Assignment Data Analytics for Non-Life Insurance – Collective Insurance Risk Models

Introduction

In this assignment you will apply some collective insurance risk models to car insurance data. Most of the material is covered in the first five weeks of the course Data Analytics for Non-Life Insurance.

The assignment can be done in groups of 3 or 4 persons. Each group will have a group number that is linked to a specific dataset on Canvas. Please subscribe to a group on Canvas.

The group number then determines which dataset you will have to use. For each group number, data will be provided in a comma separated csv-file. It includes headings in the first row. An accompanying document explaining the variables in the data is also provided. The data represent a cross-section of previous claim amounts on several policies that can be used to estimate parameters or to sample from. You can assume that this sample is representative for your own balance sheet and we further assume that the composition of your policies beginning 2023 is identical to the composition in year 2022 (so you can just filter your historical data for the year 2022 to analyze your aggregate claim amount over 2023).

Note that the volume v_i in your dataset differs per policy.

To describe your research question, let *S* denote your aggregate claim amount over 2023 and define

Collective model: $S = \sum_{i=1}^{N} Y_i$, where *N* is the random variable for the number of non-zero claims and Y_i is the claim amount for claim *i*.

Use your dataset and this collective model to estimate the values for *s* such that $P[S \leq s]$ equals 0.9, 0.95 and 0.99, respectively.

In some approaches below, simulation will be used. Let *J* denote the number of scenarios in your simulation and let *j=1,…,J*. Then use the following notation for the simulated aggregate claim amount in scenario *j*:

Collective model: $S_j = \sum_{i=1}^{N_j} Y_{ij}$, where N_j is the number of non-zero claims in scenario *j* and Y_{ij} is the claim amount per claim *i* in scenario *j*.

Questions

Use your dataset to:

Estimate lambda by using the technique in Section 2.3.3.

Determine and report *n*, the number of policies on your balance sheet at the beginning of 2023, and $v = \sum_{i=1}^{n} v_i$,

Estimate the values for *s* such that $P[S \leq s]$ equals 0.9, 0.95 and 0.99, respectively, by applying the following approaches:

- Collective model
  - Compound Poisson combined with
    - Normal Approximation (Chapter 4)
    - Simulation (not explicitly covered in class)
- Your initial capital *u* is such that without reinsurance your ruin probability equals 1% (See Chapter 4). Assuming your safety loading *θ* equals 0.05, determine *u* using the normal

approximation. You will now use proportional reinsurance to reduce the default probability to 0.5%. A reinsurer has offered you a proportional reinsurance contract on all your policies with proportionality factor α (so you keep fraction α of the claims) and *ξ=0.07*. Clearly explain all the steps you take to determine the optimal α and provide a plot containing your ruin probability and profit as a function of α.

Present your results in a structured way, e.g. by using tables. Also include *E[S]*, $\sigma[S]$ and other building blocks used in determining $P[S \leq s]$. Explain, interpret and compare the results based on the various methodologies. Based on your data, try to find out why the different methods (dis-)agree in the way they do. Which method do you consider most convincing? How could it be improved further using the additional information at hand?

Important remarks

- Deadline for handing in the report: **Thursday 12 October 2023, 4 pm, via Canvas.** Submit your **report** and your **code** (including comments!) for replication of results.
- See below for some guidelines on writing the report.
- The questions leave some room for interpretation of what needs/can be done. This often occurs in practice and using this room is part of the assignment.
- Use of simulation techniques might be new so allow some time to figure that out.
- Document all the checks you performed on your computations.
- You can use any programming language you want. Some support can be provided on Python (Javier Garcia Gonzalez) and R (Erwin Charlier).

Some useful libraries/commands when using R (or Rstudio)

read.table	Reading in data
plot	for graphs

Writing the report

The maximum word count is 5,000 (excluding table of contents, tables, figures, appendices).

For the structure of the report (in English, including page numbers) you can use the following:

- Cover Page including your **group number**, names and student number (**SIS id**) of group members
- Table of Contents
- Introduction
- Research Question
- Data Description and Analyses
- Models and Assumptions
- Methods
- Results
- Possible Improvements
- Conclusion
- Appendices (optional)