Chapter 2
The Basics of Pricing with GLMs

As described in the previous section, the goal of a tariff analysis is to determine how one or more key ratios Y vary with a number of rating factors. This is reminiscent of analyzing how the dependent variable Y varies with the covariates (explanatory variables) x in a multiple linear regression. Linear regression, or the slightly larger general linear model, is not fully suitable for non-life insurance pricing, though, since: (i) it assumes normally distributed random errors, while the number of insurance claims follows a discrete probability distribution on the non-negative integers, and claim costs are non-negative and often skewed to the right; (ii) in linear models, the mean is a linear function of the covariates, while multiplicative models are usually more reasonable for pricing, cf. Sect. 1.3.

Generalized linear models (GLMs) is a rich class of statistical methods, which generalizes the ordinary linear models in two directions, each of which takes care of one of the above mentioned problems:

- *Probability distribution.* Instead of assuming the normal distribution, GLMs work with a general class of distributions, which contains a number of discrete and continuous distributions as special cases, in particular the normal, Poisson and gamma distributions.
- *Model for the mean.* In linear models the mean is a linear function of the covariates x. In GLMs some monotone transformation of the mean is a linear function of the x's, with the linear and multiplicative models as special cases.

These two generalization steps are discussed in Sects. 2.1 and 2.2, respectively.

GLM theory is quite recent—the basic ideas were introduced by Nelder and Wedderburn [NW72]. Already in the first 1983 edition of the standard reference by McCullagh and Nelder there is an example using motor insurance data; in the second edition [MN89] this example can be found in Sects. 8.4.1 and 12.8.3. But it was not until the second half of the 90's that the use of GLMs really started spreading, partly in response to the extended needs for tariff analysis due to the deregulation of the insurance markets in many countries. This process was facilitated by the publication of some influential papers by British actuaries, such as [BW92, Re94, HR96]; see also [MBL00], written for the US Casualty Actuarial Society a few years later.

E. Ohlsson, B. Johansson, *Non-Life Insurance Pricing with Generalized Linear Models,* EAA Lecture Notes,
DOI 10.1007/978-3-642-10791-7_2,

Some advantages of using GLMs over earlier methods for rate making are:

- GLMs constitute a general statistical theory, which has well established techniques for estimating standard errors, constructing confidence intervals, testing, model selection and other statistical features.
- GLMs are used in many areas of statistics, so that we can draw on developments both within and without of actuarial science.
- There is standard software for fitting GLMs that can easily be used for a tariff analysis, such as the SAS, GLIM, R or GenStat software packages.

In spite of the possibility to use standard software, many insurance companies use specialized commercial software for rate making that is provided by major consulting firms.

2.1 Exponential Dispersion Models

Here we describe the *exponential dispersion models* (EDMs) of GLMs, which generalize the normal distribution used in the linear models.

In our discussion of the multiplicative model in (1.3) and (1.4) we used a way of organizing the data with the observations, $y_{i_1,i_2,\ldots,i_K}$, having one index per rating factor. This is suitable for displaying the data in a table, especially in the two-dimensional case y_{ij}, and will therefore be called *tabular form*.

In our general presentation of GLMs, we rather assume that the data are organized on *list form*, with the n observations organized as a column vector $\mathbf{y}' = (y_1, y_2, \ldots, y_n)$. Besides the key ratio y_i, each row i of the list contains the exposure weight w_i of the tariff cell, as well as the values of the rating factors. The transition from tabular form to list form amounts to deciding on an order to display the tariff cells; a simple example is given in Table 2.1. As further illustration, consider again the moped insurance example, for which Table 1.2 contains an implicit list form; in Table 2.2 we repeat the part of that table that gives the list form for analysis of claim frequency.

List form corresponds to the way we organize the data in a data base, such as a SAS table. Tabular form, on the other hand, is useful for demonstrative purposes; hence, the reader should try to get used to both forms.

By the assumptions in Sect. 1.2, the variables $Y_1, \ldots, Y_n$ are independent, as required in general GLM theory. The probability distribution of an EDM is given

Table 2.1 Transition from tabular form to list form in a 2 by 2 case

Married	Male	Female
Yes	y_{11}	y_{12}
No	y_{21}	y_{22}

$\Longrightarrow$

i	Married	Gender	Observation
1	Yes	M	y_1
2	Yes	F	y_2
3	No	M	y_3
4	No	F	y_4

Table 2.2 Moped tariff on list form (claim frequency per mille)

Tariff cell i	Covariates Class x_{i1}	Age x_{i2}	Zone x_{i3}	Duration (exposure) w_i	Claim frequency y_i
1	1	1	1	62.9	270
2	1	1	2	112.9	62
3	1	1	3	133.1	68
4	1	1	4	376.6	19
5	1	1	5	9.4	0
6	1	1	6	70.8	14
7	1	1	7	4.4	228
8	1	2	1	352.1	148
9	1	2	2	840.1	82
⋮	⋮	⋮	⋮	⋮	⋮
21	2	1	7	14.5	0
22	2	2	1	844.8	111
23	2	2	2	1 296.0	76
24	2	2	3	1 214.9	30
25	2	2	4	3 740.7	15
26	2	2	5	109.4	37
27	2	2	6	404.7	12
28	2	2	7	66.3	15

by the following frequency function, specializing to a probability density function in the continuous case and a probability mass function in the discrete case,

$$f_{Y_i}(y_i;\theta_i,\phi) = \exp\left\{\frac{y_i\theta_i - b(\theta_i)}{\phi/w_i} + c(y_i,\phi,w_i)\right\}. \tag{2.1}$$

Here θ_i is a parameter that is allowed to depend on i, while the *dispersion parameter* $\phi > 0$ is the same for all i. The so called *cumulant function* $b(\theta_i)$ is assumed twice continuously differentiable, with invertible second derivative. For every choice of such a function, we get a family of probability distributions, e.g. the normal, Poisson and gamma distributions, see Example 2.1 and (2.3) and (2.6) below. Given the choice of $b(\cdot)$, the distribution is completely specified by the parameters θ_i and ϕ. The function $c(\cdot,\cdot,\cdot)$, which does not depend on θ_i, is of little interest in GLM theory.

Of course, the expression above is only valid for the y_i that are possible outcomes of Y_i—the *support*; for other values of y_i we tacitly assume $f_{Y_i}(y_i) = 0$. Examples of support we will encounter is $(0,\infty)$, $(-\infty,\infty)$ and the non-negative integers.

Further technical restrictions are that $\phi > 0$, $w_i \geq 0$ and that the parameter space must be open, i.e., θ_i takes values in an open set, such as $0 < \theta_i < 1$ (while a closed set such as $0 \leq \theta_i \leq 1$ is not allowed).

An overview of the theory of the so defined exponential dispersion models can be found in Jörgensen [Jö97]. We will presuppose no knowledge of the theory for EDMs, and our interest in them is restricted only to the role they play in GLMs.

Remark 2.1 If ϕ was regarded as fixed, (2.1) would define a so called one-parameter exponential family, see, e.g., Lindgren [Li93, p. 188]. If, on the other hand, ϕ is unknown then we usually do *not* have a two-parameter exponential family, but we do have an EDM.

