Risk classification in insurance

Emiliano A. Valdez, Ph.D., F.S.A.
Michigan State University

joint work with K. Antonio*

* K.U. Leuven

Universidad Nacional de Colombia, Bogota
23-25 April 2014

The business of insurance

- Risks (unexpected events): we face them everyday.
  - all kinds, different kinds
  - some just cause slight irritation, some with huge financial consequences
- Insurance
  - a form of transferring some or all of the financial consequences associated with uncertain events
  - pooling similar, independent risks forms the basis of actuarial practice
  - Lloyd's of London: "the contributions of the many to the misfortunes of the few"
- Earliest form of insurance
  - 1700 BC: Babylonian traders insured losses from shipment of goods against catastrophe (e.g. theft)
  - even believed to be inscripted in the early written laws of Hammurabi's code

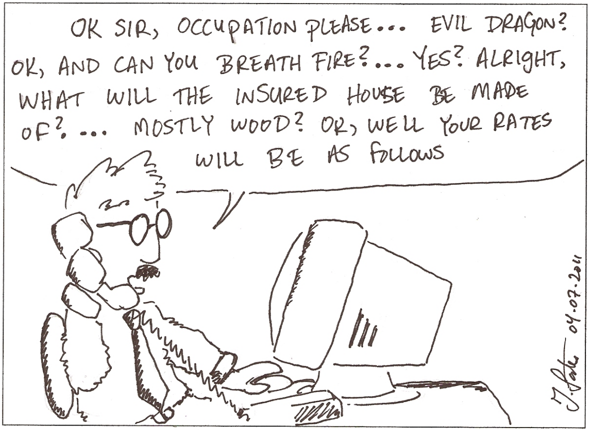
Ratemaking and risk classification

MICHIGAN STATE UNIVERSITY
Actuarial Science

- Ratemaking (or pricing): a major task of an actuary
  - calculate a predetermined price in exchange for the uncertainty
  - probability of occurrence, timing, financial impact
- Risk classification
  - the art and science of grouping insureds into homogeneous (similar), independent risks
  - the same premium cannot be applied for all insured risks in the portfolio
  - ‘good risks’ may feel paying too much and leave the company; ‘bad risks’ may favor uniform price and prefer to stay
  - spiral effect of having a disproportionate number of ‘bad risks’
  - to stay in business, you keep increasing premium

Risk classification

- Risk classification system must:
    - lead to fairness among insured individuals
    - ensure the financial soundness of the insurance company
- What risk classification is not:
    - about predicting the experience for an individual risk: impossible and unnecessary
    - should not reward or penalize certain classes of individuals at the expense of others
- See American Academy of Actuaries (AAA) Risk Classification Statement of Principles



Good underwriting—essential for fairy tale characters.

* courtesy of J. Lautier

Statistical or actuarial considerations

Constructing a risk classification system involves the selection of classifying or rating variables which must meet certain actuarial criteria:

- the rating variable must be accurate in the sense that it has a direct impact on costs
- the rating variable must meet homogeneity requirement in the sense that the resulting expected costs within a class are reasonably similar
- the rating variable must be statistically credible and reliable

a priori vs *a posteriori*

With *a priori* risk classification, the actuary lacks (individual) measurable information about the policyholder to make a more informed decision:

- unable to identify all possible important factors
- especially the unobservable or the unmeasurable
- makes it more difficult to achieve a more homogeneous classification

With *a posteriori* risk classification, the actuary makes use of an experience rating mechanism:

- premiums are re-evaluated by taking into account the history of claims of the insured
- the history of claims provide additional information about the driver's unobservable factors

Statistical techniques of risk classification

a priori techniques:

- (ordinary) linear regression, e.g. Lemaire (1985) on automobile insurance
- Generalized Linear Models (GLMs)
- Generalized Additive Models (GAMs)
- Generalized count distribution models and heavy-tailed regression

a posteriori techniques:

- experience rating schemes: No Claim Discounts, Bonus-Malus
- models for clustered data (panel data, multilevel data models)
- estimation methods: likelihood-based, Bayesian
- use of Markov chain models

