Generalized Linear Models for a Dependent Aggregate Claims Model

Juliana Schulz

A Thesis

for The Department of

Mathematics and Statistics

Presented in Partial Fulfillment of the Requirements

for the Degree of Master of Science (Mathematics) at

Concordia University

Montréal, Québec, Canada

September 2013

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: **Juliana Schulz**

Entitled: **Generalized Linear Models for a Dependent Aggregate Claims Model**

and submitted in partial fulfillment of the requirements for the degree of

Master of Science (Mathematics)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

________________________________ Examiner

Dr. C. Genest

________________________________ Examiner

Dr. Y. P. Chaubey

________________________________ Thesis Supervisor

Dr. J. Garrido

Approved by ____

Chair of Department or Graduate Program Director

____

Dean of Faculty

Date ____

Abstract

Generalized Linear Models for a Dependent Aggregate Claims Model

Juliana Schulz

Key words: Aggregate Claims Model, Frequency, Severity, Loss Cost, Dependence, Exponential Dispersion Family, Generalized Linear Models, Correlated Responses, Conditional Model, Marginal Model

This thesis develops an alternative approach to modelling the expected loss cost of an insurance portfolio that allows for dependence between the frequency and severity components of the aggregate claims process. The traditionally used independent aggregate claims model is extended to define a dependent model, thus allowing for a correlation between the claim counts and claim amounts. A Generalized Linear Model framework is developed for the aggregate claims model in the dependent setting using a conditional severity model and marginal frequency model. We find that the pure premium in the dependent aggregate claims model is the product of a marginal mean frequency, a modified marginal mean severity and a correction term. This dependent modelling approach is then compared with the independent aggregate claims model GLM structure. It is shown that the expected total loss amount derived in the independent model is in fact a special case of the dependent model.

Acknowledgments

First, I would like to express my sincere gratitude to my supervisor Dr. José Garrido for all his help and guidance throughout my Master's degree. Thank you, as well, to all the wonderful professors I had at Concordia University who helped me to further my knowledge and passion in math. I would also like to thank Dr. Christian Genest and Dr. Yogendra Chaubey for reviewing my thesis.

This work was funded by Le Fonds de recherche du Québec Nature et technologies (FRQNT), the Department of Mathematics and Statistics at Concordia University, and Mprime.

To my grandfather: Edwin Schulz

Contents

List of Tables ix

List of Figures x

Introduction 1

1 The Aggregate Claims Model 4

1.1 Aggregate Claims Under Independence 5

1.2 Aggregate Claims Under Dependence 9

2 Dispersion Models 11

2.1 Definitions 12

2.2 Exponential Dispersion Models 13

2.2.1 Reproductive and Additive ED Models 16

2.2.2 Properties of the ED models 17

2.2.3 Tweedie Models 22

3 Generalized Linear Models 25

3.1 The Model 26

3.1.1 Maximum Likelihood Estimation for the ED Family 27

3.1.2 Link Function 29

3.1.3 Asymptotic Results for MLEs 33

3.1.4 Goodness of Fit 34

3.1.5 Residuals 36

3.2 GLMs for Insurance Data 37

4 GLMs for Aggregate Claims Under Independence 39

4.1 The Independent Model 39

4.1.1 Higher Moments in the Independent Model 42

4.2 MLEs in the Independent Model 44

4.3 Tweedie Modelling 47

5 GLMs for Aggregate Claims Under Dependence 48

5.1 The Dependent Model 49

5.1.1 Higher Moments for the Dependent Model 52

5.2 MLEs in the Dependent Model 55

6 Example 62

6.1 Data Description 62

6.2 Modelling the Data 66

6.3 Independent Model 69

6.3.1 Frequency Model I 70

6.3.2 Severity Model I 71

6.3.3 Aggregate Claims Model I 72

6.4 Dependent Model 73

6.4.1 Frequency Model D 73

6.4.2 Severity Model D 73

6.4.3 Aggregate Claims Model D 75

6.5 Model Comparisons 75

6.5.1 Residuals 79

6.5.2 Multicollinearity 84

6.5.3 Hold Out Dataset 85

6.5.4 Regression Parameter Estimates 90

Conclusion 94

Bibliography **95**

List of Tables

2.1 Examples of Dispersion Models 14

2.2 Some Continuous Exponential Dispersion Models 22

2.3 Some Discrete Exponential Dispersion Models 22

2.4 Tweedie Exponential Dispersion Models 24

3.1 Canonical Link Function of Some Common ED Models 31

3.2 Deviances of Some Common ED Models 35

6.1 Claim Count Distribution 63

6.2 Average Claim Severity by Claim Count 64

6.3 Continuous Covariates Correlation Matrix 66

6.4 Model I - Severity GLMs 72

6.5 Model D - Severity GLMs 74

6.6 Deviance Comparison 77

6.7 Deviance Comparison - Hold Out Dataset 85

6.8 Regression Parameter Estimates - Frequency Model 91

6.9 Regression Parameter Estimates - Severity Model I 92

6.10 Regression Parameter Estimates - Severity Model D 93

List of Figures

6.1 Average Severity Histogram 63
6.2 Model I Modified Pearson Residuals 79
6.3 Model D Modified Pearson Residuals 80
6.4 Model I Modified Pearson Residuals (zoom) 80
6.5 Model D Modified Pearson Residuals (zoom) 81
6.6 Frequency Model Deviance Residuals 82
6.7 Severity Model I Deviance Residuals 82
6.8 Severity Model D Deviance Residuals 83
6.9 Loss Cost Model I Deviance Residuals 83
6.10 Loss Cost Model D Deviance Residuals 84
6.11 Model I Modified Pearson Residuals Hold Out Dataset 86
6.12 Model D Modified Pearson Residuals Hold Out Dataset 86
6.13 Model I Modified Pearson Residuals Hold Out Dataset (zoom) 87
6.14 Model D Modified Pearson Residuals Hold Out Dataset (zoom) . . . 87
6.15 Frequency Model Deviance Residuals Hold Out Dataset 88
6.16 Severity Model I Deviance Residuals Hold Out Dataset 88
6.17 Severity Model D Deviance Residuals Hold Out Dataset 89
6.18 Loss Cost Model I Deviance Residuals Hold Out Dataset 89
6.19 Loss Cost Model D Deviance Residuals Hold Out Dataset 90

Introduction

A primary objective of property casualty insurers is to adequately price the risk inherent in their portfolio by estimating the expected value of all future costs associated with the insurance protection provided. The standard approach in the industry is to study both the frequency (the number of claims) and severity (the claim amounts) separately. The product of the expected claim frequency and severity then yields the expected loss cost (or pure premium), which represents the total cost of all claims.

In recent years, several insurers have adopted the use of Generalized Linear Models (GLMs) for modelling both the frequency and the severity of the claims process. The GLM approach allows for the mean of the response variable to be expressed in terms of a linear combination of covariates via a link function (McCullagh and Nelder 1989). This method requires that the response variable distribution be a member of the Exponential Family (EF). The EF distribution structure results in a particular mean-variance relation that allows to further characterize the response variable. Specifically, under the EF, the variance is a function of the mean.

The standard approach in the insurance industry is to develop a GLM for the claim frequency separately from the GLM fitted to the claim severity, and then calculate the pure premium as the product of the expected frequency and the expected average severity. This approach inherently assumes independence between the frequency and the severity of the claims process, an assumption that is unrealistic: Both GLMs share common explanatory variables and are fitted to the same portfolio data. An alternative approach is to model the total loss cost directly by means of the Tweedie distribution, which models the aggregate claims as a Compound Poisson-gamma sum. However, this method also assumes independence between the claim counts and claim sizes.

In order to address the dependence between the frequency and the severity in the collective risk model, a multivariate modelling approach for correlated data must be used. Fahrmeir and Tutz (2001) provide an overview of various techniques that can be used within the multivariate modelling framework for dealing with dependent response variables; namely conditional models, marginal models and random effects models (Neuhaus, Hauck and Kalbfleish 1991; Agresti 1993; and Diggle, Liang and Zeger 1994). Each of these approaches presents different methods for dealing with the dependence between the responses, in our case frequency and severity, and entails different inference techniques. Song (2007) also describes several multivariate modelling techniques for correlated responses, namely, quasi-likelihood modelling, conditional modelling and joint modelling approaches.

In the aggregate claims model, the compound sum provides a particular mean structure within the GLM framework that allows a different approach to modelling the frequency and severity components under the assumption of dependence between the two processes. Specifically, this thesis develops a multivariate modelling approach via a modified conditional GLM. Without any assumption of independence between the frequency and severity components, it is shown that the expected loss cost can be written in terms of the marginal mean claim frequency, a modified marginal mean severity and a correction term. The structure obtained for the mean total claims cost includes the independent model as a special case.

This research provides an alternative approach for establishing insurance premiums which allows for dependence between the claim frequency and severity. The structure obtained is simple to implement and allows for a straightforward comparison of the dependent model with the traditionally used independent model. Moreover, this dependent GLM approach for correlated responses provides a more accurate representation of the insurance data and ultimately leads to more precise insurance premiums.

This thesis is organized as follows: Chapter 1 provides an introduction to the independent aggregate claims model and then extends the assumptions to define a dependent aggregate claims model that allows for correlation between the claim amounts and the claim counts. In Chapter 2, an overview of dispersion models, and in particular the sub-class of the exponential dispersion family, is provided. Generalized linear models as well as their application to insurance data are discussed in depth in Chapter 3. Chapter 4 then goes into detail on the GLM approach for the aggregate claims model in the independent setting while Chapter 5 extends this framework to the dependent model. Finally, Chapter 6 provides an application of the GLM structure for the dependent aggregate claims model using car insurance data and compares these results to the independent model approach.

Chapter 1

The Aggregate Claims Model

The aggregate losses incurred by an insurer represent the total claim amount paid out over a fixed time period. The aggregate claims models can be defined as the random sum

$$S = \sum_{i=1}^{N} Y_i \ ,$$

where N represents the number of claims incurred, or the claim frequency, and the $Y_i's$ represent the individual claim amounts, or claim severities. Both of the components of the aggregate claims, namely the frequency process and the severity process, are random. Thus we have that the aggregate claims random variable S is defined in terms of the random vector $(N, Y_1, ..., Y_N)$.

Consider the aggregate claims S on the individual level, that is, the total claim cost for an individual policyholder. It is a reasonable assumption that for a given policyholder, the individual claim severities, Y_i, will be independent and identically distributed. Often, for simplicity, it is further assumed that the claim counts, N, and the individual claim amounts, Y_i, are also independent, thus yielding the independent aggregate claims model. However, this simplifying assumption is unrealistic and does not provide an accurate representation of the total loss amount as the claim severities are likely to be dependent on the claim counts. For example, a policyholder that submits several claims might only generate small claim amounts while an insured who makes only one claim might in fact submit a higher-than-average claim amount.

Such associations between claim frequency and severity are not accounted for in the independent aggregate claims model. Accordingly, there is a need to extend the aggregate loss model to the dependent case.

The goal of an insurer is to charge an adequate premium for the insurance coverage provided to policyholders. Consequently, an insurer is interested in estimating the expected value of the aggregate claims amount for each individual, as well as the variance of the loss cost so as to quantify the risk ensued by the policyholder. Obviously, premiums will differ according to the assumptions made in the model as well as the modelling techniques used for the estimation. As previously mentioned, typically, the estimated expected loss cost is derived under the assumption that the frequency and severity components are independent. Under these assumptions, we have what we will refer to as the independent aggregate claims model. This chapter will begin by defining the aggregate claims model under independence and then extend this definition to a dependent model by allowing the individual claim amounts to be dependent on the claim counts.

1.1 Aggregate Claims Under Independence

Under the assumption of independence in the aggregate claims model, the components of the random vector $(N, Y_1, ..., Y_N)$ are assumed to be mutually independent. Here we will use the model formulation as defined in Klugman, Panjer, and Willmot (2008). More formally, suppose that the following assumptions hold:

1. Given the claim count, the claim severities are conditionally i.i.d.; that is, conditional on $N = n$, the random variables $Y_1, ..., Y_n$ are i.i.d.

2. The claim severities are independent of the claim frequency. Thus, conditional on $N = n$, the random claim amounts $Y_1, ..., Y_n$ will not depend on N and, moreover, the distribution of N does not depend on the values of the claim amounts $Y_1, ..., Y_N$.

Under the above independence assumption on the frequency and severity components of the aggregate claims model, the random variable S is simply a compound sum. This compound sum is defined in terms of a counting process and a jump process. In the insurance setting, the standard approach to modelling the aggregate losses is to model the claim frequency as a Poisson random variable and the claim severities by a gamma distribution. In the particular case where $N \sim Poisson$ and $Y_i \sim gamma$, we have that S follows a Compound Poisson-gamma (CPG) distribution.

In the independence setting, the distribution of the aggregate loss random variable S can be obtained directly from the marginals, that is, from the marginal distribution of N and the marginal distribution of the $Y_i's$. Once a separate model has been developed for both the frequency and severity components, the distribution of the aggregate losses can be derived by conditioning on N. It follows from the assumptions of mutual independence that the cumulative distribution function of S is:

$$F_S(s) = \mathbb{P}(S \leq s) = \sum_{n=0}^{\infty} \mathbb{P}(S \leq s \mid N = n) \times \mathbb{P}(N = n), \quad s \geq 0,$$

where the probability $\mathbb{P}(S \leq s \mid N = n)$ is often simplified for certain choices of distribution for Y_i. If we return to the CPG case, we have that conditional on $N = n$, S is the sum of n i.i.d. gamma random variables so that conditionally, S is also gamma distributed.

Similarly, under the assumption of mutual independence and i.i.d. claim severities, the probability generating function of S can be derived as follows:

$$\begin{aligned}
P_S(t) &= \mathbb{E}[t^S] = \mathbb{E}\left[t^{\sum_{i=1}^N Y_i}\right] = \mathbb{E}\left[\mathbb{E}\left[t^{\sum_{i=1}^N Y_i}|N\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[t^{Y_1}\cdots t^{Y_N}|N\right]\right] = \mathbb{E}\left[\prod_{i=1}^{N}\mathbb{E}\left[t^{Y_i}|N\right]\right] = \mathbb{E}\left[\left[\mathbb{E}[t^Y|N]\right]^N\right] \\
&= \mathbb{E}\left[\left[\mathbb{E}[t^Y]\right]^N\right] = \mathbb{E}\left[[P_Y(t)]^N\right] = P_N\left(P_Y(t)\right),
\end{aligned}$$

for $t \in \mathbb{R}$ such that t^S has finite expectation.

In the same way, it follows that the moment generating function of S is:

$$M_S(t) = M_N[\ln M_Y(t)],$$

wherever $M_Y(t)$ exists.

Using the moment generating function, we can derive the first and second moments of the aggregate claims random variable:

$$
\begin{aligned}
\mathbb{E}(S) &= \frac{d}{dt}M_S(t)\Big|_{t=0} = \frac{d}{dt}\left\{M_N[\ln M_Y(t)]\right\}\Big|_{t=0} = \left\{M_N'[\ln M_Y(t)] \times \frac{M_Y'(t)}{M_Y(t)}\right\}\Big|_{t=0} \\
&= M_N'[\ln M_Y(0)] \times \frac{M_Y'(0)}{M_Y(0)} = M_N'[\ln(1))] \times \frac{M_Y'(t)}{1} = M_N'(0) \times M_Y'(0) \\
&= \mathbb{E}(N)\,\mathbb{E}(Y)
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{E}(S^2) &= \frac{d^2}{dt^2}M_S(t)\Big|_{t=0} = \frac{d^2}{dt^2}\left\{M_N[\ln M_Y(t)]\right\}\Big|_{t=0} \\
&= \frac{d}{dt}\left\{M_N'[\ln M_Y(t)] \times \frac{M_Y'(t)}{M_Y(t)}\right\}\Big|_{t=0} \\
&= \Big\{M_N''[\ln M_Y(t)] \times \left[\frac{M_Y'(t)}{M_Y(t)}\right]^2 + M_N'[\ln M_Y(t)] \\
&\quad \times \frac{M_Y''(t)M_Y(t) - [M_Y'(t)]^2}{[M_Y(t)]^2}\Big\}\Big|_{t=0} \\
&= M_N''[\ln M_Y(0)] \times \left[\frac{M_Y'(0)}{M_Y(0)}\right]^2 + M_N'[\ln M_Y(0)] \\
&\quad \times \frac{M_Y''(0)M_Y(0) - [M_Y'(0)]^2}{[M_Y(0)]^2} \\
&= M_N''[\ln(1)] \times \left[\frac{\mathbb{E}(Y)}{1}\right]^2 + M_N'[\ln(1))] \times \frac{\mathbb{E}(Y^2) \times 1 - [\mathbb{E}(Y)]^2}{1^2} \\
&= \mathbb{E}(N^2) \times \mathbb{E}(Y)^2 + \mathbb{E}(N) \times [E(Y^2) - [\mathbb{E}(Y)]^2] \\
&= \mathbb{E}(N^2)\mathbb{E}(Y)^2 + \mathbb{E}(N)\mathbb{V}\text{ar}(Y),
\end{aligned}
$$

which implies that

$$
\begin{aligned}
\mathbb{V}\text{ar}(S) &= \mathbb{E}(S^2) - [\mathbb{E}(S)]^2 = \mathbb{E}(N^2)\mathbb{E}(Y)^2 + \mathbb{E}(N)\mathbb{V}\text{ar}(Y) - [\mathbb{E}(N)\,\mathbb{E}(Y)]^2 \\
&= \mathbb{E}(Y)^2[\mathbb{E}(N^2) - (\mathbb{E}(N))^2] + \mathbb{E}(N)\mathbb{V}\text{ar}(Y) \\
&= [\mathbb{E}(Y)^2]\mathbb{V}\text{ar}(N) + \mathbb{E}(N)\mathbb{V}\text{ar}(Y).
\end{aligned}
$$

Thus, we have that the first two moments of S are determined by the first two moments of the frequency and severity respectively.

Example 1.1.1. ***Compound Poisson-gamma***

Return to the case where the aggregate losses follow a Compound Poisson gamma with $N \sim Poisson(\lambda)$ and $Y_i \sim gamma(\alpha, \beta)$. Conditional on $N = n$, S is the sum of n i.i.d. gamma distributed random variables and so it is distributed as a $gamma(n\alpha, \beta)$. Thus, we have the following results for the aggregate loss random variable:

i) the cumulative distribution function of S is given by

$$F_S(s) = \sum_{n=0}^{\infty} \int_0^s \frac{\beta^{\alpha n}}{y\Gamma(\alpha n)} y^{\alpha n} e^{-y\beta} \times \frac{\lambda^n e^{-\lambda}}{n!} dy, \quad s > 0$$

and

$$F_S(0) = \mathbb{P}(N = 0) = e^{-\lambda}.$$

ii) the moment generating function of S is given by

$$M_S(t) = M_N[\ln M_Y(t)],$$

where $M_N(t) = \exp(\lambda(e^t - 1))$, and $M_Y(t) = \left(1 - \frac{t}{\beta}\right)^{-\alpha}$ for $t < \beta$; thus

$$M_S(t) = \exp[\lambda\{(1 - t/\beta)^{-\alpha} - 1\}], \quad t < \beta.$$

iii) the first and second moments of S are

$$\mathbb{E}(S) = \mathbb{E}(N)\,\mathbb{E}(Y) = \lambda\frac{\alpha}{\beta},$$

and

$$\mathbb{E}(S^2) = \mathbb{E}(N^2)\mathbb{E}(Y)^2 + \mathbb{E}(N)\mathbb{V}ar(Y) = (\lambda + \lambda^2) \times \left(\frac{\alpha}{\beta}\right)^2 + \lambda\left(\frac{\alpha}{\beta^2}\right),$$

implying that the variance is

$$\mathbb{V}ar(S) = [\mathbb{E}(Y)^2]\mathbb{V}ar(N) + \mathbb{E}(N)\mathbb{V}ar(Y) = \lambda\left(\frac{\alpha + \alpha^2}{\beta^2}\right).$$

The simplified expressions obtained for the first two moments of the random variable S in the independent model make the estimation of the expected loss cost for insurance purposes more straightforward. A model for the claim counts and claim amounts can be developed separately to ultimately obtain estimates for $\mathbb{E}[N]$ and $\mathbb{E}[Y_i]$, respectively. Then, under the assumption of the independent aggregate claims model, the pure premium is simply $\mathbb{E}[S] = \mathbb{E}[N]\mathbb{E}[Y_i]$. Obviously, the simplicity of this model makes it appealing and practical to implement.

Nonetheless, this model fails to account for the potential correlation between the severities Y_i and the frequency N. Extending the independent model to a dependent model will allow to better quantify the risk involved in the aggregate claims model, thus allowing to obtain a more accurate estimation of the expected loss cost, as well as its variance.

1.2 Aggregate Claims Under Dependence

Let us now define the aggregate claims model under dependence, thus relaxing the assumption of independence between the claim sizes and claim counts. We will continue to assume that the individual claim amounts $Y_1, ..., Y_N$ are conditionally i.i.d. given N, however, now these individual severities are assumed to be dependent on the claim count N.

Note that while modelling the claim frequency and severity separately, as done in the independent model, allows for greater insight into each of these processes, this approach ignores the possible association between the two components. In cases where there is a dependence between the claim counts and amounts, the independent model approach can lead to inaccurate results. Without any assumptions of independence between the frequency and severity, the first and second moments of the aggregate claims cannot simply be written in terms of the marginal moments of the claim counts and claim amounts.

Allowing for dependence complicates the model in the sense that knowing the marginal distributions of the claim frequency and severity is no longer sufficient to define the distribution of the aggregate losses. Since S is defined in terms of the random vector $(N, Y_1, ..., Y_N)$, a model for S can be defined through the joint distribution of the frequency and severity components such that the joint density integrates to the appropriate marginal models.

The aggregate claims size can be regarded as a bivariate random vector $\mathbf{S} = (N, Y)$ where N represents the claim frequency and Y, the claim severity. In modelling the aggregate losses, we are then assuming that the observations are realizations of the bivariate random vector $\mathbf{S} \sim p(\mathbf{s}; \boldsymbol{\theta})$. As discussed in Song (2007), the modelling objective is then to find estimates of the parameter vector $\boldsymbol{\theta}$. In the dependence setting, the parameter vector $\boldsymbol{\theta}$ must include a correlation structure to characterize the dependence between the components.

Note that for the multivariate normal distribution, the joint density is fully defined in terms of the first and second moments, i.e. a mean vector $\boldsymbol{\mu}$ and a covariance matrix $\boldsymbol{\Sigma}$. As Song (2007) points out, for non-normal correlated random variables, which is the case for insurance data, it is generally not possible to determine the joint distribution based on only the first and second moments.

As mentioned in the introduction, there are several approaches that can be taken in developing the aggregate loss cost in the dependent model, both in the assumptions that define the model as well as in the actual modelling techniques used. In this thesis, we will assume that the claim amounts Y_i are again conditionally i.i.d., however, now we will assume that they are dependent on the claim count N.

In modelling the expected aggregate claims size, the focus is on obtaining an estimate of the mean rather than defining a multivariate probability density for the aggregate amount. Thus, rather than defining a joint density, we can instead use an inference approach that will allow for modelling the mean while incorporating a dependence structure between the marginal components. This will be accomplished through a generalized linear modelling framework. More specifically, we will construct a particular GLM structure for the mean loss cost in the dependent aggregate claims model involving conditional and marginal means along with a correction term that encompasses the dependence.

Chapter 2

Dispersion Models

The family of dispersion models (DM) is an important class of probability distributions that encompasses many commonly used random variables, including the Normal distribution. The structure of the DM density function is flexible and implies many nice properties. The exponential dispersion models are an important subclass of the DM family that is of particular importance for modelling insurance data. This chapter will provide an introduction to dispersion models as well as the subclass of exponential dispersion models and some important properties of these models, as defined by Jørgensen (1997). Note that Song (2007) also provides a good overview of dispersion models.

Let us first consider the normal distribution: the density of $Y \sim \mathcal{N}(\mu, \sigma^2)$ is

$$f_Y(y; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y-\mu)^2\right\}, \qquad y \in \mathbb{R}. \tag{2.1}$$

Notice that the quantity $(y-\mu)^2$ essentially measures the distance of the observation y from its mean μ. Moreover, the quantity $\frac{1}{\sqrt{2\pi\sigma^2}}$ does not depend on the mean μ. Building on this particular structure, Jørgensen extends the discrepancy $(y-\mu)^2$ to a more general deviance function $d(y;\mu)$ as to define a broader class of distributions known as Dispersion Models (DM). The class of DM includes several distributions, both discrete and continuous, including the Poisson, binomial and negative binomial as well as the gamma, inverse Gaussian, and normal distributions, to name a few. This family of distributions is of particular importance in the generalized linear model framework, as will be shown in Chapter 3.

2.1 Definitions

Definition 2.1.1. *The (reproductive) dispersion model $DM(\mu, \sigma^2)$ is a family of distributions with probability density functions defined as*

$$f_Y(y; \mu, \sigma^2) = a(y; \sigma^2) \exp\left\{-\frac{1}{2\sigma^2} d(y, \mu)\right\}, \qquad y \in \mathcal{C}, \tag{2.2}$$

where

$\mu \in \Omega$ *is called the location parameter,*

$\sigma^2 > 0$ *is called the dispersion parameter,*

$a(y; \sigma^2)$ *is a normalizing term, and*

$d(y, \mu)$ *is called the deviance function.*

Note that the normalizing term $a(y; \sigma^2)$ is independent of μ, thus allowing for inference on μ to be carried out separately from that on σ^2. This follows from the orthogonality of the likelihood, that is, the Fisher Information matrix for the parameters (μ, σ^2) is diagonal. This property will greatly facilitate the estimation of parameters for members of the DM family.

Each distribution that belongs to the DM family is uniquely determined by the deviance function $d(y; \mu)$ and is fully parametrized by the location parameter, μ, and the dispersion parameter, σ^2. Many commonly used distributions can be re-parametrized as a DM.

The deviance function $d(\cdot; \cdot)$ on $(y, \mu) \in \mathcal{C} \times \Omega$ is referred to as a *unit deviance* function if:

i) $d(y; y) = 0, \quad \forall y \in \Omega;$

ii) $d(y; \mu) > 0, \quad \forall y \neq \mu.$

The unit deviance function is *regular* if it is twice continuously differentiable with respect to (y, μ) on $\mathcal{C} \times \Omega$ and

$$\frac{\partial^2}{\partial \mu^2} d(y; y) = \frac{\partial^2}{\partial \mu^2} d(y; \mu)|_{\mu=y} > 0, \qquad \forall y \in \mathcal{C}.$$

The *unit variance function* $V : \Omega \longrightarrow (0, \infty)$ for a regular unit deviance function $d(y; \mu)$ is defined as

$$V(\mu) = \frac{2}{\frac{\partial^2}{\partial \mu^2} d(y; \mu)\Big|_{y=\mu}}, \qquad \forall \mu \in \Omega.$$

Example 2.1.1. ***Normal***
For $Y \sim \mathcal{N}(\mu, \sigma^2)$, we have that the unit deviance is $d(y; \mu) = (y - \mu)^2$ for $y \in \mathbb{R}$ and $\mu \in \mathbb{R}$. It is easy to verify that the unit deviance is regular. The unit deviance is then $V(\mu) = \frac{2}{2} = 1$.

As previously mentioned, the dispersion models have many nice properties. The following propositions present a few of them.

Proposition 2.1.1. *If d is a regular unit deviance function, then*

$$\frac{\partial^2}{\partial y^2} d(\mu; \mu) = \frac{\partial^2}{\partial \mu^2} d(\mu; \mu) = -\frac{\partial^2}{\partial y \partial \mu} d(\mu; \mu), \quad \forall \mu \in \Omega.$$

Proposition 2.1.2. ***Saddlepoint Approximation*** *As $\sigma^2 \to 0$, the density of a regular DM can be approximated by:*

$$f_Y(y; \mu, \sigma^2) \simeq \left\{2\pi\sigma^2 V(y)\right\}^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2} d(y; \mu)\right\}.$$

Table 2.1 provides details on some members of the Dispersion Models family.