Example 2.1 (The normal distribution) Here we show that the normal distribution used in (weighted) linear models is a member of the EDM class; note, though, that the normal distribution is seldom used in the applications we have in mind. Nevertheless, assume for the moment that we have a normally distributed key ratio Y_i. The expectation of observation i is denoted μ_i, i.e., $\mu_i = E(Y_i)$. Lemma 1.1 shows that the variance must be w_i-weighted; the σ^2 of that lemma is assumed to be the same for all i in linear models. We conclude that $Y_i \sim N(\mu_i, \sigma^2/w_i)$, where w_i is the exposure. Then the frequency function is

$$f_{Y_i}(y_i) = \exp\left\{\frac{y_i\mu_i - \mu_i^2/2}{\sigma^2/w_i} + c(y_i, \sigma^2, w_i)\right\}, \tag{2.2}$$

were we have separated out the part of the density not depending on μ_i,

$$c(y_i, \sigma^2, w_i) = -\frac{1}{2}\left(\frac{w_i y_i^2}{\sigma^2} + \log(2\pi\sigma^2/w_i)\right).$$

This is an EDM with $\theta_i = \mu_i$, $\phi = \sigma^2$ and $b(\theta_i) = \theta_i^2/2$. Hence, the normal distribution used in (weighted) linear models is an EDM; the unweighted case is, of course, obtained by letting $w_i \equiv 1$.

2.1.1 Probability Distribution of the Claim Frequency

Let $N(t)$ be the number of claims for an individual policy during the time interval $[0, t]$, with $N(0) = 0$. The stochastic process $\{N(t); t \geq 0\}$ is called the *claims process*. Beard, Pentikäinen and Pesonen [BPP84, Appendix 4] show that under assumptions that are close to our Assumptions 1.2–1.3, plus an assumption that claims do not cluster, the claims process is a *Poisson process*. This motivates us to assume a Poisson distribution for the number of claims of an individual policy during any given period of time. By the independence of policies, Assumption 1.1, we get a Poisson distribution also at the aggregate level of all policies in a tariff cell.

So let X_i be the number of claims in a tariff cell with duration w_i and let μ_i denote the expectation when $w_i = 1$. Then by Lemma 1.1 we have $E(X_i) = w_i\mu_i$, and so X_i follows a Poisson distribution with frequency function

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

We are more interested in the distribution of the claim frequency $Y_i = X_i/w_i$; in the literature, this case is often (vaguely) referred to as Poisson, too, but since it is rather a transformation of that distribution we give it a special name, the *relative Poisson distribution*. The frequency function is, for y_i such that $w_i y_i$ is a non-negative integer,

$$f_{Y_i}(y_i;\mu_i) = P(Y_i = y_i) = P(X_i = w_i y_i) = \mathrm{e}^{-w_i\mu_i}\frac{(w_i\mu_i)^{w_i y_i}}{(w_i y_i)!}$$
$$= \exp\{w_i[y_i \log(\mu_i) - \mu_i] + c(y_i, w_i)\}, \tag{2.3}$$

where $c(y_i, w_i) = w_i y_i \log(w_i) - \log(w_i y_i!)$. This is an EDM, as can be seen by reparameterizing it through $\theta_i = \log(\mu_i)$,

$$f_{Y_i}(y_i;\theta_i) = \exp\{w_i(y_i\theta_i - \mathrm{e}^{\theta_i}) + c(y_i, w_i)\}.$$

This is of the form given in (2.1), with $\phi = 1$ and the cumulant function $b(\theta_i) = \mathrm{e}^{\theta_i}$. The parameter space is $\mu_i > 0$, i.e., the open set $-\infty < \theta_i < \infty$.

Remark 2.2 Is the Poisson distribution realistic? In practice, the homogeneity within cells is hard to achieve. The expected claim frequency μ_i of the Poisson process may vary with time, but this is not necessarily a problem since the number of claims during a year will still be Poisson distributed. A more serious problem is that there is often considerable variation left *between* policies within cells. This can be modeled by letting the risk parameter μ_i itself be the realization of a random variable. This leads to a so called *mixed* Poisson distribution, with larger variance than standard Poisson, see, e.g., [KPW04, Sect. 4.6.3]; such models often fit insurance data better than the standard Poisson. We will return to this problem in Sects. 3.4 and 3.5.

It is a reasonable requirement for a probabilistic model of a claim frequency to be *reproductive*, in the following sense. Suppose that we work under the assumption that the claim frequency in each cell has a relative Poisson distribution. If a tariff analysis shows that two cells have similar expectation, we might decide to merge them into just one cell. Then it would be very strange if we got another probability distribution in the new cell. Luckily, this problem will not arise—on the contrary, the relative Poisson distribution is reproduced on the aggregated level, as we shall now show.

Let Y_1 and Y_2 be the claim frequency in two cells with exposures w_1 and w_2, respectively, and let both follow a relative Poisson distribution with parameter μ. If we merge these two cells, the claim frequency in the new cell will be the weighted average

$$Y = \frac{w_1 Y_1 + w_2 Y_2}{w_1 + w_2}.$$

Since $w_1 Y_1 + w_2 Y_2$ is the sum of two independent Poisson distributed variables, it is itself Poisson distributed. Hence Y follows a relative Poisson distribution with exposure $w_1 + w_2$ and the parameter is, by elementary rules for the expectation of a linear expression, μ.

As indicated above, a parametric distribution that is closed under this type of averaging will be called *reproductive*, a concept coined by Jörgensen [Jö97]. In fact, this is a natural requirement for any key ratio; fortunately, we will see below in Theorem 2.2 that all EDMs are reproductive.

2.1.2 A Model for Claim Severity

We now turn to claim severity, and again we shall build a model for each tariff cell, but for the sake of simplicity, let us temporarily drop the index i. The exposure for claim severity, the number of claims, is then written w. Recall that in this analysis, we *condition* on the number of claims so that the exposure weight is non-random, as it should be. The idea is that we first analyze claim frequency with the number of claims as the outcome of a random variable; once this is done, we condition on the number of claims in analyzing claim severity. Here, the total claim cost in the cell is X and the claim severity $Y = X/w$.

In the previous section, we presented a plausible motivation for using the Poisson distribution, under the assumptions on independence and homogeneity. However, in the claim severity case it is not at all obvious which distribution we should assume for X. The distribution should be positive and skewed to the right, so the normal distribution is not suitable, but there are several other candidates that fulfill the requirements. However, the gamma distribution has become more or less a de facto standard in GLM analysis of claim severity, see, e.g., [MBL00, p. 10] or [BW92, Sect. 3]. As we will show in Sect. 2.3.3, the gamma assumption implies that the standard deviation is proportional to μ, i.e., we have a constant coefficient of variation; this seems quite plausible for claim severity. In Sect. 3.5 we will discuss the possibility of constructing estimators without assuming a specific distribution, starting from assumptions for the mean and variance structure only.

