AStA Adv Stat Anal (2012) 96:187–224
DOI 10.1007/s10182-011-0152-7

ORIGINAL PAPER

Statistical concepts of *a priori* and *a posteriori* risk classification in insurance

Katrien Antonio · Emiliano A. Valdez

Received: 24 September 2010 / Accepted: 9 January 2011 / Published online: 2 February 2011

Abstract Everyday we face all kinds of risks, and insurance is in the business of providing us a means to transfer or share these risks, usually to eliminate or reduce the resulting financial burden, in exchange for a predetermined price or tariff. Actuaries are considered professional experts in the economic assessment of uncertain events, and equipped with many statistical tools for analytics, they help formulate a fair and reasonable tariff associated with these risks. An important part of the process of establishing fair insurance tariffs is risk classification, which involves the grouping of risks into various classes that share a homogeneous set of characteristics allowing the actuary to reasonably price discriminate. This article is a survey paper on the statistical tools for risk classification used in insurance. Because of recent availability of more complex data in the industry together with the technology to analyze these data, we additionally discuss modern techniques that have recently emerged in the statistics discipline and can be used for risk classification. While several of the illustrations discussed in the paper focus on general, or non-life, insurance, several of the principles we examine can be similarly applied to life insurance. Furthermore, we also distinguish between *a priori* and *a posteriori* ratemaking. The former is a process which forms the basis for ratemaking when a policyholder is new and insufficient information may be available. The latter process uses additional historical

Katrien Antonio acknowledges financial support from NWO through a Veni 2009 grant.

K. Antonio (✉)
University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands
e-mail: k.antonio@uva.nl

K. Antonio
Katholieke Universiteit Leuven, Naamsestraat 69, 3000 Leuven, Belgium

E.A. Valdez
Department of Mathematics, University of Connecticut, 196 Auditorium Road, Storrs, CT 06269-3009, USA
e-mail: emiliano.valdez@uconn.edu

information about policyholder claims when this becomes available. In effect, the resulting *a posteriori* premium allows one to correct and adjust the previous *a priori* premium making the price discrimination even more fair and reasonable.

Keywords Actuarial science · Regression and credibility models · Bonus–Malus systems

1 Introduction

Facing risks, or unexpected events, is part of our everyday life. A wide range of situations can be enumerated to illustrate this statement. Rain on your wedding day, a traffic jam when you are already late for an appointment or a black fly in your Chardonnay are examples of pernicious scenarios that simply cause some annoying irritation, but which may also be classified as risks. On the other hand, risks resulting into more tragic disaster such as a fire, an accident or unemployment, illustrate examples that can have a huge impact on one's economic and personal situation.

Many of us are aware of these risks we face everyday, especially those with possible huge economic impact. There is some level of risk aversion in all of us in the sense that to avoid being affected when these risks occur, we look for possible ways to transfer some or all of these risks to other economic agents in the market willing to assume them. In light of this risk aversion brought the birth of a whole sector of activities that we find today willing to provide us the needed protection: the business of insurance. Risk compensation by grouping similar, independent risks forms the basis of actuarial practice: *"The contributions of the many to the misfortunes of the few"* as the motto of Lloyd's of London says.

The discipline of actuarial science deals with uncertain events where clearly the concepts of probability and statistics provide for an indispensable instrument in the measurement and management of risks. An important aspect of the business of insurance is the determination of the price, typically called premium but preferably called tariff in this paper, to pay in exchange for the transfer of risks; it is the job of the actuary to evaluate a fair price given the nature of the risk. In this article, we provide for a survey and discussion of contemporary statistical techniques that can be practically implemented for pricing risks through ratemaking based on risk classification.

1.1 Ratemaking and risk classification

Clearly then within the actuarial profession, a major challenge can be found in the measurement and construction of a fair tariff structure. This is the objective of a *ratemaking* process. Pricing risks based upon certain specific characteristics has a long history in actuarial science, e.g. McClenahan (2001) observed that 18th century fire insurance rates for dwellings in the United States were based upon roof type and basic construction. Premium rates for marine insurance, believed to be the oldest form of insurance, are based on characteristics of the design and built-in protection of each ship and several ships vary in design.

Indeed, in light of the heterogeneity within an insurance portfolio, an insurance company should not apply the same premium for all insured risks in the portfolio.

Otherwise the so-called concept of *adverse selection* could undermine the solvency of the company and possibly lead to the collapse of the insurance market. On one hand, 'good' risks, with low risk profiles, could pay too much and eventually, prefer to leave the company. On the other hand, 'bad' risks may find a uniform tariff to be in their favor and therefore, prefer to stay with the company. This could lead to a spiral effect where the insurer could end up with a disproportionate number of 'bad' risks in its portfolio and to remain solvent, it may have to keep increasing its premium rates. Therefore, it is important for the insurer to optimally group the risks in the portfolio so that those insureds with similar risk profile pay the same reasonable premium rate. Such is the idea behind *risk classification* within the ratemaking process. A risk classification system should not only allow insurers to price discriminate their products in a fair and equitable manner, but should also be constructed based on a sound statistical basis.

To construct a tariff structure that reflects the various risk profiles in a portfolio within a reasonable and statistically sound basis, actuaries usually rely on *regression* techniques. Such techniques allow for the inclusion of various explanatory (also called classifying or rating) variables so that the actuary is able to construct risk classes with more or less similar risk profiles. For non-life (also called: property and casualty or general) insurance, typical response variables in these regression models are the number of claims (or claim frequency) per unit of exposure on the one hand, and the corresponding amount of loss, given a claim (or claim severity) on the other hand. A formal discussion of these actuarial terminologies will follow in Sect. 2.1. That section also explains how regression models for claims frequency and severity allow us to estimate the price of risk.

Different classifying variables impacting either frequency or severity (or both) can be found in all forms of insurance. For example, in automobile insurance, a non-life product, it is typical to find the following classifying variables used: age-gender-marital status, use of the car, geography (location of garage) and other factors such as whether the vehicle is a sports car or not. The cost of claims may be influenced among others by factors such as the use of the car (more driving implies more exposure to claims, driving conditions: time of day, weather, area), driving ability (experience and training, reaction time, eyesight and hearing, condition of the car, driving style), interaction with the claims mechanism and the extent of damages (crashworthiness of the car where certain brands are able to withstand severity of accidents better than others, use of safety devices). See Finger (2001) for further discussion. In fire insurance, studies have shown that restaurants have a higher frequency of accidents than stores; the presence of a sprinkler system, the value of the building and contents being insured all can impact the amount of damage in the event of a fire. For workers' compensation, a form of insurance that provides protection for injuries in the course of employment, statistics show significant difference in claims for various sectors of employment, e.g. manufacturing versus education, with employees in manufacturing firms indicating larger claims frequencies. While several characteristics may impact both frequency and severity, some affect severity but not frequency, and vice versa. To illustrate, the presence of a sprinkler system may not affect the frequency of claims but clearly affects the severity. Finally, risk classification systems are also found in life insurance. During the underwriting process for the purchase of life insurance, the insurer collects information on the applicant's risk factors (e.g. age, gender, smoking habits, occupation, any dangerous hobbies, and personal and family health history) through a questionnaire and possibly, a medical examination. The information is then used to classify policyholders into risk classes and to price their policies accordingly. See Dickson et al. (2009) for a discussion. However, risk classification for life insurance products is beyond the scope of this article.

In the construction of a risk classification system, the statistical considerations, often referred to as the actuarial criteria, are but one of several, but of utmost importance, criteria for selecting classifying or rating variables. This survey paper focuses on these criteria. Here, we ensure that the classifying variable must be considered to: (1) be accurate in the sense that it has a direct impact on costs, (2) meet homogeneity requirement in the sense that the resulting expected costs within a class are reasonably similar, (3) be statistically credible and reliable. Other considerations for the selection of variables are practical or operational implementation, "social acceptability", and legal considerations. To illustrate for example, the use of "gender", even if statistically sound, as a classifying variable may be restricted for certain forms of insurance because sex discrimination is prohibited according to some constitutions. For further discussion of these criteria, we refer the reader to Finger (2001).

1.2 *A priori* and *a posteriori* ratemaking

When the explanatory variables used as rating factors express *a priori* correctly measurable information about the policyholder (or, for instance, the vehicle or the insured building), the system is said to be an *a priori* classification scheme. A discussion of *a priori* rating will follow in Sect. 2. However, an *a priori* system is unable to identify all the possible important factors because some of them are either unmeasurable or unobservable. Take the case for example of automobile insurance where the insurer is unable to detect the driver's aggressiveness behind the wheel or the quickness of his reflexes to avoid a possible accident. Thus, tariff cells within an *a priori* rating system will never be completely homogeneous. For that reason, an *a posteriori* or an *experience* rating system is necessary to allow for the re-evaluation of the premium by taking into account the history of claims of the insured as it becomes available. The statistical philosophy behind this is that the best (or optimal in some sense) predictor for the future number of claims that an insured will report is conditionally based on the number of claims reported in the past. The actuarial *credibility* systems discussed in Sects. 3.1 and 3.2 are examples of *a posteriori* rating systems that take into account the history of claims as it emerges for an individual risk. Commercial versions of these experience rating schemes are more widely known in practice as *Bonus–Malus* scales. Rating schemes according to these Bonus–Malus scales is the topic of Sect. 3.3.

1.3 Statistical techniques for risk classification

Since the early development of constructing *a priori* and *a posteriori* rating schemes, ordinary regression techniques based on the assumption of Normal data has been the standard norm in practice. See, for example, Lemaire (1985). Several papers based on this technique have appeared in the actuarial literature, but actuaries have realized that insurance data usually violate the assumption of the Normal distribution. Recent advances in statistical modeling, spurred both by the availability of more data and computing technologies to analyze these data, have emerged. The applications of these up-to-date statistical techniques in the analysis of insurance data have provided avenues for further developing skills for actuarial modeling. This article highlights some of these advances in the literature and those that are slowly crawling up in practice. Within the context of *a priori* ratemaking, for example, it is becoming a standard norm in practice to use Generalized Linear Models (GLMs) where data is modeled within the class of exponential dispersion distributions. Other advances, which include Generalized Additive Models (GAMs), regression models based on generalized count distributions and heavy-tailed regression models for claims severity, can be used in the context of *a priori* ratemaking. These topics are discussed in Sect. 2.

In Sect. 3, we emphasize our discussion on risk classification and *a posteriori* ratemaking. Several of the statistical models discussed in this section cover topics that are lesser known in practice, for example, models for clustered data (panel data and multilevel data) are discussed. In addition, we also consider the two types of model estimation: likelihood-based and Bayesian estimation methods. The latter method has the advantage of allowing the analyst to construct a full predictive distribution of quantities of interest. In our discussion of Bonus–Malus schemes, some concepts of Markov chains necessarily appear because of the possible transition through the various Bonus–Malus scales.

We present numerous examples based on real-life actuarial data throughout the paper to illustrate several of the methodologies. These illustrations can help practitioners implement these techniques in practice. Denuit et al. (2007) provides for an excellent and comprehensive reference on several aspects of *a priori* and *a posteriori* risk classification, with an emphasis on claim frequency. Frees (2010) is a useful reference that contains several case-studies to illustrate statistical regression models for insurance rating.

2 Regression models for *a priori* risk classification

Regression techniques are indispensable tools for the pricing actuary. Here in this section, we focus on two types of response variables that are important in pricing for short term insurance contracts: the number of claims (or claim frequency) per unit of exposure and the amount of claim given a claim occurs (or claim severity). In Sect. 2.1 definitions follow of actuarial concepts relevant in ratemaking (e.g. exposure, frequency and severity). It is also explained how risk classification models for both frequency and severity are combined to calculate a so-called pure premium. Section 2.2 presents current industry practice for risk classification based on generalized linear models. More advanced statistical methods are illustrated in Sect. 2.3.

2.1 Frequency, severity and pure premium for cross-sectional data

Some actuarial concepts, crucial for ratemaking, are discussed below. See McClenahan (2001) and Frees (2010) for further details. The basic rating unit underlying an insurance premium is called a unit of *exposure*. In our examples from automobile insurance, earned exposure is used and this refers to the fraction of the year for which premium is paid and therefore coverage is provided. Another example is workers compensation where company payroll is typically used as exposure base. When an insured demands for a payment under the terms and conditions of an insurance contract, this is referred to as a *claim*. Claim *frequency* refers to the number of times a claim is made during a (calendar) year and is expressed in terms of frequency per exposure unit. The amounts paid to a claimant under the terms of an insurance contract are called *losses* and *severity* is the term used for the amount of loss per claim.

Insurers typically keep track of frequency and severity data in separate files. In the *policyholder file*, underwriting information is registered about the insured (e.g. age, gender, policy information such as coverage, deductibles and limitations) and additional information may be kept about the claims event. In the *claims file*, information is recorded about the claims filed to the insurer together with the amounts and payments made. With *a priori* rating each individual risk is priced based on the history of frequency and severity observed usually from a cross-sectional data set. Cross-sectional means that the data base contains information from N policyholders, but the time series structure of the data is ignored.

Typically from these files, we find that for each insured i the observable responses are:

- N_i: the number of claims and the total period of exposure E_i during which these claims were observed; and
- C_{ij} the loss corresponding to each claim made (with $j = 1, \ldots, N_i$).

The set $\{C_{ij}\}$ is empty when $N_i = 0$. The so-called aggregate loss L_i is defined as $L_i := C_{i1} + \cdots + C_{iN_i}$ and refers to the total amount of claims paid during the period of observation. The data available then to the insurer for *a priori* rating typically have one of the following formats:

(1) $\{N_i, E_i, C_{i1}, \ldots, C_{iN_i}\}$, thus the vector of individual losses $\boldsymbol{C}_i := (C_{i1}, \ldots, C_{iN_i})'$ is registered;
(2) $\{N_i, E_i, L_i\}$, only aggregate losses are available.

