

On the cost of risk misspecification in insurance pricing

D. Finger* H. Albrecher† L. Wilhelmy‡

Abstract. In the non-life insurance industry, pricing is often done relative to individual criteria of policyholders. Various classification algorithms are in use to categorize policyholders into risk classes defined by the insurer, but classification errors may result from this process. In the light of recent automatic classification practices, it becomes important to assess the risks caused by such errors. In this paper we examine the impact of risk class misspecifications for a simple situation with two risk types. We provide a mean-variance framework for quantitatively studying the insurer’s optimization problem of specifying premiums and we analyze the tradeoff of costs and benefits when classification error probabilities are known.

1 Introduction

Risk classification is a classical tool in actuarial practice. Indeed, the distribution of individual risks will often differ substantially, depending on personal characteristics, different exposure, environmental conditions etc. If an insurer applies the same premium across all such categories, this may lead to adverse selection, in the sense that individuals who face a lower premium than appropriate for their true risk will massively enter the contract, raising the price and squeezing out rational individuals with lower risk (see e.g. [4] for a general discussion). Classification is commonly used in insurance during the underwriting process and the tarification, cf. [32, 42] for health and life insurance and [43] for property lines. Different types of variables, such as quantitative or categorical can be used in various algorithms for actuarial classification purposes [38]. For an overview of classification methods we refer to [21]. In the literature, numerous authors discuss risk classification in view of adverse selection and efficiency in the Pareto sense, see [13] for a survey on this topic. In [11, 12, 22] authors study the efficiency of imperfect categorization in a utility setup. In [14, 34], authors consider the effect of bans on classification on the market efficiency. [39] deals with the customers’ perspective on adverse selection and efficiency of risk classification. The contributions [11, 34] consider also costs of categorization. Indeed, risk classification may be costly both computationally (in terms of resource-consuming algorithms in the presence of huge datasets) and

*Department of Actuarial Science, Faculty of Business and Economics, University of Lausanne, CH-1015 Lausanne. Email: dina.finger@unil.ch

†Department of Actuarial Science, Faculty of Business and Economics, University of Lausanne, CH-1015 Lausanne and Swiss Finance Institute. Email: hansjoerg.albrecher@unil.ch

‡Swiss Re, Mythenquai 50/60, CH-8002 Zurich. Email: Lutz_Wilhelmy@swissre.com

in monetary terms (e.g. involving data acquisition from official statistics or from competitors).

The insurer hence needs to select appropriate criteria for rating variables, which may take different forms and are not always motivated by actuarial drivers. For instance, [17] distinguishes actuarial, operational, social and legal criteria. In some cases, classification may be inaccurate when the main criterion is unobservable or there are legal constraints to use it, e.g. for social or political reasons. For instance, gender-based discrimination is nowadays forbidden in the European Union (see e.g. [36]), even if it is widely considered a relevant characteristic from the statistical point of view. In [37], authors discuss the perceptions of gender as pricing criteria by the customer and explain the complex interplay between anti-discrimination laws and actuarial principles within the insurance industry, see [27] for suggestions to deal with this issue. We refer to the recent book [9] for a rich source of information and ideas on this topic.

Although criteria to distinguish risk profiles will exist, they may often be unknown to the insurer due to the asymmetry of information, see e.g. [1, 3]. At the same time, recent years have seen a dramatic increase in both the amount of risk information available and the ability to analyse it statistically. In many situations, risks can be analysed on an almost individual basis, but in any case as elements of much smaller rating pools, i.e. pools of risks that share characteristics such that they can be assumed to have the same loss distribution. Offered insurance cover may also take into account the current risk situation, i.e. environmental variables (e.g. time, place) or even behavioural variables. The small size of rating pools leads to larger estimation errors in estimating the expected cost of insurance cover. In the presence of competing insurance providers and given the transparency of the prices of their insurance offers, customers may tend to choose the cheapest offer, i.e. an offer that is too low compared to the true but unknown production costs. From the perspective of the insurance company, this phenomenon is known as the "winner's curse". In order to avoid the negative economic consequences of the winner's curse, insurance companies like to apply tailor-made surcharges to the offered premium. It is likely that these surcharges, in addition to the higher cost of more granular risk assessment, will lead to higher overall costs for the entire insured portfolio of risks. Hence, the total welfare of the community is reduced. In addition, the overall coverage ratio may be reduced as some risk owners may find the increased cost of insurance too high. An overly granular approach may therefore be counterproductive and there may be an optimal level of rating granularity, together with legal boundaries, see for instance [17], [2] and [40].

Even when the resulting loss distributions are assumed to be known, each classification method will inherently contain classification errors, which may lead to unprofitable decision making. Note that error rates tend to be higher when one class is less represented in the population, cf. [24, 44]. For a classical reference to empirical evidence on concrete values for error probabilities arising from frequently used statistical methods, see e.g. [6]. One can also apply empirical methods for the estimation of the misclassification probabilities. For instance, an empirical estimate arises from historical records, when after the loss occurrences, one obtains more information on the risk type of the policyholders. The insured individual can be reclassified using, for example, maximum likelihood techniques, and thus proportions

of misclassified individuals can be identified.

In this paper, we would like to quantitatively study the effects of such classification errors in the context of a simple model with only two possible risk classes. In that case one faces two types of errors: assigning a risk of the first class as one of the second and vice versa (a *false positive* and *false negative* in statistical terms, or also *sensitivity* vs. *specificity*, see e.g. [24, 42]). We compare three scenarios, one where the insurer does not classify the heterogeneous risks and two others where the risks are differentiated, but once with perfect knowledge and once with some probability of misspecification of policyholder types. While the introduction of the classification mechanism allows to price the insurance risks more efficiently, it entails certain fixed costs, and the potential misspecification can impact the insurer's profit further. In [8], the authors study a similar framework with incomplete information from the insurer's side, whereas the customer knows their true risk type. During the underwriting process, the customer may be required to take a test and the result will reveal for what type of coverage they are eligible. In contrast to our present work, the authors consider a plurality of insurers and customers choosing their coverages, while the present paper deals with a one-insurer setting and customers being provided with one single offer. In a similar spirit to [18], we develop a simple framework to assess the respective trade-off between costs of classification and profits. We consider linear and sigmoid-type demand functions of the premium for the probability that a customer accepts an offered policy (rather than deterministic demand functions as in [18]). Our main optimization target is the expected profit for the insurer (see e.g. [25]), and we consider several risk measures to assess the risk part in the analysis.

The remainder of the paper is structured as follows: In Section 2, we start with an insurer's expected profit approach and a piece-wise linear demand setting. We then consider different scenarios for this setup in Section 3. In Section 4, we then include the variance in our considerations, and establish mean-variance frontiers for the profit. We illustrate the results and its main drivers in Section 5. In order to assess the sensitivity of the results, Section 6 then develops a number of extensions, namely a sigmoid-type rather than piece-wise linear demand function, a lower semi-variance and a value-at-risk concept for replacing the variance in the risk assessment of the strategies, as well as a utility function approach to unite the consideration of profitability and variability in one function, together with numerical illustrations for each of these cases. Finally, Section 7 concludes.

2 Model Setting

Assume that there is only one insurer present in the market, so there is no competition. Let us further assume a population of n individuals, who are all willing to contract insurance, and independently from each other choose whether they enter the insurance contract for a given premium or not. The individuals fall into two types: low risk type with loss random variable L and high risk type with loss random variable H , with underlying cumulative loss distribution functions $F_L(x)$ and $F_H(x)$, $x \geq 0$ respectively (both L and H will typically have atoms at 0, signifying the case of no claim in the considered time period). Define the respective means

and variances by

$$\mathbb{E}(L) = \mu_L, \quad \mathbb{V}\text{ar}(L) = \sigma_L^2, \quad (1)$$

$$\mathbb{E}(H) = \mu_H, \quad \mathbb{V}\text{ar}(H) = \sigma_H^2, \quad \mu_H > \mu_L, \quad (2)$$

which are all assumed to be finite. All risks are assumed to be independent and identically distributed within each type. Let $\{p_L, p_H \geq 0\}$ be the actual proportion of the low- and high-risk type among the policyholders ($p_L + p_H = 1$).

Define an acceptance function f_i which for any proposed premium P gives the probability for an individual to enter the contract; the form of this function differs for each risk type i . Each individual is assumed to take the decision about entering independent of all the others. If m individuals are offered a premium P and all use the same acceptance function f_i , we then expect $m \cdot f_i(P)$ individuals to enter the contract. Let us first assume that f_i is piece-wise linear

$$f_i(P) = \begin{cases} 1 - \frac{P}{P_i^{max}}, & 0 \leq P \leq P_i^{max}, \quad i \in \{L, H\} \\ 0, & P_i^{max} < P, \end{cases} \quad (3)$$

where P_i^{max} is the so-called *reservation price*, cf. Figure 1. Clearly, f_i is non-increasing in P . We set the condition $\mu_i < P_i^{max}$ to ensure the possibility of positive expected profits for the company.

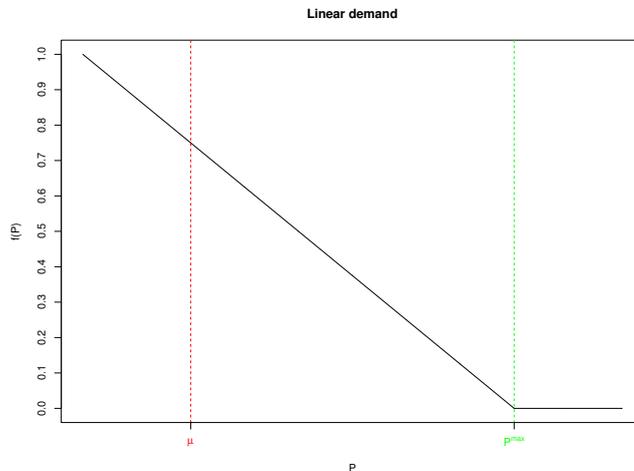


Figure 1: Acceptance probability $f_i(P)$ as a function of offered premium P .

Our approach represents a stochastic setup for the acceptance of an offered contract, in contrast to [18] who work with (3) as a deterministic demand function for a given price. While for expected profits as dealt with in Section 3 this difference does not matter, it will be important for the risk considerations in the subsequent sections. Despite the non-differentiability of the piece-wise linear f_i at P_i^{max} , this form will lead to simple local solutions. An extension of the results to a more complex, but analytically better tractable sigmoid form will be considered in the Appendix.

Remark 2.1. In the classical work of Rothschild and Stiglitz [35] as well as further models based on it, in a comparable setting of identifying contracts to sell, the

authors find optimal solutions in terms of premiums and levels of coverage. For the present purpose and simplicity we, however, prefer to use the concept of the acceptance function, with individuals facing a binary choice of entering the contract or not. That is, we do not allow a partial coverage or deductibles here.

In the following sections, we introduce three different scenarios that the insurer may face. We start with the case where the insurer can observe the risk type of each individual, which we refer to as the full information case. We then consider the situation where the insurer can not distinguish between the two types ex-ante at all. In that case the only possible method of pricing is to not differentiate individuals, and the profitability of the insurance business will then depend on the empirical fraction of each risk type in the population. Finally, the possibility to observe and measure a certain characteristic, which can be discrete or continuous, allows us to classify an observation, but with a certain error probability. This probability depends on the true class of the observation. The introduction of the classification mechanism has certain costs, but allows to better price according to the true risk class. We are interested to quantitatively assess the respective trade-off in this simple model setup.

3 Expected profit in three scenarios

In this section, we focus on the expected profit only. Let us introduce three different scenarios that the insurer may face, starting with the case where the insurer can actually observe the risk type of each individual.

3.1 Full information

The benchmark for our analysis is the situation with no asymmetry of information. Here the insurer can observe the risk type of each individual and therefore price according to the true type. Recall that we know the actual proportion $\{p_L, p_H\}$ of the population in each class. Therefore, we can maximize profit by differentiating between groups. If, for a given individual, the price is higher than its true risk premium, his/her willingness to accept the contract decreases. Let us denote by $X_j^{(L)}$ the j^{th} loss random variable of the low risk type (independent copies of L) and by $X_j^{(H)}$ the j^{th} high-risk loss variable (independent copies of H). Then, the (random) profit is given by

$$\Pi = \sum_{j=1}^{n \cdot p_L} I_j^L (P_L - X_j^{(L)}) + \sum_{j=1}^{n \cdot p_H} I_j^H (P_H - X_j^{(H)}), \quad (4)$$

where I_j^L and I_j^H are independent Bernoulli random variables with probabilities $f_L(P_L)$ and $f_H(P_H)$, respectively. In this case, an adaptation of [18] establishes that the optimal premiums are independent of n , p_L and p_H , and they are simply the average of the mean claim size and the maximum premium that the policyholders are willing to pay.

Theorem 3.1. *In the full information case, the expected profit of the insurer is maximized by the premium choice*

$$P_L = \frac{1}{2}(\mu_L + P_L^{max}) \quad (5)$$

and

$$P_H = \frac{1}{2}(\mu_H + P_H^{max}) \quad (6)$$

for the two risk classes.

