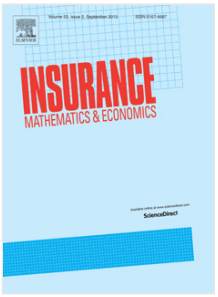
Accepted Manuscript



A survey of personalized treatment models for pricing strategies in insurance

Leo Guelman, Montserrat Guillén, Ana M. Pérez-Marín

PII: S0167-6687(14)00071-7
DOI: http://dx.doi.org/10.1016/j.insmatheco.2014.06.009
Reference: INSUMA 1956

To appear in: *Insurance: Mathematics and Economics*

Received date: April 2014
Revised date: June 2014
Accepted date: 8 June 2014

Please cite this article as: Guelman, L., Guillén, M., Pérez-Marín, A.M., A survey of personalized treatment models for pricing strategies in insurance. *Insurance: Mathematics and Economics* (2014), http://dx.doi.org/10.1016/j.insmatheco.2014.06.009

A survey of personalized treatment models for pricing strategies in insurance

Leo Guelman[a], Montserrat Guillén[b,*], Ana M. Pérez-Marín[b]

[a] *Royal Bank of Canada, RBC Insurance, 6880 Financial Drive, Mississauga, Ontario L5N 7Y5, Canada*
[b] *Dept. Econometrics, Riskcenter-IREA, University of Barcelona, Diagonal 690, E-08034 Barcelona, Spain*

Abstract

We consider a model for price calculations based on three components: a fair premium; price loadings reflecting general expenses and solvency requirements; and profit. The first two components are typically evaluated on a yearly basis, while the third is viewed from a longer perspective. When considering the value of customers over a period of several years, and examining policy renewals and cross-selling in relation to price adjustments, many insurers may prefer to reduce their short-term benefits so as to focus on their most profitable customers and the long-term value. We show how models of personalized treatment learning can be used to select the policy holders that should be targeted in a company's marketing strategies. An empirical application of the causal conditional inference tree method illustrates how best to implement a personalized cross-sell marketing campaign in this framework.

Keywords: rate making, cross-selling in insurance, predictive models, causal inference

1. Introduction

Actuarial science in nonlife insurance deals with the problem of pricing by calculating the fair premium each customer should pay for an insurance contract. Nevertheless, the final price offered to the customer is also markedly influenced by the specific conditions prevailing in the insurance market and by each company's commercial strategy over time. Examples of such strategies can be readily consulted in the literature: for example, De Kok (2003) argues that a company offering a

*Corresponding author.
Email addresses: leo.guelman@rbc.com (Leo Guelman), mguillen@ub.edu (Montserrat Guillén), amperez@ub.edu (Ana M. Pérez-Marín)

new product has marketing instruments that can influence market growth until a specific target market share has been reached for the new product. Additionally, at a given moment in time, a company observing an increase in customer lapses could offer its policies at a reduced rate so as to attract new customers and avoid losing its market share; alternatively, it might offer a multi-product discount to existing customers when buying additional insurance products from the same insurer. Nevertheless, the insurer must be careful when implementing such actions, as they will have a direct impact on the company's level of risk and solvency. In this context, predictive purchasing models can provide some orientation to the company as to which action should be adopted in each situation, but the particular nature of the insurance business has yet to be properly examined and discussed in the marketing literature. A number of recent papers have focused primarily on the reactions of those holding insurance to price changes, but none of them presents a comprehensive framework for addressing the problem of pricing, renewal and cross-selling (Donnelly et al., 2013; Kaishev et al., 2013; Guillén et al., 2011, 2012; Thuring et al., 2012).

Pricing products is an aspect of the activity of an insurance company that, on top of fair premium calculation, seeks to satisfy many fundamental objectives: to avoid the lapses of its profitable customers (retention), to attract new customers away from its competitors (market share growth), and to sell more policies to existing customers in the portfolio (cross-selling). The bonus-malus system, which is at the core of a posteriori ratemaking, is sometimes used as part of an overall strategy aimed at retaining good customers, as has been evidenced by Pitrebois et al. (2003), where "good" customers can be defined as being equivalent to profitable customers, i.e. those for whom observed losses over the years are lower than expected.

Customer retention and marketing interventions are generally expensive and if unsuccessful, they may represent a loss for the insurance company. In this sense, the presence of marketing costs in insurance has been noted in the literature (Broeders et al., 2011; Faust et al., 2012). The success of a campaign depends on a number of factors, but perhaps the most important is the way target customers are selected. For instance, it is clear that not all customers will respond in the same way to a specific price increase. Additionally, the expected profit generated by the customer as a result of a policy renewal or a new policy underwriting also needs to be taken into consideration.

Evidence of the losses that accrue when targeting the wrong policy holder for a marketing intervention can also be found in several articles. Thuring et al. (2012) are able to identify policy holders who have up to 64% more claims than a priori expected in a new insurance policy which might have been wrongly offered to them by the insurer. In a similar context, Kaishev et al. (2013) analyze the magnitude of that loss and report that it might be attributable to just a few policy holders in the portfolio. Finally, Guillén et al. (2011) identify policy holders that were negatively impacted by a retention program, i.e., they present a higher probability of canceling their policy after being targeted by a retention program. In all these cases, the insurance company could have avoided the loss suffered if they had correctly selected the customers to be contacted for a marketing action.

Optimal cross-selling involves experience rating or other ways of incorporating historical knowledge of the customer. It also involves a good underlying pricing model. Finally, it involves a good model for the probability of a sale. The *personalized treatment learning* framework discussed in this paper can incorporate all these components simultaneously. However, experimental data are required for that purpose. That is, the assignment of policy holders to the cross-sell marketing action must be performed using a chance procedure. On the contrary, if policy holders are exposed to the action based on some underlying non-random mechanism, the personalized treatment learning formulation is not directly applicable.

In this paper we present a summary of alternative methods formulated in the context of personalized treatment learning and we describe them in the framework of a model that considers the price of a policy as the sum of three components, namely fair premium, loadings and profits, where profits are related to the value of the customer. We show the implementation of causal conditional inference trees in a real insurance dataset to identify policy holders who should be targeted by a cross-sell campaign. This new method identifies the customers who are likely to be more positively influenced by a marketing effort and who will generate positive expected profits for the insurance company. The paper is organized as follows. In Section 2, the model describing the price charged to a policy holder and the value generated by a strategic action are presented. In Section 3, models for personalized treatment learning are described. An empirical application of personalized treatment

methods in cross-selling in insurance is presented in Section 4 and Section 5 concludes.

2. Model

We assume that the price $P_{\ell t}^{*}$ charged to policy holder $\ell = \{1, \ldots, L\}$ for a given contract in year $t = \{1, \ldots, T\}$ is the sum of three components:

- a) a fair premium ($LC_{\ell t}$), resulting from an evaluation of the policy holder's risk characteristics, that is, an estimation of expected loss claims;
- b) a price loading ($SR_{\ell t}$), capturing solvency requirements, managerial efficiency or caution; and, finally,
- c) profits $B_{\ell t}$, reflecting a minimum level of return to the company's shareholders or to the insurance company's owner.

