Stockholms universitet

An Extension of Generalized Linear Models for dependent frequency and severity

Masar Al-Mosawi

Masteruppsats i försäkringsmatematik
Master Thesis in Actuarial Mathematics

Masteruppsats 2017:13
Försäkringsmatematik
December 2017

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Mathematical Statistics
Stockholm University
Master Thesis **2017:13**
http://www.math.su.se

An Extension of Generalized Linear Models for dependent frequency and severity

Masar Al-Mosawi*

December 2017

Abstract

In non-life insurance pricing, Generalized Linear models are used to estimate the pure premium through the product of the claim frequency and the claim severity. Traditionally, the methods in the Generalized Linear models imply an independence between claim count and claim amount. In practice the claim counts and amounts are often dependent. In this thesis, the two models where the claim counts and amounts are classically independent and a new approach where they are dependent will be analyzed and compared. The underlying data for the models considered is derived from a Swedish motorcycle insurance.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: masar@kth.se. Supervisor: Filip Lindskog.

For my father Muhsin Al-Mosawi and my mother Suhaila Al-Fahad

Contents

1. Introduction

A non-life insurance policy is an agreement between the insurance company and the policyholder. The policyholder pays a premium to the insurance company which represents the risk that is covered by the insurance policy. In practice the insurance company add other costs such as expenses and cost of capital but in this thesis we only discuss the modeling of the risk based pure premium. The pure premium should be based on the expected loss that is transferred from the policyholder to the insurance company.

The fact that the expected losses vary between policies, can be due to: properties of the policyholder, properties of the insured objects, properties of the geographic region. This introduces the need for statistical methods and in particular Generalized Linear Models (GLMs) for pricing of non-life insurance policies. In non-life pricing the pure premium is modeled as the product of two estimates: claim frequency and claim severity. Traditionally this assumes that the claim counts and claim amounts are independent. The assumption that the claim counts and claim amounts are independent is not always vindicated. For example, motorcycle insurance policyholders who tend to file several claims per year are often associated with lesser claim amounts than policyholders who tend to file lesser claims per year.

An description in GLM can be found in Esbjorn Ohlsson and Bjorn Johansson's textbook 'Non-Life Insurance Pricing with generalized Linear Models' [1].

The goal in this thesis is to construct and analyze the classical model where the claim counts and amounts are independent, and an extension of the classical model where the claim counts and claim amounts are dependent. The extension of classical GLM analysis with independent frequency and severity to possible dependence is based on J.Garrido, C.Genest, J.Schulz's article 'Generalized linear models for dependent frequency and severity for insurance claims' [2].

2. Model building

The insurance data for the model building comes from the former Swedish insurance company Wasa and the underlying insurance product is motorcycle insurance. One of the main goal of a non-life insurance pricing is the tariff analysis, which is to determine how the key ratios varies with respect to some covariates.

2.1 Assumptions

Key ratios are needed for measuring data, and the analysis are carried out on them. A key ratio is defined in non-life pricing as the ratio between the stochastic variable response X and the volume measure, exposure w. The key ratios of interest when modeling a Generalized Linear Model (GLM) are summarized in Table 2.1 below.

The pure premium varies between policies due to for example, different properties of the policyholder, properties of the insured objects and properties of the geographic region. Hence, these variations can be estimated by a set of variables called covariates. The range of each covariate are called classes, meaning that each covariate is divided into disjoint intervals where each interval is called a class. For example, let vehicle age be a covariate, two classes could be defined as "Vehicle age at most 1 year" and "Vehicle age 2 years or more". For every unique combination of covariates and class we have a tariff cell. All the policies that are within the same tariff cell obtain the same pure premium. There are three assumptions to be explained before describing the GLMs.

- **Policy independence.** Let n be the number of policies and X_i be the response for policy i. Then $X_1, \ldots, X_n$ are independent.

- **Time independence.** Let n be the number of disjoint time intervals and X_i be the response for policy i. Then $X_1, \ldots, X_n$ are independent.

- **Homogeneity.** Let X_1 and X_2 be the response for two any policies in the same tariff cell with same exposure. Then X_1 and X_2 have the same probability distribution.

These assumptions make the foundation of the GLMs. Note that these three basic assumptions are not always true in practice: Motorcycle crashes can occur dependently of each other, which violates the policy independence. Motorcycle owners that have been in an accident tend to be more careful in the future, which violates the time independence. Homogenicity is rarely fulfilled since grouping motorcycles in different homogeneous risk groups and charge them with the same tariff is hard, this violates homogenicity. For put in best light, the deviation from the basic model assumptions in practice does not lead to any problems with the analysis of the tariff, hence we assume that the model assumptions are true trough out this thesis.

The assumptions can be read in detail in p.6 [1].

Response X	**Exposure** w	**Key ratio** $Y = X/w$
Number of claims	Duration	Claim frequency
Claim cost	Number of claims	Claim severity
Claim cost	Duration	Pure premium

Table 2.1: Key ratios in GLM

2.2 The multiplicative model

The data in the insurance business is insufficient to estimate the expected cost by observed pure premium. If we had enough claims data in each tariff cell, we could determine a premium for the cell by simply estimating the expected cost by the observed pure premium. Hence, there is a need for models giving an expected pure premium that varies more smoothly over the cells, with good precision of the cell estimates. The multiplicative model is the model to do so, in fact it is the standard model in pricing.

Let M be the number of covariates, and let m_i be the number of classes for covariate i. Then a tariff cell is denoted by the vector $(i_1,\ldots,i_M)$. Let $X_{i_1,\ldots,i_M}$ be the response and let $w_{i_1,\ldots,i_M}$ be the exposure, then this gives that the key ratio for a given tariff cell is $Y_{i_1,\ldots,i_M} = X_{i_1,\ldots,i_M}/w_{i_1,\ldots,i_M}$. Let $E(Y_{i_1,\ldots,i_M}) = \mu_{i_1,\ldots,i_M}$ be the expectation value of the key ratio for a given tariff cell. The multiplicative model is given by

$$\mu_{i_1,\ldots,i_M} = \gamma_0\gamma_{1i_1}\gamma_{2i_2}\cdots\gamma_{Mi_M}, \tag{2.1}$$

where $\gamma_{ki_k}, i_k \in \{1,\ldots,m_k\}$ are the relativities for covariate k. The model is over-parameterized at the moment. if we multiply all γ_{1i_1} with a factor $c \in \mathbb{N}$ and divide all γ_{2i_2} with the same factor c we get the same expectation value μ. Thus, to make the relativities unique we introduce a base cell $\gamma_{11} = \gamma_{21} = \cdots = \gamma_{M1} = 1$. This means that the first class in every covariate is equal to 1, and the other relativities measure the difference in relation to the base cell. In practice we let the class with the largest exposure be equal to 1. By introducing a base cell the, the number of parameters in the multiplicative model is reduced from $\prod_{j=1}^{M} m_j$ to $\sum_{i=1}^{M} m_j - M + 1$.

The multiplicative model described above is as any regression model, in any regression model the parameters or in this case the relativities should be interpreted as *ceteris paribus*. This means that the relativities measure the effect when all other variables are held constant, or sometimes stated as "other things being equal" or "holding everything else constant".

For more information see Murphy, Brockman and Lee's article [3], and p.12 [1].

