Property and Casualty Premiums based on Tweedie Families of Generalized Linear Models

Oscar Alberto Quijano Xacur

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
Montreal, Quebec, Canada

August 2011

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: **Oscar Alberto Quijano Xacur**

Entitled: **Property and Casualty Premiums based on Tweedie Families of Generalized Linear Models**

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. P. Gaillardetz

______________________________ Examiner
Dr. Y. P. Chaubey

______________________________ Thesis Supervisor
Dr. J. Garrido

Approved by ______________________________

Chair of Department or Graduate Program Director

____

Dean of Faculty

Date ______________________________

Abstract

Property and Casualty Premiums based on Tweedie Families of Generalized Linear Models

Oscar Alberto Quijano Xacur

We consider the problem of estimating accurately the pure premium of a property and casualty insurance portfolio when the individual aggregate losses are assumed to follow a compound Poisson distribution with gamma jump sizes. Generalized Linear Models (GLMs) with a Tweedie response distribution are analyzed as a method for this estimation. This approach is compared against the standard practice in the industry of combining estimations obtained separately for the frequency and severity by using GLMs with Poisson and gamma responses, respectively. We show that one important difference between these two methods is the variation of the scale parameter of the compound Poisson-gamma distribution when it is parametrized as an exponential dispersion model. We conclude that both approaches need to be considered during the process of model selection for the pure premium.

Acknowledgments

I would like to thank my supervisor, Dr. Jose Garrido. His knowledge, guidance, encouragement, patience and help have been invaluable for the preparation of this thesis and my experience as a Masters student.

I am grateful to professor Arusharka Sen who answered many of the doubts I had about estimation and regression methods. I am also grateful to Ferenc Balogh, the discussions I had with him at the early stage of this work were very useful for my understanding of some topics in the first chapter.

Thanks to Dr. Yogendra Chaubey and Dr. Patrice Gaillardetz for reviewing this work.

Thanks to my parents for their unconditional support and confidence in me.

I would also like to thank my friends, fortunately for me they are too many to mention individually, for their support, help, beers and encouragement during the hard, and not so hard, moments of this thesis; But I specially thank them for their friendship.

This work would not have been possible without the financial support of the Government of Canada awards program, the Department of Mathematics and Statistics of Concordia University, MITACS and AVIVA Canada.

To my parents: Oscar and Malka.

Contents

List of Figures

List of Tables

Introduction

Property and Casualty insurance is a cyclical business. Accurate estimation of future losses is of crucial importance in managing insurance risks and ensuring a solvent and profitable operation. In recent years, the industry has adopted Generalized Linear Models (GLMs) to improve the fit and prediction accuracy of models in insurance portfolios.

A common distribution used for the modeling of the total loss of each policyholder is the Compound Poisson distribution with gamma jump sizes (CPG). Using this distribution is equivalent to assuming that the claim frequency and the claim severity are independent and that they follow Poisson and gamma distributions, respectively. When the CPG distribution is assumed and GLMs are used for the estimation of its parameters, the standard practice is to fit separate GLMs for the claim frequency and claim severity. The Tweedie family of distributions allows to parametrize the CPG as an exponential dispersion model. This reduces the model to a single GLM for the estimation of the CPG parameters. The purpose of this thesis is to present the GLMs with Tweedie response as an option for modeling the individual total loss of each policyholder and to compare it against the separate Poisson-gamma estimation.

In the first chapter an introduction to Exponential Dispersion Families of Distributions is given. A formal presentation is made through a generating measure. The additive and reproductive versions of these families are formulated as well as their characterization through the variance function. The chapter concludes by deriving the reparametrization of these families in terms of their mean and scale parameter with the introduction of the unit deviance.

In Chapter 2, a brief introduction to GLMs is given. The distributional assumptions and maximum likelihood estimation of the coefficients are also given.

In Chapter 3 the Tweedie families of distributions are defined. Even though a general definition is given, the chapter focuses on showing that the CPG distributions are part of these families. References are given for other distributions.

Chapter 4 develops the general methodology used in insurance modeling. The classical assumptions made for the distributions of the aggregated losses are presented along with a critical analysis of these assumptions. The chapter continues by introducing the CPG as a possible distribution for the total claim size. A discussion follow on the differences in the estimation of its parameters by using a single Tweedie GLM or separated ones for the frequency and severity. The chapter concludes with two examples of applications.

Chapter 1

The Exponential Dispersion Families

Exponential dispersion families (EDFs) are sets of probability distributions. Many widely used densities can be parametrized as an EDF. EDFs are also the basis for Generalized Linear Models. This chapter presents a general construction of the univariate Additive Exponential Dispersion Families (AEDFs) and then the Reproductive Exponential Dispersion Families (REDFs) are derived from them. The goal was to show the EDFs without going first through the Natural Exponential Families. For this purpose the order of ideas in which our construction is made is based on Jørgensen (1986), but the names given to the different EDFs are taken from Jørgensen (1997).

1.1 Construction of the Additive Exponential Dispersion Families.

In this section it is shown how an AEDF can be generated from a probability measure that satisfies certain conditions.

Let M and κ be the moment generating function (mgf) and cumulant generating function (cmf) of some probability measure Q on $\mathbb{R}$, respectively, and let

$$\Theta_0 = \left\{ \theta \in \mathbb{R} : M(\theta) = \int \exp(\theta x) dQ(x) < \infty \right\}.$$

Let Θ be the interior of Θ_0 and assume that $\Theta \neq \emptyset$. In general Θ_0 is an interval that includes zero (including $\{0\}$), so Θ will always be an open interval. Throughout this chapter Q will be called the generating measure. Define the set Λ, which we will call from now on the index set of Q, as

$$\Lambda = \left\{ \lambda \in \mathbb{R} \backslash \{0\} : M_\lambda = M^\lambda \text{ is the mgf of some probability measure } Q_\lambda \right\}$$

and let k_λ be the cumulant distribution function for each Q_λ. Now, let $\lambda \in \Lambda$ be fixed and define for each $\theta \in \Theta$ the map

$$Q_{\lambda,\theta}(A) = \int_A \exp\left(\theta x - \lambda \kappa(\theta)\right) dQ_\lambda(x), \qquad A \in \mathcal{B}(\mathbb{R}), \tag{1.1}$$

where $\mathcal{B}(\mathbb{R})$ is the Borel σ-algebra in $\mathbb{R}$.

Proposition 1.1. *For every* $(\lambda, \theta) \in \Lambda \times \Theta$*, the map* $Q_{\lambda,\theta}$ *defined above is a probability measure on the measurable space* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Proof. Let $(\lambda, \theta) \in \Lambda \times \Theta$. As $\exp\left(\theta x - \lambda \kappa(\theta)\right) > 0$ for every x then $Q_{\lambda,\theta}$ is a measure on $\mathbb{R}$, thus in order to prove that it is a probability measure it is only needed to show that $Q_{\lambda,\theta}(\mathbb{R}) = 1$. Now, from the definitions of Q_λ and $Q_{\lambda,\theta}$

$$\begin{aligned} Q_{\lambda,\theta}(\mathbb{R}) &= \int \exp[\theta x - \lambda \kappa(\theta)] dQ_\lambda(x) \\ &= \frac{\displaystyle\int \exp(\theta x) dQ_\lambda(x)}{\exp[\kappa_\lambda(\theta)]} \\ &= \frac{M_\lambda(\theta)}{\exp[\ln(M_\lambda(\theta))]} \\ &= \frac{M_\lambda(\theta)}{M_\lambda(\theta)} = 1 \end{aligned}$$

□

Now it is possible to give a proper definition of an exponential dispersion family.

Definition 1.1. *With* $Q, \Theta, \Lambda, \kappa, Q_\lambda$ *and* $Q_{\lambda,\theta}$ *defined as above, let* Λ_0 *be any subset of* Λ*. The additive exponential dispersion family generated by* Q *and* Λ_0 *is defined as*

$$ED(Q, \Lambda_0) := \{Q_{\lambda,\theta} : (\lambda, \theta) \in \Lambda_0 \times \Theta\},$$

where Θ*,* κ *and* Λ_0 *are called the canonical space, the cumulant generator and the index set of the family, respectively. When working with a fixed member* $Q_{\lambda,\theta}$*,* θ *and* λ *are called the canonical and scale parameters, respectively.*

Remark 1.1. *Since* Θ *is defined as the interior of* Θ_0*, the canonical space in this definition is always open. In Jørgensen (1997) the set* Θ_0 *is the canonical space and the family is called regular when it is open. Thus, in the sense of Jørgensen (1997) all the exponential families treated here are regular.*

It is important to clarify that we are making a distinction between the index set of the generating measure and the index set of the family. This is not made in Jørgensen (1997), where the index set of the family is always the one of the generating measure. The motivation for this distinction is that sometimes it is possible to prove that a certain set is contained in Λ and then it is possible to generate an EDF with this set without the need of checking if there are more elements in Λ or not. In the following example it is proved that $\mathbb{N} \subset \Lambda$ (we use $\mathbb{N}$ to denote the set of natural numbers), and hence with $\mathbb{N}$ it is possible to generate the binomial family of distributions.

Example 1.1. *(The binomial distribution). Using the construction described above and a reparametrization shows that the binomial distribution can be seen as an AEDF. Let* Q *be the probability measure that gives probability* 0.5 *to* $\{1\}$ *and* $\{0\}$*, i.e.*

$$Q(A) = \frac{I_A(0) + I_A(1)}{2}, \qquad A \in \mathcal{B}(\mathbb{R}), \tag{1.2}$$

where I is an indicator function. Then the moment generating function of Q is given by

$$M(\theta) = \frac{1+\exp(\theta)}{2}, \qquad \theta \in \mathbb{R},$$

and thus $\Theta = \mathbb{R}$. Now, for any $n \in \mathbb{N}$,

$$M_n(\theta) = M(\theta)^n = \left(\frac{1+\exp(\theta)}{2}\right)^n, \qquad \theta \in \mathbb{R},$$

is the mgf of a Binomial$(n, \frac{1}{2})$, and therefore the corresponding measure is

$$Q_n(A) = \sum_{j=0}^{n} \binom{n}{j} \left(\frac{1}{2}\right)^n I_A(j), \qquad A \in \mathcal{B}(\mathbb{R}).$$

Then, for each $\theta \in \mathbb{R}$ and $A \in \mathcal{B}(\mathbb{R})$, $Q_{n,\theta}$ is defined as

$$\begin{aligned}
Q_{n,\theta}(A) &= \int_A \exp[\theta x - \kappa_n(\theta)] dQ_n(x) \\
&= \int_A \exp\left[\theta x - n \ln\left(\frac{1}{2} + \frac{1}{2}\exp(\theta)\right)\right] dQ_n(x) \\
&= \frac{\int_A \exp(\theta x) dQ_n(x)}{\left[\frac{1}{2} + \frac{1}{2}\exp(\theta)\right]^n} \\
&= \frac{\sum_{j=0}^{n} \binom{n}{j} \left(\frac{1}{2}\right)^n \exp(\theta j) I_A(j)}{\left(\frac{1}{2}\right)^n [1+\exp(\theta)]^n} = \frac{\sum_{j=0}^{n} \binom{n}{j} \exp(\theta j) I_A(j)}{(1+\exp(\theta))^n}.
\end{aligned}$$

Now, if we define $p := \frac{\exp(\theta)}{1+\exp(\theta)}$, then depending on the value of θ, p can take any value in $(0,1)$. Writing θ in terms of p, gives $\theta = \ln(\frac{p}{1-p})$. Applying this reparametrization *to $Q_{\lambda,\theta}$, we get*

$$Q_{\lambda,p}(A) = \frac{\sum_{j=0}^{n} \binom{n}{j} \exp\left[j \ln\left(\frac{p}{1-p}\right)\right] I_A(j)}{\left(1 + \exp\left[\ln\left(\frac{p}{1-p}\right)\right]\right)^n}$$

$$= \frac{\sum_{j=0}^{n} \binom{n}{j} \left(\frac{p}{1-p}\right)^j I_A(j)}{\left(1 + \frac{p}{1-p}\right)^n}$$

$$= \sum_{j=0}^{n} \binom{n}{j} p^j (1-p)^{n-j} I_A(j),$$

which is the measure corresponding to a Binomial(n, p). Therefore, the $ED(Q, \mathbb{N})$ with Q defined as in (1.2), corresponds to the binomial family of distributions.

In this construction of the $AEDF$ we only assumed that $\Theta \neq \emptyset$. A special important case is when $0 \in \Theta$, then Q has all finite moments, and therefore so does Q_λ. Also, assuming this we assure that κ_λ is infinitely differentiable. As these properties will be very useful later in this chapter, the following definition is made.

Definition 1.2. *A $ED(Q, \Lambda)$ with canonical space Θ is called appropriate if $0 \in \Theta$.*

1.2 General Properties

In addition to $\Theta, \Lambda, \kappa, \kappa_\lambda, Q_\lambda$ and $Q_{\lambda,\theta}$ defined in the previous section, for each $(\lambda, \theta) \in \Lambda \times \Theta$, $M_{\lambda,\theta}$ will denote the mgf of $Q_{\lambda,\theta}$.

After seeing the construction of the AEDFs, a natural question arises: Take a specific measure from the generated family, and use it as a generating measure. Does it generate new distributions or does it give back the same family? In order to answer this question the mgf of the members of the family will be needed.

Theorem 1.1. *For every* $(\lambda, \theta) \in \Lambda \times \Theta$

$$M_{\lambda,\theta}(t) = \left(\frac{M(\theta + t)}{M(\theta)}\right)^{\lambda} < \infty, \qquad \textit{for } t \in \Theta - \theta, \tag{1.3}$$

where $\Theta - \theta := \{\ x - \theta : x \in \Theta\}$.

Proof. Let $(\lambda, \theta) \in \Lambda \times \Theta$ be fixed. As for any $t \in \Theta - \theta$, $\theta + t \in \Theta$ then $M(\theta + t) < \infty$ and therefore

$$\begin{aligned} M_{\lambda,\theta}(t) &= \int \exp(xt) \exp[\theta x - \kappa_{\lambda}(\theta)] dQ_{\lambda}(x) \\ &= \frac{\int \exp[(\theta + t)x] dQ_{\lambda}(x)}{\exp(\kappa_{\lambda}(\theta))} \\ &= \left(\frac{M\,(\theta + t)}{M\,(\theta)}\right)^{\lambda} < \infty. \end{aligned}$$

□

Corollary 1.1. *Let* $ED(Q, \Lambda_0)$ *be appropriate with* $1 \in \Lambda_0$*, then* $Q \in ED(Q, \Lambda_0)$.

Proof. Taking $\theta = 0$ and $\lambda = 1$ in (1.1) the mgf of Q is obtained, thus Q is in the family. □

It is important to note that for every $\theta \in \Theta$, $\Theta - \theta$ is an open set that contains zero; Thus even if $0 \notin \Theta$, every member of the family has a finite mgf in an open set containing zero. Hence, if we show that by using a member of the $ED(Q, \Lambda)$ as a generating measure we obtain the same set of mgfs, then we can conclude that we are generating the same family. In the next theorem this is proved for appropriate exponential families. A lemma is shown first.

Lemma 1.1. *Let* $Q_{\lambda,\theta}$ *be a member of* $ED(Q, \Lambda)$*, where* Λ *is the index set of* Q*. If* Λ^* *is the index set of* $Q_{\lambda,\theta}$ *then* $\Lambda/\lambda \subset \Lambda^*$*, where* $\Lambda/\lambda = \{x/\lambda : x \in \Lambda\}$*. If in addition we assume that* $ED(Q, \Lambda)$ *is appropriate then* $\Lambda/\lambda = \Lambda^*$.

