MULTIVARIATE LATENT RISK: A CREDIBILITY APPROACH

BY

MARTIN ENGLUND[1], MONTSERRAT GUILLÉN[2], JIM GUSTAFSSON[3], LARS HOUGAARD NIELSEN[4] AND JENS PERCH NIELSEN[5]

ABSTRACT

We investigate a concept of multivariate pricing, which includes claim history for more than one line of business and is a generalization of the Bühlmann-Straub model. The multivariate credibility model is extended to allow for the age of claims to influence the estimation of future claims. The model is applied to data from a portfolio of commercial lines of business.

KEYWORDS

Bonus-Malus System, Multivariate Credibility, Time-dependent Random Effects

1. INTRODUCTION

We start by considering the simplest possible sampling scheme of the multivariate credibility approach of Bühlmann and Gisler (2005): Let X_1, X_2 be stochastic variables that are conditionally independent given an one-dimensional unobservable variable Θ. This sampling scheme is excellent for catching heterogeneity properties in correlated variables with the same underlying risk such as small claims versus big claims or claim frequencies versus claim severities. In this paper we also consider a multivariate credibility approach. We are interested in the situation where the two considered stochastic variables X_1, X_2 are correlated, not through the same unobservable underlying risk, but with an underlying unobservable risk parameter tied to each of the stochastic variables:

Let X_1, X_2 be conditionally independent stochastic variables given $\Theta = (\Theta_1, \Theta_2)$, where Θ_1 is the latent risk variable describing the severity of X_1 and Θ_2 is the latent risk variable describing the severity of X_2.

This sampling scheme is useful when analyzing more than one line of business per customer. If Θ_1 and Θ_2 are correlated, information from one line of business will impact on the knowledge of the other line of business. The sampling scheme of Bühlmann and Gisler (2005), used in this particular context, would correspond to an absolute correlation of 1 between the two latent variables (Θ_1, Θ_2). In empirical studies we see the whole spectrum of positive correlations ranging between zero and one. In this paper, we present in an empirical illustration study with correlation 0.6.

There is not a large literature on methods incorporating data from different business lines on the same policyholder. Exceptions includes the studies by Desjardins, Dionne and Pinquet (2001), in which Bonus-Malus systems for fleets of vehicles are derived from the claims or safety offences history, and Pinquet (1997 and 1998), in which examples is given of experience rating that incorporates additional claim information, from the number and the cost of claims, and claims at fault and not at fault within the same line of business. Instead of following the expected value principle, as in the latter papers, we follow the tradition of linear credibility.

Traditional approaches of credibility theory, based on the ideas of Lundberg (1966), Bichsel (1967), Bühlmann (1967) and more recently Lemaire (1995), consider one unobserved risk parameter, for each customer and treat one policy or coverage, which we here will call line of business, at a time in order to calculate a premium based on both collective and individual information. In this paper we consider two recent extensions to the traditional approach; firstly the multidimensional generalization of the credibility approach to more than one line of business, which was first introduced by Jewell (1973 and 1974), and secondly to allow the multidimensional individual latent risk parameter to change over time.

Pioneering papers on one dimensional evolutionary credibility models are Gerber and Jones (1975), Sundt (1981) and Pinquet, Guillén and Bolancé (2001). The latter shows that the date of the claim does matter because the effects of a claim on the risk evaluation diminish over time. A concise review of credibility theory can be found in Norberg (2004), and as a textbook on multiple time series analysis we recommend Lütkepohl (2005).

While our approach has an obvious potential for improved pricing, it is also promising for other purposes where a joint understanding of more than one line of business at a time is important. One such example is cross-selling where policyholders may have policies underwritten in several lines within the same company. Here the insurance company would like to use the information of a policyholders latent risk of one line of business to be able to cross-sell him a policy of another line of business at a competitive price. This requires a multidimensional model where the correlation between the latent risk parameters of the two lines of business are understood. So, for a customer with one

policy only, the experience rating mechanism should allow to incorporate the collected historical information when pricing some other policy of interest of the customer. Therefore, the insurance company can turn its data base into a competitive advantage even in situations where they do not have historical information of the customer in the line of business which they wish to cross-sell. This, however, needs a new derivation of a modified Bühlmann-Straub model, since the ordinary multidimensional model is reduces to model of lesser order when the customer lack one of the investigated lines of business.

