Multiline Experience Rating

Jim Gustafsson[*], Lars Hougaard Hansen[†],
Montserrat Guillen[‡] and Jens Perch Nielsen[§]

Draft, do not quote (November 2005)

Abstract

We investigate the concept of multivariate pricing, which includes claiming history for more than one line of business. Our approach generalises the Bühlmann-Straub model. The multivariate credibility model is extended to allow for the age of claims to have an effect on the estimated risk of future claims from the same individual within the same line or in other lines of business. The model is applied to data from a commercial portfolio.

Keywords: Bonus-Malus, Multivariate Credbility, Time-dependent Random Effects

[*]Codan Insurance, Denmark

[†]Codan Insurance, Denmark

[‡]University of Barcelona, Spain. Thanks are given to the Spanish Ministerio de Educacion y Ciencia, FEDER grant SEC2004-05052.

[§]Royal & SunAlliance, Denmark

1 Introduction

Experience rating has been profusely analyzed since the publication of seminal papers and the benchmark work by Lemaire (1995). The ideas of Lundberg (1966), Bichsel (1967) and Bühlmann (1967) established the path to the development of the subject; i.e., the application of a Bayesian approach, the assumption that risk changes over time and the usefulness of credibility theory. The literature offers a range of statistical approaches for calculating the corrections to the premium paid, once claims history knowledge is available to the insurer. The traditional approach has assumed a treatment per line of business and automobile insurance is where this type of *a posteriori* pricing correction is most frequently used.

In practice, customers may have policies underwritten in several lines within the same company. In this framework, the application of a generalised procedure that accounts for all the information on the claiming experience derived from several contracts is of particular interest. In the literature on experience rating, little attention has been paid to incorporating data from other claims reported by the same customer, under the terms of other contracts. One exception is the study by Desjardins, Dionne and Pinquet (2001), in which Bonus-Malus systems for fleets of vehicles are derived from the claims or safety offences history. An experience rating that distinguishes between different types of claims is also well documented (Pinquet, 1997 and 1998). In automobile insurance, for example, one can consider claims with and without bodily injury separately, and also split them by cost. Some authors have also shown that the date of the claim does matter because the effects of a claim on the risk evaluation diminish over time, see Pinquet *et al.* (2001). One can even consider claims affecting different guarantees in the same contract. The recent work by Bühlmann, Gisler and Kollöfen (2003) (presented at the 2003 Astin Conference in Berlin) introduces the multidimensional credibility approach to the study of large and small claims, and to deriving estimators that combine information from both sources (see also Bühlmann and Gisler, 2005). Their article has provided us with the necessary framework model to propose an experience rating scheme that will be useful when a customer has several policies, which we will call lines of business.

A number of practical problems arise in this situation where the insurer deals with customers who possibly have more than one policy. The most simple case would be one customer having two insurance contracts in two lines of business, which were issued on the same date, each with its own claims history. We would like to be able to obtain a Bonus-Malus scheme based on information from both contracts. This would mean that a claim belonging to one line of business may imply raising the premium not only for that line, but also for the other. We argue that the hidden characteristics may affect both lines, although to a lesser extent in the line where no claim was filed. This applied situation will be discussed in the latter part of this paper. We will use a real data set from commercial insurance in Denmark to see the differences between the one- and two-dimensional history rating schemes. By looking at a particular individual policyholder, we will see the effect on the premium modification of the univariate versus the multivariate setting.

Our presentation differs from the study by Bühlmann, Gisler and Kollöfen (2005) in that we also consider time dependence. We assume that the individual random effects within each line of business change over time with their own autocorrelation scheme, and with some correlation between different lines. We believe that this is a more general setting. Although we do not deal with statistical inference in this paper, we do show that considering time effects and multidimensional credibility can lead to a more general treatment of the premium charged to the costumer after knowledge of the claims history has been acquired.

As we mentioned above, there are a number of cases that can occur in real practice and require a careful statistical treatment. The first is that the number of underwritten lines of business may not be the same for all customers, but we would still like to apply the same model framework. For example, we would like to have a generalised version of one-dimensional credibility, so that when we apply the general model to one individual with just one contract, it would be similar to the one-dimensional classical credibility estimator and account covariance with other risks. Secondly, we would like to be able to cope with a policyholder who has different lines of business issued at different moments in time, so that the length of the claims history is not the same for each line. And thirdly, we have to take into account that durations of the contracts do not necessarily equal one year, because the policy may not be in force during the whole period. Our approach addresses all three issues and results in a practical scheme that can readily be applied in practice.

In actuarial science, the issue of experience rating has often been dealt with by the Bühlmann-Straub model (see, section 2), which interprets an overdispersion in frequency data as the consequence of an individual latent risk parameter. To obtain some kind of continuity between the traditional approach to experience rating and the multi-line model, we aim to develop a model which reduces to the simple Bühlmann-Straub model when the customer has only drawn one line of business. For illustrative reasons, we only present the model in two dimensions, but all methods can easily be generalised to k dimensions. This paper is organised as follows: section 2 begins with a review of the classical model, while section 3 aims to describe the extension to the two-dimensional case and discusses the credibility estimators in this context. Section 4 introduces time-dependent random effects into the model. An application to the data is presented in section 5. Conclusions and hints for further work are given in the final section.