GLM is a generalization of the ordinary linear regression models. The link function is the fundamental object in a GLM, it links the mean to the linear structure. We rewrite the equation 2.1 using the dummy variable z_{kj} defined as

$$z_{kj} = \begin{cases} 1,\ j = i_k,\ \gamma_{ki_k} \neq 1 \\ 0,\ otherwise \end{cases} \tag{2.2}$$

with a logarithm transform, we obtain

$$log(\mu_{i_1,\dots,i_M}) = log(\gamma_0) + \sum_{j=1}^{m_1} z_{1j} log(\gamma_{ij}) + \cdots + \sum_{j=1}^{m_M} z_{Mj} log(\gamma_{ij}). \tag{2.3}$$

Let

$$\begin{aligned} \beta_1 &= log(\gamma_0) \\ \beta_2 &= log(\gamma_{12}) \\ &\vdots \\ \beta_{m_1} &= log(\gamma_{1m_1}) \\ &\vdots \\ \beta_{m_1+\cdots+m_M-M+1} &= log(\gamma_{Mm_M}). \end{aligned}$$

With these variables defined above, we can rewrite the multiplicative model 2.1 as a linear model for the mean,

$$\eta_i \equiv g(\mu_i) = log(\mu_i) = \sum_{j=1}^{p} x_{ij}\beta_j, \tag{2.4}$$

where $p = m_1 + \cdots + m_M - M + 1$ and $x_{ij} \in \mathbb{R}^{\prod_j^M m_j \times p}$ are the entries for a *design matrix* with dimensions $\prod_{j=1}^{M} m_j \times p$. Note that this is the same structure as the linear model $\mu_i = \sum_{j=1}^{r} x_{ij}\beta_j$ except that the mean depends on a function $g(\mu_i)$. Since the function $g(\mu_i)$ links the mean to the linear structure, the function $\eta_i \equiv g(\mu_i)$ is called the link function. In this case, the link function is the logarithm function and will be used trough out this thesis because of the multiplicative model.

Further discussions on the link function and multiplicative models can be found in Brockman, M.J. and Wright, T.S article 'Statistical motor rating: making efficient use of your data [4], and in p.9 [1]

3. Generalized Linear Models

Generalized Linear Models (GLMs) is a rich class of statistical methods which generalizes the linear models as described above. GLMs takes care of the problems that occurs with linear regression since it is not fully suited for non-life insurance pricing: (i) GLMs assumes a general class of distributions instead of the normal distribution, it is a problem since the later assumes normally distributed random errors when in practice the number of insurance claims follows a discrete probability distribution on $\mathbb{N}$, and claim cost are non-negative and often skewed to the right. (ii) GLMSs have the link function instead of the mean being a linear function of the covariates, since the multiplicative model is more reasonable for pricing.

In this chapter we will deduce the classic GLM, that is well explained in [1]. The approach is to estimate the claim frequency and the claim severity and the estimated pure premium will be the product of the two. First, a description of the Exponential Dispersion Models (EDMs) are required since these EDMs are the probability distributions that will generalize the normal distribution used in linear models into the GLMs.

The basic ideas of GLMs were introduced in Nelder and Wederburn article 'Generalized linear models' [5]. More description on GLMs can be found in p.15 [1].

3.1 Exponential Dispersion Models

The probability distribution of an Exponential Dispersion Model (EDM) is given by the following probability density function in the continuous case and a probability mass function in the discrete case,

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

where y_i is the outcome of the key ratio stochastic variable Y_i, $\phi > 0$ is the dispersion parameter and θ_i is a parameter that is allowed to depend on i. The function $b(\theta_i)$ is assumed twice continuously differentiable, with invertible second derivative. For every choice of the function $b(\theta_i)$, we get a family of probability distributions.

We use the cumulant-generating function to derive the expected value and the variance of an EDM. The *moment generating function* of an EDM is defined as

$$E[e^{tY_i}] = \int e^{tY_i} f_Y(y,\theta_i,\phi)dy = \int exp\left\{\frac{y_i(\theta_i + t\phi/w_i) - b(\theta_i)}{\phi/w_i} + c(y_i,\phi,w_i)\right\}dy$$
$$= exp\left\{\frac{b(\theta_i + t\phi/w_i) - b(\theta_i)}{\phi/w}\right\}$$
$$* \int exp\left\{\frac{y_i(\theta_i + t\phi/w_i) - b(\theta_i + t\phi/w_i)}{\phi/w_i} + c(y_i,\phi,w_i)\right\}dy.$$

The second factor right of the last equal sign is the integral of the probability density function $f(y_i, \theta_i + t\phi/w_i, \phi)$ on the parameter space, thus it is equal to one. Taking the logarithm on the equation, we conclude that the *cumulant-generating function* of Y_i, denoted $\Psi(t)$ for an EDM exist and is given by

$$\Psi_{Y_i}(t) = \frac{b(\theta_i + t\theta_i/w_i) - b(\theta_i)}{\phi/w_i}. \tag{3.2}$$

The first moment, the expected value is given by

$$E[Y] = \Psi^{'}(0) = b^{'}(\theta). \tag{3.3}$$

And the second moment, the variance is given by

$$Var[Y] = \Psi^{''}(0) = b^{''}(\theta)\phi/w. \tag{3.4}$$

In general, it is more convenient to view the variance as a function of the mean μ. Since $\mu = E[Y] = b^{'}(\theta)$ and it is assumed that $b^{'}(\theta)$ is an invertible function. We can rearrange the variance $Var[Y] = b^{''}(\theta)\phi/w$ with $\theta = b^{'-1}(\mu)$ into

$$Var[Y] = b^{''}(b^{'-1}(\mu))\phi/w \equiv v(\mu), \tag{3.5}$$

where $v(\mu)$ is called the *variance function*. The variance function is important in GLM model building since within an EDM class, a family of probability distributions is uniquely characterized by its variance function.

We claim that the EDMs are closed under averaging two independent random variables with different weights, this is called that EDMs are *reproductive*. This result is important later on in the GLM extension in the next chapter, a fact that is emphasized by the following theorem, it can be found in [1] where the proof is left as an exercise.

Theorem 3.1.1. *Let Y_1 and Y_2 be two independent r.v from the same EDM family, i.e with the same $b(\theta_i)$ function, mean μ and dispersion parameter ϕ, with possibly different weights w_1 and w_2. Then $\frac{w_1Y_1 + w_2Y_2}{w_1+w_2}$ belongs to the same EDM family with weight $w_1 + w_2$.*

Proof. We will use the *cumulant-generating function* in this proof. It will help us find the distribution of sums of independent random variables. The cumulant-generating function is given by

$$\Psi_{Y_i}(t) = \frac{b(\theta_i + t\theta_i/w_i) - b(\theta_i)}{\phi/w_i}. \tag{3.6}$$

The cumulant-generating function has two properties: the first property is

$$\begin{aligned}\Psi_{Y_1+Y_2}(t) &= log(E[e^{t(Y_1+Y_2)}]) = log(E[e^{tY_1}]E[^{tY_2}]) \\ &= log(E[e^{tY1}]) + log(E[e^{tY1}]) = \Psi_{Y_1}(t) + \Psi_{Y_2}(t),\end{aligned}$$

and the second property is

$$\Psi_{aY_1}(t) = log(E[e^{t(aY_1)}]) = log(E[e^{(at)Y_1}]) = \Psi_{Y_1}(at),$$

where $a \in \mathbb{R}$. Combining these two properties will complete the proof

$$\begin{aligned}\Psi_{\frac{w_1Y_1+w_2Y_2}{w_1+w_2}}(t) &= \Psi_{\frac{w_1Y_1}{w_1+w_2}}(t) + \Psi_{\frac{w_2Y_2}{w_1+w_2}}(t) = \Psi_{Y_1}(\frac{w_1t}{w_1+w_2}) + \Psi_{Y_2}(\frac{w_2t}{w_1+w_2}) \\ &= \frac{b(\theta_1 + \frac{w_1t}{w_1+w_2}\theta_1/w_1) - b(\theta_1)}{\phi/w_1} + \frac{b(\theta_2 + \frac{w_2t}{w_1+w_2}\theta_2/w_2) - b(\theta_2)}{\phi/w_2}\end{aligned}$$

In the hypothesis, Y_1 and Y_2 have the same mean μ. Since $\mu = E[Y] = b^{'}(\theta)$ and $b^{'}(\theta)$ is an invertible function, having the same mean μ is equivalent to having the same θ parameter. Let $\theta_1 = \theta_2 = \theta_i$,

$$\begin{aligned}\Psi_{\frac{w_1Y_1+w_2Y_2}{w_1+w_2}}(t) &= (w_1+w_2)\frac{b(\theta_i + t\theta_i/(w_1+w_2)) - b(\theta_i)}{\phi} \\ &= \frac{b(\theta_i + t\theta_i/(w_1+w_2)) - b(\theta_i)}{\phi/(w_1+w_2)}\end{aligned}$$

Hence the weighted average $\frac{w_1Y_1+w_2Y_2}{w_1+w_2}$ belongs to the same EDM family as Y_1 and Y_2 with weight $w_1 + w_2$.

