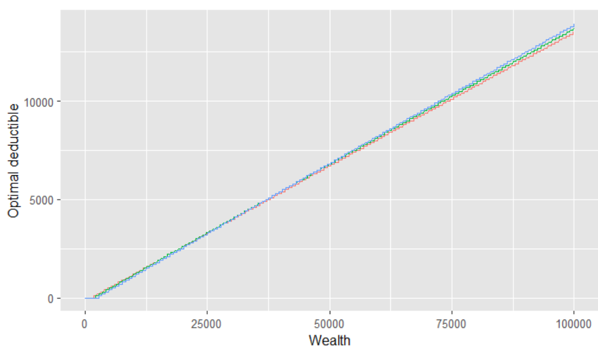

Figure 2: Optimal deductible as a function of wealth for three levels of pure premium. The red line represents the 25th percentile, green the 50th, and blue the 75th.

The variation across the different percentiles is minimal, as indicated by the overlapping curves. This implies that the model's results are relatively robust to variations in parameters such as the pure premium and expected loss, and that wealth is the dominant factor influencing the optimal deductible rather than differences in claim frequency or claim size.

Figure 3 illustrates the insurance premium as a function of the optimal deductible for three levels of pure premium (25th, 50th, and 75th percentiles). The premium decreases as the deductible increases, reflecting that higher deductibles reduce the insurer's exposure and thus the cost of insurance. This relationship mirrors the wealth effect observed in Figure 2, as individuals with greater wealth tend to select higher deductibles and consequently pay lower premiums.

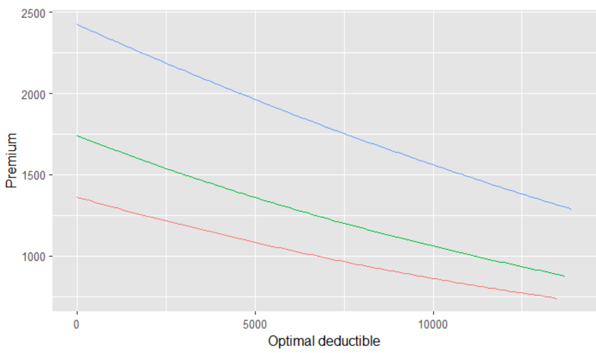

Figure 3: Premium as a function of the optimal deductible, illustrating how the premium varies with deductible levels that depend on wealth. The red line represents the 25th percentile, green the 50th, and blue the 75th.

In Figure 3, the variation across different percentiles is more pronounced. Unlike the optimal deductible, where differences between percentiles are minimal, the premium curves are clearly separated. The 25th percentile consistently pays the lowest premium, the 50th percentile an intermediate amount, and the 75th percentile the highest. While the inverse relationship between deductible and premium holds across all risk groups, the differences in intercepts suggest that factors other than wealth, such as higher expected claim frequency or claim severity, lead to increased expected loss and thus higher premiums. This indicates that although wealth primarily drives deductible choice, risk heterogeneity plays a stronger role in determining the premium.

5 Discussion

This discussion is divided into three parts. In Section 5.1, we interpret the key results obtained in this thesis. Section 5.2 provides a critical reflection on the methodological choices made throughout the work. Finally, in Section 5.3, we outline the main limitations of the study and suggest directions for future improvements.

5.1 Interpretation of Results

This study aimed to determine optimal deductible levels for an insurance portfolio using Borch's utility-based method. While Borch's framework identifies the optimal deductible for a single individual, we extended it to a full portfolio of heterogeneous policyholders to derive common deductible levels.

Based on the results seen in Figure 1, the three-level deductible structure appears to offer the most appropriate segmentation. The elbow plot of within-cluster sum of squares suggests that additional levels provide limited improvement, and a three-level structure captures most of the explanatory variation without unnecessary complexity.