For the time being, we assume that the cost of an individual claim is gamma distributed; this is the case $w = 1$. One of several equivalent parameterizations is that with a so called *index parameter* $\alpha > 0$, a *scale parameter* $\beta > 0$, and the frequency function

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \mathrm{e}^{-\beta x}; \quad x > 0. \tag{2.4}$$

We denote this distribution $G(\alpha, \beta)$ for short. It is well known that the expectation is α/β and the variance α/β^2, see Exercise 2.4. Furthermore, sums of independent gamma distributions with the same scale parameter β are gamma distributed with the same scale and an index parameter which is the sum of the individual α, see Exercise 2.11. So if X is the sum of w independent gamma distributed random variables, we conclude that $X \sim G(w\alpha, \beta)$. The frequency function for $Y = X/w$ is then

$$f_Y(y) = wf_X(wy) = \frac{(w\beta)^{w\alpha}}{\Gamma(w\alpha)} y^{w\alpha-1} \mathrm{e}^{-w\beta y}; \quad y > 0,$$

and so $Y \sim G(w\alpha, w\beta)$ with expectation α/β. Before transforming this distribution to EDM form, it is instructive to re-parameterize it through $\mu = \alpha/\beta$ and $\phi = 1/\alpha$. In Exercise 2.5 we ask the reader to verify that the new parameter space is given by $\mu > 0$ and $\phi > 0$. The frequency function is

$$\begin{aligned} f_Y(y) = f_Y(y; \mu, \phi) &= \frac{1}{\Gamma(w/\phi)} \left(\frac{w}{\mu\phi}\right)^{w/\phi} y^{(w/\phi)-1} \mathrm{e}^{-wy/(\mu\phi)} \\ &= \exp\left\{\frac{-y/\mu - \log(\mu)}{\phi/w} + c(y, \phi, w)\right\}; \quad y > 0, \end{aligned} \tag{2.5}$$

where $c(y, \phi, w) = \log(wy/\phi)w/\phi - \log(y) - \log\Gamma(w/\phi)$. We have $E(Y) = w\alpha/(w\beta) = \mu$ and $\mathrm{Var}(Y) = w\alpha/(w\beta)^2 = \phi\mu^2/w$, which is consistent with Lemma 1.1.

To show that the gamma distribution is an EDM we finally change the first parameter in (2.5) to $\theta = -1/\mu$; the new parameter takes values in the open set $\theta < 0$. Returning to the notation with index i, the frequency function of the claim severity Y_i is

$$f_{Y_i}(y_i; \theta_i, \phi) = \exp\left\{\frac{y_i\theta_i + \log(-\theta_i)}{\phi/w_i} + c(y_i, \phi, w_i)\right\}. \tag{2.6}$$

We conclude that the gamma distribution is an EDM with $b(\theta_i) = -\log(-\theta_i)$ and hence we can use it in a GLM.

Remark 2.3 By now, the reader might feel that in the right parameterization, any distribution is an EDM, but that is not the case; the log-normal distribution, e.g., can not be rearranged into an EDM.

2.1.3 Cumulant-Generating Function, Expectation and Variance

The *cumulant-generating function* is the logarithm of the moment-generating function; it is useful for computing the expectation and variance of Y_i, for finding the distribution of sums of independent random variables, and more. For simplicity, we once more refrain from writing out the subindex i for the time being. The *moment-generating function* of an EDM is defined as $M(t) = E(\mathrm{e}^{tY})$, if this expectation is finite at least for real t in a neighborhood of zero. For continuous EDMs we find, by using (2.1),

$$\begin{aligned} E(\mathrm{e}^{tY}) &= \int \mathrm{e}^{ty} f_Y(y; \theta, \phi)\, dy \\ &= \int \exp\left\{\frac{y(\theta + t\phi/w) - b(\theta)}{\phi/w} + c(y, \phi, w)\right\} dy \end{aligned}$$

$$
\begin{aligned}
&= \exp\left\{\frac{b(\theta + t\phi/w) - b(\theta)}{\phi/w}\right\} \\
&\quad \times \int \exp\left\{\frac{y(\theta + t\phi/w) - b(\theta + t\phi/w)}{\phi/w} + c(y, \phi, w)\right\} dy. \qquad (2.7)
\end{aligned}
$$

Recall the assumption that the parameter space of an EDM must be open. It follows, at least for t in a neighborhood of 0, i.e., for $|t| < \delta$ for some $\delta > 0$, that $\theta + t\phi/w$ is in the parameter space. Thereby, the last integral equals one and the factor preceding it is the moment-generating function, which thus exists for $|t| < \delta$.

In the discrete case, we get the same expression by changing the integrals in (2.7) to sums. By taking logarithms, we conclude that the cumulant-generating function (CGF), denoted $\Psi(t)$, exists for any EDM and is given by

$$
\Psi(t) = \frac{b(\theta + t\phi/w) - b(\theta)}{\phi/w}, \qquad (2.8)
$$

at least for t in some neighborhood of 0. This is the reason for the name *cumulant function* that we have already used for $b(\theta)$.

As the name suggests, the CGF can be used to derive the so called cumulants; this is done by differentiating and setting $t = 0$. The first cumulant is the expected value; the second cumulant is the variance; the reader who is not familiar with this result is invited to derive it from well-known results for moment-generating functions in Exercise 2.8.

We use the above property to derive the expected value of an EDM as follows, recalling that we have assumed that $b(\cdot)$ is twice differentiable,

$$
\Psi'(t) = b'(\theta + t\phi/w); \qquad E(Y) = \Psi'(0) = b'(\theta).
$$

The second cumulant, the variance, is given by

$$
\Psi''(t) = b''(\theta + t\phi/w)\phi/w; \qquad \mathrm{Var}(Y) = \Psi''(0) = b''(\theta)\phi/w.
$$

As a check, let us see what this yields in the case of a normal distribution: here $b(\theta) = \theta^2/2$, $b'(\theta) = \theta$ and $b''(\theta) = 1$, whence $E(Y) = \theta = \mu$ and $\mathrm{Var}(Y) = \phi/w = \sigma^2/w$ as it should be.

In general, it is more convenient to view the variance as a function of the mean μ. We have just seen that $\mu = E(Y) = b'(\theta)$, and since this is assumed to be an invertible function, we may insert the inverse relationship $\theta = b'^{-1}(\mu)$ into $b''(\theta)$ to get the so called *variance function* $v(\mu) \doteq b''(b'^{-1}(\mu))$. Now we can express $\mathrm{Var}(Y)$ as the product of the variance function $v(\mu)$ and a scaling and weighting factor ϕ/w. Here are some examples.

Example 2.2 (Variance functions) For the relative Poisson distribution in Sect. 2.1.1 we have $b(\theta_i) = \exp(\theta_i)$, by which $\mu_i = b'(\theta_i) = \exp(\theta_i)$. Furthermore $b''(\theta_i) = \exp(\theta_i) = \mu_i$, so that $v(\mu_i) = \mu_i$ and $\mathrm{Var}(Y) = \mu_i/w_i$, since $\phi = 1$. In the case $w_i = 1$ this is just the well-known result that the Poisson distribution has variance equal to the mean.

Table 2.3 Example of variance functions

Distribution	Normal	Poisson	Gamma	Binomial
$v(\mu)$	1	μ	μ^2	$\mu(1-\mu)$

The gamma distribution has $b(\theta_i) = -\log(-\theta_i)$ and $b'(\theta_i) = -1/\theta_i$ so that $\mu_i = -1/\theta_i$, as we already knew. Furthermore, $b''(\theta_i) = 1/\theta_i^2 = \mu_i^2$ so that $\mathrm{Var}(Y_i) = \phi\mu_i^2/w_i$.

The variance functions we have seen so far are collected in Table 2.3, together with the binomial distribution that is treated in Exercises 2.6 and 2.9.

We summarize the results in the following lemma, returning to our usual notation with index i for the observation number.

Lemma 2.1 *Suppose that Y_i follows an EDM, with frequency function given in* (2.1). *Then the cumulant generating function exists and is given by*

$$\Psi(t) = \frac{b(\theta_i + t\phi/w_i) - b(\theta_i)}{\phi/w_i},$$

and

$$\mu_i \doteq E(Y_i) = b'(\theta_i);$$
$$\mathrm{Var}(Y_i) = \phi v(\mu_i)/w_i,$$

where the variance function $v(\mu_i)$ is $b''(\cdot)$ expressed as a function of μ_i, i.e., $v(\mu_i) = b''(b'^{-1}(\mu_i))$.