Proof. Let $\lambda_0 \in \Lambda/\lambda$ then there exists $\lambda_1 \in \Lambda$ such that $\lambda_0 = \frac{\lambda_1}{\lambda}$, and then

$$M_{\lambda,\theta}^{\lambda_0}(t) = \left[\left(\frac{M(\theta+t)}{M(\theta)}\right)^{\lambda}\right]^{\lambda_0} = \left[\left(\frac{M(\theta+t)}{M(\theta)}\right)^{\lambda}\right]^{\frac{\lambda_1}{\lambda}} = \left(\frac{M(\theta+t)}{M(\theta)}\right)^{\lambda_1}, \qquad t \in \Theta - \theta,$$

which is an mgf corresponding to $Q_{\lambda_1,\theta}$. Thus $\lambda_0 \in \Lambda^*$, and therefore $\Lambda/\lambda \subset \Lambda^*$. Assume now that $ED(Q,\Lambda)$ is appropriate. From Theorem 1.1 the canonical set of an AEDF, using $Q_{\lambda,\theta}$ as generating measure, is $\Theta - \theta$. Let $\lambda_0 \in \Lambda^*$, then for any $\theta_1 \in \Theta - \theta$,

$$M_{\lambda_0,\theta_1}^*(t) = \left(\frac{M_{\lambda,\theta}(\theta_1+t)}{M_{\lambda,\theta}(\theta_1)}\right)^{\lambda_0} = \left(\frac{M(\theta+\theta_1+t)}{M(\theta+\theta_1)}\right)^{\lambda_0\lambda}, \qquad t \in \Theta - \theta - \theta_1 \tag{1.4}$$

is an mgf. As $ED(Q,\Lambda)$ is appropriate then $-\theta \in \Theta - \theta$, therefore (1.4) is also an mgf for $\theta_1 = -\theta$, *i.e.*

$$M_{\lambda_0,-\theta}^*(t) = \left(\frac{M(t)}{M(0)}\right)^{\lambda_0\lambda} = M(t)^{\lambda_0\lambda}, \qquad t \in \Theta$$

is an mgf; This implies that $\lambda_0\lambda \in \Lambda$ and therefore that $\lambda_0 \in \Lambda/\lambda$, thus we have $\Lambda^* \subset \Lambda/\lambda$.

□

Theorem 1.2. *Let $Q_{\lambda,\theta}$ be any member of the appropriate $ED(Q,\Lambda_0)$, then for any $(\lambda,\theta) \in \Lambda_0 \times \Theta$, $ED(Q,\Lambda_0) = ED(Q_{\lambda,\theta},\Lambda_0/\lambda)$.*

Proof. Let $(\lambda,\theta) \in \Lambda_0 \times \Theta$ be fixed. From Lemma 1.1 we know that Λ_0/λ is a subset of the index set of $Q_{\lambda,\theta}$, thus $ED(Q_{\lambda,\theta},\Lambda_0/\lambda)$ is well defined. Denote with Q_{λ^*,θ^*}^* the members of $ED(Q_{\lambda,\theta},\Lambda_0/\lambda)$ and let M_{λ^*,θ^*}^* be their corresponding mgfs.
Let $(\lambda_1,\theta_1) \in \Lambda \times \Theta$, then the mgf of Q_{λ_1,θ_1} is given by

$$M_{\lambda_1,\theta_1}(t) = \left(\frac{M(\theta_1+t)}{M(\theta_1)}\right)^{\lambda_1}, \qquad t \in \Theta - \theta_1. \tag{1.5}$$

On the other hand, for any $(\lambda^*, \theta^*) \in (\Theta - \theta) \times \Lambda_0/\lambda$, the mgf of $Q^*_{\lambda^*,\theta^*}$ is given by

$$M^*_{\lambda^*,\theta^*}(t) = \left(\frac{M_{\lambda,\theta}(\theta^* + t)}{M_{\lambda,\theta}(\theta^*)}\right)^{\lambda^*} = \left(\frac{M(\theta + \theta^* + t)}{M(\theta + \theta^*)}\right)^{\lambda\lambda^*}, \qquad t \in \Theta - (\theta + \theta^*).$$

Then, taking $\lambda^* = \frac{\lambda_1}{\lambda}$ and $\theta^* = \theta_1 - \theta$ we get that for every $t \in \Theta - \theta_1$,

$$M^*_{\frac{\lambda_1}{\lambda},\theta_1-\theta}(t) = \left(\frac{M(\theta + \theta_1 - \theta + t)}{M(\theta + \theta_1 - \theta)}\right)^{\lambda\left(\frac{\lambda_1}{\lambda}\right)} = \left(\frac{M(\theta_1 + t)}{M(\theta_1)}\right)^{\lambda_1} = M_{\lambda_1,\theta_1}(t),$$

which implies that $Q_{\lambda_1,\theta_1} \in ED(Q_{\lambda,\theta}, \Lambda_0/\lambda)$, and therefore one family is included in the other: $ED(Q, \Lambda_0) \subset ED(Q_{\lambda,\theta}, \Lambda_0/\lambda)$. To prove the converse, let $(\lambda^*, \theta^*) \in \Lambda_0/\lambda \times \Theta - \theta$. Then there exist $\theta_0 \in \Theta$ and $\lambda_0 \in \Lambda$ such that $\theta^* = \theta_0 - \theta$ and $\lambda\lambda^* = \lambda_0$. Then for every $t \in \Theta - \theta_0$

$$M^*_{\lambda^*,\theta^*}(t) = \left(\frac{M_{\lambda,\theta}(\theta^* + t)}{M_{\lambda,\theta}(\theta^*)}\right)^{\lambda^*} = \left(\frac{M(\theta + \theta^* + t)}{M(\theta + \theta^*)}\right)^{\lambda\lambda^*} = \left(\frac{M(\theta_0 + t)}{M(\theta_0)}\right)^{\lambda_0} = M_{\lambda_0,\theta_0}(t),$$

which implies that $Q^*_{\lambda^*,\theta^*} \in ED(Q, \Lambda_0)$, and therefore $ED(Q_{\lambda,\theta}, \Lambda_0/\lambda) \subset ED(Q, \Lambda_0)$.

□

Theorem 1.3. *Let $ED(Q, \Lambda)$ be a appropriate AEDF, $(\lambda, \theta) \in \Lambda \times \Theta$ and let X be a random variable with probability law $Q_{\lambda,\theta}$; Then $\mathbb{E}[X] = \kappa'_\lambda(\theta) = \lambda\kappa'(\theta)$ and $\mathbb{V}[X] = \kappa''_\lambda(\theta) = \lambda\kappa''(\theta)$.*

Proof. As $ED(Q, \Lambda)$ is appropriate then κ is infinitely differentiable. Taking derivatives from equation (1.3) the following expressions are obtained

$$M'_{\lambda,\theta}(t) = \lambda \left(\frac{M(\theta + t)}{M(\theta)}\right)^{\lambda} \kappa'(\theta + t),$$
$$M''_{\lambda,\theta}(t) = \lambda \left[\left(\frac{M(\theta + t)}{M(\theta)}\right)^{\lambda} \kappa''(\theta + t) + \lambda \left(\frac{M(\theta + t)}{M(\theta)}\right)^{\lambda-1} \frac{M'(\theta + t)}{M(\theta)} \kappa'(\theta + t)\right],$$

and, in turn,

$$\mathbb{E}[X] = M'_{\lambda,\theta}(0) = \lambda\kappa'(\theta)$$
$$\mathbb{E}[X^2] = M''_{\lambda,\theta}(0) = \lambda\kappa''(\theta) + \lambda^2\kappa'(\theta)^2$$

implying that

$$\mathbb{V}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \lambda\kappa''(\theta)$$

□

Theorem 1.4. *Let μ and ν be two finite measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that*

$$\nu(A) = \int_A f d\mu,$$

where f is a positive function. Then ν is concentrated at one point if and only if μ is concentrated at one point.

Proof. Suppose that μ is concentrated at a, then for any $A \in \mathcal{B}(\mathbb{R})$,

$$\nu(A) = \begin{cases} f(a)\mu(\{a\}) & \text{if } a \in A \\ 0 & \text{if } a \notin A \end{cases}$$

therefore ν is also concentrated at a. On the other hand, assume that ν is concentrated at a, then

$$0 = \nu(\{a\}^c) = \int_{\{a\}^c} f d\mu = \int f I_{\{a\}^c} d\mu,$$

as $fI_{\{a\}^c} \geq 0$. This implies that $fI_{\{a\}^c} = 0$ μ-a.e., but since $f > 0$ this means that $I_{\{a\}^c} = 0$ μ-a.e., which implies that $\mu(\{a\}^c) = 0$, therefore μ is concentrated at a. □

Theorem 1.5. *Let $ED(Q, \Lambda_0)$ be a appropriate AEDF where Q is not concentrated at one point. Then for every $\theta \in \Theta$, $\kappa''(\theta) > 0$.*

Proof. Without loss of generality we can assume that $1 \in \Lambda_0$ (because if it were not we could just take the family $ED(Q, \Lambda_0 \bigcup \{1\})$). Let $\theta \in \Theta$ and X be a random variable with law $Q_{1,\theta}$, then from the definition of $Q_{1,\theta}$ and Theorem 1.4 we know that $Q_{1,\theta}$ is not concentrated in one point. Thus $\mathbb{V}[X] > 0$ and then from Theorem 1.3

$$\kappa''(\theta) = \mathbb{V}[X] > 0.$$

□

Theorem 1.6. *Let Q be a probability measure that is not concentrated in one point. Assume that zero is in its canonical set and let Λ be its index set, then*

(i) If $\lambda_1, \lambda_2 \in \Lambda$ then $\lambda_1 + \lambda_2 \in \Lambda$.

(ii) $\mathbb{N} \subset \Lambda$.

(iii) $\Lambda \subset (0, \infty)$.

Proof. For part *(i)* let λ_1 and λ_2 be in Λ, then M^{λ_1} and M^{λ_2} are the mgfs of Q_{λ_1} and Q_{λ_2} respectively. Thus $M^{\lambda_1+\lambda_2}$ is the mgf of $Q_{\lambda_1} * Q_{\lambda_2}$(the convolution of Q_{λ_1} and Q_{λ_2}) and therefore, from the definition of Λ, $\lambda_1 + \lambda_2 \in \Lambda$. Part *(ii)* follows from *(i)* and the fact that $1 \in \Lambda$. For part *(iii)*, let $(\lambda, \theta) \in \Lambda \times \Theta$ and X be a random variable with law $Q_{\lambda,\theta}$, then from Theorem 1.3

$$\mathbb{V}[X] = \lambda\kappa''(\theta).$$

Now, as Q is not concentrated in one point then $\mathbb{V}[X] > 0$ and $\kappa''(\theta) > 0$, which implies that

$$\lambda = \frac{\mathbb{V}[X]}{\kappa''(\theta)} > 0.$$

As λ was chosen arbitrarily, this shows that $\Lambda \subset (0, \infty)$. □

The theorems from this section show many properties of the appropriate exponential families. But, is it always possible to work with appropriate $AEDFs$ or there are cases where it is necessary to work with a non appropriate family? The following theorem helps answer this question.

Theorem 1.7. *Let Q be a probability measure and let Θ and Λ be its canonical space and index set respectively. Assume also that $0 \notin \Theta$. Let $\theta \in \Theta$ be fixed and let Λ^* be the index set of $Q_{1,\theta}$, then $ED(Q,\Lambda) \subset ED(Q_{1,\theta},\Lambda^*)$.*

Proof. From Lemma 1.1 we have that $\Lambda \subset \Lambda^*$. Denote with $Q^*_{\lambda^*,\theta^*}$ the elements of $ED(Q_{1,\theta},\Lambda^*)$ and let $M^*_{\lambda^*,\theta^*}$ represent the respective mgfs. Then for any $(\lambda^*,\theta^*) \in \Lambda^* \times (\Theta - \theta)$

$$M^*_{\lambda^*,\theta^*}(t) = \left(\frac{M_{1,\theta}\left(\theta^* + t\right)}{M_{1,\theta}\left(\theta^*\right)}\right)^{\lambda^*} = \left(\frac{M(\theta + \theta^* + t)}{M(\theta + \theta^*)}\right)^{\lambda^*}, \qquad t \in (\Theta - \theta - \theta^*).$$

Now, let Q_{λ^*,θ^*} be any element of $ED(\Theta,\Lambda)$. Taking $\lambda^* = \lambda_1$ and $\theta^* = \theta_1 - \theta$ implies that for $t \in \Theta - \theta$

$$M^*_{\lambda_1,\theta_1-\theta}(t) = \left(\frac{M(\theta_1 + t)}{M(\theta_1)}\right)^{\lambda_1} = M_{\lambda_1,\theta_1}(t),$$

which implies that $Q_{\lambda_1,\theta_1} \in ED(Q_{1,\theta},\Lambda^*)$, therefore $ED(Q,\Lambda) \subset ED(Q_{1,\theta},\Lambda^*)$. □

Corollary 1.2. *For every non appropriate $AEDF$, there exists an appropriate one that contains it.*

Proof. It follows from the previous theorem and the fact that every member of an $AEDF$ (appropriate or not) generates an appropriate $AEDF$. □

Thus, the appropriate $AEDF$s are richer than the non appropriate ones. So from a practical point of view it is always possible to assume that we are working with a appropriate $AEDF$. In the rest of this chapter every $AEDF$ will be assumed to be appropriate.

1.3 The Reproductive Exponential Dispersion Family

The Reproductive Exponential Dispersion Families (REDFs) are formed by a reparametrization of the AEDFs. The development of Generalized Linear Models is based on this version of the EDFs. Its definition and some of its main properties are presented in this section.

Definition 1.3. *A random variable X is said to belong to the $ED(Q, \Lambda_0)$ if its probability law is in $ED(Q, \Lambda_0)$.*

Definition 1.4. *Let $(\lambda, \theta) \in \Lambda \times \Theta$, then the measures P_λ and $P_{\lambda,\theta}$ are defined as*

$$\begin{aligned} P_\lambda(A) &= Q_\lambda(\lambda A) \\ P_{\lambda,\theta}(A) &= \int_A \exp\left(\lambda\left\{\theta x - \kappa(\theta)\right\}\right) dP_\lambda(x), \end{aligned}$$

for every $A \in \mathcal{B}(\mathbb{R})$, where $\lambda A = \{\lambda x : x \in A\}$.

Theorem 1.8. *Let X be a random variable in $ED(Q, \Lambda_0)$ with probability law $Q_{\lambda,\theta}$ and let $Y = \frac{X}{\lambda}$, then the probability law of Y is $P_{\lambda,\theta}$.*

Proof. Let P_X and P_Y be the probability distributions of X and Y, respectively, $h(x) = \frac{x}{\lambda}$ and $A \in \mathcal{B}(\mathbb{R})$, then

$$\begin{aligned} P_Y(A) = P_X(h^{-1}(A)) &= P_X(\lambda A) \\ &= Q_{\lambda,\theta}(\lambda A) \\ &= \int_{\lambda A} \exp[\theta x - \kappa_\lambda(\theta)] dQ_\lambda(x) \\ &= \int \exp[\theta x - \lambda\kappa(\theta)] I_{\lambda A}(x) dQ_\lambda(x) \\ &= \int \exp[\theta x - \lambda\kappa(\theta)] I_A(x/\lambda) dQ_\lambda(x), \end{aligned}$$

then, by the change of variable formula, applying the transformation $y = \frac{x}{\lambda}$ we get that

$$\begin{aligned} P_Y(A) &= \int \exp[\lambda\theta y - \lambda\kappa(\theta)] I_A(x) dP_\lambda(x) \\ &= \int \exp[\lambda\{\theta y - \kappa(\theta)\}] I_A(x) dP_\lambda(x) \\ &= P_{\lambda,\theta}(A). \end{aligned}$$

□

Definition 1.5. *The REDF generated by Q and $\Lambda_0 \subset \Lambda$ is defined as*

$$ED^*(Q, \Lambda_0) = \{P_{\lambda,\theta} : (\lambda, \theta) \in \Lambda_0 \times \Theta\},$$

for each $(\lambda, \theta) \in \Lambda \times \Theta$ and $\overline{M}_{\lambda,\theta}$ denotes the mgf of $P_{\lambda,\theta}$.

