NOVA IMS Information Management School

MDSAA

Master Degree Program in

Data Science and Advanced Analytics

Price elasticity in motor insurance

Defining price elasticity with causal inference for a major Portuguese insurer

Miguel Ângelo Nunes Lince

Dissertation

presented as partial requirement for obtaining the Master Degree Program in Data Science and Advanced Analytics

NOVA Information Management School

Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

PRICE ELASTICITY IN MOTOR INSURANCE

by

Miguel Ângelo Nunes Lince

Dissertation presented as partial requirement for obtaining the Master's degree in Advanced Analytics, with a Specialization in Data Science

Supervisor: Roberto Henriques

November 2023

DEDICATION

To my mother, hope to make you proud.

ACKNOWLEDGEMENTS

First, I would like to express my gratitude to my girlfriend Claire for all the support and patience during the past 2 years.

Secondly, I would like to thank Generali Portugal for the funding of my Master program and their continuous contribution to my personal and professional development, specifically to my former Director João Gonçalves, who gave me the opportunity to achieve this remarkable milestone.

Last, but not least, I would like to thank my supervisor Roberto Henriques for the guidance during my research.

ABSTRACT

Demand-based pricing allows insurance companies to improve their profitability by setting their prices accordantly with customer's willingness to pay. To employ such price strategy a reliable estimation of customers' price elasticity is paramount. This dissertation explores the construction of price elasticity functions by applying a causal inference framework which permits reducing the risks of extrapolation of price changes effects on an observational study. The causal inference framework presented leverages from a matching technique which restricts the inference of alternative price changes effect on churn from comparable customers only. The price elasticity functions were estimated by policy degree of coverage since we believe that price sensitivity between Third-party liability (TPL) and Own damage (OD) policies are very distinct. As result, we were able to draw the price elasticity functions for specific subpopulations of customers based on tenure, premium and the existence of claims in the past year.

KEYWORDS

Casual inference; Price elasticity; Churn prediction; Machine Learning; Motor insurance; Pricing optimization

Sustainable Development Goals (SGD):

INDEX

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS AND ACRONYMS

TPL	Third-party Liability
OD	Own Damage
MDM	Mahalanobis Distance Metric
PSM	Propensity Score Metric
SMD	Standard Mean Difference
GPS	Generalized Propensity Score
GLM	Generalized Linear Model
GAM	Generalized Addictive Model
GBM	Gradient Boosting Machine
ASAM	Average Standardized Absolute Mean
SMD	Standard Mean difference

1. INTRODUCTION

Historically, insurance companies have applied a cost-based pricing strategy when producing their tariffs combining the estimation of frequency and severity to estimate the cost incurred in future claims from contracting the customer, known as the pure premium, plus certain loadings like operational costs and margin which are defined by the company (Antonio & Valdez, 2012). Due to market competitiveness, companies are willing to lose a significant portion of their margins to attract new customers by applying considerable discounts in the first year. To compensate for that, existing policies have their premiums increased every renewal, intending to reduce the deviation between the first-year premium and the company's commercial tariff (Guy Thomas, 2012).

Suppose the proposed price change at renewal is excessive. In that case, customers might search the market for a cheaper alternative and consequently cancel their policy, not giving time for the company to recover the margin lost and negatively impacting its profitability (Guelman & Guillén, 2014). On the other hand, demand-based pricing aims to infer the optimal price to offer each customer, subject to their characteristics and price sensitivity. According to Guven and McPhail (2013), price sensitivity measures the impact on demand for a given change in price, ranging from high sensitivity (elastic) to low sensitivity (inelastic). Demand-based pricing allows companies to maximize profitability, controlling customer retention, by setting the price according to customers' willingness to pay (Calabrese & De Francesco, 2014). Moreover, adding of market competitiveness data is a standard approach in demand-based pricing (Guven & McPhail, 2013; Guelman & Guillén, 2014; Spedicato, 2021; Verschuren, 2022).

In practice, estimate customers price sensitivity means that we are interested in the impact of price changes in customer retention. Causal inference aims to study the impact of interventions, i.e., treatments, on an outcome. That is, contrary to traditional statistical methods which measure the association between covariates, causal inference is related with the inference of not only the likelihood of the event but also the impact of changes in treatments (Pearl, 2010). If we define different price changes as treatments, we are looking to estimate what is the effect on customer retention from applying different treatments, i.e., price change levels.

Randomized controlled experiments (RCT) are the standard approach for causal effects estimation since it ensures that treatment assignment will not be influenced by any other factors (Austin, 2011). Stuart (2010) claims that any study carried with the intent of estimate causal effects have two main stages, the design and analysis of the outcome. Matching methods are essential to design an observational study which maintain the benefits from randomization, since it helps to identify areas of reduced overlap in covariates distributions and where without such techniques treatment effects estimates would highly depend on extrapolation.

Insurance companies store historical data about renewal price offers for existing customers, composed of a change in the price and the respective acceptance or not of such proposal. Since most insurance companies follow cost-based pricing, those price changes are most likely influenced by the evolution of customers' risk characteristics, making their assignment deterministic. The issue is that if price changes are correlated with customer risk, their effect is not comparable among different customer risk profiles since the customers who receive those changes are very different in their covariates.

Guelman and Guillén (2014) propose a causal inference approach based on Rubin's causal effects model (Rosenbaum & Rubin, 1983), to infer the causal effect of price change in customer churn. This approach allows for a more robust estimation of the effect of alternative price changes on customer churn, by restricting the inference only to similar customers across different treatments (price changes). In other words, to infer the churn probability of a customer for a certain range of alternative price changes, which were not observed, one needs to match that customer with similar ones who received that price change and reduce the risk of extrapolating such effects.

The selection of a matching method usually involves the choice of a distance measure to calculate similarity, a matching algorithm, and the structure of the match. Mahalanobis distance (MDM) and propensity score (PSM) are the main distances researchers use. MDM is appropriate for multivariate matching when the matching covariates are few and follow a normal distribution, otherwise, the matching procedure will result in a small number of matches and in more considerable bias (Guelman & Guillén, 2014; Sekhon, 2008; Stuart, 2010). On the other hand, PSM allows to perform matching only using a scalar, the propensity score, instead matching the subjects on their entire set of covariates (Rubin, 1997; Guelman & Guillen, 2014; Stuart, 2010). The propensity score estimates the likelihood of treatment assignment conditioned on the subject's covariates. The logic behind it is that two subjects with similar covariates will have similar treatment assignment probability (Guelman & Guillén, 2014; Stuart, 2010). A downside of PSM is that even though matched subjects have similar propensity scores, a significant difference in the covariates can exist. A more robust alternative might be to combine both, i.e., incorporating the propensity score as a matching covariate and allowing the matches only within a certain radius (caliper) of the covariates (Guelman & Guillén, 2014; Rubin and Thomas, 2000; Stuart, 2010). To select the caliper widths Rosenbaum and Rubin (1985) argue that since the logit of the propensity score is likely to follow a normal distribution, bias can be reduced by using proportions of standard deviations of the propensity score. Austin (2011) suggests that researchers use a caliper width of approximately 0.2 of the pooled standard deviation of the logit of the propensity score, since it reduced the mean difference of covariates in several studies.

After selecting of the distance measure, choice a matching algorithm is needed. Greedy matching and optimal pair matching are the algorithms commonly used. The greedy algorithm matches each subject in the treatment group with his most similar match from the subjects still available from the control group, not accounting for the order of the matches. Optimal matching, on the other hand, contrary to the greedy algorithm, aims to optimize the matches by minimizing the overall distance between all matches (Rosenbaum, 1989; Guelman & Guillén, 2014; Stuart, 2010). Although optimal matching can produce better overall matched pairs, both algorithms are similarly effective in producing balanced samples (Gu & Rosenbaum, 1993). Lastly, a common matching structure used is 1:1 pair matching, where 1 treatment subject is paired with 1 control. One drawback of 1:1 pair matching is that a relevant number of similar subjects may be discarded. An alternative is using full matching, where at least one treatment subject is matched with at least one control subject, maximizing the usage data available (Stuart, 2010; Guelman & Guillen, 2014).

