

Case 2  
MT7028 - Non-life Insurance Pricing

Majken Gunnarsson - [majken.gunnarsson@gmail.com](mailto:majken.gunnarsson@gmail.com)  
Oscar Potts - [oscar.n.potts@gmail.com](mailto:oscar.n.potts@gmail.com)

December 11, 2024

## 1 Resulterande faktor specifikation

En beskrivning av klasserna som används i de resulterande tarifferna i avsnitten nedan ges i tabell 1. Standardkategorin är (3,1,3,3,2) vilket är landsbygdsområden med medelstora fordon i åldern 5-9 år som körs av personer som är över 40 år gamla. Observera att bonusklassen togs bort från denna modell då ingen signifikans hittades. Tariffen med de preliminära relativa faktorerna innehåller zon, MC klass och fordonsålder medan den nya tariffen även innehåller ägarålder samt en kombinationsvariabel mellan ägarålder och kön.

Faktor	Klass	Klass beskrivning
Zon	1	Centrala och semicentrala delar av Sveriges tre största städer
	2	Förorter och medelstora städer.
	3	Resten av Sverige
MC klass	1	Andra EV-ratios
	2	EV ratio 20-24
Fordon ålder	1	Under 1 år
	2	1-5 år
	3	5-9 år
	4	9-13 år
	5	Över 13 år
Ägarålder & Kön	1	Man mellan 0-25 år
	2	Man mellan 25-30 år
	3	Andra förar kombinationer.
Ägarålder	1	Förare mellan 0-40 år
	2	Förare över 40 år

Table 1: Beskrivning av klasserna som används i de slutgiltiga tarifferna.

## 2 Preliminära Tariff Relativiteter

I denna uppgift har vi utgått från de klasser och aggregerad data som togs fram i Case 1. Uppgiften var därefter att utveckla en tariff där det är möjligt att exkludera/gruppera klasser som inte är signifikanta för tariffen. Till skillnad från Case 1 så använder vi en quasi-Poisson-fördelning i GLM-modellerna istället för en Poisson-fördelning. Den används för att tillåta eventuell överdispersion, vilket sker om variansen är större än medelvärdet. I denna preliminära tariff fick vi  $\rho = 1.06$  vilket är nära 1 medan i den slutgiltiga fick vi  $\rho = 1.474$  vilket var mer signifikant.

Den resulterande tariffen togs fram genom att iterera igenom olika alternativ på tariffer där de klasser som ej är signifikanta (har ett p-värde över 0,05 i GLM modellen) grupperades

ihop eller exkluderades från modellen. Den slutgiltiga modellen innehåller inte bonusklass vilket var inkluderat i tariffen i Case 1. Andra ändringar som skett är att zon har gått från 7 klasser till 3 där de mindre städerna, norra Sverige samt Gotland grupperats ihop. MC klass har gått från 7 klasser till två där alla fordon som inte har en EV ratio mellan 20-24 har grupperats ihop. Fordonålder har fått fler klasser och det är framförallt de äldre åldrarna som blivit uppdelade i fler klasser. Tariffen samt de relativa frekvensfaktorerna och relativa skadekostnadsfaktorerna finns i figur 5. Vad vi kan notera i denna tariff jämfört med den som togs fram i Case 1 är:

- De klasser med högst relativ risk premie har fått en något lägre relativ faktor. En anledning till detta kan vara att standardkategorin har ändrats när klasser grupperats ihop eller delats upp mer.
- Konfidensintervallen har blivit mindre. Att vi ändrade från Poisson till Quasi-poisson i GLM modellen borde bidra till att konfidensintervallen blir större eftersom att vi tar hänsyn till överdispersion. Att konfidensintervallet istället blivit mindre kan bero på att vi bara inkluderar de klasser som är signifikanta i den nya tariffen, vilket minskar brus.