Proof. The optimal premium choice is the solution of the following optimization problem:

$$\{P_L, P_H\} = \underset{x,y}{\operatorname{argmax}} \mathbb{E}(\Pi) = \underset{x,y}{\operatorname{argmax}} np_L f_L(x)(x - \mu_L) + np_H f_H(y)(y - \mu_H). \quad (7)$$

We notice that $\mathbb{E}(\Pi)$ is a continuous function of $\{x, y\}$ and that for $\{x < \mu_L, y < \mu_H\}$, $\mathbb{E}(\Pi) < 0$; $\{x = \mu_L, y = \mu_H\}$, $\mathbb{E}(\Pi) = 0$; $\{x > \mu_L, y > \mu_H\}$, $\mathbb{E}(\Pi) > 0$. Also, $\lim_{x \rightarrow +\infty, y \rightarrow +\infty} \mathbb{E}(\Pi) = 0$, which means that $\mathbb{E}(\Pi)$ admits a strictly positive maximum for some $\{x > \mu_L, y > \mu_H\}$ (and the optimization in x and y can in fact be separated). We can characterize this point by the following equations:

$$\frac{\partial \mathbb{E}(\Pi)}{\partial x} = np_L (f'_L(x)x + f_L(x)) - np_L f'_L(x)\mu_L \stackrel{!}{=} 0 \quad (8)$$

$$\frac{\partial \mathbb{E}(\Pi)}{\partial y} = np_H (f'_H(y)y + f_H(y)) - np_H f'_H(y)\mu_H \stackrel{!}{=} 0, \quad (9)$$

where the $\stackrel{!}{=}$ operator denotes a necessary condition. From (8) we have

$$\begin{aligned} \frac{\partial \mathbb{E}(\Pi)}{\partial x} &= np_L (f'_L(x)x + f_L(x)) - np_L f'_L(x)\mu_L = 0 \\ &\iff np_L \left(-\frac{1}{P_L^{max}} \right) (x - \mu_L) + np_L \left(1 - \frac{x}{P_L^{max}} \right) = 0 \\ &\iff x = \frac{1}{2}(\mu_L + P_L^{max}). \end{aligned}$$

Since the optimal solution P_L respects the condition $P_L \leq P_L^{max}$, as $\mu_L < P_L^{max}$, it is not necessary to distinguish cases of the piece-wise function. To prove it is indeed a local and global maximum, we can easily prove that the second derivative is negative for a linear f :

$$\begin{aligned} \frac{\partial^2 \mathbb{E}(\Pi)}{\partial x^2} &= np_L f''_L(x)(x - \mu_L) + 2np_L f'_L(x) \\ &= 2np_L \left(-\frac{1}{P_L^{max}} \right) < 0. \end{aligned}$$

The same reasoning holds for the first and second derivative w.r.t. y . \square

Remark 3.2. It is easy to check that if

$$\mu_L + P_L^{max} < 2\mu_H, \quad (10)$$

then $P_L < \mu_H$, in which case charging a low-risk type premium to a high-risk type customer results in an expected loss on the individual level. \square

In terms of sensitivities, we simply see from (5) and (6) that

$$\frac{\partial P_i}{\partial \mu_i} = \frac{\partial P_i}{\partial P_i^{max}} = \frac{1}{2}$$

for both risk types $i \in \{L, H\}$. That is, the reactivity of the optimal premium is constant for variation in the mean loss. Consequently, in case of increasing losses, the increase in premium will only cover half of the increase in losses, thus decreasing profits by the double effect of smaller margins and smaller acceptance rate. Similarly, shifting the endpoint P_i^{max} of the acceptance function (for invariant μ_i) also increases the optimal chargeable premium linearly with slope $1/2$.

3.2 No differentiation

Next, let us consider the situation where the insurer has no possibility to distinguish between risk types on the individual level, but still has an estimate for the fractions $\{p_L, p_H\}$ of the population in each class (e.g. through some historical figures). So we assume these numbers to be known ($p_L + p_H = 1$). In this scenario, the insurer proposes an identical premium P to every individual. This has the advantage that one saves the cost of identification of risk types, and provides another benchmark for the sequel. The profit in this case is

$$\Pi = \sum_{j=1}^{np_L} I_j^L(P - X_j^{(L)}) + \sum_{j=1}^{np_H} I_j^H(P - X_j^{(H)}),$$

where I_j^L and I_j^H are Bernoulli random variables with probabilities $f_L(P)$ and $f_H(P)$ respectively, and the optimal premium then amounts to

$$P = \underset{z}{\operatorname{argmax}} \mathbb{E}(\Pi) = \underset{z}{\operatorname{argmax}} n(p_L f_L(z) + p_H f_H(z))z - np_L f_L(z)\mu_L - np_H f_H(z)\mu_H. \quad (11)$$

Comparing the resulting optimization problem

$$\max_z np_L f_L(z)(z - \mu_L) + np_H f_H(z)(z - \mu_H) \quad (12)$$

with (7), we see from $P_L \neq P_H$ (which itself is due to $\mu_H > \mu_L$) there that the optimal solution P to (12) will now yield a smaller profit (this is intuitive, since we have less information available than in the setup of Section 3.1).

Theorem 3.3. *In the no differentiation case, the expected profit of the insurer is maximized by the premium choice*

$$P = a^* P_L + (1 - a^*) P_H, \quad (13)$$

where P_L and P_H are the optimal premiums of the full information case given in (5) and (6) and $a^* = \frac{p_L/P_L^{max}}{p_L/P_L^{max} + p_H/P_H^{max}}$.

Proof. Problem (12) can be solved using the first order condition

$$\begin{aligned} \frac{\partial \mathbb{E}(\Pi)}{\partial z} &= np_L (f'_L(z)z + f_L(z)) + np_H (f'_H(z)z + f_H(z)) \\ &\quad - np_L f'_L(z)\mu_L - np_H f'_H(z)\mu_H \stackrel{!}{=} 0. \end{aligned}$$

Using $z < \min\{P_L^{max}, P_H^{max}\}$, plugging in the function f yields

$$\begin{aligned} np_L \left(-\frac{1}{P_L^{max}} \right) (z - \mu_L) + np_L \left(1 - \frac{z}{P_L^{max}} \right) \\ + np_H \left(-\frac{1}{P_H^{max}} \right) (z - \mu_H) + np_H \left(1 - \frac{z}{P_H^{max}} \right) = 0, \end{aligned}$$

which leads to

$$z = \frac{\frac{p_L}{P_L^{max}} \frac{\mu_L + P_L^{max}}{2} + \frac{p_H}{P_H^{max}} \frac{\mu_H + P_H^{max}}{2}}{\frac{p_L}{P_L^{max}} + \frac{p_H}{P_H^{max}}}$$

and finally (13).

The second order condition yields a strictly negative result, thus confirming the global maximum. \square

Expression (13) shows that P is the average of the optimal premiums under full information, weighted by the proportions in the population and the maximum affordable premiums. Under the assumption (10), this also establishes

$$P_L \leq P \leq P_H.$$

Remark 3.4. One should be careful to check whether $P > P_L^{max}$: in that case, L type customers do not enter the contract. This happens if $a^* < \frac{P_H - P_L^{max}}{P_H - P_L}$. Consequently, the optimal premium is that for higher risk types only, meaning $P = P_H$. If the expected profit for $P = P_H$ is greater than the one found above, then the optimal premium will be P_H and only H types will enter the contract. \square

The change of the expected profit when compared to the case of full information can now also be expressed as

$$\begin{aligned} & \underbrace{np_L (f_L(P) - f_L(P_L)) (P_L - \mu_L)}_{\text{loss on } L \text{ not entering the contract}} + \overbrace{np_L f_L(P) (P - P_L)}^{\text{gain on extra margin on } L} \\ & + \underbrace{np_H (f_H(P) - f_H(P_H)) (P_H - \mu_H)}_{\text{gain on more } H \text{ entering the contract}} + \overbrace{np_H f_H(P) (P - P_H)}^{\text{loss on reduced margin on } H} < 0. \quad (14) \end{aligned}$$

In particular, for low risk types the proposed premium P is higher than their appropriate optimal premium P_L under full information. Thus, with the decreasing shape of the acceptance function, on average the insurer loses low-risk type customers and the associated expected profit (negative first term in (14)). At the same time, those who remain bring higher profits (the second term in (14)). Correspondingly, due to the cheaper than appropriate premium P_H , more high-risk type customers join (positive third term in (14)), but they pay less premium now (negative fourth term).

3.3 Differentiation in two classes

Assume now that the insurer does not know the individuals' risk type, but has access to a mechanism that can assign (*classify*) the risk types correctly with a certain probability. Assume that the probability of misclassification is the same for each policyholder of the same type and given by

$$\begin{aligned} p_{H|L} &:= \mathbb{P}(i \text{ is classified as } H \mid i \in L), \\ p_{L|H} &:= \mathbb{P}(i \text{ is classified as } L \mid i \in H). \end{aligned}$$

Remark 3.5. If $p_{H|L} = p_{L|H} = 0$, we get back to the full information setting, as there is no classification error. If $p_{H|L} = p_{L|H} = 1$, all the true H end up in the L group and all the true L are classified in the H group (which would also result in knowing the true type of each one, but having to switch the categories). In the cases when $p_{H|L} = 1$ and $p_{L|H} = 0$ or $p_{H|L} = 0$ and $p_{L|H} = 1$, all individuals are classified in the same group. Typically, there is a tradeoff between the two error types: in an attempt to classify one risk more accurately, the precision on the other one will go down. For instance, in order to minimize $p_{L|H}$, we could simply attribute all observations to group H , which indeed gives $p_{L|H} = 0$, but $p_{H|L}$ would increase drastically as all L observations are then erroneously identified as H .

The cost $c(n)$ of applying the classification algorithm will increase with population size n (the computational cost of different algorithms is increasing in the sample size (take for instance the simplest Bayesian classifier [45] with linear complexity), the human time invested in analysing data and making decisions increases, and more powerful machines may be needed to run the algorithms, just to name a few reasons). At the same time, the marginal cost is likely to decrease in n (fixed costs in the process can be divided onto more policyholders, the insurer gains experience and recognizes patterns etc.). Hence, we define

$$c : \mathbb{R}^+ \mapsto \mathbb{R}^+, \quad c'(n) \geq 0, \quad c''(n) \leq 0.$$

A mathematically simple candidate for such a function is

$$c(n) = c_0 \log(\gamma n),$$

where γ offsets for the minimal cost amount and c_0 scales for the intensity of the effect of the population size.

The insurer will propose premiums, P_L^* and P_H^* , different from the ones in Section 3.1 under full information, and some customers receive 'wrong' offers, leading to a different customer behaviour with respect to accepting the contract. Figure 2 visualizes the pricing process. An initial population of n customers is subdivided into groups by their true risk type, rather than their identified risk type, and finally the insurer loses some customers because of the entailed acceptance patterns of policies.

In this situation, the profit is given by

$$\begin{aligned} \Pi = & \sum_{j=1}^{n \cdot p_L \cdot (1-p_{H|L})} I_j^{L|L} \left(P_L^* - X_j^{(L)} \right) + \sum_{j=1}^{n \cdot p_L \cdot p_{H|L}} I_j^{H|L} \left(P_H^* - X_j^{(L)} \right) \\ & + \sum_{j=1}^{n \cdot p_H \cdot (1-p_{L|H})} I_j^{H|H} \left(P_H^* - X_j^{(H)} \right) + \sum_{j=1}^{n \cdot p_H \cdot p_{L|H}} I_j^{L|H} \left(P_L^* - X_j^{(H)} \right), \end{aligned}$$

where $I_j^{L|L}$, $I_j^{H|L}$, $I_j^{H|H}$ and $I_j^{L|H}$ are Bernoulli random variables with parameters $f_L(P_L^*)$, $f_L(P_H^*)$, $f_H(P_H^*)$ and $f_H(P_L^*)$ respectively. The optimization procedure now amounts to

$$\begin{aligned} \{P_L^*, P_H^*\} = \operatorname{argmax}_{v,w} \mathbb{E}(\Pi) = \operatorname{argmax}_{v,w} & \quad n p_L (1 - p_{H|L}) f_L(v) (v - \mu_L) \\ & + n p_L p_{H|L} f_L(w) (w - \mu_L) + n p_H (1 - p_{L|H}) f_H(w) (w - \mu_H) \\ & + n p_H p_{L|H} f_H(v) (v - \mu_H) - c(n). \end{aligned} \quad (15)$$

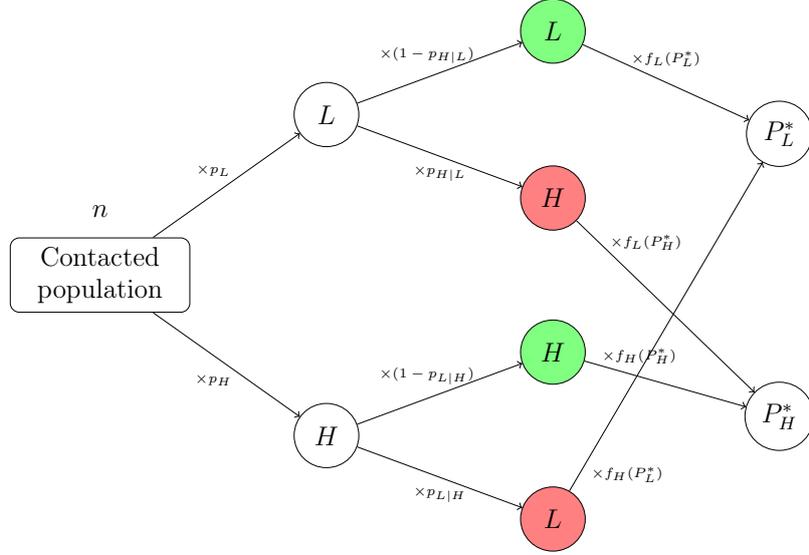


Figure 2: Visualisation of the pricing process.