The fair premium is readily calculated via predictive modeling combined with loss models. Price loadings and profits are not so straightforward, because they need to be viewed from a broader perspective. Part of the explanation might lie in the fact that solvency requirements and strategic actions are designed at the macro level, and as such involve the whole portfolio and not just each individual policy. As a result, price loadings have been studied to some extent in the context of regulatory capital, management science and marketing. Unfortunately, contributions in management science and marketing rarely address insurance applications; moreover, they are generally concerned with consumer preferences, that is, with demand as opposed to supply.

A major drawback when addressing the three components of price is the lack of a common time horizon. The fair premium calculation is usually performed on a one-year basis, as are solvency regulations. Yet, managers and shareholders may prefer a longer perspective and request a multi-year evaluation of risk. However, if a period of several years is considered, then policy renewal has to be taken into account as do a number of other questions. Many insurers, in fact, appear to prefer to reduce their profits a little in order to raise renewal rates, since overall company profits may in fact rise higher than the level obtained when customers are lost, especially if good customers (those

for whom observed losses over the years are below expectations) opt to move to a rival company[1]. Here, we define renewal $D_{\ell t}$ as a binary variable which equals 1 if policy holder ℓ renews his policy in year t, and 0 otherwise. In the multi-period setting, renewal ($D_{\ell t}$) and price ($P^*_{\ell t}$) are mutually dependent. In fact, it is intuitive that if the price increases many policy holders will abandon the company, but if the price falls then renewal is more likely than lapsing.

The interaction between price and renewal has been studied previously by Guelman and Guillén (2014), the authors concluding that price elasticity is not constant throughout the portfolio. Here, in addition, we argue that dependence between renewals and price adjustments is not stable over time, because the relationship may depend on external factors such as the prices offered by competitors in the insurance market.

Customer value is important when evaluating lapses. Typically, only renewal rates are reported; however it might be preferable to report the "retained value". Instead of looking at lapse rates, managers might consider the value of the customers that have lapsed and compare this to the value of those who have stayed.

We assume that there are L policy holders in a portfolio and that they may hold more than one policy. We indicate each type of insurance product by j, where $j = 1, .., K$ and K is the total number of possible insurance products. The company can control prices, so let us call $A_{\ell jt}$ the price change to be offered to policy holder ℓ in year t for policy j before renewal. We then estimate the expected change in customer value due to this potential price change.

Interestingly, price changes do not need to be homogeneous across all customers. For instance, not all prices have to increase by 5%, since not all customers will react to this price increase in the same way. An optimal strategy for a company would be to define a vector of price change functions $A_{\ell j} = (A_{\ell j1}, ..., A_{\ell jT})$ for a time horizon of T years, where $A_{\ell jt}$, $t = 1, ..., T$ is a function that assigns an optimal price change for policy j for customer ℓ in year t. Recently, Landriault et al. (2012) proposed a model in which the premium rate was reviewed each time the surplus of the company dropped below its previously reached minimum, a new rate then being chosen from

[1]Note that it has been argued that some optimal pricing techniques may imply that insurance policies are sold below the fair premium prices. This could be critical for the insurance company. Instead, we only consider a reduction of profits as we argue that pricing below actuarial fairness should be penalized by regulators.

among a certain number of possible rates. In this way, premium rates were fixed ahead of time, which is viewed by Landriault et al. (2012) as positive from a marketing point of view for the insurance company.

We define the set of all individual strategies as $A_t = \{A_{\ell jt};\ \ell = 1, ..., L\, j = 1, ..., K\}$. The total value at t of such strategy, $V(A_t)$, is the sum of the expected benefits over all customers generated from year t to T. The indicator $I_{\ell t}(j)$ equals one if policy holder ℓ holds product j in year t, and 0 otherwise. Additionally, let $S_{\ell js}$ be the probability that customer ℓ keeps policy j in year s, namely $P(I_{\ell s}(j) = 1)$ for $s = t, ..., T$. Finally, let $B_{\ell jt}$ be the benefit of policy j from policy holder ℓ in year t, and r is the interest discount factor. So the total value at t is:

$$V(A_t) = \sum_{\ell=1}^{L} \sum_{j=1}^{K} I_{\ell t}(j) B_{\ell jt} \sum_{s=t}^{T} S_{\ell js} r^{s-t}. \tag{1}$$

Note that the total value at t is calculated by assuming that each policy held by the customer in year t will generate a yearly profit equal to that corresponding to year t, should the customer decide to renew it in subsequent years. This is a simplification, which might be even further simplified if we assume a constant profit margin for all products and policy holders. Also, from the practical perspective, notice that formulation (1) requires knowing the entire set of insurance product holdings for each customer at a given point in time. This can be a challenging endeavour for some insurers, as products may have been sold historically under different aging legacy systems, and the identification of common clients across different platforms is sometimes not feasible.

The goal is to find an optimal A_t that maximizes (1), i.e. the aggregate value. Personalized methods aim at identifying the customers that should be targeted in cross-selling actions and the price adjustments that are recommended to increase the overall value. Personalized methods usually involve many possible actions and the complexity of this approach is proportional to the number of possible treatments. In the section that follows we survey a list of personalized treatment models.

3. Models for personalized treatment learning with applications to insurance

Recently, new algorithms and statistical models have been developed in response to the development of data storage technology, which allows insurance companies, for example, to store vast databases containing historical information about their policy holders. Efforts aimed at *learning from data* (Abu-Mostafa et al., 2012) refer to tasks of extracting "implicit, previously unknown, and potentially useful information from data" (Frawley, 1991). In this context, a distinction is required between *supervised* and *unsupervised* learning. The goal of *supervised* learning is to predict the value of a response variable based on some *observable* covariates. By contrast, in *unsupervised* learning no response variable is present to "supervise" the learning process and the goal is to identify structures and patterns among the covariates.

In insurance, in addition to observable policy holder characteristics (e.g., age, gender, territory, etc.), the values of certain covariates can be proactively selected by the insurer. For example, the new premium offered to policy holders for policy renewal at the end of the coverage period. In this instance, we are interested in selecting the optimal premium level or "treatment" for each policy holder based on their characteristics. Here, we consider a treatment to be optimal if it maximizes the probability of a desirable outcome. This task of learning the optimal personalized treatment is referred to as *personalized treatment learning*.

As each policy holder receives a single treatment (or yearly premium increase/decrease), the value of the response under alternative treatments is unobserved. Therefore, the quantity we seek to predict (i.e., the optimal personalized treatment) is unknown on a given training data set. This problem is referred to as *the fundamental problem of causal inference* (Holland, 1986).