The goal of matching is to produce balanced samples, in the sense that, the covariates of the matched sample will present a more similar distribution between treated and control groups compared with the original one. Austin (2011) outlines several methods, in the context of 1:1 pair matching, to evaluate the quality of such matching procedures. One of them is to compare means of continuous covariates through the usage of standardize differences, and according to the author, such metric is

not influenced by sample size and permits the comparison of covariates in different units of measure. The standardized mean difference (SMD) can be defined as

$$d = \frac{(\bar{x}_{treatment} - \bar{x}_{control})}{\sqrt{\frac{s^2_{treatment} + s^2_{control}}{2}}},$$

Equation 1 - Standard mean difference (SMD)

where $\bar{x}$ represents the sample mean and s^2 the sample variance.

The objective of this dissertation is to estimate the price elasticity of a motor insurance portfolio from a major Portuguese insurer by applying a causal inference framework inspired on the one proposed by Guelman and Guillén (2014). This dissertation adds to the existing literature by estimating customer sensitivity according to policy coverage, i.e., splitting the analysis between third-party liability and own damage coverages. Moreover, we demonstrate that using Euclidean distance for the matching presents a benefit over Mahalanobis distance, by being more efficient, producing similarly balanced samples in a significant smaller fraction of the time, especially for large sample sizes.

The remainder of this document presents in Section 2 the literature review on the application of causal inference in motor insurance and details about the framework proposed in this dissertation. Section 3 presents the elements developed to apply the framework detailed in Section 2 and the results obtained.

2. CAUSAL INFERENCE FRAMEWORK

Guelman and Guillén (2014) propose a causal inference approach with propensity scores, to estimate price elasticity in motor insurance, applied to 320,000 policy renewals from a Canadian direct insurer, based on Rubin's causal effects model (Rosenbaum & Rubin, 1983). Treatment was defined as the percentage price change which customers were exposed to, discretized into 5 ordered levels, ranging from -20% to 40% price change.

Figure 1 presents the causal inference framework proposed by Guelman and Guillén (2014) where the goal is to obtain estimates of churn probabilities for each customer.

Figure 1 - Diagram of Guelman and Guillén's causal inference framework

Guelman and Guillén's (2014) causal inference process starts by fitting a churn model for each treatment, i.e., discretized price change level, only with customers who received that treatment to estimate their churn probability. After, propensity score models are fitted to estimate treatment assignment probability of each customer for each treatment level dichotomy (j, k), (for any $j \neq k$). Having the propensity score the matching algorithm can be applied to identify similar subjects between each treatment level dichotomy (j, k) and impute the churn estimates to those customers for price change levels which were not observed. So, for the subject who received price change level *j*, the churn probability estimated at stage 1 of price change *k* is imputed. The logic of this is that, since the effect of alternative price change levels was not observed our best estimate is inferring from another customer who received it and as similar covariates. The last stage is the development a global churn model using the churn estimates of the observed outcomes and the imputed ones to estimate the churn probability under all treatment levels, subject to the matches found. This means that each customer will be repeated in the global dataset for each treatment level that the matching procedure allowed, with whether the churn estimate from the observed treatment level or the imputed one.

Verschuren (2022) adopted the work of Guelman and Guillén (2014) and applied it to a Dutch motor insurance portfolio. Moreover, the author introduces the usage of a continuous treatment approach based on Generalized Propensity Score (GPS) (Hirano & Imbens, 2004), allowing the selection of the exact optimal price change value instead of the optimal price change range as in the discrete approach. Additionally, Verschuren (2022) included competitiveness data by incorporating the relativity of the price offered to the competitors offers.

To fit the propensity score models, Guelman and Guillén (2014) used Gradient Boosting Models (GBM) (Friedman, 2002), whereas Verschuren (2022) adopts a more flexible and faster extension of GBM called Extreme Gradient Boosting (XGBoost) (Chen & Guestrin, 2016). The authors argue that GBMs are more suited to fit these propensity models compared with traditional GLM because it allows for a faster training process, since GBM possesses native feature selection and its able to detect non-linear

effects and interactions, whereas using GLM, one needs to make sure the model is well specified. Contrary to usual classification tasks, when estimating the propensity score, the authors didn't choose model hyperparameters to minimize prediction error, but instead to improve covariates balance between treatment and control groups, by minimizing the Average Standardized Absolute Mean difference in the covariates (ASAM). The ASAM can be calculated by averaging the absolute different between the mean of the treatment group and the mean of the control group, weighted by the odds of assignment to the treatment group, divided by the standard deviation of the treatment group, across all covariates.

The ASAM for one covariate can be defined as:

$$ASAM = \frac{\bar{x}_{treatment} - \bar{x}_{control} \times \omega}{s_{treatment}}$$

Equation 2 - Average Standardized Absolute Mean

Where $\omega = \hat{\pi}(x)/(1 - \hat{\pi}(x))$, the odds that subject with covariates x will be assigned to treatment.

Verschuren (2022) compares the balance achieved after matching using different algorithms to estimate the propensity score, namely random forests, neural networks, and Generalized Additive Models (GAM) (Hastie and Tibshirani, 1990), and concludes that XGBoost was the algorithm that improved balance the most. Moreover, balance was significantly higher using all the other algorithms rather than GAM, suggesting that more complex and non-linear algorithms are necessary to better estimate treatment assignment.

Analyzing the distributions of the estimated propensity score for each treatment dichotomy the authors drawn two interesting conclusions:

1. The overlap between distributions is more noticeable for price changes that are closer, compared to the ones further apart, suggesting that it's easier to find similar customers when examining treatment levels that represent price changes that are closer.
2. For most treatment dichotomies there are subjects on opposite extremes of the propensity score distribution, leading to the conclusion that exists a combination of covariates which are not present for both groups.

After applying the matching algorithm Guelman and Guillén (2014) concluded that the highest balance was obtain by optimal pair matching, using Mahalanobis distance, and adding the propensity score within calipers to the remaining matching covariates, where the width of the calipers was selected to obtain a good trade-off between balance and number of matches. On the other hand, Verschuren (2022) chose to multiple impute the counterfactual responses by sampling with replacement from the 10 closest subjects using absolute distance on the propensity score and hoping to better account for the uncertainty of the response compared with Guelman and Gillen (2014).

Concerning the global models Guelman and Guillén (2014) fitted a GAM with cubic splines for all continuous covariates and interaction between rate change and specific covariates, allowing the detection of non-linear patterns with the churn rates. Verschuren (2022), on the other hand, opted to fit a GLM with LASSO penalization (Tibshirani, 1996) for the discrete approach and XGBoost for the continuous one. The reason for the author to use XGBoost to fit the global model for the continuous

approach is that, since the model has no causal interpretation due to the fixed GPS estimates, XGBoost allows for flexible and accurate estimation of the churn probabilities (Hirano & Imbens, 2004).

Having obtained the global churn probabilities for all the portfolio, Guelman and Guillén (2014) concluded that customers with accidents during the last period tend to have a lower sensitivity compared with customers with no accidents since they already expect the price increase and probably is harder to find a cheaper alternative in the market. Moreover, younger customers with higher premiums and with lower tenure with the company have higher price sensitivity. Verschuren (2022) identifies that the probability of churn increases with competitiveness, i.e., customers are more likely to churn if their policy premium is already cheap in the market. Furthermore, the author observes that higher rate increases and customers with lesser and higher risk classes are also more likely to lapse their policies.

In this dissertation the estimation of the price sensitivity function will be inspired on proposal of Guelman and Guillén (2014) detailed earlier, with the following differences:

1. Distinct price elasticity function for Third-party Liability (TPL) policies, and Own Damage (OD) policies. TPL covers the insured customer against damage caused on third-party property, whereas OD covers third-party property and insured customer property. Since TPL is mandatory by law in Portugal, a higher churn rate is expected, and consequently higher price sensitivity compared with policies which have OD cover contracted (which are optional). Based on that, we believe that distinct price sensitivity functions should be estimated for those groups.
2. Consider 6 treatment levels instead of 5 and allow for different number of policies in each treatment. This was due to the distribution of price increases in the Portuguese insurer portfolio and low representation of policies outside of certain increase intervals. For example, if a price optimization would be developed discretizing price change into quintiles as in Guelman and Guillén (2014), the maximum increase would be 6%, completely infeasible policies with claims. Considering 5 categories with the same boundaries of Guelman and Guillén (2014), the minimum price change would be 0%, since we can only use the medians of each treatment, and only 0.81% of policies would belong to the price change level [0.1, 0.2). Therefore, for this dissertation, price change was discretized in 6 treatments with different number of policies. This means that by increasing the number of treatments to 6, we will have a total of 15 (T*(T-1)/2) propensity score models fitted, 5 more models compared with Guelman and Guillén (2014).
3. Fit the churn models and propensity score models using an innovative version of GBM, called LightGBM (Ke, G. et al., 2017). LightGBM improves memory consumption and presents a higher computational speed than XGBoost and GBM, achieving similar performance.
4. Addition of relevant covariates. The data used to train the models includes more relevant explanatory features. Therefore, we believe that it will improve the propensity score estimation, consequently achieving a better balance and a better estimation of global churn, leading to a finer estimation of customers' price sensitivity and superior results when performing pricing optimization.