2.2 Exponential Dispersion Models

The exponential dispersion (ED) family of models are an important subclass of the DM family of distributions. It includes both continuous and discrete distributions, such as the gamma and Poisson distributions. The ED models are a special case of

Table 2.1: Examples of Dispersion Models

Distribution	$d(y;\mu)$	$V(\mu)$	$\mathcal{C}$	Ω
Normal	$(y-\mu)^2$	1	$(-\infty,\infty)$	$(-\infty,\infty)$
Poisson	$2\left(y\ln\frac{y}{\mu}-y+\mu\right)$	μ	$\{0,1,...\}$	$(0,\infty)$
Binomial	$2\left\{y\ln\frac{y}{\mu}+(n-y)\ln\frac{n-y}{n-\mu}\right\}$	$\mu(1-\mu)$	$\{0,1,...,n\}$	$(0,1)$
Negative Binomial	$2\left\{y\ln\frac{y}{\mu}+(1-y)\ln\frac{1-y}{1-\mu}\right\}$	$\mu(1+\mu)$	$\{0,1,...\}$	$(0,\infty)$
Gamma	$2\left(\frac{y}{\mu}-\ln\frac{y}{\mu}-1\right)$	μ^2	$(0,\infty)$	$(0,\infty)$
Inverse Gaussian	$\frac{(y-\mu)^2}{y\mu^2}$	μ^3	$(0,\infty)$	$(0,\infty)$

the DM, where the probability density function takes on a certain form. As previously mentioned, the exponential dispersion models are of particular importance for studying insurance data. Both count data, taking on positive integer values, and loss data, taking on values on the positive real line, can be written in terms of a ED model. Furthermore, the structure of the ED density allows for data to be modelled in the more flexible framework of generalized linear models (GLMs), which will be discussed in detail in the next chapter.

The framework of GLMs, as defined by McCullagh and Nelder (1989), is based on the assumption that the response variable Y is a member of the exponential family. They define the pdf of the response variable Y as follows:

Definition 2.2.1. *A response variable $Y \sim ED(\theta, \phi)$ has density*

$$f_Y(y;\theta,\phi) = \exp\left[\frac{y\theta-\kappa(\theta)}{a(\phi)} + C(y,\phi)\right], \qquad y \in \mathcal{C}, \tag{2.3}$$

for specific functions κ, a, C where:

θ is called the canonical parameter,

ϕ is called the dispersion parameter with $\phi > 0$,

$\kappa(\theta)$ is called the cumulant function.

When ϕ is known, this distribution is a member of the exponential family while when ϕ is unknown it is part of the exponential dispersion (ED) family.

It can be shown that for a certain unit deviance function $d(y, \mu)$ and normalizing term $C(y, \phi)$, the ED density can be rewritten in the same form of the DM as in definition (2.1.1). Let $\lambda = \dfrac{1}{a(\phi)}$, then the ED density can be rewritten as:

$$f_Y(y; \theta, \lambda) = c(y; \lambda) \exp[\lambda\{\theta y - \kappa(\theta)\}], \qquad y \in \mathcal{C}, \tag{2.4}$$

where the function $c(y; \lambda)$ is a normalizing term and the parameter λ is referred to as the *index parameter* with index set $\Lambda = \{\lambda > 0\}$.

For $Y \sim ED$, there is a relationship between the mean and variance of Y. Let $l(\theta, \phi; y) = \ln f_Y(y; \theta, \phi)$ denote the log likelihood function. Under standard conditions, we have that:

i) $\mathbb{E}_Y\left[\dfrac{\partial}{\partial\theta} l(\theta, \phi; y)\right] = \mathbb{E}_Y[\dot{l}(\theta, \phi; y)] = 0,$

ii) $\mathbb{E}_Y\left[\dfrac{\partial^2}{\partial\theta^2} l(\theta, \phi; y)\right] = \mathbb{E}_Y[\ddot{l}(\theta, \phi; y)] = -\mathbb{E}_Y[\dot{l}(\theta, \phi; y)^2].$

Thus, for the exponential family we have that:

$$l(\theta, \phi; y) = \frac{y\theta - \kappa(\theta)}{a(\phi)} + C(y, \phi),$$

$$\dot{l}(\theta, \phi; y) = \frac{y - \dot{\kappa}(\theta)}{a(\phi)},$$

$$\ddot{l}(\theta, \phi; y) = \frac{-\ddot{\kappa}(\theta)}{a(\phi)}.$$

It then follows from i) that:

$$\mathbb{E}_Y\left[\dot{l}(\theta, \phi; y)\right] = \mathbb{E}_Y\left[\frac{Y - \dot{\kappa}(\theta)}{a(\phi)}\right] = \frac{\mathbb{E}(Y) - \dot{\kappa}(\theta)}{a(\phi)} = 0,$$

which implies that

$$\mathbb{E}(Y) = \dot{\kappa}(\theta).$$

In the same way, from ii) we have that:

$$\begin{aligned}
\mathbb{E}_Y\left[\ddot{l}(\theta,\phi;y)\right]+\mathbb{E}_Y\left[\dot{l}(\theta,\phi;y)^2\right] &= \mathbb{E}_Y\left[\frac{-\ddot{\kappa}(\theta)}{a(\phi)}\right]+\mathbb{E}_Y\left[\left(\frac{Y-\dot{\kappa}(\theta)}{a(\phi)}\right)^2\right]\\
&= \frac{-\ddot{\kappa}(\theta)}{a(\phi)}+\mathbb{E}_Y\left[\left(\frac{Y-\mathbb{E}(Y)}{a(\phi)}\right)^2\right]\\
&= \frac{-\ddot{\kappa}(\theta)}{a(\phi)}+\frac{\mathbb{V}\text{ar}(Y)}{a(\phi)^2}=0,
\end{aligned}$$

which implies that

$$\mathbb{V}\text{ar}(Y)=a(\phi)\ddot{\kappa}(\theta).$$

We refer to $\tau(\theta)=\dot{\kappa}(\theta)=\mathbb{E}(Y)=\mu$ as the mean function. Notice that the variance is a function of the canonical parameter θ, and thus it is also a function of the mean μ. It follows that we can write $\mathbb{V}\text{ar}(Y)=a(\phi)V(\mu)$ where $V(\mu)=\ddot{\kappa}(\theta)$ is referred to as the variance function, which coincides with the variance function previously discussed.

Note that the mean mapping $\tau(\theta)=\dot{\kappa}(\theta)=\mu$ is strictly increasing since the variance $\mathbb{V}\text{ar}(Y)=\lambda\ddot{\kappa}(\theta)=\lambda\dot{\tau}(\theta)>0$ and thus $\dot{\tau}(\theta)>0$. Consequently, the inverse of the mean mapping exists and so we can write the canonical parameter θ in terms of the mean μ as $\theta=\tau^{-1}(\mu)$. The ED density can then be re-parametrized in terms of the parameters (μ,σ^2) for $\mu=\tau(\theta)$ and $\sigma^2=\dfrac{1}{\lambda}$ as follows:

$$f_Y(y;\mu,\sigma^2)=c(y;\sigma^2)\exp\left[\frac{1}{\sigma^2}\{y\tau^{-1}(\mu)-\kappa(\tau^{-1}(\mu))\}\right],\qquad y\in\mathcal{C}. \tag{2.5}$$

We also have that for the ED family, the unit variance function can be written in terms of the mean mapping function:

$$V(\mu)=\dot{\tau}(\tau^{-1}(\mu)),\quad \mu\in\Omega.$$

2.2.1 Reproductive and Additive ED Models

The form of the ED in (2.5) is referred to as the *reproductive exponential dispersion model*, denoted $ED(\mu,\sigma^2)$. Another form of the ED family is the *additive exponential dispersion model*, denoted $ED^*(\theta,\lambda)$ with density taking the form:

$$f_Z^*(z;\theta,\lambda)=c^*(z;\lambda)\exp\{\theta z-\lambda\kappa(\theta)\},\quad y\in\mathcal{C}. \tag{2.6}$$

These two representations of the ED family are essentially equivalent. The *duality transformation* links the $ED(\mu, \sigma^2)$ model to the $ED^*(\theta, \lambda)$:

$$Z \sim ED^*(\theta, \lambda) \quad \Rightarrow \quad Y = Z/\lambda \sim ED(\mu, \sigma^2)$$

for $\mu = \tau(\theta), \sigma^2 = 1/\lambda$, and

$$Y \sim ED(\mu, \sigma^2) \quad \Rightarrow \quad Z = Y/\sigma^2 \sim ED^*(\theta, \lambda)$$

for $\theta = \tau^{-1}(\mu), \lambda = 1/\sigma^2$. Thus we have that the mean and variance of the additive exponential dispersion family are given by:

$$\mathbb{E}[Z] = \mu^* = \lambda\tau(\theta),$$

$$\mathbb{V}\text{ar}[Z] = \lambda V(\mu^*/\lambda).$$

2.2.2 Properties of the ED models

Convolution

A nice property of the ED family is that it is closed under convolutions.

Proposition 2.2.1. *For the additive exponential dispersion family, if $Z_1, ..., Z_n$ are independent with $Z_i \sim ED^*(\theta, \lambda_i)$ then*

$$Z_+ = Z_1 + \cdots + Z_n \sim ED^*(\theta, \lambda_1 + \cdots + \lambda_n).$$

Proposition 2.2.2. *For the reproductive exponential dispersion family, if $Y_1, ..., Y_n$ are mutually independent with $Y_i \sim ED\left(\mu, \frac{\sigma^2}{w_i}\right)$ where the w_i are positive weights, then*

$$\frac{1}{w_+}\sum_{i=1}^{n} w_i Y_i \sim ED\left(\mu, \frac{\sigma^2}{w_+}\right),$$

where $w_+ = w_1 + \cdots + w_n$.

Furthermore, we have the following proposition concerning the deconvolution of the additive exponential dispersion models.

Proposition 2.2.3. *The family of additive exponential dispersion models is infinitely divisible if and only if the index parameter set* $\Lambda = (0, \infty)$*. So for* $Z \sim ED^*(\theta, \lambda)$ *we have that there exists i.i.d. random variables* $Z_1, ..., Z_n$ *such that* $Z \stackrel{d}{=} Z_1 + \cdots + Z_n$*, where each* $Z_i \sim ED^*(\theta, \lambda/n)$*.*

Moment Generating Function

For a random variable Y, we denote the moment generating function as $M_Y(t) = \mathbb{E}[e^{tY}]$ and the *cumulant generating function* as $K_Y(t) = \log M_Y(t)$. For the family of natural exponential models (i.e. exponential dispersion model with λ known) with density of the form

$$f_Y(y;\theta) = c(y)\exp\{\theta y - \kappa(\theta)\}, \quad y \in \mathcal{C},$$

the cumulant generating function can be found through the cumulant function $\kappa(\theta)$ as:

$$K_Y(t;\theta) = \kappa(\theta + t) - \kappa(\theta).$$

It then follows that the moment generating function is:

$$M_Y(t;\theta) = \exp\{K_\theta(t)\} = \exp\{\kappa(\theta + t) - \kappa(\theta)\}.$$

Recall that for the additive exponential family, the exponent of the pdf has the form $\exp\{\theta z - \lambda\kappa(\theta)\}$, thus corresponding to a cumulant function $\lambda\kappa(\theta)$. It then follows that for the $ED^*(\theta, \lambda)$ family, the cumulant generating function is

$$K^*(t;\theta,\lambda) = \lambda\{\kappa(\theta + t) - \kappa(\theta)\}$$

and the moment generating function is then

$$M^*(t;\theta,\lambda) = \exp[K^*(t;\theta,\lambda] = \exp[\lambda\{\kappa(\theta + t) - \kappa(\theta)\}].$$

Then, for the reproductive exponential family, by the duality transformation we have that $Y = Z/\lambda \sim ED(\mu, \sigma^2)$ and so the cumulant generating function of Y is then

$$K(t;\theta,\lambda) = \lambda\{\kappa(\theta + t/\lambda) - \kappa(\theta)\}$$

and the moment generating function is

$$M(t;\theta,\lambda) = \exp[K(t;\theta,\lambda)] = \exp[\lambda\{\kappa(\theta + t/\lambda) - \kappa(\theta)\}].$$

Note that the domain of the generating functions $K_Y(t;\theta)$ and $M_Y(t;\theta)$ is the set $\{t \in \mathbb{R} : \mathbb{E}[e^{tY}] < \infty\}$. If the expectation $\mathbb{E}[e^{tY}]$ is not finite, then $K_Y(t;\theta)$ and $M_Y(t;\theta)$ are defined as infinity.

We can see that the mgf is a function of both the canonical parameter θ and the dispersion parameter λ. Since the mean μ can be written in terms of the canonical parameter θ via the mean mapping τ, the moment generating function of the ED family is also a function of the mean. It follows that the moment generating function for the ED family can be obtained directly from the cumulant function κ. This allows to define higher moments in terms of the mean.

Example 2.2.1. ***Poisson***

$Y \sim Poisson(\mu)$ has pdf:

$$f_Y(y;\mu) = \frac{e^{-\mu}\mu^y}{y!} = \frac{1}{y!}\exp\{y\ln(\mu) - \mu\}, \quad y \in \mathbb{N}.$$

Hence the Poisson distribution is a member of the natural exponential family with known dispersion parameter $\lambda = 1$, where:

$\theta = \ln(\mu)$ *is the canonical parameter,*

$\kappa(\theta) = \mu = e^\theta$ *is the cumulant function,*

$d(y;\mu) = 2\left\{y\ln\left(\frac{y}{\mu}\right) - y + \mu\right\}$ *is the unit deviance,*

$V(\mu) = \frac{2}{\frac{\partial^2}{\partial\mu^2}d(y;\mu)\big|_{\mu=y}} = \frac{2}{2/\mu} = \mu$ *is the unit variance function.*

The pdf can be rewritten in terms of the canonical parameter θ as

$$f_Y(y;\theta) = \frac{1}{y!}\exp\left\{y\theta - e^\theta\right\}, \quad y \in \mathbb{N}.$$

The mean and variance can be derived through the cumulant function:

$$\mathbb{E}(Y) = \dot{\kappa}(\theta) = e^{\theta} = \mu,$$

$$\mathbb{V}ar(Y) = a(\phi)\ddot{\kappa}(\theta) = e^{\theta} = \mu,$$

implying that $V(\mu) = \mu$ *is the variance function, as derived above.*

Since Y *is a member of the natural exponential family, the cumulant generating function can be derived through the cumulant function as:*

$$\begin{aligned} K(t;\theta) &= \kappa(\theta + t) - \kappa(\theta) = \exp(\theta + t) - \exp(\theta) \\ &= \exp(\theta)\{\exp(t) - 1\}, \quad t \in \mathbb{R}, \end{aligned}$$

which in terms of $\mu = e^{\theta}$ *is* $K(t;\mu) = \mu\{\exp(t) - 1\}$*. The moment generating function is thus*

$$M(t;\theta) = \exp\{K(t;\theta)\} = \exp\Big\{\exp(\theta)\{\exp(t) - 1\}\Big\}, \quad t \in \mathbb{R}$$

or, in terms of μ*,*

$$M(t;\mu) = \exp\Big\{\mu\{\exp(t) - 1\}\Big\}, \quad t \in \mathbb{R}.$$

Note that the Poisson distribution is also a member of the family of additive exponential dispersion models.

Example 2.2.2. ***Gamma***

$Y \sim$ *gamma*(α, β)*, has pdf:*

$$f_Y(y;\alpha,\beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y} = \exp\Big\{(\alpha - 1)\ln y - \beta y + \alpha \ln(\beta) - \ln\Gamma(\alpha)\Big\}, \quad y > 0.$$

Consider the following re-parametrization: $\lambda = \alpha$*,* $\mu = \frac{\alpha}{\beta}$*, then the probability density function can be written as:*

$$f_Y(y;\mu,\lambda) = \frac{1}{\Gamma(\lambda)} \left(\frac{\lambda}{\mu}\right)^{\lambda} y^{\lambda-1} \exp\left(-\frac{\lambda y}{\mu}\right), \quad y > 0.$$

For $\sigma^2 = \dfrac{1}{\lambda}$*, the density can again be rewritten in terms of the parameters* μ *and* σ^2 *as follows:*

$$f_Y(y;\mu,\sigma^2) = a(y;\sigma^2)\exp\left\{-\frac{1}{\sigma^2}\left(\frac{y}{\mu} - \ln\frac{y}{\mu} - 1\right)\right\}, \quad y > 0,$$

where the function

$$a(y;\sigma^2) = a(y;1/\lambda) = \frac{\lambda^\lambda e^{-\lambda}}{y\Gamma(\lambda)}, \quad y > 0,$$

which shows that the gamma distribution is a member of the reproductive exponential dispersion family with:

$\theta = -\dfrac{1}{\mu}$ *is the canonical parameter,*

$\kappa(\theta) = -\ln\left(\dfrac{1}{\mu}\right) = -\ln(-\theta)$ *is the cumulant function,*

$d(y;\mu) = 2\left(\dfrac{y}{\mu} - \ln\dfrac{y}{\mu} - 1\right)$ *is the unit deviance,*

$V(\mu) = \dfrac{2}{\dfrac{\partial^2}{\partial\mu^2} d(y;\mu)\Big|_{\mu=y}} = \dfrac{2}{2/\mu^2} = \mu^2$ *is the unit variance function.*

The mean and variance can be derived through the cumulant function:

$$\mathbb{E}(Y) = \dot{\kappa}(\theta) = (-1)\frac{-1}{-\theta} = \frac{-1}{\theta} = \mu = \frac{\alpha}{\beta},$$

$$\mathbb{V}ar(Y) = a(\phi)\ddot{\kappa}(\theta) = \frac{1}{\nu}\frac{1}{\theta^2} = \frac{1}{\alpha}\mu^2 = \frac{\alpha}{\beta^2},$$

implying that $V(\mu) = \mu^2$ is the variance function, as shown further above.

Since Y is a member of the reproductive exponential dispersion family, the cumulant generating function is derived through the cumulant function as:

$$\begin{aligned}
K(t;\theta,\lambda) &= \lambda\left\{\kappa(\theta + t/\lambda) - \kappa(\theta)\right\} = \lambda\left\{-\ln(-(\theta + t/\lambda)) - \big(-\ln(-\theta)\big)\right\} \\
&= -\lambda\left\{\ln(-(\theta + t/\lambda)) - \ln(-\theta)\right\} = -\lambda\ln\left(\frac{-(\theta + t/\lambda)}{-\theta}\right) \\
&= -\lambda\ln\left(\frac{\lambda\theta + t}{\lambda\theta}\right) = \ln\left\{\left(\frac{\lambda\theta + t}{\lambda\theta}\right)^{-\lambda}\right\} \\
&= \ln\left\{\left(1 + \frac{t}{\lambda\theta}\right)^{-\lambda}\right\}, \quad t > -\lambda\theta.
\end{aligned}$$

The moment generating function is then

$$M(t;\theta,\lambda) = \exp K(t;\theta,\lambda) = \left(1 + \frac{t}{\lambda\theta}\right)^{-\lambda}, \quad t > -\lambda\theta.$$

Tables 2.2 and 2.3 list some continuous and discrete members of the exponential dispersion models family.

Table 2.2: Some Continuous Exponential Dispersion Models

Distribution	$c(y;\lambda)$	$\kappa(\theta)$	$\tau(\theta)$	$V(\mu)$
Normal $\mathcal{N}(\mu,\sigma^2)$	$\sqrt{\frac{\lambda}{2\pi}}e^{\frac{-\lambda y^2}{2}}$	$\frac{\theta^2}{2}$	θ	1
Gamma $Ga(\mu,\sigma^2)$	$\frac{\lambda^\lambda y^{\lambda-1}}{\Gamma(\lambda)}$	$-\ln(-\theta)$	$\frac{-1}{\theta}$	μ^2

Table 2.3: Some Discrete Exponential Dispersion Models

Distribution	$c^*(y;\lambda)$	$\kappa(\theta)$	$\tau(\theta)$	$V(\mu)$
Poisson $Po(e^\theta)$	$\frac{1}{z!}$	e^θ	e^θ	μ
Binomial $Bi(\lambda,\mu)$	$\binom{\lambda}{z}$	$\ln(1+e^\theta)$	$\frac{e^\theta}{1+e^\theta}$	$\mu(1-\mu)$
Negative Binomial $Nb(p,\lambda)$	$\binom{\lambda+z-1}{z}$	$-\ln(1-e^\theta)$	$\frac{e^\theta}{1-e^\theta}$	$\mu(1+\mu)$

2.2.3 Tweedie Models

The Tweedie model is a subclass of the exponential dispersion family characterized as being closed under scale transformations. The Tweedie family, denoted by $Tw_p(\mu,\sigma^2)$ in terms of the reproductive exponential dispersion model, has unit variance function defined as:

$$V_p(\mu) = \mu^p, \quad \mu \in \Omega_p,$$

where the parameter p is referred to as the *shape parameter*. It follows that for $Y \sim Tw_p(\mu,\sigma^2)$, Y has mean μ and variance $\mathbb{V}\text{ar}(Y) = \sigma^2\mu^p$. Jørgensen (1997) shows that there exists exponential dispersion models with unit variance functions defined as the power function $V(\mu) = \mu^p$ for all $p \in \mathbb{R}$ except $0 < p < 1$.

The following theorem from Jørgensen characterizes the Tweedie class of distributions.

Theorem 2.2.1. *Let $ED(\mu,\sigma^2)$ denote a reproductive ED model such that $1 \in \Omega$ and*

and $V(1) = 1$*. If the model is closed with respect to scale transformations, such that there exists a function* $f : \mathbb{R}_+ \times \Lambda^{-1} \to \Lambda^{-1}$ *for which*

$$c \times ED(\mu, \sigma^2) = ED\{c\mu, f(c, \sigma^2)\}, \quad \forall c > 0,$$

then:

1. $ED(\mu, \sigma^2)$ *is a Tweedie model for some* $p \in \mathbb{R}$*,*

2. $f(c, \sigma^2) = c^{2-p}\sigma^2$*,*

3. *The mean domain is* $\Omega = \mathbb{R}$ *for* $p = 0$ *and* $\Omega = \mathbb{R}_+$ *for* $p \neq 0$*,*

4. *The model is infinitely divisible.*

Table 2.4 provides a summary of the distributions that belong to the Tweedie subclass of exponential dispersion models.

The Tweedie ED models are of particular interest for the analysis of the aggregate loss cost in the independent model as described in Chapter 1. For $N \sim Poisson$ and $Y_i \sim gamma$ the aggregate claims S is a Compound Poisson-gamma and follows a Tweedie distribution with $1 < p < 2$.

Table 2.4: Tweedie Exponential Dispersion Models

Distribution	**p**	**Support**	**Ω Mean Domain**	**Θ Canonical Parameter Domain**
Extreme stable	$p < 0$	$\mathbb{R}$	$\mathbb{R}_+$	$\mathbb{R}_0$
Normal	$p = 0$	$\mathbb{R}$	$\mathbb{R}$	$\mathbb{R}$
N/A	$0 < p < 1$	-	$\mathbb{R}_+$	$\mathbb{R}_0$
Poisson	$p = 1$	$\mathbb{N}_0$	$\mathbb{R}_+$	$\mathbb{R}$
Compound Poisson	$1 < p < 2$	$\mathbb{R}_0$	$\mathbb{R}_+$	$\mathbb{R}_-$
Gamma	$p = 2$	$\mathbb{R}_+$	$\mathbb{R}_+$	$\mathbb{R}_-$
Positive stable	$2 < p < 3$	$\mathbb{R}_+$	$\mathbb{R}_+$	$-\mathbb{R}_0$
Inverse Gaussian	$p = 3$	$\mathbb{R}_+$	$\mathbb{R}_+$	$-\mathbb{R}_0$
Positive stable	$p > 3$	$\mathbb{R}_+$	$\mathbb{R}_+$	$-\mathbb{R}_0$
Extreme stable	$p = \infty$	$\mathbb{R}$	$\mathbb{R}$	$\mathbb{R}_-$

Note: $-\mathbb{R}_0 = (-\infty, 0]$

Chapter 3

Generalized Linear Models

Classical linear models attempt to fit a model to the mean response of some observed variable Y in the form of a linear predictor. Generalized linear models (GLMs) are an extension to this approach. The GLM framework allows for greater flexibility in modelling observations in several ways. Firstly, rather than writing the mean as a simple linear function of covariates and regression parameters, GLMs allow for a non-linear function of the mean to be modelled in terms of a linear predictor. Secondly, classical linear regression assumes that the error terms are normally distributed with mean zero and constant variance. Generalized linear models relax this assumption by allowing the error distribution to be a member of the exponential dispersion family, thus greatly broadening the set of distributions that can be fit to the data. Moreover, classical linear models treat the mean and variance structure of the response variable separately. Generalized linear models, on the other hand, allow for a mean-variance relation which is inherent in the exponential dispersion models density structure. Thus, in modelling the mean through a GLM, we are also indirectly modelling the variance.

GLMs are of particular importance for insurance data as this framework allows to model non-normal observations. For instance, in the aggregate claims model, both the frequency and severity components do not follow a normal distribution: claim counts are positive integer valued random quantities (e.g. Poisson distributed observations) and claim amounts can take on positive, continuous, right-skewed values (e.g. gamma distributed observations).

Recall the representation in McCullagh and Nelder (1989) of the exponential dispersion family given in Chapter 2:

$$f_Y(y;\theta,\phi) = \exp\left[\frac{y\theta - \kappa(\theta)}{a(\phi)} + C(y,\phi)\right].$$

The GLM framework assumes that the response variable Y is a member of the $ED(\mu, \phi)$ family. In both the classical linear model and the generalized linear model, the goal is to model the mean response, conditional on a given set of covariates. That is, for a $p \times 1$ vector of known covariates $\mathbf{X} = (x_1, ..., x_p)^\top$, the model will define $\mathbb{E}[Y|\mathbf{X}]$ in terms of a linear predictor η such that:

$$\eta = \mathbf{X}^\top \boldsymbol{\beta} = \sum_{k=1}^{p} x_k \beta_k,$$

where $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown regression parameters.

As previously mentioned, the classical linear model assumes normal observations and models the mean response as $\mathbb{E}[Y|\mathbf{X}] = \mu = \eta = \mathbf{X}^\top \boldsymbol{\beta}$. GLMs no longer restrict the responses to be linearly associated to the predictor η, but rather allow for a function of the mean to be modelled in terms of the linear predictor. For a link function g, the GLM formulation is $g\{\mathbb{E}[Y|\mathbf{X}]\} = g\{\mu\} = \eta = \mathbf{X}^\top \boldsymbol{\beta}$. This added flexibility allows GLMs to fit a variety of data, particularly insurance data.

3.1 The Model

The GLM framework, as discussed in McCullagh and Nelder(1989), assumes that the observations $y_1, ..., y_n$ are independent and that $Y_i \sim ED(\mu_i, \phi)$, so the mean varies with each observation, while the dispersion is assumed the same for all observations, but unknown. The model expresses the conditional mean of the response Y, given the corresponding vector of covariates $\mathbf{X}$ via a known link function g as:

$$g\{\mathbb{E}[Y \mid \mathbf{X}]\} = g\{\mu\} = \eta = \mathbf{X}^\top \boldsymbol{\beta} = \sum_{k=1}^{p} x_k \beta_k,$$

where the link g is any monotonic differentiable function. Thus, the mean is a function of the linear predictor:

$$\mu = \mathbb{E}[Y \mid \mathbf{X}] = g^{-1}(\eta) = g^{-1}(\mathbf{X}^\top \boldsymbol{\beta}).$$

The goal of the GLM approach is to estimate the regression parameters $\boldsymbol{\beta}$ to ultimately predict the response variable Y.