The underlying motivation for *personalized treatment learning* is that individuals can show significant heterogeneity in response to treatments, thereby making an appropriate treatment choice for each individual is essential. As discussed, a premium increase of 5% affecting all policy holders is likely to lead some to switch to an alternative insurer, while others will opt to remain with the same company. In this context, Guelman et al. (2012) proposed a method based on an adapted version of *random forests* to identify policy holders that are positively/negatively impacted by a client retention program. Similarly, Guelman and Guillén (2014) describe a framework for determining

the optimal rate change (i.e., playing the role of the treatment) for each individual policy holder so as to maximize the overall expected profitability of an insurance portfolio.

A key consideration in building models of personalized treatment learning concerns the mechanism by which the treatment is assigned to the observational units. Treatments may be assigned as a result of a careful randomized experimental design or, alternatively, in accordance with a non-random mechanism. For instance, in the insurance business, experimental designs in marketing applications are more common than in pricing applications. In this latter case, understanding the exact impact of a change in premium on the likelihood of policy renewal requires special modeling considerations (Guelman and Guillén, 2014). The methods discussed in this paper for building personalized treatment learning models assume that experimental data are available.

3.1. The simplest problem: two possible treatments

Upper case letters denote random variables and lower case letters denote the values of the random variables. We assume that a sample of individuals is randomly assigned to two treatment arms, denoted by A, $A \in \{0,1\}$, also known as control and treatment states, respectively. Let $Y(a) \in \{0,1\}$ denote a potential binary outcome of an individual if assigned to treatment state $A = a$, $a = \{0,1\}$. An outcome is, for instance, a renewal versus a lapse or, alternatively, the buying or not buying of a new product. In the insurance business, $A = 1$ may, for example, represent a 5% premium increase and $A = 0$ may represent no premium increase. Alternatively, in a cross-sell campaign conducted by an insurance company $A = 1$ may represent the offering of an additional coverage to an existing policy holder and $A = 0$ may represent no cross-sell attempt.

Therefore, $Y = AY(1) + (1-A)Y(0)$ represents the observed outcome. We assume that a value of $Y = 1$ is more desirable than $Y = 0$. Let $\mathbf{X} = (X_1, \ldots, X_p)^\top$ be the p-dimensional vector of covariates describing the characteristics of the policy holders. We also assume that our data consist of L independent and identically distributed realizations of $(Y, A, \mathbf{X})$, $\{(Y_\ell, A_\ell, \mathbf{X}_\ell),\ \ell = 1, \ldots, L\}$.

Under the randomization assumption we assume that treatment assignment A is independent of the outcomes $Y(0)$ and $Y(1)$, which we denote by $\{Y_\ell(0), Y_\ell(1) \perp A_\ell\}$. Under these assumptions, the *average treatment effect* (ATE) is assumed to be constant for all policy holders and is defined by

$$\begin{aligned}\tau &= E[Y_\ell(1) - Y_\ell(0)] \\ &= E[Y_\ell|A_\ell = 1] - E[Y_\ell|A_\ell = 0].\end{aligned} \tag{2}$$

As discussed, individuals are highly heterogeneous with regard to their response to treatments. In order to identify subgroups of subjects for which the treatment is most beneficial/harmful, the *individual treatment effect* (ITE), defined as $Y_\ell(1) - Y_\ell(0)$, as opposed to the ATE, should be used. However, as a subject is never observed simultaneously in both treatment states, the ITE is unobservable. In practice, we use the *subpopulation treatment effect* (STE) as an approximation to the ITE. For a subject with covariates $\mathbf{X}_\ell = \mathbf{x}$, the STE is given by

$$\begin{aligned}\tau(\mathbf{x}) &= E[Y_\ell(1) - Y_\ell(0)|\mathbf{X}_\ell = \mathbf{x}] \\ &= E[Y_\ell|\mathbf{X}_\ell = \mathbf{x}, A_\ell = 1] - E[Y_\ell|\mathbf{X}_\ell = \mathbf{x}, A_\ell = 0].\end{aligned} \tag{3}$$

Here, we use the term *personalized treatment effect* (PTE) to refer to the STE. A *personalized treatment rule* $\mathcal{H}$ is a map from the space of covariates $\mathbf{X}$ to the space of treatments A, $\mathcal{H}(\mathbf{X}) : \mathbb{R}^p \to \{0, 1\}$. An *optimal treatment rule* is the one that maximizes the expected outcome if the personalized treatment rule is implemented for the whole population[2], i.e., $E[Y(\mathcal{H}(\mathbf{X}))]$. The optimal personalized treatment rule is given by $\mathcal{H}^* = \text{argmax}_{\mathcal{H}}\, E[Y(\mathcal{H}(\mathbf{X}))]$ for a subject with covariates $\mathbf{X}_\ell = \mathbf{x}$ as $\mathcal{H}^* = 1$ if $\tau(\mathbf{x}) > 0$, and $\mathcal{H}^* = 0$ otherwise.

Next, we describe alternative methods to estimate (3), which can be used by the insurance company to select customers for specific campaigns.

3.2. Indirect estimation methods

In this section we present three methods to estimate the STE indirectly. In the context of personalized treatment learning, such indirect estimation methods are those that use a two-stage

[2]Notice that since Y is binary, this expectation has a probabilistic interpretation. That is, $E[Y(\mathcal{H}(\mathbf{X}))] = \text{Prob}\,(Y(\mathcal{H}(\mathbf{X})) = 1)$.

procedure in order to obtain an estimation of (3). In the first stage, these methods predict the outcome Y conditional on the covariates $\mathbf{X}$ and treatment A. In the second stage, they subtract the predicted value of Y under each treatment to obtain an estimate of the STE.

The first method, the *difference score method*, was presented by Larsen (2009), who discussed estimating two independent models for the response Y, one based on the treated subjects, $E[Y|\mathbf{X}, A = 1]$, and one based on the control subjects, $E[Y|\mathbf{X}, A = 0]$. By subtracting the estimated value of the response from the two models an estimation of the PTE for a subject with covariate $\mathbf{X}_\ell = \mathbf{x}$ is obtained. The two models can be estimated by using any conventional bivariate response model or binary classification algorithm.

The second method, the *interaction* approach, was proposed by Lo (2002). This method involves fitting a single model to the response on the main effects and adding interaction terms between each covariate $\mathbf{X} = (X_1, \ldots, X_p)^\top$ and the treatment indicator A. A standard logistic regression is used to fit the model. The additional effect of each covariate due to treatment is measured by the estimated parameters of the interaction terms. By setting $A_\ell = 1$ and $A_\ell = 0$ in the fitted model and by subtracting the corresponding predicted probabilities, an estimate of the PTE for a subject with covariates $\mathbf{X}_\ell = \mathbf{x}$ is obtained. Note that the interaction method represents an improvement on the difference score method (as the interaction between treatment and covariates can be formally tested), but that it also suffers from overfitting problems (Zhao and Zeng, 2012). Overfitting may be prevented by using least absolute shrinkage and selection operator (LASSO) logistic regression (Tibshirani, 1996) for variable selection and shrinkage, but the same LASSO constraints apply to the main and treatment heterogeneity parameters. Moreover, as the variability in the response due to the interaction effects is usually smaller than the variability in the response due to the main effects, this gives rise to a problem.

