# Case

Lukas Andersson

## Problem

You are working as an analyst for an insurance company. A delivery company is seeking to insure their car fleet. You are given statistics for the previous five years presented in table 1 below as well as data for the individual claims containing information about date of accident and cost.

Additional information and conditions given in the case:

- The number of insured vehicles in 2025 will be the same amount of vehicles as in 2024.
- Assume a 3% yearly inflation.
- Assume operations costs are 0.
- The closest competitor is offering a premium at 525 000.

| Year | 2020 | 2021 | 2022 | 2023 | 2024 | 2025 |
|---|---|---|---|---|---|---|
| Total claims cost | 240 336 | 140 368 | 356 549 | 284 752 | 246 378 | ?? |
| Insurance Claims | 30 | 36 | 37 | 41 | 39 | ?? |
| Insured vehicles | 90 | 101 | 113 | 113 | 114 | 114 |

Table 1: Given insurance claims statistics for years 2020-2024. Part of the task is to fill out the missing values in this table for year 2025.

The questions that need to be answered are the following:

- How many accidents will happen in 2025?
- What will the total claims cost be in 2025?
- What premium should we offer?
- Should inflation be taken into consideration in the calculations?

# Solution

Let's start by answering one of the given questions; whether or not inflation should be taken into consideration in the calculations. The aswer to this is **yes**. As we seek to accurately predict the expected costs for year 2025, we need to adjust the price of the historical data to reflect potential costs if the claims were to be made in 2025 instead. This transformation is shown in equation 1 below, where we have the inflation adjusted cost $C_{Y=2025}$ on the left hand side, determined by historical cost $C_Y$, yearly inflation rate $I$ and year of data point $Y$.

$$C_{Y=2025} = C_Y \cdot (1+I)^{2025-Y} \tag{1}$$

## Data analysis

Let's start with analysing the individual claims data. After adjusting for inflation we are left with the updated table 2. A new row has been added, labeled *liable claims*. These refer to the number of non-zero valued claims costs, which I interpret as the cases where the insurance company was found liable to pay for the claim. I interpret the zero-valued claims as cases where the claimant was found to not be covered by the insurance.

| | 2020 | 2021 | 2022 | 2023 | 2024 |
|---|---|---|---|---|---|
| Total claims cost | 278615 | 157985 | 389611 | 302093 | 253769 |
| Claims | 30 | 36 | 37 | 41 | 39 |
| Liable claims | 17 | 18 | 21 | 21 | 20 |
| Insured vehicles | 90 | 101 | 113 | 113 | 114 |

Table 2: Updated table of information regarding vehicles, accidents and costs.

There are some major jumps in the total claims cost from year to year, especially from 2020 to 2021 and from 2021 to 2022. There are no clear linear trends in the data, other than the fact that number of insured vehicles is strictly increasing. In figure 1 you can find data for the individual claims by year.

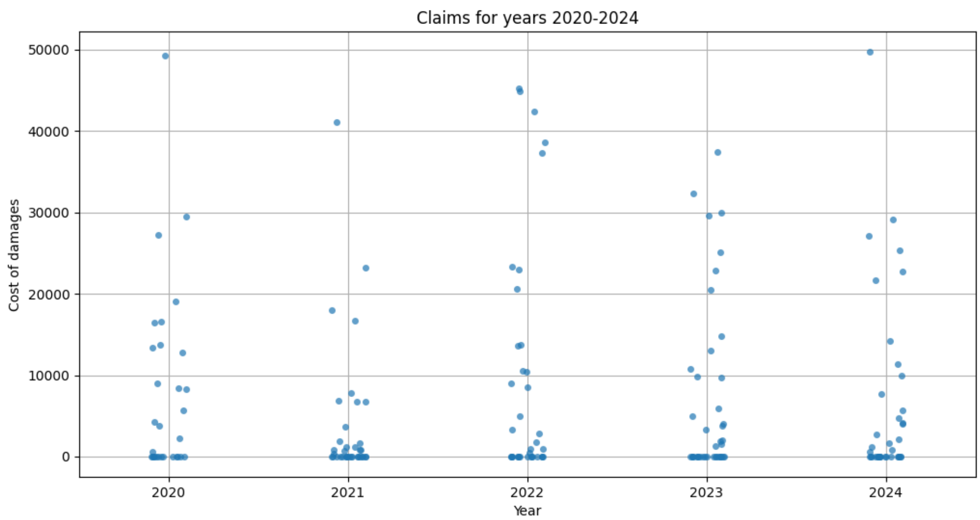


Figure 1: Inflation adjusted data for the individual claims by year.