3.1.1 Maximum Likelihood Estimation for the ED Family

Following the notation of McCullagh and Nelder (1989) for the exponential dispersion family, we have that for independent observations $(y_i, \mathbf{x}_i)$; for $i = 1, ..., n$, where the y_i are independent realizations of $Y_i \sim ED(\mu_i, \phi)$, the likelihood function for the canonical parameter vector $\boldsymbol{\theta} = (\theta_1, ..., \theta_n)^\top$ and the dispersion parameter ϕ is given as:

$$L(\boldsymbol{\theta}, \phi; y) = \prod_{i=1}^{n} f_Y(y_i; \theta_i, \phi) = \prod_{i=1}^{n} \exp\left[\frac{y_i\theta_i - \kappa(\theta_i)}{a_i(\phi)} + C(y_i, \phi)\right].$$

It then follows that the log-likelihood function is

$$\begin{aligned} \ell(\boldsymbol{\theta}, \phi; y) &= \ln L(\boldsymbol{\theta}, \phi; y) = \sum_{i=1}^{n} f_Y(y_i; \theta_i, \phi) = \sum_{i=1}^{n} \left[\frac{y_i\theta_i - \kappa(\theta_i)}{a_i(\phi)} + C(y_i, \phi)\right] \\ &= \sum_{i=1}^{n} \left\{\frac{y_i\theta_i - \kappa(\theta_i)}{a_i(\phi)}\right\} + \sum_{i=1}^{n} C(y_i, \phi). \end{aligned}$$

As mentioned in Chapter 2, in taking derivatives of $\ell(\boldsymbol{\theta}, \phi; y)$ with respect to the $\theta_i' s$, the dispersion term ϕ factors out for some $a_i(\phi)$ functions so that the estimation of the canonical parameters θ_i can be carried out separately from that for the parameter ϕ.

The score functions are then defined as the system of partial derivatives of the log-likelihood:

$$s(\boldsymbol{\theta}; y) = \frac{\partial}{\partial \boldsymbol{\theta}} \ell(\boldsymbol{\theta}, \phi; y).$$

The maximum likelihood estimate $\hat{\boldsymbol{\theta}}$ is then found as the solution to the system of equations $s(\boldsymbol{\theta}; y) = 0$.

For a given observation y_i, write $\theta_i = \theta$. Then the canonical parameter θ is related to the mean μ through the mean mapping $\tau(\theta) = \dot{\kappa}(\theta) = \mu$. Thus, by the invariance property of maximum likelihood estimators, the MLE $\hat{\theta}$ will also give the MLE for the mean as $\hat{\mu} = \tau(\hat{\theta})$.

The goal of the GLM is to find the maximum likelihood estimates for the regression parameters $\boldsymbol{\beta}$. The model framework relates the mean μ to the linear predictor $\eta = \mathbf{X}^\top\boldsymbol{\beta}$ through the link function. Thus we have the following relation:

$$\mu = \tau(\theta) = g^{-1}(\eta) = g^{-1}\{\mathbf{X}^\top\boldsymbol{\beta}\},$$

which implies that

$$\theta = \tau^{-1}(\mu) = \tau^{-1}\left[g^{-1}(\eta)\right] = \tau^{-1}\left[g^{-1}(\mathbf{X}^\top\boldsymbol{\beta})\right], \text{ and}$$

$$\kappa(\theta) = \kappa\left\{\tau^{-1}\left[g^{-1}(\mathbf{X}^\top\boldsymbol{\beta})\right]\right\}.$$

The log-likelihood function in terms of the regression parameters then becomes:

$$\begin{aligned}\ell(\boldsymbol{\beta};\phi,\mathbf{y}) &= \sum_{i=1}^{n}\frac{\{y_i\theta_i - \kappa(\theta_i)\}}{a_i(\phi)} + \sum_{i=1}^{n}C(y_i,\phi)\\ &= \sum_{i=1}^{n}\frac{\{y_i\tau^{-1}\left[g^{-1}(\mathbf{x}_i\boldsymbol{\beta})\right] - \kappa(\tau^{-1}\left[g^{-1}(\mathbf{x}_i\boldsymbol{\beta})\right])\}}{a_i(\phi)} + \sum_{i=1}^{n}C(y_i,\phi).\end{aligned}$$

Then the MLE for regression parameter β_j associated with the covariate x_{ij} is the solution to the score equation:

$$\frac{\partial}{\partial\beta_j}\ell(\boldsymbol{\beta};\phi,\mathbf{y}) = 0.$$

The GLM with link function g sets $g(\mu_i) = g(\dot{\kappa}(\theta_i)) = \eta_i = x_i\beta$. Recall from Chapter 2 that $\mu_i = \tau(\theta_i) \Leftrightarrow \theta_i = \tau^{-1}(\mu_i)$. It follows that $\frac{\partial}{\partial\theta_i}\tau(\theta_i) = \frac{\partial\mu_i}{\partial\theta_i}$ so that $\frac{\partial\mu_i}{\partial\theta_i} = \dot{\tau}(\theta_i) = \dot{\tau}(\tau^{-1}(\mu_i))$. We also showed in Chapter 2 that $V(\mu_i) = \ddot{\kappa}(\theta_i) = \dot{\tau}(\theta_i)$ so that $V(\mu_i) = \dot{\tau}(\tau^{-1}(\mu_i))$ which is also equal to $V(\mu_i) = \frac{\partial\mu_i}{\partial\theta_i}$. Thus we have that:

$$\frac{\partial g(\mu_i)}{\partial\beta_j} = \frac{\partial g(\mu_i)}{\partial\mu_i}\frac{\partial\mu_i}{\partial\theta_i}\frac{\partial\theta_i}{\partial\beta_j} = \dot{g}(\mu_i)V(\mu_i)\frac{\partial\theta_i}{\partial\beta_j},$$

which implies that

$$\frac{\partial\theta_i}{\partial\beta_j} = \frac{1}{\dot{g}(\mu_i)V(\mu_i)}\frac{\partial g(\mu_i)}{\partial\beta_j} = \frac{1}{\dot{g}(\mu_i)V(\mu_i)}\frac{\partial\eta_i}{\partial\beta_j} = \frac{x_{ij}}{\dot{g}(\mu_i)V(\mu_i)}.$$

The resulting score equation for parameter β_j can then be simplified:

$$
\begin{aligned}
s(\beta_j;\phi,\mathbf{y}) &= \frac{\partial}{\partial\beta_j}\ell(\boldsymbol{\beta};\mathbf{y}) = \frac{\partial}{\partial\beta_j}\left[\sum_{i=1}^{n}\frac{1}{a_i(\phi)}\{y_i\theta_i-\kappa(\theta_i)\}+\sum_{i=1}^{n}C(y_i,\phi)\right] \\
&= \sum_{i=1}^{n}\frac{1}{a_i(\phi)}\left(y_i\frac{\partial\theta_i}{\partial\beta_j}-\frac{\partial\kappa(\theta_i)}{\partial\beta_j}\right) = \sum_{i=1}^{n}\frac{1}{a_i(\phi)}\left(y_i-\frac{\partial\kappa(\theta_i)}{\partial\theta_i}\right)\frac{\partial\theta_i}{\partial\beta_j} \\
&= \sum_{i=1}^{n}\frac{1}{a_i(\phi)}\left(y_i-\dot{\kappa}(\theta_i)\right)\frac{x_{ij}}{\dot{g}(\mu_i)V(\mu_i)} \\
&= \sum_{i=1}^{n}\frac{(y_i-\mu_i)}{a_i(\phi)V(\mu_i)}\frac{x_{ij}}{\dot{g}(\mu_i)}. \qquad (3.1)
\end{aligned}
$$

The MLE for the regression parameters is then found as the solution to

$$
s(\beta_j;\phi,\mathbf{y}) = \sum_{i=1}^{n}\frac{(y_i-\mu_i)}{a(\phi_i)V(\mu_i)}\frac{x_{ij}}{\dot{g}(\mu_i)} = 0, \qquad j=1,...,p.
$$

If we now suppose that the observations $\mathbf{y}=(y_1,...,y_n)^\top$ have known prior weights $w_1,...,w_n$ such that the function $a_i(\phi)$ has the form $a_i(\phi)=\phi/w_i$, then the score function can be further simplified:

$$
s(\beta_j;\phi,\mathbf{y}) = \sum_{i=1}^{n}\frac{w_i(y_i-\mu_i)}{\phi V(\mu_i)}\frac{x_{ij}}{\dot{g}(\mu_i)} = 0.
$$

The dispersion parameter ϕ now factors out so that the equation becomes:

$$
s(\beta_j;\mathbf{y}) = \sum_{i=1}^{n}\frac{w_i(y_i-\mu_i)}{V(\mu_i)}\frac{x_{ij}}{\dot{g}(\mu_i)} = 0. \qquad (3.2)
$$

Thus, $\hat{\beta}_j$ is such that $s(\beta_j;\mathbf{y})=0$, and $\hat{\boldsymbol{\beta}}$ is the solution to the $p\times 1$ system of equations $s(\boldsymbol{\beta};\phi,\mathbf{y})=0$.

3.1.2 Link Function

In the classical linear regression model, the link function is the identity and hence $\mu=\eta$. That is, the mean is a linear function of the regression parameters and covariates. The linear predictor η can lie anywhere on the real line, i.e. $-\infty<\eta<\infty$. Consequently, in classical linear regression, the mean model can map μ anywhere in the interval $(-\infty,\infty)$. GLMs provide an improved modelling framework to that of the linear model in the sense that the range of μ is not necessarily the interval $(-\infty, \infty)$. The link function essentially allows to define how the expected response will be mapped from the linear predictor scale to the mean scale through its inverse: $g^{-1} : \eta \to \mu$. Consequently, for a particular choice of link function, one can ensure that the mean is mapped to the proper mean range Ω. For example, if the response variable is assumed to follow a gamma distribution, Y has support $(0, \infty)$ and so the mean $\mathbb{E}[Y]$ must also fall in $(0, \infty)$. Clearly, the choice of link function used for modelling gamma responses must ensure that $g^{-1} : (-\infty, \infty) \to (0, \infty)$. It then follows that the identity link is not an appropriate choice whereas the log link, with inverse being exponential, would properly map the mean.

Canonical Link

A convenient choice of link function is such that the linear predictor is set equal to the canonical parameter, so that $\eta = \theta$. This choice of link function is referred to as the *canonical link function*. Recall that in the GLM framework the response variable is a member of the exponential dispersion family with parameters (θ, ϕ), where the mean function $\tau(\theta) = \dot{\kappa}(\theta) = \mu$ relates the mean to the canonical parameter. Thus, using the canonical link function, g_c, in the GLM gives:

$$g_c\{\mathbb{E}[Y \mid \mathbf{X}]\} = g_c\{\mu\} = \eta = \theta = \tau^{-1}(\mu),$$

which implies that

$$\mu = g_c^{-1}(\eta) = \tau(\eta).$$

Some examples of common EF models and their canonical link functions are provided in Table 3.1.

Using the canonical link function simplifies the estimation of the regression parameters $\boldsymbol{\beta}$ since the log-likelihood function then becomes a simplified function in

Table 3.1: Canonical Link Function of Some Common ED Models

Distribution	**Canonical Parameter**	**Canonical Link Function**	**Link Name**
Normal	$\theta = \mu$	$\mu = \eta$	identity
Poisson	$\theta = \ln\{\mu\}$	$\ln\{\mu\} = \eta$	log
Binomial	$\theta = \ln\left(\dfrac{\mu}{1-\mu}\right)$	$\ln\left(\dfrac{\mu}{1-\mu}\right) = \eta$	logit
Gamma	$\theta = \dfrac{-1}{\mu}$	$\dfrac{-1}{\mu} = \eta$	reciprocal
Inverse Gaussian	$\theta = \dfrac{-1}{2\mu^2}$	$\dfrac{-1}{2\mu^2} = \eta$	inverse squared

terms of β:

$$\begin{aligned}
\ell(\boldsymbol{\beta};\phi,\mathbf{y}) &= \sum_{i=1}^{n}\frac{1}{a_i(\phi)}\{y_i\theta_i - \kappa(\theta_i)\} + \sum_{i=1}^{n} C(y_i,\phi) \\
&= \sum_{i=1}^{n}\frac{1}{a_i(\phi)}\{y_i\eta_i - \kappa(\eta_i)\} + \sum_{i=1}^{n} C(y_i,\phi) \\
&= \sum_{i=1}^{n}\frac{1}{a_i(\phi)}\{y_i\mathbf{x}_i^\top\boldsymbol{\beta} - \kappa(\mathbf{x}_i^\top\boldsymbol{\beta})\} + \sum_{i=1}^{n} C(y_i,\phi).
\end{aligned}$$

The score function is then:

$$s(\boldsymbol{\beta};\phi,\mathbf{y}) = \frac{\partial}{\partial\boldsymbol{\beta}}\ell(\boldsymbol{\beta};\phi,\mathbf{y}) = \sum_{i=1}^{n}\frac{1}{a_i(\phi)}\{y_i\mathbf{x}_i - \dot{\kappa}(\mathbf{x}_i^\top\boldsymbol{\beta})\} = \mathbf{0}.$$

The maximum likelihood estimates for the regression parameters are found as the solution to the score equation. Thus, the MLE for the regression parameter associated with covariate x_{ij} is $\hat{\beta}_j$ and is the solution to the score equation:

$$\begin{aligned}
s(\beta_j;\phi,\mathbf{y}) &= \frac{\partial}{\partial\beta_j}\ell(\boldsymbol{\beta};\phi,\mathbf{y}) = \frac{\partial}{\partial\beta_j}\Big\{\sum_{i=1}^{n}\frac{1}{a_i(\phi)}\{y_i\left(\mathbf{x}_i^\top\boldsymbol{\beta}\right) - \kappa(\mathbf{x}_i^\top\boldsymbol{\beta})\} + \sum_{i=1}^{n} C(y_i,\phi)\Big\} \\
&= \sum_{i=1}^{n}\frac{1}{a_i(\phi)}\left\{y_i x_{ij} - \dot{\kappa}(\mathbf{x}_i^\top\boldsymbol{\beta})x_{ij}\right\} = \sum_{i=1}^{n}\frac{1}{a_i(\phi)}x_{ij}\left\{y_i - \dot{\kappa}(\mathbf{x}_i^\top\boldsymbol{\beta})\right\}.
\end{aligned}$$

Note that $\dot{\kappa}(\mathbf{x}_i^\top\boldsymbol{\beta}) = \dot{\kappa}(\eta_i) = \dot{\kappa}(\theta_i) = \tau(\theta_i) = \mu_i$. Thus we have:

$$s(\beta_j;\phi,\mathbf{y}) = \sum_{i=1}^{n}\frac{1}{a_i(\phi)}x_{ij}\{y_i - \mu_i\} = 0.$$

If we write $a_i(\phi) = \phi/w_i$, as before, for weights $w_1, ..., w_n$, then we have:

$$s(\beta_j; \phi, \mathbf{y}) = \sum_{i=1}^{n} \frac{1}{\phi/w_i} x_{ij} \{y_i - \mu_i\} = \sum_{i=1}^{n} \frac{1}{\phi} w_i (y_i - \mu_i) x_{ij}.$$

The dispersion parameter ϕ can then be factored out so that the score equation is:

$$s(\beta_j; \mathbf{y}) = \sum_{i=1}^{n} w_i (y_i - \mu_i) x_{ij},$$

and the maximum likelihood estimates are then the solution to $s(\beta_j; \mathbf{y}) = 0$ for $j = 1, ..., p$.

Example 3.1.1. ***Poisson***

Consider a GLM with Poisson responses $Y_i \sim Poisson(\mu_i)$ and weights $w_i = 1$. Then using the canonical link (log link) we have that:

The likelihood is:

$$L(\boldsymbol{\beta}; \boldsymbol{y}) = \prod_{i=1}^{n} \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!}.$$

The log-likelihood is:

$$\ell(\boldsymbol{\beta}; \boldsymbol{y}) = \sum_{i=1}^{n} \{-\mu_i + y_i \ln(\mu_i)\} + \sum_{i=1}^{n} - \ln(y_i!).$$

The score equation for $\beta_1, ..., \beta_p$ is:

$$\begin{aligned}
s(\beta_j; \boldsymbol{y}) &= \frac{\partial}{\partial \beta_j} \ell(\boldsymbol{\beta}; \boldsymbol{y}) \\
&= \sum_{i=1}^{n} \left\{ \frac{-\partial}{\partial \beta_j} \mu_i + \frac{\partial}{\partial \beta_j} y_i \ln(\mu_i) \right\} = \sum_{i=1}^{n} \left\{ \frac{-\partial}{\partial \beta_j} \mu_i + \frac{y_i}{\mu_i} \frac{\partial}{\partial \beta_j} \mu_i \right\} \\
&= \sum_{i=1}^{n} \left\{ \frac{-\partial}{\partial \beta_j} \exp(\boldsymbol{x}_i^\top \boldsymbol{\beta}) + \frac{y_i}{\mu_i} \frac{\partial}{\partial \beta_j} \exp(\boldsymbol{x}_i^\top \boldsymbol{\beta}) \right\} \\
&= \sum_{i=1}^{n} \left\{ -x_{ij} \exp(\boldsymbol{x}_i^\top \boldsymbol{\beta}) + \frac{y_i}{\mu_i} x_{ij} \exp(\boldsymbol{x}_i^\top \boldsymbol{\beta}) \right\} \\
&= \sum_{i=1}^{n} \left\{ -x_{ij} \mu_i + \frac{y_i}{\mu_i} x_{ij} \mu_i \right\} = \sum_{i=1}^{n} \{-x_{ij} \mu_i + y_i x_{ij}\} \\
&= \sum_{i=1}^{n} x_{ij} \{\mu_i - y_i\} = 0.
\end{aligned}$$

3.1.3 Asymptotic Results for MLEs

Suppose that the data $\mathbf{y}$ are generated from the true distribution with $p \times 1$ vector of canonical parameters $\boldsymbol{\theta}_0$, denoted $f_Y(y_i; \theta_{0i})$. As mentioned in Chapter 2, we have that:

$$\mathbb{E}_Y[\dot{\ell}(\boldsymbol{\theta}_0; \mathbf{Y})] = \mathbb{E}_Y[s(\boldsymbol{\theta}_0; \mathbf{Y})] = 0.$$

Furthermore, we have the relation:

$$\mathbb{E}_Y[s(\boldsymbol{\theta}_0; \mathbf{Y}) s(\boldsymbol{\theta}_0; \mathbf{Y})^\top] = -\mathbb{E}[\Psi(\boldsymbol{\theta}_0; \mathbf{Y})], \tag{3.3}$$

where $\Psi(\boldsymbol{\theta}_0; \mathbf{Y})$ is the matrix of first derivatives of the score equation, or, equivalently, the matrix of second derivatives of the log-likelihood function, with (j, k)th element equal to:

$$\frac{\partial^2 \ell(\boldsymbol{\theta}_0; \mathbf{Y})}{\partial \theta_j \partial \theta_k}, \qquad j, k = 1, ..., p.$$

The $p \times p$ matrix defined by equation (3.3), denoted $I(\boldsymbol{\theta}_0)$, is referred to as the Fisher Information. Under certain regularity conditions, this matrix is symmetric and positive definite.

For a maximum likelihood estimate based on n observations, denoted $\hat{\theta}$, we have the following key results:

i) Consistency: As $n \to \infty$, $\hat{\theta} \xrightarrow{p} \theta_0$.

ii) Asymptotic Normality:

   ⋄ As $n \to \infty$, $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, \{I(\boldsymbol{\theta}_0)\}^{-1})$,

   ⋄ As $n \to \infty$, $\frac{1}{\sqrt{n}} \dot{\ell}(\boldsymbol{\theta}_0; \mathbf{y}) \xrightarrow{d} \mathcal{N}(0, I(\boldsymbol{\theta}_0))$.

iii) Score Test: as $n \to \infty$, $\dot{\ell}(\boldsymbol{\theta}_0; \mathbf{y})^\top \{I(\boldsymbol{\theta}_0)\} \dot{\ell}(\boldsymbol{\theta}_0; \mathbf{y}) \sim \chi^2_p$.

iv) Observed Information: It is often necessary to find an estimate for the Fisher Information $I(\theta_0)$. We can estimate $I(\theta_0)$ by $\hat{I}_n$, where the quantity $\hat{I}_n$ is referred to as the *observed information*. Some estimates for $I(\theta_0)$ include :

   ⋄ $\hat{I}_n = I(\hat{\theta})$,

◇ $\hat{I}_n = \frac{1}{n}\sum_{i=1}^n s(\hat{\boldsymbol{\theta}}; y_i)s(\hat{\boldsymbol{\theta}}; y_i)^\top$,

◇ $\hat{I}_n = -\frac{1}{n}\sum_{i=1}^n \Psi(\hat{\boldsymbol{\theta}}; y_i)$.

3.1.4 Goodness of Fit

The goal of modelling data is to obtain fitted values, $\hat{\boldsymbol{\mu}}$, for the mean of the response values $\mathbf{y}$. Generally, the fitted values will not exactly coincide with the actual data values. The significance of the discrepancy between the actual values $\mathbf{y}$ and the estimated expected values $\hat{\boldsymbol{\mu}}$ can be measured and analysed through the *deviance*.

Denote $\hat{\boldsymbol{\mu}}_{\boldsymbol{i}} = \hat{\boldsymbol{\mu}}_{\mathbf{0}}$ as the fitted values obtained in the null model, i.e. the simplest model that contains only an intercept, and let $\bar{\boldsymbol{\theta}}$ be the estimated canonical parameter associated with the null model estimates. Let $\hat{\boldsymbol{\mu}}_{\boldsymbol{i}} = \mathbf{y}_i$ denote the fitted values under the full model, i.e. the most complex model, and $\tilde{\boldsymbol{\theta}}$ denote the resulting canonical parameter. Note that the full model is fully saturated so that the fitted values are exactly equal to the data values. Finally, denote the intermediate model fitted values by $\hat{\boldsymbol{\mu}}_{\boldsymbol{i}} = \boldsymbol{\mu}(\mathbf{X}; \boldsymbol{\beta})$ and the estimated canonical parameters by $\hat{\boldsymbol{\theta}}$. The discrepancy between the fitted values and the data values can then be measured as twice the difference between the log-likelihood under the full and intermediate, or fitted, models. If we again suppose that $a_i(\phi) = \phi/w_i$, then we have that:

$$\begin{aligned} 2\{\ell(\tilde{\boldsymbol{\theta}}; \phi, \mathbf{y}) - \ell(\hat{\boldsymbol{\theta}}; \phi, \mathbf{y})\} &= \sum_{i=1}^n \frac{w_i}{\phi}\left\{y_i(\tilde{\theta}_i - \hat{\theta}_i) - \left(b(\tilde{\theta}_i) - b(\hat{\theta}_i)\right)\right\} \\ &= \frac{D(\mathbf{y}, \hat{\boldsymbol{\mu}})}{\phi} = D^*(\mathbf{y}, \hat{\boldsymbol{\mu}}). \end{aligned} \tag{3.4}$$

where $D(\mathbf{y}, \hat{\boldsymbol{\mu}})$ is referred to as the *deviance* for the intermediate model defined by $\hat{\boldsymbol{\mu}}$ and $D^*(\mathbf{y}, \hat{\boldsymbol{\mu}})$ is the *scaled deviance*.

The deviance measures the discrepancy in the model fit between the full model and the fitted model. Suppose that the fitted model has p parameters. Then, according to the likelihood ratio theorem, we have that

$$\frac{D(\mathbf{y}, \hat{\boldsymbol{\mu}})}{\phi} = 2\{\ell(\tilde{\boldsymbol{\theta}}; \phi, \mathbf{y}) - \ell(\hat{\boldsymbol{\theta}}; \phi, \mathbf{y})\} \sim \chi^2_{n-p}.$$

Thus, in expectation, we have that $\mathbb{E}\left[\frac{D(\mathbf{y}, \hat{\boldsymbol{\mu}})}{\phi}\right] = n - p$.

Table 3.2: Deviances of Some Common ED Models

Distribution	**Deviance $\boldsymbol{D(y,\hat{\mu})}$**
Normal	$\sum_{i=1}^{n}(y_i - \hat{\mu}_i)$
Poisson	$2\sum_{i=1}^{n}\{y_i \ln(y_i/\hat{\mu}_i) - (y_i - \hat{\mu}_i)\}$
Binomial	$2\sum_{i=1}^{n}\{y_i \ln(y_i/\hat{\mu}_i) + (m - y_i)\ln[(m - y_i)/(m - \hat{\mu}_i)]\}$
Gamma	$2\sum_{i=1}^{n}\{-\ln(y_i/\hat{\mu}_i) + (y_i - \hat{\mu}_i)/\hat{\mu}_i\}$
Inverse Gaussian	$2\sum_{i=1}^{n}(y_i - \hat{\mu}_i)^2/(\hat{\mu}_i^2 y_i)$

The deviance can also be used to compare nested models. Supposed we have that model A with p_A parameters is nested in model B with p_B parameters, with $p_B > p_A$, so that model A can be obtained from model B by applying some equality constraints, i.e. in fixing $p_B - p_A$ parameters of model B we retrieve model A. Then if we wish to test whether model A is an adequate simplification of model B, we can use the difference of the scaled deviances from these models as the test statistic:

$$\frac{D(\mathbf{y},\hat{\boldsymbol{\mu}}_{\boldsymbol{A}}) - D(\mathbf{y},\hat{\boldsymbol{\mu}}_{\boldsymbol{B}})}{\phi} = \frac{2\{\ell^{(B)}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{B}};\phi,\mathbf{y}) - \ell^{(A)}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{A}};\phi,\mathbf{y})\}}{\phi} \sim \chi^2_{p_B - p_A}.$$

To then test the hypothesis that these $p_B - p_A$ parameters of model B should be zero, we can compare the above test statistic to, say, the 95th percentile of the Chi-Square distribution with degrees of freedom equal to $p_B - p_A$. If the test statistic is larger than the percentile, we can then conclude that model A is not an adequate simplification of the more complex model B, that is, we reject the hypothesis that the parameters should be set to zero.

Note that the analysis of deviance, as described above, relies on the χ^2 approximation for the difference of deviances of nested models. McCullagh and Nelder (1989) point out that these approximations are, in general, not very good, even as $n \to \infty$. Nonetheless, the analysis of deviances provides a good approach for selecting the best model to explain the data under question.

3.1.5 Residuals

Residuals provide another way to assess the adequacy of a model. In the classical linear regression model, we define the residual vector as $\mathbf{y} - \hat{\boldsymbol{\mu}}$. Plotting the residuals versus the fitted values $\hat{\boldsymbol{\mu}}$ should produce a band around zero. These residuals are used to assess the local fit of the model. In the case of generalized linear models, the definition of residuals must be extended as we must now assume a non-Gaussian probability model. There are several forms of generalized residuals, namely the Pearson residual, Anscombe residual and deviance residual. Each provide different advantages in assessing the adequacy of a model fit.

The **Pearson residual** is defined as:

$$r_P = \frac{y - \hat{\mu}}{\sqrt{V(\hat{\mu})}}.$$

These residuals will have mean zero and unit variance if the model is correct, so that r_P is standardized. Note that while the Pearson residual is standardized, it is not normalized. Although this form of residual has zero-mean, the distribution of r_P itself may be highly skewed for non-normal distributions.

The Anscombe residual addresses the issue of skewness by considering a transformation of the data $A(y)$. That is, by choosing a function $A(\cdot)$ such that $A(Y)$ is approximately normally distributed, the skewness will essentially be removed. It can be shown that for a member of the exponential dispersion family, the optimal transformation is defined as

$$A(t) = \int_{-\infty}^{t} \frac{1}{V^{1/3}(s)} ds,$$

where $V(\cdot)$ is the variance function of Y as defined in Chapter 2. The variance of $A(Y)$ can be approximated as $\mathbb{V}\text{ar}\,(A(Y)) \approx \{\dot{A}(\mu)\}^2 V(\mu)$. The **Anscombe residual** is then defined as:

$$r_A = \frac{A(y) - A(\hat{\mu})}{\sqrt{\dot{A}(\hat{\mu})\}^2 V(\hat{\mu})}},$$

where r_A is standardized and the skewness is removed.

