Modeling and Performance of Bonus-Malus Systems: Stationarity versus Age-Correction

Søren Asmussen

13

Modeling and Performance of Bonus-Malus Systems: Stationarity versus Age-Correction

Søren Asmussen

Department of Mathematics, Aarhus University, asmus@imf.au.dk

Abstract

In a bonus-malus system in car insurance, the bonus class of a customer is updated from a year to the next as a function of the current class and the number of claims in the year (assumed Poisson). Thus the sequence of classes of a customer in consecutive years forms a Markov chain, and most of the literature measures performance of the system in terms of the stationary characteristics of this Markov chain. However, the rate of convergence to stationarity may be slow in comparison to the typical sojourn time of a customer in the portfolio. We suggest an age-correction to the stationary distribution and present an extensive numerical study of its effects. An important feature of the modeling is a Bayesian view, where the Poisson rate according to which claims are generated for a customer is the outcome of a random variable specific to the customer.

Keywords: actuarial mathematics; Bayes premium; equilibrium distribution; experience rating; insurance portfolio; Markov chain; motor insurance; Poisson claims; stationary distribution

1   Introduction

In the classical actuarial model for bonus-malus systems in automobile insurance (Denuit *et al.* [8] or Lemaire [14], for example), there is a finite set of bonus classes $\ell = 1, \ldots, K$. A customer having n claims and bonus class ℓ in a given year has bonus class $b(\ell, n)$ in the next for some deterministic function b (the *bonus rule*; claim sizes are ignored and only claim numbers counted). The customer has a risk parameter λ, such that the number of claims $N_0, N_1, \ldots$ in consecutive years are i.i.d. Poisson(λ), and so the sequence $L_0, L_1, \ldots$ of bonus classes is a time-homogeneous Markov chain with transition matrix $\boldsymbol{P}(\lambda) = \big(p_{k\ell}(\lambda)\big)_{k,\ell=1,\ldots,K}$ where

$$p_{\ell,\ell'}(\lambda) = \sum_{n=0}^{\infty} \mathrm{e}^{-\lambda} \frac{\lambda^n}{n!} I\big(b(\ell, n) = \ell'\big)$$

(such a customer we denote a λ-*customer*). Customers pay premium r_ℓ when in class ℓ and enter the system in some fixed class ℓ_0. The r_ℓ and ℓ_0 may be chosen

according to certain optimality and/or financial equilibrium principles (see below) or arbitrarily.

For a simple example of a bonus rule, consider the $-1/+2$ rule. Here each claim causes the bonus level to increase by 2, whereas it decreases by 1 for each claim-free year (obvious boundary modifications apply to levels $1, K-1, K$). The systems in use are often substantially more detailed, with K of order 15–25.

Much of the discussion of the literature employs stationarity modeling, measuring characteristics of the system via the stationary distribution $\boldsymbol{\pi}(\lambda) = (\pi_\ell(\lambda))_{\ell=1,\dots,K}$ (existing under the weak assumptions of irreducibility and aperiodicity). In particular, the average premium $\overline{r}(\lambda)$ of a customer with risk parameter λ is defined by

$$\overline{r}(\lambda) = \sum_{\ell_1}^{K} r_\ell \pi_\ell(\lambda) = \sum_{\ell=1}^{K} \pi_\ell \cdot \lim_{n\to\infty} p_{\ell_0,\ell}^n(\lambda) \tag{1.1}$$

where the p^n are the n-step transition probabilities. From the $\overline{r}(\lambda)$, one often proceeds to calculate the *Loimaranta efficiency*

$$e(\lambda) = \frac{\mathrm{d}\log\overline{r}(\lambda)}{\mathrm{d}\log\lambda} = \frac{\lambda\overline{r}'(\ell)}{\overline{r}(\lambda)}$$

at λ (denoted *elasticity* outside the actuarial sciences); it measures to which extent $\overline{r}(\lambda)$ is linear at λ (as should ideally be the case), with $e(\lambda) = 1$ expressing 'local linearity' at λ.

Such stationary performance measures are only meaningful if the Markov chain L attains (approximate) stationarity within the typical time a customer spends in the portfolio. For this reason much attention has been given to studying the approach of the $p_{\ell_0,\ell}^n(\lambda)$ to $\pi_\ell(\lambda)$. The rate of convergence is known to be geometric, with decay parameter the second largest eigenvalue of the transition matrix $\boldsymbol{P}(\lambda)$. However, this is an asymptotic result and so the studies are most often numerical, depicting for example the mean annual premium $\mathbb{E}_\lambda r_{L_n}$ or the total variation (t.v.) distance

$$d_{\mathrm{TV}}\big(\boldsymbol{p}_{\ell_0 \cdot}(\lambda), \boldsymbol{\pi}(\lambda)\big) = \sum_{\ell=1}^{K} \big| p_{\ell_0,\ell}^n(\lambda) - \pi_\ell(\lambda) \big| \tag{1.2}$$

as function of n (see, for example, Denuit *et al.* [8, p. 183ff]). The results are sometimes encouraging: for some bonus systems, $d_{\mathrm{TV}}\big(\boldsymbol{p}_{\ell_0 \cdot}(\lambda), \boldsymbol{\pi}(\lambda)\big) = 0$ already for $n = 4$–6. However, these are typically simple-minded systems, and for the more realistic ones, one often sees a substantial value of the t.v. distance for say $n = 30$, a value exceeding the time span a customer can be expected to stay in the portfolio. Nevertheless, the studies of the effects of the sojourn time A in the portfolio being finite are remarkably few, with Borgan *et al.* [5] being the main exception. One purpose of this paper is to go deeper into this direction and to formulate an alternative (which we call *age-correction*) to the stationarity point of view.

Bonus-malus systems may be seen as an example of *experience rating* which has as aim to calculate the premium on an individual basis by using the information available to the company. In the automobile insurance setting, we ignore in this paper profit, administration costs etc., and take the average claim size equal to 1,

so that in after a given year m, the company would want to compute its net premium c_m for year m as its best guess of the customer's λ as function of the numbers $N_0, N_1, \ldots, N_{m-1}$ of claims filed in years $0, 1, \ldots, m-1$. The naive guess is of course the average $\widehat{\lambda}_m = (N_0 + N_1 + \cdots + N_{m-1})/m$. However, a high value of $\widehat{\lambda}_m$ could be due to bad luck of an otherwise good driver and a low value luck of an otherwise bad driver. An estimate which is more fair to the customer is therefore obtained by a Bayesian view where one involves information on the population of customers in form of a prior distribution, say U, of λ, views the particular customer's λ as the outcome of a r.v. with distribution U, calculates the posterior distribution U_m and takes c_m as the mean of U_m, the *Bayes premium*.

Example 1.1. A popular choice for the prior U is a Gamma(b, μ) with density $\mu(\mu u)^{b-1}\mathrm{e}^{-\mu u}/\Gamma(b)$, in which case the posterior U_m is again Gamma with parameters

$$ b_m = b + N_0 + N_1 + \cdots + N_{m-1}, \qquad \mu_m = \mu + m, $$

and one gets the Bayes premium as the posterior mean

$$ \frac{b_m}{\mu_m} = \frac{b + N_0 + N_1 + \cdots + N_{m-1}}{\mu + m} = \frac{\mu}{\mu + m} \cdot \frac{b}{\mu} + \frac{m}{\mu + m} \cdot \widehat{\lambda}_m. $$

This has a neat interpretation as a weighted average of the population mean b/μ (the premium the company will charge without access to claims statistics) and the mean $\widehat{\lambda}_m$ of the claims, with the weight $m/(\mu + m)$ of $\widehat{\lambda}_m$ increasing to 1 as years go by and the information on the customer accumulates. $\qquad \square$

The Bayes premium enjoys the optimality property of minimizing the quadratic loss $\mathbb{E}(c_m - \Lambda)^2$ in the class of all functions c_m of $N_0, N_1, \ldots, N_{m-1}$, the natural class of predictors of Λ using information on $N_0, N_1, \ldots, N_{m-1}$. In fact, it is standard that the solution to this minimization problem is $c_m = \mathbb{E}[\Lambda \mid N_0, N_1, \ldots, N_{m-1}]$. For these facts and the general theory of Bayes premiums, see Bühlmann [6], Bühlmann & Gisler [7] and Denuit *et al.* [8, Ch. 3].[1]

Given the above optimality property, the Bayes premium can be viewed as the optimal fair choice of the insurer's premium. Presumably, the reason that bonus-malus systems are used instead in practice is that they are better understandable to the average customer who would not know about prior and posterior distributions. Nevertheless, as noted by Norberg [19], the Bayesian view is highly relevant also in bonus-malus systems but for a different purpose, to compute the premiums $r_1, \ldots, r_K$ in the different bonus levels. To this end, the idea of basing the premium on the bonus level means that one chooses the minimizer of $\mathbb{E}(c - \Lambda)^2$ in the class of all functions of the bonus level. One then needs to specify what is meant by this level, and and to avoid the dependency on m, the choice of [19] and much subsequent literature is a r.v. L_∞ distributed as $\boldsymbol{\pi}(\Lambda)$ (recall that $\boldsymbol{\pi}(\lambda)$ is the stationary distribution of the

---

[1]These references have as their main theme not the Bayes premium but rather the *credibility premium*, also called the *linear Bayes premium*, computed as the minimizer of $\mathbb{E}(a_m - \Lambda)^2$ in the class of all *linear* functions a_m of $N_0, N_1, \ldots, N_{m-1}$; for the Gamma example and many others, the Bayes premium and the linear Bayes premium coincide. The motivation for considering linear predictors only is computational ease.

