NON-LIFE INSURANCE PRICING USING THE GENERALIZED ADDITIVE MODEL, SMOOTHING SPLINES AND L-CURVES

Kivan Kaivanipour

A thesis submitted for the degree of
Master of Science in Engineering Physics

Department of Mathematics
ROYAL INSTITUTE OF TECHNOLOGY
Stockholm, Sweden
May 2015

Abstract

In non-life insurance, almost every tariff analysis involves continuous rating variables, such as the age of the policyholder or the weight of the insured vehicle. In the generalized linear model, continuous rating variables are categorized into intervals and all values within an interval are treated as identical. By using the generalized additive model, the categorization part of the generalized linear model can be avoided. This thesis will treat different methods for finding the optimal smoothing parameter within the generalized additive model. While the method of cross validation is commonly used for this purpose, a more uncommon method, the L-curve method, is investigated for its performance in comparison to the method of cross validation. Numerical computations on test data show that the L-curve method is significantly faster than the method of cross validation, but suffers from heavy under-smoothing and is thus not a suitable method for estimating the optimal smoothing parameter.

Acknowledgements

First of all I would like to thank my supervisor Jonathan Fransson, for your helpful discussions on this subject, and for giving me the possibility of working with you. You taught me insurance pricing theory when we were co-workers at If P&C Insurance. I really appreciate how you always were available for fun discussions about insurance mathematics and eSports. Thanks also to my other colleagues at If P&C Insurance and Vardia Insurance Group for fruitful collaborations. I want to thank Camilla Landén, Rasmus Janse and Andreas Runnemo for your helpful comments on the drafts of this thesis. Finally, I would like to use this oppurtunity to thank my friends and family, for your support and for all the good times we have shared.

Stockholm, May 2015
Kivan Kaivanipour

Contents

Chapter 1

Introduction

In non-life insurance, almost every tariff analysis involves continuous rating variables, such as the age of the policyholder or the weight of the insured vehicle. In the generalized linear model (GLM), continuous rating variables are categorized into intervals and all values within an interval are treated as identical.

Tariff cell	Age (years)	Weight (kg)	Premium
1	$18-30$	$0-1\,500$	P_1
2	$18-30$	$1\,500-2\,500$	P_2
3	$18-30$	$2\,500-3\,500$	P_3
4	$30-99$	$0-1\,500$	P_4
5	$30-99$	$1\,500-2\,500$	P_5
6	$30-99$	$2\,500-3\,500$	P_6

Table 1.1: Illustration of a simple tariff with two categorized rating variables.

This method is simple and often works well enough. However, a disadvantage of categorizing a rating variable is that two policies with different but close values for the rating variable may get significantly different premiums if the values happen to belong to different intervals. Also, finding a good subdivision into intervals can be time consuming and tedious. The intervals must be large enough to achieve good precision of the price relativities, but at the same time they have to be small if the effect of the rating variable varies much. Sometimes it is difficult to fulfill both these requirements.

With this in mind, an alternative modelling approach can be used. By using the generalized additive model (GAM), the categorization part of the GLM can be avoided. This thesis will treat different methods for finding the optimal smoothing parameter within the GAM. While the method of cross validation is commonly used for this purpose, a more uncommon method, the L-curve method, is investigated for its performance in comparison to the method of cross validation.

Chapter 2

Insurance

An insurance policy is a contract between an insurer, e.g. an insurance company, and a policyholder, e.g. a consumer or a company. An insurance policy can be considered as a promise made by the insurer to cover certain unexpected losses that the policyholder may encounter. The policyholder pays the insurer a fee for this contract, called the premium. An event reported by the policyholder where he or she demands economic compensation is called a claim.

Insurance policies are commonly categorized into life and non-life insurance. A life insurance policy covers future financial losses and the insurer pays a sum of money to a designated recipient, called the beneficiary. A non-life insurance policy covers damage incurred to the policyholder's possesion or property, and the policyholder receives compensation so that the property can be recovered to the previous physical condition. However, there are counterexamples, such as business interruption policies (non-life insurance), where the policyholder (usually a company) is being covered for future financial losses.

The strategic business idea of an insurer has its origin in ancient history where the community supported individual losses [13]. Back then, the premiums were distributed uniformly, but as time has passed, the realization of a modern insurance company's success is through risk differentiation. Statistics and probability theory can be used to estimate the risk an insurer is exposed to and the main theorem which is the cornerstone of insurance mathematics is the Law of Large Numbers [3]. The theorem states that the expected value μ of a sequence of independent and identically distributed random variables Y_i can be approximated by the sample average of $\{Y_1, \ldots, Y_n\}$. Applied to insurance mathematics, this means that the total loss for a large set of customers should be close to its expected value and the premium for a policy should be based on the expected average loss that is transferred from the policyholder to the insurer [12, §1]. In this thesis we will omit non-risk related costs, such as administration costs, and only discuss the part of the premium that is directly connected to the losses.

2.1 Tariff analysis

A tariff is a formula, by which the premium for any policy can be computed. The underlying work for designing a tariff is in the insurance business known as a tariff analysis. The data material for a tariff analysis is historical data with information about policies and claims. We limit the definition of a tariff to explain the expected average loss, which will hereafter be referred to as the pure premium. The pure premium can be expressed as the product of the claim frequency (how often a claim occurs) and the claim severity (how much a claim will cost on average) [12, §1.1], such that

$$\text{Pure premium} = \text{Claim frequency} \times \text{Claim severity}. \tag{2.1}$$

An analysis can be performed on the pure premium directly, but there are advantages in splitting the model to the product above, generating two sets of analyzes. First, the claim frequency and the claim severity may have different dependencies on different variables. Second, the amount of data is sometimes low and might give inconclusive statistical results, whereas the model split realizes the weak link. The weak link in the split model is almost always the claim severity analysis due to the fact that the only relevant data for this analysis are policies where a claim has occured and that is merely a subset of all the available data.

2.2 Rating variables

The claim frequency and the claim severity varies between policies and can be estimated based on a set of parameters, called the rating variables. A rating variable usually describes an attribute of either the policyholder or the insured object. The age of the policyholder, the weight of the insured vehicle and the value of the insured property are a few examples of rating variables.

A rating variable can be either continuous or categorical. In a tariff analysis, it is common to categorize continuous rating variables into intervals and to treat them as categorical rating variables [12, §1.1]. This is done to improve the significance of the statistical results. Policies within the same interval for each rating variable are said to belong to the same tariff cell and share the same premium.

Tariff cell	Age (years)	Weight (kg)	Premium
1	$18-30$	$0-1\,500$	P_1
2	$18-30$	$1\,500-2\,500$	P_2
3	$18-30$	$2\,500-3\,500$	P_3
4	$30-99$	$0-1\,500$	P_4
5	$30-99$	$1\,500-2\,500$	P_5
6	$30-99$	$2\,500-3\,500$	P_6

Table 2.1: Illustration of a simple tariff with two categorized rating variables.

2.3 The multiplicative model

There are basically two ways the elements of a tariff could operate when computing the premium. The model can be additive or multiplicative, whereas the latter model is nowadays considered as standard for insurance pricing [11]. In order to describe the multiplicative model we need some preliminaries. Let M be the number of rating variables, let m_k be the number of intervals for rating variable k and let i_k denote the interval of rating variable k.

Tariff cell	Age (years)	Weight (kg)	i_1	i_2
1	$18-30$	$0-1\,500$	1	1
2	$18-30$	$1\,500-2\,500$	1	2
3	$18-30$	$2\,500-3\,500$	1	3
4	$30-99$	$0-1\,500$	2	1
5	$30-99$	$1\,500-2\,500$	2	2
6	$30-99$	$2\,500-3\,500$	2	3

Table 2.2: An illustrative example with $M=2$, $m_1=2$ and $m_2=3$.

Let μ be the mean of a key ratio Y, where the key ratio can be either the claim frequency, the claim severity or the pure premium. The mean μ for a policy with rating variables pertaining to the intervals $i_1, i_2, \ldots, i_M$ is then given by

$$\mu_{i_1,i_2,\ldots,i_M} = \gamma_0 \prod_{k=1}^{M} \gamma_{k,i_k}, \tag{2.2}$$

where γ_0 is the base value and $\{\gamma_{k,i_k},\ i_k = 1,2,\ldots,m_k\}$ are the price relativities for rating variable k [12, §1.3].

The model is over-parameterized since if we multiply all γ_{1,i_1} with any number α and divide all γ_{2,i_2} with the same α we get the same μ's as before. To make the price relativities unique we specify a base cell $\{i_1 = b_1, i_2 = b_2, \ldots, i_M = b_M\}$ and set $\{\gamma_{k,b_k} = 1,\ k = 1,2,\ldots,M\}$. The base value γ_0 can now be interpreted as the mean in the base cell and the price relativities measure the relative difference in relation to the base cell. In Chapter 3 we will get back to the multiplicative model and show how to determine the base value and the price relativities.

Tariff cell	Age (years)	Weight (kg)	i_1	i_2	Base cell	$\mu_{i_1.i_2}$
1	$18-30$	$0-1\,500$	1	1	x	γ_0
2	$18-30$	$1\,500-2\,500$	1	2		$\gamma_0\gamma_{2,2}$
3	$18-30$	$2\,500-3\,500$	1	3		$\gamma_0\gamma_{2,3}$
4	$30-99$	$0-1\,500$	2	1		$\gamma_0\gamma_{1,2}$
5	$30-99$	$1\,500-2\,500$	2	2		$\gamma_0\gamma_{1,2}\gamma_{2,2}$
6	$30-99$	$2\,500-3\,500$	2	3		$\gamma_0\gamma_{1,2}\gamma_{2,3}$

Table 2.3: A multiplicative model with $b_1 = b_2 = 1$ and $\gamma_{1,1} = \gamma_{2,1} = 1$.

Chapter 3

Non-life insurance pricing using the GLM

As discussed in the previous chapter, the main goal of a tariff analysis is to determine how a key ratio Y varies with a number of rating variables. As mentioned, the key ratio can be either the claim frequency, the claim severity or the pure premium.

The linear model (LM) is not suitable for insurance pricing due to;

i) Poor distribution approximation

   In the LM the response variable Y follows a normal distribution, whereas in insurance, the number of claims follows a discrete probability distribution on the non-negative integers. Furthermore, the distribution of the claim cost is non-negative and often skewed to the right [12, §2].

ii) Non-reasonable mean modeling

   In the LM the mean μ is a linear function of the covariates, whereas multiplicative models are usually more reasonable for insurance pricing.

The GLM [10] generalizes the LM in two different ways and is more suitable for insurance pricing due to;

i) Good distribution approximation

   In the GLM the response variable Y can follow any distribution that belongs to the class of the exponential dispersion models (EDMs). The normal, Poisson and gamma distributions are all members of this class.

ii) Reasonable mean modeling

   In the GLM some monotone transformation of the mean $g(\mu)$ is a linear function of the covariates, with the linear and multiplicative models as special cases [12, §2].

These generalizations are discussed in detail in Sections 3.1 and 3.2, respectively.

3.1 Exponential dispersion models

In the GLM the response variable Y can follow any distribution that belongs to the class of the EDMs [8]. The probability distribution of an EDM is given by the frequency function

$$f_{Y_i}(y_i, \theta_i, \phi) = \exp\left\{\frac{y_i\theta_i - b(\theta_i)}{\phi} + c(y_i, \phi)\right\}, \tag{3.1}$$

where y_i is a possible outcome of the response variable Y_i, θ_i is a parameter that may depend on i, ϕ is called the dispersion parameter and b is called the cumulant function. The function c is not of interest in GLM theory. We will now show that the normal, Poisson and gamma distributions all are members of the EDM class.

