Mathematics of Non-life Insurance 2
lecture notes

Lucie Mazurová

December 2018

Contents

Chapter 1

Rate-making in a Multiplicative Tariff Structure

1.1 Introduction

Rate-making is a process of determining insurance premium in a heterogeneous portfolio that is divided into sufficiently homogeneous **risk classes**. The premium is usually expressed in a form of a premium rate, i.e. as premium per unit of **risk exposure**. The choice of an appropriate exposure base depends on the type of the insurance (examples are the value of the insured property, one car-year in automobile insurance, sum insured,...)

We assume that the risks are classified in the risk classes on the basis of a number of **tariff variables** - characteristics known for each risk and chosen so that the resulting segmentation provides us with satisfactory homogenity and size of risk classes.

Without loss of generality we explain the rate-making methods in the simple case with only two tariff variables. The portfolio can be divided into I classes according to the first variable and into J classes according to the second variable. Thus, we have $M = I\,J$ different risk classes, they are denoted by pairs (i, j), where $i \in \{1, \dots, I\}$, $j \in \{1, \dots, J\}$.

For each class we observe yearly aggregate losses (or alternatively, yearly numbers of claims). Let S_{ij} denote the total loss in class (i, j).

We make some assumption concerning the dependence of the distribution of losses on the values of tariff variables. In what follows, we work with **multiplicative model**

$$\mathrm{E}\, S_{ij} = \mu\, v_{ij}\, \varphi_{1i}\, \varphi_{2j},\ \ i = 1, \dots, I,\ \ j = 1, \dots, J. \tag{1.1}$$

Here v_{ij} stands for the total risk exposure in class (i, j) (it can be for example the number of policies) and it is assumed to be known. Our goal is to estimate parameters μ, φ_{1i}, $i = 1, \ldots, I$, φ_{2j}, $j = 1, \ldots, J$, from the observed data.

First, we introduce some simple rate-making procedures, later we explain the application of generalized linear models in the above described context.

1.2 Simple Methods

1.2.1 Method of Bailey and Simon

This method consists in estimating parameters μ, φ_{1i}, φ_{2i} by positive values minimizing the expression

$$\chi^2 = \sum_{i,j} \frac{\left(S_{ij} - \mu\, v_{ij}\, \varphi_{1i}\, \varphi_{2j}\right)^2}{\mu\, v_{ij}\, \varphi_{1i}\, \varphi_{2j}}. \tag{1.2}$$

Remark. (1.2) is minimized numerically. Reduction of the number of parameters is necessary. We can for example set

$$\mu = \widehat{\varphi_{11}} = 1.$$

A disadvantage of the Bailey-Simon method is a (systematic) positive bias of the estimates $\widehat{\mu}$, $\widehat{\varphi_{1i}}$, $\widehat{\varphi_{2j}}$, $i = 1, \ldots, I$, $j = 1, \ldots, J$, obtained by minimizing (1.2). It holds

$$\sum_{i,j} v_{ij}\, \widehat{\mu}\, \widehat{\varphi_{1i}}\, \widehat{\varphi_{2j}} \geq \sum_{i,j} S_{ij}. \tag{1.3}$$

Proof. To prove (1.3) we differentiate ith summand in (1.2) with respect to φ_{2j}. We obtain

$$\begin{aligned}\frac{\partial}{\partial \varphi_{2j}} \frac{\left(S_{ij} - v_{ij}\, \mu\, \varphi_{1i}\, \varphi_{2j}\right)^2}{v_{ij}\, \mu\, \varphi_{1i}\, \varphi_{2j}} &= \frac{\partial}{\partial \varphi_{2j}} \left[\frac{S_{ij}^2}{v_{ij}\, \mu\, \varphi_{1i} \varphi_{2j}} - 2\, S_{ij} + v_{ij}\, \mu\, \varphi_{1i}\, \varphi_{2j}\right] \\ &= -\frac{S_{ij}^2}{v_{ij}\, \mu\, \varphi_{1i}\, \varphi_{2j}^2} + v_{ij}\, \mu \varphi_{1i}.\end{aligned}$$

The estimates $\widehat{\mu}$, $\widehat{\varphi_{1i}}$ and $\widehat{\varphi_{2j}}$ satisfy

$$\frac{\partial \chi^2}{\partial \varphi_{2j}} = 0,$$

which means

$$\sum_i \frac{S_{ij}^2}{v_{ij}\,\widehat{\mu}\,\widehat{\varphi_{1i}}\,\widehat{\varphi_{2j}}^2} = \sum_i v_{ij}\,\widehat{\mu}\,\widehat{\varphi_{1i}}.$$

From here it follows

$$\widehat{\varphi_{2j}} = \left(\frac{\sum_i \frac{S_{ij}^2}{v_{ij}\,\widehat{\mu}\,\widehat{\varphi_{1i}}}}{\sum_i v_{ij}\,\widehat{\mu}\,\widehat{\varphi_{1i}}}\right)^{1/2}.$$

Finally,

$$\sum_i v_{ij}\,\widehat{\mu}\,\widehat{\varphi_{1i}}\,\widehat{\varphi_{2j}} = \left(\sum_i v_{ij}\,\widehat{\mu}\,\widehat{\varphi_{1i}}\right)^{1/2} \left(\sum_i \frac{S_{ij}^2}{v_{ij}\,\widehat{\mu}\,\widehat{\varphi_{1i}}}\right)^{1/2}.$$

Applying Schwarz inequality to the terms on the right-hand side provides the lower bound

$$\sum_i v_{ij}\,\widehat{\mu}\,\widehat{\varphi_{1i}}\,\widehat{\varphi_{2j}} \geq \sum_i (v_{ij}\,\widehat{\mu}\,\widehat{\varphi_{1i}})^{1/2} \left(\frac{S_{ij}^2}{v_{ij}\,\widehat{\mu}\,\widehat{\varphi_{1i}}}\right)^{1/2} = \sum_i S_{ij}.$$

(1.3) is proved by summing over j.

□

1.2.2 Method of Total Marginal Sums

This method (also called method of Bailey and Jung) derives the estimates of parameters μ, φ_{1i} and φ_{2j} as positive solutions of the following system of equations

$$\sum_{j=1}^{J} v_{ij}\,\mu\,\varphi_{1i}\,\varphi_{2j} = \sum_{j=1}^{J} S_{ij},\ i = 1, \ldots, I, \tag{1.4}$$

$$\sum_{i=1}^{I} v_{ij}\,\mu\,\varphi_{1i}\,\varphi_{2j} = \sum_{i=1}^{I} S_{ij},\ j = 1, \ldots, J. \tag{1.5}$$

Remark. It can be shown that when we suppose that S_{ij} are random variables with Poisson distribution with mean values given by (1.1), the solutions of the system (1.4)-(1.5) coincide with the estimates obtained by the method of maximum likelihood.

Chapter 2

Generalized Linear Models

2.1 Introduction

Generalized linear models (GLM) represent a class of statistical models generalizing classical linear regression model: the data may come from a broader family of probability distributions (not necessarily normal). A transformation of the expected value of observed data is expressed in the form of a linear function of regression parameters.

We start with basic definitions and relations, later we explain the application of GLM to parameter estimation in the multiplicative tariff structure.

Definition 2.1. *Probability density function f is a member of* ***exponential dispersion family*** *(EDF) if it has a form*

$$f(x, \theta, \varphi) = \exp\left\{\frac{x\,\theta - b(\theta)}{\varphi/w} + c(x, \varphi, w)\right\}, \tag{2.1}$$

In (2.1) $w > 0$ *is a given weight,* $\varphi > 0$ *is so called* ***dispersion parameter****,* θ *is a parameter satisfying* $\theta \in \Theta$*, where* Θ *is an open set in* $\mathbb{R}$*. The function* $b(\theta)$ *(**cumulant function**) is supposed to be twice differentiable function such that* $(b')^{-1}$ *exists.*

Remark. (2.1) can be a density of an absolutely continuous distribution, it can also be a probability function of a discrete distribution.

Examples of well known members of EDF are normal, gamma, exponential, Poisson distribution.

The following result concerning a distribution with density (2.1) can be used to express its expected value and variance.

Lemma 2.1. *Suppose that for given cumulant function b (2.1) defines densities with identical supports for all parameters $\theta \in \Theta$. Assume further that for any $\theta \in \Theta$ the moment generating function $M(r)$ of (2.1) exists at some neighborhood of zero. Then for all $\theta \in \Theta$ and for r sufficiently close to zero, it holds*

$$M(r) = \exp\left\{\frac{b\,(\theta + r\,\varphi/w) - b(\theta)}{\varphi/w}\right\}. \tag{2.2}$$

Proof. Choose $\theta \in \Theta$ and r such that $M(r)$ exists. Then

$$\begin{aligned}
M(r) &= \int e^{rx} \exp\left\{\frac{x\theta - b(\theta)}{\varphi/w} + c(x, \varphi, x)\right\} \mathrm{d}x \\
&= \int \exp\left\{\frac{x\,(\theta + r\,\varphi/w) - b(\theta)}{\varphi/w} + c(x, \varphi, w)\right\} \mathrm{d}x \\
&= \exp\left\{\frac{b(\theta + r\,\varphi/w) - b(\theta)}{\varphi/w}\right\} \\
&\quad \times \int \exp\left\{\frac{x\,(\theta + r\,\varphi/w) - b(\theta + r\,\varphi/w)}{\varphi/w} + c(x, \varphi, w)\right\} \mathrm{d}x.
\end{aligned}$$

For r sufficiently close to zero it holds $\theta + r\,\varphi/w \in \Theta$, thanks to the assumption that Θ is an open set. The integrand in the last integral is a density of the type (2.1) with the same support as the one we consider. Therefore, the last integral is equal to one.