Finally, we can define a third type of residual based on the model deviance. Recall the definition of deviance for the GLM; for $a_i(\phi) = \phi/w_i$ we have that:

$$\begin{aligned}\frac{D(\mathbf{y}, \hat{\boldsymbol{\mu}})}{\phi} &= 2\{\ell(\tilde{\boldsymbol{\theta}}; \phi, \mathbf{y}) - \ell(\hat{\boldsymbol{\theta}}; \phi, \mathbf{y})\} \\ &= \frac{2}{\phi}\sum_{i=1}^{n} w_i \left\{ y_i(\tilde{\theta}_i - \hat{\theta}_i) - \left(b(\tilde{\theta}_i) - b(\hat{\theta}_i) \right) \right\} \\ &= \frac{1}{\phi}\sum_{i=1}^{n} D_i,\end{aligned}$$

where $D_i = 2w_i \left\{ y_i(\tilde{\theta}_i - \hat{\theta}_i) - \left(b(\tilde{\theta}_i) - b(\hat{\theta}_i) \right) \right\}$. The quantity D_i essentially measures the discrepancy contribution of datum i. We can then define the **deviance residual** as follows:

$$r_D = \text{sign}(y_i - \hat{\mu}_i)\sqrt{D_i}.$$

For more detail on the GLM residuals, see McCullagh and Nelder (1989) and the references therein.

3.2 GLMs for Insurance Data

Generalized linear models have become an important modelling tool for the insurance industry as the flexibility of the GLM framework and use of exponential dispersion models allow for a better representation of insurance data. As previously mentioned, an insurer must set premiums in accordance with the expected total claim cost. In Chapter 2, we showed that both the frequency and severity components could be modelled in terms of an ED model. Moreover, in the special case of a Compound Poisson model assumption, the total loss follows a Tweedie distribution, $S \sim Tw_p(\mu, \sigma^2)$, which is also a member of the ED family. Thus, GLMs can be used to model the marginal means of the claim counts and claim amounts respectively, or even to directly model the expected loss cost using a Tweedie distribution. Note that both of the approaches mentioned here allude to the aggregate claims model under the independence assumption. As we will see in Chapter 5, this modelling approach needs to be modified in order to allow for dependence between the claim frequency and claim severity.

In the analysis of insurance data, covariates are often referred to as *rating variables* and are used to characterize the risk ensued by the insured individual. For car insurance data, covariates are characteristics related to both the driver and vehicle, and may include the driver's age and gender, the distance driven, marital status, vehicle model, etc. For home insurance, common rating variables include location, construction year, amount of insurance, etc.

It is common practice when modelling the mean response in an insurance setting to use a log-link in the GLM. This is done as it yields a multiplicative rating structure. That is, if

$$\ln(\mu) = \sum_{i=1}^{p} x_i\beta_i$$

then, on the mean scale, we have that

$$\mu = \prod_{i=1}^{n} \exp(x_i\beta_i) = \prod_{i=1}^{n} \psi_i.$$

These multiplicative factors associated with each rating variable, i.e. ψ_i, are referred to as differentials or relativities. It follows that the expected loss cost for a policyholder can then be determined by multiplying the base rate (i.e. the intercept $\exp(\beta_0) = \psi_0$) by the differentials that correspond to the individual's rating variable levels.

Chapter 4

GLMs for Aggregate Claims Under Independence

The common approach to estimating premiums in the property and casualty insurance industry is to consider the aggregate claims cost in the independent setting on the individual level. For more detail see, for example, Anderson et al. (2007). In finding an estimate for the individual's expected loss cost, we are essentially estimating the pure premium for that policyholder. As we have seen in the previous chapters, assuming that the claim frequency and severity are independent allows for simplified results and makes the estimation of the expected loss cost more accessible. We now look in detail at the generalized linear model framework for modelling the pure premium in the independent aggregate claims model.

4.1 The Independent Model

Consider the aggregate claims on the individual level. For policyholder i, we have that the loss cost is

$$S_i = \sum_{j=1}^{N_i} Y_{ij} \tag{4.1}$$

where

(1) N_i is the claim count,

(2) $Y_{ij},\ j = 1, ..., N_i$ are the individual claim amounts,

(3) $Y_{ij},\ j = 1, ..., N_i$ are conditionally i.i.d., given N_i.

Let $\bar{Y}_i = \dfrac{1}{N_i} \sum_{j=1}^{N_i} Y_{ij}$ be the average claim size, or severity, with $\bar{Y}_i = 0$ when $N_i = 0$. This average severity clearly depends on N_i. Then the aggregate claims can be rewritten as:

$$S_i = \sum_{j=1}^{N_i} Y_{ij} = \frac{1}{N_i} \sum_{j=1}^{N_i} Y_{ij} = N_i\, \bar{Y}_i. \tag{4.2}$$

Thus we have that the aggregate loss cost is the product of the claim frequency and severity.

If it is further assumed that the claim frequency is independent of the individual claim amounts, a restrictive assumption, then the GLM structure for the aggregate claims is simplified. In this setting, we have that N_i is independent of $\bar{Y}_i$. If we assumed that at the individual policyholder level, the claim amounts Y_{ij} are i.i.d. with $Y_{ij} \overset{d}{=} Y_i$, then we can write the mean aggregate claim amount as:

$$\begin{aligned}
\mathbb{E}[S_i] = \mathbb{E}[\mathbb{E}(S_i \mid N_i)] &= \mathbb{E}\left[\mathbb{E}\left(\sum_{j=1}^{N_i} Y_{ij} \mid N_i\right)\right] \\
&= \mathbb{E}\left[\sum_{j=1}^{N_i} \mathbb{E}\left(Y_{ij} \mid N_i\right)\right] = \mathbb{E}\left[\sum_{j=1}^{N_i} \mathbb{E}(Y_{ij})\right] \\
&= \mathbb{E}\left[N_i \mathbb{E}(Y_i)\right] = \mathbb{E}(N_i)\mathbb{E}(Y_i).
\end{aligned} \tag{4.3}$$

Equivalently, in terms of the average claim severity $\bar{Y}_i$, we have that the mean aggregate claim cost can be written as:

$$\begin{aligned}
\mathbb{E}[S_i] = \mathbb{E}[\mathbb{E}(S_i \mid N_i)] &= \mathbb{E}\left[\mathbb{E}\left(\sum_{j=1}^{N_i} Y_{ij} \mid N_i\right)\right] \\
&= \mathbb{E}\left[\mathbb{E}\left(N_i \times \bar{Y}_i \mid N_i\right)\right] = \mathbb{E}[N_i \times \mathbb{E}(\bar{Y}_i)] = \mathbb{E}[N_i]\mathbb{E}[\bar{Y}_i].
\end{aligned} \tag{4.4}$$

Since the mean claim cost is the product of the mean frequency and mean severity in the independent model, then in a GLM framework, the model for S_i is simply the product of the marginal GLMs for N_i and $\bar{Y}_i$ respectively.

In Chapter 2, it was shown that the gamma distribution is a member of the reproductive exponential dispersion family. Then by the convolution property (see

Proposition 2.2.2), for independent claim sizes $Y_{ij} \sim ED\left(\mu_i, \dfrac{\sigma^2}{w_{ij}}\right)$ with equal weights $w_{ij} = 1$ we have that

$$\mathbb{E}[\bar{Y}_i] = \mathbb{E}\left[\frac{1}{N_i}\sum_{j=1}^{N_i} Y_{ij}\right] = \mu_i.$$

It follows that modelling the means of the individual claim amounts Y_{ij} is equivalent to modelling the mean of the average severity $\bar{Y}_i$.

Denote μ_{i1} as the mean frequency given the covariates $\mathbf{X}_i$ $\mathbb{E}[N_i|\mathbf{X}_i]$, and μ_{i2} as the mean severity given the covariate $\mathbf{X}_i$ $\mathbb{E}[\bar{Y}_i|\mathbf{X}_i]$. For a $p \times 1$ vector of covariates $\mathbf{X}_i = (X_{i1}, ..., X_{ip})^\top$ and link functions g_1, g_2, we have that the marginal GLMs are defined as:

(i) $g_1\left(\mathbb{E}[N_i|\mathbf{X}_i]\right) = g_1(\mu_{i1}) = \eta_{i1} = \mathbf{X}_{i1}^\top \boldsymbol{\beta}_1 \quad \Leftrightarrow \quad \mu_{i1} = g_1^{-1}(\mathbf{X}_{i1}^\top \boldsymbol{\beta}_1),$

(ii) $g_2\left(\mathbb{E}[\bar{Y}_i|\mathbf{X}_i]\right) = g_2(\mu_{i2}) = \eta_{i2} = \mathbf{X}_{i2}^\top \boldsymbol{\beta}_2 \quad \Leftrightarrow \quad \mu_{i2} = g_2^{-1}(\mathbf{X}_{i2}^\top \boldsymbol{\beta}_2),$

where both $\mathbf{X}_{i1}$ and $\mathbf{X}_{i2}$ are subsets of the covariate vector $\mathbf{X}_i$ and $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are vectors of unknown regression parameters derived from the frequency and severity GLM, respectively.

It follows that the expected aggregate claim cost in the GLM framework for the independent models is simply:

$$\mathbb{E}[S_i|\mathbf{X}_i] = \mu_i^I = \mu_{i1} \times \mu_{i2} = g_1^{-1}(\mathbf{X}_{i1}^\top \boldsymbol{\beta}_1) \times g_2^{-1}(\mathbf{X}_{i2}^\top \boldsymbol{\beta}_2). \tag{4.5}$$

We will refer to this GLM for S_i as the independent model, or Model I, denoted μ_i^I.

In the particular case where both marginal GLMs use a log-link, we have the following simplifications:

$$\ln(\mu_{i1}) = \mathbf{X}_{i1}^\top \boldsymbol{\beta}_1 \quad \Leftrightarrow \quad \mu_{i1} = \exp(\mathbf{X}_{i1}^\top \boldsymbol{\beta}_1),$$

$$\ln(\mu_{i2}) = \mathbf{X}_{i2}^\top \boldsymbol{\beta}_2 \quad \Leftrightarrow \quad \mu_{i2} = \exp(\mathbf{X}_{i2}^\top \boldsymbol{\beta}_2),$$

which then gives the expected total claims costs as:

$$\mu_i^I = \exp(\mathbf{X}_{i1}^\top \boldsymbol{\beta}_1) \times \exp(\mathbf{X}_{i2}^\top \boldsymbol{\beta}_2) = \exp\left\{\mathbf{X}_{i1}^\top \boldsymbol{\beta}_1 + \mathbf{X}_{i2}^\top \boldsymbol{\beta}_2\right\}.$$

Note that using a log-link in the GLM provides many advantages. Firstly, it ensures that both the mean frequency and mean severity are positive. Often, the claim frequency is assumed to follow a Poisson distribution, for which the log-link is in fact the canonical link. Moreover, the log-link yields a non-linear but simple rating structure with multiplication factors associated with each covariate (or rating variable), as discussed in Chapter 3.

4.1.1 Higher Moments in the Independent Model

Variance

As outlined in Chapter 1, when the frequency and severity are considered to be mutually independent, the variance of the aggregate claims can easily be obtained. Using the marginal means derived by the GLM and the properties of the ED models, it follows that the variance for the aggregate losses at the individual level is:

$$\begin{aligned}\mathbb{V}\text{ar}(S_i|X_i) &= \left[\mathbb{E}(Y_i|X_i)\right]^2 \mathbb{V}\text{ar}(N_i|X_i) + \mathbb{V}\text{ar}(Y_i|X_i)\mathbb{E}(N_i|X_i) \\ &= \mu_{i2}^2\phi_1 V_1(\mu_1) + \phi_2 V_2(\mu_{i2})\mu_{i1}, \qquad (4.6)\end{aligned}$$

where the individual claims Y_{ij} are i.i.d. with $Y_{ij} \overset{d}{\sim} Y_i$.

In the particular case where S_i follows a Compound-Poisson-gamma, the variance can be further simplified to:

$$\mathbb{V}\text{ar}(S_i|X_i) = \mu_{i2}^2\mu_{i1} + \phi\mu_{i2}^2\mu_{i1} = \mu_{i1}\mu_{i2}^2(\phi + 1). \qquad (4.7)$$

Recall that the individual claim amounts are such that $Y_{ij} \sim gamma(\mu_{i2}, \phi)$, and thus by the reproductive exponential dispersion model convolution property the mean severity is such that $\bar{Y}_i \sim gamma\left(\mu_{i2}, \dfrac{\phi}{N_i}\right)$. Then, using the mean severity, an alternative way to arrive at the variance in the independent model setting is as follows:

$$\begin{aligned}
\mathbb{V}\text{ar}(S_i|\mathbf{X}_i) &= \mathbb{V}\text{ar}(N_i\bar{Y}_i|\mathbf{X}_i) = \mathbb{V}\text{ar}\left[\mathbb{E}[N_i\bar{Y}_i|N_i,\mathbf{X}_i]|\mathbf{X}_i\right] + \mathbb{E}\left[\mathbb{V}\text{ar}[N_i\bar{Y}_i|N_i,\mathbf{X}_i]|\mathbf{X}_i\right] \\
&= \mathbb{V}\text{ar}\left[N_i\mathbb{E}[\bar{Y}_i|\mathbf{X}_i]|\mathbf{X}_i\right] + \mathbb{E}\left[N_i^2\mathbb{V}\text{ar}[\bar{Y}_i|\mathbf{X}_i]|\mathbf{X}_i\right] \\
&= \mathbb{V}\text{ar}\left[N_i\mu_{i2}|\mathbf{X}_i\right] + \mathbb{E}\left[N_i^2\left(\frac{\phi}{N_i}\right)V(\mu_{i2})\big|\mathbf{X}_i\right] \\
&= \mu_{i2}^2\mathbb{V}\text{ar}[N_i|\mathbf{X}_i] + \mathbb{E}[N_i\phi\mu_{i2}^2|\mathbf{X}_i] \\
&= \mu_{i2}^2\mathbb{V}\text{ar}[N_i|\mathbf{X}_i] + \phi\mu_{i2}^2\mathbb{E}[N_i|\mathbf{X}_i] \\
&= \mu_{i2}^2\mu_{i1} + \phi\mu_{i2}^2\mu_{i1} \\
&= \mu_{i1}\mu_{i2}^2(\phi+1),
\end{aligned}$$

which is the same as the equation (4.7).

Moment Generating Function

Recall from Chapter 1, the moment generating function for S_i in the independent model is:

$$M_{S_i}(t) = M_{N_i}\left[\ln M_{Y_{ij}}(t)\right].$$

Thus, for the Compound-Poisson-gamma model, we have:

$$\begin{aligned}
M_{S_i}(t) &= M_{N_i}\left[\ln(1-t\phi\mu_{i2})^{-\dfrac{1}{\phi}}\right] = M_{N_i}\left[-\frac{1}{\phi}\ln(1-t\phi\mu_{i2})\right] \\
&= \exp\left\{\mu_{i1}(e^{\ln(1-t\phi\mu_{i2})^{-\dfrac{1}{\phi}}}-1)\right\} \\
&= \exp\left\{\mu_{i1}\left((1-t\phi\mu_{i2})^{-\dfrac{1}{\phi}}-1\right)\right\}, \quad t < (1/\phi\mu_{i2}).
\end{aligned}$$

This allows to find higher moments for the aggregate claims S_i and allows to further characterize the aggregate claims distribution.

4.2 MLEs in the Independent Model

In the independent model, the frequency and severity components are considered separately. That is, a GLM is developed for N_i separately from that for $\bar{Y}_i$ and consequently inference for $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are done separately.

As previously mentioned, by the convolution properties of the exponential dispersion models, modelling $\bar{Y}_i$ is equivalent to modelling Y_{ij}. Here, we will consider the maximum likelihood estimation on the individual claim amounts data y_{ij}.

In Chapter 3 equation (3.1), we derived the score equations for the regression parameters in the GLM structure and found the simplified expression:

$$s(\beta_j;\phi,\mathbf{y}) = \sum_{i=1}^{n} \frac{(y_i-\mu_i)}{a_i(\phi)V(\mu_i)}\frac{x_{ij}}{\dot{g}(\mu_i)} = 0. \tag{4.8}$$

Thus, the score equations for the regression parameters $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are as follows:

i) $s(\beta_{1k};\phi_1,\mathbf{n}) = \sum_{i=1}^{m} \dfrac{(n_i-\mu_{i1})}{a_i(\phi_1)V(\mu_{i1})}\dfrac{x_{i1k}}{\dot{g}_1(\mu_{i1})} = 0$ for $k = 1,...,p_1$,

ii) $s(\beta_{2k};\phi_2,\mathbf{y}) = \sum_{i=1}^{m}\sum_{j=1}^{n_i} \dfrac{(y_{ij}-\mu_{i2})}{a_i(\phi_2)V(\mu_{i2})}\dfrac{x_{i2k}}{\dot{g}_2(\mu_{i2})} = 0$ for $k = 1,...,p_2$.

If we return to the Compound-Poisson-gamma case, as is the usual distribution assumptions for the aggregate claims model in the insurance industry, we have that $N_i \sim Poisson(\mu_{i1})$ and $Y_{ij} \sim gamma(\mu_{i2},\phi)$. In Chapter 2, both the Poisson and gamma models were studied in detail. It was shown that for Poisson responses N_i, the dispersion is $a_i(\phi) = 1$ and the variance function is $V(\mu_{i1}) = \mu_{i1}$; while for gamma responses Y_{ij}, $a_{ij}(\phi) = \dfrac{\phi}{w_{ij}}$, for weights $w_{i1},...,w_{in_i}$, and the variance function is $V(\mu_{i2}) = \mu_{i2}^2$. Using equation (4.8), we then have that for a portfolio of m policyholders, the score equations for the frequency and severity parameters, $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ respectively, are as follows:

i) $s(\beta_{1k};\mathbf{n}) = \sum_{i=1}^{m} \dfrac{(n_i-\mu_{i1})}{\mu_{i1}}\dfrac{x_{i1k}}{\dot{g}_1(\mu_{i1})} = 0$ for $k = 1,...,p_1$,

ii) $s(\beta_{2k};\phi,\mathbf{y}) = \sum_{i=1}^{m}\sum_{j=1}^{n_i} \dfrac{(y_{ij}-\mu_{i2})}{\phi/w_{ij}\mu_{i2}^2}\dfrac{x_{i2k}}{\dot{g}_2(\mu_{i2})} = 0$ for $k = 1,...,p_2$.

Using log-link functions for both the frequency and severity models, the framework of Model I gives the following:

◇ $g_1(\mu_{i1}) = \ln(\mu_{i1}) \quad \Rightarrow \quad \dot{g}_1(\mu_{i1}) = \frac{1}{\mu_{i1}}$,

◇ $g_2(\mu_{i2}) = \ln(\mu_{i2}) \quad \Rightarrow \quad \dot{g}_2(\mu_{i2}) = \frac{1}{\mu_{i2}}$.

It follows that the score equations can be further simplified:

i) $s(\beta_{1k};\mathbf{n}) = \sum_{i=1}^{m} \frac{(n_i - \mu_{i1})}{\mu_{i1}} x_{i1k}\mu_{i1} = \sum_{i=1}^{m} x_{i1k}(n_i - \mu_{i1}) = 0$ for $k = 1, ..., p_1$,

ii) $s(\beta_{2k};\phi,\mathbf{y}) = \sum_{i=1}^{m}\sum_{j=1}^{n_i} \frac{(y_{ij} - \mu_{i2})}{\phi/w_{ij}\mu_{i2}^2} x_{i2k}\mu_{i2} = \sum_{i=1}^{m}\sum_{j=1}^{n_i} \frac{w_{ij}}{\phi}\frac{x_{i2k}}{\mu_{i2}}(y_{ij} - \mu_{i2}) = 0$ for $k = 1, ..., p_2$.

Note, that while the individual claim severities are $Y_{ij} \sim gamma(\mu_i, \phi)$, $j = 1, ..., N_i$, the average claim severity is $\bar{Y}_i \sim gamma\left(\mu_i, \frac{\phi}{N_i}\right)$. The score equations derived for the regression parameters $\boldsymbol{\beta}_2$ are equivalent whether we consider the individual claim amounts data y_{ij} or the averaged claim amounts $\bar{y}_i$. If we take the weights to be $w_{ij} = 1$ for all i, j, then we have:

$$\begin{aligned} s(\beta_{2k};\phi,\mathbf{y}) &= \sum_{i=1}^{m}\sum_{j=1}^{n_i} \frac{w_{ij}}{\phi}\frac{x_{i2k}}{\mu_{i2}}(y_{ij} - \mu_{i2}) = \sum_{i=1}^{m}\sum_{j=1}^{n_i} \frac{1}{\phi}\frac{x_{i2k}}{\mu_{i2}}(y_{ij} - \mu_{i2}) \\ &= \sum_{i=1}^{m} \frac{1}{\phi}\frac{x_{i2k}}{\mu_{i2}}\sum_{j=1}^{n_i}(y_{ij} - \mu_{i2}) = \sum_{i=1}^{m} \frac{1}{\phi}\frac{x_{i2k}}{\mu_{i2}} n_i(\bar{y}_i - \mu_{i2}). \end{aligned}$$

Similarly, if we consider the score equations directly for $\bar{Y}_i \sim gamma\left(\mu_{i2}, \frac{\phi}{N_i}\right)$, the dispersion function is $a_i(\phi) = \frac{\phi}{N_i}$ so that the weights used in the model are $w_i = n_i$, for $i = 1, ..., m$. Then based on equation (4.8) we have:

$$s(\beta_{2k};\phi,\mathbf{y}) = \sum_{i=1}^{m} \frac{n_i}{\phi}\frac{x_{i2k}}{\mu_{i2}}(\bar{y}_i - \mu_{i2}).$$

Therefore, the two approaches produce the same score equations.

As mentioned in Chapter 3, by the asymptotic results for maximum likelihood estimates, we have that the MLE based on m observations, $\hat{\boldsymbol{\theta}}_m$, is asymptotically normally distributed with $\sqrt{m}(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_0) \sim \mathcal{N}(\mathbf{0}, \{I(\boldsymbol{\theta}_0)\}^{-1})$. Here, we can estimate the Fisher Information matrix $I(\boldsymbol{\theta}_0)$ by the observed information $\hat{I}_m = I(\hat{\boldsymbol{\theta}}_m)$, where

$$I(\hat{\boldsymbol{\theta}}_m) = -\mathbb{E}[\Psi(\hat{\boldsymbol{\theta}}_m;\mathbf{Y})]$$

and $\Psi(\hat{\boldsymbol{\theta}}_m;\mathbf{Y})$ is a matrix of second partial derivatives of the log-likelihood.

In the Compound Poisson-gamma case, we have:

i) $\diamond$ $\frac{\partial^2}{\partial \beta_{1k}^2}\ell(\boldsymbol{\beta}_1;\mathbf{n}) = \frac{\partial}{\partial \beta_{1k}} s(\beta_{1k};\mathbf{n}) = \sum_{i=1}^m -x_{i1k}\frac{\partial \mu_{i1}}{\partial \beta_{1k}} = \sum_{i=1}^m -x_{i1k}^2\mu_{i1}$

$\Rightarrow -\mathbb{E}[\frac{\partial^2}{\partial \beta_{1k}^2}\ell(\boldsymbol{\beta}_1;\mathbf{n})] = \sum_{i=1}^m x_{i1k}^2\mu_{i1},$

$\diamond$ $\frac{\partial^2}{\partial \beta_{1k}\partial \beta_{1l}}\ell(\boldsymbol{\beta}_1;\mathbf{n}) = \sum_{i=1}^m -x_{i1k}x_{i1l}\mu_{i1}$

$\Rightarrow -\mathbb{E}[\frac{\partial^2}{\partial \beta_{1k}\partial \beta_{1l}}\ell(\boldsymbol{\beta}_1;\mathbf{n})] = \sum_{i=1}^m x_{i1k}x_{i1l}\mu_{i1}.$

ii) $\diamond$ $\frac{\partial^2}{\partial \beta_{2k}^2}\ell(\boldsymbol{\beta}_2;\mathbf{y}) = \sum_{i=1}^m\sum_{j=1}^{n_i}\frac{\partial^2}{\partial \beta_{2k}^2}\frac{1}{\phi}x_{i2k}\left(\frac{y_{ij}}{\mu_{i2}}-1\right)$

$= \sum_{i=1}^m\sum_{j=1}^{n_i}\frac{1}{\phi}x_{i2k} - \frac{y_{ij}}{\mu_{i2}^2}\frac{\partial \mu_{i2}}{\partial \beta_{2k}}$

$= \sum_{i=1}^m\sum_{j=1}^{n_i}\frac{1}{\phi}x_{i2k}^2 - \frac{y_{ij}}{\mu_{i2}}$

$\Rightarrow -\mathbb{E}[\frac{\partial^2}{\partial \beta_{2k}^2}\ell(\boldsymbol{\beta}_2;\mathbf{y})] = \sum_{i=1}^m\sum_{j=1}^{n_i}\frac{1}{\phi}x_{i2k}^2,$

$\diamond$ $\frac{\partial^2}{\partial \beta_{2k}\partial \beta_{2l}}\ell(\boldsymbol{\beta}_2;\mathbf{y}) = \sum_{i=1}^m\sum_{j=1}^{n_i}\frac{1}{\phi}x_{i2k}x_{i2l} - \frac{y_{ij}}{\mu_{i2}}$

$\Rightarrow -\mathbb{E}[\frac{\partial^2}{\partial \beta_{2k}\partial \beta_{2l}}\ell(\boldsymbol{\beta}_2;\mathbf{y})] = \sum_{i=1}^m\sum_{j=1}^{n_i}\frac{1}{\phi}x_{i2k}x_{i2l}.$

The observed information matrices then become:

$$\text{i) } I(\hat{\boldsymbol{\beta}}_1)_{p_1,p_1} = \begin{pmatrix} \sum_{i=1}^m x_{i11}^2\hat{\mu}_{i1} & \sum_{i=1}^m x_{i11}x_{i12}\hat{\mu}_{i1} & \cdots & \sum_{i=1}^m x_{i11}x_{i1p_1}\hat{\mu}_{i1} \\ \sum_{i=1}^m x_{i12}x_{i11}\hat{\mu}_{i1} & \sum_{i=1}^m x_{i12}^2\hat{\mu}_{i1} & \cdots & \sum_{i=1}^m x_{i12}x_{i1p_1}\hat{\mu}_{i1} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^m x_{i1p_1}x_{i11}\hat{\mu}_{i1} & \sum_{i=1}^m x_{i1p_1}x_{i12}\hat{\mu}_{i1} & \cdots & \sum_{i=1}^m x_{i1p_1}^2\hat{\mu}_{i1} \end{pmatrix}$$

$$\text{ii) } I(\hat{\boldsymbol{\beta}}_2)_{p_2,p_2} = \frac{1}{\phi}\begin{pmatrix} \sum_{i=1}^m x_{i21}^2 & \sum_{i=1}^m x_{i21}x_{i22} & \cdots & \sum_{i=1}^m x_{i21}x_{i2p_2} \\ \sum_{i=1}^m x_{i22}x_{i21} & \sum_{i=1}^m x_{i22}^2 & \cdots & \sum_{i=1}^m x_{i22}x_{i2p_2} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^m x_{i2p_2}x_{i21} & \sum_{i=1}^m x_{i2p_2}x_{i22} & \cdots & \sum_{i=1}^m x_{i2p_2}^2 \end{pmatrix}$$

Note that in the matrix $I(\hat{\boldsymbol{\beta}}_2)_{p_2,p_2}$, the dispersion parameter ϕ is generally unknown and must be estimated. For example, we can use the estimate

$$\hat{\phi}_{p_2} = \frac{1}{m-p_2}\sum_{i=1}^m\left(\frac{\bar{y}_i-\hat{\mu}_{i2}}{\hat{\mu}_{i2}}\right)^2.$$

Therefore, we have that the MLE's are asymptotically normally distributed with $\sqrt{m}(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) \sim \mathcal{N}\left(\mathbf{0}, \{I(\boldsymbol{\beta}_1)\}^{-1}\right)$ and $\sqrt{m}(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2) \sim \mathcal{N}\left(\mathbf{0}, \{I(\boldsymbol{\beta}_2)\}^{-1}\right)$ where $I(\boldsymbol{\beta}_1)$ and $I(\boldsymbol{\beta}_2)$ can be estimated by $I(\hat{\boldsymbol{\beta}}_1)$ and $I(\hat{\boldsymbol{\beta}}_2)$, respectively.