Observable data for *a priori* rating

For existing portfolios, insurers typically keep track of frequency and severity data:

Policyholder file:

- underwriting information about the insured and its coverage (e.g. age, gender, policy information such as coverage, deductibles and limitations)

Claims file:

- information about claims filed to the insurer together with amounts and payments made

For each insured i, we can write the observable data as

$$\{N_i, E_i, \boldsymbol{y}_i, \boldsymbol{x}_i\}$$

where N_i is the number of claims and the total period of exposure E_i during which these claims were observed, $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{iN_i})'$ is the vector of individual losses, and $\boldsymbol{x}_i$ is the set of potential explanatory variables.

Pure premium: claim frequency and claim severity

Define the aggregate loss as

$$L_i = y_{i1} + \cdots + y_{iN_i}$$

so that frequency and severity data can be combined into a pure premium as

$$P_i = \frac{L_i}{E_i} = \frac{N_i}{E_i} \times \frac{L_i}{N_i} = F_i \times S_i,$$

where F_i refers to the claim frequency per unit of exposure and S_i is the claim severity for a given loss.

To determine the price, some premium principle can be applied (e.g. expected value):

$$\pi[P_i] = \mathrm{E}[P_i] = \mathrm{E}[F_i] \times \mathrm{E}[S_i].$$

For each frequency and severity component, the explanatory variables will be injected.

Current practice: generalized linear models

Canonical density from the exponential family:

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

where $\psi(\cdot)$ and $c(\cdot)$ are known functions, θ and ϕ are the natural and scale parameters, respectively.

Members include, but not limited to, the Normal, Poisson, Binomial and the Gamma distributions.

May be used to model either the frequency (count) or the severity (amount).

The following are well-known:

$$\mu = \mathrm{E}[Y] = \psi^{'}(\theta) \;\; \text{and} \;\; \mathrm{Var}[Y] = \phi\psi^{''}(\theta) = \phi V(\mu),$$

where the derivatives are with respect to θ and $V(\cdot)$ is the variance function.

Claim frequency models

The Poisson distribution model:

$$\Pr(N_i = n_i) = \frac{\exp(-\lambda_i)\lambda_i^{n_i}}{n_i!},$$

Risk classification variables can be introduced through the mean parameter

$$\lambda_i = E_i \exp(\boldsymbol{x}_i^{\prime}\boldsymbol{\beta}).$$

The Negative Binomial model:

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

where $\alpha = \tau/\mu$. Risk classification variables can be built through $\mu_i = E_i \exp(\boldsymbol{x}_i^{\prime}\boldsymbol{\beta})$, or through the use of a Poisson mixture with $N_i \sim \text{Poi}(\lambda_i\theta)$ with $\lambda_i = E_i \exp(\boldsymbol{x}_i^{\prime}\boldsymbol{\beta})$ and $\theta \sim \Gamma(\tau/\mu, \tau/\mu)$.

Illustration for claim counts

Claim counts are modeled for an automobile insurance data set with 159,947 policies.

No classification variables considered here.

No. 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

Generalized count distributions

MICHIGAN STATE UNIVERSITY Actuarial Science

Mixtures The NB distribution is indeed a mixture of Poisson. Other continuous mixtures of the Poisson include the Poisson-Inverse Gaussian ('PIG') distribution and the Poisson-LogNormal ('PLN') distribution. Panjer and Willmot (1992).

Zero-inflated models Here, $N = 0$ with probability p and N has distribution $\Pr(N = n|\boldsymbol{\theta})$ with probability $1 - p$. This gives the following ZI distributional specification:

$$\Pr_{\text{ZI}}(N = n|p, \boldsymbol{\theta}) = \begin{cases} p + (1-p)\Pr(N = 0|\boldsymbol{\theta}), \ n = 0, \\ (1-p)\Pr(N = n|\boldsymbol{\theta}), \ n > 0. \end{cases}$$

Hurdle models For hurdle models,

$$\begin{aligned} \Pr_{\text{Hur}}(N = 0|p, \boldsymbol{\theta}) &= p, \\ \Pr_{\text{Hur}}(N = n|p, \boldsymbol{\theta}) &= \frac{1-p}{1 - \Pr(0|\boldsymbol{\theta})}\Pr(N = n|\boldsymbol{\theta}), \ n > 0 \end{aligned}$$

Illustration with ZI and hurdle Poisson models

Using the same set of data earlier introduced.