The resulting deductibles in the three-level structure align reasonably well with real-world products, though they tend to be slightly higher, likely due to uncertainty in wealth estimates affecting the utility calculations. Therefore, results should be interpreted cautiously, especially where reliable wealth data is unavailable. As shown in Section 4.3, wealth has a significant effect on the optimal deductible choice, with deductibles increasing nearly linearly with wealth. Interestingly, the model suggests that different risk profiles receive similar deductible recommendations at the same level of wealth, indicating that wealth dominates deductible preferences. However, this does not imply that wealth and risk are uncorrelated in reality. On the contrary, a policyholder's risk (e.g., claim frequency or severity) may correlate with their wealth. For example, more expensive vehicles may result in costlier claims. Nevertheless, the method holds promise when wealth data or close proxies exist.

A notable pattern, when examining the individual-specific optimal deductibles, is that individuals with very high wealth receive correspondingly high deductibles, reflecting low marginal utility of wealth and near indifference to risk transfer. Applying Equation (1), we found that their expected utility of bearing the risk themselves, $\mathbb{E}[u(w-X)]$, often exceeds the utility of purchasing insurance, $u(w-m-P)$, confirming that self-insurance is preferable. While these extreme cases could be excluded, we retained them to capture the full behavioral spectrum. They also constitute a small fraction of policyholders and have negligible impact on the overall results.

5.2 Methodological Reflections

In this thesis, wealth is a central concept, since it plays a key role in the framework by Borch (1975) for determining the optimal deductible. While Borch defines wealth as an individual's total available assets, particularly liquid assets, we instead use monthly income as a proxy for wealth. This choice is motivated by the idea that individuals are unlikely to use their accumulated assets to cover recurring expenses such as insurance premiums. Instead, monthly income better reflects the financial resources they have at their disposal on a regular basis. Another alternative available to insurance companies could be the value of the insured asset, such as the vehicle's market value, as this information is typically accessible to them. Using it either as a proxy for wealth or as a tool for approximating it could improve the precision of the model. However, we do not have access to such data in this thesis, and this limitation is discussed further in the limitations section.

Wealth is only one component of the framework. Another important modeling choice lies in the selection of an appropriate distribution for claim amounts. As shown in the results, the Compound Poisson-Gamma distribution proved to be the best-fitting model. This outcome is reasonable given the nature of the data: as observed in the data section (see 3.1), a significant proportion of the observations are zero claims. A Gamma distribution alone cannot account for this mass of zeros, whereas the Compound Poisson component naturally incorporates the frequency of zero claims through the Poisson process. This makes the Compound Poisson-Gamma a more realistic choice in modeling insurance losses, where many policyholders may not report a claim within the observed time frame. In contrast, relying solely on a Gamma distribution led to unrealistic results across a wide range of wealth levels, with the optimal deductible frequently converging to zero for many policyholders. An explanation for this behavior is the following:

The Gamma distribution models only positive claim sizes, implicitly assuming a claim occurs in every period. Consequently, the expected loss remains consistently high, resulting in a high premium, as demonstrated by the results. This premium, calculated as:

$$P(m) = (1 + \lambda)\mathbb{E}[(X - m)_+]$$

remains elevated even for substantial deductibles. This occurs because the Gamma distribution has both a long right tail, which implies a non-negligible probability of very large claims, and a significant density near zero, indicating a large number of small claims. Increasing the deductible m excludes these small claims from coverage. However, since their amounts are low, the marginal reduction in expected indemnity $\mathbb{E}[(X - m)_+]$ is minimal. In other words, although a significant portion of claims lies just above zero, the difference $(X - m)_+$ for these claims is close to zero, contributing little to the decrease in expected indemnity as m increases.

As a result, for policyholders with lower wealth, the high premium makes it optimal, under log utility, to choose no deductible, thereby avoiding out-of-pocket costs. Although this behavior is derived from log-utility, similar patterns may occur with other utility functions as well. Investigating alternative utility specifications could therefore be a relevant direction for further analysis. This limitation illustrates why the Gamma distribution does not yield meaningful results across the entire wealth spectrum.