Markov chain $L_0, L_1, \ldots$ when the customer's Poisson parameter is λ). Using the general $\mathcal{L}_2$-theory quoted above, the minimizer is

$$c = \mathbb{E}[\Lambda \mid L_\infty] = \sum_{\ell=1}^{K} I(L_\infty = \ell) \, \mathbb{E}[\Lambda \mid L_\infty = \ell] \,,$$

so that the optimal bonus level r_ℓ in class ℓ is $\mathbb{E}[\Lambda \mid L_\infty = \ell]$. Evaluating the conditional expectation, we get

$$r_\ell \;=\; \frac{\displaystyle\int_0^\infty \lambda \pi_\ell(\lambda) \, U(\mathrm{d}\lambda)}{\displaystyle\int_0^\infty \pi_\ell(\lambda) \, U(\mathrm{d}\lambda)} \tag{1.3}$$

Formula (1.3) gives the premium rule of the bonus system which is optimal from the point of view of minimizing the error in predicting a customer's Λ.[2]

The rule (1.3) enjoys in a certain sense the principle of *financial equilibrium* (for the company) which asserts that on average, premium incomes and payments of claims should balance. Namely, the expected claims in a year of a typical customer is $\mathbb{E}\Lambda$ and his expected premiums are $\mathbb{E}r_{L_\infty}$ *assuming that a typical customers bonus class is distributed as the stationary r.v. L_∞.* By the tower property of conditional expectations, these expressions coincide. However, the point we take in this paper is that this assumption is questionable and needs further discussion.

Remark 1.1. The above set-up ignores claim amounts (for examples of discussion involving also claim severity, see Frangos & Vontos [9] and Mahmoudvand & Hassani [16]). For simplicity, we will assume that the monetary scale is chosen such that the mean claim size is one. □

Remark 1.2. Car insurance in the real world involves several features not included in the above discussion. Examples are discounts for female drivers, penalties for young drivers, premium adjusted according to further covariates such as mileage, engine size of the car (or even color!), geographical region etc. □

Remark 1.3. Regulations vary greatly from country to country. At one extreme, all insurance companies are obliged to use the same bonus-malus system, at the other they have complete freedom. The general tendency has gone towards deregulation. A detailed survey of the situation in Europe as of year 2000 concerning such rules (as well as the use of covariates as in Remark 1.2) is in Meyer [18]. Of course, much has changed since then but still, [18] will serve to give an impression of many practical issues connected with motor insurance. □

The paper is organized as follows. In Section 2, we introduce our age-correction approach. and give some of its simple properties. Sections 3–6 then contain an extensive numerical study of its behavior in concrete case and how it compares to the traditional stationarity-based approach. A concluding discussion, including more careful references to the literature, is in Section 7, and the Appendix contains some complements as well as an outline of a more general modeling approach via Markov chains.

---

[2]Once this rule is chosen, one can also assert what is the optimal initial bonus level ℓ_0 by a further mean square error minimization, cf. [19]. We do not go into this here

2 Independent sojourn times

Motivated from the criticism of the traditional use of stationarity, we now assume that a customer stays in the portfolio only for a finite number of years A. I.e., he is in the portfolio in years $0, 1, \dots, A-1$ after entering. We further assume that A is independent of his Λ and his claim sequence $N_0, N_1, \dots$ and thereby his sequence of bonus levels, and that $\mu_A = \mathbb{E}A < \infty$. The distribution of A is denoted by F, the point probabilities by $f_a = \mathbb{P}(A = a)$, and we write $f_a^e = (f_{a+1} + f_{a+2} + \cdots)/\mu_A$, $n = 0, 1, \dots$ (f^e is the *equilibrium distribution* familiar from renewal theory, cf. [2, V.3]).

Much of the discussion of Section 1 remains relevant, only do we need for each value λ of the Poisson parameter to replace the stationary distribution $\boldsymbol{\pi}(\lambda)$ of the bonus level by the distribution $\boldsymbol{\pi}^*(\lambda)$ of the typical bonus level L^*.

We then need to specify what is meant by the 'typical bonus level' of a λ-customer, and our suggestion is to define this as a r.v. L^* with distribution

$$\pi_\ell^*(\lambda) = \mathbb{P}_\lambda(L^* = \ell) = \frac{1}{\mu_A} \mathbb{E}_\lambda \sum_{n=0}^{A-1} I(L_n = \ell) = \sum_{a=0}^{\infty} f_a^e \mathbb{P}_\lambda(L_a = \ell), \qquad (2.1)$$

denoted the *age-corrected distribution* in the rest of the paper. Expression (2.1) is fundamental for the paper and may be approached in various ways. We choose here the set-up in the following Theorem 2.1 where the interpretation is as a limiting long-term average of bonus classes of λ-customers seen by the company (for an alternative, see Appendix B).

Before stating the Theorem, we need some notation and assumptions. Let M_y denote the number of λ-customers in the portfolio in year $y = 0, \pm 1, \pm 2, \dots$, E_y the number of λ-customers entering the portfolio[3] [4] and $L_{y,c}$, $c = 1, \dots, M_y$ the bonus classes of the λ-customers, $T_{y,c} \in \{0, 1, \dots\}$ the time since they entered the portfolio. Assume that the E_y are i.i.d. with finite mean μ_E and independent of the $\left(M_{y'}, L_{y',1}, \dots, L_{y',M_{y'}}\right)$ with $y' < y$, and that that the $A, \Lambda, N_0, N_1, \dots$ for different customers are i.i.d. and independent of the E_y.

Theorem 2.1. *Under the above assumptions,*

$$\frac{1}{M_0 + \cdots + M_Y} \sum_{y=0}^{Y} \sum_{c=1}^{M_y} I(T_{y,c} = a) \to f_a^e, \qquad (2.2)$$

$$\frac{1}{M_0 + \cdots + M_Y} \sum_{y=0}^{Y} \sum_{c=1}^{M_y} I(L_{y,c} = \ell) \to \pi_\ell^*(\lambda) \qquad (2.3)$$

a.s. as $Y \to \infty$ for all ℓ, no matter initial conditions.

---

[3]For convenience, the dependence on λ is suppressed in the notation. In the Bayesian set-up, E_y a r.v. with a continuous distribution so that strictly speaking, $E_y = M_y = 0$ a.s.; the set-up and statements then have to be understood in a suitable conditional sense.

[4]To be strict, one needs also to define some ordering of customers. These matters are to our mind of formal nature rather than intrinsically difficult, and so we omit the details.

Proof. We can write

$$M_0 + \cdots + M_Y = C' + C'' + C''',$$

where C' is the time-in-portfolio-before-Y of λ-customers that first entered the portfolio at some time $y' < 0$, C'' the similar time of those that entered at some time $y = 0, \ldots, Y$ and left before $Y + 1$, and C''' the time of those that entered at some time $y = 0, \ldots, Y$ but still remain in the portfolio at time $Y + 1$ (and possibly after). Here we can bound C''' by $\widetilde{C}'''$, the total-time-in-portfolio (not necessarily before Y) of λ-customers that first entered at some time $y = 0, \ldots, Y$ but remained in the portfolio after Y. By the law of large numbers, $C'' + \widetilde{C}''' \sim Y \mu_E \mu_A$. Further,

$$C' = M_0 - E_0 = \mathrm{O}(1) = \mathrm{o}(Y),$$

$$\mathbb{E}C''' = \mu_E \sum_{y=0}^{Y} \mathbb{P}(A > Y - y) = \mu_E \sum_{y=0}^{Y} \mathbb{P}(A > y) \uparrow \mu_E \mu_A = \mathrm{o}(Y).$$

Combining these facts gives

$$\frac{1}{Y}(M_0 + \cdots + M_Y) \to \mu_E \mu_A. \tag{2.4}$$

A similar argument shows that the total time C'' or C''' λ-customers ever spend is bonus class ℓ (not necessarily before Y) is of order

$$\mu_E Y \sum_{n=0}^{\infty} \mathbb{P}_\lambda(L_n = \ell, A > n) = \mu_E Y \mu_A \pi_\ell^*(\lambda)$$

and that the C'' contribution to the l.h.s. of (2.3) dominates the C' and C''' contributions. Combining with (2.4) gives (2.3). The proof of (2.2) is similar, though slightly easier. $\qquad\square$

Remark 2.1. The analysis in the proof of Theorem 2.1 is similar to the one of a discrete time G/G/∞ queue. M_y then plays the role of the queue length at time y, and $T_{y,c}$ as the elapsed service time of customer y, c. We will not use this connection and hence leave out further details. $\qquad\square$

It may be noted that expression (2.1) may be evaluated in closed analytical form. To this end, we need the *fundamental matrix* $\boldsymbol{Z}(\lambda)$ of the Markov chain given by $\boldsymbol{Z}(\lambda) = (\boldsymbol{I} - \boldsymbol{P}(\lambda) + \mathbf{1}\boldsymbol{\pi}(\lambda))^{-1}$ where $\mathbf{1}$ is the (row) vector with 1 at all entries ([2, p. 31]). Further let $\widehat{f}[z] = \sum_0^\infty f_n z^n$ denote p.g.f. of F and use the same notation $\widehat{f}[\boldsymbol{C}] = \sum_0^\infty f_n \boldsymbol{C}^n$ for a square matrix $\boldsymbol{C}$. Then, with $\boldsymbol{e}_\ell$ the ℓth (row) unit vector:

Theorem 2.2. *The distribution $\boldsymbol{\pi}^*(\lambda)$ with point probabilities (2.1) is given by*

$$\boldsymbol{\pi}^*(\lambda) = \boldsymbol{\pi}(\lambda) + \frac{1}{\mu_A} \boldsymbol{e}_{\ell_0}\big(\boldsymbol{Z}(\lambda)(\boldsymbol{I} - \widehat{f}[\boldsymbol{P}(\lambda)])\big).$$

For the proof, see the Appendix.