□

Corollary. Under the same assumptions as in the previous lemma we obtain for the moments of the distribution with density (2.1)

$$\mathrm{E}\,X = b'(\theta), \tag{2.3}$$

$$\mathrm{Var}\,X = \frac{\varphi}{w} b\text{''}(\theta). \tag{2.4}$$

Proof. It holds

$$\mathrm{E}\,X = \frac{\mathrm{d}}{\mathrm{d}r} \log M_X(r)\,|_{r=0}$$

$$\mathrm{Var}\,X = \frac{\mathrm{d}^2}{\mathrm{d}r^2} \log M_X(r)\,|_{r=0}$$

Denote

$$\psi(r) = \frac{b(\theta + r\,\varphi/w) - b(\theta)}{\varphi/w}.$$

From the previous lemma it follows

$$\mathrm{E}\,X = \frac{\mathrm{d}\,\psi(r)}{\mathrm{d}r}\,|_{r=0} = b'\,(\theta + r\,\varphi/w)\,|_{r=0} = b'(\theta),$$

$$\mathrm{Var}\,X = \frac{\mathrm{d}^2\,\psi(r)}{\mathrm{d}r^2}\,|_{r=0} = \frac{\varphi}{w}\,b"\,(\theta + r\,\varphi/w)\,|_{r=0} = \frac{\varphi}{w}\,b"\,(\theta).$$

□

Denoting $\mu = \mathrm{E}\,X$ we can write

$$\mathrm{Var}\,X = \frac{\varphi}{w}\,v(\mu), \tag{2.5}$$

where

$$v(\mu) = b"\,\left((b')^{-1}(\mu)\right)$$

is so called **variance function**. It describes the relationship between the variance and the expected value of the EDF distribution.

Consider independent random variables $Y_1, \ldots, Y_n$ such that Y_i has a density

$$f_i(y) = \exp\left\{\frac{y\,\theta_i - b\,(\theta_i)}{\varphi/w_i} + c\,(y, \varphi, w_i)\right\}. \tag{2.6}$$

(We assume a distribution of Y_i from EDF with the same dispersion parameter φ for all the observations.) The assumption for the expected value of the vector $\mathbf{Y} = (Y_1, \ldots, Y_n)'$ from linear regression, $\mathrm{E}\,\mathbf{Y} = \mathbf{x}\,\boldsymbol{\beta}$, where $\mathbf{x}$ is a known matrix of type $n \times (k+1)$ and $\boldsymbol{\beta}$ is a vector of regression parameters, is now generalized to

$$g\,(\mathrm{E}\,Y_i) = \eta_i = \sum_{j=0}^{k} x_{ij}\,\beta_j, \;\; i = 1, \ldots, n, \tag{2.7}$$

where g is a strictly monotone and differentiable function, called **link function**.

When we choose link function in the form

$$g(z) = (b')^{-1}(z), \tag{2.8}$$

we obtain (denoting $\mu_i = \mathrm{E}\,Y_i$)

$$\eta_i = g\,(\mu_i) = (b')^{-1}\,(\mu_i) = (b')^{-1}\,(b'\,(\theta_i)) = \theta_i.$$

(2.8) is called **canonical link function**.

Using the canonical link can make the estimation computationally easier, but it is not always the link that suits best to the modelled situation. Later we will see that in the multiplicative structure the logarithmic link $g(y) = \log y$ is the natural choice.

2.2 Estimation of Parameters in GLM

We look for the estimates of $\beta_0, \dots, \beta_k$ based on observations

$$Y_1 = y_1, \dots, Y_n = y_n. \tag{2.9}$$

For the sake of maximum likelihood estimation we derive the logarithmic likelihood function

$$l(\theta_1, \dots, \theta_n, \varphi, y_1, \dots, y_n) = \frac{1}{\varphi} \sum_{i=1}^{n} w_i (y_i \theta_i - b(\theta_i)) + \sum_{i=1}^{n} c(y_i, \varphi, w_i). \tag{2.10}$$

We aim at maximizing (2.10) with respect to unknown parameters $\beta_0, \dots, \beta_k$.

When differentiating (2.10) with respect to β_j we use the following relations:

$$\mu_i = b'(\theta_i), \ \eta_i = g(\mu_i) \tag{2.11}$$

$$\theta_i = (b')^{-1}(\mu_i), \ \mu_i = g^{-1}(\eta_i) \tag{2.12}$$

$$\eta_i = \sum_{j=0}^{k} x_{ij} \beta_j. \tag{2.13}$$

According to the rule for differentiating inverse function,

$$\left(f^{-1}\right)'(y) = \frac{1}{f'(f^{-1}(y))},$$

we obtain

$$\frac{\partial \theta_i}{\partial \mu_i} = \frac{1}{\frac{\partial \mu_i}{\partial \theta_i}((b')^{-1}(\mu_i))} = \frac{1}{b"((b')^{-1}(\mu_i))} = \frac{1}{v(\mu_i)},$$

where $v(\mu)$ is the variance function introduced in (2.5). Further,

$$\frac{\partial \mu_i}{\partial \eta_i} = \frac{1}{\frac{\partial \eta_i}{\partial \mu_i}(g^{-1}(\eta_i))} = \frac{1}{g'(\mu_i)},$$

$$\frac{\partial \eta_i}{\partial \beta_j} = x_{ij}.$$

Finally,

$$\begin{aligned}\frac{\partial l}{\partial \beta_j} &= \sum_{i=1}^{n} \frac{\partial l}{\partial \theta_i} \frac{\partial \theta_i}{\partial \beta_j} \\ &= \frac{1}{\varphi} \sum_{1=1}^{n} w_i \left(y_i - b'\left(\theta_i\right)\right) \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} \\ &= \frac{1}{\varphi} \sum_{i=1}^{n} w_i \frac{y_i - \mu_i}{v\left(\mu_i\right)\, g'\left(\mu_i\right)} x_{ij}.\end{aligned}$$

Maximum likelihood estimators (MLE) of $\beta_0, \ldots, \beta_k$ are obtained from the solution of

$$\sum_{i=1}^{n} w_i \frac{y_i - \mu_i}{v\left(\mu_i\right)\, g'\left(\mu_i\right)} x_{ij} = 0,\ j = 0, \ldots, k, \tag{2.14}$$

where

$$\mu_i = g^{-1} \left(\sum_{j=0}^{k} x_{ij}\, \beta_j \right).$$

Remark. When the number of observations is equal to the number of estimated parameters, $n = k + 1$, we speak about a **saturated model**. In such a (theoretical) case we could find the solution of the system (2.14) by setting

$$y_i = \mu_i,\ i = 1, \ldots, k + 1.$$

In a general case, the system (2.14) can be rewritten in a matrix form

$$\mathbf{x}'\mathbf{W}\mathbf{y} = \mathbf{x}'\mathbf{W}\boldsymbol{\mu}, \tag{2.15}$$

where

$$\begin{aligned}\mathbf{x} &= \left(x_{ij}\right)_{i=1,\ldots,n, j=0,\ldots,k}, \\ \mathbf{y} &= \left(y_1, \ldots, y_n\right)', \\ \boldsymbol{\mu} &= \left(\mu_1, \ldots, \mu_n\right)',\end{aligned}$$

and $\mathbf{W}$ is a diagonal matrix with diagonal elements

$$w_{ii} = \frac{w_i}{v\left(\mu_i\right)\, g'\left(\mu_i\right)},\ i = 1 \ldots, n.$$

Example. In classical linear regression we assume normally distributed observations with mean values μ_i and the same variance σ^2, the density of Y_i is therefore

$$
\begin{aligned}
f(y) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(y-\mu_i)^2}{2\sigma^2}\right\} \\
&= \exp\left\{\frac{y\,\mu_i - \mu_i^2/2}{\sigma^2} - \frac{y^2}{2\sigma^2} - \log\left(\sqrt{2\pi}\sigma\right)\right\}.
\end{aligned}
$$

Comparing this with the general form of EDF density (2.1) we conclude

$$
\begin{aligned}
&\theta = \mu,\ b(\theta) = \frac{\theta^2}{2},\ b'(\theta) = \theta,\ b"(\theta) = 1, \\
&w = 1,\ \varphi = \sigma^2, \\
&v(\mu) = \frac{\varphi}{w} = \sigma^2.
\end{aligned}
$$

The link function now is identical function, $g(y) = y$.

It holds $\mathbf{W} = \sigma^{-2}\,\mathbf{I}$, where $\mathbf{I}$ is identity matrix and (2.15) is rewritten as

$$
\mathbf{x}'\mathbf{y} = \mathbf{x}'\mathbf{x}\,\boldsymbol{\beta}.
$$

In general, the solution of (2.15) is found by an iterative numerical algorithm. We briefly discuss two possible ways of finding the solution of (2.15).

Newton-Raphson algorithm

The iterative solution of an equation of the form $f(x) = 0$ is based on the consecutive computation given by

$$
x^{[n+1]} = x^{[n]} - f'\left(x^{[n]}\right)^{-1} f\left(x^{[n]}\right)
$$

with a starting point $x^{[0]}$.

For the generalization to the case of a system of equations for unknown $\beta_0, \ldots, \beta_k$,

$$
\begin{aligned}
f_0(\boldsymbol{\beta}) &= 0, \\
&\vdots \\
f_k(\boldsymbol{\beta}) &= 0,
\end{aligned}
$$

the derivative f' is substituted by the matrix $\left(\frac{\partial f_i}{\partial \beta_j}\right)_{i=0,\ldots,k,j=0,\ldots,k}$.

In our case, f_j is the first derivative of the logarithmic likelihood function with respect to β_j. Therefore, we need the matrix of second derivatives

$$\frac{\partial^2 l}{\partial\beta_r\,\partial\beta_j} = \frac{1}{\varphi}\sum_{i=1}^{n} w_i \frac{\partial}{\partial\mu_i}\left[\frac{y_i-\mu_i}{v\left(\mu_i\right)g'\left(\mu_i\right)}\right]\frac{\partial\mu_i}{\partial\eta_i}\frac{\partial\eta_i}{\partial\beta_r}\,x_{ij}$$
$$= \frac{1}{\varphi}\sum_{i=1}^{n} w_i\,\frac{\partial}{\partial\mu_i}\left[\frac{y_i-\mu_i}{v\left(\mu_i\right)g'\left(\mu_i\right)}\right]\frac{1}{g'\left(\mu_i\right)}\,x_{ir}\,x_{ij}.$$

We can write

$$\frac{\partial}{\partial\mu_i}\left[\frac{y_i-\mu_i}{v\left(\mu_i\right)\,g'\left(\mu_i\right)}\right]\frac{1}{g'\left(\mu_i\right)}$$
$$= \frac{1}{g'\left(\mu_i\right)}\left[-\frac{1}{v\left(\mu_i\right)\,g'\left(\mu_i\right)} - \left(y_i-\mu_i\right)\frac{v\left(\mu_i\right)\,g"\left(\mu_i\right)+v'\left(\mu_i\right)\,g'\left(\mu_i\right)}{\left[v\left(\mu_i\right)\,g'\left(\mu_i\right)\right]^2}\right]$$
$$= -\frac{1}{v\left(\mu_i\right)\left[g'\left(\mu_i\right)\right]^2}\left[1+\left(y_i-\mu_i\right)\frac{v\left(\mu_i\right)\,g"\left(\mu_i\right)+v'\left(\mu_i\right)\,g'\left(\mu_i\right)}{v\left(\mu_i\right)\,g'\left(\mu_i\right)}\right]. \qquad (2.16)$$

Denoting (2.16) as $-a_i$, we write

$$\frac{\partial^2 l}{\partial\beta_r\,\partial\beta_j} = -\sum_{i=1}^{n} x_{ij}\,a_i\,x_{ir}.$$

For the matrix of second derivatives of l,

$$\mathbf{H} = \left(\frac{\partial^2 l}{\partial\beta_r\,\partial\beta_j}\right)_{r,j=0,\ldots,k},$$

it holds

$$\mathbf{H} = -\mathbf{x}'\,\mathbf{A}\,\mathbf{x},$$

Where $\mathbf{A}$ is a diagonal matrix with diagonal elements $a_i,\ i=0,\ldots,k$.

