

Assignment: Industry Sandbox Group Project

Multi-coverage Claim Modelling for Insurance Packaged Products

1. Background

This group project is to provide you with a real-world project working experience. Imagine you are a group of newly employed actuarial analysts of IAG, part of the commercial line pricing team. Industry partners from IAG are your managers and colleagues, who are willing to answer any questions that you might have. Your lecturer-in-charge is your adviser of this project, who will also be willing to provide advice for your group during this project.

IAG offers insurance packaged products which cover various risks. In this project, you are assigned a project to analyse a property owner package product covering two risks: *property damage* and *business interruption* for small and median enterprises (SMEs). Australian Tax Office (ATO) defines small business in Australia as individual, partnership, company or trust that has an aggregated annual turnover of less than \$10 million. These businesses are common to us in our everyday lives such as the pubs, restaurants, hair salons and cinemas or the plumbers, cleaners and gardeners you might engage with. We often call these businesses SME, standing for Small Medium Enterprises. These SME businesses would usually require several types of insurance covers to protect them against the risks that can severely impact their operations.

In the case of a Property owner package product offered by IAG, association exists between the two coverages (property damage and business interruption) within the package in that claim occurrence of one coverage can be highly dependent on claim occurrence of another. For example, it is usually the case that a Business Interruption (loss of rental) can be triggered by a Property claim (damage to building).

For an insurance packaged product, it is important to be able to understand how a particular claim from one risk coverage (say property damage) might trigger claim for another risk coverage (say business interruption) covered in a packaged product. And incorporating the association between different coverages in claims modelling could make the technical pricing (pure premium) more accurate. However, not allowing for the association between coverages

may lead to underestimation of claims cost and hence not charging enough premium to remain sustainable. Therefore, it is important for IAG to be able to accurately model claims of multiple coverages of insurance packaged products.

2. Business Considerations

The current insurance pricing practice is to model the claims of these coverages independently. However, your managers believe that not allowing for association between the coverages can cause under-pricing, hence deteriorating performance over time. Hence your group is given this task to model the claims of the two associated coverages while considering the business constraints (such as interpretability).

Notes:

- You only need to consider the technical pricing (pure premium) without expenses, profits, tax assumptions. Each coverage should have their own pure premium model.
- Interpretability constraint: the estimated pure premium should be able to be written as a combination of multiplicative and/or additive mathematical functions of rating factors. For example, under the GLM framework, the estimated technical premiums can be naturally written as a product of the exponential functions of rating factors, which satisfies the interpretability requirements.

3. Resources

- 1) [A Practical Guide to Commercial Insurance Pricing](#) provides useful resources to understand the commercial line insurance pricing practice.
- 2) There are generally two methods to model pure premium. For example, pure premium can be modelled as a product of frequency and severity. In Week 5, we will also introduce using a single Tweedie distribution to model a mixture of discrete (Poisson) and continuous response (Gamma) under the GLM framework. You are welcome to consider any approach (and potential models) you prefer to use. We summarize the mathematical formulae below.

General Insurance Pricing: Pure Premium

If the number of policies in a collection, n , is large, then the average provides a good approximation of the expected loss

$$E(X) \approx \frac{\sum_{i=1}^n X_i}{n} = \frac{\text{Loss}}{\text{Exposure}} = \text{Pure Premium.}$$

We define the 'pure premium' to be the sum of losses divided by the exposure; it is also known as a 'loss cost'.

We can multiply and divide by the number of claims, **claim count**, to get

$$\text{Pure Premium} = \frac{\text{claim count}}{\text{Exposure}} \times \frac{\text{Loss}}{\text{claim count}} = \text{frequency} \times \text{severity.}$$

So, when premiums are determined using the pure premium method, we can either take the average loss (loss cost) or use the frequency-severity approach.

- 3) We have purchased RStudio Cloud Instructor account service for this course to support your group collaboration. For those groups who would like to use RStudio Cloud to complete this group project please fill out this [online form](#) asap by 5pm 27 September. Please make sure to sign up an [RStudio Cloud](#) account before submitting the online form.
- 4) Consultation with Industry Partners and Lecturer-in-charge. Workshops, one-on-one consultations, Sandbox Q&A Teams channel, and check-in meetings will be organised to support you developing your own innovative solutions during this term. Different from traditional assignments which is usually developed in a closed environment without consultation or sharing, this group project will provide you with an open and engaging learning environment with an aim to help you achieve your personal best. Both industry partners and educators are your resources, who are endeavour to help you throughout this process. If there is anything unclear for this assignment, or if you have any question, please don't hesitate to consult with them for clarification or advice.