In the figure we see that the claims are fairly similarly distributed by year with a large amount of zero-valued claims, most non-zero claims having a value of < 10 000, and a few comparatively large claims. From the figure we see that for 2021 specifically we have less extreme values compared to other years. Looking at the right-skewedness of the data, I think a Gamma distribution might fit the non-zero values of the claims. Although 2021 differs a fair bit from the other years I can not think of a reason why that would be and it does not look to be a trend but more a deviation. Therefore I will try to estimate a distribution for the entire data set of non-zero values, using Python.

Assumptions made for the data:

- All vehicles are the same, implying that risk of accident and average claim cost would not differ from year to year for this reason (instead, detected claims cost differences would have other reasons)
- The average claims cost is not affected by year of accident (after adjusting for inflation) meaning the entire data set can be included

In figure 2 you can find the estimated density function and a histogram of the claims data. To determine the goodness of fit, a Kolgomorov-Smirnov test was applied, and the result from the test is presented in table 3. The test statistic and p-value imply that the estimated distribution has a good fit to the observed data and is probable, and can therefore be used to improve calculations of insurance pricing.

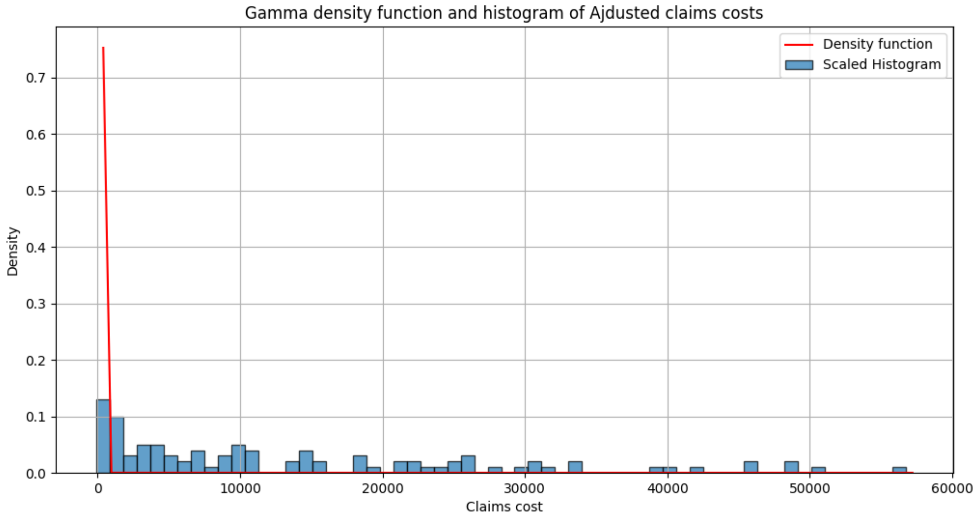


Figure 2: Estimated probability density function and histogram of the observed non-zero claims.

| D | p |
|---|---|
| 0.0766 | 0.592 |

Table 3: Kolmogorov-Smirnovtest to measure the goodness of fit of the estimated Gamma distribution on the observed claims. The test statistic $D$ is a measurement of largest distance between orbserved value and estimated distribution, where a smaller value is better and value of 0.0766 is fairly good. the p-value is a measurement of the probability of observing the data based on the hypothesis of it originating from a process with the estimated distribution. A p-value of $0.592>0.05$ means that the hypothesis cannot be rejected, and that the estimated distribution is probable.

## Pricing

To determine a fair price for the insurance, first we must determine the expected cost. As previously mentioned, alot of the claims in the data are zero-valued. I assume that means that the insurer was not found liable to pay for those claims, and to accurately price the expected cost, the probability of being found liable should be taken into consideration. The expected cost of a claim is calculated using equations 2-4, and in table 4 you can find the results for years 2020-2024 as well as average values. There is no clear observable trend in the data in regards to claim frequency or severity, and compared to the average of the data set we see that the claims cost for year 2022 deviates approximately as much as for year 2021, the previously mentioned outlier.