Remark 2.2. If the univariate p.g.f. $\widehat{f}[z]$ is available in closed form, then so is usually the matrix version $\widehat{f}[\boldsymbol{C}]$ by obvious changes in the expression for $\widehat{f}[z]$. For example for a negative binomial distribution of order 2 with point probabilities $f_n = (1 - \rho)^2 (n + 1)\rho^n$, $n = 0, 1, \ldots$, we have $\widehat{f}[z] = (1 - \rho)^2/(1 - \rho z)^2$ and $\widehat{f}[\boldsymbol{C}] = (1 - \rho)^2(\boldsymbol{I} - \rho\boldsymbol{C})^{-2}$. $\qquad\square$

6

3 Numerical set-up

3.1 The bonus systems

We have selected three rather different systems for our numerical studies. Doing so, our source has been the survey in Meyer [18] treating the situation in most European countries (as well as Japan and the US) around 1999. Important characteristics of a bonus system is the number K of classes and the *spread factor*, defined as the ratio between the highest premium r_K and the lowest r_1. We chose systems from three countries, Ireland, Italy and Germany. Ireland has a small number $K = 6$ of classes and a low spread factor of 2, Germany has a high number $K = 29$ of classes and a high spread factor of 8.2, whereas Italy is intermediate with $K = 18$ of classes and spread factor 4. More detail on the various systems are given below. It should be noted that each system may only be one among several in the particular country when [18] was published and that much may have changed since then. However, our point is not to analyze systems that are necessarily in current use but rather that our examples both show diversity and are typical of many other systems.

The premiums r_ℓ, often called *relativities*, are traditionally given in percent of the premium in the initial class ℓ_0 or some other reference class, and when presenting the three systems, we follow that tradition (as does Meyer [18]). However, later we shall renormalize to get financial equilibrium.

3.2 Trial sojourn time distributions

We selected four trial distributions for the distribution of the sojourn time A of a customer in the portfolio. Two are negative binomial with point probabilities

$$f_n = \binom{n+1}{2}(1-\rho)^3\rho^n, \qquad n = 1, 2, \ldots \tag{3.1}$$

(i.e., A is distributed as $1+B_1+B_2+B_3$ where B_1, B_2, B_3 are independent geometric(ρ) on $\{0, 1, \ldots\}$). Here ρ was chosen to make the means $1 + 3\rho/(1-\rho)$ equal to 7 and 13, and the two distributions are denoted $NB(7)$, resp. $NB(13)$. The two other, denoted $U(6.5)$ and $U(12.5)$, were taken as uniform distributions on $\{1, 2, \ldots, 12\}$, resp. $\{1, 2, \ldots, 24\}$, i.e. with roughly the same means. The distributions and their equilibrium distributions given by (2.2) are illustrated in Figure 1.

In the numerical calculations, the distributions were truncated at $n = 50$, except for Figure 14 where the truncation point was $n = 2000$.

3.3 Bayesian assumptions

We have taken the distribution U of the customers' λ parameter to be exponential with mean 0.1. The value 0.1 is from Lemaire & Zi [13, p. 288] who argue this to be typical in many countries. The exponential assumption is from Bichsel [3], who fitted a gamma distribution to data and found the shape parameter to be close to 1.

Motivated by this assumption, we have in many of the illustrations selected four values of λ, 0.04, 0.08, 0.16 and 0.32, i.e. two below the population mean and
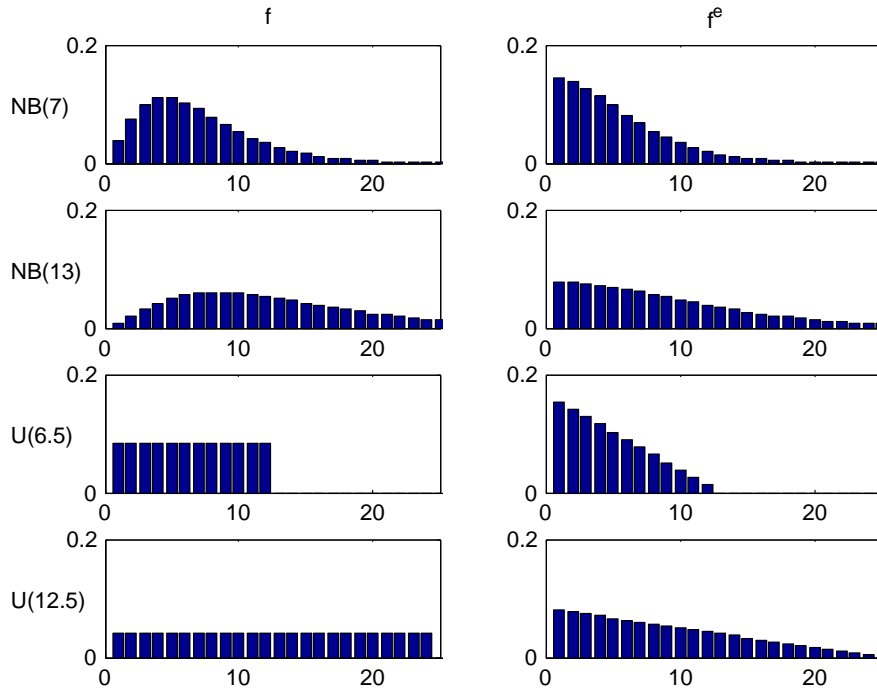
Figure 1: Trial distributions and their equilibrium distributions

two above. For the spread, note that 0.04 and 0.32 roughly correspond to the 5%, resp. 95%, quantiles in the exponential distribution with mean 0.1.

4 Convergence to stationarity. Age-corrected distributions

4.1 Ireland

The Irish system is very simple with $K = 6$ classes and transition rules as in Table 1. The initial class is $\ell_0 = 6$.

Table 1: Bonus rules, Ireland

ℓ	r_ℓ	$n = 0$	$n = 1$	$n = 2+$
6	100	5	6	6
5	90	4	6	6
4	80	3	6	6
3	70	2	5	6
2	60	1	4	6
1	50	1	3	6

The convergence speed to the stationary distribution is illustrated in two figures. The first, Figure 2, shows the shape of the transient λ-distributions of L_n for four selected values $n = 5, 10, 15, \infty$ of n ($n = \infty$ corresponds to the stationary distribu-

8

tion) and the four selected values $0.04, 0.08, 0.16, 0.32$ of λ, and the next, Figure 3, plots the t.v. distance (1.2) to the stationary distribution.
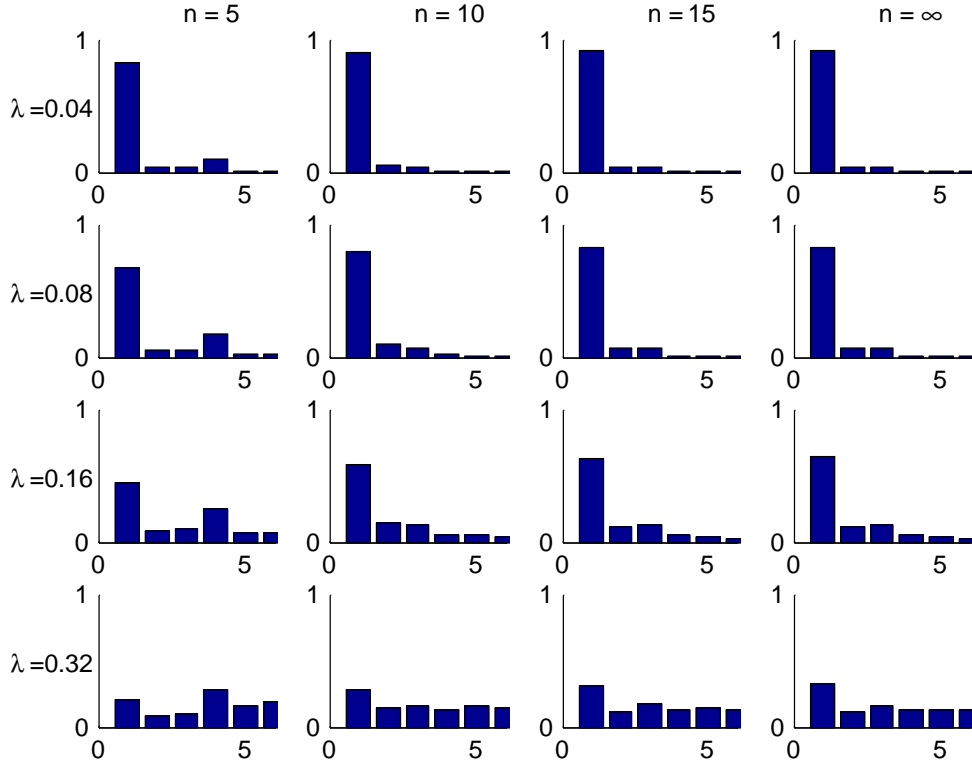

Figure 2: Transient distributions, Ireland

The shape of these figures may be understood from the transition rules. Consider for example a customer with $\lambda = 0.04$. Here most of the mass of π is concentrated in class 1, but class 1 can at earliest been reached in year 5. This explains the steep drop in Figure 3 in the t.v. distance between years 4 and 5. When looking at the bar plots in Figure 2 for the distribution of his class in different years, consider for example year 5 and note that w.p. $e^{-0.2} = 0.82$ he have will have no claims in the first 5 years, so 0.82 is precisely the mass at class 1. W.p. $0.2e^{-0.2} = 0.16$ he have will have exactly one claim. If this happens in year 0, his sequence of states in years $1, 2, 3, 4, 5$ is $6, 5, 4, 3, 2$. The similar sequences for a claim in year $1, 2, 3$, resp. 4 are

$$5, 6, 5, 4, 3, \qquad 5, 4, 6, 5, 4, \qquad 5, 4, 3, 5, 4, \qquad 5, 4, 3, 2, 4$$

Since any of the years 0,1,2,3,4 are equally likely for the claim. this explains that class 4 is more likely than classes 2,3, which is of course not the case for a good customer in stationarity ($n = \infty$). The possibility of two or more claims giving mass in states 5,6 is just 0.02 and hence negligible.