Remark 2.4 Recall that in linear regression, we have a constant variance $\mathrm{Var}(Y_i) = \phi$, plus possibly a weight w_i which we disregard for a moment. The most general assumption would be to allow $\mathrm{Var}(Y_i) = \phi_i$, but this would make the model heavily over-parameterized. In GLMs we are somewhere in-between these extremes, since the variance function $v(\mu_i)$ allows the variance to vary over the cells i, but without introducing any new parameters.

The variance function is important in GLM model building, a fact that is emphasized by the following theorem.

Theorem 2.1 *Within the EDM class, a family of probability distributions is uniquely characterized by its variance function.*

The practical implication of this theorem is that if you have decided to use a GLM, and hence an EDM, you only have to determine the variance function; then you know the precise probability distribution within the EDM class. This is an interesting result: only having to model the mean and variance is much simpler than having to specify an entire distribution.

The core in the proof of this theorem is to notice that since $v(\cdot)$ is a function of the derivatives of $b(\cdot)$, the latter can be determined from $v(\cdot)$ by solving a pair of differential equations—but $b(\cdot)$ is all we need to specify the EDM distribution in (2.1). The proof can be found in Jörgensen [Jö87, Theorem 1].

We saw in Sect. 2.1.1 that the relative Poisson distribution had the appealing property of being reproductive. The next theorem shows that this holds for all distributions within the EDM class.

Theorem 2.2 (EDMs are reproductive) *Suppose we have two independent random variables Y_1 and Y_2 from the same EDM family, i.e., with the same $b(\cdot)$, that have the same mean μ and dispersion parameter ϕ, but possibly different weights w_1 and w_2. Then their w-weighted average $Y = (w_1 Y_1 + w_2 Y_2)/(w_1 + w_2)$ belongs to the same EDM distribution, but with weight $w_\cdot = w_1 + w_2$.*

The proof is left as Exercise 2.12, using some basic properties of CGFs derived in Exercise 2.10. From an applied point of view the importance of this theorem is that if we merge two tariff cells, with good reason to assume they have the same mean, we will stay within the same family of distributions. From this point of view, EDMs behave the way we want a probability distribution in pricing to behave. One may show that the log-normal and the (generalized) Pareto distribution do not have this property; our conclusion is that, useful as they may be in other actuarial applications, they are not well suited for use in a tariff analysis.

2.1.4 Tweedie Models

In non-life actuarial applications, it is often desirable to work with probability distributions that are closed with respect to scale transformations, or *scale invariant*. Let c be a positive constant, $c > 0$, and Y a random variable from a certain family of distributions; we say that this family is scale invariant if cY follows a distribution in the same family. This property is desirable if Y is measured in a monetary unit: if we convert the data from one currency to another, we want to stay within the same family of distributions—the result of a tariff analysis should not depend on the currency used. Similarly, the inference should not depend on whether we measure the claim frequency in per cent or per mille. We conclude that scale invariance is desirable for all the key ratios in Table 1.3, except for the proportion of large claims, for which scale is not relevant.

It can be shown that the only EDMs that are scale invariant are the so called Tweedie models, which are defined as having variance function

$$v(\mu) = \mu^p \tag{2.9}$$

for some p. The proof can be found in Jörgensen [Jö97, Chap. 4], upon which much of the present section is based, without further reference.

Table 2.4 Overview of Tweedie models

	Type	Name	Key ratio
$p < 0$	Continuous	–	–
$p = 0$	Continuous	Normal	–
$0 < p < 1$	Non-existing	–	–
$p = 1$	Discrete	Poisson	Claim frequency
$1 < p < 2$	Mixed, non-negative	Compound Poisson	Pure premium
$p = 2$	Continuous, positive	Gamma	Claim severity
$2 < p < 3$	Continuous, positive	–	Claim severity
$p = 3$	Continuous, positive	Inverse normal	Claim severity
$p > 3$	Continuous, positive	–	Claim severity

With the notable exception of the relative binomial in Exercise 2.6, for which scale invariance is not required since y is a proportion, the EDMs in this book are all Tweedie models. In Table 2.4 we give a list of all Tweedie models and the key ratios for which they might be useful. The cases $p = 0, 1, 2$ have already been discussed. Tweedie models with $p \geq 2$ are often suggested as distributions for the claim severity, especially $p = 2$, but also the *inverse normal distribution* ($p = 3$) is sometimes mentioned in the literature.

The class of models with $1 < p < 2$ is interesting: these so called *compound Poisson distributions* arise as the distribution of a sum of a Poisson distributed number of claims which follow a gamma distribution; hence, they are proper for modeling the pure premium, without doing a separate analysis of claim frequency and severity. Note that the compound Poisson is mixed (neither purely discrete nor purely continuous), having positive probability at zero plus a continuous distribution on the positive real numbers.

Jörgensen and Souza [JS94] analyze the pure premium for a private motor insurance portfolio in Brazil, and get the value $p = 1.37$ from an algorithm they designed for maximum likelihood estimation of p.

A bit surprising is that for $0 < p < 1$ no EDM exists. Negative values of p, finally, are allowed and give continuous distributions on the whole real axis, but to the best of our knowledge no application in insurance has been proposed.

From now on, we only discuss the case $p \geq 1$, which covers our applications, and we start by presenting the corresponding cumulant function $b(\theta)$.

$$b(\theta) = \begin{cases} e^{\theta}, & \text{for } p = 1; \\ -\log(-\theta), & \text{for } p = 2; \\ -\frac{1}{p-2}[-(p-1)\theta]^{(p-2)/(p-1)}, & \text{for } 1 < p < 2 \text{ and } p > 2. \end{cases} \tag{2.10}$$

The parameter space M_θ is

$$M_\theta = \begin{cases} -\infty < \theta < \infty, & \text{for } p = 1; \\ -\infty < \theta < 0, & \text{for } p > 1. \end{cases} \tag{2.11}$$

The derivative $b'(\theta)$ is given by

$$b'(\theta) = \begin{cases} e^{\theta}, & \text{for } p = 1; \\ [-(p-1)\theta]^{-1/(p-1)}, & \text{for } p > 1, \end{cases} \tag{2.12}$$

with inverse

$$h(\mu) = \begin{cases} \log(\mu), & p = 1; \\ -\frac{1}{p-1}\mu^{-(p-1)}, & p > 1. \end{cases} \tag{2.13}$$

These results are taken from Jörgensen [Jö97]; some are also verified in Exercise 2.13.

2.2 The Link Function

We have seen how to generalize the normal error distribution to the EDM class, and now turn to the other generalization of ordinary linear models, concerning the linear structure of the mean.

We start by discussing a simple example, in which we only have two rating factors, one with two classes and one with three classes. On tabular form, we let μ_{ij} denote the expectation of the key ratio in cell (i, j), where the first factor is in class i and the second is in class j. Linear models assume an *additive model* structure for the mean:

$$\mu_{ij} = \gamma_0 + \gamma_{1i} + \gamma_{2j}. \tag{2.14}$$

We recognize this model from the analysis of variance (ANOVA), and recall that it is over-parameterized, unless we add some restrictions. In an ANOVA the usual restriction is that marginal sums should be zero, but here we chose the restriction to force the parameters of some *base cell* to be zero. Say that (1, 1) is the base cell; then we let $\gamma_{11} = \gamma_{21} = 0$, so that $\mu_{11} = \gamma_0$, and the other parameters measure the mean departure from this cell. Next we rewrite the model on list form by sorting the cells in the order (1, 1); (1, 2); (1, 3); (2, 1); (2, 2); (2, 3) and renaming the parameters,

$$\begin{aligned} \beta_1 &\doteq \gamma_0, \\ \beta_2 &\doteq \gamma_{12}, \\ \beta_3 &\doteq \gamma_{22}, \\ \beta_4 &\doteq \gamma_{23}. \end{aligned}$$

With these parameters the expected values in the cells are as listed in Table 2.5.

