UMEÅ UNIVERSITY

The deductibles impact on the risk premium

Ludvig Bergman

Umeå University
Department of Mathematics and Mathematical st atistics
Master thesis in mathematical statistics, 30 ECTS
Supervisor: Markus Ådahl
Spring 2023

Abstract

The aim of this master thesis is to derive methods that assesses the impact the deductible has on the risk premium of an insurance contract. The additive structure of a deductible necessitates approaches beyond treating it as a regular covariate in a generalized linear model for predicting the risk premium. Using simulated data, three methods are implemented to estimate a parameter, denoted as $\beta_d \in [0,1]$, which quantifies the proportion of the risk premium that remains after imposing a deductible d on the insurance contract.

The three implemented methods involve: 1) Regularized generalized linear models, 2) Utilizing the cumulative density function of the insurance contract, and 3) Estimating a discrete probability distribution with K-means clustering and a classifier. To compare the performance of these methods, they are tested against each other through a competition. The results reveal that the method employing the fitting of a discrete distribution yielded the best performance.

Självriskens påverkan på riskpremien

I den här uppsatsen härleds tre olika metoder där syftet är att skapa en oberoende variabel $\beta_d \in [0,1]$ som beskriver hur stor del av ett försäkringskontrakts riskpremie som kvarstår, efter att en självrisk d ålagts på försäkringskontraktet. Självrisken kan inte användas som vilken oberoende variabel som helst på grund av den additiva struktur som självrisken innehar. De tre olika metoderna som har implementerats skiljer sig åt genom att 1) använder sig av upprepade regulariserad GLM-modeller för olika självrisknivåer, 2) nyttjar försäkringskontraktets fördelningsfunktion, och 3) skattar en diskret sannolikhetsfördelning med hjälp av klustring och en klassificerare. De tre olika metoderna testades sedan mot varandra i ett spel. Metod tre presterade bäst i spelet, då den hade lägst procentuell avvikelse från den sanna riskpremien.

Acknowledgments

I would like to express my gratitude to Per Wilhelmsson and Johan Palmquist from Länsförsäkringar AB for our cheerful collaboration. Their knowledge and interest within the subject of insurance pricing have been extremely helpful. Also, the culture at the department of mathematics and mathematical statistics at Umeå University must be acknowledged. A very welcoming environment where the staff is only a door knock away.

Contents

1 Introduction

An insurance contract is an agreement between a policyholder and an insurer, were the insurer compensates the policyholder when a certain event of loss occurs. The compensation is commonly called a *claim*. For this, the policyholder pays a premium to the insurer. Hence, the economic risk is transferred from the policyholder to the insurer. Since the loss of a policyholder is unpredictable for the insurer, one may consider this as a random variable. The total cost for the insurer is the sum of many losses, and by the law of large numbers, this is much more predictable than the expected cost of one insurance contract. So, the insurance company can predict their total expected cost with adequate precision, assuming no extreme events occur – the pricing model is globally unbiased. To cover the cost, the insurer has to charge a sufficiently high premium. Without taking vulnerability to insolvency, costs for staff and a mark-up for profits into consideration, the aggregated premium should be equal to the total expected cost, and thereafter allocated over all customers. An easy method to solve this is by dividing the total premium with the number of customers, letting every customer pay the average premium.

But an insurance company often sells a various amount of types of insurance contracts, and the expected cost for a house insurance is likely to be higher than for a bicycle, for example. This makes it unreasonable to assume that the premium for the bicycle insurance should be equal to the house insurance premium. The portfolio of insurance contracts could therefore be divided into segments, with different insurance types in each segment. Further, each segment may be divided into sub-segments, that each contains different characteristics specific for that sub-segment. For instance, a car insurance is a very wide segment. The expected cost varies with different types of cars, drivers and geographic location. One sub-segment may be young males between 18-22 years who owns a sedan with horse powers in some interval. Each sub-segment should face their fair premium. If one sub-segment is paying more than their fair premium, and conversely, another subgroup paying less than their fair premium, the insurance company may loose the customers who are overpriced to a competitor with a more efficient pricing model, and are left with the under priced segment, leading to increasing losses. Hence, the insurance company wants a locally unbiased pricing model.

The premium payed by the customer consists of several parts. The insurance company has to pay for costs such as staff, office rent, taxes etc. Also, the company may have some mark-up to make a profit. But the major part of the premium is dedicated to cover for the expected loss of an insurance contract. This part is called the *risk premium*, here denoted by π.

1.1 Multiplicative models and pricing with GLMs

The risk premiums are determined with the use of a number of *rating factors*. These can describe characteristics of the insurance, the policy holder or the product that the policy holder insures; anything that has a relation to the risk premium is a candidate to be included as a rating factor. To estimate the risk premium, it is useful to represent the target variable with relative changes. It would be more fair to have a percentage increase instead of an absolute increase in the premium between two rating factors. This is a motivation for the use of multiplicative models. Let μ_{ij} be the expectation for the target variable, which could be the size of the claim, or the number of claims. Let γ_{ij} be a rating factor with $i = 0,1,2...$ as the number of factors and $j = 1,2...$ the number of insurance contracts. A multiplicative model with three rating factors is:

$$\mu_{ij} = \gamma_0 \gamma_{1i} \gamma_{2j} \quad (1)$$

If γ_{12} is 1.25, it would mean that μ_{12} is 25 percent higher compared to μ_{11}, if $\mu_{11} = 1$. Thus, the percentage increase will be equal over all rating factors, no matter the current level of premium. If instead an additive model were used, i.e., $\mu_{ij} = \gamma_0 + \gamma_{1i} + \gamma_{2j}$, an absolute increase in the premium will have a large effect when the premium is small, but a very negligible effect when the premium already is large (Ohlsson and Johansson (2010)).

Now, take the natural logarithm of both sides of (1), and let:

$$\beta_1 = \ln(\gamma_0)$$
$$\beta_2 = \ln(\gamma_{1i})$$
$$\beta_3 = \ln(\gamma_{2j})$$

If $\ln(\mu)$ is the dependent variable, it can be described by covariates $\mathbf{x} = [x_1\ x_2\ x_3]^T$ which, when multiplied with $\boldsymbol{\beta} = [\beta_1\ \beta_2\ \beta_3]$ gives:

$$\ln(\mu) = g(\mu) = \eta = \boldsymbol{\beta}\mathbf{x} \quad (2)$$

With g(·) as the *link function*; here, the log-link (Olsson (2002)). When a distribution for μ is specified, and belongs to the exponential family of distributions (EDF), (2) is a generalized linear model (GLM). The EDF takes the general form:

$$f(y|\theta,\phi) = exp\left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right]$$

where θ is the canonical parameter, representing the location, and ϕ the dispersion parameter that represents the scale (Olsson (2002)).

Actuaries often use GLMs for pricing insurance contracts and to determine the risk premium. Let $Z \in [0,\infty)$ and $N \in \mathbb{N}$ be two random variables each describing the claim severity and the number of claims, respectively. Then the total claim amount for an insurance contract S is given by the compound random variable S:

$$S = \sum_{j=1}^{N} Z_j$$

Here, Z and N are assumed to be independent, i.e. the loss of each claim does not depend on the number of claims. The individual claim sizes $Z_j, j = 1,2,..$ is assumed to be i.i.d. The number of claims N is a counting process – it is industry standard to let N follow a Poisson distribution. That is, $N \sim Poi(\theta)$.

To model the claim severity Z, the distribution must be continuous and often right-skewed. A very popular choice for this purpose is to use the gamma distribution, although log-normal and inverse Gaussian are two other possible distributions. (Ohlsson and Johansson, 2010). Using gamma distribution for the claim severity is just as common as using the Poisson distribution for the claim frequency.

Now, due to the independence assumptions and the distributions of the underlying random variables, the expected value of S is:

$$\mathbb{E}[S] = \mathbb{E}[N]\mathbb{E}[Z]$$

Let $\mathbf{X_i} = [X_{1i}, ..., X_{pi}]$ be a vector of covariates that describes all the available information regarding insurance contract i. As described above, pricing actuaries often seek to use a multiplicative structure to model the effect of the covariates on the dependent variable, which can be the claim size Z, or claim frequency N. The purpose for this is to obtain relative differences between the covariates. If the premium to own a car for one gender group is 20 percent higher than for the other group, then this relation should hold for another covariate as well. If an additive structure were used, the difference between covariates would not be relative, making the premiums very unfair for large differences (Ohlsson and Johansson, 2010). This motivates the use of GLMs with a log-link, which provides a multiplicative model for both the claim size and claim frequency. Another perk of using GLMs is the ability to draw inference on what factors that affect the risk premium, which is an important part for actuaries. Thus, an expression for the mean claim severity for insurance contract i, with the use of log-link is:

$$\mu_{Z_i} = \exp(\beta_{z,0} + \beta_{z,1} x_{1,i} + ... + \beta_{z,p} x_{p,i})$$

Using the corresponding notation and link function, the expression for the mean claim frequency for insurance contract i:

$$\mu_{N_i} = \exp(\beta_{n,0} + \beta_{n,1} x_{1,i} + ... + \beta_{n,p} x_{p,i})$$

Since the length of insurance contracts can be different between customers, one must take this into account when predicting the claim frequency by controlling for the exposure w of the insurance contract. Under the log-link, the Poisson regression is:

$$\ln(\mu) = \boldsymbol{\beta}\mathbf{x}$$

Dividing μ with the exposure generates:

$$\ln\left(\frac{\mu}{w}\right) = \boldsymbol{\beta}\mathbf{x}$$

Using logarithm rules, moving $\ln(w)$ to the other side of the equality sign and exponentiate leads to:

$$\mu = w \exp(\boldsymbol{\beta}\mathbf{x})$$

Controlling for the exposure of an insurance contract in this way is called using an *offset*. Now, with μ_{N_i}, μ_{Z_i} and w_i, the expression for the risk premium for insurance contract i can be defined as:

$$\pi_i = w_i \mathbb{E}[N|\mathbf{X} = \mathbf{x_i}]\mathbb{E}[Z|\mathbf{X} = \mathbf{x_i}] = w_i \mu_{Ni} \mu_{Zi} \tag{3}$$

1.2 Deductibles and the purpose of the study

A deductible is the part of a damage that is not covered by the insurer, but by the policy holder. There are multiple reasons for the use of deductibles: It decreases the premium payed by the policy holder, since the policy holder is paying a share of the expected cost of the damage. This also creates risk-minimizing incentives for the policy holder, since part of the risk is transferred back to the policy holder. For the insurer, a deductible decreases the total loss obligation, which has a large implication in situations when many losses are occurring simultaneous, for example after extreme floods. When a loss is below the deductible, the insurance company will not cover the loss. This, combined with the possible risk-minimizing incentives from the customer leads to a smaller administrative burden for the insurance company, as not as many claims will be reported to the insurer.

