COMPARISON OF SOME METHODS TO FIT
A MULTIPLICATIVE TARIFF STRUCTURE
TO OBSERVED RISK DATA

By B. Ajne

Skandia, Stockholm

Abstract

Three methods for fitting multiplicative models to observed, cross-classified risk data are compared. They are the method of Bailey–Simon, the method of marginal totals and a maximum likelihood method. The methods are applied to a number of risk data sets and compared with respect to balance and goodness-of-fit.

Keywords

Multiplicative models, Tariff structures.

1. INTRODUCTION

The setting is as follows.

(a) For a certain insurance portfolio we have at our disposal a number of rating factors or *tariff arguments* $U, V, W, \ldots$. E.g. in a motor portfolio we could have $U =$ the age of the car, $V =$ the home district of the owner of the car, $W =$ the mileage of the car per policy year.

(b) Each tariff argument assumes a finite number of values or *levels*, which may be denoted by consecutive integers so that

$$U = 1, 2, \ldots, m_u$$
$$V = 1, 2, \ldots, m_v$$
$$W = 1, 2, \ldots, m_w \quad \text{etc.}$$

The insurance portfolio is thus divided into

$$m = m_u m_v m_w \cdots$$

disjoint classes or cells, a cell being defined as all members of the portfolio corresponding to a certain combination of levels of the tariff arguments. E.g., with 14 age classes, 7 home districts and 5 mileage classes, we will in the example mentioned above have

$$m_u = 14, \qquad m_v = 7, \qquad m_w = 5 \quad \text{and} \quad m = 490.$$

A typical cell will be denoted by c. The corresponding values of $U, V, W, \ldots$ will always be denoted by $i, j, k, \ldots$ respectively.

(c) For each cell c we have observed risk data consisting of

—an exposure n_c (e.g., number of policy years or insurance sum under risk)

—a relative risk measure p_c consisting of a claims total for the cell (e.g., number of claims or claims amount) divided by the exposure

(d) To the observed relative risk measures p_c we want to fit numbers f_c of the form

$$f_c = \bar{p} u_i v_j w_k \ldots, \qquad \bar{p} = \sum n_c p_c / \sum n_c.$$

In other words, we want to impose a multiplicative tariff structure on the insurance portfolio. Putting all u_i, v_j, $w_k \ldots$ equal to one yields the simplest possible tariff structure, where all cells get the same premium $\bar{p}$.

2. METHODS COMPARED

Formally, the multiplicative structure has $m_u + m_v + m_w + \cdots$ parameters u_i, v_j, w_k, ... i.e., one for each possible level of the tariff arguments. The number of free parameters must, however, be less. This is shown by the fact that we can multiply all u-values and divide all v-values by the same positive number without affecting the set of numbers f_c. The number of free parameters is $1 + (m_u - 1) + (m_v - 1) + (m_w - 1) + \cdots$ i.e., one parameter for the overall level and the remaining ones for the "relativities" u_2/u_1, $u_3/u_1 \ldots$, v_2/v_1, $v_3/v_1, \ldots$, w_2/w_1, $w_3/w_1, \ldots$.

This is considerably less than m, in the example given above 24 as compared to $m = 490$, so a perfect fit of the f_c to the p_c cannot be expected. This, on the other hand, is of course the very idea in introducing a tariff structure. It should graduate, i.e., simplify, observed risk data.

In the following we will study three methods of fitting the multiplicative structure to the observed risk data. The first two of these have also been discussed by VAN EEGHEN, NIJSSEN and RUYGT (1982). Reference may also be made to VAN EEGHEN, GREUP and NIJSSEN (1983) and to the further references given there.

We will now briefly describe the three methods for estimating the parameters u_i, v_j, w_k, ...

(1) *Minimum Chi-Square (or Bailey-Simon)*:

Put chi-square $= \sum_c n_c (p_c - f_c)^2 / f_c$. The parameters are determined so that chi-square is minimized.

(2) *Marginal Totals*

The parameters are determined so that

$$\sum_M n_c f_c = \sum_M n_c p_c$$

for each marginal M, i.e., for each fixed i, each fixed j, each fixed k, ...

In the case of numbers of claims as observed claims totals in the cells, and under the assumption that they are stochastically independent and Poisson-distributed with respective parameters $n_c f_c$, this coincides with the method of maximum-likelihood (JUNG, 1968).