Next, we introduce so called *dummy variables* through the relation

$$x_{ij} = \begin{cases} 1, & \text{if } \beta_j \text{ is included in } \mu_i, \\ 0, & \text{else.} \end{cases}$$

Table 2.5 Parameterization of a two-way additive model on list form

i	Tariff cell		μ_i			
1	1	1	β_1			
2	1	2	β_1		$+\beta_3$	
3	1	3	β_1			$+\beta_4$
4	2	1	β_1	$+\beta_2$		
5	2	2	β_1	$+\beta_2$	$+\beta_3$	
6	2	3	β_1	$+\beta_2$		$+\beta_4$

Table 2.6 Dummy variables in a two-way additive model

i	Tariff cell		x_{i1}	x_{i2}	x_{i3}	x_{i4}
1	1	1	1	0	0	0
2	1	2	1	0	1	0
3	1	3	1	0	0	1
4	2	1	1	1	0	0
5	2	2	1	1	1	0
6	2	3	1	1	0	1

The values of the dummy variables in this example are given in Table 2.6. Note the similarity to Table 2.5.

With these variables, the linear model for the mean can be rewritten

$$\mu_i = \sum_{j=1}^{4} x_{ij}\beta_j \quad i = 1, 2, \ldots, 6, \tag{2.15}$$

or on matrix form $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$, where $\mathbf{X}$ is called the *design matrix*, or *model matrix*, and

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \\ \mu_5 \\ \mu_6 \end{pmatrix}, \qquad \mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & x_{13} & x_{14} \\ x_{21} & x_{22} & x_{23} & x_{24} \\ x_{31} & x_{32} & x_{33} & x_{34} \\ x_{41} & x_{42} & x_{43} & x_{44} \\ x_{51} & x_{52} & x_{53} & x_{54} \\ x_{61} & x_{62} & x_{63} & x_{64} \end{pmatrix}, \qquad \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix}. \tag{2.16}$$

So far the additive model for the mean has been used; next we turn to the multiplicative model $\mu_{ij} = \gamma_0 \gamma_{1i} \gamma_{2j}$ that was introduced in Sect. 1.3. By taking logarithms we get

$$\log(\mu_{ij}) = \log(\gamma_0) + \log(\gamma_{1i}) + \log(\gamma_{2j}), \tag{2.17}$$

and again we must select a base cell, say $(1, 1)$, where $\gamma_{11} = \gamma_{21} = 1$. Let us now do a transition to list form similar to the one we just performed for the additive model, by first letting

$$\beta_1 \doteq \log \gamma_0,$$
$$\beta_2 \doteq \log \gamma_{12},$$
$$\beta_3 \doteq \log \gamma_{22},$$
$$\beta_4 \doteq \log \gamma_{23}.$$

By the aid of the dummy variables in Table 2.6, we have

$$\log(\mu_i) = \sum_{j=1}^{4} x_{ij}\beta_j; \quad i = 1, 2, \ldots, 6. \tag{2.18}$$

This is the same linear structure as in (2.15), the only difference being that the left-hand side is $\log(\mu_i)$ instead of just μ_i. In general GLMs this is further generalized to allow the left-hand side to be any monotone function $g(\cdot)$ of μ_i.

Leaving the simple two-way example behind, the general tariff analysis problem is to investigate how the response Y_i is influenced by r covariates $x_1, x_2, \ldots, x_r$. Introduce

$$\eta_i = \sum_{j=1}^{r} x_{ij}\beta_j; \quad i = 1, 2, \ldots, n, \tag{2.19}$$

where x_{ij} as before is the value of the covariate x_j for observation i.

In the ordinary linear model, $\mu_i \equiv \eta_i$; in a GLM this is generalized to an arbitrary relation $g(\mu_i) = \eta_i$, with the restriction that $g(\cdot)$ must be a monotone, differentiable function. This fundamental object in a GLM is called the *link function*, since it links the mean to the linear structure through

$$g(\mu_i) = \eta_i = \sum_{j=1}^{r} x_{ij}\beta_j. \tag{2.20}$$

We have seen that multiplicative models correspond to a logarithmic link function, a *log link*,

$$g(\mu_i) = \log(\mu_i),$$

while the linear model uses the *identity link* $\mu_i = \eta_i$, i.e., $g(\mu_i) = \mu_i$. Note that the link function is not allowed to depend on i.

For the analysis of proportions, it is common to use a *logit link*,

$$\eta_i = g(\mu_i) = \log\left(\frac{\mu_i}{1-\mu_i}\right). \tag{2.21}$$

This link guarantees that the mean will stay between zero and one, as required in a model where μ_i is a proportion, such as the relative binomial model for the key ratio *proportion of large claims*. The corresponding GLM analysis goes under the name *logistic regression*.

In a GLM, the link function is part of the model specification. In non-life insurance pricing, the log link is by far the most common one, since a multiplicative model is often reasonable. For a discussion of the merits of the multiplicative model, see Sect. 1.3.

Interactions may cause departure from multiplicativity. An example is motor insurance, where young men are more accident-prone than young women, while in midlife there is little gender difference in driving behavior. In Sect. 3.6.2 we will show how this problem can be handled within the general multiplicative framework. For further discussion on multiplicative models, see [BW92, Sects. 2.1 and 3.1.1]. With the exception of the analysis of proportions, we will assume a multiplicative model throughout this text, with possible interactions handled by combining variables as in the age and gender example.

We end this section by summarizing the GLM generalization of the ordinary linear model and the particular case that is most used in our applications.

Weighted linear regression models:
- Y_i follows a normal distribution with $\mathrm{Var}(Y_i) = \sigma^2/w_i$;
- The mean follows the additive model $\mu_i = \sum_j x_{ij}\beta_j$.

Generalized Linear Models (GLMs):
- Y_i follows an EDM with $\mathrm{Var}(Y_i) = \phi v(\mu_i)/w_i$;
- The mean satisfies $g(\mu_i) = \sum_j x_{ij}\beta_j$, where g is a monotone function.

Multiplicative Tweedie model, subclass of GLMs:
- Y_i follows a Tweedie EDM with $\mathrm{Var}(Y_i) = \phi\mu_i^p/w_i$, $p \geq 1$;
- The mean follows the multiplicative model $\log(\mu_i) = \sum_j x_{ij}\beta_j$.

2.2.1 Canonical Link*

In the GLM literature one encounters the concept of a *canonical link*; this concept is not so important in our applications, but for the sake of completeness we shall give a brief orientation on this subject. First we note that we have been working with several different parameterizations of a GLM; these parameters are unique functions of each other as illustrated in the following figure:

$$\theta \xrightarrow{b'(\cdot)} \mu \xrightarrow{g(\cdot)} \eta. \tag{2.22}$$

Here $b(\cdot)$, and hence $b'(\cdot)$, are determined by the structure of the random component, uniquely determined by the choice of variance function. The link function $g(\cdot)$, on the other hand, is part of the modeling of the mean and in some applications there are several reasonable choices. The special choice $g(\cdot) = b'^{-1}(\cdot)$ is the *canonical link*. From (2.22) we find that the canonical link makes $\theta = \eta$. It turns out that using the canonical link simplifies some computations, but the name is somewhat misleading since it may not be the natural choice in a particular application. On the other hand, some of the most common models actually use canonical links.