□

An overview of the theory of EDM can be found in Jorgensen's article 'The theory of Dispersion Models' [6], and in p.17 [1].

3.2 Inference

To estimate the β parameters in GLM we use the maximum-likelihood estimation (ML). Let m be the number of classes. The log-likelihood function of θ is by independence,

$$l(\theta,\phi,y) = l(y_1,\ldots,y_n,\theta_1,\ldots,\theta_n,\phi) = log(f_{Y_1,\ldots,Y_n}(y_1,\ldots,y_n,\theta_1,\ldots,\theta_n,\phi))$$
$$= \sum_{i=1}^{m} log(f_{Y_i}(y_i,\theta_i,\phi)) = \frac{1}{\phi}\sum_{i=1}^{m} w_i(y_i\theta_i - b(\theta_i)) + \sum_{i=1}^{m} c(y_i,\phi,w_i).$$

We are interested in finding the likelihood as a function of β rather than θ. It is achieved through the relations $\mu_i = b^{'}(\theta_i)$, $g(\mu_i) = \eta_i = \sum_i x_{ij}\beta_j$ and the chain rule. Fisher's score function of θ becomes

$$\frac{\partial l}{\partial \beta_j} = \sum_{i=1}^{m} \frac{\partial l}{\partial \theta_i}\frac{\partial \theta_i}{\partial \beta_j} = \frac{1}{\phi}\sum_{i=1}^{m}(w_i y_i - w_i b^{'}(\theta_i))\frac{\partial \theta_i}{\partial \beta_j} \tag{3.7}$$

$$= \frac{1}{\phi}\sum_{i=1}^{m}(w_i y_i - w_i b^{'}(\theta_i))\frac{\partial \theta_i}{\partial \mu_i}\frac{\partial \mu_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial \beta_j}. \tag{3.8}$$

If we use the property that $(f^{-1}(x))^{'} = \frac{1}{f^{'}(f^{-1}(x))}$ then the log-likelihood of β is

$$\frac{\partial l}{\partial \beta_j} = \frac{1}{\phi}\sum_{i=1}^{m}(w_i y_i - w_i b^{'}(\theta_i))\frac{1}{b^{''}(b^{'-1}(\mu_i)}\frac{1}{g^{'}(g^{-1}(\eta_i))}x_{ij} \tag{3.9}$$

$$= \frac{1}{\phi}\sum_{i=1}^{m} w_i(y_i - \mu_i)\frac{x_{ij}}{v(\mu_i)g^{'}(\mu_i)}. \tag{3.10}$$

By setting all these r partial derivatives equal to zero and multiplying by ϕ, we get the ML equation:

$$\sum_{i=1}^{m} w_i \frac{y_i - \mu_i}{v(\mu_i)g^{'}(\mu_i)} x_{ij} = 0, \quad \forall j. \tag{3.11}$$

And thus, we can estimate the $\beta = (\beta_1,\ldots,\beta_p)^T$ through $\mu_i = g^{-1}(x\beta)$ where $x \in \mathbb{R}^{\prod_j^M m_j \times p}$ is the design matrix with dimensions $\prod_j^M m_j \times p$.

For test of null hypotheses on parameters that have been estimated by maximum likelihood we use the Wald statistic. Let $\hat{\theta}$ be the estimate of the parameters θ obtained by maximizing the log-likelihood over the parameter space $\Theta \subseteq \mathbb{R}^p$,

$$\hat{\theta} = arg\max_{\theta\in\Theta} l(\theta,\phi,y). \tag{3.12}$$

We want to test the null hypothesis $H_0 : \theta_o \in \Theta_R$, where Θ_R is the restricted parameter space, a proper subset of Θ. The restriction being tested can be written as

$$\Theta_R = \{\theta \in \Theta : g(\theta) = 0\}, \tag{3.13}$$

where $g : \mathbb{R}^p \to \mathbb{R}^r$ is a vector valued function, $r \leq p$, all the entries of g are continuously differentiable with respect to its arguments. If we want to test the null hypothesis $H_0 : g(\theta_0) = 0$, where $g : \mathbb{R}^p \to \mathbb{R}^r$, $r \leq p$, then the Wald statistic is

$$W = ng(\hat{\theta})^T [J_g(\hat{\theta})\hat{V}J_g(\hat{\theta})^T]^{-1} g(\hat{\theta}), \tag{3.14}$$

where p is the number of parameters in the estimate θ, n is the number of observations, r is the number of arguments in the function $g(\theta_0)$, $J_g(\theta)$ is the Jacobian of g, i.e the $r \times p$ matrix of partial derivates of $g(\theta)$ with respect to the entries of θ with rank r and $\hat{V}$ is a consistent estimate of the asymptotic covariance matrix of $\hat{\theta}$. Asymptotically, the Wald statistic has a Chi-square distribution with r degrees of freedom, $W \sim \chi^2_{1-p}(r)$ where $1-p$ is the confidence level. The null hypothesis is rejected if

$$W > z, \tag{3.15}$$

where z is a pre-determined critical value. The test of the null hypothesis can be approximated by its asymptotic value,

$$p = P(W > z) = 1 - P(W \leq z) \approx 1 - F(z), \tag{3.16}$$

where $F(z)$ is the Chi-square distribution function with r degrees of freedom. The critical value z can be chosen as

$$z = F^{-1}(1-p). \tag{3.17}$$

In the univariate case, i.e the case where only one parameter is tested, the Wald statistic under the null hypothesis $H_0 : \theta = \theta_0$ with confidence level of $1-p$ becomes

$$W = \frac{(\hat{\theta} - \theta_0)^2}{Var(\hat{\theta})} \sim \chi^2_{1-p}(1). \tag{3.18}$$

In GLM a generalization of the idea of using the sum of squares of residuals for a good measure of goodness-of-fit is the *deviance function*. One can view the deviance function as a distance between two probability distributions and can be used to perform model comparison. We will use the deviance function, essentially the distance of the data from the predicted values as model validation. The smaller the mean of the deviance function the better the fit of the model compared to the data. Let the number r be the number of non-redundant parameters and let n equal the number of observations. If $r = n$, then the model is called a *saturated model* and is the perfect fit by setting all $\mu_i = y_i$. This model is trivial and is of non-practical interest, it is used as a benchmark in measuring the goodness-of-fit of other models, since it has the perfect fit. Define the scaled deviance D^* and the unscaled deviance D as,

$$D^*(y,\mu) = \frac{1}{\phi}D(y,\mu) = \frac{2}{\phi}(l(\theta,\phi,y) - l(\theta,\phi,\mu)). \tag{3.19}$$

The deviance functions will generate deviance plots for model validation, they can asses which model fits the data best. Meaning that for each probability distribution, the deviance function needs to be derived for model comparison.