The deductibles can have different structures – a common option is that the customer gets to choose a fixed level of deductible among pre-specified amounts. Another choice is that the customer pays a fixed amount and some percentage of the exceeding amount.

Given the information above, it is a reasonable assumption that the deductible affects the risk premium, and should therefore be used as some sort of input for the model. But to use it as just another covariate in the GLM is necessary, but not enough, since a deductible does not follow the multiplicative structure as most other covariates. Rather, the deductible has an additive structure. If the deductible followed a multiplicative structure, all customers with a deductible of 10 000 SEK should get the same relative rating as the customers with a deductible of 5000 SEK. Assume that two customers have an expected claim severity of 25 000 and 12 000 SEK, respectively, and are given the two options of deductibles mentioned above. Then, the expected claim severity after subtracting the deductible is 15 000 and 20 000 SEK for the first customer, and correspondingly 2 000 and 7 000 SEK for the second customer. This gives a ratio of $\frac{3}{4}$ for the first customer, and $\frac{2}{7}$ for the second. Under a multiplicative structure, these ratios would be equal for both customers.

So, the deductible should be incorporated as an input to estimate the risk premium, but it is not sufficient to use it as any other covariate due to its additive structure. Let d be the deductible for an insurance contract. Define $Z_d = max(0, Z - d)$ as the claim severity after the deductible has been subtracted, i.e., the actual cost for the insurance company. Then the *deductible risk factor* for a given deductible d can be defined as

$$\beta_d = \frac{\mathbb{E}[Z_d|\mathbf{X} = \mathbf{x}]}{\mathbb{E}[Z|\mathbf{X} = \mathbf{x}]}$$

The purpose of this study is to evaluate a theoretically and numerically correct method to estimate β_d that is user-friendly and easy to implement. If β_d is estimated properly, the interpretation of β_d would be that with a deductible of size d, the risk premium decreases by $(1 - \beta_d) \cdot 100$ percent, and hence, follows a multiplicative structure. β_d can then be incorporated in (3) as a covariate. This will give a new expression for the risk premium: for insurance contract i, with the deductible d, the risk premium can be defined as:

$$\pi_i = w_i \mathbb{E}[N|\mathbf{X} = \mathbf{x_i}]\mathbb{E}[Z_d|\mathbf{X} = \mathbf{x_i}] = w_i \beta_d \mu_{Ni} \mu_{Zi}$$

As of the spring of 2023, there are no known previous studies within this topic that quantifies an expression for β_d. Hence, the research gap within this area is large, and hopefully, this master thesis can make a contribution to the research of insurance mathematics.

The methods have been implemented on actual insurance data in addition to the simulated data presented in this study. Due to corporate confidentiality, these results will not be described here. But the methods can be generalized to other types of data within property insurance. The structure of the deductibles in this study is only expressed in monetary terms. That is, in percentage of price base amount that can be converted to SEK or directly in SEK. The target of this study is property insurance, and limited to using quantitative methods. As will be described later, the psychological aspects of insurances is large, and the reader is urged to keep this in mind.

1.3 The psychology of insurance contracts

When the buyer of the insurance contract has more information regarding their risk behaviour than the insurer does, adverse selection exists between the policy holder and the insurance company. With this information asymmetry, the policy holder can use this information to take more optimal decisions. The adverse selection can take its shape in different forms, such as information about the probability of a claim, or the severity of the claim. The classic assumption is that for policyholders with a riskier behaviour than the average policyholder, who are aware of this, have incentives to choose a lower deductible. The opposite would hold for the less riskier policyholder. There have been many attempts to confirm this empirically. A literature review by Cohen and Siegelman (2010) found various results for different types of insurance contracts.

Adverse selection can be viewed as an unfavourable selection, in the view of the insurer. It is not easy to prove that adverse selection exists. One reason for this may be due to the theory of propitious selection, a term first coined by Hemenway, that would work as a counter force to adverse selection. Propitious selection is a favourable selection, meaning that risk-averse people will buy more insurance coverage while at the same time be less prone to taking risks, and vice-versa for people with a riskier behaviour. This is a favourable behaviour since the insurer receives more premium with less claim costs.

Moral hazard on the other hand, can be described as the change in human behaviour to prevent economic loss after the purchase of insurance. The theory is that when an individual purchases insurance with complete coverage (without deductible) to a loss, he does no longer have the same incentives to take precautionary actions. Since the beginning of the 1960s, economists have believed that two possible solutions to moral hazard are: 1) the insurance company observes to which extent the insurer tries to prevent loss, and 2) imposing a deductible on the insurance contract, leaving the policy holder with an incomplete coverage against a loss (Shavell (1979)). The first solution is complicated and inefficient, while the second is well established and already in use.

Does moral hazard exist in the real world? Just as with adverse selection, it is hard to prove, since it involves human behaviour, which can be tedious to quantify. Wang et al. (2008) studied the impact of increasing deductibles on automobile insurance contracts in Taiwan. A large majority of these contracts has the structure where the deductible increases after a claim, which makes it suitable for analyzing any existence of moral hazard. The authors found evidence that moral hazard exists in the automobile insurance industry in Taiwan. It is important to note that the evidence is only for that market, and more specific, regarding that type of product. A completely different product, say housing apartment buildings, probably exhibits a completely different behaviour.

The remaining part of this thesis is organised as follows. In section two, relevant theories regarding the methods will be explained. Section three describes how this study is accomplished. Thereafter, the methods will be evaluated in section four, and discussed in section five.

2 Theory

2.1 The distribution of claim severity

When a claim is reported, the insurance company will only see the actual claim severity. One can append the deductible to see the total cost of the damage, but as written in the section about deductibles, the insurance company will never see the severity of the claims which does not exceed the given level of deductible. When the data has this structure, the distribution is *lower-truncated*. If Z is the total cost of a damage with density $f(z;\mathbf{x})$, and lower-truncate Z as $Z_d = max(0, Z-d) = (Z-d)_+$, with d as the truncation level i.e. the deductible, then the density of Z_d is:

$$f_{(d,\infty)}(z;\mathbf{x}) = \frac{f(z;\mathbf{x})\mathbb{1}_{[z>d]}}{1-F(d,\mathbf{x})}$$

Were $\mathbb{1}_{[z>d]}$ is an indicator function that takes value one if the realized claim z exceeds the deductible d, and zero otherwise, and $F(d,\mathbf{x})$ is the CDF of Z. The expected value of the truncated random variable is:

$$\mathbb{E}[(Z-d)_+] = \int_d^\infty (z-d) f_{(d,\infty)}(z;\mathbf{x})\, dz = \mathbb{P}[Z>d](\mathbb{E}[Z|Z>d]-d) = \mathbb{P}[Z>d]\mathbb{E}[Z-d|Z>d]$$

If the distribution is truncated, the risk of including a biased model is severe. Figure 1 shows two histograms generated from a gamma distribution. The lower figure contains all observations. The upper figure contains only the observations who fulfills $(Z-d)_+$, i.e, the distribution is left-truncated with d as the threshold. The mean of the two distributions are 7403 for the full distribution, and 4359 for the left-truncated distribution when excluding all zeros. This highlights the problem one is facing when handling a truncated distribution; the original distribution can be very different from the observed. To connect this problem with the thesis subject, the distribution of the claims below the lowest deductible is unknown. However, it is assumed that the frequency model that estimate the number of damages captures this uncertainty. In the end, it is the risk premium that is of interest, which is determined by the product of the claim frequency and the claim severity. This mean that below a certain deductible, it is assumed that the frequency model incorporates this uncertainty.

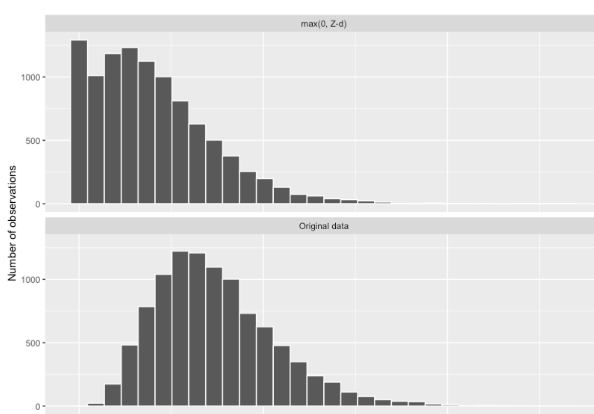


Figure 1 *Comparison between the left-truncated and original distribution*

2.2 Balance property and the choice of link function

Using the log-link is favourable since it leads to multiplicative models. However, it is not always the most appropriate choice of link function in terms of statistical properties.The *canonical link* is a link function such that $\eta = g(\mu) = \theta$. The canonical link for the gamma and Poisson distributions is $\eta = \mu^{-1}$ and $\ln(\mu)$, respectively. Thus, for the Poisson distribution, the canonical link is equal to the log-link, while this is not true under the gamma distribution.