Theorem 3.6. *In the differentiation case, the expected profit of the insurer is maximized by the premium choice*

$$P_L^* = b^* P_L + (1 - b^*) P_H \quad (16)$$

and

$$P_H^* = c^* P_L + (1 - c^*) P_H \quad (17)$$

for the two classified risk classes, where P_L and P_H are the optimal premiums (5) and (6) of the full information case, $b^* = \frac{p_L(1-p_{H|L})/P_L^{max}}{p_L(1-p_{H|L})/P_L^{max} + p_H p_{L|H}/P_H^{max}}$ and $c^* = \frac{p_L p_{H|L}/P_L^{max}}{p_H(1-p_{L|H})/P_H^{max} + p_L p_{H|L}/P_L^{max}}$.

Proof. We make use of the following first order conditions from (15) to determine the optimal solution:

$$\begin{aligned} \frac{\partial \mathbb{E}(\Pi)}{\partial v} &= n p_L (1 - p_{H|L}) (f'_L(v)v + f_L(v)) + n p_H p_{L|H} (f'_H(v)x + f_H(v)) \\ &\quad - n p_L (1 - p_{H|L}) f'_L(v) \mu_L - n p_H p_{L|H} f'_H(v) \mu_H \stackrel{!}{=} 0 \end{aligned} \quad (18)$$

$$\begin{aligned} \frac{\partial \mathbb{E}(\Pi)}{\partial w} &= n p_L p_{H|L} (f'_L(w)w + f_L(w)) + n p_H (1 - p_{L|H}) (f'_H(w)w + f_H(w)) \\ &\quad - n p_L p_{H|L} f'_L(w) \mu_L - n p_H (1 - p_{L|H}) f'_H(w) \mu_H \stackrel{!}{=} 0 \end{aligned} \quad (19)$$

Equation (18) yields

$$\begin{aligned} n p_L (1 - p_{H|L}) \left(-\frac{1}{P_L^{max}} \right) (v - \mu_L) + n p_L (1 - p_{H|L}) \left(1 - \frac{v}{P_L^{max}} \right) \\ + n p_H p_{L|H} \left(-\frac{1}{P_H^{max}} \right) (v - \mu_H) + n p_H p_{L|H} \left(1 - \frac{v}{P_H^{max}} \right) = 0, \end{aligned} \quad (20)$$

leading to (16). Formula (17) is obtained in a completely analogous way from (19).

The second order conditions are strictly negative and thus confirm the maximum. \square

Like in the no differentiation case, the optimal premiums can again be expressed simply as a weighted average of the optimal premiums from the full information case, and the weights now involve the error probabilities.

Remark 3.7. One needs to verify the limiting case of $P_H^* = P_L^{max}$ and $P_L^* = P_L^{max}$ to obtain the true maximum, since misclassified L individuals may not enter the contract after the limiting premium. This is the case when $c^* < \frac{P_H - P_L^{max}}{P_H - P_L}$ and $b^* < \frac{P_H - P_L^{max}}{P_H - P_L}$. One can distinguish three cases. Firstly, if both $P_H^* < P_L^{max}$ and $P_L^* < P_L^{max}$, then the optimal solutions are given by Equations (16) and (17). Secondly, if only $P_H^* > P_L^{max}$, then the correct premium for the proposed H contract should be $P_H^* = P_H$, since we correctly price for only H types entering the group. H types will always enter the contract since their premium is a weighted average of P_L and P_H , and both are smaller than P_H^{max} from Section 3.1. Thirdly, if both $P_H^* > P_L^{max}$ and $P_L^* > P_L^{max}$, which could happen with a high proportion of misclassified H individuals, then the optimal solution would be to offer the contract only to H types by setting $P_L^* = P_H$ and $P_H^* = P_H$. It is worthwhile to notice that the second-order mixed partial derivatives $\frac{\partial^2 \mathbb{E}(\Pi)}{\partial v \partial w} = \frac{\partial^2 \mathbb{E}(\Pi)}{\partial w \partial v} = 0$, and therefore the optimal price for the low risk types does not depend on the optimal price for the high risk types and vice versa. \square

What is of particular interest is the situation where H individuals are wrongly classified as L . Indeed, since $P_i > \mu_i$, this is the only situation where the insurer makes losses, so it is important to maintain control over this group. The loss (presented here as a negative gain) compared to the benchmark of the situation of full information can be decomposed into

$$\begin{aligned}
& \underbrace{np_L(1 - p_{H|L}) [(f_L(P_L^*) - f_L(P_L))(P_L - \mu_L) + f_L(P_L^*)(P_L^* - P_L)]}_{\text{True L: Loss on L not entering the contract and gain on those who remain}} \\
& + \underbrace{np_L p_{H|L} [(f_L(P_H^*) - f_L(P_L))(P_L - \mu_L) + f_L(P_H^*)(P_H^* - P_L)]}_{\text{False H: Loss on L not entering the contract and gain on those who remain}} \\
& + \underbrace{np_H(1 - p_{L|H}) [(f_H(P_H^*) - f_H(P_H))(P_H - \mu_H) + f_H(P_H^*)(P_H^* - P_H)]}_{\text{True H: Gain on extra H entering the contract and loss on them underpriced}} \\
& + \underbrace{np_H p_{L|H} [(f_H(P_L^*) - f_H(P_H))(P_H - \mu_H) + f_H(P_L^*)(P_L^* - P_H)]}_{\text{False L: Gain on extra H entering the contract and loss on them underpriced}} \\
& - \underbrace{c(n)}_{\text{Invested cost}}.
\end{aligned}$$

Recall that P was the optimal uniform premium for the case without differentiation. Differentiation of risk types only makes sense, if the resulting premiums P_L^*, P_H^* satisfy $P_L^* \leq P \leq P_H^*$ (cf. Figure 3). From Equations (13), (16) and (17), this

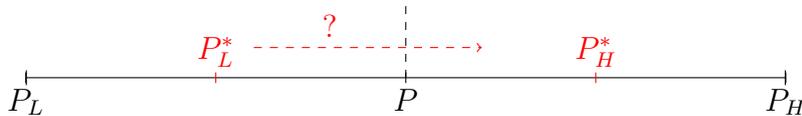


Figure 3: Illustration of premiums in different scenarios.

amounts to the condition $c^* \leq a^* \leq b^*$ which can easily translated to the following condition on the error probabilities:

$$p_{H|L} + p_{L|H} \leq 1.$$

This will always be fulfilled in practically relevant situations.

4 A mean-variance analysis

Proposing a unique premium P to both categories of risks attracts a higher relative proportion of H than the differentiating strategies. This heterogeneity in the portfolio composition generates a higher level of risk, which one should consider in the underwriting process. As the expected profit considered in Section 3 does not capture this aspect of the problem, we introduce the variance of the profit as a simple indicator that can be easily implemented in practical settings, as one only needs estimates for the first two moments of the underlying claim distributions for the analysis.

Define by N_L, N_H the (random) number of insured persons of risk type L and H , respectively, entering the contract. Their first two moments are summarized in Table 1.

Table 1: Expected value and variance of the number of insured for each scenario.

	Full information	No differentiation	Differentiation
$\mathbb{E}(N_L)$	$np_L f_L(P_L)$	$np_L f_L(P)$	$np_L(1 - p_{H L})f_L(P_L^*)$ $+ np_L p_{H L} f_L(P_H^*)$
$\text{Var}(N_L)$	$np_L f_L(P_L)(1 - f_L(P_L))$	$np_L f_L(P)(1 - f_L(P))$	$np_L(1 - p_{H L})f_L(P_L^*)(1 - f_L(P_L^*))$ $+ np_L p_{H L} f_L(P_H^*)(1 - f_L(P_H^*))$
$\mathbb{E}(N_H)$	$np_H f_H(P_H)$	$np_H f_H(P)$	$np_H(1 - p_{L H})f_H(P_H^*)$ $+ np_H p_{L H} f_H(P_L^*)$
$\text{Var}(N_H)$	$np_H f_H(P_H)(1 - f_H(P_H))$	$np_H f_H(P)(1 - f_H(P))$	$np_H(1 - p_{L H})f_H(P_H^*)(1 - f_H(P_H^*))$ $+ np_H p_{L H} f_H(P_L^*)(1 - f_H(P_L^*))$

For each risk type i , $N_i = \sum_{j=1}^n I_j^i$, where I_j^i are independent Bernoulli random variables with probability $f_i(P_i)$, so $\mathbb{E}(N_i) = n f_i(P_i)$ and variance $\text{Var}(N_i) = n f_i(P_i)(1 - f_i(P_i))$. The claim sizes are $X_j^{(i)}$ and the premium is P_i . From (4), we then get

$$\begin{aligned} \text{Var}(\Pi_i) &= \text{Var}\left(\sum_{j=1}^{N_i} (P_i - X_j^{(i)})\right) \\ &= \mathbb{E}\left(\text{Var}\left(\sum_{j=1}^{N_i} (P_i - X_j^{(i)}) \mid N_i\right)\right) + \text{Var}\left(\mathbb{E}\left(\sum_{j=1}^{N_i} (P_i - X_j^{(i)}) \mid N_i\right)\right) \\ &= \mathbb{E}(N_i) \cdot \sigma_i^2 + (P_i - \mu_i)^2 \cdot \text{Var}(N_i). \end{aligned} \tag{21}$$

With this ingredient, we can now derive the variance of the profit in our three scenarios introduced in the previous section.

- Full information:

$$\text{Var}(\Pi) = \mathbb{E}(N_L) \sigma_L^2 + (P_L - \mu_L)^2 \text{Var}(N_L) + \mathbb{E}(N_H) \sigma_H^2 + (P_H - \mu_H)^2 \text{Var}(N_H).$$

- No differentiation:

$$\text{Var}(\Pi) = \mathbb{E}(N_L) \sigma_L^2 + (P - \mu_L)^2 \text{Var}(N_L) + \mathbb{E}(N_H) \sigma_H^2 + (P - \mu_H)^2 \text{Var}(N_H).$$

- Differentiation:

$$\begin{aligned} \text{Var}(\Pi) = & np_L(1 - p_{H|L})f_L(P_L^*)\sigma_L^2 + (P_L^* - \mu_L)^2 np_L(1 - p_{H|L})f_L(P_L^*)(1 - f_L(P_L^*)) \\ & + np_L p_{H|L} f_L(P_H^*)\sigma_L^2 + (P_H^* - \mu_L)^2 np_L p_{H|L} f_L(P_H^*)(1 - f_L(P_H^*)) \\ & + np_H(1 - p_{L|H})f_H(P_H^*)\sigma_H^2 + (P_H^* - \mu_H)^2 np_H(1 - p_{L|H})f_H(P_H^*)(1 - f_H(P_H^*)) \\ & + np_H p_{L|H} f_H(P_L^*)\sigma_H^2 + (P_L^* - \mu_H)^2 np_H p_{L|H} f_H(P_L^*)(1 - f_H(P_L^*)). \end{aligned}$$

Remark 4.1. *Note that in all of the above expressions a term containing the variance due to the randomness of claim sizes is followed by one with the variance due to the randomness of underwriting, i.e. the customer's probability to enter the contract or not. In Section 5, we will illustrate this decomposition with the help of a numerical example.*

For both the insurance company itself and the regulator it will be natural to also include a risk constraint into the problem of optimizing profits. Consequently, we introduce a variance constraint in the optimization problem, modifying the problem from Section 3 to

$$\begin{aligned} & \max \mathbb{E}(\Pi) \\ & \text{s.t. } \text{Var}(\Pi) \leq \bar{\sigma}^2. \end{aligned}$$

Varying the value of $\bar{\sigma}$ will lead to a mean-variance efficient frontier in the spirit of Markowitz [29]. Introduce the Lagrange multipliers

$$\mathcal{L}(P_i, \lambda) = \mathbb{E}(\Pi(P_i)) + \lambda (\bar{\sigma}^2 - \text{Var}(\Pi(P_i))) \quad (22)$$

for the premium P_i in any of the optimization programs (7), (11) and (15). The optimal premiums are then obtained by the first order conditions

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial P_i} &= \frac{\partial \mathbb{E}(\Pi)}{\partial P_i} - \lambda \frac{\partial \text{Var}(\Pi)}{\partial P_i} \stackrel{!}{=} 0, \\ \frac{\partial \mathcal{L}}{\partial \lambda} &= \bar{\sigma}^2 - \text{Var}(\Pi) \stackrel{!}{=} 0. \end{aligned}$$