Another criteria for estimating the quality of models in purpose for model selection is the Akaike information criteria (AIC). AIC finds among candidate models, the model that minimizes the information loss compared to the "true" model (i.e the process that generates the data). Define AIC as,

$$AIC = -2l(\hat{\theta},\phi,y) + 2K, \tag{3.20}$$

where K is the number of estimated parameters and $l(\hat{\theta},\phi,y)$ is the maximum value of the log-likelihood function for the model. The estimated parameters $\hat{\theta}$ is estimated using the ML-equations 3.11, see definition 3.12. Since $l(\hat{\theta},\phi,y)$ is the maximum log-likelihood, it chooses the parameters $\hat{\theta}$ s.t it maximizes the probability that the parameters $\hat{\theta}$ explains the observations y. The log-likelihood can thus be used as a measure of fit, higher log-likelihood gives better model fit. This gives us that the lower AIC value compared to another model the better the fit. But in equation 3.20 there is a positive parameter K, meaning that the more parameters a model has the higher AIC value. In summary AIC rewards goodness of fit (as assessed by the log-likelihood function), but it also includes a penalty that is an increasing function of the number of estimated parameters.

$\Delta_i = AIC_i - AIC_{min}$ is a measure of each model relative to the best model, where AIC_{min} is the minimum AIC value and represents the "best" model.

For further reading on scaled deviance testing see p.39 [1], and on the Wald statistica and Akaike information criteria further reading can be done in Leonhard Held and Daniel Sabanes Bove's Applied Statistical Inference: Likelihood and Bayes [7] and in Marco Taboga's Lectures on Probability Theory and Mathematical Statistics [8].

3.3 Frequency and severity models

The aggregate losses incurred by an insurer is the total amount paid out in claims over a fixed time period, $S = \sum_{j=1}^{N} Y_j$. S is the aggregate losses incurred, N is the number of claims, Y_j is the claim amount for the jth incurred claim and $j \in \{1,\dots,N\}$. It is assumed that conditionally on N, the individual claim amounts $Y_1, \ldots, Y_N$ are mutually independent, identically distributed. It is further assumed that N is independent on the values of the claim amounts, meaning that their distribution does not depend on N.

Because the frequency and severity are assumed to be independent one can write the mean aggregate claims in terms of the two mean models,

$$E[S] = E[N]E[Y]. \tag{3.21}$$

For further details on independence see, Renshaw A.E's article 'Modelling the claims process in the presence of covariates' [9], [1] and [2].

Let N_i be the number of claims in a tariff cell and let v_i denote the expectation, $v_i = E[N_i]$. We assume that N_i is a poisson distribution of an individual policy during any given period of time, $N_i \sim P(v_i)$. Let N_i follow a poisson distribution with probability mass function

$$f_{N_i}(n_i, v_i) = e^{-v_i} \frac{(v_i)^{n_i}}{n_i!}, \quad n_i = 0, 1, 2, \ldots \tag{3.22}$$

This probability distribution is a member of the EDM family,

$$f_{N_i}(n_i, v_i) = e^{-v_i} \frac{(v_i)^{n_i}}{n_i!} = exp\{(n_i log(v_i) - v_i) + c(n_i, w_i)\},$$

where $c(n_i, v_i) = -log(n_i!)$. By reparameterizing it through $\theta_i = log(v_i)$ it becomes clear that the poisson distribution is an EDM of the form of equation 3.1,

$$f_{N_i}(n_i, v_i) = exp\{(n_i \theta - e^{\theta_i}) + c(n_i, w_i)\}, \tag{3.23}$$

where $\phi = 1$ and $b(\theta_i) = e^{\theta_i}$. The general ML-equations 3.11 becomes for the poisson distribution,

$$\sum_i x_{ij}(n_i - v_i) = 0. \tag{3.24}$$

Since the variance function for the poisson distribution is obtained when $v(E[N]) = E[N]^{p-1}, p = 1$, see equation 3.5, the ML-equation is easily simplified into the above equation.

The unscaled deviance function of the poisson distribution can be obtained by taking the logarithm from equation 3.23 onto equation 3.19,

$$D(n, v) = 2 \sum_i (n_i log(n_i / v_i) + (v_i - n_i)). \tag{3.25}$$

For more detail see Beard and Pesonen's textbook 'Risk Theory' [10] and [1].

We assume that the cost of an individual claim Y_i is gamma distributed and let μ_i denote the expectation $\mu_i = E[Y_i]$. Let $Y_i \sim G(\alpha, \beta)$ s.t the density function is

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

where, $\alpha > 0$ and $\beta > 0$. A parameterization is required to show that the distribution is a member of the EDM family. Let $\mu_i = E[Y_i] = \alpha/\beta$ and $\phi = 1/\alpha$. The density function becomes

$$\begin{aligned} f_{Y_i}(y) = f_{Y_i}(y_i, \mu_i, \phi) &= \frac{1}{\Gamma(1/\phi)} \left(\frac{1}{\mu_i \phi} \right)^{1/\phi} y^{1/\phi - 1} e^{-y_i/(\mu_i \phi)} \\ &= exp\left\{ \frac{-y_i/\mu_i - log(\mu_i)}{\phi} + c(y_i, \phi, w_i) \right\}, \end{aligned}$$

where $c(y_i, \phi, w_i) = log(y_i/\phi)/\phi - log(y_i) - log\Gamma(1/\phi)$. To show that the distribution is an EDM in the form of 3.1.1 we do the parameterization $\theta = -1/\mu$. This gives us that the gamma distribution is,

$$f_{Y_i}(y_i, \theta_i, \phi) = exp\left\{ \frac{y_i \theta_i + log(-\theta_i)}{\phi} + c(y_i, \phi, w_i) \right\}. \tag{3.27}$$

The gamma function is thus an EDM with $b(\theta_i) = -log(-\theta_i)$. The ML-equations for the gamma distribution can be obtained as with the poisson distribution, but with the variance function $v(E[Y]) = E[Y]^{p-1}, p = 3$, see equation 3.5. The ML-equation 3.11 is then simplified into,

$$\sum_i \frac{x_{ij}}{\mu_i} (y_i - \mu_i) = 0. \tag{3.28}$$

The unscaled deviance function for the gamma distribution is given by

$$D(y, \mu) = 2 \sum_i (-1 + \frac{y_i}{\mu_i} + log(\frac{\mu_i}{y_i})), \tag{3.29}$$

if we simplify equation 3.19 from equation 3.27 when $\mu = -1/\theta$. For further details, see p.30 [1] and [3].

4. Generalized Linear Models extension

In this chapter an extension to the classic GLM will be put forth. Although it is convenient to assume independent GLMs for claim counts and amount, these variable are often associated in practice. We consider here a conditional approach in which the GLM for the claim severity is allowed to depend on the claim count. Two general approaches have been proposed to account for dependence between frequency and severity, Frees and Wang's article 'Copula credibility for aggregate loss models. Insurance Math' [11], and Gschlobl and Czado [12] together with Frees, E.W., Gao, J., Rosenberg, M.A's article 'Predicting the frequency and amount of health care expenditures' [13] . In this paper we follow the dependency approach following the proceedings of [2] which is based on [13] .

For a given class of policyholders, we continue to assume that conditionally on N, the individual claim amounts $Y_1,\ldots,Y_N$ are mutually independent and identically distributed but conditionally on N. The distribution of N itself does not depend on the values of the claim amounts. *To account for dependence, the mean of the severity distribution is allowed to depend on N*. This gives that,

$$E[S] = E[NE[\overline{Y}|N]], \tag{4.1}$$

where $\overline{Y}|N = (Y_1+\cdots+Y_N)/N$ is the average claim severity, S is the aggregate losses incurred and N is the number of claims. Conventionally $S \equiv 0$ when $N = 0$.