Our ultimate goal is to price the risk using *a priori* measurable characteristics. To achieve that, the frequency and severity data will be combined into a *pure premium* P_i defined as

$$P_i = \frac{L_i}{E_i} = \frac{N_i}{E_i} \times \frac{L_i}{N_i} \tag{1}$$

$$= F_i \times S_i, \tag{2}$$

with F_i, the claim frequency per unit of exposure and S_i, the severity. Depending on the format of the available data, regression models for F_i and the individual losses $\{C_{i1}, \ldots, C_{iN_i}\}$ (format (1)) or F_i and the severity S_i (format (2)) will be specified. Ultimately, policy i is priced by applying a *premium principle* $\pi(\cdot)$ to the random variable P_i. In this paper, we essentially focus on the expected value principle, which leads us to the net premium:

$$\pi(P_i) = E[P_i] = E[F_i] \times E[C_{ij}] \quad \text{with format (1), or} \tag{3}$$

$$\pi(P_i) = E[P_i] = E[F_i] \times E[S_i] \quad \text{with format (2),} \tag{4}$$

under the (traditional) assumption of (for format (1)) independent and identically distributed individual losses $\{C_{ij}\}$ and independence between claim frequency and losses, and (for format (2)) independence between claim frequency and severity. Here, the term 'net premium' is often referred to that portion of the premium that considers only the benefits to be paid under the terms of the contract. Typically, insurers add a 'risk loading' to cover other items: administrative expenses, profits, margins for contingencies. If a 'risk loading' is added to the net premium, the term 'contract premium' or 'gross premium' would be used. Other premium calculation principles may be used, but these are beyond the scope of this paper. See, for example, Kaas et al. (2008).

2.2 Current practice: generalized linear models (GLMs)

GLMs are nowadays standard industry practice for pricing risks and this topic has been an addition to the syllabus for actuarial examinations in several countries. The paper by Haberman and Renshaw (1996) provides an overview of its applications in actuarial science. Additional discussion of GLMs with an actuarial bent can be found in de Jong and Heller (2008), Frees (2010) and Kaas et al. (2008). The early formulation of GLMs in the statistics literature can be found in Nelder and Wedderburn (1972).

GLMs extend the framework of ordinary (normal) linear models to the class of distributions derived from the exponential. A whole variety of possible outcome measures such as counts, binary and skewed data that are relevant in actuarial science, can be modeled within this framework. The canonical specification of densities from the exponential family can be expressed as

$$f(y) = \exp\left[\frac{y\theta - \psi(\theta)}{\phi} + c(y, \phi)\right], \tag{5}$$

where $\psi(\cdot)$ and $c(\cdot,\cdot)$ are known functions, θ and ϕ are the natural and scale parameters, respectively. Members belonging to this family include, but not limited to, the Normal, Poisson, Binomial and the Gamma distributions. Let $Y_1, \ldots, Y_n$ be independent random variables with a distribution from this family. The following well-known relations hold for these distributions:

$$\mu_i = \mathrm{E}[Y_i] = \psi'(\theta_i) \quad \text{and} \quad \mathrm{Var}[Y_i] = \phi\psi''(\theta_i) = \phi V(\mu_i), \tag{6}$$

where the derivatives are with respect to θ and $V(\cdot)$ is referred to as the variance function. This function captures the relationship, if any exists, between the mean and variance of Y.

2.2.1 Claim frequency models

When the actuary is interested in risk classification systems for claim frequencies, the following two GLMs are of special interest: Poisson regression and Negative Binomial (NB) regression. Both distributions belong to the framework specified in (5) and (6) (see e.g. Kaas et al. 2008 for a proof).

Poisson distribution Ratemaking with a Poisson regression model—as an archetype of GLM—is illustrated here with a case-study from automobile insurance. This example will be reconsidered in Illustrations 2.2–2.6.

Illustration 2.1 (Poisson distribution for claim counts) Claim counts are modeled for an automobile insurance data set with 159,947 policies. The response variable is the total number of claims registered for each insured vehicle in the data set. Next to a set of explanatory variables, an exposure variable is available which reflects the period during which premiums are paid and the claim counts are registered. In this example, the exposure period is expressed in years. Total exposure is 101,914 years. The data are overdispersed: the empirical variance exceeds the empirical mean (exposure taken into account). Table 1 illustrates the fit of the Poisson distribution to the 'raw' data (i.e. no regression taken into account). Under these assumptions $\Pr(N_i = n_i) = \frac{\exp(-\lambda_i)\lambda_i^{n_i}}{n_i!}$ and $\lambda_i = e_i \exp(\beta_0)$ with e_i the exposure and n_i the claim counts registered for policyholder i. We reconsider these data again in Illustrations 2.2, 2.3, 2.4, 2.5 and 2.6 where the fit of other count distributions as well as the construction of regression models for risk classification are discussed.

Negative Binomial distribution A generalization of the Poisson distribution, formulated as a continuous mixture, is the NB distribution. A mixture model is a common method to deal with heterogeneity and the overdispersion resulting from it, which is often apparent in actuarial data on claim counts. This approach will be re-introduced in the discussion of *a posteriori* ratemaking in Sect. 3, where random effects are used

Table 1 Empirical distribution and Poisson fit for claim frequencies from the Illustration 2.1 data

Number of Claims	Observed Frequency	Fitted Frequency	
0	145,683	145,141	
1	12,910	13,902	
2	1,234	863	
3	107	39	
4	12	1.4	
5	1	0.04	
Mean	0.1546	−2 log Lik.	101,668
Variance	0.1628	AIC	101,670

to deal with neglected or unobservable covariates at various levels. The NB distribution is constructed as

$$\Pr(Y = y|\boldsymbol{\theta}) = \int_0^{\infty} \frac{\exp(-\lambda)\lambda^y}{y!} f(\lambda|\boldsymbol{\theta})\, d\lambda, \tag{7}$$

with λ varying stochastically according to a $\Gamma(\tau, \tau/\mu)$-distribution.[1] This yields to the NB distribution of the form

$$\Pr(Y = y|\mu, \tau) = \frac{\Gamma(y+\tau)}{y!\Gamma(\tau)} \left(\frac{\tau}{\mu+\tau}\right)^{\tau} \left(\frac{\mu}{\mu+\tau}\right)^{y}. \tag{8}$$

Under (8), $\mathrm{E}[Y] = \mu$ and $\mathrm{Var}[Y] = \mu + \frac{\mu^2}{\tau}$, thus overdispersion is clearly present. It can be shown that the NB distribution belongs to the exponential family and as such is an example of a GLM.

Illustration 2.2 (Negative binomial distribution for claim counts) The data from Illustration 2.1 are reconsidered. We extend the analysis by fitting the NB distribution to the raw data. As demonstrated in Table 2, the NB distribution provides for an improved fit of the empirical distribution of the claim frequencies.

Risk classification Including risk factors in the Poisson or NB distribution allows one to build classification systems for the frequency component of the data. This is done with regression techniques. The expressions given below are for cross-sectional data with sample size m.

– *Poisson regression*: specify a log–linear structure for the mean, namely $\lambda_i = \exp(\boldsymbol{x}_i'\boldsymbol{\beta})$, then

$$\log \mathcal{L}(\boldsymbol{\beta}) = \sum_{i=1}^{m} \left\{ n_i \boldsymbol{x}_i'\boldsymbol{\beta} - \exp(\boldsymbol{x}_i'\boldsymbol{\beta}) - \log(n_i!) \right\}; \tag{9}$$

Table 2 Empirical distribution and Poisson and NB fit for claim frequencies from the Illustration 2.1 data

Number of Claims	Observed Frequency	Poisson Frequency	NB Frequency
0	145,683	145,141	145,690
1	12,910	13,902	12,899
2	1,234	863	1,225
3	107	39	119
4	12	1.4	12
>4	1	0.04	1
	−2 log Lik.	101,668	101,314
	AIC	101,670	101,318

[1] $X \sim \Gamma(\alpha, \beta)$ means $f_X(x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$.

– *NB regression*: for the specification in (8) $\mathrm{E}[Y]=\mu$ and regression models can be built through $\mu_i=\exp(\boldsymbol{x}_i'\boldsymbol{\beta})$. Or $N_i\sim\mathrm{Poi}(\lambda_i\theta)$ with $\lambda_i=\exp(\boldsymbol{x}_i'\boldsymbol{\beta})$ and $\theta\sim\Gamma(\tau/\mu,\tau/\mu)$, which leads to

$$\Pr[N_i=n_i]=\frac{\Gamma(\alpha+n_i)}{\Gamma(\alpha)n_i!}\left(\frac{\alpha}{\lambda_i+\alpha}\right)^{\alpha}\left(\frac{\lambda_i}{\lambda_i+\alpha}\right)^{n_i}, \tag{10}$$

with $\alpha:=\tau/\mu$. Both approaches are similar.

Using these expressions, maximum-likelihood and Bayesian estimation are straightforward. The advantage of using Bayesian statistics is that one can simulate from the posterior distribution of N_i, conditional on the risk characteristics. Besides the expected value, other summary measures can be calculated for this posterior distribution. This enables the application of a whole range of premium principles.

Illustration 2.3 (Regression models for claim counts) Using covariate information a risk classification system is built for the data from Illustration 2.1. The risk variables used in the construction of this *a priori* model are enumerated in Table 3. Their categorization is in Table 14. Corresponding parameter estimates are in Table 14 at the end of this article. Selected risk profiles and their corresponding a priori premium (i.e. $\mathrm{E}[N_i]$) are in Table 4. Exposure for these profiles is assumed to be one full year. A description of these profiles are given following the table below. Note that the variables 'Age of insured' and 'Driving experience' may be strongly correlated and, for reasons of multicollinearity, it may not be possible to retain both variables in the same regression model. In this example however, both had a significant effect on the response and, for a particular age, insureds with driving experiences between 0 and age-18 years could be observed.

– *Low*: a 45 years old male driver with a driving experience of 19 years and a NCD = 40. He's driving a 1,166 cc Toyota Corolla that is 22 years old. He only has a theft cover. The car is for private use.

Table 3 List of explanatory variables used for the data in Illustration 2.1

Covariate	Description
Vehicle Age	The age of the vehicle in years.
Cubic Capacity	Vehicle capacity for cars and motors.
Tonnage	Vehicle capacity for trucks.
Private	1 if vehicle is used for private purpose, 0 otherwise.
CompCov	1 if cover is comprehensive, 0 otherwise.
SexIns	1 if driver is female, 0 if male.
AgeIns	Age of the insured.
Experience	Driving experience of the insured.
NCD	1 if there is no 'No Claims Discount', 0 if discount is present. This is based on previous accident record of the policyholder. The higher the discount, the better the prior accident record.
TLength	(Exposure) Number of calendar years during which claim counts are registered.

Table 4 A priori risk premium for a selection of risk profiles

Risk Profile	Poisson Distribution	NB Distribution
Low	0.0460	0.0454
Medium	0.1541	0.1541
High	0.3727	0.3732

Table 5 A summary of some severity distributions considered in this section

Distribution	Density	Conditional Mean
Gamma	$f(y)=\frac{1}{\Gamma(\alpha)}\beta^{\alpha}y^{\alpha-1}e^{-\beta y}$	$\mathrm{E}[Y]=\frac{\alpha}{\beta}=\exp(\boldsymbol{x}'\boldsymbol{\gamma})$
Inverse Gaussian	$f(y)=\left(\frac{\lambda}{2\pi y^3}\right)^{1/2}\exp\left[\frac{-\lambda(y-\mu)^2}{2\mu^2 y}\right]$	$\mathrm{E}[Y]=\mu=\exp(\boldsymbol{x}'\boldsymbol{\gamma})$
Lognormal	$f(y)=\frac{1}{\sqrt{2\pi}\sigma y}\exp\left[-\frac{1}{2}\left(\frac{\log y-\mu}{\sigma}\right)^2\right]$	$\mathrm{E}[Y]=\exp\left(\mu+\frac{1}{2}\sigma^2\right)$ with $\mu=\exp(\boldsymbol{x}'\boldsymbol{\gamma})$

- *Medium*: a 43 years old male driver with a driving experience of 11 years and a NCD = 50. He's driving a 1,995 cc Nissan Cefiro that is 2 years old. He has a comprehensive cover and the car is for private use.
- *High*: a 21 years old male driver with a driving experience of 3 years and a NCD = 0. He's driving a 1,597 cc Nissan that is 4 years old. His cover is comprehensive and the car is for private use.

2.2.2 Claim severity models

Actuarial data on severities are (usually) positive and (very often) are skewed to the right, exhibiting a long right tail. Distributions from the exponential family suitable to model severity data are the Gamma and Inverse Gaussian distribution. After applying a log transformation to the losses, the Normal distribution is popular as well among practitioners. Specification of these distributions as members of the exponential family is documented in Kaas et al. (2008). In Illustration 2.4 risk classification based on these severity distributions is demonstrated. Covariate information is incorporated in the likelihood specifications according to the table summarized below.

Severity distributions from the GLM framework can be used to fit the whole data set with a single distribution. However, very often, when claim sizes are modeled, right tails heavier than those from the GLM framework are encountered. It is then useful to specify a mixture of distributions; one for the body of the data and another one for the tail. For the latter extreme value methods can be used (e.g. using Pareto-type modeling of the tail). See Beirlant et al. (2004) for an example demonstrating its usefulness using losses from a reinsurance company.

Illustration 2.4 (Severity models) We consider the severities corresponding to the data analyzed in Illustration 2.1. These severities are obtained as the ratio "aggregate loss/total number of claims" for each policyholder i and for each time period.