Theorem 1.9. *For every $(\lambda, \theta) \in \Lambda \times \Theta$,*

$$\overline{M}_{\lambda,\theta}(t) = \left(\frac{M(\theta + \frac{t}{\lambda})}{M(\theta)}\right)^{\lambda} < \infty, \qquad t \in \lambda(\Theta - \theta). \tag{1.6}$$

Proof. From the definition of an mgf we have that

$$\overline{M}_{\lambda,\theta}(t) = \int \exp(xt) \exp(\lambda(\theta x - \kappa(\theta))) dP_\lambda(x)$$

then, applying the change of variable $y = \lambda x$ in this last expression we get that

$$\overline{M}_{\lambda,\theta}(t) = \int \exp\left(\frac{y}{\lambda} t\right) \exp[\theta y - \lambda\kappa(\theta)] dQ_\Lambda(y) = M_{\lambda,\theta}\left(\frac{t}{\lambda}\right) = \left(\frac{M(\theta + \frac{t}{\lambda})}{M(\theta)}\right)^{\lambda},$$

which, from Theorem 1.1, is finite for $t \in \lambda(\Theta - \theta)$. □

Theorem 1.10. *In $ED^*(Q, \Lambda_0)$, let $(\lambda, \theta) \in \Lambda \times \Theta$ and Y be a random variable with law $P_{\lambda,\theta}$, then $\mathbb{E}[Y] = \kappa'(\theta)$ and $\mathbb{V}[Y] = \kappa''(\theta)/\lambda$.*

Proof. From Theorem 1.8 we have that $Y = X/\lambda$ where X has law $Q_{\lambda,\theta}$, and then from Theorem 1.3

$$\mathbb{E}[Y] = \mathbb{E}\left[\frac{X}{\lambda}\right] = \frac{1}{\lambda}\mathbb{E}[X] = \frac{1}{\lambda}(\lambda\kappa'(\theta)) = \kappa'(\theta)$$
$$\mathbb{V}[Y] = \mathbb{V}\left[\frac{X}{\lambda}\right] = \frac{1}{\lambda^2}\mathbb{V}[X] = \frac{1}{\lambda^2}(\lambda\kappa''(\theta)) = \frac{\kappa''(\theta)}{\lambda}$$

□

1.4 The Variance Function

Definition 1.6. *The mean-space of a $ED^*(Q, \Lambda_0)$ is defined as*

$$\Omega = \kappa'(\Theta).$$

As a consequence of Theorem 1.5 we have that when Q is not concentrated at one point, κ' is strictly increasing and therefore invertible. When this happens we can write the canonical parameter as a function of the mean in the following way: Let X be in $ED^*(Q, \Lambda_0)$ with probability law $P_{\lambda,\theta}$ and let $\tau = \kappa'$ and $\mu = \mathbb{E}[X]$. Then we have that $\theta = \tau^{-1}(\mu)$, and then the variance of X can be written as $\mathbb{V}[X] = \frac{1}{\lambda}(\kappa'' \circ \tau^{-1})(\mu)$, which motivates the following definition.

Definition 1.7. *Let $ED^*(Q, \Lambda_0)$ be such that Q is not concentrated at one point and let $\tau = \kappa'$, then the variance function of the family is defined as $\mathbf{V} : \Omega \to (0, \infty)$,*

$$\mathbf{V}(\mu) = \kappa'' \circ \tau^{-1}(\mu).$$

After this definition we see that $\mathbb{V}[X] = \mathbf{V}(\mu)/\lambda$, thus the variance function allows us to express the variance of a random variable in an exponential family as the scale parameter times a function of the mean. Moreover, the variance function is important as it characterizes the family. This will be shown in the following theorems.

Theorem 1.11. *Let $\theta_0 \in \Theta$, $\mu_0 = \tau(\theta_0)$ and θ be another element in Θ, then there exists $\mu_1 \in \Omega$ such that*

$$\theta = \theta_0 + \int_{\mu_0}^{\mu_1} \frac{1}{\mathbf{V}(m)} dm$$

Proof. Let μ be any member of Ω, then as τ is invertible and $\tau' \neq 0$ by the inverse function theorem we have that

$$(\tau^{-1})'(\mu) = \frac{1}{\tau'(\tau^{-1}(\mu))}.$$

But $\tau'(\tau^{-1}(\mu)) = \kappa''(\tau^{-1}(\mu)) = \mathbf{V}(\mu)$, then we have that

$$(\tau^{-1})'(\mu) = \frac{1}{\mathbf{V}(\mu)}. \tag{1.7}$$

Thus

$$\tau^{-1}(\mu) - \tau^{-1}(\mu_0) = \int_{\mu_0}^{\mu} (\tau^{-1})'(m) dm = \int_{\mu_0}^{\mu} \frac{1}{\mathbf{V}(m)} dm,$$

as this last expression is true for every $\mu \in \Omega$ then it is true for $\mu_1 = \tau(\theta)$. Thus substituting $\theta = \tau^{-1}(\mu_1)$ and $\theta_0 = \tau^{-1}(\mu_0)$ in the last expression we get that

$$\theta = \theta_0 + \int_{\mu_0}^{\tau(\theta)} \frac{1}{\mathbf{V}(m)} dm = \theta_0 + \int_{\mu_0}^{\mu_1} \frac{1}{\mathbf{V}(m)} dm.$$

□

Theorem 1.12. *Let θ_0 and θ be elements of Θ, then*

$$\kappa(\theta) = \kappa(\theta_0) + \int_{\mu_0}^{\mu} \frac{m}{\mathbf{V}(m)} dm,$$

where $\mu_0 = \tau(\theta_0)$ and $\mu = \tau(\theta)$.

Proof. Define the function $h : \Theta \longrightarrow \mathbb{R}$ as

$$h(\theta) = \int_{\mu_0}^{\tau(\theta)} \frac{m}{\mathbf{V}(m)} dm.$$

By the fundamental theorem of calculus and the chain rule we get that

$$h'(\theta) = \frac{\tau(\theta)}{\mathbf{V}(\tau(\theta))}\tau'(\theta) = \frac{\tau(\theta)\tau'(\theta)}{(\tau' \circ \tau^{-1})(\tau(\theta))} = \frac{\tau(\theta)\tau'(\theta)}{\tau'(\theta)} = \tau(\theta)$$

and then, from the definition of h,

$$\int_{\mu_0}^{\mu} \frac{m}{\mathbf{V}(m)} dm - 0 = h(\theta) - h(\theta_0) = \int_{\theta_0}^{\theta} h'(x)dx = \int_{\theta_0}^{\theta} \tau(x)dx = \kappa(\theta) - \kappa(\theta_0),$$

which implies that

$$\kappa(\theta) = \kappa(\theta_0) + \int_{\mu_0}^{\mu} \frac{m}{\mathbf{V}(m)} dm.$$

□

Remark 1.2. *The proof of Theorems 1.11 and 1.12 depends on the fact that Ω is an interval. This is true because κ' is a continuous function and $\Omega = \kappa'(\Theta)$, where Θ is an interval.*

The last two theorems show how, given the variance function of an appropriate Exponential Dispersion Family, it is possible to obtain Θ and κ. Once κ is known then M is also known and as it is defined in an open interval that contains zero (because we are in an appropriate EDF) then Q can be obtained from M. Once Q is known then Λ, the index set, is also characterized. Thus, given the variance function, the family is characterized except for the subset of the index set that is used to generate a given family.

1.5 The Unit Deviance

Theorem 1.13. *Fix $\lambda \in \Lambda$ and let θ be any element from Θ. Then for every $A \in \mathcal{B}(\mathbb{R})$, $P_{\lambda,\theta}(A) = 0$ if and only if $P_\lambda(A) = 0$.*

Proof. From the definition of $P_{\lambda,\theta}$ (Definition 1.4) we immediately have that for every $\theta \in \Theta$, the value of $P_\lambda(A)$ implies that $P_{\lambda,\theta}(A) = 0$. Now, in order to prove the converse implication, suppose that $P_{\lambda,\theta}(A) = 0$ for some $A \in \mathcal{B}(\mathbb{R})$. Define $f : \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}$ with $f(x) = \exp\{\lambda[\theta x - \kappa(\theta)]\}$, then we have that

$$\int_A f dP_\lambda = \int f I_A dP_\lambda = 0.$$

As f is non-negative, then $fI_A = 0$ almost surely with respect to P_λ. Then $f(x) > 0$, for every x, implies that $P_\lambda(A) = 0$ and hence $P_{\lambda,\theta}(A) = 0$. Thus $P_{\lambda,\theta}(A) = 0$ implies $P_\lambda(A) = 0$. □

The previous theorem shows that the support of the members of a $REDF$ varies only with λ, which justifies the following definition.

Definition 1.8. *Let $ED^*(Q, \Lambda)$ be a $REDF$ and let C_λ be the support of P_λ for each $\lambda \in \Lambda$. Then, the support of the family is given by*

$$C = \bigcup_{\lambda \in \Lambda} C_\lambda.$$

Definition 1.9. *Let $ED^*(Q, \Lambda)$ be a $REDF$. The family deviance is defined as the function $d : C \times \Omega \longrightarrow \mathbb{R}$ with*

$$d(y, \mu) = 2\left[\sup_{\theta \in \Theta}\{y\theta - \kappa(\theta)\} - y\tau^{-1}(\mu) + \kappa(\tau^{-1}(\mu))\right].$$

Theorem 1.14. *Let $ED^*(Q, \Lambda)$ be a $REDF$. Then there exists $a : \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}$ such that for every $(\lambda, \theta) \in \Lambda \times \Theta$*

$$P_{\lambda,\theta}(A) = \int_A a(y, \lambda) \exp\left(-\frac{\lambda}{2} d(x, \mu)\right) dP_\lambda(x). \tag{1.8}$$

Proof. Let $(\lambda, \theta) \in \Lambda \times \Theta$ and let $f : C_\lambda \longrightarrow \mathbb{R}$ with

$$f(x) = \exp\{\lambda[\theta x - \kappa(\theta)]\},$$

then we have that

$$P_{\lambda,\theta}(A) = \int_A f dP_\lambda.$$

From the definition of τ (Definition 1.7), we have that $\theta = \tau^{-1}(\mu)$. Substituting in the previous equation gives

$$\begin{aligned}
f(x) &= \exp\left(\lambda\{\tau^{-1}(\mu)x - \kappa(\tau^{-1}(\mu))\}\right) \\
&= \exp\left(-\lambda\{-\tau^{-1}(\mu)x + \kappa(\tau^{-1}(\mu))\}\right) \\
&= \exp\left(-\lambda\left\{-\sup_{\theta\in\Theta}\{x\theta - \kappa(\theta)\} + \sup_{\theta\in\Theta}\{x\theta - \kappa(\theta)\} - \tau^{-1}(\mu)x + \kappa(\tau^{-1}(\mu))\right\}\right) \\
&= \exp\left(\lambda\sup_{\theta\in\Theta}\{x\theta - \kappa(\theta)\}\right)\exp\left(-\frac{\lambda}{2}\left\{2\left[\sup_{\theta\in\Theta}\{x\theta - \kappa(\theta)\} - \tau^{-1}(\mu)x + \kappa(\tau^{-1}(\mu))\right]\right\}\right) \\
&= \exp\left(\lambda\sup_{\theta\in\Theta}\{x\theta - \kappa(\theta)\}\right)\exp\left(-\frac{\lambda}{2}d(x,\mu)\right).
\end{aligned}$$

Now, if we define

$$a(x,\lambda) = \exp\left(\lambda\sup_{\theta\in\Theta}\{x\theta - \kappa(\theta)\}\right),$$

then

$$f(x) = a(x,\lambda)\exp\left(-\frac{\lambda}{2}d(x,\mu)\right).$$

□

Corollary 1.3. *Every REDF can be parametrized in terms of its mean-space (see Definition 1.6) and index set (see Definition 1.1).*

Chapter 2

The Generalized Linear Models

2.1 Distributional Assumptions

Let Y, X and W be random variables with supports on $\mathbb{R}$, $\mathbb{R}^p$ and $\mathbb{R}$ respectively. Suppose that there exists a probability measure Q with index set Λ and $\Lambda_0 \subset \Lambda$ such that for every $\boldsymbol{x}$ in the support of X (supp(X)) and $w \in \text{supp}(W)$, the conditional distribution of Y, given $X = \boldsymbol{x}$ and $W = w$, is in $ED^*(Q, \Lambda_0)$, *i.e.* it belongs to a fixed REDF. Assume that there exists $c : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ such that each P_λ has density function $c(\cdot, \lambda)$ with respect to some fixed measure. If $f_{Y|X,W}$ denotes the density of Y given $X = \boldsymbol{x}, W = w$, then

$$f_{Y|X,W}(y|\boldsymbol{x}, w) = c(y, \lambda) \exp(\lambda \{\theta y - \kappa(\theta)\}), \tag{2.1}$$

for some $(\lambda, \theta) \in \Lambda_0 \times \Theta$ that depend on x and w in the following way:

1. There exists σ^2 such that for every x and w,

$$\lambda = \frac{w}{\sigma^2}.$$

   This implies that λ varies only with the value of W.

2. Let $\mu_{\boldsymbol{x},w} = \mathbb{E}\left[Y|X = \boldsymbol{x}, W = w\right]$. There exists a fixed vector $\boldsymbol{\beta} = (\beta_1, ..., \beta_p) \in \mathbb{R}^p$ and a one-to-one differentiable function h such that

$$\mu_{\boldsymbol{x},w} = h(\boldsymbol{x^T\beta})$$

If g denotes the inverse of h, then writing the last equation in terms of g gives

$$g(\mu) = \boldsymbol{x^T\beta}.$$

In the jargon of GLMs, Y is called the response variable, the components $X_1, ..., X_p$ of X are called the covariates, σ^2 is called the dispersion parameter, g is called the link function and the conditional distribution of Y, given $X = \boldsymbol{x}, W = w$, is called the response distribution.

Remark 2.1. *In practice, besides the response distribution, the link function has to be chosen. There is substantial freedom for this choice but in order for this construction to make sense, g should be chosen such that* $\text{range}(h) \subset \Omega$ *(the mean-space).*

2.2 Estimation

The estimation part of the GLMs consists in estimating $\boldsymbol{\beta}$ and σ from a given sample. In many cases one is interested in estimating $\mu_{\boldsymbol{x},w}$, for which only a $\boldsymbol{\beta}$ estimation is necessary. The following theorem will be useful to find the Maximum Likelihood Estimator(MLE) of $\boldsymbol{\beta}$.

Theorem 2.1. *Suppose a population that follows the distributional assumptions for some fixed REDF and link function. Let $\{(y_i, \boldsymbol{x_i}, w_i)\}_{i=1}^n$ be a sample of (Y, X, W),* *where* $\boldsymbol{x_i} = (x_{i1}, ..., x_{ip})$*. Let*

$$\boldsymbol{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \qquad \boldsymbol{X} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ & \vdots & \\ x_{np} & \cdots & x_{np} \end{pmatrix}$$

and let $\boldsymbol{H}$*,* $\boldsymbol{\Sigma}$ *and* $\boldsymbol{\mu}$ *be depend on* $\boldsymbol{\beta}$ *with*

$$\boldsymbol{H}(\boldsymbol{\beta}) = \operatorname{diag}(h'(\boldsymbol{x_i}^T\boldsymbol{\beta}))_{i=1}^n, \qquad \boldsymbol{\Sigma}(\boldsymbol{\beta}) = \operatorname{diag}\left(\frac{w_i}{V(h(\boldsymbol{x_i}^T\boldsymbol{\beta}))}\right),$$

$$\text{and} \qquad \boldsymbol{\mu}(\boldsymbol{\beta}) = \begin{pmatrix} h(\boldsymbol{x_1}^T\boldsymbol{\beta}) \\ \vdots \\ h(\boldsymbol{x_n}^T\boldsymbol{\beta}) \end{pmatrix}.$$

Assume also that the MLE of $\boldsymbol{\beta}$ *exists and lets denote it with* $\hat{\boldsymbol{\beta}}$*. Then*

$$\boldsymbol{X}^T\boldsymbol{H}(\hat{\boldsymbol{\beta}})\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})(\boldsymbol{Y} - \boldsymbol{\mu}(\hat{\boldsymbol{\beta}})) = \boldsymbol{0}, \tag{2.2}$$

where $\boldsymbol{0}$ *is a vector of size* p *whose entries are all zero.*

Proof. For each i, the conditional density function of Y, given $X = \boldsymbol{x_i}$ and $W = w_i$, evaluated at y_i is given by

$$f_i(y_i) = f_{Y|X,W}(y_i|\boldsymbol{x_i}, w_i) = c\left(y_i, \frac{w_i}{\sigma^2}\right)\exp\left(\frac{w_i}{\sigma^2}\{\theta_i y_i - \kappa(\theta_i)\}\right), \qquad \text{for some } \theta_i \in \Theta. \tag{2.3}$$

We have that for each i

$$\theta_i = \tau^{-1}(\mu_{\boldsymbol{x_i},w}) = (\tau^{-1} \circ h)(\boldsymbol{x_i}^T\boldsymbol{\beta}),$$

substituting this in (2.3) we obtain

$$f_i(y_i) = c\left(y_i, \frac{w_i}{\sigma^2}\right) \exp\left(\frac{w_i}{\sigma^2}\left\{(\tau^{-1}\circ h)(\boldsymbol{x_i}^T\boldsymbol{\beta})y_i - \left((\kappa\circ\tau^{-1}\circ h)(\boldsymbol{x_i}^T\boldsymbol{\beta})\right)\right\}\right).$$