Similar remarks apply to other values of λ and n as well as the parallel figures for Italy and Germany to follow, but we shall not give the details.

The figures shows the fastest convergence rate among our three selected systems, and also that the rate is not that crucially depending on the value of λ. The explanation could be related to the simplicity of the Irish system.

Our next figure, Figure 4, plots the age-corrected distribution $\boldsymbol{\pi}^*(\lambda)$ for our four selected values of λ (one in each column) and our four trial distributions together
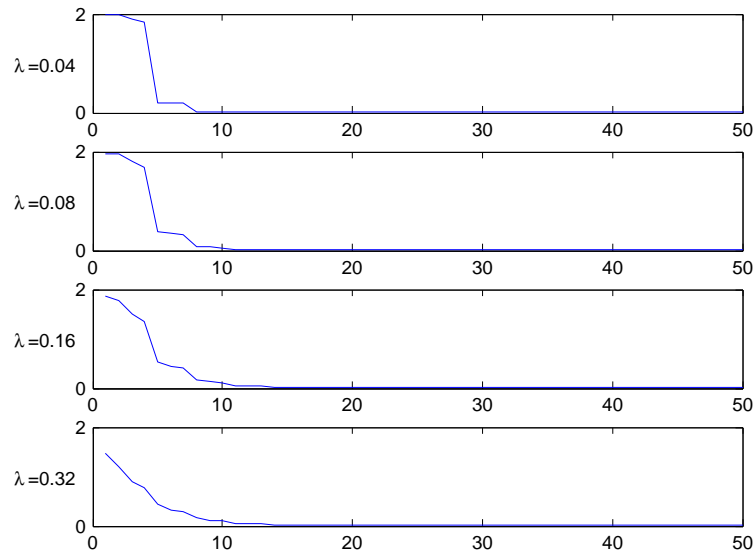
Figure 3: T.v. convergence rate, Ireland
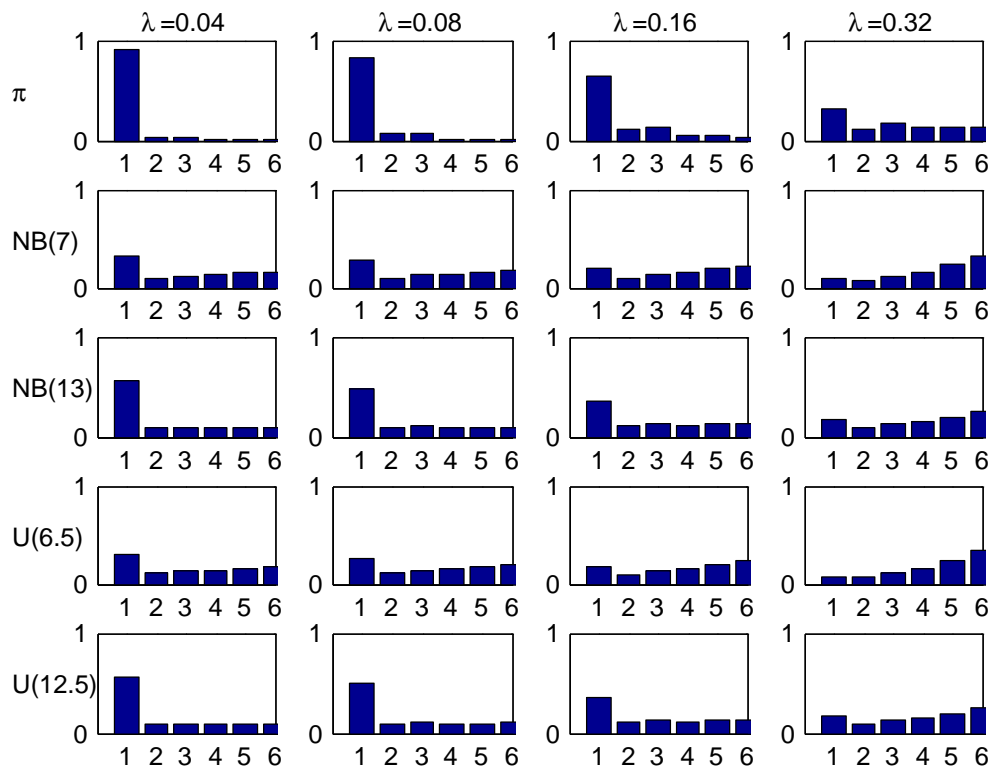

Figure 4: Stationary and age-corrected distributions, Ireland

with the stationary distribution $\boldsymbol{\pi}(\lambda)$ (the benchmark of much literature) on top. It is seen that the agreement within columns is relatively good, with the most marked differences for small values of λ. The explanation could be the relatively fast convergence rate in the Irish system

4.2 Italy

The Italian system is intermediate with $K = 18$ classes and transition rules as in Table 2. The initial class is $\ell_0 = 14$.

Table 2: Bonus rules, Italy

ℓ	r_ℓ	$n = 0$	$n = 1$	$n = 2$	$n = 3$	$n = 4+$
18	200	17	18	18	18	18
17	175	16	18	18	18	18
16	150	15	18	18	18	18
15	130	14	17	18	18	18
14	115	13	16	18	18	18
13	100	12	15	18	18	18
12	94	11	14	17	18	18
11	88	10	13	16	18	18
10	82	9	12	15	18	18
9	78	8	11	14	17	18
8	74	7	10	13	16	18
7	70	6	9	12	15	18
6	66	5	8	11	14	17
5	62	4	7	10	13	16
4	59	3	6	9	12	15
3	56	2	5	8	11	14
2	53	1	4	7	10	13
1	50	1	3	6	9	12

Figures 5 and 6 are parallel to Figures 2 and 3 for Ireland, illustrating the convergence speed to stationarity. As a new feature, we observe some gaps in the transient distributions. Consider for example $n = 5$ where Figure 6 shows that classes 10 and 11 can not be attained. The explanation is that with 0 claims in years $0, 1, 2, 3, 4$ one will go down 5 classes from 14 to 9, but with 1 claim one goes down 4 and up two, so to 12, and with more than two of course even higher. One sees also a somewhat slower rate of convergence to stationarity.

Finally the age-corrected distribution $\boldsymbol{\pi}^*(\lambda)$ are plotted in Figure 7 together with the stationary distribution $\boldsymbol{\pi}(\lambda)$. One sees a marked worse agreement within columns than for Ireland. The most marked differences occur for small values of λ, with one feature being a considerable concentration of the $\boldsymbol{\pi}^*(\lambda)$ (but not of $\boldsymbol{\pi}(\lambda)$) close to the inital class 14. Again, the most natural explanation is the slow convergence rate.
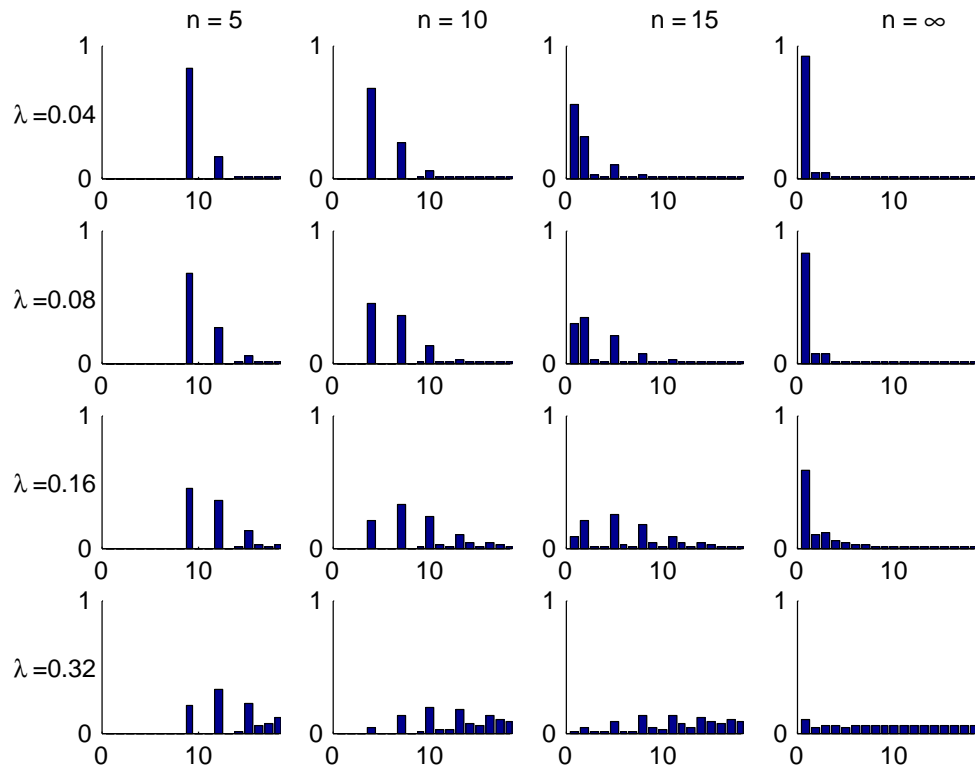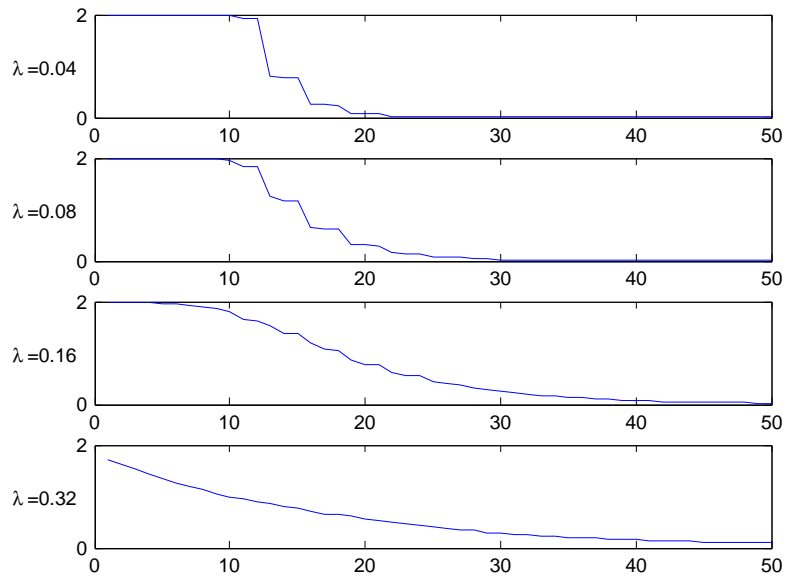
Figure 5: Transient distributions, Italy

