Discussion (and demonstration) of the safety loading calculation.

The risk premium of a policy can be split up in two main parts, a pure premium and a risk loading. The pure premium which is calculated in the previous part, is used to pay the future losses of the policy. The risk loading which will be discussed in this part, has the purpose of covering excess future losses that are not covered by the pure premium. (Yang et al., 2020)

The most used approach for calculating the risk premium is by separately analysing the pure premium and the risk loading. Traditionally generalized linear models or GLMs are used for this analysis. Risk loadings can be derived in this traditional way by using the expected value premium principle or standard deviation premium principle. (Yang et al., 2020)

Expected value premium principle: $$ H(Y\_i) = E(Y\_i) + phi\*E(Y\_i) $$

Standard deviation premium principle: $$ H(Y\_i) = E(Y\_i) \* phi sqrt(VAR(Y\_i) $$

GLMs are considered the industry standard (Baione & Biancalana, 2019) although there are some downsides. Traditional regression models like GLMs often have to rely on assumptions (Kudryavtsev, 2009). According to Kudryavtsev the following problems can occur with GLMS. An Inaccurate estimate of loss distribution may occur. This estimation of the loss distribution could be very different from the real one. It could be difficult to give larger weights to extreme values, thus making it difficult to work with loss distributions that have heavy tails. Working with a number of outliers in the sample and dependence structure of the data also could cause problems.

From the previous part there can be concluded that traditional regression methods like GLMs are not always the ideal match for real word situations. Therefore there should be looked at some alternative approaches (Kudryavtsev, 2009). Kudryavtsev and Yang both propose a quantile regression method. Using a quantile regression method has some advantages. It is a distribution-free method, that is robust to outliers. The quantile regression doesn’t require independence Kudryavtsev, 2009). But because GLMs are well known and are the current standard in actuarial sciences this will be the focus here for the discussion of the safety loading (Baione & Biancalana, 2019). Keeping in mind the criticism on these GLMs.

To calculate the risk loading the risk loading parameters need to be determined. In practice the risk loading parameters need to be determined in advance. This can be done in several ways. This can be done in a top-down way as proposed by Buhlmann. Were premiums are related to stability criterion of portfolio risks and dividend requirements for the capital invested into the insurance operation (Buhlmann, z.d.). The risk loading parameters can also be determined by a down-top-down. Here they first calculate the risk premium for each policy with the expected value premium principle at the individual level. They then obtain the total risk premium by putting together these risk premiums. The risk loading parameter can then be determined by making sure that the total of al risk premiums is big enough to cover the total expected losses (Yang et al., 2020).

The top down approach for calculating the risk loading parameters, suggested by Yang in the paper “Risk loadings in classification ratemaking” seems to be the best method. First the risk premium of the entire portfolio will be calculated by applying the bootstrap method. This enables the possibility to get the total risk premium at the collective level. The risk premiums of different tariff classes can then be calculated by GLMs or quantile regressions. Like stated earlier this discussion will only focus on GLMs.

For the calculation of the risk premium the expected value premium principle will be used. This was chosen because it is easy to understand and calculate and often used in actuarial practice (Heras et al., 2018). Before the expected value premium principle can be used the risk loading parameter phi has to be calculated. To calculate phi the earlier mentioned top down approach suggest by Yang will be used. First a predictive distribution of the future losses of the policy will be calculated by using a bootstrap with 2500 repetitions. Yang suggests to use 10000 repetitions this was not possible with the used PC. The bootstrap was run over the pure premium which was calculated in the previous parts. These pure premiums were calculated with a two-part GLM regression. The 99.5% quantile of the predictive distribution of the bootstrap equals $36 214 173. The 99.5% quantile was used, because this is conform the Solvency II regulation(Yang et al., 2020). This 99.5% quantile and the sum of all pure premiums calculated in the previous part can then be used to calculate phi.

$$ Phi = (C – sum\_{E(Y\_i)})/(sum{E(Y\_i}) $$

Phi can then be filled in in the expected value premium principle, so that the risk loading and total risk premium can be calculated for each individual policy.

Baione, F., & Biancalana, D. (2019). An Individual Risk Model for Premium Calculation Based on Quantile: A Comparison between Generalized Linear Models and Quantile Regression. *North American Actuarial Journal*, *23*(4), 573–590. <https://doi.org/10.1080/10920277.2019.1604238>

Buhlmann, H. (z.d.). *PREMIUM CALCULATION FROM TOP DOWN*. 13.

Frees, E. W., Jin, X., & Lin, X. (2013). Actuarial Applications of Multivariate Two-Part Regression Models. *Annals of Actuarial Science*, *7*(2), 258–287. <https://doi.org/10.1017/S1748499512000346>

Heras, A., Moreno, I., & Vilar-Zanón, J. L. (2018). An application of two-stage quantile regression to insurance ratemaking. *Scandinavian Actuarial Journal*, *2018*(9), 753–769. <https://doi.org/10.1080/03461238.2018.1452786>

Kudryavtsev, A. A. (2009). Using quantile regression for rate-making. *Insurance: Mathematics and Economics*, *45*(2), 296–304. <https://doi.org/10.1016/j.insmatheco.2009.07.010>

Yang, L., Li, Z., & Meng, S. (2020). Risk Loadings in Classification Ratemaking. *ArXiv:2002.01798 [q-Fin, Stat]*. <http://arxiv.org/abs/2002.01798>

Conclusion:

After thoroughly analysing the data, a model is build to predict the future losses. This model has a frequency and severity part. To model the frequency a GLM and GBM were constructed. Out of the GLM and GBM, the GLM was most suitable to model the frequency. Both the GLM and GBM were an equally good fit, but the GLM was less complex. To model the severity a GLM and Random Forest were used. The GLM preformed better on the trainset than the Random Forest. Thus the GLM was preferred to model the severity. These GLMs were then used to calculate a pure premium. On this pure premium a risk loading was added. This risk loading was calculated using the expected value premium principle. A top down approach suggest by Yang was used to calculate the risk loading parameters.