The Newton-Raphson algorithm is now described by

$$\boldsymbol{\beta}^{[n+1]} = \boldsymbol{\beta}^{[n]} - \left(\mathbf{H}^{-1}\right)^{[n]}\left(\mathbf{x}'\,\mathbf{W}^{[n]}\,\mathbf{y} - \mathbf{x}'\,\mathbf{W}^{[n]}\,\boldsymbol{\mu}^{[n]}\right),$$

where $\mathbf{H}$, $\mathbf{W}$ and $\boldsymbol{\mu}$ are recalculated at each step.

Fisher method of scores

This algorithm is based on the substitution of the matrix $\mathbf{H}$ by $-\mathbf{J}$, where $\mathbf{J}$ is Fisher information matrix defined by

$$\mathbf{J} = -\,\mathrm{E}\,\mathbf{H},$$

while taking the random variables $\{Y_i\}$ instead of their realizations $\{y_i\}$ in (2.16).

Instead of a_i introduced in (2.16), we use

$$\begin{aligned} d_i &= \mathrm{E}\, \frac{1}{v(\mu_i)\left[g'(\mu_i)\right]^2} \left[1 + (Y_i - \mu_i)\, \frac{v(\mu_i)\; g"(\mu_i) + v'(\mu_i)\; g'(\mu_i)}{v(\mu_i)\; g'(\mu_i)}\right] \\ &= \frac{1}{v(\mu_i)\left[g'(\mu_i)\right]^2}. \end{aligned}$$

The iteration step is now given by

$$\boldsymbol{\beta}^{[n+1]} = \boldsymbol{\beta}^{[n]} + \left(\mathbf{J}^{-1}\right)^{[n]} \left(\mathbf{x}'\, \mathbf{W}^{[n]}\, \mathbf{y} - \mathbf{x}'\, \mathbf{W}^{[n]}\, \boldsymbol{\mu}^{[n]}\right).$$

Remark. With canonical link, $g(\mu) = (b')^{-1}(\mu)$, we obtain

$$g'(\mu) = \frac{1}{b"\left((b')^{-1}(\mu)\right)} = \frac{1}{v(\mu)}$$

and also

$$v(\mu)\, g"(\mu) = -\frac{v'(\mu)}{v(\mu)} = -g'(\mu)\, v'(\mu).$$

Then $a_i = d_i$ and both Newton-Raphson algorithm and Fisher method of scores are identical.

2.3 GLM in Multiplicative Tariff Structure

We explain how GLM is applied to loss data in a portfolio classified by two tariff variables.

We recall formula (1.1) and we introduce the ratio

$$Y_{ij} = \frac{S_{ij}}{v_{ij}}.$$

We consider two possible interpretations of variables S_{ij} and exposures v_{ij}:

1. Let S_{ij} be the total number of claims in class (i, j). When v_{ij} stands for the number of policies in class (i, j), Y_{ij} represents claims frequency (average number of claims for one policy) of class (i, j).

2. In case S_{ij} is the aggregate loss in class (i, j), we treat v_{ij} as the total number of losses (not random in this situation). Then Y_{ij} is interpreted as the average loss per claim in class (i, j).

In what follows, we describe the stochastic models suitable for both above mentioned settings. In both cases we start with renumbering the classes so as to obtain observations in the form of a vector

$$\mathbf{Y} = (Y_1, \ldots, Y_M)' = (Y_{11}, \ldots, Y_{1J}, Y_{21}, \ldots, Y_{IJ}),$$

where $Y_m = Y_{ij}$ for $m = (i-1)\,J + j$ and $M = I\,J$.

We model the variables Y_{ij} in both situations 1. and 2. by probability distributions that are members of EDF family.

For the multiplicative model (1.1) the logarithmic link function $g(y) = \log y$ is an appropriate choice. Indeed,

$$g\left(\mathrm{E}\, Y_m\right) = g\left(\mathrm{E}\, Y_{ij}\right) = \log \mu + \log \varphi_{1i} + \log \varphi_{2j},$$

which can be rewritten as

$$g\left(\mathrm{E}\, Y_{ij}\right) = \beta_0 + \beta_{1i} + \beta_{2j}. \tag{2.17}$$

As the vector of regression parameters we choose

$$\boldsymbol{\beta} = (\beta_0, \beta_{12}, \ldots, \beta_{1I}, \beta_{22}, \ldots, \beta_{2J})' = (\beta_0, \beta_1, \ldots, \beta_k)',$$

where $k = I + J - 1$.

Using the notation

$$g\left(\mathrm{E}\,\mathbf{Y}\right) = \left(g\left(\mathrm{E}\, Y_1\right), \ldots, g\left(\mathrm{E}\, Y_m\right)\right)'$$

we obtain from (2.17)

$$g\left(\mathrm{E}\,\mathbf{Y}\right) = \mathbf{Z}\,\boldsymbol{\beta},$$

where $\mathbf{Z}$ is a $M \times (I + J - 1)$ matrix of assumed full rank (that is why we set $\beta_{11} = \beta_{21} = 0$) consisting of elements equal to 0 or 1.

Example. For $M = 9$ classes $(1,1),\ (1,2), \ldots, (3,3)$ we have

$$\boldsymbol{\beta} = (\beta_0, \beta_{12}, \beta_{13}, \beta_{22}, \beta_{23}) = (\beta_0, \ldots, \beta_4)$$

and

$$\mathbf{Z} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{bmatrix}.$$

The expected value of Y_{ij} is estimated by

$$\mathrm{E}\, Y_{ij} = \mathrm{e}^{\widehat{\beta_0}}\, \mathrm{e}^{\widehat{\beta_{1i}}}\, \mathrm{e}^{\widehat{\beta_{2j}}},\ i = 1, \ldots, I, j = 1, \ldots, J,$$

where $\widehat{\beta_{11}} = \widehat{\beta_{21}} = 0$ and $\widehat{\beta_0}$, $\widehat{\beta_{12}}, \ldots,$ $\widehat{\beta_{2J}}$ are maximum likelihood estimates obtained as described above.

2.3.1 Poisson Model for Claim Frequency

We suppose that S_{ij} in (1.1) is the total number of claims observed in class (i, j) and v_{ij} is the number of policies in class (i, j). We assume that S_{ij} is Poisson distributed with

$$\mathrm{E}\, S_{ij} = \lambda_{ij}\, v_{ij}.$$

Random variable Y_{ij} than takes on values from the set

$$\frac{\mathbf{N}_0}{v_{ij}} = \left\{0, \frac{1}{v_{ij}}, \frac{2}{v_{ij}}, \ldots\right\}.$$

For $y \in \frac{\mathbf{N}_0}{v_{ij}}$ we derive the probability function of Y_{ij} (we omit the subscript ij in the sequel):

$$\begin{aligned} \mathrm{P}\,(Y = y) = \mathrm{P}\,(S = v\, y) &= \frac{(\lambda\, v)^{vy}}{(v\, y)!}\, \mathrm{e}^{-\lambda v} \\ &= \exp\left\{v\, y\, (\log\lambda + \log v) - \lambda\, v - \log\left[(v\, y)!\right]\right\} \\ &= \exp\left\{\frac{y\, \log\lambda - \lambda}{1/v} + v\, y \log v - \log\left[(v\, y)!\right]\right\}. \end{aligned} \tag{2.18}$$

Probability distribution (2.18) has the form of (2.1) with

$$\theta = \log\lambda,\ b(\theta) = \mathrm{e}^{\theta},\ w = v,\ \varphi = 1.$$

Note that $\log b'(\theta) = \theta$, logarithmic link in case of Poisson model is canonical.

After renumbering the classes we see that parameter θ_m is the m-th element of the vector $\mathbf{Z}\, \boldsymbol{\beta}$:

$$\theta_m = (\mathbf{Z}\, \boldsymbol{\beta})_m\, .$$

The logarithmic likelihood is

$$l\,(\boldsymbol{\beta}) = \sum_m \frac{y_m\, \theta_m - \mathrm{e}^{\theta_m}}{1/v_m} + \sum_m \left(v_m\, y_m \log v_m - \log\left[(v_m\, y_m)!\right]\right).$$

Differentiating it with respect to β_r we obtain

$$\frac{\partial}{\partial \beta_l} l\,(\boldsymbol{\beta}) = \sum_m \frac{y_m - \mathrm{e}^{\theta_m}}{1/v_m} \frac{\partial \theta_m}{\partial \beta_l}$$
$$= \sum_m \frac{y_m - \exp\left\{(\mathbf{Z}\,\boldsymbol{\beta})_m\right\}}{1/v_m} z_{ml},$$

where z_{ml} is the corresponding element of matrix $\mathbf{Z}$.

The system of likelihood equations now can be expressed as

$$\sum_m v_m\, y_m\, z_{ml} = \sum_m v_m \exp\left\{(\mathbf{Z}\,\boldsymbol{\beta})_m\right\} z_{ml},\ \ l = 0\ldots,k. \tag{2.19}$$

The matrix form of (2.19) is

$$\mathbf{Z}'\mathbf{V} \exp\left\{\mathbf{Z}\,\boldsymbol{\beta}\right\} = \mathbf{Z}'\mathbf{V}\,\mathbf{y}, \tag{2.20}$$

where $\mathbf{V}$ is a diagonal matrix with diagonal elements $v_m,\ m = 1,\ldots,M$ and we denote by $\exp\left\{\mathbf{Z}\,\boldsymbol{\beta}\right\}$ a vector with elements $\exp\left\{(\mathbf{Z}\,\boldsymbol{\beta})_m\right\},\ m = 1\ldots,M$.