The Bühlmann-Straub credibility model assumes all customers to be a priori equal, which is not the case for most insurance companies. So, to obtain some kind of continuity between the present actuarial practice and our approach, we aim to develop our models with prior known differences.

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.

We show through our practical example that modern extensions of credibility theory, incorporating multidimensional credibility and time effects, can lead to improved pricing of the customer risk after some knowledge of the claims history has been acquired.

This paper is organized as follows, section 2 begins with a review of the classical Bühlmann-Straub credibility model and the extension to the multidimensional case. Section 3 introduces time-dependent random effects into the model. An application to the data is presented in section 4. Conclusions are given in the final section.

2. MULTIDIMENSIONAL BÜHLMANN-STRAUB CREDIBILITY MODEL

In this section, we provide a method to obtain a multidimensional credibility estimator for claim frequencies, with prior known differences of the individuals. An excellent and comprehensive treatment of credibility theory is found in Bühlmann and Gisler (2005). We will use similar notation and refer to the results therein.

Let us in the univariate case consider the number of claims N_{ij} of customer i in insurance period j, usually years, on one particular coverage, with $\mathbf{N}_i = (N_{ij})_{j=1,\ldots,J}$. We assume that every customer i has its individual risk profile θ_i, a realization of the random variable Θ_i. We also assume that we have a priori knowledge on the expected number of claims from customer i in period j, λ_{ij}, which depends on regression components represented by the vector x_{ij} and the duration ω_{ij}: $\lambda_{ij} = \omega_{ij}\exp(x_{ij}a)$; $a \in \mathbb{R}^k$, where a is a vector of parameters and where k is the number of regression components. This is particularly useful in practice because Bonus-Malus schemes are frequently superimposed on tariffs given by regression components, or rating factors, that may change over time, e.g. age.

Formally, we assume that $N_{ij} | \Theta_i = \theta_i$ is Poisson distributed with mean $\theta_i \lambda_{ij}$, and moreover we assume that $(\Theta_1, \mathbf{N}_1), (\Theta_2, \mathbf{N}_2), \ldots$ are independent random

vectors, where $\Theta_1, \Theta_2, \ldots$ are *iid* with $\mathbb{E}(\Theta_i) = 1$ and $\mathbb{V}(\Theta_i) = \tau^2$. This model corresponds to the case of known a priori differences, which we allow to change over time, studied in Bühlmann and Gisler (2005) where $Y_{ij} = N_{ij}/\lambda_{ij}$. Since Y_{ij} satisfy the conditions of the Bühlmann-Straub model, and due to the linearity property of the projections, the best linear predictor of N_{iJ+1} given $N_{i1}, \ldots, N_{iJ}$ and λ_{iJ+1} is $\hat{\theta}_i \lambda_{iJ+1}$, where

$$\hat{\theta}_i = \mathbb{E}(\Theta_i) + \mathbb{COV}(\Theta_i, \bar{Y}_i)^{-1}(\bar{Y}_i - \mathbb{E}(\bar{Y}_i))$$

$$= 1 + \alpha_i(\bar{Y}_i - 1) = 1 + \alpha_i\left(\frac{N_{i\cdot}}{\lambda_{i\cdot}} - 1\right), \tag{1}$$

and $N_{i\cdot} = \sum_j N_{ij}$, $\lambda_{i\cdot} = \sum_j \lambda_{ij}$, $\bar{Y}_i = \frac{1}{\lambda_{i\cdot}}\sum_j \lambda_{ij} Y_{ij}$ and $\alpha_i = \lambda_{i\cdot}/\left(\lambda_{i\cdot} + \tau^{-2}\right)$.

The first expression follows easily from standard linear regression theory and the equivalence to orthogonal projection to minimizing the sum of squares (see Norberg, 2004).