In order to construct the efficient frontier, we maximize the expected profit subject to the constraint of the variance being smaller than a certain level ($\bar{\sigma}^2$). We use the Lagrange multiplier method in order to perform the optimization under this constraint. Thus, we obtain one point of the frontier defined by the coordinated μ and $\bar{\sigma}^2$. To obtain more points and draw the frontier, we augment the $\bar{\sigma}^2$ and redo the analysis each time. We give here the corresponding equations for the full information case, the other cases follow in an analogous way. Equation (22)

translates into

$$\begin{aligned}
\mathcal{L}(P_L, P_H) = & n \left(p_L \left(1 - \frac{1}{P_L^{max}} P_L \right) (P_L - \mu_L) + p_H \left(1 - \frac{1}{P_H^{max}} P_H \right) (P_H - \mu_H) \right) \\
& + \lambda \left\{ \bar{\sigma}^2 - \left[\left(n p_L \left(1 - \frac{1}{P_L^{max}} P_L \right) \right)^2 + n p_L \left(1 - \frac{1}{P_L^{max}} P_L \right) \right] \sigma_L^2 \right. \\
& - (P_L - \mu_L)^2 n p_L \left(1 - \frac{1}{P_L^{max}} P_L \right) \frac{P_L}{P_L^{max}} \\
& - \left[\left(n p_H \left(1 - \frac{1}{P_H^{max}} P_H \right) \right)^2 + n p_H \left(1 - \frac{1}{P_H^{max}} P_H \right) \right] \sigma_H^2 \\
& \left. - (P_H - \mu_H)^2 n p_H \left(1 - \frac{1}{P_H^{max}} P_H \right) \frac{P_H}{P_H^{max}} \right\}.
\end{aligned}$$

The first order conditions are given by

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial P_L} = & n p_L \left(-\frac{1}{P_L^{max}} \right) (P_L - \mu_L) + n p_L \left(1 - \frac{1}{P_L^{max}} P_L \right) \\
& - \lambda \left(n p_L \left(-\frac{1}{P_L^{max}} \right) \sigma_L^2 + 2(P_L - \mu_L) n p_L \left(1 - \frac{1}{P_L^{max}} P_L \right) \frac{P_L}{P_L^{max}} \right. \\
& \left. + (P_L - \mu_L)^2 n p_L \frac{1}{P_L^{max}} \left(1 - 2 \frac{P_L}{P_L^{max}} \right) \right) \stackrel{!}{=} 0, \\
\frac{\partial \mathcal{L}}{\partial P_H} = & n p_H \left(-\frac{1}{P_H^{max}} \right) (P_H - \mu_H) + n p_H \left(1 - \frac{1}{P_H^{max}} P_H \right) \\
& - \lambda \left(n p_H \left(-\frac{1}{P_H^{max}} \right) \sigma_H^2 + 2(P_H - \mu_H) n p_H \left(1 - \frac{1}{P_H^{max}} P_H \right) \frac{P_H}{P_H^{max}} \right. \\
& \left. + (P_H - \mu_H)^2 n p_H \frac{1}{P_H^{max}} \left(1 - 2 \frac{P_H}{P_H^{max}} \right) \right) \stackrel{!}{=} 0, \\
\frac{\partial \mathcal{L}}{\partial \lambda} = & \bar{\sigma}^2 - \left(n p_L \left(1 - \frac{1}{P_L^{max}} P_L \right) \right) \sigma_L^2 + (P_L - \mu_L)^2 n p_L \left(1 - \frac{1}{P_L^{max}} P_L \right) \frac{P_L}{P_L^{max}} \\
& + n p_H \left(1 - \frac{1}{P_H^{max}} P_H \right) \sigma_H^2 + (P_H - \mu_H)^2 n p_H \left(1 - \frac{1}{P_H^{max}} P_H \right) \frac{P_H}{P_H^{max}} \stackrel{!}{=} 0.
\end{aligned}$$

This results in a system of three equations for the three unknowns P_L, P_H and λ which can be solved numerically for every choice of involved parameters.

5 Numerical illustrations

Let us now consider concrete numerical illustrations of the results of the previous sections. The following parametrization will be used throughout this section unless otherwise stated:

$$\mu_L = 1, \mu_H = 5, \sigma_L^2 = 1, \sigma_H^2 = 10, p_L = 0.9, p_{H|L} = p_{L|H} = 0.1. \quad (23)$$

For the shape of the cost function we assume $c(n) = 20 \log n$ in the plots, but note that any other choice would be feasible as well.

5.1 Expected profit

Let $f_i(P)$ have the form (3) with $P_L^{max} = 4\mu_L = 4$ and $P_H^{max} = 4\mu_H = 20$. Then we get from the respective formulas of Section 3:

- Full information:

$$P_L = 2.5, \quad P_H = 12.5, \quad \mathbb{E}(\Pi) = 0.788n.$$

- No differentiation:

$$P \approx 2.717, \quad \mathbb{E}(\Pi) \approx 0.298n.$$

Note that $\mu_L < P < \mu_H$. In this case, the insurer targets the low-risk L type customers because their proportion in the population is large enough to compensate for the losses on the H types.

- Differentiation in two classes:

$$P_L^* \approx 2.525, \quad P_H^* = 12.5, \quad \mathbb{E}(\Pi) \approx 0.687n - c(n).$$

Note that $P_H^* > P_L^{max}$.

Applying the classification is hence only an advantage if

$$0.687n - c(n) \geq 0.298n.$$

Conversely, the maximum cost which the insurer will be willing to pay for the classification, given a population of size n , is

$$c(n) < 0.388n.$$

It is instructive to look into the sensitivity of the results. Let us first explore the variability of the profit under different error probabilities, which can be a helpful decision tool in case of limited investment resources. For each level of error probabilities, we recompute the optimal premiums. Figure 4 features the sensitivity of the expected profit with respect to both error probabilities. The classification cost still needs to be deducted here from the expected profit. If the insurer is given a choice of different classification algorithms or investment possibilities for improvement of precision with known resulting error probabilities, one can verify whether that investment is worthwhile. This figure may help the decision makers to judge whether with given error probabilities, a refinement in the classification may be of added value to the company.

5.2 Variance

Numerically, with the parameters defined in (23), we obtain the following results for the three cases of the linear acceptance function:

- Full information:

$$\text{Var}(\Pi) = 2.505469n.$$

- No differentiation:

$$\text{Var}(\Pi) = 1.79213n.$$

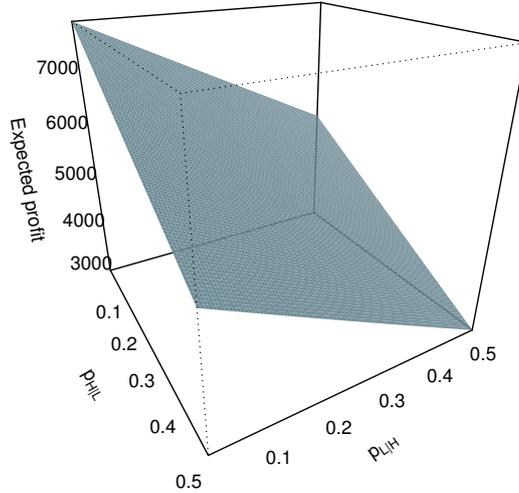


Figure 4: Expected profit as function of error probabilities.

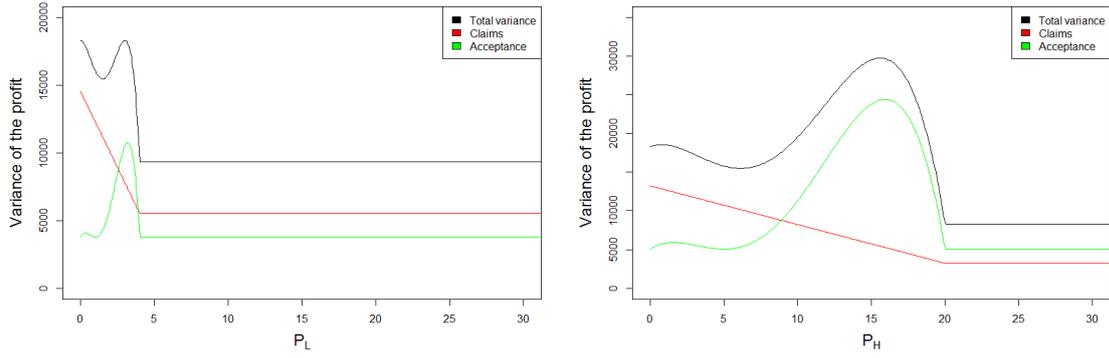
- Differentiation:

$$\text{Var}(\Pi) = 2.355267n.$$

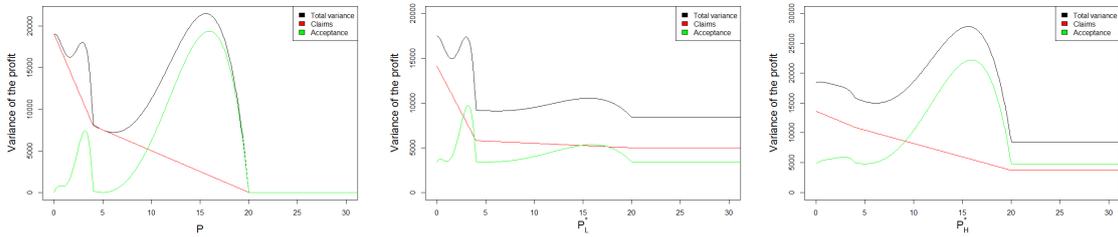
As the total variance is an increasing function of the number of policyholders, it will naturally be higher under a differentiation strategy, as the insurer gets more market share. But the structure of the variance will be different. Without differentiation, the variance inside the group is much higher than the average of internal group variances from the differentiation case, that difference being larger when the two distributions are further apart.

In Figure 5, we plot the variances in the three scenarios to illustrate their forms as a function of chosen premium. We split the variances according to the part stemming from the variability of claims (in red) and from the one of acceptance of contracts (in green). The humps indicate the region where the increase of variance due to the increasing deviation from the mean is compensated by the decrease in the number of underwritten policies. In Figures 5b and 5c, we can observe two humps, appearing because of the mixture of two risk types. With the help of this decomposition, we can clearly see that the humps in the plots come from the acceptance behaviour. In Figure 5a, we observe that if the premium becomes too high, the total variance decreases as the population does not enter the contract any more.

In Figures 6-8, we show the variances when varying one parameter at a time. In Figure 6, one can observe the variance shapes for a small range acceptance function. In this case, the variance is mostly defined by the claims behaviour, since the acceptance rate remains low and the humps are less pronounced. For an acceptance function ranging up to high premiums as shown in Figure 7, the acceptance variance dominates. The humps are more pronounced as policyholders exist in a



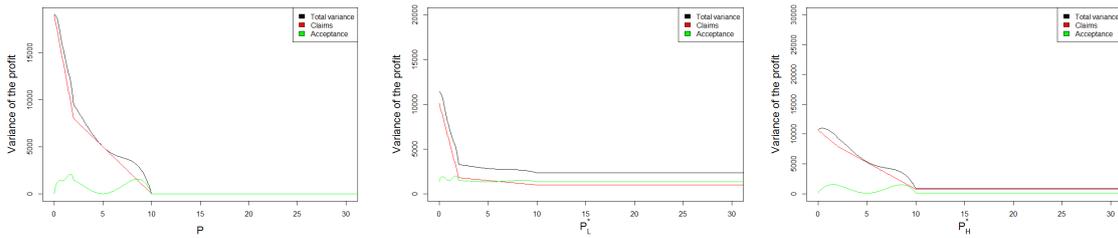
(a) Perfect information case. Left-hand side: $\text{Var}(\Pi)$ as a function of P_L . Right-hand side: $\text{Var}(\Pi)$ as a function of P_H .



(b) No differentiation case. (c) Differentiation case. Left-hand side: $\text{Var}(\Pi)$ as a function of P . Right-hand side: $\text{Var}(\Pi)$ as a function of P_H^* .

Figure 5: Decomposition of the variance.

broader range. Finally, with a bigger expected claim size difference between risk types, the relationship between the humps and the risk types becomes clearer as they are further apart, see Figure 8.



(a) No differentiation case.

(b) Differentiation case.

Figure 6: Decomposition of the variance, parameter $P_i^{max} = 2\mu_i$.

5.3 Mean-variance efficient frontier

To complete the numerical part, we now address the illustration of the mean-variance frontier as defined in Section 4 for a population size of $n = 10,000$. We see in Figure 9 that up to a certain variance level, the non-differentiation strategy dominates differentiation in terms of expected profit. This breaking point depends on the cost function $c(n)$ and the error probabilities. One may also want to consider limiting

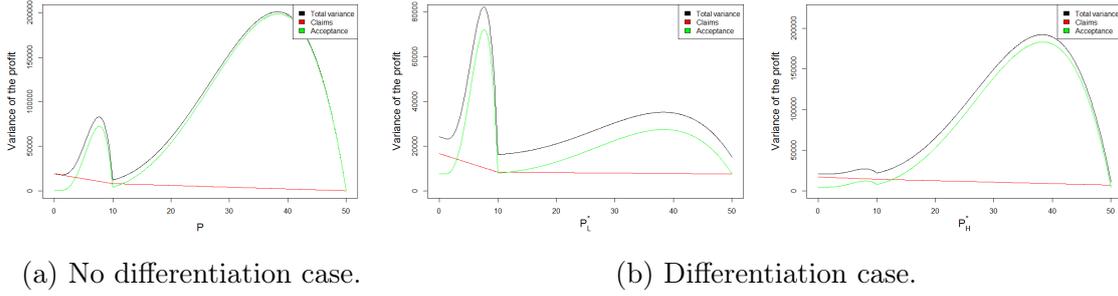


Figure 7: Decomposition of the variance, parameter $P_i^{max} = 10\mu_i$.