2.3.2 Gamma Model for Claim Severity

Now we take S_{ij} in (1.1) as the aggregate loss in class (i,j), the number of claims is taken as non-random weight v_{ij}. Random variable Y_{ij} now represents the average claim size in class (i,j).

We derive the distribution of Y_{ij} in a given class, we omit the subscripts i,j. We start with the assumption that the size of an individual claim in the given class follows gamma distribution $\Gamma(\alpha,\beta)$ with the density

$$g(x) = \frac{\beta^\alpha}{\Gamma(\alpha)}\, x^{\alpha-1}\, \mathrm{e}^{-\beta\, x},\ \ x > 0, \alpha > 0, \beta > 0. \tag{2.21}$$

We further assume that individual claim sizes are independent. When the number of claims in the class is fixed, then the aggregate loss S has also gamma distribution, $S \sim \Gamma(v\,\alpha,\beta)$.

This is easily shown by considering the moment generating function. The m.g.f. of $\Gamma(\alpha,\beta)$ is

$$M_X(r) = \left(\frac{\beta}{r-\beta}\right)^\alpha,\ \ r < \beta.$$

For the sum S of v independent gamma distributed random variables we obtain

$$M_S(r) = M_X(r)^v = \left(\frac{\beta}{r-\beta}\right)^{\alpha\, v}.$$

The density of Y is derived from

$$\mathrm{P}(Y \le y) = \mathrm{P}(S \le v\,y)$$

as

$$\begin{aligned} f(y) = v\,h(v\,y) &= v\,\frac{\beta^{v\,\alpha}}{\Gamma(v\,\alpha)}\;\mathrm{e}^{-\beta\,v\,y}\,(v\,y)^{\alpha\,v-1} \\ &= \frac{(\beta\,v)^{v\,\alpha}}{\Gamma(v\,\alpha)}\;\mathrm{e}^{-\beta\,v\,y}\;y^{\alpha\,v-1}. \end{aligned} \tag{2.22}$$

Y is also gamma distributed, $Y \sim \Gamma(\alpha\,v, \beta\,v)$. We rewrite the density (2.22) in a form corresponding to EDF density (2.1).

$$\begin{aligned} f(y) &= \exp\left\{\alpha\,v\,\log\beta\,v - \log\Gamma(\alpha\,v) - \beta\,v\,y + (\alpha\,v - 1)\log y\right\} \\ &= \exp\left\{\alpha\,v\left(\log\beta \pm \log\alpha + \log v - \frac{\beta}{\alpha}\,y\right) - \log\Gamma(\alpha\,v) + (\alpha\,v-1)\log y\right\}. \end{aligned}$$

We introduce parameters

$$\mu = \frac{\alpha}{\beta},\;\; \varphi = \frac{1}{\alpha}$$

and we write

$$f(y) = \exp\left\{\frac{-y/\mu - \log\mu}{\varphi/v} + c(y, \varphi, v)\right\}.$$

In the notation of (2.1) we have

$$\theta = -\frac{1}{\mu},\;\; b(\theta) = -\log(-\theta),\;\; w = v.$$

After renumbering classes we have the density of the average claim in class m in the form

$$f(y) = \exp\left\{\frac{y\,\theta_m + \log\left(-\theta_m\right)}{\varphi/v_m} + c(y, \varphi, v_m)\right\}.$$

Remark. In accordance with (2.6) we assume that parameter α is the same for claim sizes in all classes. The differences in expected claim size are due to different values of parameter β in (2.21).

The logarithmic link function in this case is

$$g\left(\mathrm{E}\,Y_m\right) = \log\left(-\frac{1}{\theta_m}\right) = -\log\left(-\theta_m\right).$$

From

$$g\left(\mathrm{E}\, Y_m\right) = \left(\mathbf{Z}\,\boldsymbol{\beta}\right)_m$$

we obtain

$$\theta_m = -\exp\left(-\left(\mathbf{Z}\,\boldsymbol{\beta}\right)_m\right).$$

The logarithmic likelihood function is

$$\begin{aligned} l(\boldsymbol{\beta}) &= \sum_m \frac{\theta_m\, y_m + \log\left(-\theta_m\right)}{\varphi / v_m} + \sum_m \log c\left(y_m, \varphi, v_m\right) \\ &= \sum_m \frac{v_m}{\varphi}\left[y_m \exp\left(-\left(\mathbf{Z}\,\boldsymbol{\beta}\right)_m\right) - \left(\mathbf{Z}\,\boldsymbol{\beta}\right)_m\right] + \sum_m \log c\left(y_m, \varphi, v_m\right). \end{aligned}$$

Its derivative with respect to β_l is

$$\begin{aligned} \frac{\partial}{\partial \beta_l} l(\boldsymbol{\beta}) &= \sum_m \frac{v_m}{\varphi}\left[y_m \exp\left(-\left(\mathbf{Z}\,\boldsymbol{\beta}\right)_m\right) - 1\right]\, z_{ml} \\ &= \sum_m \frac{v_m}{\varphi}\left[-y_m\, \theta_m - 1\right]\, z_{ml}. \end{aligned}$$

The system of likelihood equations written in a matrix form is

$$\mathbf{Z}'\mathbf{V}_\theta \exp\left\{\mathbf{Z}\,\boldsymbol{\beta}\right\} = \mathbf{Z}'\mathbf{V}_\theta \mathbf{y}, \tag{2.23}$$

where $\exp\{\mathbf{Z}\,\boldsymbol{\beta}\}$ has the same meaning as in (2.20) and $\mathbf{V}_\theta$ is a diagonal matrix with diagonal elements $-\theta_m\, v_m/\varphi,\ m = 1\ldots, M$. The substantial difference between (2.20) and (2.23) is in the dependence of $\mathbf{V}_\theta$ on the estimated regression parameters (through θ).

2.3.3 Variable Reduction Analysis

We consider the general case of independent observations $\mathbf{Y} = (Y_1, \ldots, Y_M)'$ of random varibles from EDF, the density of Y_m being

$$f\left(y, \theta_m, \varphi, w_m\right) = \exp\left\{\frac{y\,\theta_m - b\left(\theta_m\right)}{\varphi / w_m} + c\left(y, \varphi, w_m\right)\right\},$$

i.e. we suppose that all observed variables have the same dispersion parameter φ and cumulant function b. When the logarithmic link function is used, we have the estimate of $\mathrm{E}\, Y_m$ in the form

$$\widehat{\mu_m} = \exp\left\{\left(\mathbf{Z}\,\widehat{\boldsymbol{\beta}}\right)_m\right\},$$

where $\mathbf{Z}$ is a matrix consisting of 0 and 1 and $\widehat{\boldsymbol{\beta}}$ is the MLE of the vector

$$\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_k)' .$$

In the previous subsections we obtained this model by renumbering classes in the multiplicative structure with two tariff variables. We can imagine that regression parameters correspond to different values of any number of tariff variables (l):

$$\boldsymbol{\beta} = (\beta_0, \beta_{12}, \ldots, \beta_{1I}, \beta_{22}, \ldots, \beta_{2J}, \ldots, \beta_{l1}, \ldots, \beta_{lK}) .$$

We deal with the problem of possible reduction of the number of parameters in $\boldsymbol{\beta}$ - the question is whether we can eliminate a tariff variable or reduce the number of possible levels for a tariff variable (by merging classes).

It holds

$$\widehat{\mu_m} = b'\left(\widehat{\theta_m}\right) .$$

We denote $h = (b')^{-1}$ and we write

$$\widehat{\theta_m} = h\left(\widehat{\mu_m}\right) .$$

The value of the log-likelihood at this estimate is

$$l\left(\widehat{\mu}\right) = \sum_m \frac{Y_m\, h\left(\widehat{\mu_m}\right) - b\left(h\left(\widehat{\mu_m}\right)\right)}{\varphi / w_m} + c\left(Y_m, \varphi, w_m\right) . \tag{2.24}$$

(2.24) maximizes the likelihood over all possible choices of $\boldsymbol{\beta}$ (under given design matrix $\mathbf{Z}$ and cumulant function b). Statistical tests for GLM often use a statistics called **deviance**. It is based on the ratio of likelihood functions in the considered model and in the saturated model.

Recall that in the saturated model there is the same number of observations as is the number of estimated regression parameters ($M = k + 1$) and the MLE of $\boldsymbol{\beta}$ is obtained by setting $\widehat{\mu_m} = Y_m$. The value of log-likelihood in this case is

$$l\left(\mathbf{Y}\right) = \sum_m \frac{Y_m\, h\left(Y_m\right) - b\left(h\left(Y_m\right)\right)}{\varphi / w_m} + c\left(Y_m, \varphi, w_m\right) .$$

The **scaled deviance** is defined by

$$\begin{aligned} D^*\left(\mathbf{Y}, \widehat{\boldsymbol{\mu}}\right) &= 2\,\left(l\left(\mathbf{Y}\right) - l\left(\widehat{\boldsymbol{\mu}}\right)\right) \\ &= \frac{2}{\varphi} \sum_m w_m\,\left[Y_m\, b\left(Y_m\right) - b\left(h\left(Y_m\right)\right) - Y_m\, h\left(\widehat{\mu_m}\right) + b\left(h\left(\widehat{\mu_m}\right)\right)\right] . \end{aligned} \tag{2.25}$$

The **deviance** statistics is defined by

$$D\left(\mathbf{Y}, \widehat{\boldsymbol{\mu}}\right) = \varphi \, D^{*}\left(\mathbf{Y}, \widehat{\boldsymbol{\mu}}\right). \tag{2.26}$$

It does not depend on φ.