The parameters β in a GLM are estimated using the method of maximum likelihood. One favourable property of using a canonical link is that the log-likelihood function is strictly concave, implying that, if it exists, the critical point is unique. Another favourable implication is that, using a canonical link, the *balance property* is always fulfilled: Assume that Z_i, $i = 1,2,...,N$ are independent random variables, and let β^{MLE} denote the parameter estimated with maximum likelihood. If $\hat{Z}$ is modelled with a GLM under the canonical link, then there is balance property on portfolio level, meaning that the sum of the predicted values of the training data is equal to the true values of the training data:

$$\sum_{i=1}^{N} \mathbf{E}_{\hat{\beta}^{MLE}}[w_i \hat{Z}_i] = \sum_{i=1}^{N} w_i Z_i$$

with w as the weight (Wüthrich and Merz (2023)). This means that the portfolio mean is unbiased against the training data as long as the criterions above are fulfilled, no matter how poorly estimated the GLM is. A way to see this is to predict new values with the GLM using the training data. If the sum of the predicted values equals the sum of the target parameter values, then the balance property is fulfilled. Note that this does not say anything regarding the performance of the GLM on local level, nor on the test data.

If the balance property is not fulfilled, the estimated intercept $\hat{\beta}_0$ must be re-balanced. This can be done by shifting the intercept: Let $\hat{\beta}_0^{MLE}$ denote the intercept of a GLM regression, were the canonical link has not been used, implying that the balance property is not fulfilled. Predict new values values from the GLM regression, from the training data. These are denoted *predicted values*. By shifting $\hat{\beta}_0^{MLE}$ with $-\ln \frac{\sum \textit{Predicted values}}{\sum \textit{True values}}$, the balance property is met. The *true values* are the dependent values that were used to train the GLM regression. The model is now globally unbiased on the training data (Wüthrich and Merz (2023)).

2.3 Regularization and elastic net

The objective of a GLM regression is to maximize the likelihood and get a good fit to the target variable. If the target variable is quantitative like the claim severity, this is equivalent to minimizing the squared-error loss. This solution often has a low bias, in exchange for a large variance, so the mean-squared error could potentially be improved. Also, when fitting a GLM, one has to consider which covariates to include in the model. If the number of possible covariates to include in the model is large, evaluating all possible combinations of covariates is infeasible. By shrinking the regression parameters, or setting some of them to zero, model selection is performed while simultaneously possibly decreasing the mean-squared error (Hastie et al. (2015)). This is called regularized, or penalized regression.

Regularized regression introduces a penalty term λ that controls the size of the regression parameters. λ can be considered as a budget that the parameters have to relate to in how well the model can fit the data. With a tight budget, λ is small, and the regression parameters can not afford to be large. This gives a low variance, but a high bias. Conversely, with a large budget, the regression parameters can afford to take large values, which could possibly lead to a low bias, but with an increased risk of overfitting. Using a budget to the regression in this way can control the variance of the predictions. λ is determined by choosing a grid of possible values, and then performing k-fold cross-validation to determine the cross-validation error. The λ that gives the lowest cross-validation error is then selected as the final candidate for the model.

Before fitting the parameters, the covariates are standardized to have mean zero and unit variance. The two most common regularization techniques are ridge regression and lasso. They can be generalized as:

$$\tilde{\beta} = \underset{\beta}{argmin}\left\{\sum_{i=1}^{N}(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda\sum_{j=1}^{p}|\beta_j|^q\right\} \tag{4}$$

With $\beta_j, j = 1,2,...,p$ as coefficient parameters, x_j as the covariates, $q \in \{1,2\}$ and y_i as the dependent variable. From (4) it can be seen that as λ increases, the coefficients shrinks toward zero. When $q = 1$, (4) is a lasso regularization, as the second term is $\lambda\sum_{j=1}^{p}|\beta_j|$. This implies that some of the coefficients can take values exactly equal to zero, and thus, lasso can be considered as a type of variable selection. On the other hand, if $q = 2$, the second term is $\lambda\sum_{j=1}^{p}\beta^2$, and (4) turns into ridge regression. Ridge regression differs from lasso in the way that all coefficients shrink towards zero as λ increases, while the lasso forces some of the coefficients to be exactly zero when λ is sufficiently large (James et al. (2021)). The advantage of lasso compared to ridge regression is that it automatically performs subset selection. A disadvantage of lasso is that if there is a group of highly correlated covariates, then the lasso tends to select one variable from the group without any consideration of which covariate that is being chosen. Under this situation, when the number of observations exceeds the number of predictors, the predictive performance of lasso is often dominated in favor of ridge regression (Zou and Hastie (2005)). However, a disadvantage of ridge regression is that it does not perform variable selection as lasso does.

A combination of lasso and ridge regression is the *elastic net*, invented by Zou and Hastie (2005). Replace the penalty term in (4) with:

$$\lambda\sum_{j=1}^{p}(\alpha|\beta_j| + (1-\alpha)\beta_j^2 \tag{5}$$

The elastic net penalty performs both variable selection and continuous shrinkage of the covariates, while also handling groups of correlated covariates. The elastic net has two hyperparameters to be tuned, $\alpha \in [0,1]$ and λ, which are chosen with k-fold cross-validation. One can see that if $\alpha = 1$, (5) is the standard lasso. Conversely, if $\alpha = 0$, then (5) is a ridge regression.

2.4 Generalized additive models

It is unlikely that the continuous covariates exhibits a completely linear relationship with the target variable, whether it is claim severity or claim frequency. Hence, it could be useful to allow for non-linearity in the model to find this relationship. By splitting the domain of one single predictor into multiple continuous intervals, and use a *basis function* for each interval, the relationship between the target variable and the covariate will be non-linear depending on the basis functions. A regression spline is a type of basis function that puts constraints on the knots, where the intervals intersect each other. When the first and second derivative are continuous at each knot, the regression spline is called a cubic spline. These constraints on the knots leads to a very smooth function that can capture a high degree of non-linearity. One problem that may arise with a cubic spline is that the fit gets very uncontrolled and wiggly at the boundaries. A natural cubic spline handles this by invoking a third constraint, that the basis functions are linear beyond the boundary knots (Hastie et al. (2009)).

Figure 2 shows a scatter plot of claim severity and log-transformed insured amount. The line in the figure is the fit from a natural cubic spline with five degrees of freedom. A non-linear relationship seems to exist.

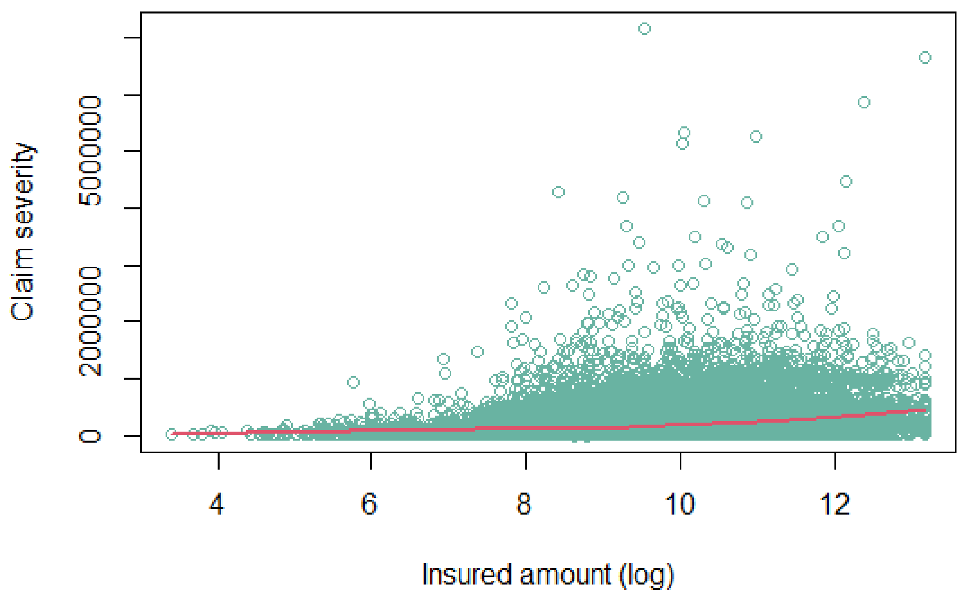


Figure 2 *Scatter plot with a natural cubic spline*

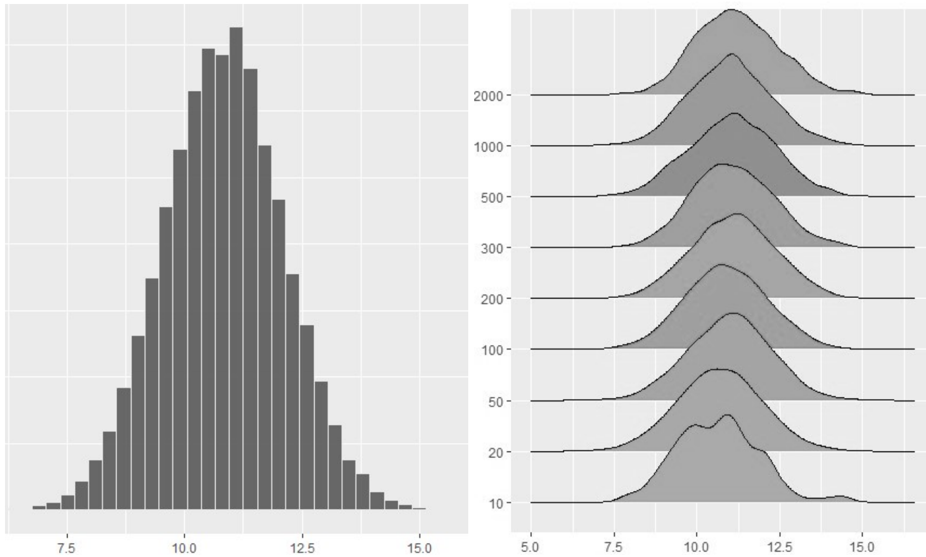
3 Method

To derive $\beta_{Z,d}$, three different methods are used. These methods differs in various ways, while also having some similarities. The first two methods are based on an approximation of the claim severity μ_Z. Recall from (3) that as the end goal is to predict the risk premium π, the claim frequency μ_N must also be estimated. This section will describe the three methods, along with the estimation of μ_Z and μ_N. The programming language R has been used throughout the entire study.

