On the Optimal Baseline Auto Insurance Premium

PATRICK HOSEIN, The University of the West Indies, Trinidad

The automobile insurance business is quite competitive with narrow profit margins. It is therefore important to determine accurate premiums to charge customers. If premium values are too low then the company may run at a loss. On the hand if premiums are too high, then customers may switch to competitors and business is lost. A premium has three major components, the administrative and operational costs to manage the customer, the expected losses that must be paid out in claims and the profit expected to be made. Many papers have been written on assessment of risk (i.e., predicting claims) but our focus is on the determination of a suitable profit margin for a given competitive environment. Ideally one would like to try different premiums and evaluate the overall profit given the resulting customer base. The premium that provides the largest total profit can then be used. Unfortunately such an approach is impractical. Instead we obtain samples of premium and market share values for multiple insurance companies and then use regression to determine the best trade-off between price and market share. In other words, each company acts as a sample in determining the price elasticity of demand function.

CCS Concepts: • **Computing methodologies → Machine learning**.

Additional Key Words and Phrases: auto insurance, premium optimization, price elasticity, competitive pricing, machine learning

ACM Reference Format:
Patrick Hosein. 2020. On the Optimal Baseline Auto Insurance Premium. In *Proceedings of Workshop on Machine Learning in Finance (KDD 2020).* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

The insurance business is very competitive with typically low profit margins. Therefore it is vital that companies choose proper premium rates to maximize total profits. If rates are too low then they may be operating at a loss. On the other hand, if rates are too high then they may not attract sufficient customers and this too can lead to operating at a loss. Note that the premium to be charged includes several factors. The risk associated with the customer must be taken into account and riskier customers should be charged more. The total operating cost of managing a customer must be taken into account. Our focus is on the profit margin that must be included to allow the company to maximize its overall profit. Many papers have been written on customer risk and predicting customer claims (e.g., [4, 7, 8]) but this is not our focus.

In a competitive market with multiple players it is difficult to determine whether a company's premium is too low or too high. Theoretically one could obtain the optimal point by trying different values of premiums, each leading to a different market-share, and picking the one that gives the maximum total profit. However, in practice, this is not a feasible approach. We instead do the following. Suppose we consider a single customer with particular features and determined the premium that would be charged for this customer by several insurance companies. Furthermore, assume that we can also determine the present market-share of customers with those features. One would expect that

Manuscript submitted to ACM

companies with higher premiums would have fewer customers and vice-versa. We can therefore use these samples to estimate the trade-off between premium value and market-share. From this model we can then determine which premium maximizes total profit. The data required for this exercise was collected from insurance companies as well as from an insurance company association that provides aggregate statistical data (market share) of its members.

We were unable to find papers on similar approaches to what is being proposed. In [6] the authors determine the degree of competition among insurance companies in Brazil, and secondly, they estimate the demand for automobile insurance. Aggregate data was used in their study but the approach is formula based. In [1], the authors study the competition between two insurance companies by using a Stochastic Differential game model. Our contribution is an evaluation of price elasticity using data rather than traditional formulas. The paper [2] investigates price elasticity of demand but in the context of whole life insurance. In [5] the authors address price elasticity for auto insurance but they use an inference model to estimate price-elasticity functions at the individual policyholder level and not at the company level as we do. The paper [9] investigates retention rates of existing customers while our objective is setting appropriate premiums for new customers.

The paper [3] also addresses price elasticity but from a different perspective. They considered individual customers and investigated their willingness to purchase an automobile policy as a function of price. They propose an analytical model and use an experiment to determine the parameters of their model. In the experiment, the normal process was used to determine the premium and then 80% of customers were offered this premium while 10% were offered a premium that was 5% higher and 10% were offered a premium that was 10% lower. Our approach captures user preferences but at an aggregate level using the different premiums (for a specific individual) offered by different companies and the resulting number of customers they attract.

In the next section we provide a description of the data that was used for this study. Premiums for several customer profiles were collected and, using publicly available market share estimates, we then investigate the price elasticity of demand. These functions can then be used to inform individual companies on changes they can make to improve their competitiveness. Note that the model can be enhanced with additional data such as market share values for different customer profiles as well as profit margins and expenses for individual companies. In spite of these limitations we believe that the approach outlined can benefit the automobile insurance industry.