Example. Poisson model. In the Poisson model we have $b(\theta) = \mathrm{e}^{\theta}$, so $h(\mu) = \log \mu$. Also, $\varphi = 1$, i.e. $D = D^{*}$. We have

$$\begin{aligned} D\left(\mathbf{Y}, \widehat{\boldsymbol{\mu}}\right) &= 2 \sum_{m} w_m \left[Y_m \log\left(Y_m\right) - Y_m\right] - 2 \sum_{m} w_m \left[Y_m \log\left(\widehat{\mu_m}\right) - \widehat{\mu_m}\right] \\ &= 2 \sum_{m} \left[w_m \, Y_m \log\left(\frac{Y_m}{\widehat{\mu_m}}\right) + w_m \left(\widehat{\mu_m} - Y_m\right) \right]. \end{aligned}$$

We want to see from the data $\mathbf{Y}$ wheter we can reduce the number of parameters in $\boldsymbol{\beta}$. For given $p < k + 1$ we formulate the null hypothesis:

$$H_0 : \; \beta_0 = \cdots = \beta_{p-1} = 0.$$

We calculate the deviance statistics $D\left(\mathbf{Y}, \widehat{\boldsymbol{\mu}}_{\text{full}}\right)$ in the full model with $\boldsymbol{\beta} \in \mathbb{R}^{k+1}$ and the deviance statistics $D\left(\mathbf{Y}, \widehat{\boldsymbol{\mu}}_{H_0}\right)$ under the null hypothesis H_0. We define the test statistics

$$F = \frac{D\left(\mathbf{Y}, \widehat{\boldsymbol{\mu}}_{H_0}\right) - D\left(\mathbf{Y}, \widehat{\boldsymbol{\mu}}_{\text{full}}\right)}{D\left(\mathbf{Y}, \widehat{\boldsymbol{\mu}}_{\text{full}}\right)} \, \frac{M - k - 1}{p}. \tag{2.27}$$

Under the null hypothesis, F has approximately F-distribution with p and $M - k - 1$ degrees of freedom.

Another test is based on the statistics

$$X^2 = D^{*}\left(\mathbf{Y}, \widehat{\boldsymbol{\mu}}_{H_0}\right) - D^{*}\left(\mathbf{Y}, \widehat{\boldsymbol{\mu}}_{\text{full}}\right). \tag{2.28}$$

Under the null hypothesis, the distribution of (2.28) is approximated by χ^2-distribution with p degrees of freedom.

For calculation of (2.28) an estimate of the dispersion parameter φ is needed. Note that when we assume the same value of patameter φ for all the observations, the estimates $\widehat{\theta_m}$, obtained by maximizing the log-likelihood, do not depend on φ. We use the estimates $\widehat{\theta_m}$ to compute **Pearson's residuals**

$$r_m^P = \frac{Y_m - b'\left(\widehat{\theta_m}\right)}{\sqrt{b"\left(\widehat{\theta_m}\right)/w_m}}. \tag{2.29}$$

The residuals (2.29) can be used to assess the accuracy of the GLM model. The sum of squared Pearson's residuals also serves for estimation of the dispersion parameter φ.

$$\widehat{\varphi}_P = \frac{1}{M-k-1} \sum_m w_m \frac{\left(Y_m - b'\left(\widehat{\theta_m}\right)\right)^2}{b"\left(\widehat{\theta_m}\right)}$$

is an approximately unbiased estimate of φ. It follows from the fact that

$$\widehat{\varphi}_P \frac{M-k-1}{\varphi} = \sum_m w_m \frac{\left(Y_m - \widehat{\mu_m}\right)^2}{\varphi\, v\left(\widehat{\mu_m}\right)}$$

has approximately χ^2-distribution with $M-k-1$ degrees of freedom with the expected value equal to $M-k-1$.

An alternative estimator of φ can be based on the deviance,

$$\widehat{\varphi}_D = \frac{\varphi\, D^*\left(\mathbf{Y}, \widehat{\boldsymbol{\mu}}\right)}{M-k-1} = \frac{D\left(\mathbf{Y}, \widehat{\boldsymbol{\mu}}\right)}{M-k-1}.$$

2.4 GLM in Loss Reserving

In this section we explain the main principles of the application of GLM in the field of estimating claims reserves in non-life insurance. At the beginning we recall the usual notation used in analysis of claim development.

2.4.1 Notation

We separate loss data according to the **accident year** (year of ocurrence) i and the **development year** j. We assume $i \in \{0, \ldots, I\}$, $j \in \{0, \ldots, J\}$, where I denotes the most recent accident year and J denotes the last development year. We further assume that $I = J$.

We denote the **incremental** data X_{ij}: X_{ij} stands for payments for claims in accident year i made in year $i+j$ (alternatively, X_{ij} may denote the number of reported claims with delay j or the change of the reported claims amount).

The **cumulative** amount C_{ij} for accident year i after j development years is then given by

$$C_{ij} = \sum_{k=0}^{j} X_{ik}. \tag{2.30}$$

The observations available at time I are represented by the set

$$D_I = \{X_{ij} : i + j \leq I,\ 0 \leq j \leq J\}. \tag{2.31}$$

They are usually represented in the form of a **development triangle**. Data in (2.31) arranged in rows according to accident years and in columns according to development years represent the incremental development triangle. The information available at time I is equivalently described by the cumulative development triangle formed by the data

$$\{C_{ij} : i + j \le I,\ 0 \le j \le J\}.$$

The values $X_{i,j}$ (or $C_{i,j}$) for $i + j > I$ need to be estimated or predicted.

We suppose that there is no further development after development year J, so C_{iJ} represents aggregate loss from all claims occurred in accident year i. Then the **outstanding loss liabilities** (claims reserve) for accident year i at time I are given by

$$R_i = \sum_{j=I-i+1}^{J} X_{ij} = C_{iJ} - C_{i,I-i}. \tag{2.32}$$

2.4.2 GLM Model for Incremental Claims

We assume that the data in the incremental development triangle (2.31) are independent random variables and that X_{ij} has a distribution from EDF as defined by (2.1). We suppose that all variables in (2.31) have a density with the same cumulant function b, unlike in the multiplicative model for rate-making we work here not only with specific canonical parameters θ_{ij} and weights w_{ij}, but we also have different values of dispersion parameters φ_{ij}. The density of X_{ij} has the general form

$$f\left(x, \theta_{ij}, \varphi_{ij}, w_{ij}\right) = \exp\left\{\frac{x\,\theta_{ij} - b\left(\theta_{ij}\right)}{\varphi_{ij}/w_{ij}} + c\left(x, \varphi_{ij}, w_{ij}\right)\right\}. \tag{2.33}$$

We know that

$$\mathrm{E}\,X_{ij} = b'\left(\theta_{ij}\right), \tag{2.34}$$

$$\mathrm{Var}\,X_{ij} = \frac{\varphi_{ij}}{w_{ij}\,b''\left(\theta_{ij}\right)} = \frac{\varphi_{ij}}{w_{ij}}\,v\left(\mathrm{E}\,X_{ij}\right), \tag{2.35}$$

where $v(\mu) = b''\,(b')^{-1})$ is the variance function introduced in (2.5).

We use the parametrization of a multiplicative structure

$$\mathrm{E}\,X_{ij} = \mu_i\,\gamma_j \tag{2.36}$$

with μ_i standing for the exposure of the accident year i (e.g. the expected number of claims, total number of policies,...) and γ_j defining an expected reporting/payment pattern over the different development periods j.

We use the logarithmic link $g(\mu) = \log \mu$, i.e.

$$\eta_{ij} = g\left(\mathrm{E}\, X_{ij}\right) = \log \mu_i + \log \gamma_j. \tag{2.37}$$

We set $\mu_0 = 1$ and denote

$$\begin{aligned}\boldsymbol{\beta} &= \left(\beta_1, \ldots, \beta_{I+J-1}\right)' \\ &= \left(\log \mu_1, \ldots, \log \mu_I, \log \gamma_0, \ldots, \log \gamma_J\right)'.\end{aligned}$$

It is seen from (2.37) that

$$\eta_{ij} = \mathbf{z}_{ij}\, \boldsymbol{\beta},$$

where $\mathbf{z}_{ij}$ is $1 \times (I + J + 1)$ vector with elements equal to 0 or 1: $\mathbf{z}_{0j}$ has one unit element on the position $I + j + 1$. For $i \neq 0$, the vector $\mathbf{z}_{ij}$ has two unit entries - on positions i and $I + j + 1$, the other elements are equal to zero.

We form from the incremental claims in the triangle (2.31) one column vector

$$\mathbf{X} = \left(X_{00}, X_{01}, \ldots, X_{I0}\right)'$$

and we write

$$g\left(\mathrm{E}\, \mathbf{X}\right) = \mathbf{Z}\, \boldsymbol{\beta},$$

where

$$g\left(\mathrm{E}\, \mathbf{X}\right) = \left(g\left(\mathrm{E}\, X_{00}\right), \ldots, g\left(\mathrm{E}\, X_{I0}\right)\right)'$$

and $\mathbf{Z}$ is a matrix with rows $\mathbf{z}_{00}, \ldots, \mathbf{z}_{I0}$.

We estimate the elements of $\boldsymbol{\beta}$ by maximum likelihood method and we obtain

$$\widehat{\boldsymbol{\beta}} = \left(\widehat{\log \mu_1}, \ldots, \widehat{\log \mu_I}, \widehat{\log \gamma_0}, \ldots, \widehat{\log \gamma_I}\right)'.$$

Then

$$\begin{aligned}\widehat{X_{ij}}^{\mathrm{GLM}} &= \mathrm{E}\,[\widehat{X_{ij} \mid D_I}] = \exp\left\{\widehat{\log \mu_i} + \widehat{\log \gamma_j}\right\} \\ &\overset{\text{denote}}{=} \widehat{\mu_i}\, \widehat{\gamma_j}, \text{ for } i + j > I.\end{aligned} \tag{2.38}$$

The estimate of the total claims in accident year i is derived from (2.38) as

$$\widehat{C_{iI}}^{\mathrm{GLM}} = \mathrm{E}\,[\widehat{C_{iI} \mid D_I}] = C_{i,I-i} + \sum_{j=I-i+1}^{I} \widehat{X_{ij}}^{\mathrm{GLM}} = C_{i,I-i} + \widehat{R_i}^{\mathrm{GLM}}. \tag{2.39}$$

2.4.3 Mean Squared Error of Prediction

For the estimated total loss from years $i = 1, \ldots, I$, we define conditional **mean square error of prediction** (MSEP) as

$$\mathrm{mse}_{|D_I}\left(\sum_{i=1}^{I} \widehat{C_{iI}} \mid D_I\right) = \mathrm{E}\left[\left(\sum_{i=1}^{I} \widehat{C_{iI}} - \sum_{i=1}^{I} C_{iI}\right)^2 \mid D_I\right]. \qquad (2.40)$$

Clearly,

$$\begin{aligned}\mathrm{mse}_{|D_I}\left(\sum_{i=1}^{I} \widehat{C_{iJ}} \mid D_I\right) &= \mathrm{mse}\left(\sum_{i=1}^{I} \widehat{R_i} \mid D_I\right) \\ &= \mathrm{E}\left[\left(\sum_{i+j>I} \widehat{X_{ij}} - \sum_{i+j>I} X_{ij}\right)^2 \mid D_I\right]. \qquad (2.41)\end{aligned}$$

(2.41) can be decomposed in two parts:

$$\begin{aligned}&\mathrm{mse}_{|D_I}\left(\sum_{i=1}^{I} \widehat{C_{iI}} \mid D_I\right) \\ &= \mathrm{E}\left[\left(\sum_{i+j>I}\left(\widehat{X_{ij}} - \mathrm{E}\left[X_{ij} \mid D_I\right] + \mathrm{E}\left[X_{ij} \mid D_I\right] - X_{ij}\right)\right)^2 \mid D_I\right] \\ &= \mathrm{Var}\left(\sum_{i+j>I} X_{ij} \mid D_I\right) + \left(\sum_{i+j>I}\left(\widehat{X_{ij}} - \mathrm{E}\left[X_{ij} \mid D_I\right]\right)\right)^2. \qquad (2.42)\end{aligned}$$

The first term on the right hand side of (2.42) represents the (conditional) process variance, the second term on the right hand side of (2.42) expresses the (conditional) estimation error.