2 The Bühlmann-Straub model

Let N_{it} be the number of claims and λ_{it} the expected number of claims for customer i in year t ($t = 1, \ldots, T$), and suppose that

$$N_{it} \mid \theta_i \sim \mathsf{Pois}\left(\lambda_{it}\theta_i\right). \tag{1}$$

Let us also suppose that $\mathbb{E}\left(\theta_i\right) = 1$ and $\mathbb{V}\left(\theta_i\right) = \sigma_{11}^2$. The interpretation of the latent risk parameter θ_i is that it represents the random effect, i.e. the unobserved risk characteristics for customer i. Furthermore, let

$$X_{it} = N_{it}/\lambda_{it}, \quad \overline{X}_i = \tfrac{1}{\lambda_{i\cdot}} \textstyle\sum_t \lambda_{it} X_{it} \ ,$$

where $\lambda_{i\cdot} = \sum_t \lambda_{it}$ and we implicitly assume that $\lambda_{it} > 0$ for both all i and t. We can interpret $\overline{X}_i$ as a weighted average of observed frequency excess. For example, if $\overline{X}_i = 1.5$ then that customer has had 50% more claims on average than the expected figure. In a model without smoothing, one would raise his premium in year $T + 1$ by 50 %, but the Bühlmann-Straub model smoothes this loading. It holds that $\mathbb{E}\left(X_{it} \mid \theta_i\right) = \theta_i$ and $\mathbb{V}\left(X_{it} \mid \theta_i\right) = \theta_i/\lambda_{it}$.

The Best Linear Predictor (BLP) of N_{iT+1} based on the claims history $N_{i1}, \ldots, N_{iT}$, is equivalent to finding the best linear predictor of θ_i based on $\overline{X}_i$. Using standard theory it follows that the best linear predictor of θ_i given $\overline{X}_i$, which we will call BS_i in the one dimension model, is

$$\begin{aligned} BS_i &= \mathbb{E}\left(\theta_i\right) + \mathbb{COV}\left(\theta_i, \overline{X}_i\right) \mathbb{V}\left(\overline{X}_i\right)^{-1} \left(\overline{X}_i - \mathbb{E}\left(\overline{X}_i\right)\right) && (2) \\ &= 1 + z_i\left(\overline{X}_i - 1\right) = 1 + z_i\left(\frac{N_{i\cdot}}{\lambda_{i\cdot}} - 1\right), && (3) \end{aligned}$$

where $N_{i\cdot} = \sum_t N_{it}$ and $z_i = \lambda_{i\cdot}/(\lambda_{i\cdot} + \sigma_{11}^{-2})$. Finally, we can write

$$\mathrm{BLP}(N_{iT+1} \mid N_{i1}, \ldots, N_{iT}) = BS_i \cdot \lambda_{iT+1}.$$

3 Two-Dimensional Bühlmann-Straub model

In this section we assume that every individual may have several contracts. Each contract covers a risk defined under line k. We will present the two-dimensional situation, but it would be the same for higher dimensions. Let N_{ikt} be the number of claims and let λ_{ikt} be the expected number of claims for customer i in line of business k $(k = 1, 2)$ in year t $(t = 1, \ldots, T)$. Let us suppose, that

$$N_{it} = (N_{i1t}, N_{i2t})' \mid (\theta_{i1}, \theta_{i2})' \sim \bigotimes_{k=1}^{2} \mathsf{Pois}\left(\lambda_{ikt}\theta_{ik}\right)$$

and let $\theta_i = (\theta_{i1}, \theta_{i2})'$. We assume that

$$\mathbb{E}\left(\theta_i\right) = \mathbf{1}, \quad \mathbb{V}\left(\theta_i\right) = A = \begin{pmatrix} \sigma_{11}^2 & \sigma_{12}^2 \\ \sigma_{12}^2 & \sigma_{22}^2 \end{pmatrix}.$$

We assume that $\lambda_{ikt} > 0$ for all i, k and t, and introduce the notation $X_{ikt} = N_{ikt}/\lambda_{ikt}$ and $\overline{X}_{ik} = \frac{1}{\lambda_{ik\cdot}} \sum_t \lambda_{ikt} X_{ikt}$.

We seek the Best Linear Predictor of N_{iT+1} given the observed claims history. The multivariate approach leads to the best linear predictor of θ_{ik}, $k = 1, 2$, given $\overline{X}_i = \{\overline{X}_{i1}, \overline{X}_{i2}\}$, which we will call MBS_{ik} as follows:

$$MBS_{ik} = \mathbb{E}\left(\theta_{ik}\right) + \mathbb{COV}\left(\theta_{ik}, \left(\overline{X}_{i1}, \overline{X}_{i2}\right)\right) \left(\mathbb{V}\begin{pmatrix} \overline{X}_{i1} \\ \overline{X}_{i2} \end{pmatrix}\right)^{-1} \left(\begin{pmatrix} \overline{X}_{i1} \\ \overline{X}_{i2} \end{pmatrix} - \mathbb{E}\begin{pmatrix} \overline{X}_{i1} \\ \overline{X}_{i2} \end{pmatrix}\right) \tag{4}$$