Then the likelihood function for $\boldsymbol{\beta}$ is given by

$$\begin{aligned} L(\boldsymbol{\beta}) &= \prod_{i=1}^{n}\left\{c\left(y_i, \frac{w_i}{\sigma^2}\right) \exp\left(\frac{w_i}{\sigma^2}\left\{(\tau^{-1}\circ h)(\boldsymbol{x_i}^T\boldsymbol{\beta})y_i - \left((\kappa\circ\tau^{-1}\circ h)(\boldsymbol{x_i}^T\boldsymbol{\beta})\right)\right\}\right)\right\} \\ &= \left(\prod_{i=1}^{n} c\left(y_i, \frac{w_i}{\sigma^2}\right)\right) \exp\left(\frac{1}{\sigma^2}\sum_{i=1}^{n} w_i\left\{(\tau^{-1}\circ h)(\boldsymbol{x_i}^T\boldsymbol{\beta})y_i - \left((\kappa\circ\tau^{-1}\circ h)(\boldsymbol{x_i}^T\boldsymbol{\beta})\right)\right\}\right). \end{aligned}$$

Thus, the log-likelihood function is

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{n} c\left(y_i, \frac{w_i}{\sigma^2}\right) + \frac{1}{\sigma^2}\sum_{i=1}^{n} w_i\left\{(\tau^{-1}\circ h)(\boldsymbol{x_i}^T\boldsymbol{\beta})y_i - \left((\kappa\circ\tau^{-1}\circ h)(\boldsymbol{x_i}^T\boldsymbol{\beta})\right)\right\}. \quad (2.4)$$

Let $k \in \{1, 2, ..., p\}$, then

$$\frac{\partial}{\partial\beta_k}\ell(\boldsymbol{\beta}) = \frac{1}{\sigma^2}\sum_{i=1}^{n} w_i\left\{y_i\left(\frac{\partial}{\partial\beta_k}(\tau^{-1}\circ h)(\boldsymbol{x_i}^T\boldsymbol{\beta})\right) - \frac{\partial}{\partial\beta_k}(\kappa\circ\tau^{-1}\circ h)(\boldsymbol{x_i}^T\boldsymbol{\beta})\right\}. \quad (2.5)$$

Now, for each $i = 1, ..., n$

$$\begin{aligned} \frac{\partial}{\partial\beta_k}(\tau^{-1}\circ h)(\boldsymbol{x_i}^T\boldsymbol{\beta}) &= \frac{\partial}{\partial\beta_k}(\tau^{-1}\circ h)\left(\sum_{j=i}^{p} x_{ij}\beta_j\right) \\ &= (\tau^{-1})'\left(h\left(\sum_{j=i}^{p} x_{ij}\beta_j\right)\right) h'\left(\sum_{j=i}^{p} x_{ij}\beta_j\right) x_{ik} \\ &= (\tau^{-1})'\left(h\left(\boldsymbol{x_i}^T\boldsymbol{\beta}\right)\right) h'\left(\boldsymbol{x_i}^T\boldsymbol{\beta}\right) x_{ik}. \end{aligned}$$

But, from equation (1.7) we know that $(\tau^{-1})'(h(\boldsymbol{x_i}^T\boldsymbol{\beta})) = \frac{1}{V(h(\boldsymbol{x_i}^T\boldsymbol{\beta}))}$, thus

$$\frac{\partial}{\partial\beta_k}(\tau^{-1}\circ h)(\boldsymbol{x_i}^T\boldsymbol{\beta}) = \frac{h'(\boldsymbol{x_i}^T\boldsymbol{\beta})}{V(h(\boldsymbol{x_i}^T\boldsymbol{\beta}))}x_{ik}. \quad (2.6)$$

On the other hand,

$$\frac{\partial}{\partial \beta_k}(\kappa \circ \tau^{-1} \circ h)(\boldsymbol{x_i}^T \boldsymbol{\beta}) = \kappa'((\tau^{-1} \circ h)(\boldsymbol{x_i}^T \boldsymbol{\beta}))\frac{\partial}{\partial \beta_k}(\tau^{-1} \circ h)(\boldsymbol{x_i}^T \boldsymbol{\beta}),$$

and then, from (2.6) and as $k' = \tau$,

$$\frac{\partial}{\partial \beta_k}(\kappa \circ \tau^{-1} \circ h)(\boldsymbol{x_i}^T \boldsymbol{\beta}) = \frac{h(\boldsymbol{x_i}^T \boldsymbol{\beta})h'(\boldsymbol{x_i}^T \boldsymbol{\beta})}{V(h(\boldsymbol{x_i}^T \boldsymbol{\beta}))} x_{ik}. \tag{2.7}$$

By substituting equations (2.6) and (2.7) in (2.5), the following equation can be obtained

$$\frac{\partial}{\partial \beta_k}\ell(\boldsymbol{\beta}) = \frac{1}{\sigma^2}\sum_{i=1}^{n} w_i \left(\frac{y_i - h(\boldsymbol{x_i}^T \boldsymbol{\beta})}{V(h(\boldsymbol{x_i}^T \boldsymbol{\beta}))}\right) h'(\boldsymbol{x_i}^T \boldsymbol{\beta}) x_{ik}. \tag{2.8}$$

Now, as V, h and h' are continuous (V is continuous as it is the composition of two continuous functions: $\kappa'' \circ \tau^{-1}$) the last expression implies that $\frac{\partial}{\partial \beta_k}\ell$ is a continuous function whose domain is $\mathbb{R}^p$ for each $k \in \{1, 2, ..., p\}$. This, combined with the fact that $\frac{\partial}{\partial \beta_k}\ell(\boldsymbol{\beta})$ reaches a maximum at $\hat{\boldsymbol{\beta}}$ implies that

$$\frac{\partial}{\partial \beta_k}\ell(\hat{\boldsymbol{\beta}}) = 0, \qquad \text{for } k = 1, ..., p.$$

From equation (2.8) and by performing standard operations it is possible to see that for every $\boldsymbol{\beta} \in \mathbb{R}^p$,

$$\sigma^2 \begin{pmatrix} \frac{\partial}{\partial \beta_1}\ell(\boldsymbol{\beta}) \\ \vdots \\ \frac{\partial}{\partial \beta_k}\ell(\boldsymbol{\beta}) \end{pmatrix} = \boldsymbol{X}^T \boldsymbol{H}(\hat{\boldsymbol{\beta}})\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})(\boldsymbol{Y} - \boldsymbol{\mu}(\hat{\boldsymbol{\beta}})),$$

which, combined with the fact that $\sigma^2 > 0$ implies that

$$\boldsymbol{X}^T \boldsymbol{H}(\hat{\boldsymbol{\beta}})\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})(\boldsymbol{Y} - \boldsymbol{\mu}(\hat{\boldsymbol{\beta}})) = \boldsymbol{0}.$$

□

An alternative proof of this theorem can be found in page 104 of Madsen and Thyregod (2011).

Thus, we have that if the MLE exists it can be found by solving equation (2.2). This can be done by using iteratively reweighted least squares. This method will not be explained here, but a detailed account can be found in Chapter 4 of Björk (1996).

Definition 2.1. *Let* $\{(y_i, \boldsymbol{x_i}, w_i)\}_{i=1}^n$ *be a sample of* (Y, X, W)*, where* Y*,* X *and* W *follow the distributional assumptions from the previous section, with* $X \in \mathbb{R}^p$*. Then, for each* $\boldsymbol{\beta} \in \mathbb{R}^p$ *the residual deviance is defined as*

$$D(\boldsymbol{\beta}) = \sum_{i=1}^{n} w_i d(y_i, h(\boldsymbol{x}_i^T \boldsymbol{\beta})).$$

Theorem 2.2. *Let* $\{(y_i, \boldsymbol{x_i}, w_i)\}_{i=1}^p$ *be a sample of* (Y, X, W)*. If* $\hat{\beta}$*, the MLE of* $\boldsymbol{\beta}$*, exists then it minimizes the residual deviance.*

Proof. By using the parametrization from (1.8), (2.1) can be written as

$$f_{Y|X,W}(y|\boldsymbol{x}, w) = c(y, \lambda) a(y, \lambda) \exp\left(-\frac{\lambda}{2} d(y, h(\boldsymbol{x}^T \boldsymbol{\beta}))\right).$$

Let $c^*(y, \lambda) = c(y, \lambda) a(y, \lambda)$, then the likelihood function for $\boldsymbol{\beta}$ can be written as

$$L(\boldsymbol{\beta}) = \left(\prod_{i=1}^{n} c^*(y, \lambda)\right) \exp\left(-\frac{1}{\sigma^2} \sum_{i=1}^{n} w_i d(y_i, h(\boldsymbol{x_i}^T \boldsymbol{\beta}))\right)$$

and the corresponding log-likelihood function as

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{n} c^* \left(y_i, \frac{w_i}{\sigma^2}\right) - \frac{1}{\sigma^2} D(\boldsymbol{\beta}).$$

As the first term of the right hand side of the above equation is the same for every value of $\boldsymbol{\beta}$, from this equation it is possible to see that $\ell(\boldsymbol{\beta_1}) \leq \ell(\boldsymbol{\beta_2})$ if and only if $D(\boldsymbol{\beta_1}) \geq D(\boldsymbol{\beta_2})$, which implies that if $\hat{\boldsymbol{\beta}}$ exists then it minimizes D. □

Chapter 3

The Tweedie Families of Distributions

Definition 3.1. *A REDF is called a Tweedie Family if the domain of its variance function V is $(0, \infty)$ with*

$$V(\mu) = \mu^p,$$

for some $p \in \mathbb{R}$.

The Tweedie families contain many distributions that are characterized by the value of p. The following table presents the well known distributions that can be seen as a Tweedie family for different values of p.

Value of p	**Distribution**
$p = 0$	Normal
$p = 1$	Poisson
$p \in (1, 2)$	Compound Poisson - Gamma
$p = 2$	Gamma
$p = 3$	Inverse Gaussian

In addition to this, it is known that for $p < \infty$ the Tweedie families characterize distributions that are supported on $\mathbb{R}$, while for $p > 2$ it characterizes distributions that are supported on $(0, \infty)$. The case $p \in (0, 1)$ does not correspond to any probability measure. In this chapter it will only be proved that the compound Poisson-gamma can be reparametrized to be a Tweedie family with $p \in (1, 2)$. A proof for other cases can be found in Chapter 4 of Jørgensen (1997).

3.1 The Compound Poisson-Gamma Distribution

The goal of this section is to define and derive the mgf of the compound Poisson-gamma distribution. This will be useful to prove the existence of the Tweedie families for $p \in (1, 2)$. A few necessary definitions and theorems are stated first.

Definition 3.2. *Let N be a random variable with probability measure*

$$P(A) = \sum_{n=0}^{\infty} \frac{e^{-m} m^n}{n!} I_A(n), \qquad A \in \mathcal{B}(\mathbb{R}),$$

where $m > 0$. Then N is said to follow a Poisson distribution with parameter m and it is denoted $N \sim Poisson(m)$

Definition 3.3. *Let $N \sim Poisson(m)$, $X_0 = 0$ and $X_1, X_2, ...$ be independent and identically distributed (iid) random variables with probability distribution F and also independent from N. The probability distribution of*

$$S = \sum_{i=0}^{N} X_i$$

is called a compound Poisson distribution with rate m and jump size distribution F.

The proof of the following lemma can be found in Chapter 1 of Gerber (1979).

Lemma 3.1. *Let S have a compound Poisson distribution with rate m and jump distribution F. Assume also that F has mgf M_F, then the mgf of S is given by*

$$M(t) = \exp[m(M_F(t) - 1)]. \tag{3.1}$$

Definition 3.4. *Let X be a random variable with probability measure*

$$P(A) = \int_{A\bigcap(0,\infty)} \frac{x^{\alpha-1}\exp(-\beta x)\beta^{\alpha}}{\Gamma(\alpha)}dx, \qquad A \in \mathcal{B}(\mathbb{R}),$$

where $\alpha, \beta > 0$. Then X is said to follow a gamma distribution with shape parameter α and scale parameter β and it is denoted with $X \sim gamma(\alpha, \beta)$.

The following lemma can be found in the Appendix A of Klugman and Willmot (2004).

Lemma 3.2. *Let $X \sim gamma(\alpha, \beta)$, then the mgf of X is given by*

$$M(t) = \left(\frac{1}{1-\frac{t}{\beta}}\right)^{\alpha}, \qquad t < \beta. \tag{3.2}$$

Even though the Poisson, the compound Poisson and the gamma are well known distributions, their definitions are given here in order to set the parametrization used here, and for the sake of completeness. Now it is possible to state the definition of the compound Poisson-gamma distribution.

Definition 3.5. *Let S have a compound Poisson distribution with rate m and jump size distribution $gamma(\alpha, \beta)$. Then S is said to follow a compound Poisson-gamma distribution with parameters m, α, β and it will be denoted $S \sim CPG(m, \alpha, \beta)$.*

Theorem 3.1. *Let $S \sim CPG(m, \alpha, \beta)$, then the mgf of S is given by*

$$M(t) = \exp\left(m\left\{\left(\frac{1}{1-\frac{t}{\beta}}\right)^{\alpha} - 1\right\}\right), \qquad t < \beta. \tag{3.3}$$

Proof. It follows from substituting equation (3.2) in (3.1). □

3.2 General Properties

In this section some of the properties of the Tweedie family for the case $p \in (1,2)$ are presented.

Theorem 3.2. *Suppose that $ED^*(Q,\Lambda)$ is an appropriate Reproductive Exponential Dispersion Family. Also assume that the domain of its variance function is $(0,\infty)$ with*

$$V(\mu) = \mu^p,$$

for some $p \in (1,2)$. Then there exists $a > 0$ such that the mgf of Q has domain $\Theta = (-\infty, (1-\xi)a^{\frac{1}{\xi-1}})$ with

$$M(\theta) = \exp\left\{\left(\frac{\xi-1}{\xi}\right)\left[\left(a^{\frac{1}{\xi-1}} + \frac{\theta}{\xi-1}\right)^{\xi} - a^{\frac{\xi}{\xi-1}}\right]\right\}, \tag{3.4}$$

where $\xi = \frac{p-2}{p-1}$.

Proof. By the assumption that $ED^*(Q,\Lambda)$ is appropriate, k is differentiable at 0 so we can define $a = \tau(0)$ $(\tau = \kappa')$. Since the domain of V is $(0,\infty)$ we have that $a > 0$. From the definition of the variance function, for every $m \in \Omega$

$$(\tau')^{-1}(m) = m^p.$$

Let θ be any member of Θ and $\mu = \tau(\theta)$. Then by integrating both sides of the last equation above from 0 to μ, and from the definition of τ the following relation can be obtained

$$\theta = \frac{\kappa'(\theta)^{1-p} - a^{1-p}}{1-p},$$

which implies that

$$\kappa'(\theta) = \left(a^{1-p} + \theta(1-p)\right)^{\frac{1}{1-p}}. \tag{3.5}$$

In addition, as the domain of V is $(0,\infty)$ then the canonical space is given by

$$\Theta = \{\theta \in \mathbb{R} : \kappa'(\theta) > 0\}.$$

Thus, $\theta \in \Theta$ if and only if $(a^{1-p} + \theta(1-p))^{\frac{1}{1-p}} > 0$ therefore

$$\Theta = \left\{\theta \in \mathbb{R} : \theta < \frac{a^{1-p}}{p-1}\right\}.$$

Now, integrating κ' from 0 to θ it is possible to obtain that

$$\kappa(\theta) = \frac{(a^{1-p} + \theta(1-p))^{\frac{2-p}{1-p}} - a^{2-p}}{2-p}, \qquad \theta < \frac{a^{1-p}}{p-1}.$$

Finally, by defining $\xi = \frac{p-2}{p-1}$ and rewriting the last expression in terms of ξ we get that

$$\kappa(\theta) = \left(\frac{\xi-1}{\xi}\right)\left[\left(a^{\frac{1}{\xi-1}} + \frac{\theta}{\xi-1}\right)^{\xi} - a^{\frac{\xi}{\xi-1}}\right], \qquad \theta < (1-\xi)a^{\frac{1}{\xi-1}}$$

and therefore

$$M(\theta) = \exp\left\{\left(\frac{\xi-1}{\xi}\right)\left[\left(a^{\frac{1}{\xi-1}} + \frac{\theta}{\xi-1}\right)^{\xi} - a^{\frac{\xi}{\xi-1}}\right]\right\}, \qquad \theta \in (-\infty, (1-\xi)a^{\frac{1}{\xi-1}}).$$

□

It is important to emphasize that the previous theorem does not prove the existence of the Tweedie families for $p \in (1,2)$. It just gives the mgf of a generator of the family assuming that such a family exists and that it is appropriate. The following corollary goes in the same direction, *i.e.* it builds assuming the existence of the family.