3.1 The data

One of the data sets used in this study is a simulated data set, with characteristics based on a real data set from Länsförsäkringar AB. The data set consists of insurance contracts on apartments buildings with 19 variables. The number of observations is 1 200 000, and the reported claims are 38 205. The deductible is in percentage of a price base amount, ranging from ten to 2000 percent, with a median of 50 percent. The lowest value of the net claim severity $Z_d = max(0, Z-d)$ is 321 SEK, while the highest value is 7 103 987 SEK, with a median of 50 477, indicating a positively skewness. Figure 3a shows a histogram of the claim severity in log-scale, confirming the positively skewed distribution. Figure 3b visualizes the densities for the claim severity by all levels of deductibles. From this, it can be noted that the range of Z_d is quite wide over all levels of deductibles, although it seems that as the deductible increases, the densities tends to be more right-oriented. Descriptive statistics in table 1 is confirming that the mean and the median claim severity grows as the deductible increases, with some exceptions. The column "Deductible" in table 1 represents the claim severity Z_d with nine different choices of deductibles, except for the last row, which is the total claim severity Z i.e., when the deductible has been added to the claim severity.

The covariates are mostly variables that describe the buildings, such as number of floors, building class, wether the apartments are rentals or condos, and gross floor area; other types of covariates are which company that handles the insurance policy (Länsförsäkringar consists of a group of smaller companies, separated by geographic regions). Of the 19 variables, 14 are considered as possible covariates, and of these, five are discrete, four continuous, and the other five are categorical. Since the focus of this study is not about the covariates itself, a more thorough explanation will not be given here. The simulated data were "sampled" from between years 2013 and 2022. The data from 2022 were used as test data, with the remaining part used as training set.



(a) *Histogram of the claim severity Z_d (log scale)* **(b)** *Claim severity Z_d by level of deductible (log scale)*

Figure 3

Table 1 *Descriptive statistics of claim severity*

Deductible	Min	Median	Mean	Max	n
10 %	2401	41 134	102 290	2 008 645	158
20 %	321	43 639	98 564	5 844 560	20 835
50 %	615	60 184	127 659	5 246 948	4839
100 %	538	52 385	118 880	7 103 987	4968
200 %	1031	67 209	147 313	6 575 980	3959
300 %	1219	62 758	139 678	2 143 082	720
500 %	1289	66 049	143 594	3 258 134	1071
1000 %	1139	60 473	131 783	3 218 331	1206
2000 %	2159	66 832	165 913	2 903 249	449
Total claim severity Z	7131	89 380	172 068	7 150 487	38 205

3.2 Estimating the total claim severity

The estimation of the claim severity with the deductible added, is accomplished using a regularized elastic net regression. 15 covariates are included in the original model, and of those, three were continuous. The variable "insured amount" ranged from 30 to 531 251. Due to the large spread, this variable was log-transformed. Figure 2 shows the relation between total claim severity and insured amount after the log-transformation, showing a non-linear behavior. Therefore, a natural spline of insured amount with five degrees of freedom is used as a predictor on claim severity. This was the only variable that was transformed.

The model was fitted using five-fold cross-validation, and the intercept was shifted to fulfill the balance property. The α with the lowest cross-validation error turned out to be 0.7, implying that the elastic net tends toward a regular lasso, while still maintaining some characteristics from ridge-regression. Figure 4 shows the cross-validation deviance from the saturated model, for different values of λ. The two vertical lines show the λ that returns the lowest deviance of all values (λ_{min}), and the value that returns the model with the largest penalty while still maintaining a cross-validated deviance error within one standard deviation (λ_{1SD}). Choosing the latter λ is often preferred since it gives a less complex model without sacrificing too much predictive performance (Hastie et al. (2009)). But after predicting $\hat{\mu}_Z$ on the training data, using λ_{min} generated a lower RMSE than using λ_{1SD}. Looking at descriptive statistics of the predictions and comparing with the test data, the conclusion is that λ_{min} predicts the true values with a greater precision, and will therefore be used.

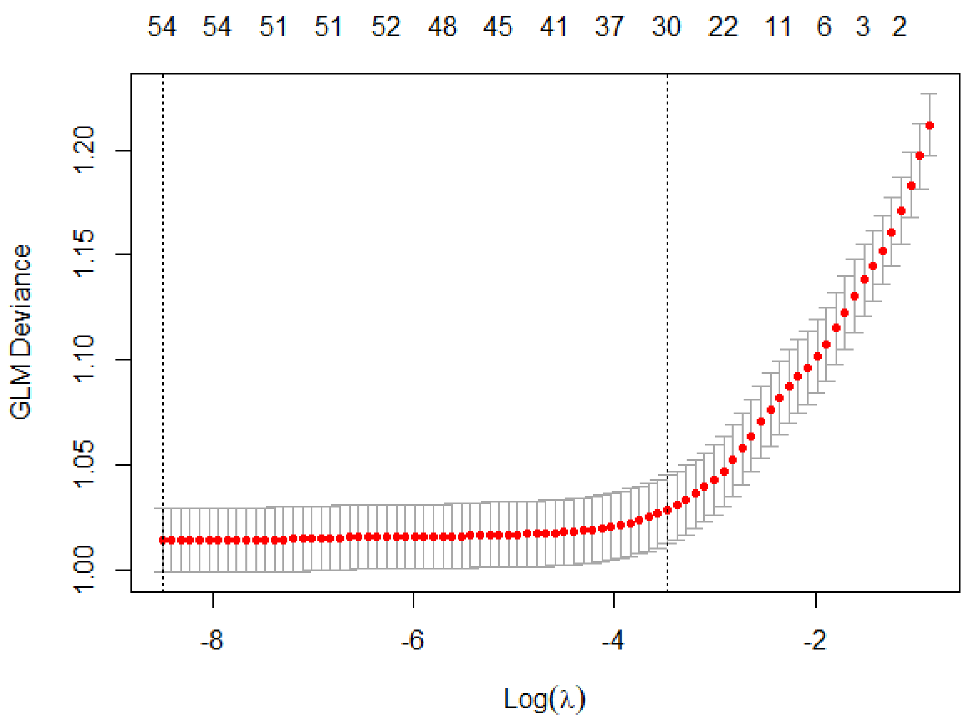


Figure 4 *Cross-validation results. The left vertical line is λ_{min}, and the right λ_{1SD}*

3.3 Estimating the claim frequency

The method to approximate $\hat{\mu}_N$ is very similar to the modelling of $\hat{\mu}_Z$, The obvious change is that a Poisson GLM is fitted using penalized maximum likelihood instead of a gamma GLM. Another difference is that the deductible is included as a covariate. Also, the duration is used as an offset. The α that generated the lowest mean cross-validated error was 0.729. Since the canonical link for a Poisson model is the log-link, the balance property is already fulfilled.

The claim frequency was predicted for each deductible in the vector $\mathbf{d}$ = [0, 20, 40, 60, 80, 100, 150, 250, 350, 450, 600, 1200, 1500, 1800, 2100, 2400] The same vector will be used in the upcoming methods. For each element in **d**, $\beta_{N,d} = \frac{\hat{\mu}_{N,d}}{\hat{\mu}_{N,0}}$ was derived at deductible d. This $\beta_{N,d}$ is explaining how much of the claim frequency that remains at the deductible d. $\beta_{N,d}$ will later be multiplied with $\beta_{Z,d}$ from the three methods that will be described, generating a final β_d that describes how much of the predicted risk premium is remaining after a deductible d has been imposed in the insurance contract.

Since the decrease in claim frequency at deductible d is already incorporated in β_d, using the deductible as a covariate in the Poisson model to predict claim frequency would be equivalent of using the same effect twice, and it would probably underestimate π_i. Therefore, the deductible coefficient in the frequency model is null-ed, leaving the other covariates untouched.

3.4 Method one – Estimate $\beta_{Z,d}$ using multiple GLMs

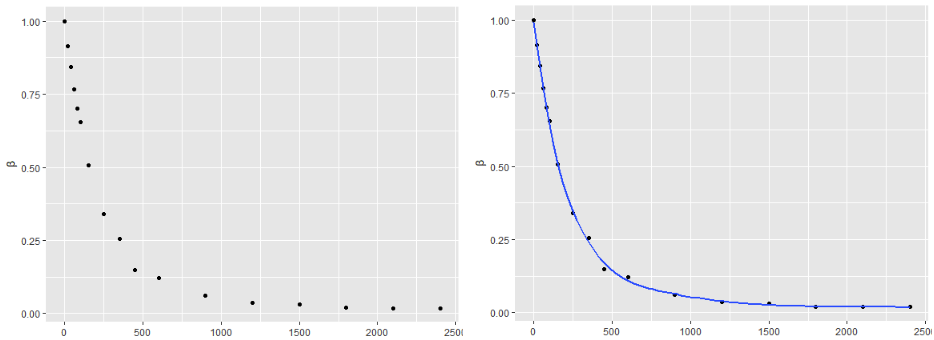
By fitting a GLM with the claim severity Z as target variable, and fitting a similar GLM but with Z replaced with the net claim severity Z_d, the deductible risk factor $\beta_{Z,d}$ can be estimated by taking the quotient of the predicted values $\hat{\mu}_{Zd}$ and $\hat{\mu}_Z$. This would give a unique $\beta_{Z,d}$ for each customer for a given choice of deductible, which is favourable since it will provide the insurance company with information on how much the deductible will decrease the expected risk premium for the specific customer. The GLMs are fitted through the use of regularized elastic nets with gamma as the distribution and with log-link. α is tuned by five-fold cross-validation, and thereafter the optimal choice of λ is decided, also by the use of five-fold cross-validation. After the model is fitted, the intercept is shifted in order to fulfill the balance property.