4. The Tasks

The objective of this exercise can be separated into the following tasks. The audience of the project is the IAG industry partners who you will have met in class.

- **Task 1:** Exploratory data analysis, such as data quality checking, data visualization, data manipulation, and getting data ready for next steps.
- **Task 2:** Model building. How to model the pure premium for each of two coverages (Property Damage and Business Interruption)? How to take the association of the two coverages into consideration? This task focuses on the model design, techniques, and the modelling process.
- **Task 3:** Model interpretation, evaluation and validation in the business context. This task includes model outcome interpretation in the business context, model validation

and evaluation (testing). This task also includes summarising the results and provide practical suggestion to IAG.

5. You will develop the following skills through the project

- Business Understanding
- Data Understanding
- Exploratory Data Analysis
- Modelling
- Communication
- Team Collaboration

6. Data Descriptions

Data Structure

The dataset is at an individual property (situation) and accident month level. Each individual property has a unique policy number and situation number. There could be more than one property covered within one policy. These are differentiated by situation number within the policy. For example, a policy covering for two properties will have the first property have situation number as '0' and the second property will have situation number as '1'.

As demonstrated in the example below, this policy has two properties covered. This first property has 'situation_num' = 0 and the second property has 'situation_num' = 1. The first property has sum insured = 743700, whereas the second property has sum insured = 900000. In general, different properties can have different characteristics even if it's covered in the same policy.

Example

effectdate	expirydate	ym	xm	situation_num	start	suminsured_prop
18/03/2015	18/03/2016	2015	3	0	18/03/2015	743700
18/03/2015	18/03/2016	2015	4	0	31/03/2015	743700
18/03/2015	18/03/2016	2015	5	0	30/04/2015	743700
18/03/2015	18/03/2016	2015	3	1	18/03/2015	900000
18/03/2015	18/03/2016	2015	4	1	31/03/2015	900000
18/03/2015	18/03/2016	2015	5	1	30/04/2015	900000

The policy term is divided into individual accident months (allowing for any cancellations), so an annual policy will generally have 13 records in a term. Exposure is calculated as the number of days of cover in each accident month divided by 365. If a claim exists for a property, it will be recorded against the accident month that the loss occurred in.

Data Headers

Column Name	Description
polycyno	Policy number
effectdate	Effective date of the policy term

expirydate	Expiry date of the policy term
ym	Accident year
xm	Accident months
situation_num	Unique property identifier within one policy
riskpostcode	Postcode where the property is located
state	State where the is vehicle located
start	Start date for that accident month (usually it is the last day of the previous month, but if the policy is only effective from mid-month, start date = effective date)
Suminsured_prop	Sum insured for property cover
epy	Number of days of cover in each accident month divided by 365
Building_age	The age of the building
Building_type	Type of the building (have 4 categories)
Construction_walls	Construction material of the walls (have 5 categories)
Construction_floor	Construction material of the flooe (have 3 categories)
Sprinkler_type	The type of sprinkler system the insured building has (have 4 categories)
Grossincurred_prop	Claim cost for property damage cover including payments and case estimates. This amount is <u>gross</u> of excess and recoveries.
Indem_per_grp	Indemnity period for loss of income cover <i>*Indemnity Period means the period beginning with the date of the occurrence of the physical loss, destruction or damage during the Period of Insurance and ending not later than the last day of the 'Indemnity Period' shown in the Schedule, during which period the results of the Business are affected as a consequence of the physical loss, destruction or damage.</i>
Suminsured_lossofinc	Sum insured for loss of income cover
Grossincurred_lossofinc	Claim cost for loss of income cover including payments and case estimates. This amount is <u>gross</u> of excess and recoveries.
occupation	The occupation of the tenant who is occupying the commercial property (have 59 categories)
Property_cover	A flag indicating whether the policy has property damage cover
LossofInc_cover	A flag indicating whether the policy has loss of income cover