4.3 Tweedie Modelling

An alternative approach to modelling the marginal means of the frequency and severity components is to directly model the claims cost as a Tweedie distribution. As previously mentioned, the Tweedie distribution corresponds to a compound Poisson process. In the case of insurance data, the model assumptions are that the claim counts follow a Poisson distribution while the jumps, which represent the claim sizes, follow a gamma distribution. Since the Tweedie model is a member of the Exponential Dispersion models family, it can also be modelled in the Generalized Linear Model framework. Jørgensen provides details on using the Tweedie distribution for modelling claims; see Jørgensen and De Souzaa (1994) for details. Jørgensen and Smyth (2002) also revisit the problem to include a dispersion component to the model framework.

Chapter 5

GLMs for Aggregate Claims Under Dependence

Several approaches can be taken to address the dependence between the frequency and severity components in the aggregate claims model. Gschlößl and Czado (2007), in particular, investigate this problem using a fully Bayesian approach and estimate parameters using Markov Chain Monte Carlo under a slightly different model specification than what we present here. Sarabia and Guillén (2008) consider the joint distribution of (S, N) using the conditional distributions S given N and of N given S. Jørgensen (2011) provides yet another approach to this problem. Although not specifically addressing the issue of dependency in the aggregate claims model, Jørgensen extends the univariate dispersion models definition to a multivariate model. In particular, he introduces a construction of multivariate exponential dispersion models and also briefly introduces the concept of multivariate generalized linear models. Similarly, Iwasaki and Tsubaki (2005) construct a bivariate distribution in the natural exponential family, which could be used to define the bivariate distribution of the frequency and severity components of the aggregate claims.

This thesis will provide an alternative framework by allowing for the dependence between the severity and frequency components through a conditional GLM. The goal of this approach is to allow for dependence between the claim amounts Y_{ij} and the claim counts N_i, at the individual level, by assuming that the conditional mean severity $\mathbb{E}[\bar{Y}_i|N_i]$ (which is equivalent to $\mathbb{E}[Y_{ij}|N_i]$) is a function of the claim count N_i. This is achieved by including the claim count as a covariate in the conditional mean severity GLM for $\mathbb{E}[\bar{Y}_i|N_i]$. This chapter provides the details of this model formulation.

5.1 The Dependent Model

Now let us relax the assumption that the claim frequency and claim severity are independent. Then the mean aggregate loss cost is no longer simply the product of the marginal means for the frequency and severity components respectively. In this model we have that:

$$\begin{aligned}\mathbb{E}[S_i|\mathbf{X}_i] &= \mathbb{E}[N_i\bar{Y}_i|\mathbf{X}_i] = \mathbb{E}\big[\mathbb{E}[N_i\bar{Y}_i|\mathbf{X}_i, N_i]\big|\mathbf{X}_i\big] \\ &= \mathbb{E}\big[N_i\mathbb{E}[\bar{Y}_i|\mathbf{X}_i, N_i]\big|\mathbf{X}_i\big] \\ &\neq \mathbb{E}[N_i|\mathbf{X}_i]\ \mathbb{E}[\bar{Y}_i|\mathbf{X}_i].\end{aligned}$$

We will refer to this GLM for S_i as the dependent model, or Model D and denote its mean by μ_i^D.

Here $\mathbb{E}[\bar{Y}_i|\mathbf{X}_i, N_i]$ is a function of both N_i and $\mathbf{X}_i$, which can be defined through a conditional marginal GLM as:

$$g\{\mathbb{E}[\bar{Y}_i|\mathbf{X}_i, N_i]\} = g\{\mu_{i2}^D\} = \tilde{\mathbf{X}}_{i2}^\top\tilde{\boldsymbol{\beta}}_2 + N_i\beta_N. \tag{5.1}$$

Note that the regression parameters here, $\tilde{\boldsymbol{\beta}}_2$, are different than the regression parameters in Model I, $\boldsymbol{\beta}_2$, since the presence of N_i as a covariate in this GLM affects the regression parameters and their estimates. Similarly, the covariates used in the marginal severity GLM could be different in Model D as compared to Model I, thus we have that the covariates $\tilde{\mathbf{X}}_{i2}$ are not necessarily the same as $\mathbf{X}_{i2}$.

Using a log-link in the GLM for $\mathbb{E}[\bar{Y}_i|\mathbf{X}_i, N_i]$ gives:

$$\ln\big\{\mathbb{E}[\bar{Y}_i|\mathbf{X}_i, N_i]\big\} = \ln(\mu_{i2}^D) = \tilde{X}_{i2}^\top\tilde{\boldsymbol{\beta}}_2 + N_i\beta_N,$$

which implies that the conditional mean severity is given by

$$\mathbb{E}[\bar{Y}_i|\mathbf{X}_i, N_i] = \exp(\tilde{\mathbf{X}}_{i2}^\top\tilde{\boldsymbol{\beta}}_2 + N_i\beta_N) = \exp(\tilde{\mathbf{X}}_{i2}^\top\tilde{\boldsymbol{\beta}}_2)\exp(N_i\beta_N) = \tilde{\mu}_{i2}\exp(N_i\beta_N).$$

It follows that the mean aggregate loss cost becomes:

$$\begin{aligned}
\mu_i^D &= \mathbb{E}[S_i|\mathbf{X}_i] = \mathbb{E}\big[N_i\mathbb{E}\left(\bar{Y}_i|\mathbf{X}_i, N_i\right)\big|\mathbf{X}_i\big] = \mathbb{E}\big[N_i\tilde{\mu}_{i2}\exp(N_i\beta_N)\big|\mathbf{X}_i\big] \\
&= \tilde{\mu}_{i2}\mathbb{E}\big[N_i\exp(N_i\beta_N)\big|\mathbf{X}_i\big] = \tilde{\mu}_{i2}\mathbb{E}\left[\frac{\partial}{\partial\beta_N}\exp(N_i\beta_N)\big|\mathbf{X}_i\right] \\
&= \tilde{\mu}_{i2}\frac{\partial}{\partial\beta_N}\mathbb{E}\big[\exp(N_i\beta_N)\big|\mathbf{X}_i\big] = \tilde{\mu}_{i2}\frac{\partial}{\partial\beta_N}M_{N_i|X_i}(\beta_n),
\end{aligned} \tag{5.2}$$

where $M_{N_i|X_i}(\cdot)$ is the conditional moment generating function of N_i given the vector of covariates X_i.

The frequency, N_i, is also modelled through a GLM. Since we are modelling $\mathbb{E}[N_i|\mathbf{X}_i]$ as was done in the independent model, it follows that the marginal GLM for the claim frequency in Model D is equivalent to that in Model I. Thus we have:

$$\begin{aligned}
g_1\{\mathbb{E}[N_i|\mathbf{X}_i]\} &= g_1\{\mu_{i1}\} = \eta_{i1} = \mathbf{X}_{i1}^\top\boldsymbol{\beta}_1 \\
\Leftrightarrow \mu_{i1} &= g_1^{-1}(\mathbf{X}_{i1}^\top\boldsymbol{\beta}_1).
\end{aligned}$$

By the distributional structure of the Exponential Dispersion models, the moment generating function of the response variable is a function of the canonical parameter, and thus a function of the mean, as well as a function of the dispersion parameter. Hence, assuming that the dispersion parameter is known, the GLM on μ_{i1} allows to define the moment generating function of N_i through the cumulant function $\kappa(\theta)$, as shown in Chapter 2.

Thus from (5.2) we have:

$$\mu_i^D = \mathbb{E}[S_i|\mathbf{X}_i] = \tilde{\mu}_{i2}\frac{\partial}{\partial\beta_N}M_{N_i|X_i}(\beta_n) = \tilde{\mu}_{i2}\frac{\partial}{\partial\beta_N}h(\beta_N;\mu_{i1},\phi_1),$$

where $M_{N_i|X_i}(\beta_n)$ is a function of the mean μ_{i1} and the dispersion parameter ϕ_1 as defined by the frequency response distribution.

In the case where N_i is assumed to follow a Poisson distribution, then using the log-link, which is also the canonical link function for the Poisson model, the marginal GLM for the frequency is the same as that in the independent model:

$$\mu_{i1} = \exp(\mathbf{X}_{i1}^\top\boldsymbol{\beta}_1).$$

For Poisson responses, the dispersion is $\phi_1 = 1$ and the moment generating function is $M_{N_i|X_i}(t) = \exp\{\mu_{i1}(e^t - 1)\}, \quad \text{for } t \in \mathbb{R}.$

Returning to equation (5.2), for a Poisson distributed marginal frequency, the expected total loss cost can be further simplified to:

$$\begin{aligned}
\mu_i^D = \mathbb{E}[S_i|\mathbf{X}_i] &= \tilde{\mu}_{i2}\frac{\partial}{\partial \beta_N} M_{N_i|X_i}(\beta_n) \\
&= \tilde{\mu}_{i2}\frac{\partial}{\partial \beta_N} \exp\left\{\mu_{i1}(e^{\beta_N} - 1)\right\} \\
&= \tilde{\mu}_{i2} \exp\left\{\mu_{i1}(e^{\beta_N} - 1)\right\} \mu_{i1} \exp(\beta_N) \\
&= \tilde{\mu}_{i2}\mu_{i1} \exp\left\{\mu_{i1}(e^{\beta_N} - 1) + \beta_N\right\}. \qquad (5.3)
\end{aligned}$$

Notice that this final formulation of μ_i^D makes no distributional assumptions for the severity $\bar{Y}_i$. The only restriction is that $\bar{Y}_i$ be a member of the Exponential Dispersion family so that the mean can be modelled via a GLM. There is, however, a restriction on the choice of link function since this formulation relies on the use of the log link in the conditional mean severity GLM. The model assumption made for the severity component $\bar{Y}_i$ will only affect the estimation of the regression parameters $\tilde{\boldsymbol{\beta}}_2$ since the distribution will define the score function used in the maximum likelihood estimation.

Note that when $\beta_N = 0$ we retrieve the independent case. The marginal mean frequency μ_{i1} remains the same in both the dependent and independent cases. If $\beta_N = 0$, then the regression parameters in the marginal GLM for severity will be identical under the dependence and independence assumptions since both means $\mathbb{E}[\bar{Y}_i|\mathbf{X}_i]$ and $\mathbb{E}[\bar{Y}_i|\mathbf{X}_i, N_i]$ will be modelled using the same covariates. That is, if $\beta_N = 0$, then $\tilde{\mathbf{X}}_{i2} = \mathbf{X}_{i2}$, and since the data will then be modelled using the same covariate in both models, this in turn implies that $\tilde{\boldsymbol{\beta}}_2 = \boldsymbol{\beta}_2$ and so $\mu_{i2} = \exp(\mathbf{X}_{i2}^\top \boldsymbol{\beta}_2) = \tilde{\mu}_{i2} = \exp(\tilde{\mathbf{X}}_{i2}^\top \tilde{\boldsymbol{\beta}}_2)$. The remaining correction term in the equation, $\exp\left\{\mu_{i1}(e^{\beta_N} - 1) + \beta_N\right\}$, is equal to 1 if $\beta_N = 0$. Thus, for $\beta_N = 0$ we have that

$$\mu_i^D = \exp(\mathbf{X}_{i2}^\top \boldsymbol{\beta}_2) \exp(\mathbf{X}_{i1}^\top \boldsymbol{\beta}_1) = \mu_{i2}\mu_{i1} = \mu_i^I.$$

Hence, the expected loss costs under Model I and Model D are identical for $\beta_N = 0$.

Although the model formulation is straightforward, the interpretation of the effect of the dependence between N_i and $\bar{Y}_i$ is not so clear. First consider the model for the modified marginal mean severity $\tilde{\mu}_{i2} = \exp(\tilde{\mathbf{X}}_{i2}^\top \tilde{\boldsymbol{\beta}}_2)$. If β_N is a significant regression parameter in the conditional severity GLM, we can conclude that the severity $\bar{Y}_i$ is indeed significantly influenced by the frequency N_i. However, the impact on the modified mean severity $\tilde{\mu}_{i2}$ is not so obvious since including N_i as an extra covariate in the model will change the remaining regression parameters $\tilde{\beta}_2$ and their estimates. That is, the influence of the remaining covariates $\tilde{\mathbf{X}}_{i2}$ on the modified mean severity $\tilde{\mu}_{i2}$ will be different if N_i is included as an additional covariate in the GLM. In Model D, the GLM on $\mathbb{E}[\bar{Y}_i|\mathbf{X}_i, N_i]$ is a function of both N_i and $\tilde{\mathbf{X}}_{i2}$. Thus, even if $\beta_N > 0$ ($\beta_N < 0$) leads to a factor $\exp(N_i\beta_N) > 1$ ($\exp(N_i\beta_N) < 1$), the mean severity $\mu_{i2}^D = \tilde{\mu}_{i2} \exp\{N\beta_N\}$ might not necessarily increase (decrease) since this effect might be offset by the change in the remaining regression parameters from β_2 in Model I to $\tilde{\beta}_2$ in Model D. This ambiguity is also inherent in the interpretation of the effect of dependency between N_i and $\bar{Y}_i$ on the expected loss cost $\mathbb{E}[S_i|\mathbf{X}_i]$.

5.1.1 Higher Moments for the Dependent Model

Variance

As was the case for the first moment of the total loss cost, $\mathbb{E}[S_i|\mathbf{X}_i]$, the variance of the aggregate claims in the dependent model can no longer be written in terms of the marginal moments of the frequency and severity. Recall that under the assumptions and GLM results of Model D, $N_i \sim \mathit{Poisson}(\mu_{i1})$ and conditional on N_i the average claim severity $\bar{Y}_i \sim \mathit{gamma}\left(\mu_{i2}^D, \dfrac{\phi}{N_i}\right)$. The variance can then be derived as follows:

$$\begin{aligned}
\mathbb{V}\mathrm{ar}[S_i|\mathbf{X}_i] &= \mathbb{V}\mathrm{ar}\big[\mathbb{E}[S_i|\mathbf{X}_i, N_i]\big|\mathbf{X}_i\big] + \mathbb{E}\big[\mathbb{V}\mathrm{ar}[S_i|\mathbf{X}_i, N_i]\big|X_i\big] \\
&= \mathbb{V}\mathrm{ar}\big[\mathbb{E}[N_i\bar{Y}_i|\mathbf{X}_i, N_i]\big|\mathbf{X}_i\big] + \mathbb{E}\big[\mathbb{V}\mathrm{ar}[N_i\bar{Y}_i|\mathbf{X}_i, N_i]\big|\mathbf{X}_i\big] \\
&= \mathbb{V}\mathrm{ar}\big[N_i\mathbb{E}[\bar{Y}_i|\mathbf{X}_i, N_i]\big|\mathbf{X}_i\big] + \mathbb{E}\big[N_i^2\mathbb{V}\mathrm{ar}[\bar{Y}_i|\mathbf{X}_i, N_i]\big|\mathbf{X}_i\big].
\end{aligned}$$

Recall that for members of the ED family with $Z_i \sim ED(\mu_i, \phi)$, the variance is $\mathbb{V}\mathrm{ar}(Z_i) = a_i(\phi)V(\mu_i)$. In particular, for gamma responses $\bar{Y}_i$, we have that conditional on N_i, $a_i(\phi) = \dfrac{\phi}{N_i}$ and $V(\mu_i) = \mu_i^2$. It follows from the conditional severity GLM results that

$$\mathbb{E}[\bar{Y}_i|\mathbf{X}_i, N_i] = \mu_{i2}^D,$$

and

$$\mathbb{V}\text{ar}[\bar{Y}_i|\mathbf{X}_i, N_i] = \left(\frac{\phi}{N_i}\right) V(\mu_{i2}^D) = \left(\frac{\phi}{N_i}\right) (\mu_{i2})^2.$$

Thus, we can further simplify the expression:

$$\begin{aligned}
\mathbb{V}\text{ar}[S_i|\mathbf{X}_i] &= \mathbb{V}\text{ar}\big[N_i\mu_{i2}^D\big|\mathbf{X}_i\big] + \mathbb{E}\Big[N_i^2\left(\frac{\phi}{N_i}\right)V(\mu_{i2}^D)\Big|X_i\Big] \\
&= \mathbb{V}\text{ar}\big[N_i\mu_{i2}^D\big|\mathbf{X}_i\big] + \mathbb{E}\big[\phi N_i(\mu_{i2}^D)^2\big|\mathbf{X}_i\big].
\end{aligned}$$

Finally, we can use the results from the severity GLM: Conditional on N_i, the severity mean is $\mu_{i2}^D = \exp\{\tilde{\mathbf{X}}_{i2}^\top\tilde{\boldsymbol{\beta}}_2 + N_i\beta_N\}$. Thus we have:

$$\begin{aligned}
\mathbb{V}\text{ar}[S_i|\mathbf{X}_i] &= \mathbb{V}\text{ar}[N_i\exp\{\tilde{\mathbf{X}}_{i2}^\top\tilde{\boldsymbol{\beta}}_2 + N_i\beta_N\}|\mathbf{X}_i] + \mathbb{E}[\phi N_i\exp\{2\tilde{\mathbf{X}}_{i2}^\top\tilde{\boldsymbol{\beta}}_2 + 2N_i\beta_N\}|\mathbf{X}_i] \\
&= \exp\{2\tilde{\mathbf{X}}_{i2}^\top\tilde{\boldsymbol{\beta}}_2\}\mathbb{V}\text{ar}[N_i\exp\{N_i\beta_N\}|\mathbf{X}_i] \\
&\quad + \phi\exp\{2\tilde{\mathbf{X}}_{i2}^\top\tilde{\boldsymbol{\beta}}_2\}\mathbb{E}[N_i\exp\{2N_i\beta_N\}|\mathbf{X}_i] \\
&= (\tilde{\mu}_{i2})^2\mathbb{V}\text{ar}[N_i\exp\{N_i\beta_N\}|\mathbf{X}_i] + \phi(\tilde{\mu}_{i2})^2\mathbb{E}[N_i\exp\{2N_i\beta_N\}|\mathbf{X}_i]. \qquad (5.4)
\end{aligned}$$

We can then further simplify the above expressions using the following results:

$$\begin{aligned}
(i)\quad \mathbb{E}[N_i\exp(N_i\beta_N)|\mathbf{X}_i] &= \mathbb{E}\left[\frac{\partial}{\partial\beta_N}\exp(N_i\beta_N)|\mathbf{X}_i\right] = \frac{\partial}{\partial\beta_N}\mathbb{E}\left[\exp(N_i\beta_N)|\mathbf{X}_i\right] \\
&= \frac{\partial}{\partial\beta_N}M_{N_i|X_i}(\beta_N) = \frac{\partial}{\partial\beta_N}\exp\left\{\mu_{i1}(e^{\beta_N}-1)\right\} \\
&= \exp\left\{\mu_{i1}(e^{\beta_N}-1)\right\}\mu_{i1}e^{\beta_N} \\
&= \mu_{i1}\exp\left\{\mu_{i1}(e^{\beta_N}-1)+\beta_N\right\}.
\end{aligned}$$

$$\begin{aligned}
(ii)\quad \mathbb{E}[N_i\exp(2N_i\beta_N)|\mathbf{X}_i] &= \mathbb{E}\left[\frac{1}{2}\frac{\partial}{\partial\beta_N}\exp(2N_i\beta_N)|\mathbf{X}_i\right] = \frac{1}{2}\frac{\partial}{\partial\beta_N}M_{N_i|X_i}(2\beta_N) \\
&= \mu_{i1}\exp\left\{\mu_{i1}(e^{2\beta_N}-1)+2\beta_N\right\}.
\end{aligned}$$

$$
\begin{aligned}
(iii) \quad \mathbb{V}\mathrm{ar}[N_i \exp\{N_i\beta_N\}|\mathbf{X}_i] &= \mathbb{E}\Big[\,[N_i \exp(N_i\beta_N) - \mathbb{E}[N_i \exp(N_i\beta_N)|\mathbf{X}_i]]^2\,|\mathbf{X}_i\Big] \\
&= \mathbb{E}\left[N_i^2 \exp(2N_i\beta_N)|\mathbf{X}_i\right] - \left[\mathbb{E}[N_i \exp(N_i\beta_N)|\mathbf{X}_i]\right]^2 \\
&= \mathbb{E}\left[\frac{1}{4}\frac{\partial^2}{\partial\beta_N^2}\exp(N_i\beta_N)|\mathbf{X}_i\right] \\
&\qquad - \mu_{i1}^2 \exp\left\{2\mu_{i1}(e^{\beta_N}-1) + 2\beta_N\right\} \\
&= \frac{1}{4}\frac{\partial^2}{\partial\beta_N^2}\mathbb{E}\left[\exp(N_i\beta_N)|\mathbf{X}_i\right] \\
&\qquad - \mu_{i1}^2 \exp\left\{2\mu_{i1}(e^{\beta_N}-1) + 2\beta_N\right\} \\
&= \frac{1}{4}\frac{\partial^2}{\partial\beta_N^2}M_{N_i|X_i}(2\beta_N) \\
&\qquad - \mu_{i1}^2 \exp\left\{2\mu_{i1}(e^{\beta_N}-1) + 2\beta_N\right\} \\
&= \frac{1}{4}\frac{\partial^2}{\partial\beta_N^2}\exp\left\{\mu_{i1}(e^{2\beta_N}-1)\right\} \\
&\qquad - \mu_{i1}^2 \exp\left\{2\mu_{i1}(e^{\beta_N}-1) + 2\beta_N\right\} \\
&= \mu_{i1}^2 \exp\left\{\mu_{i1}(e^{2\beta_N}-1) + 4\beta_N\right\} \\
&\qquad + \mu_{i1} \exp\left\{\mu_{i1}(e^{2\beta_N}-1) + 2\beta_N\right\} \\
&\qquad - \mu_{i1}^2 \exp\left\{\mu_{i1}(e^{2\beta_N}-1) + 2\beta_N\right\}.
\end{aligned}
$$

Going back to equation (5.4) then gives:

$$
\begin{aligned}
\mathbb{V}\mathrm{ar}[S_i|\mathbf{X}_i] = (\tilde{\mu}_{i2})^2\Big[&\mu_{i1}^2 \exp\left\{\mu_{i1}(e^{2\beta_N}-1) + 4\beta_N\right\} \\
&+ \mu_{i1} \exp\left\{\mu_{i1}(e^{2\beta_N}-1) + 2\beta_N\right\} \\
&- \mu_{i1}^2 \exp\left\{\mu_{i1}(e^{\beta_N}-1) + 2\beta_N\right\}\Big] \\
&+ \phi(\tilde{\mu}_{i2})^2\mu_{i1} \exp\left\{\mu_{i1}(e^{2\beta_N}-1) + 2\beta_N\right\} \\
= \mu_{i1}(\tilde{\mu}_{i2})^2\Big[&\mu_{i1} \exp\left\{\mu_{i1}(e^{2\beta_N}-1) + 4\beta_N\right\} \\
&+ (\phi+1) \exp\left\{\mu_{i1}(e^{2\beta_N}-1) + 2\beta_N\right\} \\
&- \mu_{i1} \exp\left\{\mu_{i1}(e^{\beta_N}-1) + 2\beta_N\right\}\Big]. \qquad (5.5)
\end{aligned}
$$

Note that if $\beta_N = 0$ then this reduces to

$$
\mathbb{V}\mathrm{ar}[S_i|\mathbf{X}_i] = \mu_{i1}\mu_{i2}^2[\mu_{i1}\exp(0) + (\phi+1)\exp(0) - \mu_{i1}\exp(0)] = \mu_{i1}\mu_{i2}^2(\phi+1),
$$

which is the same variance equation obtained in the independent model (see equation (4.7)). Thus again, $\beta_N = 0$ recovers the independent case.

Moment Generating Function

In the dependent model, the moment generating function is more complicated. We have that:

$$\begin{aligned} M_{S_i|\mathbf{X}_i}(t) &= \mathbb{E}\left[e^{S_i t}|\mathbf{X}_i\right] = \mathbb{E}\left[e^{N_i \bar{Y}_i t}|\mathbf{X}_i\right] \\ &= \mathbb{E}\left[\mathbb{E}[e^{N_i \bar{Y}_i t}|N_i, \mathbf{X}_i]\big|\mathbf{X}_i\right] = \mathbb{E}\big[M_{\bar{Y}_i|N_i,X_i}(N_i t)|\mathbf{X}_i\big]. \end{aligned}$$

If we return to the Compound Poisson gamma case, we have that conditionally on N_i and $\mathbf{X}_i$, $\bar{Y}_i$ is *gamma* $\left(\mu_{i2}^D, \dfrac{\phi}{N_i}\right)$ and so $M_{\bar{Y}_i|N_i,X_i}(t) = \left(1 - \dfrac{\phi}{N_i}\mu_{i2}^D t\right)^{-N_i/\phi}$. Thus, we have that

$$\begin{aligned} M_{S_i|X_i}(t) &= \mathbb{E}\big[M_{\bar{Y}_i|N_i,X_i}(N_i t)\big|\mathbf{X}_i\big] = \mathbb{E}\left[\left(1 - \frac{\phi}{N_i}\mu_{i2}^D N_i t\right)^{-N_i/\phi}\big|\mathbf{X}_i\right] \\ &= \mathbb{E}\left[\left(1 - \phi\mu_{i2}^D t\right)^{-N_i/\phi}|\mathbf{X}_i\right] \\ &= \mathbb{E}\left[\left(1 - \phi\exp\{\tilde{X}_{i2}^\top \tilde{\beta}_2\}\exp\{N_i\beta_N\}t\right)^{-N_i/\phi}|X_i\right] \\ &= \mathbb{E}\left[\left(1 - \phi\tilde{\mu}_{i2}\exp\{N_i\beta_N\}t\right)^{-N_i/\phi}|X_i\right] \\ &= \sum_{n=0}^{\infty}(1 - \phi\tilde{\mu}_{i2}\exp\{n\beta_N\}t)^{-n/\phi}\,\frac{e^{-\mu_{i1}}\mu_{i1}^n}{n!}. \end{aligned}$$

Unlike in the independent case, this moment generating function has no closed form.

5.2 MLEs in the Dependent Model

Consider the joint distribution of the frequency and severity components of the aggregate losses in Model D:

$$f_{\bar{Y},N}(y,n) = f_{\bar{Y}|N}(y|n)f_N(n).$$

Assuming the dispersion parameter ϕ is known, the marginal means obtained from the GLM for $\mathbb{E}[N] = \mu_1$ and $\mathbb{E}[\bar{Y}|N] = \mu_2^D$ allow to fully parametrize the distributions $f_{\bar{Y}|N}$ and f_N, respectively.

As discussed in Chapter 2, both the additive and reproductive forms of the exponential dispersion models are closed under convolution. Thus, for individual claim amounts Y_{ij} with means μ_i, we have that $\bar{Y}_i$ also has mean μ_i. More precisely, if the individual severities belong to the additive ED family with Y_{ij} given N_i being $ED^*(\theta_i, \lambda)$ then the distribution of the sum $\sum_{j=1}^{N_i} Y_{ij} = N_i \times \bar{Y}_i$, conditional on N_i, is also a member of the additive ED models with $N_i\bar{Y}_i$ given N_i being $ED^*(\theta_i, N_i\lambda)$. If we choose the severity to belong to the reproductive ED family, then we have that Y_{ij} given N_i is $ED(\mu_i, \sigma^2)$ and so $\bar{Y}_i$ given N_i is $ED\left(\mu_i, \dfrac{\sigma^2}{N_i}\right)$.

