TARIFF THEORY

PAAVO PITKÄNEN
Finland

I. INTRODUCTION

I 1. *The requirements on tariffs*

When an insurance company accepts new insurances or when the premiums of earlier accepted insurances have to be changed on renewal the company has to

— search for the factors that influence the premium and
— calculate the premium according to the values of these factors.
In order to calculate the premiums the company gathers data consisting of factors eventually influencing the amount of claims. On the basis of these data the company calculates the tariff which has to fulfill the following general principles:

1. The tariff has to be as correct as possible in relation to different risk groups.
2. The structure of the tariff has to be such that the calculation of the insurance premium is quite straightforward. With this in mind the factors influencing the tariff have to be few enough and the structure of the tariff has to be the simple (e.g. linear or multiplicative) function of the factors or it should be rather easy to put them into tabular form.

These principles are partly contradictory. If the premium is correct, the structure of the tariff is not usually simple.

I 2. *Formulation of the problem*

Let us assume that the total amount of the claims on a certain risk time is the random variable Y.

To be able to calculate the tariff we have gathered the variables $x_1, \ldots, x_n$ which may have an influence on the amount of the claims.

https://doi.org/10.1017/S0515036100009338 Published online by Cambridge University Press

These variables can be of a qualitative or a quantitative nature. Into this group of possible risk variables are to be included the variables which are known to be related to the total amount of the claims and furthermore the variables that could be related to it. To take an example: In motor car insurance x_1 can indicate the area where the vehicle is driven, x_2 the sex, x_3 the engine's stroke capacity, x_4 the age of the vehicle, etc.

Now have we two main problems:

1. *The selection problem*

From the group of the possible risk variables $x_1, \ldots, x_n$ we must select the variables $x_{i_1}, \ldots, x_{i_k}$ which have a significant influence on the amount of the claims Y.

We denote these variables by $x_1, \ldots, x_k$ and let us call them tariff variables.

As soon as these variables have been fixed, every risk can be represented as a point $(x_1, \ldots, x_k)$ in *a* k-dimensional space.

The most difficult question concerned in this problem is to specify which is a significant influence. This question will come up later in this presentation in chapter III.

2. *The tariff construction problem*

We have to calculate the premium as a function of the tariff variables chosen in accordance with the above mentioned factors, i.e.

$$P = P(x_1, \ldots, x_k)$$

Our aim is to solve the problem exactly in this order, i.e. we first have to search for the tariff factors and then construct the tariff.

The research concerning tariff theory usually only deals with problem 2. In this presentation we first analyze previous publications concerning tariff construction and then we examine the problem from the viewpoint of this presentation. In searching for the factors influencing the tariff we make no assumptions concerning the structure of the tariff.

I 3. *The risk premium and the collective premium*

Let us assume that we have already solved the selection problem presented in the previous chapter. Every risk can thus be indicated individually using the values of the tariff factors $x_1, \ldots, x_k$. Thus we get $X = (x_1, \ldots, x_k)$ as a combination of these values. According to Bühlmann's [5] practice we can now make a definition as follows:

1. *The risk premium* is the premium $P(X)$ corresponding to the value combination of the tariff factors, and thus it is defined for each value combination separately.
2. *The collective premium* is the combined premium calculated on the different value combination of the different tariff factors. In practice it is usually difficult to use all the k tariff factors. Correspondingly if some of the tariff factors can have many different values, it is in practice preferable to classify the values into a few classes. Thus many different value groups form a value class $\{X_\nu\}$, $\nu = 1, 2, \ldots$ and all the risks belonging to this class have the same premium, which is a collective one.

   All insurance premiums can in practice be considered as collective premiums, as not all tariff factors can for practical purposes be counted as factors influencing the premium. The collective premium thus depends on the distribution of the unused tariff factors. If the distribution of these factors changes, the collective premium should also be revised. Let's take an example:

   Assume that, in motor insurance the type of brakes (disc brakes, drum brakes) influences the amount of the claims, but for practical purposes this variable is not a tariff factor. If all motor vehicle manufacturers started to produce only disc brakes, this should influence the collective premium too.

I 4. *The premium*

If the distribution of the amount of the claims Y upon the risk X is known to be $F_X(y)$, the amount of the premium can be calculated according to the following principles:

1. The expected value principle

   $P(X) = (1 + \lambda)EY = (1 + \lambda) \int y dF_X(y)$,

   where λEY is the safety loading.

2. The standard deviation principle

   $P(X) = EY + \alpha\sigma(Y)$,

   where $\sigma^2(Y) = \int (y - EY)^2 \, dF_X(y)$ and α the safety loading.

3. The variance principle

   $P(X) = EY + \beta\sigma^2(Y)$,

   where β is the safety loading.

4. The utility function principle

   The premium $P(X)$ is arrived at as a result of the equation

   $E[u(P(X) - Y)] = \gamma$

   where $u(x)$ is the utility function of the company profit.

   This function should usually fulfil the following requirements:

   — $u(x)$ has to be continuous
   — $u(x)$ has to be non-decreasing
   — $u'(x)$ has to be non-increasing.

   Thus this function measures the profit achieved by the company. The constant γ also represents the safety loading including the profit desired or expected by the company.

Let us have a closer look at these principles.

The principle of calculating the premium is additive, if the premium assigned to the sum of two independent risks is the sum of the premiums that are assigned to the two risks individually. E.g. the premium in fire insurance is additive, if the total of two houses insured separately is equal to the premium for the two houses insured as one object only.

This principle is to be considered as very practical both in regard of practice as in theory.

1. The expected value principle is the most common principle of premium calculation and it is easy to see that it fulfills the requirement of additivity.