We rewrite (2.42) using the assumption that X_{ij} and D_I are independent for $i + j > I$ and also the assumption of mutual independence of X_{ij}:

$$\begin{aligned}&\mathrm{mse}_{|D_I}\left(\sum_{i=1}^{I} \widehat{C_{iI}} \mid D_I\right) \\ &= \sum_{i+j>I} \mathrm{Var}\left(X_{ij}\right) + \left(\sum_{i+j>I}\left(\widehat{X_{ij}} - \mathrm{E}\, X_{ij}\right)\right)^2. \qquad (2.43)\end{aligned}$$

The first term on the right hand side of (2.43) is estimated using (2.35) for a given variance function.

We define unconditional MSEP by

$$\begin{aligned}\operatorname{mse}\left(\sum_{i=1}^{I}\widehat{C_{iI}}\right) &= \operatorname{E}\operatorname{mse}\left(\sum_{i=1}^{I}\widehat{C_{iI}} \mid D_I\right)\\ &= \sum_{i+j>I}\frac{\varphi_{ij}}{w_{ij}}\,v\left(\operatorname{E}X_{ij}\right) + \operatorname{E}\left(\sum_{i+j>I}\left(\widehat{X_{ij}} - \operatorname{E}X_{ij}\right)\right)^2 . \qquad (2.44)\end{aligned}$$

The last term in (2.44) is rewritten as

$$\sum_{i+j>I,m+n>I}\operatorname{E}\left[\left(\widehat{X_{ij}} - \operatorname{E}X_{ij}\right)\left(\widehat{X_{mn}} - \operatorname{E}X_{mn}\right)\right]. \qquad (2.45)$$

We approximate the summands in (2.45) by $\operatorname{Cov}\left(\widehat{X_{ij}}, \widehat{X_{mn}}\right)$. Note that

$$\operatorname{E}\widehat{X_{ij}} \neq \operatorname{E}X_{ij}$$

in general, but the bias is usually neglected. For the GLM estimates we have

$$\begin{aligned}\operatorname{Cov}\left(\widehat{X_{ij}}, \widehat{X_{mn}}\right) &= \operatorname{Cov}\left(\mathrm{e}^{\mathbf{z}_{ij}\widehat{\boldsymbol{\beta}}}, \mathrm{e}^{\mathbf{z}_{mn}\widehat{\boldsymbol{\beta}}}\right)\\ &= \mathrm{e}^{(\mathbf{z}_{ij}+\mathbf{z}_{mn})\boldsymbol{\beta}}\operatorname{Cov}\left(\mathrm{e}^{\mathbf{z}_{ij}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)}, \mathrm{e}^{\mathbf{z}_{mn}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)}\right).\end{aligned}$$

Further, we use

$$\mathrm{e}^{\mathbf{z}_{ij}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)} \approx 1 + \mathbf{z}_{ij}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right).$$

From here and from the fact that $\boldsymbol{\beta}$ is asymptotically unbiased we deduce

$$\operatorname{Cov}\left(\widehat{X_{ij}}, \widehat{X_{mn}}\right) \approx \operatorname{E}X_{ij}\ \operatorname{E}X_{mn}\,\mathbf{z}_{ij}\ \operatorname{Var}\left(\widehat{\boldsymbol{\beta}}\right)\ \mathbf{z}'_{mn}. \qquad (2.46)$$

In section 2.2 we mentioned the Fisher method of scores as a possible way how to calculate the estimate $\widehat{\boldsymbol{\beta}}$. It can be shown that the inverse of the Fisher information matrix is a reasonable estimator for the variance matrix $\operatorname{Var}\left(\widehat{\boldsymbol{\beta}}\right)$.

Recall the definition of the Fisher information matrix,

$$\mathbf{J}(\boldsymbol{\beta}) = -\operatorname{E}\mathbf{H}(\boldsymbol{\beta}),$$

where the matrix of second derivatives of the log-likelihood function $\mathbf{H}(\boldsymbol{\beta})$ now consists of the elements

$$\sum_{i+j\leq I}\frac{\partial^2}{\partial\beta_k\,\partial\beta_l}\log f\left(X_{ij}, \theta_{ij}, \varphi_{ij}, w_{ij}\right) \qquad (2.47)$$

for $k,l = 1,\dots,I+J+1$. f in (2.47) is the density (2.33). $\mathbf{H}(\boldsymbol{\beta})$ is a function of incremental claims observed up to time I. When we substitute the parameters in (2.47) by the estimates calculated from the data in the development triangle (2.31), we have an estimate of the variance matrix in (2.46) in the form

$$\widehat{\mathrm{Var}\left(\widehat{\boldsymbol{\beta}}\right)} = \mathbf{J}\left(\widehat{\boldsymbol{\beta}}\right)^{-1}.$$

To sum up, we can calculate the mean square error of the GLM prediction of ultimate claims as

$$\widehat{\mathrm{mse}}\left(\sum_{i=1}^{I} \widehat{C_{iJ}}^{\mathrm{GLM}}\right) = \sum_{i+j>I} \frac{\widehat{\varphi_{ij}}}{w_{ij}} v\left(\widehat{X_{ij}}\right) + \sum_{i+j>I,m+n>I} \widehat{X_{ij}}\,\widehat{X_{mn}}\,\mathbf{z}_{ij}\,\mathbf{J}\left(\widehat{\boldsymbol{\beta}}\right)^{-1}\mathbf{z}'_{mn}.$$

Chapter 3

Credibility Theory

Credibility models are used in situations where the data used for estimation (e.g. claim experience of a given tariff class) are not sufficient and we want to utilize also the information contained in the data from other tariff classes or any external information we have.

Modern credibility theory is based on the principles of bayesian statistics and its origins are connected with the works of Hans Bühlmann. First we explain the asic concept using a model for just one insurance policy. THen we present the most widely used Bülhman-Straub model. Finally, we mention the application of credibility approach in the area of rate-making in multiplicative tariff structure.

3.1 Bühlmann's Model

We consider one insurance policy with unknown and fixed risk parameter θ, during a period of t years. The yearly claim amounts are $X_1, \dots, X_t$. We make the following assumptions:

1. θ is a realization of a r.v. Θ with a density $u(\theta)$.

2. For given $\Theta = \theta$, the claims are conditionally independent and identically distributed, with known density $f(x, \theta)$.

We want to estimate $\mu(\Theta) = \mathrm{E}[X|\Theta]$. We look for an estimator based on $X_1, \dots, X_t$, solving minimization

$$\min_g \mathrm{E}\left(\mu(\Theta) - g\left(X_1, \dots, X_t\right)\right)^2. \tag{3.1}$$

In statistics (3.1) is called the **Bayes estimator** with respect to the quadratic loss function.

With given $u(\theta)$ and $f(x, \theta)$, (3.1) can be rewritten using Bayes theorem

$$\mathrm{E}\left[\mu(\theta) \mid X_1, \ldots, X_t\right] = \frac{\int \mu(\theta)\, f(x_1, \theta) \cdots f(x_t, \theta)\, u(\theta)\ \mathrm{d}\theta}{\int f(x_1, \theta) \cdots f(x_t, \theta)\, u(\theta)\ \mathrm{d}\theta}. \tag{3.2}$$

Linear credibility estimator is obtained by minimization of (3.1) under the restriction that g is a linear combination of $X_1, \ldots, X_t$. The minimization problem is then expressed by

$$\min_{c_o, \ldots, c_t} \mathrm{E}\left(\mu(\theta) - c_0 - \sum_{s=1}^{t} c_s\, X_s\right)^2. \tag{3.3}$$

The solution of (3.3) is found by differentiating with respect to the coefficients $c_0, \ldots, c_t$ and equating the derivatives to zero. The resulting estimate is written in the form

$$tM^a = (1 - z)\, m + z\, \bar{X}, \tag{3.4}$$

where

$$z = \frac{t\, a}{t\, a + s^2}$$

is **credibility factor**,

$$\begin{aligned} m &= \mathrm{E}\, \mu(\Theta) = \mathrm{E}\, X_s \\ a &= \mathrm{Var}\, \mu(\Theta) \\ s^2 &= \mathrm{E}\, \mathrm{Var}\, (X_s \mid \Theta) \end{aligned}$$

are **structural parameters**.

In the situation when the linear credibility estimator (3.4) coincides with the best estimator (3.2) we speak about **exact credibility**.

Assume that the distribution of $X_1, \ldots, X_t$ for given value θ has the density

$$f(x, \theta) = p(x)\ \mathrm{e}^{-\theta\, x}\, / q(\theta),\ x > 0,\ \theta > 0 \tag{3.5}$$

where $p(x)$ is a non-negative function. Note that (3.5) contains the same exponential term with the product of the argument x and the parameter θ, as in the EDF density (2.1). By assuming (3.5) we take the a distribution from EDF as the appropriate model for our data.

Assume further that the density of r.v. Θ is

$$u(\theta) = q(\theta)^{-t_0}\ \mathrm{e}^{-\theta\, x_0}\, / c\, (t_0, x_0)\,,\ \theta > 0, \tag{3.6}$$

for some positive constants t_0, x_0.