The claim count N is handled in the same manner as in the classic GLM approach explained above. Thus, the claim frequency is modeled with a poisson distribution,

$$f_{N_i}(n_i,v_i) = exp\{(n_i\theta - e^{\theta_i}) + c(n_i,w_i)\}, \tag{4.2}$$

where $\phi = 1$ and $b(\theta_i) = e^{\theta_i}$.

The deviance function is the same as 3.23 since it is the same probability distribution. For more discussion on the average claim severity see [13] and p.207 [2].

Let $\overline{Y} = (Y_1+\cdots+Y_N)/N$ where $N > 0$ denote the average claim severity. In the proposed GLM extension model, the dependent setup requires modeling the average claim severity using claim count N as both covariate and weight factor in the GLM with a logarithm link function. It is worth noting that $\overline{Y}$ is functionally dependent on N.

We need the probability distribution for the average claim severity $\overline{Y}$ conditionally on number of claims incurred, N, i.e an EDM for $\overline{Y}|N$. Theorem 3.1.1 helps us prove that $\overline{Y} \sim EDM(\mu, \phi/N)$. It can be found in p.207 [2].

Corollary 4.0.0.1. *If $Y_j \sim EDM(\mu, \phi)$ then $\overline{Y_j}|N_j \sim EDM(\mu, \phi/N_j)$, where $\overline{Y_j} = (Y_{j1} + \cdots + Y_{jN_j})/N_j$*

Proof. Let $Y_j \sim EDM(\mu, \phi)$, and $Y_j^{'}|N_j = \frac{w_1 Y_{j1} + \cdots + w_N Y_{jN_j}}{w_1 + \cdots + w_N}|N_j$. Let all the weights equal to one, i.e $w_j = 1, \forall j$. Then $Y_j^{'}|N_j = \overline{Y_j}|N_j = (Y_{j1} + \cdots + Y_{jN_j})/N_j|N_j$. According to Theorem 3.1.1, $\overline{Y_j}|N_j = (Y_{j1} + \cdots + Y_{jN_j})/N_j|N_j$ belongs to the same EDM family with weight N_j. □

Thus, modeling individual claim amounts is equivalent to modeling the average severity only when N is included as a weight in the model for $\overline{Y}$. Hence, the probability distribution is as in the independent case, the gamma distribution but with a weight . This gives us that the probability distribution for the average claim severity is

$$f_{\overline{Y}_i|N_i}(\overline{y}_i, \theta_i, \phi/n_i) = exp\left\{\frac{\overline{y}_i\theta_i + log(-\theta_i)}{(\phi/n_i)} + c(\overline{y}_i, \phi/n_i, w_i)\right\} \tag{4.3}$$

$$= exp\left\{\frac{\overline{y}_i\theta_i + log(-\theta_i)}{\phi} + c(\overline{y}_i, \phi/n_i, w_i)\right\} exp\{n_i\}. \tag{4.4}$$

GLMs for N and $\overline{Y}$ are considered in order to incorporate the effect of dependency. To be specific, for the logarithmic link function g_N and g_Y we have that,

$$v = E[N] = g_N^{-1}(xa), \quad \mu_\theta = E[\overline{Y}|N] = g_Y^{-1}(x\beta + \theta N), \tag{4.5}$$

where α and β are vectors of regression coefficients, and $\theta \in \mathbb{R}$ induces a degree of dependence between claim counts and amounts. The model formulation is such that when $\theta = 0$, $\mu_\theta = \mu_0 = g_Y^{-1}(x\beta) = \mu$, i.e the model for the average claim severity using N as a weight is equivalent to modeling the individual claim severities as is done in the independent claims model. With g_Y being a logarithmic link function, one gets that $\mu_\theta = e^{x\beta + \theta N} \equiv \mu e^{\theta N}$. Thus unless $\theta = 0$, one has

$$E[S] = E[NE[\overline{Y}|N]] \neq E[N]E[Y]. \tag{4.6}$$

In particular, the expected value of the aggregate claims can be written as

$$\begin{aligned}E[S] &= E[E[S|N]] = E[E[N\overline{Y}|N]] = E[NE[\overline{Y}|N]] \\ &= E[Ng_Y^{-1}(x\beta+\theta N)] = E[Ne^{x\beta+\theta N}] = E[N\mu e^{\theta N}] \\ &= \mu E[\frac{\partial}{\partial\theta}e^{\theta N}] = \mu\frac{\partial}{\partial\theta}E[e^{\theta N}] = \mu M_N^{'}(\theta),\end{aligned}$$

where $M_N^{'}$ is the moment generating function of N, which is poisson distributed. Hence, by definition the moment generating function of the Poisson distribution is given by $M_N(\theta) = \sum_n e^{\theta n}\frac{v^n}{n!}e^{-v} = e^{v(e^{\theta}-1)}$. Thus the expected value of the aggregate claims is

$$E[S] = g_Y^{-1}(x\beta)M_N^{'}(\theta) = v\mu e^{v(e^{\theta}-1)+\theta}. \tag{4.7}$$

If we compare the equation 4.6 with the above GLM extension equation 4.7 one can see that the only difference is the multiplicative factor

$$exp\{v(e^{\theta}-1)+\theta\}, \tag{4.8}$$

which can be regarded as correction term for dependence. When $\theta = 0$ that multiplicative factor becomes equal to one, and the mean of the aggregate claims $E[S]$ becomes as in the classic GLM.

The parameters α and β from equations 4.5 is estimated using maximum likelihood estimation (ML). For class $i \in \{1,\ldots,m\}$, let all policyholders in the class share the same vector $x_i = (x_{i1},\ldots,x_{ip})$ of rating variables. Then N_i and $\overline{Y}_i|N_i$ can be expressed as,

$$v_i = e^{x_i\alpha}, \quad \mu_{\theta i} = e^{x_i\beta+n_i\theta}, \tag{4.9}$$

where $\alpha = (\alpha_1,\ldots,\alpha_p)^T$, $\beta = (\beta_1,\ldots,\beta_p)^T$ and $\theta \in \mathbb{R}$. The joint log-likelihood becomes then,

$$l(\alpha,\beta,\theta) = \sum_i^m l_N(\alpha;n_i) + \sum_i^m l_{\overline{Y}|N}(\beta,\theta;\overline{y}_i|n_i). \tag{4.10}$$

Since claim count is modeled exactly in the same manner as in the classic GLM, it has the poisson distribution $N_i \sim P(v_i)$ that is described in equation 3.23. Thus, in the GLM extension claim count has the same ML-equations for α as in the classical GLM,

$$\sum_i^m x_{ij}(n_i - v_i) = 0. \tag{4.11}$$

The average claim severity conditionally on number of claims incurred has the gamma distribution with a weight N, $\overline{Y}_i|N_i \sim G(\mu_{\theta i},\phi/N_i)$, its probability distribution is the equation 4.3. From equation 3.10 one can obtain Fisher's score function,

$$\frac{\partial l_{\overline{Y}|N}(\beta,\theta;\overline{y}_i|n_i)}{\partial \beta_j} = \frac{1}{\phi}\sum_{i=1}^{m}(\overline{y}_i - \mu_{\theta i})\frac{x_{ij}}{v(\mu_{\theta i})g^{'}(\mu_{\theta i})}. \tag{4.12}$$

With the variance function $v(\mu_{\theta i}) = b^{''}(b^{'}(\mu_{\theta i})^{-1})\phi/n_i$, the ML-equation for β_j becomes

$$\sum_{i}^{m}\frac{n_i x_{ij}}{\mu_{\theta i}}(\overline{y}_i - \mu_{\theta i}) = 0. \tag{4.13}$$