Table 6 Descriptive statistics of empirical distribution automobile insurance severity data

Mean	StdDev	Min	10%	25%	50%	75%	90%	97.5%	99.5%	Max
4,439	7,100	1	610	1,229	2,575	5,337	9,835	19,205	37,344	306,304

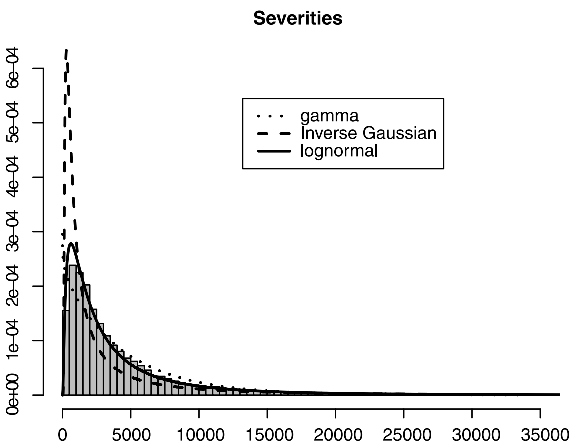

Fig. 1 Automobile insurance severity data; histogram and densities of gamma, Inverse Gaussian and lognormal

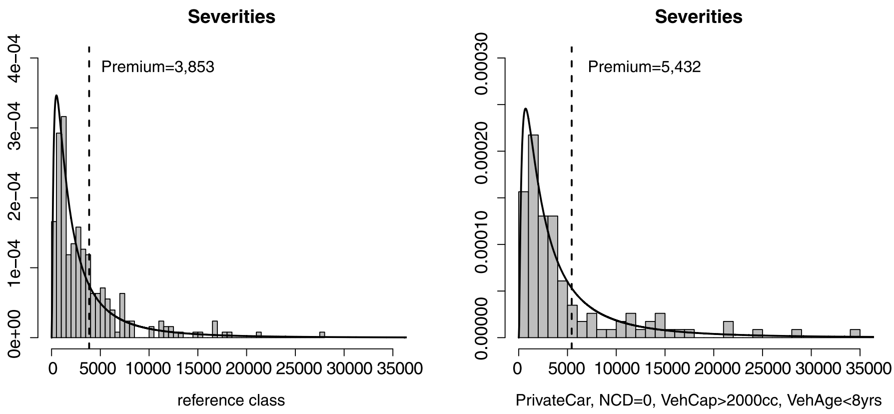

Fig. 2 Automobile insurance severity data; histogram, fitted Lognormal density and risk premium for selected groups of insureds

Characteristics of these severities are displayed in Table 6. Ignoring risk classifying information, the quality of the fit of the Gamma, Inverse Gaussian and Lognormal distributions to the data is illustrated in Fig. 1.

The results of including risk classifying characteristics in these severity models are tabulated in Table 15. Figure 2 compares the histogram of severities with the fitted Lognormal density for groups of insureds (based on risk classifying characteristics). For each group, the corresponding risk premium—determined as the expected value of the Lognormal distribution—is indicated with a vertical line.

2.3 Advanced models

Empirical distributions of claim counts often reveal an inflated number of zeros (i.e. no claim reported) and overdispersion (i.e. variance exceeding the mean). To cope with these characteristics, regression models beyond the Poisson and NB regression models are discussed in Sect. 2.3.1. In addition, a typical feature of the empirical distribution of claim severities is its right-skewness and long- or heavy-tailed character. To appropriately model such response variables, distributions outside the exponential family may be necessary. We provide two examples of flexible, parametric families of distributions, namely the Burr XII and Generalized Beta of the Second Kind (GB2). This is the topic of Sect. 2.3.2.

2.3.1 Frequency models: generalized count distributions

The recent works of Yip and Yau (2005) and Boucher et al. (2007) in the actuarial literature highlight the use of parametric distributions other than the Poisson and the NB distributions to accommodate specific features of actuarial data. Cameron and Trivedi (1998), Winkelmann (2003), Yau et al. (2003), and Lee et al. (2006) similarly discuss regression modeling of data with excess zeros found in econometrics and medical statistics.

Mixtures The NB distribution was our first example of a mixture of the Poisson distribution in this paper. Other continuous mixtures of the Poisson distribution that have been studied in the actuarial literature include, among others, the Poisson-Inverse Gaussian ('PIG') distribution and the Poisson-LogNormal ('PLN') distribution. These and other types of mixture models may be found in Panjer and Willmot (1992).

Zero-inflation The use of a discrete or finite mixture to model count data, like a zero-inflated Poisson or a zero-inflated NB distribution, recently gained popularity in actuarial statistics. See, for example, Yip and Yau (2005), Boucher et al. (2007) for cross-sectional and Boucher et al. (2009) for longitudinal data. A zero-inflated distribution is a mixture of a standard count distribution with a degenerate distribution concentrated at zero. Its use is primarily motivated by the inflated number of zeros (i.e. no claims reported) commonly found in actuarial data for claim count statistics.

Say $\Pr(Y=y|\boldsymbol{\theta})$ is the standard count distribution (e.g. Poisson or NB) and $p \in (0,1)$ denotes the extra proportion of zeros, then

$$Y \sim \begin{cases} 0 & \text{with probability } p, \\ \Pr(Y=y|\boldsymbol{\theta} & \text{with probability } 1-p. \end{cases} \tag{11}$$

This gives the following distributional specification (with 'ZI' for zero-inflated)

$$\Pr_{\text{ZI}}(Y=y|p,\boldsymbol{\theta}) = \begin{cases} p+(1-p)\Pr(Y=0|\boldsymbol{\theta}), & y=0, \\ (1-p)\Pr(Y=y|\boldsymbol{\theta}), & y>0. \end{cases} \tag{12}$$

In countries where experience rating systems are operational, reporting a claim can generally increase the insurance premium in the following years. Policyholders therefore have the tendency to avoid reporting all incurred claims, which motivates the presence of inflated probability mass at zero. Zero-inflated distributions take this into account.

Hurdle models The so-called *hurdle* models (Mullahy 1986) provide another possibility to deal with extra zeros in empirical data. Hurdle models are two-part models that will be used here with the hurdle at zero. As such we discriminate between policies with a claim and policies without a claim. Zeros are then modeled as a binary component and non-zero observations with a truncated count distribution or a count distribution defined for strictly positive integers. Thus, when a truncated count distribution is used over the hurdle, we get

$$\begin{aligned}\mathrm{Pr}_{\mathrm{Hur}}(Y=0|p,\boldsymbol{\theta}) &= p,\\ \mathrm{Pr}_{\mathrm{Hur}}(Y=y|p,\boldsymbol{\theta}) &= \frac{1-p}{1-\mathrm{Pr}(0|\boldsymbol{\theta})}\mathrm{Pr}(Y=y|\boldsymbol{\theta}), \quad y>0,\end{aligned} \tag{13}$$

where p is the probability of zero claims and $\mathrm{Pr}(\cdot)$ is the standard count distribution from which the truncated form $\frac{1}{1-\mathrm{Pr}(0)}\mathrm{Pr}(Y=y)$ is derived. For a count distribution specified on the strictly positive integers, one similarly gets

$$\begin{aligned}\mathrm{Pr}_{\mathrm{Hur}}(Y=0|p,\boldsymbol{\theta}) &= p,\\ \mathrm{Pr}_{\mathrm{Hur}}(Y=y|p,\boldsymbol{\theta}) &= (1-p)\mathrm{Pr}(Y=y), \quad y>0.\end{aligned} \tag{14}$$

$\mathrm{Pr}(\cdot)$ now denotes a count distribution with the strictly positive integers as support.

As stated in Boucher et al. (2007), the belief that insureds behave differently when they already have reported a claim (as compared to when they are still claim free) motivates the use of a hurdle model with the hurdle at zero.

Illustration 2.5 (Generalized count distributions for claim counts) We now fit the generalized count distributions discussed in this section to the data set introduced in Illustration 2.1. Regression models based on these distributions are discussed in Illustration 2.6.

Risk classification The inclusion of risk factors in the count distributions from Sect. 2.3.1 is done with regression techniques. The expressions given below are for cross-sectional data with sample size n.

– *Zero-inflated Poisson regression*: use again a log-linear structure for the mean of the Poisson part and a logistic regression for the extra zeros. Thus,

$$\begin{aligned}\log\mathcal{L}(\boldsymbol{\beta},\boldsymbol{\gamma}) &= \sum_{i=1}^{n}\big\{u_i\log\big(p_i+(1-p_i)\exp(-\lambda_i)\big)\big\}\\ &\quad+\sum_{i=1}^{n}(1-u_i)\log\left((1-p_i)\frac{\exp(-\lambda_i)\lambda_i^{y_i}}{y_i!}\right),\end{aligned} \tag{15}$$

Table 7 Empirical distribution and Negative Binomial, zero-inflated Poisson and hurdle Poisson fit for claim frequencies corresponding to the data introduced in Illustration 2.1

Number of Claims	Observed Frequency	NB Frequency	ZIP Frequency	Hurdle Poisson Frequency
0	145,683	145,690	145,692	145,683
1	12,910	12,899	12,858	13,161
2	1,234	1,225	1,295	1,030
3	107	119	96	69
4	12	12	6	4
>4	1	1	0.28	0.18
	$-2\log$ Lik.	101,314	101,326	105,910
	AIC	101,318	101,330	105,914

where $p_i = \frac{\exp(z_i'\boldsymbol{\gamma})}{1+\exp(z_i'\boldsymbol{\gamma})}$, $\lambda_i = \exp(\boldsymbol{x}_i'\boldsymbol{\beta})$ and $u_i = I(y_i = 0)$. The log-likelihood then becomes

$$\log\mathcal{L}(\boldsymbol{\beta},\boldsymbol{\gamma}) = \sum_{i=1}^{n} u_i \log\big(\exp(z_i'\boldsymbol{\gamma}) + \exp\big(-\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\big)\big) + \sum_{i=1}^{n}(1-u_i)\big(y_i\boldsymbol{x}_i'\boldsymbol{\beta} - \exp(\boldsymbol{x}_i\boldsymbol{\beta})\big) - \sum_{i=1}^{n}\big\{\log\big(1+\exp(z_i'\boldsymbol{\gamma})\big) - (1-u_i)\log(y_i!)\big\}. \quad (16)$$

– *Hurdle Poisson regression:*

$$\log\mathcal{L}(\boldsymbol{\beta},\boldsymbol{\gamma}) = \sum_{i=1}^{n}\big\{u_i\log(p_i) + (1-u_i)\log(1-p_i)\big\} + \sum_{i=1}^{n}(1-u_i)\big[y_i\log(\lambda_i) - \log\big(1-\exp(-\lambda_i)\big) - \log(y_i!) - \lambda_i\big], \quad (17)$$

where similarly $p_i = \frac{\exp(z_i'\boldsymbol{\gamma})}{1+\exp(z_i'\boldsymbol{\gamma})}$ and $\lambda_i = \exp(\boldsymbol{x}_i'\boldsymbol{\beta})$.

Illustration 2.6 (Regression models for claim counts) Illustration 2.3 is continued. Table 14 shows parameter estimates for a risk classification system based on the ZIP. Corresponding risk premiums are in Table 8.

Table 8 A priori risk premium for a selection of risk profiles

Risk Profile	Poisson distribution	NB distribution	ZIP distribution
Low	0.0460	0.0454	0.0455
Medium	0.1541	0.1541	0.1537
High	0.3727	0.3732	0.3715

2.3.2 Flexible, parametric models for claim severity

There are peculiar characteristics of insurance claim amounts such as skewness and heavy-tailness that usually cannot be accommodated by classes of distributions belonging to the exponential family. Thus, modeling the severity of claims as a function of their risk characteristics, in the form of covariate information, may require statistical distributions outside those belonging to the GLM class. Principles of regression within other flexible parametric families of distributions are illustrated in this section. In particular, we find the GB2 class of distributions to be indeed quite flexible. Introduced in economics to model distributions of income by McDonald (1984), the GB2 class has four parameters that allow for this extreme flexibility and is able to accommodate covariates as used in Sun et al. (2008). A special case of the GB2 class is the Burr Type XII distribution. The work of Beirlant et al. (1998) introduces Burr regression in the actuarial literature.

Other flexible parametric models also appear in Klugman et al. (2008), many of which are tabulated in the appendix of the book. Several of these loss distributions can be similarly extendable for regression purposes. All these parametric models usually try to fit the whole body of the data with a single distribution. In Sect. 2.2.2 an alternative approach was mentioned, mixing a distribution for the body with one for the tail.

Illustration 2.7 (Fire insurance portfolio) The cumulative distribution functions for the Burr Type XII and the GB2 distribution are given, respectively by

$$F_{\text{Burr},Y}(y) = 1 - \left(\frac{\beta}{\beta + y^{\tau}}\right)^{\lambda}, \quad y > 0,\ \beta, \lambda, \tau > 0, \tag{18}$$

and

$$F_{\text{GB2},Y}(y) = B\left(\frac{(y/b)^a}{1 + (y/b)^a}; p, q\right), \quad y > 0, a \neq 0, b, p, q > 0, \tag{19}$$

where $B(\cdot,\cdot)$ is the incomplete Beta function. If the available covariate information is denoted by $\mathbf{x}$ $(1 \times p)$, it is straightforward to allow one or more of the parameters in (18) or (19) to vary with $\mathbf{x}$. The result can be called a Burr or a GB2 regression model.