3.1.1 Normal distribution

Let us first show that the normal distribution used in the LM is a member of the EDM class. Let $Y_i \sim N\big(\mu_i, \sigma^2\big)$. The frequency function is then given by

$$\begin{aligned} f_{Y_i}(y_i) &= \sqrt{\frac{1}{2\pi\sigma^2}}e^{-(y_i-\mu_i)^2/2\sigma^2} \\ &= \exp\left\{\log\sqrt{\frac{1}{2\pi\sigma^2}} - \frac{1}{2\sigma^2}\big(y_i^2 - 2y_i\mu_i + \mu_i^2\big)\right\} \\ &= \exp\left\{\frac{y_i\mu_i - \mu_i^2/2}{\sigma^2} - \frac{1}{2}\left(\frac{y_i^2}{\sigma^2} + \log\big(2\pi\sigma^2\big)\right)\right\}. \end{aligned} \tag{3.2}$$

The normal distribution is thus an EDM with $\theta_i = \mu_i$, $\phi = \sigma^2$, $b(\theta_i) = \theta_i^2/2$ and

$$c(y_i, \phi) = -\frac{1}{2}\left(\frac{y_i^2}{\phi} + \log(2\pi\phi)\right). \tag{3.3}$$

3.1.2 Poisson distribution

Let us also show that the Poisson distribution is a member of the EDM class. Let $Y_i \sim Po(\mu_i)$. The frequency function is then given by

$$\begin{aligned} f_{Y_i}(y_i) &= \frac{\mu_i^{y_i}}{y_i!}e^{-\mu_i} \\ &= \exp\{y_i \log\mu_i - \mu_i - \log(y_i!)\}. \end{aligned} \tag{3.4}$$

The Poisson distribution is thus an EDM with $\theta_i = \log\mu_i$, $\phi = 1$, $b(\theta_i) = e^{\theta_i}$ and

$$c(y_i, \phi) = -\log(y_i!). \tag{3.5}$$

3.1.3 Gamma distribution

Let us finally show that the gamma distribution also is a member of the EDM class. Let $Y_i \sim G(\alpha, \beta_i)$. The frequency function is then given by

$$
\begin{aligned}
f_{Y_i}(y_i) &= \frac{\beta_i^\alpha}{\Gamma(\alpha)} y_i^{\alpha-1} e^{-\beta_i y_i} \\
&= \exp\bigl\{\log(\beta_i^\alpha) + \log\bigl(y_i^{\alpha-1}\bigr) - \log\Gamma(\alpha) - \beta_i y_i\bigr\}.
\end{aligned}
\tag{3.6}
$$

Before transforming this expression to EDM form we need to re-parameterize it through $\mu_i = \alpha/\beta_i$ and $\phi = 1/\alpha$. The frequency function is then given by

$$
\begin{aligned}
f_{Y_i}(y_i) &= \exp\left\{\log\left(\left(\frac{1}{\mu_i\phi}\right)^{1/\phi}\right) + \log\left(y_i^{1/\phi-1}\right) - \log\Gamma(1/\phi) - \frac{y_i}{\mu_i\phi}\right\} \\
&= \exp\left\{\frac{1}{\phi}\log\frac{1}{\mu_i\phi} + (1/\phi - 1)\log y_i - \log\Gamma(1/\phi) - \frac{y_i}{\mu_i\phi}\right\} \\
&= \exp\left\{\frac{-y_i/\mu_i - \log\mu_i}{\phi} + \frac{\log(y_i/\phi)}{\phi} - \log y_i - \log\Gamma(1/\phi)\right\}.
\end{aligned}
\tag{3.7}
$$

The gamma distribution is thus an EDM with $\theta_i = -1/\mu_i = -\beta_i/\alpha$, $\phi = 1/\alpha$, $b(\theta_i) = -\log(-\theta_i)$ and

$$
c(y_i, \phi) = \frac{\log(y_i/\phi)}{\phi} - \log y_i - \log\Gamma(1/\phi). \tag{3.8}
$$

3.2 The link function

In the previous section we described how an EDM generalizes the normal distribution used in the LM, and we now turn to the other generalization, which concerns the linear structure of the mean. Let us start by looking at a simple example, in which we only have two rating variables, one with two intervals and one with three intervals. Using the multiplicative model, the mean is given by

$$\mu_{i_1,i_2} = \gamma_0 \gamma_{1,i_1} \gamma_{2,i_2}. \tag{3.9}$$

By taking logarithms we can transform the above model to linear form

$$\log \mu_{i_1,i_2} = \log \gamma_0 + \log \gamma_{1,i_1} + \log \gamma_{2,i_2}. \tag{3.10}$$

We recall from Section 2.3 that the model is over-parameterized so we choose a base cell, $b_1 = b_2 = 1$ and set $\gamma_{1,1} = \gamma_{2,1} = 1$. We realize that $\mu_{1,1} = \gamma_0$ and that the price relativities measure the relative difference in relation to the base cell.

Tariff cell	i_1	i_2	Base cell	$\log \mu_{i_1,i_2}$			
1	1	1	x	$\log \gamma_0$			
2	1	2		$\log \gamma_0$		$+\log \gamma_{2,2}$	
3	1	3		$\log \gamma_0$			$+\log \gamma_{2,3}$
4	2	1		$\log \gamma_0$	$+\log \gamma_{1,2}$		
5	2	2		$\log \gamma_0$	$+\log \gamma_{1,2}$	$+\log \gamma_{2,2}$	
6	2	3		$\log \gamma_0$	$+\log \gamma_{1,2}$		$+\log \gamma_{2,3}$

Table 3.1: A simple example with $b_1 = b_2 = 1$ and $\gamma_{1,1} = \gamma_{2,1} = 1$.

To simplify the notation, we change the index of the mean so it represents the tariff cell instead of the interval for each rating variable. We also introduce the regression parameters $\beta_0 = \log \gamma_0$, $\beta_1 = \log \gamma_{1,2}$, $\beta_2 = \log \gamma_{2,2}$ and $\beta_3 = \log \gamma_{2,3}$.

Tariff cell (i)	i_1	i_2	Base cell	$\log \mu_i$			
1	1	1	x	β_0			
2	1	2		β_0		$+\beta_2$	
3	1	3		β_0			$+\beta_3$
4	2	1		β_0	$+\beta_1$		
5	2	2		β_0	$+\beta_1$	$+\beta_2$	
6	2	3		β_0	$+\beta_1$		$+\beta_3$

Table 3.2: Parameterization of a two-way multiplicative model.

Next, we introduce the dummy variables x'_{ij} through the relation

$$x'_{ij} = \begin{cases} 1, & \text{if } \beta_j \text{ is included in } \log \mu_i, \\ 0, & \text{otherwise.} \end{cases} \tag{3.11}$$

Tariff cell (i)	i_1	i_2	Base cell	x'_{i0}	x'_{i1}	x'_{i2}	x'_{i3}
1	1	1	x	1	0	0	0
2	1	2		1	0	1	0
3	1	3		1	0	0	1
4	2	1		1	1	0	0
5	2	2		1	1	1	0
6	2	3		1	1	0	1

Table 3.3: Dummy variables in a two-way multiplicative model.

Using the dummy variables x'_{ij} and the regression parameters β_j that we have introduced, we finally obtain the following useful equation for the logarithm of the mean

$$\log \mu_i = \sum_{j=0}^{3} x'_{ij}\beta_j, \quad i = 1, 2, \ldots, 6. \tag{3.12}$$

We now leave our simple example behind and look at the general problem of how a key ratio Y is affected by M rating variables. We let m_k be the number of intervals for rating variable k and introduce

$$\eta_i = \sum_{j=0}^{r} x'_{ij}\beta_j, \quad i = 1, 2, \ldots, n, \tag{3.13}$$

where x'_{ij} and β_j are defined similarly as in the simple example, r is the total number of regression parameters (the base value parameter β_0 not included) and n is the total number of tariff cells. The values r and n are given by

$$r = \sum_{j=1}^{M} m_j - M, \qquad\qquad n = \prod_{j=1}^{M} m_j. \tag{3.14}$$

In the LM, $\mu_i = \eta_i$, but in the GLM it can be any arbitrary function $g(\mu_i) = \eta_i$ as long as it is monotone and differentiable [12, §2.2]. The function g is called the link function [10], since it links the mean to the linear structure through

$$g(\mu_i) = \eta_i = \sum_{j=0}^{r} x'_{ij}\beta_j. \tag{3.15}$$

Since we are solely working with multiplicative models we will only be using a logarithmic link function from now on

$$g(\mu_i) = \log \mu_i. \tag{3.16}$$

3.3 Price relativities

So far we have only discussed some properties of the GLM, but now it is time for the most important step, the estimation of the base value and the price relativities. We will treat two different cases. For the claim frequency we will assume a Poisson distribution and for the claim severity we will assume a gamma distribution. These assumptions are reasonable and considered as standard for insurance pricing [1, 2].

3.3.1 The claim frequency

For the number of claims of an individual policy during any period of time we assume a Poisson distribution. Furthermore, we assume that policies are independent and we use the fact that the sum of independent Poisson distributed random variables is also Poisson distributed [4]. We therefore get a Poisson distribution also at the aggregate level of the total number of claims in a tariff cell. Let X_i be the number of claims in a tariff cell with duration w_i and let μ_i denote the expected value of X_i when $w_i = 1$. Then $\mathbb{E}(X_i) = w_i\mu_i$ and X_i follows a Poisson distribution with frequency function

$$f_{X_i}(x_i, \mu_i) = \frac{(w_i\mu_i)^{x_i}}{x_i!} e^{-w_i\mu_i}, \quad x_i = 0, 1, 2 \ldots . \tag{3.17}$$

As discussed earlier, we want to model the claim frequency $Y_i = X_i/w_i$, and it follows that the frequency function of the claim frequency is given by

$$f_{Y_i}(y_i, \mu_i) = \frac{(w_i\mu_i)^{w_i y_i}}{(w_i y_i)!} e^{-w_i\mu_i}, \quad w_i y_i = 0, 1, 2 \ldots . \tag{3.18}$$

We are now ready to use maximum likelihood estimation (MLE) [10] to estimate the regression parameters β_j. Since we have assumed that all policies are independent it follows that the log-likelihood of the whole sample is the sum of the log-likelihoods in all the tariff cells [12, §2.3.1]

$$\ell(y_i, \mu_i) = \sum_{i=1}^{n} w_i(y_i \log \mu_i - \mu_i + y_i \log w_i) - \log(w_i y_i!). \tag{3.19}$$

From Equation 3.15 and 3.16 we deduce that

$$\mu_i = \exp\left\{\sum_{j=0}^{r} x'_{ij}\beta_j\right\}, \tag{3.20}$$

which inserted to Equation 3.19 gives

$$\ell = \sum_{i=1}^{n} w_i \left(y_i \sum_{j=0}^{r} x'_{ij}\beta_j - \exp\left\{\sum_{j=0}^{r} x'_{ij}\beta_j\right\} + y_i \log w_i \right) - \log(w_i y_i!). \tag{3.21}$$

We then differentiate the log-likelihood with respect to every β_j to find the regression parameters that maximize the expression. The partial derivatives are given by

$$\begin{aligned}\frac{\partial \ell}{\partial \beta_j} &= \sum_{i=1}^{n} w_i \left(y_i - \exp\left\{ \sum_{j=0}^{r} x'_{ij}\beta_j \right\} \right) x'_{ij} \\ &= \sum_{i=1}^{n} w_i (y_i - \mu_i) x'_{ij}.\end{aligned} \tag{3.22}$$