A comparable, but manageable, issue occurred in the simulations using the compound Poisson distribution, where we encountered both numerical and economic difficulties when the deductible m approached or exceeded the initial wealth of the policyholder w. In such cases, the residual wealth:

$$w - P(m) - m$$

which represents the worst-case outcome for the insured, could become negative. This poses a fundamental issue when using a log-utility function, as its marginal utility is only defined for strictly positive wealth. While it is mathematically possible to compute values outside the domain, the results are economically meaningless and lead to misleading behavior in the optimization. Specifically, when residual wealth is negative, the marginal utility:

$$u'(w - P(m) - m) = \frac{1}{w - P(m) - m}$$

becomes a small negative number. This may cause the left- and right-hand sides of the Borch equilibrium condition to appear numerically close, falsely suggesting optimality. To prevent this, we imposed a constraint on the deductible by limiting the search space to values $m \leq 0.9 \cdot w$, ensuring that the policyholder's post-loss wealth remains strictly positive and that the utility function remains well-defined.

5.3 Limitations

This study is subject to several limitations related to both data availability and methodological choices.

First, individual wealth levels were assigned randomly due to the absence of detailed data connecting policyholders to financial characteristics. In particular, the dataset does not contain information on vehicle value, such as purchase price, make, or model, which could have served as a proxy or as a tool when approximating wealth. This simplification may be misleading, as wealth plays a central role in determining the optimal deductible. For instance, policyholders with high-value vehicles may have been assigned unrealistically low wealth, or vice versa, leading to inaccurate estimates of optimal deductibles. This introduces noise by causing a misalignment between assumed wealth and actual financial capacity, ultimately reducing the model's accuracy. Future studies could improve precision by incorporating vehicle characteristics or other relevant factors to better approximate wealth.

Second, the approximation of expected values using Monte Carlo simulation, as described in Section 3.4, introduces some numerical error and randomness in the results. Although a large number of simulations were performed to ensure stability, some variability remains between repeated runs. This stochastic error may affect the precision of the estimated optimal deductibles, especially in scenarios involving extreme values or rare events. Future research could explore variance reduction techniques such as antithetic or control variates, or alternative numerical methods like Panjer recursion. However, while Panjer recursion can give exact results for certain aggregate claim distributions, it generally requires discrete claim sizes and can be computationally demanding or impractical when applied to continuous or complex severity distributions such as the Gamma (Dickson, 1995).

Lastly, in determining deductible levels across the population, the k-means clustering algorithm was applied using the optimal deductible values themselves as inputs, and the within-cluster sum of squares as the criterion. An alternative approach would be to cluster based on individual utility values instead, potentially leading to more utility-optimal segmentation. However, since the utility function is already incorporated into the optimization procedure used to determine each individual's deductible, we reasoned that clustering on the deductible values indirectly reflects the underlying utility.

6 Conclusion

This thesis applied Borch's utility-based framework to determine individuals' optimal deductibles, and expanded it to obtain optimal deductible levels for a heterogeneous insurance portfolio using k-means clustering. While Borch defines wealth as an individual's total available assets, we used simulated monthly income as a proxy for wealth.

References

Knut K. Aase. On the st. petersburg paradox. *Scandinavian Actuarial Journal*, 2001(1):69–78, 2001. doi: 10.1080/034612301750077356.

Kenneth J. Arrow. Optimal insurance and generalized deductibles. *Scandinavian Actuarial Journal*, (1):1–42, 1974. doi: 10.1080/03461238.1974.10408659.

Daniel Bernoulli. Exposition of a new theory on the measurement of risk. *Econometrica*, 22(1):23–36, 1954. URL `http://www.jstor.org/stable/1909829`.

Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer Science+Business Media, New York, NY, USA, 2006. ISBN 978-0-387-31073-2. URL `https://www.springer.com/gp/book/9780387310732`.

Karl Borch. Optimal insurance arrangements. *ASTIN Bulletin*, 8(3):284–290, 1975. doi: 10.1017/S0515036100011223.

D. C. M. Dickson. A review of panjer's recursion formula and its applications. *British Actuarial Journal*, 1(1):107–124, 1995. URL `https://www.jstor.org/stable/41141204`.

Rob Kaas, Marc Goovaerts, Jan Dhaene, and Michel Denuit. *Modern Actuarial Risk Theory*. Springer, Heidelberg, 2008. ISBN 978-3-540-70992-3. doi: 10.1007/978-3-540-70998-5.

Mark J. Machina. Choice under uncertainty: Problems solved and unsolved. *The Journal of Economic Perspectives*, 1(1):121–154, 1987. URL `http://www.jstor.org/stable/1942952`.