Still no classification variables considered here.

No. of Claims	Observed	NB	ZI Poisson	Hurdle Poisson
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

Introducing risk classification in ZI and hurdle models

The common procedure is to introduce regressor variables through the mean parameter using for example

$$\mu_i = E_i \exp\left(\boldsymbol{x}_i^{'}\boldsymbol{\beta}\right)$$

and for the zero-part, use a logistic regression of the form

$$p_i = \frac{\exp\left(\boldsymbol{z}_i^{'}\boldsymbol{\gamma}\right)}{1 + \exp\left(\boldsymbol{z}_i^{'}\boldsymbol{\gamma}\right)}$$

where $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$ are sets of regressor variables.

Risk classification variables

For the automobile insurance data, description of covariates used:

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.

Parameter estimates for various count models

MICHIGAN STATE UNIVERSITY Actuarial Science

	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

Case examples

Consider the following selection of risk profiles:

- **Low**: a 45 years old male driver with a driving experience of 19 years and a NCD=40. He drives a 1,166 cc Toyota Corolla that is 22 years old. He only has a theft cover. The car is for private use.
- **Medium**: a 43 years old male driver with a driving experience of 11 years and a NCD=50. He drives 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 drives a 1,597 cc Nissan that is 4 years old. His cover is comprehensive and the car is for private use.

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

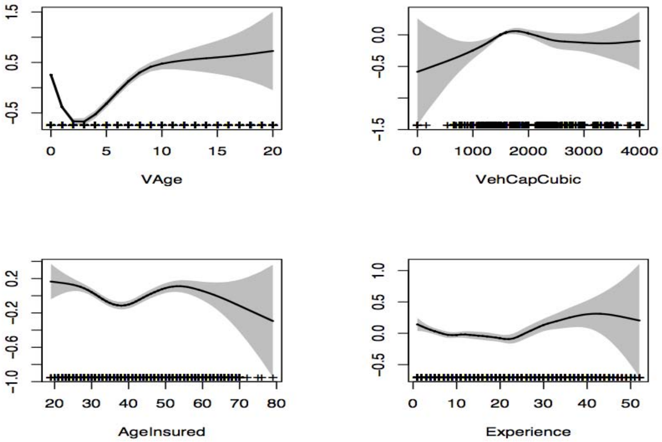
Additive regression models

Generalized additive models (GAMs) allow for more flexible relations between the response and a set of covariates.

For example:

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

Additive effects in a Poisson GAM - illustration



Some claim severity models

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

Parameter estimates for various severity models

	Gamma	Inverse Gaussian	Lognormal
Parameter	Estimate (s.e.)	Estimate (s.e.)	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			
$\leq$ *2 years*	ref. group		
$>$ *2 and* $\leq$ *8 years*	ref. group		
$>$ *8 years*	-0.1075 (0.02)	-0.103 (0.0428)	-0.1146 (0.0229)
Age Insured			
$\leq$ *28 years*	not sign.	not sign.	not sign.
$>$ *28 years and* $\leq$ *35 years*			
$>$ *35 and* $\leq$ *68 years*			
$>$ *68 years*			
Private Car			
Yes	0.1376 (0.0348)	0.1355 (0.0697)	0.1443 (0.04)
Capacity of Car			
$\leq$ 1500	ref. group	ref. group	ref. group
$>$ 1500 and $\leq$ 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			
$\leq$ 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			
$\leq$ 5 years	not sign.	not sign.	not sign.
$>$ 5 and $\leq$ 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

Other flexible parametric models for claim severity

MICHIGAN STATE UNIVERSITY Actuarial Science

The cumulative distribution functions for the Burr Type XII and the GB2 distribution are given, respectively by

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

and

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

where $B(\cdot,\cdot)$ is the incomplete Beta function.