3. EMPIRICAL STUDY

In this chapter, the data used will be outlined in terms of scope, feature domains and initial figures of portfolio composition. After the results of the propensity score models, the evaluation of the matching procedure and global churn model assessment will be demonstrated.

3.1. THE DATA

To train the different models, data was collected from a Portuguese insurer concerning motor light vehicle insurance policies, up to renewal in 2020 and 2021. The insurer sends renewal price offers to the customers 30 days before renewal date. Considering the risk characteristics, policy level of discounts, and customer segment, the definitive price change for the renewal is defined. So, covariates concerning policy, vehicle, customer details and the respective price change offered were collected for each policy in the dataset. Table 1 details some examples of features extracted by each domain. Moreover, the renewal outcome was extracted 30 days after the renewal date, a binary variable with indicates if the policy churned (1) or not (0).

Domain	Features
Customer data	Age of the customer Age of driver's license Municipality Total of annual premiums Number of policies
Policy data	Policy annual premium Annual commissions Price change Tenure of the policy Payment Frequency Number of claims in the past year Bonus Malus
Vehicle data	Age of the vehicle Brand Cylinder capacity Engine power Gross weight

Table 1 - Features extracted by domain (not exhaustive).

The dataset contains over 1.4 million motor light vehicle insurance policies, with 163.211 (11,63%) canceled by the customers. Table 2 shows that most of the policies in the portfolio (79%) are TPL only policies with a churn rate of 12,4%, whilst for OD policies a churn rate of 9% was observed.

Coverage	Nº Observations	Proportion Observations (%)	Churn rate (%)
Third-party liability (TPL)	1 105 703	78.81	12.37
Own damage (OD)	297 310	21.19	8.90
Total	**1 403 013**	**100.00**	**11.63**

Table 2 - Distribution of policies and churn rate by coverage.

Figure 2 shows the distribution of price changes on the left and churn proportions by price change on the right. Analyzing the distribution of price changes, we observe that rare price changes occur as in Guelman and Guillén (2014) and Verschuren (2022), with most of the observations concentrated between -8% and 12%. Moreover, churn proportions show that the inflection point observed in Verschuren (2022) close to 0% price change is also observed (not observed in Guelman and Guillén (2014)). Verschuren (2022) argues that the inflection point exists because customers are likely to react to the fact that the price is changed somewhat to the size of the change. In fact, the author claims that

insurance companies commonly maintain the same price instead of applying small price reductions to not raise customers' price awareness. On top of the inflection point at 0%, we can also observe an inflection point around 25%.

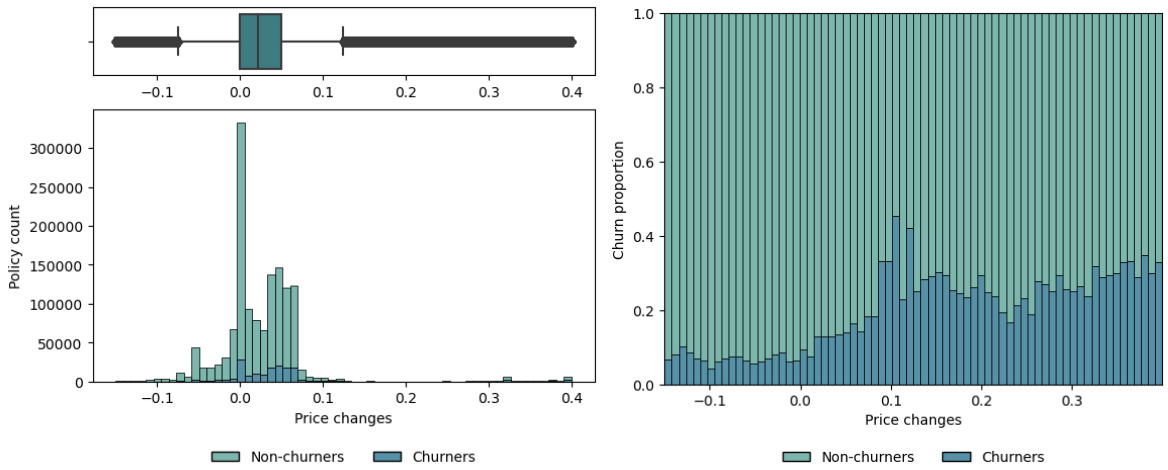

Figure 2 - Frequency of price changes on the left and churn proportion by price change on the right.

Figure 3 shows the proportion of policies with and without claims by price change on the left and a boxplot of price change for split by coverage for policies with and without claims. We can see that more than 80% of policies with increases higher than 25% have at least one claim in the previous annuity, suggesting that the effect of price changes on churn for policies without claims is observed for price changes lower than 25% and of policies with claims above that point. Indeed, the boxplot shows that policies with claims have a median price change of ~2% compared with the median increase of ~33% of policies with claims. Interestingly, we can also observe that policies with OD have a slightly lower median increase for policies without claims and even policies with claims receive lower increases compared with TPL policies.

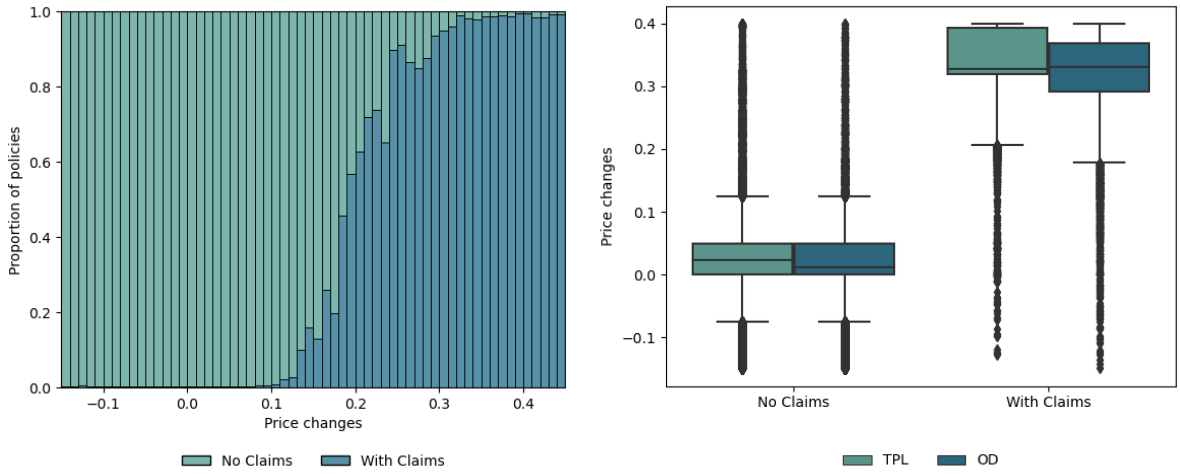

Figure 3 - Proportion of policies with claims on the left and distribution of price change split by coverage for policies with or no claims in the previous year.

The following tables show the distribution of policies, the median increase and the churn rate by price change interval applied to the TPL and OD portfolios respectively. Some insights can be drawn from analyzing Tables 3 and 4, for example, around 28% of the portfolio of renewals receive a price decrease

compared with the previous year, this may have to do with retention discounts or due to a decrease in customer's risk. In line to what we saw in Figure 1 frequently the insurer proposes a 0% price change to the customers on renewal, around 50% of the times. An interesting fact is that for TPL, it seems that when customers receive a price change higher than 25%, they churn less than when a price change between 8% and 25% is offered. This effect does not seem to happen for OD policies since prince change appears to have a monotonic increasing impact on churn, i.e., the higher the price change, the higher the churn rate. After data preparation, missing values dropped and the price increase discretized in 6 treatments Table 3 and 4 present proportions of observations, median price change and churn rate in each treatment for TPL and OD, respectively.