By setting all these $r+1$ partial derivatives equal to zero we find the stationary point, and the maximum likelihood (ML) equations can be expressed as

$$\sum_{i=1}^{n} w_i (y_i - \mu_i) x'_{ij} = 0, \quad j = 0, 1, \ldots, r. \tag{3.23}$$

The ML equations must almost always be solved numerically. Newton-Raphson's method and Fisher's scoring method are two well-known methods for solving the ML equations [12, §3.2.3]. When the ML equations have been solved we obtain the base value and the price relativities by using the fact that

$$\gamma_j = e^{\beta_j}, \quad j = 0, 1, \ldots, r. \tag{3.24}$$

3.3.2 The claim severity

For the cost of an individual claim we assume a gamma distribution and we use the fact that the sum of independent gamma distributed random variables is also gamma distributed [4]. We therefore get a gamma distribution also at the aggregate level of the total claim cost in a tariff cell. Let X_i be the claim cost in a tariff cell with duration w_i and let μ_i denote the expected value of X_i when $w_i = 1$. Then $\mathbb{E}(X_i) = w_i \mu_i$ and X_i follows a gamma distribution with frequency function

$$f_{X_i}(x_i) = \frac{\beta_i^{w_i \alpha}}{\Gamma(w_i \alpha)} x_i^{w_i \alpha - 1} e^{-\beta_i x_i}, \quad x_i > 0. \tag{3.25}$$

As discussed earlier, we want to model the claim severity $Y_i = X_i / w_i$, and it follows that the frequency function of the claim severity is given by

$$f_{Y_i}(y_i) = \frac{(w_i \beta_i)^{w_i \alpha}}{\Gamma(w_i \alpha)} y_i^{w_i \alpha - 1} e^{-w_i \beta_i y_i}, \quad y_i > 0. \tag{3.26}$$

We re-parameterize this expression through $\mu_i = \alpha / \beta_i$ and $\phi = 1/\alpha$, and obtain

$$f_{Y_i}(y_i, \mu_i) = \frac{1}{\Gamma(w_i/\phi)} \left(\frac{w_i}{\mu_i \phi} \right)^{w_i/\phi} y_i^{w_i/\phi - 1} e^{-w_i y_i / (\mu_i \phi)}, \quad y_i > 0. \tag{3.27}$$

We are now ready to use MLE to estimate the regression parameters β_j. Since we have assumed that all policies are independent it follows that the log-likelihood of the whole sample is the sum of the log-likelihoods in all the tariff cells

$$\ell = \frac{1}{\phi}\sum_{i=1}^{n} w_i\left(\log\frac{1}{\mu_i} - \frac{y_i}{\mu_i} + \log\frac{w_i y_i}{\phi} - \frac{\phi}{w_i}\left(\log y_i + \log\Gamma\left(\frac{w_i}{\phi}\right)\right)\right). \tag{3.28}$$

We replace μ_i with the expression from Equation 3.20, and the log-likelihood can be rewritten as

$$\begin{aligned}\ell = \frac{1}{\phi}\sum_{i=1}^{n} w_i\Bigg(-\sum_{j=0}^{r} x'_{ij}\beta_j - y_i/\exp\left\{\sum_{j=0}^{r} x'_{ij}\beta_j\right\} + \log\frac{w_i y_i}{\phi}\\ -\frac{\phi}{w_i}\left(\log y_i + \log\Gamma\left(\frac{w_i}{\phi}\right)\right)\Bigg).\end{aligned} \tag{3.29}$$

We then differentiate the log-likelihood with respect to every β_j to find the regression parameters that maximize the expression. The partial derivatives are given by

$$\begin{aligned}\frac{\partial\ell}{\partial\beta_j} &= \frac{1}{\phi}\sum_{i=1}^{n} w_i\left(y_i/\exp\left\{\sum_{j=0}^{r} x'_{ij}\beta_j\right\} - 1\right)x'_{ij}\\ &= \frac{1}{\phi}\sum_{i=1}^{n} w_i\left(\frac{y_i}{\mu_i} - 1\right)x'_{ij}.\end{aligned} \tag{3.30}$$

By setting all the partial derivatives equal to zero and multiplying by ϕ, we get the ML equations

$$\sum_{i=1}^{n} w_i\left(\frac{y_i}{\mu_i} - 1\right)x'_{ij} = 0, \quad j = 0, 1, \ldots, r. \tag{3.31}$$

As mentioned before, the ML equations must almost always be solved numerically. When the ML equations have been solved we obtain the base value and the price relativities by using the fact that

$$\gamma_j = e^{\beta_j}, \quad j = 0, 1, \ldots, r. \tag{3.32}$$

Chapter 4

Non-life insurance pricing using the GAM

In the 1980's Hastie and Tibshirani analyzed the effect of continuous variables and introduced the GAM [6]. There are several possible approaches when using the GAM. In this thesis we will only be using smoothing splines [14].

4.1 Penalized deviances

Let us once again look at the general problem of how a key ratio Y is affected by M rating variables. In Section 3.2 we introduced the mean

$$\eta_i = \sum_{j=0}^{r} x'_{ij}\beta_j, \quad i = 1, 2, \ldots, n. \tag{4.1}$$

Hastie and Tibshirani introduced the GAM where instead of Equation 4.1 they assumed that

$$\eta_i = \beta_0 + \sum_{j=1}^{M} f_j(x_{ij}), \quad i = 1, 2, \ldots, n, \tag{4.2}$$

for some functions f_j. Here, x_{ij} is the value of rating variable j for observation i and should not be mixed up with the dummy variable x'_{ij}. Also, n is the total number of observations, not the total number of tariff cells. This model is more general since the mean depends more freely on the value of each rating variable. Let us consider a simple example where we model all the rating variables, except one, in the usual way. The mean is then given by

$$\eta_i = \sum_{j=0}^{r} x'_{ij}\beta_j + f(x_{i1}), \quad i = 1, 2, \ldots, n. \tag{4.3}$$

Our goal is now to find the function f that describes the effect of rating variable x_{i1} in the best way. We want the function to have a good fit to the data but we also want it to be smooth and not to vary wildly.

4.1.1 The deviance

To measure the goodness of fit of some estimated means $\hat{\mu} = \exp\{f(x)\}$ to the data y, we use the deviance $D(y, \hat{\mu})$ [12, §3.1], which is defined as

$$D(y, \hat{\mu}) = 2\phi \sum_{i=1}^{n} (\ell(y_i, y_i) - \ell(y_i, \hat{\mu})), \tag{4.4}$$

where ϕ is the dispersion parameter introduced in Equation 3.1. The deviance may be interpreted as the weighted sums of distances of the estimated means $\hat{\mu}$ from the data y. For the Poisson and gamma distributions, we get

$$D(y, \exp\{f(x)\}) = 2 \sum_{i=1}^{n} w_i \Big(y_i \log y_i - y_i f(x_i) - y_i + e^{f(x_i)} \Big), \tag{4.5}$$

$$D(y, \exp\{f(x)\}) = 2 \sum_{i=1}^{n} w_i \Big(y_i / e^{f(x_i)} - 1 - \log y_i + f(x_i) \Big). \tag{4.6}$$

4.1.2 The regularization

As a measure of the variability of the function f, we use the regularization $R(f(x))$ [12, §5.1], which is defined as

$$R(f(x)) = \int_a^b \left(f''(x)\right)^2 dx, \tag{4.7}$$

where a is a value lower than the lowest possible value of x and b is a value higher than the highest possible value of x. For a function that varies wildly the regularization will be high, whereas for a function with little variation the regularization will be low.

4.1.3 The smoothing parameter

Now when the deviance and the regularization have been defined, we are looking for the function f that minimizes the penalized deviance

$$\Delta(f(x)) = D(y, \exp\{f(x)\}) + \lambda R(f(x)), \tag{4.8}$$

where λ is called the smoothing parameter. The smoothing parameter creates a trade-off between good fit to the data and low variability of the function f. A small value of λ would increase the weight put on data, letting the function f vary freely (Figure 4.1), whereas a large value of λ would decrease the weight put on data, forcing the integrated squared second derivative of the function f to be small (Figure 4.2). In Chapter 5 we will compare two different methods for finding the optimal smoothing parameter.

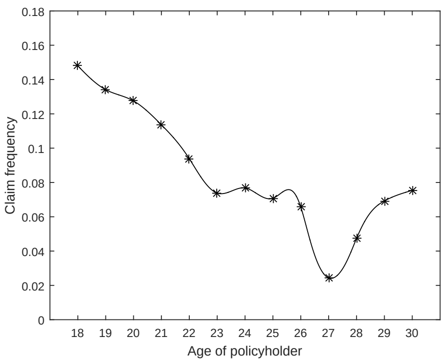

Figure 4.1: A small value of λ would increase the weight put on data, letting the function f vary freely.

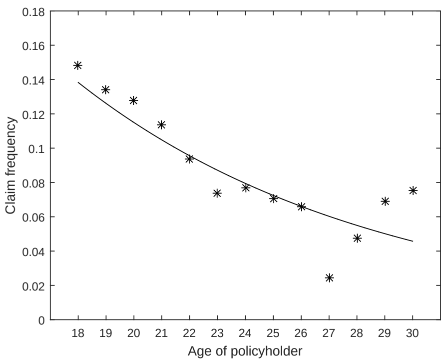

Figure 4.2: A large value of λ would decrease the weight put on data, forcing the integrated squared second derivative of the function f to be small.

4.2 Smoothing splines

Smoothing splines have important properties in connection to penalized deviances and it can be shown that among all twice continuously differentiable functions, the natural cubic spline minimizes the integrated squared second derivative [12, Appendix B.1]. This important feature is contained in the following theorem:

Theorem 4.1 *For any points $u_1 < \cdots < u_m$ and real numbers $y_1, \ldots, y_m$, there exists a unique natural cubic spline s, such that $s(u_k) = y_k,\ k = 1, \ldots, m$. Furthermore, if f is any twice continuously differentiable function such that $f(u_k) = y_k$, $k = 1, \ldots, m$, then for any $a \leq u_1$ and $b \geq u_m$,*

$$\int_a^b (s''(x))^2\, dx \leq \int_a^b (f''(x))^2\, dx. \tag{4.9}$$

By Theorem 4.1 there exists a unique natural cubic spline s, such that $\{s(u_k) = f(u_k),\ k = 1, \ldots, m\}$, and since $D(y, \exp\{s(x)\}) = D(y, \exp\{f(x)\})$ it follows that

$$\Delta(s(x)) \leq \Delta(f(x)). \tag{4.10}$$

This means that when looking for the twice continuously differentiable function f that minimizes Equation 4.8 we only need to consider the set of natural cubic splines. Since any spline function of given degree can be expressed as a linear combination of B-splines of that degree [12, Appendix B.2], we are going to use B-splines to parameterize the set of natural cubic splines. Let $B_{3k}(x)$ be the k:th B-spline of order 3. The following useful theorem will be used extensively throughout the rest of this paper.

Theorem 4.2 *For a given set of m knots, a cubic spline s may be written as*

$$s(x) = \sum_{k=1}^{m+2} \beta_k B_{3k}(x), \tag{4.11}$$

for unique constants $\beta_1, \dots, \beta_{m+2}$.