Corollary 3.1. *Let $ED^*(Q,\Lambda)$ be appropriate. Suppose that the domain of its variance function is $(0,\infty)$ with $V(\mu) = \mu^p$ for some $p \in (1,2)$. Let $(\theta,\lambda) \in \Theta \times \Lambda$ and*

$\xi = \frac{p-2}{p-1}$, *then there exists* $a \in \mathbb{R}$ *such that*

$$M_{\lambda,\theta}(t) = \exp\left(\lambda\left(\frac{\xi-1}{\xi}\right)\left(a^{\frac{1}{\xi-1}} + \frac{\theta}{\xi-1}\right)^{\xi}\left[\left(1 + \frac{t}{\lambda((\xi-1)a^{\frac{1}{\xi-1}}+\theta)}\right)^{\xi} - 1\right]\right), \quad (3.6)$$

for $t \in (-\infty, \lambda((1-\xi)a^{\frac{1}{\xi-1}} - \theta))$.

Proof. It follows from substituting the $M(\theta)$ found in the previous theorem in equation (1.6). □

Theorem 3.3. *For every* $p \in (1,2)$ *there exists a probability measure* Q *with index set* $\Lambda = (0,\infty)$ *such that* $ED^*(Q,\Lambda)$ *is an appropriate REDF with variance function*

$$V(\mu) = \mu^p, \qquad \mu \in (0,\infty).$$

Moreover, the elements of $ED^*(Q,\Lambda)$ *are all compound Poisson-gamma.*

Proof. Suppose that it has been proved that equation (3.6) corresponds to a compound Poisson-gamma distribution for every $\lambda < 0$, $\theta \in \Theta$ and $p \in (1,2)$. In this case, as equation (3.4) is a especial case of (3.6) when $\lambda = 1$ and $\theta = 0$, then we would immediately have the existence of Q. Furthermore, as this would have been proved for every $\lambda > 0$ we would also have that $\Lambda = (0,\infty)$ and we would be done.

Let us now prove this. Let $\lambda > 0$, $\theta \in \Theta$ and $p \in (1,2)$. Define $\alpha = -\xi$ and notice that if $\xi < 0$ then $\alpha > 0$. Define $\beta = \lambda((1-\xi)a^{\frac{1}{\xi-1}} - \theta)$, as $\Theta = (-\infty, (1-\xi)a^{\frac{1}{\xi-1}})$ then $\theta < (1-\xi)a^{\frac{1}{\xi-1}}$ and therefore $(1-\xi)a^{\frac{1}{\xi-1}} - \theta > 0$. Since also $\lambda > 0$, we have that $\beta > 0$. Now define $m = \lambda\left(\frac{\xi-1}{\xi}\right)\left(a^{\frac{1}{\xi-1}} + \frac{\theta}{\xi-1}\right)^{\xi}$, as $\xi < 0$ then $\xi - 1 < 0$ and therefore $\frac{\xi-1}{\xi} > 0$. Similarly as $\theta > 0$ and $a > 0$, then $a^{\frac{1}{\xi-1}}$ and $\frac{\theta}{\xi-1} > 0$, thus $\left(a^{\frac{1}{\xi-1}} + \frac{\theta}{\xi-1}\right)^{\xi} > 0$ and therefore $m > 0$. Inserting these expressions for α, β and m in (3.3), we obtain equation (3.6). This proves that (3.6) is a compound Poisson-gamma (CPG) for every $\lambda > 0$, $\theta \in \Theta$ and $p \in (1,2)$. □

The previous theorems guarantee the existence of the Tweedie families for $p \in (1,2)$ and show that they are equivalent to a CPG distribution. From the proof of this theorem, we see that in order to find the corresponding parameters of the CPG distribution, we need the mean of the generating measure at $\theta = 0$. It is useful to express this equivalence without the need of the generating measure. This can be achieved by reparametrizing the family in terms of the mean-space. Thus, we denote by $Tw(p, \mu, \lambda)$ a Tweedie distribution with variance function power p, mean μ and scale parameter λ. Notice that this parametrization is justified by Corollary 1.3. By using this parametrization, the following theorem allows us to express the equivalence between the Tweedie family and the CPG without using the generator.

Theorem 3.4. *Let $p \in (1, 2)$, $\mu > 0$ and $\lambda > 0$, then*

$$Tw(p, \mu, \lambda) = CPG\left(\frac{\lambda\mu^{2-p}}{2-p}, -\frac{p-2}{p-1}, \frac{\lambda\mu^{1-p}}{p-1}\right).$$

Similarly, for $m, \alpha, \beta > 0$,

$$CPG(m, \alpha, \beta) = Tw\left(\frac{\alpha+2}{\alpha+1}, \frac{m\alpha}{\beta}, \frac{(m\alpha)^{\frac{\alpha+2}{\alpha+1}-1}\beta^{2-\frac{\alpha+2}{\alpha+1}}}{\alpha+1}\right).$$

Proof. Let $p \in (1, 2)$, $\mu > 0$ and $\lambda > 0$. Theorem 3.3 implies that there exists m, α and p such that $Tw(p, \mu, \alpha) = CPG(m, \alpha, \beta)$. Let X and Y be random variables such that $X \sim Tw(p, \mu, \alpha)$ and $Y \sim CPG(m, \alpha, \beta)$. As these two variables are equally distributed, we have that $\mathbb{E}[X] = \mathbb{E}[Y]$ and $\mathbb{V}[X] = \mathbb{V}[Y]$. By using the mgf from Lemma 3.2 it is possible to see that $\mathbb{E}[Y] = \frac{m\alpha}{\beta}$ and $\mathbb{V}[Y] = m\left(\frac{\alpha}{\beta^2} + \frac{\alpha^2}{\beta^2}\right)$. On the other hand, from the definition of the Tweedie distribution we have that $\mathbb{E}[X] = \mu$ and $\mathbb{V}[X] = \frac{\mu^p}{\lambda}$. In addition, from the proof of Theorem 3.3 we have that $\alpha = -\frac{p-2}{p-1}$. Thus we have the following set of equations

$$\alpha = -\frac{p-2}{p-1} \qquad \mu = \frac{m\alpha}{\beta} \qquad \frac{\mu^p}{\lambda} = m\left(\frac{\alpha}{\beta^2} + \frac{\alpha^2}{\beta^2}\right) \tag{3.7}$$

by solving for m and β, we get that $m = \frac{\lambda\mu^{2-p}}{2-p}$ and $\beta = \frac{\lambda\mu^{1-p}}{p-1}$. Similarly, let $m, \alpha, \beta > 0$. By solving (3.7) for p, μ and α we obtain the second distribution equivalence in the theorem. □

3.3 GLMs for the Tweedie Families

As EDFs, the Tweedie families are suitable as densities for GLMs. In this case, the methods presented in Chapter 2 can be used to find estimators for $\boldsymbol{\beta}$. The case when p needs to be estimated is more complex and it will not be treated here. A way to do it is explained in Gilchrist and Drinkwater (2000). This method has been implemented in R (see R Development Core Team, 2010) in the Tweedie package (see Dunn, 2010).

Chapter 4

Modeling Premiums with Tweedie Families using GLMs

The goal of this chapter is to show how the Tweedie distribution can be used for modeling insurance premiums with the aid of GLMs. The first sections introduce to non-life insurance and the basic assumptions that are made for premium modeling.

4.1 Non-life Insurance

Definition 4.1. *A non-life insurance policy is an agreement between two parties, in which one of them engages to compensate the other party for certain unpredictable losses during a fixed time period, in exchange of a fee. The compensating party is called the insurer, the other party is called the policyholder and the fee is called the insurance premium.*

The main idea behind insurance is that every policyholder pays a premium, which is considerably smaller than the potential loss, but not all of them claim for financial compensation. Thus, intuitively, the money of many pays for the losses of few.

In order to be solvent, an insurance company needs to charge enough money in premiums to face its liabilities. Deciding how much the premium should be is not an easy task, and this is where probability and statistical models play a decisive role. Several models have been used for this and there is a common background strategy that most of them share. In order to introduce this strategy, some definitions are needed first.

Definition 4.2. *A claim is an event for which the policyholder demands financial compensation.*

Definition 4.3. *The money paid by the insurer to the policyholder as the result of a claim is called the size of the claim or simply the claim size.*

Definition 4.4. *The total claim size of a policy is the sum of the sizes of all the claims made during the validity of the policy. The total claim size of a group of policies if the sum of the total sizes of each policy in the group.*

Suppose you are an insurer and you have a group of n policyholders. Let S_i, for each $i = 1, ..., n$, be the total size of the i-th policy. As these amounts are not known at the beginning of the policy, we will treat them as random variables. Assume also that $\{S_i\}_{i=1}^n$ are iid with $\mu = \mathbb{E}[S_1] < \infty$. Then, by the law of large numbers

$$\lim_{n\to\infty} \frac{S_1 + S_2 + ... + S_n}{n} = \mu, \quad a.s.$$

which means that for large enough n,

$$\frac{S_1 + S_2 + ... + S_n}{n} \approx \mu$$

and therefore

$$S_1 + S_2 + ... + S_n \approx n\mu.$$

This implies that the total claim size of this group will be approximately $n\mu$. Thus, the amount of money charged to each policyholder should be based on μ. Something noticeable from this is that n is not needed to be known, it just has to be large.

Definition 4.5. *Suppose you have a group of policyholders. Let S_n be total claim size of the nth policyholder. The group is called homogeneous if $\{S_n\}$ are iid with $\mathbb{E}[S_1] < \infty$.*

As a consequence of the previous reasoning and definition, it is possible to summarize the usual steps followed by insurance companies to analyze what premium to charge:

1. Divide the policyholders into large enough homogeneous groups.

2. Estimate μ for each group.

3. Charge μ plus a risk factor plus fees (administration fees, profit margin, etc).

These steps carry implicitly the following two assumptions (referred from now on as the insurability assumptions):

1. There is a set of observable characteristics of the policyholders that allows to divide them into large enough homogeneous groups.

2. The mean of each homogeneous group can be accurately estimated.

It is not hard to find cases where these assumptions do not hold. For example in catastrophes like hurricanes or earthquakes, several policyholders will be affected at the same time; in this case it doesn't seem reasonable to consider independence between the S_is. Also, nothing guarantees that the segmenting characteristics exist or that they are observable.

When these assumptions are accepted, there is no unique way to follow the steps described to compute the premium. The purpose of this chapter is to show how this can be done with the GLMs and the Tweedie family of distributions.

4.2 The Key Ratios

Insurance companies usually offer their non-life policies with a term of one year. Nevertheless the actual in force time of the policy can be less (for example if a policyholder decides to cancel the policy).

Definition 4.6. *The duration of a policy is the amount of time a policy is in force. It is usually measured in years.*

One of the key assumptions described in the previous section was that $\{S_n\}$ are iid. Nevertheless, even within the same homogeneous group if a policy has a duration of 0.1 and another one has a duration of 0.9, it is not reasonable to consider that their respective total claim sizes are equally distributed. In this section, some key ratios are introduced along with some assumptions on them. This will help us deal with policies that have different durations. The following definitions are valid in the context of individual policies or groups of policies.

Definition 4.7. *The claim frequency is the number of claims divided by the duration, i.e. the average number of claims per unit time.*

Definition 4.8. *The claim severity is the total claim size divided by the amount of claims, i.e. the average size per claim.*

Definition 4.9. *The pure premium is the total claim size divided by the duration, i.e. the average amount paid per unit time.*

Definition 4.10. *The earned premium is the duration times the annual premium.*

Definition 4.11. *The loss ratio is the total claim size divided by the earned premium.*

Notice that the claim frequency, claim severity, pure premium and loss ratio are the result of a random outcome divided by a volume measure. As such they are called the key ratios. The volume measure is called the exposure for each case. The analysis of the earned premium is not the subject of this thesis, but is included here for the sake of completeness.

It is important to introduce some assumptions regarding the key ratios in order to see why they are useful. For this, let us introduce some notation. Consider a fixed policyholder and let N be his/her number of claims, X_i be the size of his/her $i-th$ claim and S be his/her total claim size. Notice that they are related by

$$S = \sum_{i=1}^{N} X_i \tag{4.1}$$

where $X_0 = 0$. As S, N and the X_i's are not known at the beginning of the policy, we treat them as random variables. We are interested in the distribution of these random variables.

Certainly the duration of the policy affects these distributions, thus we include it in the analysis. It is natural to think that the higher the duration, the more likely the policyholder is to have more accidents, i.e. the distribution of N is influenced by the duration. So we modify the notation to N_w, to recognize the dependency on the duration. Then equation (4.1) makes clear that the distribution of S also depends on the duration, so we will write S_w, but use S and N for $w = 1$.

Before considering the X_n's, let us see some of the implications of the introduction of the duration in the model. The notation N_w implies that the distribution of the number of claims depends only on the duration. This appears reasonable, but has some strong implications. In order to enumerate some, let w_1 and w_2 be the durations of two time intervals. The following statements follow:

1. If $w_1 = w_2$, then $N_{w_1} \stackrel{D}{=} N_{w_2}$(here $\stackrel{D}{=}$ means equal in distribution).

2. If w_1 and w_2 correspond to non-overlapping periods of time, then N_{w_1} is independent of N_{w_2}.

3. If w_1 and w_2 correspond to non-overlapping periods of time, then $N_{w_1} + N_{w_2} \stackrel{D}{=}$

$N_{w_1+w_2}$.

4. $\mathbb{E}[N_w] = w\mathbb{E}[N]$.

It is possible to find cases in which these implications do not hold. For example, in car insurance it is not likely that the distribution of the number of accidents for 1 month, during the winter, is the same that of 1 month during the summer. It could also happen that a driver will drive more carefully after a first accident, which will change the probability of having a second accident; such cases do not satisfy the independence assumptions for non-overlapping periods of time.

Now consider the claim sizes X_i. By definition, the distribution of X_i is conditioned to the occurrence of an i-th accident. Thus, it seems reasonable to assume that the distribution of the X_i's is not affected by the duration. Furthermore, as the X_i's denote payments for the same risk, we will consider them iid.

One last assumptions is needed: The X_i's and N_w are independent. At first, this seems as a natural assumption; Why would the number of claims affect the size of the payments? Nevertheless experience shows that this is not as natural as it seems. Often, events that produce more frequent claims are associated with lower size payments, while rare events are associated with large size claims.

To conclude this section, we put together the assumptions that have been introduced for the key ratios. They will be referred from now on as the key ratio assumptions:

1. For each policyholder, the distribution of the number of claims depends only on the duration.

2. The X_i's are iid, their distribution is not affected by the duration and they are independent of the number of claims.

4.3 Group Estimation

Several assumptions have been introduced in this chapter, along with examples that show how they are a simplification to reality. As with every theoretical model, we have to wonder if these simplifications approximate reality sufficiently to be useful. As it turns out these have been useful in several situations and form part of classical insurance models. For instance these are found found in Chapter 6 of Klugman and Willmot (2004) and Chapter 1 of Ohlsson and Johansson (2010). There have been extensions of these assumptions, for example copulas have been used to describe the dependency between the number and the size of the claims. Nevertheless, the results in this and the next section depend on these classical assumptions. Thus, from here on the insurability and the key ratio assumptions are considered to hold.

When the strategy used by insurance companies was that described in Section 4.1, the duration was not being considered. The duration complicates the analysis in two ways:

1. The duration for new policyholders is not known in advance.
2. In historical data for specific homogeneous groups, policyholders have different durations and therefore their total claim sizes are not equally distributed. This is a complication because the sample is then not iid.

The first complication is handled by insurers with a practical approach. All the policies are standardized to a duration of 1 and policyholders are charged accordingly. If, for some reason, the policy is terminated before the stipulated date, a portion of the pure premium is reimbursed to the policyholder, which usually is equal to the pure premium minus the earned premium. As a consequence, the estimation should be made based on a duration of 1.

In order to develop a way of handling the second complication, let us assume that we are in a homogeneous group with n policyholders, and adopt the following notation:

- w_i denotes the duration of the i-th policy.

- S_w^i, N_w^i, X_j^i denote the random variables for the total claim size, number and size of the j-th claim, respectively, for the i-th policyholder assuming, a duration of w. For $w = 1$, S^1 and N^1 are used for S_1^1 and N_1^1.