for $k = 1, 2$. In the following example, we will fix $k = 1$, all arguments can be applied to $k = 2$ owing to symmetry. In order to maintain the focus on the results, the calculations have been placed in the Appendix. An alternative expression for (4) is

$$MBS_{i1} = 1 + \alpha_{i1}\left(\overline{X}_{i1} - 1\right) + \alpha_{i2}\left(\overline{X}_{i2} - 1\right), \tag{5}$$

where the credibility weights are defined as

$$\begin{aligned} \alpha_{i1} &= \frac{\sigma_{11}^2/\lambda_{i2\cdot} + \det(A)}{\det(\mathbb{V}\left(\overline{X}_i\right))} \\ \alpha_{i2} &= \frac{\sigma_{12}^2/\lambda_{i1\cdot}}{\det(\mathbb{V}\left(\overline{X}_i\right))} \end{aligned}$$

and where

$$\det(\mathbb{V}\left(\overline{X}_i\right)) = \left(\frac{1}{\lambda_{i1\cdot}} + \sigma_{11}^2\right)\left(\frac{1}{\lambda_{i2\cdot}} + \sigma_{22}^2\right) - \left(\sigma_{12}^2\right)^2.$$

The expression (5) shows that the BLP in two dimensions has the same form as the simple Bühlmann-Straub estimator (3).

3.1 Discussion of the credibility estimators

The two-dimensional model collapses to the simple Bühlmann-Straub if the policy only has one line of business. Before investigating this issue, we will make some other comparisons with the one dimensional Bühlmann-Straub. It is appealing to compare the weights z_i and α_{i1} (from (3) and (5)), since these weights control the level of smoothing of the excess $\left(\overline{X}_{i1} - 1\right)$. It can easily be seen that

$$MBS_{i1} = 1 + \beta_{i1}\left(BS_{i1} - 1\right) + \alpha_{i2}\left(\overline{X}_{i2} - 1\right),$$

where BS_{i1} is the one dimensional Bühlmann-Straub model applied on line 1, and $\beta_{i1} = \alpha_{i1}/z_{i1}$. Indeed, in the multivariate setting there is a modification of the one dimensional credibility estimator BS_{i1}, and some information from the excess in line 2, $\left(\overline{X}_{i2} - 1\right)$, is transferred to line 1. Furthermore,

$$\begin{aligned}
\beta_{i1} &= \frac{\left(\sigma_{11}^2/\lambda_{i2\cdot} + \det(A)\right)\left(\lambda_{i1\cdot} + \sigma_{11}^{-2}\right)}{\lambda_{i1\cdot}\det(\mathbb{V}\left(\overline{X}_i\right))} \\
&= \frac{\left(\sigma_{11}^2/\lambda_{i2\cdot} + \det(A)\right)\left(\lambda_{i1\cdot} + \sigma_{11}^{-2}\right)}{\lambda_{i1\cdot}\left(1/\lambda_{i1\cdot}\lambda_{i2\cdot} + \sigma_{11}^2/\lambda_{i2\cdot} + \sigma_{22}^2/\lambda_{i1\cdot} + \det(A)\right)} \\
&= \frac{1/\lambda_{i2\cdot} + \sigma_{11}^2\lambda_{i1\cdot}/\lambda_{i2\cdot} + \det(A)\left(\lambda_{i1\cdot} + \sigma_{11}^{-2}\right)}{1/\lambda_{i2\cdot} + \sigma_{11}^2\lambda_{i1\cdot}/\lambda_{i2\cdot} + \sigma_{22}^2 + \lambda_{i1\cdot}\det(A)} \\
&= \frac{C_1 + \sigma_{22}^2 - \left(\sigma_{12}^2\right)^2\sigma_{11}^2}{C_1 + \sigma_{22}^2} \\
&= \frac{C_1 + \sigma_{22}^2\left(1 - \rho^2\right)}{C_1 + \sigma_{22}^2},
\end{aligned}$$

with $C_1 = 1/\lambda_{i2\cdot} + \sigma_{11}^2\lambda_{i1\cdot}/\lambda_{i2\cdot} + \lambda_{i1\cdot}\det(A)$ and $\rho = \sigma_{12}^2/(\sigma_{11}\sigma_{22})$, showing that $\beta_{i1} \in [0, 1]$. This means that the multivariate credibility estimator smoothes the prediction from the one dimensional model, and the level of smoothing is driven by the correlation between the two lines of business. For $\rho = 0$, we have $\beta_{i1} = 1$ and $\alpha_{i2} = 0$ giving $MBS_{i1} = BS_{i1}$, i.e. the two-dimensional estimator reduces to the simple Bühlmann-Straub estimator. Note that

$$\begin{aligned}
\alpha_{i2} &= \frac{\sigma_{12}^2/\lambda_{i1\cdot}}{1/\lambda_{i1\cdot}\lambda_{i2\cdot} + \sigma_{11}^2/\lambda_{i2\cdot} + \sigma_{22}^2/\lambda_{i1\cdot} + \det(A)} \\
&= \frac{\rho}{\sigma_{11}^{-1}\sigma_{22}^{-1}\lambda_{i2\cdot}^{-1} + \sigma_{11}^{-1}\sigma_{22} + \sigma_{11}\lambda_{i1\cdot}\sigma_{22}^{-1}\lambda_{i2\cdot}^{-1} + \lambda_{i1\cdot}\sigma_{11}^{-1}\sigma_{22}^{-1}\det(A)}
\end{aligned}$$