Since we assume a conditional Poisson distribution of the individual claims N_{ij}, given that the a priori differences are known and given the individual risk profile, $\mathbb{V}(N_{ij}|\theta_i) = \mathbb{E}(N_{ij}|\theta_i) = \lambda_{ij}\theta_i$. This implies that we assume equidispersion at the individual level, but it does not imply that there is a restriction on the dispersion in the portfolio. Moreover, we believe that Poisson is a model framework that is easy to apply in practice, since modelling the number of claims with the Poisson regression model leads to consistent estimates of λ_{ij} under very general specification conditions (Gouriéroux, Monfort and Trognon (1984)). In practice, however, we will need to estimate τ^2 before we obtain the credibility estimator $\hat{\theta}_i$.

One obvious generalizations of the presented model is to increase the dimensionality of N_{ij}, e.g. we may consider different types of claims related to the same policy or several types of claims arising from several policies held by the same costumer. This is the studied by Bühlmann, Gisler and Kollöffel (2003). In the multivariate model presented below, we depart from their setting in that we assume that there are a priori differences, which are not necessarily the same, in each dimension, and, more importantly, we assume one specific risk parameter for each line of business.

In the multivariate case, we need to consider P lines of business. $\theta_i = (\theta_{i1}, \ldots, \theta_{iP})$ is then a vector of random variables, where each component produces a particular individual risk profile for each type of insurance held by the same customer i. We assume $\mathbb{E}(\theta_i) = \mathbf{1}$ and $\mathbb{COV}(\theta_i) = T$, with element τ_{kl}^2 on row k and column l. The covariance matrix of the risk profile random variables is constant for all individuals in the portfolio. No other restriction is initially imposed on the covariance between the risk profile of different components. The a priori differences now are λ_{ijp} where subscript i refers to the customer, j refers to the year and p refers to the type of claim, and correspondingly, the risk profiles are θ_{ip} for every individual and product.

Again, using the best linear predictor principle in the multivariate setting, the credibility estimator follow immediately as:

$$\hat{\theta}_i = \mathbb{E}(\theta_i) + \mathbb{COV}(\theta_i, \bar{\mathbf{Y}}_i)\, \mathbb{V}(\bar{\mathbf{Y}}_i)^{-1}\,(\bar{\mathbf{Y}}_i - \mathbb{E}(\bar{\mathbf{Y}}_i)), \qquad (2)$$

where bold symbols are used to denote vectors. $\bar{\mathbf{Y}}_i$ is P-dimensional and each component equals $\frac{1}{\lambda_{i\cdot p}} \sum_j \lambda_{ijp} Y_{ijp}$, with $Y_{ijp} = N_{ijp}/\lambda_{ijp}$ and $\lambda_{i\cdot p} = \sum_j \lambda_{ijp}$ and $\hat{\theta}_i = (\hat{\theta}_{i1}, \dots, \hat{\theta}_{iP})$.

3. Multidimensional Bühlmann-Straub credibility model with time dependence

Time dependence can be introduced in the model by letting the individual risk profiles vary over time. Let us assume that $\boldsymbol{\Theta}_{ij}$ is the (multivariate) random vector of individual risk profiles of customer i in period j. We assume that $\boldsymbol{\Theta}_{ij}$ are independent in i with $\mathbb{E}(\boldsymbol{\Theta}_{ij}) = \mathbf{1}$. We assume a vector stationary process with covariance structure both between the P types of policies and between the J periods. To focus on the covariation between lines of business we follow and generalize Pinquet et al. (2001). We denote the covariances and autocorrelation coefficients as $\mathbb{COV}(\theta_{ijp}, \theta_{i^*j^*p^*}) = \rho_{pp^*}^{|j-j^*|} \tau_{pp^*}^2$ if $i = i^*$ and $\mathbb{COV}(\theta_{ijp}, \theta_{i^*j^*p^*}) = 0$ otherwise, with $|\rho| \le 1$. This holds for all individuals in the portfolio, $j, j^* = 1, \dots, J$ and $p, p^* = 1, \dots, P$. Theoretical, or intuitive, there are no restrictions for ρ_{pp^*} and ρ_{p^*p} to differ, but in our application below we put $\rho_{pp^*} = \rho_{p^*p}$, for simplicity.