As an illustration of this approach, we consider a fire insurance portfolio (see Beirlant et al. 1998) that consists of 1,823 observations. We want to assess how the loss distribution changes with the sum insured and the type of building. Claims expressed as a fraction of the sum insured (‘SI’) are used as the response variable. Explanatory

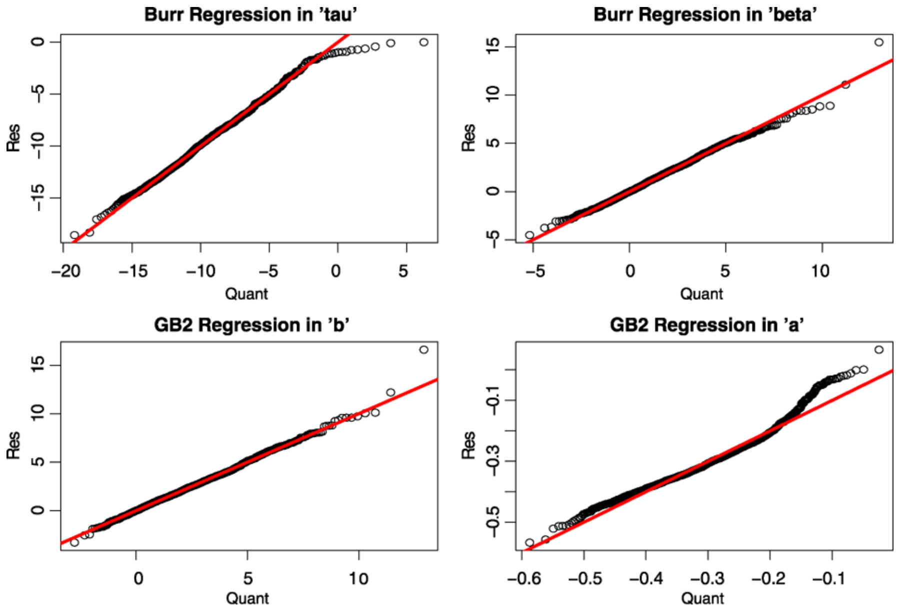

Fig. 3 Fire Insurance Portfolio: residual QQ plots for Burr and GB2 regression

variables are the type of building and the sum insured. Residual Q–Q plots like those in Fig. 3 are used to judge the goodness-of-fit of the proposed regression models. Details about the construction of these residual Q–Q plots can be found in the above-mentioned references. A summary of the resulting parameter estimates together with their respective standard errors are tabulated in Table 9.

2.3.3 Additive regression models

Thus far, only regression models with a linear structure for the mean, a transformation of the mean or a parameter in the distribution have been considered. Generalized additive models (GAMs) allow for more flexible relations between a response and a set of covariates. Without going into details, Fig. 4 shows the additive effect of the age of the vehicle ('VAge'), its cubic capacity ('VehCapCubic') and the age of the driver ('AgeInsured') and his driving experience ('Experience') in a Poisson additive model with predictor

$$\begin{aligned}\log \mu_i = \eta_i &= \text{Exposure} + \beta_0 + \beta_1 * I(\text{Sex} = \text{F}) + \beta_2 * I(\text{NCD} = 0)\\ &\quad + \beta_3 * I(\text{Cover} = \text{C}) + \beta_4 * I(\text{Private} = 1) + f_1(\text{VAge})\\ &\quad + f_2(\text{VehCapCubic}) + f_3(\text{Experience}) + f_4(\text{AgeInsured}). \quad (20)\end{aligned}$$

The model was fitted to the data from Illustration 2.1. GAMs are an alternative for GLMs. As industry practice requires, our examples of GLMs use categorizations of continuous risk factors. An interesting feature of GAMs is that they provide insight

Table 9 Fire insurance portfolio: maximum-likelihood estimates and standard errors for Burr and GB2 regression models for observed losses. 'SI' stands for sum insured. For numerical reasons, losses are in 10^2 for the Burr analysis and 10^3 for the GB2 regression

Parameter	Burr (τ)	Burr (β)	GB2 (b)	GB2 (a)
	Estimate (s.e.)	Estimate (s.e.)	Estimate (s.e.)	Estimate (s.e.)
Intercept	0.46 (0.073)	−4.921 (0.316)	−8.446 (0.349)	0.049 (0.002)
Type 1	−0.327 (0.058)	−2.521 (0.326)	−2.5 (0.327)	−0.012 (0.002)
2	−0.097 (0.06)	−0.855 (0.325)	−0.867 (0.317)	−0.001 (0.002)
3	−0.184 (0.17)	−1.167 (0.627)	−1.477 (0.682)	−0.003 (0.003)
4	−0.28 (0.055)	−2.074 (0.303)	−2.056 (0.3)	−0.01 (0.002)
5	−0.091 (0.067)	−0.628 (0.376)	−0.651 (0.37)	−0.003 (0.003)
Type 1*SI	−0.049 (0.025)	−0.383 (0.152)	−0.384 (0.154)	−0.002 (0.001)
2*SI	0.028 (0.028)	0.252 (0.174)	0.248 (0.18)	0.001 (0.001)
3*SI	−0.51 (0.067)	−2.098 (0.345)	−2.079 (0.326)	−0.006 (0.001)
4*SI	−0.954 (0.464)	−5.242 (1.429)	−6.079 (1.626)	−0.025 (0.006)
5*SI	−0.074 (0.027)	−0.614 (0.17)	−0.598 (0.169)	−0.001 (0.001)
6*SI	−0.024 (0.037)	−0.21 (0.223)	−0.183 (0.235)	−0.001 (0.001)
β	0.00023 (0.00013)			
λ	0.457 (0.04)	0.444 (0.037)		
τ		1.428 (0.071)		
a			0.735 (0.045)	
b				0.969 (0.114)
p			3.817 (0.12)	263.53 (0.099)
q			1.006 (0.12)	357 (0.132)

to the specification of meaningful categories. Denuit and Lang (2004) provides for an excellent discussion of ratemaking with additive regression models. At the same time these authors incorporate postcode information in their rating system. This involves concepts from spatial statistics, a topic which is outside the scope of this paper.

3 Risk classification and *a posteriori* rating

During the construction of an *a priori* tariff structure, not all important risk factors may be observable. This is usually the situation for a new policyholder or even an existing one where there may be insufficient information to account for as many important risk factors to meet the homogeneity requirement of an efficient risk classification system. As a result, tariff cells will not be completely homogeneous. Take the case of a driver insured in an automobile insurance, his aggressiveness behind the wheel and the swiftness of his reflexes to be able to avoid possible accidents are difficult to assess. However, as the driver's claims history become more available to the insurer, this provides an additional, important revelation of the true riskiness of the policyholder. The insurer has the need to continually assess the efficiency of its risk classification scheme, and as such, the additional information provided by the

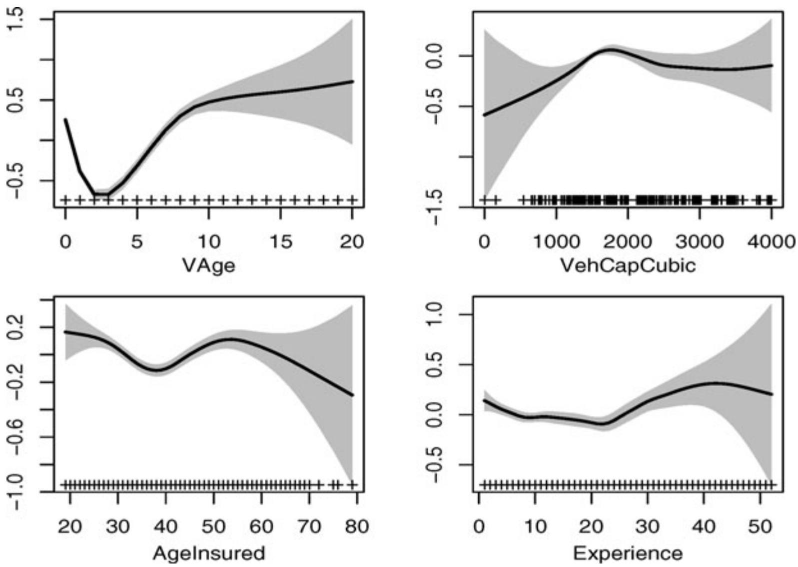

Fig. 4 Data from Illustration 2.1: additive effects in a Poisson GAM as specified in (20)

history of claims as they emerge must be taken into account. Thus, *a posteriori* statistical models are necessary in taking into account the history of reported claims and in adjusting the *a priori* premium accordingly.

Experience rating has a long tradition in actuarial science. It is a way to penalize ‘bad risks’ and reward ‘good risks’. Here we find that the premium for an insurance contract is calculated, after some claims history has been revealed, by accounting for both the experience of the individual policyholder together with that of the whole insurance portfolio to which the contract belongs. How much weight is assigned to the policyholder’s own experience is termed ‘credibility factor’ in actuarial science. In the statistics discipline, this revised estimate of the premium is based on what is called a ‘shrinkage estimator’. In actuarial science, this topic is called *credibility theory* and several works on this concept have appeared in the literature. Typically, we start with citation of the formulation of the classic actuarial credibility models developed by Bühlmann (1967) and Bühlmann (1969), which provided the profession with a theoretical justification of the underlying principles of experience or *a posteriori* rating. Professionals and researchers extended his fundamental works in several directions, e.g. Jewell (1975) presented credibility models for hierarchically structured portfolios and Hachemeister (1975) combined concepts of *a priori* risk classification with credibility. For a detailed and recent overview of credibility in actuarial science, see Bühlmann and Gisler (2005).

When updating the *a priori* tariff based on historical claims, actuaries are clearly dealing with *panel* (or longitudinal) data. This is different from the cross-sectional setting in Sect. 2.1 and involves repeated measurement on a group of ‘subjects’ (in this case: policies or policyholders) over time. Since they share subject-specific characteristics, observations on the same subject over time often are substantively correlated and require a different toolbox for statistical modeling. For instance, in the case of *a posteriori* rating, we will use Generalized Linear Mixed Models (GLMMs) instead of the GLMs from Sect. 2.2. GLMMs extend GLMs by including random effects in the linear predictor. The random effects do not only determine the correlation structure between observations on the same subject, but also take heterogeneity among subjects, due to unobserved characteristics, into account. Whereas the above-mentioned papers on credibility are theoretically oriented, contemporary statistical and econometric models for panel data and other types of clustered data allow the actuary to construct computationally driven *a posteriori* rating models, which can be implemented in standard statistical software packages. Throughout the various examples, we present this kind of statistical tools for *a posteriori* rating. Useful references that discuss various aspects of *a posteriori* rating include, but are not limited to, Pinquet (1997, 1998), Frees et al. (1999), Pinquet et al. (2001), Bolancé et al. (2003) and Antonio and Beirlant (2007).

Apart from the time component of recorded data, it is possible that a company may have several other layers of observations. To illustrate, it is not uncommon to find an automobile insurer to provide "fleet" coverages where a single insurance policy covers more than a single vehicle. Thus, the company has observations for a vehicle over a period of time under a "fleet" coverage. An additional level may be introduced when the observations are aggregated in the form of an "intercompany" experience, where the data comes from several companies. Such is usually the case for data obtained by reinsurers, companies that provide insurance protection to insurance companies. The analysis of data with such a multilevel structure is the topic of Sect. 3.2.

Experience rating based on multilevel (panel or higher order) models as discussed in Sects. 3.1 and 3.2 poses a challenge to the insurer when it comes to communicating the predictive results of these models to the policyholders. Customers may find it difficult to understand. It is not readily transparent to an ordinary policyholder how the surcharges (*maluses*) for reported claims and the discounts (*bonuses*) for claim-free periods are evaluated. In order to establish an experience rating system where insureds can easily understand the effect of reported claims or periods without claims, Bonus–Malus scales have been developed. Examples of these are discussed in Sect. 3.3.

3.1 *A priori* and *a posteriori* rating with credibility models for panel data

Continuing with our discussion of GLMs started in Sect. 2.2, we now focus on extending this type of models for *a posteriori* rating. GLMMs extend GLMs by allowing for random, or subject-specific, effects in the linear predictor (6). Including random effects in the linear predictor reflects the idea that there is a natural heterogeneity across subjects (policyholders, in this case) and that the observations on the same subject share common characteristics. This is the idea behind *a posteriori* ratemaking.

To fix ideas, suppose we have a data set consisting of N subjects. For each subject i $(1 \le i \le N)$ T_i observations are available. Given the vector $\boldsymbol{b}_i$ with the random effects for subject (or cluster) i, the repeated measurements $Y_{i1}, \ldots, Y_{iT_i}$ are assumed to be independent with a density from the exponential family

$$f(y_{it}|\boldsymbol{b}_i, \boldsymbol{\beta}, \phi) = \exp\left(\frac{y_{it}\theta_{it} - \psi(\theta_{it})}{\phi} + c(y_{it}, \phi)\right), \quad t = 1, \ldots, T_i. \tag{21}$$

Similar to (6), the following (conditional) relations hold

$$\mu_{it} = \mathrm{E}[Y_{it}|\boldsymbol{b}_i] = \psi'(\theta_{it}) \quad \text{and} \quad \mathrm{Var}[Y_{it}|\boldsymbol{b}_i] = \phi\psi''(\theta_{it}) = \phi V(\mu_{it}), \tag{22}$$

where $g(\mu_{it}) = \boldsymbol{x}_{it}'\boldsymbol{\beta} + \boldsymbol{z}_{it}'\boldsymbol{b}_i$. As before, $g(\cdot)$ is called the link and $V(\cdot)$, the variance function. $\boldsymbol{\beta}$ $(p \times 1)$ denotes the fixed effects parameter vector (governing *a priori* rating) and $\boldsymbol{b}_i$ $(q \times 1)$ the random effects vector. $\boldsymbol{x}_{it}$ $(p \times 1)$ and $\boldsymbol{z}_{it}$ $(q \times 1)$ contain subject i's covariate information for the fixed and random effects, respectively. The specification of the GLMM is completed by assuming that the random effects, $\boldsymbol{b}_i$ $(i = 1, \ldots, N)$, are mutually independent and identically distributed with density function $f(\boldsymbol{b}_i|\boldsymbol{\alpha})$. Herewith $\boldsymbol{\alpha}$ denotes the unknown parameters in the density. It is not uncommon to assume that the random effects have a (multivariate) normal distribution with zero mean and covariance matrix determined by $\boldsymbol{\alpha}$. Dependence between observations on the same subject arises because they share the same random effects $\boldsymbol{b}_i$.

The likelihood function for the unknown parameters $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$ and ϕ then becomes

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \phi; \boldsymbol{y}) = \prod_{i=1}^{N} f(\boldsymbol{y}_i|\boldsymbol{\alpha}, \boldsymbol{\beta}, \phi) = \prod_{i=1}^{N} \int \prod_{t=1}^{T_i} f(y_{it}|\boldsymbol{b}_i, \boldsymbol{\beta}, \phi) f(\boldsymbol{b}_i|\boldsymbol{\alpha})\, d\boldsymbol{b}_i, \tag{23}$$

where $\boldsymbol{y} = (\boldsymbol{y}_1', \ldots, \boldsymbol{y}_N')'$ and the integral is with respect to the q dimensional vector $\boldsymbol{b}_i$. For instance, when both the data and the random effects are normally distributed, the integral can be worked out analytically and explicit expressions exist for the maximum-likelihood estimator of $\boldsymbol{\beta}$ and the Best Linear Unbiased Predictor ('BLUP') for $\boldsymbol{b}_i$. For more general GLMMs, however, approximations to the likelihood or numerical integration techniques are required to maximize (23) with respect to the unknown parameters.