$$P(Liable) = \frac{Number\ of\ non-zero\ claims}{Total\ number\ of\ claims} \tag{2}$$

$$E(Claims\ Cost|Liable) = \frac{\sum Claims\ cost}{Number\ of\ non-zero\ claims} \tag{3}$$

$$E(Claims\ Cost) = E(Claims\ Cost|Liable) \cdot P(Liable) \tag{4}$$

| | 2020 | 2021 | 2022 | 2023 | 2024 | 2020-2024 |
|---|---|---|---|---|---|---|
| Claims | 30 | 36 | 37 | 41 | 39 | 183 |
| Claims per vehicle | 0.33 | 0.36 | 0.33 | 0.36 | 0.34 | 0.35 |
| Liable claims | 17 | 18 | 21 | 21 | 20 | 97 |
| P(Liable) | 0.57 | 0.50 | 0.57 | 0.51 | 0.51 | 0.53 |
| E(Claims Cost\|Liable) | 16 389 | 8 777 | 18 552 | 14 385 | 12 668 | 14 248 |
| E(Claims Cost) | 9 342 | 4 389 | 10 575 | 7 336 | 6 461 | 7 551 |

Table 4: Expected value of claims costs for years 2020-2024 and average for the same time period.

With the expected value for cost per claim calculated, we can move on to estimate the total claims cost for year 2025. As previously mentioned, there is no observed trend in the amount of claims per vehicle by year. What one could expect is a seasonal trend with the weather conditions affecting probability of accident, but since the insurance is to cover the whole year it is at the time not necessary to take into consideration. As no linear trend has been observed, we can once again use the entire data set to estimate the 2025 values. The total cost for the year can then be calculated by equations 5-6, using values from the rightmost column.

$$E(Claims) = Claims\ per\ vehicle \cdot Insured\ vehicles = 0.35 \cdot 114 \approx 40 \tag{5}$$

$$E(Cost) = E(Claims) \cdot E(Claims\ cost) = 40 \cdot 7\ 551 = 302\ 040 \tag{6}$$

With this estimated claims cost , the competitor is offering to insure the vehicles at a $\frac{525\ 000-302\ 040}{302\ 040} \approx 0.74 = 74\%$ markup, which sounds like quite the large profit margin. To better understand the pricing, let's incorporate the previously estimated probability distribution of liable claims costs. With this distribution, we can estimate probabilities of the average liable claims cost exceeding certain thresholds, and with this we can better understand the risk of offering a premium of a certain price. Table 5 below shows the results from these calculations.

| E(Cost\|Liable) | >20 000 | >25 000 | >30 000 | >35 000 | >40 000 | >45 000 | >50 000 |
|---|---|---|---|---|---|---|---|
| Probability | 22.58% | 16.30% | 11.82% | 8.61% | 6.28% | 4.60% | 3.37% |

Table 5: Probabilities of certain average claims costs being exceeded given the estimated Gamma distribution.

In this table we see that the probability of the average liable claims cost exceeding 25 000 is 16.3%. Assuming the estimated number of claims and probability of being found liable calculated for 2025, the should be 21 liable claims. Therefore there is a 16.3% risk of the claims exceeding a total cost of 525 000, the premium offered by the competitor. To offer a more competitive premium one would have to assume an even larger risk. Let's assume capital allows for a loss risk of 20%. At this level of risk, we get the calculations

$$E(Claims\ Cost|Liable) = Q(0.2) = F^{-1}_{E(Cost|Liable)}(0.2) = 21\ 850 \tag{7}$$

$$E(Claims\ cost) = 21850 \cdot 0.53 = 11\ 580.5 \tag{8}$$

$$E(Cost) = 11\ 580.5 \cdot 40 = 463220 \tag{9}$$

With a 20 % risk of costs exceeding 463 220 in year 2025, I therefore suggest offering a insurance premium of 475 000. Another way to look at the chances of profits and risk of loss is throught equations 10 and 11, which shows that with this pricing there is approximately a 74% chance of making a minimum profit of 100 000 and a 14% risk of making a minimum loss of 100 000.

$$P(Profit \geq 100000) = F_{E(Cost|Liable)}(\frac{475\ 000 - 100\ 000}{Liable\ claims}) \approx 0.7368 \tag{10}$$

$$P(Loss \geq 100000) = 1 - F_{E(Cost|Liable)}(\frac{475\ 000 + 100\ 000}{Liable\ claims}) \approx 0.1421 \tag{11}$$

# Conclusion

- 40 accidents are estimated to occur in 2025
- Total claims cost will be 302 040
- Offer a premium of 475 000
- Inflation needs to be taken into consideration in the calculations. If it is to be used to estimate prices in 2025, historical prices need to be adjusted to what it would cost in 2025 if the same claim were to be made.