and it follows that the weight $\alpha_{i2} \in [0, \infty)$. We get $\alpha_{i2} > 1$, if for example ρ is close to 1, $\lambda_{i2\cdot}$ is large and larger than $\lambda_{i1\cdot}$ and finally $\sigma_{11} > \sigma_{22}$. Let us end this section by looking at some limit features of the weights. If we suppose that $|\rho| < 1$, then $\lim_{\lambda_{i1}\to\infty} \beta_{i1} = 1$ and $\lim_{\lambda_{i1}\to\infty} \alpha_{i2} = 0$, showing that the credibility estimator collapses to the Bühlmann-Straub estimator in one dimension, when the observed amount of exposure in line 1 goes to infinity.

3.2 Only one line of business

As mentioned in the beginning at this section, we need to see how the model reacts when a costumer only has one line of business. In this case $\lambda_{i2\cdot} = 0$ (assuming that it is line 2 which has not been active), and $\overline{X}_{i2}$ is not defined. Hence, the BLP is not given by (5), and formally, we need to re-calculate it under these assumptions. The best linear predictor of θ_{ik} $(k = 1, 2)$ given $\overline{X}_{i1}$ alone, can be calculated analogously to (4) (see the appendix for details). We obtain the best linear predictor for θ_{ik}, which is equal to

$$1 + \frac{\sigma_{1k}^2}{\sigma_{11}^2 + \lambda_{i1\cdot}^{-1}} \left(\overline{X}_{i1} - 1\right).$$

We observed that for θ_{i1}, the credibility estimator is equivalent to the simple Bühlmann-Straub case. For θ_{i2}, the BLP also depends on $\overline{X}_{i1}$. This can then be used to set a price for line 2 based on the claims history from line 1 and the covariance between those lines which can be estimated from other customers.

4 Multivariate model with time dependence

In this section we will extend the two dimensional model to take the age of claims into account. This is done for one dimension in Pinquet *et al.* (2001). Firstly, we define the hist-operator

$$\mathsf{hist}\left(X_{iT}\right) = \left(X_{i11}, \ldots, X_{i1T}, X_{i21}, \ldots, X_{i2T}\right)' \tag{6}$$

which returns the full history before time $T+1$. As in Pinquet *et al.* (2001), we denote $\theta_{it} = (\theta_{i1t}, \theta_{i2t})'$ and we assume $\mathbb{E}\left(\theta_{it}\right) = (1,1)'$ for all t, and we will assume that the process (θ_{it}) is stationary in the sense that

$$\mathbb{COV}\left(\theta_{ikr}, \theta_{ils}\right) = \sigma_{kl}^2 \left(\rho_{kl}\right)^{|s-r|}$$

with $\sigma_{12}^2 = \sigma_{21}^2$ and $\rho_{12} = \rho_{21}$. Assuming this model for θ_{it}, the best linear predictor of θ_{iT+1} based on the claims history becomes

$$\begin{aligned}\mathrm{BLP}\left(\theta_{iT+1}|\mathsf{hist}(X_{iT})\right) &= \mathbb{E}\left(\theta_{Tt+1}\right)+\mathbb{COV}\left(\theta_{iT+1},\mathsf{hist}\left(X_{iT}\right)'\right)\mathbb{V}\left(\mathsf{hist}\left(X_{iT}\right)\right)^{-1}\\ &\quad\times\left(\mathsf{hist}\left(X_{iT}\right)-\mathbb{E}\left(\mathsf{hist}\left(X_{iT}\right)\right)\right)\\ &= \mathbf{1}+A\cdot\left(B+S_i\right)\left(\mathsf{hist}\left(X_{iT}\right)-\ \mathbf{1}\right) \qquad (7)\end{aligned}$$

where A is the $2\times 2T$ matrix

$$\begin{aligned}A &= \mathbb{COV}\left(\theta_{iT+1},\mathsf{hist}\left(X_{iT}\right)'\right)\\ &= \begin{pmatrix}(\sigma^2_{11s})_{s=1,\dots T} & (\sigma^2_{12s})_{s=1,\dots T}\\ (\sigma^2_{12s})_{s=1,\dots T} & (\sigma^2_{22s})_{s=1,\dots T}\end{pmatrix} \qquad (8)\end{aligned}$$

with $\sigma^2_{qrs}=\sigma^2_{qr}\left(\rho_{qr}\right)^{s-1}$. The $2T\times 2T$ matrix B has the form

$$B=\begin{pmatrix}b_{11(s,t)} & b_{12(s,t)}\\ b_{12(s,t)} & b_{22(s,t)}\end{pmatrix} \qquad (9)$$

with $b_{qr(s,t)}=\sigma^2_{qr}\left(\rho_{qr}\right)^{|s-t|}$ being the elements of a $T\times T$ block matrix. And finally,

$$\begin{aligned}S_i &= \mathbb{E}\left(\mathbb{V}\left(\mathsf{hist}\left(X_{it}\right)\ \mid \mathsf{hist}\left(\theta_{it}\right)\right)\right)\\ &= \left(\mathsf{diag}\left(\frac{1}{\lambda_{i11}},\dots,\frac{1}{\lambda_{i1T}},\frac{1}{\lambda_{i21}},\dots,\frac{1}{\lambda_{i2T}}\right)\right). \qquad (10)\end{aligned}$$

The expression (7) for the best linear predictor is analogue to (2) and (4).