2 DATA-SET DESCRIPTION

We obtained the premium values for 16 different customer profiles from 8 insurance companies. This dataset was compiled by having each customer call each insurance company and asking for the premium given their specific personal information (e.g., gender, age, number of claim free years, value of insured automobile, etc.). In Table 1 the first two rows list the premiums for the various insurance companies for two such customer profiles in which the only difference in attributes was gender and age. Using demographics and other publicly available insurance statistics we also determined the percentage of available customers for each profile and we then averaged premiums over all profiles for each company. This is provided in row 3. Finally we were able to obtain market shares for each of the companies and this is provided in row 4. The market-share percentages do not add to 100 because we did not have premiums for all companies.

3 NONLINEAR REGRESSION FOR MALE PROFILE

We assume that the profit achieved by a company is a percentage of the premium charged. Furthermore, we assume that the market share values hold for each of the profiles. For a given profile, assume that the total customer population

Table 1. Premiums (in USD) for various Insurance Companies

Company	A	B	C	D	E	F	G	H
Male, 30 yrs	2,064	2,188	2,845	1,279	1,881	1,770	1,671	1,313
Female, 50 yrs	2,064	2,610	2,777	1,298	2,149	2,655	1,713	1,224
Average Premium	422	485	729	594	630	578	539	481
Market Share	9%	16%	12%	6%	5%	4%	15%	8%

is N, the market share for a particular company is α, the premium charged per customer by that company is r and the profit margin for that company is κ. The total profit P is therefore given by $P = \alpha N \kappa r$. We assume that N and κ are approximately the same for all companies and so normalize profit by the quantity $N\kappa$. If we plot this normalized total profit $\hat{P} = \alpha r$ versus r then the maximum point will provide the optimal r.

Note that we are using the information for each company as samples for the trade-off between profit and premium. We then use nonlinear regression to determine this function. In Figure 1 we provide the sample points for the Male profile together with the regression function. Note that the regression function is a close match to the sample points. We left out the points associated with A, E and F when performing the regression because these are outliers. One can perform a more formal analysis to show that these points are indeed outliers but a simple visual inspection conveys the information. These companies are not operating at optimal efficiency since they can potentially achieve a much higher market share for the premium they charge. This could simply be due to the fact that these companies have decided not to target customers with this particular profile. The regression function is given by

$$\hat{P} = -1476.56r + 4625.14\sqrt{r} - 3260.24. \tag{1}$$

The maximum total profit is achieved with a premium of $2,453 and a market-share of 14.7%. However, note that a company cannot simply change their premium to this optimal value and obtain the corresponding market-share and hence optimal profit. Since this is a closed system, any change made by one company affects the others and hence a company would need to make small changes in premiums, achieve the required market-share and once the market settles then repeat the process. However, one can deduce the effect of small changes to the system. For example, consider company B. If this company were to increase the premium for this profile by 5% then, assuming that they maintain the same level of efficiency, their market-share will drop by 2.7% but their total profit will increase by 2.1%.

4 NONLINEAR REGRESSION FOR AVERAGED PROFILE

In this section we consider the premiums that were obtained by taking a weighted average of premiums over the 16 profiles for each company. The resulting regression function is plotted in Figure 2. The outliers D, E and F were left out when determining the regression curve which, in this case, is the following quadratic equation.

$$\hat{P} = -0.001017r^2 + 1.3423r - 350.1289 \tag{2}$$

As mentioned previously, whenever one company makes premium changes then potentially market shares for all companies may change. Let us assume that all companies except D, E and F, continue to make small adjustments towards the optimal point. Assuming that all companies are equally efficient and have the same profit margins, expenses and claims per customer then they should all converge to the same premium and market share. Let us assume that the total market share of these companies remains the same then we can compute the resulting common premium to be