In our application we have used a Weighted Least Squares (WLS) method, with the duration as weight:

$$\min_{\rho, \tau^2} \sum_{ijp} \left(N_{ijp} - \lambda_{ijp}\, \hat{\theta}_{ijp} \right)^2 \omega_{ijp} \qquad (3)$$

where $\hat{\theta}_{ip}$ is the credibility estimator expression, with or without time dependence for individual i, year j and product type p, which depends on the structural parameters.

4. Data study

In this section, we apply the models introduced in this paper to a portfolio of a Danish insurance company. By way of introduction, we need to discuss a few theoretical and implementational challenges, which arise when going from one-dimensional to multidimensional experience rating. Firstly, the number of underwritten lines of business may not be the same for all customers, but we would still like to apply the same framework. Secondly, we would like to be able to cope with a policyholder who has different lines of business issued at different points in time, so that the length of the claims history varies between lines.

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.

Actually, the first and third problem are the same in the ordinary multidimensional Bühlmann-Straub context, and is solved through using the estimated expected claim frequency as weight. The second problem is only a concern in time-dependent models and the problem is solved by treating the unobserved period as a missing value and replacing the missing information with full collective information.

The data for the study come from an active Danish insurance company's commercial portfolio and contain information on claim frequencies for various types of coverage: Fire, water damage, fungus and insect damage, glass damage and theft. For illustrative reasons, we have chosen to focus on only two types of coverage – lines of business – water damage and theft. The original data set contained information from 1995 to 2003 on 19,270 policies having both lines of business. For sake of simplicity (see, discussion above), we excluded policies with less than 4 years of claims history in one of the business lines. This restriction reduced the number of policies to 10,212. Finally, we saved a random subsample of 25% of the policies for testing the estimated models, leaving 7,656 policies for analysis. In this final version, we had 288 claims on water damage and 739 on theft.

The data also contained an estimate of the annually a priori expected number of claims for each policy on each coverage λ_{ijp}, which were originally calculated using a Poisson regression model on the entire data set (see Dionne and Vanasse (1992) or McCullagh and Nelder (1989) for generalized linear models).

We estimated the unknown structural parameters in each credibility model by minimizing (3), and the results are shown in Table 1. Numerical minimization was achieved by an iterative method, with the one-dimensional estimates as starting values for the two-dimensional model parameters.

The estimates of Table 1 reveals a positive correlation between the two lines of business within the same policy ($\tau_{12}^2 = 0.619$). This coincides with our expectations, since it corresponds to unmeasured risk factors which cannot be captured by the a priori differences alone. Further more, expected results of the time-correlation parameters were obtained. Risk profiles for consecutive time periods show positive correlation, as in Pinquet et al. (2001) and even for different types of coverage a positive correlation between consecutive periods was found ($\rho_{12} = 0.351$).

A comparison between the different models was made by applying the estimates of Table 1 to the testing sample containing 2,556 policies. The prediction power of the different models was assessed by calculating the weighted sum of squared residuals for the time-period $J + 1$. All models performed better than the a priori differences λ_{ijp} alone, and the two-dimensional models showed slightly better results than the one-dimensional models.

TABLE 1

ESTIMATED PARAMETERS FOR EACH MODEL USING WEIGHTED LEAST SQUARES

Model	τ_{11}^2	τ_{22}^2	τ_{12}^2	ρ_{11}	ρ_{22}	ρ_{12}
Simple one-dimensional, Theft	0.377	·	·	·	·	·
One-dimensional with time, Theft	0.412	·	·	0.721	·	·
Simple one-dimensional, Water	·	1.686	·	·	·	·
One-dimensional with time, Water	·	1.712	·	·	0.811	·
Simple two-dimensional	0.447	1.702	0.619	·	·	·
Two-dimensional with time	0.461	1.922	0.863	0.865	0.922	0.351

TABLE 2

HISTORY OF FIVE ARBITRARILY CHOSEN POLICIES

Policy	N_{ij1} $\{N_{i11}, N_{i21}, N_{i31}\}$	N_{ij2} $\{N_{i12}, N_{i22}, N_{i32}\}$	λ_{ij1} $\{\lambda_{i11}, \lambda_{i21}, \lambda_{i31}\}$	λ_{ij2} $\{\lambda_{i12}, \lambda_{i22}, \lambda_{i32}\}$
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 3

