NMFM402 – Mathematics of Non-Life Insurance 2

Simple tariffication methods Practical 1

Exercise 1:
We have the following tariff factors:
$\chi_{1,1} \dots$ passenger car,
$\chi_{1,2} \dots$ delivery van,
$\chi_{1,3} \dots$ truck,
$\chi_{2,1} \dots$ 21-30y,
$\chi_{2,2} \dots$ 31-40y,
$\chi_{2,3} \dots$ 41-50y,
$\chi_{2,4} \dots$ 51-60y, and
$\mu \dots$the average claim per policy.

The multiplicative tariff structure means this model (recall that $v_{i,j} = 1$):

$$\mathbb{E}S_{i,j} = \mu\, \chi_{1,i}\, \chi_{2,j}$$

To avoid over-parametrization (over-identification), we set $\mu = 1$ and $\chi_{1,1} = 1$.

(a) The function to be minimized is

$$X^2 = \sum_{j=1}^{4} \frac{(S_{1,j} - \chi_{2,j})^2}{\chi_{2,j}} + \sum_{i=2}^{3}\sum_{j=1}^{4} \frac{(S_{i,j} - \chi_{1,i}\, \chi_{2,j})^2}{\chi_{1,i}\, \chi_{2,j}}$$

In particular, we get:

$$\begin{aligned} X^2 &= \frac{(2000 - \chi_{2,1})^2}{\chi_{2,1}} + \frac{(1800 - \chi_{2,2})^2}{\chi_{2,2}} + \dots + \frac{(1600 - \chi_{2,4})^2}{\chi_{2,4}} + \\ &= \frac{(2200 - \chi_{1,2}\, \chi_{2,1})^2}{\chi_{1,2}\, \chi_{2,1}} + \frac{(1600 - \chi_{1,2}\, \chi_{2,2})^2}{\chi_{1,2}\, \chi_{2,2}} + \dots + \frac{(1600 - \chi_{1,3}\, \chi_{2,4})^2}{\chi_{1,3}\, \chi_{2,4}}. \end{aligned} \tag{1}$$

The resulting tariff factors and tariffs are summarized in the table below (recall that we set $\mu = 1$ and $\chi_{1,1} = 1$):

	21-30y	31-40y	41-50y	51-60y	$\hat{\chi}_{1,i}$
passenger car	2176	1751	1491	1493	1
delivery van	2079	1674	1425	1427	0.96
truck	2456	1977	1684	1686	1.13
$\hat{\chi}_{2,j}$	2176	1751	1491	1493	

(b) The system of equations is:

$$\sum_{j=1}^{4} \chi_{1,i}\, \chi_{2,j} = \sum_{j=1}^{4} S_{i,j} \quad i = 2, 3$$

$$\chi_{2,j} + \sum_{i=2}^{3} \chi_{1,i}\, \chi_{2,j} = \sum_{i=1}^{3} S_{i,j} \quad j = 1, .., 4$$

In particular, we get:

$$\sum_{j=1}^{4} \chi_{1,2}\ \chi_{2,j} = 6600$$

$$\sum_{j=1}^{4} \chi_{1,3}\ \chi_{2,j} = 7800$$

$$\chi_{2,1} + \sum_{i=2}^{3} \chi_{1,i}\ \chi_{2,1} = 6700$$

$$\ldots$$

$$\chi_{2,4} + \sum_{i=2}^{3} \chi_{1,i}\ \chi_{2,4} = 4600$$

Note that the equation for the sum in the first row was dropped, because we set $\chi_{1,i} = 1$. The first row sum must be satisfied, if all the equations above hold. The solution and resulting tariffs are summarized below:

	21-30y	31-40y	41-50y	51-60y	$\hat{\chi}_{1,i}$
passenger car	2170	1749	1490	1490	1
delivery van	2076	1673	1425	1425	0.96
truck	2454	1977	1685	1685	1.13
$\hat{\chi}_{2,j}$	2170	1749	1490	1490	

(c) Recall that we assume

$$X_{i,j} = \log(S_{i,j}) \sim \mathcal{N}(\beta_0 + \beta_{1,i} + \beta_{2,j}, \sigma^2). \tag{2}$$

This implies the desired multiplicative structure for the expected values

$$\mathbb{E}S_{i,j} = \mathrm{e}^{\beta_0 + \frac{\sigma^2}{2}}\ \mathrm{e}^{\beta_{1,i}}\ \mathrm{e}^{\beta_{2,j}}$$

so that we can identify

$$\chi_{1,i} = \mathrm{e}^{\beta_{1,i}}, \quad \chi_{2,j} = \mathrm{e}^{\beta_{2,j}}, \quad \mu = \mathrm{e}^{\beta_0 + \frac{\sigma^2}{2}} \approx \mathrm{e}^{\beta_0}.$$

