

HOW TO FIND THE RIGHT SUBDIVISION INTO TARIFF CLASSES

H. SCHMITTER AND E. STRAUB

Zurich

I. INTRODUCTORY REMARKS

By intuition, the subdivision of an insurance portfolio into a number of classes is said to be good if it reflects the heterogeneity of the portfolio in an efficient way. To illustrate this rather vague statement we take the following very simple example:

The portfolio consists of 20 independent risks, 10 of them producing an expected loss ratio of say 30% each (type A risks) and 80% each (type B risks) respectively.

This "natural" subdivision is certainly better than, for instance,
no subdivision at all;
or, the finest possible subdivision with 20 classes consisting of only 1 risk each (because there is no point in differentiating between risks of the same type);
or, 5 classes each containing two A- and two B-risks (here the number of classes is unnecessarily high and the heterogeneity has been completely wiped out; statistics based on this subdivision would even make us believe that the portfolio is totally homogeneous).

As a matter of fact, the above "natural" subdivision is—of course!—the best of all subdivisions, it is the optimal subdivision in this case.

In practice, however, as we all know, it is not easy to find the optimal subdivision. For one thing, the inherent structure or "natural" subdivision is not known a priori and secondly, for many different reasons, we can only choose from a limited number of subdivisions and not from all theoretically possible solutions. Note that even with only 20 risks there are 58.10¹² possibilities of subdividing the portfolio. Thus, in practice, there is only a relatively

small number of *admissible* subdivisions and the optimal one may not be among them, but we still need some sort of statistical criterion to choose the best one from these admissible subdivisions.

2. THE MODEL

Let the structure (= "natural" subdivision) of the portfolio be given by N classes, assume the existence of claims statistics over the last n years and let class no. j be characterized by the risk parameter θ_j . If P_{hj} denotes the number of risks (or total sum insured or underlying premium volume, or any other measure of volume) and X_{hj} the loss ratio of class j in year h respectively, we further assume independence of the X_{hj} and

$$E[X_{hj} | \theta_j = \lambda_j] = \mu(\lambda_j), \text{Var}[X_{hj} | \theta_j = \lambda_j] = \frac{\sigma^2(\lambda_j)}{P_{hj}}.$$

We now consider the above-mentioned admissible subdivisions of the portfolio: Let there be L of such admissible subdivisions which we can choose from, and let them be numbered from $g = 1$ to $g = L$. If now $P_{hr}^{(g)}$ and $X_{hr}^{(g)}$ denote the volume and loss ratios of year h and class r of the subdivision g , we then—under the above assumptions—have for

$$\underline{\theta} = (\theta_1, \theta_2, \dots, \theta_N) \text{ and } \underline{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_N)$$

$$E[X_{hr}^{(g)} | \underline{\theta} = \underline{\lambda}] = \sum_{j=1}^N \frac{P_{hrj}^{(g)}}{P_{hr}^{(g)}} \mu(\lambda_j), \text{Var}[X_{hr}^{(g)} | \underline{\theta} = \underline{\lambda}] = \sum_{j=1}^N \frac{P_{hrj}^{(g)}}{P_{hr}^{(g)}} \sigma(\lambda_j)$$

where $P_{hrj}^{(g)}$ denotes the number of risks in the r -th class of subdivision g which at the same time belong to the j -th class of the natural subdivision.

These equations hold for $h = 1, 2, \dots, n$ and $r = 1, 2, \dots, N_g$ (N_g being the number of classes within subdivision g) and $g = 1, 2, \dots, L$.

Here the vector $\underline{\theta}$ of risk parameters is regarded as a stochastic vector of N independent and identically distributed components.

Actually, it is only necessary to assume

$$E[\mu(\theta_j)] = m, E[\sigma^2(\theta_j)] = v \text{ and } \text{Var}[\mu(\theta_j)] = w \text{ independent of } j$$

in order to get

$$E[\text{Var} [X_{hr}^{(g)} | \underline{\theta}]] = \frac{v}{P^{(g)}_{hr}} \text{ and } \text{Var} [E[X_{hr}^{(g)} | \underline{\theta}]] = w \sum_{j=1}^N \frac{P^{2(g)}_{hrj}}{P^{2(g)}_{hr}} (\leq w).$$

For averages over time, i.e. for

$$X_{.r}^{(g)} = \sum_{h=1}^n \frac{P^{(g)}_{hr}}{P^{(g)}_{.r}} X_{hr}^{(g)} \text{ with } P_{.r}^{(g)} = \sum_{h=1}^n P_{hr}^{(g)}$$

we get the same type of formulae, namely

$$E[\text{Var} [X_{.r}^{(g)} | \underline{\theta}]] = \frac{v}{P_{.r}^{(g)}} \text{ and } \text{Var}[E[X_{.r}^{(g)} | \underline{\theta}]] = w \sum_{j=1}^N \frac{P_{.rj}^{2(g)}}{P_{.r}^{2(g)}} \leq w$$