After subtracting the deductible from the claim severity Z, an issue arises in that only the insurance contracts greater than the deductible will be used to train the GLM, resulting in a subset of the original dataframe. After rebalancing the GLM on this subset, only the characteristics associated with the damages larger than the deductible gets incorporated in the model. Thus, the predictions $\hat{\mu}_{Zd}$ may be inflated. This bias is problematic since the claim severity $Z_d = \max(0, Z - d)$ after removing the deductible cannot be higher than the original claim severity Z. By only considering insurance contracts with a claim severity that exceeds the deductible, the model may overestimate the predicted claim severity $\hat{\mu}_{Z,d}$.

To handle this overestimation, a binary variable is defined in the training data that take the value one if Z exceeds the deductible d, and zero otherwise. In addition to the gamma GLM fitted for each deductible level, a logistic regression is fitted using the same regularization technique with the same covariates but with the binary variable as target variable. The link function for the logistic regression is a logit-link. The purpose of this is that with the logistic regression model, predict the probability that the claim severity Z_d of a new insurance contract exceeds the deductible level d. This probability is then multiplied with the overestimated predicted claim severity $\hat{\mu}_{Zd}$. This generates an updated predicted claim severity such that $\hat{\mu}_{Zd} < \hat{\mu}_Z$. Algorithm 1 describes method one in general.

After iterating this procedure for vector of deductibles, fitting both a logistic regression and a gamma GLM for each deductible, one receives points of $\beta_{Z,d}$, as can be seen in figure 5a. To achieve a continuous curve over an interval $[0, \infty)$, a natural spline is fitted for the points against the deductibles. This is illustrated in figure 5b, with the deductibles expressed in percentage of price base amount. Note that figure 5 shows the $\beta_{Z,d}$ curve for one (1) insurance contract, i.e, the curve may look different for another insurance contract with other characteristics. The figure illustrates that when the deductible is zero, $\beta_{Z,d}$

is equal to one. As the deductible increases, $\beta_{Z,d}$ decreases towards zero. The curve is strictly decreasing, as expected. The curves for the upcoming two methods will have a similar appearance.



(a) *β_d for different deductibles* **(b)** *Creating a continuous $\beta_{Z,d}$ using a spline*

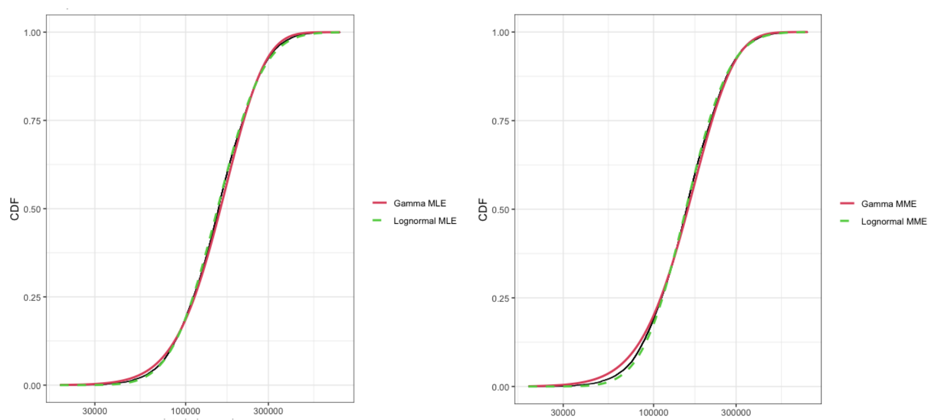
Figure 5

Algorithm 1 Method one

for all insurance contracts with reported claims **do**
  Fit a regularized gamma GLM with Z as target value
  Rebalance the GLM
  for *deductible* $= d$ **do**
    $Z_d = max(0, Z - d)$
    if $Z_d > 0$ **then**
      Set binary variable = 1
    else if $Z_d = 0$ **then**
      Set binary variable = 0
    end if
    for insurance contracts with $Z_d > 0$ **do**
      Fit a regularized gamma GLM with Z_d as target value
      Rebalance the GLM
    end for
    Fit a regularized logistic regression with the binary variable as target value
  end for
end for
for a new insurance contract i **do**
  Predict $\hat{\mu}_{Z,i}$ and $\hat{\mu}_{Zd,i}$
  Estimate $\mathbb{P}(Z_i > d)$ from the logistic regression
  $\beta_{Z,d} = \frac{\mathbb{P}(Z_i > d) \cdot \hat{\mu}_{Zd,i}}{\hat{\mu}_{Z,i}}$

3.5 Method two – Estimate $\beta_{Z,d}$ with the distribution function

Since the predicted claim severity $\hat{Z}$ is a random variable with mean $\hat{\mu}_Z$ and standard deviation σ, $\beta_{Z,d}$ could be estimated using the distribution function of the random variable. This requires an assumption of the distribution of $\hat{Z}$. It is known from beforehand that two candidates are the log-normal and gamma distribution. Since $\hat{\mu}_Z$ is predicted from a gamma GLM, it is not surprising that the data fit well to these two distributions. Figure 6 represents the estimation for the distribution of $\hat{\mu}_Z$ for all predictions of the test data, fitted using a log-normal and gamma distribution, with both maximum likelihood estimation and moments matching estimation. The interpretation of the figures is that the best fit to the data is a close call. The two distributions fitted with maximum likelihood estimation look almost identical. However, the log-normal distribution fitted using maximum likelihood gives the lowest AIC of these four estimates. Therefore, moving forward, it will be assumed that $\hat{Z} \in LN(\hat{\mu}_z, \sigma^2)$, with σ^2 defined as the variance of the residuals.



(a) *Maximum likelihood estimation* **(b)** *Moments matching estimation*

Figure 6 *Empirical and theoretical CDFs for $\hat{\mu}_Z$*

The idea of method two is to derive $\beta_{Z,d}$ by estimating the probability that the claim severity Z for an insurance policy i is larger than the deductible. So, $\beta_{Z,d} = 1 - F_{\hat{Z}}(d) = \mathbb{P}(\hat{Z} > d)$

Taking the natural logarithm of $\hat{Z}$, $\hat{Y} = \ln \hat{Z}$ is normally distributed with mean $\hat{\mu}_Y$ and variance σ^2. σ^2 is assumed to be equal for all observations, an assumption that simplifies calculations, but may be flawed. Including a covariate dependent variance $\sigma^2(\mathbf{x}_i)$ for observation i would be useful, although difficult, as described by Wüthrich and Buser (2023). Since $\hat{Y}$ is transformed, it must be re-transformed. Due to Jensen's inequality, $\mathbb{E}[\hat{Y}] = GM(\hat{Y}) \cdot \sqrt{GVar(\hat{Y})} = \exp(\mu + \frac{1}{2}\sigma^2)$ where $GM(\cdot)$ and $GVar(\cdot)$ represents the geometric mean and variance, respectively (Friedrich and Heil (2013)). This convexity bias is corrected for by multiplying $\mathbb{E}[\hat{Y}]$ with $\exp(-\frac{1}{2}\sigma^2)$.

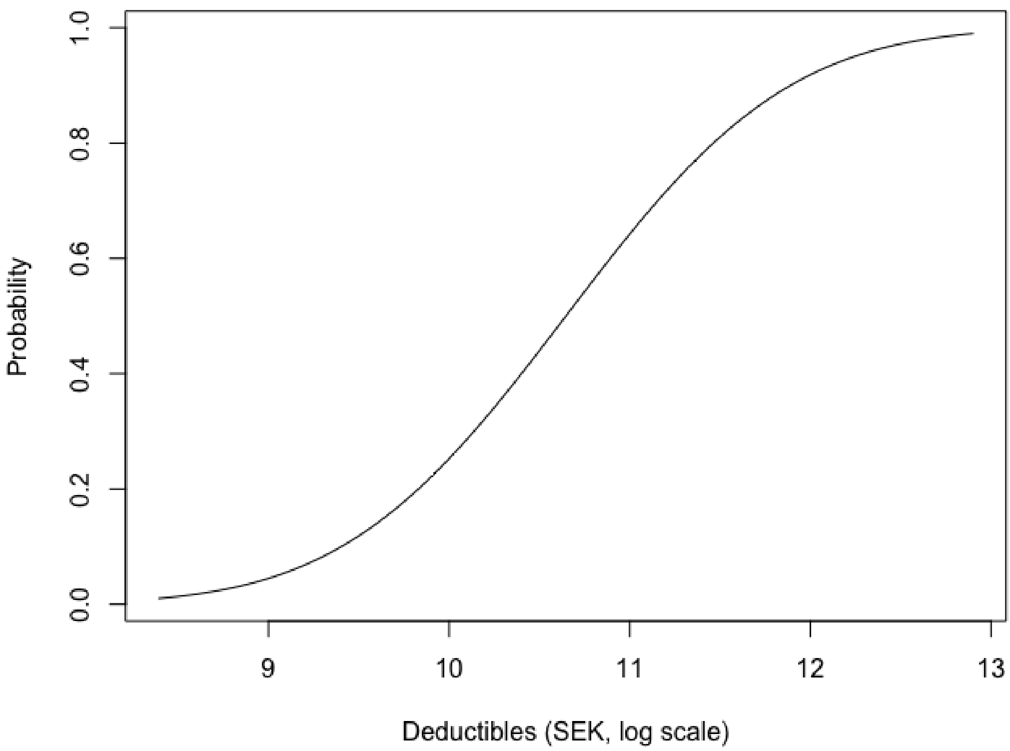


Figure 7 *Distribution plot for* $\hat{\mu}_Z = 67\ 000$

Figure 7 represents a cumulative distribution function for an insurance contract with a predicted claim severity of 67 000 SEK. The deductibles are represented in SEK, in log-scale. The interpretation of this CDF plot is that for a given level of deductible x, the probability that the predicted claim severity will take values lower than the deductible is equal to y. To connect this with the subject, the y-axis describes how much of the estimated risk premium that disappears at a deductible of x, hence, $\beta_{Z,d}$ is represented as 1 - y.