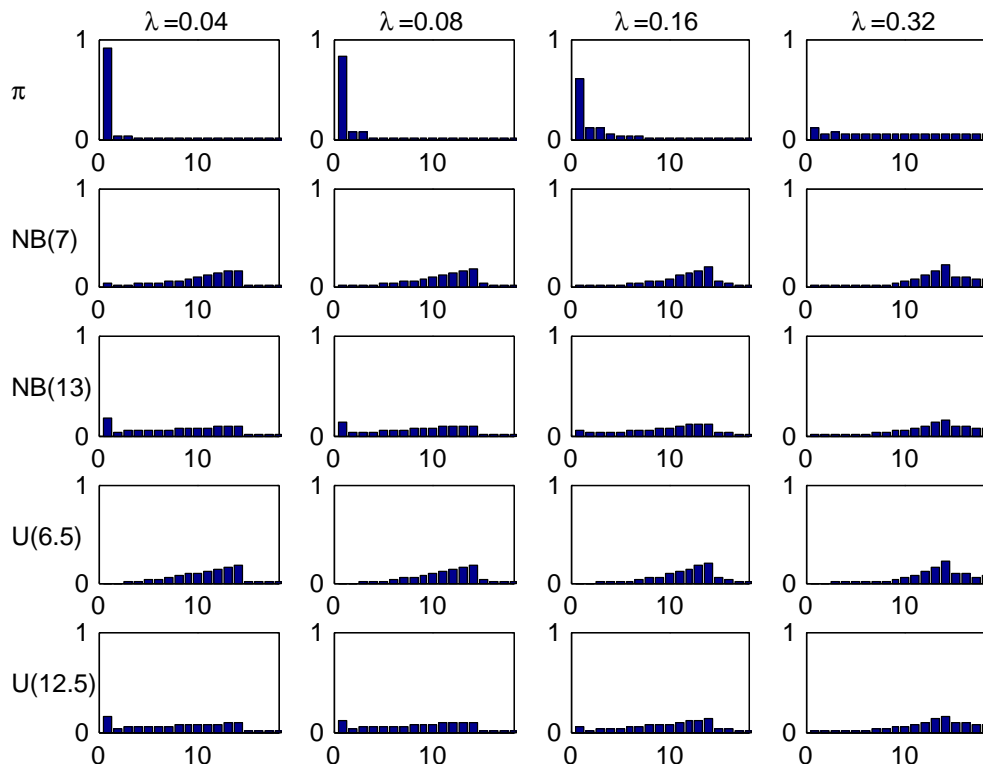
Figure 6: T.v. convergence rate, Italy

Figure 7: Stationary and age-corrected distributions, Italy

4.3 Germany

The German system is rather elaborate. It has a large number of classes, $K = 29$, initial class 26, and quite detailed rules for the new class after one or more claims. For example, after one claim the customer moves up 14 classes when in class 1, always to class 17 when in classes 6–11, and up 3 classes when in classes 19–22. The rules for some selected cases are given in Table 3; for full details, see Meyer [18] or Mahmoudvand *et al.* [17].

Table 3: Bonus rules, Germany

ℓ	r_ℓ	$n = 0$	$n = 1$	$n = 2$	$n = 3$	$n = 4+$
29	245	25	29	29	29	29
25	100	24	26	29	29	29
20	55	19	23	26	27	29
15	40	14	21	25	27	29
10	35	9	17	24	26	29
5	30	4	16	22	24	29
1	30	1	15	22	24	29

A quite special feature of the German system is the very high initial class, 26, meaning that a customer at earliest can reach the lowest premium level in class 1 after 25 years! This clearly shows up in the following Figures 8, 9, 10, for example in the $\lambda = 0.04$ row in Figure 9 where the t.v. distance from the stationary distri-

bution is substantial up to time $n = 25$, and in the comparisons of age-corrected distributions in Fig. 10 which shows the same phenomenon as for Italy, a strong concentration of the $\boldsymbol{\pi}^*(\lambda)$ (but not of $\boldsymbol{\pi}(\lambda)$) close to the inital class 26.
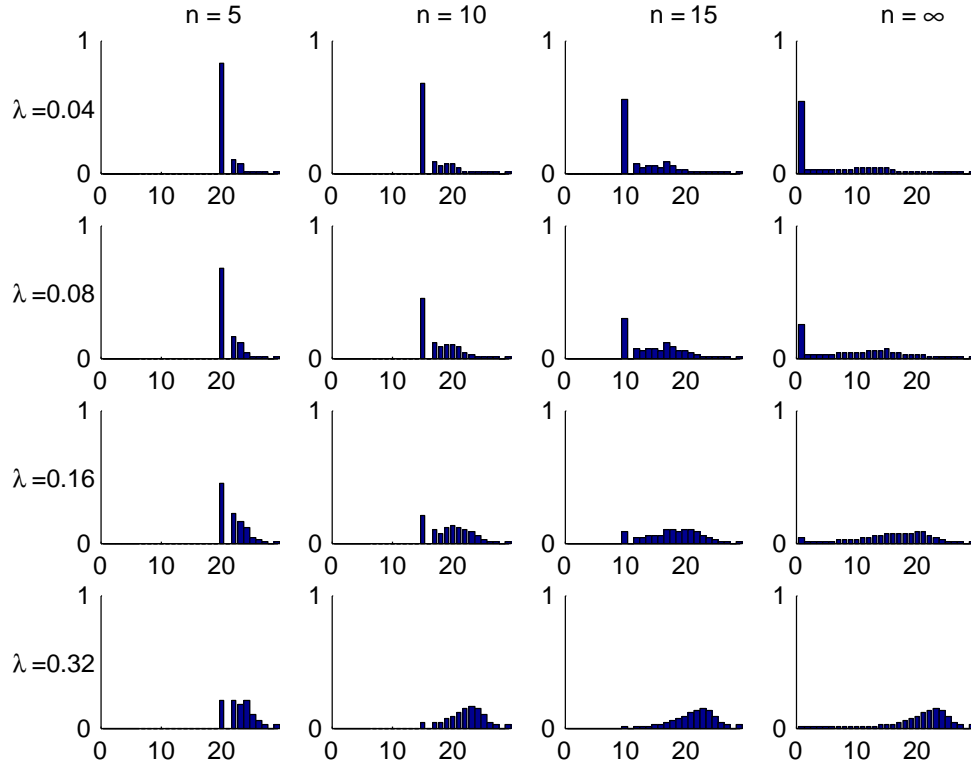