2. The standard deviation principle does not fulfil the requirement of additivity, if both of the variances differ from 0, because

$$\begin{aligned} P(X_1 + X_2) &= E(Y_1 + Y_2) + \alpha \sqrt{\sigma^2(Y_1) + \sigma^2(Y_2)} \\ &\neq [EY_1 + \alpha\sigma(Y_1)] + [EY_2 + \alpha\sigma(Y_2)] \\ &= P(X_1) + P(X_2), \end{aligned}$$

   where $X_1 + X_2$ means the risk which is the sum of the risks X_1 and X_2.

Instead it can be established that if the distribution of the amount of the claims is normal, then

$$P\{Y - P(X) \leq k\sigma(Y)\} = P\{\xi \leq \alpha + k\} = \text{constant},$$

where $\xi \sim N(0, 1)$.

This characteristic feature does not however apply to the insurance business as a whole.

It is also possible to assume that the safety loading $\alpha\sigma(Y)$ covers reinsurance expenses. Furthermore it can be assumed that large changes in certain risks demand a higher premium, as such a business involves a threat to the company's security.

3. The variance principle fulfils the additivity requirement, as

$$\begin{aligned} P(X_1 + X_2) &= E(Y_1 + Y_2) + \beta\sigma^2(Y_1 + Y_2) \\ &= EY_1 + EY_2 + \beta[\sigma^2(Y_1) + \sigma^2(Y_2)] \\ &= P(X_1) + P(X_2), \end{aligned}$$

as it is assumed that the risks are independent.

The safety loading $\beta\sigma^2(Y)$ is equivalent to the safety loading $\alpha\sigma(Y)$.

4. The utility function principle is very interesting in theory, but in practice may be of very little importance. The fulfillment of the additivity requirement depends on the utility function. Here it may be established that if the utility function is linear, the result will be $P(X) = EY$.

Usually $\gamma = 0$, but according as $\lambda > 0$, $\alpha > 0$, $\beta > 0$ we can require that $\gamma > 0$.

In this presentation we only examine the calculation of the expected value EY of the amount of the claims. The safety loading λEY, $\alpha\sigma(Y)$, $\beta\sigma^2(Y)$ or γ should be examined in connection with the risk theoretic research of the company.

II. Premium Calculation with given Tariff Factors

II 1. *The expected value of the amount of the claims*

In this chapter we examine the methods that have been used in calculating the premium when the tariff factors are given.

Usually the estimate of the expected value of the amount of the claims is calculated by multiplying the estimated number of claims

https://doi.org/10.1017/S0515036100009338 Published online by Cambridge University Press

by the estimated average claim. This method can be motivated by the following theoretical formula:

It is assumed that

$Y = Z_0 + Z_1 + \ldots + Z_n$ the total amount of the claims, if there have been n claims during the period of insurance.

$Z_0 = 0$

$Z_i =$ size of the i:th claim ($i \geq 1$).

$N =$ number of the claims during the period of insurance.

If the variables Z_i are independent and identically distributed, $EY = EZ \,.\, EN$.

Similarly we may expect that the same equation is also valid for the estimates.

II 2. *Construction models of the tariff*

We usually try to keep the structure of the tariff simple, so that the premium may be easily calculated. In the following we assume that the tariff variables $x_1, \ldots, x_k$ are given. The premium means here the estimate of the expected value of the amount of the claims.

The following models may be used:

1. The additive model

$$P(X) = K \,.\, \sum_{i=1}^{k} f_i(x_i),$$

where K is constant, e.g. the mean of the claim or the amount insured, and the functions $f_i(x_i)$ represent the influence of each tariff variable.

According to this model it is assumed that a change in the value of the tariff variable also causes a certain absolute change of the premium, independent of the values of the other variables, i.e.

$$EY = \sum_{i=1}^{k} \alpha_{ij_i},$$

where α_{ij_i} represents the influence caused by the factor i's class j_i.

This model thus provides that two or more factors do not interact in influencing the amount of the premium. It can be assumed that one risk consists of k small risks which are independent of one another and which are indicated as the tariff variables.

2. The multiplicative model

$$P(X) = K \cdot \prod_{i=1}^{k} f_i(x_i)$$

According to this model a change in one of the tariff variables causes a proportional change in the premium, independent of the value of the other variables, i.e.

$$EY = \prod_{i=1}^{k} \alpha_{ij_i}$$

Increase in the risk caused by the change in value of one variable influences the whole risk in the same proportion.
This can be so for example in fire insurance, where the method of heating greatly affects the probability of fire breaking out. If a change in the heating system reduces the probability of a fire by half and if other factors remain unchanged, the premium must also be cut by half.

3. The mixed model

$$P(X) = \sum_{i=1}^{k} f_{1i}(x_i) + \sum_{i \neq j} f_{2i}(x_i) f_{2j}(x_j) + \ldots + \prod_{i=1}^{k} f_{ki}(x_i)$$

where the functions f_{ji} indicate the influence of the factor i.

4. The general model

$$P(X) = K \cdot f(x_1, \ldots, x_k)$$

In this model each of the tariffs in the risk group is calculated separately. This model is difficult in practice, as the tariff structure becomes complicated, if there are many tariff variables.

To force the tariff into the form of the multiplicative model or the additive model naturally simplifies the procedure considerably. In particular this does not account for interaction between variables. Thus if two or more variables have a strong interaction, they must be combined into one tariff variable only. This of course makes the structure of the tariff more complicated.

II 3. *The tariff construction*

Let us look at the tariff calculation, assuming that the structure is given. Boehm [4] has handled the calculation of functions $f_i(x_i)$ in the additive and multiplicative model. His presentation is based on Almer's [1], Bailey-Simon's [2], Jung's [8] and Mehring's [9] previous research. Seal [10] has also done research on this subject.