To illustrate the concepts described above, we now consider a Poisson GLMM with normally distributed random intercept. This GLMM allows for explicit calculation of the marginal mean and covariance matrix. In this way, one can clearly see how in this example the inclusion of the random effect leads to overdispersion and within-subject covariance.

Illustration 3.1 (A Poisson GLMM) Let N_{it} denote the claim frequency registered in year t for policyholder i. Assume that, conditional on b_i, N_{it} follows a Poisson distribution with mean $\mathrm{E}[N_{it}|b_i] = \exp(\boldsymbol{x}_{it}'\boldsymbol{\beta} + b_i)$ and that $b_i \sim N(0, \sigma_b^2)$. Straightforward calculations lead to

$$\begin{aligned}\mathrm{Var}(N_{it}) &= \mathrm{Var}\big(\mathrm{E}(N_{it}|b_i)\big) + \mathrm{E}\big(\mathrm{Var}(N_{it}|b_i)\big)\\ &= \mathrm{E}(N_{it})\big(\exp(\boldsymbol{x}_{it}'\boldsymbol{\beta})\big[\exp\big(3\sigma_b^2/2\big) - \exp\big(\sigma_b^2/2\big)\big] + 1\big),\end{aligned} \tag{24}$$

and

$$\begin{aligned}\mathrm{Cov}(N_{it_1}, N_{it_2}) &= \mathrm{Cov}\big(\mathrm{E}(N_{it_1}|b_i), \mathrm{E}(N_{it_2}|b_i)\big) + \mathrm{E}\big(\mathrm{Cov}(N_{it_1}, N_{it_2}|b_i)\big)\\ &= \exp\big(\boldsymbol{x}_{it_1}'\boldsymbol{\beta}\big)\exp\big(\boldsymbol{x}_{it_2}'\boldsymbol{\beta}\big)\big(\exp\big(2\sigma_b^2\big) - \exp\big(\sigma_b^2\big)\big).\end{aligned} \tag{25}$$

Hereby we used the expressions for the mean and variance of a Lognormal distribution. In the expression for the covariance we used the fact that, given the random effect b_i, N_{it_1} and N_{it_2} are independent. We see that the expression inside the parentheses in (24) is always bigger than 1. Thus, although $N_{it}|b_i$ follows a regular Poisson distribution, the marginal distribution of N_{it} is overdispersed. According to (25), due to the random intercept, observations on the same subject are no longer independent.

Illustration 3.2 (A Poisson GLMM—continued) Let N_{it} again denote the claim frequency for policyholder i in year t. Assume that, conditional on b_i, N_{it} follows a Poisson distribution with mean $\text{E}[N_{it}|b_i] = \exp(\boldsymbol{x}'_{it}\boldsymbol{\beta} + b_i)$ and that $b_i \sim N(-\frac{\sigma_b^2}{2}, \sigma_b^2)$. This re-parameterization is commonly used in ratemaking. Indeed, we now get

$$\text{E}[N_{it}] = \text{E}\big[\text{E}[N_{it}|b_i]\big] = \exp\left(\boldsymbol{x}'_{it}\boldsymbol{\beta} - \frac{\sigma_b^2}{2} + \frac{\sigma_b^2}{2}\right) = \exp(\boldsymbol{x}'_{it}\boldsymbol{\beta}), \tag{26}$$

and

$$\text{E}[N_{it}|b_i] = \exp(\boldsymbol{x}'_{it}\boldsymbol{\beta} + b_i). \tag{27}$$

This specification shows that the *a priori* premium, given by $\exp(\boldsymbol{x}'_{it}\boldsymbol{\beta})$, is correct on the average. The *a posteriori* correction to this premium is determined by $\exp(b_i)$.

Besides the Lognormal distribution used in the above examples, other mixing distributions can be used. In the Poisson-Gamma framework for instance, the conjugacy of these distributions allows for explicit calculation of the predictive premium. This is demonstrated in Illustration 3.3.

Illustration 3.3 (A Poisson-Gamma model) A simple and classical random effects Poisson model for panel data (see e.g. Hausman et al. 1984) is constructed with the assumptions

$$N_{it} \sim \text{Poi}(b_i\lambda_{it}), \quad \text{where } \lambda_{it} = \exp(\boldsymbol{x}'_{it}\boldsymbol{\beta}) \text{ and } b_i \sim \Gamma(\alpha, \alpha).$$

It follows that $E[b_i] = 1$ and the resulting joint, unconditional distribution then becomes

$$\begin{aligned}&\Pr(N_{i1} = n_{i1}, \ldots, N_{iT_i} = n_{iT_i})\\ &= \left(\prod_{t=1}^{T_i} \frac{\lambda_{it}^{n_{it}}}{n_{it}!}\right) \frac{\Gamma(\sum_{t=1}^{T_i} n_{it} + \alpha)}{\Gamma(\alpha)} \left(\frac{\alpha}{\sum_{t=1}^{T_i} n_{it} + \alpha}\right)^{\alpha} \times \left(\sum_{t=1}^{T_i} n_{it} + \alpha\right)^{-\sum_{t=1}^{T_i} n_{it}},\end{aligned} \tag{28}$$

with $\text{E}[N_{it}] = \text{E}[\text{E}[N_{it}|b_i]] = \lambda_{it}$ and $\text{Var}[N_{it}] = \text{E}[\text{Var}[N_{it}|b_i]] + \text{Var}[\text{E}[N_{it}|b_i]] = \lambda_{it} + \frac{1}{\alpha}\lambda_{it}^2$.

For the specification in (28), the posterior distribution of the random intercept b_i has again a Gamma distribution with

$$f(b_i|N_{i1}=n_{i1},\ldots,N_{iT_i}=n_{iT_i}) \propto \Gamma\left(\sum_{t=1}^{T_i} n_{it}+\alpha, \sum_{t=1}^{T_i}\lambda_{it}+\alpha\right). \quad (29)$$

The (conditional) mean and variance of this posterior distribution are given, respectively, by

$$\mathrm{E}[b_i|N_{it}=n_{it},\ t=1,\ldots,T_i] = \frac{\alpha+\sum_{t=1}^{T_i} n_{it}}{\alpha+\sum_{t=1}^{T_i}\lambda_{it}} \quad \text{and} \quad (30)$$

$$\mathrm{Var}[b_i|N_{it}=n_{it},\ t=1,\ldots,T_i] = \frac{\alpha+\sum_{t=1}^{T_i} n_{it}}{\left(\alpha+\sum_{t=1}^{T_i}\lambda_{it}\right)^2}. \quad (31)$$

This leads to the following *a posteriori* premium:

$$\begin{aligned}\mathrm{E}[N_{i,T_i+1}|N_{it}=n_{it},\ t=1,\ldots,T_i] &= \lambda_{i,T_i+1}\mathrm{E}[b_i|N_{it}=n_{it},\ t=1,\ldots,T_i]\\ &= \lambda_{i,T_i+1}\left\{\frac{\alpha+\sum_{t=1}^{T_i} n_{it}}{\alpha+\sum_{t=1}^{T_i}\lambda_{it}}\right\}. \quad (32)\end{aligned}$$

The above credibility premium is optimal when a quadratic loss function is used. Indeed, as is known in mathematical statistics, the conditional expectation minimizes a mean squared error criterion.

It is now demonstrated how credibility calculations for a panel data set can be done with standard statistical software packages.

Illustration 3.4 (A numerical example of *a posteriori* rating with a Poisson GLMM) We illustrate the statistical tools introduced above with a numerical example. Data consist of 12,893 policyholders who were observed during (fractions of) the period 1993–2003. Let N_{it} be the number of claims registered for policyholder i in period t. The following model specification is used:

$$N_{it}|b_i \sim \mathrm{Poi}(\mu_{it}|b_i) \quad \text{and} \quad \mu_{it}|b_i = e_{it}\exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}+b_i), \quad (33)$$

$$b_i \sim N\left(-\sigma^2/2, \sigma^2\right), \quad (34)$$

where e_{it} is the exposure for policyholder i in year t (expressed in years). If the rating model only uses observable risk characteristics (in $\mathbf{x}_{it}$) and not the claims history of the insured, the *a priori* premium is given by

$$(a\ priori) \quad E[N_{it}] = e_{it}\exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}). \quad (35)$$

An experience rating system will adopt the *a priori* premium based on the claims reported by the insured. This results in the following *a posteriori* premium:

$$(a\ posteriori) \quad E[N_{it}|b_i] = e_{it}\exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}+b_i). \quad (36)$$

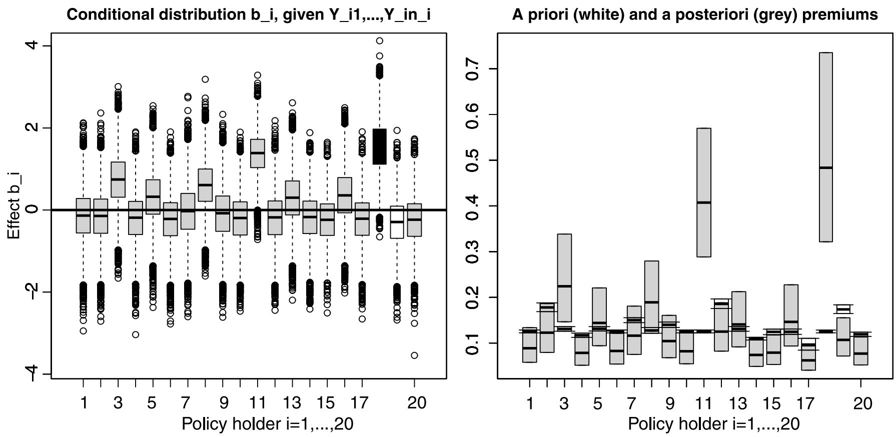

Fig. 5 (*Left*) Boxplot of the conditional distribution of b_i, given the history $N_{i1}, \ldots, N_{iT_i}$, for a random selection of 20 policyholders. (*Right*) For the same selection of policyholders: boxplots with simulations from the a priori (*white*) and a posteriori (*gray*) premium

The ratio (36)/(35) is called the theoretical Bonus–Malus Factor (BMF). It reflects the extent to which the policyholder is rewarded or penalized for his past claims. Using standard statistical software (like `Proc NlMixed` in `SAS`) a point estimate for (35), (36) and the BMF can be obtained easily. In Fig. 5 (left) random draws from the conditional distributions of b_i (given $N_{i1}, \ldots, N_{iT_i}$) are shown for a random selection of policyholders. These results can be obtained from a Bayesian analysis of the model (using, for example, `WinBugs`). The conditional distributions reflect the heterogeneity between policyholders as well as their risk behavior. For instance, the black boxplot represents a policyholder who reported four claims during an insured period of 0.67 years, while the white boxplot represents a policyholder with zero claims during 6.4 years of exposure. The right panel of Fig. 5 shows boxplots of simulated values of the *a priori* ((35)) and *a posteriori* ((36)) premiums of N_{i,T_i+1} (for the selection of policyholders under consideration). One can see how the *a priori* premiums are corrected based on observed claims.

More supporting graphs are given in Figs. 6 and 7. Such graphs help the actuary to gain further insight into the portfolio. The vertical axis in both figures gives the *a posteriori* premium, expressed as a percentage of the *a priori* premium. The horizontal axis in Fig. 6 gives the average number of claims reported by the policyholder per year. This average is the ratio of total reported claims and the total period of exposure (in years). For instance, for the policyholders highlighted in Fig. 6 (left panel) we observed four claims on 0.64 years (purple), two on 0.36 (years) (blue) and one on 0.167 year. These claim histories result in *a posteriori* 462%, 220% and 150% of the *a priori* premium. The right panel in Fig. 6 is a detail of the left panel. In this panel we only display policyholders with a total period of exposure of 3 years. For instance, the green triangles all represent policyholders who reported the same number of claims (i.e. one claim) during the same period of exposure (i.e. 3 years). Nevertheless, the *a posteriori* corrections that apply to this group of policyholders differ within the group. The dashed lines in the plot connect policyholders with a low (respectively

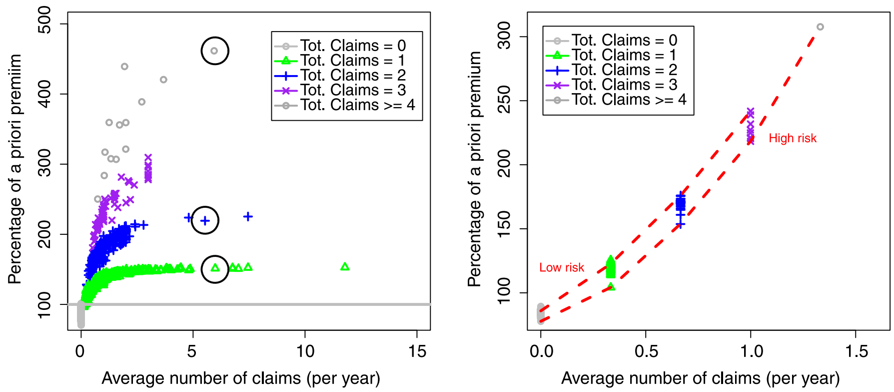

Fig. 6 A posteriori premium expressed as percentage of the a priori premium (y-axis) versus the average number of claims

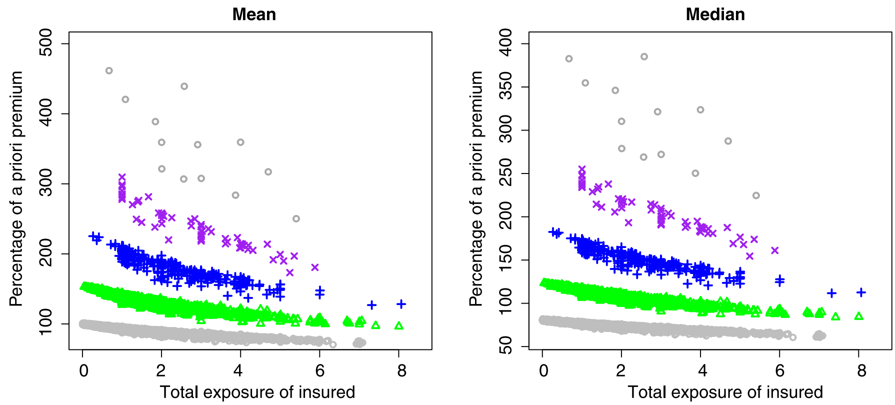

Fig. 7 A posteriori premium expressed as percentage of the a priori premium (y-axis) versus the total period of insurance. *Left panel* uses the mean and *right panel* the median of the conditional distribution of b_i, given $N_{i1}, \ldots, N_{iT_i}$

high) *a priori* risk premium. It becomes clear from the plot that corrections for high *a priori* risks are softer (i.e. penalties are lower and discounts are higher) than those for low *a priori* risks.