In the GLM formulation, the marginal means of the frequency and severity components are a function of the regression parameters $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \beta_N)$. We are thus interested in obtaining maximum likelihood estimates for these parameters as it will allow ultimately to estimate the expected loss cost. The MLEs for the regression parameters can be found through the joint likelihood of N and $\bar{Y}$. For m policyholders, we have that the joint likelihood function is:

$$L(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \beta_N; \mathbf{y}, \mathbf{n}) = \prod_{i=1}^{m} f_{\bar{Y},N}(y_i, n_i) = \prod_{i=1}^{m} f_{\bar{Y}|N}(y_i|n_i) f_N(n_i),$$

and the joint log-likelihood is:

$$\ell(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \beta_N; \mathbf{y}, \mathbf{n}) = \ln L(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \beta_N; \mathbf{y}, \mathbf{n}),$$

where $f_{\bar{Y}|N}$ is a function of the parameters (μ_{i2}^D, ϕ_2) and f_N is a function of (μ_{i1}, ϕ_1).

Recall from Chapter 3 that the likelihood and log-likelihood functions for the exponential dispersion models could be written in terms of the canonical parameters θ_i as:

$$\begin{aligned} L(\boldsymbol{\theta}, \phi; y) &= \prod_{i=1}^{m} \exp\left[\frac{y_i\theta_i - \kappa(\theta_i)}{a_i(\phi)} + C(y_i, \phi)\right] \\ &= \exp\left[\sum_{i=1}^{m}\left\{\frac{y_i\theta_i - \kappa(\theta_i)}{a_i(\phi)}\right\}\right] \times \prod_{i=1}^{m} \exp\left\{C(y_i, \phi)\right\}, \end{aligned}$$

and

$$\ell(\boldsymbol{\theta}, \phi; y) = \ln L(\boldsymbol{\theta}, \phi; y) = \sum_{i=1}^{m}\left\{\frac{y_i\theta_i - \kappa(\theta_i)}{a_i(\phi)}\right\} + \sum_{i=1}^{m} C(y_i, \phi).$$

Thus, we can write the joint likelihood of $(\bar{Y}, N)$ in terms of the canonical parameter $\boldsymbol{\theta} = (\theta_1, \theta_2)$ as follows:

$$\begin{aligned} L(\boldsymbol{\theta}, \boldsymbol{\phi}; \mathbf{y}, \mathbf{n}) &= \exp\left\{\sum_{i=1}^{m}\left\{\frac{y_i\theta_{i2} - \kappa_2(\theta_{i2})}{a_{i2}(\phi_2)}\right\}\right\} \times \prod_{i=1}^{m} \exp\left\{C(y_i, \phi_2)\right\} \\ &\quad \times \exp\left\{\sum_{i=1}^{m}\left\{\frac{n_i\theta_{i1} - \kappa_1(\theta_{i1})}{a_{i1}(\phi_1)}\right\}\right\} \times \prod_{i=1}^{m} \exp\left\{C(n_i, \phi_1)\right\}, \end{aligned}$$

and the log-likelihood as:

$$\begin{aligned} \ell(\boldsymbol{\theta}, \boldsymbol{\phi}; \mathbf{y}, \mathbf{n}) &= \sum_{i=1}^{m}\left\{\frac{y_i\theta_{i2} - \kappa_2(\theta_{i2})}{a_{i2}(\phi_2)}\right\} + \sum_{i=1}^{m} C(y_i, \phi_2) \\ &\quad + \sum_{i=1}^{m}\left\{\frac{n_i\theta_{i1} - \kappa_1(\theta_{i1})}{a_{i1}(\phi_1)}\right\} + \sum_{i=1}^{m} C(n_i, \phi_1). \end{aligned}$$

By the GLM structure used in Model D with log-link functions, we have that the mean mapping function is $\mu = \tau(\theta) = \exp(\theta)$ for both the frequency and severity components. Thus, we have that $\mu_{i2}^D = \tau_2(\theta_{i2}) = \exp(\tilde{\mathbf{X}}_{i2}^\top\tilde{\boldsymbol{\beta}}_2 + N_i\beta_N)$ and $\mu_{i1} = \tau_1(\theta_{i1}) = \exp(\mathbf{X}_{i1}^\top\boldsymbol{\beta}_1)$. It follows that all of the information for the regression parameter vector $\boldsymbol{\beta}_1$ is contained in the portion of the likelihood contributed by the marginal probability density function f_N while the information for $(\tilde{\boldsymbol{\beta}}_2, \beta_N)$ is in the portion from the conditional density $f_{\bar{Y}|N}$. We can then write the log-likelihood as:

$$\ell(\boldsymbol{\theta}; \mathbf{y}, \mathbf{n}) = \ell_{\bar{Y}|N}(\tilde{\boldsymbol{\beta}}_2, \beta_N; \mathbf{y}|\mathbf{n}) + \ell_N(\boldsymbol{\beta}_1; \mathbf{n}). \tag{5.6}$$

It follows that the information for $\boldsymbol{\beta}_1$ is contained in the marginal log-likelihood $\ell_N(\boldsymbol{\beta}_1; \mathbf{n})$ while the information for $(\tilde{\boldsymbol{\beta}}_2, \beta_N)$ is contained in the conditional log-likelihood $\ell_{\bar{Y}|N}(\tilde{\boldsymbol{\beta}}_2, \beta_N; \mathbf{y}|\mathbf{n})$. Due to the separable nature of the likelihood, we have that $(\tilde{\boldsymbol{\beta}}_2, \beta_N)$ and $\boldsymbol{\beta}_1$ are orthogonal parameters, that is, the Fisher Information matrix is diagonal. Thus, we can consider inference on $(\tilde{\boldsymbol{\beta}}_2, \beta_N)$ and $\boldsymbol{\beta}_1$ separately.

By the assumptions of Model D, we have that

$\mu_{i1} = \exp(\mathbf{X}_{i1}^\top\boldsymbol{\beta}_1)$ where $\boldsymbol{\beta}_1$ is a $p_1 \times 1$ vector of parameters

$\mu_{i2}^D = \exp(\tilde{\mathbf{X}}_{i2}^\top\tilde{\boldsymbol{\beta}}_2 + N_i\beta_N)$ where $\boldsymbol{\beta}_2$ is a $p_2 \times 1$ vector of parameters.

We then have that:

◇ $\dfrac{\partial \mu_{i1}}{\partial \beta_{1k}} = x_{i1k}\mu_{i1}, \quad \dfrac{\partial \mu_{i2}^D}{\partial \beta_{1k}} = 0 \quad \text{for } k = 1, ..., p_1,$

◇ $\dfrac{\partial \mu_{i2}^D}{\partial \beta_{2t}} = \tilde{x}_{i2t}\mu_{i2}^D, \quad \dfrac{\partial \mu_{i1}}{\partial \beta_{2t}} = 0, \quad \text{for } t = 1, ..., p_2,$

◇ $\dfrac{\partial \mu_{i2}^D}{\partial \beta_N} = N_i\mu_{i2}^D, \quad \dfrac{\partial \mu_{i1}}{\partial \beta_N} = 0.$

The score equations for the regression parameters $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \beta_N$ are thus:

$$\begin{aligned}
(i) \quad s(\beta_{1k}; \mathbf{y}, \mathbf{n}) &= \frac{\partial}{\partial \beta_{1k}} \ell(\boldsymbol{\beta}_1, \tilde{\boldsymbol{\beta}}_2, \beta_N; \mathbf{y}, \mathbf{n}) \\
&= \frac{\partial}{\partial \beta_{1k}} \left\{ \ell_{\bar{Y}|N}(\tilde{\boldsymbol{\beta}}_2, \beta_N; \mathbf{y}|\mathbf{n}) + \ell_N(\boldsymbol{\beta}_1; \mathbf{n}) \right\} \\
&= \frac{\partial}{\partial \beta_{1k}} \ell_N(\boldsymbol{\beta}_1; \mathbf{n}) \\
&= s(\beta_{1k}; \mathbf{n}), \quad k = 1, ..., p_1. \qquad (5.7)
\end{aligned}$$

$$\begin{aligned}
(ii) \quad s(\tilde{\beta}_{2t}; \mathbf{y}, \mathbf{n}) &= \frac{\partial}{\partial \tilde{\beta}_{2t}} \ell(\boldsymbol{\beta}_1, \tilde{\boldsymbol{\beta}}_2, \beta_N; \mathbf{y}, \mathbf{n}) \\
&= \frac{\partial}{\partial \tilde{\beta}_{2t}} \left\{ \ell_{\bar{Y}|N}(\tilde{\boldsymbol{\beta}}_2, \beta_N; \mathbf{y}|\mathbf{n}) + \ell_N(\boldsymbol{\beta}_1; \mathbf{n}) \right\} \\
&= \frac{\partial}{\partial \tilde{\beta}_{2t}} \ell_{\bar{Y}|N}(\tilde{\boldsymbol{\beta}}_2, \beta_N; \mathbf{y}|\mathbf{n}) \\
&= s(\tilde{\beta}_{2t}; \mathbf{y}|\mathbf{n}), \quad t = 1, ..., p_2. \qquad (5.8)
\end{aligned}$$

$$\begin{aligned}
(iii) \quad s(\beta_N; \mathbf{y}, \mathbf{n}) &= \frac{\partial}{\partial \beta_N} \ell(\boldsymbol{\beta}_1, \tilde{\boldsymbol{\beta}}_2, \beta_N; \mathbf{y}, \mathbf{n}) \\
&= \frac{\partial}{\partial \beta_N} \left\{ \ell_{\bar{Y}|N}(\tilde{\boldsymbol{\beta}}_2, \beta_N; \mathbf{y}|\mathbf{n}) + \ell_N(\boldsymbol{\beta}_1; \mathbf{n}) \right\} \\
&= \frac{\partial}{\partial \beta_N} \ell_{\bar{Y}|N}(\tilde{\boldsymbol{\beta}}_2, \beta_N; \mathbf{y}|\mathbf{n}) \\
&= s(\beta_N; \mathbf{y}|\mathbf{n}). \qquad (5.9)
\end{aligned}$$

In the particular case where the frequency distribution is Poisson and the severity distribution is gamma, the score equations can be simplified as follows:

$$
\begin{aligned}
(i) \quad s(\beta_{1k};\mathbf{y},\mathbf{n}) &= s(\beta_{1k};\mathbf{n}) \\
&= \sum_{i=1}^{m}\left[n_i\frac{1}{\mu_{i1}}\frac{\partial\mu_{i1}}{\partial\beta_{1k}}-\frac{\partial\mu_{i1}}{\partial\beta_{1k}}\right]=\sum_{i=1}^{m}\left[\frac{n_i}{\mu_{i1}x_{i1k}\mu_{i1}-x_{i1k}\mu_{i1}}\right] \\
&= \sum_{i=1}^{m}x_{i1k}(n_i-\mu_{i1})=0, \qquad k=1,...,p_1. \qquad (5.10)
\end{aligned}
$$

$$
\begin{aligned}
(ii) \quad s(\tilde{\beta}_{2t};\mathbf{y},\mathbf{n}) &= s(\tilde{\beta}_{2t};\mathbf{y}|\mathbf{n}) \\
&= \frac{\partial}{\partial\tilde{\beta}_{2t}}\sum_{i=1}^{m}\frac{n_i}{\phi}\left[\frac{-y_i}{\mu_{i2}^D}+\ln\left(\frac{1}{\mu_{i2}^D}\right)\right]=\frac{\partial}{\partial\tilde{\beta}_{2t}}\sum_{i=1}^{m}\frac{n_i}{\phi}\left[\frac{-y_i}{\mu_{i2}^D}-\ln(\mu_{i2}^D)\right] \\
&= \sum_{i=1}^{m}\frac{n_i}{\phi}\left[-y_i\frac{-1}{(\mu_{i2}^D)^2}\frac{\partial\mu_{i2}^D}{\partial\tilde{\beta}_{2t}}-\frac{1}{\mu_{i2}^D}\frac{\partial\mu_{i2}^D}{\partial\tilde{\beta}_{2t}}\right] \\
&= \sum_{i=1}^{m}\frac{n_i}{\phi}\left[\frac{y_i}{\mu_{i2}^D}\tilde{x}_{i2t}\mu_{i2}^D-\frac{1}{\mu_{i2}^D}\tilde{x}_{i2t}\mu_{i2}^D\right]=\sum_{i=1}^{m}\frac{n_i}{\phi}\tilde{x}_{i2t}\left[\frac{y_i}{\mu_{i2}^D}-1\right] \\
&= \sum_{i=1}^{m}\frac{n_i}{\phi}\frac{\tilde{x}_{i2t}}{\mu_{i2}^D}(y_i-\mu_{i2}^D)=0, \qquad t=1,...,p_2. \qquad (5.11)
\end{aligned}
$$

$$
\begin{aligned}
(iii) \quad s(\beta_N;\mathbf{y},\mathbf{n}) &= s(\beta_N;\mathbf{y}|\mathbf{n}) \\
&= \frac{\partial}{\partial\beta_N}\sum_{i=1}^{m}\frac{n_i}{\phi}\left[\frac{-y_i}{\mu_{i2}^D}+\ln\left(\frac{1}{\mu_{i2}^D}\right)\right]=\frac{\partial}{\partial\beta_N}\sum_{i=1}^{m}\frac{n_i}{\phi}\left[\frac{-y_i}{\mu_{i2}^D}-\ln(\mu_{i2}^D)\right] \\
&= \sum_{i=1}^{m}\frac{n_i}{\phi}\left[-y_i\frac{-1}{(\mu_{i2}^D)^2}\frac{\partial\mu_{i2}^D}{\partial\beta_N}-\frac{1}{\mu_{i2}^D}\frac{\partial\mu_{i2}^D}{\partial\beta_N}\right] \\
&= \sum_{i=1}^{m}\frac{n_i}{\phi}\left[\frac{y_i}{\mu_{i2}^D}n_i\mu_{i2}^D-\frac{1}{\mu_{i2}^D}n_i\mu_{i2}^D\right]=\sum_{i=1}^{m}\frac{n_i}{\phi}n_i\left[\frac{y_i}{\mu_{i2}^D}-1\right] \\
&= \sum_{i=1}^{m}\frac{n_i}{\phi}\frac{n_i}{\mu_{i2}^D}(y_i-\mu_{i2}^D)=0. \qquad (5.12)
\end{aligned}
$$

Note that the score equation for the regression parameters associated with the marginal frequency GLM, e.g. $s(\beta_{1k};\mathbf{n})$ for $k=1,...,p_1$, is identical to that obtained under the assumption of independence as in the Model I. Thus, the regular maximum likelihood properties hold and so $\hat{\boldsymbol{\beta}}_1$ is asymptotically normally distributed with $\sqrt{m}(\hat{\boldsymbol{\beta}}_1-\boldsymbol{\beta}_1)\sim\mathcal{N}(\mathbf{0},\{I(\boldsymbol{\beta}_1)\}^{-1})$, where $I(\boldsymbol{\beta}_1)$ can be estimated by $I(\hat{\boldsymbol{\beta}}_1)$.

The maximum likelihood estimates for the severity parameters $\hat{\beta}_N$ and $\hat{\beta}_{21}, ..., \hat{\beta}_{2p_2}$ are not based on a regular likelihood equation but rather on a conditional likelihood. It follows that these estimates are conditional maximum-likelihood estimates. Andersen (1970) discusses the asymptotic properties of conditional maximum likelihood estimators (CMLE) and shows that, under some regularity assumptions, the CMLE $\hat{\theta}$ is consistent and is asymptotically normally distributed, with mean equal to the true parameter value θ_0 and asymptotic variance equal to

$$\sum_{i=1}^{m} \mathbb{E}_Y \left[\left\{ \frac{\partial \ln f_{\bar{Y}|N}(y_i|\theta, n_i)}{\partial \theta} \right\}^2 \right] = -\sum_{i=1}^{m} \mathbb{E}_Y \left[\frac{\partial^2 \ln f_{\bar{Y}|N}(y_i|\theta, n_i)}{\partial \theta^2} \right].$$

It follows from equations (5.11) and (5.12) that:

$$\triangleright \quad \frac{\partial^2}{\partial \beta_{2k}^2} \ell_{\bar{Y}|N}(\tilde{\beta}_2, \beta_N; \mathbf{y}|\mathbf{n}) = \sum_{i=1}^{m} \frac{n_i}{\phi} \tilde{x}_{i2k} \left(-\frac{y_i}{(\mu_{i2}^D)^2} \frac{\partial \mu_{i2}^D}{\partial \tilde{\beta}_{2k}} \right)$$
$$= \sum_{i=1}^{m} \frac{n_i}{\phi} \tilde{x}_{i2k} \left(-\frac{y_i}{(\mu_{i2}^D)^2} \tilde{x}_{i2k} \mu_{i2}^D \right)$$
$$= \sum_{i=1}^{m} \frac{n_i}{\phi} \tilde{x}_{i2k}^2 \left(-\frac{y_i}{\mu_{i2}^D} \right),$$

$$\triangleright \quad \frac{\partial^2}{\partial \tilde{\beta}_{2k} \tilde{\beta}_{2l}} \ell_{\bar{Y}|N}(\tilde{\beta}_2, \beta_N; \mathbf{y}|\mathbf{n}) = \sum_{i=1}^{m} \frac{n_i}{\phi} \tilde{x}_{i2k} \tilde{x}_{i2l} \left(-\frac{y_i}{\mu_{i2}^D} \right),$$

$$\triangleright \quad \frac{\partial^2}{\partial \beta_N^2} \ell_{\bar{Y}|N}(\tilde{\beta}_2, \beta_N; \mathbf{y}|\mathbf{n}) = \sum_{i=1}^{m} \frac{n_i}{\phi} n_i \left(-\frac{y_i}{(\mu_{i2}^D)^2} \frac{\partial \mu_{i2}^D}{\partial \tilde{\beta}_N} \right)$$
$$= \sum_{i=1}^{m} \frac{n_i^2}{\phi} n_i \left(-\frac{y_i}{\mu_{i2}^D} \right)$$
$$= \sum_{i=1}^{m} \frac{n_i^3}{\phi} \left(-\frac{y_i}{\mu_{i2}^D} \right),$$

$$\triangleright \quad \frac{\partial^2}{\partial \beta_N \tilde{\beta}_{2k}} \ell_{\bar{Y}|N}(\tilde{\beta}_2, \beta_N; \mathbf{y}|\mathbf{n}) = \sum_{i=1}^{m} \frac{n_i^2}{\phi} \tilde{x}_{i2k} \tilde{x}_{i2l} \left(-\frac{y_i}{\mu_{i2}^D} \right).$$

In expectation, the above second partial derivatives are:

$$\triangleright \quad \mathbb{E}_{\bar{Y}|N}\Big[\frac{\partial^2}{\partial\tilde{\beta}_{2k}^2}\ell_{\bar{Y}|N}(\tilde{\beta}_2,\beta_N;\mathbf{y}|\mathbf{n})\Big] = -\sum_{i=1}^{m}\frac{n_i}{\phi}\tilde{x}_{i2k}^2,$$

$$\triangleright \quad \mathbb{E}_{\bar{Y}|N}\Big[\frac{\partial^2}{\partial\tilde{\beta}_{2k}\tilde{\beta}_{2l}}\ell_{\bar{Y}|N}(\tilde{\beta}_2,\beta_N;\mathbf{y}|\mathbf{n})\Big] = -\sum_{i=1}^{m}\frac{n_i}{\phi}\tilde{x}_{i2k}\tilde{x}_{i2l},$$

$$\triangleright \quad \mathbb{E}_{\bar{Y}|N}\Big[\frac{\partial^2}{\partial\beta_N}\ell_{\bar{Y}|N}(\tilde{\beta}_2,\beta_N;\mathbf{y}|\mathbf{n})\Big] = -\sum_{i=1}^{m}\frac{n_i^3}{\phi},$$

$$\triangleright \quad \mathbb{E}_{\bar{Y}|N}\Big[\frac{\partial^2}{\partial\tilde{\beta}_{2k}\beta_N}\ell_{\bar{Y}|N}(\tilde{\beta}_2,\beta_N;\mathbf{y}|\mathbf{n})\Big] = -\sum_{i=1}^{m}\frac{n_i^2}{\phi}\tilde{x}_{i2k}.$$

It follows that the observed information matrix is then:

$$I(\hat{\boldsymbol{\beta}}_2,\hat{\beta}_N)_{p_2+1,p_2+1} = \frac{1}{\phi}\begin{pmatrix} \sum_{i=1}^m n_i\tilde{x}_{i21}^2 & \sum_{i=1}^m n_i\tilde{x}_{i21}\tilde{x}_{i22} & \cdots & \cdots \\ \sum_{i=1}^m n_i\tilde{x}_{i22}\tilde{x}_{i21} & \sum_{i=1}^m n_i\tilde{x}_{i22}^2 & \cdots & \cdots \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^m n_i\tilde{x}_{i2p_2}\tilde{x}_{i21} & \sum_{i=1}^m n_i\tilde{x}_{i2p_2}\tilde{x}_{i22} & \cdots & \cdots \\ \sum_{i=1}^m n_i^2\tilde{x}_{i21} & \sum_{i=1}^m n_i^2\tilde{x}_{i22} & \cdots & \sum_{i=1}^m n_i^3 \end{pmatrix}$$

It follows that the MLE $(\hat{\boldsymbol{\beta}}_2,\hat{\beta}_N)$ is asymptotically normally distributed with

$$\sqrt{m}((\hat{\boldsymbol{\beta}}_2,\hat{\beta}_N)-(\boldsymbol{\beta}_2,\beta_N)) \sim \mathcal{N}\left(\mathbf{0},\{I(\boldsymbol{\beta}_2,\beta_N)\}^{-1}\right),$$

where $I(\boldsymbol{\beta}_2,\beta_N)$ can be estimated by $I(\hat{\boldsymbol{\beta}}_2,\hat{\beta}_N)$.

Chapter 6

Example

We will now apply the GLM models for the aggregate claims in both the independent and dependent cases, as developed in Chapters 4 and 5. Both of these models will be tested using insurance data and then compared in order to quantify the effect of dependence in the aggregate claims models.

6.1 Data Description

The dataset consists of automobile insurance policies in Canada. The model was fit to the collision claims experience for the years 2003 through 2005. Note that the claims experience for years 2006 through 2008 was used as a hold out dataset for cross-validation and to further test the fit of the models derived from the first dataset.

As mentioned above, the models were fit to the claims categorized under the collision insurance coverage. The collision coverage, in particular, reimburses the policyholder for car damages caused by an accident with another vehicle or object. The claim payment will only cover damages caused from an actual car collision and does not include damages due to theft, vandalism, weather, etc.

From the original dataset, only those policies with at least two weeks of exposure were kept as to avoid spurious observations. The final dataset for the frequency model was comprised of $799,877$ observations. The data used to fit the severity models, which is the subset of the frequency dataset with positive claim counts, consisted of

18, 895 observations.

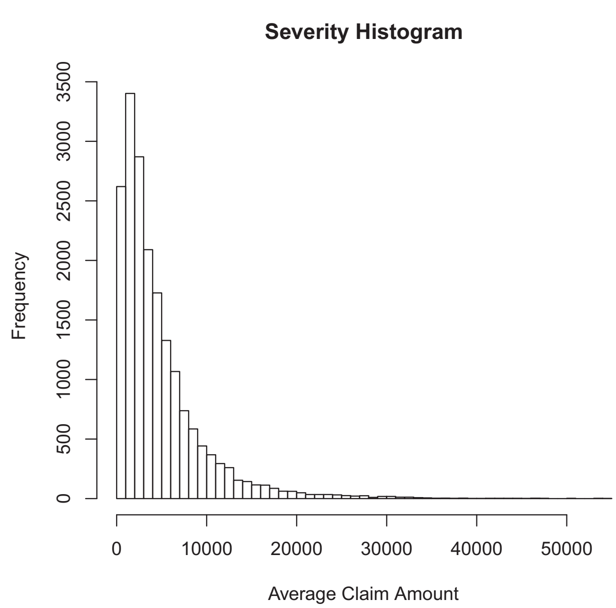
As expected in a car insurance portfolio, very few policies made any claims. The claim counts ranged from 0 to a maximum of 3 claims. The distribution of the claim counts is given in Table 6.1.

Table 6.1: Claim Count Distribution

Claim Count	**Frequency**	**Percent**
0	780,982	97.64%
1	18,584	2.32%
2	301	0.04%
3	10	0.00%
Total	799,877	100.00%

Given a claim had occurred, the average claim severity ranged from \$5.65 to \$54, 998.75. Figure 6.1 provides the histogram of the average severity for positive claim counts.

Figure 6.1: Average Severity Histogram



We also have that the distribution of average claim severity by claim count is as in Table 6.2.

Table 6.2: Average Claim Severity by Claim Count

Claim Count	**Average Claim Severity**
0	–
1	\$4,757.34
2	\$4,120.52
3	\$3,600.08
Overall	\$4,746.59

We can see that as the number of claims increases, the average claim amount decreases. This suggests that we can expect the regression parameter β_N to be negative.

Note that when the claim count is zero, the aggregate claim amount is also zero by definition. Thus, an average claim severity is only defined for policies with positive claim counts and is otherwise set to zero. If we consider the dataset used for the frequency model, that is, policies with both zero and positive claim counts, the following correlation statistics were obtained between the frequency and severity components:

- Pearson's: $\rho^{P}_{N,\bar{Y}} = 0.6814$,
- Spearman's: $\rho^{S}_{N,\bar{Y}} = 0.9999$.

These correlation statistics suggest that the claim counts and amounts are strongly positively correlated. This could be due to the fact that when there are no claims the severity is set to zero and for this dataset in particular, 98% of observations are at $(N_i = 0, \bar{Y}_i = 0)$.

When we consider only the severity model subset of the data, that is, those policies with positive claim counts, the correlations are significantly different:

- Pearson's: $\rho^{P}_{N,\bar{Y}} = -0.0170$,
- Spearman's: $\rho^{S}_{N,\bar{Y}} = 0.0045$.

Note that similar correlation values were obtained when considering the entire dataset over the years 2003 through 2008. The correlations for all policies is:

- ⋄ Pearson's: $\rho^P_{N,\bar{Y}} = 0.6882$,
- ⋄ Spearman's: $\rho^S_{N,\bar{Y}} = 0.9999$.

While the correlations for policies with positive claim counts is:

- ⋄ Pearson's: $\rho^P_{N,\bar{Y}} = -0.0151$,
- ⋄ Spearman's: $\rho^S_{N,\bar{Y}} = 0.0046$.

Here the correlations are essentially indicating no relation between the frequency and severity components. However, we must bear in mind that in the severity sub-dataset, 98% of the policies make only one claim. Perhaps there are not enough observations at the higher claim counts to reflect the relation between the frequency and severity components.

It is also important to note that the illustration done here is with one particular coverage in auto insurance, namely collision. We can expect the relation between the claim frequency and severity to be different for other car insurance coverages, as well as for different lines of business, such as home insurance.

Several rating variables were included in the dataset, specifically, the deductible, the driver's age, gender, and marital status, the number of years the driver has been licensed, the number of years the policy has been with the company, the vehicle type and finally the vehicle age. The gender, marital status and vehicle type variables were used as factor covariates in the model, while the others were considered continuous covariates.

When using GLMs to model data, it is important to ensure that the columns of the design matrix are orthogonal so that there are no issues of multicollinearity. If we consider the continuous rating variables included in the collision car insurance claims, we can check for multicollinearity by considering the correlation between rating variables. Table 6.3 provides the correlation matrix of the continuous covariates included in the dataset, with X_1 = *deductible*, X_2 = *age*, X_3 = *number of years licensed* X_4 = *number years policy with company*, and X_5 = *vehicle age*.

Table 6.3: Continuous Covariates Correlation Matrix

	X_1	X_2	X_3	X_4	X_5
X_1	1	-0.086	-0.077	-0.081	-0.160
X_2	-0.086	1	0.842	0.338	0.071
X_3	-0.077	0.842	1	0.384	0.040
X_4	-0.081	0.338	0.384	1	0.087
X_5	-0.160	0.071	0.040	0.087	1

Naturally, there is a high correlation between the number of years the driver is licensed and the driver's age. This correlation between explanatory variables could cause multicollinearity and lead to inaccurate regression parameter estimates. We can then consider the impact on the regression parameter estimates and standard errors for correlated variables by comparing a model with both variables included with a model that only includes one of the correlated rating variables. This issue will be further investigated with the frequency and severity GLMs selected in both the independent and dependent models.