To determine θ an additional score function is required,

$$\frac{\partial l_{\overline{Y}|N}(\beta,\theta;\overline{y}_i|n_i)}{\partial \theta} = \frac{1}{\phi}\sum_{i=1}^{m}(\overline{y}_i - \mu_{\theta i})\frac{n_i}{v(\mu_{\theta i})g^{'}(\mu_{\theta i})}, \tag{4.14}$$

with the same variance function $v = v(\mu_{\theta i})$ as above. the ML-equation for θ is

$$\sum_{i}^{m}\frac{n_i^2}{\mu_{\theta i}}(\overline{y}_i - \mu_{\theta i}) = 0. \tag{4.15}$$

The deviance function for the GLM extension severity model differs by a multiplicative factor due to the weight N that is described in corollary 4.0.0.1. Thus, the unscaled deviance can be simplfied from the probability distribution of $\overline{Y}|N$,

$$D(y,\mu) = 2\sum_{i}^{m} n_i(-1 + \frac{\overline{y}_i}{\mu_{\theta i}} + log(\frac{\mu_{\theta i}}{\overline{y}_i})). \tag{4.16}$$

Note that the difference from the classical GLM approach is the factor n_i in the sum. For further reading on the average claim severity, the correction term for dependence and ML-equations see p.207 [2].

5. Results

The analysis was carried out on data from the former Swedish insurance company Wasa, and concerns partial casco insurance for motorcycles. It contains aggregated data on all insurance policies and claims during 1994-1998. The data can be downloaded from here [1]. Further analysis of the data where made. The covariates for the GLM analysis is a subset due to correlation between some of the covariates. Some classes were merged due to low duration and small claim cost. Table 5.1 shows the covariates and classes together with a short description. We find that the covariates and classes in Table 5.1 are best

Covariates	**Description**	**Classes**
Zon	Geographic zone	(1,2,3,4,5)
MC class	Mc class	(1,2,3,4)
Vehicle age	The vehicle age	(1,2,3,4)

Table 5.1: Data description with covariates and classes. These covariates and corresponding classes are found independent and significant, and being used for the analysis between the classic GLM and its extension.

suited for the GLM analysis since they are both independent and significant.

Table 5.2 gives a small description of the distribution of the data with respect on claim count. One can see that the majority with 67% have claim count equal to zero and hence the corresponding aggregate losses incurred is also zero. Also 4% of the observations have claim count equal to 2. We see a small positive association with claim count and average amount gives that it may exist a positive correlation between frequency and severity. Consider-

Claim count	**Frequency**	**Percent**	**Average amount (Kr)**
0	412	67 %	0
1	178	29%	83 372
2	26	4%	84 674

Table 5.2: Data distribution on claim count with a short description of the data. 4% of observations have claim count 2, 67% have zero in claim count.

ing the data at hand see Table 5.2, we have that claim count is either 1 or 2.

When the claim count is 1 the extended GLM model, the average claim severity equals the classic GLMs claim amount, $\overline{Y_i} = Y_i$ when $N_i = 1$. Note that the subset of the data where claim count is 2 is relativily small, only 4% of the set of observations consists of $N_i = 2$, or 14% of claim count greater than zero.

It is natural to focus on the effect of dependency on the expected total loss cost, in order to compare classic GLM with its extension. Secondly a comparison on the estimates of the average severity between the classic GLM and GLM extension is necessary, since they both have exactly the same frequency model and estimates.

In theory, dependency between the claim counts and amounts is reflected only through the correction term $exp\{v(e^{\theta} - 1) + \theta\}$. We use the ML-equations 4.15 and 4.13 to obtain the estimate $\hat{\mu}_{\theta i}$, and through equation 4.9 we get $\hat{\beta}_j$ and the estimated dependence parameter $\hat{\theta}$. From the extension of the GLM model the dependence parameter was estimated to

$$\hat{\theta} = -0.3472. \tag{5.1}$$

We use the Wald statistic for testing the null hypothesis that the claim count has not a significant effect on confidence level 97.5% in the GLM extension, $H_0 : \hat{\theta} = \theta_0 = 0$. Equation 3.18 and 3.16 gives us that the Wald statistic is $W_{\theta} = 5.057$. With confidence level of 97.5%, the Chi-square distribution function with 1 degree of freedom gives us $\chi^2_{0,025}(1) = 5.024$. The Wald statistic is greater than the Chi-square distribution function with confidence level of 97.5%,

$$W_{\theta} > \chi^2_{0.025}(1). \tag{5.2}$$

And by equation 3.15 we deduce that we can reject the null hypothesis and thus the parameter θ is of significant effect of confidence level 97.5% in the GLM extension model with p-value of $0,0245$, this leads to the conclusion that there is dependence between claim counts and claim amounts.

Fig 5.1, 5.2 shows the relativities for the frequency and severity models both from the classical GLM model and the GLM extended model. All of the estimated parameters are defined as in equation 3.12 since they are estimated through their respective ML-equation. For the classical GLM model, the parameter estimates $\hat{\beta}_j$ and $\hat{\alpha}_j$ are deduced from the ML-equations 3.24 and 3.28. For the GLM extended model, the ML-equations 4.13, 4.11 and 4.15 are used for calculating the estimated parameters $\hat{\beta}_j$, $\hat{\alpha}_j$ and $\hat{\theta}$. The relativities are obtained through the link-function 2.4. Also in Fig 5.1 you can find the correction term 4.8. Fig 5.3 describes the pure premium for the classical GLM model with a red color, and the GLM extension with a green color. Their x-axes represent the total classes over the covariates, and the y-axies represent the relativities.

One can conclude that since $\hat{\theta} < 0$ gives the correction term to be < 1, leading to lesser values for all policies for the severity estimates. This can be misleading since the estimates of the severity also differ from one another due to the absence/presence of the claim count N as a covariate. Fig 5.1 show the relativities of the severity model between the approaches classic GLM and GLM extension together with the correction term. One can see that the relativities of the GLM extension is always equal or less than the classic GLM, this is due to the presence of the claim count N as a covariate in the severity model in the GLM extension, but it is absent in the classic GLM severity model. The correction term is less than 1 everywhere as predicted and is shown in Fig 5.1. Fig 5.2 shows that the relativities of the claim frequency is exactly the same between the two approaches, since the claim frequency is modeled in the same way. Fig 5.3 captures the difference in pure premium between the two approaches. It follows that for the GLM extension model, the pure premium which is the product of the claim frequency and average claim severity, is lesser than the classical GLM approach due to the difference in claim severity model and the correction term that emerges in the extension of the GLM model.

The deviance residuals describes the difference in observed value and predicted value and are calculated through the deviance function. Fig 5.4, 5.5 and 5.6 describes the fitted values on the x-axis and deviance residuals on the y-axis. The mean of the deviance residuals plotted against the fitted value should be scattered around randomly neighboring zero, which the figures 5.4, 5.5 and 5.6 are showing. If the mean of the deviance residuals are equal to zero then the predicted values show a good fit against the observed values. All the models 5.4, 5.5 and 5.6 seems to show that the mean of the deviance residuals are equal to zero, but with different variances. Fig 5.4 shows how closely our model's predictions are to the observed outcomes, and are calculated in equation 3.25. Fig 5.5 is calculated in equation 3.29, and Fig 5.6 is calculated trough 4.16

One can see that the deviance residuals are scattered more closely around zero, giving lower variance for the severity model in the GLM extension in Fig 5.6 compared to Fig 5.5 which is the severity model for the classical GLM. This indicates the predicted values make a better fit against the observed values for the GLM extension, and one can make the conclusion that claim count as a covariate does make the severity model a better fit.