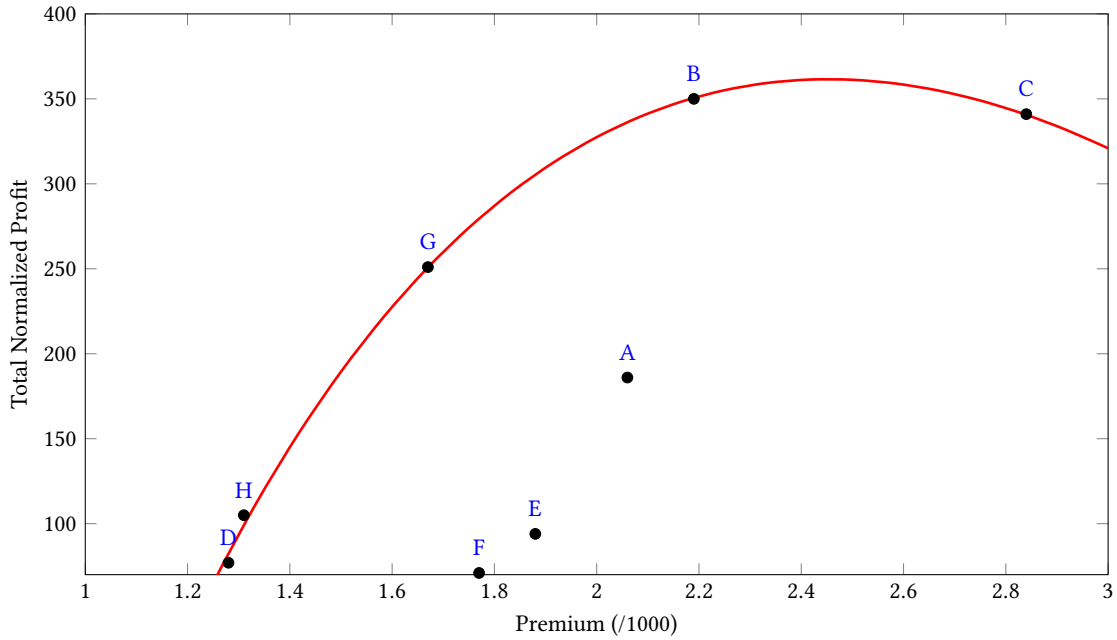

Fig. 1. Regression Function for Total Profit for Male profile

$P^* = 537$. However, this scenario is unlikely because the more efficient companies will tend to remain more efficient than the less efficient ones. If the less efficient ones are making a profit then they may not do much more to be more competitive to take away market-share.

5 DISCUSSION

Note that more data is required for better accuracy but the analysis provided can give some indication of actions that should be taken by each company. Once companies take actions then of course the ecosystem is affected and the price elasticity curve changes. However if a single company makes a small change then the effect should be small. Let us consider what small steps each company should take.

Company A is clearly charging too little and is also not efficient since its market-share is relatively small. They can increase premiums by 1.4% while maintaining the same total profit but with a small loss in market-share. This will bring them on the red line where further modifications can be made. Company B is in a better position. They can increase premiums by 11.1% and maintain the same total profit with only a small loss in market-share. They can then move along the red curve. Company C is already in a very profitable position but can increase their profits further. If they were to decrease their premium by 5% they can increase market-share and increase overall total profit by 10%. This is the only company that can potentially reduce premiums and increase profit. Companies D, E and F are clearly not running at optimal efficiency. For the premiums that they charge their market shares can be significantly higher. This can probably be accomplished by reducing their operational costs and profit margins while keeping premium values fixed. With these changes they can invest more in attracting customers by offering more free additional features as well as using targeted marketing campaigns. Company G is already in a good position but they can also make a small increase in premiums