Example 2.3 (Some canonical links) The normal distribution has $\mu = b'(\theta) = \theta$ and the identity link $g(\mu) = \mu$ that is used in the linear model is the canonical one.

In the Poisson case we have $\mu = b'(\theta) = e^{\theta}$, by which the log link $g(\mu) = \log \mu$ is canonical. So for this important EDM, the multiplicative model is canonical.

In the case of the gamma distribution $b'(\theta) = -1/\theta$ and so the canonical link is the inverse link $g(\mu) = -1/\mu$. A problem with this link is that, as opposed to the log link, it may cause the mean to take on negative values. McCullagh and Nelder [MN89, Sects. 8.4.1 and 12.8.3], suggests using this link for claim severity, but as far as we know it is not used in practice.

2.3 Parameter Estimation

So far, we have defined GLMs and studied some of their properties. It is now time for the most important step: the estimation of the regression parameters in (2.20), from which we will get the relativities—the basic building blocks of the tariff. Before deriving a general result, we will give an introduction to the subject by treating an important special case.

2.3.1 The Multiplicative Poisson Model

We return to the simple case in Sect. 1.3.1, with just two rating factors. For easy reference, we repeat the multiplicative model on tabular form from (1.3)

$$\mu_{ij} = \gamma_0 \gamma_{1i} \gamma_{2j}. \tag{2.23}$$

In GLM terms, we say that we use a log link. In addition to this, we assume that the claim frequency Y_{ij} has a relative Poisson distribution, with frequency function given on list form in (2.3). We rewrite this on tabular form and note that the cells are independent due to Assumption 1.1, implying that the log-likelihood of the whole sample is the sum of the log-likelihoods in the individual cells.

$$\ell = \sum_i \sum_j w_{ij}\{y_{ij}[\log(\gamma_0) + \log(\gamma_{1i}) + \log(\gamma_{2j})] - \gamma_0 \gamma_{1i} \gamma_{2j}\} + c, \tag{2.24}$$

where c does not depend on the γ parameters. By in turn differentiating ℓ with respect to each γ, we get a system of equations for the maximum likelihood estimates (MLEs), the so called ML equations. The result is exactly the MMT equations in (1.6)–(1.8); this is no surprise if we recall that Jung [Ju68] derived these equations as the MLEs of a Poisson model, cf. Remark 1.1.

In general, under a GLM model for the claim frequency with a relative Poisson distribution and log link (multiplicative model), the ML equations are equal to the equations of the method of marginal totals (MMT), and hence the resulting estimates are the same, see Exercise 2.16.

2.3.2 General Result

We now turn to the general case of finding the MLEs of the $\boldsymbol{\beta}$-parameters in a GLM. The estimates are based on a sample of n observations. The individual observations follow the EDM distribution given on list form in (2.1) and by independence, the log-likelihood as a function of $\boldsymbol{\theta}$ is

$$\ell(\boldsymbol{\theta}; \phi, \mathbf{y}) = \frac{1}{\phi} \sum_i w_i (y_i \theta_i - b(\theta_i)) + \sum_i c(y_i, \phi, w_i). \tag{2.25}$$

It is clear that the dispersion parameter ϕ does not affect the maximization of ℓ with respect to $\boldsymbol{\theta}$—a similar observation should be familiar from the linear regression model, where ϕ is denoted σ^2. We can therefore ignore ϕ here, but we will return to the question of how to estimate it later, in Sect. 3.1.1.

The likelihood as a function of $\boldsymbol{\beta}$, rather than $\boldsymbol{\theta}$, could be found by the inverse of the relation $\mu_i = b'(\theta_i)$, combined with the link $g(\mu_i) = \eta_i = \sum_j x_{ij}\beta_j$. The derivative of ℓ with respect to β_j is, by the chain rule,

$$\begin{aligned} \frac{\partial \ell}{\partial \beta_j} &= \sum_i \frac{\partial \ell}{\partial \theta_i}\frac{\partial \theta_i}{\partial \beta_j} = \frac{1}{\phi}\sum_i \left(w_i y_i - w_i b'(\theta_i)\right)\frac{\partial \theta_i}{\partial \beta_j} \\ &= \frac{1}{\phi}\sum_i \left(w_i y_i - w_i b'(\theta_i)\right)\frac{\partial \theta_i}{\partial \mu_i}\frac{\partial \mu_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial \beta_j}. \end{aligned} \tag{2.26}$$

By the mentioned relation $\mu_i = b'(\theta_i)$ we have $\partial\mu_i/\partial\theta_i = b''(\theta_i)$. The derivative of the inverse relation is simply the inverse of the derivative, and so $\partial\theta_i/\partial\mu_i = 1/v(\mu_i)$, since by definition $v(\mu_i) = b''(\theta_i)$.

Furthermore, $\partial\mu_i/\partial\eta_i = [\partial\eta_i/\partial\mu_i]^{-1} = 1/g'(\mu_i)$. Finally, from $\eta_i = \sum_j x_{ij}\beta_j$ we get $\partial\eta_i/\partial\beta_j = x_{ij}$.

Putting all this together gives the result,

$$\frac{\partial \ell}{\partial \beta_j} = \frac{1}{\phi}\sum_i w_i \frac{y_i - \mu_i}{v(\mu_i)g'(\mu_i)} x_{ij}, \tag{2.27}$$

the so called *score function*. By setting all these r partial derivatives equal to zero and multiplying by ϕ, we get the ML equations

$$\sum_i w_i \frac{y_i - \mu_i}{v(\mu_i)g'(\mu_i)} x_{ij} = 0, \quad j = 1, 2, \ldots, r. \tag{2.28}$$

It might look as if the solution is simply $\mu_i = y_i$, but then we forget that $\mu_i = \mu_i(\boldsymbol{\beta})$ also has to satisfy the relation given by the regression on the x's, i.e.,

$$\mu_i = g^{-1}(\eta_i) = g^{-1}\left(\sum_j x_{ij}\beta_j\right). \tag{2.29}$$

It is only the so called *saturated model*, where there are as many parameters as there are observations, that allows the solution $\mu_i = y_i$.

It is interesting to note that the only property of the probability distribution that affects the ML equations (2.28) is the mean and the variance, through the link function g and the variance function v; for further discussion on this issue, see Sect. 3.5.

Example 2.4 (Tweedie models) In a tariff analysis, we typically use the Tweedie models of Sect. 2.1.4, where $v(\mu) = \mu^p$, in connection with a multiplicative model for the mean, i.e., $g(\mu_i) = \log(\mu_i)$, implying that $g'(\mu_i) = 1/\mu_i$. Then the general ML equations in (2.28) simplify to

$$\sum_i w_i \frac{y_i - \mu_i}{\mu_i^{p-1}} x_{ij} = 0 \quad \Longleftrightarrow \quad \sum_i \frac{w_i}{\mu_i^{p-1}} y_i x_{ij} = \sum_i \frac{w_i}{\mu_i^{p-1}} \mu_i x_{ij}, \tag{2.30}$$

where the μ's are connected to the β's through

$$\mu_i = \exp\left\{\sum_j x_{ij}\beta_j\right\}. \tag{2.31}$$

Compared to the Poisson case $p = 1$, models with $p > 1$ will downweigh both sides of the right-most equation in (2.30) by μ^{p-1}, giving less weight to cells with large expectation.