In this context, Imai and Ratkovic (2012) proposed a third method, known as *L2-SVM*, which is an adapted version of the support vector machine (SVM) classifier (Vapnik, 1995). Adopting the SVM approach, it is possible to formulate separate LASSO constraints over the main and treatment heterogeneity parameters. Specifically, let $Y_\ell^* = 2Y_\ell - 1 \in \{-1, 1\}$ and consider the following optimization problem

$$\min_{(\alpha,\theta)} \sum_{\ell=1}^{L} \left|1 - Y_\ell^*(\mu + \alpha^\top \mathbf{X}_\ell + \theta^\top \mathbf{X}_\ell A_\ell)\right|_+^2 + \lambda_{\mathbf{X}} \sum_{m=1}^{p} |\alpha_m| + \lambda_{\mathbf{XA}} \sum_{m=1}^{p} |\theta_m| \tag{4}$$

where $\lambda_{\mathbf{X}}$ and $\lambda_{\mathbf{XA}}$ are separate pre-specified LASSO penalties for the main effect parameters α and treatment heterogeneity parameters θ, respectively, $|x|_+ \equiv \max(x, 0)$ is the hinge-loss (Wahba, 2002), and μ is a constant term. Once model (4) has been estimated, the PTE can be obtained by calculating the difference in the truncated values of the predicted response under each treatment condition.

The main criticism of indirect estimation methods is that they emphasize the prediction accuracy of the response, as opposed to the accuracy in estimating the change in the response *caused* by the treatment at the individual level. Therefore, even when the indirect estimation methods are correctly specified to predict Y_ℓ conditional on covariates $\mathbf{X}_\ell = \mathbf{x}$ and treatment A, there is no guarantee that these models can accurately predict $Y_\ell(1) - Y_\ell(0)$ conditional on the same covariates.

3.3. Modified covariate method

The modified covariate method was proposed by Tian et al. (2012). After modifying the covariates in a simple manner, an appropriate regression model can be estimated using these modified covariates. In so doing, the main effects do not have to be modeled directly. The following steps have to be taken in applying this method: i) the treatment indicator is transformed as $A_\ell^* = 2A_\ell - 1 \in \{-1, 1\}$, ii) each covariate in $\mathbf{X}_\ell$ is also transformed as $\mathbf{Z}_\ell = \mathbf{X}_\ell^* A_\ell^*/2$, where $\mathbf{X}^*$ is the centered version of $\mathbf{X}$, and iii) a regression model is estimated to predict Y on the modified covariates $\mathbf{Z}$, for example, by using a logistic regression model.

3.4. Modified outcome method

This method was proposed by Jaśkowski and Jaroszewicz (2012). First, a new outcome variable W is defined such that

$$W_\ell = \begin{cases} 1 & \text{if } A_\ell = 1 \text{ and } Y_\ell = 1 \\ 1 & \text{if } A_\ell = 0 \text{ and } Y_\ell = 0 \\ 0 & \text{otherwise.} \end{cases}$$

Second, a binary regression model is fitted to W on covariates $\mathbf{X}$. We assume that a value of $Y = 1$ is more desirable than that of $Y = 0$, and, thus, we can intuitively think of $W = 1$ as the event of obtaining a potential outcome under treatment that is at least as good as the observed outcome. The probability of this event is given by

$$P(W_\ell = 1|\mathbf{X}_\ell = \mathbf{x}) = P(Y_\ell = 1|\mathbf{X}_\ell = \mathbf{x}, A_\ell = 1)P(A_\ell = 1) + P(Y_\ell = 0|\mathbf{X}_\ell = \mathbf{x}, A_\ell = 0)P(A_\ell = 0).$$

By assuming that $P(A = 1) = P(A = 0) = 1/2$, the same assumption as in the modified covariate method, it can be proved that

$$\tau(\mathbf{x}) = 2P(W_\ell = 1|\mathbf{X}_\ell = \mathbf{x}) - 1.$$

Then a logistic regression model can be used to fit $P(W = 1|X)$ and to obtain an estimate of the PTE.

3.5. Causal conditional inference trees

Uplift random forests directly predict the expected change in the outcome as a result of the treatment, as opposed to predicting the outcome itself. Further details about uplift random forests can be found in Guelman et al. (2014). In short, an ensemble of B trees are grown, each built on a fraction ν of the training data (which include both treatment and control observations). The sampling, in line with Friedman (2002), incorporates randomness as an integral part of the fitting procedure. This not only reduces the correlation between the trees in the sequence, but also reduces the computing time by the same fraction ν. A typical value for ν might be $1/2$, although for large data, it can be substantially smaller. The tree-growing process involves selecting $n \leq p$ covariates at random as candidates for splitting. This adds another layer of randomness, which further reduces

the correlation between trees, hence reducing the variance of the ensemble. The split rule is based on a measure of distributional divergence, as defined in Rzepakowski and Jaroszewicz (2012). The individual trees are grown to maximal depth (i.e., no pruning is done). The estimated personalized treatment effect is obtained by averaging the predictions of the individual trees in the ensemble.

Causal conditional inference trees represent an improvement on uplift random forests in two respects: overfitting and the selection bias towards covariates with many possible splits. The development of this new approach was motivated by the *unbiased recursive partitioning* method proposed by Hothorn et al. (2006).

The flow diagram of the new algorithm is presented in Figure 1. Its most relevant features are discussed here. Specifically, for each terminal node in the tree the algorithm tests the global null hypothesis of no interaction effect between treatment A and *any* of the n covariates randomly selected from the set of p covariates. The global hypothesis of no interaction is formulated by using n partial hypotheses $H_0^m : E[W|X_m] = E[W],\ m = \{1, \ldots, n\}$, with the global null hypothesis $H_0 = \cap_{m=1}^{n} H_0^m$, where W is defined as in the modified outcome method discussed in Section 3.4. Thus, a conditional independence test of W and X_m has a causal interpretation for the treatment effect for subjects with covariate X_m. Bonferroni-adjusted p-values or alternative adjustment procedures (Wright, 1992; Shaffer, 1995; Benjamini and Hochberg, 1995) can be used for handling multiplicity in testing. The algorithm stops the splitting process at that node when it is not able to reject H_0 at a pre-specified significance level α. Otherwise, the m^*th covariate X_{m*} with the smallest adjusted p-value is selected. The algorithm then creates a partition Ω^* of the covariate X_{m*} in two disjoint sets $\mathcal{M} \subset X_{m*}$ and $X_{m*} \setminus \mathcal{M}$.