The Akaike's information criteria (AIC) values have been used for determining which of the two models would best approximate reality given the data at hand. In other words, which model minimizes the loss of information. For each parameter estimate defined in 3.12 for a specific model the AIC value have been calculated using equation 3.20. For the GLM extension, the AIC value is $AIC_{GLMextension} = 2637$, but when we drop the claim count as an covariate the AIC value then becomes $AIC_i = 2641$, hence

$$AIC_{min} = AIC_{GLMextension} < AIC_i. \tag{5.3}$$

This gives us that

$$\Delta_i = AIC_i - AIC_{min} = 4. \tag{5.4}$$

Meaning that the GLM extension have a lower AIC value compered with the same model but without claim count as a covariate which makes the GLM extension model a better model with respect to the goodness of fit and the number of parameters used.

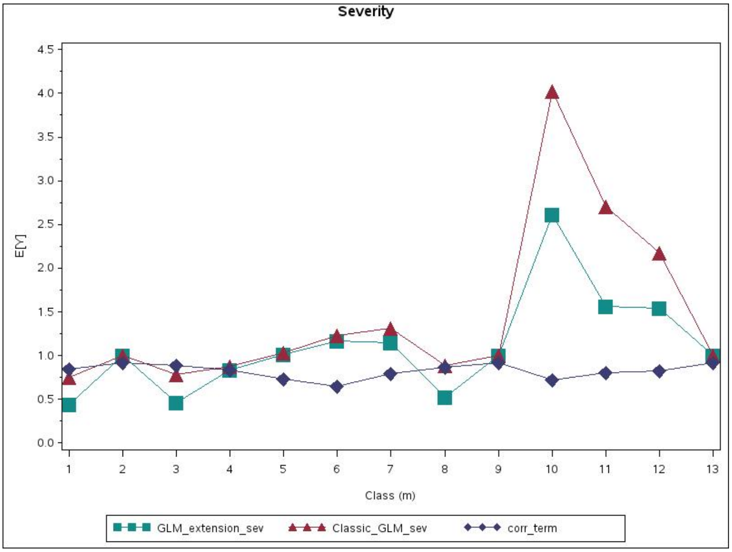

Figure 5.1: Comparison of the claim severity between the classic GLM and the GLM extension. The x-axis equals the classes, and the y-axis the relativities. The red represent the classical GLM severity model 3.27, the green represents the GLM extension which has a dependent setup that requires modeling the average using the claim count N as both a covariate and weight factor in the GLM 4.3. Equation 4.8 is used for the correction term which represents the purple graph.

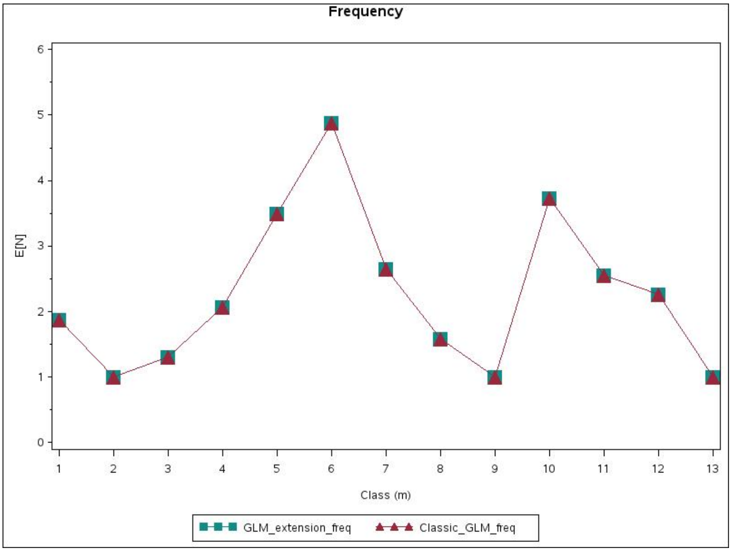

Figure 5.2: Comparison of the claim frequency between the classic GLM and the GLM extension. The x-axis equals the classes, and the y-axis the relativities. The claim frequency is handled in exactly the same way between the two approaches. The relativties is shown in this plot, and they are indeed exactly the same. Equation 3.23 is used both for the classical GLM model, and for the GLM extension equation since they are the same.

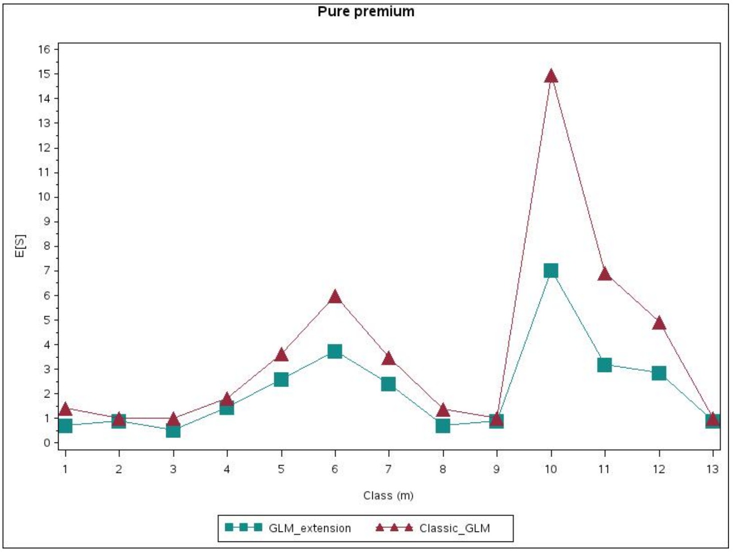

Figure 5.3: Comparison of the pure premium between the classic GLM and the the GLM extension. The x-axis equals the classes, and the y-axis the relativities. The pure premium is almost lower everywhere except for class $m = 1$ in the GLM extension than the classic GLM, due to the difference in the severity and correction term. For the corresponding model the pure premium is the multiplication between the corresponding models frequency and severity.

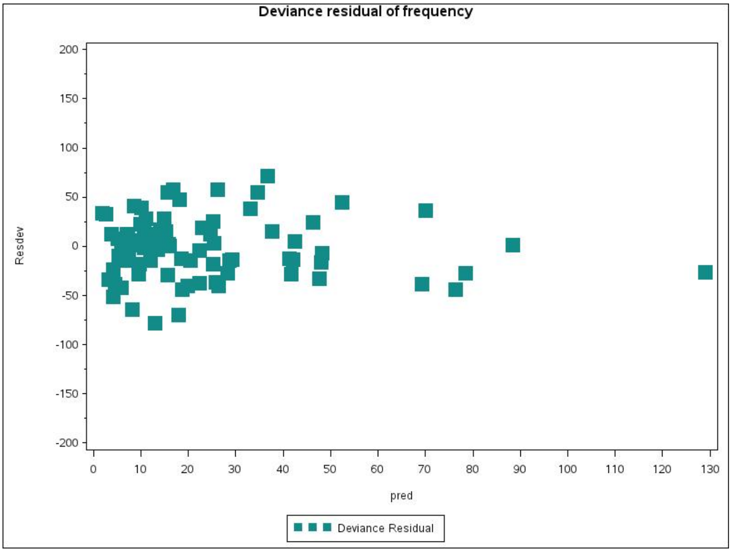

Figure 5.4: The deviance of the claim frequency model. The x-axis represent the fitted values, the y-axis represents deviance residuals. The mean of the deviance is zero, and the deviances seem to be scattered randomly around zero. The deviance of the frequency is obtained through equation 3.25.