3. CHOOSING THE BEST SUBDIVISION FOR THE TARIFF

We shall see in the sequel that the statistical value

$$T^{(g)} = (N_g - 1) (W^{(g)} - V^{(g)})$$

with
$$W^{(g)} = \frac{1}{N_g - 1} \sum_{r=1}^{N_g} \frac{P_{.r}^{(g)}}{P} (X_{.r}^{(g)} - X)^2$$

and
$$V^{(g)} = \frac{1}{N_g} \sum_{r=1}^{N_g} \frac{1}{n-1} \sum_{h=1}^n \frac{P_{hr}^{(g)}}{P} (X_{hr}^{(g)} - X_{.r}^{(g)})^2$$

where
$$X = \sum_{r=1}^{N_g} \frac{P_{.r}^{(g)}}{P} X_{.r}^{(g)} \text{ and } P = \sum_{r=1}^{N_g} P_{.r}^{(g)},$$

has the property that its mathematical expectation is maximum if g is the natural subdivision.

For a proof of the above statement, we

write
$$W = \sum_{j=1}^N \frac{P_{.j}}{P} (X_{.j} - X)^2, \quad V = \frac{1}{N} \sum_{j=1}^N \frac{1}{n-1} \sum_{h=1}^n \frac{P_{hj}}{P} (X_{hj} - X_{.j})^2$$

and $T = (N - 1) (W - V)$ for the natural subdivision and show

$$\begin{aligned} E[T^{(g)}] &= E[(N_g - 1) (W^{(g)} - V^{(g)})] \leq E[(N - 1) (W - V)] = \\ &= E[T] = w \sum_{j=1}^N \frac{P_j}{P} \left(1 - \frac{P_j}{P} \right) \end{aligned}$$

where equality holds if g is the natural subdivision or subdivides it further. In the latter case $E[W^{(g)}] < E[W]$. In practice it therefore is reasonable to choose the subdivision g that shows the highest value of $W^{(g)}$ among those with high $T^{(g)}$ -values. Then one could hope that it will come closest to the optimal, i.e. the natural subdivision of the portfolio.

We first note that

$$E[X_{hr}^{(g)^2}] = m^2 + w \sum_{j=1}^N \frac{P_{hrj}^2(g)}{P_{hr}^2(g)} + \frac{v}{P_{hr}^{(g)}}$$

and

$$E[X_r^{(g)^2}] = m^2 + w \sum_{j=1}^N \frac{P_{rj}^2(g)}{P_r^2(g)} + \frac{v}{P_r^{(g)}}$$

(In the following we write P_r and P_{rj} instead of $P_{.r}$ and $P_{.rj}$) and therefore

$$E[W^{(g)}] = \frac{v}{P} + \frac{w}{N_g - 1} \sum_{j=1}^N \left(\sum_{r=1}^{N_g} \frac{P_{rj}^2(g)}{PP_r^{(g)}} - \frac{P_j^2}{P^2} \right)$$

and

$$E[V^{(g)}] = \frac{v}{P} + \frac{w}{PN_g(n-1)} \sum_{j=1}^N \sum_{r=1}^{N_g} \left[\left(\sum_{h=1}^n \frac{P_{hrj}^2(g)}{P_{hr}^{(g)}} - \frac{P_{rj}^2(g)}{P_r^{(g)}} \right) \right]$$

Next we prove that

$$E[V^{(g)}] \geq \frac{v}{P}$$

and for this purpose define

$$\varepsilon_h = \frac{P_{hrj}^{(g)}}{P_{hr}^{(g)}} - \frac{P_{rj}^{(g)}}{P_r^{(g)}}$$

Summing up the squares ε_h^2 multiplied with $P_{hr}^{(g)}$

we get

$$\sum_{h=1}^n \varepsilon_h^2 P_{hr}^{(g)} = \sum_{h=1}^n \frac{P_{hrj}^2(g)}{P_{hrj}^{(g)}} - \frac{P_{rj}^2(g)}{P_r^{(g)}} \geq 0$$

showing that the second term in $E[V^{(g)}]$ is nonnegative and equal to zero if and only if for each r and j

$$\frac{P_{hrj}^{(g)}}{P_r^{(g)}} = \frac{P_{rj}^{(g)}}{P_r^{(g)}}.$$

Looking at $E[W^{(g)}]$ again, we have

$$\sum_{j=1}^N \sum_{r=1}^{N_g} \frac{P_{rj}^{2(g)}}{P_r P_r^{(g)}} = \sum_{r=1}^{N_g} \sum_{j=1}^N \frac{P_{rj}^{2(g)}}{P_r P_r^{(g)}} \leq \sum_{r=1}^{N_g} \frac{1}{P_r P_r^{(g)}} \left(\sum_{j=1}^N P_{rj}^{(g)} \right)^2 = 1$$