Thus we assume that the tariff variables $x_1, \ldots, x_k$ are given. Furthermore we assume that the values of each tariff variable x_i have been classified into a limited number of classes. We denote the risk $X = (i_1, \ldots, i_k)$ if the value of the variable x_j belongs to the class i_j.

For the tariff calculation, data have been collected.

Let us denote as follows:

$y(i_1, \ldots, i_k)$ = observed amount of the claims on the risks $(i_1, \ldots, i_k)$

$n(i_1, \ldots, i_k)$ = observed number of risks $(i_1, \ldots, i_k)$ weighted by the period of insurance. Thus if a year is taken as time period, insurances that have been valid only half a year are counted only for the half.*)

$P(i_1, \ldots, i_k)$ = the premium of the risk $(i_1, \ldots, i_k)$.

Now the observed mean amount of the claims of the risk $(i_1, \ldots, i_k)$ is

$$r(i_1, \ldots, i_k) = \frac{y(i_1, \ldots, i_k)}{n(i_1, \ldots, i_k)}$$

The functions $f_i(x_i)$ or accordingly $f_{ji}(x_i)$ can be found using the following methods:

1. Method of the least squares

   The function f is calculated from the equation

   $$\sum_{i_1, \ldots, i_k} n(i_1, \ldots, i_k) [r(i_1, \ldots, i_k) - P(i_1, \ldots, i_k)]^2 = \min,$$

   where the addition is made over all possible value combinations of the tariff variables.

2. χ^2-minimum method.

   We try to fix the functions f so that the premiums fit together with the data as well as possible according to the X^2-test (comp. Cramer [ ]). The criterion is thus

$$\chi^2 = \sum_{i_1, \ldots, i_k} \frac{n(i_1, \ldots, i_k) [r(i_1, \ldots, i_k) - P(i_1, \ldots, i_k)]^2}{P(i_1, \ldots, i_k)} = \min$$

*) Using this method errors may be made e.g. owing to seasonal variations.

3. Modified χ^2-minimum method.

As the equations arrived at by the above method are mostly difficult to solve, the method can be modified so that instead of the premium the denominator is changed to the observed mean amount of the claims. The criterion is thus

$$R = \sum_{i_1, \ldots, i_k} \frac{n(i_1, \ldots, i_k)\,[r(i_1, \ldots, i_k) - P(i_1, \ldots, i_k)]^2}{r(i_1, \ldots, i_k)} = \min$$

4. Moment method.

This method requires that when each class of each tariff variable is examined separately, the amount of premiums belonging to this class is equal to the corresponding observed amount of the claims. Thus it is indicated

$y_j^i = \sum_{\neq i} y(i_1, \ldots, i_{j-1}, i, i_{j+1}, \ldots, i_k)$, where the addition is made over all class combinations of all variables $x_n (n \neq j)$. Thus we get the total of all the claims on these risks, where the value of the variable x_j belongs to the class i.

The requirement can now be indicated by the following formula:

$$u_j^i = \frac{\sum_{\neq i} n(i_1, \ldots, i_{j-1}, i, i_{j+1}, \ldots, i_k) \cdot P(i_1, \ldots, i_{j-1}, i, i_{j+1}, \ldots, i_k)}{y_j^i} = 1$$

This method gives a solvable group of equations, if the structure model of the tariff is a multiplicative model.

The equations derived by the above methods can be solved iteratively.

When we try to investigate the fitness of the premiums calculated in accordance with the above methods, we can use as one criterion the values of u_j^i. The closer these values are to 1, the better the model is according to this criterion. It can be proved that in the sum model the method of the least squares gives premiums, for which $u_j^i = 1$ for all i, j.

Finally we see by changing the data, complicated models can be modified to simpler ones. For example it is possible to step from the multiplicative model to the sum model just by using the logarithms of the observations. In this way the margin totals y_j^i are lost, and as a consequence the values u_j^i might differ considerably from 1.

III. Selection Problem

III 1. *General features*

In this presentation we have so far concentrated our attention on previously published works.

Let us now proceed to examine ways in which the tariff variables may be selected.

If the structure of the tariff is given and the possible tariff variables are quantitative, the tariff variables can be calculated using the step-wise regression analysis (c.f. [7] Draper-Smith). At the same time it is also possible to calculate the tariff.

The fact that the tariff structure is given, naturally influences the selection of the tariff variables. A more correct procedure is to try to select the tariff variables without any previous assumptions about the tariff structure. Then the aim is to find the best construction model, the parametres of which are decided e.g. according to the methods in chapter II 3.

When selecting the tariff variables in this way, certain criteria have to be created, on the basis of which the selection is made. The selection is made one at a time in the same way as in the step-wise regression analysis.

III 2. *The degree of influence*

Let us assume that the values of each possible tariff variable $x_i(i = 1, \dots n)$ have been classified into a closed number of classes. Each risk is thus indicated by the classes $(i_1, \dots, i_n)$ of possible tariff variables.

Furthermore we assume that the mean of the total amount of the claims of each risk (Y) is proportional to the period of time of insurance t. Now we can denote this as follows:

$$EY = \Pi(i_1, \dots, i_n) \cdot t,$$

where the parametre $\Pi(i_1, \dots, i_n)$ only depends on the values of the possible tariff variables. In practice, however, this assumption is not always accurate. For example concerning motor insurance we can expect that more accidents occur in winter than in summer

https://doi.org/10.1017/S0515036100009338 Published online by Cambridge University Press

and accordingly the premium for winter should be higher than for summer.