- μ denotes the pure premium of a policy with duration 1 for the policyholders in the group, i.e., $\mu = \mathbb{E}[S^1]$.

It is important to distinguish μ from the observed pure premium from a sample, thus, we give the following definition.

Definition 4.12. *The empirical pure premium of a homogeneous group of policies is defined as*

$$M = \frac{\sum_{i=1}^{n} S_{w_i}^i}{\sum_{i=1}^{n} w_i}. \tag{4.2}$$

Notice that in this definition, the numerator and denominator correspond, respectively, to the total claim size and duration of the group.

The proofs of the following lemmas can be found in the Appendix.

Lemma 4.1. *For every i and w,*

$$E[N_w^i] = wE[N^i].$$

Lemma 4.2. *For every i,*

$$\mu = \mathbb{E}[N^i]\mathbb{E}[X_1].$$

Lemma 4.3. *For every i and w,*

$$\mathbb{E}[S_w^i] = w\mu.$$

Now, it is possible to prove the main result from this section, the unbiasedness of M in (4.2).

Theorem 4.1. *Suppose you have a homogeneous group of n policyholders with known durations* $w_1, w_2, ..., w_n$, *then*

$$\mathbb{E}[M] = \mu.$$

Proof. From Lemma 4.3 and the properties of the expectation, we have that

$$\mathbb{E}[M] = \mathbb{E}\left[\frac{\sum_{i=1}^n S_{w_i}^i}{\sum_{i=1}^n w_i}\right] = \frac{\sum_{i=1}^n E\left[S_{w_i}^i\right]}{\sum_{i=1}^n w_i} = \frac{\sum_{i=1}^n w_i\mu}{\sum_{i=1}^n w_i} = \frac{\mu\left(\sum_{i=1}^n w_i\right)}{\sum_{i=1}^n w_i} = \mu.$$

□

This theorem is very important, as it solves the second complication. Now we shall concentrate on the estimation of the mean of the empirical pure premium, for which we do not need to worry about the different durations of the policyholders in the group. This theorem also allows an important data simplification, as now we only need to keep the total claim size and duration of the group instead of keeping a record of the individual values.

4.4 The Tweedie as a Total Claim Size Distribution

The purpose of this section is to analyze some distributional assumptions for S^i, N^i and X_j^i.

We start with the discrete distribution for the number of claims N^i. When at most one claim can occur, like in life insurance the support of its distribution should be $\{0, 1\}$ and the Bernoulli distribution can be used. When several claims are possible, the support should be $\mathbb{N}\bigcup\{0\}$ for which the Poisson or the negative binomial distributions can be used.

The claim size distribution X_j^i, should be a continuous distribution with support on $(0, \infty)$. Commonly used distributions for it are the gamma, the log-normal and the inverse Gaussian. Alternatively, any Tweedie distribution in Chapter 3 with $p \geq 2$ can be used.

As a consequence of the previous two paragraphs, we have that the distribution of S^i should be continuous on $(0, \infty)$ with a mass at zero that corresponds to the probability of not having a claim.

In this thesis we analyze cases in which N^i and X_j^i can be modeled with the Poisson and gamma distributions respectively. This implies that S^i is Tweedie distributed with $p \in (1, 2)$, i.e., compound Poisson.

In the previous section, it was shown that $\mathbb{E}[M] = \mu$. When the Tweedie distributions is assumed for each S^i it turns out that M is also Tweedie distributed (the proof of this can be found in the Appendix). Thus, not only M has the same mean, but it also belongs to the same family of distributions.

It is also important to mention the limitation of these distributions. The Poisson distribution has variance equal to its mean. Thus it should be used when there is a strong evidence of over-dispersion in the number of claims. The gamma distribution is light tailed, which is not useful to model large claim payments.

4.5 Estimation of the Pure Premium with GLMs

GLMs provide a practical methodology for the segmentation and the estimation of the pure premium for policyholders. In this context, the total claim size acts as the response variable, the segmenting characteristics of the population as the covariates and the duration as the weight. Some characteristics of these elements are analyzed in the following paragraphs.

GLMs allow categorical and continuous covariates. In an insurance context their role is to divide the population into sufficiently large homogeneous groups. This is impossible when a continuous covariate is used. Thus, the standard practice is to divide continuous variables into intervals so they can be considered as categorical.

As in classical regression models, for a covariate with n possible levels, $n-1$ dummy variables are introduced in the model.

Here we consider the case where the response distribution is Tweedie with $p \in (1,2)$, i.e. we assume that the total claim size of each policyholder is Tweedie distributed. In the previous section it was mentioned that when this is the case then so is M, i.e the empirical pure premium is Tweedie distributed. This is useful when working with grouped data.

Another element needed for GLMs is the link function. Ideally, the link function should be chosen with the objective of linearizing the data. Nevertheless, it is a standard practice in pure premium modeling to use the natural logarithm as the link function. This is because it yields a multiplicative rating structure, which is considered more fair. An argument about this can be found in Sections 2.1 and 2.3 of Brockman and Wright (1992).

To fit the model, we need a sample (historical data), that either comes from the databases of the insurance company, from publicly available data, or from information that is bought from private companies or associations.

In using GLMs, the insurability and key ratio assumptions, as well as the distributional assumptions from Chapter 2 are implicitly assumed. The use of GLMs is not exclusive to the pure premium, they can also be used to estimate all the key ratios. In this case, the response distribution and the weight are the key ratio and its exposure, respectively. In fact, it is not common for insurers to model the pure premium directly.

When the total claim size distribution is assumed to be compound Poisson-gamma, the standard practice is to multiply separate estimations for the claim frequency and claim severity. These estimations are obtained with GLMs assuming Poisson and gamma response distributions, respectively. We will refer to this method of estimation as the separated Poisson-gamma approach (SPGA). Let us analyze the differences between this method and a Tweedie GLM.

Suppose that you have divided your portfolio of policyholders into n homogeneous groups. When using the SPGA, the distribution of each group is CPG, or equivalently Tweedie with $p \in (1, 2)$. Thus, when using the SPGA, the distribution of the individual total claim size of each policyholder in the i-th group is $Tw(p', \mu_i', \lambda_i')$ for some $p', \mu_1, ..., \mu_n, \lambda_1, ..., \lambda_n$. The fact that all the groups have the same variance function power is proved in the Appendix. On the other hand, with the Tweedie GLM, each group has distribution $Tw(p, \mu_i, \lambda)$ for some $p, \mu_1, ..., \mu_n, \lambda$. Thus, in both cases the individual claim size of each group is Tweedie distributed, with the important difference that with the Tweedie GLM, all the groups have the same dispersion parameter, while with the SPGA, it differs for each group. In order to have a more intuitive understanding of this, let us see it in terms of the variance.

With the Tweedie GLM, the i-th group has variance $\lambda \mu_i^p$ while with the SPGA it is $\lambda_i' \mu_i'^{p'}$. Thus, in both models we get different variances between groups, but the SPGA has more potential variability for the variances due to the different λ_i's. Nevertheless there is a price to pay for this extra variability: a larger number of parameters.

Suppose that you use q covariates to segment your population. For the SPGA there are $2q+1$ parameters, q for the β_is of the Poisson GLM plus q for the β_is of the gamma, plus 1 for the dispersion parameter of the gamma. On the other hand, with the Tweedie GLM there are $q+2$ parameters. The q β_is, the dispersion parameter and p. Thus, the SPGA has $q-1$ more parameters than the Tweedie GLM.

From this analysis we see that the Tweedie GLM and the SPGA are appropriate for different situations. In those cases where one common dispersion parameter is sufficient to explain the variances for the different groups, the Tweedie GLM is more appropriate. For such cases a SPGA would be overparametrizing. On the other hand, a SPGA is more adequate in cases where several dispersion parameters are needed to explain the variances among the different homogeneous groups.

There are differences worth mentioning about the model fit assessment for each option. When using a Tweedie GLM, the goodness of fit and the distributional assumptions of the model can be analyzed with the theory of GLMs. On the other hand, when using the SPGA, the theory of GLMs can be used separately for testing the Poisson and the gamma models, but not for the final estimation of the pure premium. In this case, in order to asses the goodness of fit of the pure premium estimation, different techniques have to be used.

4.6 Model Assessment

As explained in the previous section, when separate GLMs for the claim frequency and claim severity are used, the theory of GLMs is not that useful in assessing the goodness of fit of the combined model. In recent years, data mining techniques have gained popularity in the insurance modeling world. For instance, the lift and gain charts are now routinely used for model selection and assessment.

4.6.1 The Lift Chart

Suppose that you have a model created to predict a certain phenomenon and suppose that you have some observations of this phenomenon. In order to create a lift chart based on this, the following steps should be followed:

1. By using the model generate predictions for the observations.
2. Order the observations increasingly with respect to the predictions.
3. Divide the ordered data in groups that have equal number of predictions.
4. Plot the mean of the observations and the mean of the predictions for each group.

When this chart is made for a GLM, it is also common to add bars that correspond the exposure of each group. An example of a lift curve with 20 groups is shown in Figure 4.1. In this graph, the scale of the vertical axis on the left corresponds to the mean computed for each group and the scale of the vertical axis on the right corresponds to the exposure.

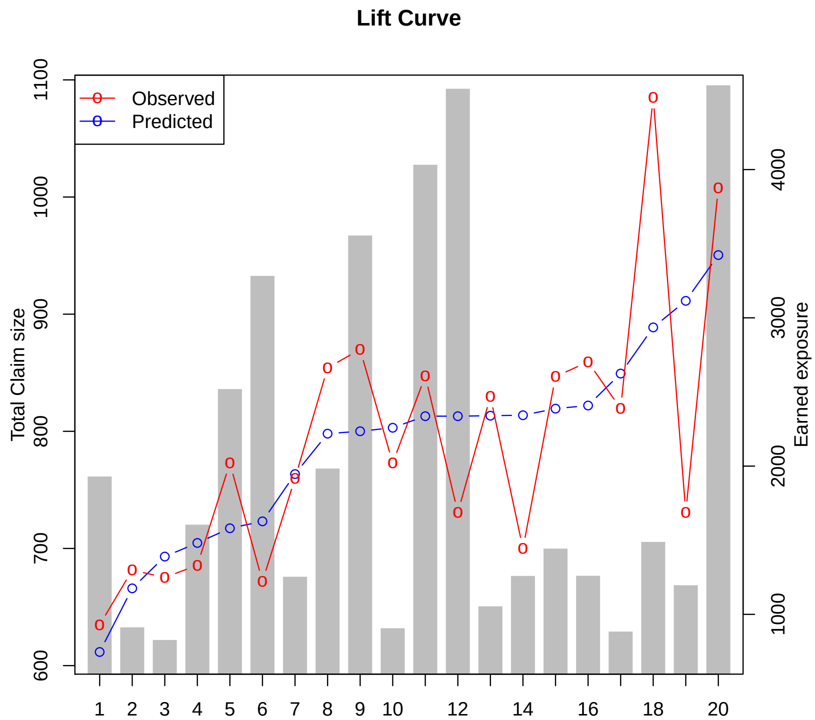


Figure 4.1: Example of Lift Chart

This kind of chart gives information about two aspects of the model. On the one hand, by seeing the trend on the curve for the observed means it is possible to see if the model more or less identifies the groups that are more costly. On the other hand, the vertical distance between the predicted mean and the observed mean gives an idea of how far the predictions are from the observations.

4.6.2 The Gain Chart

As with the previous chart, suppose that you have a model with some observations. In order to create the gains chart, the following steps should be followed:

1. By using the model, generate predictions for the observations.

2. Order the observations decreasingly with respect to the predictions, and compute their cumulative percentages.

3. Let n be the number of observations. Draw the lines that connect the points $\{(\frac{1}{i}, x_i)\}_{i=1}^{n}$, where x_i is the i-th cumulative percentage.

4. Draw two reference lines. The identity line and the upper bound. The upper bound graph is obtained by following steps 2 and 3, but ordering the observations decreasingly (this time not with respect to the model).

An example of a gain chart is shown in Figure 4.2. This graph gives a visual representation of how well the model distinguishes the groups that are more expensive than others. The closer the model line to the upper bound the better.

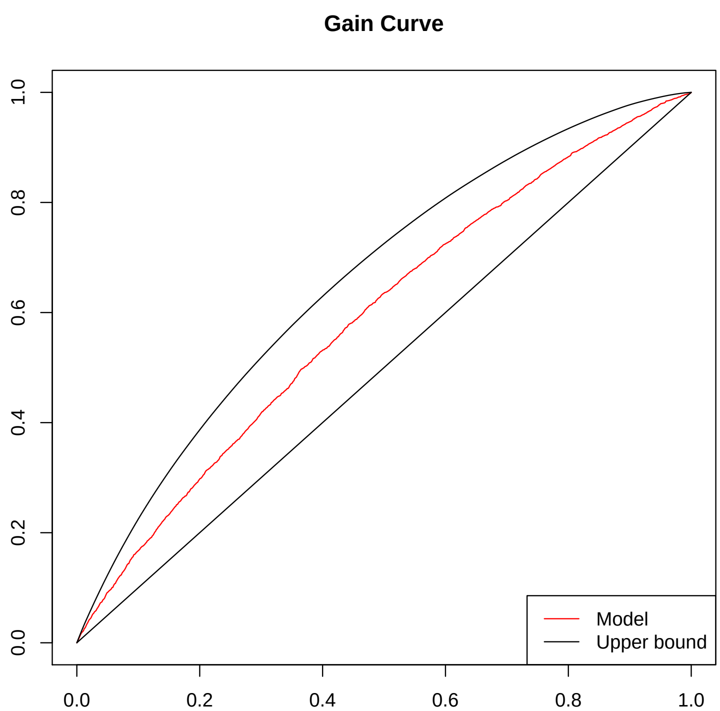


Figure 4.2: Example of Gain Chart

In the following two sections, two examples are presented. The goal of these examples is to show comparisons between the direct analysis of the pure premium with the Tweedie family and the separate Poisson-gamma approach. This is done by using the graphs introduced in this section.

4.7 Example: Third Party Motor Insurance in Sweden

In Section 4.5 it was mentioned that a log-link function yields a multiplicative rating structure. Before presenting our first example we explain the meaning of this. Suppose that you are modeling a key ratio with a GLM for which you only use categorical covariates. Assume that the classes of all your covariates are represented in your model by n dummy variables, say $X_1, ..., X_n$. Let β_i be the coefficient of X_i for $i = 1, ..., n$ and β_0 be the intercept of the model. Then the predicted mean $\hat{\mu}$ of a given policyholder satisfies the equation

$$\ln(\hat{\mu}) = \beta_0 + \sum_{\{i:X_i=1\}} \beta_i,$$

and therefore

$$\hat{\mu} = \exp\left(\beta_0 + \sum_{\{i:X_i=1\}} \beta_i\right) = \exp\left(\beta_0\right) \prod_{\{i:X_i=1\}} \exp(\beta_i).$$

Now, define $R_i = \exp(\beta_i)$ for $i = 0, ..., n$. Then

$$\hat{\mu} = R_0 \prod_{\{i:X_i=1\}} R_i.$$

The R_i's are called the relativities of the model. From the equation above we see that for each case, the predicted mean is obtained by multiplying the appropriate relativities. It is because of this that the model is said to have a multiplicative rating structure.

Kilometres	Zone	Bonus	Make	Insured	Claims	Payment
1	1	1	1	455.13	108	392491
1	1	1	2	69.17	19	46221
1	1	1	3	72.88	13	15694
1	1	1	4	1292.39	124	422201
1	1	1	5	191.01	40	119373
1	1	1	6	477.66	57	170913
1	1	1	7	105.58	23	56940
1	1	1	8	32.55	14	77487
1	1	1	9	9998.46	1704	6805992
1	1	2	1	314.58	45	214011

Table 4.1: Some Observations from the Third Party Motor Insurance in Sweden Dataset

In this section a Tweedie response GLM is fitted to a publicly available motor insurance dataset. All the computations shown in this section were done using R with the package tweedie (see R Development Core Team, 2010 and Dunn, 2010). The data was taken from Smyth (2011) at the address `http://www.statsci.org/data/general/motorins.html`. Some information from this dataset is summarized in Table 4.1.

