Kungliga Tekniska Högskolan

SF2930 Regression Analysis

Report II

Isac Karlsson
Ludvig Wärnberg Gerdin

Examiner
Tatjana Pavlenko

March 10, 2020

Contents

1 Introduction

1.1 Background

Most of the tractors in Sweden have a third party liability insurance, because they are required by law. In southern Europe a few large players have dominated the sales of tractor insurances. Our main task this project is to create our own tractor tariff on the form:

$$\text{Price} = \gamma_0 \prod_{k=1}^{M} \gamma_{k,i} \tag{1}$$

Here y0 corresponds to the base level and yki are the risk factors and corresponds to each individual tractor.

1.2 Data

If P&C have provided us with data to train this price model, example given in the table below.

1.3 Problem description

1.3.1 Risk Differentiation and Grouping

Using GLM analysis we aim to make each group Risk homogeneous and that they contain enough data to get a stable GLM analysis, meanwhile handling imperfections in the dataset.

1.3.2 Levelling

Here we aim to calculate yo such that the forecasted claim costs for each insurance are covered by the the price for each insurance, on a full year basis. We us a ratio between the estimated claim cost and the total premium of 90%. Lastly we calculate the base level yo from the formula given in (1).

2 Methods and Methodological Considerations

2.1 Grouping and Risk Differentiation

The criteria on which we based our groups were that

1. Each group should be risk homogeneous, and

2. Each group should have enough data to make the GLM estimates stable.

Greater emphasis were placed on fulfilling criteria 2) due to it being more concrete. In order to do that we considered cut-offs that placed a fairly equal shares of data in each risk group.

In addition to the above, we adjusted the data slightly to adjust for some odd rows. For example, the tractors with a weight of 0 were placed in a category of its own. Also, we noticed the use of (Other) and Other as different factor levels for the ActivityCode regressor. These levels, however, were left as they are. The rationale was that we assume a tractor with ActivityCode = Other and ActivityCode = (Other) are mapped to more specific types of businesses internally by If. In a future version of this model, the model could input more granular groups of ActivityCode to potentially improve the performance of the model.

The resulting cut-offs and risk groups are found in section 3.

2.2 Levelling

From the results of section 2.1, we get risk factor estimates for each level of each predictor. These are henceforth referred to as the "group factors".

For each corresponding assignment presented under Levelling in the project description, we conducted the following:

1. From the original data, we selected those rows (or tractors) that had a `RiskYear` 2016. That way the GLM analysis were only conducted on the active customers, leaving out those that weren't customers to If anymore.

   Following the GLM-script each row were aggregated to include one row per combination of variables. From the aggregated data, we calculated the expected yearly claim-cost per tractor by dividing the duration by the corresponding claim cost for each row. The rationale was to enable us to analyse the yearly cost, even if the insurances had not been active for all of 2016.

2.

In pseudo-code, the process was

```
for each
```

3 Results

3.1 Grouping and Risk Differentiation

3.2 Levelling

Table 1: AIC for model with and without ActivityCode predictor

	With ActivityCode	Without ActivityCode
Frequency model	808.121	851.2993
Severity model	9573.452	9614.1782

Table 2: Variable groups and corresponding risk factors

rating.factor	class	duration	n.claims	cost	rels.frequency	rels.severity	rels.risk
Weight	<1000kg	21995.8389	79	1365398.63	0.2869079	0.7915870	0.2271125
Weight	1000-5000	28175.4419	313	8393343.76	1.0000000	1.0000000	1.0000000
Weight	>5000kg	4259.9868	96	4144599.90	1.7337642	2.0072123	3.4800328
Climate	Middle	21991.9321	188	3777858.34	0.9575653	0.7009944	0.6712480
Climate	North	8887.6077	89	2752275.02	1.1743785	0.8473031	0.9950545
Climate	South	23551.7278	211	7373208.93	1.0000000	1.0000000	1.0000000
ActivityCode	A - Agriculture, Hunting and Forestry	9530.0000	106	3498769.47	1.3550861	0.9076681	1.2299684
ActivityCode	C - Mining and quarrying	1324.3479	13	411922.13	1.2602618	1.4183414	1.7874814
ActivityCode	F - Construction	2504.5092	44	1783928.34	2.1622092	1.8357804	3.9693412
ActivityCode	G - Wholesale & retail trade; repair of motor vehicles, household	1353.1258	14	426094.59	1.4354881	1.2457129	1.7882060
ActivityCode	H - Hotels and restaurants	1245.1478	19	452105.30	1.8504858	0.9574236	1.7716987
ActivityCode	I - Transport, storage and communication	475.7573	2	36313.86	0.6831100	0.7708704	0.5265893
ActivityCode	L - Public administration and defence; compulsory social security	5639.0372	48	927661.90	1.4491391	0.8133608	1.1786730
ActivityCode	M - Education	749.2439	5	94636.54	0.9021751	0.8200808	0.7398565
ActivityCode	Missing	27273.8746	151	4390569.38	1.0000000	1.0000000	1.0000000
ActivityCode	N - Health and social work	2661.5378	66	1407760.95	3.2385437	0.8298863	2.6876230
ActivityCode	Other	1674.6861	20	473579.83	1.4014307	0.8318882	1.1658337
VehicleAge	01_<4years	12597.1703	167	4689891.27	3.2788492	1.0586949	3.4713010
VehicleAge	02_4-8years	12133.4954	126	3744243.69	2.5973204	1.0347108	2.6874754
VehicleAge	03_9-15years	14189.8670	108	3487266.62	1.7561213	1.2226605	2.1471402
VehicleAge	04_>15years	15510.7349	87	1981940.71	1.0000000	1.0000000	1.0000000

4 Conclusion