Here, $\beta_1, \dots, \beta_{m+2}$ are constants which decide the weight of each B-spline and should not be mixed up with the regression parameters defined in Section 3.2.

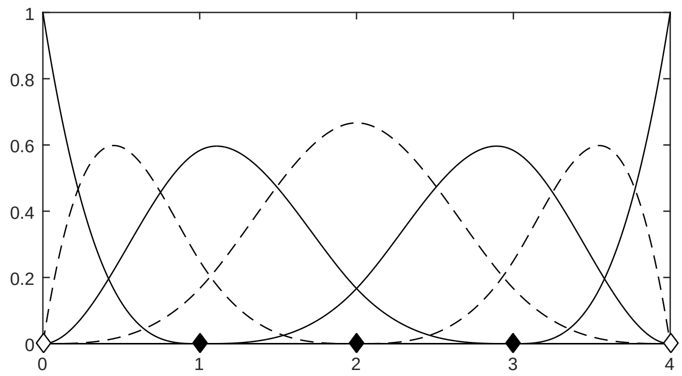

Figure 4.3: B-splines of order 3. The three internal knots are marked with black diamonds and the two boundary knots are marked with white diamonds.

4.3 Price relativities - one rating variable

It is now time to estimate the price relativities by finding the natural cubic spline s that minimizes the penalized deviance

$$\Delta(s(x)) = D(y, \exp\{s(x)\}) + \lambda R(s(x)). \tag{4.12}$$

We will start by considering one rating variable and we will treat two different cases, one for the claim frequency and one for the claim severity.

4.3.1 The claim frequency

Let us assume that we have one continuous rating variable, and that the observations of the key ratio Y are Poisson distributed. When storing insurance data, continuous rating variables are usually rounded, so instead of assuming millions of different values, they usually assume a much smaller number of different values. For instance, the age of the policyholder is usually stored in years and rounded down to the nearest integer, so less than a hundred different values occur.

Let x_i be the value of the rating variable for observation i and let $z_1, \dots, z_m$ denote the possible values of x_i, in ascending order. For the Poisson distribution, the deviance is given by

$$D(y, \exp\{s(x)\}) = 2\sum_{i=1}^{n} w_i \Big(y_i \log y_i - y_i s(x_i) - y_i + e^{s(x_i)} \Big). \tag{4.13}$$

Since $a \leq z_1$ and $b \geq z_m$ the regularization can be expressed as

$$R(s(x)) = \int_a^{z_1} \left(s''(x)\right)^2 dx + \int_{z_1}^{z_m} \left(s''(x)\right)^2 dx + \int_{z_m}^{b} \left(s''(x)\right)^2 dx. \tag{4.14}$$

The natural cubic spline is linear outside $[z_1, z_m]$ so the first and third term of Equation 4.14 equals zero and the regularization can be simplified to

$$R(s(x)) = \int_{z_1}^{z_m} \left(s''(x)\right)^2 dx. \tag{4.15}$$

The penalized deviance can now be expressed as

$$\begin{aligned} \Delta(s(x)) = 2\sum_{i=1}^{n} w_i \Big(y_i \log y_i - y_i s(x_i) - y_i + e^{s(x_i)} \Big) \\ + \lambda \int_{z_1}^{z_m} \left(s''(x)\right)^2 dx. \end{aligned} \tag{4.16}$$

It is time to use the fact that the natural cubic spline s can be expressed as a sum of B-splines (Theorem 4.2), such that

$$s(x) = \sum_{j=1}^{m+2} \beta_j B_j(x), \tag{4.17}$$

where $\beta_1, \dots, \beta_{m+2}$ are unique parameters and $B_1(x), \dots, B_{m+2}(x)$ are the cubic B-splines with knots $z_1, \dots, z_m$. The penalized deviance can then be expressed as a function of the parameters $\boldsymbol{\beta} = [\beta_1, \dots, \beta_{m+2}]$ and is given by

$$\begin{aligned} \Delta(\boldsymbol{\beta}) = 2\sum_{i=1}^{n} w_i \Bigg(y_i \log y_i - y_i \sum_{j=1}^{m+2} \beta_j B_j(x_i) - y_i \\ + \exp\left\{ \sum_{j=1}^{m+2} \beta_j B_j(x_i) \right\} \Bigg) + \lambda \sum_{j=1}^{m+2} \sum_{k=1}^{m+2} \beta_j \beta_k \Omega_{jk}, \end{aligned} \tag{4.18}$$

where

$$\Omega_{jk} = \int_{z_1}^{z_m} B_j''(x) B_k''(x)\, dx. \tag{4.19}$$

The numbers Ω_{jk} can be computed using the basic properties of B-splines [12, Appendix B.2]. We are now ready to use partial derivatives to find the minimizing parameters $\beta_1, \ldots, \beta_{m+2}$. The partial derivatives are given by

$$\begin{aligned}
\frac{\partial \Delta}{\partial \beta_\ell} &= 2\sum_{i=1}^{n} w_i \left(-y_i B_\ell(x_i) + \exp\left\{ \sum_{j=1}^{m+2} \beta_j B_j(x_i) \right\} B_\ell(x_i) \right) + 2\lambda \sum_{j=1}^{m+2} \beta_j \Omega_{j\ell} \\
&= \{\text{Letting } I_k \text{ denote the set of } i \text{ for which } x_i = z_k\} \\
&= 2\sum_{k=1}^{m} \sum_{i \in I_k} w_i \left(-y_i B_\ell(z_k) + \exp\left\{ \sum_{j=1}^{m+2} \beta_j B_j(z_k) \right\} B_\ell(z_k) \right) + 2\lambda \sum_{j=1}^{m+2} \beta_j \Omega_{j\ell} \\
&= 2\sum_{k=1}^{m} \tilde{w}_k \left(-\tilde{y}_k + \exp\left\{ \sum_{j=1}^{m+2} \beta_j B_j(z_k) \right\} \right) B_\ell(z_k) + 2\lambda \sum_{j=1}^{m+2} \beta_j \Omega_{j\ell},
\end{aligned} \tag{4.20}$$

where

$$\tilde{w}_k = \sum_{i \in I_k} w_i, \qquad \tilde{y}_k = \frac{1}{\tilde{w}_k} \sum_{i \in I_k} w_i y_i. \tag{4.21}$$

In the next step, we set the partial derivatives equal to zero

$$\begin{aligned}
\frac{\partial \Delta}{\partial \beta_\ell} = 2\sum_{k=1}^{m} \tilde{w}_k \left(-\tilde{y}_k + \exp\left\{ \sum_{j=1}^{m+2} \beta_j B_j(z_k) \right\} \right) B_\ell(z_k) \\
+2\lambda \sum_{j=1}^{m+2} \beta_j \Omega_{j\ell} = 0, \quad \ell = 1, 2, \ldots, m+2,
\end{aligned} \tag{4.22}$$

and obtain the equations

$$\begin{aligned}
\sum_{k=1}^{m} \tilde{w}_k \exp\left\{ \sum_{j=1}^{m+2} \beta_j B_j(z_k) \right\} B_\ell(z_k) - \sum_{k=1}^{m} \tilde{w}_k \tilde{y}_k B_\ell(z_k) \\
+\lambda \sum_{j=1}^{m+2} \beta_j \Omega_{j\ell} = 0, \quad \ell = 1, 2, \ldots, m+2.
\end{aligned} \tag{4.23}$$

Since the first term in Equation 4.23 depends in a non-linear way on $\beta_1, \ldots, \beta_{m+2}$, we will use Newton-Raphson's method [9] to solve the equation system. We begin by defining $h_\ell(\beta_1, \ldots, \beta_{m+2})$, such that

$$\begin{aligned}
h_\ell(\beta_1, \ldots, \beta_{m+2}) = \sum_{k=1}^{m} \tilde{w}_k \exp\left\{ \sum_{j=1}^{m+2} \beta_j B_j(z_k) \right\} B_\ell(z_k) \\
-\sum_{k=1}^{m} \tilde{w}_k \tilde{y}_k B_\ell(z_k) + \lambda \sum_{j=1}^{m+2} \beta_j \Omega_{j\ell}, \quad \ell = 1, 2, \ldots, m+2.
\end{aligned} \tag{4.24}$$

To solve the equations $h_\ell(\beta_1, \ldots, \beta_{m+2}) = 0$ we proceed by iteratively solving the linear equation systems for the unknowns $\beta_1^{(n+1)}, \ldots, \beta_{m+2}^{(n+1)}$

$$h_\ell\Big(\beta_1^{(n)}, \ldots, \beta_{m+2}^{(n)}\Big) + \sum_{j=1}^{m+2}\Big(\beta_j^{(n+1)} - \beta_j^{(n)}\Big)\frac{\partial h_\ell}{\partial \beta_j}\Big(\beta_1^{(n)}, \ldots, \beta_{m+2}^{(n)}\Big) = 0, \qquad \ell = 1, 2, \ldots, m+2, \tag{4.25}$$

where

$$\frac{\partial h_\ell}{\partial \beta_j} = \sum_{k=1}^{m} \tilde{w}_k \exp\left\{\sum_{j=1}^{m+2} \beta_j B_j(z_k)\right\} B_j(z_k) B_\ell(z_k) + \lambda \Omega_{j\ell}. \tag{4.26}$$

To make the notation a bit easier, we define

$$\gamma_k^{(n)} = \exp\left\{\sum_{j=1}^{m+2} \beta_j^{(n)} B_j(z_k)\right\}. \tag{4.27}$$

Using the properties of Equations 4.24, 4.26 and 4.27, Equation 4.25 can be simplified to

$$\begin{aligned}\sum_{k=1}^{m} \tilde{w}_k \gamma_k^{(n)} B_\ell(z_k) - \sum_{k=1}^{m} \tilde{w}_k \tilde{y}_k B_\ell(z_k) + \lambda \sum_{j=1}^{m+2} \beta_j^{(n)} \Omega_{j\ell} \\ + \sum_{j=1}^{m+2}\Big(\beta_j^{(n+1)} - \beta_j^{(n)}\Big)\left(\sum_{k=1}^{m} \tilde{w}_k \gamma_k^{(n)} B_j(z_k) B_\ell(z_k) + \lambda \Omega_{j\ell}\right) = 0, \\ \ell = 1, 2, \ldots, m+2.\end{aligned} \tag{4.28}$$

Collecting $\beta_j^{(n+1)}$ on one side of the equality and $\beta_j^{(n)}$ on the other side, we get

$$\begin{aligned}\sum_{j=1}^{m+2}\sum_{k=1}^{m} \tilde{w}_k \gamma_k^{(n)} B_j(z_k) B_\ell(z_k) \beta_j^{(n+1)} + \lambda \sum_{j=1}^{m+2} \beta_j^{(n+1)} \Omega_{j\ell} \\ = \sum_{k=1}^{m} \tilde{w}_k \gamma_k^{(n)}\left(\tilde{y}_k/\gamma_k^{(n)} - 1 + \sum_{j=1}^{m+2} \beta_j^{(n)} B_j(z_k)\right) B_\ell(z_k), \\ \ell = 1, 2, \ldots, m+2.\end{aligned} \tag{4.29}$$

Let us now introduce the $m \times (m+2)$ matrix $\mathbf{B}$ by

$$\mathbf{B} = \begin{pmatrix} B_1(z_1) & B_2(z_1) & \cdots & B_{m+2}(z_1) \\ B_1(z_2) & B_2(z_2) & \cdots & B_{m+2}(z_2) \\ \vdots & \vdots & \ddots & \vdots \\ B_1(z_m) & B_2(z_m) & \cdots & B_{m+2}(z_m) \end{pmatrix}, \tag{4.30}$$

and the $m \times m$ diagonal matrix $\mathbf{W}^{(n)}$ by

$$\mathbf{W}^{(n)} = \begin{pmatrix} \tilde{w}_1\gamma_1^{(n)} & 0 & \cdots & 0 \\ 0 & \tilde{w}_2\gamma_2^{(n)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \tilde{w}_m\gamma_m^{(n)} \end{pmatrix}. \tag{4.31}$$

Furthermore, let $\mathbf{\Omega}$ denote the symmetric $(m+2) \times (m+2)$ matrix with elements Ω_{jk}. Also, let $\boldsymbol{\beta}^{(n)}$ denote the vector with elements $\beta_j^{(n)}$ and $\mathbf{y}^{(n)}$ the vector with elements $\tilde{y}_k/\gamma_k^{(n)} - 1 + \sum_{j=1}^{m+2} \beta_j^{(n)} B_j(z_k)$. The system of linear equations may now be written on matrix form as

$$\Big(\mathbf{B}'\mathbf{W}^{(n)}\mathbf{B} + \lambda\mathbf{\Omega}\Big)\boldsymbol{\beta}^{(n+1)} = \mathbf{B}'\mathbf{W}^{(n)}\mathbf{y}^{(n)}, \tag{4.32}$$

and the equation system can be solved using Newton-Raphson's method to obtain $\boldsymbol{\beta}$. The natural cubic spline $s(x)$ can then be calculated using Equation 4.17 and the price relativities are given by $e^{s(z_1)}, \ldots, e^{s(z_m)}$.