There are 4 segmenting variables available in this dataset. *Kilometers*, with 5 categories, corresponds to different intervals of kilometers traveled per year. *Zone*, with 7 categories, corresponds to different geographical zones. The classes 1 to 8 of *Make* correspond to different car models, while the 9-th category corresponds to any other make of car. Finally, *Bonus*, with 7 classes, corresponds to the number of years

Kilometers	Zone	Bonus	Make	Observed total claim size	Observed claim frequency	Observed claim severity
1	1	1	1	862.371	0.237	3634.176
1	1	1	2	668.223	0.275	2432.684
1	1	1	3	215.340	0.178	1207.231
1	1	1	4	326.682	0.096	3404.847
1	1	1	5	624.957	0.209	2984.325
1	1	1	6	357.813	0.119	2998.474
1	1	1	7	539.307	0.218	2475.652
1	1	1	8	2380.553	0.430	5534.786
1	1	1	9	680.704	0.170	3994.127
1	1	2	1	680.307	0.143	4755.800

Table 4.2: Observed Key Ratios for some Homogeneous Groups

plus one since the last claim. In this example it is assumed that each combination of values of these variables defines a homogeneous group.

Variable	**Poisson coefficients**	**Gamma coefficients**	**Combined Poisson-Gamma coefficients**	**Tweedie coefficients**	**Combined Poisson-Gamma relativities**	**Tweedie relativities**
Intercept	-1.813	8.397	6.585	6.565	723.811	709.781
Make 1	0.000	0.000	0.000	0.000	1.000	1.000
2	0.076	-0.038	0.038	0.034	1.039	1.035
3	-0.247	0.081	-0.166	-0.173	0.847	0.841
4	-0.654	-0.166	-0.820	-0.807	0.441	0.446
5	0.155	-0.089	0.066	0.053	1.068	1.055
6	-0.336	-0.039	-0.375	-0.354	0.687	0.702
7	-0.056	-0.120	-0.176	-0.148	0.839	0.862
8	-0.044	0.212	0.168	0.165	1.183	1.179
9	-0.068	-0.056	-0.124	-0.113	0.883	0.893
Bonus 1	0.000	0.000	0.000	0.000	1.000	1.000
2	-0.479	0.044	-0.435	-0.435	0.647	0.647
3	-0.693	0.070	-0.623	-0.625	0.536	0.535
4	-0.827	0.055	-0.773	-0.771	0.462	0.462
5	-0.926	0.036	-0.889	-0.882	0.411	0.414
6	-0.993	0.072	-0.922	-0.917	0.398	0.400
7	-1.327	0.116	-1.211	-1.203	0.298	0.300
Zone 1	0.000	0.000	0.000	0.000	1.000	1.000
2	-0.238	0.022	-0.216	-0.207	0.806	0.813
3	-0.386	0.048	-0.339	-0.325	0.713	0.722
4	-0.582	0.128	-0.454	-0.442	0.635	0.643
5	-0.326	0.052	-0.274	-0.258	0.760	0.773
6	-0.526	0.143	-0.383	-0.360	0.682	0.698
7	-0.731	0.022	-0.709	-0.670	0.492	0.511
Kilometres 1	0.000	0.000	0.000	0.000	1.000	1.000
2	0.213	0.023	0.235	0.219	1.265	1.244
3	0.320	0.019	0.339	0.337	1.404	1.400
4	0.405	0.040	0.445	0.456	1.561	1.577
5	0.576	0.037	0.613	0.612	1.846	1.844

Table 4.3: Coefficients and Relativities for the Different Models

The observations in this dataset are grouped. For each homogeneous group, the columns Insured, Claim and Payment correspond to the observed duration, number of claims and total claim size of the group respectively. In order to fit GLMs for the pure premium, claim frequency and claim severity, the observed values of these quantities for each group have to be computed. Some of these computations are shown in Table 4.2.

A GLM with a Tweedie response and logarithmic link function was applied to the whole dataset to estimate the pure premium. The MLE found for p was 1.471429. Similarly, GLMs with Poisson and gamma responses, both with the log-link function,

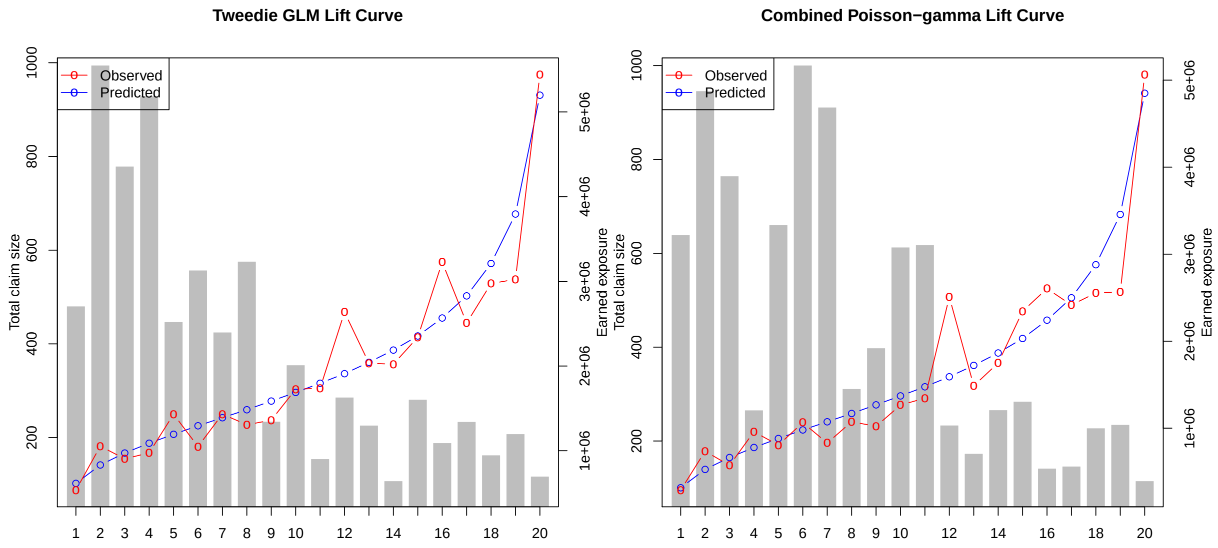


Figure 4.3: Lift Charts

were fitted to the claim frequency and claim severity respectively. The fitted values obtained for the coefficients of the covariates are summarized in Table 4.3.

The columns Poisson coefficients, gamma coefficients and Tweedie coefficients correspond to the estimation of the betas for each of the classes in each model. The

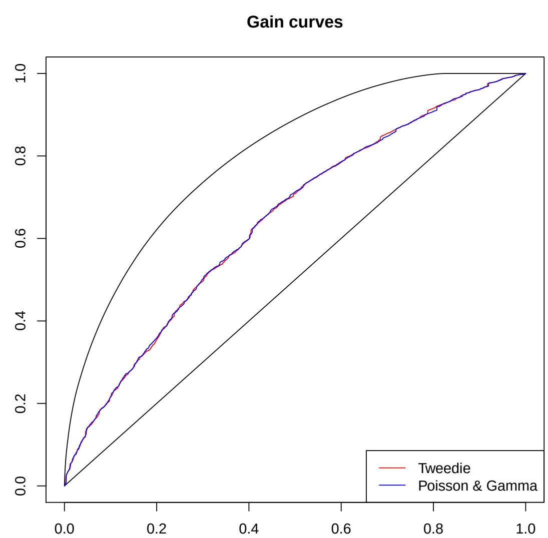


Figure 4.4: Gain Curve for Tweedie GLM and Combined Poisson-gamma

combined Poisson-gamma coefficient is equal to the sum of the Poisson and gamma coefficients. There is a justification for this. On the one hand, the estimation of the pure premium based on the separate Poisson and gamma models, is obtained by multiplying the estimations of these two. On the other hand there is the fact that the log-link function is being used. This implies that the contribution of each class to the log of the pure premium is equal to the sum of the coefficients from each model.

The lift charts with 20 groups, for both methods of estimation compared to the observations are given in Figure 4.3. From the exposure bars in these charts, we can see that the models have ordered the groups differently. Nevertheless, in both cases the observed means curve is increasing and the predicted means curve is not far from it. Figure 4.4 shows the gain charts for both models in the same graph. In this chart we see that the gain curves are almost superimposed. Thus, the lift and gain charts indicate that the two models fit the data more or less equally well. Now, in order to see how different are the predictions between both models, Figure 4.5 shows the PP-plot and the QQ-plot of the models. From these graphs we see that these two models produce very similar predictions. Thus, in this example modeling the pure premium

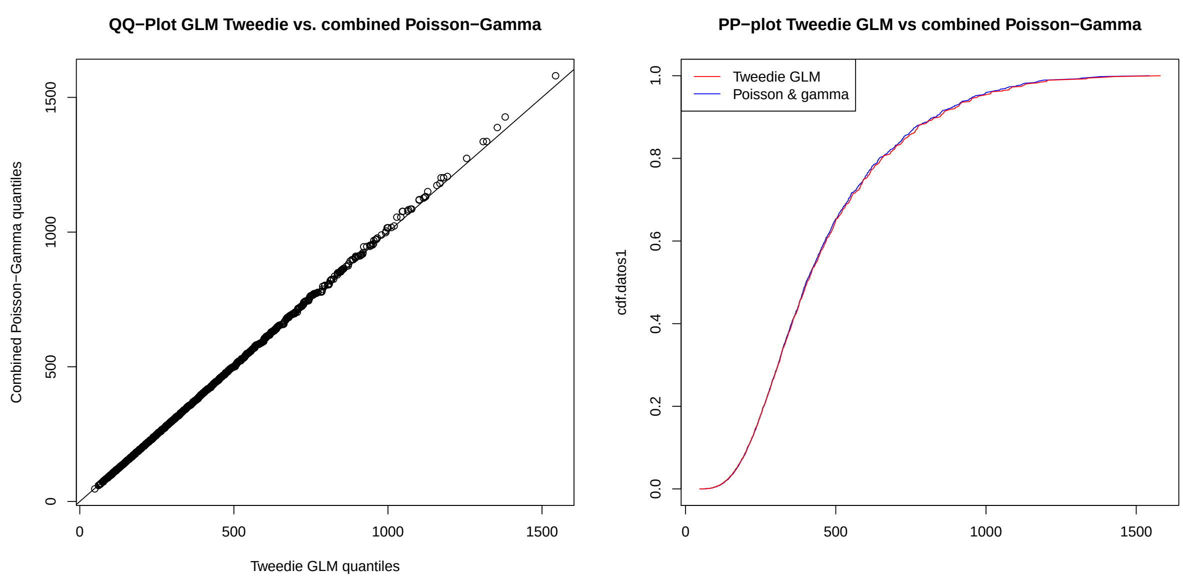


Figure 4.5: Comparison Between the Tweedie GLM and SPGA Predictions

directly or doing it through the claim frequency and claim severity separately gives very similar results.

4.8 Example: Car Insurance in Toronto

The results in this section were obtained from a 4 months internship at AVIVA Canada. AVIVA has allowed to include some graphs as long as they do not give detailed information about the variables used in the model.

The purpose of the project was to build confidence intervals around pure premium estimators for certain car insurance covers. These coverages are Accident Benefits (AB), Bodily Injury (BI), Direct Compensation (DC) and Collision (COL).

The model being used to estimate the pure premium consists of a tree structure with a GLM for each node. In effect, this tree is equivalent to having separate Poisson gamma estimators for the claim frequency and severity, respectively, for each cover. In order to be able to use the theory of GLMs to build confidence intervals, it is necessary to have all the covers under one single GLM, that would include claim frequency and severity for all the covers. The Tweedie distribution was selected as the response distribution. The purpose of this section is to compare the results of the Tweedie model against the tree structure.

First we describe the tree model. We cannot show the actual tree used at AVIVA, but we explain its functioning with a simplified example. Then we will show how this is equivalent to separate Poisson gamma analysis for each different cover.

To simplify assume that the insurance only has 2 covers, say Cover 1 and Cover 2. A possible estimation of the pure premium for these two covers is to find the pure premium for each independently, and then to sum them. The problem with this procedure, is that it assumes that when they both occur, their respective claim sizes behave similarly to when they occur separately. This is not always the case and therefore sometimes it is necessary to model their joint behavior. A way to include all cases in the model is to consider the following tree:

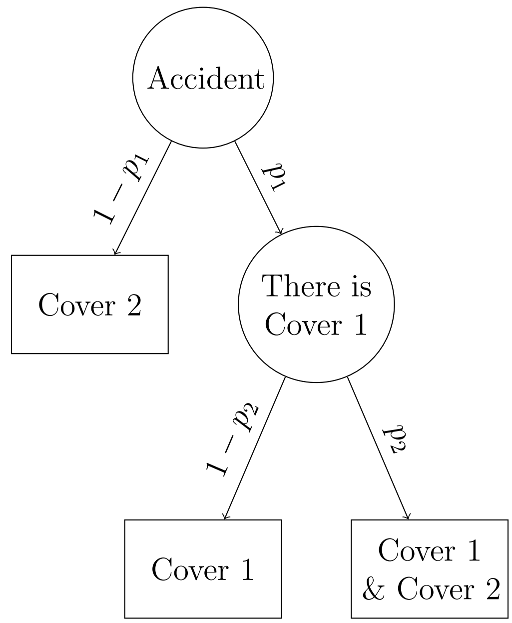


Figure 4.6: Tree for 2 Covers

The node on top, labeled Accident, represents the event of a claim. Then p_1 represents the probability of having a claim resulting from Cover 1. Similarly, p_2 represents the probability of a claim from Cover 2, given that Cover 1 is also part of the claim. The node labeled Cover 1 represents the size of a claim related to Cover 1 only. Similarly, the node labeled Cover 2 represents the size of a claim in which only Cover 2 applies. Finally, the node labeled Cover 1 & Cover 2 represents the size of a claim in which both covers apply.

Let us assume now that the number of accidents is Poisson distributed with parameter λ and that the claim sizes of Cover 1, Cover 2 and Cover 1 & Cover 2 are gamma distributed with parameters (α_1, β_1), (α_2, β_2) and $(\alpha_{12}, \beta_{12})$ respectively. Then, from the properties of the compound Poisson distribution (see Section 7 from Chapter 1 of Gerber, 1979), the total claim size for Cover 1 only, Cover 2 only and Cover 1 & Cover 2 are all CPG distributed. Table 4.4 shows the parameters of the distribution for each of these cases. Hence, the tree in Figure 4.6 combined with the above assumptions is equivalent to separate Poisson-gamma models for both, the

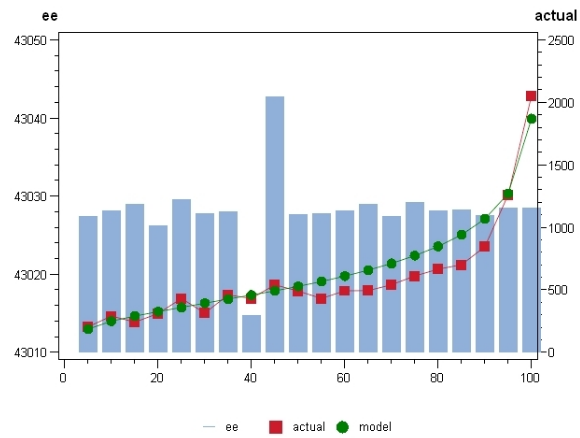
Table 4.4: Parameters for the Different Combinations of Covers

Coverage	**Parameters of CPG**
Cover 1 only	$\lambda p_1(1-p_2), \alpha_1, \beta_1$
Cover 2 only	$\lambda(1-p_1), \alpha_2, \beta_2$
Cover 1 & Cover 2	$\lambda p_1 p_2, \alpha_{12}, \beta_{12}$

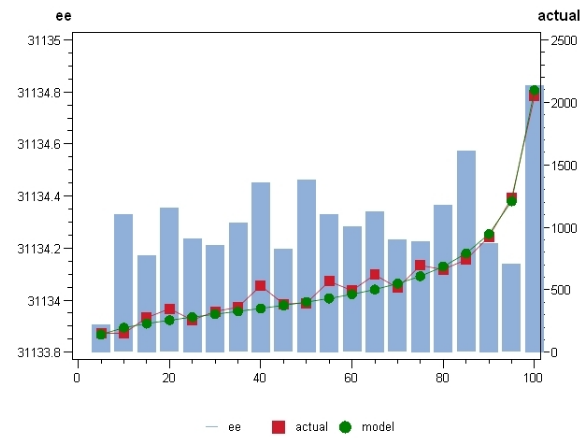
claim frequency and claim severity.