If the available covariate information is denoted by $\mathbf{x}$, it is straightforward to allow one or more of the parameters to vary with $\mathbf{x}$.

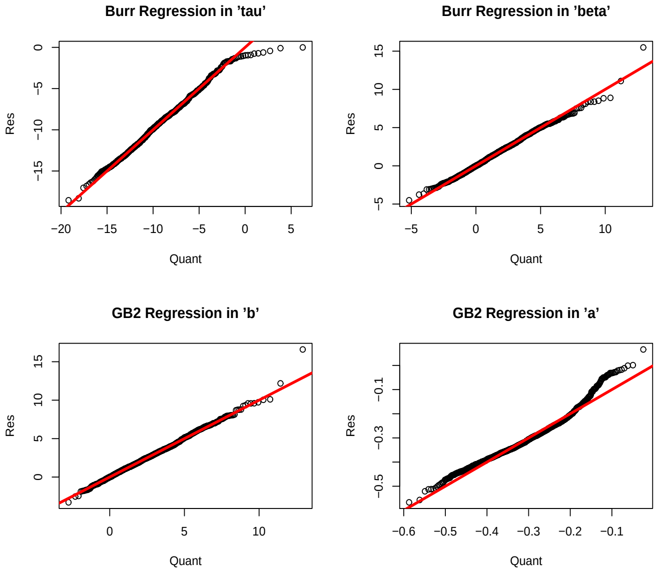
The result can be called a Burr or a GB2 regression model.

Fire insurance portfolio

	Burr (τ)	Burr (β)	GB2 (b)	GB2 (a)
Parameter	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)

Fire insurance portfolio: residual QQ plots

MICHIGAN STATE
UNIVERSITY
Actuarial Science



A posteriori risk classification

MICHIGAN STATE UNIVERSITY Actuarial Science

- When constructing an *a priori* tariff structure, not all important risk factors may be observable.
  - usually the situation for either a new policyholder or an existing one with insufficient information
  - the result is lack of many important risk factors to meet the homogeneity requirement
- For *a posteriori* risk classification, the premiums are adjusted to account for the available history of claims experience.
  - use of an experience rating mechanism - a long tradition in actuarial science
  - the premise is that the claims history reveals more of the factors or characteristics that were previously unobservable
  - the challenge is to optimally mix the individual claims experience and that of the group to which the individual belongs
  - credibility theory - a well developed area of study in actuarial science

Generalized linear mixed models

GLMMs are extensions to GLMs allowing for random, or subject-specific, effects in the linear predictor.

Consider M subjects with each subject i $(1 \leq i \leq M)$, T_i observations are available. Given the vector $\boldsymbol{b}_i$, the random effects for subject (or cluster) i, the repeated measurements $Y_{i1}, \ldots, Y_{iT_i}$ are assumed independent with density from the exponential family

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

and the following (conditional) relations hold

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

where $g(\mu_{it}) = \boldsymbol{x}^{'}_{it}\boldsymbol{\beta} + \boldsymbol{z}^{'}_{it}\boldsymbol{b}_i$.

The random effects

- Specification of the GLMM is completed by assuming that $\boldsymbol{b}_i$ $(i = 1, \ldots, M)$ are mutually independent and identically distributed with density
$$f(\boldsymbol{b}_i|\boldsymbol{\alpha}).$$
- $\boldsymbol{\alpha}$ denotes the unknown parameters in the density.
    - common to assume 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 is

$$\begin{aligned}\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \phi; \boldsymbol{y}) &= \prod_{i=1}^{M} f(\boldsymbol{y}_i|\boldsymbol{\alpha}, \boldsymbol{\beta}, \phi) \\ &= \prod_{i=1}^{M} \int \prod_{t=1}^{T_i} f(y_{it}|\boldsymbol{b}_i, \boldsymbol{\beta}, \phi) f(\boldsymbol{b}_i|\boldsymbol{\alpha}) d\boldsymbol{b}_i.\end{aligned}$$