In a handy, menu-operated APL-program at our disposal we have included a third method,

(3) ML-special

This is the maximum-likelihood method under the assumption that the relative risk measures are independently distributed according to the normal distribution and that

$$E(p_c) = f_c, \qquad \text{Var}(p_c) = \sigma^2 f_c / n_c.$$

Here σ^2 is an unknown proportionality factor common to all cells. It may be noted that in the case of Poisson distributed numbers of claims, referred to above, these equations hold with $\sigma^2 = 1$. They also imply that chi-square/m is an unbiased estimator of σ^2, chi-square being defined under (1) above.

In the case of observed claims amounts, generated as sums of independent and identically distributed individual claims, the equations above will hold true i.e.,

—if the number of individual claims is Poisson distributed and if the sizes of the individual claims are independent of the number of claims and if

$$\text{Var}(X_c) = E(X_c)(\sigma^2 - E(X_c))$$

where X_c denotes a typical individual claims size for cell c.
—if the number of individual claims is deterministic and if

$$\text{Var}(X_c) = \sigma^2 E(X_c).$$

By using method (3) we should maximize the likelihood function corresponding to the assumptions made. This is equivalent to minimizing chi-square/$\sigma^2 +$ $m \log \sigma^2 + \sum_c \log f_c$ with respect to u_i, v_j, $w_k \ldots$ and σ.

3. RESULTS

The table in the appendix summarizes some experience in using the above mentioned methods to fit multiplicative tariff structures to observed risk data.

The first column gives a brief description of the risk data, the number of tariff arguments, the number of levels of each argument and the total number of cells.

In column "Cat" an "a" denotes that claims amounts are observed and a "n" denotes that claims numbers are observed (only two cases).

The "Size req" states how many of the observed marginal totals do *not* fulfil the size requirement

$$\sum_M n_c p_c \geq 9 \text{ chi-square}/m.$$

The total number of marginal totals is given within parenthesis. The size requirement is a very approximate rule of thumb. It expresses the desire that each observed marginal total be equal to at least three times its estimated standard

deviation according to the model under (3) above. In the size requirement the chi-square of method (3) is thus used. For materials not analyzed by method (3), the chi-square of method (2) was used instead.

The three colums headed "Balance" are computed from quotients

$$S = \sum n_c f_c / \sum n_c p_c$$

where the estimated parameter values are inserted into f_c. Each S-value thus is a quotient between graduated and observed claims totals. The S-values are computed for all marginals (the largest and the smallest marginal S is given in the table) and for the whole material ("Total").

The remaining four columns describe the goodness-of-fit. Var red (variance reduction) is computed, for marginals and the total as $1 - \sum n_c (p_c - f_c)^2 / \sum n_c (p_c - \bar{p})^2$ i.e., gives the variance reduction relative to the structure with all cell premiums equal to $\bar{p}$.

The column "chi-2" refers to values of chi-square, computed by inserting estimated parameter values into f_c. The values for methods (2) and (3) are compared to the value for method (1), which is the minimum value under the multiplicative structure.

If claims numbers are observed and if they are independent and Poisson distributed with parameters according to the multiplicative structure, minimum chi-square is for large exposures approximately chi-square distributed with degrees of freedom equal to the number of cells minus the number of free parameters. This may be used to investigate departures from the hypothesis of a multiplicative structure.

It can be proved that for method (1)

$$\text{Minimum chi-square} = 2(\text{Total } S - 1) \times \sum n_c p_c$$

so that investigations based on minimum chi-square may as well be based on the total balance of this method.

4. DISCUSSION

The general impression is that method (2) is the best one of those three studied. It is, by its very definition, marginally and totally balanced. It gives variance reductions superior to method (1) and superior or equivalent to those of method (3). Of course, it gives a higher chi-square than does method (1), but where the difference is great, the latter method tends to show a disturbing lack of balance.

Method (3) sometimes has a smaller chi-square than method (2). It is also balanced for the total, but it may give, occasionally, rather low balance values for marginals with a small claims total.

Method (1) always has marginal and total balance greater than or equal to one. The safety margin tends to be larger for marginals with small claims totals. An appealing feature of this method is the earlier mentioned possibility to interpret its (lack of) total balance as a measure of the departure from the multiplicative structure.