6.2 Modelling the Data

The *glm* function in R was used to model and analyse the dataset. A Poisson model was used for the marginal frequency GLM while gamma responses were assumed for the severity models. For both the frequency and severity models, a log link function was used. The analysis of deviance was used to determine the best model for both the frequency and severity marginal models, as described in Chapter 3, by comparing nested models. Only main effects models were analysed and interactions were not considered so as to simplify the process and allow for a clearer interpretation of the effect of dependence in the aggregate claims model. Given the limited rating variables included in the dataset in the first place, the goal here was not necessarily to find the best possible model to describe the data but rather to compare the effects of extending the independent model to the dependent setting.

When modelling the dataset, we had to ensure that all observations were treated in a consistent manner by taking into account each individual's exposure to risk. Since not all policies were necessarily insured for a full year, we had to make adjustments such that all observations were considered in a congruent way. This adjustment was done by using an offset for the exposure variable in the frequency model, where the exposure variable indicates what portion of the year the individual was insured for. Thus, an exposure unit equal to 1 means that the policy was insured for a full year, while an exposure unit of, say, 0.5 implies the individual was insured for only half a year. Note that the claim counts N_i in the dataset represent the total number of claims incurred over the insured time period. Here, we assume that the claim counts follow a Poisson distribution. It then follows that as the exposure increases, the expected claim count will increase proportionally. In using an offset in the GLM for N_i, we are essentially including the exposure variable as a fixed effect with regression coefficient equal to 1. If we denote the exposure variable by t_i, then the GLM for N_i with an exposure offset is as follows:

$$\ln\left(\frac{\mu_{i1}}{t_i}\right) = \mathbf{X}_{i1}^{\top}\boldsymbol{\beta}_1$$

$$\Rightarrow \ln(\mu_{i1}) = \ln(t_i) + \mathbf{X}_{i1}^{\top}\boldsymbol{\beta}_1,$$

where $\ln(t_i)$ is the exposure offset term. Thus, on the mean scale we have:

$$\mu_{i1} = t_i \, \exp(\mathbf{X}_{i1}^{\top}\boldsymbol{\beta}_1),$$

so that we can interpret the term $\exp(\mathbf{X}_{i1}^{\top}\boldsymbol{\beta}_1)$ as a yearly expected claim count.

Note that it was not necessary to use an offset for the severity models since we are modelling $\bar{Y}_i$, that is, the average loss amount per claim occurrence. It follows that the exposure variable will not effect the expected average claim severity in a fixed proportional manner. Thus, there is no need to adjust the expected claim severity for the exposure.

The weights to be given to each observation in the frequency models were taken to be 1. However, in the severity models, the claim counts were used as weights. Recall from Chapters 4 and 5, when modelling the average claim amounts, we have that $\bar{Y}_i \sim ED\left(\mu_{i2}, \frac{\phi}{N_i}\right)$. It follows that we have $a_i(\phi) = \phi/n_i$ and thus the claim counts n_i must be used as weights in the GLM.

Note that using the standard residuals as defined in Chapter 3 to test the model fit and adequacy was not straightforward in this analysis. The standard definitions described in Chapter 3 are appropriate for response variables that are members of the ED family, however, here the variable of interest is the aggregate losses S_i. Although both the frequency and severity components of the aggregate claims are assumed to follow ED models, we cannot conclude the same for the total claims by only analysing the first and second moments of S_i. We could use any type of residual on the marginal components, N_i and $\bar{Y}_i$, however the interpretation for the final model on S_i is not so clear. Moreover, in this analysis, the marginal frequency mean $\mathbb{E}[N_i|\mathbf{X}_i]$ and the marginal severity means in the independent model and dependent models, $\mathbb{E}[\bar{Y}_i|\mathbf{X}_i]$ and $\mathbb{E}[\bar{Y}_i|\mathbf{X}_i, N_i]$ respectively, are only of secondary interest and are more of an intermediate step. Consequently, the interpretation of the residuals on the claim counts and average claim amounts is not so straightforward with respect to the expected loss cost. It is also not clear how to define a standard residual directly for the model on the mean aggregate claim amount.

We can, however, consider the simple response residual as the discrepancy between the fitted value and the observation:

$$r_{Ri} = s_i - \hat{\mu}_i.$$

We can also consider a modified Pearson residual with the denominator being the variance of the response $\mathbb{V}\text{ar}[S_i]$ rather than the variance function $V(\mu_i)$ so that the residual becomes:

$$r^*_{Pi} = \frac{s_i - \hat{\mu}_i}{\sqrt{\mathbb{V}\text{ar}[S_i]}}.$$

Another alternative is to use a modified version of the deviance residual: rather than using the deviance D_i, we can consider the scaled deviance D_i^* where

$$\begin{aligned}
D^* &= \frac{1}{\phi}D = 2\{\ell(\tilde{\boldsymbol{\theta}};\phi,\mathbf{y}) - \ell(\hat{\boldsymbol{\theta}};\phi,\mathbf{y})\} \\
&= 2\left\{\left[\ell_{\bar{Y}|N}(\tilde{\boldsymbol{\theta}}_2;\phi_2,\mathbf{y}|\mathbf{n}) + \ell_N(\tilde{\boldsymbol{\theta}}_2;\phi_1,\mathbf{n})\right] - \left[\ell_{\bar{Y}|N}(\hat{\boldsymbol{\theta}}_2;\phi_2,\mathbf{y}|\mathbf{n}) + \ell_N(\hat{\boldsymbol{\theta}}_1;\phi_1,\mathbf{n})\right]\right\} \\
&= \left[\ell_{\bar{Y}|N}(\tilde{\boldsymbol{\theta}}_2;\phi_2,\mathbf{y}|\mathbf{n}) - \ell_{\bar{Y}|N}(\hat{\boldsymbol{\theta}}_2;\phi_2,\mathbf{y}|\mathbf{n})\right] + \left[\ell_N(\tilde{\boldsymbol{\theta}}_2;\phi_1,\mathbf{n}) - \ell_N(\hat{\boldsymbol{\theta}}_1;\phi_1,\mathbf{n})\right] \\
&= D^*_{\bar{Y}|N} + D^*_N.
\end{aligned}$$

Then the scaled deviance residual can be defined as:

$$r^*_{Di} = \text{sign}(s_i - \hat{\mu}_i)\sqrt{D_i^*} = \text{sign}(s_i - \hat{\mu}_i)\sqrt{D^*_{\bar{Y}|N,i} + D^*_{N,i}}.$$

Note that in the independent model, we will have $D^* = D^*_{\bar{Y}} + D^*_N$ and then the scaled deviance residual is $r^*_{D_i} = \text{sign}(s_i - \hat{\mu}_i)\sqrt{D_i^*} = sign(s_i - \hat{\mu}_i)\sqrt{D^*_{\bar{Y},i} + D^*_{N,i}}$.

The analysis of deviance, as described in Chapter 3, was used to establish the best fitting models for the marginal frequency and severity models respectively. Recall from equation (3.1.4):

$$D^*(\mathbf{y},\hat{\boldsymbol{\mu}}) = \frac{D(\mathbf{y},\hat{\boldsymbol{\mu}})}{\phi} \sim \chi^2_{m-p}.$$

This can be used to assess the significance of an individual model. We also have from equation (3.1.4) that:

$$\frac{D(\mathbf{y},\hat{\boldsymbol{\mu}}_{\boldsymbol{A}}) - D(\mathbf{y},\hat{\boldsymbol{\mu}}_{\boldsymbol{B}})}{\phi} \sim \chi^2_{p_B - p_A}.$$

This equation allows to compare nested models and assess whether the simpler model is an acceptable simplification of the more complex model. The deviance statistic thus allowed us to assess the model adequacy and ultimately conclude which model was most suitable for the data.

6.3 Independent Model

The independent model, as described in Chapter 4, was fit to the data assuming a Poisson frequency distribution and a gamma severity distribution.

6.3.1 Frequency Model I

The full main effects model is as follows:

$$\begin{aligned}\ln\left(\mathbb{E}[N_i|\mathbf{X}_i]\right) = \ln(exposure)\ &+\ deductible\ +\ age\ +\ gender\ +\ marital\ status \\ &+\ number\ of\ years\ licensed\ +\ number\ years\ policy\ with\ insurer \\ &+\ vehicle\ type\ +\ vehicle\ age,\end{aligned}$$

where the term $\ln(exposure)$ is the exposure offset previously mentioned. All terms, with the exception of gender, were found to be significant in the model at the 0.1% level. The next model fit to the data dropped the gender term from the GLM. Since this updated model is nested in the main effects model, we can use the analysis of deviance to assess whether the simplified model is significant. The change in residual deviance between the two models is -0.40623 and the difference in the degree of freedom is 1. Comparing the difference in deviance with the 95^{th} percentile of a Chi-Square distribution with 1 degree of freedom, 3.841459, confirms that the simplified model is indeed an adequate simplification. All terms in this model were found to be significant at the 5% level and consequently dropping any additional terms would not produce an adequate simplification to the model. Thus, we can conclude that the best fitting model for the mean frequency is:

$$\begin{aligned}\hat{\mu}_{i1}^{I} = \exp\{&\beta_{1,0} + \beta_{1,1}\ deductible + \beta_{1,2}\ age + \beta_{1,3}\ marital\ status \\ &+ \beta_{1,4}\ number\ of\ years\ licensed + \beta_{1,5}\ number\ years\ policy\ with\ insurer \\ &+ \beta_{1,6}\ vehicle\ type + \beta_{1,7}\ vehicle\ age\}. \end{aligned} \tag{6.1}$$

Note that since we are assuming Poisson responses, the dispersion parameter should be equal to 1. A rough estimate for the dispersion parameter can be taken as $D(\mathbf{y};\hat{\mu}_{i1})/m - p_1 \approx \phi_1$ since $\mathbb{E}\left[\frac{D(\mathbf{y};\hat{\mu}_{i1})}{\phi_1}\right] = m - p_1$. However, for this GLM we have $D(\mathbf{y};\hat{\mu}_{i1})/m - p_1 = 145568/799848 = 0.1819946$. This is evidence of underdispersion in the model. This could imply that the distribution assumption of Poisson responses is perhaps inadequate. Since there is a substantial amount of policies with zero claim count, a zero-inflated Poisson distribution assumption could potentially provide a better fit.

6.3.2 Severity Model I

Similarly, for the severity model in the independent case, a series of nested models were fit until an adequate simplification was found. The following models were fit to the data:

i) main effects model - I (full model)

$$\begin{aligned}\ln\left(\mathbb{E}[\bar{Y}_i|\mathbf{X}_i]\right) &= \textit{deductible coll} \ + \ \textit{age} \ + \ \textit{gender} + \ \textit{marital status} \\ &+ \ \ \textit{number of years licensed} \ + \ \textit{number years policy with insurer} \\ &+ \ \textit{vehicle type} \ + \ \textit{vehicle age}.\end{aligned}$$

ii) model m1-I

$$\begin{aligned}\ln\left(\mathbb{E}[\bar{Y}_i|\mathbf{X}_i]\right) &= \textit{deductible coll} \ + \ \textit{gender} + \ \textit{marital status} \\ &+ \ \ \textit{number of years licensed} \ + \ \textit{number years policy with insurer} \\ &+ \ \textit{vehicle type} \ + \ \textit{vehicle age}.\end{aligned}$$

iii) model m2-I

$$\begin{aligned}\ln\left(\mathbb{E}[\bar{Y}_i|\mathbf{X}_i]\right) &= \textit{deductible coll} \ + \ \textit{marital status} \ + \ \ \textit{number of years licensed} \\ &+ \ \textit{number years policy with insurer} \ + \ \textit{vehicle type} \ + \ \textit{vehicle age}.\end{aligned}$$

iv) model m3-I

$$\begin{aligned}\ln\left(\mathbb{E}[\bar{Y}_i|\mathbf{X}_i]\right) &= \textit{deductible coll} \ + \ \textit{marital status} \ + \ \ \textit{number of years licensed} \\ &+ \ \textit{vehicle type} \ + \ \textit{vehicle age}.\end{aligned}$$

v) model m4-I

$$\begin{aligned}\ln\left(\mathbb{E}[\bar{Y}_i|\mathbf{X}_i]\right) &= \textit{deductible coll} \ + \ \ \textit{number of years licensed} \\ &+ \ \textit{vehicle type} \ + \ \textit{vehicle age}.\end{aligned}$$

The analysis of scaled deviance was used to compare these nested models and also to test the final model (item v) with the null model (which fits only an intercept). Table 6.4 list the deviances obtained for each of the above mentioned models.

Table 6.4: Model I - Severity GLMs

Model	Deviance	Degress of Freedom	Dispersion
sev main effects-I	16161	18865	0.7982221
m1-I	16161	18866	0.7980804
m2-I	16162	18867	0.7981926
m3-I	16163	18868	0.7976447
m4-I	16168	18871	0.7976825
null model	17633	18894	1.092249

We can see that the differences in the deviance and dispersion between the nested models are very small. This indicates that the simpler model is adequate. For example, comparing models m3-I and m4-I, we have that the difference in scaled deviance is roughly 6.27. Comparing this with the 95^{th} percentile of the χ^2 distribution with 3 degrees of freedom, 7.81, we can conclude that model m4-I is an adequate simplification of m3-I. That is, setting the regression parameter for the marital status covariate to zero is an acceptable hypothesis.

The remaining covariates in model m4-I were all found to be significant at the 0.1% level. It follows that the best fit for the severity component in the independent model is m4-I:

$$\begin{aligned} \hat{\mu}_{i2}^{I} = \exp\big\{&\beta_{2,0} + \beta_{2,1}\, deductible + \beta_{2,2}\, number\ of\ years\ licensed \\ &+ \beta_{2,3}\, vehicle\ type + \beta_{2,4}\, vehicle\ age\big\}. \end{aligned} \tag{6.2}$$

6.3.3 Aggregate Claims Model I

By the independence assumption in model I, it follows that the expected aggregate claims is the product of the marginal mean frequency and mean severity so that:

$$\mathbb{E}[S_i|\mathbf{X}_i] = \mathbb{E}[\bar{Y}_i|\mathbf{X}_i] \times \mathbb{E}[N_i|\mathbf{X}_i] = \hat{\mu}_{i1}^{I} \times \hat{\mu}_{i2}^{I} = \hat{\mu}_{i}^{I}. \tag{6.3}$$

6.4 Dependent Model

The dependent model, as described in Chapter 5, was fit to the same dataset used for fitting the independent model. Again, we assumed a Poisson frequency distribution and a gamma severity distribution.

6.4.1 Frequency Model D

In the dependent model, the marginal frequency GLM is modelled the same way as in the independent model. Thus, the final model for $\hat{\mu}_{i1}^{D}$ is taken as $\hat{\mu}_{i1}^{I}$:

$$\begin{aligned}\hat{\mu}_{i1}^{D} = \exp\{&\beta_{1,0} + \beta_{1,1}\, deductible + \beta_{1,2}\, age + \beta_{1,3}\, marital\ status \\ &+ \beta_{1,4}\, number\ of\ years\ licensed + \beta_{1,5}\, number\ years\ policy\ with\ insurer \\ &+ \beta_{1,6}\, vehicle\ type + \beta_{1,7}\, vehicle\ age\}. \qquad (6.4)\end{aligned}$$

6.4.2 Severity Model D

The procedure for determining the best fitting severity model in the dependent setting is similar to that of the independent model, however, now an additional covariate is included for the claim counts N_i. Again, a series of nested models were fit to the data, as follows:

i) main effects model - D (full model)

$$\begin{aligned}\ln\left(\mathbb{E}[\bar{Y}_i|\mathbf{X}_i, N_i]\right) = N_i + &\ deductible\ coll\ +\ age\ +\ gender +\ marital\ status \\ &+\ number\ of\ years\ licensed\ +\ number\ years\ policy\ with\ insurer \\ &+\ vehicle\ type\ +\ vehicle\ age.\end{aligned}$$

ii) m1-D

$$\begin{aligned}\ln\left(\mathbb{E}[\bar{Y}_i|\mathbf{X}_i, N_i]\right) = N_i + &\ deductible\ coll\ +\ gender +\ marital\ status \\ &+\ number\ of\ years\ licensed\ +\ number\ years\ policy\ with\ insurer \\ &+\ vehicle\ type\ +\ vehicle\ age.\end{aligned}$$

iii) m2-D

$$\ln\left(\mathbb{E}[\bar{Y}_i|\mathbf{X}_i, N_i]\right) = N_i + \textit{ deductible coll } + \textit{ marital status}$$
$$+ \textit{ number of years licensed } + \textit{ number years policy with insurer}$$
$$+ \textit{ vehicle type } + \textit{ vehicle age.}$$

vi) m3-D

$$\ln\left(\mathbb{E}[\bar{Y}_i|\mathbf{X}_i, N_i]\right) = N_i + \textit{ deductible coll } + \textit{ marital status}$$
$$+ \textit{ number of years licensed } + \textit{ vehicle type } + \textit{ vehicle age.}$$

v) m4-D

$$\ln\left(\mathbb{E}[\bar{Y}_i|\mathbf{X}_i, N_i]\right) = N_i + \textit{ deductible coll } + \textit{ number of years licensed}$$
$$+ \textit{ vehicle type } + \textit{ vehicle age.}$$

Similarly to what was done in the independent model setting, the analysis of scaled deviance was used to assess the model fit. Table 6.5 list the deviances obtained for each of the above mentioned models.

Table 6.5: Model D - Severity GLMs

Model	**Deviance**	**Degrees of Freedom**	**Dispersion**
main effects - D	16148	18864	0.7946048
m1-D	16149	18865	0.7944659
m2-D	16149	18866	0.7946156
m3-D	16151	18867	0.794058
m4-D	16155	18870	0.7940439
null model	17633	18894	1.092249

Similar to in the independent case, we have that the differences in the deviance and dispersion between the nested models are small. Based on the results in Table 6.5, we can conclude that model m4-D is the best fit for the severity model in the dependent case. Note that the covariates in model m4-D were all found to be significant at the 0.1% level. It follows that the conditional mean severity in the dependent model is:

$$\hat{\mu}_{i2}^{D} = \exp\left\{\tilde{\mathbf{X}}_{i2}^{\top}\tilde{\boldsymbol{\beta}}_2 + N_i\beta_N\right\},$$

and thus the modified marginal severity mean, $\tilde{\mu}_{i2} = \exp\left(\tilde{X}_{i2}^{\top}\tilde{\boldsymbol{\beta}}_2\right)$, is:

$$\begin{aligned}\tilde{\mu}_{i2} = \exp\big\{&\tilde{\beta}_{2,0} + \tilde{\beta}_{2,1}\, deductible + \tilde{\beta}_{2,2}\, number\ of\ years\ licensed \\ &+ \tilde{\beta}_{2,3}\, vehicle\, type + \tilde{\beta}_{2,4}\, vehicle\ age\big\}. \end{aligned} \tag{6.5}$$

and the correction term for the claim counts is:

$$\exp\left\{\mu_{i1}\left(e^{\beta_N} - 1\right) + \beta_N\right\}. \tag{6.6}$$

6.4.3 Aggregate Claims Model D

It follows from the dependent model assumption that the expected aggregate claims in Model D is then:

$$\hat{\mu}_{i}^{D} = \hat{\mu}_{i1}\hat{\tilde{\mu}}_{i2}\exp\left\{\hat{\mu}_{i1}\left(e^{\hat{\beta}_N} - 1\right) + \hat{\beta}_N\right\}. \tag{6.7}$$

6.5 Model Comparisons

The goal of this thesis was to assess the effect of extending the independent aggregate claims model to the dependent case. In particular, we focused on the effect of dependence between frequency and severity on the expected total loss cost. Thus, we are interested in the effect of the model formulation on $\mathbb{E}[S_i|\mathbf{X}_i] = \mu_i$. In the independent model, we have that the expected aggregate claims is the product of the marginal frequency and severity means, as determined in a GLM framework:

$$\mu_i^I = \mu_{i1} \times \mu_{i2}.$$

In the dependent model, the expected total claims becomes the product of the marginal frequency mean, a modified severity mean and a correction term:

$$\mu_i^D = \mu_{i1} \times \tilde{\mu}_{i2} \times \exp\left\{\mu_{i1}\left(e^{\beta_N} - 1\right) + \beta_N\right\}.$$

Thus, any difference between μ_i^I and μ_i^D is generated by the change in the marginal mean severity as well as the correction term. The difference between the mean severity in the independent and dependent cases, $\hat{\mu}_{i2}^I$ and $\hat{\mu}_{i2}^D$ respectively, is caused by the presence of the claim count N_i as a covariate in the GLM for model D. As discussed in Chapter 5, the difference between the expected loss cost in the independent and dependent models is not straightforward as $\beta_N > 0$ ($\beta_n < 0$) will cause the correction term to be greater than 1 (less than 1), but this increase (decrease) could be offset by the change in the marginal severity mean from μ_{i2} to the modified $\tilde{\mu}_{i2}$.

In order to evaluate whether the dependent model is indeed significant, we can compare it with the independent model. Note that testing whether $\beta_N = 0$ can be done by comparing the final severity models in the independent and dependent cases respectively since the final independent severity model (m4-I) is nested in the final dependent severity model (m4-D). That is, setting $\beta_N = 0$ in the dependent model m4-D will retrieve the independent model m4-I. We can then compare the change in deviance with the χ_1^2 distribution. Here, taking $\hat{\phi} \approx 0.7940439$, we have that:

$$\frac{\left\{D(\mathbf{y}, \boldsymbol{\hat{\mu}_{i2}^I}) - D(\mathbf{y}, \boldsymbol{\hat{\mu}_{i2}^D})\right\}}{\hat{\phi}} = \frac{(16168 - 16155)}{0.7940439} = 16.37189.$$

If we compare this with the 95^{th} percentile of the χ_1^2 distribution, we have that the test statistic is much larger that the χ^2 statistic: $16.37189 > 3.841459$. Since the change in deviance is significantly larger than the percentile, we can conclude that the independent model is not an adequate simplification of the dependent model. This is evidence that there is a need for a dependent model for the aggregate claims.

Similarly, if we consider the Wald test statistic for the null hypothesis $H_0 : \beta_{0,N} = 0$ we obtain:

$$\frac{\hat{\beta}_N - \beta_{0,N}}{\sqrt{\mathbb{V}\text{ar}(\hat{\beta}_N)}} = \frac{-0.1397 - 0}{0.03377} = -4.136808.$$

We can thus conclude that the coefficient for the frequency covariate is strongly significant in the model.

Table 6.6 shows the deviances obtained for the fitted models. Note that the loss cost model deviance is the sum of the frequency and severity deviances. We can see that the dependent model has a lower deviance than the independent model.

Table 6.6: Deviance Comparison

Model	**Deviance**	**Scaled Deviance**
Frequency Model	$145,567.60$	$145,567.60$
Severity Model I	$16,167.56$	$20,361.04$
Severity Model D	$16,154.62$	$20,344.74$
Loss Cost Model I	$161,735.16$	$165,928.64$
Loss Cost Model D	$161,722.22$	$165,912.34$

On average, the coefficients for the modified severity GLM, $\hat{\hat{\boldsymbol{\beta}}}_2$, were increased by a mere 0.53% as compared to the independent marginal severity GLM regression parameters $\hat{\boldsymbol{\beta}}_2$. Some parameters were increased by as much as 5.38% and decreased by as much as −4.91%.

The regression parameter generated from the presence of the claim count in the modified severity GLM was estimated as $\hat{\beta}_N \approx -0.1396594$. This implies that the correction term will be less than 1 for all cases, and can be thought of as a discount multiplicative factor that will be applied to all policies. It follows that the increase seen for some of the regression parameters in the modified severity model could be offset by the correction term, while those coefficients that produced a decrease will be further decreased by the correction term.

Overall, the extent of the change in expectations between model I and model D was minimal. The impact of extending the independent model to the dependent model produced, on average, a 0.1037% increase in pure premiums, while the minimum percent difference was −1.622% and the maximum was 1.361%. (Note here we are considering the percent difference as the dependent model fit divided by the independent model fit). If we further analyse the extremes of the impact, we have that the exposure with the 1.361% change is a 19 year old single female, with 1 year of experience with the company, 1 year of driving experience, a \$300 deductible, vehicle type G and vehicle age 13. This policyholder in particular had experienced a claim yielding a loss of \$3242.73. At the other extreme, the exposure who experienced a −1.622% change is a 69 year old married male with 14 years experience with the company, 51 years of driving experience, a \$5000 deductible, vehicle type X and vehicle age -1 (a new car). This insured, on the other hand, did not have any claims.

Apart from modelling pure premiums, we can also consider the effect of dependence on the variance of the aggregate claim mount. On average, the variance of the aggregate claims, $\mathbb{V}\text{ar}[S_i|\mathbf{X}_i]$, decreased by -0.285% in the dependent model as compared to the independent model. On the extremes, the variance was increased by as much as 2.514% and decreased by as much as -3.417%. Although the impact on the variance was small, we can nonetheless conclude that overall, extending the model to the dependent case allowed for a more precise estimation of the aggregate claims. It follows that the dependent aggregate claims models provides a more accurate representation of the risk of the insurance portfolio.

It is important to note that for the collision coverage in particular, the association between the frequency and severity components of the aggregate loss amount was found to be negative. It follows that $\mathbb{E}[N_i]$ and $\mathbb{E}[\bar{Y}_i]$ somewhat counteract each other; e.g. if a certain rating variable implies a discount for the expected claim frequency, it might imply a surcharge for the expected claim severity. Thus, the overall impact of dependence on the expected total loss cost, $\mathbb{E}[S_i]$, could be relatively small when the claim frequency and severity have a negative correlation. Nonetheless, Model D allows for a more accurate representation of $\mathbb{E}[S_i]$.

The dependent model framework introduced in this thesis could have a much greater impact for other car insurance coverages or lines of business where there is a positive association between the claim frequency and severity. In this case, in considering the mean claim count and amount separately, the independent model could be double counting the effects of certain rating variables. That is, Model I will give a double discount for good risks and double penalize bad risks by considering the effects of rating variables on $\mathbb{E}[N_i]$ and $\mathbb{E}[\bar{Y}_i]$ separately. The framework of Model D will avoid this double counting effect by accurately reflecting the dependence between N_i and $\bar{Y}_i$ and adjusting the expected loss cost $\mathbb{E}[S_i]$ accordingly.

6.5.1 Residuals

In order to further assess the model fit, we considered the modified Pearson residuals, as described in Section 6.2. The residual plots for both the independent and dependent models showed a similar pattern, with most residuals centered around zero with the exception of a few outliers. Figures 6.2, 6.3, 6.4, and 6.5 provide the residual plots for both model I and model D.

Notice that as we narrow the range of the residuals (see Figures 6.4 and 6.5), there seems to be a slight decreasing trend in the residuals. This could suggest that perhaps a predictor variable is missing in the models. As previously mentioned, the dataset only contains a few rating variables typically used in the property casualty insurance industry. It follows that the models fit here are rather simple compared to what is done in the industry. Potentially significant rating variables not available in the dataset include annual distance driven, territory, driving record, etc.