Esbjörn Ohlsson and Björn Johansson. *Non-Life Insurance Pricing with Generalized Linear Models*. Springer, Heidelberg, 2010. ISBN 978-3-642-10790-0. doi: 10.1007/978-3-642-10791-7.

Murray H. Protter and Charles B. Jr. Morrey. *Intermediate Calculus*. Springer, New York, second edition, 1985. ISBN 978-0-387-96058-6. doi: 10.1007/978-1-4612-1086-3.

Sheldon M. Ross. *Introduction to Probability Models*. Academic Press, Boston, 11 edition, 2014. ISBN 9780124079489.

Dennis D. Wackerly, William Mendenhall, and Richard L. Scheaffer. *Mathematical Statistics with Applications*. Cengage Learning, Boston, MA, 8 edition, 2021. ISBN 9780357026554.

7 Appendix

Derivation of the Premium and Utility

In this section, we provide the full derivation of $\frac{\mathrm{d}P}{\mathrm{d}m}$ and $\frac{\mathrm{d}U}{\mathrm{d}m}$, as used in the Borch model described in Section 2.3. To differentiate integrals with variable limits or parameters, we apply Leibniz's integral rule, which states (Protter and Morrey, 1985):

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}m}\left(\int_{a(m)}^{b(m)} f(m,x)\,\mathrm{d}x\right) &= f(m,b(m))\cdot b'(m) - f(m,a(m))\cdot a'(m)\\ &\quad + \int_{a(m)}^{b(m)} \frac{\partial}{\partial m} f(m,x)\,\mathrm{d}x.\end{aligned} \tag{5}$$

We start by recalling the expression for the premium P, which depends on the deductible m:

$$P(m) = (1+\lambda)\int_m^\infty (x-m)\,\mathrm{d}F(x) = (1+\lambda)\int_m^\infty (x-m)f(x)\,\mathrm{d}x,$$

where $f(x)$ is the probability density function associated with the cumulative distribution function $F(x)$, i.e. $f(x) = \frac{\mathrm{d}F}{\mathrm{d}x}$. We now apply Equation (5) to the integral. Let:

$$\begin{aligned} &f(m,x) = (x-m)f(x),\\ &a(m) = m \quad \Rightarrow \quad a'(m) = 1,\\ &b(m) = \infty \quad \Rightarrow \quad b'(m) = 0.\end{aligned}$$

Note that the density $f(x)$ does not depend on m. Then we obtain:

$$\frac{\mathrm{d}P}{\mathrm{d}m} = (1+\lambda)\left[0 - f(m,m)\cdot 1 + \int_m^\infty \frac{\partial}{\partial m}\big((x-m)f(x)\big)\,\mathrm{d}x\right].$$

We compute the terms. For the first one, we have $f(m,m) = (m-m)f(m) = 0$,. For the latter term, we compute:

$$\frac{\partial}{\partial m}\big((x-m)f(x)\big) = -f(x),$$

since $f(x)$ does not depend on m, and the derivative of $x-m$ with respect to m is -1. Substituting into the expression, we obtain:

$$\begin{aligned}\frac{\mathrm{d}P}{\mathrm{d}m} &= (1+\lambda)\int_m^\infty (-f(x))\,\mathrm{d}x\\ &= -(1+\lambda)\int_m^\infty f(x)\,\mathrm{d}x.\end{aligned}$$

We now simplify the integral:

$$\int_{m}^{\infty} f(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} f(x)\,\mathrm{d}x - \int_{-\infty}^{m} f(x)\,\mathrm{d}x$$
$$= 1 - F(m),$$

where we have used the fact that the total area under a probability density function is 1, and that (Wackerly et al., 2021):

$$F(m) = \int_{-\infty}^{m} f(x)\,\mathrm{d}x.$$

Hence, we conclude:

$$\frac{\mathrm{d}P}{\mathrm{d}m} = -(1+\lambda)(1-F(m)),$$

which matches the expression presented in Borch (1975). We now proceed to derive the full expression for $\frac{\mathrm{d}U}{\mathrm{d}m}$. Recall the definition of $U(m)$:

$$U(m) = \int_{0}^{m} u(w-P-x)\,\mathrm{d}F(x) + \int_{m}^{\infty} u(w-P-m)\,\mathrm{d}F(x).$$