Figure 8: Transient distributions, Germany
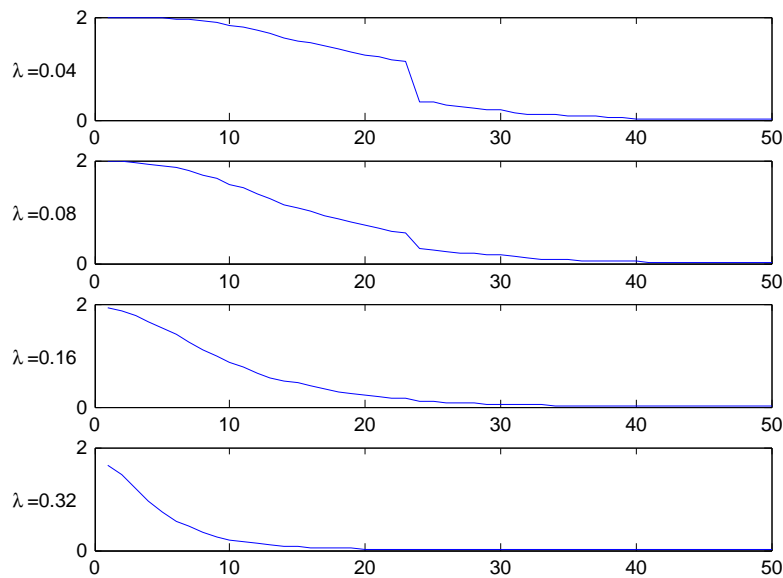
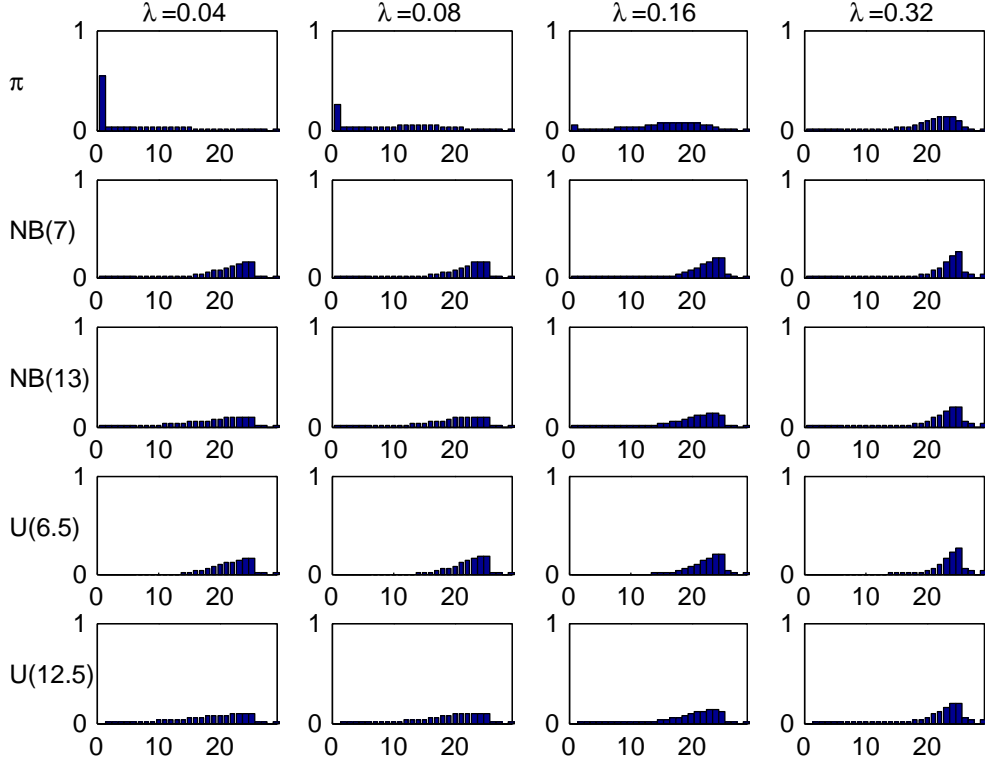**Figure 9:** T.v. convergence rate, Germany

14

Figure 10: Stationary and age-corrected distributions, Germany

5 Relativities

We now turn to the influence of finite customer sojourn times on the Bayes premium, proceeding as follows. For each of the three selected bonus systems and of the four trial sojourn time distributions, we first compute our age-corrected alternatives $\pi_\ell^*(\lambda)$ to the stationary distribution $\pi_\ell(\lambda)$ by means of (2.1) and next the Bayes premium r_ℓ^* in bonus class ℓ by means of the analogue of (1.3)

$$r_\ell^* = \frac{\displaystyle\int_0^\infty \lambda \pi_\ell^*(\lambda)\, U(\mathrm{d}\lambda)}{\displaystyle\int_0^\infty \pi_\ell^*(\lambda)\, U(\mathrm{d}\lambda)} \tag{5.1}$$

The results are in the following three Figures 11, 12, 13. The legends are solid red for distribution $NB(6)$, dotted red for $NB(13)$, solid blue for $U(6)$, and dotted blue for $U(12)$. As supplement we also compute the Bayes premium corresponding to the stationary distribution $\pi_\ell(\lambda)$ (dotted black) and supplement with the premium corresponding to the given relativities for the bonus system (e.g. 50, 60, 70, 80, 90, 100 for Italy) in solid black; whereas the Bayes premium automatically yields financial equilibrium, cf. the discussion following (1.3), we here need to normalize to satisfy this requirement.

When interpreting the figures, we first note that it does not contradict financial equilibrium that for a given country, one set of of relativities is below the other. For example, all relativities corresponding to one of our four trial sojourn time
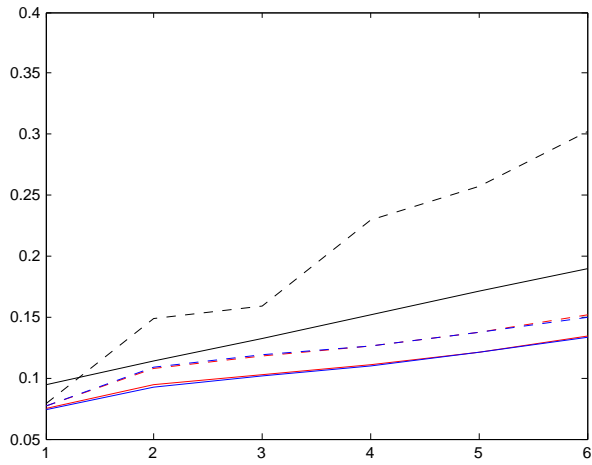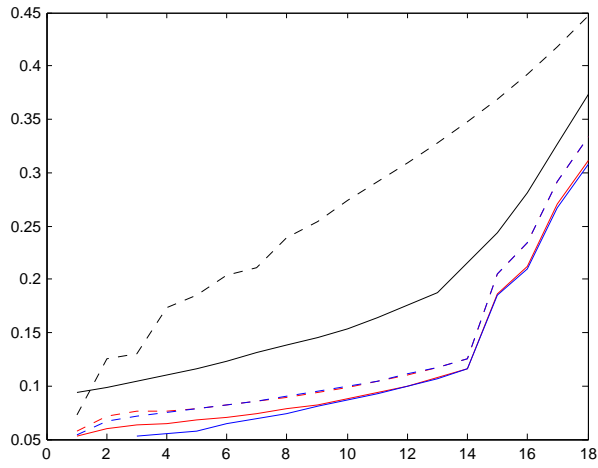
Figure 11: Relativities, Ireland

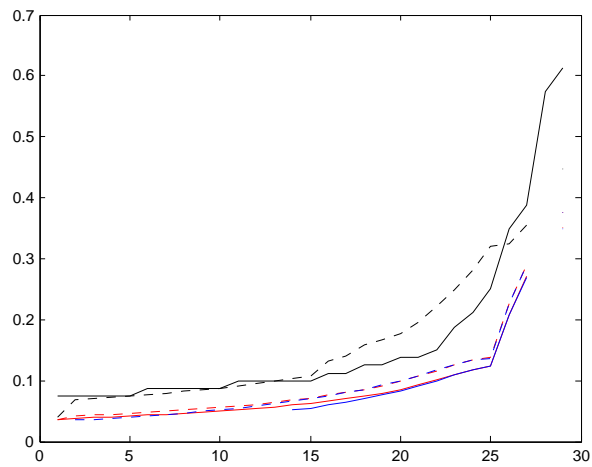**Figure 12:** Relativities, Italy

Figure 13: Relativities, Germany

16

distributions (colored graphs) are below the given relativities (solid black graph). But the explanation is simply that the relativities should be weighted with the age corrected, resp. the stationary distribution, and the age corrected distributions have a region of importance which is more shifted towards high classes.

We next note that the two distributions $NB(7)$, $U(6.5)$ with the low mean are quite close, in some cases even hard to distinguish. Dirstributions $NB(13)$, $U(12.5)$ have a roughly doubled mean. As could be expected, this puts them closer to the Bayesian relativity computed w.r.t. $\boldsymbol{\pi}(\lambda)$. The convergence rate appears quite slow, however, and this is further illustrated in the following Figure 14. We took here the Irish system and compared the stationarity-based Bayesian relativity (solid black) to those of four versions of the negative binomial distribution (3.1), one with mean 10 (dotted red), one with mean 100 (dashed red), one with mean 200 (dash-dotted red), and one with mean 400 (solid red). The figure confirms the expectation of convergence, but shows also that (as just noted) it is slow.
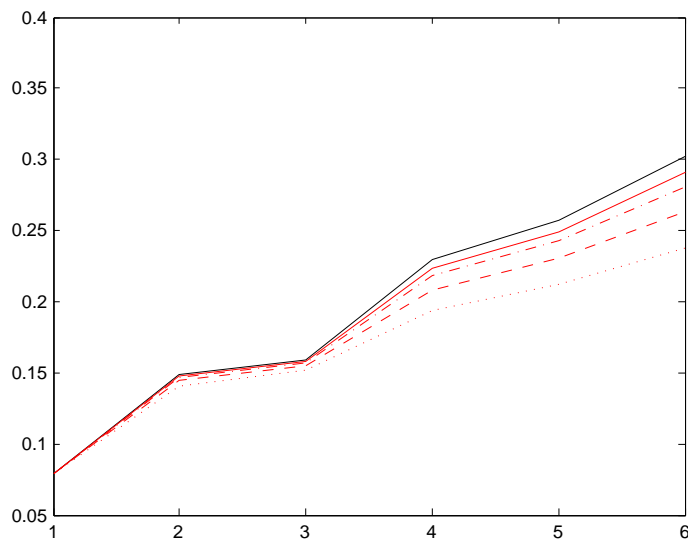

Figure 14: Convergence of relativities, Ireland

6 Age-corrected average premiums

Analogous with the stationarity-based definition (1.1) of the average premium $\bar{r}(\lambda)$ of a λ-customer, we define the age-corrected version as

$$\bar{r}^*(\lambda) = \sum_{\ell_1}^{K} r_\ell \pi_\ell^*(\lambda) \tag{6.1}$$

The $\bar{r}^*(\lambda)$ are plotted in Figures 15, 16, 17 one for each of the three bonus systems and the same 6 cases as for the relativities in Section 5, with the same legends.