As in chapter II 3. the notations used are as follows:

$y(i_1, \ldots, i_n)$ = the observed loss amount on the risks $(i_1, \ldots, i_n)$.
$n(i_1, \ldots, i_n)$ = the observed number of risks $(i_1, \ldots, i_n)$ weighted by the period of insurance.
$y_j^a = \sum_{\neq i} y(i_1, \ldots, i_{j-1}, a, i_{j+1}, \ldots, i_n)$ the total of all the claims on the risks where the value of the variable x_j belongs to class a.
n_j^a = the number of the above risks weighted by the period of insurance.

Accordingly we introduce the following notation:

$y_{jm}^{ab} = \sum_{\neq a, \neq b} y(i_1, \ldots, i_{j-1}, a, i_{j+1}, \ldots, i_{m-1}, b, i_{m+1}, \ldots, n)$ the total of all the claims on the risks, where the value of the variable x_j belongs to class a and the value of the variable x_m to class b.

n_{jm}^{ab} = the number of the above risks weighted by the period of insurance.

Furthermore we define

$$\Pi = \frac{E \sum_{i_1, \ldots, i_n} y(i_1, \ldots, i_n)}{\sum_{i_1, \ldots, i_n} n(i_1, \ldots, i_n)}$$

$$\Pi_j^a = \frac{Ey_j^a}{n_j^a}$$

and

$$\Pi_{jm}^{ab} = \frac{Ey_{jm}^{ab}}{n_{jm}^{ab}}$$

For the selection procedure we make the following definition:
Definition: The variable x_j does not have an influence of the 1st degree on the amount of the claims, if
$\Pi_j^a = \Pi$ in all classes a of the variable x_j.
According to the definition the variable x_j does not thus have an influence of the 1st degree on the amount of the claims, if the expected loss ratio Π_j^a is equal to the expected loss ratio of the total data Π in all classes of the variable x_j.

https://doi.org/10.1017/S0515036100009338 Published online by Cambridge University Press

Definition: The variable x_m does not have an influence of the 2nd degree on the amount of the claims, when the variable x_j is selected, if

$\Pi_{jm}^{ab} = \Pi_j^a$ in all the classes a of the variable x_j and in classes b of the variable x_m.

If the condition of this definition is valid, the value of the variable x_j fully defines the expected loss ratio Π_{jm}^{ab} and the value of the variable x_m does not give any additional information about the amount of the claims.

Proceeding in this way we get the following general definition:

Definition: The variable x_m does not have an influence of the *i:th* degree on the amount of the claims, when the variables $x_{(1)}, \ldots, x_{(i-1)}$ are selected, if

$$\prod_{(1), \ldots, (i-1), m}^{a_{(1)}, \ldots, a_{(m)}} = \prod_{(1), \ldots, (i-1)}^{a_{(1)}, \ldots, a_{i-1}}$$

in all classes

$$a_{(1)}, \ldots, a_{(i-1)}, a_m \text{ of variables } x_{(1)}, \ldots, x_{(i-1)}, x_m.$$

Thus the expected loss ratio in the classes of the variables $x_{(1)}, \ldots, x_{(i-1)}, x_m$ depends only on the variables $x_{(1)}, \ldots, x_{(i-1)}$.

It might be interesting to look at this definition a little closer. Let us assume that the dependence of the expected loss ratio $\Pi(i_1, \ldots, i_n)$ on the tariff variables is known in full and we want to arrange the variables in order according to their influence. We assume that we have selected $(i - 1)$ variables on the basis of the intensity of the influence measured in one way or another. When selecting the following variable we naturally leave out at least those which do not have the influence of the *i:th* degree when the $(i - 1)$ previously selected variables are given. This fact does not of course mean that a variable of this kind should not have an influence on the expected loss ratio, but selecting it does not make the construction of the expected loss ratio more accurate.

Accordingly the fact that some of the variables have an influence of the 1st degree might be due to the alterations caused by other tariff variables on the expected loss ratio and the unequal distribution of the variables in classes of first variables.

Let us take an example:

x_2 / x_1	1		2		3		
	Ey	n	Ey	n	Ey	n	Π_2^j
1	50	50	100	100	150	150	1.0
2	300	150	800	400	100	50	2.0
Π_1^j	1.75		1.80		1.25		

In this example the variable x_1 has an influence of the 1st degree on the expected ratio claim, but not an influence of the 2nd degree, when x_2 is given.

Although the influence defined in this way handles the situation only with respect to already selected variables, this procedure allows a possible method of selecting the most important variables.

III 3. *Selection of the tariff variables*

The selection of the tariff variables is made one by one. For each selection the influence of the previously selected variables on the ratio claim is taken into consideration. The most difficult problem is measuring the significance of the influence of the different variables. One solution is given below.

Selection of the first variable

In the case where the variable x_j does not have the influence of the 1st degree on the expected ratio claim, the equation $\Pi_j^a = \Pi$ is true in all classes a of the variable x_j. On the basis of the data we try to investigate which of the variables differs most from this hypothesis.

Accordingly we set up a test, where the null hypothesis is $H_0 : \Pi_j^a = \Pi$ in all classes a of the variable x_j.

The test variable is:

$$\chi_j^2 = C \sum_a \frac{n_j^a \left[\dfrac{y_j^a}{n_j^a} - \dfrac{\sum\limits_{i_1, \ldots, i_n} y(i_1, \ldots, i_n)}{\sum\limits_{i_1, \ldots, i_n} n(i_1, \ldots, i_n)} \right]^2}{\dfrac{\sum\limits_{i_1, \ldots, i_n} y(i_1, \ldots, i_n)}{\sum\limits_{i_1, \ldots, i_n} n(i_1, \ldots, i_n)}}$$

https://doi.org/10.1017/S0515036100009338 Published online by Cambridge University Press

where the addition is made over the classes of the variable x_j. We have supposed, that the variance of the observed ratio loss is approximately $(1/Cn_j^a)x$ the expected loss ratio (Bailey-Simon [2]).