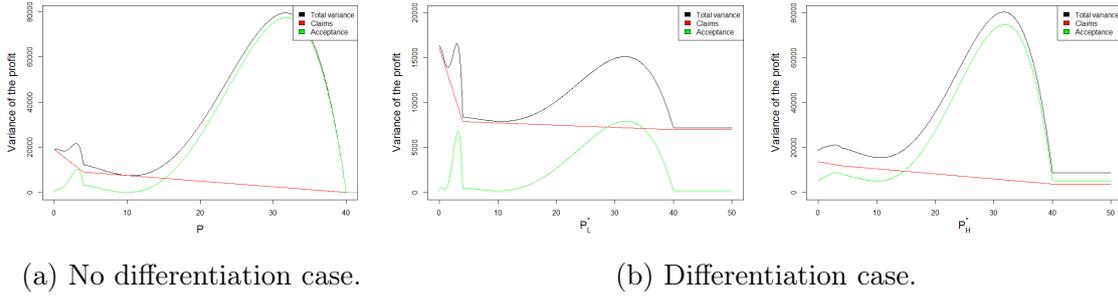


Figure 8: Decomposition of the variance, parameter $\mu_H = 10$.

constraints in practice such as regulatory constraints or the demands of stakeholders. The kinks in the frontier arise from the fact that for different variance limitations, a different portfolio composition becomes optimal. In other words, the optimal strategy switches in the points of the kinks by letting more of a lower or higher risk type entering the contract.

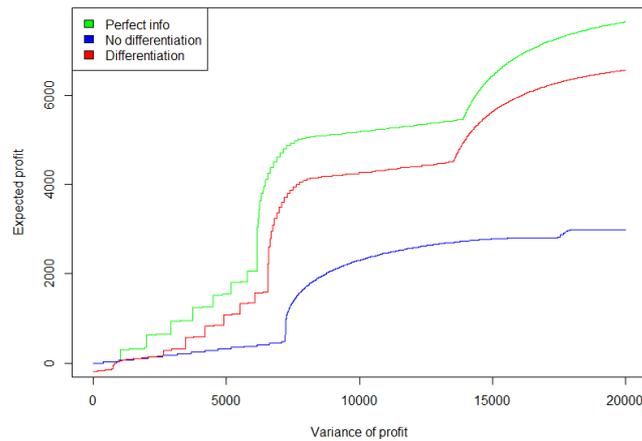


Figure 9: Mean-variance frontier with linear demand function.

The mean-variance approach assumes variations to both sides as equally weighted since the variance is a symmetric risk measure. This framework can be extended to other risk measures, such as the lower semi-variance to take in account only

one-sided deviations from the mean or the value-at-risk to consider minimal profit requirements. These adaptations are developed in Section 6, where we also present an alternative approach for the risk assessment based on utility functions.

6 Extensions

6.1 A sigmoid-type acceptance function

While the piece-wise linear acceptance functions used in this paper allow for intuitive and transparent results, one may want to challenge this simplistic assumption. In this section we would like to extend the previous analysis to a possibly more realistic shape that still allows for an explicit treatment. Concretely, assume that f belongs to the class of sigmoid functions, namely the logistic functions, which are smooth and monotone, thus suitable for our situation [26]. This form of function appears when applying a logit lapsing model with different risk factors, see e.g. [15, 19]. An example of a model using premiums as risk factors can be found in [7, 20] and particularly in [16]. Consider the following concrete shape of the acceptance function f_i of an individual of risk type i :

$$f_i(P) = \frac{1}{1 + e^{a_i(P-b_i)}}, \quad a_i \in \mathbb{R}^+, \quad b_i \in \mathbb{R}, \quad i \in \{L, H\}, \quad (24)$$

where the parameters a_i and b_i need to be calibrated. We can suppose $b_i > \mu_i$, so that the function reaches value 1/2 for premiums that are higher than the actuarially fair premium, cf. Figure 10. As a grows, the curve becomes steeper around the pivotal position determined by the parameter b (note that the choice of b also determines the value of f for $P = 0$ which will typically be smaller than 1). From an analytical point of view, the form (24) is more attractive than the piece-wise linear shape considered in the previous sections, as it is differentiable everywhere. Clearly, f_i is

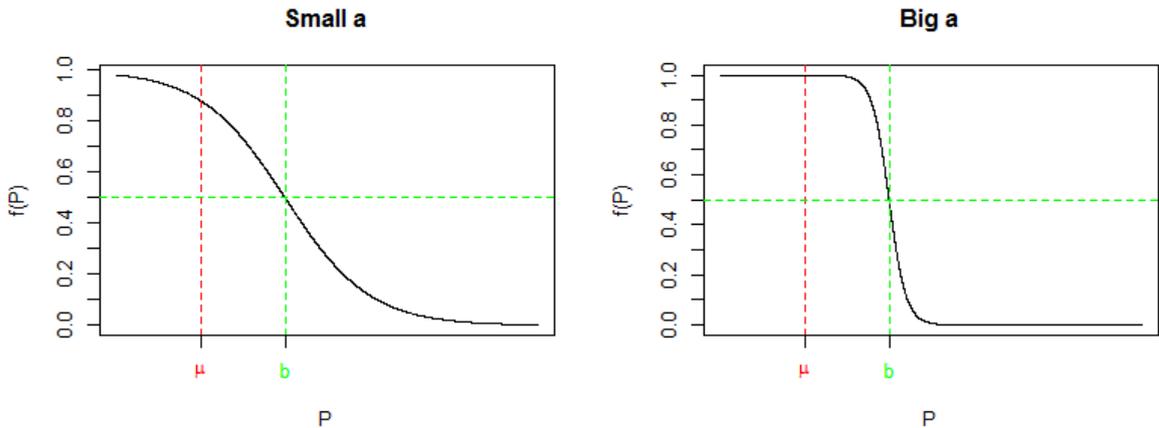


Figure 10: Acceptance function f for small (left) and large (right) parameter a .

strictly decreasing in P :

$$\frac{\partial f_i(P)}{\partial P} = \frac{-a_i e^{a_i(P-b_i)}}{(1 + e^{a_i(P-b_i)})^2} < 0.$$

Define further the *price elasticity* of a risk type as the change of the number of customers entering the contract with respect to the price variation:

$$E_P = \frac{\partial f_i(P)}{\partial P} \frac{P}{f_i} = \frac{-a_i e^{a_i(P-b_i)} P}{1 + e^{a_i(P-b_i)}} = -a_i e^{a_i(P-b_i)} P f_i(P).$$

This measure illustrates the reactivity of the portfolio size to the variation of premium, cf. for instance [41, Ch.15].

6.1.1 Theoretical results

We first derive the analogous results to the ones in Sections 3, 4, under the sigmoid acceptance function. The full information case still leads to an explicit formula:

Theorem 6.1. *In the full information case, the expected profit of the insurer is maximized for the premium choice*

$$P_L = \mu_L + \frac{1}{a_L} + \frac{1}{a_L} W(e^{a_L b_L - a_L \mu_L - 1}) \quad (25)$$

and

$$P_H = \mu_H + \frac{1}{a_H} + \frac{1}{a_H} W(e^{a_H b_H - a_H \mu_H - 1}), \quad (26)$$

where $W(z)$ denotes the (principal branch of the) Lambert W function, which is the inverse function of $g(x) = x e^x$ (cf. [10]).

Proof. The optimal premium choice is the solution of the optimization problem

$$\{P_L, P_H\} = \operatorname{argmax}_{x,y} \mathbb{E}(\Pi) = \operatorname{argmax}_{x,y} n(p_L f_L(x)(x - \mu_L) + p_H f_H(y)(y - \mu_H)).$$

We can characterize the maxima by the equations

$$\frac{\partial \mathbb{E}(\Pi)}{\partial x} = n p_L (f'_L(x)x + f_L(x)) - n p_L f'_L(x) \mu_L \stackrel{!}{=} 0, \quad (27)$$

$$\frac{\partial \mathbb{E}(\Pi)}{\partial y} = n p_H (f'_H(y)y + f_H(y)) - n p_H f'_H(y) \mu_H \stackrel{!}{=} 0. \quad (28)$$

From (27) we have

$$\begin{aligned} n p_L \left(\frac{\partial \left(1 - \frac{1}{1 + e^{-a_L(x-b_L)}} \right)}{\partial x} (x - \mu_L) + 1 - \frac{1}{1 + e^{-a_L(x-b_L)}} \right) &= 0 \\ \iff a_L(x - \mu_L) &= 1 + e^{-a_L(x-b_L)} \end{aligned}$$

leading to (25). Equation (26) for the high-risk individuals is then obtained in a completely analogous way.

To see that the extremal point is indeed a local maximum, one needs to verify

$$\frac{\partial^2 \mathbb{E}(\Pi)}{\partial x^2} = n p_L f''_L(x)(x - \mu_L) + 2 n p_L f'_L(x) < 0.$$

Using $\{P_L > \mu_L, P_H > \mu_H\}$ and

$$\frac{\partial^2 f(P)}{\partial P^2} = \frac{a_i^2 e^{-a_i(P-b_i)} (1 - e^{-a_i(P-b_i)})}{(1 + e^{-a_i(P-b_i)})^3},$$

we can rearrange the previous condition as

$$f_L''(x)(x - \mu_L) < -2f_L'(x).$$

Since $-2f_L'(x)$ is always positive, $f_L''(x) < 0$ for all $x < b_L$ and $x = P_L > \mu_L$. The same conclusion holds for the second derivative w.r.t. y . \square

Remark 6.2. Note that we can provide a necessary condition for P_L to be smaller than μ_H :

$$e^{a_L(b_L - \mu_H)} < a_L(\mu_H - \mu_L) - 1. \quad (29)$$

This condition is of interest for analysing the case when the low-risk type premium yields losses in absolute terms if sold to a high-risk type. Also, one can easily obtain sensitivities of the premium with respect to the parameters by means of first derivatives.

$$\frac{\partial P_i}{\partial \mu_i} = 1 + \frac{1}{a_i} \frac{\partial W}{\partial \mu_i} = 1 - \frac{W(e^{a_i b_i - a_i \mu_i - 1})}{1 + W(e^{a_i b_i - a_i \mu_i - 1})}, \in [0, 1] > 0,$$

$$\frac{\partial P_i}{\partial b_i} = \frac{1}{a_i} \frac{\partial W}{\partial b_i} = \frac{W(e^{a_i b_i - a_i \mu_i - 1})}{1 + W(e^{a_i b_i - a_i \mu_i - 1})}, \in [0, 1] > 0,$$

$$\frac{\partial P_i}{\partial a_i} = -\frac{1}{a_i^2} - \frac{1}{a_i^2} W + \frac{1}{a_i} \frac{\partial W}{\partial \mu_i} = \frac{b_i - \mu_i}{a_i} \frac{W(e^{a_i b_i - a_i \mu_i - 1})}{1 + W(e^{a_i b_i - a_i \mu_i - 1})} - \frac{1}{a_i^2} (1 + W(e^{a_i b_i - a_i \mu_i - 1})).$$

Concerning the sign of the last term, under $b_i > \mu_i$ the first term is positive and the second is negative. The overall difference is negative for small values of a_i , but positive for larger a_i , that effect manifesting itself sooner if the difference $b_i - \mu_i$ is larger.

In case of no differentiation, we proceed as before by taking first order conditions of the expected profit defined above in Equation (11):

$$\begin{aligned} \frac{\partial \mathbb{E}(\Pi)}{\partial z} &= np_L (f_L'(z)z + f_L(z)) + np_H (f_H'(z)z + f_H(z)) \\ &\quad - np_L f_L'(z)\mu_L - np_H f_H'(z)\mu_H \stackrel{!}{=} 0. \end{aligned}$$

Plugging in the sigmoid function f yields

$$\begin{aligned} & p_L \left(\frac{-a_L e^{-a_L(z-b_L)}}{(1 + e^{-a_L(z-b_L)})^2} (z - \mu_L) + 1 - \frac{1}{1 + e^{-a_L(z-b_L)}} \right) \\ & + p_H \left(\frac{-a_H e^{-a_H(z-b_H)}}{(1 + e^{-a_H(z-b_H)})^2} (z - \mu_H) + 1 - \frac{1}{1 + e^{-a_H(z-b_H)}} \right) = 0, \end{aligned}$$

which can be solved numerically. We establish that $P_L \leq P \leq P_H$, following the assumption in (29).