We see a considerable difference between the two stationarity-based average premiums (solid black and dotted black) for Ireland and Italy, whereas they appear almost identical for Germany. The age-corrected average premiums are again quite different, and exhibit somewhat similar behavior as the relativities in Figures 11, 12, 13.
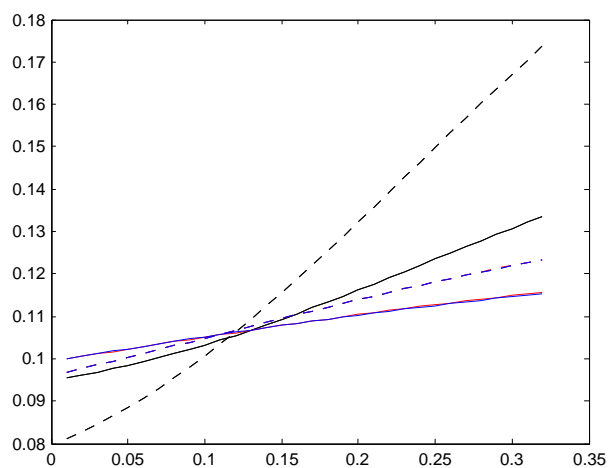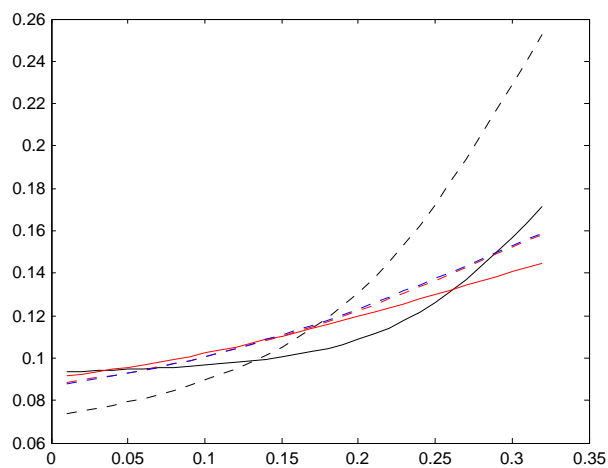
Figure 15: The $\bar{r}^*(\lambda)$, Ireland

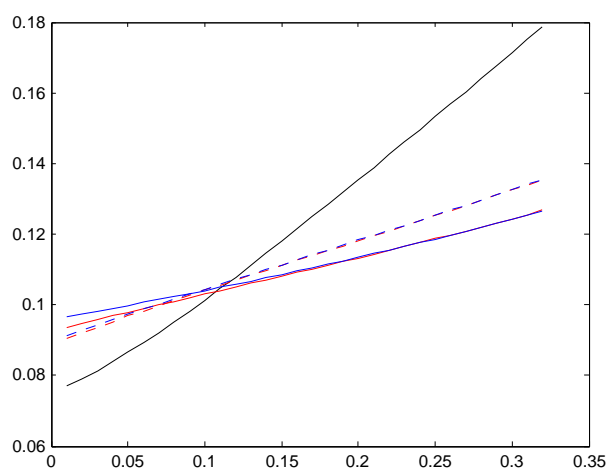**Figure 16:** The $\bar{r}^*(\lambda)$, Italy

Figure 17: The $\bar{r}^*(\lambda)$, Germany

18

The ideal fairness criterion for a Bayesian premium rule is that the premium for a λ-customer should come as close to λ as possible. This can never be perfectly achieved: since the premium in the lowest bonus class is non-zero, a customer with a small λ will always pay too much, and since the premium in the highest class is finite, a customer with a large λ will always pay too little. The figures show that this effect is substantially more marked for the age-corrected average premiums than for the stationarity-based ones. The explanation is natural: if the customer has a finite sojourn time, the system will have less time to learn about his risk characteristics in the form of λ than if he had been there for ever, as is the false assumption underlying the stationarity-based calculations.

7 Concluding remarks

In this paper, we have inspected how reasonable it is to view bonus-malus system via the stationary distribution, as is usually done. The conclusion is that in many cases the transient distributions are quite far from the stationary ones, and that this has considerable consequences on the computation of such quantities as Bayesian relativities and average premiums.

We do not necessarily insist that our trial distributions for the sojourn time in the portfolio have the relevant time span. A motor insurance may be terminated for example just if the insured gets a new car. In that case, he will typically continue with a new policy in the same company, but not enter in the same level ℓ_0 as completely new insurers. Similar remarks to change of company, where usually some information on present bonus class or general previous claim statistics is passed from the new company to the old.

Examples of numerical studies of special bonus-malus systems are, for example, in Lemaire [12], Lemaire & Zi [13] and Mahmoudvand et $al.$ [17]. These papers differ from the present one by not going into the Bayesian aspect. Here the more closely related literature is Norberg [19] and Borgan et $al.$ [5]. In particular, [5] contains ideas on how to get away from the stationary point of view. As analogue of our $\boldsymbol{\pi}^*(\lambda)$, [5] suggests a distribution of the form $\sum_0^\infty w_n \boldsymbol{e}_{\ell_0} \boldsymbol{P}(\lambda)^n$ where the w_n are suitable weights summing to one. It is also briefly mentioned that one interpretation corresponds to sampling a customer at random from the portfolio, but the connection to our $\boldsymbol{\pi}^*(\lambda)$ which is obtained by taking $w_n = f_n^e$ is not given. Also the concept of sampling a customer at random is not explained very clearly, cf., e.g., our Theorem 2.1 and Appendix B below, and in key examples the w_n are taken constant on an interval whereas f_n^e is decreasing. Nevertheless, [5] contains some key ideas related to this paper, and to our mind, the paper has received surprisingly little attention in subsequent literature (but see Denuit et $al.$ [8, Ch. 8]).

Of further classical references in the bonus-malus area not cited elsewhere in the text, we mention in particular (in chronological order) Grenander [10], Loimaranta [15], Bonsdorff [4] and Rolski et $al.$ [20].

References

1. Asmussen, S; Albrecher. H. *Ruin Probabilities* (2nd ed.). World Scientific: Singapore, 2010.

2. Asmussen, S. *Applied Probability and Queues* (2nd ed.). Springer-Verlag: New York, 2003.

3. Bichsel, F. Erfahrungstariffierung in der Motorfahrzeug-Haftphlicht-Versicherung. *Mitt. Verein. Schweiz. Versich. Math.* **1964**, 119–130.

4. Bonsdorff, H. On the convergence rate of bonus-malus systems. *ASTIN Bull.* **1992**, *22*, 217–223.

5. Borgan, Ø.; Hoem, J.M.; Norberg, R. A nonasymptotic criterion for the evaluation of automobile bonus systems. *Scand. Actuar. J.* **1981**, 165–178.

6. Bühlmann, H. *Mathematical Methods in Risk Theory.* Springer-Verlag: Berlin Heidelberg,1970.

7. Bühlmann, H.; Gisler, A. *A Course in Credibility Theory and its Applications.* Springer-Verlag: Berlin Heidelberg, 2005.

8. Denuit, M.; Marchal, X.; Pitrebois, S.; Walhin, J.F. *Actuarial Modelling of Claims Count.* Wiley: Chichester, 2007.

9. Frangos, N.E.; Vontos, S.D. Design of optimal bonus-malus systems with a frequency and a severity component on an individual basis in automobile insurance. *ASTIN Bull.* **2001**, *31*, 1–22.

10. Grenander, U. Some remarks on bonus systems in automobile insurance. *Scand. Actuar. J.* **1957**, *40*, 180–197.

11. Lemaire, J. *Automobile Insurance: Actuarial Models.* Kluwer: Boston, 1985.

12. Lemaire, J. A comparative analysis of most European and Japanese Bonus-malus systems. *J. of Risk and Insurance* **1988**, *LV*, 660–681.

13. Lemaire, J.; Zi, H. A comparative analysis of 30 bonus-malus systems. *ASTIN Bull.* **1994**, *24*, 287–309.

14. Lemaire, J. *Bonus-Malus Systems in Automobile Insurance.* Kluwer: Boston, 1995.

15. Loimaranta, K. Some asymptotic properties of bonus systems. *ASTIN Bull.* **1972**, *6*, 233–245.

16. Mahmoudvand,R.; Hassani , H. (2009) Generalized bonus-malus system with a frequency and a severity component on an individual basis in automobile insurance. *ASTIN Bull.* **2009**, *39*, 307–315.

17. Mahmoudvand, R.; Edalati, A.; Shokooi, F. (2013) Bonus-malus system in Iran: an empirical evaluation. *J. Data Sci.* **2013**, *11*, 29–41.

18. Meyer, U. *Third Party Motor Insurance in Europe. Comparative Study of the Economical-Statistical Situation.* University of Bamberg: 2000.