Treatment	Proportion Observations (%)	Median (%)	Churn rate (%)
(-inf, -0.01]	11.12	-4.39	7.52
(-0.01, 0.01]	29.22	0.00	8.76
(0.01, 0.05]	38.94	3.97	13.11
(0.05, 0.08]	16.57	5.35	15.70
(0.08, 0.25]	2.12	9.70	32.03
(0.25, inf]	2.00	34.13	30.18
Total	**100.00**	**2.39**	**12.39**

Table 3 - Distribution of policies, median increase, and churn rate by treatment for TPL for policies without claims.

Treatment	Proportion Observations (%)	Median (%)	Churn rate (%)
(-inf, -0.01]	17.25	-4.94	5.49
(-0.01, 0.01]	27.96	0.00	7.15
(0.01, 0.05]	30.87	2.27	8.27
(0.05, 0.08]	19.43	6.48	11.57
(0.08, 0.25]	2.56	10.26	22.32
(0.25, inf]	1.91	34.14	28.87
Total	**100.00**	**1.34**	**8.87**

Table 4 - Distribution of policies, median increase, and churn rate by treatment for OD for policies without claims.

Analyzing Tables 3 and 4, we can observe that churn rate increases with the increase in price change, but as Guelman and Guillen concluded, the treatment (0.25, inf], which is mainly composed of policies with claims, churn rate is lower compared with policies which receive price change between 8% and 25%, for TPL. In OD this effect is not observed, as the churn rate increases monotonically with the increase of price change.

3.2. TREATMENT CHURN MODELS

Having defined the levels of price change, individual models were fitted using LGBM to each of the treatments, only with subjects belonging to that treatment, to estimate churn propensity for TPL and OD. Table 5 shows the hyperparameter space used to search for the combination of parameters which maximizes the average precision (AP) of model predictions, with a reasonable amount of overfitting. The search was conducted with a random search cross-validation, with 100 iterations. The choice of maximizing average was because it evaluates the model along all the precision-recall curve without the need to define a probability threshold. AP can be defined as the mean of precision achieved at each threshold, weighted by the increase in recall from the previous threshold,

$$AP = \sum_n (R_n - R_{n-1}) P_n$$

Equation 3 - Average precision formula (source: Scikit-learn)

where R_n and P_n are the recall and the precision at threshold n, respectively.

After selecting the best hyperparameter combination, a model was fitted to the entire subset of data to evaluate the model performance. Models were evaluated by:

- Ability to order customers. Estimated probabilities were discretized in 10 deciles. A lift curve was plotted to evaluate how well the model ordered the customers based on the average observed churn rate achieved in each decile. Figure 4 shows the lift curve plotted of the model fitted for the treatment (-inf, -0.01] of TPL. Being decile 1 the most propense to churn and decile 10 having the least propense, we can see that the model can order the customers correctly, showing a monotonic increase in churn rate. The customers on decile 1 present a 3.3 times higher churn rate compared to overall churn rate of the subsample. More details of the remaining models can be found in Annexes 1 and 2.
- Identification of the number of churners. By summing the propensities of the models and comparing it to the true number of customers who cancel their policies, we could evaluate how close the predictions were. Table 5 shows true and predicted number churners by price change level for TPL and OD, respectively. One can conclude that the predicted cancelations are close to the observed ones, existing a small overestimation of the total number of cancelations for both models (11 customers for TPL and 12 customers for OD).

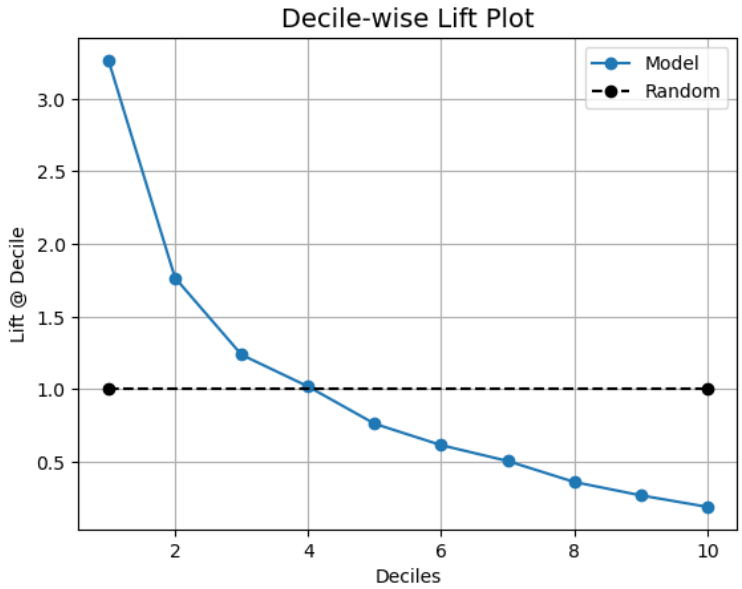

Figure 4 - Lift curve for treatment (-inf, -0.01] of OD, where decile 1 is the most propense and decile 10 the least. Analyzing the curve, one can conclude that the model predictions are correctly ordered presenting a monotonic increasing observed churn rate by decile.

	TPL		OD	
Treatment	**Observed # cancelations**	**Predicted # cancelations**	**Observed # cancelations**	**Predicted # cancelations**
(-inf, -0.01]	8 993	8 995	2 695	2 694
(-0.01, 0.01]	9 984	9 983	5 687	5 690
(0.01, 0.05]	27 506	27 507	7 263	7 268
(0.05, 0.08]	54 864	54 868	6 396	6 395
(0.08, 0.25]	27 965	27 969	1 625	1 630
(0.25, inf]	7 319	7 320	1 567	1 568
Total	**136 631**	**136 642**	**25 233**	**25 245**

Table 5 - Observed and predicted number of cancelations for TPL and OD.

3.3. PROPENSITY SCORE MODELS

A propensity model was trained for each of the 15 dichotomies (j, k), to obtain the log-odds of each subject to belong to the treatment group compared with the control group, where j is the price change level with fewer observations representing the treatment and k as the control. The best hyperparameter combination of the LGBM was selected to minimize the Average Standardized Absolute Mean difference (ASAM) of the covariates as in Guelman and Guillén (2014) and Verschuren (2022), this means, the combination of hyperparameters was selected to maximize the overlap of the propensity scores between treatment and control, thus maximizing common support and consequently the number of matches. In Table 6, the search space of the grid search is presented.

Parameter	Values
Number of estimators	[50,100,150,200]
Max depth	[3,6,9,12]
Reg. alpha	[0,1,3,6,9]
Learning rate	[0.01,0.1,0.5,0.8]
Number of leaves	[10,20,30]
Min. child samples	[100,200,500]

Table 6 - Grid search parameters space to fit the LGBM propensity models.

After fitting the models with the best hyperparameters, analyzing the propensity score distributions, a higher common support (overlap) exists in treatments that are closer compared with the ones further apart, like what Guelman and Guillén (2014) observed. This means that it will be easier to identify similar subjects in the former case compared with the latter. In Figure 5, we compare the distribution of the log-odds of the propensity scores for dichotomy (1,2) on the left and (1,4) on the right and we conclude that a higher overlap exists between closer dichotomies.

