A SEMI-PARAMETRIC PREDICTOR OF THE IBNR RESERVE

LOUIS G. DORAY[1]

Université de Montréal

ABSTRACT

We develop a semi-parametric predictor of the IBNR reserve in a macro-model when the claim amount for a certain accident and development year can be expressed in a loglinear form composed of a deterministic part and a random error. We need to make assumptions only on the first two moments of the error, without any specified parametric assumption on its distribution. We give its properties, present its advantages and compare the estimates obtained with various predictors of the IBNR reserve, parametric and non-parametric, using a data set.

KEYWORDS

Chain-ladder; regression; least-squares; smearing estimator.

1. INTRODUCTION

In a macro-model, claims are grouped by accident year (year in which the accident giving rise to a claim occurrs) and development year (number of years elapsed since the accident), and data are presented in a trapezoidal array. Taylor (1986) presents a comprehensive survey of various macro methods and models, both deterministic and stochastic, developed to predict incurred but not reported (IBNR) reserves; it is usually assumed that the pattern of cumulative claims incurred or paid is stable across the development years, for each accident year. The problem of setting IBNR reserves consists in predicting for each accident year, the ultimate amount of claims incurred and subtracting the amount already paid by the insurer.

To illustrate the predictor proposed in this paper, we will use the cumulative claims appearing in Doray (1996), which represent the liability claims in thousands of dollars incurred by a Canadian insurance company over the ten-year period 1978-1987. We will perform the analysis on the incremental claims (see Table 1), obtained by differencing successive cumulative amounts, and assume that they are independent. Section 2 presents the loglinear model used, and section 3 the semi-parametric predictor of the IBNR reserve; finally, we compare various predictors of the reserve with the claims of Table 1.

[1] The author gratefully acknowledges the financial support of the Natural Sciences and Engineering Research Council of Canada.

TABLE 1

INCREMENTAL CLAIMS INCURRED

	Development year					
Accident year	1	2	3	4	5	6
1978	8489	1296	924	580	246	126
1979	12970	1796	1435	859	654	265
1980	17522	2783	1469	1023	423	652
1981	21754	2584	1163	783	887	355
1982	19208	2341	1220	619	841	703
1983	19604	2469	1223	1247	612	
1984	21922	2311	1141	1508		
1985	25038	3363	2144			
1986	32532	4474				
1987	39862					

2. A LOGLINEAR MODEL

We consider models of the form $Y_i = \exp(X_i\beta + \sigma\varepsilon_i)$, or expressed as a loglinear regression model,

$$Z_i = \ln Y_i = X_i\beta + \sigma\varepsilon_i, \quad Y_i > 0 \tag{2.1}$$

where Y_i is the ith element of the data vector Y, of dimension n, X is the regression matrix of dimension n x p, whose ith row is the vector X_i, element (i, j) is denoted X_{ij}, and where we assume that the unit vector is in the column space of X, β is the vector (of dimension p) of unknown parameters to be estimated, and ε_i are independent random errors with mean 0 and variance 1.

For the regression parameters, various choices are possible, for example $\alpha_i + \beta_j$ for the stochastic chain ladder model, where i is the accident year and j, the development year, or $\alpha + \beta \ln j + \gamma j + \iota(i + j - 2)$, as in Zehnwirth (1990).

This paper does not study models which rely on parametric assumptions for the distribution of the error ε; instead, we present a semi-parametric regression model which does not assume any particular density for ε, but uses its first two moments only.

3. A PREDICTOR IMPLIED BY THE SMEARING ESTIMATOR

Let us represent by Y_k a value to be predicted, corresponding to a cell in the lower right unobserved triangle of Table 1 (i = 6, ..., 10 and j = 12 – i, ..., 6). Doray (1996) analyzed the two types of errors involved in the prediction of the value Y_k by its expected value, the estimation error on the parameter β from past values and the process error ε_k for a future value, yielding $X_k\tilde{\beta} + \tilde{\sigma}\varepsilon_k$, where X_k is the vector of coefficients of the parameters corresponding to Z_k.

According to Gauss-Markov theory, the least-square estimator $\tilde{\beta} = (X'X)^{-1}X'Z$ is the minimum variance linear unbiased estimator of β, for any distribution of ε such that $E(\varepsilon) = 0$ and $Var(\varepsilon) = 1$. The variance σ^2 is estimated by the mean-square error

$\tilde{\sigma}^2 = (Z - X\tilde{\beta})'(Z - X\tilde{\beta})/(n-p)$. For a fixed vector X_k, $X_k\tilde{\beta}$ is an unbiased and consistent estimator of $X_k\beta$, but $\exp(X_k\tilde{\beta})$ is not in general an unbiased or consistent estimator of $E(Y_k)$. The assumption that ε is normal influences only the efficiency of the estimator $\tilde{\beta}$; if the true error is not normal, the estimator $\tilde{\beta}$ is still consistent and minimum variance linear unbiased. If ε is normal, $\exp(X_k\tilde{\beta} + \tilde{\sigma}^2/2)$ is a consistent estimator for $E(Y_k)$; however, the predictor for Y_k will not be consistent if the assumption that ε is $N(0, 1)$ is wrong.