Algorithm 2 Method two

```
for all insurance contracts with reported claims do
    Fit a regularized gamma GLM with Z as target value
    Rebalance the GLM
    for the insurance contracts in the test data do
        Predict the mean claim severity μ̂_Z
        Find the best distribution of μ̂_Z, (log-normal in this case)
        Extract the residuals of μ̂_Z
        Let σ^2 = ln(residuals)
        for a new insurance contract i and a deductible d do
            Ŷ_i = ln(Ẑ_i)
            Ŷ_i ∈ N(μ̂_{Y,i} − σ^2/2, σ^2)
            β_{Z,d} = P(Ŷ_i > d)
```

3.6 Method three – A distribution free estimate of $\beta_{Z,d}$

By splitting the claim severity into multiple groups, and using a classifier to generate probabilities for one insurance contract to fall into one of these groups, one can create a discrete distribution for the claim severity, with an expected value. Subtracting the deductible d from the group means and using the same probabilities, a new discrete distribution with a smaller expected value is produced. Taking the quotient between these expected values generates an estimate for $\beta_{Z,d}$. A way to achieve this is by using a clustering method to find the groups. In this study, K-means clustering is used, but other methods are also possible.

The object of clustering is to group the distribution of the claim severity into clusters, such that the insurance contracts within each cluster are more closely related to each other, than claim severities assigned to other clusters. K-means clustering starts by randomly assigning each observation a cluster. The cluster centroid is calculated for each cluster, being the mean claim severity for the claims belonging to the specific cluster. Each observation is then reassigned to the cluster with the closest centroid in terms of euclidean distance. This process is then repeated until no observation changes clusters. The target of K-means clustering is to have the within-cluster variation as small as possible (James et al. (2021)). Here, K-means clustering is a one-dimensional problem since it is only the claim severity that is clustered. This means that the clustering algorithm calculates the cluster mean for each cluster over the line that corresponds to claim severity. Choosing the number of clusters can be done by inspecting how the total within-clusters sum of squares (SST) changes over the number of clusters. This can be viewed in figure (8). The decrease in SST after appending another cluster is large for the first four or five clusters. After that, the effect diminishes. Six clusters are used in this method.

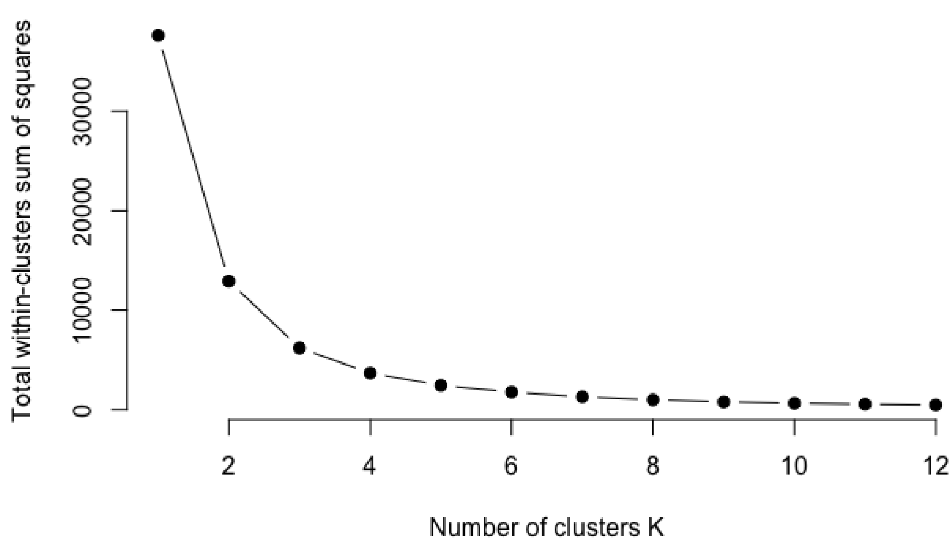


Figure 8 *Within-clusters sum of squares over number of clusters*

After generating the clusters, each observation in the data set is assigned a predicted cluster. For each cluster group C, the mean claim severity is calculated, denoted by $\bar{C}$. To estimate the probabilities for each observation, a random forest classification is conducted with the clusters as dependent variables. Any classifier would fulfill the purpose – as long as it takes the covariates for a single insurance contract as input.

A classification tree is a classifier that segments the predictor space into regions by using splitting rules. The classifier is built by starting with a first split (the top of the tree) into branches, and successively splitting the branches until some stopping criterion is met. The terminal nodes are called leaves. Figure 9 shows a classification tree with three terminal nodes that represents which cluster a claim will fall in to. All covariates were used as input, but the tree only used two covariates; insured amount (fbel) and if the building consisted of rental apartments or condos (BOFORM). A favourable part about using trees is the ability to understand and interpret the decision, but a disadvantage of trees is the risk of overfitting to the data. With only one tree, small changes in the data can lead to a completely different result, so the tree can be rather fragile.

A potential solution to this problem is random forest, a tree based method that builds many repeated trees by using bootstrap re-sampling. That is, for each tree, take a random sample from the training data, with replacement. From this random sample, a classification tree is built. The final prediction for a new insurance contract is decided by letting each tree predict the cluster, and the cluster that receives the most votes is the final prediction. In this way, the high variance that one classification tree suffers from is decreased. Random forest can improve the performance further by using a random sample of $m = \sqrt{p}$ of the p covariates when considering a split in a tree. For each split, the classification tree only considers one of the m candidates. From figure 9, it seems that the importance the covariate insured amount (fbel) has on the classification is large; of all possible combination of covariates, the tree used fbel as a predictor. If this covariate is allowed to participate in the decision of every tree, all trees will exhibit a high correlation due to the large influence that insured amount has. With random forest, insured amount will not be allowed to be considered for all trees, making the result of the classification more stable. This will decrease the risk of overfitting (James et al. (2021)).

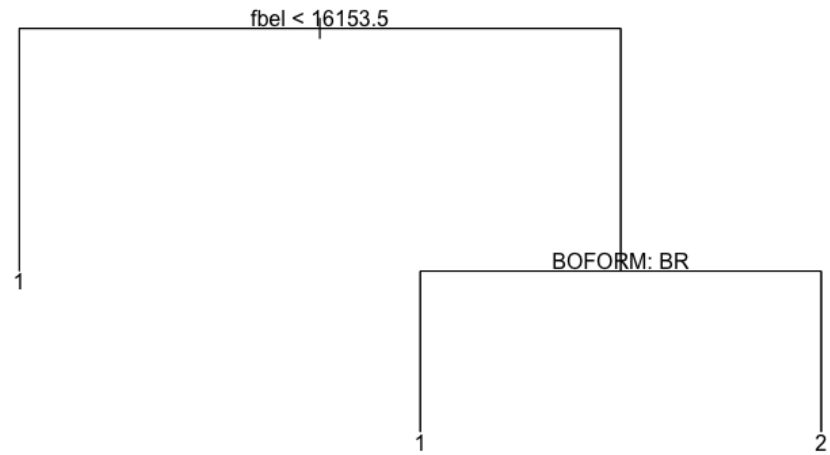


Figure 9 *Classification tree with three terminal nodes*

The probabilities for each cluster is then multiplied with its corresponding cluster mean and summed, generating the expected claim severity in the discrete case. The same is performed for the net claim severity Z_d, with a deductible d subtracted from the cluster means $\bar{C}$. By taking the quotient of these, the deductible risk factor $\beta_{Z,d}$ for insurance contract Z can be estimated, with six clusters:

$$\beta_{Z,d} = \frac{\mathbb{E}[Z_d|\mathbf{X} = \mathbf{x}]}{\mathbb{E}[Z|\mathbf{X} = \mathbf{x}]} = \frac{\sum_{i=1}^{6} \mathbb{P}(Z \in C_i) \cdot (\bar{C}_i - d)}{\sum_{i=1}^{6} \mathbb{P}(Z \in C_i) \cdot \bar{C}_i}$$

Figure 10 below shows the predicted probabilities to fall into the clusters for one insurance contract of the test data; plotted against the cluster means (triangle), and the cluster means after subtracting a deductible of 150 percent of a price base amount, corresponding to 72 450 SEK. From the plot, it can be seen that the random forest classifier predicts this specific insurance policy to fall into one of the smaller cluster means. The probabilities are strictly decreasing as the cluster means increases, and for the cluster mean above 3 000 000 SEK, the predicted probability to fall into that cluster is close to zero. For this specific insurance policy, at the deductible level mentioned above, $\beta_{Z,d} = \frac{95501}{167951} = 0.569$. This means that at a deductible of 150 percent, 56.9 percent of the claim severity is remaining after the deductible is imposed on the insurance contract. Algorithm 3 describes method three in general.