Figure 7 has the policyholder's total exposure on the x-axis. In the left panel we use the mean of the conditional distribution of b_i (given $N_{i1}, \ldots, N_{iT_i}$) as predictor for b_i in (36). The right panel uses the median. We see that the median results in a less severe experience rating: smaller penalizations for past claims and bigger rewards for claim-free policies. At the same time Fig. 7 illustrates graphically how many claim-free years a policyholder needs to get rid of his penalty, e.g. after reporting one claim, and pay the *a priori* premium again. Graphically this is represented by the intersection of the green cloud and a horizontal line at $y = 100$.

3.2 *A priori* and *a posteriori* rating with credibility models for clustered data

Whereas Sect. 3.1 is an overview of techniques involved in the rating of data bases with a panel structure, actuaries may have data sets at their disposal with a more difficult structure. This is illustrated with an example of multilevel or hierarchical rating that first appeared in the actuarial literature in Antonio et al. (2010).

Illustration 3.5 (A multilevel model for intercompany claim counts) Data are available on claim counts registered for automobile insurance policies over a period of nine years (1993–2002). The source contains a pooled experience of several insurers. Moreover, vehicles under consideration are insured under a 'fleet' policy. Fleet policies are umbrella-type policies issued to customers whose insurance covers more than a single vehicle. The hierarchical or multilevel structure of the data is as follows: vehicles (v) observed over time (t) that are nested within fleets (f), with policies issued by insurance companies (c). Multilevel statistical models allow one to incorporate the hierarchical structure of the data by specifying random effects at the vehicle, fleet and company levels. These random effects represent unobservable characteristics at each level. At the vehicle level, the missions assigned to a vehicle or unobserved driver behavior may influence the riskiness of a vehicle. At the fleet level, guidelines on driving hours, mechanical check-ups, loading instructions and so on, may influence the number of accidents reported. At the insurance company level, underwriting and claim settlement practices may affect claims. Moreover, random effects allow one to update an *a priori* tariff, taking into account the past performance of vehicle, fleet and company. As such, they are relevant for *a posteriori* rating with clustered data.

Antonio et al. (2010) compare the performance of *a posteriori* rating models (incorporating *a priori* characteristics) based on various count distributions (namely Poisson, negative binomial, zero-inflated Poisson and hurdle Poisson). Denote $Y_{c,f,v,t}$ the number of claims in period t for vehicle v insured under fleet f by company c. With the Poisson distribution the *a priori* tariff is expressed as

$$\begin{aligned} N_{c,f,v,t} &\sim \text{Poi}\big(\mu^{\text{prior}}_{c,f,v,t}\big), \\ \mu^{\text{prior}}_{c,f,v,t} &= e_{c,f,v,t}\exp(\eta_{c,f,v,t}), \qquad (37) \\ \eta_{c,f,v,t} &= \beta_0 + \boldsymbol{x}'_c\boldsymbol{\beta}_4 + \boldsymbol{x}'_{cf}\boldsymbol{\beta}_3 + \boldsymbol{x}'_{cfv}\boldsymbol{\beta}_2 + \boldsymbol{x}'_{cfvt}\boldsymbol{\beta}_1. \qquad (38) \end{aligned}$$

Hereby $\boldsymbol{x}_c$, $\boldsymbol{x}_{cf}$, $\boldsymbol{x}_{cfv}$ and $\boldsymbol{x}_{cfvt}$ contain observable covariate information registered at the level of company, fleet, vehicle and period, respectively. $e_{c,f,v,t}$ denotes the exposure. *A posteriori* the tariff is updated as follows:

$$\begin{aligned} N_{c,f,v,t}|b_c; b_{c,f}; b_{c,f,v} &\sim \text{Poi}(\mu_{c,f,v,t}|b_c; b_{c,f}; b_{c,f,v}), \\ \mu_{c,f,v,t}|b_c; b_{c,f}; b_{c,f,v} &= \mu^{\text{prior}}_{c,f,v,t} \times \exp(b_c + b_{c,f} + b_{c,f,v}), \qquad (39) \end{aligned}$$

$$b_c \sim N\left(-\frac{\sigma_c^2}{2}, \sigma_c^2\right) \quad \text{and} \quad b_{c,f} \sim N\left(-\frac{\sigma_{c,f}^2}{2}, \sigma_{c,f}^2\right) \quad \text{and}$$

$$b_{c,f,v} \sim N\left(-\frac{\sigma_{c,f,v}^2}{2}, \sigma_{c,f,v}^2\right). \qquad (40)$$

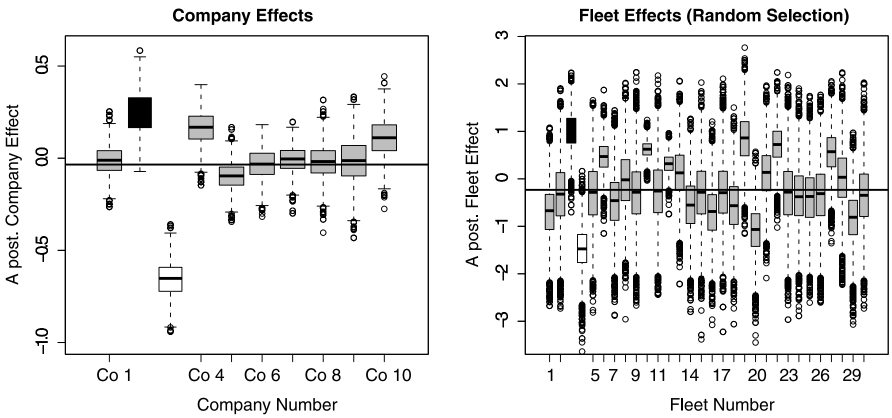

Fig. 8 Illustration of posterior distributions of company effects and a random selection of fleet effects. A *horizontal line* is plotted at the mean of the random effects distribution

Equation (37) gives the *a priori* premium and (39) the *a posteriori* tariff. The ratio (*a posteriori* premium/*a priori* premium) is the theoretical Bonus–Malus Factor (BMF). Differences between companies and fleets are revealed by the posterior distributions of the random effects at company and fleet level. See Fig. 8 for an illustration. The left panel in this figure reflects the company effects. For instance, the black boxplot represents a company where 1,096 claims were registered on a total exposed period of 4,440 years. The white boxplot corresponds with a company having 191 claims on a period of 2,480 years of exposure. The right panel in Fig. 8 reflects fleet effects. The white boxplot is from a fleet with zero claims on 53 years of exposure, whereas the past experience represented by the black boxplot is much worse: nine claims on a period of 9 years of exposure.

Antonio et al. (2010) investigate how various claim count distributions (see Sect. 2.3.1) perform in *a posteriori* rating systems. In Tables 10 and 11 we follow three vehicles to illustrate the mechanism of experience rating with each of the model specifications under investigation. The first illustration from Table 10 uses a Poisson hierarchical model with random effects for company, fleet and vehicle. In this illustration the BMFs for all vehicles are above 1, but the BMF for the vehicle that reports one claim is much higher (2.05) than the BMF for the claim-free vehicles (1.56 and 1.58). Checking the corresponding results for the hierarchical negative binomial model and the ZIP with fixed p (see distribution specification (11)), the BMF for all vehicles is >1 and in between those reported in the first illustration in Table 10. The latter models calculate BMFs at the fleet level, a natural point in the hierarchy because it is at this level where an insurance contract between a fleet and insurance company is written. Hence, fleet level BMFs can be used for premium renewals. The first illustration in Table 10 shows BMFs calculated at the vehicle level. This information could also be used for contracts written at the fleet level; as the fleet composition changes through the retirement or sale of vehicles, the total fleet premium should reflect the changing composition of vehicles. Vehicle level BMFs will allow prices to depend on the vehicle composition of fleets.

Table 10 Effects of different models on premiums for selected vehicles. Results for hierarchical Poisson, NB and ZIP with fixed p regression models

Vehicle Number	*A Priori* (Exp.)	*A Posteriori*	BMF	Acc. Cl. Fleet (Exp.)	Acc. Cl. Veh. (Exp.)
Hierarchical Poisson with random effects for vehicle, fleet and company					
6,645	0.08435 (0.5038)	0.1725	2.05	6 (18.5)	1 (1)
7,006	0.08435 (0.5038)	0.1316	1.56		0 (1)
6,500	0.08435 (0.5038)	0.1329	1.58		0 (1)
Hierarchical NB with random effects for fleet and company					
6,645	0.08383 (0.5038)	0.1435	1.71	6 (18.5)	1 (1)
7,006	0.08383 (0.5038)	0.1435			0 (1)
6,500	0.08383 (0.5038)	0.1435			0 (1)
Hierarchical ZIP with random effects for fleet and company, fixed p					
6,645	0.08241 (0.5038)	0.1484	1.8	6 (18.5)	1 (1)
7,006	0.08241 (0.5038)	0.1484			0 (1)
6,500	0.08241 (0.5038)	0.1484			0 (1)

Note: 'Acc. Cl. Fleet' and 'Acc. Cl. Veh.' are accumulated number of claims at fleet and vehicle levels, respectively. 'Exp.' is exposure at year level, in parentheses

Table 11 Effects of different models on premiums for selected vehicles. Results for ZIP with fleet-specific p and hurdle Poisson model

Vehicle	*A Priori* (Exp.)	*A Posteriori*	BMF	Acc. Cl. (Exp.)	Claim-Free Years
Hierarchical ZIP with random effects for fleet and company, fleet-specific p					
6,645	0.09051 (0.5038)	0.1306	1.37	6 (18.5)	17
7,006	0.09051 (0.5038)	0.1306			
6,500	0.09051 (0.5038)	0.1306			
Hierarchical hurdle Poisson with random effects for fleet and company					
6,645	0.1098 (0.5038)	0.11	1	6 (18.5)	17
7,006	0.1098 (0.5038)	0.11			
6,500	0.1098 (0.5038)	0.11			

Note: 'Acc. Cl. Fleet' and 'Acc. Cl. Veh.' are accumulated number of claims at fleet and vehicle levels, respectively. 'Exp.' is exposure at year level, in parentheses

Comparing the results in Tables 10 and 11 we see that *a priori* premiums obtained with the different model specifications closely correspond. The zero-inflated model with fleet-specific $p_{c,f}$ and the hurdle Poisson model take the claim-free period of a fleet into account. For panel data, this feature was made explicit in Boucher et al. (2009) and Boucher et al. (2008). Compare the results for fleet in Tables 10 and 11 between the various specifications: in the NB and ZIP with p fixed, the BMF for this fleet is $1.71/1.8$. In ZIP model with random p this drops to 1.37 and in the hurdle model even to 1. That is because these last two model specifications not only use the number of registered claims, but also the claim-free periods (which is here 17 out of a total of 18.5 years).

3.3 Experience rating with Bonus–Malus scales

Illustrations 3.4 and 3.5 show that credibility models for hierarchically structured data are statistically challenging. To the insureds, it is not obvious how penalties for past claims and discounts based on claim-free periods are calculated. Within automobile insurance Bonus–Malus (BM) scales are a well-known and widely used commercial alternative for the credibility type rating systems discussed above. Their commercial attractiveness lies in the fact that insureds are able to understand how information on the number of claims reported in year t (N_{it}) will change the premium they have to pay in year $t+1$ for their automobile insurance. Our main reference in this section is Denuit et al. (2007). To discuss the probabilistic and statistical aspects of Bonus–Malus scales, a *credibility model* similar to the one in Illustration 3.3 is assumed with the following specification:

- policy i of the portfolio ($i = 1, \ldots, n$) is represented by a sequence $(\Theta_i, \boldsymbol{N}_i)$ where $\boldsymbol{N}_i = (N_{i1}, N_{i2}, \ldots)$ and Θ_i represents unexplained heterogeneity and has mean 1;
- given $\Theta_i = \theta$ the random variables N_{it} ($t = 1, 2, \ldots$) are independent and $\text{Poi}(\lambda_{it}\theta)$ distributed; and
- the sequences $(\Theta_i, \boldsymbol{N}_i)$ ($i = 1, \ldots, n$) are assumed to be independent.

3.3.1 Bonus–Malus scales

A BM scale consists of a certain number of levels, say $s+1$, which are numbered from $0, \ldots, s$. A new driver will enter the scale at a specified level, say ℓ_0. Drivers will transition up and down the scale according to the number of claims reported in each year of insurance. A claim-free year results in a bonus point, which implies that the driver goes one level down (0 being the best scale). Claims are penalized by malus points, meaning that for each claim filed, the driver goes up a certain number of levels. Denote this number by ‘pen’, the penalty. The trajectory of a driver through the scale can be represented by a sequence of random variables: $\{L_1, L_2, \ldots\}$ where L_k takes values in $\{0, \ldots, s\}$ and represents the level occupied in the time interval $(k, k+1)$. With N_k the number of claims reported by the insured in the period $(k-1, k)$, the future level of an insured L_k is obtained from the present level L_{k-1} and the number of claims reported during the present year N_k. This is at the heart of Markov models: the future depends on the present and not on the past. The L_k’s obey the recursion:

$$L_k = \begin{cases} \max(L_{k-1} - 1, 0) & \text{if } N_k = 0, \\ \min(L_{k-1} + N_k \times \text{pen}, s) & \text{if } N_k \geq 1, \end{cases} \tag{41}$$

assuming independence of $N_1, N_2, \ldots$. With each level ℓ in the scale a so-called relativity r_ℓ is associated. A policyholder who has at present *a priori* premium λ_{it} and is in scale ℓ, has to pay $r_\ell \times \lambda_{it}$. The relativities, together with the transition rules in the scale, are the commercial alternative for the credibility type models discussed before.

Table 12 Transitions in the (−1/Top Scale) BM system

Starting level	Level 0 claim	occupied if ≥1 is reported
0	0	5
1	0	5
2	1	5
3	2	5
4	3	5
5	4	5

Illustration 3.6 (−1/Top Scale) Throughout this section fundamentals of BM scales are illustrated with a simple example of such a scale: the (−1/Top Scale). This scale has six levels, numbered $0, 1, \dots, 5$. Starting class is level 5. Each claim-free year is rewarded by one bonus class. When an accident is reported the policyholder is transferred to scale 5. Table 12 represents these transitions.

3.3.2 Transition rules, transition probabilities and stationary distribution

To enable the calculation of the relativity corresponding with each level ℓ, some probabilistic concepts associated with BM scales have to be introduced. The *transition rules* corresponding with a certain BM scale are indicator variables $t_{ij}(k)$ such that

$$t_{ij}(k) = \begin{cases} 1 & \text{if the policy transfers from } i \text{ to } j \text{ when } k \text{ claims are reported,} \\ 0 & \text{otherwise.} \end{cases} \tag{42}$$

They can be summarized in the matrix $\boldsymbol{T}(k)$:

$$\boldsymbol{T}(k) = \begin{pmatrix} t_{00}(k) & t_{01}(k) & \dots & t_{0s}(k) \\ t_{10}(k) & t_{11}(k) & \dots & t_{1s}(k) \\ \vdots & \vdots & \ddots & \vdots \\ t_{s0}(k) & t_{s1}(k) & \dots & t_{ss}(k) \end{pmatrix}, \tag{43}$$

which is a 0–1 matrix where each row has exactly one 1.

Assuming $N_1, N_2, \dots$ are independent and Poi(θ) distributed, the trajectory this driver follows through the scale will be represented as $\{L_1(\theta), L_2(\theta), \dots\}$. The transition probability of this driver to go from level ℓ_1 to ℓ_2 in a single step is

$$p_{\ell_1\ell_2}(\theta) = \Pr[L_{k+1}(\theta) = \ell_2 | L_k(\theta) = \ell_1] \tag{44}$$

$$= \sum_{n=0}^{+\infty} \Pr\big[L_{k+1}(\theta)$$

$$= \ell_2 | N_{k+1} = n, L_k(\theta) = \ell_1\big] \Pr\big[N_{k+1} = n | L_k(\theta) = \ell_1\big] \tag{45}$$

$$= \sum_{n=0}^{+\infty} \frac{\theta^n}{n!} \exp(-\theta) t_{\ell_1\ell_2}(n), \tag{46}$$

where we used the independence of N_{k+1} and L_k. The corresponding *one-step transition matrix* $\boldsymbol{P}(\theta)$ is given by

$$\boldsymbol{P}(\theta) = \begin{pmatrix} p_{00}(\theta) & p_{01}(\theta) & \dots & p_{0s}(\theta) \\ p_{10}(\theta) & p_{11}(\theta) & \dots & p_{1s}(\theta) \\ \vdots & \vdots & \ddots & \vdots \\ p_{s0}(\theta) & p_{s1}(\theta) & \dots & p_{ss}(\theta) \end{pmatrix}. \tag{47}$$

The n-step transition probability $p_{ij}^{(n)}$ gives the probability of being transferred from level i to level j in n steps:

$$\begin{aligned} p_{ij}^{(n)}(\theta) &= \Pr\big[L_{k+n}(\theta) = j | L_k(\theta) = i\big] \\ &= \sum_{i_1=0}^{s} \sum_{i_2=0}^{s} \cdots \sum_{i_{n-1}=0}^{s} p_{ii_1}(\theta) p_{i_1 i_2}(\theta) \cdots p_{i_{n-1} j}(\theta), \end{aligned} \tag{48}$$

where the last expression includes all possible paths between i and j in n steps and the probability of their occurrence. These probabilities are summarized in the ***n**-step transition matrix* $\boldsymbol{P}^{(n)}(\theta)$:

$$\boldsymbol{P}^{(n)}(\theta) = \begin{pmatrix} p_{00}^{(n)}(\theta) & p_{01}^{(n)}(\theta) & \dots & p_{0s}^{(n)}(\theta) \\ p_{10}^{(n)}(\theta) & p_{11}^{(n)}(\theta) & \dots & p_{1s}^{(n)}(\theta) \\ \vdots & \vdots & \ddots & \vdots \\ p_{s0}^{(n)}(\theta) & p_{s1}^{(n)}(\theta) & \dots & p_{ss}^{(n)}(\theta) \end{pmatrix}. \tag{49}$$

The following relation holds between the 1- and n-step transition matrices: $\boldsymbol{P}^{(n)}(\theta) = \boldsymbol{P}^n(\theta)$.

Ultimately, the BM system will stabilize and the proportion of policyholders occupying each level of the scale will remain unchanged. These proportions are captured in the *stationary distribution* $\boldsymbol{\pi}(\theta) = (\boldsymbol{\pi}_0(\theta), \dots, \boldsymbol{\pi}_s(\theta))'$, which are defined as

$$\boldsymbol{\pi}_{\ell_2}(\theta) = \lim_{n \to +\infty} p_{\ell_1 \ell_2}^{(n)}(\theta). \tag{50}$$

Correspondingly, $\boldsymbol{P}^{(n)}(\theta)$ converges to $\boldsymbol{\Pi}(\theta)$ defined as

$$\lim_{n \to +\infty} \boldsymbol{P}^{(n)}(\theta) = \boldsymbol{\Pi}(\theta) = \begin{pmatrix} \boldsymbol{\pi}'(\theta) \\ \boldsymbol{\pi}'(\theta) \\ \vdots \\ \boldsymbol{\pi}'(\theta) \end{pmatrix}. \tag{51}$$

Illustration 3.7 (-1/Top Scale—continued) For the BM scale introduced in Illustration 3.6 the transition and one-step probability matrices are given as follows:

$$\boldsymbol{T}(0) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \quad \text{and} \quad \boldsymbol{T}(1) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \tag{52}$$

$$\boldsymbol{P}(\theta) = \begin{pmatrix} \exp(-\theta) & 0 & 0 & 0 & 0 & 1-\exp(-\theta) \\ \exp(-\theta) & 0 & 0 & 0 & 0 & 1-\exp(-\theta) \\ 0 & \exp(-\theta) & 0 & 0 & 0 & 1-\exp(-\theta) \\ 0 & 0 & \exp(-\theta) & 0 & 0 & 1-\exp(-\theta) \\ 0 & 0 & 0 & \exp(-\theta) & 0 & 1-\exp(-\theta) \\ 0 & 0 & 0 & 0 & \exp(-\theta) & 1-\exp(-\theta) \end{pmatrix}. \tag{53}$$

Using a result from Rolski et al. (1999) (also see Denuit et al. 2007) the stationary distribution $\boldsymbol{\pi}(\theta)$ can be obtained as $\boldsymbol{\pi}'(\theta) = \boldsymbol{e}'(\boldsymbol{I} - \boldsymbol{P}(\theta) + \boldsymbol{E})^{-1}$, with $\boldsymbol{E}$ the $(s+1) \times (s+1)$ matrix with all entries 1. For the (-1/Top Scale) this results in:

$$\boldsymbol{\pi}'(\theta) = (1,1,1,1,1,1) \times \begin{pmatrix} 2-\exp(-\theta) & 1 & 1 & 1 & 1 & \exp(-\theta) \\ 1-\exp(-\theta) & 2 & 1 & 1 & 1 & \exp(-\theta) \\ 1 & 1-\exp(-\theta) & 2 & 1 & 1 & \exp(-\theta) \\ 1 & 1 & 1-\exp(-\theta) & 2 & 1 & \exp(-\theta) \\ 1 & 1 & 1 & 1-\exp(-\theta) & 2 & \exp(-\theta) \\ 1 & 1 & 1 & 1 & 1-\exp(-\theta) & 1+\exp(-\theta) \end{pmatrix}^{-1}. \tag{54}$$

For instance, with $\theta = 0.1546$ (the annual claim frequency from Illustration 2.1) the stationary distribution becomes:

$$\boldsymbol{\pi}'(0.1546) = \begin{pmatrix} 0.46416 & 0.0772 & 0.0901 & 0.1051 & 0.1227 & 0.1432 \end{pmatrix}. \tag{55}$$

3.3.3 Relativities

In a BM scale the relativity r_ℓ corresponding with scale ℓ corrects the *a priori* premium: *a posteriori* the policyholder will pay r_ℓ% of the *a priori* premium. The calculation of the relativities, given *a priori* risk characteristics, is one of the main tasks of the actuary. This type of calculations shows a lot of similarities with explicit credibility type calculations (as in Illustration 3.3). Following Norberg (1976) with the number of levels and transition rules being fixed, the optimal relativity r_ℓ, corresponding with level ℓ, is determined by maximizing the asymptotic predictive accuracy. This implies that one tries to minimize

$$E\big[(\Theta - r_L)^2\big], \tag{56}$$

the difference between the relativity r_L and the 'true' relative premium Θ, under the assumptions of our credibility model. Simplifying the notation in this model, the *a priori* premium of a random policyholder is denoted with Λ and the residual effect of unknown risk characteristics with Θ. The policyholder then has (unknown) annual expected claim frequency $\Lambda\Theta$, where Λ and Θ are assumed to be independent. The weights of different risk classes follow from the *a priori* system with $\Pr[\Lambda = \lambda_k] = w_k$.

Calculation of the r_ℓ's goes as follows:

$$\begin{aligned} \min E\big[(\Theta - r_L)^2\big] &= \sum_{\ell=0}^{s} E\big[(\Theta - r_\ell)^2 | L = \ell\big] \Pr[L = \ell] \qquad (57) \\ &= \sum_{\ell=0}^{s} \int_0^{+\infty} (\theta - r_\ell)^2 \Pr[L = \ell | \Theta = \theta]\, dF_\Theta(\theta) \\ &= \sum_k w_k \int_0^{+\infty} \sum_{\ell=0}^{s} (\theta - r_\ell)^2 \pi_\ell(\lambda_k \theta)\, dF_\Theta(\theta), \qquad (58) \end{aligned}$$

where $\Pr[\Lambda = \lambda_k] = w_k$. In the last step of the derivation conditioning is on Λ. It is straightforward to obtain the optimal relativities by solving

$$\frac{\partial E[(\Theta - r_L)^2]}{\partial r_j} = 0 \quad \text{with } j = 0, \ldots, s. \tag{59}$$

Alternatively, from mathematical statistics it is well-known that for a quadratic loss function (see (57)) the optimal $r_\ell = E[\Theta | L = \ell]$. This is calculated as follows:

$$\begin{aligned} r_\ell &= E[\Theta | L = \ell] \\ &= E\big[E[\Theta | L = \ell, \Lambda] | L = \ell\big] \\ &= \sum_k E[\Theta | L = \ell, \Lambda = \lambda_k] \Pr[\Lambda = \lambda_k | L = \ell] \\ &= \sum_k \int_0^{+\infty} \theta \frac{\Pr[L = \ell | \Theta = \theta, \Lambda = \lambda_k] w_k}{\Pr[L = \ell, \Lambda = \lambda_k]}\, dF_\Theta(\theta) \frac{\Pr[\Lambda = \lambda_k, L = \ell]}{\Pr[L = \ell]}, \qquad (60) \end{aligned}$$

where the relation $f_{\Theta | L=\ell, \Lambda=\lambda_k}(\theta | \ell, \lambda_k) = \frac{\Pr[L=\ell | \Theta=\theta, \Lambda=\lambda_k] \times w_k \times f_\Theta(\theta)}{\Pr[\Lambda=\lambda_k, L=\ell]}$ is used. The optimal relativities are given by

$$r_\ell = \frac{\sum_k w_k \int_0^{+\infty} \theta \pi_\ell(\lambda_k \theta)\, dF_\Theta(\theta)}{\sum_k w_k \int_0^{+\infty} \pi_\ell(\lambda_k \theta)\, dF_\Theta(\theta)}. \tag{61}$$

Table 13 Numerical characteristics for the (−1/Top Scale) and the portfolio from Illustration 2.1, without and with *a priori* rating taken into account

Level ℓ	$\Pr[L=\ell]$	$r_\ell = E[\Theta \mid L=\ell]$	
		without *a priori*	with *a priori*
5	13.67%	160%	136.7%
4	10.79%	145.6%	127.7%
3	8.7%	133.9%	120.5%
2	7.14%	123.1%	114.4%
1	5.94%	114.2%	109.2%
0	53.75%	65.47%	78.9%

When no *a priori* rating system is used, all the λ_ks are equal (estimated by $\hat{\lambda}$) and the relativities reduce to

$$r_\ell = \frac{\int_0^{+\infty} \theta \pi_\ell(\hat{\lambda}\theta)\, dF_\Theta(\theta)}{\int_0^{+\infty} \pi_\ell(\hat{\lambda}\theta)\, dF_\Theta(\theta)}. \tag{62}$$

Illustration 3.8 ((−1/Top Scale) (continued)) The relativities are calculated for the data set introduced in Illustration 2.1 and the (−1/Top Scale) from Illustration 3.6. Numerical integration was done with the function `integrate` in `R`. Without *a priori* ratemaking the relativities are calculated with $\hat{\lambda} = 0.1546$ and $\Theta_i \sim \Gamma(\alpha, \alpha)$ with $\hat{\alpha} = 1.4658$. Results are in Table 13. Incorporating the *a priori* rating system developed in Illustration 2.3 and tabulated in Table 14, relativities are adapted as in Table 13. By taking *a priori* characteristics into account, the *a posteriori* corrections are softened.