For the differentiation case, we have the problem defined in Equation (15) to solve. Once again, as the first order conditions are symmetric, we will detail only one of them.

$$\begin{aligned}
\frac{\partial \mathbb{E}(\Pi)}{\partial v} &= np_L(1 - p_{H|L})(f'_L(v)v + f_L(v)) + np_H p_{L|H}(f'_H(v)v + f_H(v)) \\
&\quad - np_L(1 - p_{H|L})f'_L(v)\mu_L - np_H p_{L|H}f'_H(v)\mu_H = 0 \\
\iff p_L(1 - p_{H|L}) &\left(\frac{-a_L e^{-a_L(v-b_L)}}{(1 + e^{-a_L(v-b_L)})^2}(v - \mu_L) + 1 - \frac{1}{1 + e^{-a_L(v-b_L)}} \right) \\
&+ p_H p_{L|H} \left(\frac{-a_H e^{-a_H(v-b_H)}}{(1 + e^{-a_H(v-b_H)})^2}(v - \mu_H) + 1 - \frac{1}{1 + e^{-a_H(v-b_H)}} \right) = 0.
\end{aligned} \tag{30}$$

Similarly, we also get

$$\begin{aligned}
\frac{\partial \mathbb{E}(\Pi)}{\partial w} = 0 &\iff p_L(p_{H|L}) \left(\frac{-a_L e^{-a_L(w-b_L)}}{(1 + e^{-a_L(w-b_L)})^2}(w - \mu_L) + 1 - \frac{1}{1 + e^{-a_L(w-b_L)}} \right) \\
&+ p_H(1 - p_{L|H}) \left(\frac{-a_H e^{-a_H(w-b_H)}}{(1 + e^{-a_H(w-b_H)})^2}(w - \mu_H) + 1 - \frac{1}{1 + e^{-a_H(w-b_H)}} \right) = 0.
\end{aligned} \tag{31}$$

These conditions characterize the optimum, which is then solved numerically.

6.1.2 Numerical illustrations

Let us look into the case of a sigmoid acceptance function (24) with parameters

$$a_L = a_H = 1, b_i = 2\mu_i, i \in \{L, H\}.$$

All other parameters remaining identical to those from Section 5, we obtain the following results:

- Full information:

$$\begin{aligned}
P_L &\approx 2.567203, \\
P_H &\approx 8.926367, \\
\mathbb{E}(\Pi) &\approx 0.8030561n.
\end{aligned}$$

- No differentiation:

$$\begin{aligned}
P &\approx 8.836827, \\
\mathbb{E}(\Pi) &\approx 0.2998947n.
\end{aligned}$$

Note that $P > \mu_H > \mu_L$. In this case, the insurer targets the high risk type audience because even if its size is smaller, with this acceptance function form he can make higher margins on them, thus compensating their smaller size.

- Differentiation:

$$\begin{aligned} P_L^* &\approx 2.601853, \\ P_H^* &\approx 8.91731, \\ \mathbb{E}(\Pi) &\approx 0.6993114n - c(n). \end{aligned}$$

Note that $\mu_L < P_L^* < \mu_H < P_H^*$.

Consequently, differentiation is only preferable here if the population size satisfies

$$0.6993114n - c(n) \geq 0.2998947n, \quad \text{that is } n/\ln(\gamma n) \geq 2.503651c_0,$$

in our case $n \geq 282.617$. Conversely, the maximum cost the insurer is willing to pay given a population size is given by:

$$c(n) < 0.3994167n.$$

For the variances, the results are as follows:

- Full information:

$$\text{Var}(\Pi) = 1.874096n.$$

- No differentiation:

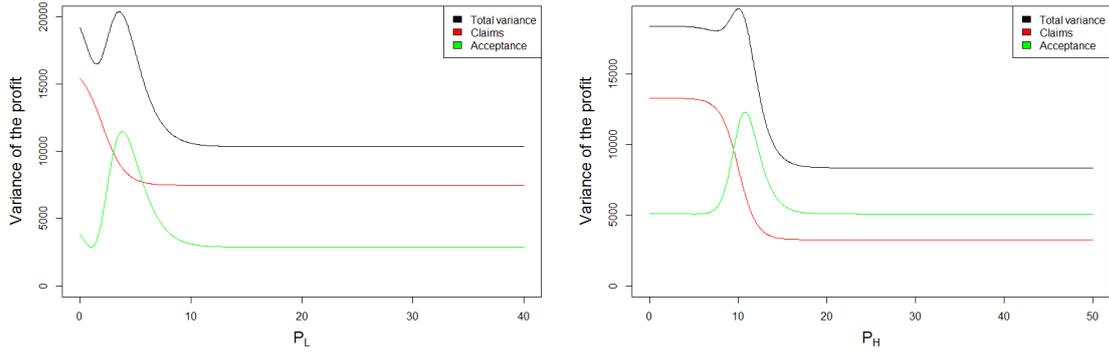
$$\text{Var}(\Pi) = 1.089133n.$$

- Differentiation:

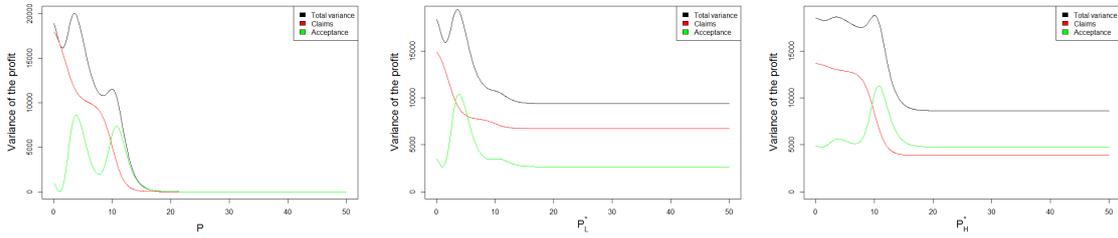
$$\text{Var}(\Pi) = 1.800876n.$$

Figure 11 depicts the form of the variance as function of the proposed premiums in the different scenarios and its decomposition into the two parts (the variance arising from the acceptance function and the one from the claim size variability), showing again a hump pattern. Figures 12-14 illustrate the sensitivity of the variances of the profit with and without differentiation, when varying one of the parameters. In Figure 12, we observe that under small price elasticity, with more customers entering the contract, the hump behaviour disappears, since at the limit there is no gap in different risk types behaviour. The total variance is now mostly due to the underwriting process via the acceptance function, and claim size variance has little effect on the total variance. In contrast, a high price elasticity pushes different risk types to stabilize around their pivotal point of their respective acceptance function, accepting contracts only below this point, see Figure 13. All the variance of the profit can then be explained by the claim size variance. Finally, Figure 15 gives the mean-variance frontier in case of this sigmoid acceptance function.

We observe that the no-differentiation strategy changes depending on the form of the acceptance function used in the analysis. This can be particularly relevant in the case when a company conducts a study using a simplified linear form instead of a more realistic logistic approximation.

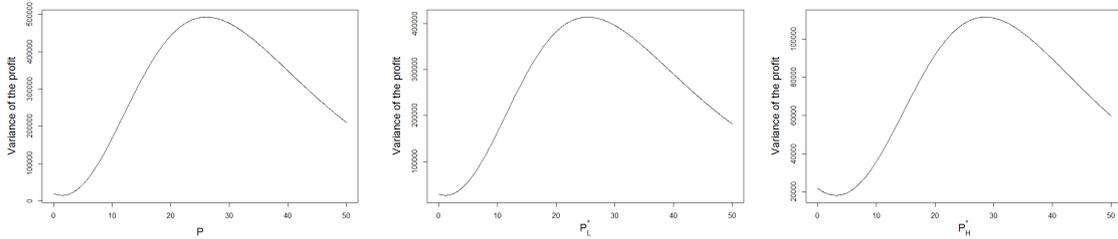


(a) Perfect information case. Left-hand side: $\text{Var}(\Pi)$ as function of P_L . Right-hand side: $\text{Var}(\Pi)$ as function of P_H .



(b) No differentiation case. (c) Differentiation case. Left-hand side: $\text{Var}(\Pi)$ as function of P . Middle: $\text{Var}(\Pi)$ as function of P_L^* . Right-hand side: $\text{Var}(\Pi)$ as function of P_H^* .

Figure 11: Decomposition of the variance.



(a) No differentiation case.

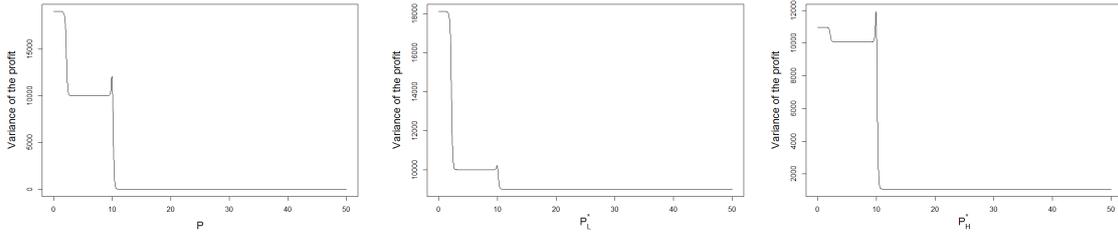
(b) Differentiation case.

Figure 12: Parameter $a = 0.1$. Under small price-elasticity of demand, we observe higher levels of underwriting for both risk types, hence higher and smoother variance.

6.1.3 Sensitivities

Let us now investigate the sensitivity of the expected profit w.r.t. to each parameter. The change in expected profits will allow to compute the variation in the maximal cost for which the differentiation policy is still advantageous.

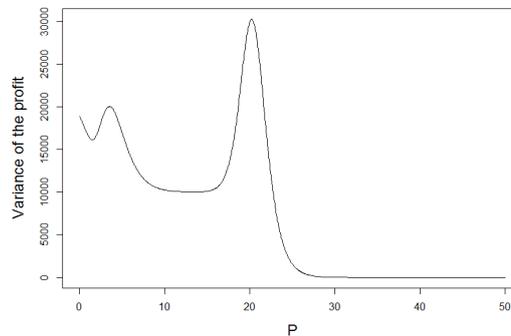
Parameter a : In the sigmoid curve of the acceptance function f , a represents the steepness, giving the speed at which the function changes around its central point (see Figure 10). We choose $a_L = a_H = a$, so we will vary it as a unique parameter. We observe in Figure 16 that for higher levels of a , the steepness of the twist increases, meaning that the values of the acceptance function grow closer to the



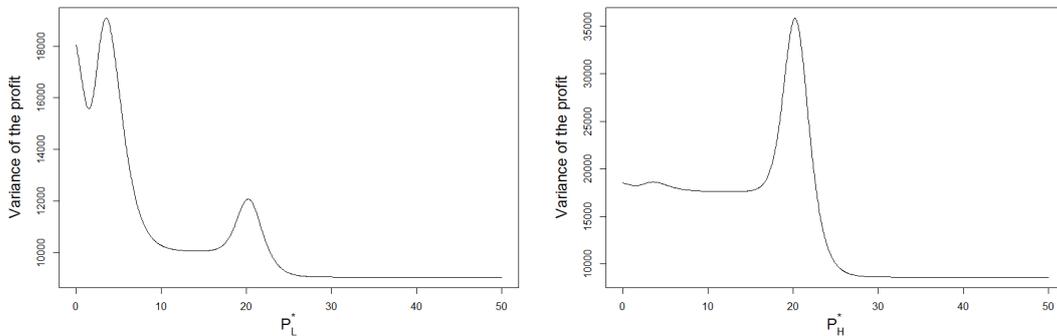
(a) No differentiation case.

(b) Differentiation case.

Figure 13: Parameter $a = 10$. Under big price-elasticity of demand, demand concentrates around pivotal points, thus different risk types only enter contract until their pivotal point price, therefore steps are noticeable.



(a) No differentiation case.



(b) Differentiation case.

Figure 14: Parameter $\mu_H = 10$. With a bigger expected claim size difference between risk types, the relationship between the humps and the risk types becomes clearer.

points b_L and b_H . Thus, the price can be set closer to the twisting point, allowing a higher proportion of individuals to enter the contract. As in our initial parametrization $b_i > \mu_i$, we gain strictly positive profit when pricing around b_i . In the case of no differentiation, we cannot entirely benefit from this feature, as one of our types twisting point will end up far from the unique price P . Therefore, the maximum cost the insurer is willing to invest into the classification method is increasing in the parameter a . We can determine the limit:

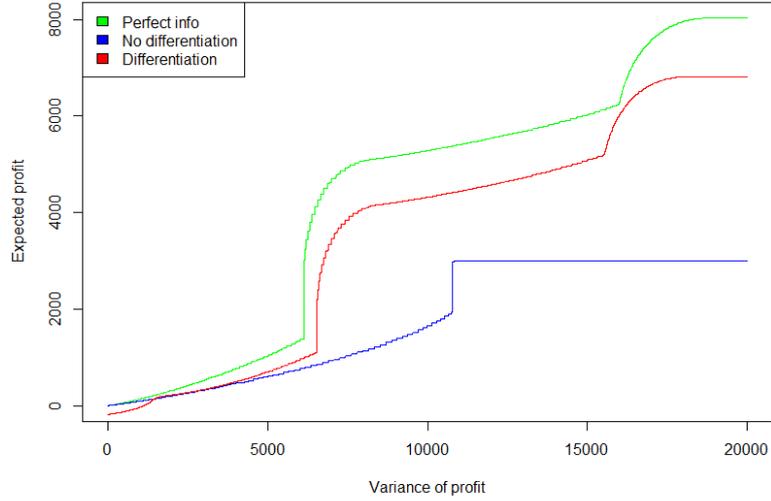
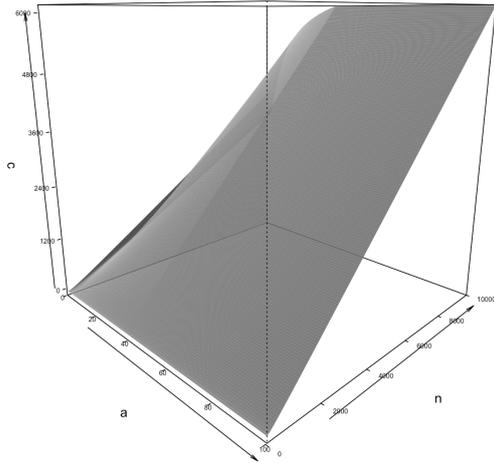