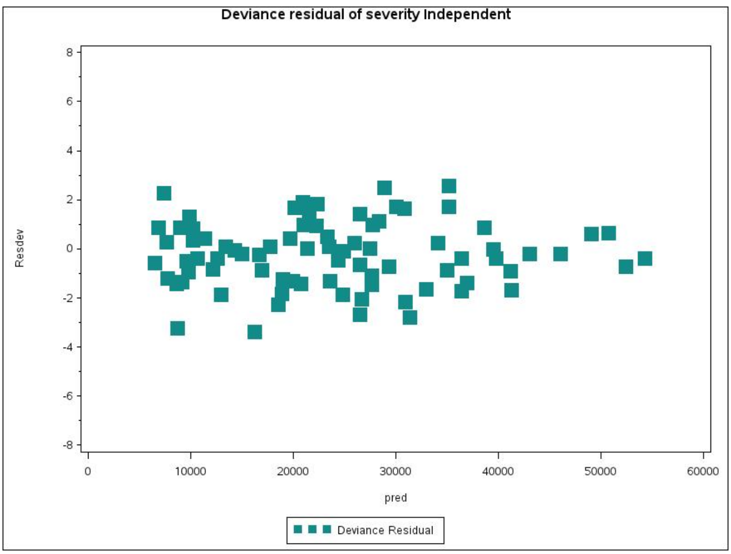

Figure 5.5: The deviance of the claim severity for the classical GLM. The x-axis represent the fitted values, the y-axis represents deviance residuals. The mean is also zero and the deviances seem to be scattered randomly. The deviance of the severity for the classical GLM is calculated in equation 3.29.

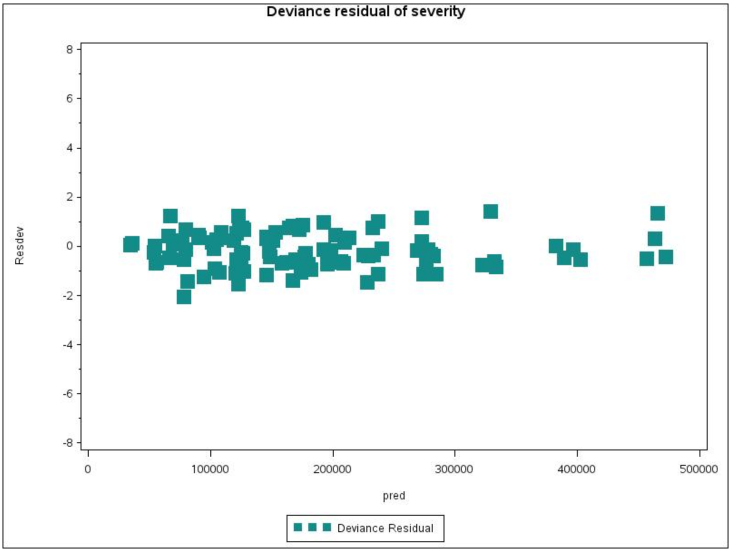

Figure 5.6: The deviance of the claim severity model for the GLM extension. The x-axis represent the fitted values, the y-axis represents deviance residuals. Also here one can see that the mean of the deviance is zero, and the deviances seem to be scattered randomly around zero. Compared to fig 5.5 the deviances are less scattered and more centered around zero, indicates that the GLM extension fit better to the data than the classical GLM. The deviance of the severity for the GLM extension is obtained in equation 4.16

6. Conclusion

In this thesis we have discussed a proposal for how to extend the classical GLM model by assuming that the claim count and claim amounts are independent for estimating the aggregate losses incurred by an insurer. We found that the dependence parameter 5.1 in the GLM extension was significant on the confidence level 97.5%, see 5.2, which indicates that claim count is a significant covariate in the GLM extension model.

When comparing the GLM extension model with and without the claim amount as a covariate, the Akaike information criteria (AIC) found that the GLM extension with the claim count is the best model to approximate reality, see 5.3. A rule of thumb for interpreting Δ_i is if $3 < \Delta_i < 7$ then there is less support for the GLM extension model without the claim count as a covariate than the GLM extension model, see 5.4. The value Δ_i in our case equals 4 and thus there is less support for the GLM extension model without the claim count. This result is in alignment with the previous result, that the dependence parameter 5.1 was significant on the confidence level 97,5 % 5.2. Note that $\Delta_i = 4$ is in the lower bound of $(3,7)$ meaning that Δ_i is not large enough to fully accept claim count, nor is it small enough to omit it. The data at hand may be to small to draw any conclusion from the AIC for testing the GLM extension since only 4% of the data consists of claim counts greater than one, see 5.2. This means that the severity model in the GLM extension approach only differs from the classical GLM in these cases. This may be the reason why the AIC value does not lead to a firm conclusion. Despite the small data for the extended GLM model, we have a significant effect on dependence between claim count and amount, and an AIC value that indicates that the GLM extension model is the better one.

The deviance figures 5.6 and 5.5 both show that the mean of the deviance residuals are equal to zero, but Fig 5.6 show that the deviance residual are scattered more close to zero, the variance is lower compared to Fig 5.5. This shows that the GLM extension model makes a better fit to the observations compared to the classical GLM model. Fig 5.3 shows that the pure premium will decrease if we assume a dependence between claim count and amount, meaning that the risk which the insurance company takes is valued too high. The reason is due to the fact that the severity model differs from the classical GLM model, and the emergence of a correction term, both can bee seen in Fig

5.1.

For insurance companies where the data is greater, the impact of dependence may be great. However, ignoring dependence could have more repercussions.

The structure for the dependence approaches makes it very easy to implement. The frequency model is exactly the same. For the severity model, the dependent setup requires modeling the average severity using claim count as a covariate and a weight factor in the GLM. The proposed extension makes it an appealing way for account for dependence, and is valid irrespectively of the choice of distribution for the claim counts and claim amounts.

Further studies can be done for different distributions for claim counts and amounts and with greater data, where the share of claim count larger than one is greater.

Acknowledgements

I would like to thank my supervisor Filip Lindskog, for your helpful discussions on this subject.

References

[1] BJORN JOHANSSON ESBJORN OHLSSON. *Non-Life Insurance Pricing with Generalized Linear Models*. 2009. 9, 12, 13, 14, 15, 16, 17, 20, 21, 22, 27

[2] J.SCHULZ J.GARRIDO, C.GENEST. **Generalized linear models for dependent frequency and severity of insurance claims**. 2016. 9, 21, 23, 24, 26

[3] BROCKMAN M.J. LEE R.W.M MURPHI, K.P. **Using generalized linear models to build dynamig pricing systems for personal lines insurance.** 2000. 13, 22

[4] WRIGHT T.S BROCKMAN, M.J. **Statistical motor rating: making efficient use of your data . J. Inst. Actuar**. 1992. 14

[5] WEDDERBURN R.W.M NEDLER, J.A. **Generalized linear models.** 1972. 15

[6] B JORGENSEN. **The theory of Dispersion Models**. 1987. 17

[7] DANIEL SABANES BOVE LEONHARD HELD. *Applied Statistical Inference: Likelihood and Bayes*. 2013. 20

[8] MARCO TABOGA. *Lectures on probability theory and mathematical statistics*. 2012. 20

[9] A.E RENSHAW. **Modelling the claims process in the presence of covariates**. 1994. 21

[10] PENTIKAINEN T. PESONEN E BEARD, R.E. *Risk Theory*. 1984. 21

[11] GAO J. ROSENBERG M.A FREES, E.W. **Copula credibility for aggregate loss models, Insurance Math. Econom**. 2006. 23

[12] CZADO C.. GSCHLOBL, S. **Spatial modelling of claim frequency and claim size in non-life-insurance. Scand. Actuar. J.2007**. 2007. 23

[13] GAO J. ROSENBERG M.A FREES, E.W. **Predicting the frequency and the amount of health care expenditures. N. Am. Actuar. J**. 2011. 23