The distribution with density $u(\theta)$ is called **prior distribution**. It describes our idea about the distribution of the random risk parameter independently on the observations. **Posterior distribution** of Θ is the conditional distribution for given realizations of $X_1, \ldots, X_t$. We denote the density of the posterior distribution as $u(\theta|x_1, \ldots, x_t)$. From the Bayes theorem we know

$$u(\theta|x_1, \ldots, x_t) = \frac{f(x_1, \ldots, x_t|\theta)\, u(\theta)}{\int f(x_1, \theta) \ldots, f(x_t, \theta)\, u(\theta)\, \mathrm{d}\theta}. \tag{3.7}$$

After inserting (3.5) and (3.6) into (3.7) we see that the posterior distribution is of the same type as the prior distribution when we replace t_0 with $t_0 + t$ and x_0 with $x_0 + x_1 + \cdots x_t$. The system of densities obtained from (3.6) for different values of parameters t_0 and x_0 is referred to as **conjugate prior** to the system of densities (3.5). Choosing conjugate prior distribution we get the posterior distribution of the unknown parameter that is of the same type as the prior distribution.

We use this property to show that with the densities (3.5) and (3.6) we arrive at the exact credibility.

Proof. From (3.5) it is seen

$$q(\theta) = \int_0^\infty p(x)\ \mathrm{e}^{-\theta\, x}\, \mathrm{d}x,$$

and by differentiating

$$q'(\theta) = -\int_0^\infty x\, p(x)\ \mathrm{e}^{-\theta\, x}\, \mathrm{d}x = -q(\theta)\, \mu(\theta).$$

From (3.6) we derive

$$u'(\theta) = (t_0\, \mu(\theta) - x_0)\ u(\theta).$$

We assume $u(0) = u(\infty) = 0$, so we can write

$$\int_0^\infty u'(\theta)\, \mathrm{d}\theta = 0,$$

which together with the previous result gives

$$m = \int_0^\infty \mu(\theta)\, u(\theta)\, \mathrm{d}\theta = \frac{x_0}{t_0}. \tag{3.8}$$

We get the expression for the expected value of the prior distribution of Θ. To express the conditional expectation (3.2) it suffices to replace in (3.8) the parameters of the prior distribution with the corresponding parameters of the posterior distribution. We see that the Bayes estimator (3.2) is a linear function of the observed variables.

$$\mathrm{E}\left[\mu(\theta) \mid X_1, \ldots, X_t\right] = \frac{x_0 + X_1 + \cdots + X_t}{t_0 + t}. \tag{3.9}$$

(3.9) is easily rearranged to

$$\mathrm{E}\left[\mu(\theta) \mid X_1, \ldots, X_t\right] = (1 - z)\, m + z\, \bar{X},$$

where

$$z = \frac{t}{t + t_0}.$$

□

3.2 The Bülhman-Straub Model

We consider a portfolio consisting of J groups of policies for which we observe loss variables in time periods $1, \ldots, t$. We denote by X_{js} the observation from group j in period s. Moreover, for each group we know the weight w_{js} reflecting the size of risk exposure for given group and given time period. The distribution of the loss variable in group j depends on an unknown parameter Θ_j which is again considered as a random variable.

The **Bühlmann-Straub model** works wih the following assumptions:

1. The groups, i.e. the random vectors

$$(X_{j1}, \ldots, X_{jt}, \Theta_j)', \ j = 1, \ldots, J,$$

   are independent.

2. $\Theta_1, \ldots, \Theta_J$ are identically distributed.

3. For any j, conditional on Θ_j, the variables $X_{j1}, \ldots, X_{jt}$ are uncorrelated, i.e.

$$\mathrm{Cov}\left(X_{js}, X_{jr} \mid \Theta_j\right) = 0 \text{ for } s \neq r.$$

We specify only the dependence of the conditional expected value and the conditional variance of X_{js} on Θ_j. We assume for all j and s

$$\mathrm{E}\left[X_{js} \mid \Theta_j\right] = \mu\left(\Theta_j\right), \tag{3.10}$$

$$\mathrm{Var}\left[X_{js} \mid \Theta_j\right] = \frac{\sigma^2\left(\Theta_j\right)}{w_{js}}. \tag{3.11}$$

Similarly as in the previous section we denote

$$m = \mathrm{E}\,\mu\left(\Theta_j\right) = \mathrm{E}\,X_{js}, \tag{3.12}$$

$$s^2 = \mathrm{E}\,\sigma^2\left(\Theta_j\right), \tag{3.13}$$

$$a = \mathrm{Var}\,\mu\left(\Theta_j\right). \tag{3.14}$$

For given group j we seek an estimator of $\mu(\Theta_j)$ based on the data from all groups. We want the estimate to be a non-homogeneous linear combination

$$\widehat{\mu\left(\Theta_j\right)} = c_0 + \sum_l \sum_r c_{lr}\,X_{lr} \tag{3.15}$$

minimizing

$$\mathrm{E}\left(\mu\left(\Theta_j\right) - c_0 - \sum_l \sum_r c_{lr}\,X_{lr}\right)^2.$$

The solution of the minimization problem is

$$\widehat{\mu\left(\Theta_j\right)} = z_j\,\bar{X}_{j.} + \left(1 - z_j\right) m, \tag{3.16}$$

where

$$z_j = \frac{w_{j.}}{w_{j.} + s^2/a},\ w_{j.} = \sum_{s=1}^{t} w_{js},\ \bar{X}_{j.} = \frac{\sum_s w_{js}\,X_{js}}{w_{j.}}.$$

Proof. To show the optimality of (3.16), we take a non-homogeneous linear combination of the form (3.15) with arbitrary coefficients. We denote it $h(\mathbf{X})$. We write

$$\begin{aligned}
\mathrm{E}\left[h(\mathbf{X}) - \mu\left(\Theta_j\right)\right]^2 &= \mathrm{E}\left[h(\mathbf{X}) - \widehat{\mu\left(\Theta_j\right)} + \widehat{\mu\left(\Theta_j\right)} - \mu\left(\Theta_j\right)\right]^2 \\
&= \mathrm{Var}\left[h(\mathbf{X}) - \widehat{\mu\left(\Theta_j\right)} + \widehat{\mu\left(\Theta_j\right)} - \mu\left(\Theta_j\right)\right] \\
&+ \left(\mathrm{E}\left[h(\mathbf{X}) - \widehat{\mu\left(\Theta_j\right)} + \widehat{\mu\left(\Theta_j\right)} - \mu\left(\Theta_j\right)\right]\right)^2 \\
&= \mathrm{Var}\left(h(\mathbf{X}) - \widehat{\mu\left(\Theta_j\right)}\right) + \mathrm{Var}\left(\widehat{\mu\left(\Theta_j\right)} - \mu\left(\Theta_j\right)\right) \\
&+ 2\ \mathrm{Cov}\left[\left(h(\mathbf{X}) - \widehat{\mu\left(\Theta_j\right)}\right)\ \left(\widehat{\mu\left(\Theta_j\right)} - \mu\left(\Theta_j\right)\right)\right] \\
&+ \left[\mathrm{E}\,h(\mathbf{X}) - \mathrm{E}\,\mu\left(\Theta_j\right)\right]^2.
\end{aligned} \tag{3.17}$$

It holds

$$\mathrm{Cov}\left(X_{jr}, \widehat{\mu\left(\Theta_j\right)}\right) = z_j \sum_s \frac{w_{js}}{w_{j.}}\,\mathrm{Cov}\left(X_{js}, X_{jr}\right), \tag{3.18}$$

$$\mathrm{Cov}\left(X_{jr}, X_{js}\right) = \mathrm{Var}\,\Theta_j = a,\ r \neq s,$$

$$\mathrm{Cov}\left(X_{jr}, X_{js}\right) = \mathrm{Var}\,X_{js} = a + \frac{s^2}{w_{jr}},\ r = s.$$

Inserting into (3.18) we get

$$\operatorname{Cov}\left(X_{jr}, \widehat{\mu(\Theta_j)}\right) = z_j \left(a + \frac{w_{jr}}{w_{j.}} \frac{s^2}{w_{jr}}\right) = a.$$

At the same time,

$$\operatorname{Cov}\left(X_{lr}, \widehat{\mu(\Theta_j)}\right) = 0,\ l \neq j$$

$$\operatorname{Cov}\left(X_{lr}, \mu(\Theta_j)\right) = \operatorname{Cov}\left(\mu(\Theta_l), \mu(\Theta_j)\right) \begin{cases} = 0, & l \neq j, \\ = a, & l = j. \end{cases}.$$

It is

$$\operatorname{Cov}\left(\widehat{\mu(\Theta_j)}, X_{lr}\right) = \operatorname{Cov}\left(\mu(\Theta_j), X_{lr}\right),\ l = 1, \ldots, J,\ r = 1, \ldots, t.$$

$h(\mathbf{X}) - \widehat{\mu(\Theta_j)}$ is a linear combination of the elements of $\mathbf{X}$. Therefore, the covariance on the right hand side of (3.17) is equal to zero. Also,

$$\mathrm{E}\,\widehat{\mu(\Theta_j)} = \mathrm{E}\,\mu(\Theta_j) = m.$$

The right hand side of (3.17) attains its minimum in the case $h(\mathbf{X}) = \widehat{\mu(\Theta_j)}$.
□

The estimate (3.16) cannot be computed exactly just from the observed values and known weights, it depends on unknown structural parameters (3.12)-(3.14). They are substituted in (3.16) by their estimates.

