1 Exposure

General claim frequency modeling has traditionally assumed that the occurrence of claims follows a Poisson distribution, implying that claims are proportional to the time at risk and independent of the observation period. However, there is a noticeable seasonality effect: November exhibits the highest claim frequency, while January and February show the lowest. Additionally, the impact of COVID-19 curfews has introduced anomalies that are difficult to explain. Post-COVID, there has been a consistent decline in claim frequency.

To address seasonality and the pre/post-COVID differences, we leverage country-wide claim count summaries available on a quarterly basis.

2 Postcode-Level Territorial Modelling

Since some postcodes have limited exposure due to biased data, we model the expected claim count for each postcode and interpret the deviation as the territorial impact.

Observations indicate a negative exponential relationship between the geographical distribution of claims and the travel distance (measured in hours under optimal traffic conditions). To incorporate this, we employ a distance-weighted Bayesian smoothing approach using the following function:

$$f_p = \frac{c_p + \alpha \sum_i c_i e^{-d_{i,p}}}{e_p + \alpha \sum_i e_i e^{-d_{i,p}}}$$

where:

- f_p = smoothed claim frequency for postal code p
- c_p = observed number of claims in postal code p
- e_p = exposure (weight) for postal code p
- $d_{i,p}$ = distance between postal codes i and p (in hours, optimal traffic conditions)
- α = smoothing parameter

This approach allows for a more stable estimation of claim frequencies for postcodes with limited data by borrowing strength from neighboring areas.

3 Custom Loss Function

To optimize profit, we introduce a custom loss function that adjusts the objective function based on predicted price (P), observed claim (A), and market price (M). This function prioritizes cases that have a more significant financial impact using the following formula:

$$L(P) = \max(M, A) - P + (M - A - (\max(M, A) - P)) \cdot \sigma(P, M - \epsilon, k)$$

where:

- k = sharpness parameter controlling the transition speed
- ϵ = small positive value to define the transition point
- $\sigma(P, M - \epsilon, k)$ = smooth step function defined as:

$$\sigma(P, M - \epsilon, k) = \frac{1}{1 + e^{-k(P-(M-\epsilon))}}$$

This loss function design aims to balance underpricing and overpricing risks, focusing the model's attention on cases with a higher impact on profitability. The smooth step function σ enables a gradual transition, making the optimization process more stable.