CREDIBILITY ESTIMATES USING THE PROPOSED MODELS FOR THE FIVE POLICIES

Policy	Line of Business	Crude estimate	One-dimensional		Two-dimensional	
			Simple	With time	Simple	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

We end this section with an investigation of the consequences of implementing the introduced models. This is done by choosing five arbitrary policies and realise how the different models respond to their claims history. Characteristics on the five policies can be seen in Table 2. Table 3 displays estimates of $\hat{\theta}_{ip}$ for the five policies using the different models. The column 'Crude estimate' shows the unsmoothed estimate $\frac{N_{i \cdot p}}{\lambda_{i \cdot p}}$.

Focusing on the first policy, we see that for the theft coverage the one-dimensional models give a rebate on the premium, while the two-dimensional models suggest a raise in the premium. This is because the individual history from the second coverage, water damage, 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 appeared in the previous year (the two-dimensional model without time-dependence weighs all claims equally). If we consider the water damage line for the same policy, both time-dependence models suggest an increase of more than 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 experienced 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.

5. 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. 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 customer performs badly in one line, we will expect her to perform badly in all lines of business, to some extent. 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 multiline experience rating in this study. The model is applied for two lines of business, but can easily be extended to arbitrary numbers of 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 slightly better than the one-dimensional Bühlmann-Straub credibility model in terms of predicting error in a testing sample.

References

BICHSEL, F. (1967) Experience rating in subsets of risks. *ASTIN Bulletin*, **4**, 210-227.

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, Heidelberg.

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

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.

GERBER, H.U. and JONES, D.A. (1975) Credibility Formulas of the Updating Type. *Transaction of the Society of Actuaries*, **27**, 31-52.

GOURIÉROUX, C., MONFORT, A. and TROGNON, A. (1984) Pseudo maximum likelihood methods: applications to Poisson models. *Econometrica*, **52**, 701-720.

JEWELL, W.S. (1973) Multidimensional Credibility. *Operations Research Center* 73-7, Berkeley.

JEWELL, W.S. (1974) Exact multidimensional credibility. *Bulletin of Swiss Association of Actuaries*, **74**, 193-214.

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

LUNDBERG, O. (1966) Une note sur des systèmes de tarification basés sur des modèles du type Poisson composé, *ASTIN Bulletin*, **4**, 49-58.

LÜTKEPOHL, H. (2005) *New Introduction to Multiple Time Series Analysis*, Springer Verlag, Heidelberg.

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

NORBERG, R. (2004) Credibility theory. In *Encyclopaedia of Actuarial Science* (ed. Sundt B. and J. Teugels), Wiley, New York.

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.

SUNDT, B. (1981) Recursive credibility estimation. *Scandinavian Actuarial Journal*, 3-21.

MARTIN ENGLUND
Dept. Research & Development, BI SMC
Royal & SunAlliance Scandinavia
Gammel Kongevej 60
DK-1790 Copenhagen V, Denmark
E-mail: eld@codan.dk

MONTSERRAT GUILLÉN
Dept. Econometrics, RFA-IREA
University of Barcelona
Diagonal, 690
08034 Barcelona, Spain
E-mail: mguillen@ub.edu

JIM GUSTAFSSON
Dept. Research & Development, BI SMC
Royal & SunAlliance Scandinavia
Gammel Kongevej 60
DK-1790 Copenhagen V, Denmark
E-mail: jgu@codan.dk

LARS HOUGAARD NIELSEN
Dept. Biostatistics, Institute of Public Health
University of Copenhagen
Øster Farimagsgade 5B
P.O.Box 20299
DK-1014 Copenhagen K, Denmark
E-mail: l.h.nielsen@biostat.ku.dk

JENS PERCH NIELSEN
Cass Business School
106 Bunhill Row
London EC1Y8TE, United Kingdom
E-mail: Jens.Nielsen.1@city.ac.uk