4.3.2 The claim severity

Let us again assume that we have one continuous rating variable, but this time we assume that the observations of the key ratio Y are gamma distributed. For the gamma distribution, the deviance is given by

$$D(y, \exp\{s(x)\}) = 2\sum_{i=1}^{n} w_i \Big(y_i/e^{s(x_i)} - 1 - \log y_i + s(x_i)\Big). \tag{4.33}$$

If we combine the expression above with the conclusion from Equation 4.15 we can express the penalized deviance as

$$\Delta(s(x)) = 2\sum_{i=1}^{n} w_i \Big(y_i/e^{s(x_i)} - 1 - \log y_i + s(x_i)\Big) + \lambda \int_{z_1}^{z_m} \left(s''(x)\right)^2 dx. \tag{4.34}$$

It is once again time to use the fact that the natural cubic spline s can be expressed as a sum of B-splines. The penalized deviance can then be expressed as a function of the parameters $\beta_1, \ldots, \beta_{m+2}$ and is given by

$$\begin{aligned} \Delta(\boldsymbol{\beta}) = 2\sum_{i=1}^{n} w_i \Bigg(y_i/\exp\Bigg\{\sum_{j=1}^{m+2} \beta_j B_j(x_i)\Bigg\} - 1 - \log y_i \\ + \sum_{j=1}^{m+2} \beta_j B_j(x_i)\Bigg) + \lambda \sum_{j=1}^{m+2}\sum_{k=1}^{m+2} \beta_j\beta_k\Omega_{jk}. \end{aligned} \tag{4.35}$$

We are now ready to use partial derivatives to find the minimizing parameters $\beta_1, \ldots, \beta_{m+2}$. The partial derivatives are given by

$$
\begin{aligned}
\frac{\partial \Delta}{\partial \beta_\ell} &= 2\sum_{i=1}^{n} w_i \left(-y_i / \exp\left\{ \sum_{j=1}^{m+2} \beta_j B_j(x_i) \right\} B_\ell(x_i) + B_\ell(x_i) \right) + 2\lambda \sum_{j=1}^{m+2} \beta_j \Omega_{j\ell} \\
&= 2\sum_{k=1}^{m} \sum_{i \in I_k} w_i \left(-y_i / \exp\left\{ \sum_{j=1}^{m+2} \beta_j B_j(z_k) \right\} B_\ell(z_k) + B_\ell(z_k) \right) + 2\lambda \sum_{j=1}^{m+2} \beta_j \Omega_{j\ell} \\
&= 2\sum_{k=1}^{m} \tilde{w}_k \left(-\tilde{y}_k / \exp\left\{ \sum_{j=1}^{m+2} \beta_j B_j(z_k) \right\} + 1 \right) B_\ell(z_k) + 2\lambda \sum_{j=1}^{m+2} \beta_j \Omega_{j\ell}.
\end{aligned} \tag{4.36}
$$

In the next step, we set the partial derivatives equal to zero

$$
\begin{aligned}
\frac{\partial \Delta}{\partial \beta_\ell} = 2\sum_{k=1}^{m} \tilde{w}_k \left(-\tilde{y}_k / \exp\left\{ \sum_{j=1}^{m+2} \beta_j B_j(z_k) \right\} + 1 \right) B_\ell(z_k) \\
+2\lambda \sum_{j=1}^{m+2} \beta_j \Omega_{j\ell} = 0, \quad \ell = 1, 2, \ldots, m+2,
\end{aligned} \tag{4.37}
$$

and obtain the equations

$$
\begin{aligned}
-\sum_{k=1}^{m} \tilde{w}_k \tilde{y}_k B_\ell(z_k) / \exp\left\{ \sum_{j=1}^{m+2} \beta_j B_j(z_k) \right\} + \sum_{k=1}^{m} \tilde{w}_k B_\ell(z_k) \\
+\lambda \sum_{j=1}^{m+2} \beta_j \Omega_{j\ell} = 0, \quad \ell = 1, 2, \ldots, m+2.
\end{aligned} \tag{4.38}
$$

Since the first term in Equation 4.38 depends in a non-linear way on $\beta_1, \ldots, \beta_{m+2}$, we will once again use Newton-Raphson's method to solve the equation system. We begin by defining $h_\ell(\beta_1, \ldots, \beta_{m+2})$, such that

$$
\begin{aligned}
h_\ell(\beta_1, \ldots, \beta_{m+2}) = -\sum_{k=1}^{m} \tilde{w}_k \tilde{y}_k B_\ell(z_k) / \exp\left\{ \sum_{j=1}^{m+2} \beta_j B_j(z_k) \right\} \\
+\sum_{k=1}^{m} \tilde{w}_k B_\ell(z_k) + \lambda \sum_{j=1}^{m+2} \beta_j \Omega_{j\ell}, \quad \ell = 1, 2, \ldots, m+2.
\end{aligned} \tag{4.39}
$$

To solve the equations $h_\ell(\beta_1, \ldots, \beta_{m+2}) = 0$ we proceed by iteratively solving the linear equation systems for the unknowns $\beta_1^{(n+1)}, \ldots, \beta_{m+2}^{(n+1)}$

$$
\begin{aligned}
h_\ell\left(\beta_1^{(n)}, \ldots, \beta_{m+2}^{(n)}\right) + \sum_{j=1}^{m+2} \left(\beta_j^{(n+1)} - \beta_j^{(n)}\right) \frac{\partial h_\ell}{\partial \beta_j} \left(\beta_1^{(n)}, \ldots, \beta_{m+2}^{(n)}\right) = 0, \\
\ell = 1, 2, \ldots, m+2,
\end{aligned} \tag{4.40}
$$

where

$$\frac{\partial h_\ell}{\partial \beta_j} = \sum_{k=1}^{m} \tilde{w}_k \tilde{y}_k B_j(z_k) B_\ell(z_k) / \exp\left\{ \sum_{j=1}^{m+2} \beta_j B_j(z_k) \right\} + \lambda \mathit{\Omega}_{j\ell}. \tag{4.41}$$

Using the properties of Equations 4.27, 4.39 and 4.41, Equation 4.40 can be simplified to

$$\begin{aligned} &-\sum_{k=1}^{m} \tilde{w}_k \tilde{y}_k B_\ell(z_k)/\gamma_k^{(n)} + \sum_{k=1}^{m} \tilde{w}_k B_\ell(z_k) + \lambda \sum_{j=1}^{m+2} \beta_j^{(n)} \mathit{\Omega}_{j\ell} \\ &+ \sum_{j=1}^{m+2} \left(\beta_j^{(n+1)} - \beta_j^{(n)} \right) \left(\sum_{k=1}^{m} \tilde{w}_k \tilde{y}_k B_j(z_k) B_\ell(z_k)/\gamma_k^{(n)} + \lambda \mathit{\Omega}_{j\ell} \right) = 0, \\ &\ell = 1, 2, \ldots, m+2. \end{aligned} \tag{4.42}$$

Collecting $\beta_j^{(n+1)}$ on one side of the equality and $\beta_j^{(n)}$ on the other side, we get

$$\begin{aligned} &\sum_{j=1}^{m+2} \sum_{k=1}^{m} \frac{\tilde{w}_k \tilde{y}_k}{\gamma_k^{(n)}} B_j(z_k) B_\ell(z_k) \beta_j^{(n+1)} + \lambda \sum_{j=1}^{m+2} \beta_j^{(n+1)} \mathit{\Omega}_{j\ell} \\ &= \sum_{k=1}^{m} \frac{\tilde{w}_k \tilde{y}_k}{\gamma_k^{(n)}} \left(1 - \frac{\gamma_k^{(n)}}{\tilde{y}_k} + \sum_{j=1}^{m+2} \beta_j^{(n)} B_j(z_k) \right) B_\ell(z_k), \\ &\ell = 1, 2, \ldots, m+2. \end{aligned} \tag{4.43}$$

Let us now introduce the $m \times m$ diagonal matrix $\mathbf{W}^{(n)}$ by

$$\mathbf{W}^{(n)} = \begin{pmatrix} \tilde{w}_1 \tilde{y}_1/\gamma_1^{(n)} & 0 & \cdots & 0 \\ 0 & \tilde{w}_2 \tilde{y}_2/\gamma_2^{(n)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \tilde{w}_m \tilde{y}_m/\gamma_m^{(n)} \end{pmatrix}. \tag{4.44}$$

Furthermore, let $\boldsymbol{\beta}^{(n)}$ denote the vector with elements $\beta_j^{(n)}$ and $\mathbf{y}^{(n)}$ the vector with elements $1 - \gamma_k^{(n)}/\tilde{y}_k + \sum_{j=1}^{m+2} \beta_j^{(n)} B_j(z_k)$. The system of linear equations may now be written on matrix form as

$$\left(\mathbf{B}' \mathbf{W}^{(n)} \mathbf{B} + \lambda \mathbf{\Omega} \right) \boldsymbol{\beta}^{(n+1)} = \mathbf{B}' \mathbf{W}^{(n)} \mathbf{y}^{(n)}, \tag{4.45}$$

and the equation system can be solved using Newton-Raphson's method to obtain $\boldsymbol{\beta}$. The natural cubic spline $s(x)$ can then be calculated using Equation 4.17 and the price relativities are given by $e^{s(z_1)}, \ldots, e^{s(z_m)}$.

4.4 Price relativities - several rating variables

We will now continue by considering several rating variables. The backfitting algorithm is an iterative procedure used for estimation of the price relativities within the GAM [7]. The idea behind the backfitting algorithm is to reduce the estimation problem to a one dimensional problem and only consider one continuous rating variable at a time. Let us assume that we have a large number of categorical rating variables together with two continuous rating variables. The generalization to the case with an arbitrary number of continuous rating variables is completely straightforward. We let x_{1i} and x_{2i} be the values of observation i for the first and second continuous rating variable, respectively. We also let $z_{11}, \ldots, z_{1m_1}$ and $z_{21}, \ldots, z_{2m_2}$ denote the possible values for the continuous rating variables. The mean for this model can then be expressed as

$$\eta_i = \sum_{j=0}^{r} x'_{ij}\beta_j + \sum_{k=1}^{m_1+2} \beta_{1k}B_{1k}(x_{1i}) + \sum_{\ell=1}^{m_2+2} \beta_{2\ell}B_{2\ell}(x_{2i}), \quad i = 1, 2, \ldots, n. \tag{4.46}$$

We will once again treat two different cases, one for the claim frequency and one for the claim severity.