The tree at AVIVA has several more covers, but the distributional assumptions are similar, i.e. Poisson for the claim frequency and different gamma distributions for the nodes that correspond to claim severities. Therefore, the total claim size for each combination of covers corresponds to a CPG distribution. GLMs are used to estimate the parameters of the distribution at each node. A pure premium for each combination is estimated by multiplying the relevant parameters from the tree. The overall pure premium is estimated by summing the pure premium of each combination.

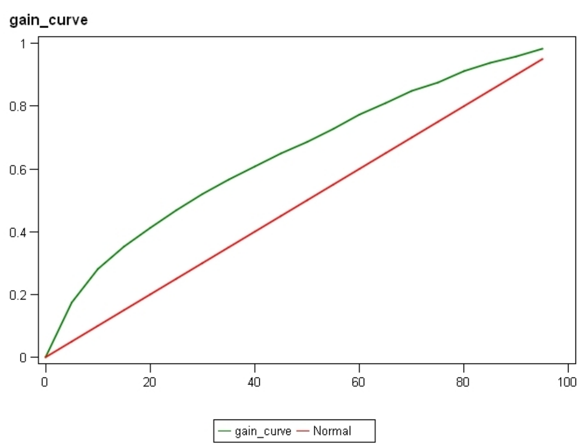
It should be pointed out that this example differs from the previous one. Here, the Tweedie is being compared with the sum of several Poisson-gamma combinations, which has much more parameters. Both models were fitted using data collected from 2006 to 2008. Their respective lift and gain charts applied to the data from 2009 are shown in Figure 4.7. These charts show a slightly better fit for the tree as the Tweedie overestimates the pure premium for groups 55 to 85. A possible way to improve the Tweedie fit in this case would be to make separate analyses for those groups with observed pure premium above 500. Nevertheless, due to the limited time from the internship this is the version last worked on.



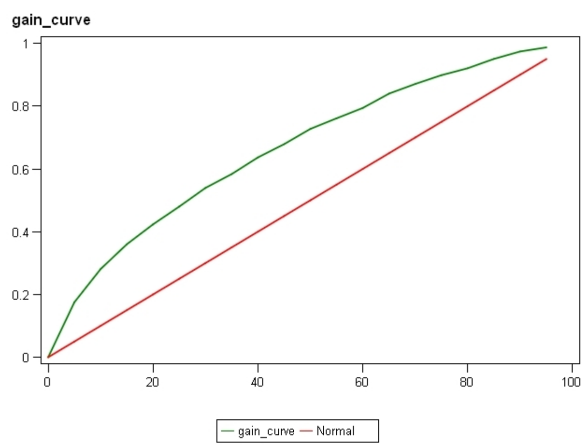
(a) Tweedie Lift chart



(b) Tree Lift chart



(c) Tweedie Gain chart



(d) Tree Gain chart

Figure 4.7: Lift Charts for the AVIVA Data

Conclusion

The compound Poisson-gamma distribution is often used to model the total claim size of individual policyholders. When this is the case, GLMs with a Tweedie response provide a convenient method for the segmentation and pure premium estimation of a property and casualty insurance portfolio. Nevertheless, the standard practice in the industry is to combine estimations obtained separately for the claim frequency and claim severity by using GLMs with Poisson and gamma responses, respectively. Both approaches imply a Tweedie distribution for the total claim size of the different homogeneous groups. The main difference between these two methods is that in the Tweedie GLM all the groups have the same dispersion parameter, while the standard method assigns different values to each group. This implies that, in a given situation, one method will fit better than the other, depending if one common dispersion parameter is sufficient to fit the variance of the different homogeneous groups or not.

From the point of view of model assessment, the unified approach has the advantage that its goodness of fit and distributional assumptions can be tested using the theory of GLMs. On the other hand, the standard approach requires different techniques for its assessment. In conclusion, the Tweedie GLM is a good competitor to the standard estimation method used in the industry. It should always be considered in the set of models to evaluate when the total claim size is believed to follow a compound Poisson-gamma distribution.

Bibliography

Björk, A. (1996). *Numerical Methods for Least Squares Problems.* SIAM.

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

Dunn, P. K. (2010). *tweedie: Tweedie exponential family models.* R package version 2.0.7.

Gerber, H. U. (1979). *An Introduction to Mathematical Risk Theory.* S.S. Huebner Foundation for Insurance Education.

Gilchrist, R. and D. Drinkwater (2000). Fitting tweedie models to data with probability of zero responses. *Proceedings of the 14th International Workshop on Statistical Modelling*, 207–214.

Jørgensen, B. (1986). Some properties of exponential dispersion models. *Scandinavian Journal of Statistics 13*(3), 187–197.

Jørgensen, B. (1997). *The Theory of Dispersion Models* (First ed.). Chapman & Hall.

Klugman, S.A., P. H. and G. Willmot (2004). *Loss Models: From Data to Decisions* (second ed.). Wiley.

Madsen, H. and P. Thyregod (2011). *Introduction to General and Generalized Linear Models.* CRC Press.

Ohlsson, E. and B. Johansson (2010). *Non-Life Insurance Princing with Generalized Linear Models*. Springer-Verlag Berlin Heidelberg.

R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0. `http://www.R-project.org`.

Smyth, G. (2011). *Australasian Data and Story Library (OzDASL)*. `http://www.statsci.org/data`.

Appendix A

Proof of Propositions from Chapter 4

In Chapter 4, Lemmas 4.1, 4.2 and 4.3 are for every i. Since proving it for one specific value of i implies that it is true for all the others, the i will not be written in the following proofs. Before proving the above mentioned lemmas, a first auxiliary result is proved.

Lemma A.1. *Let w_1 and w_2 be two durations with $w_1 \leq w_2$, then*

$$\mathbb{E}[N_{w_1}] \leq \mathbb{E}[N_{w_2}].$$

Proof. Let $w = w_2 - w_1$ and consider two non-overlapping time intervals of length w_1 and w. Then,

$$N_{w_1} + N_w \stackrel{D}{=} N_{w_2}$$

and therefore

$$\mathbb{E}[N_{w_1}] + \mathbb{E}[N_w] = \mathbb{E}[N_{w_2}].$$

Now, as N_w is a non-negative random variable we have $\mathbb{E}[N_{w_1}] \geq 0$. Thus

$$\mathbb{E}[N_{w_1}] \leq \mathbb{E}[N_{w_2}].$$

□

Lemma A.2. *For every $w > 0$*

$$\mathbb{E}[N_w] = w\mathbb{E}[N].$$

Proof. The proof is divided in three parts. First it is proved for $w = \frac{1}{n}$, where $n \in \mathbb{N}$, then for $w \in \mathbb{Q}\bigcap(0,\infty)$ and finally for $w \in (0,\infty)$.

Let $n \in \mathbb{N}$, and $w_1, ..., w_n$ be durations that correspond to non-overlapping time intervals with $w_i = \frac{1}{n}$ for $i = 1, ..., m$. Then

$$\sum_{i=1}^{n} N_{w_i} \stackrel{D}{=} N \qquad \text{and} \qquad \mathbb{E}[N_{w_i}] = \mathbb{E}[N_{\frac{1}{n}}] \qquad i = 1, ..., n,$$

which implies

$$\mathbb{E}[N] = \mathbb{E}\left[\sum_{i=1}^{n} N_{w_i}\right] = \sum_{i=1}^{n} \mathbb{E}[N_{w_i}] = \sum_{i=1}^{n} \mathbb{E}\left[N_{\frac{1}{n}}\right] = n\mathbb{E}\left[N_{\frac{1}{n}}\right]$$

and therefore,

$$\frac{1}{n}\mathbb{E}[N] = \mathbb{E}\left[N_{\frac{1}{n}}\right].$$

This proves the lemma for $w = \frac{1}{n}$. Now, let $m, n \in \mathbb{N}$ and $w_1, ..., w_m$ be durations that correspond to non-overlapping periods of time with $w_i = \frac{1}{n}$ for $i = 1, ..., m$. Then,

$$\sum_{i=1}^{m} N_{w_i} \stackrel{D}{=} N_{\frac{m}{n}} \qquad \text{and} \qquad \mathbb{E}[N_{w_i}] = \mathbb{E}\left[N_{\frac{1}{n}}\right], \qquad i = 1, ..., m,$$

and therefore,

$$\mathbb{E}\left[N_{\frac{m}{n}}\right] = \sum_{i=1}^{m} \mathbb{E}\left[N_{w_i}\right] = \sum_{i=1}^{m} \mathbb{E}\left[N_{\frac{1}{n}}\right] = m\mathbb{E}\left[N_{\frac{1}{n}}\right] = \frac{m}{n}\mathbb{E}[N].$$

Let $w \in (0,\infty)$, $\{p_n\}$ and $\{q_n\}$ be sequences in $\mathbb{Q}\bigcap(0,\infty)$ with $p_n \uparrow w$ and $q_n \downarrow w$.

Then from Lemma A.1, we have that for every $n \in \mathbb{N}$

$$p_n\mathbb{E}[N] = \mathbb{E}\left[N_{p_n}\right] \leq \mathbb{E}\left[N_w\right] \leq \mathbb{E}\left[N_{q_n}\right] = q_n\mathbb{E}[N],$$

then

$$\lim_{n\to\infty} p_n\mathbb{E}[N] \leq \mathbb{E}[N_w] \leq \lim_{n\to\infty} q_n\mathbb{E}[N],$$

which implies,

$$\mathbb{E}[N_w] = w\mathbb{E}[N].$$

This proves the theorem for $w \in (0, \infty)$. □

Lemma A.3. *Let* $\mu = \mathbb{E}[S]$. *Then,*

$$\mu = \mathbb{E}[N]\mathbb{E}[X_1]$$

and for every $w > 0$*,*

$$\mathbb{E}[S_w] = w\mu.$$

Proof. Let $w > 0$. As the X_i's are iid and independent of N_w, we have that

$$\begin{aligned}\mathbb{E}[S_w] = \mathbb{E}\left[\sum_{i=1}^{N_w} X_i\right] &= \mathbb{E}\left[\mathbb{E}\left[\sum_{i=1}^{N_w} X_i | N_w\right]\right] \\ &= \mathbb{E}\big[N_w\mathbb{E}\left[X_1\right]\big] \\ &= \mathbb{E}[N_w]\mathbb{E}[X_1] = w\mathbb{E}[N]\mathbb{E}[X_1].\end{aligned}$$

which implies $\mu = \mathbb{E}[S] = \mathbb{E}[N]\mathbb{E}[X_1]$ and $\mathbb{E}[S_w] = w\mu$. □

In Chapter 4, in the example shown for the Swedish motor insurance, we worked with grouped data. Originally we assumed that the total claim size of each policyholder follows a Tweedie distribution. Nevertheless, in building the model the response variable was always the empirical total claim size of each policyholder, and as such we assumed it to also follow a Tweedie distribution. This needs to be justified, which is the purpose of the following theorem. Two lemmas are first needed for this.

The proof of the following lemma can be found in Chapter 1 of Gerber (1979).

Lemma A.4. *Let S_1 and S_2 be two independent compound Poisson distributions with rates λ_1 and λ_2 respectively. Assume that F_1 is the jump size distribution function of S_1 and F_2 the one of S_2. Then, the random variable $S = S_1 + S_2$ has a compound Poisson distribution with jump size distribution function*

$$F(x) = \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right) F_1(x) + \left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right) F_2(x), \qquad x \in \mathbb{R}.$$

Lemma A.5. *Let X be a random variable with distribution $CPG(m, \alpha, \beta)$, $k > 0$ and $Y = kX$. Then $Y \sim CPG(m, \alpha, \frac{\beta}{k})$.*

Proof. Let M_X and M_Y be the mgfs of X and Y respectively. From the definition of mgf and Theorem 3.3, we have that

$$M_Y(t) = \mathbb{E}[\exp(tY)] = \mathbb{E}[\exp(tkX)] = M_X(tk) = \exp\left(m\left\{\left(\frac{1}{1 - \frac{t}{\beta/k}}\right)^{\alpha} - 1\right\}\right),$$

which corresponds to the mgf of a $CPG(m, \alpha, \frac{\beta}{k})$. □

Theorem A.1. *Suppose you have a homogeneous group with n policyholders for which the total claim size distribution, assuming a duration of 1, is $Tw(p, \mu, \lambda)$. Suppose that the durations $w_1, ..., w_n$ for each policyholder are known. Then M, the empirical total claim size of the group, has distribution $Tw(p, \mu, w_+\lambda)$, where $w_+ = \sum_{i=1}^n w_i$.*

Proof. Let $S^i_{w_i}$ represent the total claim size of the i-th policyholder taking into consideration his/her distribution. From the hypothesis of the theorem we have that $S^i \sim Tw(p, \mu, \lambda)$. By Theorem 3.4 we can equivalently say that $S^i \sim CPG(m, \alpha, \beta)$, where $m = \frac{\lambda\mu^{2-p}}{2-p}$, $\alpha = -\frac{p-2}{p-1}$ and $\beta = \frac{\lambda\mu^{1-p}}{p-1}$. We have then that N^i, the claim frequency for a duration of 1, follows a Poisson distribution with parameter m, and therefore $N_{w_i}^i \sim Poisson(w_i m)$. Thus $S_{w_i}^i \sim CPG(w_i m, \alpha, \beta)$. Define $S = \sum_{i=1}^n S_{w_i}$, $w_+ = \sum_{i=1}^n w_i$ and let G be the distribution function of a gamma(α, β). Then, as a consequence of Lemma A.4, S has a compound Poisson distribution with rate w_+ and distribution function

$$F(x) = \frac{\sum_{i=1}^n w_i G(x)}{w_+} = \frac{G(x) \sum_{i=1}^n w_i}{w_+} = G(x),$$

which implies that F is a distribution function of a gamma(α, β) and therefore $S \sim CPG(w_+ m, \alpha, \beta)$. By definition, $M = \frac{S}{w_+}$ and then by Lemma A.5, $M \sim CPF(w_+ m, \alpha, w_+ \beta)$. Now, by writing m, α and β in terms of p, μ and λ and using Theorem 3.4 we obtain that equivalently $M \sim Tw(p, \mu, w_+ \lambda)$. □

The purpose of the following theorem is to derive the Tweedie parameters for the different homogeneous groups when the SPGA is used.

Theorem A.2. *Suppose you have divided a portfolio of policyholders into homogeneous groups by using q explicative variables. Let $\boldsymbol{x_i} \in \mathbb{R}^q$ be the values of the explicative variables for the i-th group. Let $\boldsymbol{\beta_p}, \boldsymbol{\beta_g} \in \mathbb{R}^q$ and $\lambda > 0$, and define for each i, $m_i = \exp(\boldsymbol{x_i}^T \boldsymbol{\beta_p})$ and $\mu_i = \exp(\boldsymbol{x_i}^T \boldsymbol{\beta_g})$. If for the i-th group, the distribution of the number of claims is $Poisson(m_i)$ and the distribution of the claim size is gamma with mean μ_i and dispersion parameter λ, then*

$$S_i \sim Tw\left(\frac{\lambda+2}{\lambda+1}, m_i \mu_i, \left(\frac{\lambda}{\lambda+1}\right) \frac{m_i^{\frac{1}{\lambda+1}}}{\mu_i^{\frac{\lambda}{\lambda+1}}}\right),$$

where S_i is the total claim size of a member of the i-th group.

Proof. For each i, let X_i be a random variable with a gamma distribution of mean μ_i and dispersion parameter λ. Let us find the parameters α_i and β_i of this distribution. Under the parametrization α and β of the gamma distribution we have that $\mathbb{E}[X_i] = \frac{\alpha}{\beta}$ and $\mathbb{V}[X_i] = \frac{\alpha}{\beta^2}$. On the other hand, from the *REDF* parametrization we have that

$\mathbb{E}[X_i] = \mu_i$ and $\mathbb{V}[X_i] = \frac{\mu_i^2}{\lambda}$. Thus, we have the following equations

$$\frac{\alpha_i}{\beta_i} = \mu_i, \qquad \frac{\alpha_i}{\beta_i^2} = \frac{\mu_i^2}{\lambda}.$$

Solving for α_i and β_i we get that $X_i \sim \text{gamma}(\lambda, \frac{\lambda}{\mu_i})$. Therefore $S_i \sim CPG(m_i, \lambda, \frac{\lambda}{\mu_i})$. Then, by Theorem 3.4, we have that equivalently,

$$S_i \sim Tw\left(\frac{\lambda+2}{\lambda+1}, m_i\mu_i, \left(\frac{\lambda}{\lambda+1}\right)\frac{m_i^{\frac{1}{\lambda+1}}}{\mu_i^{\frac{\lambda}{\lambda+1}}}\right).$$

□