If the null hypothesis is true, the test variable is approximately χ^2-distributed with $(I_j - 1)$ degrees of freedom, where I_j is the number of classes of the variable x_j.

The measure of the significance of the influence of the variable x_j is the fractile of this test variable.

Consequently, the tariff variable is selected as the first one, for which

$$F(\chi_j^2) = P(\chi^2 \leq \chi^2) = \max_j$$

The variable thus selected is denoted as $x_{(1)}$.

Selection of the 2nd variable

If the variable x_j does not have an influence of the 2nd degree, when $x_{(1)}$ has been selected, $\Pi_{(1)j}^{ab} = \Pi_{(1)}^{a}$ in all classes of the variables $x_{(1)}$ and x_j.

As in selecting the first tariff variable let us set the null hypothesis as follows:

$H_0 : \Pi_{(1)j}^{ab} = \Pi_{(1)}^{a}$ in all the classes of the variables $x_{(1)}$ and x_j.

The test variable is in this case

$$\chi^2 = C \sum_{a,b} \frac{n_{(1)j}^{ab} \left[\frac{y_{(1)j}^{ab}}{n_{(1)j}^{ab}} - \frac{y_{(1)}^{a}}{n_{(1)}^{a}}\right]^2}{\frac{y_{(1)}^{a}}{n_{(1)}^{a}}}$$

If the null hypothesis is true, the test variable is approximately χ^2-distributed with $I_{(1)}\, x(I_k - 1)$ degrees of freedom.

The measure of the significance of the influence of the variable x_j is the fractile of this test variable.

Thus, the tariff variable is selected as the second one, for which

$$\nu(x_j \mid x_{(1)}) = F(\chi_j^2) = P(\chi^2 \leq \chi_j^2) = \max_j,$$

and where accordingly $\nu(x_j|x_{(1)})$ describes the influence of the variable $x_{(j)}$ when $x_{(1)}$ is given.

The selected variable is denoted $x_{(2)}$.

https://doi.org/10.1017/S0515036100009338 Published online by Cambridge University Press

Proceeding in this way the selection of $x_{(3)}$, $x_{(4)}$ etc. is made. Generalizing we get

the selection of the p:th variable

So far we have selected $(p - 1)$ tariff variables. To be able to select the next variable we have to set the null hypothesis for each remaining possible tariff variable x_j:

$$H_0 : \Pi^{ab \dots gh}_{(1)(2)\dots(p-1)j} = \Pi^{ab \dots g}_{(1)(2)\dots(p-1)}$$

in all cells indicated by $x_{(1)}, \dots, x_{(p-1)}, x_j$

The test variable in this case is

$$\chi_j^2 = C \sum_{a,b,\dots,g,h} \frac{n^{a\dots gh}_{(1),\dots,(p-1)j} \left[\dfrac{y^{a\dots gh}_{(1),\dots,(p-1)j}}{n^{a\dots gh}_{(1),\dots,(p-1)j}} - \dfrac{y^{a\dots g}_{(1),\dots,(p-1)}}{n^{a\dots g}_{(1),\dots,(p-1)}} \right]^2}{\dfrac{y^{a\dots g}_{(1),\dots,(p-1)}}{n^{a\dots g}_{(1),\dots,(p-1)}}},$$

where the addition is made over all the cells indicated by the variables $x_{(1)}, \dots, x_{(p-1)}, x_j$.

The test variable is approximately χ^2-distributed, the degrees of freedom being the number of all possible class combinations of the variables $x_{(1)}, \dots, x_{(p-1)}, x_j$ minus the number of the possible class combinations of the variables $x_{(1)}, \dots, x_{(p-1)}$.

The next tariff variable selected is the one for which

$$\nu(x_j \mid x_{(1)}, \dots, x_{(p-1)}) = F(\chi_j^2) = P(\chi^2 \leq \chi_j^2) = \max_j$$

In this way we can continue until the selection must stop on the basis of one criterion or another.

The number of tariff variables can be decided in advance.

Another way of limiting the selection of the tariff variables is to give the constant E in advance so that the selection can be stopped when

$$\max_j F(\chi_j^2) \leq E$$

If many tariff variables have to be selected, the number of the class combinations to be handled will grow considerably.

https://doi.org/10.1017/S0515036100009338 Published online by Cambridge University Press

In this case the techniques of the selection method can be changed so that after three steps of selection it is continued according to the criterion

$$\min \nu(x_j \mid x_{(1)}, x_{(2)}) = \max_j$$

where $x_{(1)}$ and $x_{(2)}$ are already selected tariff variables.

Accordingly the selection can be made by using a quite simple calculation. On the other hand the method takes into consideration the interaction of three variables but not the interaction of four variables.

IV. A Tariff Construction Method

In starting to construct the tariff, the variables have somehow been put into the best order $x_{(1)}, x_{(2)}, \ldots, x_{(k)}$, and the variables that are left, the number of which is accordingly $(n - k)$, can be ignored.

As previously stated, the most common tariff models are multiplicative models or sum models. Deciding which one is more fitting for the problem in question or if perhaps both are unsuitable, is often difficult. Consequently the best procedure seems to be to look at the problem in a more liberal way that allows the existence of both the above models. In this way we get a model which includes parts of both models.

