Report I
MT7028 - Non-life Insurance Pricing

Majken Gunnarsson - majken.gunnarsson@gmail.com
Oscar Potts - oscar.n.potts@gmail.com

December 11, 2024

In this lab we have fitted a severity-frequency model according to a set of given rating factors. We aggregated the data in tariff cells and made regression on number of claims and the average claim size per tariff cell. We observe over-dispersion for a poisson frequency fit and through Akaike information criterion we decided on a lognormal severity model. The resulting factors can be seen in figure 1 where we have joint the resulting fits into a single table. The baseline frequency is 0.00224 claims per full year and an baseline claim size of 10106 SEK.

1. *How data is distributed over the rating factor classes as an indication of precision of the estimated relativities.*

   - The baseline category (4,3,3,3) is not the biggest category with only 9 claims and a duration of 2432 years. One may want to change the base category.
   - Bonus category 1 is the biggest bucket of data, note that this is the amount of portfolio with bonus category 1 which overlaps to other dissections.
   - Low amount of duration leads to high variance in estimates, for instance is zone 7 (Gotland) very low amounts of data with no clear significance, also the single claim observed on Gotland makes the relative factor near 0. Similar observations may be said for mcklass7 (EV ratio 25+)
   - Since factors are shown to have significant large factors. Specifically inner city zones (zon1), new vehicles (vehage_ cat1) and EV ratios of 20-24 (mcklass6).
   - Many relative factors are shown to be non significant, specifically many zone variables (zon3,zon5,zon6,zon7), class variables (mcklass1,mcklass2,mcklass4,mcklass5,mcklass7) and bonus class 3-4. One should consider if these categories could be grouped to achieve better estimates.

2. *How the results compare with the current tariff. Is there a need to change the tariff?*

   As we likely have have more data to support the pricing, there is a need to change the tariff. But more concisely we have that:

   - Only 14 out of 16 factors are inside the confidence interval of the new estimates, meaning that we can not dismiss current tariff factors statistically.
   - For 2 out 16 factors we can dismiss the current tariff relative factors which specifically is the factor related to vehicle age. These factors are outside the lower interval compared the confidence interval.
   - For 10 out the remaining 14 categories we can not dismiss the factor to actually be 1, meaning that the tariff factor could in fact be be 1.

   There is a likely need to update the tariff because of this inaccuracies. In doing so one should maybe rethink the grouping and fit homogeneous groups to attain better support for current factors.

3. *How the claim frequency and claim severity relativities differ.*

   For the frequency have significant 1 differed factors for 13 out 16 factors in in which we specifically have zone 5,6,7 are found non-significant, which as previously argued would benefit a new type of grouping. For severity it is the opposite, and we have no significant factors in 3 out of 16 factors in which specifically vehicle ages and Gotland (zon7). One may argue that zon7 is an outlier with a single very small claim, which may therefore be dismissed.

   Comparing frequency and severity we find from our estimates that.

   - mcklass 1,2,3 have a positive relative frequency factor and lower relative severity factor. Likely as the smaller vehicles are cheaper to repair but tend to be in more accidents.
   - mcklass 5,6.7 have high frequency and positive severity factors, like being that large vehicles are more likely to be in accidents and also be more expensive.
   - Inner city zones have a high frequency of 5.13 and 2.73 and also higher severity, which may be because of higher inner city tendencies for accidents and more expensive vehicles.
   - Newer vehicles tend to have both increased frequency and also severity, meaning they have a higher tendency for accidents but also be more expensive to repair.
   - Bonus class 1,2 have a higher frequency and severity than bonus class 3. Meaning if you have not had claims you are less likely to have claims.

4. *The ratio between the highest pure premium and the lowest.*

   Based on the estimates given, the highest relative factor one could attain is the combination $(1, 6, 1, 1)$ which is a large (EV 20-24) a new inner city vehicle with 1-2 claim free years which would correspond to a factor $f_{high} = 443$ which given the previous base-level of 185 gives the maximum pure premium of $185 f_{high} = 82\ 112$ SEK.

   Compare this to the combination of lowest relative factors which is $(7, 1, 3, 3)$ which is an older, small vehicle (EV ratio less than 5) driving on Gotland with 5-7 claim free years. The relative factor for this combination $f_{low} = 0.0279$ which would yield an annual premium of $185 f_{low} = 5.16$ SEK. This premium is non-sensible and like an overfit towards the Gotland outlier observation. One should as previously stated, group Gotland with other zones and create more rigid estimates.

   Finally we have the ratio $\frac{f_{high}}{f_{low}} = 15878$ which as a consequence from the Gotland observation, is a very high factor, likely not sensible for a future tariff model.

Variable	Rel Freq Factor	lower	upper	Rel Sev Factor	lower	upper	Relative Factor	lower	upper	Current Factor	Duration	Number of Claims	Total Cost	Average Cost
Baseline	1.00			1.00			1.00			1	2 432	9	216 896	1 952 064
bonus_cat1	1.34	1.12	1.60	1.46	1.00	2.14	1.96	1.12	3.42	1.25	25 044	264	6 307 773	23 893
bonus_cat2	1.37	1.13	1.68	1.10	0.74	1.63	1.51	0.83	2.74	1.125	12 296	152	4 502 968	29 625
mcklass1	1.48	1.06	2.05	0.63	0.35	1.15	0.93	0.37	2.36	0.625	5 190	46	993 062	21 588
mcklass2	2.10	1.56	2.85	0.72	0.40	1.31	1.52	0.62	3.73	0.769	3 990	57	883 137	15 494
mcklass4	1.31	1.02	1.69	0.84	0.49	1.43	1.10	0.50	2.42	1.406	11 740	98	2 191 578	22 363
mcklass5	2.04	1.63	2.56	0.96	0.57	1.62	1.96	0.93	4.14	1.875	13 440	149	3 297 119	22 128
mcklass6	3.98	3.19	4.98	1.36	0.80	2.31	5.41	2.55	11.50	4.062	8 880	175	4 160 776	23 776
mcklass7	3.28	1.45	7.41	1.36	0.47	3.90	4.46	0.69	28.93	6.873	331	6	144 605	24 101
vehage_cat1	3.26	2.66	4.00	2.00	1.32	3.02	6.52	3.52	12.05	2	4 955	126	4 964 419	39 400
vehage_cat2	1.90	1.57	2.30	1.84	1.25	2.71	3.50	1.96	6.24	1.2	9 754	145	5 506 945	37 979
zon1	5.13	4.18	6.29	1.25	0.79	2.00	6.42	3.29	12.55	7.678	6 205	183	5 539 963	30 273
zon2	2.72	2.21	3.34	1.28	0.79	2.06	3.48	1.75	6.89	4.227	10 103	167	4 811 166	28 809
zon3	1.71	1.36	2.14	0.83	0.51	1.35	1.42	0.70	2.89	1.336	11 677	123	2 522 628	20 509
zon5	0.91	0.47	1.77	0.57	0.21	1.53	0.52	0.10	2.72	1.734	1 582	9	104 739	11 638
zon6	1.04	0.64	1.68	0.66	0.32	1.36	0.68	0.20	2.28	1.402	2 800	18	288 045	16 003
zon7	0.72	0.10	5.14	0.03	0.00	0.33	0.03	0.00	1.70	1.402	241	1	650	650

Figure 1: Resulting relative factors from fitting a poisson glm and a lognormal severity glm model base on given data. The lower and upper intervals are in the 95% normal percentile.