5 Data study

The data come from an active Danish insurance company's commercial fire portfolio. Despite the fire portfolio label a variety of risks other than fire are also covered, including water damage to both buildings and contents, fungus and insect damage, glass damage and theft. We have chosen to focus on two types of coverage (referred to below as lines of business) - Building Water and Theft. The data set contains information from the years 1995-2003 on 19,270 policies where all have both lines of business. For sake of simplicity, we have only included 4 years of claims history for each policy, meaning that policies with less than 4 years of history are excluded. This restriction means that the number of policies is reduced to 10,212. Thereafter, we randomly split the data into two sets. The first data set (D_1) includes 75% of the policies and is used to estimate the parameters involved in each model. The second data set (D_2) is used to evaluate the prediction power of the models amid the global parameters. Some characteristics of data are given in table 1.

TABLE 1

Sample statistics for Theft and Water lines of business

Data	Size	Line of Business	Total number of claims ($N_{\cdot k \cdot}$)	Expected number of claims ($\lambda_{\cdot k \cdot}$)
D_1	7,656	Theft	739	694.0
		Water	288	333.2
D_2	2,556	Theft	239	250.8
		Water	103	120.8

It can be seen from table 1 that the Theft line of business causes a higher exposure to risk than the Water line. There is a considerable difference between the observed number of claims and the expected number of claims. This happens because the expected number of claims is calculated on an individual basis and then aggregated. The expected number of claims is calculated with a Poisson regression model (see Dionne and Vanasse, 1992). McCullagh and Nelder (1989) give a general introduction to generalized linear models.

We use the weighted least squares method to estimate the unknown parameters in each model. In short, the parameters are estimated by the numerical minimisation of the sum of squared residuals. The following expression is therefore minimised

$$\sum_{i,t} \left(N_{ikt} - \lambda_{ikt} F_{ikt}\right)^2 d_{ikt} \text{ for } k = 1, 2,$$

where F_{ikt} stands for the relativity credibility estimator that will change depending on the model framework (and depend on the parameters accordingly), d_{ikt} is the duration of policyholder i and the line of business k, during the observed period t. When $F_{ikt} \equiv 1$, we have the baseline model where no credibility is used and no time-dependence. We will use this as a benchmark model. We will consider the univariate credibility model with and without time-dependence and with time-dependence for both lines. BS_i (which depends on σ^2_{kk} for line k) thus replaces F_{ikt} when no time-dependence is considered and we are working in the one-dimensional model. If we consider the time-dependence model while still in the one-dimensional model, then F_{ikt} depends on two parameters σ^2_{kk} and ρ_{kk} for the line k. In the two-dimensional model F_{ikt} equals MBS_{ik} without any time-dependence and the parameters to be estimated are σ_{11}^2, σ_{22}^2 and σ_{12}^2. If time-dependence is introduced, three more parameters need to be estimated, namely ρ_{11}, ρ_{22} and ρ_{12}. These estimates are shown in table 2. Numerical minimisation is achieved with an iterative method. The one-dimensional estimates are used as starting values for the two-dimensional model parameters.

TABLE 2

Estimated parameters for each model using weighted least squares

Model	σ_{11}^2	σ_{22}^2	σ_{12}^2	ρ_{11}	ρ_{22}	ρ_{12}
One-dimensional, Theft with no time	0.377	·	·	·	·	·
with time	0.412	·	·	0.721	·	·
Water, with no time	·	1.686	·	·	·	·
with time	·	1.712	·	·	0.811	·
Two-dimensional with no time	0.447	1.702	0.619	·	·	·
with time	0.461	1.922	0.863	0.865	0.922	0.351

Using the estimated parameters from table 2, the prediction power can assessed by the weighted sum of squares residuals for the time-period $t+1$. The prediction results for the validation data could can be found in table 3.

TABLE 3

Sum squared residuals in the testing sample

		One-dimensional		Two-dimensional	
Line of Business	Baseline	No time	With time	No time	With time
Theft	0.1269	0.1211	0.1224	0.1204	0.1208
Water	0.0434	0.0346	0.0336	0.0352	0.0334

It can be seen that the largest difference in the prediction error is between the models with and without a latent risk parameter. The two-dimensional model is the best predictor when considering the most risk exposed business - Theft. The two-dimensional with time-dependence model also has a prediction error that is only marginally larger. However, when the Water line of business is considered, the performance of the time-dependence model compared to the other models improves, with the two-dimensional model giving the best results.