In the end, we are not interested in the β's, but rather the relativities γ. These are found by the relation $\gamma_j = \exp\{\beta_j\}$.

Introduce the diagonal matrix $\mathbf{W}$ with the parameter-dependent "weights"

$$\tilde{w}_i = \frac{w_i}{v(\mu_i)g'(\mu_i)},$$

on the diagonal and zeroes off the diagonal, $\mathbf{W} = \mathrm{diag}(\tilde{w}_i; i = 1, \ldots, n)$. Then (2.28) may be written on matrix form

$$\mathbf{X}'\mathbf{W}\mathbf{y} = \mathbf{X}'\mathbf{W}\boldsymbol{\mu}, \tag{2.32}$$

where $\mathbf{X}$ is the design matrix, cf. the example in (2.16). In the case of the normal distribution with identity link, it is readily seen that $\tilde{w}_i = w_i$, and if we furthermore let all weights $w_i = 1$, then (2.32) is reduced to the well-known "normal equations" of the linear model $\mathbf{X}'\mathbf{y} = \mathbf{X}'\mathbf{X}\boldsymbol{\beta}$, as we would expect since the linear model is a special case of the GLM. Here we have used the fact that $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ when we use the identity link.

Except for a few special cases, the ML equations must be solved numerically. The interested reader is referred to Sect. 3.2.3 for an introduction to numerical methods for solving the ML equations and to Sect. 3.2.4 for the question whether ML equations really give a (unique) maximum of the likelihood.

2.3.3 Multiplicative Gamma Model for Claim Severity

In Sect. 2.3.1 we discussed estimation for claim frequency; our next important special case is claim severity. With the data on list form, in cell i we have w_i claims and the claim severity Y_i, which is assumed to be relative gamma distributed with density given by (2.5). Let us have a closer look at the relation between the mean μ_i and the variance in this case. By Lemma 2.1 and Table 2.3 we have $E(Y_i) = \mu_i$ and $\mathrm{Var}(Y_i) = \phi\mu_i^2/w_i$. Hence,

$$\frac{\mathrm{Var}(Y_i)}{E(Y_i)^2} = \frac{\phi}{w_i}, \tag{2.33}$$

which means that the coefficient of variation (*CV*) is constant over cells with the same exposure w_i. This is, of course, a consequence of having a constant CV in the underlying model for *individual* claims. Such a constant CV, i.e., a standard deviation proportional to the mean, is plausible and in any case much more realistic than having a constant standard deviation: say that we have a tariff cell with mean 20 and standard deviation 4, then in another cell with the same exposure but with mean 200 we would rather expect a standard deviation of 40 than of 4.

From (2.30) with $p = 2$ we get the ML equations, which in this case are

$$\sum_i w_i \frac{y_i}{\mu_i} x_{ij} = \sum_i w_i x_{ij}. \tag{2.34}$$

It is instructive to write these equations on *tabular form* in the simple case with just two rating factors, using the multiplicative model in (2.23):

$$\begin{aligned}
\sum_i \sum_j \frac{w_{ij} y_{ij}}{\gamma_0 \gamma_{1i} \gamma_{2j}} &= \sum_i \sum_j w_{ij}; \\
\sum_j \frac{w_{ij} y_{ij}}{\gamma_0 \gamma_{1i} \gamma_{2j}} &= \sum_j w_{ij} \quad i = 2, \ldots, k_1; \\
\sum_i \frac{w_{ij} y_{ij}}{\gamma_0 \gamma_{1i} \gamma_{2j}} &= \sum_i w_{ij} \quad j = 2, \ldots, k_2.
\end{aligned} \tag{2.35}$$

This system of equations provides a quite natural algorithm for estimating relativities, giving a marginal balance in the sum of the relative deviations of the observations from their means: the average relative deviation equals one for each rating factor level i or j. Mack [Ma91] refers to work by Eeghen, Nijssen and Ruygt from 1982 where these equations are used for the pure premium, under the name *the direct method*. They found that the estimates of this method were very close to those given by the MMT. This is consistent with our general experience that the choice of p in the Tweedie models is not that important for estimating relativities.

2.3.4 Modeling the Pure Premium

In the end, it is the model for the pure premium that gives the tariff. One might consider using a Tweedie model with $1 < p < 2$ to analyze the pure premium directly, as demonstrated by Jörgensen and Souza [JS94]. However, the standard GLM tariff analysis is to do separate analyses for claim frequency and claim severity, and then relativities for the pure premium are found by multiplying the results. The reason for this separation into two GLMs is:

(i) Claim frequency is usually much more stable than claim severity and often much of the power of rating factors is related to claim frequency: these factors can then be estimated with greater accuracy;
(ii) A separate analysis gives more insight into how a rating factor affects the pure premium.

See also [BW92, MBL00].

Example 2.5 (Moped insurance contd.) We return once more to Example 1.1, with the aggregated data given in Table 1.2. In Example 1.3 we applied the MMT directly to the pure premium. We now carry out a separate analysis for claim frequency and severity, obtaining the relativities for the pure premium by multiplying the factors from these two analyses. The results are listed in Table 2.7.

We observe that the two variables *vehicle class* and *vehicle age* affect claim frequency and severity in the same direction, which means that newer and stronger vehicles are not only more expensive to replace when stolen, but are also stolen more often. The geographic zone has a large impact on the claim frequency, but once a moped is stolen the cost of replacing it is not necessarily larger in one zone than another, with a possible exception for the largest cities in zone 1. Note that these interesting conclusions could not have been drawn, had we analyzed pure premium only. On the other hand, the resulting estimates for pure premium are quite close to those obtained by the MMT method, see Table 1.4.

For zone 5 and 7, it is rather obvious that no conclusion can be drawn due to the very small number of claims, and hence very uncertain estimates of claim severity. But how can we know if the 23 observations in zone 6 or the 132 in zone 3 are enough to draw significant conclusions? This is one of the themes in the next chapter, where we go further into GLM theory and practice.

Table 2.7 Moped insurance: relativities from a multiplicative Poisson GLM for claim frequency and a gamma GLM for claim severity

Rating factor	Class	Duration	No. claims	Relativities, frequency	Relativities, severity	Relativities, pure premium
Vehicle class	1	9833	391	1.00	1.00	1.00
	2	8824	395	0.78	0.55	0.42
Vehicle age	1	1918	141	1.55	1.79	2.78
	2	16740	645	1.00	1.00	1.00
Zone	1	1451	206	7.10	1.21	8.62
	2	2486	209	4.17	1.07	4.48
	3	2889	132	2.23	1.07	2.38
	4	10069	207	1.00	1.00	1.00
	5	246	6	1.20	1.21	1.46
	6	1369	23	0.79	0.98	0.78
	7	147	3	1.00	1.20	1.20

2.4 Case Study: Motorcycle Insurance

Under the headline "case study" we will present larger exercises using authentic insurance data; for solving the case studies, a computer equipped with SAS or other suitable software is required.