An unbiased estimator of s^2 is

$$\widehat{s^2} = \frac{1}{J} \sum_j \widehat{s_j^2},$$

where $\widehat{s_j^2}$ is an unbiased estimator of s^2 based on the data from group j:

$$\widehat{s_j^2} = \frac{1}{t-1} \sum_r w_{jr} \left(X_{jr} - \bar{X}_{j.}\right)^2.$$

An unbiased estimator of a is

$$\widehat{a} = \frac{\sum_j w_{j.} \left(\bar{X}_{j.} - \bar{X}_{..}\right)^2 - (J-1)\,\widehat{s^2}}{w_{..} - \sum_j w_{j.}^2 / w_{..}}.$$

The mean m is estimated by a weighted average of all the observed values, e.g.

$$\widehat{m} = \bar{X}_{..} = \frac{1}{w_{..}} \sum_j w_{j.}\,\bar{X}_{j.}.$$

3.3 Credibility Models in Classification Rate-making

We explain the application of the Bühlmann-Straub credibility model in the rate-making in the portfolio of risks classified according to values of several tariff variables. In section 2.3 we worked with the multiplicative model based on two tariff variables. We recall the formula for expected claims.

$$\mathrm{E}\,Y_{ij} = \mu\,\varphi_{1i}\,\varphi_{2j},\ i = 1, \ldots, I,\ j = 1, \ldots, J. \tag{3.19}$$

If we set $\varphi_{11} = \varphi_{21} = 1$, then φ_{1i} is interpreted as a **relativity** - it determines the ratio of the tariff rate in the class where the value of tariff variable 1 is equal to i to the tariff rate in the class with the value of the first tariff variable equal to 1 (always for the same level of the second tariff variable), i.e.

$$\varphi_{1i} = \frac{\mathrm{E}\,Y_{ij}}{\mathrm{E}\,Y_{1j}},\ j = 1, \ldots, J.$$

After renumbering the classes we obtained the vector of observations

$$\mathbf{Y} = (Y_1, \ldots, Y_M)',$$

where

$$\mathrm{E}\,Y_m = \mu\,\gamma_m,$$

with

$$\gamma_m = \gamma_{1m}\,\gamma_{2m},$$

where $\gamma_{1m} = \varphi_{1i}$, $\gamma_{2m} = \varphi_{2j}$, when the class m corresponds to the pair (i, j). We want to work with more general model with R tariff variables, then

$$\mathrm{E}\,Y_{i_1,\ldots,i_R} = \mu\,\varphi_{1i_1}\cdots\varphi_{Ri_R},\ i_k = 1, \ldots, I_k,$$

where I_K denotes the number of possible values for the tariff variable k.

After renumbering we have

$$\mathrm{E}\,Y_m = \mu\,\gamma_{1m}\ldots\gamma_{Rm}.$$

Assume that the numbers of possible values of the R tariff variables considered are relatively small and the available data are sufficient for estimating parameters $\mu, \gamma_{1m}, \ldots, \gamma_{Rm}$ based on GLM.

3.3.1 Multi-level Factors

Now let us introduce a new tariff variable with much higher range of possible levels (such a variable is sometimes called **multi-level factor**). There are levels of the multi-level factor, for which we have very few or none observations in our dataset. The estimation of parameters by GLM procedure is not possible due to the lack of data and also the intractability for so many different parameters. The idea of credibility theory is in exploring the information available in all classes to estimate the expected losses even in the classes with the factor levels that are very rarely met.

We denote a level of the multi-level factor by subscript j, while m stands for the classification according to "standard" tariff variables. For the sake of credibility aproach we also assume that our observations come from different time periods. This will be denoted by the third subscript, t. We work with the data in the form

$$Y_{mjt},\ m = 1, \ldots, M,\ j = 1, \ldots, J,\ t = 1 \ldots, n_j,$$

with known weights w_{mjt}.

We assume

$$\mathrm{E}\left(Y_{mjt} \mid U_j\right) = \mu\, \gamma_m\, U_j, \tag{3.20}$$

where

$$\gamma_m = \gamma_{1m} \cdots \gamma_{Rm}$$

is the product of the relativities of the class m with respect to "standard" tariff variables $1, \ldots, R$ and U_j is the random relativity of the multi-level factor class j. We also assume

$$\mathrm{E}\, U_j = 1.$$

If we denote

$$V_j = \mu\, U_j,$$

we can write

$$\mathrm{E}\left[Y_{mjt} \mid V_j\right] = \gamma_m\, V_j. \tag{3.21}$$

We use more specific assumptions concerning the conditional variances of the observations than in section 2.2. We suppose

$$\mathrm{Var}\left[Y_{mjt} \mid V_j\right] = \frac{\sigma^2\left(V_j\right)}{w_{ijt}} \tag{3.22}$$

with

$$\sigma^2\left(V_j\right) = \varphi\, \left(\gamma_m\, V_j\right)^p. \tag{3.23}$$

φ and p in (3.23) are parameters.

Remark. R.v. Y with the distribution from EDF with the density (2.1) has the variance

$$\operatorname{Var} Y = \frac{\varphi\, v(\mu)}{w},$$

where $v(\mu)$ is the variance function. (3.22) with σ^2 as specified in (3.23) is a special case with $v(\mu) = \mu^p$. This particular form of the variance function defines the class of **Tweedie distributions** in the framework of EDF. Important special cases are Poisson distribution with $p = 1$ or gamma distribution with $p = 2$.

Denoting

$$\sigma^2 = \varphi \operatorname{E} V_j^p, \tag{3.24}$$

we obtain from (3.23)

$$\operatorname{E}\left[\operatorname{Var}\left(Y_{mjt} \mid V_j\right)\right] = \frac{\gamma_m^p\, \sigma^2}{w_{mjt}}.$$

We want to use the Bühlmann-Straub credibility model to cope with the stochastic relativity representing the multi-level factor. For this we need the following assumptions to be fulfilled.

1. The classes given by random vectors

   $$\left(\{Y_{mjt}\}_{mt}, V_j\right), \; j = 1, \ldots, J,$$

   are independent.

2. Random variables V_j, $j = 1, \ldots, J$ are identically distributed with $\operatorname{E} V_j = \mu > 0$ and

   $$\operatorname{Var} V_j = \tau^2 > 0. \tag{3.25}$$

3. For every j, random variables Y_{mjt} are conditionally independent when V_j is given, and

   $$\operatorname{E}\left[Y_{mjt} \mid V_j\right] = \gamma_m V_j, \tag{3.26}$$

   $$\operatorname{E}\left(\operatorname{Var}\left[Y_{mjt} \mid V_j\right]\right) = \frac{\gamma_m^p\, \sigma^2}{w_{mjt}}. \tag{3.27}$$

We need to transform the observed variables to

$$\tilde{Y}_{mjt} = \frac{Y_{mjt}}{\gamma_m}. \tag{3.28}$$

Together with the transformed loss variables we use transformed weights given by

$$\tilde{w}_{mjt} = w_{mjt}\, \gamma_m^{2-p}. \tag{3.29}$$

Then we have from (3.26) and (3.27)

$$\mathrm{E}\left[\tilde{Y}_{mjt} \mid V_j\right] = V_j \tag{3.30}$$

$$\mathrm{E}\left(\mathrm{Var}\left[\tilde{Y}_{mjt} \mid V_j\right]\right) = \frac{\sigma^2}{\tilde{w}_{mjt}}. \tag{3.31}$$

It is seen that (3.30) and (3.31) formally correspond to the assumptions of the Bühlmann-Straub model of section 2.2. The results presented there can be applied to the transformed data under the assumption that the parameters $\gamma_m,\ m = 1, \ldots, M$ are treated as known values. Then we obtain from (3.16) the estimator

$$\widehat{V}_j = \tilde{z}_j\, \bar{\bar{\tilde{Y}}}_{.j.} + (1 - \tilde{z}_j)\ \mu, \tag{3.32}$$

or equivalently,

$$\widehat{U}_j = \tilde{z}_j\, \frac{\bar{\bar{\tilde{Y}}}_{.j.}}{\mu} + (1 - \tilde{z}_j)\,, \tag{3.33}$$

where

$$\tilde{z}_j = \frac{\tilde{w}_{.j.}}{\tilde{W}_{.j.} + \sigma^2/\tau^2}$$

and

$$\bar{\bar{\tilde{Y}}}_{.j.} = \frac{\sum_{m,t} \tilde{w}_{mjt}\, \tilde{Y}_{mjt}}{\sum_{m,t} \tilde{w}_{mjt}}.$$

The structural parameters σ^2 and τ^2 are replaced in (3.32) or (3.33) by the estimates calculated by the same means as described in section 2.2.

3.3.2 Combination of GLM and Credibility Models

In the final subsection we explain an iterative algorithm that enables us to estimate simultaneously the parameters concerning the "standard" tariff classes and the parameters of the Bühlmann-Straub model used to deal with multi-level factors. For (3.32) and (3.33) the parameters μ and $\{\gamma_m\}$ are assumed to be known.

On the other hand, when the estimates (3.33) are inserted as known values into (3.20), we can apply maximum likelihood estimation in an appropriate GLM model (with a Tweeedie distribution for Y_{mjt}). Since now the value U_j is treated as a known constant, before the application of methods described in section 1.3 we need a slight modification, called in the GLM context **offsetting**. If part of the multiplicative expression of the expected value in (1.1) is known, the estimation procedure is applied to modified data obtained by dividing the original values by known values of offsets.

In the case of (3.20) we denote $u_j = \widehat{U}_j$ and we work with new observations Y_{mjt}/u_j. In the multiplicative model with logarithmic link this results in subtracting $\log u_j$ from the linear predictor η_i.

Since we use the Tweedie model as specified by (3.23), the offsetting will change the weights in (3.22) to $w_{mjt}\, u_j^{2-p}$.

The iterative algorithm that applies both GLM and credibility estimation is described in the following steps:

- Step 0. Initially, set $\widehat{U_j} = 1$ for all $j = 1, \ldots, J$.
- Step 1. Estimate the parameters

$$\widehat{\mu},\ \widehat{\gamma}_1, \ldots, \widehat{\gamma_M} \tag{3.34}$$

  for the "standard" tariff variables by a GLM model with a Tweedie distribution, using $\log \widehat{U_j}$ as an offset variable.
- Step 2. Use (3.34) to compute the estimates of the structural parameters σ^2 and τ^2 given in (3.24) and (3.25).
- Step 3. Use (3.33) to compute $\widehat{U_j}$ using the estimates from Step 1 and 2.
- Step 4. Return to Step 1 with the new value of $\widehat{U_j}$ from Step 3.

Steps 1 - 4 are repeated until the estimates converge with required precision.

Bibliography

[1] E. Ohlsson, B. Johansson *Non-life insurance pricing with generalized linear models*, Springer, Berlin, 2010.

[2] M. V. Wüthrich *Non-life Insurance Mathematics & Statistics*, December 21, 2017. Available at SSRN: https://ssrn.com/abstract=2319328

[3] M.V.Wüthrich, M.Merz *Stochastic claims reserving methods in insurance*, John Wiley & Sons, 2008.

[4] R. Kaas, M. Goovaerts, J. Dhaene, M. Denuit *Modern Actuarial Risk Theory*, Kluwer Academic Publishers, Boston, 2001.

[5] H. Bühlmann, A. Gisler *A course in credibility theory and its applications*, Springer, Berlin, 2005.