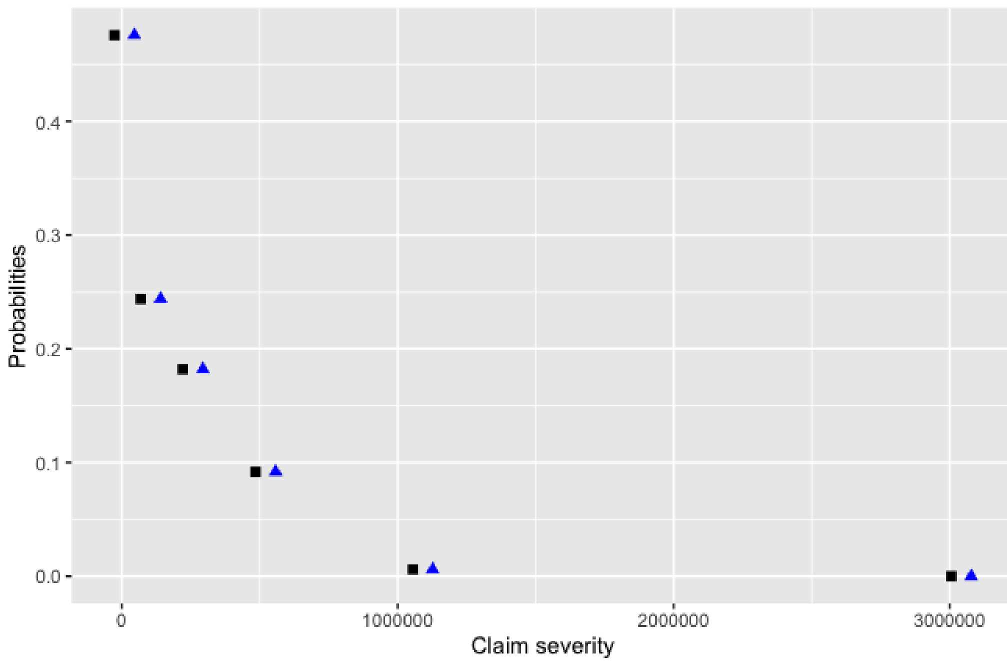


Figure 10 *Predicted probabilities and the cluster means*

Algorithm 3 Method three

for all insurance contracts with reported claims **do**
  K-means clustering on the claim severity with k = 6
  Assign each insurance contract a cluster C
  Run a classifier with the clusters as target variable
  for $deductible = d$ **do**
    for new insurance contract i **do**
      Predict $\mathbb{P}(Z_i \in C_j),\ j = 1, 2, \ldots, 6$
      $\mathbb{E}[Z_d|\mathbf{X} = \mathbf{x}] = \sum_{j=1}^{6} \mathbb{P}(Z_i \in C_j) \cdot (\bar{C}_j - d)$
      $\mathbb{E}[Z|\mathbf{X} = \mathbf{x}] = \sum_{j=1}^{6} \mathbb{P}(Z_i \in C_j) \cdot (\bar{C}_j)$
      $\beta_{Z,d} = \frac{\mathbb{E}[Z_d|\mathbf{X}=\mathbf{x}]}{\mathbb{E}[Z|\mathbf{X}=\mathbf{x}]}$

Method two and three are both trying to estimate the distribution of the claim severity $\hat{Z}$. The main difference between these two methods is that method two is based on assumptions of the distribution, compared to method three, that does not make any assumptions regarding the distribution, while instead trying to estimate the distribution itself. The advantage of using an assumption free method is that one does not have to take the true distribution of the data into account, which makes the method more flexible. On the other hand, when using method two, if the assumptions are satisfied, then the implementation is simple and the result is easy to interpret.

4 Results

The evaluation of the three methods is achieved by predicting new risk premiums $\hat{\pi}_i = w_i \beta_d \hat{\mu}_{Ni} \hat{\mu}_{Zi}$ using the three methods. Denote β_d calculated using method one $\beta_{d,1}$, method two $\beta_{d,2}$ and so on, one can see that the only difference between $\hat{\pi}_i$ is by which method β_d is calculated with, since the predicted claim severity, claim frequency and the duration of each insurance contract are equal regardless of $\beta_{d,i}, i = 1,2,3$. The three methods compete against each other, with the rule that the method giving the lowest β_d for each insurance policy "wins" that insurance contract. The purpose of the competition is to simulate a competitive market, exposing the three methods to competition. If the methods are performing well, they should be locally unbiased in the market such that when the true risk premium is low, then the methods predicts a low risk premium as well, and so on. As described in the introduction, a pricing model may be globally unbiased, but there is no guarantee that the model performs well on a local basis, which may include unattractive customers that are not priced correctly.

The deductibles are randomly generated with probabilities for the deductibles equal to how frequent the respective deductible occurs in the test data. This can be seen in table 2. The reason for generating deductibles with a probability according to how frequent occurring they are in the test data, is that this minimizes the noise that arises when comparing against the test data. Since the test data is large, the sampled deductibles should reflect the actual deductibles fairly good. If the deductibles instead were sampled with uniform probability for each deductible, then the result may be biased on a global level. For example, if the probability of getting 2000 percent of a price base amount as a deductible would be equal to that of getting a 20 percent deductible, it would not reflect the true structure of the test data, since the latter deductible occurs roughly 18 times more frequent. The consequence would be that the risk premium is underestimated. A simple example shows why this is the case. Assume that one price base amount is 50 000 SEK. Take the two deductibles 2000 percent and 20 percent of a PBB. In monetary terms, the deductibles are $50000 \cdot \frac{2000}{100} = 1\,000\,000$ SEK and $50000 \cdot \frac{20}{100} = 10\,000$ SEK. From this, it can be seen that:

$$Z_{1\,000\,000} = max(0, Z - 1\,000\,000) \leq Z_{10\,000} = max(0, Z - 10\,000)$$

Table 2 *Deductibles in the test data*

Deductible (% of PBB)	10	20	50	100	200	300	500	1000	2000	Sum
n	517	69 687	29 153	21 618	14 353	3 253	5 570	18 391	3 824	166 366
%	0.3	41.9	17.5	13.0	8.6	1.9	3.3	11.0	2.3	100

Each β_d is calculated and predicted from a natural spline, as described in method one, generating a continuous β_d. The risk premiums for the insurance contracts are then grouped by all methods and summed, generating the aggregated predicted risk premium for each method. If $j = 1,2,3$ denotes the three methods and N the number of insurance contracts in the test data, then the aggregated predicted risk premium for each method is:

$$\sum_{i=1}^{N} min(\hat{\pi}_i^j, j \in \{1,2,3\})$$

The percentage deviation from the true risk premium is calculated for each method. Since the deductibles are based on a random event, this procedure is repeated ten times to get a stable result. Table 3 shows an example from one competition. Note the large difference in β_d for the different deductibles, that describes how large influence the deductible has on the risk premium.

Table 3 *Three rows from one competition*

Deductible (% of PBB)	$\beta_{d,1}$	$\beta_{d,2}$	$\beta_{d,3}$	Winner	$\hat{\pi}$
200	0.414	0.361	0.330	Method 3	1446 SEK
1000	0.054	0.020	0.025	Method 2	5 SEK
20	0.875	0.953	0.867	Method 3	2622 SEK

Table 4 shows the average percentage deviation from the true risk premium for each method, and the average percent of insurance contracts won after ten rounds. The standard deviation is reported within the parenthesis. The interpretation of the result is that, for the contracts that method one won, it on average underestimated the true risk premium by 35.3 percent. Method two and three both overestimated the risk premium, and the standard deviation of method two was a lot higher than method three. So, method three performed better than the other two methods, with more stable result between competitions. The optimal result for a method would be to have zero deviation from the true risk premium. If that was the case, it indicates that the model is locally unbiased in the market.

Table 4 *Average result from ten competitions (standard deviation)*

Method 1		Method 2		Method 3	
Deviance	Won	Deviance	Won	Deviance	Won
-35.3 % (3.0)	45.1 % (1.1)	14.3 % (15.5)	12.2 % (1.7)	8.9 % (2.2)	42.7 % (2.4)

To see how each method performed without the competition, the simulation was repeated ten times for each method, without letting the methods compete against each other. This means that each method was tested against the entire test data, and not just against the insurance contracts where they gave the lowest price. The equivalent situation in the real world is how each method behaves when there is no competition in the market. The results are presented in table 5. From the table, it can be seen that method one is underpricing by -11.37 percent on average, compared to the true risk premium. Method two performed best, overestimating the risk premium by 1.30 percent. Method three also underpriced against the true risk premium, with -6.74 percent.

Table 5 *Results without competition*

Method 1	Method 2	Method 3
Deviance	Deviance	Deviance
-11.37 % (2.27)	1.30 % (1.01)	-6.74 % (0.98)

Figure 11 represents the mean β_d for each method. Note that the x-axis is represented in log-scale, since it is more interesting to see how each β_d curve behave at lower deductibles. From the figure, it can be noted that $\beta_{d,3}$ is lower for deductibles between zero up and until around 120 percent of a price base amount. After that, method one intersects the curve and replaces $\beta_{d,3}$ as the lowest curve, on average. The gap in y-direction between $\beta_{d,1}$ and the other curves is quite large. This has a large impact on the predicted risk premium between the three methods. A possible explanation to the large underestimation of the risk premium for method one is that when the predicted risk premiums are high, the method heavily underprices insurance contracts.

Method two wins a lot fewer contracts compared to the other two methods, and figure 11 provides an intuitive reason for that. On average, $\beta_{d,2}$ is a lot higher compared to the other two methods which makes it harder to win contracts.