The data for this case study comes from the former Swedish insurance company *Wasa*, and concerns partial casco insurance, for *motorcycles* this time. It contains aggregated data on all insurance policies and claims during 1994–1998; the reason for using this rather old data set is confidentiality—more recent data for ongoing business can not be disclosed. The data set `mccase.txt`, available at www.math.su.se/GLMbook, contains the following variables (with Swedish acronyms):

- *AGARALD*: The owners age, between 0 and 99.
- *KON*: The owners gender, M (male) or K (female).
- *ZON*: Geographic zone numbered from 1 to 7, in a standard classification of all Swedish parishes. The zones are the same as in the moped Example 1.1.
- *MCKLASS*: MC class, a classification by the so called EV ratio, defined as (Engine power in kW $\times$ 100) / (Vehicle weight in kg $+$ 75), rounded to the nearest lower integer. The 75 kg represent the average driver weight. The EV ratios are divided into seven classes as seen in Table 2.8.
- *FORDALD*: Vehicle age, between 0 and 99.
- *BONUSKL*: Bonus class, taking values from 1 to 7. A new driver starts with bonus class 1; for each claim-free year the bonus class is increased by 1. After the first claim the bonus is decreased by 2; the driver can not return to class 7 with less than 6 consecutive claim free years.

Table 2.8 Motorcycle insurance: rating factors and relativities in current tariff

Rating factor	Class	Class description	Relativity
Geographic zone	1	Central and semi-central parts of Sweden's three largest cities	7.678
	2	Suburbs plus middle-sized cities	4.227
	3	Lesser towns, except those in 5 or 7	1.336
	4	Small towns and countryside, except 5–7	1.000
	5	Northern towns	1.734
	6	Northern countryside	1.402
	7	Gotland (Sweden's largest island)	1.402
MC class	1	EV ratio −5	0.625
	2	EV ratio 6–8	0.769
	3	EV ratio 9–12	1.000
	4	EV ratio 13–15	1.406
	5	EV ratio 16–19	1.875
	6	EV ratio 20–24	4.062
	7	EV ratio 25–	6.873
Vehicle age	1	0–1 years	2.000
	2	2–4 years	1.200
	3	5– years	1.000
Bonus class	1	1–2	1.250
	2	3–4	1.125
	3	5–7	1.000

- *DURATION*: the number of policy years.
- *ANTSKAD*: the number of claims.
- *SKADKOST*: the claim cost.

The "current" tariff, the actual tariff from 1995, is based on just four rating factors, described in Table 2.8, where their current relativities are also given.

For each rating factor, we chose the class with the highest duration as base class. For the interested reader, we mention that the annual premium in the base cell $(4, 3, 3, 3)$ is 183 SEK, while the highest premium is 24 156 SEK (!).

Problem 1: Aggregate the data to the cells of the current tariff. Compute the empirical claim frequency and severity at this level.

Problem 2: Determine how the duration and number of claims is distributed for each of the rating factor classes, as an indication of the accuracy of the statistical analysis.

Problem 3: Determine the relativities for claim frequency and severity separately, by using GLMs; use the result to get relativities for the pure premium.

Problem 4: Compare the results in 3 to the current tariff. Is there a need to change the tariff? Which new conclusions can be drawn from the separate analysis in 3? Can we trust these estimates? With your estimates, what is the ratio between the highest pure premium and the lowest?

Exercises

2.1 (Section 2.1) Work out the details in the derivation of (2.2).

2.2 (Section 2.1) Suppose we have three rating factors, with two, three and five levels respectively. In how many ways could the tariff be written on list form? Each ordering of the cells is counted as one way of writing the tariff.

2.3 (Section 2.1) Show directly, without using the results in Sect. 2.1.3, that the normal distribution of Example 2.1 is reproductive; what is the value of $\mathrm{Var}(Y)$?

2.4 (Section 2.1) Verify that the expectation and variance of the gamma distribution in (2.4) are α/β and α/β^2, respectively.

2.5 (Section 2.1) Consider the reparameterization of the gamma distribution just before (2.5). Show that when ϕ and μ run through the first quadrant, α and β run through their parameter space, so that the same family of distributions is covered by the new parameterization.

2.6 (Section 2.1) An actuary studies the probability of customer renewal—that the customer chooses to stay with the insurance company for one more year—and how it varies between different groups. Let w_i be the number of customers in group i and X_i the number of renewals among these, so that by policy independence $X_i \sim Bin(w_i, p_i)$. Furthermore, let p_i be the probability under study, estimated by the key ratio $Y_i = X_i/w_i$.

Is the distribution of Y_i an EDM? If the answer is yes, determine the "canonical" parameter θ_i, an expression for ϕ, plus the functions $b(\cdot)$ and $c(\cdot)$. (In any case, we might call this the *relative binomial distribution*.)

2.7 (Section 2.1) The Generalized Pareto Distribution (GPD) is commonly used for large claims, especially in the area of reinsurance. The frequency function can be written, see e.g. Klugman et al. [KPW04, Appendix A.2.3.1],

$$f(y) = \frac{\gamma\alpha^{\gamma}}{(\alpha + y)^{\gamma+1}} \quad y > 0,$$

where $\alpha > 0$ and $\gamma > 0$. Is this distribution an EDM? If so, determine the canonical parameter θ_i and in addition ϕ and the function $b(\cdot)$.

2.8 (Section 2.1) Use well-known results for moment-generating functions, see e.g. Gut [Gu95, Theorem 3.3], to show that if the cumulant-generating function $\Psi(t)$ exists, then $E(Y) = \Psi'(0)$ and $\mathrm{Var}(Y) = \Psi''(0)$.

2.9 (Section 2.1) Derive the variance function of the relative binomial distribution, defined in Exercise 2.6.

2.10 (Section 2.1) Let X and Y be two independent random variables and let c be a constant. Show that the CGF has the following properties

(a) $\Psi_{cX}(t) = \Psi_X(ct)$,
(b) $\Psi_{X+Y}(t) = \Psi_X(t) + \Psi_Y(t)$.

2.11 (Section 2.1) Prove that the sum of independent gamma distributions with the same scale parameter β are again gamma distributed and determine the parameters. *Hint*: use the moment generating function.

2.12 (Section 2.1) Prove Theorem 2.2. *Hint*: Use the result in Exercise 2.10.

2.13 (Section 2.1)

(a) Derive $b'(\theta)$ in (2.12) from the $b(\theta)$ given by (2.10). Then derive $b''(\theta)$ for $p \geq 1$.
(b) Show that $h(\mu)$ in (2.13) is the inverse to $b'(\theta)$ in (2.12).
(c) Show that an EDM with the cumulant function $b(\theta)$ taken from (2.10) has variance function $v(\mu) = \mu^p$, so that it is a Tweedie model.

2.14 (Section 2.2) Use a multiplicative model in the moped example, Example 1.2, and write down the design matrix $\mathbf{X}$ for the list form that is implicit in Table 1.2. For the sake of simplicity, use just three zones, instead of seven.

2.15 (Section 2.2*) Derive the canonical link for the relative binomial distribution, defined in Exercise 2.6 and treated further in Exercise 2.9.

2.16 (Section 2.3) Derive (2.24) and differentiate ℓ with respect to each parameter to show that the ML equations are given by (1.6)–(1.8).

2.17 (Section 2.3) Derive the system of equations in (2.35) directly, in the special case with just two rating factors, by maximizing the log-likelihood. *Hint*: use (2.5) and (2.23).

http://www.springer.com/978-3-642-10790-0

Non-Life Insurance Pricing with Generalized Linear Models
Ohlsson, E.; Johansson, B.
2010, XIII, 174 p., Softcover
ISBN: 978-3-642-10790-0