Strasser and Weber (1999) developed a comprehensive theory based on a general functional form of multivariate linear statistics appropriate for arbitrary independence problems. Specifically, to test the null hypothesis of independence between W and $X_m,\ m = \{1, \ldots, n\}$, we define

$$\mathcal{T}_m = \text{vec}\left(\sum_{\ell=1}^{L} g(X_{m\ell}) h(W_\ell, (W_1, \ldots, W_L))^\top \right) \in \mathbb{R}^{u_m v \times 1} \tag{5}$$

where $g : X_m \to \mathbb{R}^{u_m \times 1}$ is a transformation of the covariate X_m and $h : W \to \mathbb{R}^{v \times 1}$ is known as the *influence function*. The "vec" operator transforms the $u_m \times v$ matrix into a $u_m v \times 1$ column

vector. By fixing $X_{m1}, \ldots, X_{mL}$ and conditioning on all possible permutations Z of the responses $W_1, \ldots, W_L$, the distribution of $\mathcal{T}_m$ under the null hypothesis can be obtained. A univariate test statistic c is then obtained by standardizing $\mathcal{T}_m \in \mathbb{R}^{u_m v \times 1}$ based on its conditional expectations $\mu_m \in \mathbb{R}^{u_m v \times 1}$ and covariance $\Sigma_m \in \mathbb{R}^{u_m v \times u_m v}$, as derived by Strasser and Weber (1999). A common choice is the maximum of the absolute values of the standardized linear statistic

$$c_{\max}(\mathcal{T}, \mu, \Sigma) = \max|\frac{\mathcal{T} - \mu}{\text{diag}(\Sigma)^{1/2}}|, \tag{6}$$

or a quadratic form

$$c_{\text{quad}}(\mathcal{T}, \mu, \Sigma) = (\mathcal{T} - \mu)\Sigma^{+}(\mathcal{T} - \mu)^{\top}, \tag{7}$$

where Σ^{+} is the Moore-Penrose inverse of Σ.

In the flow diagram in Figure 1, we select the covariate X_m^* with the smallest adjusted p-value. The p-value P_m is given by the number of permutations $z \in Z$ of the data with corresponding test statistic exceeding the observed test statistic $q_m \in \mathbb{R}^{u_m v \times 1}$. Specifically,

$$P_m = \mathbf{P}(c(\mathcal{T}_m, \mu_m, \Sigma_m) \geq c(q_m, \mu_m, \Sigma_m)|Z).$$

For large sample sizes, it might not be possible to obtain the exact distribution of the test statistic. In that case, we propose approximating the distribution by using a random sample of the set of all permutations Z. Additionally, the asymptotic distribution of the test statistic given by (6) tends to multivariate normal with parameters μ and Σ as $L \to \infty$ (see Strasser and Weber (1999)). The test statistic (7) follows an asymptotic chi-square distribution with degrees of freedom given by the rank of Σ. Therefore, asymptotic p-values can be computed for these test statistics.

After selecting the covariate X_{m*} to split, the next step is to use a split criterion which explicitly attempts to find subgroups with heterogeneous treatment effects. The following measure proposed by Su et al. (2009) and subsequently implemented by Radcliffe and Surry (2011) can be used for assessing the personalized treatment effect from a split Ω

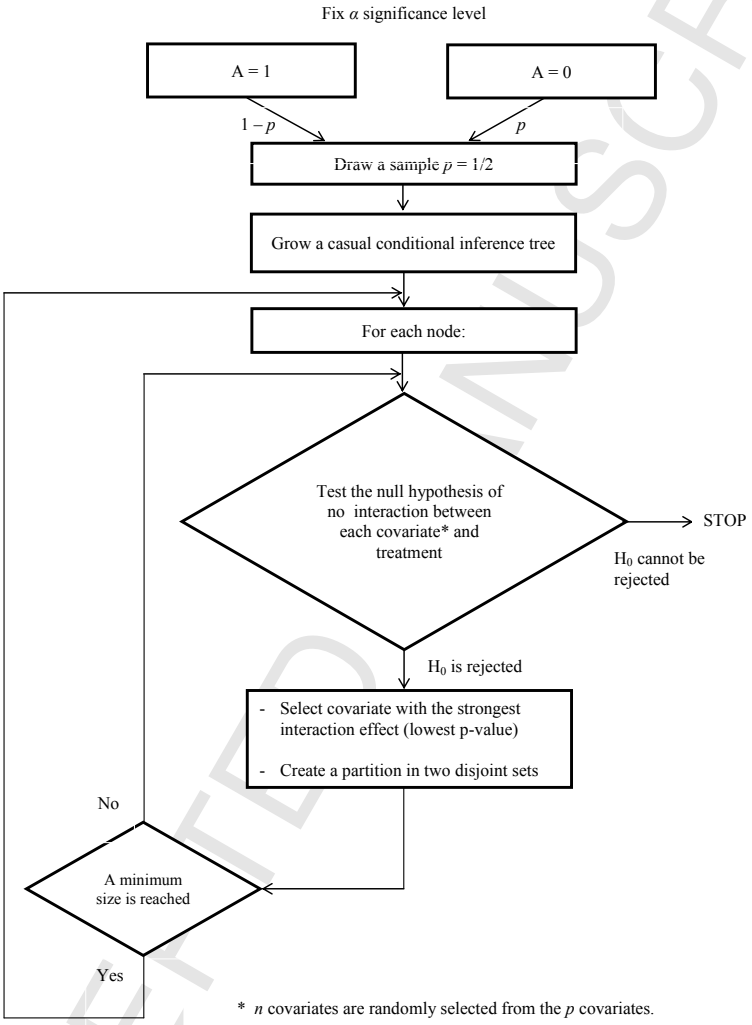


Figure 1: *Algorithm for causal conditional inference forests.*

$$G^2(\Omega) = \frac{(L-4)\{(\bar{Y}_{n_L}(1) - \bar{Y}_{n_L}(0)) - (\bar{Y}_{n_R}(1) - \bar{Y}_{n_R}(0))\}^2}{\hat{\sigma}^2\{1/L_{n_L}(1) + 1/L_{n_L}(0) + 1/L_{n_R}(1) + 1/L_{n_R}(0)\}} \tag{8}$$

where n_L and n_R denote the left and right child nodes, respectively, $L_{i\in\{n_L,n_R\}}(A)$ denotes the number of observations in child node i exposed to treatment $A \in \{0,1\}$, and

$$\bar{Y}_{i\in\{n_L,n_R\}}(1) = \frac{\sum_{\forall \ell \in i} Y_\ell A_\ell}{\sum_{\forall \ell \in i} A_\ell}, \tag{9}$$

$$\bar{Y}_{i\in\{n_L,n_R\}}(0) = \frac{\sum_{\forall \ell \in i} Y_\ell (1 - A_\ell)}{\sum_{\forall \ell \in i} (1 - A_\ell)}, \tag{10}$$

$$\hat{\sigma}^2 = \sum_{A\in\{0,1\}} \sum_{i\in\{n_L,n_R\}} L_i(A)\bar{Y}_i(A)(1 - \bar{Y}_i(A)). \tag{11}$$

The best split is the one that maximizes the criterion $G^2(\Omega)$ among all possible splits, given by $G^2(\Omega^*) = \max_\Omega G^2(\Omega)$. It can be proved that the split criterion given in (8) is equivalent to a chi-square test for testing the interaction effect between the treatment and the covariate X_{m*} dichotomized at the value given by the split Ω (Su et al., 2009).