Material m	Cat	Size Req	Method	Balance			Var Red %		Chi-2 Relat. to 1)
				Marginal Max	Min	Total	Least Marg	Total	
1. Motor, glass damage	a	0(26)	(1)	1.26	1	1 005	63.1	94 5	1
$14 \times 7 \times 5 = 490$			(2)	1	1	1	65 4	94 5	1 01
			(3)	1	0 88	1	63.5	94.5	1 02
2 Motor, hull	n	0(25)	(1)	1 51	1 01	1 030	−1 5	62 1	1
$13 \times 7 \times 5 = 455$			(2)	1	1	1	8 9	66 6	1 11
			(3)	1 21	0 92	1	9 8	66 2	1 10
3. Motor, hull	n	0(19)	(1)	1 15	1	1 009	10 8	88.2	1
$7 \times 5 \times 7 = 245$			(2)	1	1	1	19.9	88 8	1.03
			(3)	1	0 97	1	20 6	88 8	1 03
4 "–	a	0(19)	(1)	1 28	1 01	1.019	17 0	82 2	1
			(2)	1	1	1	24 9	83 6	1.04
			(3)	1.01	0.98	1	25 2	83 6	1 05
5 Motor, hull-large vehicles	a	3(6)	(1)	1.82	1.02	1 120	−2180	53 2	1
$2 \times 2 \times 3 = 12$			(2)	1	1	1	−574	70 7	1.34
			(3)	1 26	0 72	1	−958	62 8	1.16
6. Property, contents	a	1(15)	(1)	2.14	1.05	1.156	−102	24 9	1
$5 \times 4 \times 6 = 120$			(2)	1	1	1	−7 6	40.1	1 92
			(3)	1.60	0.06	1	−112	34.7	1 47
7 Property, buildings	a	2(15)	(1)	2.07	1 12	1 239	−58 7	31.7	1
$5 \times 4 \times 6 = 120$			(2)	1	1	1	−12 7	44 3	1.36
8. Property, buildings—fire	a	0(36)	(1)	6 20	2.05	2.701	−169	−65	1
$7 \times 2 \times 3 \times 10 \times 6 \times 8 = 20\ 160$			(2)	1	1	1	−0.8	1.3	2 57
9 Property, buildings-not fire	a	0(36)	(1)	5.73	1.39	1.760	−92	−34	1
$7 \times 2 \times 3 \times 10 \times 6 \times 8 = 20\ 160$			(2)	1	1	1	0.9	4 3	2.62
10 Motor, motorcycles	n	1(16)	(1)	1 58	1	1.019	19.5	88.6	1
$6 \times 10 = 60$			(2)	1	1	1	38 3	89.3	1.05
			(3)	1.03	0 66	1	37 1	88.9	1 13

As a support of the propositions above it may be mentioned that for materials 1-4 and 10 max marginal balance for method (1) and min marginal balance for method (3) occur at marginals with claims totals ranging from one to six percent of the average claims total for marginals belonging to the same tariff argument. Except for material No. 2, the marginal concerned is the same for both methods. Also the low value for method (3) in material No. 6 occurs for a very small risk group.

Method (2) also seems to be less sensitive to outlying observations. Actually, one of the motor materials had such an observation (caused by input error). After correction it turned out that this had disturbed the results of methods (1) and (3) much more than that of method (2).

As to the materials, the motor data 1-4 are well-bahaved and show a good fit to the multiplicative structure. The three methods also generally gave very similar results for the relative sizes of the factor parameters.

Material No. 10 is taken from FOLKESSON, NEUHAUS and NORBERG (1985). It also shows an acceptable fit to the multiplicative structure.

For materials No. 3 and 10 the total number of claims is 15 797 and 3027 respectively. Minimum chi-square is 284 (d.f. = 228) and 115 (d.f. = 45). The values are significant on the respective levels 1% and 0.05% giving rather strong evidence against an *exact* multiplicative Poisson model.

The remaining materials do not behave that well. This may be explained by

(a) the size requirement is not fulfilled, and/or
(b) more dangerous distributions of individual claims sizes, and, in cases 7 and 8,
(c) the very drastic reduction in the number of free parameters, from 20 160 cells to $36 - 6 + 1 = 31$ free parameters in the multiplicative structure. This results in a very low variance reduction relative to the single premium structure, which still may be significant as judged by an F-test.

6. ACKNOWLEDGEMENT

REFERENCES

VAN EEGHEN, J , GREUP E K and NIJSSEN, J A (1983) *Rate Making, Surveys of Actuarial Studies No 2* Nationale-Nederlande N.V Rotterdam.

VAN EEGHEN, J , NIJSSEN, J A. and RUYGT, F. A. M. (1982) Interdependence of Risk Factors Application on some Models In *New Motor Rating Structure in the Netherlands* ASTIN-group Nederland

FOLKESON, M., NEUHAUS, W. and NORBERG, R. (1985) A Hierarchical Credibility Model Applied to Insurance data Paper presented at the 18th ASTIN Colloquium at Biarritz

JUNG, J. (1968) On Automobile Insurance Rate Making, *ASTIN Bulletin* 5 (1)

B. AJNE
Skandia, Sveavägen 44, *S*-103 50 *Stockholm, Sweden.*