Thus another possible tariff construction model can be offered as follows:

Step 1: The tariff in the class a of the variable $x_{(1)}$ is

$$f_1 = \alpha^a_{(1)} = \frac{y^a_{(1)}}{n^a_{(1)}}$$

Step 2: Let us assume that in the class (a, b) of the variables $x_{(1)}$ and $x_{(2)}$ the tariff is

$$f_2 = \alpha^b_{(2)}\, \alpha^a_{(1)} + \beta^b_{(2)},$$

where $\alpha^a_{(1)}$ has been reached as a result of the first step.

So we have the regression model of one independent variable, where the loss ratio in the class (a, b) is the

https://doi.org/10.1017/S0515036100009338 Published online by Cambridge University Press

dependent variable and the tariff which is a result of the previous step is the independent variable. The parameters $\alpha^b_{(2)}$ and $\beta^b_{(2)}$ can be set for each class of the variable $x_{(2)}$ using the method of the least squares, i.e. the criterion being

$$\sum_a n^{ab}_{(1)\,(2)} \left[\frac{y^{ab}_{(1)\,(2)}}{n^{ab}_{(1)\,(2)}} - (\alpha^b_{(2)}\,\alpha^a_{(1)} + \beta^b_{(2)}) \right] = \min_{\alpha^b_{(2)},\,\beta^b_{(2)}}$$

The addition is made over the classes of the variable $x_{(1)}$.

Accordingly a general step can be formed:

Step ν: Let us assume that the tariff has the model

$$f_\nu = \alpha_{(\nu)}\, f_{\nu-1} + \beta_{(\nu)},$$

where $f_{\nu-1}$ is the tariff set at the previous step for the classes fixed by the variables $x_{(1)}, \ldots, x_{(\nu-1)}$.

For each class of the variable $x_{(\nu)}$ the parametres $\alpha_{(\nu)}$ and $\beta_{(\nu)}$ can be fixed using the method of the least squares:

$$\sum n^{ab\ldots}_{(1)\ldots(\nu)} \left[\frac{y^{ab\ldots}_{(1)\ldots(\nu)}}{n^{ab\ldots}_{(1)\ldots(\nu)}} - (\alpha_{(\nu)}\, f_{\nu-1} + \beta_\nu) \right]^2 = \min$$

The addition is made over the classes of the variables $x_{(1)}, \ldots, x_{(\nu-1)}$.

The advantage of this method is considered to be the fact that it is not very strongly tied to the model, as it combines in itself the multiplicative and sum models. The final tariff looks as follows:

$$\begin{aligned} f_k &= \alpha_{(k)}\alpha_{(k-1)} \cdot \ldots \cdot \alpha_{(1)} + \alpha_{(k)}\alpha_{(k-1)} \cdot \ldots \cdot \alpha_{(3)} + \\ &\quad + \beta_{(2)} + \alpha_{(k)}\alpha_{(k-1)} \cdot \ldots \cdot \alpha_{(4)}\beta_{(3)} + \ldots \\ &= \prod_{(i)=1}^{(k)} \alpha_{(i)} + \sum_{(j)=2}^{(k-1)} \beta_{(j)} \prod_{(j+1)}^{(k)} \alpha_{(i)} + \beta_{(k)}. \end{aligned}$$

If $(k) = 3$, the tariff will be

$$\begin{aligned} f_3 &= \alpha_{(3)}\alpha_{(2)}\alpha_{(1)} + \alpha_{(3)}\beta_{(2)} + \beta_{(3)} \\ &= \alpha_{(3)}[\alpha_{(2)}\alpha_{(1)} + \beta_{(2)}] + \beta_{(3)} \end{aligned}$$

and the tariff is easy to put into tabular form.

If $(k) = 4$, the tariff will be

$$f_4 = \alpha_{(4)}\alpha_{(3)}\alpha_{(2)}\alpha_{(1)} + \alpha_{(4)}\alpha_{(3)}\beta_{(2)} + \alpha_{(4)}\beta_{(3)} + \beta_{(4)},$$

https://doi.org/10.1017/S0515036100009338 Published online by Cambridge University Press

which in this form is rather difficult to put into tabular form, but changing the tariff to the form

$$f_4 = \alpha_{(4)}\alpha_{(3)} \left[(\alpha_{(2)}\alpha_{(1)} + \beta_{(2)}) + \left(\frac{\beta_{(3)}}{\alpha_{(3)}} + \frac{\beta_{(4)}}{\alpha_{(4)}\alpha_{(3)}} \right) \right]$$

we reach the conclusion that the tariff can be calculated using three two-dimensional tables.

References

[1] ALMER: Risk Analysis in theory and practical Statistics. Transactions XV:th International Congress of Actuaries, Vol. II, pp. 314-353. New York 1957.

[2] BAILEY-SIMON: Two Studies in Automobile Insurance Ratemaking, The ASTIN Bulletin, Vol. I, Part IV, 1960, pp. 192-217.

[3] BEARD-PENTIKÄINEN-PESONEN: Risk Theory, Methuen & Co Ltd, 1969.

[4] BOEHM: Das Faktoren- und das Summande-Modell, Blätter der Deutschen Gesellschaft für Versicherungmathematik, Band X, Heft 1, 1971.

[5] BÜHLMANN: Mathematical Methods in Risk Theory, Springer Verlag, 1970.

[6] CRAMER: Mathematical Methods of Statistics, Princeton University Press 1946.

[7] DRAPER-SMITH: Applied Regression Analysis, John Wiley & Sons, Inc. 1966.

[8] JUNG: On Automobile Insurance Ratemaking, The ASTIN Bulletin, Vol. V, Part I, 1968, pp. 41-47.

[9] MEHRING: Ein mathematisches Hilfsmittel für Statistik- und Tariffragen in der Kraftfahrtversicherung, Blätter der Deutschen Gesellschaft für Versicherungmathematik, Band VII, Heft I, 1964.

[10] SEAL: Stochastic Theory of a Risk Business, John Wiley & Sons, Inc., 1969.

Application of the Theory to Third Party Motor Insurance

To investigate the methods of selecting the tariff variables and the construction of tariff that have been presented in chapters 3 and 4, research was made into the motor cars registered in Finland on December 31, 1972. The accidents in which these vehicles were involved during the period July 1, 1972 to June 30, 1973 were gathered and put together with other information available. A better way of handling the matter would have been to investigate first e.g. the total amount of vehicles at December 31, 1971 and to examine the accidents in which they were involved during 1972, but due to the insufficiency of the data in the registers the research had to be completed using the above mentioned method.

We did not know all the claims, which had been paid and should be paid. Nevertheless we can expect, that the selection of tariff variables can be correctly done. The tariff based on these data is not precise, but we can however get some information concerning relations between different tariff classes.

The following facts were collected concerning each vehicle during the period July 1, 1972, - June 30, 1973:

Variable	Data
x_1	Number of accidents
x_2	Total amount of claims
x_3	Insurance period
x_4	Description of holder — private — other
x_5	Language of holder — Finnish — Swedish
x_6	Is the holder the owner
x_7	Domicile of the vehicle
x_8	The economic area of the vehicle
x_9	The type of commune of the vehicle's domicile — town — country town — rural commune
x_{10}	The tariff area according to tariff classification to day (four districts)
x_{11}	The economic geographical classification of the commune, in which the vehicle is kept (7 classes)
x_{12}	Country of origin
x_{13}	Age of the vehicle
x_{14}	Motive power — petrol — other
x_{15}	Usage
x_{16}	Number of passenger places
x_{17}	Time of possession by latest owner
x_{18}	Bonus-class according to current bonus system

x_{19} Premium class of motor insurance defined for each make of car
x_{20} Front wheel drive?
x_{21} Engine stroke capacity cc
x_{22} Weight of vehicle

A group of vehicles to be analysed was selected such that all vehicles which had had an accident were counted, and of the remainder that had had no accidents, only every 10th was counted. The time for which the latter vehicles had been insured was multiplied by 10. In this way the analysis takes into account about 100,000 vehicles.

Selection Problem

The method for selecting tariff variables presented in chapter 3 was applied to these data. The calculations were made for both the total amounts claimed and the number of claimed. The following results were reached:

Selection of 1st tariff variable

	y = total amount of claims		y = number of claims	
Variable	χ_j^2	Number of classes	χ_j^2	Number of classes
x_4	10610	2	1088	2
x_5	410	2	88	2
x_6	10650	2	1708	2
x_7	16675	12	1609	12
x_8	14170	15	1369	15
x_9	7900	3	1333	3
x_{10}	14170	4	1283	4
x_{11}	16325	7	2071	7
x_{12}	10655	12	707	12
x_{13}	1330	16	72	16
x_{14}	5400	2	278	2
x_{15}	12570	4	712	4
x_{16}	12720	7	775	7
x_{17}	12825	16	1649	16
x_{18}	39245	18	6077	18
x_{19}	28385	14	1775	14
x_{20}	15	2	3	2
x_{21}	27415	11	1614	11
x_{22}	22465	12	1498	12

The first selected variable was $x_{(1)} = x_{18}$ = bonus-class according to the current bonus system.

Selection of 2nd tariff variable

Variable	y = total amount of claims χ_j^2	Number of classes	y = number of claims χ_j^2	Number of classes
x_4	10375	36	1006	36
x_5	825	36	85	36
x_6	5850	36	873	36
x_7	21290	216	1766	216
x_8	21060	270	1816	270
x_9	8435	54	1324	54
x_{10}	15705	72	1309	72
x_{11}	18630	126	2113	126
x_{12}	19930	216	1619	216
x_{13}	10710	288	762	288
x_{14}	7795	36	445	36
x_{15}	25270	72	1739	72
x_{16}	31400	126	1653	126
x_{17}	13030	288	1228	288
x_{18}	0		0	
x_{19}	46640	252	3028	252
x_{20}	895	36	88	36
x_{21}	45015	198	2722	198
x_{22}	37870	216	2530	216

The second selected variable was $x_{(2)} = x_{19}$ = premium class of motor insurance.

Selection of 3rd tariff variable

Once the data was classified according to three variables, we had many classes, in which there were only a few vehicles. Therefore we did not take those classes into account, for which the expected number of claims was less than 5 or the expected total amount claimed was less than 5000 Fmk.

So we selected $x_{(3)} = x_{11}$ = the economic geographical classification of the vehicle's domicile.

Selection of 4th tariff variable

The selected variable was $x_{(4)} = x_{13}$ = age of the vehicle.

https://doi.org/10.1017/S0515036100009338 Published online by Cambridge University Press

Selection of 5th tariff variable

We reached the following results:

	y = total amount of claims		y = number of claims	
Variable	χ_j^2	Number of classes	χ_j^2	Number of classes
x_4	4700	2669	107	1219
x_5	4550	2654	90	1208
x_6	31645	2596	527	1122
x_7	34015	1567	292	375
x_8	23775	1406	113	386
x_9	9150	2568	144	1139
x_{10}	29915	2549	459	970
x_{11}	0		0	
x_{12}	48510	2225	496	632
x_{13}	0		0	
x_{14}	2560	2702	21	1234
x_{15}	2540	2707	37	1232
x_{16}	28995	2518	391	1085
x_{17}	40810	2005	403	730
x_{18}	0		0	
x_{19}	0		0	
x_{20}	14200	2653	276	1118
x_{21}	47545	2237	441	664
x_{22}	45385	2094	419	568

So, we made no further selections after the 4th variable, because the χ^2-values of the number of claims were so small.