Duan (1983) proposed the following smearing estimator for the expected value of Y_k, $\frac{1}{n}\sum_{i=1}^{n}\exp(X_k\tilde{\beta} + \tilde{\sigma}\tilde{\varepsilon}_i)$, where $\tilde{\varepsilon}_i = Z_i - X_i\tilde{\beta}$ denotes the least-squares residual. He shows that under certain regularity conditions, the smearing estimator of $E(Y_k)$ is weakly consistent and notes that for small σ^2, its relative efficiency compared to the simple estimator $\exp(X_k\tilde{\beta} + \tilde{\sigma}^2/2)$ is very high when the error distribution is normal (for $\sigma^2 \leq 1.00$ and rank $(X) \geq 3$, it is at least 94%). This efficiency increases as σ^2 decreases or rank (X) increases.

Using the smearing estimator, we can define the following semi-parametric predictor of the IBNR reserve,

$$\hat{\theta}_{SP} = \sum_k \frac{1}{n}\sum_{i=1}^{n}\exp(X_k\tilde{\beta} + \tilde{\sigma}\tilde{\varepsilon}_i) = \left(\sum_k \exp(X_k\tilde{\beta})\right) \times \left(\frac{1}{n}\sum_{i=1}^{n}\exp(\tilde{\sigma}\tilde{\varepsilon}_i)\right),$$

where Σ_k denotes a summation over all cells in the lower triangle to be predicted.

4. COMPARISON OF VARIOUS PREDICTORS

We can obtain a simple approximation for $\hat{\theta}_{SP}$ when σ^2 is small by using the first three terms of the Taylor's series expansion for $\exp(\tilde{\sigma}\tilde{\varepsilon}_i)$, and the facts that $\sum_{i=1}^{n}\tilde{\varepsilon}_i = 0$ and $\sum_{i=1}^{n}\tilde{\sigma}^2\tilde{\varepsilon}_i^2/2 = (n-p)\tilde{\sigma}^4/2$,

$$\hat{\theta}_{SP} \cong \hat{\theta}_A = \left(\sum_k \exp(X_k\tilde{\beta})\right) \times \left[1 + (n-p)\tilde{\sigma}^4/2n\right].$$

In Table 2, we compare the predicted values of the IBNR reserve obtained with the non-parametric predictors, $\hat{\theta}_{SP}$, $\hat{\theta}_A$, the chain-ladder ($\hat{\theta}_{CL}$), and predictors obtained when ε_i's in (2.1) are assumed to be i.i.d. $N(0, 1)$, the uniformly minimum variance unbiased predictor of Doray (1996)

$$\hat{\theta}_U = {}_0F_1\left(\frac{n-p}{2}; \frac{n-p}{4}\tilde{\sigma}^2\right)\sum_k \exp(X_k\tilde{\beta}),$$

where ${}_0F_1(\alpha; z)$ is the hypergeometric function defined as

$${}_0F_1(\alpha; z) = \sum_{j=0}^{\infty}\frac{z^j}{j!(\alpha)_j}, \text{ with } (\alpha)_j = \alpha(\alpha+1)\ldots(\alpha+j-1),\ j \geq 1, \text{ and } (\alpha)_0 = 1,$$

the predictor of Kremer (1982), $\hat{\theta}_K = \sum_k \exp(X_k \tilde{\beta})$, and the simple estimator $\hat{\theta}_1 = \sum_k \exp(X_k \tilde{\beta} + \tilde{\sigma}^2 / 2)$. The model used was the stochastic chain ladder model $(\alpha_i + \beta_j)$, on the claims of Table 1. We notice that $\hat{\theta}_A$, $\hat{\theta}_U$ and $\hat{\theta}_1$ are of the form C x $\hat{\theta}_K$, where C is a factor depending only on $\tilde{\sigma}^2$.

In conclusion, the smearing estimator possesses four important properties. It is easily calculated, consistent, highly efficient if the error ε has a normal distribution and robust against departure from the assumed parametric distribution for ε. It can also be used with transformations other than exponential. The semi-parametric predictor of the IBNR reserve based on the smearing estimator will share those properties and present a worthwhile alternative to predictors based on full parametric assumptions.

TABLE 2

PREDICTION OF THE IBNR RESERVE

Predictor	Predicted value
$\hat{\theta}_{SP}$	23,552
$\hat{\theta}_A$	23,589
$\hat{\theta}_{CL}$	23,919
$\hat{\theta}_U$	24,403
$\hat{\theta}_K$	23,549
$\hat{\theta}_1$	24,404

REFERENCES

DORAY, L. G. (1996). UMVUE of the IBNR Reserve in a Lognormal Linear Regression Model, *Insurance: Mathematics and Economics.* **18,** 43-57.

DUAN, N. (1983). Smearing Estimate: A Nonparametric Retransformation Method. *Journal of the American Statistical Association.* **78,** 605-610.

KREMER, E. (1982). IBNR claims and the two-way model of ANOVA. *Scandinavian Actuarial Journal,* 47-55.

TAYLOR, G. C. (1986). Claims reserving in non-life insurance. North-Holland, Amsterdam.

ZEHNWIRTH, B. (1990). Interactive Claims Reserving Forecasting System, Version 5.3, Manual Volume 1, Benhar Nominees Pty. Ltd., Australia.

DR. LOUIS DORAY

Département de mathématiques et de statistique

Université de Montréal

C.P. 6128, Succ. Centre-ville. Montréal, Qué., Canada, H3C 3J7