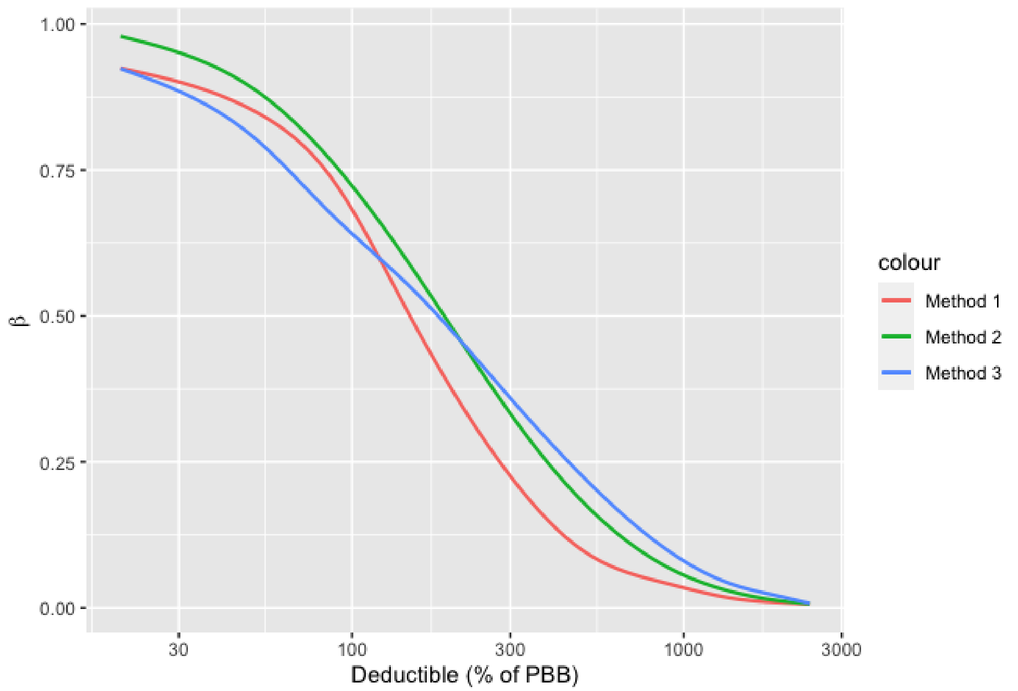


Figure 11 *Mean β_d for all methods (x-axis in log-scale)*

5 Discussion

Method one

Method one underestimated the risk premium with 35 percent, which will lead to inclusion of unattractive customers, and in the end, bankruptcy. The model underestimated the total risk premium at zero percent deductible, and as long as that is the case, it will be difficult not to underprice. Neither did the method perform well without competitors. There is a large degree of complexity involved in the model: For each level of deductible, a logistic regression and a gamma GLM is fitted, both using penalized regression with cross-validation. It is unlikely that the overall use of regularization methods is to blame for the poor performance. Compared to using a regular GLM, it should improve the predictive performance. However, an issue with the combination of using the gamma distribution and regularization is that it does not handle large claims very well. The gamma distribution may not be the best solution to model extreme values. In combination with the fact that as the deductible increases, the number of coefficients in the penalized regression decreased. For a sufficiently high deductible, the model only used the mean as a prediction for all insurance contracts. Hence, the method may not be performing well locally. Lets say that the method overestimates β_d for deductibles below 120 percent of PBB, and underestimates those above. From the data, it could be noted that the distribution of damages with a higher deductible tended to be higher, all else equal. This would mean that for a larger predicted claim severity, method one wins more insurances wrongfully.

A suggestion of further studies is to improve the predictive performance of the claim severity for large claims, possibly by evaluating other distributions. Maybe this could improve the performance of method one. The method should also be easy to implement. Besides from the long computation time of fitting a large amount of regularized models, it is quite easy to compute a way to automate the method, that can be generalized to other insurance types, and this is favourable.

Method two

Method two was the most theoretical method, using the distribution function of the insurance contracts to generate the probability that the claim severity exceeds the deductible. Due to the underlying assumptions of method two, it was also the method with the biggest limitations. It is unrealistic that all insurance contracts have the same variance. Different types of insurance objects may not have equal variance, for example if the property consists of rental or condominium apartments. This could possibly also explain the high standard deviation in the results of the competition. Another issue of method two is that the distribution of the claim severity is based on predictions from a penalized GLM. Just as with method one, it is likely that the predictions suffered from local bias. Predicting claims with a locally high precision is hard due to the low signal-to-noise ratio that exists in insurance contracts. The computational complexity of method two was only a very small fraction of that of method one, and the method itself was rather easy to implement computationally. A favourable property of method two is that the computational cost to predict β_d for a new deductible is very low, compared to method one, were one has to fit two new GLMs (both the logistic and the gamma regression). Method two has a large potential, and a suggestion to further studies is to develop method two by including a covariate unique variance $\sigma^2(\mathbf{x})$. The interesting thing about method two is the intuitive logic behind it. Being able to describe the result such that non-mathematicians understand is important, and method two provides a simple explanation for $\beta_{Z,d}$ by using the probability that the claim severity exceeds a deductible.

Method three

The best performing method when it came to the competition was method three, with a mean overestimation of the risk premium of 8.9 percent. There are two probable reasons for this. First, method three did not assume a distribution of the claim severity, which makes the model all-round and unrestricted. Second, the foundations of the method is that it does not depend on predicted values of the claim severity. Rather, it uses the training data to find groups of insurance contracts that clusters together. Actual claim severities in the training data is probably a better local estimate to the test data in comparison to predicted claim severities. So the issue of suffering from a locally biased regularized model is avoided using method three, while still being able to use the characteristics of the individual insurance contracts to predict the probability of falling into all clusters. One potential issue of using k-means clustering to group the claim severity is that since the distribution is so right-skewed, most of the clusters will be close to each other, while one cluster will handle the very large claims. This can be seen in figure 8, with the clusters handling the large claims located at around 3 000 000 SEK. The claims that fall into this cluster are very few, which may give rise to overfitting.

The method is very easy to implement. One needs to group the claim severity into groups, and fit a classifier to generate the probability for every insurance contract to be placed in each group. The method could be implemented using a large variation of binning methods and classification techniques, from simple methods such as multinomial regression to more advanced techniques like neural networks. It would be interesting to see what combination of grouping technique and classifier that gives the best result. This could be something to look into further. Another interesting topic is to see how the result would be if the groups were created so that there are equal numbers of claims in each group. Using this method, the potential problem with overfitting will probably decrease since the number of observations in each group will be larger in comparison to the current solution.

6 Conclusion

As the insurance business currently works, when a potential customer contacts the insurance company with the interest of buying insurance, he is offered insurance with a couple of deductibles to choose from, and depending on the choice of deductible, he will get a premium to pay. What if instead, the customer gets to choose how much premium he is interested to pay, and from that, will get a deductible? This could be a new way to sell insurance, a new way to attract customers. Instead of working as a pricing actuary, the new title becomes a "deductible actuary".

One problem that all methods have in common is perhaps also the hardest to work around. Insurance contracts is about human behaviour, which is complicated to quantify. Especially when it comes to deductibles. It can be quantified how the previous data behave when the deductibles change. What cannot be predicted in the same sense is the change in human behaviour that a change in the deductible will have, and as seen from the study in Taiwan from Wang et al. (2008), deductibles can have a large impact on human behaviour. How a change in the deductible affects the risk premium can at best be guessed from β_d, but in the end, one has to wait to see the actual outcome. The best method to understand how the deductible influences the risk premium is to randomly split a large sample of insurance contracts in groups with different deductibles in each group. Setting up a randomized trial this way isolates the effect that deductibles has on the insurance buyers. But this would only be valid for that type of insurance, it is likely that other insurance products have different effects on the risk premium.

But the purpose of this study was not about the effect of a deductible. The purpose of this study was to evaluate and implement a method that quantifies how much of the risk premium that remains after a deductible have been imposed on the contract. The three methods that were evaluated did not perform perfect. There are ways to improve all three methods. Given the results, the recommendation is to use method three. It produced the best result in the market, without being restricted by assumptions on the distribution. As a whole, I believe that the purpose of this study is fulfilled, and the very large research gap in this research area can hopefully be viewed as a bit more narrow.

References

A. Cohen and P. Siegelman. Testing for Adverse Selection in Insurance Markets. *The Journal of Risk and Insurance*, 77(1):39–84, 2010. doi: 10.1111/j.1539-6975.2009.01337.x.

B. Friedrich and P. Heil. Onset-Duration Matching of Acoustic Stimuli Revisited: Conventional Arithmetic vs. Proposed Geometric Measures of Accuracy and Precision. *Frontiers in Psychology*, 7:1–18, 2013. doi: 10.3389/fpsyg.2016.02013.

T. Hastie, R. Tibshirani, and J. Friedman. *Elements of Statistical Learning*. Springer Series in Statistics. Springer New York, NY, 2009. ISBN 978-0-387-84858-7.

T. Hastie, R. Tibshirani, and M. Wainwright. *Statistical Learning with Sparsity: The Lasso and Generalizations*. Chapman Hall/CRC, 2015. ISBN 9780429171581.

D. Hemenway. Propitious Selection. *The Quarterly Journal of Economics*, 105(4):1063–1069. doi: https://doi.org/10.2307/2937886.

G. James, D. Witten, T. Hastie, and R. Tibshirani. *An Introduction to Statistical Learning*. Springer Texts in Statistics. Springer New York, NY, 2021. ISBN 978-1-0716-1418-1.

E. Ohlsson and B. Johansson. *Non-Life Insurance Pricing with Generalized Linear Models*. European Actuarial Academy. Springer, 2010. ISBN 978-3-642-10790-0.

U. Olsson. *Generalized Linear Models*. Studentlitteratur, 2002. ISBN 91-44-04155-1.

S. Shavell. On Moral Hazard and Insurance. *The Quarterly Journal of Economics*, 93(4): 541–562, 1979. doi: https://doi-org.proxy.ub.umu.se/10.2307/1884469.

J. L. Wang, C.-F. Chung, and L. Y. Tzeng. An Empirical Analysis of the Effects of Increasing Deductibles on Moral Hazard. *The Journal of Risk and Insurance*, 75(3):551–566, 2008. doi: http://www.jstor.org/stable/25145295.

M. V. Wüthrich and C. Buser. Data Analytics for Non-Life Insurance Princing. *Swiss Finance Institute Research Paper*, 16(68), 2023. doi: https://ssrn.com/abstract=2870308.

M. V. Wüthrich and M. Merz. *Statistical Foundations of Actuarial Learning and its Applications*. Springer Actuarial. Springer, 2023. ISBN 978-3-031-12408-2.

H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 67(2):301–320, 2005. doi: https://hastie.su.domains/Papers/B67.2\%20(2005)\%20301-320%20Zou%20&%20Hastie.pdf.