4.4.1 The claim frequency

Let us first assume that the observations of the key ratio Y are Poisson distributed. The penalized deviance can then be expressed as

$$\begin{aligned} \Delta(\boldsymbol{\beta_1}, \boldsymbol{\beta_2}) &= 2\sum_{i=1}^{n} w_i(y_i \log y_i - y_i\eta_i - y_i + e^{\eta_i}) \\ &+\lambda_1 \sum_{j=1}^{m_1+2}\sum_{k=1}^{m_1+2} \beta_{1j}\beta_{1k}\Omega^{(1)}_{jk} + \lambda_2 \sum_{j=1}^{m_2+2}\sum_{k=1}^{m_2+2} \beta_{2j}\beta_{2k}\Omega^{(2)}_{jk}, \end{aligned} \tag{4.47}$$

where λ_1 and λ_2 are the smoothing parameters for the first and second continuous rating variable, respectively. Let us also introduce the notation

$$\nu_{0i} = \exp\left\{\sum_{j=0}^{r} x'_{ij}\beta_j\right\}, \tag{4.48}$$

$$\nu_{1i} = \exp\left\{\sum_{j=1}^{m_1+2} \beta_{1j}B_{1j}(x_{1i})\right\}, \tag{4.49}$$

$$\nu_{2i} = \exp\left\{\sum_{j=1}^{m_2+2} \beta_{2j}B_{2j}(x_{2i})\right\}. \tag{4.50}$$

The deviance can then be written as

$$
\begin{aligned}
D &= 2\sum_{i=1}^{n} w_i(y_i \log y_i - y_i\eta_i - y_i + e^{\eta_i}) \\
&= 2\sum_{i=1}^{n} w_i(y_i \log y_i - y_i \log \mu_i - y_i + \mu_i) \\
&= 2\sum_{i=1}^{n} w_i(y_i \log y_i - y_i \log(\nu_{0i}\nu_{1i}\nu_{2i}) - y_i + \nu_{0i}\nu_{1i}\nu_{2i}) \\
&= 2\sum_{i=1}^{n} w_i\nu_{1i}\nu_{2i}\left(\frac{y_i}{\nu_{1i}\nu_{2i}} \log \frac{y_i}{\nu_{1i}\nu_{2i}} - \frac{y_i}{\nu_{1i}\nu_{2i}} \log \nu_{0i} - \frac{y_i}{\nu_{1i}\nu_{2i}} + \nu_{0i}\right) \\
&= 2\sum_{i=1}^{n} w_i\nu_{0i}\nu_{2i}\left(\frac{y_i}{\nu_{0i}\nu_{2i}} \log \frac{y_i}{\nu_{0i}\nu_{2i}} - \frac{y_i}{\nu_{0i}\nu_{2i}} \log \nu_{1i} - \frac{y_i}{\nu_{0i}\nu_{2i}} + \nu_{1i}\right) \\
&= 2\sum_{i=1}^{n} w_i\nu_{0i}\nu_{1i}\left(\frac{y_i}{\nu_{0i}\nu_{1i}} \log \frac{y_i}{\nu_{0i}\nu_{1i}} - \frac{y_i}{\nu_{0i}\nu_{1i}} \log \nu_{2i} - \frac{y_i}{\nu_{0i}\nu_{1i}} + \nu_{2i}\right).
\end{aligned}
\tag{4.51}
$$

If $\beta_{11}, \ldots, \beta_{1,m_1+2}$ and $\beta_{21}, \ldots, \beta_{2,m_2+2}$ were known, the problem of estimating $\beta_0, \ldots, \beta_r$ would be identical to the problem with only categorical rating variables (the regularization terms in Equation 4.47 are only constants in this case), but with observations $y_i/(\nu_{1i}\nu_{2i})$ and weights $w_i\nu_{1i}\nu_{2i}$ as can be seen in Equation 4.51. Similarly, if $\beta_0, \ldots, \beta_r$ and $\beta_{21}, \ldots, \beta_{2,m_2+2}$ were known, the problem of estimating $\beta_{11}, \ldots, \beta_{1,m_1+2}$ is precisely the one treated in Section 4.3.1, with the observations $y_i/(\nu_{0i}\nu_{2i})$ and the weights $w_i\nu_{0i}\nu_{2i}$. The corresponding statement of course holds for the second continuous rating variable as well.

To be able to use the backfitting algorithm we need some initial estimates. A good initial estimate would be to derive $\hat{\beta}_0^{(0)}, \ldots, \hat{\beta}_r^{(0)}$ by analyzing the data with the continuous rating variables excluded and to set all $\hat{\beta}_{11}^{(0)}, \ldots, \hat{\beta}_{1,m_1+2}^{(0)}$ and $\hat{\beta}_{21}^{(0)}, \ldots, \hat{\beta}_{2,m_2+2}^{(0)}$ to zero. Given this set of initial estimates we define

$$
\hat{\nu}_{0i}^{(0)} = \exp\left\{\sum_{j=0}^{r} x'_{ij}\hat{\beta}_j^{(0)}\right\}, \tag{4.52}
$$

$$
\hat{\nu}_{1i}^{(0)} = \exp\left\{\sum_{j=1}^{m_1+2} \hat{\beta}_{1j}^{(0)} B_{1j}(x_{1i})\right\} = 1, \tag{4.53}
$$

$$
\hat{\nu}_{2i}^{(0)} = \exp\left\{\sum_{j=1}^{m_2+2} \hat{\beta}_{2j}^{(0)} B_{2j}(x_{2i})\right\} = 1. \tag{4.54}
$$

We continue by iterating the following three steps until the estimates have converged (we have used the tolerance 1e-10 throughout this paper). For iteration q we have:

- Step 1

Compute $\hat{\beta}_{11}^{(q)}, \ldots, \hat{\beta}_{1,m_1+2}^{(q)}$ based on the observations $y_i / \Big(\hat{\nu}_{0i}^{(q-1)} \hat{\nu}_{2i}^{(q-1)} \Big)$ and the weights $w_i \hat{\nu}_{0i}^{(q-1)} \hat{\nu}_{2i}^{(q-1)}$ with x_{1i} being the only explanatory rating variable. Calculate $\hat{\nu}_{1i}^{(q)}$ with the formula $\hat{\nu}_{1i}^{(q)} = \exp\Big\{ \sum_{j=1}^{m_1+2} \hat{\beta}_{1j}^{(q)} B_{1j}(x_{1i}) \Big\}$.

- Step 2

Compute $\hat{\beta}_{21}^{(q)}, \ldots, \hat{\beta}_{2,m_2+2}^{(q)}$ based on the observations $y_i / \Big(\hat{\nu}_{0i}^{(q-1)} \hat{\nu}_{1i}^{(q)} \Big)$ and the weights $w_i \hat{\nu}_{0i}^{(q-1)} \hat{\nu}_{1i}^{(q)}$ with x_{2i} being the only explanatory rating variable. Calculate $\hat{\nu}_{2i}^{(q)}$ with the formula $\hat{\nu}_{2i}^{(q)} = \exp\Big\{ \sum_{j=1}^{m_2+2} \hat{\beta}_{2j}^{(q)} B_{2j}(x_{2i}) \Big\}$.

- Step 3

Compute $\hat{\beta}_0^{(q)}, \ldots, \hat{\beta}_r^{(q)}$ based on the observations $y_i / \Big(\hat{\nu}_{1i}^{(q)} \hat{\nu}_{2i}^{(q)} \Big)$ and the weights $w_i \hat{\nu}_{1i}^{(q)} \hat{\nu}_{2i}^{(q)}$ using only the categorical rating variables. Calculate $\hat{\nu}_{0i}^{(q)}$ with the formula $\hat{\nu}_{0i}^{(q)} = \exp\Big\{ \sum_{j=0}^{r} x'_{ij} \hat{\beta}_j^{(q)} \Big\}$.

When the estimates have converged, we obtain the price relativities for the categorical rating variables, using the fact that

$$\gamma_j = e^{\beta_j}, \quad j = 0, 1, \ldots, r. \tag{4.55}$$

We obtain the natural cubic splines for the continuous rating variables by

$$s_1(x) = \sum_{j=1}^{m_1+2} \beta_{1j} B_{1j}(x), \tag{4.56}$$

$$s_2(x) = \sum_{j=1}^{m_2+2} \beta_{2j} B_{2j}(x), \tag{4.57}$$

and the price relativities for the continuous rating variables are given by $e^{s_1(z_{11})}, \ldots, e^{s_1\left(z_{1m_1}\right)}$ and $e^{s_2(z_{21})}, \ldots, e^{s_2\left(z_{2m_2}\right)}$.

4.4.2 The claim severity

Let us now instead assume that the observations of the key ratio Y are gamma distributed. The penalized deviance can then be expressed as

$$\begin{aligned}\Delta(\boldsymbol{\beta_1}, \boldsymbol{\beta_2}) = 2\sum_{i=1}^{n} w_i(y_i/e^{\eta_i} - 1 - \log(y_i/e^{\eta_i}))\\ +\lambda_1 \sum_{j=1}^{m_1+2}\sum_{k=1}^{m_1+2} \beta_{1j}\beta_{1k}\Omega_{jk}^{(1)} + \lambda_2 \sum_{j=1}^{m_2+2}\sum_{k=1}^{m_2+2} \beta_{2j}\beta_{2k}\Omega_{jk}^{(2)}.\end{aligned} \tag{4.58}$$

The deviance can be written as

$$\begin{aligned} D &= 2\sum_{i=1}^{n} w_i\Big(\frac{y_i}{e^{\eta_i}} - 1 - \log\frac{y_i}{e^{\eta_i}}\Big)\\ &= 2\sum_{i=1}^{n} w_i\Big(\frac{y_i}{\mu_i} - 1 - \log\frac{y_i}{\mu_i}\Big)\\ &= 2\sum_{i=1}^{n} w_i\Big(\frac{y_i}{\nu_{0i}\nu_{1i}\nu_{2i}} - 1 - \log\frac{y_i}{\nu_{0i}\nu_{1i}\nu_{2i}}\Big)\\ &= 2\sum_{i=1}^{n} w_i\Big(\frac{y_i/\nu_{1i}\nu_{2i}}{\nu_{0i}} - 1 - \log\frac{y_i/\nu_{1i}\nu_{2i}}{\nu_{0i}}\Big)\\ &= 2\sum_{i=1}^{n} w_i\Big(\frac{y_i/\nu_{0i}\nu_{2i}}{\nu_{1i}} - 1 - \log\frac{y_i/\nu_{0i}\nu_{2i}}{\nu_{1i}}\Big)\\ &= 2\sum_{i=1}^{n} w_i\Big(\frac{y_i/\nu_{0i}\nu_{1i}}{\nu_{2i}} - 1 - \log\frac{y_i/\nu_{0i}\nu_{1i}}{\nu_{2i}}\Big). \end{aligned} \tag{4.59}$$

If $\beta_{11}, \ldots, \beta_{1,m_1+2}$ and $\beta_{21}, \ldots, \beta_{2,m_2+2}$ were known, the problem of estimating $\beta_0, \ldots, \beta_r$ would be identical to the problem with only categorical rating variables (the regularization terms in Equation 4.58 are only constants in this case), but with observations $y_i/(\nu_{1i}\nu_{2i})$ as can be seen in Equation 4.59. Similarly, if $\beta_0, \ldots, \beta_r$ and $\beta_{21}, \ldots, \beta_{2,m_2+2}$ were known, the problem of estimating $\beta_{11}, \ldots, \beta_{1,m_1+2}$ is precisely the one treated in Section 4.3.2, with the observations $y_i/(\nu_{0i}\nu_{2i})$. The corresponding statement of course holds for the second continuous rating variable as well.