If we set $\beta_{1,1} = 0$ and $\beta_{2,1} = 0$ (to avoid over-parametrization), we can rewrite model (2) in the matrix form

$$X \sim \mathcal{N}(Z\beta, \sigma^2 \mathbb{I}),$$

with parameter vector

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_{1,2} \\ \beta_{1,3} \\ \beta_{2,2} \\ \beta_{2,3} \\ \beta_{2,4} \end{pmatrix}$$

and design matrix

$$Z = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}$$

The MLE for β in this setting has the simple closed formula:

$$\hat{\beta} = (Z^T Z)^{-1} Z^T X$$

The resulting parameters, tariff factors and tariffs are summarized in the table below:

$\hat{\beta}_0 = 7.688$	21-30y	31-40y	41-50y	51-60y	$\hat{\chi}_{1,i} = e^{\hat{\beta}_{1,i}}$	$\hat{\beta}_{1,i}$
passenger car	2182	1759	1500	1501	1	0
delivery van	2063	1663	1417	1419	0.946	-0.056
truck	2444	1970	1680	1682	1.120	0.113
$\hat{\chi}_{2,j} = e^{\hat{\beta}_{2,j}}$	1	0.806	0.687	0.688		
$\hat{\beta}_{2,j}$	0	-0.216	-0.375	-0.374		

You may want to check the following R code for log-linear model (taken from here):

```
###  Load  the  observed  claim  amounts  into a matrix
S <- matrix(c(2000 ,2200 ,2500 ,1800 ,1600 ,2000 ,1500 ,1400 ,1700 ,1600 ,1400 ,1600) ,  nrow =3)

###  Define  the  design  matrix Z
Z <- matrix(c(rep(1,12),rep(0,4),rep(1,4),rep(0,12),rep(1,4),rep(c(0,1,0,0),3),
             repscriptsize(c(0,0,1,0),3),rep(c(0,0,0,1),3)),  nrow =12)

###  Store  design  matrix Z and  log(S_{i,j}) in one  dataset
data  <- as.data.frame(cbind(Z[,-1],matrix(log(t(S)),nrow =12)))
colnames(data) <- c("van", "truck", "X31_40y", "X41_50y", "X51_60y", "observation ")

###  Apply  the  regression  model
linear.model1  <- lm(formula = observation ~ van+truck+X31_40y+X41_50y+X51_60y , data=data)
summary(linear.model1)

###  Fitted  values
matrix(exp(fitted(linear.model1)), byrow=TRUE , nrow =3)

### We can  also  get  the  parameters  by  applying  the closed formula
solve(t(Z)%*%Z) %*% t(Z) %*%  matrix(log(t(S)), nrow =12)

### We can  also  use R directly  on the  data (it  finds  the  design  matrix  internally)
car  <- c(" passenger  car", "van", "truck")
age  <- c(" X21_30y", "X31_40y", "X41_50y", "X51_60y ")
dat  <- expand.grid(car , age)
colnames(dat) <- c("car","age")
dat$observation  <- as.vector(log(S))
linear.model1.direct  <- lm(formula = observation ~ car+age , data=dat)
summary(linear.model1.direct
```

(d) Let us compare the tariffs determined by the three methods

Bailey & Simon:

	21-30y	31-40y	41-50y	51-60y	$\hat{\chi}_{1,i}$
passenger car	2176	1751	1491	1493	1
delivery van	2079	1674	1425	1427	0.96
truck	2456	1977	1684	1686	1.13
$\hat{\chi}_{2,j}$	2176	1751	1491	1493	

Total marginal sums:

	21-30y	31-40y	41-50y	51-60y	$\hat{\chi}_{1,i}$
passenger car	2170	1749	1490	1490	1
delivery van	2076	1673	1425	1425	0.96
truck	2454	1977	1685	1685	1.13
$\hat{\chi}_{2,j}$	2170	1749	1490	1490	

log-linear model (reparametrized so that $\mu = 1$ and $\chi_{1,1} = 1$):

	21-30y	31-40y	41-50y	51-60y	$\hat{\chi}_{1,i}$
passenger car	2182	1759	1500	1501	1
delivery van	2063	1663	1417	1419	0.946
truck	2444	1970	1680	1682	1.120
$\hat{\chi}_{2,j}$	2182	1759	1499	1501	

All three methods provide similar tariffs. Tariffs by the method Bailey & Simon are slightly overestimated (which corresponds to the general theory). Log-linear model generates somewhat higher tariffs for passenger cars and lower for the other vehicles (compared to the other two methods). For all three methords, the least risky car is delivery van and the most risky age group is 21-30y (this age group has significantly higher tariffs compared to the other age groups).