Thus, the maximum of $E[W^{(g)}]$ is attained for $N_g \geq N$ if and only if for each $r = 1, 2, \dots, N_g$ the equality

$$\sum_{j=1}^N P_{rj}^{2(g)} = \left(\sum_{j=1}^N P_{rj}^{(g)} \right)^2$$

holds, i.e. if—apart from possible permutations of the subscripts r and j — $P_{rj}^{(g)} = \delta_{rj} P_j$ and $N_g = N$, where $\delta_{rj} =$ Kronecker symbol, that is to say, if g is the natural subdivision. But in this case $E[V^{(g)}]$ is a minimum at the same time, since $P_{rj}^{(g)} = \delta_{rj} P_j$ involves

$$\frac{P_{hrj}^{(g)}}{P_r^{(g)}} = \frac{P_{rj}^{(g)}}{P_r^{(g)}} = \delta_{rj}.$$

Also note that the maximum value of $E[T^{(g)}]$ is given by

$$\max E[T^{(g)}] = w \sum_{j=1}^N \frac{P_j}{P} \left(1 - \frac{P_j}{P} \right).$$

Conclusion: Choosing from different possible subdivisions of a portfolio into risk classes, we take among those which maximize (apart from an unknown factor) an unbiased estimator for the variance w of the individual loss ratio the one that maximizes $W^{(g)}$.

A NUMERICAL EXAMPLE

R. GRÜNIG

Zürich

Let us look at a motor portfolio for which we possess data of five years, broken down according to two criteria: Age of driver and

horsepower. We have three age groups A_1, A_2, A_3 , and three horsepower groups H_1, H_2, H_3 , i.e. 9 classes in all.

We now seek the best subdivision of our portfolio.

As admissible subdivisions we allow all combinations of the age groups (5 possibilities) and combinations of neighbouring horsepower groups (4 possibilities). We get $5 \cdot 4 = 20$ cases.

To test the method, we have drawn the loss ratios for each class from lognormal distributions with the following means μ and standard deviations σ

	H_1	H_2	H_3	
μ	1.00	0.55	0.55	A_1
σ	0.25	0.33	0.33	
μ	0.70	0.85	0.85	A_2
σ	0.21	0.425	0.425	
μ	1.00	0.55	0.55	A_3
σ	0.25	0.33	0.33	

As can be seen from the table, we took the same distributions for age groups A_1 and A_3 and for the horsepower groups H_2 and H_3 . We can therefore combine these groups which leads to the theoretically "natural" subdivision ($A_1 + A_3, H_2 + H_3$).

	H_1	$H_2 + H_3$ combined	
μ	1.00	0.55	$A_1 + A_3$ combined
σ	0.25	0.33	
μ	0.70	0.85	A_2
σ	0.21	0.425	

For the five year premiums (which can also be interpreted as number of risks or sum insured) we used

H_1	H_2	H_3	
110	134	247	A_1
445	700	482	A_2
249	367	424	A_3

The simulations led to the following five years loss ratio

	H_1	H_2	H_3	
	1.018	0.381	0.437	A_1
	0.690	0.966	0.672	A_2
	1.040	0.398	0.528	A_3

Based on these figures, the statistics $T^{(g)}$ and $W^{(g)}$ for the 20 admissible cases were calculated as:

no of subdivision	N_g	Age groups combined	HP groups combined	1000 $T^{(g)}$	1000 $W^{(g)}$
1	9	—	—	<u>39.8</u>	<u>6.6</u>
2	6	1 + 2	—	26.5	6.9
3	6	1 + 3	—	<u>41.4</u>	<u>10.5</u>
4	6	2 + 3	—	12.7	4.1
5	6	—	1 + 2	9.5	3.8
6	6	—	2 + 3	<u>35.9</u>	<u>8.8</u>
7	4	1 + 2	1 + 2	8.2	4.6
8	4	1 + 2	2 + 3	18.6	8.0
9	4	1 + 3	1 + 2	10.9	6.2
10	4	1 + 3	2 + 3	<u>36.8</u>	<u>14.5</u>
11	4	2 + 3	1 + 2	7.2	4.1
12	4	2 + 3	2 + 3	12.1	5.4
13	3	1 + 2 + 3	—	7.8	5.9
14	3	—	1 + 2 + 3	6.7	5.9
15	2	1 + 2 + 3	1 + 2	7.6	9.7
16	2	1 + 2 + 3	2 + 3	5.0	7.1
17	2	1 + 2	1 + 2 + 3	1.4	4.3
18	2	1 + 3	1 + 2 + 3	7.9	11.6
19	2	2 + 3	1 + 2 + 3	2.0	3.4
20	1	1 + 2 + 3	1 + 2 + 3	0	0

Comment:

The subdivisions $g = 1, 3, 6$ (all three being subdivisions of $g = 10$) and 10 (natural one) show high $T^{(g)}$ -values. Among these we find the highest $W^{(g)}$ -value for $g = 10$.