I figur 1 och 2 kan man se hur signifikanta variablerna är för skadekostnad respektive frekvens i GLM-modellerna. Vi kan se att för frekvensen så har alla variabler ett p-värde under 0,001. För skadekostnad har fordonålder ett p-värde under 0,001 medan zon har ett p-värde under 0,01 och MC klass har ett p-värde under 0,05. MC-klassens signifikans för skadekostnad är alltså den variabel som är minst signifikant, men då de valda grupperingarna inom MC-klass är väldigt signifikanta för skadefrekvensen så är vi nöjda med de valda grupperingarna.

```
Model:
skadkost ~ zon + mcklass + fordald_grupp
              Df Deviance   AIC scaled dev.   Pr(>Chi)
<none>                1128.5 14556
zon                2    1147.2 14562    10.533 0.005161 **
mcklass            1    1134.7 14557     3.518 0.060710 .
fordald_grupp      4    1293.4 14641    93.017 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 1: Signifikans för faktorer i skadekostnad modellen för tariffen med preliminära relativiteter.

```
Model:
mopfreq ~ zon + mcklass + fordald_grupp
              Df Deviance scaled dev.   Pr(>Chi)
<none>                6130.3
zon                2    6345.9    95.224 < 2.2e-16 ***
mcklass            1    6206.2    33.508 7.098e-09 ***
fordald_grupp      4    6329.7    88.101 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 2: Signifikans för faktorer i skadefrekvens modellen för tariffen med preliminära relativiteter.

### 3 Nya Tariff Relativiteter

I denna del har vi undersökt ifall det finns ytterligare variabler som är signifikanta och bör ingå i tariffen. Variablerna vi undersökt är kön, ägarålder samt en kombinerad faktor mellan kön och ägarålder. De slutgiltiga klasserna togs fram genom att iterera och kolla på

p-värdet för klassen i GLM-modellerna. Slutsatserna kring signifikansen för dessa variabler är:

- Det finns en signifikant skillnad på förare under 40 år och förare över 40 år.
- Vi hittade ingen signifikant skillnad mellan kön.
- Det är en signifikant skillnad för män under 25 år, män mellan 25-30 år samt kvinnor och män över 30 år.

Den resulterande tariffen som innehåller även dessa variabler kan ses i tabell 6. Det man kan notera från dessa nya variabler är att förare under 40 år har en dubbelt så stor relativ faktor jämfört med de över 40 år. Ifall föraren även är man och under 30 år så ökar den relativa faktorn ungefär 3 gånger. Det är mestadels frekvensen som drar upp den relativa faktorn för män under 30 år, men för män mellan 25-30 år är även den relativa skadekostnaden större än standardkategorin (vilket är kvinnliga förare och män över 30 år).

I figur 3 och 4 kan vi se signifikansen för variablerna som används i GLM-modellerna. Precis som för tariffen utan förarålder och kön så är alla variabler som används väldigt signifikanta i GLM-modellen som beräknar frekvensen (alla variabler har ett p-värde under 0,001). I skadekostnadsmodellen kan vi se att MC-klass inte längre är signifikant och inte heller ägarålder är signifikant för skadekostnad. Den kombinerade variabeln kön och ägarålder är däremot signifikant för skadekostnad.

```
Model:
skadkost ~ zon + mcklass + fordald_grupp + agarald_grupp + kon_agarald
Df Deviance AIC scaled dev. Pr(>Chi)
<none> 1108.3 14547
zon 2 1128.6 14554 11.171 0.003751 **
mcklass 1 1112.2 14547 2.122 0.145239
fordald_grupp 4 1266.4 14626 87.042 < 2.2e-16 ***
agarald_grupp 1 1109.9 14546 0.859 0.354029
kon_agarald 2 1125.4 14552 9.418 0.009015 **
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 3: Signifikans för faktorer i skadekostnad modellen för tariffen med nya relativiteter.

```
Model:
mopfreq ~ zon + mcklass + fordald_grupp + agarald_grupp + kon_agarald
Df Deviance scaled dev. Pr(>Chi)
<none> 5758.5
zon 2 5953.0 113.696 < 2.2e-16 ***
mcklass 1 5818.5 35.078 3.167e-09 ***
fordald_grupp 4 5933.9 102.497 < 2.2e-16 ***
agarald_grupp 1 5779.9 12.509 0.000405 ***
kon_agarald 2 5867.3 63.576 1.566e-14 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 4: Signifikans för faktorer i skadefrekvens modellen för tariffen med nya relativiteter.