All the methods described in this section have been developed by the authors in the **uplift** R package (Guelman, 2014), now available from CRAN (R Development Core Team, 2013).

4. Empirical application

In the context of insurance, estimating the profitability of a given cross-sell attempt requires considering three key components. First, we need to consider a campaign-specific *response model*. The methods proposed in the insurance literature to tackle this problem involve building a model for the conditional probability of a sale given the marketing action (Donkers et al., 2007; Kamakura et al., 2003; Harrison and Ansell, 2002). This approach is unlikely to maximize the expected profitability of the action as it does not recognize the policy holder purchase behavior in the absence of the action. Some customers will buy independently of the action while others may

be neutral or even turned-off by it. In this paper, we propose using the personalized treatment learning methods discussed in Section 3 to estimate the impact of marketing interventions at the individual policy holder level. Second, once we have estimated the impact of actions in terms of their response outcomes, we need to consider the expected economic benefits from those actions. This was addressed in Section 2 and involves considering the profitability of the insurance product holdings and their estimated retention over a given time horizon. Lastly, we need to take into account the fixed and variable expenses involved in the action itself. Clearly, certain actions may drive a large number of incremental sales, but their associated cost could make them economically inviable.

In this section, we present an empirical application in which policy holders are selected for a cross-sell campaign by considering the three key elements discussed above. In this application, we use causal conditional inference forests as the personalized treatment learning method. A brief introduction to cross-selling in insurance is provided and then a real dataset example is discussed.

4.1. Cross-selling in insurance: background

Recent contributions (see, Thuring et al. (2012) and Kaishev et al. (2013)) consider the potential heterogeneity in the profitability of the cross-sell effort, i.e., an action designed to sell additional products to existing customers. Thuring et al. (2012) implement a method for selecting policy holders to whom a company should seek to cross-sell additional insurance contracts using multivariate credibility. The method is based on the idea that the claims behavior of an individual customer in relation to a policy not yet owned by that customer should be similar to their claims behavior in relation to their existing policies. The authors use multivariate credibility theory to estimate a customer-specific risk profile and to evaluate whether selling an additional product to that customer is expected to contribute positively to the company's profits. In this context, profits are measured by the deviation between the a priori expected number of claims and the corresponding observed number of claims. Thuring et al. (2012) apply their method to an actual portfolio of customers who owned motor, building and content insurance policies between 1999 and 2004, although not for the same length of time. In this way, the authors identify the 20% of customers the company should not target in their marketing program. However, their approach fails to take

into account the probability of a positive response to their cross-sell effort and the impact of the cost of the marketing campaign itself on company profits.

In Kaishev et al. (2013) a new method is proposed for optimal cross-sales customer selection based on expected profit maximization and mean-variance optimization. The authors propose a model for the stochastic cross-sales profit which depends on the cross-sell probability, the stochastic price, the costs of the new cross-sell policy, and the deterministic cost of the cross-sell attempt. They propose estimating the probability of a successful cross-sell attempt by using a regression model based on collateral data from the insurance company, collected from past cross-sell campaigns. They apply their method to a dataset of customers that owned household insurance coverage and who were offered a car insurance policy. In this application, the authors obtained a mean value of the predicted probability of a successful cross-sell attempt (6.9%) that was markedly higher than the actual proportion of customers that decided to underwrite car insurance (4%); as such, the company expected to cross-sell the car insurance policy to more customers than actually ended up purchasing coverage.

It is worth stressing that both Thuring et al. (2012) and Kaishev et al. (2013) make a relevant contribution insofar as they consider the potential heterogeneity in the profitability of the cross-sell attempt from an actuarial perspective by considering the number of expected claims from a customer under the new cross-sold policy. This had not previously been discussed in the marketing literature on cross-selling (see, for example, Kamakura et al. (1991); Knott et al. (2002); Li et al. (2005)). Here, we seek to show that *personalized treatment learning models* can be used to approximate the specific response of each individual to a particular treatment or marketing campaign.

4.2. Data and results

The dataset used in this empirical study corresponds to a direct mail campaign conducted by a leading Canadian insurer between June 2012 and May 2013. The objective of this campaign was to cross-sell a home insurance policy to existing auto insurance customers. Attempts were made to persuade the latter to buy a new policy by offering them a multi-product discount. The cross-sell campaign was conducted principally via direct mail, but customers also received a telephone call

to further convince them to underwrite a home insurance policy. Customer response to the cross-sell campaign was measured by observing whether the client actually purchased the home policy in the three-month period immediately following receipt of the offer. In addition, the dataset collected information on approximately 50 auto insurance policy covariates (namely, driver and vehicle characteristics) as well as general policy information. A randomized group of clients that were not targeted in the cross-sell campaign were included in the analysis as a control group.

A summary of cross-sell rates is given in Table 1. The cross-sell campaign targeted 30,973 policy holders, of whom just 789 opted to buy home insurance, giving a cross-sell rate of 2.55%. The policy holders not included in the campaign comprised 3,397 individuals, of whom 75 purchased home insurance, giving a cross-sell rate of 2.21% for the control group. Thus, the average treatment effect was 0.34% (2.55% - 2.21%), an outcome that is not statistically significant with a p-value of 0.23 based on a chi-square test.

Table 1: Cross-sell rates by group

	Treatment	Control
No. of customers not purchasing home insurance	30,184	3,322
No. of customers purchasing home insurance	789	75
Cross-sell rate	2.55%	2.21%

However, as discussed, the average treatment effect is of limited value to us when policy holders are significantly different in the way they react to a cross-sell attempt. Additionally, the traditional method employed by most insurers in their cross-sell efforts is to build a propensity to buy (or a customer response) model to determine which customers are most likely to buy the target product given that they own the base product. Hence, their marketing activity is directed towards the "most likely to buy customers". However, some customers may purchase the target product independently of the marketing incentive, while others may be turned off by the campaign. Hence, estimating the personalized treatment effect is more relevant as it seeks to determine which customers are most likely to be influenced positively by the specific marketing campaign. In order to verify the performance of this method, 70% of the data was used for building the model and the other 30% was used for validation. A causal conditional inference forest was estimated to the training

dataset using the default parameter values ($B = 500$, $n = 16$, and a p-value $= 0.05$ as the level of significance α). The policy holders in the validation data set were ranked (from high to low) according to their estimated personalized treatment effect and grouped into deciles. We then calculated the actual average treatment effect within each decile (i.e., the difference in cross-sell rates between the treatment and control groups).