Figure 6.2: Model I Modified Pearson Residuals

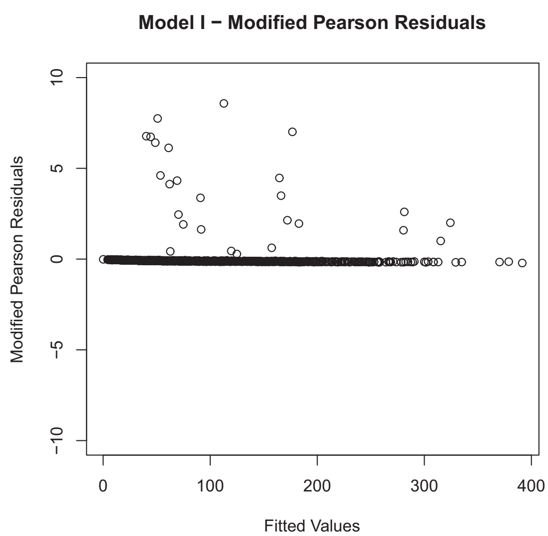


Figure 6.3: Model D Modified Pearson Residuals

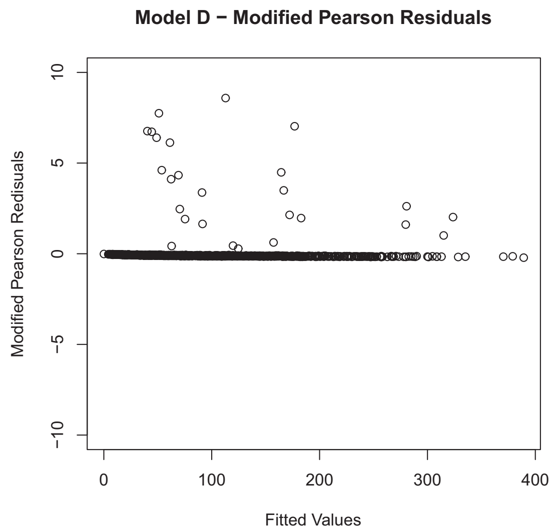


Figure 6.4: Model I Modified Pearson Residuals (zoom)

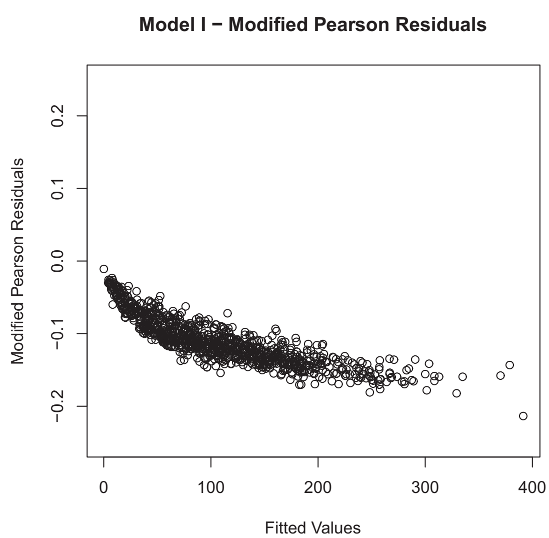
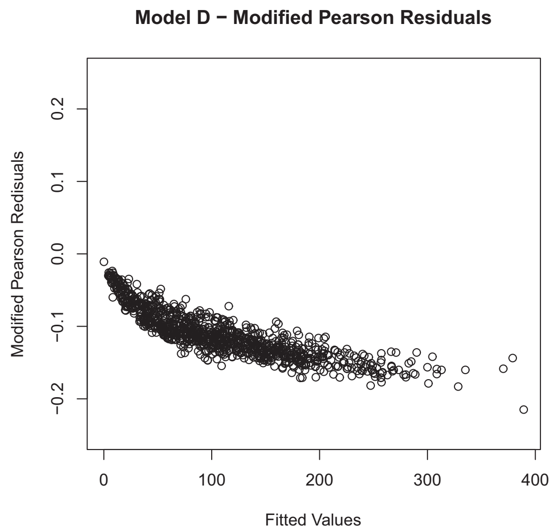


Figure 6.5: Model D Modified Pearson Residuals (zoom)



We also plotted the scaled deviance residuals, as described in Section 6.2, for the frequency model, the severity models and loss cost models; see Figures 6.6, 6.7, 6.8, 6.9, and 6.10. Notice that in the frequency model, there are two clouds of data points; one above 2 and one around 0. Note that the residuals that are significantly above the 0 level correspond to those observations with claim count of 1 or greater. This residual plot provides further evidence that the Poisson distribution assumption for the claim counts is inadequate. Since there are so many observations at $(N_i = 0, \bar{Y}_i = 0)$, the model is unable to properly fit those policies with positive claim occurrences. As previously mentioned, a zero-inflated Poisson model might provide a better fit. This pattern of two bands of residual points is also apparent in the loss cost models deviance residual plots.

Figure 6.6: Frequency Model Deviance Residuals

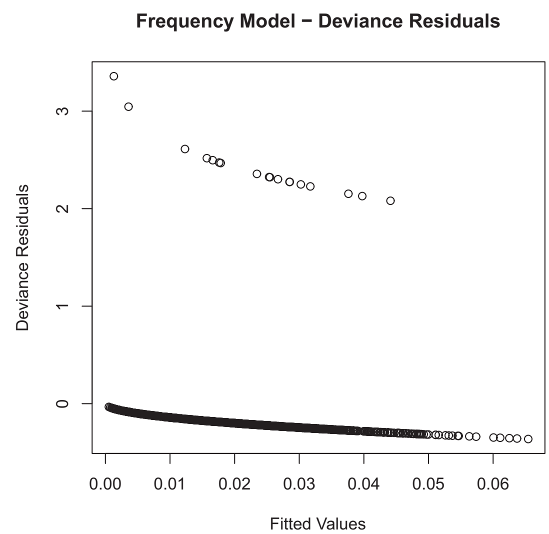


Figure 6.7: Severity Model I Deviance Residuals

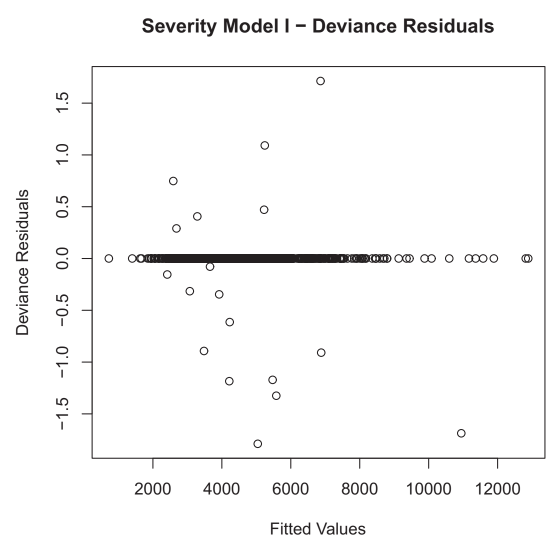


Figure 6.8: Severity Model D Deviance Residuals

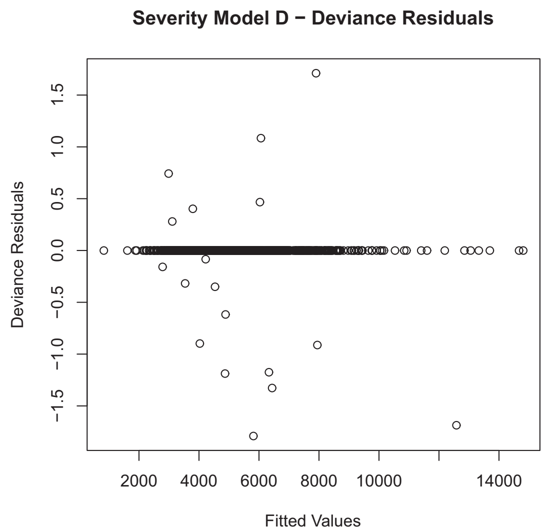


Figure 6.9: Loss Cost Model I Deviance Residuals

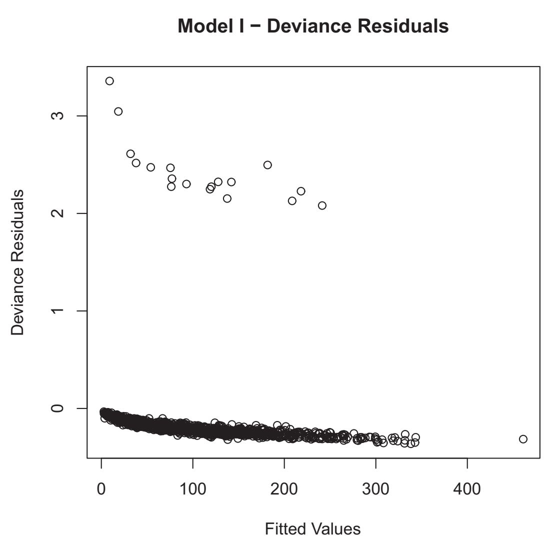
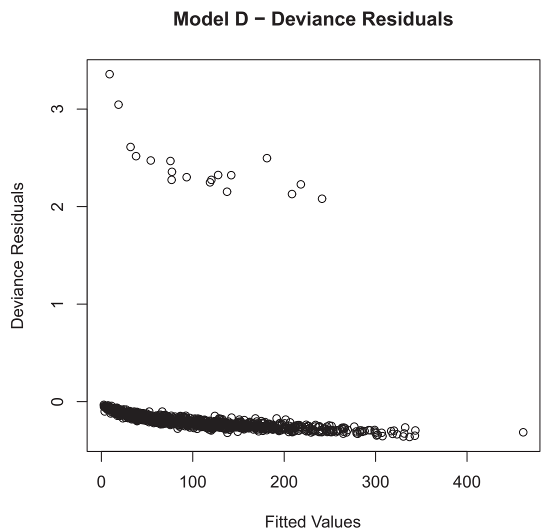


Figure 6.10: Loss Cost Model D Deviance Residuals



6.5.2 Multicollinearity

As mentioned at the beginning of this chapter, there is a strong correlation between the driver's age and the number of years the driver is licensed. In the final severity models in both the independent and dependent case, only the rating variable number of years licensed was present, thus there should not be any issues with multicollinearity. However, in the final frequency model, both age and years licensed are present. Using an analysis of deviance, the adequacy of removing both rating variables was assessed. Comparing the final frequency model with a nested model where only the age variable was removed caused a change in deviance of 118, thus indicating that we cannot remove age from the model. Similarly, removing only number of years licensed from the model caused a change in deviance of 542, thus again indicating that the years licensed variable is significant. We can thus conclude that both age and years licensed are needed in the frequency model.

6.5.3 Hold Out Dataset

Note that the model was also tested on a hold out dataset, which consisted of the collision claim experience from years 2006 through 2008. Similar conclusions were drawn from the out of sample analysis. On average, the impact on the pure premium from moving to the dependent setting was 0.1249 %, while the maximum percent difference was 1.318 % and the minimum was -1.656 %. If we consider the variance of the aggregate claim amount, the dependent model caused on average a -0.1886 % change, while the maximum impact was 2.430 % and the minimum impact was -3.516 %.

Table 6.7 provides the deviances obtained for the frequency, severity and loss cost models for the hold out dataset. Similar to the results obtained on the model dataset, the dependent model deviance is lower than the independent model.

Table 6.7: Deviance Comparison - Hold Out Dataset

Model	**Deviance**	**Scaled Deviance**
Frequency Model	118,625.40	118,625.40
Severity Model I	12,785.63	16,101.92
Severity Model D	12,778.85	16,093.38
Loss Cost Model I	131,411.03	134,727.32
Loss Cost Model D	131,404.25	134,718.78

The residual plots for the hold out dataset were again similar to those from the original dataset, as shown in Figures 6.11 through 6.19.

Figure 6.11: Model I Modified Pearson Residuals Hold Out Dataset

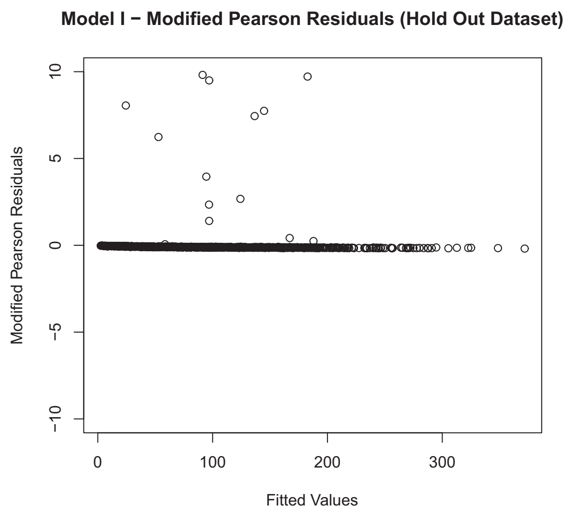


Figure 6.12: Model D Modified Pearson Residuals Hold Out Dataset

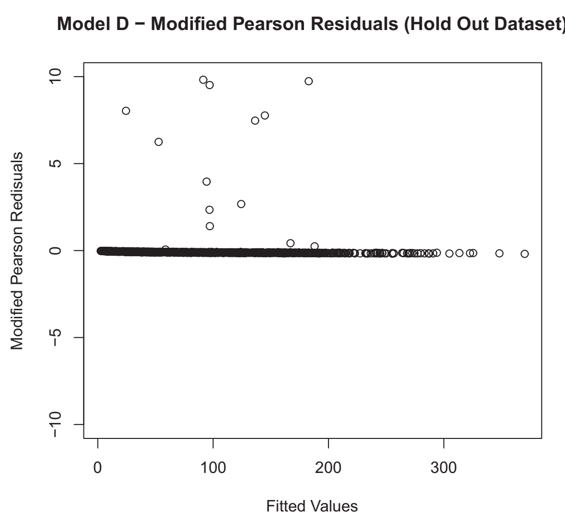


Figure 6.13: Model I Modified Pearson Residuals Hold Out Dataset (zoom)

Figure 6.14: Model D Modified Pearson Residuals Hold Out Dataset (zoom)

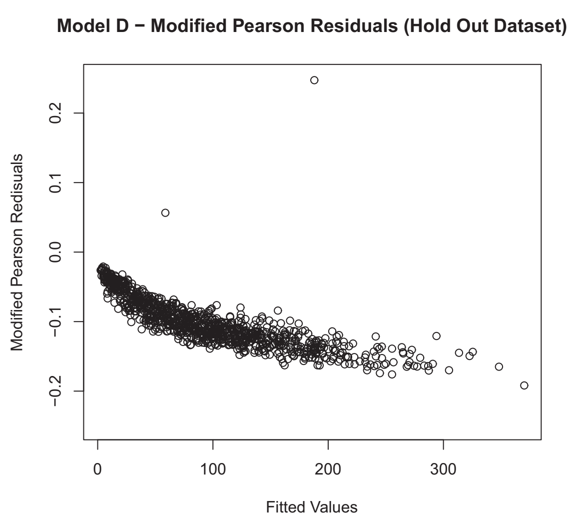


Figure 6.15: Frequency Model Deviance Residuals Hold Out Dataset

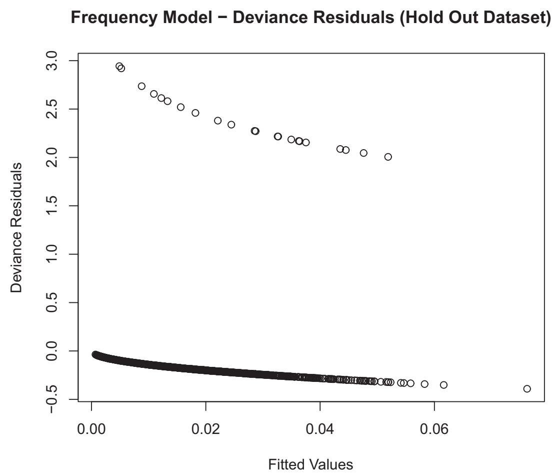


Figure 6.16: Severity Model I Deviance Residuals Hold Out Dataset

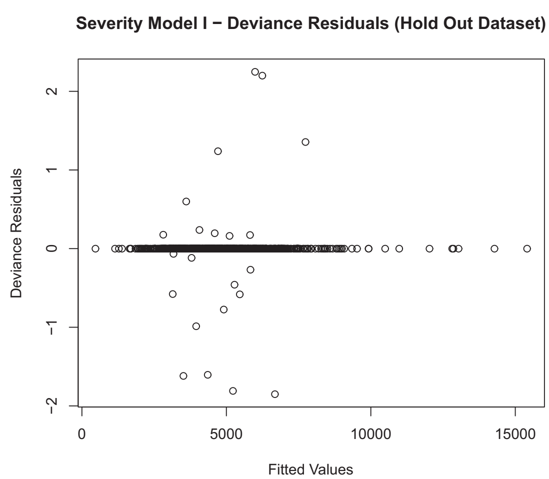


Figure 6.17: Severity Model D Deviance Residuals Hold Out Dataset

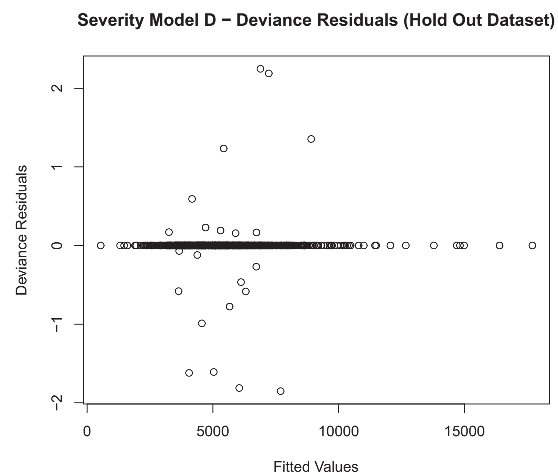


Figure 6.18: Loss Cost Model I Deviance Residuals Hold Out Dataset

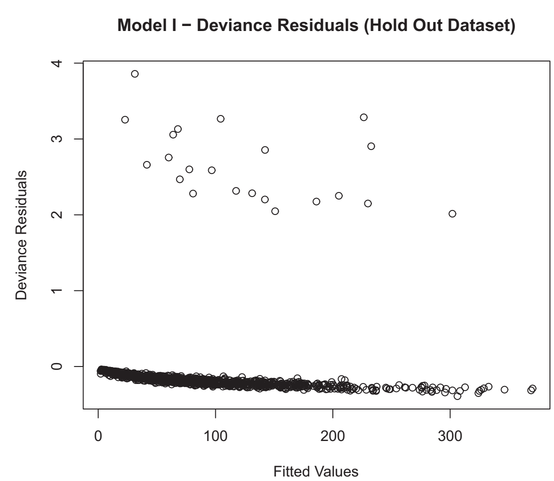
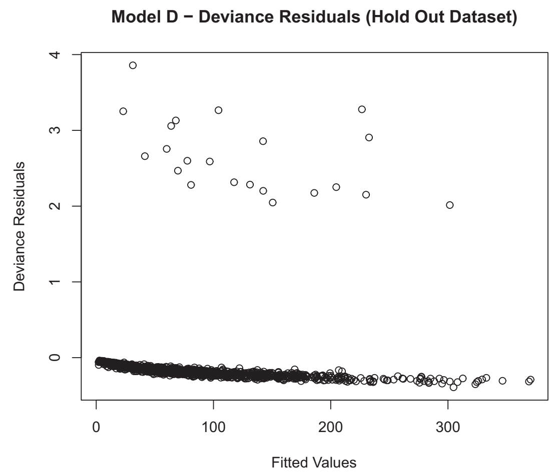


Figure 6.19: Loss Cost Model D Deviance Residuals Hold Out Dataset



6.5.4 Regression Parameter Estimates

Tables 6.8, 6.9, and 6.10 provide the details on the estimated parameters and standard errors obtained for the frequency and severity GLMs in both the independent and dependent cases. We can see that the standard errors for all of the parameter estimates are very low. In particular, in each of the final models, all of the explanatory variables are significant at the 99.9 % level.

In comparing the independent and dependent severity models, we can see that there are significant changes in the parameter estimates, while the standard errors remain low in both cases. Note that although many of the estimates seem close to zero, on the mean scale there will indeed be a considerable impact on the expected value since we are using a log link function. For example, if we consider a continuous covariate X with regression parameter estimates $\hat{\beta}$, then on the mean scale the multiplicative factor associated with that predictor variable becomes $\left(\exp\{\hat{\beta}\}\right)^X$.

Table 6.8: Regression Parameter Estimates - Frequency Model

Regression Parameter	**Estimate**	**Standard Error**
Intercept	-3.08E+00	8.50E-02
Deductible coll	-4.98E-04	4.10E-05
Driver age	1.08E-02	9.76E-04
Driver marital status Divorced	7.29E-02	3.14E-02
Driver marital status Married	-1.67E-01	1.83E-02
Driver marital status Widowed	2.03E-01	5.70E-02
Driver number years licensed	-2.40E-02	9.98E-04
Number years policy here	-1.32E-02	1.70E-03
Vehicle type A	3.12E-01	9.20E-02
Vehicle type B	2.63E-01	7.50E-02
Vehicle type C	2.71E-01	7.98E-02
Vehicle type D	3.01E-01	7.50E-02
Vehicle type E	2.66E-01	9.53E-02
Vehicle type F	3.63E-01	8.23E-02
Vehicle type G	1.66E-01	1.26E-01
Vehicle type H	3.76E-01	1.21E-01
Vehicle type I	3.32E-01	1.81E-01
Vehicle type J	3.67E-01	8.02E-02
Vehicle type K	2.98E-01	1.01E-01
Vehicle type L	2.74E-01	9.43E-02
Vehicle type M	1.67E-02	1.25E-01
Vehicle type N	1.18E-01	7.60E-02
Vehicle type O	8.81E-02	8.16E-02
Vehicle type P	2.28E-01	8.01E-02
Vehicle type Q	4.87E-02	1.20E-01
Vehicle type R	1.68E-01	9.26E-02
Vehicle type S	-9.52E-02	8.77E-02
Vehicle type X	-3.98E-01	3.83E-01
Vehicle age	-4.18E-02	2.09E-03

Table 6.9: Regression Parameter Estimates - Severity Model I

Regression Parameter	**Estimate**	**Standard Error**
Intercept	8.67E+00	7.24E-02
Deductible	2.63E-04	3.67E-05
Driver number years licensed	-2.92E-03	5.04E-04
Vehicle type A	-3.08E-01	8.22E-02
Vehicle type B	-1.02E-01	6.70E-02
Vehicle type C	-8.13E-02	7.13E-02
Vehicle type D	-4.49E-02	6.68E-02
Vehicle type E	8.69E-02	8.51E-02
Vehicle type F	2.68E-01	7.33E-02
Vehicle type G	2.01E-01	1.13E-01
Vehicle type H	3.22E-01	1.08E-01
Vehicle type I	5.71E-01	1.62E-01
Vehicle type J	5.17E-02	7.16E-02
Vehicle type K	1.92E-01	9.04E-02
Vehicle type L	8.45E-02	8.41E-02
Vehicle type M	4.90E-01	1.12E-01
Vehicle type N	-7.44E-02	6.78E-02
Vehicle type O	1.08E-01	7.28E-02
Vehicle type P	1.42E-01	7.14E-02
Vehicle type Q	3.49E-01	1.07E-01
Vehicle type R	1.34E-01	8.28E-02
Vehicle type S	2.40E-01	7.84E-02
Vehicle type X	-3.25E-01	3.44E-01
Vehicle age	-6.45E-02	1.91E-03

Table 6.10: Regression Parameter Estimates - Severity Model D

Regression Parameter	**Estimate**	**Standard Error**
Intercept	8.82E+00	8.05E-02
Claim Count (N_i)	-1.40E-01	3.38E-02
Deductible	2.60E-04	3.66E-05
Driver number years licensed	-3.01E-03	5.03E-04
Vehicle type A	-3.09E-01	8.20E-02
Vehicle type B	-1.01E-01	6.69E-02
Vehicle type C	-7.73E-02	7.11E-02
Vehicle type D	-4.34E-02	6.66E-02
Vehicle type E	8.81E-02	8.49E-02
Vehicle type F	2.71E-01	7.31E-02
Vehicle type G	2.09E-01	1.12E-01
Vehicle type H	3.25E-01	1.08E-01
Vehicle type I	5.70E-01	1.62E-01
Vehicle type J	5.45E-02	7.15E-02
Vehicle type K	1.97E-01	9.02E-02
Vehicle type L	8.43E-02	8.39E-02
Vehicle type M	4.90E-01	1.11E-01
Vehicle type N	-7.38E-02	6.77E-02
Vehicle type O	1.09E-01	7.26E-02
Vehicle type P	1.43E-01	7.12E-02
Vehicle type Q	3.54E-01	1.07E-01
Vehicle type R	1.34E-01	8.26E-02
Vehicle type S	2.39E-01	7.82E-02
Vehicle type X	-3.28E-01	3.43E-01
Vehicle age	-6.45E-02	1.90E-03

Conclusion

In the insurance industry, the common approach for modelling the aggregate claims amount is to assume that the claim frequency and severity components are independent. This independent model allows for a more simplistic representation of the total loss amount and allows to analyse the frequency and severity processes separately. Although this approach perhaps provides greater insight into the two separate processes, it fundamentally ignores the dependence that may exist between the claim frequency and claim severity. This thesis proposes a new model formulation that allows for a correlation between the claim counts and claim amounts. In this particular dependence setting, the independent model is nested inside the dependent model; that is, the independent model is actually a special case of the dependent model.

The focus in this thesis is to provide a model for the expected total loss cost on the individual level, as is done for insurance pricing. We found a closed form formula for both the first and second moments of the aggregate claims while allowing for dependence between the counting process and the jump process of the aggregate claims.

Further work to be done on the subject includes the specification of higher moments as well as the specification of the joint probability density function for the random vector $(N_i, \bar{Y}_i)$.

Bibliography

[1] Agresti, A., (2002) *Categorical Data Analysis*, Wiley, New Jersey.

[2] Agresti, A. (1993) "Distribution-free Fitting of Logit Models with Random Effects for Repeated Categorical Responses", *Statistics in Medicine*, 12, pp. 1969–1988.

[3] Andersen, E.B., (1970) "Asymptotic properties of conditional maximum-likelihood estimators", *Journal of the Royal Statistical Society. Series B (Methodological)*, 32, pp. 283–301.

[4] Anderson, D., Feldblum, S., Modlin, C., Schirmacher, D., Schirmacher, E., and Thandi, N., (2007). "A Practitioner's Guide to Generalized Linear Models", *CAS Study Note*, 3rd ed., pp. 4–39.

[5] Diggle, P.J., Liang, K., and Zeger, S. (1994) *Analysis of Longitudinal Data*, Chapman and Hall, London.

[6] Fahrmeir, L. and Tutz, G. (2001) *Multivariate Statistical Modelling Based on Generalized Linear Models (Second Edition)*, Springer–Verlag, New York.

[7] Gschlößl, S. and Czado, C. (2007) "Spatial modelling of claim frequency and claim size in non–life insurance", *Scandinavian Actuarial Journal*, 3, pp. 202–225.

[8] Iwasaki, M. and Tsubaki, H. (2005) "A new bivariate distribution in natural exponential family", *Metrika*, 61, pp. 323–336.

[9] Jørgensen, B. (1997) *The Theory of Dispersion Models*, Chapman and Hall, London.

[10] Jørgensen, B. (2013) “Construction of multivariate dispersion models”, *Brazilian Journal of Probability and Statistics*, 27, pp. 285–309

[11] Jørgensen, B. and De Souzaa, M. (1994) “Fitting Tweedie's compound Poisson model to insurance claims data”, *Scandinavian Actuarial Journal*, 1, pp. 69–93

[12] Jørgensen, B. and Smyth, G. (2002) “Fitting Tweedie's compound Poisson model to insurance claims data: dispersion modelling”, *ASTIN Bullentin*, 32, pp. 143–157

[13] Klugman, S.A., Panjer, H.H. and G.E. Willmot (2008) *Loss Models; From Data to Decisions*, 3rd ed., Wiley, New Jersey.

[14] McCullagh, P. and Nelder, J.A. (1989) *Generalized Linear Models (Second Edition)*, Chapman and Hall, London.

[15] Neuhaus, J.M., Kalbfleish, J.D., and Hauck, W.W. (1991) “A Comparison of Cluster-Specific and Population-Averaged approaches for Analyzing Correlated Binary Data”, *International Statistical Review*, 59, pp. 25–35.

[16] Sarabia, J.M., Guillén M. (2008) “Joint modelling of the total amount and the number of claims by conditionals”, *Insurance: Mathematics and Economics*, 43, pp. 466–473.

[17] Song, P. (2007) *Correlated Data Analysis: Modeling, Analytics, and Applications*, Springer, New York.

[18] Stephens, D. A. (Winter 2012) *MATH 523 Generalized Linear Models: Course Notes*, McGill University, Montréal.