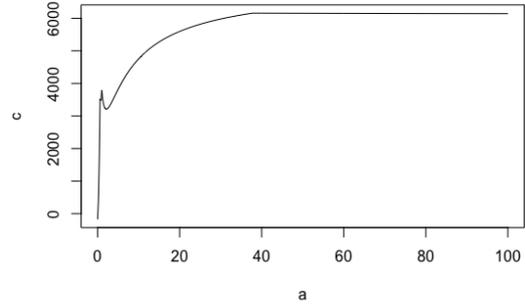
Figure 15: Efficient frontiers in the mean-variance setup for the three scenarios.

$$\begin{aligned}
\lim_{a \rightarrow +\infty} c(a) &= \lim_{a \rightarrow +\infty} \mathbb{E}(\Pi(P_L^*(a), P_H^*(a))) - \lim_{a \rightarrow +\infty} \mathbb{E}(\Pi(P(a))) \\
&= \lim_{\{P_L^* \rightarrow b_L^-, P_H^* \rightarrow b_H^-\}} \mathbb{E}(\Pi(P_L^*(a), P_H^*(a))) - \max \left(\lim_{P \rightarrow b_L^-} \mathbb{E}(\Pi(P(a))), \lim_{P \rightarrow b_H^-} \mathbb{E}(\Pi(P(a))) \right) \\
&= p_L(1 - p_{H|L})(b_L - \mu_L)n + p_H(1 - p_{L|H})(b_H - \mu_H)n + p_H p_{L|H}(b_L - \mu_H)n \\
&\quad - \max(p_L(b_L - \mu_L)n + p_H(b_L - \mu_H)n, p_H(b_H - \mu_H)n),
\end{aligned}$$

which in our case gives $0.63n$.



(a) Function of a and n .



(b) n fixed at 10000.

Figure 16: Maximum affordable investment cost for implementation of a differentiation mechanism as a function of a and n .

Parameters b_L and b_H : Now we simultaneously vary the parameters b_H and b_L (the

central points of the acceptance functions), see Figure 17. Naturally, the higher b_i , the higher will be the overall profit, as customers accept premiums until higher thresholds. Therefore, it becomes more and more attractive to differentiate customers to actually get this profit. Conversely, if b_H grows ceteris paribus, the profit increase becomes smaller with differentiation as the proportion of high-risk types is too low to strongly influence the non-differentiation premium.

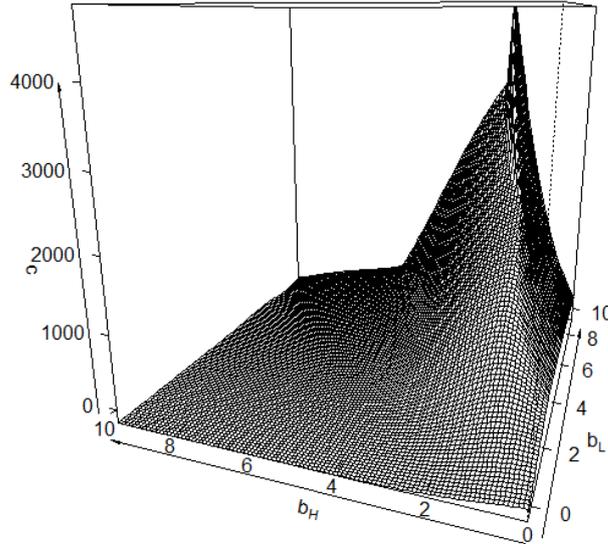


Figure 17: Maximum affordable cost as a function of b_H , b_L and $n = 10000$

6.2 Other risk measures

We give a short comparative analysis for two other risk measures replacing the variance criterion (see Pflug and Römisch [33, Ch.5] for a more extensive list of possible alternatives in the context of efficient frontier studies in decision making).

Lower semi-variance

The main drawback of a variance risk constraint is that positive deviations from the mean are also penalized. In [28], Markowitz suggests the concept of the lower semi-variance (LSV)

$$\text{Var}^- (\Pi) := \mathbb{E} \left((\min(0, \Pi - \mathbb{E}(\Pi)))^2 \right)$$

of the profit Π to account for asymmetry of positive and negative deviations from the profit target. In this case, analytical formulas are not feasible any more, but one can obtain similar results by Monte Carlo simulation, using 1000 simulation runs. For that purpose, rather than only specifying two moments, we need to take an assumption of the entire distribution of claim sizes. The left plot in Figure 18 depicts the resulting efficient frontiers for the three scenarios for an assumption

of Gamma distributions for the individual claim sizes with an additional atom at 0 with probability 0.25 (parameters consistent with their first two moments from (23)) and all other parameters chosen as in (23). The right plot in Figure 18 shows

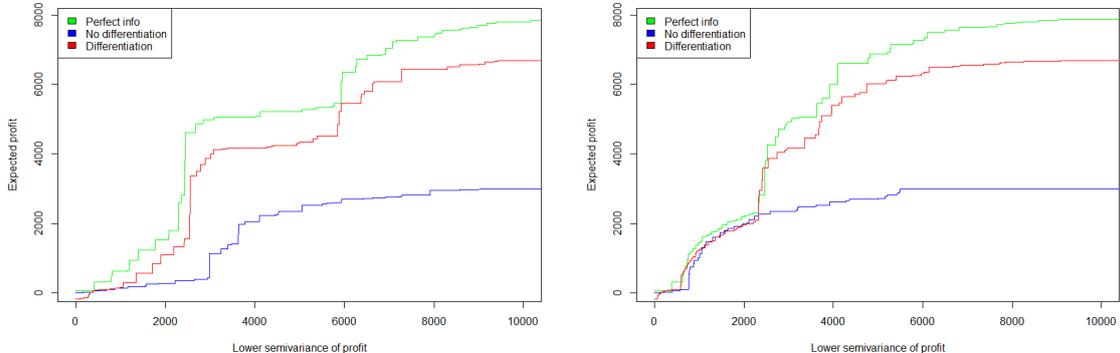


Figure 18: Efficient frontiers in the mean-LSV setup for the three scenarios under the assumption of Gamma-distributed H (left) and Log-Normal H (right).

the results for H being log-normally distributed risks (and again matching the first two moments). For instance, the intersection between the no-differentiation and differentiation scenario takes place at a much higher threshold.

Value-at-risk

Let us now instead consider the Value-at-risk

$$\text{VaR}_\alpha(\Pi) := \inf \{x \in \mathbb{R} : F_\Pi(x) > \alpha\}$$

for some level $0 < \alpha < 1$. This measure is particularly focusing on the tail of the loss (negative profit), when using small values of α . As for the LSV, we depict Monte Carlo results for the case of Gamma-distributed H and Log-normal H (Figure 19) risk types, where $\alpha = 0.025$. That is, the profit can be lower than the value of the abscissa in Figure 19 only with probability $\alpha = 0.025$, so that the more left in the abscissa one gets, the more risk-averse the strategy is. One observes that high values of $\text{VaR}_{0.025}(\Pi)$ can only be obtained by the no-differentiation case. In regions where that VaR-value can be attained by all strategies, the differentiation strategy always dominates the one without differentiation. Note that for this level of α , one virtually does not observe any difference between the case of light-tailed and heavy-tailed losses, which is also due to the size of the portfolio.

6.3 Utility functions

Utility theory is a classical tool to combine risk and profitability of an insurance undertaking in one function (see e.g. [35, 33]), so in this subsection we would like to briefly look at the problem posed in this paper from the utility point of view. Note that in this case the knowledge of the full loss distribution is needed, and not only the first two moments as in Section 4. Assume that the insurer bases decisions on a risk-averse (i.e., increasing and concave) utility function $u(x)$. The insurer's

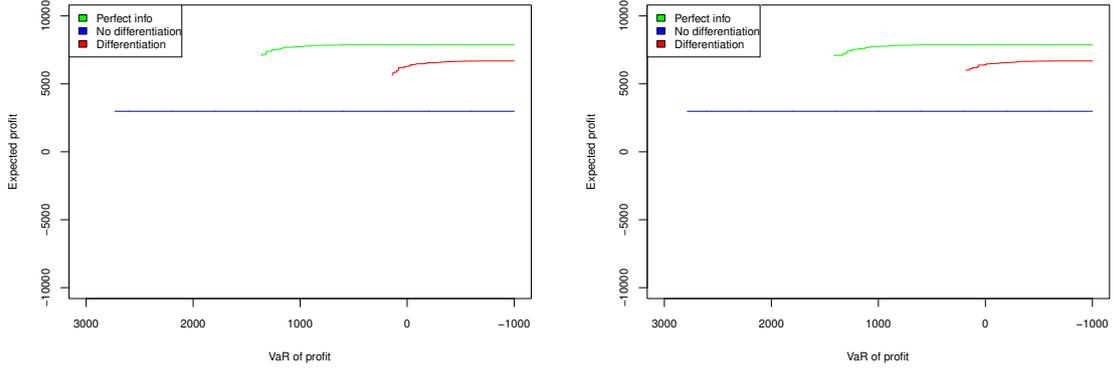


Figure 19: Efficient frontiers in the mean-VaR setup for the three scenarios for Gamma-distributed H (left) and Log-Normal H (right).

optimization problem is then modified as follows:

$$\max_{P_L, P_H} \mathbb{E}(u(\Pi)), \quad (32)$$

where the profit Π is given by the (4), which we can also write as

$$\Pi = \sum_{j=1}^{N_L} \Pi_j^L + \sum_{j=1}^{N_H} \Pi_j^H. \quad (33)$$

Firstly, the moment-generating function of each Π_j^L is

$$M_{\Pi_j^L}(t) = \mathbb{E}\left(e^{t\Pi_j^L}\right) = \mathbb{E}\left(e^{t(P_L - L)}\right) = e^{tP_L} M_L(-t).$$

Analogously, $M_{\Pi_j^H}(t) = e^{tP_H} M_H(-t)$. By independence and classical collective risk theory calculations (cf. [23]), we can then determine the moment generating function of Π :

$$M_{\Pi}(t) = M_{N_L}(\log M_{\Pi_j^L}(t)) \cdot M_{N_H}(\log M_{\Pi_j^H}(t)).$$

The same reasoning applies to the non-differentiation case with setting $P_L = P_H = P$. Finally, for differentiating pricing, an analogous derivation gives

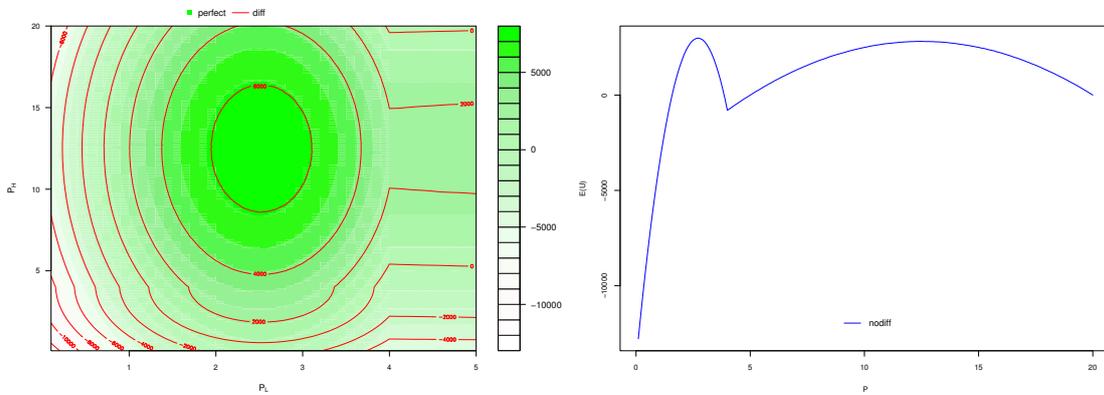
$$\begin{aligned} M_{\Pi}(t) = & M_{N_{L|L}}(\log M_{\Pi^{L|L}}(t)) \cdot M_{N_{H|L}}(\log M_{\Pi^{H|L}}(t)) \\ & \times M_{N_{H|H}}(\log M_{\Pi^{H|H}}(t)) \cdot M_{N_{L|H}}(\log M_{\Pi^{L|H}}(t)) e^{-tc(n)}. \end{aligned}$$

In each of the cases, $M_{\Pi}(t)$ can be inverted to obtain the c.d.f. $F_{\Pi}(x)$ of the profit, and the expected utility is then given by $\mathbb{E}(u(\Pi)) = \int_x u(\Pi(x)) dF_{\Pi}(x)$.

For a numerical illustration, assume now that $L \sim \text{Exp}(\alpha_L)$ and $H \sim \Gamma(\alpha_H, \lambda_H)$. To be consistent with (23), we choose $\alpha_L = \mu_L = 1$, $\alpha_H = \mu_H^2 / \sigma_H^2$, $\lambda_H = \mu_H / \sigma_H^2$. Since an explicit calculation of $\mathbb{E}(u(\Pi))$ is not feasible, we add here numerical results from a Monte Carlo simulation, simulating its value for each choice of P_L, P_H (across a discrete grid of mesh size 0.05) using 1000 runs. For the sake of comparison, we use three popular utility functions:

- linear utility $u(x) = x$ (leading to simply the expected value of the profit);
- exponential utility $u(x) = -e^{-Ax}$ for some risk aversion coefficient $A > 0$;
- quadratic utility $u(x) = x - Bx^2$, $x \leq \frac{1}{2B}$.