We use the same initial estimates as in the Poisson case and we continue by iterating the following three steps until the estimates have converged. For iteration q we have:

- Step 1
Compute $\hat{\beta}_{11}^{(q)}, \ldots, \hat{\beta}_{1,m_1+2}^{(q)}$ based on the observations $y_i / \left(\hat{\nu}_{0i}^{(q-1)} \hat{\nu}_{2i}^{(q-1)}\right)$ with x_{1i} being the only explanatory rating variable. Calculate $\hat{\nu}_{1i}^{(q)}$ with the formula $\hat{\nu}_{1i}^{(q)} = \exp\left\{\sum_{j=1}^{m_1+2} \hat{\beta}_{1j}^{(q)} B_{1j}(x_{1i})\right\}$.

- Step 2
Compute $\hat{\beta}_{21}^{(q)}, \ldots, \hat{\beta}_{2,m_2+2}^{(q)}$ based on the observations $y_i / \left(\hat{\nu}_{0i}^{(q-1)} \hat{\nu}_{1i}^{(q)}\right)$ with x_{2i} being the only explanatory rating variable. Calculate $\hat{\nu}_{2i}^{(q)}$ with the formula $\hat{\nu}_{2i}^{(q)} = \exp\left\{\sum_{j=1}^{m_2+2} \hat{\beta}_{2j}^{(q)} B_{2j}(x_{2i})\right\}$.

- Step 3
Compute $\hat{\beta}_{0}^{(q)}, \ldots, \hat{\beta}_{r}^{(q)}$ based on the observations $y_i / \left(\hat{\nu}_{1i}^{(q)} \hat{\nu}_{2i}^{(q)}\right)$ using only the categorical rating variables. Calculate $\hat{\nu}_{0i}^{(q)}$ with the formula
$\hat{\nu}_{0i}^{(q)} = \exp\left\{\sum_{j=0}^{r} x'_{ij} \hat{\beta}_{j}^{(q)}\right\}$.

When the estimates have converged we obtain the price relativities for the categorical rating variables by using the fact that

$$\gamma_j = e^{\beta_j}, \quad j = 0, 1, \ldots, r. \tag{4.60}$$

We obtain the natural cubic splines for the continuous rating variables by

$$s_1(x) = \sum_{j=1}^{m_1+2} \beta_{1j} B_{1j}(x), \tag{4.61}$$

$$s_2(x) = \sum_{j=1}^{m_2+2} \beta_{2j} B_{2j}(x), \tag{4.62}$$

and the price relativities for the continuous rating variables are given by
$e^{s_1(z_{11})}, \ldots, e^{s_1\left(z_{1m_1}\right)}$ and $e^{s_2(z_{21})}, \ldots, e^{s_2\left(z_{2m_2}\right)}$.

Chapter 5

Optimal choice of the smoothing parameter

As mentioned in Section 4.1.3, the smoothing parameter creates a trade-off between good fit to the data and low variability of the function f. A small value of λ would increase the weight put on data, letting the function f vary freely, whereas a large value of λ would decrease the weight put on data, forcing the integrated squared second derivative of the function f to be small (Figure 5.1).

In this chapter we will compare two different methods for finding the optimal smoothing parameter. While the method of cross validation is commonly used for this purpose, a more uncommon method, the L-curve method, is investigated for its performance in comparison to the method of cross validation.

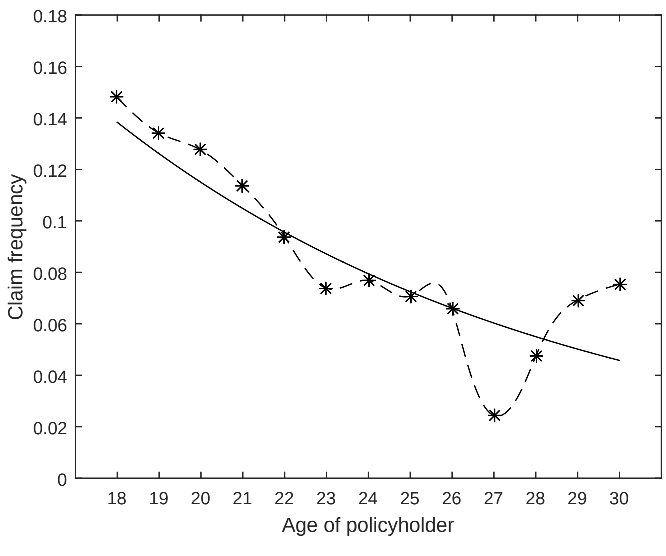

Figure 5.1: A small value of λ would increase the weight put on data, letting the function f vary freely (dashed curve), whereas a large value of λ would decrease the weight put on data, forcing the integrated squared second derivative of the function f to be small (solid curve).

5.1 Cross validation

Cross validation is a way of measuring the predictive performance of a statistical model. There are plenty of different variations of cross validation but due to its simple properties we will only be using the variation that goes under the name of leave-one-out cross validation [7]. When describing the method of cross validation we will only be considering one continuous rating variable. In the case of several continuous rating variables, we can just proceed and use the approach described below every time we update one of the continuous rating variables in the backfitting algorithm. As the calculations for the Poisson case and the gamma case are very similar we will only consider the case where the observations of the key ratio Y are Poisson distributed. For the Poisson distribution, the fitted natural cubic spline s minimizes the expression

$$\begin{aligned}\Delta(s(x)) = 2\sum_{k=1}^{m} \tilde{w}_k\Big(\tilde{y}_k \log \tilde{y}_k - \tilde{y}_k s\Big(z_k\Big) - \tilde{y}_k + e^{s(z_k)}\Big) \\ + \lambda \int_{z_1}^{z_m} \left(s''(x)\right)^2 dx.\end{aligned} \tag{5.1}$$

Now suppose we remove one particular z_k and the corresponding $\tilde{y}_k$ from the data. We can then, for any λ, calculate the minimizing natural cubic spline $s_k^\lambda(x)$ for this new data set. For a good value of λ, $s_k^\lambda(z_k)$ should be a good predictor of the deleted data point $\tilde{y}_k$. This should be true for any k. The cross validation score measures the overall ability of predicting all the removed data points (a smaller value of the cross validation score means better ability to predict the removed data points) and is defined as

$$C(\lambda) = 2\sum_{k=1}^{m} \tilde{w}_k\Big(\tilde{y}_k \log \tilde{y}_k - \tilde{y}_k s_k^\lambda(z_k) - \tilde{y}_k + e^{s_k^\lambda(z_k)}\Big). \tag{5.2}$$

The idea of cross validation is to choose λ as the value for which the cross validation score $C(\lambda)$ is minimized. We compute $C(\lambda)$ by finding all the minimizing splines $s_1^\lambda, \dots s_m^\lambda$. Since this can be very time consuming, approximative methods have been developed but these will not be used in this thesis [12, §5.5].

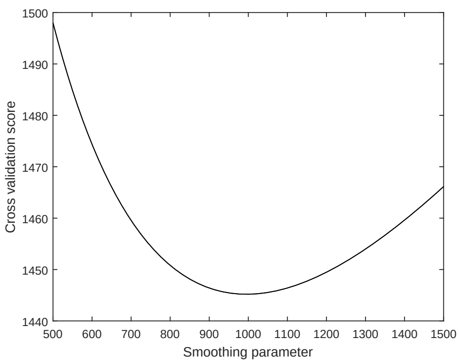

Figure 5.2: The cross validation score $C(\lambda)$ for different values of λ.

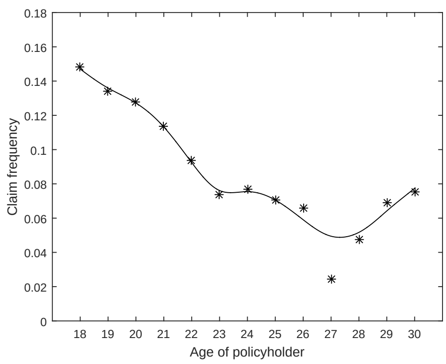

Figure 5.3: The natural cubic spline based on the λ that minimizes the cross validation score $C(\lambda)$.

5.2 The L-curve method

An L-curve is a convenient graphical tool for displaying the size of the regularization versus the size of the deviance, as the smoothing parameter varies. A standard L-curve is visualized in a log-log plot and has the shape of an "L", but for insurance (noise-free) data it has been shown that the L-curve always becomes concave [5].

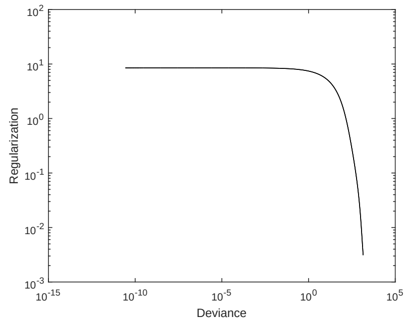

Figure 5.4: The L-curve becomes concave for insurance data.

The idea of the L-curve method is to find the point on the "corner" of the L-curve and to choose the underlying smoothing parameter. The reason behind this choice is that the corner separates the flat and vertical parts of the curve where the solution is dominated by the regularization and the deviance, respectively. For standard L-curves with the shape of an "L", this makes perfect sense, but for concave L-curves it is not clear whether the corner represents a good value of the smoothing parameter or not. This will be evaluated in Chapter 6. There are several different ways of defining the corner but we will choose the method proposed by Hansen and O'Leary [5]. We define $h_1(\lambda) = \log(\Delta - \lambda R)$ and $h_2(\lambda) = \log((\Delta - D)/\lambda)$ and we note that both h_1 and h_2 are twice differentiable with respect to λ. The corner of the L-curve is then found by minimizing the curvature $\kappa(\lambda)$, where

$$\kappa(\lambda) = \frac{h_1' h_2'' - h_1'' h_2'}{\left((h_1')^2 + (h_2')^2\right)^{3/2}}. \tag{5.3}$$

Any simple optimizing method will do for finding the minimal $\kappa(\lambda)$.

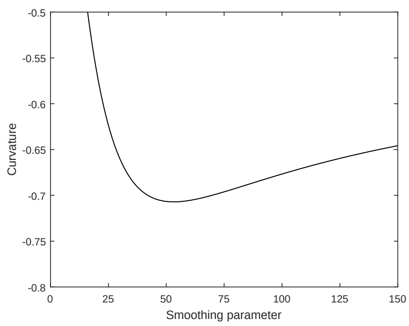

Figure 5.5: The curvature $\kappa(\lambda)$ for different values of λ.

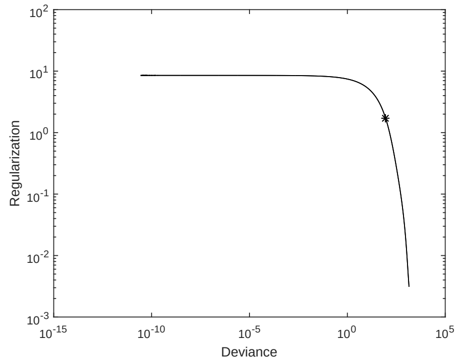

Figure 5.6: The point on the L-curve with the minimum curvature $\kappa(\lambda)$.