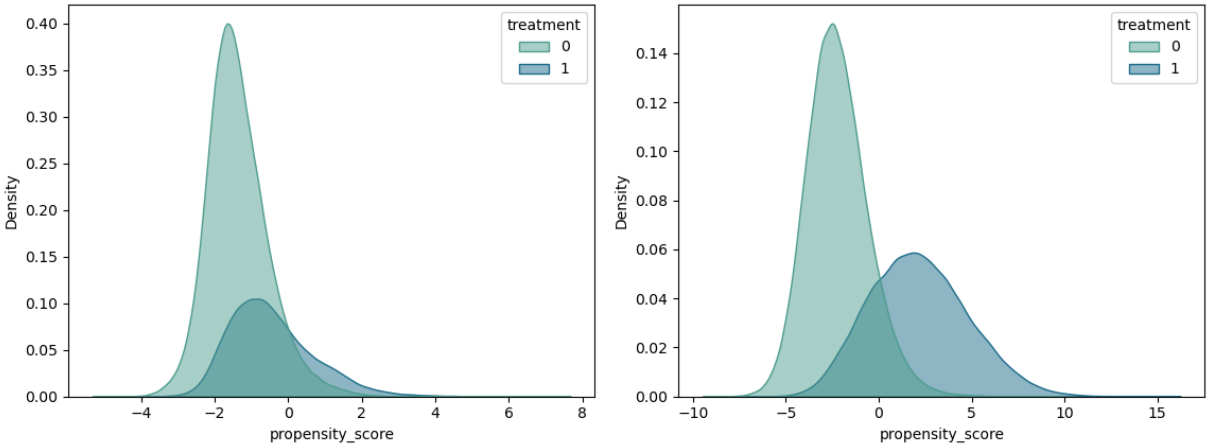

Figure 5 - Propensity score logit distributions for dichotomy (1,2) on the left and (1,4) on the right for TP. One can conclude that closer the treatments higher the overlap.

3.4. MATCHING

After fitting the propensity score models, a matching algorithm was applied, based on Nearest Neighbors (Goldberger, J. et. al, 2005), to each of the 15 treatment dichotomies. The structure of the match chosen was 1:1 pair matching with replacement, using the propensity score as matching covariate and considering a radius (caliper) of 30% of standard deviation. Other structures were tested in terms of treatment control ratio, allowing to match more than 1 control for each treatment unit, and variations of the radius in percentage of standard deviation. One could conclude that increasing the number of controls and the radius would result and more matches but poorer balance between treated and untreated samples. The final structure was chosen considering this trade-off between balance and sample size as in Guelman and Guillén (2014).

To evaluate the quality of the match, in this dissertation we have followed the proposal of Austin (2011) to calculate the Standard Mean Difference (SMD) between treatment and control groups before and after matching for all the matching covariates. Figure 5 shows the average SMD for TPL, before and after matching using both Mahalanobis and Euclidean distances. Analyzing the results, one concludes that after matching the samples have improved balance significantly with all the dichotomies presenting an average SMD lower than 0.1 – according to Austin (2011) the threshold to indicate negligible difference. Once more, average SMD before matching show us that in general dichotomies further apart from one other, higher is the unbalance in the covariates, agreeing with what was observed in the distributions of the propensity scores. In annexes 3 and 4, one can find the comparison of performance and balance of all treatment dichotomies between Euclidean and Mahalanobis distances for TPL and OD respectively.

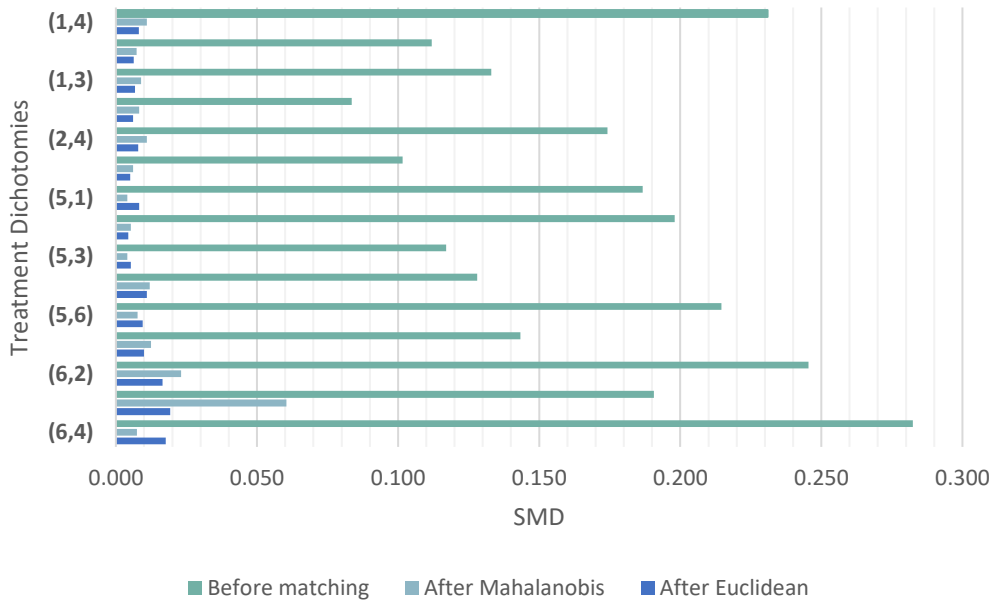

Figure 6 - Average SMD for each dichotomy before and after matching using Euclidean and Mahalanobis as matching distances. Balance is largely achieved after matching using both distance measures.

We concluded that samples were equivalently balanced using Euclidean or Mahalanobis distances, but Euclidean distance produced larger matched samples and presented presented a more efficient computation (up to 125 times faster) over Mahalanobis distance. Figure 5 shows a comparison of wall time in seconds between executing scikit-learn's Nearest Neighbors with both distances to find the closest control for each treated subject depending on sample size. Specifically with sample size of 732 258 subjects, Mahalanobis distance took approximately 1 hour and 57 minutes to execute, achieving an average SMD of 0.008, resulting in 249 154 matches, whereas Euclidean distance execution took only 56 seconds, with an average SMD equal to 0.006 and was able to match 269 165 subjects (refer to annexes 1 and 2 for more details).

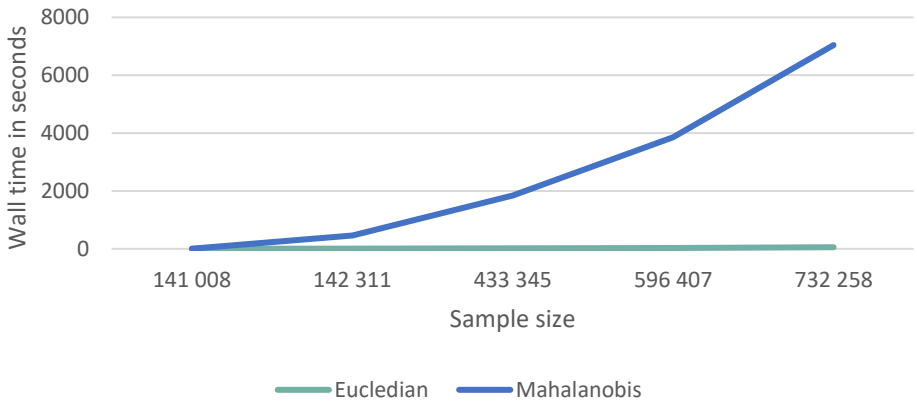

Figure 7 - Wall time comparison in seconds of Nearest Neighbors execution between Euclidean and Mahalanobis distances. Wall increases exponentially given the increase of sample size for Mahalanobis distance, whereas with Euclidean the difference in wall time is neglectable.

3.5. GLOBAL CHURN MODELS

After performing the matching procedure and the imputation of the treatment churn probabilities for unobserved price changes, the global churn model was built using a Logistic Regression (LR) to estimate the churn probabilities. The model as the following theoretical specification,

$$\hat{y} = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \varepsilon$$

Equation 4 – Theoretical model specification of the Logistic Regression.

Where $\hat{y}$ is the estimated churn prediction based on the observed and imputed responses.

Moreover, to improve nonlinearity, each one of the numerical covariates was discretized by building a Decision Tree (Breiman, L., 2017) to predict the observed and imputed churn probabilities using only that covariate, minimizing the mean squared error (MSE). Additionally, interactions between the covariates were also created. The selection of the covariates of the model was done in two stages:

1. Prior to the model fitting, univariate feature selection was applied based on F1-test and Mutual Information (Kraskov et al., 2004), since using both methods one can capture any kind of association between two numerical random variables.
2. Covariates levels were further aggregated or removed from the model specification by analyzing the p-values, allowing for a maximum p-value of 5%.