## A Bilagor

	Duration	Antal_skador	Relativ_frekvens	Frekvens_2.5_CI	Frekvens_97.5_CI	Relativ_skadestnad	Skadestnad_2.5_CI	Skadestnad_97.5_CI	Relativ_pure_premium
fordald_grupp1	4955.403	125	1.8588510	1.2750405	2.7105965	1.5920189	1.1287403	2.2471164	2.9593259
fordald_grupp2	12273.052	188	1.1336305	0.8062948	1.6065452	1.3751430	1.0033109	1.8736752	1.5589040
fordald_grupp3	8569.066	124	1.0000000	N/A	N/A	1.0000000	N/A	N/A	1.0000000
fordald_grupp4	9873.093	125	0.8527952	0.5857171	1.2418881	0.4458143	0.3176169	0.6256411	0.3801883
fordald_grupp5	29566.197	131	0.3653506	0.2517541	0.5308789	0.4546791	0.3236645	0.6378429	0.1661173
mcklass1	56356.677	519	1.0000000	N/A	N/A	1.0000000	N/A	N/A	1.0000000
mcklass2	8880.134	174	2.3288779	1.7724306	3.0282367	1.2682327	0.9895339	1.6393274	2.9535591
zon1	6205.310	182	4.0198500	3.0462632	5.2658855	1.3125806	1.0194360	1.6995348	5.2763771
zon2	10103.090	166	2.2358568	1.6827849	2.9439521	1.4873542	1.1561002	1.9270664	3.3255110
zon3	48928.411	345	1.0000000	N/A	N/A	1.0000000	N/A	N/A	1.0000000

Figure 5: De relativa faktorerna för skadefrekvens och skadestnad samt deras 2,5% och 97,5% konfidensintervall och även den relativa risk premien för den slutgiltiga tariff som innehåller de preliminära relativiteterna.

	Duration	Antal_skador	Relativ_frekvens	Frekvens_2.5_CI	Frekvens_97.5_CI	Relativ_skadestnad	Skadestnad_2.5_CI	Skadestnad_97.5_CI	Relativ_pure_premium
agardald_grupp1	19364.918	432	1.7698169	1.2985006	2.3806135	1.1602453	0.8500404	1.6064477	2.0534218
agardald_grupp2	45871.893	261	1.0000000	N/A	N/A	1.0000000	N/A	N/A	1.0000000
fordald_grupp1	4955.403	125	2.5035776	1.7958242	3.4913009	1.5217702	1.0704478	2.1647879	3.8098698
fordald_grupp2	12273.052	188	1.3782353	1.0227822	1.8682599	1.3709023	0.9944871	1.8786711	1.8894260
fordald_grupp3	8569.066	124	1.0000000	N/A	N/A	1.0000000	N/A	N/A	1.0000000
fordald_grupp4	9873.093	125	0.9339534	0.6739020	1.2945338	0.4375581	0.3090559	0.6193895	0.4086589
fordald_grupp5	29566.197	131	0.4984875	0.3588187	0.6932187	0.4533624	0.3177897	0.6460968	0.2259955
kon_agardald1	4943.474	185	3.4326060	2.4985008	4.7617275	0.7906284	0.5594682	1.1058793	2.7139158
kon_agardald2	4639.551	146	2.4090963	1.7288253	3.3793229	1.2675114	0.8848636	1.8030178	3.0535571
kon_agardald3	55653.786	362	1.0000000	N/A	N/A	1.0000000	N/A	N/A	1.0000000
mcklass1	56356.677	519	1.0000000	N/A	N/A	1.0000000	N/A	N/A	1.0000000
mcklass2	8880.134	174	2.1192890	1.6693116	2.6697409	1.2078308	0.9377386	1.5687319	2.5597426
zon1	6205.310	182	3.7154127	2.9126956	4.7137668	1.2809378	0.9898077	1.6668445	4.7592125
zon2	10103.090	166	2.2302442	1.7424193	2.8352431	1.5367205	1.1861402	2.0052839	3.4272619
zon3	48928.411	345	1.0000000	N/A	N/A	1.0000000	N/A	N/A	1.0000000

Figure 6: De relativa faktorerna för skadefrekvens och skadestnad samt deras 2,5% och 97,5% konfidensintervall och även den relativa risk premien för den slutgiltiga tariffen med nya relativiteter.