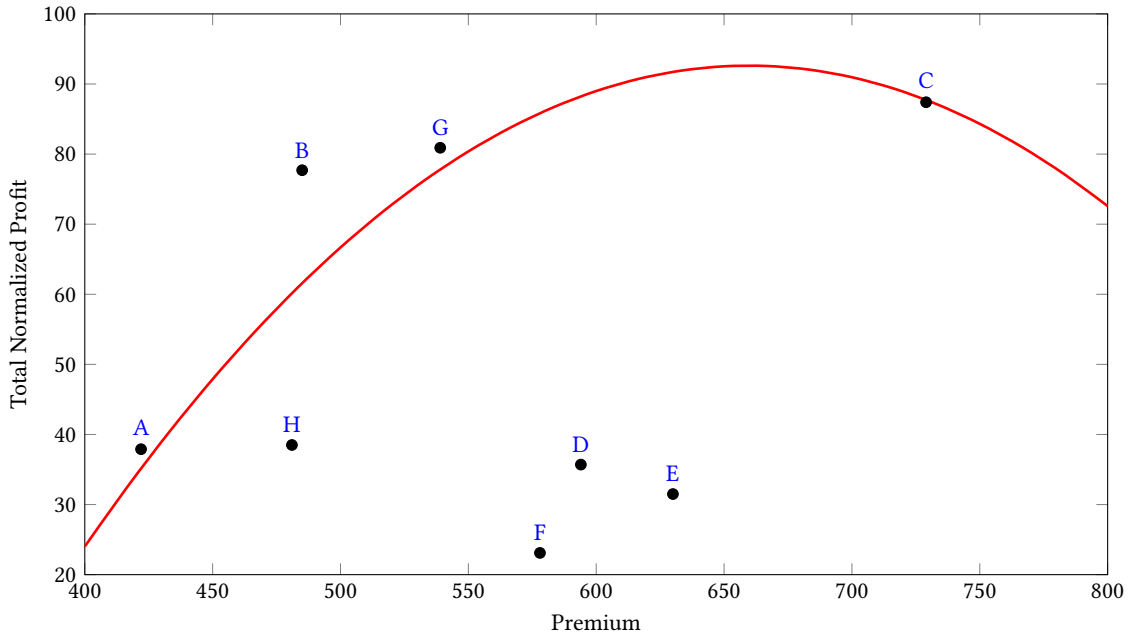

Fig. 2. Regression Function for Averaged Premiums

while maintaining the same profit margin. Finally company H should also try to increase operational efficiencies and re-invest profits into increasing profit share. Future work will entail gathering more company specific data, such as costs, profit margins and market-share, to provide a more comprehensive analysis. We also believe that this work can help consumers make better policy decisions especially when companies include personalized premium offerings.

REFERENCES

[1] Søren Asmussen, Bent Jesper Christensen, and Julie Thøgersen. 2019. Nash equilibrium premium strategies for push–pull competition in a frictional non-life insurance market. *Insurance: Mathematics and Economics* 87 (2019), 92 – 100. https://doi.org/10.1016/j.insmatheco.2019.02.002

[2] DAVID F. BABBEL. 1985. The Price Elasticity of Demand for Whole Life Insurance. *The Journal of Finance* 40, 1 (1985), 225–239. https://doi.org/10.1111/j.1540-6261.1985.tb04946.x arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.1985.tb04946.x

[3] Catalina Bolancé, Montserrat Guillen, Jens Nielsen, and Fredrik Thuring. 2018. Price and Profit Optimization for Financial Services. *Risks* 6, 1 (Feb 2018), 9. https://doi.org/10.3390/risks6010009

[4] Massimo De Felice and Franco Moriconi. 2019. Claim watching and individual claims reserving using classification and regression trees. *Risks* 7, 4 (2019), 102.

[5] Leo Guelman and Montserrat Guillén. 2014. A causal inference approach to measure price elasticity in Automobile Insurance. *Expert Systems with Applications* 41, 2 (2014), 387 – 396. https://doi.org/10.1016/j.eswa.2013.07.059

[6] Vivileine Maria Peres, Wilfredo L. Maldonado, and Osvaldo Candido. 2019. Automobile insurance in Brazil: market concentration and demand. *Revista Contabilidade & FinanÃSas* 30 (12 2019), 396 – 408. http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1519-70772019000300396&nrm=iso

[7] Jessica Pesantez-Narvaez, Montserrat Guillen, and Manuela Alcañiz. 2019. Predicting motor insurance claims using telematics data—XGBoost versus logistic regression. *Risks* 7, 2 (2019), 70.

[8] Yi Yang, Wei Qian, and Hui Zou. 2018. Insurance premium prediction via gradient tree-boosted Tweedie compound Poisson models. *Journal of Business & Economic Statistics* 36, 3 (2018), 456–470.

[9] Ai Cheo Yeo, Kate A. Smith, Robert J. Willis, and Malcolm Brooks. 2001. Modeling the Effect of Premium Changes on Motor Insurance Customer Retention Rates Using Neural Networks. In *Computational Science - ICCS 2001*, Vassil N. Alexandrov, Jack J. Dongarra, Benjoe A. Juliano, René S. Renner, and C. J. Kenneth Tan (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 390–399.