Non-Life Insurance: Final Project

Academic Year 2020-2021

University of Ljubljana

Univerza *v Ljubljani*

KU LEUVEN

LRISK

Instructions for the Assignment

You should provide answers in English. You get points for the methods you use, for your clear explanation and discussion of them and the overall quality of your report, not just for the final answer. You must use R or Python for your calculations and graphics.

Success!

Deliverables for the Assignments

Please hand in on or before June 28, 2021 via Canvas:

1. A report (format: pdf file or html) in which you answer the *Assignment Questions* stated below and which contains a selection of relevant figures. The figures should be labeled properly and integrated in your report.

2. The R or Python script (or notebook) that you have used for all your calculations. Your code should be well-organized and easy to read.

Please mention the names and student numbers of your team members on both items. It is allowed to work in teams (with three students maximum); it suffices to submit one solution per team.

Each team of students will deliver an (online) pitch presentation (schedule to be determined). This allows the teaching team to give feedback on the report, the models constructed and the presentation.

Assignment Questions

You analyze the data set (in `.csv`) that is available on Canvas. This data set contains observations on the variables listed in the table printed below. Your report should document the following steps:

1. An exploratory data analysis.

2. The construction of a (technical) tariff structure for a car insurance product. Hereto you analyze both the frequency and severity information in the data with (at least) two of the methods/algorithms discussed in the lectures (GLM, GAM, regression tree, bagging, random forest, gradient boosting, ...). You combine frequency and severity models appropriately into a technical pure premium. You compare the performance of the constructed models, based on your own defined set of criteria. You discuss the resulting (pure premium) pricing structure.

3. As an extra step you will discuss (and demonstrate) the calculation of a safety (or risk) loading on top of the pure premiums. To calculate these risk loadings you explore the literature on insurance pricing and propose a suitable strategy. Yang et al. (2020) is a useful starting point.

There is no need to answer the above questions separately (question by question) in your report. A well structured text that covers the above items is preferred. Be creative and rigorous!

`ageph`	age of the policyholder
`CODPOSS`	postal code in Belgium
`duree`	exposure, fraction of the year the insured is covered
`lnexpo`	log of exposure
`nbrtotc`	total number of claims during period of exposure
`chargtot`	total claim amount
`agecar`	age of the car: $0-1$, $2-5$, $6-10$, > 10
`sexp`	sex of the policyholder: male or female
`fuelc`	type of fuel: petrol or gasoil
`split`	split of the premium: monthly, once, twice, three times per year
`usec`	use of the car: private or professional
`fleetc`	car belonging to a fleet: yes or no
`sportc`	sport car: yes or no
`coverp`	coverage: MTPL, MTPL+, MTPL+++
`powerc`	power of the car: < 66, 66-110, >110

The coordinates in the file `inspost.xls` on Canvas can be used to find the (`lat, long`) of the center of a postal code area. Alternatively, you can calculate these directly from the Belgian shape file, using dedicated `R` or `Python` packages and functions.

L. Yang, Z. Li, and S. Meng. Risk loadings in classification ratemaking. https://arxiv.org/abs/2002.01798, 2020.