

Assignment: Industry Sandbox Group Project

Predicting Claims Inflation for Commercial Auto Insurance Pricing

1. Background

For technical pricing purposes, insurers need to model and predict total claims costs based on the available rating factors and claims inflation factor. For the insurance sector, claims inflation is one of the critical inputs for pricing. Actuaries generally find it very challenging to identify drivers of historical inflation and draw on the insights to predict future claims inflation.

Claims inflation, including both increases in claim frequency and/or claim size, has been a major issue for the general insurance industry in recent years, adding further pressure on performance. Within the motor insurance class, claim costs have been impacted by:

- Rising repair costs caused by worker shortages and supply chain issues
- Higher replacement cost due to rising market value of vehicles

Some of above issues are further exacerbated by the COVID environment. While claim frequency during lock down periods had reduced to partially offset the higher claims inflation, this is not expected to continue in the post-COVID environment as people and businesses gradually return to normal. Therefore, it is important for insurers to be able to accurately anticipate future claims inflation so that it can be priced appropriately.

2. The Challenge

The current pricing practice is to use historical data to set assumptions for the future and monitor claims inflation as new experience emerges. However, claim experience can take a long time to develop (due to delays in claims reporting and settlement) and can be volatile over the short term. This limits insurers' ability to react quickly to emerging trends. It would therefore be beneficial if insurers are able to identify early drivers or indicators of claims inflation. One possible approach is to understand **how external data can be leveraged to help predict claims inflation**.

3. Business Consideration

One way to estimate the claims inflation is by using historical claims data alone (e.g. by including a time factor in the model for claims cost). When using the sample claims data (see Section 6) provided in this assignment for WFI Commercial Motor Insurance to estimate the historical inflation (or inflation trend) for this portfolio, we need to be careful that

- Claims inflation in this context can include increases in claim frequency and/or claim size, so both will need to be considered;
- Any changes of policies over time will need to be accounted for when analysing the historical experience. For example, a policyholder's details may change over time (e.g. postcode, sum insured). Furthermore, not all policies would have exposure over the entire 5 year sample period – some could be new policies that started partway through the sample and existing policies could have cancelled/lapsed during this period as well.
- Care needs to be taken when interpreting the claims experience during the COVID period to ensure that it does not unduly distort the analysis (especially as claim frequency is expected to eventually return to normal levels).

Although using historical claims data along with subjective overlays based on the future inflation outlook is a common industry practice, it may not provide us with the best estimate of future claims inflation, especially during recent years. We could also consider leveraging external factors for claims inflation prediction. When using external factors to model and predict the claim inflation trends, some potential factors to consider:

- Economic variables such as interest rates and exchange rates that may directly or indirectly impact inflation;
- Factors that may influence or be indicative of the level of driving activity such as petrol prices or mobility data;
- Other factors that might be helpful - be open-minded!
- These external factors would ideally be readily observable and updated on a frequent basis so that they can be used in practice to predict future claims inflation.

External Data Resources:

- The Australian Macro Database: <http://www.ausmacrodata.org/>
- [Yahoo Finance](#) (You could download historical data.)
- [Australian Bureau of Statistics \(ABS\)](#) and the [Reserve Bank of Australia \(RBA\)](#)
- Other datasets that you could find

For claims cost modelling, insurers would generally require interpretability of the model so that they could explain the different factors affecting the pricing outcome to stakeholders (such as regulators, managers, and customers).

4. The Tasks

The objective of this exercise can be separated into the following tasks

- **Task 1:** Exploratory data analysis, such as data quality checking, data visualization, and getting data ready for next steps. This step also includes external data collection, understanding, and preprocessing. Think about what questions you would like to ask and explore the data to find the answers. (Please specify the external datasets and variables you used to complete this assignments.)
- **Task 2:** Model building. How to model claims inflation using both internal and external data for future claims cost prediction? This task focuses on the model design, techniques, and the modelling process.
- **Task 3:** Model interpretation, evaluation and comparison. This task includes model outcome interpretation in the business context, model comparison (if multiple models are explored), model validation and evaluation (testing). This task also includes summarising the results.

5. You will develop the following skills through the project

- Business Understanding
- Data Understanding
- Exploratory Data Analysis
- Modelling
- Communication skills

6. Data Descriptions

The dataset provides WFI Commoncial Motor Insurance claims data at an individual vehicle and accident month level. Each individual vehicle has a unique policy ID. The policy term is divided into individual accident months (allowing for any cancellations), so an annual policy will generally have 13 or more records in a term (the sum of exposures is 1). Exposure is calculated as the number of days of cover in each accident month divided by 365.

If a claim exists for a vehicle, it will be recorded against the accident month that the loss occurred in. However, if there are multiple claims in the same accident month for that vehicle, the second or subsequent claim will be recorded on a separate row. Exposure for the additional rows of claims is set to zero to avoid any double counting.

In the example below, a vehicle had 2 separate claims in the accident month ending 30/04/2021. The first loss occurred on 15/04/2021 and was for \$500. The second loss occurred on 20/04/2021 and was for \$11,000. The second row for that accident month is to record the second claim, but its exposure is set to 0 because that has already been counted by the first record.

Example

- The table below contains claims information (14 records) for a policyholder in one term (the sum of exposures in the table below is 1), and the policyholder started the contract from 16/06/2020.

accident_month	policy_id	exposure	term_start_date	term_expiry_date	claim_loss_date	total_claims_cost
30/06/2020	10101	0.0411	16/06/2020 0:00	16/06/2021 0:00		
31/07/2020	10101	0.08493	16/06/2020 0:00	16/06/2021 0:00		
31/08/2020	10101	0.08493	16/06/2020 0:00	16/06/2021 0:00		
30/09/2020	10101	0.08219	16/06/2020 0:00	16/06/2021 0:00		
31/10/2020	10101	0.08493	16/06/2020 0:00	16/06/2021 0:00		
30/11/2020	10101	0.08219	16/06/2020 0:00	16/06/2021 0:00		
31/12/2020	10101	0.08493	16/06/2020 0:00	16/06/2021 0:00		
31/01/2021	10101	0.08493	16/06/2020 0:00	16/06/2021 0:00		
28/02/2021	10101	0.07671	16/06/2020 0:00	16/06/2021 0:00		
31/03/2021	10101	0.08493	16/06/2020 0:00	16/06/2021 0:00		
30/04/2021	10101	0.08219	16/06/2020 0:00	16/06/2021 0:00	15/04/2021	\$500
30/04/2021	10101	0	16/06/2020 0:00	16/06/2021 0:00	20/04/2021	\$11000
31/05/2021	10101	0.08493	16/06/2020 0:00	16/06/2021 0:00		
30/06/2021	10101	0.0411	16/06/2020 0:00	16/06/2021 0:00		

Data Dictionary:

Column Name	Description
accident_month	Month in which any losses incurred on the insured vehicle would be covered
term_start_date	Start date of the policy term
term_expiry_date	Expiry date of the policy term
policy_id	Unique policy identifier
policy_tenure	Number of years that the policy has been with the insurer
cover_sum_insured	Sum insured of the vehicle
risk_state_name	State where the vehicle is located
risk_postcode	Postcode where the vehicle is located
year_of_manufacture	Year of manufacture of the vehicle
class_description	Vehicle classification (there are 15 different categories; there is no order to them)
claim_loss_date	Date of loss for a claim
exposure	Number of days of cover in each accident month divided by 365
total_claims_cost	Claim cost including payments and case estimates. This amount is gross of excess and recoveries.