We denote the first integral by $U_A(m)$ and the second by $U_B(m)$, and compute their derivatives separately. To compute $\frac{\mathrm{d}U_A}{\mathrm{d}m}$, we apply Leibniz's rule as stated in Equation (5). Define:

$$\begin{aligned}
&f(m,x) = u(w-P-x)f(x),\\
&a(m) = 0 \quad \Rightarrow \quad a'(m) = 0,\\
&b(m) = m \quad \Rightarrow \quad b'(m) = 1.
\end{aligned}$$

Hence, the derivative becomes:

$$\frac{\mathrm{d}U_A}{\mathrm{d}m} = f(m,m)\cdot 1 + \int_{0}^{m} \frac{\partial}{\partial m}\left(u(w-P-x)f(x)\right)\mathrm{d}x.$$

Here, the first term is $f(m,m) = u(w-P-m)f(m)$, and for the integral term, we apply the chain rule to account for the m-dependence in P:

$$\frac{\partial}{\partial m}(u(w-P-x)f(x)) = u'(w-P-x)\frac{\mathrm{d}}{\mathrm{d}m}(w-P-x)f(x)$$
$$= -u'(w-P-x)\frac{\mathrm{d}P}{\mathrm{d}m}f(x).$$

Substituting into the expression gives:

$$\frac{\mathrm{d}U_A}{\mathrm{d}m} = u(w - P - m)f(m) - \frac{\mathrm{d}P}{\mathrm{d}m}\int_0^m u'(w - P - x)f(x)\,\mathrm{d}x.$$

We now differentiate the second term $U_B(m)$. Again, applying Equation (5), define:

$$\begin{aligned} &f(m, x) = u(w - P - m)f(x),\\ &a(m) = m \quad \Rightarrow \quad a'(m) = 1,\\ &b(m) = \infty \quad \Rightarrow \quad b'(m) = 0. \end{aligned}$$

Hence, the derivative becomes:

$$\frac{\mathrm{d}U_B}{\mathrm{d}m} = -f(m, m)\cdot 1 + \int_m^\infty \frac{\partial}{\partial m}\big(u(w - P - m)f(x)\big)\,\mathrm{d}x.$$

The first term is $f(m, m) = -u(w - P - m)f(m)$, and the integrand now depends on m via both P and m. Hence:

$$\begin{aligned} \frac{\partial}{\partial m}(u(w - P - m)f(x)) &= u'(w - P - m)\frac{\mathrm{d}}{\mathrm{d}m}(w - P - m)f(x)\\ &= -u'(w - P - m)\left(1 + \frac{\mathrm{d}P}{\mathrm{d}m}\right)f(x). \end{aligned}$$

Thus:

$$\frac{\mathrm{d}U_B}{\mathrm{d}m} = -u(w - P - m)f(m) - \left(1 + \frac{\mathrm{d}P}{\mathrm{d}m}\right)u'(w - P - m)\int_m^\infty f(x)\mathrm{d}x,$$

where we have moved out $\left(1 + \frac{\mathrm{d}P}{\mathrm{d}m}\right)u'(w - P - m)$ from the integral since they do not depend on x. As noted previously when differentiating P, we can use the identity:

$$\int_m^\infty f(x) = 1 - F(m).$$

Adding the expressions for $\frac{\mathrm{d}U_A}{\mathrm{d}m}$ and $\frac{\mathrm{d}U_B}{\mathrm{d}m}$, we obtain:

$$\frac{\mathrm{d}U}{\mathrm{d}m} = -\frac{\mathrm{d}P}{\mathrm{d}m}\int_0^m u'(w-P-x)\mathrm{d}F(x) - \left(1 + \frac{\mathrm{d}P}{\mathrm{d}m}\right)u'(w-P-m)(1-F(m)).$$

This completes the derivation of the premium and utility derivatives with respect to the deductible m, confirming the results in Borch (1975). Note that the boundary terms $f(m, m)$ from the two parts cancel each other out.