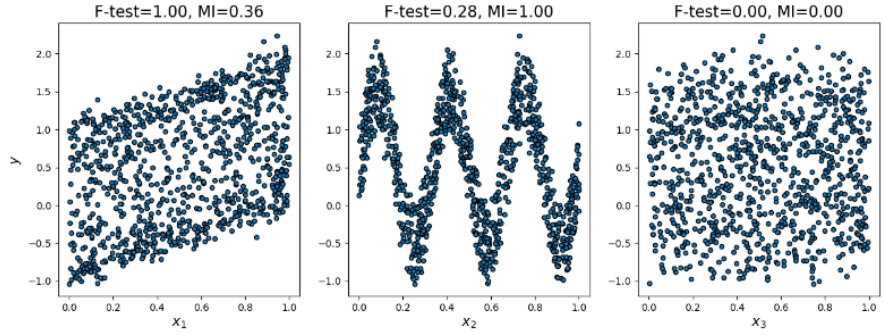

Figure 8 - Mutual Information and F-test comparison. F-test is only able to detect that y is dependent on x1, since there is a linear dependency, but is not able to detect the non-linear dependency of x2. Mutual Information on the other hand identifies that y and x2 are strongly dependent (source: Scikit-learn).

The choice of the best model specification was decided based on a validation sample, not used for model fitting, composed with 30% of the renewals, evaluating the goodness-of-fit and comparing the number of predicted cancelations and average churn probability estimated by the model.

Metrics	TPL	OD
Predicted cancellations	120 772	22 392
Estimation of predicted cancellations	120 769	22 489
Predicted churn rate	13.0%	8.9%
Estimation of predicted churn rate	13.0%	8.9%

Table 7 – Predicted and estimated predicted number of cancellations churn rate of TPL and OD global models.

After selection the best model specifications, the price elasticity of subpopulations of customers were evaluated. Figure 9 shows that in general price elasticity increases with price change. Moreover, recent customers show higher price sensitivity compared older ones, and customers who pay higher

premiums are also more sensitive to price changes compared to lower premiums. Additionally, customers with claims tend to be slightly more sensitive for smaller price changes but less when presented with a high price change compared with customers without claims in the previous year.

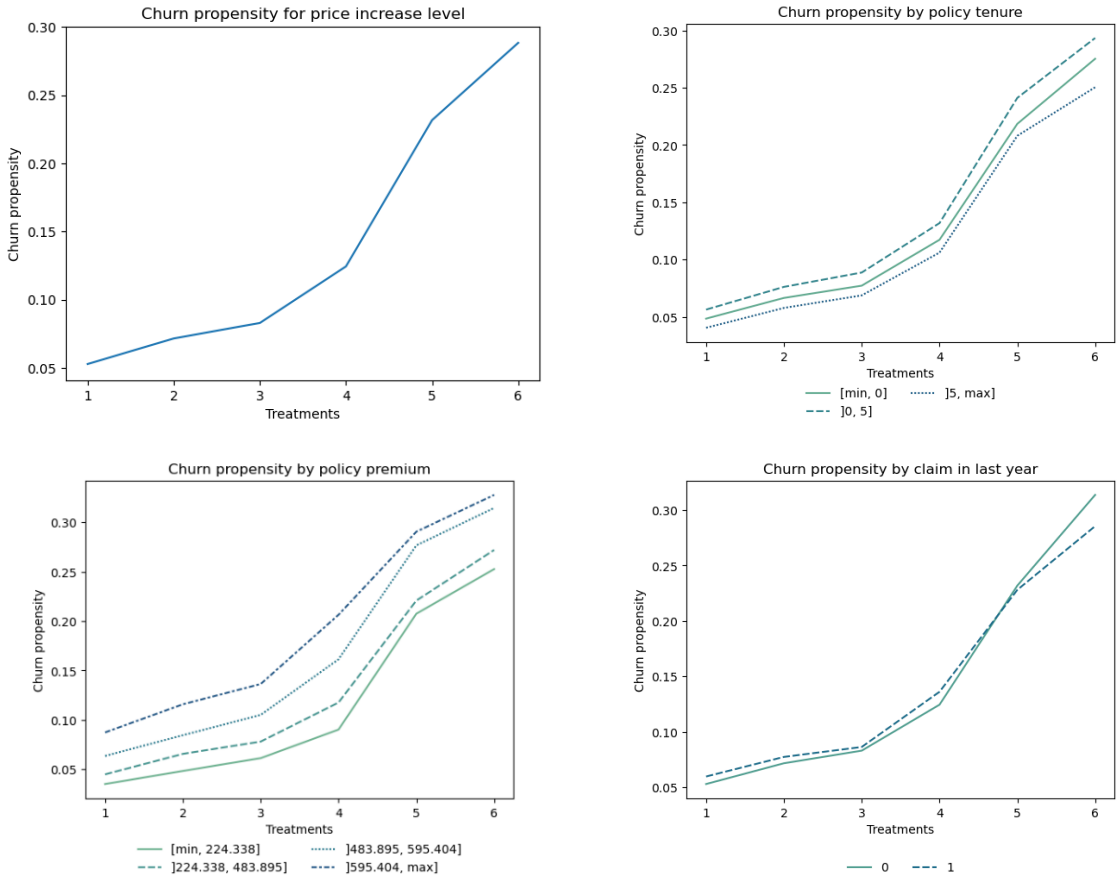

Figure 9 - Examples of estimated price elasticity functions of OD policies for each subpopulation of customers by computing the average churn rate for each price change level.

The global model churn predictions were then assessed on a holdout sample of renewals of 2022 containing over 860.000 policies. Figures 10 and 11 show that the models are overestimating the true proportion of churners in the holdout sample, this may also happen due to the decrease in the observed churn rate of 2022, for example for TPL the churn rate in 2022 was11.25% compared with 12.7% in 2020 and 2021 (12.7%), years used for model training.

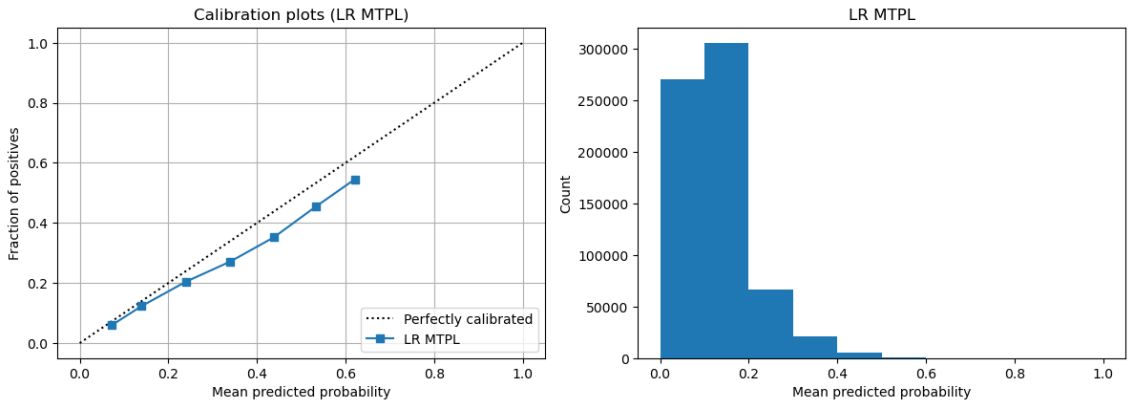

Figure 10 – Calibration plot for MPTL global models' predictions on the left and predictions distribution on the right.

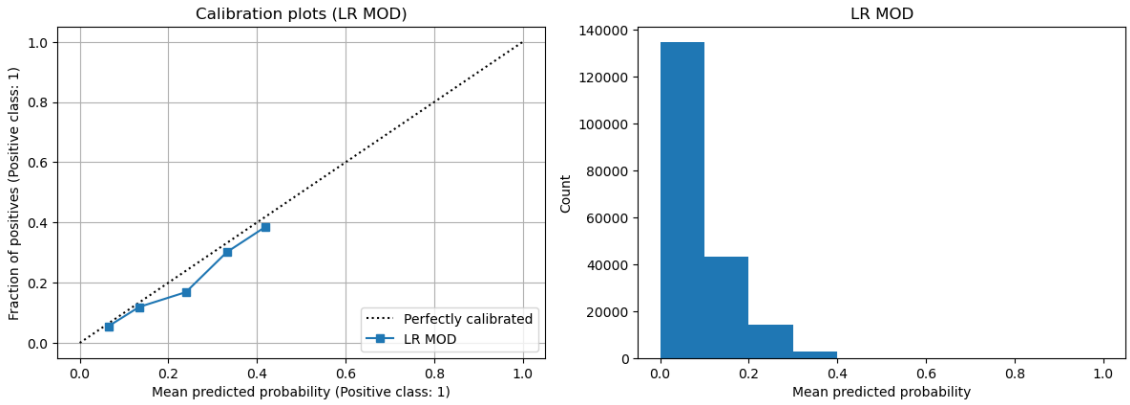

Figure 11 - Calibration plot for OD global models' predictions on the left and predictions distribution on the right.