Construction of the Tariff

We constructed the tariff according to the method in chapter IV. The tariff variables were the selected variables $x_{(1)}$, $x_{(2)}$, $x_{(3)}$ and $x_{(4)}$. As we remarked earlier, the tariff is not exact, because our data were incomplete, but we can get some information concerning the relations between different premium classes.

The first selected variable was $x_{(1)} = x_{18}$ = bonus class according to the current bonus-system. We reached the following results:

Step 1

a	$\alpha^a{}_{(1)}$	the premium % of the basic premium
1	107,7	100
2	91,3	80

https://doi.org/10.1017/S0515036100009338 Published online by Cambridge University Press

a	$\alpha^a{}_{(1)}$	the premium % of the basic premium
3	91,8	70
4	82,2	60
5	75,5	60
6	72,1	50
7	67,0	50
8	67,5	50
9	60,6	50
10	59,4	40
11	61,5	40
12	55,4	40
13	82,5	40
14	459,0	150
15	250,3	130
16	145,5	120
17	169,0	110
18	77,2	100

Step 2

The second selected variable was $x_{(2)} = x_{19}$ = premium class of motor insurance defined for each make of car.

All makes of car are classified into 14 classes. The lower the premium class, the lower the premium. Generally the smallest cars are in the first premium class and the biggest cars in the fourteenth premium class.

We reached the following results:

b	$\alpha^b{}_{(2)}$	$\beta^b{}_{(2)}$	r^2	t
1	0,5039	3,3	0,679	5,82
2	0,9347	—19,3	0,897	11,78
3	0,9419	—10,1	0,832	8,91
4	0,8471	5,0	0,885	11,10
5	1,0112	—0,7	0,904	12,29
6	1,1647	—10,6	0,944	16,48
7	1,2229	7,3	0,747	6,87
8	1,4885	—7,9	0,813	8,34
9	2,1469	—30,1	0,910	12,70
10	1,2023	37,0	0,532	4,26
11	1,9831	8,8	0,710	6,26
12	3,9761	—112,9	0,743	6,80
13	0,2415	153,1	0,005	0,28
14	2,9183	—29,3	0,125	1,51

https://doi.org/10.1017/S0515036100009338 Published online by Cambridge University Press

Here $r^2 = \frac{\text{sum of squares due to regression}}{\text{sum of squares about mean}}$ = variation explained and t = the value of test variable when we test the hypothesis that the regression coefficient $\beta^b_{(2)} = 0$.

So if a car is in bonus class 4 and in premium class 6 in motor insurance and we do not take into consideration other tariff variables, the premium is

$$P = 1{,}1647 \cdot 82{,}2 - 10{,}6 = 85{,}1$$

Step 3

The third selected variable was $x_{(3)} = x_{11}$ = the economic geographical class of the vehicle's domicile. This classification is one of four possible ways of classifying communes. It is based on a study, in which all Finnish communes are classified according to their services, distances from other centres etc. Helsinki belongs to the seventh class and the smallest and the most out-of-the-way communes belong to the first class.

We reached the following results:

c	$\alpha^c{}_{(3)}$	$\beta^c{}_{(3)}$	r^2	t
1	0,7410	9,0	0,229	8,62
2	0,7343	8,5	0,274	9,71
3	0,7977	2,4	0,218	8,35
4	0,9565	2,6	0,232	8,68
5	0,7223	22,5	0,044	3,40
6	0,8991	15,5	0,337	11,28
7	1,4905	—10,8	0,560	17,84

Step 4

The fourth selected variable was $x_{(4)}$ = age of the vehicle. All cars more than 16 years old are in the sixteenth class. The results were as follows:

d	$\alpha^d{}_{(4)}$	$\beta^d{}_{(4)}$	r^2	t
1	1,1969	13,5	0,072	11,6
2	0,8263	18,7	0,024	6,5
3	0,9681	2,9	0,073	11,8
4	0,8753	10,5	0,037	8,3
5	1,1851	—17,2	0,024	6,6
6	1,1479	—7,4	0,055	10,1

d	$\alpha^d{}_{(4)}$	$\beta^d{}_{(4)}$	r^2	t
7	1,1478	—11,3	0,072	11,7
8	1,0467	—6,7	0,056	10,3
9	1,1064	—10,6	0,059	10,5
10	1,0488	—6,7	0,023	6,5
11	0,6572	20,6	0,014	4,9
12	0,7001	20,6	0,009	4,0
13	0,3454	40,2	0,002	1,7
14	0,7470	—8,6	0,011	4,5
15	0,2864	46,4	0,001	1,6
16	0,0611	57,3	0,000	0,5

We realize, that the r^2-values are very low, but we can reject almost all hypothesis $\beta^d_{(4)} = 0$.

For example if a car belongs to the seventh bonus-class, to the second premium class of motor insurance, to the fourth commune class and its age is 5 years, the premium is

$$P = 1{,}1851 \cdot [0{,}9565 \cdot (0{,}9347 \cdot 67{,}0 - 19{,}3) + 2{,}6] - 17{,}2 = 35{,}0$$

We have also constructed the tariff in such a way, that the tariff variables were in reverse order, i.e. We calculated first the α^a for variable $x_{(4)}$, then α^b and β^b for variable $x_{(3)}$ etc. Then the r^2-values in steps 2 and 3 were much less than in the previous case but after the fourth step the r^2-values were almost as great as before.

https://doi.org/10.1017/S0515036100009338 Published online by Cambridge University Press