Poisson GLMM

Let N_{it} be the claim frequency in year t for policyholder i. Assume that, conditional on b_i, N_{it} follows a Poisson with mean $\mathrm{E}[N_{it}|b_i] = \exp\left(\boldsymbol{x}_{it}^{'}\boldsymbol{\beta} + b_i\right)$ and that $b_i \sim N(0, \sigma_b^2)$.

Straightforward calculations lead to

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

and

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

We used the expressions for the mean and variance of a Lognormal distribution. For the covariance we used the fact that, given the random effect b_i, N_{it_1} and N_{it_2} are independent.

Poisson GLMM - continued

Now, if we 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 $b_i \sim N(-\frac{\sigma_b^2}{2}, \sigma_b^2)$.

This re-parameterization is commonly used in ratemaking. Indeed, we now get

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

and

$$\mathrm{E}[N_{it}|b_i] = \exp(\boldsymbol{x}_{in}^{'}\boldsymbol{\beta} + b_i).$$

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 $\exp(b_i)$.

Poisson-Gamma model

A simple and classical random effects Poisson model for panel data is constructed with assumptions

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

Here the posterior distribution of the random intercept b_i has again a Gamma with (conditional) mean and variance:

$$\begin{aligned} \mathrm{E}[b_i | N_{it} = n_{it}] &= \frac{\alpha + \sum_{t=1}^{T_i} n_{it}}{\alpha + \sum_{t=1}^{T_i} \lambda_{it}} \text{ and} \\ \mathrm{Var}[b_i | N_{it} = n_{it}] &= \frac{\alpha + \sum_{t=1}^{T_i} n_{it}}{\left(\alpha + \sum_{t=1}^{T_i} \lambda_{it}\right)^2}. \end{aligned}$$

- continued

This leads to the following *a posteriori* premium

$$\mathrm{E}[N_{i,T_i+1}|N_{it}=n_{it}]=\lambda_{i,T_i+1}\left\{\frac{\alpha+\sum_{t=1}^{T_i} n_{it}}{\alpha+\sum_{t=1}^{T_i}\lambda_{it}}\right\}.$$

The above credibility premium is optimal when a quadratic loss function is used.

The conditional expectation minimizes a mean squared error criterion.

Numerical illustration

Data consist of 12,893 policyholders observed during (fractions of) the period 1993-2003. Let N_{it} be the number of claims registered for policyholder i in period t. The model specification:

$$\begin{aligned} N_{it}|b_i &\sim \text{Poi}(\mu_{it}|b_i) \text{ and } \mu_{it}|b_i = e_{it}\exp\left(\boldsymbol{x}_{it}^{'}\boldsymbol{\beta} + b_i\right) \\ b_i &\sim N(-\sigma^2/2, \sigma^2), \end{aligned}$$

The *a priori* premium is given by

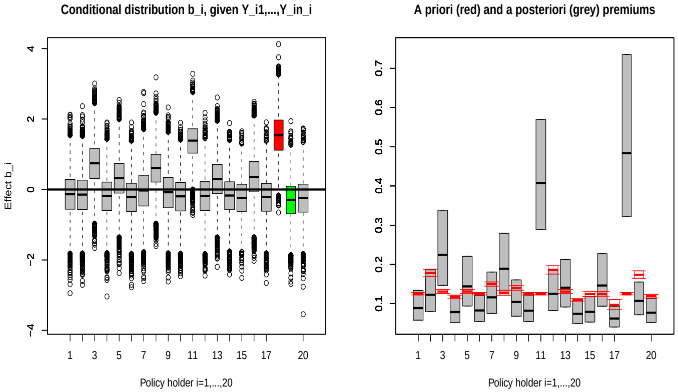
$$(\boldsymbol{a\ priori}) \qquad E[N_{it}] = e_{it}\exp\left(\boldsymbol{x}_{it}^{'}\boldsymbol{\beta}\right).$$

The *a posteriori* premium is given by:

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

The ratio of the two is called the theoretical Bonus-Malus Factor (BMF). It reflects the extent to which the policyholder is rewarded or penalized for past claims.

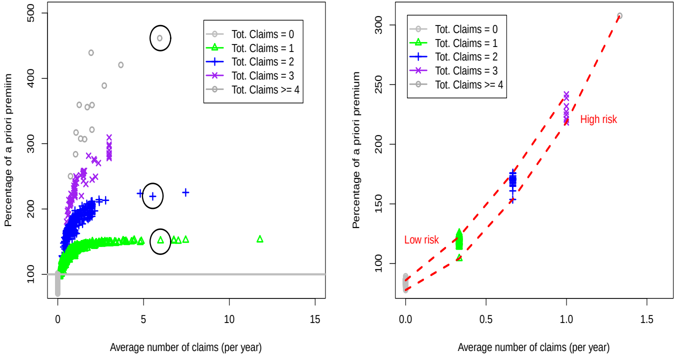
Figure 5



Left panel: Boxplot of the conditional distribution of b_i, given the history $N_{i1}, \ldots, N_{in_i}$, for a random selection of 20 policyholders. Right panel: For the same selection of policyholders: boxplots with simulations from the a priori (red) and a posteriori (grey) premium.

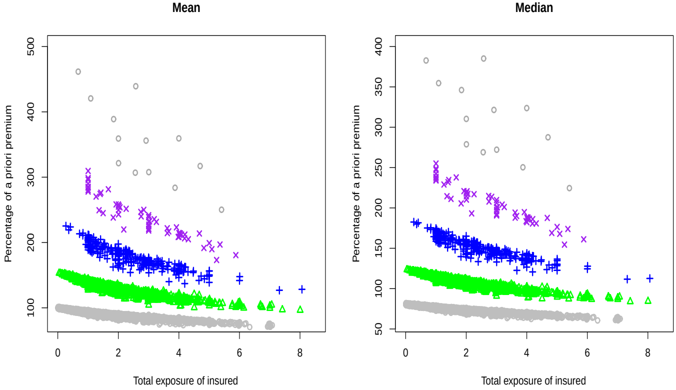
Figure 6

MICHIGAN STATE UNIVERSITY Actuarial Science



A posteriori premium expressed as percentage of the *a priori* premium (y–axis) versus the average number of claims.

Figure 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 $N_{i1}, \ldots, N_{in_i}$.

Multilevel models

- Models that are extensions to regression whereby:
  - the data are generally structured in groups, and
  - the regression coefficients may vary according to the group.
- Multilevel refers to the nested structured of the data.
- Classical examples are usually derived from educational or behavioral studies:
  - e.g. students $\in$ classes $\in$ schools $\in$ communities
- The basic unit of observation is the 'level 1' unit; then next level up is 'level 2' unit, and so on.
- Some references for multilevel models: Gelman and Hill (2007), Goldstein (2003), Raudenbusch and Byrk (2002), Kreft and De Leeuw (1995).

A multilevel model for intercompany claim counts

- We examine an intercompany database using multilevel models. We focus analysis on claim counts.
- The empirical data consists of:
    - financial records of automobile insurers over 9 years (1993-2001), and
    - policy exposure and claims experience of randomly selected 10 insurers.
- The multilevel model accommodates clustering at four levels: vehicles (v) observed over time (t) that are nested within fleets (f), with policies issued by insurance companies (c).
- More details of work are published in Antonio, Frees and Valdez (2010).

Motivation to use multilevel models

- Multilevel models allows us to account for variation in claims at the individual level as well as for clustering at the company level.
  - intercompany data models are of interest to insurers, reinsurers, and regulators.
- It also allows us to examine the variation in claims across 'fleet' policies:
  - policies whose insurance covers more than a single vehicle e.g. taxicab company.
  - possible dependence of claims of automobiles within a fleet.
- In general, it allows us to assess the importance of cross-level effects.