Figures 12 shows the decile-wise lift plot for TPL on the left and OD on the right. The models, are correctly ordering the policies, showing a monotonic increase in the observed churn rate from the most propense deciles (1) to the least propense (10). Moreover, decile 1, the top 10% of probabilities, shows over 2 times more churn rate compared with the average churn rate of each sample.

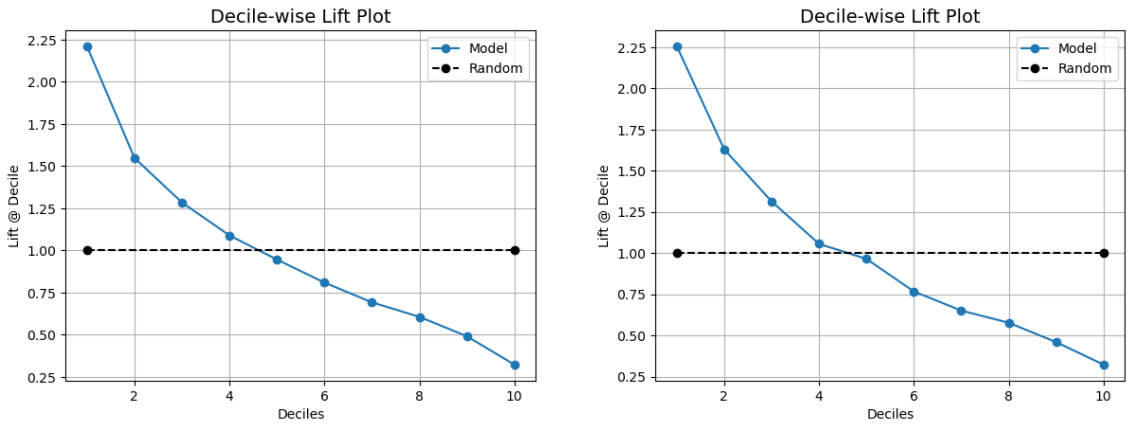

Figure 12 - Decile-wise lift plot for TPL (left) and OD (right) global models' predictions.

Table 7 shows a comparison between observed and predicted churn rate for both TPL and OD global churn models, which are in line with what was previously discussed when the calibration plots were analyzed. We can see that both models are slightly overestimating the true number of cancellations, partly due to a decrease in the company's churn rate in 2022 compared with 2020 and 2021.

Metrics	TPL	OD
Observed churn rate	11.3%	8.1%
Predicted churn rate	13.2%	9.7%

Table 8 – Comparison between observed and predicted churn rates for TPL and OD global models.

The performance of the models was also assessed by computing the precision, recall and ROC AUC. For ROC AUC the predicted probabilities were used and probability threshold, which maximizes F1-score on the training data, was defined for each model and used to calculate the model's precision and recall. Considering those thresholds, for TPL, from the 670.000 renewals, 189.000 (28%) of them were classified as churners and from those 36.581 were cancelled, representing a precision of 19.3% and a recall of 48.5%. For OD from 195.000 renewals, 22% were classified as churners, achieving a precision of 15.4% and a recall of 42%. The ability of the model to separate the true churners from non-churners was also evaluated using the Area Under the ROC curve (ROC AUC), with both models scoring around 67%.

Metrics	TPL	OD
Precision	19.3%	15.4%
Recall	48.5%	42%
ROC AUC	66.9%	67%

Table 9 - Precision, recall and ROC AUC for TPL and OD global models.

4. CONCLUSION

Insurance companies offer significant discounts to attract new customers, given the high competition in the insurance market, being forced to increase policy premiums in the following renewals to reduce the leakage to the commercial tariffs. Demand-based pricing helps insurance companies to define the optimal price change for each customer by considering his/her willingness to pay with the objective of maximizing profits for a set level of portfolio retention.

Estimate customer sensitivity to alternative price changes is not straightforward, since insurance companies have historically applied a risk-based pricing, making the alternative price changes not comparable between different customer risk segments. To solve this issue, Guelman and Guillén (2014) propose the application of a causal inference framework to determine the sensitivity function of an insurance portfolio, by restricting the analysis to similar customers only, which received alternative price changes, and like that avoid the risk of extrapolating price change effects on customer churn.

In this dissertation a causal inference model inspired on Guelman and Guillén's (2014) proposal was developed and applied to a major Portuguese insurance portfolio, composed by light duty vehicles renewing during the years of 2020 and 2021. Leveraging from a novel GBM algorithm we fitted individual churn models for each price change level and 15 propensity score models which it's estimates were used to find similar customers across different price change level dichotomies. To select the best hyperparameter combination for the individual churn models a random search with cross-validation was used to find the combination which maximized average precision (AP). For the propensity score models the average standardized absolute mean difference (ASAM) was used as evaluation metric, since we want to produce choose the combination which produced the greater overlap of the propensity score for each dichotomy.

Assessing covariate balance before matching lead us to conclude that indeed customers which receive different price changes present distinct characteristics, and it was confirmed when we analyzed the distribution of the propensity score logits and observed that there were observations with extreme values of the distributions and that higher overlap existed for closer dichotomies compared with the ones further apart. After applying the matching procedure, we concluded that greater covariate balanced was achieved compared with before the matching, proving that matching is a powerful tool to control bias in observation studies.

After producing the global sample with the observed and imputed probabilities a global churn model was fitted using Logistic Regression, discretizing continuous covariates, and allowing for interactions between them. We draw the price elasticity functions and conclude that higher price changes lead to a higher churn of the customers, unless the customer has claims in the last year, evidence also showed by Guelman and Guillén (2014). Moreover, more recent customers and customers who pay higher premiums are more price sensitive.

By assessing the model on a holdout sample from renewals of 2022 we concluded that the models were slightly overestimating the churn rate but still presenting a good ordering of the most propense customers to churn by the analysis of the decile-wise lift plots, where on top 10% of probabilities the models achieve 2 times more churn rate than the average churn rate of each sample.

5. LIMITATIONS AND RECOMMENDATIONS FOR FUTURE WORKS

In this chapter limitations and future improvements on the work developed will be outlined. The limitations are identified as factors that may have impacted the results and future improvements are described such that future research on the topic could achieve better results.

First, since this research was conducted during the years of 2022 and 2023 a choice was done to extract the most recent data available of the Portuguese insurer, as in data from the years of 2020 to 2022. In the year of 2020 the whole world faced a great humanitarian challenge with the prefoliation of the COVID-19 virus, which disrupted our life's and most of us were restricted to our homes. Because of that we believe that the results achieved may not reflect an accurate picture of company churn and consequently customers price elasticity in the years to come and the company should continue to update the models with the usage of more recent data. Moreover, in during the years of 2022 and 2023 we have witnessed to an exponential increase on inflation, reflected in the increase of costs for insurance companies and consequently policy premiums, which could have a significant impact on customer's price sensitivity.

The author chose to apply the research using Python, where causal inference modules are not as well documented and complete as in other programming languages. This fact led, for example, to the usage of only 1:1 pair matching with replacement, since it was not available any matching module in Python which allows different structures of matching as optimal matching or matching without replacement "out-of-box". Having that in mind, further structures should be explored in future work. Finally, we believe that possibly splitting the elasticity function between policies with and without claims could also be studied, since the distributions of price changes and companies' retention objectives can be very different between those groups.

6. REFERENCES

Antonio, K., & Valdez, E. A. (2012). Statistical concepts of a priori and a posteriori risk classification in insurance. AStA Advances in Statistical Analysis, 96(2), 187-224.

Ansel, J., Hong, H., & Jessie Li, A. (2018). OLS and 2SLS in randomized and conditionally randomized experiments. Jahrbücher für Nationalökonomie und Statistik, 238(3-4), 243-293.

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate behavioral research, 46(3), 399-424.

Bartlett, J. W. Covariate adjustment and estimation of mean response in randomised trials. Pharmaceutical statistics, 17(5):648–666, 2018.