From a business point of view, we do not want the premiums to be too volatile, and it is therefore of interest to have a model where the coefficient has a low variance. This is shown in the table 4

TABLE 4

Mean and variance of the relativities in the testing sample for the prediction year

		One-dimensional		Two-dimensional	
Moment	Exc	No time	With time	No time	With time
Theft, Mean	0.889	0.993	0.999	0.994	0.996
Variance	7.382	0.051	0.020	0.083	0.050
Water, Mean	0.891	1.005	1.003	0.994	1.008
Variance	16.97	0.204	0.111	0.290	0.204

As expected, it can be seen that the two-dimensional models have the most volatile coefficients. Note that the mean of $Exc = N_{ik\cdot}/\lambda_{ik\cdot}$ for Theft equals 0.889 meaning that there have been fewer claims than expected for this sample.

We then compare the results from the one-dimensional Bühlmann-Straub model with and without time dependence for each line of business against the two-dimensional models. This is done by focusing on five arbitrary policies characterised as follows.

TABLE 5

History of five arbitrarily chosen policies

Policy	N_{i1t}	N_{i2t}	λ_{i1t}	λ_{i2t}
1	$\{0,0,0\}$	$\{0,0,1\}$	$\{0.008,0.012,0.011\}$	$\{0.248,0.247,0.247\}$
2	$\{0,0,1\}$	$\{0,0,0\}$	$\{0.102,0.099,0.097\}$	$\{0.102,0.102,0.084\}$
3	$\{1,1,1\}$	$\{1,0,0\}$	$\{0.438,0.430,0.422\}$	$\{0.105,0.108,0.107\}$
4	$\{0,0,0\}$	$\{0,0,0\}$	$\{0.111,0.109,0.108\}$	$\{0.014,0.014,0.014\}$
5	$\{0,0,0\}$	$\{1,0,0\}$	$\{0.024,0.023,0.023\}$	$\{0.169,0.169,0.169\}$

TABLE 6

Estimates using their proposed models in both lines of business for the five policies

Policy	Line of Business	Exc	One-dimensional No time	One-dimensional With time	Two-dimensional No time	Two-dimensional With time
1	Theft	0	0.988	0.993	1.060	1.151
	Water	1.350	1.194	1.366	1.186	1.317
2	Theft	3.347	1.237	1.199	1.128	1.237
	Water	0	0.673	0.773	0.946	0.857
3	Theft	2.327	1.434	1.254	1.612	1.307
	Water	3.136	1.747	1.322	2.121	1.625
4	Theft	0	0.890	0.939	0.854	0.896
	Water	0	0.932	0.954	0.770	0.900
5	Theft	0	0.974	0.986	1.136	0.909
	Water	1.973	1.448	1.127	1.424	1.311

Table 5 should be interpreted as follows. For the first policy, we can see that there was only one claim in the last year in the Water line. It can also be established that according to our expectations, the second business is riskier then the firstas shown by columns 4 and 5. Table 6 gives estimates of F_{ikt} for the five policies chosen for illustrative purposes.

Table 6 shows how different models react to various claims history scenarios. If we concentrate on the first policy, we see that in the Theft line, when using both one-dimensional models one could expect a rebate on the premium, while both two-dimensional models suggest a raise in the premium. This is because the individual history from the other line, Water, is taken into account. It can also be seen that the two-dimensional model with time-dependence suggests a bigger raise in the Theft line, since the claim in the Water line was in the third year observed (the two-dimensional model without time-dependence deals with each claim in the same way). If we consider the Water line for the same policy, here both time-dependence models suggest an increase of over 30% in the premium, while models without time-dependence suggest an increase of around 19%.

An opposite scenario in the context of time-dependence structure arises in the third policy, in the Water line of business. Here, the customer made a claim during his first year. This is shown by the time-dependence models, leading to lower values than the other two models without a time-dependent structure. Policies 2, 4 and 5 should be interpreted in a similar manner.

6 Conclusions

The claims history for a customer (or policy) is commonly used to adjust the future premium, and this is often implemented using a Bonus-Malus System (see, Lemaire, 1995). The idea behind multiline experience rating is to extend this concept so that the rating also depends on performance in other lines of business. The motivation behind this is that if a costumer performs badly in one line, we will to some extent expect him to perform badly in all lines of business. The size of the impact on a particular line of business of claims made in other lines should be controlled by the correlation between the two lines, and the amount of observed exposure.

We have presented a model for multi-line experience rating in this study. The model is formulated for two lines of business, but can easily be extended to arbitrary dimensions and the time dependence can have a more general structure.

The model has been applied to data from a commercial insurance company in Denmark, and the results show that the models perform to some extent better then the one-dimensional Bühlmann-Straub model in terms of predicting error in a testing sample. However, the price one pays in terms of interpretation in formulas may not outweight the advantages of using the one-dimensional approach.

7 Appendix

In the two dimensional credibility model, the moments of X_{ik} given the latent risk parameter are

$$\mathbb{E}\left(X_{ik} \mid \theta_i\right) = \theta_i \quad \mathbb{V}\left(X_{ik} \mid \theta_i\right) = \begin{pmatrix} \frac{\theta_{i1}}{\lambda_{i1t}} & 0 \\ 0 & \frac{\theta_{i2}}{\lambda_{i2t}} \end{pmatrix}.$$

To calculate (4) we need to find

$$\mathbb{COV}\left(\theta_{i1},\left(\overline{X}_{i1},\overline{X}_{i2}\right)\right)=\underbrace{\mathbb{E}\left(\mathbb{COV}\left(\theta_{i1},\left(\overline{X}_{i1},\overline{X}_{i2}\right)\mid\theta_i\right)\right)}_{:=a}+\underbrace{\mathbb{COV}\left(\mathbb{E}\left(\theta_{i1}\mid\theta_i\right),\mathbb{E}\left(\left(\overline{X}_{i1},\overline{X}_{i2}\right)\mid\theta_i\right)\right)}_{:=b}$$

and we start by calculating a. In the same way as the 1-dimensional case

$$\mathbb{E}\left(\mathbb{COV}\left(\theta_{i1},\overline{X}_{i1}\mid\theta_i\right)\right)=\mathbb{E}\left(\theta_{i1}^2-\theta_{i1}^2\right)=0$$

and

$$\mathbb{E}\left(\mathbb{COV}\left(\theta_{i1},\overline{X}_{i2}\mid\theta_i\right)\right)=\mathbb{E}\left(\theta_{i1}\theta_{i2}-\theta_{i1}\theta_{i2}\right)=0,$$

show that term a vanishes as in the simple Bühlmann-Straub model. Similarly, term b follows the calculations from the one dimensional case. We begin by finding the expression for $k=1$

$$\begin{aligned}\mathbb{COV}\left(\mathbb{E}\left(\theta_{i1}\mid\theta_i\right),\mathbb{E}\left(\overline{X}_{i1}\mid\theta_i\right)\right) &= \mathbb{COV}\left(\theta_{i1},\theta_{i1}\right)\\ &= \mathbb{V}\left(\theta_{i1}\right)=\sigma_{11}^2\end{aligned}$$

and

$$\begin{aligned}\mathbb{COV}\left(\mathbb{E}\left(\theta_{i1}\mid\theta_i\right),\mathbb{E}\left(\overline{X}_{i2}\mid\theta_i\right)\right) &= \mathbb{COV}\left(\theta_{i1},\theta_{i2}\right)\\ &= \sigma_{12}^2.\end{aligned}$$

In short, the covariance expression is

$$\mathbb{COV}\left(\theta_{i1},\left(\overline{X}_{i1},\overline{X}_{i2}\right)\right)=\left(\sigma_{11}^2,\sigma_{12}^2\right).$$

For the variance term in (4), we obtain, recalling the notation $\overline{X}_i=\left(\overline{X}_{i1},\overline{X}_{i2}\right)^T$,

$$\begin{aligned}\mathbb{V}\left(\overline{X}_i\right) &= \mathbb{V}\left(\mathbb{E}\left(\overline{X}_i\mid\theta_i\right)\right)+\mathbb{E}\left(\mathbb{V}\left(\overline{X}_i\mid\theta_i\right)\right)\\ &= A+\mathbb{E}\begin{pmatrix}\frac{\theta_{i1}}{\lambda_{i1\cdot}} & 0\\ 0 & \frac{\theta_{i2}}{\lambda_{i2\cdot}}\end{pmatrix}\\ &= \begin{pmatrix}\frac{1}{\lambda_{i1\cdot}}+\sigma_{11}^2 & \sigma_{12}^2\\ \sigma_{12}^2 & \frac{1}{\lambda_{i2\cdot}}+\sigma_{22}^2\end{pmatrix}.\end{aligned}$$

Using standard formulas for inverse of matrices

$$\mathbb{V}\left(\overline{X}_i\right)^{-1}=\frac{1}{\det(\mathbb{V}\left(\overline{X}_i\right))}\begin{pmatrix}\frac{1}{\lambda_{i2\cdot}}+\sigma_{22}^2 & -\sigma_{12}^2\\ -\sigma_{12}^2 & \frac{1}{\lambda_{i1\cdot}}+\sigma_{11}^2\end{pmatrix}$$

where

$$\det(\mathbb{V}\left(\overline{X}_i\right)) = \left(\frac{1}{\lambda_{i1\cdot}} + \sigma_{11}^2\right)\left(\frac{1}{\lambda_{i2\cdot}} + \sigma_{22}^2\right) - \left(\sigma_{12}^2\right)^2.$$

Then, the best linear predictor is of θ_{i1} given $\left(\overline{X}_{i1}, \overline{X}_{i2}\right)$ is

$$\begin{aligned} MBS_{i1} &= 1 + \left(\sigma_{11}^2, \sigma_{12}^2\right)\frac{1}{\det(\mathbb{V}\left(\overline{X}_i\right))}\begin{pmatrix} \frac{1}{\lambda_{i2\cdot}} + \sigma_{22}^2 & -\sigma_{12}^2 \\ -\sigma_{12}^2 & \frac{1}{\lambda_{i1\cdot}} + \sigma_{11}^2 \end{pmatrix}\begin{pmatrix} \overline{X}_{i1} - 1 \\ \overline{X}_{i2} - 1 \end{pmatrix} \\ &= 1 + \alpha_{i1}\left(\overline{X}_{i1} - 1\right) + \alpha_{i2}\left(\overline{X}_{i2} - 1\right), \end{aligned}$$

where the weights are defined as

$$\begin{aligned} \alpha_{i1} &= \frac{1}{\det(\mathbb{V}\left(\overline{X}_i\right))}\left(\left(\frac{1}{\lambda_{i2\cdot}} + \sigma_{22}^2\right)\sigma_{11}^2 - \left(\sigma_{12}^2\right)^2\right) \\ &= \frac{\sigma_{11}^2/\lambda_{i2\cdot} + \det(A)}{\det(\mathbb{V}\left(\overline{X}_i\right))} \end{aligned}$$

$$\begin{aligned} \alpha_{i2} &= \frac{1}{\det(\mathbb{V}\left(\overline{X}_i\right))}\left(\left(\frac{1}{\lambda_{i1\cdot}} + \sigma_{11}^2\right)\sigma_{12}^2 - \sigma_{11}^2\sigma_{12}^2\right) \\ &= \frac{\sigma_{12}^2/\lambda_{i1\cdot}}{\det(\mathbb{V}\left(\overline{X}_i\right))}. \end{aligned}$$

When no information on $k = 2$ is available, then

$$\mathbb{COV}\left(\theta_{ik}, \overline{X}_{i1}\right) = \underbrace{\mathbb{E}\left(\mathbb{COV}\left(\theta_{ik}, \overline{X}_{i1} \mid \theta_i\right)\right)}_{:=a} + \underbrace{\mathbb{COV}\left(\mathbb{E}\left(\theta_{ik} \mid \theta_i\right), \mathbb{E}\left(\overline{X}_{i1} \mid \theta_i\right)\right)}_{:=b},$$

and to calculate a

$$\mathbb{E}\left(\mathbb{COV}\left(\theta_{ik}, \overline{X}_{i1} \mid \theta_i\right)\right) = \mathbb{E}\left(\theta_{ik}\theta_{i1} - \theta_{ik}\theta_{i1}\right) = 0,$$

and to calculate b

$$\begin{aligned} \mathbb{COV}\left(\mathbb{E}\left(\theta_{ik} \mid \theta_i\right), \mathbb{E}\left(\overline{X}_{i1} \mid \theta_i\right)\right) &= \mathbb{COV}\left(\theta_{ik}, \theta_{i1}\right) \\ &= \sigma_{1k}^2. \end{aligned}$$

In summary, the covariance expression is

$$\mathbb{COV}\left(\theta_{ik}, \overline{X}_{i1}\right) = \sigma_{1k}^2.$$

And the variance term is

$$\begin{aligned}\mathbb{V}\left(\overline{X}_{i1}\right) &= \mathbb{V}\left(\mathbb{E}\left(\overline{X}_{i1} \mid \theta_i\right)\right) + \mathbb{E}\left(\mathbb{V}\left(\overline{X}_{i1} \mid \theta_i\right)\right) \\ &= \sigma_{11}^2 + \mathbb{E}\left(\frac{\theta_{i1}}{\lambda_{i1\cdot}}\right) = \sigma_{11}^2 + \frac{1}{\lambda_{i1\cdot}}.\end{aligned}$$

8 References

Bühlmann, H. (1967) Experience rating and credibility. *Astin Bulletin*, 4, 199-207.

Bühlmann, H. and Gisler, A. (2005) *A course in credibility theory and its applications.* Springer-Verlag.

Bühlmann, H., Gisler, A. and Kollöffel, D. (2003) Multidimensional Credibility applied to estimating the frequency of big claims. http://www.astin2003.de/img/papers/gisler.pdf

Bichsel, F. (1967) Experience Rating in subsets of risks. *Astin Bulletin*, 4, 210-227.

Desjardins, D., Dionne, G. and Pinquet, J. (2001) Experience rating schemes for fleets of vehicles, *Astin Bulletin*, 31, 81-105.

Dionne, G. and Vanasse, C. (1992) Automobile Insurance Ratemaking in the Presence of Asymmetrical Information. *Journal of Applied Econometrics*, 7, 2, 149-165.

Lemaire,J. (1995) *Bonus-Malus Systems in Automobile Insurance.* Kluwer Academic Publishers.

Lundberg, O. (1966) Une note sur des systhèmes de tarification basés sur des dolèles dy type Poiisson, *Astin Bulletin*, 4, 49-58.

McCullagh, P. and Nelder, J.A. (1989) *Generalized Linear Models*, no. 37 in Monograph on Statistics and Applied Probability, Chapman & Hall, 2nd ed.

Pinquet, J. (1997) Allowance for the cost of claims in Bonus-Malus systems. *Astin Bulletin*, 27, 33-57.

Pinquet, J. (1998) Designing optimal Bonus-Malus system from different types of claims. *Astin Bulletin*, 28, 205-220.

Pinquet, J., Guillén, M. and Bolancé, C. (2001) Allowance for the age of claims in Bonus-Malus systems. *Astin Bulletin*, 31, 337-348.

Montserrat Guillen
Dept. Econometrics, RFA-IREA
University of Barcelona
Diagonal, 690
08034 Barcelona, Spain

mguillen@ub.edu