Multilevel model specification

Denote by $N_{c,f,v,t}$ 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:

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

where $\boldsymbol{x}_c$, $\boldsymbol{x}_{cf}$, $\boldsymbol{x}_{cfv}$ and $\boldsymbol{x}_{cfvt}$ are observable covariates.

A posteriori tariff is updated as follows:

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

where b_c, $b_{c,f}$ and $b_{c,f,v}$ are all assumed to have normal distributions.

The ratio (*a posteriori* premium/*a priori* premium) is the theoretical Bonus-Malus Factor (BMF).

Other count models considered

- Hierarchical Poisson models which include
    - Jewell's hierarchical model
- Hierarchical Negative Binomial model
- Hierarchical Zero-Inflated Poisson model
- Hierarchical Hurdle Poisson model

Figure 8

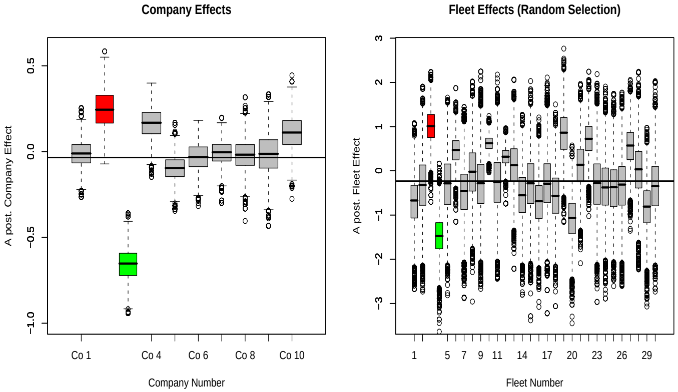


Illustration of posterior distributions of company effects and a random selection of fleet effects. A horizontal line is plotted at the mean of the random effects distribution.

Comparing the BMF factors

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					
6645	0.08435 (0.5038)	0.1725	2.05	6 (18.5)	1 (1)
7006	0.08435 (0.5038)	0.1316	1.56		0 (1)
6500	0.08435 (0.5038)	0.1329	1.58		0 (1)
Hierarchical NB with random effects for fleet and company					
6645	0.08383 (0.5038)	0.1435	1.71	6 (18.5)	1 (1)
7006	0.08383 (0.5038)	0.1435			0 (1)
6500	0.08383 (0.5038)	0.1435			0 (1)
Hierarchical ZIP with random effects for fleet and company, fixed p					
6645	0.08241 (0.5038)	0.1484	1.8	6 (18.5)	1 (1)
7006	0.08241 (0.5038)	0.1484			0 (1)
6500	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 parenthesis.

Experience rating with bonus-malus scales

A BM scale consists of a number of $s+1$ levels from $0,\ldots,s$. A new driver enters the scale at a specified level, say ℓ_0.

Drivers then transition up and down the scale according to the number of claims reported in each year.

- A claim-free year results in a bonus point where 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 the penalty by `pen'.

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 $(k, k+1)$.

- continued

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:

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

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 $r_\ell \times \lambda_{it}$.

An illustration of a BM scale

A simple example of bonus-malus scale is the so-called (-1/Top Scale).

This scale has 6 levels, numbered 0,1,. . . ,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.
- The following table represents these transitions:

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

Transition rules and probabilities

- To enable the calculation of the relativity corresponding with each level ℓ, some probabilistic concepts associated with BM scales have to be introduced.
- Details are in the paper.

Calculating the relativities

In a BM scale the relativity r_ℓ corresponding to scale ℓ corrects the *a priori* premium: *a posteriori*, the policyholder will pay $r_\ell\%$ 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.

Following Norberg (1976) with the number of levels and transition rules being fixed, the optimal relativity r_ℓ, corresponding to level ℓ, is determined by maximizing the asymptotic predictive accuracy.