Figure 2 (left) illustrates the distribution of the estimated personalized treatment effect. The first (tenth) decile represents the 10% of clients with the highest (lowest) predicted personalized treatment effect. The figure to the right shows the actual cross-sell rate by decile and target group. The results show that the actual difference in the cross-sell rate between treatment and control groups is higher for policy holders with higher estimated personalized treatment effect. There is also a subgroup of policy holders who were negatively impacted by the marketing intervention activity (deciles 8-10). Negative reactions to sale attempts have been recognized in the literature (Günes et al., 2010; Kamakura, 2008) and may happen for a variety of reasons. For instance, the marketing action may trigger a decision to shop for higher multi-product discounts among competitors. Moreover, if the client currently owns a home policy with another insurer, he may decide to switch the auto policy to that insurer instead. We found evidence of higher auto policy cancellation rates at the higher deciles. In addition, some clients tend to perceive sales calls as intrusive, generating a negative reaction.

Now we can calculate the expected profitability that would result from targeting each decile by taking into account the expected value of a home insurance policy. That is, the expected value of a home policy in decile $d = \{1, \ldots, 10\}$ is given by $V_d = B_d \sum_{t=1}^{5} S_{dt} r^t$ where B_d is the expected profit generated by a home policy in decile d, S_{dt} is the probability that a policy holder in decile d will retain the home policy beyond year t, and r^t is the interest discount factor. The expected profit B_d is estimated by calculating $B_d = \bar{P}_d^* - \hat{LC}_d - \overline{SR}_d$ where $\bar{P}_d^*$ is the average policy premium, $\hat{LC}_d$ is the predicted insurance losses per policy-year and $\overline{SR}_d$ are the fixed and variable expenses from servicing the policy, all of them in decile d. Note that, essentially, we use equation (1), albeit adapted as so to calculate the expected value of a home policy in decile d over the next five years. The expected profitability from targeting a client in decile $d = \{1, \ldots, 10\}$ is

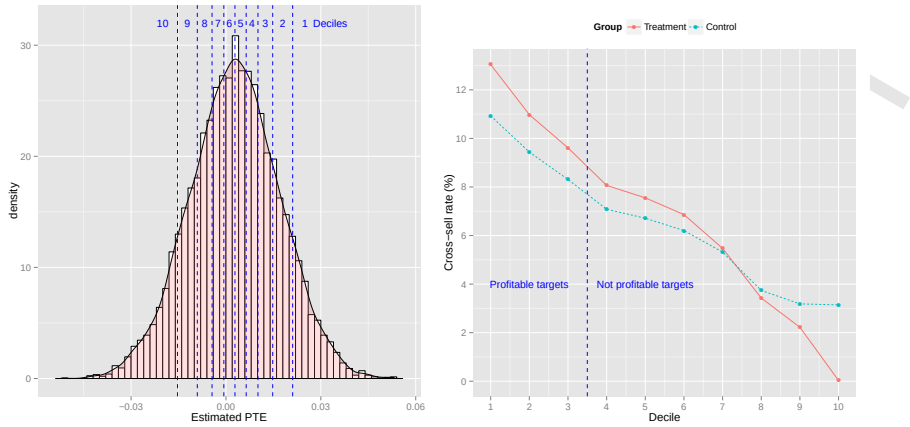


Figure 2: *The figure to the left shows the distribution of the estimated personalized treatment effect. The figure to the right shows the actual cross-sell rate by decile and target group.*

obtained by $\hat{\tau}_d(\mathbf{x})V_d - E_d$, where $\hat{\tau}_d(\mathbf{x})$ is the estimated personalized treatment effect in decile d, and E_d captures the fixed and variable campaign expenses for targeting each decile. Based on these considerations, only those clients in deciles 1-3 present positive expected profits from the marketing activity and, as such, are the customers that should be targeted. The incremental cross-sell rate of clients in deciles 4-7 is outweighed by the incremental costs, and so the company should avoid targeting these clients. Clients in deciles 8-10 have negative reactions to the campaign and clearly should not be targeted either.

5. Conclusions

In this article the concept of pricing has been reformulated by taking into account the influence that market conditions might have on the final price paid by the policy holder. To do this, we have developed a model in which the premium results from the sum of three components: a fair premium (based on an evaluation of the risk characteristics of the insured party), price loadings (capturing solvency requirements and managerial efficiency) and profits (reflecting a minimum level of return to shareholders). The calculations for the first two components are typically performed on a yearly basis; in contrast, the third component is viewed from a longer perspective, comprising a period of several years and proving a cushion for pricing strategies aimed at increasing the value of the portfolio. In the long term, policy renewal and cross-selling opportunities become essential,

since many insurers would rather reduce their profits a little so as to increase their renewal rates, especially among the group of good policy holders. Insurers may also design marketing actions to increase the number of products sold to their existing customers and to attract new clients.

Clearly, every strategy that has some impact on customer value needs to be evaluated. Here, we have introduced a model that can identify this value at a given point in time for a given strategy addressing a given group of policy holders. In making this formulation, we have taken into account the heterogeneity that policy holders can present in response to the given strategy implemented by the insurer. Specifically, we have proposed using personalized treatment learning models for selecting policy holders who should be the target of a specific strategy. Moreover, we present a summary of alternative methods formulated in the context of personalized treatment learning that can be used in the insurance business.

We also show an application to real data in which we employ the novel method of causal inference trees. Here, our objective is to determine which customers holding an automobile policy with a major Canadian insurance company should be offered a home insurance policy with the same company. The method allows us to identify the group of policy holders that can expect to make a profit thanks to the strategy and to whom, therefore, the cross-selling attempt should be addressed.

In conclusion, the methodology proposed should help insurers in taking decisions that impact their customer value. We recommend that insurers take advantage of the new algorithms associated with personalized treatment models, since they fit the nature of the problem analyzed here perfectly: that is, the heterogeneity of the response of policy holders to a given intervention, such as an increase in the premium or a multi-product discount offer. In this way, the right strategy can be used in addressing each policy holder, and ensure the company avoids losing a potential profit.

Acknowledgements

LG thanks the Royal Bank of Canada, RBC Insurance. MG and AMP-M thank ICREA Academia and the Spanish Ministry of Science / FEDER grant ECO2010-21787-C03-01. We are grateful to the editor and the reviewers for their thoughtful comments that helped us improve a

prior version of this article. All errors remain our own responsibility.

References

Abu-Mostafa, Y., Magdon-Ismail, M. and Hsuan-Tien, L. 2012. *Learning From Data.* AMLBook.

Benjamini, Y. and Hochberg, Y. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B* 57(1): 289–300.

Broeders, D., Chen, A. and Koos, B. 2011. A utility-based comparison of pension funds and life insurance companies under regulatory constraints. *Insurance: Mathematics and Economics* 49: 1–10.

De Kok, A. G. 2003. Ruin probabilities with compounding assets for discrete time finite horizon problems, independent period claim sizes and general premium structure. *Insurance: Mathematics and Economics* 33: 645–658.