4 Concluding remarks

Risk classification involves the process of grouping insurance risks into various categories or cells that share a homogeneous set of characteristics. It is an extremely important part of establishing a fair and reasonable tariff structure for a portfolio of insurance risks. Such a categorization usually leads to constructing many different cells for which members of each cell share a similar set of risk characteristics and therefore must pay the same premium rate. McClenahan (2001) outlines several considerations that must be met when selecting rating (or classifying) variables in constructing a fair risk classification scheme. Among these considerations are the so-called actuarial criteria, which require the selection to be based on sound fundamental statistical principles. It is the purpose of this paper to survey some of the old and the newer advanced statistical techniques that can be employed by the actuary to meet these criteria when establishing a risk classification system.

In selecting rating variables for risk classification purposes, the primary statistical tools fall within the class of regression-type models. These types of models have several advantages; in particular, they allow the actuary to identify statistically significant rating variables, to quantify the statistical effects of each rating variable, and to

Table 14 Parameter estimates for several regression models for the data introduced in Illustration 2.1

	Poisson	NB	ZIP
Parameter	Estimate (s.e.)	Estimate (s.e.)	Estimate (s.e.)
Regression Coefficients: Positive Part			
Intercept	−3.1697 (0.0621)	−3.1728 (0.0635)	−2.6992 (0.1311)
Sex Insured			
female	−0.1339 (0.022)	−0.1323 (0.0226)	not used
male	ref. group		
Age Vehicle			
$\leq$ *2 years*	−0.0857 (0.0195)	−0.08511 (0.02)	−0.0853 (0.02)
$>$ *2 and* $\leq$ *8 years*	ref. group		
$>$ *8 years*	−0.1325 (0.0238)	−0.1327 (0.024)	−0.1325 (0.0244)
Age Insured			
$\leq$ *28 years*	0.3407 (0.0265)	0.3415 (0.027)	0.34 (0.0273)
$>$ *28 years and* $\leq$ *35 years*	0.1047 (0.0203)	0.1044 (0.0209)	0.1051 (0.0208)
$>$ *35 and* $\leq$ *68 years*	ref. group		
$>$ *68 years*	−0.4063 (0.0882)	−0.4102 (0.0897)	−0.408 (0.0895)
Private Car			
Yes	0.2114 (0.0542)	0.2137 (0.0554)	0.2122 (0.0554)
Capacity of Car			
$\leq$ 1500	ref. group		
$>$ 1500	0.1415 (0.0168)	0.1406 (0.0173)	0.1412 (0.0172)
Capacity of Truck			
$\leq$ 1	ref. group		
$>$ 1	0.2684 (0.0635)	0.2726 (0.065)	0.272 (0.065)
Comprehensive Cover			
Yes	1.0322 (0.0321)	1.0333 (0.0327)	0.8596 (0.1201)
No Claims Discount			
No	0.2985 (0.0175)	0.2991 (0.0181)	0.2999 (0.018)
Driving Experience of Insured			
$\leq$ 5 years	0.1585 (0.0251)	0.1589 (0.0259)	0.1563 (0.0258)
$>$ 5 and $\leq$ 10 years	0.0699 (0.0202)	0.0702 (0.0207)	0.0695 (0.0207)
$>$ 10 years	ref. group		
Extra Par.		$\hat{\alpha} = 2.4212$	
Regression Coefficients: Zero Part			
Intercept			−0.5124 (0.301))
Comprehensive Cover			
Yes			−0.5325 (0.3057)
Sex Insured			
female			0.3778 (0.068)
male			ref. group
Summary			
−2 Log-Likelihood	98,326	98,161	98,167
AIC	98,356	98,191	98,199

Table 15 Parameter estimates for several regression models for the severity data introduced in Illustration 2.4

Parameter	Gamma Estimate (s.e.)	Inverse Gaussian Estimate (s.e.)	Lognormal Estimate (s.e.)
Intercept	8.1515 (0.0339)	8.1543 (0.0682)	7.5756 (0.0391)
Sex Insured			
female	not sign.	not. sign.	not sign.
male			
Age Vehicle			
≤ 2 *years*	ref. group		
> 2 *and* ≤ 8 *years*	ref. group		
> 8 *years*	−0.1075 (0.02)	−0.103 (0.0428)	−0.1146 (0.0229)
Age Insured			
≤ 28 *years*	not sign.	not sign.	not sign.
> 28 *years and* ≤ 35 *years*			
> 35 *and* ≤ 68 *years*			
> 68 *years*			
Private Car			
Yes	0.1376 (0.0348)	0.1355 (0.0697)	0.1443 (0.04)
Capacity of Car			
≤ 1500	ref. group	ref. group	ref. group
> 1500 and ≤ 2000	0.174 (0.0183)	0.1724 (0.04)	0.1384 (0.021)
> 2000	0.263 (0.043)	0.2546 (0.1016)	0.1009 (0.0498)
Capacity of Truck			
≤ 1	not sign.	not sign.	not sign.
> 1			
Comprehensive Cover			
Yes	not sign.	not sign.	not sign.
No Claims Discount			
No	0.0915 (0.0178)	0.0894 (0.039)	0.0982 (0.0205)
Driving Experience of Insured			
≤ 5 years	not sign.	not sign.	not sign.
> 5 and ≤ 10 years			
> 10 years	ref. group		
Extra Par.	$\hat{\alpha} = 0.9741$	$\hat{\lambda} = 887.82$	$\hat{\sigma} = 1.167$
Summary			
−2 Log-Likelihood	267,224	276,576	266,633
AIC	267,238	276,590	266,647

make resulting price predictions and consequently, assess the price relativities among various cells.

This paper makes several distinctions in the modeling aspects involved in ratemaking. First, it considers the distinction between *a priori* and *a posteriori* risk classification in ratemaking. In *a priori* ratemaking, the process involves establishing the premium when a policyholder is new to the company so that insufficient information may be available. As additional historical claims information becomes available to the company, an *a posteriori* ratemaking becomes necessary to correct and adjust these *a priori* premiums. Second, we separately consider modeling the claim frequency, which refers to the number of times a claim is made during a specified period (e.g. typically a calendar year), and the claim severity, which refers to the size of the claim when it occurs. When combined together, a pure premium provides an estimate of the cost of the benefit provided by the insurance coverage. Finally, we also take into consideration the form of the data that may be recorded, become available to the insurance company and are used for calibrating the statistical models. In an *a priori* ratemaking process for example, the data usually are in a cross-sectional form. Here, the common practice is to use models within the family of Generalized Linear Models. However, we also examined more advanced models of count distribution models (e.g. zero-inflated and hurdle models) to accommodate the peculiarities of some frequency data such as excess zeros. On the severity side, insurance claim sizes typically exhibit more skewness and heavier tails that members of the GLM may be unable to accommodate; we find that the class of GB2 distribution models provides greater flexibility and allows for covariates to be introduced. In an *a posteriori* ratemaking process, the recorded data can come in various layers: multilevel (e.g. longitudinal) or other types of clustering. For *a posteriori* rating with panel data credibility models as well as a commercial alternative in the form of Bonus–Malus scales are discussed.

References

Antonio, K., Beirlant, J.: Actuarial statistics with generalized linear mixed models. Insur. Math. Econ. **40**(1), 58–76 (2007)

Antonio, K., Frees, E.W., Valdez, E.A.: A multilevel analysis of intercompany claim counts. ASTIN Bull. **40**(1), 151–177 (2010)

Beirlant, J., Goegebeur, Y., Verlaak, R., Vynckier, P.: Burr regression and portfolio segmentation. Insur. Math. Econ. **23**(3), 231–250 (1998)

Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J.: Statistics of Extremes: Theory and Applications. Wiley, Chichester (2004)

Bolancé, C., Guillén, M., Pinquet, J.: Time-varying credibility for frequency risk models: estimation and tests for autoregressive specifications on random effects. Insur. Math. Econ. **33**(2), 273–282 (2003)

Boucher, J.-P., Denuit, M., Guillén, M.: Risk classification for claim counts: a comparative analysis of various zero-inflated mixed Poisson and hurdle models. N. Am. Actuar. J. **11**(4), 110–131 (2007)

Boucher, J.-P., Denuit, M., Guillén, M.: Modelling of insurance claim count with hurdle distribution for panel data. In: Arnold, B.C., Balakrishnan, N., Sarabia, J.M., Mínquez, R. (eds.) Advances in Mathematical and Statistical Modeling: Statistics for Industry and Technology, pp. 45–60. Birkhäuser, Boston (2008). Chap. 4

Boucher, J.-P., Denuit, M., Guillén, M.: Number of accidents or number of claims? J. Risk Insur. **76**(4), 821–846 (2009)

Bühlmann, H.: Experience rating and credibility I. ASTIN Bull. **4**(3), 199–207 (1967)

Bühlmann, H.: Experience rating and credibility II. ASTIN Bull. **5**(2), 157–165 (1969)

Bühlmann, H., Gisler, A.: A Course in Credibility Theory and Its Applications. Springer, Berlin (2005)

Cameron, A.C., Trivedi, P.K.: Regression Analysis of Count Data. Cambridge University Press, Cambridge (1998)

de Jong, P., Heller, G.Z.: Generalized Linear Models for Insurance Data. Cambridge University Press, Cambridge (2008)

Denuit, M., Lang, S.: Non-life ratemaking with Bayesian GAMs. Insur. Math. Econ. **35**(3), 627–647 (2004)

Denuit, M., Maréchal, X., Pitrebois, S., Walhin, J.-F.: Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus–Malus Systems. Wiley, Chichester (2007)

Dickson, D.C.M., Hardy, M.R., Waters, H.R.: Actuarial Mathematics for Life Contingent Risks. Cambridge University, Cambridge (2009)

Finger, R.J.: Risk classification. In: Foundations of Casualty Actuarial Science, 4th edn., pp. 75–148. Casualty Actuarial Society, Arlington (2001). Chap. 6

Frees, E.W.: Regression Modeling with Actuarial and Financial Applications. Cambridge University Press, Cambridge (2010)

Frees, E.W., Young, V.R., Luo, Y.: A longitudinal data analysis interpretation of credibility models. Insur. Math. Econ. **24**(3), 229–247 (1999)

Haberman, S., Renshaw, A.E.: Generalized linear models and actuarial science. Statistician **45**(4), 407–436 (1996)

Hachemeister, C.A.: Credibility for regression models with application to trend. In: Kahn, P.M. (ed.) Credibility: Theory and Applications, pp. 129–163. Academic Press, New York (1975)

Hausman, J.A., Hall, B.H., Griliches, Z.: Econometric models for count data with application to patents-R&D relationship. Econometrica **52**(4), 909–938 (1984)

Jewell, W.S.: The use of collateral data in credibility theory: a hierarchical model. G. Ist. Ital. Attuari **38**, 1–6 (1975)

Kaas, R., Goovaerts, M., Dhaene, J., Denuit, M.: Modern Actuarial Risk Theory: Using R, 2nd edn. Springer, Berlin (2008)

Klugman, S.A., Panjer, H.H., Willmot, G.E.: Loss Models: From Data to Decisions, 3rd edn. Wiley, Hoboken (2008)

Lee, A.H., Wang, K., Scott, J.A., Yau, K.K.W., McLachlan, G.J.: Multi-level zero-inflated Poisson regression modelling of correlated count data with excess zeros. Stat. Methods Med. Res. **15**(1), 47–61 (2006)

Lemaire, J.: Automobile Insurance: Actuarial Models. Kluwer Academic, Dordrecht (1985)

McClenahan, C.L.: Ratemaking. In: Foundations of Casualty Actuarial Science, 4th edn., pp. 75–148. Casualty Actuarial Society, Arlington (2001). Chap. 3

McDonald, J.B.: Some generalized functions for the size distribution of income. Econometrica **52**(3), 647–663 (1984)

Mullahy, J.: Specification and testing of some modified count data models. J. Econom. **33**(3), 341–365 (1986)

Nelder, J.A., Wedderburn, R.W.M.: Generalized linear models. J. R. Stat. Soc. A **135**(3), 370–384 (1972)

Norberg, R.: A credibility theory for automobile bonus systems. Scand. Actuar. J., 92–107 (1976)

Panjer, H.H., Willmot, G.E.: Insurance Risk Models. Society of Actuaries, Schaumburg (1992)

Pinquet, J.: Allowance for cost of claims in Bonus–Malus systems. ASTIN Bull. **27**(1), 33–57 (1997)

Pinquet, J.: Designing optimal Bonus–Malus systems from different types of claims. ASTIN Bull. **28**(2), 205–229 (1998)

Pinquet, J., Guillén, M., Bolancé, C.: Allowance for age of claims in Bonus–Malus systems. ASTIN Bull. **31**(2), 337–348 (2001)

Rolski, T., Schmidli, H., Schmidt, V., Teugels, J.: Stochastic Processes for Insurance and Finance. Wiley, Chichester (1999)

Sun, J., Frees, E.W., Rosenberg, M.A.: Heavy-tailed longitudinal data modeling using copulas. Insur. Math. Econ. **42**(2), 817–830 (2008)

Winkelmann, R.: Econometric Analysis of Count Data. Springer, Berlin (2003)

Yau, K.K.W., Wang, K., Lee, A.H.: Zero-inflated negative binomial mixed regression modeling of over-dispersed count data with extra zeros. Biom. J. **45**(4), 437–452 (2003)

Yip, K.C.H., Yau, K.K.W.: On modeling claim frequency data in general insurance with extra zeros. Insur. Math. Econ. **36**(2), 153–163 (2005)