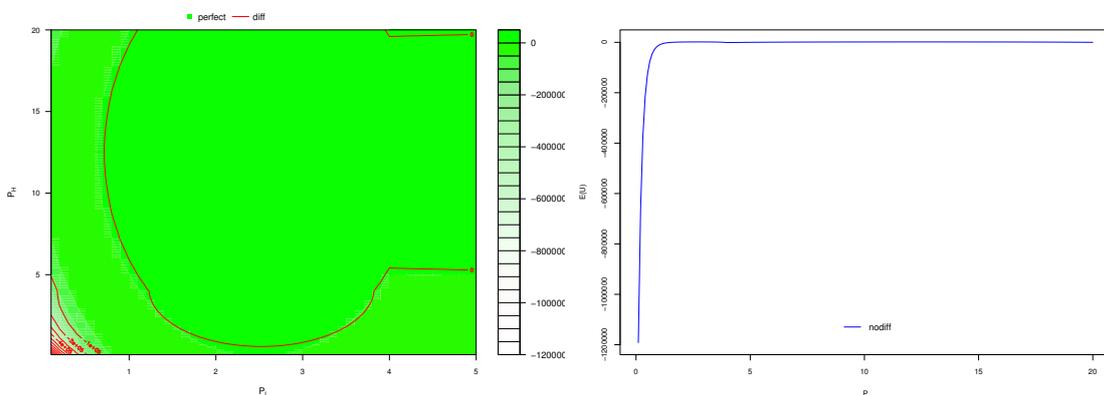
The results in Figures 20, 21, 22 show the expected utility for each of the available premium combinations and each strategy for these three utility functions. Figure 20 serves as a reference point since it represents the simple expected profit as before. One observes that the optimal solution clearly depends on the chosen utility function. With the chosen parametrization of the exponential utility function, the difference in the expected utility between the differentiation and not differentiation case is less prominent than in the quadratic utility as the marginal utility of the quadratic function is greater in this region.



(a) Perfect information (green heatmap) and differentiation (red level curves).

(b) No differentiation.

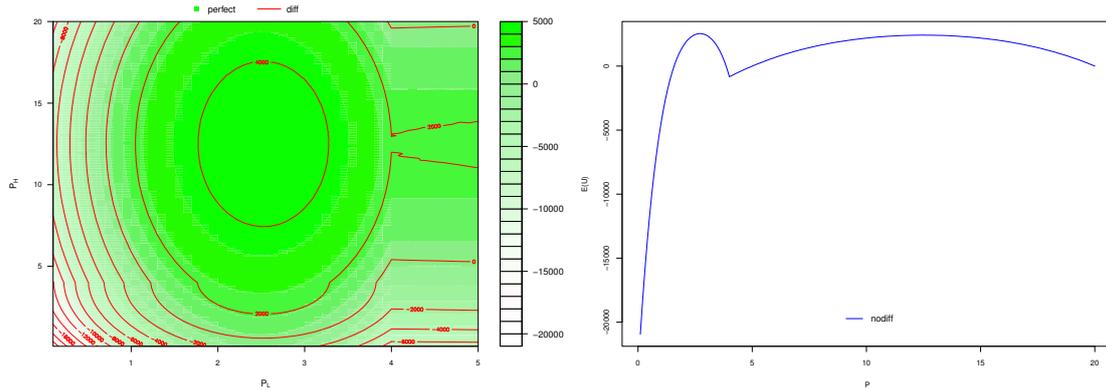
Figure 20: Expected linear utility as a function of premiums



(a) Perfect information (green heatmap) and differentiation (red level curves).

(b) No differentiation.

Figure 21: Expected exponential utility as a function of premiums ($A = 0.0005$)



(a) Perfect information (green heatmap) and differentiation (red level curves).

(b) No differentiation.

Figure 22: Expected quadratic utility as a function of premiums ($B = 0.00005$).

7 Conclusion

In this paper, we investigated the problem of risk categorization under the possibility of classification errors for an insurance company. We highlighted the impact of misspecification of risk classes on the company's profit, which is a relevant topic due to the growing use of black box techniques in classification. Resulting pricing errors may lead to adverse selection via a modified acceptance behaviour of individuals to enter a contract, potentially leading to extra costs due to lost market shares and loss of premium inflow. In a simple model with two risk types and piece-wise linear acceptance function, we distinguished three pricing scenarios: full information, undifferentiated pricing and costly price differentiation under error assumptions. In this framework, we studied the optimal solution for simply maximizing expected profit and more generally within a mean-variance framework, establishing efficient frontiers for the premium choices. The cost of the risk categorization as a function of population size will then eventually determine the optimal choice of premiums, and to what extent risk classification is profitable.

The simplicity of the introduced model allowed to quantify the effects and consequences of misspecification on the insurer's profit. Clearly, it will be of interest in future research to generalize the model assumptions in various directions. Beyond the extensions to more general acceptance functions and risk measures that we already address to a first extent in Section 6 of the paper, it will be of interest to extend the study to more than two risk categories. Another important direction will be to introduce market competition into this model (cf. [16, 31]), as well as the lapse behavior of policyholders between the different market players (see e.g. [5, 30]). Also, while our probabilistic acceptance model already covers a certain degree of randomness in the choice of insurance policies, it could be interesting to more explicitly include bounded rationality as well as other elements of inertia of policyholders and the markets in the modelling framework.

Statement: On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

- [1] H. Albrecher. Asymmetric information and insurance. In *Cahiers de l'Institut Louis Bachelier*, volume 20, pages 12–15, 2016.
- [2] H. Albrecher, A. Bommier, D. Filipović, P. Koch-Medina, S. Loisel, and H. Schmeiser. Insurance: models, digitalization, and data science. *European Actuarial Journal*, 9:349–360, 2019.
- [3] H. Albrecher and D. Daily-Amir. On effects of asymmetric information on non-life insurance prices under competition. *International Journal of Data Analysis Techniques and Strategies*, 9(4):287–299, 2017.
- [4] K. Antonio and E. A. Valdez. Statistical concepts of a priori and a posteriori risk classification in insurance. *AStA Advances in Statistical Analysis*, 96:187–224, 2012.
- [5] F. Barsotti, X. Milhaud, and Y. Salhi. Lapse risk in life insurance: Correlation and contagion effects among policyholders' behaviors. *Insurance: Mathematics and Economics*, 71:317–331, 2016.
- [6] E. Bauer and R. Kohavi. An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. *Machine learning*, 36:105–139, 1999.
- [7] P. L. Brockett, L. L. Golden, M. Guillen, J. P. Nielsen, J. Parner, and A. M. Perez-Marin. Survival analysis of a household portfolio of insurance policies: how much time do you have to stop total customer defection? *Journal of Risk and Insurance*, 75(3):713–737, 2008.
- [8] M. J. Browne and S. Kamiya. A theory of the demand for underwriting. *Journal of Risk and Insurance*, 79(2):335–349, 2012.
- [9] A. Charpentier. *Insurance, Biases, Discrimination and Fairness*. Springer, Heidelberg, 2023.
- [10] R. M. Corless, G. H. Gonnet, D. E. Hare, D. J. Jeffrey, and D. E. Knuth. On the lambert W function. *Advances in Computational Mathematics*, 5:329–359, 1996.
- [11] K. J. Crocker and A. Snow. The efficiency effects of categorical discrimination in the insurance industry. *Journal of Political Economy*, 94(2):321–344, 1986.
- [12] K. J. Crocker and A. Snow. The theory of risk classification. *Handbook of Insurance*, pages 281–313, 2013.
- [13] G. Dionne, N. Fombaron, and N. Doherty. *Adverse selection in insurance contracting*. Springer, Heidelberg, 2013.
- [14] G. Dionne and C. Rothschild. Economic effects of risk classification bans. *The Geneva Risk and Insurance Review*, 39:184–221, 2014.

- [15] C. Dutang. The customer, the insurer and the market. *Bulletin Français d'Actuariat*, 2012.
- [16] C. Dutang, H. Albrecher, and S. Loisel. Competition among non-life insurers under solvency constraints: A game-theoretic approach. *European Journal of Operational Research*, 231(3):702–711, 2013.
- [17] R. J. Finger. Chapter 6 risk classification. *Foundations of Casualty Actuarial Science*, pages 287–341, 2001.
- [18] N. Gatzert, G. Schmitt-Hoermann, and H. Schmeiser. Optimal risk classification with an application to substandard annuities. *North American Actuarial Journal*, 16(4):462–486, 2012.
- [19] M. Guillén, J. Parner, C. Densgsoe, and A. M. Perez-Marin. Using logistic regression models to predict and understand why customers leave an insurance company. In *Intelligent and other computational techniques in insurance: Theory and applications*, pages 465–490. World Scientific, 2003.
- [20] M. Guillen, A. M. Perez, and M. Alcañiz. A logistic regression approach to estimating customer profit loss due to lapses in insurance. *Document de Treball No. XREAP*, 13, 2011.
- [21] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman. *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer, 2009.
- [22] M. Hoy. Categorizing risks in the insurance industry. *The Quarterly Journal of Economics*, 97(2):321–336, 1982.
- [23] R. Kaas, M. Goovaerts, J. Dhaene, and M. Denuit. Modern actuarial risk theory. Technical report, Springer, 2008.
- [24] M. Khalilia, S. Chakraborty, and M. Popescu. Predicting disease risks from highly imbalanced data using random forest. *BMC medical informatics and decision making*, 11:1–13, 2011.
- [25] D. Kliger and B. Levikson. Pricing insurance contracts—an economic viewpoint. *Insurance: Mathematics and Economics*, 22(3):243–249, 1998.
- [26] N. Kyurkchiev and S. Markov. Sigmoid functions: some approximation and modelling aspects. *LAP LAMBERT Academic Publishing, Saarbrücken*, 4, 2015.
- [27] M. Lindholm, R. Richman, A. Tsanakas, and M. V. Wüthrich. Discrimination-free insurance pricing. *ASTIN Bulletin: The Journal of the IAA*, 52(1):55–89, 2022.
- [28] H. M. Markovitz. *Portfolio selection: Efficient diversification of investments*. John Wiley, 1959.
- [29] H. Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.

- [30] X. Milhaud and C. Dutang. Lapse tables for lapse risk management in insurance: a competing risk approach. *European Actuarial Journal*, 8:97–126, 2018.
- [31] C. Mouminoux, C. Dutang, S. Loisel, and H. Albrecher. On a Markovian game model for competitive insurance pricing. *Methodology and Computing in Applied Probability*, 24(2):1061–1091, 2021.
- [32] W. Olbricht. Tree-based methods: a useful tool for life insurance. *European Actuarial Journal*, 2:129–147, 2012.
- [33] G. C. Pflug and W. Römisch. *Modeling, measuring and managing risk*. World Scientific, Singapore, 2007.
- [34] D. Porrini. Risk classification efficiency and the insurance market regulation. *Risks*, 3(4):445–454, 2015.
- [35] M. Rothschild and J. Stiglitz. Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. In *Uncertainty in economics*, pages 257–280. Elsevier, 1978.
- [36] J. Sass and F. T. Seifried. Insurance markets and unisex tariffs: is the european court of justice improving or destroying welfare? *Scandinavian Actuarial Journal*, 2014(3):228–254, 2014.
- [37] H. Schmeiser, T. Störmer, and J. Wagner. Unisex insurance pricing: Consumers’ perception and market implications. *The Geneva Papers on Risk and Insurance-Issues and Practice*, 39:322–350, 2014.
- [38] P. Shi and K. Shi. Non-life insurance risk classification using categorical embedding. *North American Actuarial Journal*, 27(3):579–601, 2023.
- [39] R. G. Thomas. Some novel perspectives on risk classification. *The Geneva Papers on Risk and Insurance-Issues and Practice*, 32:105–132, 2007.
- [40] G. Tzougas and K. Kutzkov. Enhancing logistic regression using neural networks for classification in actuarial learning. *Algorithms*, 16(2):99, 2023.
- [41] H. R. Varian. *Intermediate microeconomics with calculus: a modern approach*. WW norton & company, 2014.
- [42] F. Vekeman, J. E. Piña-Garza, W. Y. Cheng, E. Tuttle, P. Giguère-Duval, A. Oganisian, J. Damron, M. Sheng Duh, V. Shen, T. B. Saurer, et al. Development of a classifier to identify patients with probable lennox–gastaut syndrome in health insurance claims databases via random forest methodology. *Current medical research and opinion*, 35(8):1415–1420, 2019.
- [43] W. Weidner, F. W. Transchel, and R. Weidner. Classification of scale-sensitive telematic observables for riskindividual pricing. *European Actuarial Journal*, 6:3–24, 2016.
- [44] G. M. Weiss and F. Provost. The effect of class distribution on classifier learning: an empirical study. Technical report, Rutgers University, 2001.

- [45] F. Zheng and G. I. Webb. A comparative study of semi-naive Bayes methods in classification learning. *AUSDM05*, pages 141–155, 2005.