19. Norberg, R. A credibility theory for automobile bonus system. *Scand. Actuar. J.* **1976**, 92-107.

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

A   Proof of Theorem 2.2

For ease of notation, we suppress the dependency on λ. First note that $\boldsymbol{Z} = (\boldsymbol{I} - \boldsymbol{P} + \boldsymbol{1}\boldsymbol{\pi})^{-1}$ satisfies

$$\boldsymbol{Z}\boldsymbol{\pi} = \boldsymbol{\pi}, \ \ \boldsymbol{1}\boldsymbol{Z} = \boldsymbol{1}, \ \ \boldsymbol{P}\boldsymbol{Z} = \boldsymbol{Z}\boldsymbol{P} = \boldsymbol{Z} - \boldsymbol{I} + \boldsymbol{1}\boldsymbol{\pi}$$

(multiply by $\boldsymbol{I} - \boldsymbol{P} + \boldsymbol{1}\boldsymbol{\pi}$ on both sides). From this it follows easily by induction that

$$\boldsymbol{P}^n\boldsymbol{Z} = \boldsymbol{Z}\boldsymbol{P}^n = \boldsymbol{Z} - \boldsymbol{I} - \cdots - \boldsymbol{P}^{n-1} + n\boldsymbol{1}\boldsymbol{\pi}$$

so

$$\boldsymbol{I} + \cdots + \boldsymbol{P}^{n-1} = \boldsymbol{Z}(\boldsymbol{I} - \boldsymbol{P}^n) + n\boldsymbol{1}\boldsymbol{\pi}$$

$$\boldsymbol{e}_{\ell_0}\mathbb{E}\sum_{n=0}^{A-1}\boldsymbol{P}^n = \boldsymbol{e}_{\ell_0}\mathbb{E}\big(\boldsymbol{Z}(\boldsymbol{I} - \boldsymbol{P}^A) + A\boldsymbol{1}\boldsymbol{\pi}\big)$$
$$= \boldsymbol{e}_{\ell_0}\big(\boldsymbol{Z}(\boldsymbol{I} - \widehat{f}[\boldsymbol{P}]\big) + \mu_A \cdot \boldsymbol{\pi}$$

Rewriting (2.1) in matrix notation and using the independence of A and $L_0, L_1, \ldots$ gives

$$\boldsymbol{\pi}^* = \frac{1}{\mu_A}\mathbb{E}\sum_{n=0}^{A-1}\boldsymbol{e}_{\ell_0}\boldsymbol{P}^n = \frac{1}{\mu_A}\boldsymbol{e}_{\ell_0}\mathbb{E}\big(\boldsymbol{Z}(\boldsymbol{I} - \boldsymbol{P}^A) + A\boldsymbol{1}\boldsymbol{\pi}\big)$$
$$= \frac{1}{\mu_A}\Big[\boldsymbol{e}_{\ell_0}\big(\boldsymbol{Z}(\boldsymbol{I} - \widehat{f}[\boldsymbol{P}]\big) + \mu_A \cdot \boldsymbol{\pi}\Big]$$

B   A variant of the derivation of the age-corrected distribution

A different way to arrive at distribution (2.1) as the relevant bonus class distribution in a model with finite sojourn times of customers is to 'sieve customers one-by-one through the system'. By this we mean that we consider a sequence of λ-customers such that customer n has sojourn time $A^{(n)}$ and bonus classes $L_0^{(n)}, L_1^{(n)}, \ldots, L_{A^{(n)}-1}^{(n)}$ during his time in the portfolio. We then define a $\{1, \ldots, K\}$-valued stochastic process Z as the sequence

$$L_0^{(1)}, L_1^{(1)}, \ldots, L_{A^{(1)}-1}^{(1)}, L_0^{(2)}, L_1^{(2)}, \ldots, L_{A^{(2)}-1}^{(2)}, \ldots, \tag{B.1}$$

and have:

Proposition B.1. *Assume that the distribution of A is non-lattice. Then the process Z given by* (B.1) *has limiting distribution given by* (2.1).

This follows simply by noting that the instances $A^{(1)}, A^{(1)+A^{(2)}}, \ldots$ where a new customer takes over in the construction of Z are regeneration points (the process starts afresh as from $n = 0$) and appealing to the general theory of regenerative processes ([2, Ch. 6]).

It should be noted, however, that Z is not a Markov chain. This follows by noting that

$$
\mathbb{P}(Z_{n+1} = \ell \mid Z_n = k) \tag{B.2}
$$
$$
= \sum_{a=0}^{\infty} \mathbb{P}(\xi_n = a)\Big\{ p_{k\ell}\,\mathbb{P}(\xi_n > a + 1 \mid \xi_n > a) + \delta_{\ell_0 \ell}\,\mathbb{P}(\xi_n = a + 1 \mid \xi_n > a) \Big\},
$$

where ξ_n is the time elapsed since the last regeneration point (backward recurrence time; $\xi_n = n$ for $0 \leq n < A^{(1)}$, $\xi_n = n - A^{(1)}$ for $A^{(1)} \leq n < A^{(1)} + A^{(2)}$ etc.). Here the distribution of ξ_n depends on n, except for the special case where A is geometric, and so must (B.2) do (the argument excludes time-homogeneity, but also the Markov property can be seen to fail).

C A more general model

We here suggest a model which incorporates several features not covered by the basic bonus-malus model consider in the body of the paper.

We assume that a customer is characterized by a random mark M taking value in some set $\mathcal{M}$ and a time-homogeneous Markov chain $X = X_0, X_1, \ldots$ with state space $\mathcal{X} = \mathbb{N} \times \mathcal{L} \times \mathcal{Y}$ where $\mathcal{L}$ is the finite set of bonus classes and $\mathcal{Y}$ is finite or countable. We write $X_k = (N_k, L_k, Y_k)$; N_k is the number of claims in year $k = 0, 1, \ldots$, assumed Poisson with parameter $\lambda(M, L_k, Y_k)$, and L_k is the bonus class. The initial class L_0 as well as Y_0 depends on M.

In addition to potentially influencing the Poisson parameter, the Y component also generates the sojourn time A: it is assumed that the customer still in the portfolio in year $k = 0, 1, \ldots$ will no longer be there in year $k + 1$ w.p. $\delta(M, Y_k)$, so that then $A = k + 1$. The further transition rules state that L_{k+1} is calculated as a deterministic function $b(N_k, L_k, Y_k)$ of $X_k = (N_k, L_k, Y_k)$, and given $N_k = n, L_k = \ell, Y_k = y$, one has $Y_{k+1} = y'$ w.p. $q_{y';n,\ell,y}$.

Example C.1. To cover the classical model with independent sojourn time considered in the rest of the paper, take $M = \Lambda$ as the Poisson parameter of the customer and $\lambda(m, \ell, y) = m$, $L_0 = \ell_0$. If A has some general distribution with point probabilities f_n, $n = 1, 2, \ldots$, there are at least two ways to conform to the general framework above. Both are familiar from the theory of discrete phase-type distributions, [2, III.4] or [1, IX.1 and A5] (see in particular Sections IX.1 and A5 of [1]).

In the first, we take the state space for the Y-chain to be $\mathcal{Y} = \mathbb{N} \cup \{\Delta\}$ for some extra state Δ and $Y_0 = 0$. From state $y \neq \Delta$ [corresponding to still being in the

portfolio in year y] one can go only to either $y+1$ or Δ, w.p. $f_{y+1}/\overline{f}_{y+1}$ for Δ and $1 - f_{y+1/\overline{f}_{y+1}}$ for $y+1$ (thus $q_{y';n,\ell,y}$ does not depend on n, ℓ); state Δ (the coffin state) is absorbing. In the second, we take $\mathcal{Y} = \mathbb{N}$ and $Y_0 = a$ w.p. f_a. From $Y_n = m > 0$, one goes always to $Y_{n+1} = m$, whereas state 0 is absorbing (thus $q_{y';n,\ell,y} = \delta_{y',y-1}$ if $y > 0$, $q_{y';n,\ell,y} = \delta_{y'}$, if $y = 0$). $\qquad\square$

Example C.2. Bonus hunger, i.e. the insured's aptness not to file all small enough claims in order to avoid increase in future premiums, has been studied repeatedly in the literature. E.g. Lemaire [11] (see also Lemaire & Zi [13]) calculate for a given bonus system, each class ℓ and each λ a retention level $\overline{z}_\ell$, such that the insured's costs in terms of either covering a claim of size Z himself or expected future premiums precisely balance when $Z = z_\ell$ when he is in class ℓ. He will then file the claim if $Z > \overline{z}_\ell$ and not otherwise. The calculation is based on a distribution G of a claim Z. We can model this by simply modifying Example C.1 by taking $\lambda(m, \ell, y) = m\big(1 - G(\overline{z}_\ell)\big)$ rather than $\lambda(m, \ell, y) = m$. $\qquad\square$

Example C.3. An insured may be tempted to look for another insurer if he has had many recent claims, therefore a high bonus class and so (often without reason!) believes that his present insurer's system is unfair. We can model this by modifying the first representation of A in Example C.1 by allowing direct transitions from state n of y to the coffin state Δ, occuring with a probability θ_ℓ depending on the present bonus class ℓ (typically θ_ℓ will be increasing in ℓ). Thus

$$q_{y+1;n,\ell,y} = (1 - f_{y+1}/\overline{f}_{y+1})(1 - \theta_\ell),$$
$$q_{\Delta;n,\ell,y} = f_{y+1}/\overline{f}_{y+1} + (1 - f_{y+1}/\overline{f}_{y+1})\theta_\ell = \theta_\ell + (1 - \theta_\ell)f_{y+1}/\overline{f}_{y+1},$$

all other $q_{y';n,\ell,y} = 0$. $\qquad\square$

Example C.4. Young or old drivers are generally considered to have risk characteristics different from the rest of the portfolio. We can model this by letting the mark M be the pair of the Poisson parameter Λ and (for a young driver) the year $B = 0, 1, \ldots$ after the drivers license (or the age for an old driver), as well as a state of y to be of the form (y', y''), with y' determining A as above and y'' the updated year after the license. The initial class L_0 is then chosen as function of B, and one could have, e.g., that $\lambda\big((\lambda, b), \ell, (y', y'')\big)$ has the multiplicative form $s(y'')\lambda$. $\qquad\square$

Further examples, not spelled out in detail, are M including covariates entering multiplicatively in $\lambda(m, \ell, y)$.