Optimal relativities

Calculation of the r_ℓ's is as follows:

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

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

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

- continued

Alternatively,it is well-known that for a quadratic loss function, the optimal $r_\ell = E[\Theta|L=\ell]$.

This can be shown, easily, as follows:

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

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.

Optimal solution

The optimal relativities are given by:

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

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

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

Illustration

Using the automobile insurance data set earlier introduced wtih 159,947 policies, using the (-1/Top Scale) scheme.

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 with and without *a priori* rating taken into account:

		$r_\ell = E[\Theta \mid L = \ell]$	
Level ℓ	Pr$[L = \ell]$	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%

Remarks

This paper makes several distinctions in the modeling aspects involved in ratemaking:

- *a priori* vs *a posteriori* risk classification in ratemaking
- claim frequency and claim severity make up for the calculation of a pure premium
- the form of the data that may be recorded, become available to the insurance company and are used for calibrating models:
    - *a priori*: the data usually are cross-sectional
    - *a posteriori*: the recorded data may come in various layers: multilevel (e.g. panel, longitudinal) or other types of clustering, transitions for bonus-malus schemes

Risk classification in life insurance

Gschlossl, S., Schoenmaekers, P., Denuit, M., 2011, Risk classification in life insurance: methodology and case study, *European Actuarial Journal*, 1: 23-41.

Start with n invidiuals all aged x, observed a period of time and during this period, each individual is either dead or alive:

$$\delta_i = \begin{cases} 1, & \text{if individual } i \text{ dies,} \\ 0, & \text{otherwise} \end{cases}$$

Let τ_i be the time spent by the individual i during the period. In summary, we observe n independent and identically distributed observations (δ_i, τ_i) for $i = 1, 2, \ldots, n$.

Poisson model

If the individual is alive, his contribution to the likelihood is $\exp(-\tau_i \mu_x)$. If dead, his contribution is $\mu_x \exp(-\tau_i \mu_x)$.

Thus the aggregate likelihood contribution of all individuals observed can be expressed as

$$\mathcal{L}(\mu_x) = \prod_{i=1}^{n} (\mu_x)^{\delta_i} \exp(-\tau_i \mu_x) = (\mu_x)^{d_x} \exp(-\mathsf{E}_x \mu_x),$$

where $d_x = \sum_{i=1}^{n} \delta_i$ is the total number of deaths and $\mathsf{E}_x = \sum_{i=1}^{n} \tau_i$ is the total exposure.

This likelihood is proportional to the likelihood of a Poisson number of deaths: $\mathcal{D}_x \sim$ Poisson$(\mathsf{E}_x \mu_x)$.

Poisson regression model

There is usually heterogeneity among the individual lives (age, gender, lifestyle, income, etc.) and this can be accounted for using a Poisson regression model.

In this context, we would assume we have a set of covariates say $\mathbf{x}_i = (1, x_{i1}, x_{i2}, \ldots, x_{ik})'$, which here we include an intercept.

We link these covariates to the death rates through a log-linear function as follows:

$$\log(\mu_i) = \mathbf{x}_i'\beta$$

The β coefficients in this case have the interpretation of a percentage change, in the case of a continuous covariate, or a percentage difference in the case of a binary covariate.

Selected reference

- Antonio, K. and E.A. Valdez, 2010, Statistical concepts of *a priori* and *a posteriori* risk classification in insurance, *Advances in Statistical Analysis*, forthcoming special issue on actuarial statistics.
- Antonio, K., Frees, E.W. and E.A. Valdez, 2010, A multilevel analysis of intercompany claim counts, *ASTIN Bulletin*, 40: 151-177.
- Denuit, M., Marechal, X., Pitrebois, S. and J.-F. Walhin, 2007, *Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Systems*, John Wiley, England.
- Finger, R.J., 2001, Risk Classification in *Foundations of Casualty Actuarial Science*, chapter 6, pages 287-342, Casualty Actuarial Society.
- Gschlossl, S., Schoenmaekers, P. and M. Denuit, 2011, Risk classification in life insurance: methodology and case study, *European Actuarial Journal*, 1: 23-41.