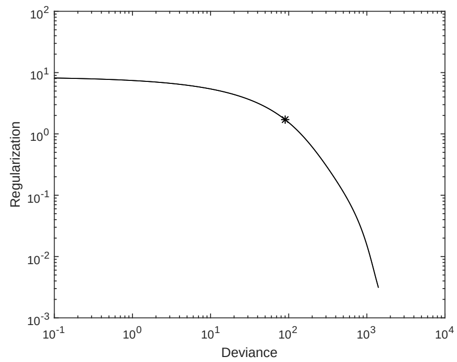

Figure 5.7: The point on the L-curve with the minimum curvature $\kappa(\lambda)$ (with axes range being of equal dimension).

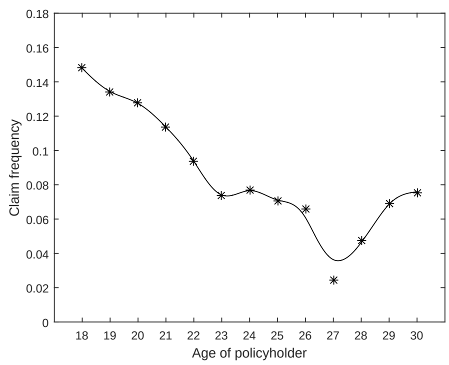

Figure 5.8: The natural cubic spline based on the λ that minimizes the curvature $\kappa(\lambda)$.

Chapter 6

Results

In this chapter we present some test results, comparing the method of cross validation with the L-curve method. The data used for this comparison is downloaded test data[1]. The test data has been modified in different ways to fully analyze and compare the different methods. The calculations have been carried out in Matlab R2015a.

6.1 Numerical computations

We recall the system of linear equations that needs to be solved in the GAM to obtain the regression parameters $\boldsymbol{\beta}$ (Equation 4.32). In the method of cross validation, the equation system needs to be solved $p \times m$ times, where p is the total number of different values that is tested for the smoothing parameter and m is the number of possible values for the rating variable. In the L-curve method the equation system only needs to be solved p times. In Table 6.1 we present a summary of running times for different choices of p. The number of possible values for the rating variable is set to $m = 12$.

	Cross validation	The L-curve method
$p = 1$	0.292	0.241
$p = 10$	0.330	0.245
$p = 100$	0.647	0.267
$p = 1\,000$	3.81	0.490
$p = 10\,000$	35.7	2.85
$p = 100\,000$	333	23.5

Table 6.1: Running times (in seconds) for different choices of p.

Note that we have only considered one continuous rating variable in the computations above. When considering several continuous rating variables these running times adds up, every time we update one of the continuous rating variables in the backfitting algorithm.

[1] http://staff.math.su.se/esbj/GLMbook/

6.2 Unmodified test data

In our first graphical example we are going to use the unmodified test data. We test $p = 10\,000$ different values of the smoothing parameter, both in the method of cross validation and in the L-curve method. The natural cubic splines based on the suggested smoothing parameters are shown in Figure 6.1. The program for the method of cross validation takes about 36 seconds to complete whereas the program for the L-curve method takes about 2.9 seconds to complete. See Section 6.1 for a more comprehensive comparison on running times for the different methods.

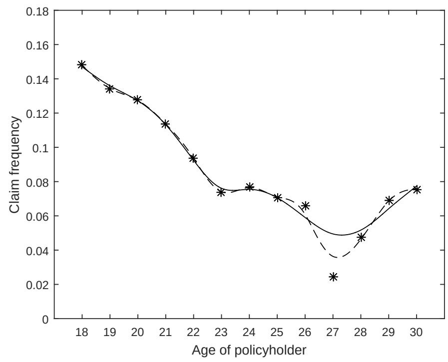

Figure 6.1: The natural cubic splines based on the smoothing parameters given by the method of cross validation (solid curve) and the L-curve method (dashed curve).

6.3 Modified test data

Now we are going to modify our original test data in different ways to be able to fully analyze and compare the method of cross validation (solid curve) and the L-curve method (dashed curve). We are going to modify data points with large duration (w_i in Section 3.3) and with small duration both upwards and downwards to be able to evaluate how the different methods react on data series with different properties. See Figures 6.2 - 6.7 below.

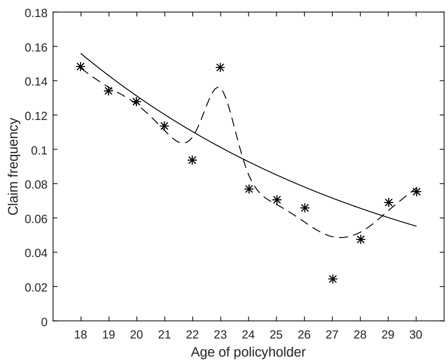

Figure 6.2: The claim frequency for an internal data point with large duration is multiplied by 2.

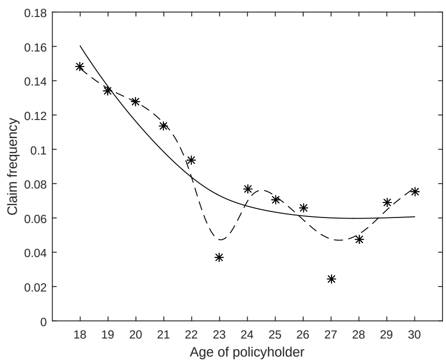

Figure 6.3: The claim frequency for an internal data point with large duration is divided by 2.

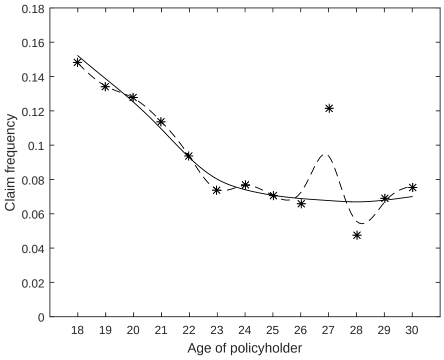

Figure 6.4: The claim frequency for an internal data point with small duration is multiplied by 5.

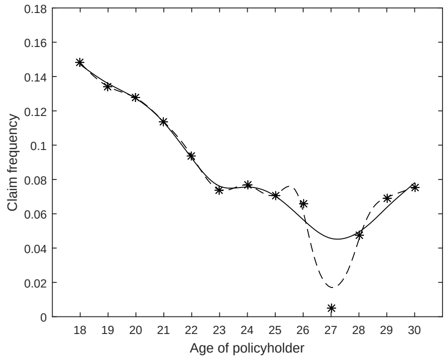

Figure 6.5: The claim frequency for an internal data point with small duration is divided by 5.

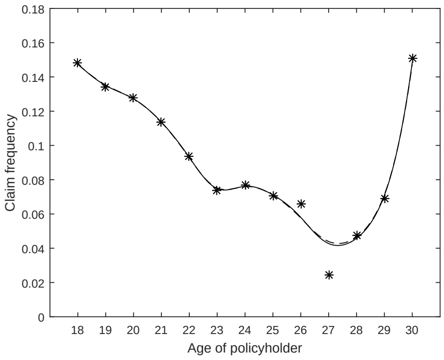

Figure 6.6: The claim frequency for a boundary data point is multiplied by 2.

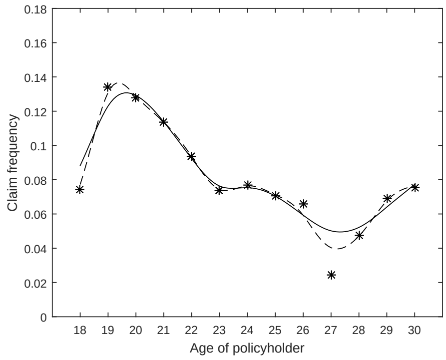

Figure 6.7: The claim frequency for a boundary data point is divided by 2.

We also present a summary of the optimal smoothing parameters, both for the method of cross validation and for the L-curve method. The optimal smoothing parameters for the examples in Figures 6.1 - 6.7 are shown in Table 6.2.

	Cross validation	The L-curve method
Figure 6.1	995	53.2
Figure 6.2	∞	999
Figure 6.3	196 000	675
Figure 6.4	16 800	78.0
Figure 6.5	1 090	13.6
Figure 6.6	290	43.7
Figure 6.7	1 190	123

Table 6.2: The optimal smoothing parameters for the examples in Figures 6.1 - 6.7.

Chapter 7

Summary and conclusion

In this thesis, we have studied two different ways of finding the optimal smoothing parameter within the GAM framework. The method of cross validation performed well in most situations, but sometimes encountered problems, for instance when no local minima could be found (see Figure 6.2). The method is also very computation heavy, especially if the number of possible values for the rating variable is large. The L-curve method is significantly faster than the method of cross validation but after visual inspection we can conclude that it suffers from heavy under-smoothing (see Figures 6.4 - 6.5 where the natural cubic spline adapts too much to internal data points with small duration). This is a subjective assessment, but it is based on years of experience from the insurance business. Although the L-curve method is significantly faster than the method of cross validation, the heavy under-smoothing rejects the L-curve method from being a suitable method for estimating the optimal smoothing parameter.

If time had permitted, we would have liked to compare the method of leave-one-out cross validation with other methods within the cross validation family, but unfortunately, that comparison is beyond the scope of this paper.

Bibliography

[1] Beard, R. E., Pentikäinen, T., and Pesonen, E. *Risk Theory: The Stochastic Basis of Insurance.* Chapman & Hall, 1984.

[2] Brockman, M. J., and Wright, T. S. Statistical motor rating: making effective use of your data. *Journal of the Institute of Actuaries 119* (1992), 457–543.

[3] Feller, W. *An Introduction to Probability Theory and Its Applications.* Wiley, 1968.

[4] Grimmett, G., and Welsh, D. *Probability: An Introduction.* Oxford University Press, 1986.

[5] Hansen, P. C., and O'Leary, D. P. The use of the L-curve in the regularization of discrete ill-posed problems. *Journal of Scientific Computing 14*, 6 (1993), 1487–1503.

[6] Hastie, T., and Tibshirani, R. Generalized Additive Models. *Statistical Science 1*, 3 (1986), 297–318.

[7] Hastie, T., and Tibshirani, R. *Generalized Additive Models.* Chapman & Hall/CRC, 1990.

[8] Jorgensen, B. *The Theory of Dispersion Models.* Chapman & Hall, 1997.

[9] Kelley, C. T. *Solving Nonlinear Equations with Newton's Method.* Society for Industrial and Applied Mathematics, 2003.

[10] McCullagh, P., and Nelder, J. A. *Generalized Linear Models.* Chapman & Hall, 1989.

[11] Murphy, K. P., Brockman, M. J., and Lee, P. K. W. Using Generalized Linear Models to Build Dynamic Pricing Systems for Personal Lines Insurance. *Casualty Actuarial Society Forum* (2000), 107–140.

[12] Ohlsson, E., and Johansson, B. *Non-Life Insurance Pricing with Generalized Linear Models.* Springer-Verlag Berlin Heidelberg, 2010.

[13] Trenerry, C. F. *The Origin and Early History of Insurance Including The Contract of Bottomry.* The Lawbook Exchange, 2009.

[14] Wood, S. *Generalized Additive Models: An Introduction with R.* Chapman & Hall/CRC, 2006.