Donkers, B., Verhoef, P. and Jong, M. 2007. Modeling CLV: A test of competing models in the insurance industry. *Quantitative Marketing and Economics* 5(2): 163–190.

Donnelly, C., Englund, M., Nielsen, J. P. and Tanggaard, C. 2013. Asymmetric information, self-selection, and pricing of insurance contracts: the simple no-claims case. *Journal of Risk and Insurance* doi: 10.1111/j.1539-6975.2013.01520.x

Faust R., Schmeiser, H. and Zemp, A. 2012. A performance analysis of participating life insurance contracts. *Insurance: Mathematics and Economics* 51: 158–171.

Frawley, W., Piatetsky-Shapiro, G. and Matheus, C. 1991. Knowledge discovery in databases – An overview. *Knowledge Discovery in Databases*: 1–30.

Friedman, J. 2002. Stochastic gradient boosting. *Computational Statistics & Data Analysis* 38: 367–378.

Guelman, L. 2014. *uplift: Uplift Modeling.* R package version 0.3.5.

Guelman, L., Guillén, M. and Pérez-Marín, A.M. 2012. Random forests for uplift modeling: an insurance customer retention case. *Lecture Notes in Business Information Processing* 115: 123–133.

Guelman, L., Guillén, M. and Pérez-Marín, A.M. 2014. Uplift random forests. *Cybernetics & Systems* , *forthcoming.*

Guelman, L. and Guillén, M. 2014. A causal inference approach to measure price elasticity in automobile insurance. *Expert Systems with Applications* 41: 387–396.

Guillén, M., Nielsen, J. P., Scheike, T. and Pérez-Marín, A. M. 2012. Time-varying effects in the analysis of customer loyalty: a case study in insurance. *Expert Systems with Applications* 39: 3551–3558.

Guillén, M.; Pérez-Marín, A.M. and Alcaniz, M. 2011. A logistic regression approach to estimating customer profit loss due to lapses in insurance. *Insurance Markets and Companies: Analyses and Actuarial Computations* 2(2): 42–54.

Günes, E., Aksin-Karaesmen, O., Örmeci, L. and Özden, H. 2010. Modeling customer reactions to sales attempts: If cross-selling backfires. *Journal of Service Research* 13(2): 168–183.

Harrison, T. and Ansell, J. 2002. Customer retention in the insurance industry: Using survival analysis to predict cross-selling opportunities. *Journal of Financial Services Marketing* 6(3): 229–239.

Holland, P. 1986. Statistics and causal inference. *Journal of the American Statistical Association* 81(396): 945–960.

Hothorn, T., Hornik, K. and Zeileis, A. 2006. Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics* 15(3): 651–674.

Imai, K. and Ratkovic, M. 2012. Estimating treatment effect heterogeneity in randomized program evaluation. *Forthcoming in Annals of Applied Statistics.*

Jaśkowski, M. and Jaroszewicz, S. 2012. Uplift modeling for clinical trial data. *ICML 2012 Workshop on Clinical Data Analysis*, Edinburgh, Scotland, UK, 2012.

Kaishev, V., Nielsen, J. and Thuring, F. 2013. Optimal customer selection for cross-selling of financial services products. *Expert Systems with Applications* 40(5): 1748–1757.

Kamakura, W., Ramaswami, S. and Srivastava, R. 1991. Applying latent trait analysis in the evaluation of prospects for cross-selling of financial services. *International Journal of Research in Marketing* 8: 326–349.

Kamakura, W., Wedel, M., Rosa, F. and Mazzon, J. 2003. Cross-selling through database marketing: A mixed data factor analyzer for data augmentation and prediction. *International Journal of Research in Marketing* 20: 45–65.

Kamakura, W. 2008. Cross-selling: Offering the right product to the right customer at the right time. *Journal of Relationship Marketing* 6(3-4): 41–58.

Knott, A., Hayes, A. and Neslin, S.A. 2002. Next-product-to-buy models for cross-selling applications. *Journal of Interactive Marketing* 16(3): 59–75.

Landriault, D., Lemieux, C. and Willmot, G.E. 2012. An adaptive premium policy with a Bayesian motivation in the classical risk model. *Insurance: Mathematics and Economics* 51: 370–378.

Larsen, K. 2009. Net models. *M2009 - 12^{th} Annual SAS Data Mining Conference.*

Li, S., Sun, B. and Wilcox, R.T. 2005. Cross-selling sequentially ordered products: an application to consumer banking services. *Journal of Marketing Research* 42: 233–239.

Lo, V. 2002. The true lift model. *ACM SIGKDD Explorations Newsletter* 4(2): 78–86.

Pitrebois, S., Denuit, M. and Walhin, J. F. 2003. Marketing and Bonus-Malus systems. *Astin Colloquium, Berlin 24-27 August 2003*

R Development Core Team 2013. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing. Vienna, Austria. http://www.R-project.org/

Radcliffe, N. and Surry, P. 2011. Real-World Uplift Modelling with Significance-Based Uplift Trees. *Portrait Technical Report* TR-2011-1.

Rzepakowski, P. and Jaroszewicz, S. 2012. Decision trees for uplift modeling with single and multiple treatments. *Knowledge and Information Systems* 32(2): 303–327

Shaffer, J. 1995. Multiple hypothesis testing. *Annual Review of Psychology* 46: 561–584.

Strasser, H. and Weber, C. 1999. On the asymptotic theory of permutation statistics. *Mathematical Methods of*

Statistics 8: 220–250.

Su, X., Tsai, C., Wang, H., Nickerson, D. and Li, B. 2009. Subgroup analysis via recursive partitioning. *Journal of Machine Learning Research* 10(2): 141–158.

Thuring, F., Nielsen, J., Guillén, M. and Bolancé, C. 2012. Selecting prospects for cross-selling financial products using multivariate credibility. *Expert Systems with Applications* 39(10): 8809–8816.

Tian, L., Alizadeh, A., Gentles, A. and Tibshirani, R. 2012. A simple method for detecting interactions between a treatment and a large number of covariates. *Submitted on Dec 2012*. arXiv:1212.2995v1 [stat.ME].

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

Vapnik, V. 1995. *The Nature of Statistical Learning Theory*. New York: Springer.

Wahba, G. 2002. Soft and hard classification by reproducing kernel Hilbert space methods. *Proceedings of the National Academy of Sciences* 99(26): 16524–16530.

Wright, P. 1992. Adjusted p-values for simultaneous inference. *Biometrics* 48: 1005–1013.

Zhao, Y. and Zeng, D. 2012. Recent development on statistical methods for personalized medicine discovery. *Frontiers of Medicine* 7(1): 102–110.

A survey of personalized treatment models for pricing strategies in insurance

Highlights

- Models of personalized treatment learning are useful for pricing and marketing
- A summary of those models is presented for estimating changes in customer value
- Insurers can adjust price to reduce short-term benefits
- Insurers can focus on most profitable customers and do efficient cross-selling
- The causal conditional inference tree method improves coss-selling rates