Breiman, L. (2017). Classification and regression trees. Routledge.

Calabrese, A., & De Francesco, F. (2014). A pricing approach for service companies: service blueprint as a tool of demand-based pricing. Business Process Management Journal, 20(6), 906-921.

Cochran, W. G., & Rubin, D. B. (1973). Controlling bias in observational studies: A review. Sankhyā: The Indian Journal of Statistics, Series A, 417-446.

Goldberger, J., Hinton, G. E., Roweis, S., & Salakhutdinov, R. R. (2004). Neighbourhood components analysis. Advances in neural information processing systems, 17.

Gu, X. S., & Rosenbaum, P. R. (1993). Comparison of multivariate matching methods: Structures, distances, and algorithms. Journal of Computational and Graphical Statistics, 2(4), 405-420.

Guelman, L., & Guillén, M. (2014). A causal inference approach to measure price elasticity in automobile insurance. Expert Systems with Applications, 41(2), 387-396.

Guo, S., & Fraser, M. W. (2014). Propensity score analysis: Statistical methods and applications (Vol. 11). SAGE publications.

Guven, S., & McPhail, M. (2013). Beyond the cost model: Understanding price elasticity and its applications. In Casualty Actuarial Society E-Forum, Spring 2013.

Guy Thomas, R. (2012). Non-risk price discrimination in insurance: market outcomes and public policy. The Geneva Papers on Risk and Insurance-Issues and Practice, 37, 27-46.

Hastie, T., & Tibshirani, R. (1990). Exploring the nature of covariate effects in the proportional hazards model. Biometrics, 1005-1016.

Hirano, K., & Imbens, G. W. (2004). The propensity score with continuous treatments. Applied Bayesian modeling and causal inference from incomplete-data perspectives, 226164, 73-84.

Holland, P. W. (1986). Statistics and causal inference. Journal of the American statistical Association, 81(396), 945-960.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. Advances in neural information processing systems, 30.

Kraskov, A., Stögbauer, H., & Grassberger, P. (2004). Estimating mutual information. Physical review E, 69(6), 066138.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. Biometrika, 70(1), 41-55.

Rosenblum, M. and Van Der Laan, M. J. Simple, efficient estimators of treatment effects in randomized trials using generalized linear models to leverage baseline variables. The international journal of biostatistics, 6(1), 2010.

Rubin, D. B. (1979). Using multivariate matched sampling and regression adjustment to control bias in observational studies. Journal of the American Statistical Association, 74(366a), 318-328.

Rubin, D. B. (2001). Using propensity scores to help design observational studies: application to the tobacco litigation. Health Services and Outcomes Research Methodology, 2(3), 169-188.

Rubin, D. B., & Thomas, N. (2000). Combining propensity score matching with additional adjustments for prognostic covariates. Journal of the American Statistical Association, 95(450), 573-585.

Pearl, J. (2010). Causal inference. Causality: objectives and assessment, 39-58.

Spedicato, G. A., Dutang, C., & Petrini, L. (2018). Machine learning methods to perform pricing optimization. A comparison with standard GLMs. Variance, 12(1), 69-89.

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. Statistical science: a review journal of the Institute of Mathematical Statistics, 25(1), 1.

Verschuren, R. M. (2021). Customer Price Sensitivities in Competitive Automobile Insurance Markets. arXiv preprint arXiv:2101.08551.

Sekhon, J. S. (2008). Multivariate and propensity score matching software with automated balance optimization: the matching package for R. Journal of Statistical Software, Forthcoming.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological), 58(1), 267-288.

Tsiatis, A. A., Davidian, M., Zhang, M., and Lu, X. Covariate adjustment for two-sample treatment comparisons in randomized clinical trials: a principled yet flexible approach. Statistics in medicine, 27(23):4658–4677, 2008.

Yang, L. and Tsiatis, A. A. Efficiency study of estimators for a treatment effect in a pretest– posttest trial. The American Statistician, 55(4):314–321, 2001.

7. ANNEXES

Annex 1 – Lift of treatment churn models by propensity decile for TPL

Cumulative Lift	Treatments					
Decile	**1**	**2**	**3**	**4**	**5**	**6**
1	3.39	3.20	2.90	2.66	2.34	2.09
2	2.55	2.48	2.31	2.19	2.04	1.86
3	2.14	2.08	1.99	1.91	1.83	1.68
4	1.83	1.81	1.75	1.70	1.67	1.54
5	1.63	1.61	1.57	1.54	1.53	1.43
6	1.46	1.44	1.42	1.41	1.40	1.34
7	1.33	1.31	1.30	1.29	1.30	1.25
8	1.20	1.19	1.19	1.18	1.19	1.17
9	1.09	1.09	1.09	1.09	1.10	1.09
10	1.00	1.00	1.00	1.00	1.00	1.00

Annex 2 – Lift of treatment churn models by propensity decile for MOD

Cumulative Lift	Treatments					
Decile	**1**	**2**	**3**	**4**	**5**	**6**
1	3.30	3.27	3.02	3.25	2.77	2.32
2	2.53	2.48	2.40	2.50	2.27	1.97
3	2.09	2.08	2.04	2.12	1.98	1.79
4	1.82	1.80	1.79	1.85	1.77	1.62
5	1.62	1.60	1.59	1.63	1.59	1.47
6	1.46	1.44	1.44	1.46	1.45	1.35
7	1.32	1.30	1.31	1.32	1.33	1.26
8	1.19	1.19	1.19	1.20	1.21	1.17
9	1.10	1.09	1.09	1.10	1.10	1.09
10	1.00	1.00	1.00	1.00	1.00	1.00

Annex 3 - Summary of matching procedure with Euclidean and Mahalanobis distances for TPL

Dichotomy	Euclidean			Mahalanobis		
	# Samples matched	**Average SMD**	**Wall time (seconds)**	**# Samples matched**	**Average SMD**	**Wall time (seconds)**
(1,3)	192 676	0.007	21.7	174 960	0.009	2 712
(2,3)	538 330	0.006	56	498 308	0.008	7 042
(4,3)	337 852	0.005	26.1	332 120	0.006	3 852
(5,3)	41 310	0.005	6.55	40 908	0.004	455
(6,3)	936	0.019	5.65	1 250	0.060	19.4
(1,4)	163 960	0.008	17.3	122 360	0.011	1 488
(2,4)	482 548	0.008	47.2	397 834	0.011	2 996
(5,4)	32 780	0.008	6.31	30 554	0.012	325
(6,4)	2 238	0.018	2.18	1 264	0.007	7.93
(1,2)	213 328	0.006	17.8	200 794	0.007	1 844
(5,2)	41 192	0.004	5.11	40 432	0.005	344
(6,2)	1 760	0.017	4.27	1 764	0.023	14.9
(5,1)	34 766	0.008	3.21	40 908	0.004	458
(6,1)	998	0.010	1.44	980	0.012	5.53
(5,6)	4 356	0.009	0.79	3 086	0.008	4.01

Annex 4 - Summary of matching procedure with Euclidean and Mahalanobis distances for OD

Dichotomy	Euclidean			Mahalanobis		
	# Samples matched	Average SMD	Wall time (seconds)	# Samples matched	Average SMD	Wall time (seconds)
(1,3)	64 028	0.011	11.2	41 428	0.015	319
(2,3)	109 774	0.010	16.8	78 924	0.012	478
(4,3)	99 008	0.005	15.4	90 900	0.008	573
(5,3)	8 604	0.007	2.84	6 610	0.014	65
(6,3)	410	0.025	1.22	230	0.025	1.84
(1,4)	53 040	0.013	7.93	25 978	0.017	197
(2,4)	98 690	0.012	18	56 442	0.014	521
(5,4)	8 754	0.007	1.85	6 616	0.010	48.8
(6,4)	32	0.056	0.72	14	0.097	1.22
(1,2)	70 244	0.011	10.3	53 458	0.012	263
(5,2)	11 226	0.014	2.9	10 078	0.016	59.4
(6,2)	284	0.017	1.06	252	0.014	1.63
(5,1)	7 510	0.007	2.01	6 610	0.014	66
(6,1)	120	0.019	0.64	78	0.061	1.07
(5,6)	452	0.027	0.22	216	0.016	0.68

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa