Uplift models for predictive modelling in finance and insurance

Montserrat Guillén[1],
Ana M. Pérez Marín[1] and Leo Guelman[1,2]

[1]University of Barcelona and [2]Royal Bank of Canada

5 June 2014

The insurance market

- ✓ Companies operate in a more competitive environment than they used to do in the past
- ✓ Customers easily switch from one service provider to another
- ✓ Nowadays, the central problem is not only to create and launch new products for the market, but additionally to achieve commercial success by retaining customers
- ✓ Policy cancellations and lapses strongly influence the position of the company in the market and its level of risk
- ✓ Understanding customer behavior is extremely valuable for insurers

Pricing, retaining, enhancing

✓ Pricing (calculating expected loss + margin + profits)

Pricing, retaining, enhancing

✓ Pricing (calculating expected loss + margin + profits)

✓ Retaining (reducing policy lapse)

Pricing, retaining, enhancing

- ✓ Pricing (calculating expected loss + margin + profits)
- ✓ Retaining (reducing policy lapse)
- ✓ Enhancing (cross-selling additional products to existing customers)

Pricing, retaining, enhancing

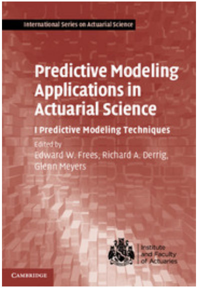
- ✓ Pricing (calculating expected loss + margin + profits)
- ✓ Retaining (reducing policy lapse)
- ✓ Enhancing (cross-selling additional products to existing customers)

  Classical approach: Predictive modeling (Negative Binomial model, Gamma model, Logistic regression, Cox,...)

Pricing, retaining, enhancing

- ✓ Pricing (calculating expected loss + margin + profits)
- ✓ Retaining (reducing policy lapse)
- ✓ Enhancing (cross-selling additional products to existing customers)

Classical approach: Predictive modeling (Negative Binomial model, Gamma model, Logistic regression, Cox,...)


International Series on Actuarial Science

Predictive Modeling Applications in Actuarial Science

I Predictive Modeling Techniques

Edited by
Edward W. Frees, Richard A. Derrig, Glenn Meyers

Institute and Faculty of Actuaries

CAMBRIDGE

Cambridge University Press, July 2014

The classical approach

✓ Price depends on expected loss and occasionally on claims experience (experience rating or bonus malus)

The classical approach

- ✓ Price depends on expected loss and occasionally on claims experience (experience rating or bonus malus)
- ✓ Price changes adjust risk transfer variations exclusively

The classical approach

- ✓ Price depends on expected loss and occasionally on claims experience (experience rating or bonus malus)
- ✓ Price changes adjust risk transfer variations exclusively
- ✓ Modeling losses by line of business is central

The classical approach

- ✓ Price depends on expected loss and occasionally on claims experience (experience rating or bonus malus)
- ✓ Price changes adjust risk transfer variations exclusively
- ✓ Modeling losses by line of business is central
- ✓ Customer retention is examined later

The classical approach

- ✓ Price depends on expected loss and occasionally on claims experience (experience rating or bonus malus)
- ✓ Price changes adjust risk transfer variations exclusively
- ✓ Modeling losses by line of business is central
- ✓ Customer retention is examined later

The classical approach

- ✓ Price depends on expected loss and occasionally on claims experience (experience rating or bonus malus)
- ✓ Price changes adjust risk transfer variations exclusively
- ✓ Modeling losses by line of business is central
- ✓ Customer retention is examined later

Observational data are available

But....do insurers have historical information that can be understood as experimental data?

Data

- ✓ Is treatment data in the insurance portfolio available?
- ✓ Have partial marketing actions been performed in the past?
- ✓ Is it possible to collect "action-response" data?

Data

- ✓ Is treatment data in the insurance portfolio available?
- ✓ Have partial marketing actions been performed in the past?
- ✓ Is it possible to collect “action-response” data?

An example:

Direct mail campaign in a bank (L=6256)

Proportion of purchase and non purchase in each treatment group

	Control	Promotion
No purchase	85.17%	61.60%
Purchase	14.83%	38.40%

Average treatment effect (uplift)=23.57%

Difficulties

- ✓ Many factors influence customer decisions, so it is difficult to predict the probability of a customer lapse and the impact of loosing a customer
- ✓ We should take into account the relationship between events affecting one particular contract and customer's decisions regarding other contracts held in the same company

- ✓ Which specific actions should a company design?
- ✓ What is the optimal price to be charged?
- ✓ Which groups of customers should be targeted in order to increase profits (reduce lapses and control price rebates)?

Model and notation

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

$$P^*_{\ell m} = LC_{\ell m} + SR_{\ell m} + B_{\ell m}, \quad \ell = \{1, \ldots, L\} \ m = \{1, \ldots, M\}$$

- a fair premium ($LC_{\ell m}$), resulting from an evaluation of the policy holder's risk characteristics, that is, an estimation of expected claims compensation or loss;
- a price loading ($SR_{\ell m}$), capturing solvency requirements, managerial efficiency or caution; and, finally,
- profits ($B_{\ell m}$), reflecting a minimum level of return to the company's shareholders or to the insurance company's owner.

Model and notation

- We define renewal $D_{\ell m}$ as a binary variable which equals 1 if policy holder ℓ renews his policy in year m, and 0 otherwise.
- Renewal $D_{\ell m}$ depends on marketing actions.
- Renewal $D_{\ell m}$ depends on external competitors.
- Renewal ($D_{\ell m}$) and price ($P^*_{\ell m}$) are mutually dependent.
- If the price increases many policy holders will abandon the company, but if the price falls then renewal is more likely than lapsing.

Our goal

- We estimate the expected change in customer value due to personalized actions (marketing campaign, price change,....).
- Or we estimate the global expected profit change due to personalized action.

We use personalized treatment models, where price change is a "treatment" (action) that predicts a "response" and combines information on :

- **risk**
- **behaviour**

General framework

- 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 $A_{\ell jm}$ the action (price change) to be offered to policy holder ℓ in year m for policy j before renewal.

General framework

- We define the set of all individual strategies as $A_m = \{A_{\ell jm}; \ell = 1, ..., L;\ \ j = 1, ..., K\}$.
- The total value at m, $V(A_m)$, is the sum of the expected profits over all customers generated from year m to M.

Value: multi-product and multi-year

- The indicator $I_{\{D_{\ell jm}=1\}}$ equals one if policy holder ℓ holds product j in year m, and 0 otherwise.
- Additionally, let $S_{\ell js}$ be the probability that customer ℓ keeps policy j in year s, namely $P(D_{\ell js} = 1)$ for $s = m, ..., M$.
- Let $B_{\ell jm}$ be the profit of policy j from policy holder ℓ in year m, and r is the interest discount factor. So the total value at m is:

$$V(A_m) = \sum_{\ell=1}^{L} \sum_{j=1}^{K} I_{\{D_{\ell jm}=1\}} B_{\ell jm} \sum_{s=m}^{M} S_{\ell js} r^{s-m}.$$

Profit: only one product and one year case

$$\underset{Z_{\ell t} \forall \ell \forall t}{\text{Max}} \sum_{\ell=1}^{L} \sum_{t=1}^{T} Z_{\ell t} \left[P_{\ell}(1 + RC_t)(1 - \hat{LR}_{\ell t})(1 - \hat{r}_{\ell t}) \right]$$

with restrictions:

$$\sum_{t=1}^{T} Z_{\ell t} = 1, \quad Z_{\ell t} \in \{0, 1\}, \quad \sum_{\ell=1}^{L} \sum_{t=1}^{T} Z_{\ell t} \hat{r}_{\ell t} / L \leq \alpha$$

where P_{ℓ} is price paid by ℓ, $\ell = \{1, 2, \ldots, L\}$, L is the total number of customers, RC_t is price change rate which is categorized in T ordered values, $t = \{1 < 2 < \ldots < T\}$, $\hat{LR}_{\ell t}$ is the loss ratio, namely, cost divided by premium, $\hat{r}_{\ell t}$ is the probability of lapse for customer ℓ if price change t is applied $(Z_{\ell t} = 1)$ and α is the maximum lapse rate that is allowed for this portfolio (so, $1 - \alpha$ is the minimum retention rate).

Motivation

- The values chosen for the actionable attributes have important implications for the ultimate profitability of the insurance company
- There is no "global" better action ⇒ Relevant in the context of treatment heterogeneity effects
- The objective is NOT to predict a response variable with high accuracy (as in predictive modeling), but to select the optimal action or treatment for each client
- Optimal personalized treatment ⇒ the one that maximizes the probability of a desirable outcome (e.g., Profits)
- Not addressed by traditional predictive modeling techniques (GLMs, CART, SVM, Neural Nets, etc.).

Background

- ✓ The demand for insurance products: Hammond, Houston and Melander (1967); Ducker (1969); Mayers and Smith (1983); Doherty (1984); Babbel (1985); Showers and Shotick (1994); Ben-Arab, Brys and Schlesinger (1996); Gandolfi and Miners (1996)
- ✓ Customer satisfaction and loyalty: Crosby and Stephens (1987); Schlesinger and Schulenburg (1993); Wells and Stafford (1995); Stafford et al. (1998); Cooley (2002), Kuo, Tsai and Chen (2003); Bozzetto et al. (2004); Guillén, Nielsen and Pérez-Marín (2006)
- ✓ Customer Lifetime Value (CLV) in insurance: Jackson (1989); Berger and Nasr (1998); Verhoef and Donkers (2001), Ryals and Knox (2005), Donkers, Verhoef and Jong (2007), Haenlein, Kaplanb and Beeserc (2007), Guillén, Nielsen and Pérez-Marín (2008)
- ✓ Cross-selling and multiple contracts: Bonato and Zweifel (2002); Guillén, Gustafsson, Hansen and Nielsen (2008), Thuring, Nielsen, Guillén and Bolancé (2012).

All approaches are based on causal/predictive modeling.

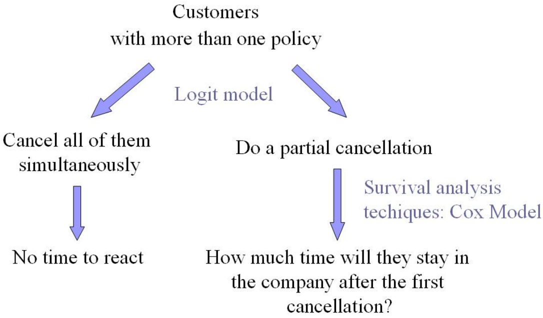
Households are customer units

- ✓ Brockett, P. L. et al. (2008) Survival Analysis of Household Insurance Policies: How Much Time Do You Have to Stop Total Customer Defection, **Journal of Risk and Insurance** 75, 3, 713-737.
- ✓ Guillen, M., Nielsen, J. P., Scheike, T. and Perez-Marin, A. M. (2011a) Time-varying effects in the analysis of customer loyalty: a case study in insurance, **Expert Systems with Applications**, 39, 3551-3558.

Households are customer units

- ✓ Using the household as the unit of analysis, we focus on the behavior of households having several policies (of three possible types: building, content and motor) in the same insurance company, and who cancel their first policy.
- ✓ How much time after the household's cancellation of the first policy does the insurer have to retain the customer and avoid customer defection on all policies to the competitor?
- ✓ What customer characteristics are associated with customer loyalty?

Methodology



Customers
with more than one policy

Logit model

Cancel all of them simultaneously

Do a partial cancellation

No time to react

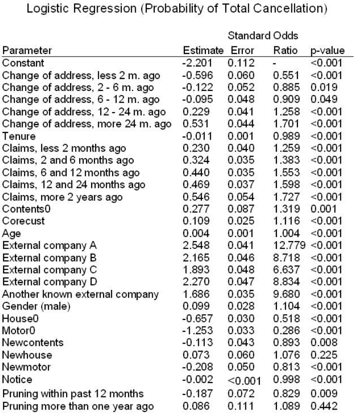
Survival analysis techiques: Cox Model

How much time will they stay in the company after the first cancellation?

Empirical study: customer information

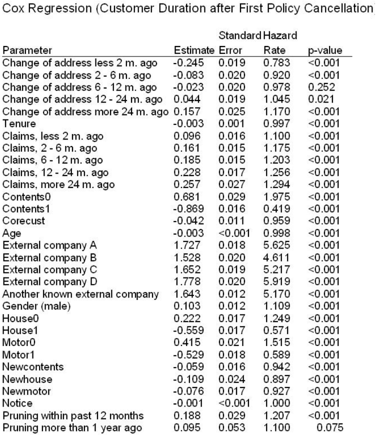
- ✓ 151,290 Danish households with several insurance policies (of three possible types: building, content and motor), who cancel their first policy between 1997-2001.
- ✓ Customer information of all household policy holders : age, gender, tenure, occurrence of a claim, change of address, intervention of a external company, premium increase, policies in force, among other covariates.

Empirical study: results



Logistic Regression (Probability of Total Cancellation)

Parameter	Estimate	Standard Error	Odds Ratio	p-value
Constant	-2.201	0.112	-	<0.001
Change of address, less 2 m. ago	-0.596	0.060	0.551	<0.001
Change of address, 2 - 6 m. ago	-0.122	0.052	0.885	0.019
Change of address, 6 - 12 m. ago	-0.095	0.048	0.909	0.049
Change of address, 12 - 24 m. ago	0.229	0.041	1.258	<0.001
Change of address, more 24 m. ago	0.531	0.044	1.701	<0.001
Tenure	-0.011	0.001	0.989	<0.001
Claims, less 2 months ago	0.230	0.040	1.259	<0.001
Claims, 2 and 6 months ago	0.324	0.035	1.383	<0.001
Claims, 6 and 12 months ago	0.440	0.035	1.553	<0.001
Claims, 12 and 24 months ago	0.469	0.037	1.598	<0.001
Claims, more 2 years ago	0.546	0.054	1.727	<0.001
Contents0	0.277	0.087	1.319	0.001
Corecust	0.109	0.025	1.116	<0.001
Age	0.004	0.001	1.004	<0.001
External company A	2.548	0.041	12.779	<0.001
External company B	2.165	0.046	8.718	<0.001
External company C	1.893	0.048	6.637	<0.001
External company D	2.270	0.047	8.834	<0.001
Another known external company	1.686	0.035	9.680	<0.001
Gender (male)	0.099	0.028	1.104	<0.001
House0	-0.657	0.030	0.518	<0.001
Motor0	-1.253	0.033	0.286	<0.001
Newcontents	-0.113	0.043	0.893	0.008
Newhouse	0.073	0.060	1.076	0.225
Newmotor	-0.208	0.050	0.813	<0.001
Notice	-0.002	<0.001	0.998	<0.001
Pruning within past 12 months	-0.187	0.072	0.829	0.009
Pruning more than one year ago	0.086	0.111	1.089	0.442



Cox Regression (Customer Duration after First Policy Cancellation

Parameter	Estimate	Standard Error	Hazard Rate	p-value
Change of address less 2 m. ago	-0.245	0.019	0.783	<0.001
Change of address 2 - 6 m. ago	-0.083	0.020	0.920	<0.001
Change of address 6 - 12 m. ago	-0.023	0.020	0.978	0.252
Change of address 12 - 24 m. ago	0.044	0.019	1.045	0.021
Change of address more 24 m. ago	0.157	0.025	1.170	<0.001
Tenure	-0.003	0.001	0.997	<0.001
Claims, less 2 m. ago	0.096	0.016	1.100	<0.001
Claims, 2 - 6 m. ago	0.161	0.015	1.175	<0.001
Claims, 6 - 12 m. ago	0.185	0.015	1.203	<0.001
Claims, 12 - 24 m. ago	0.228	0.017	1.256	<0.001
Claims, more 24 m. ago	0.257	0.027	1.294	<0.001
Contents0	0.681	0.029	1.975	<0.001
Contents1	-0.869	0.016	0.419	<0.001
Corecust	-0.042	0.011	0.959	<0.001
Age	-0.003	<0.001	0.998	<0.001
External company A	1.727	0.018	5.625	<0.001
External company B	1.528	0.020	4.611	<0.001
External company C	1.652	0.019	5.217	<0.001
External company D	1.778	0.020	5.919	<0.001
Another known external company	1.643	0.012	5.170	<0.001
Gender (male)	0.103	0.012	1.109	<0.001
House0	0.222	0.017	1.249	<0.001
House1	-0.559	0.017	0.571	<0.001
Motor0	0.415	0.021	1.515	<0.001
Motor1	-0.529	0.018	0.589	<0.001
Newcontents	-0.059	0.016	0.942	<0.001
Newhouse	-0.109	0.024	0.897	<0.001
Newmotor	-0.076	0.017	0.927	<0.001
Notice	-0.001	<0.001	1.000	<0.001
Pruning within past 12 months	0.188	0.029	1.207	<0.001
Pruning more than 1 year ago	0.095	0.053	1.100	0.075

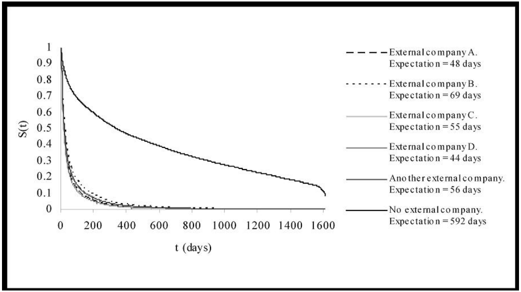
Conclusions from the empirical study

In Brockett et al. (2008) we identified relevant factors increasing the probability of total cancellation and also reducing customer duration:

- ✓ The most important factor that increases the probability of a total cancellation is the intervention of an external company.
- ✓ Claims, change of address (more than one year ago) and having a contents policy are factors that influence the probability of a total cancellation and also reduce customer duration after first policy cancellation.

Empirical study: duration after first cancellation

Example: survival function for a standard customer depending on the intervention of an external company.



Time-varying effect of covariates on loyalty

In Guillen et al. (2011a) we proved that some variables explaining customer loyalty have a time-varying effect:

- ✓ The kind of contracts held by the customer and the concurrence of an external competitor strongly influence customer loyalty right after the first cancellation, but the influence of those factors becomes much less significant some months later.
- ✓ Therefore, predictions of the probability of losing a customer can be readjusted over time to have a more precise estimation of customer duration.

Selling more policies to existing policyholders

- ✓ Guillen, M., Perez-Marin, A.M. and Alcañiz, 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.
- ✓ Thuring, F., Nielsen, J.P., Guillén, M. and Bolancé, C. (2012) Selecting prospects for cross-selling financial products using multivariate credibility, **Expert Systems with Applications**, 39, 10, 8809-8816.

Key Concepts

Guillen et al. (2011b) proposed a methodology for estimating profit loss due to policy cancellations when customers have more than one contract with the same insurer. Profits (based on the difference between premiums and cost of claims) are calculated by using three approaches:

- ✓ *Historical profit*: profit accumulated during a period of time.
- ✓ *Prospective profit*: profit that the company expects to receive if the customer keeps the policies in force. A logistic regression is proposed for estimating the probability of policy renewal.
- ✓ *Potential profit*: added profit when the customer underwrites new policies of a different type than those he currently has (cross-buying). A logistic regression is proposed for estimating the probability of selling new contracts.

Empirical application

- 79,599 customers with at least one policy in force (2005-2008).
- Two types of policies are considered: motor and diverse.
- Average profit loss due to policy lapse: 21.1 Euro per customer.

Targeting groups of customers

In Thuring et al. (2012) a method for selecting prospects for cross-selling insurance products was proposed:

- ✓ For many companies a possible way to expand its business is to sell more products to preferred customers in its portfolio.
- ✓ Historical data on customers' past behavior can be used to assess whether or not more products should be offered to a specific customer.
- ✓ In this paper we implement a method for using historical information of each individual customer, and the portfolio as a whole, to select a target group of customers to whom it would be interesting to offer more products.

Methodology

- ✓ We used multivariate credibility theory to estimate a customer specific latent risk profile and evaluate if a specific additional product, of a specific customer, is expected to contribute positively to the profit of the company, if that product is cross-sold to the customer.
- ✓ The profit is measured as the customer specific deviation between the a priori expected number of incidents (insurance claims) and the corresponding observed number.

Empirical study for cross-selling

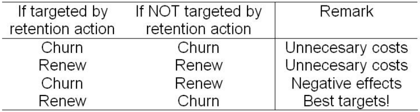
- ✓ We analyzed a sample of 3,395 customers of an insurance company (between 1999 and 2004) who have owned all of the main insurance coverages (motor, building and content).
- ✓ We find that it is easier to identify the 20 percent of the data containing customers to avoid than the 20 percent of the data containing customers to target.
- ✓ By targeting all customers but the worst 20 percent the company could expect a subset of customer associated with less claims than a priori expected indifferent of which product is considered.
- ✓ The remaining 20 percent consist of customers with up to 64 percent more claims than a priori expected.

Treatment-response: a new perspective

- ✓ Guelman, L., Guillen, M. and Perez-Marin, A. M. (2012) Random forest for uplift modeling: an insurance customer retention case, **Lecture Notes in Business Information Processing**, 115, 123-133.
- ✓ Guelman, L., Guillen, M. and Perez-Marin, A. M. (2013) Uplift random forests, **Cybernetics and Systems: an International Journal, accepted**.
- ✓ Guelman, L., Guillen, M. and Perez-Marin, A. M. (2014) A survey of personalized treatment models for pricing strategies in insurance, **Insurance: Mathematics and Economics**, under revision.

Targeting the right customers

- An insurance company is interested in increasing the retention rate of its customers.
- The point is to decide which customers should be targeted by some retention action.
- Instead of targeting the most likely to leave customers, the authors advocate that the company should target those customers with a higher expected increase in the retention probability as a result of the marketing action by using uplift modeling.



If targeted by retention action	If NOT targeted by retention action	Remark
Churn	Churn	Unnecesary costs
Renew	Renew	Unnecesary costs
Churn	Renew	Negative effects
Renew	Churn	Best targets!

Methodology:

Notation:

- $X = \{X_1, ..., X_p\}$ a vector of predictor variables,
- Y = binary response variable (1=renew, 0=lapse)
- t refers to the treatment ($t = 1$) and control ($t = 0$)
- L = a collection of observations $\{(y_\ell, x_\ell, t_\ell)\,; \ell = 1, ..., L\}$
- Uplift model $\widehat{f}^{uplift}(x_\ell) = E\left(Y_\ell | x_\ell; t_\ell = 1\right) - E\left(Y_\ell | x_\ell; t_\ell = 0\right)$

Uplift model: indirect estimation

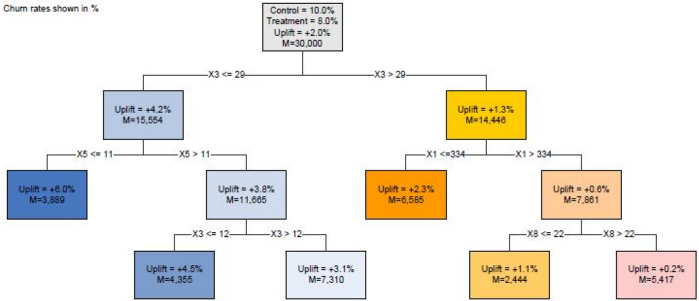
There are two general approaches: indirect and direct estimation

- Indirect uplift estimation:
    - Build two separate models, one using the treatment ($t = 1$) subset and another one using control data ($t = 0$).
      Predicted uplift is estimated by subtracting the class probabilities from the two models
      $P\left(Y = 1|x; t = 1\right) - P\left(Y = 1|x; t = 0\right)$
    - Alternatively, a single model can be obtained including an interaction term for every predictor in $X = \{X_1, ..., X_p\}$ and treatment t.
      This method does not work very well in practice, as the relevant predictors for the response are likely to be different from the relevant uplift predictors and the functional form of the predictors are likely to be different as well.

Uplift model: direct estimation

- Modeling uplift directly:
  - Requires modifying existing methods/algorithms or designing novel ones
  - Intuitively, tree-based algorithms are appropriate as they partition the input space into subgroups
  - Rzepakowski and Jaroszewicz (2011) and Radcliffe and Surry (2011) have proposed estimation algorithms
  - Our proposed algorithm: uplift Random Forests

Methodology: illustration



Methodology: uplift Random Forests

- In Guelman et al. (2012 and 2013) the proposed algorithm for modeling uplift directly is based on maximizing the distance in the class distributions between treatment and control groups
- Relative Entropy or *Kullback-Leibler distance KL* between two probability mass functions $P_t(Y)$ and $P_c(Y)$ is given by

$$KL\,(P_t(Y)||P_c(Y)) = \sum\nolimits_{y \in Y} P_t(y) \log \frac{P_t(y)}{P_c(y)}$$

.

Methodology

- Conditional on a given split Ω, KL becomes

  $$KL\left(P_t(Y)||P_c(Y)|\Omega\right) = \sum_{\omega\in\Omega}\frac{M(\omega)}{M}KL\left(P_t(Y|\omega)||P_c(Y|\omega)\right)$$

  where $M = M_t + M_c$ (the sum of the number of training cases in treatment and control groups) and $M(\omega) = M_t(\omega) + M_c(\omega)$ (the sum of the number of training cases in which the outcome of the uplift Ω is ω in treatment and control groups).
- Define KL_{gain} as the increase in the KL divergence from a split Ω relative to the KL divergence in the parent node

  $$KL_{gain}(\Omega) = KL(P_t(Y)||P_c(Y)|\Omega) - KL(P_t(Y)||P_c(Y))$$

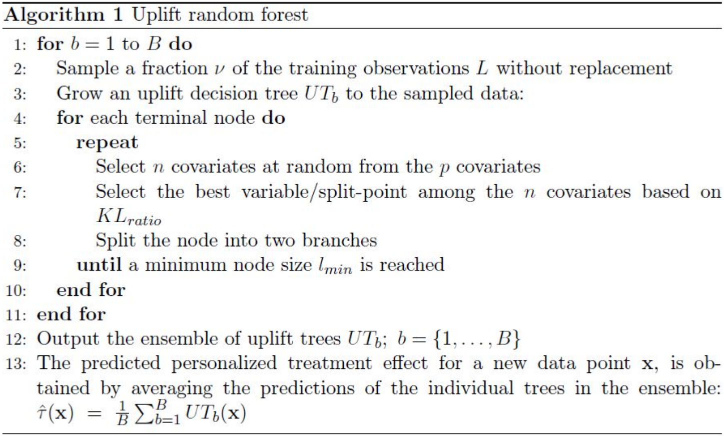
Methodology

- Final split criterion is

  $$KL_{ratio}(\Omega) = \frac{KL_{gain}(\Omega)}{KL_{norm}(\Omega)}$$

  where KL_{norm} is a normalization factor that punishes:
  - splits with different treatment/control proportions on each branch
  - splits with unbalanced number of cases on each branch

Uplift Random Forests: algorithm

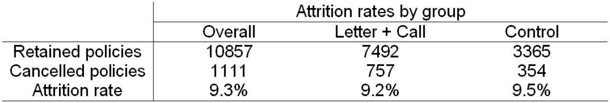


Algorithm 1 Uplift random forest

```
for $b = 1$ to $B$ do
    Sample a fraction $\nu$ of the training observations $L$ without replacement
    Grow an uplift decision tree $UT_b$ to the sampled data:
    for each terminal node do
        repeat
            Select $n$ covariates at random from the $p$ covariates
            Select the best variable/split-point among the $n$ covariates based on $KL_{ratio}$
            Split the node into two branches
        until a minimum node size $l_{min}$ is reached
    end for
end for
Output the ensemble of uplift trees $UT_b$; $b = \{1, \ldots, B\}$
The predicted personalized treatment effect for a new data point $\mathbf{x}$, is obtained by averaging the predictions of the individual trees in the ensemble: $\hat{\tau}(\mathbf{x}) = \frac{1}{B}\sum_{b=1}^{B} UT_b(\mathbf{x})$
```

Empirical study: targeting customers that react to campaigns

- Auto insurance portfolio from a large Canadian insurer
- A sample of approx. 12,000 customers coming up for renewal were randomly allocated into two groups:
    - Renewal letter+courtesy call: aim was to maximize customer retention
    - A control group: no retention efforts
    - Treatment is not much effective if targets are selected randomly

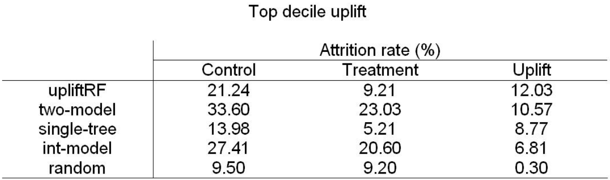


	Attrition rates by group		
	Overall	Letter + Call	Control
Retained policies	10857	7492	3365
Cancelled policies	1111	757	354
Attrition rate	9.3%	9.2%	9.5%

Empirical study

We compare four uplift models:

- Uplift Random Forest Algorithm (upliftRF)
- The Two-Model Approach by using logistic regression (two-model)
- A Single Uplift Tree with Pruning (single-tree)
- and the approach based on explicitly adding an interaction term between each predictor and the treatment indicator by using logistic regression (int-model)



Top decile uplift

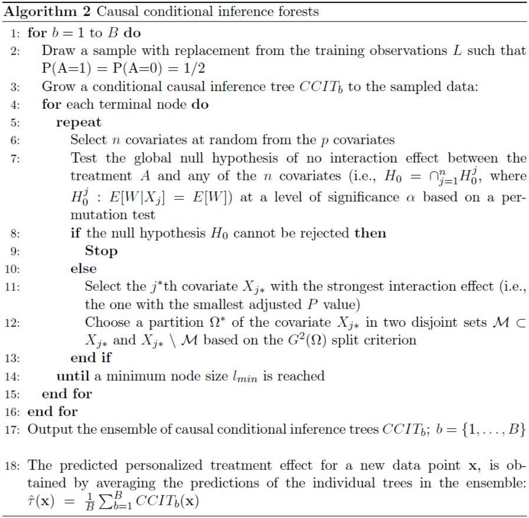
	Attrition rate (%)		
	Control	Treatment	Uplift
upliftRF	21.24	9.21	12.03
two-model	33.60	23.03	10.57
single-tree	13.98	5.21	8.77
int-model	27.41	20.60	6.81
random	9.50	9.20	0.30

Empirical study

Conclusions:

- None of the models dominates the others at all target volumes
- The *upliftRF* performs best in this application, specially for low target volumes: it is able to identify a 30 percent of customers for whom the retention program was highly effective (any additional targeted customer would result in a smaller reduction in attrition, as a result of negative effects of the campaign on the remaining customers)
- The *int-model* and *two-model* are able to identify the top 10 percent customers with highest attrition rate, but not those most impacted by the retention activity

Causal conditional inference forest: algorithm

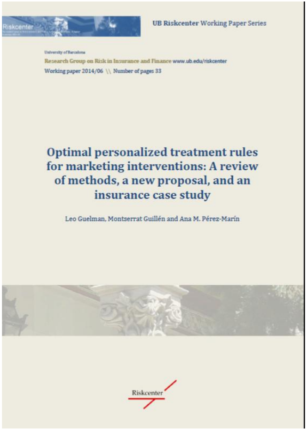


Algorithm 2 Causal conditional inference forests

```
for $b = 1$ to $B$ do
    Draw a sample with replacement from the training observations $L$ such that P(A=1) = P(A=0) = 1/2
    Grow a conditional causal inference tree $CCIT_b$ to the sampled data:
    for each terminal node do
        repeat
            Select $n$ covariates at random from the $p$ covariates
            Test the global null hypothesis of no interaction effect between the treatment $A$ and any of the $n$ covariates (i.e., $H_0 = \cap_{j=1}^{n} H_0^j$, where $H_0^j : E[W|X_j] = E[W]$) at a level of significance $\alpha$ based on a permutation test
            if the null hypothesis $H_0$ cannot be rejected then
                Stop
            else
                Select the $j^*$th covariate $X_{j*}$ with the strongest interaction effect (i.e., the one with the smallest adjusted $P$ value)
                Choose a partition $\Omega^*$ of the covariate $X_{j*}$ in two disjoint sets $\mathcal{M} \subset X_{j*}$ and $X_{j*} \setminus \mathcal{M}$ based on the $G^2(\Omega)$ split criterion
            end if
        until a minimum node size $l_{min}$ is reached
    end for
end for
Output the ensemble of causal conditional inference trees $CCIT_b$; $b = \{1, \ldots, B\}$

The predicted personalized treatment effect for a new data point $\mathbf{x}$, is obtained by averaging the predictions of the individual trees in the ensemble: $\hat{\tau}(\mathbf{x}) = \frac{1}{B} \sum_{b=1}^{B} CCIT_b(\mathbf{x})$
```

Working paper



UB Riskcenter Working Paper Series

University of Barcelona

Research Group on Risk in Insurance and Finance www.ub.edu/riskcenter

Working paper 2014/06 \\ Number of pages 33

Optimal personalized treatment rules for marketing interventions: A review of methods, a new proposal, and an insurance case study

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

Riskcenter

http://www.ub.edu/riskcenter/research/WP/UBriskcenterWP201406.pdf

Retention combined with price changes

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

The role of price in customer retention

- Understanding price sensitivities at the individual policy holder level is extremely valuable for insurers.
- A rate increase has a direct impact on the premiums customers are paying, but there is also a causal effect on the customers decision to renew the policy term.
- It is difficult to measure price elasticity from most insurance datasets, as historical rate changes are reflective of a risk-based pricing exercise, therefore they are not assigned at random across the portfolio of policyholders.
- We propose a causal inference framework to measure price elasticity in the context of auto insurance.

Data considerations

1. The gold standard for measuring causal effects (i.e., effects attributable to treatments) is to obtain experimental data
2. In the context of price-elasticity, this would involve randomizing policyholders to various rate change levels (the latter playing the role of the "treatments")
3. This condition rarely holds in practice, as usually rate changes are assigned to policyholders based on a risk-based pricing model. Thus we end up with observational data (as opposed to experimental)

Data considerations

1. The good news is that under certain data conditions (Rosenbaum and Rubin, 1983) it is still possible to obtain unbiased estimates of causal effect from observational data – that is, we can obtain unbiased estimates of price elasticities
2. Two key concepts come into play here: propensity scores and matching algorithms
3. These methods can be used to reconstruct a ”sort of” randomized study from observational data

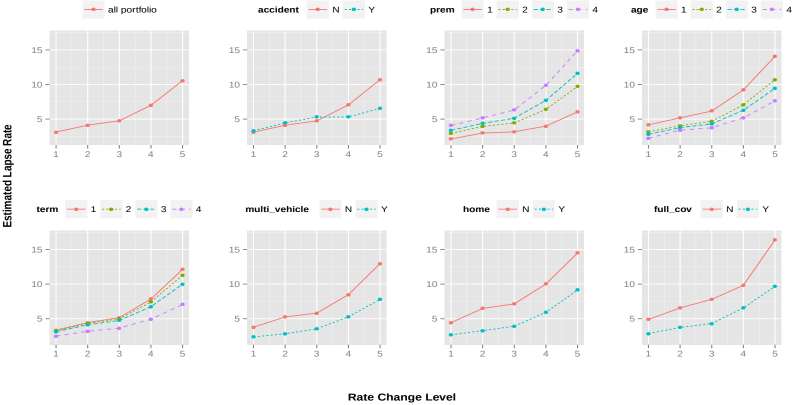
Methodology

- L policyholders, $\ell = \{1, 2, \ldots, L\}$.
- vector of pre-treatment covariates $\mathbf{x}_\ell$.
- ordered treatment variable t (rate change levels), which takes values $t = \{1 < 2 < \ldots < T\}$ on a set $\Im$.
- $Z_{\ell t}$ set of T binary treatment indicators, $Z_{\ell t} = 1$ if subject ℓ received treatment t, and $Z_{\ell t} = 0$ otherwise.
- potential responses $r_{\ell t}$, renewal outcome that would be observed from policyholder ℓ if assigned to treatment t .
- observed response for subject ℓ is $R_\ell = \sum_{t \in \Im} Z_{\ell t} r_{\ell t}$.
- Our interest is to estimate price elasticity, defined as the renewal outcomes that result and are caused by the price change interventions.

Empirical application: the data

- $L = 329,000$ auto insurance policies from a major Canadian insurer that have been given a renewal offer from June-2010 to May-2012 consisting on a new rate either lower, equal or higher than the current rate.
- more than 60 pre-treatment covariates (characteristics of the policy, the vehicle and driver).
- the treatment is the rate change: percentage change in premium from the current to the new rate, categorized into 5 ordered values $t = \{1 < 2 < \ldots < 5\}$.
- response variable: renewal outcome of the policy, measured 30 days after the effective date of the new policy term

Empirical application: estimated lapse rate



Empirical application: managerial implications

Which rate change should be applied to each policyholder to maximize the overall expected profit for the company subject to a fixed overall retention rate?

$$\underset{Z_{\ell t} \forall \ell \forall t}{\text{Max}} \sum_{\ell=1}^{L} \sum_{t=1}^{T} Z_{\ell t} \left[P_\ell (1 + RC_t)(1 - \hat{LR}_{\ell t})(1 - \hat{r}_{\ell t}) \right]$$

where P_ℓ is the current premium, RC_t is the actual rate change level associated with treatment t, $\hat{LR}_{\ell t}$ the predicted loss ratio (i.e., the ratio of the predicted insurance losses relative to premium), $\hat{r}_{\ell t}$ is the lapse probability of subject ℓ if exposed to rate change level t, and α the overall lapse rate of the portfolio.

Empirical application: managerial implications

The expected function to maximize is the expected profit of the portfolio

$$\underset{Z_{\ell t} \forall \ell \forall t}{\text{Max}} \sum_{\ell=1}^{L} \sum_{t=1}^{T} Z_{\ell t} \left[P_{\ell}(1 + RC_{t})(1 - \hat{LR}_{\ell t})(1 - \hat{r}_{\ell t}) \right]$$

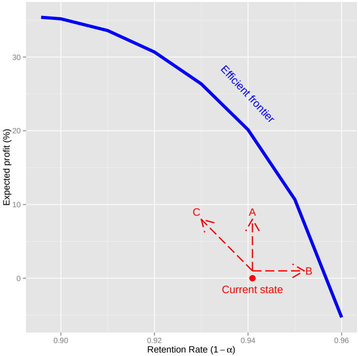
subject to the following constraints

$$\sum_{t=1}^{T} Z_{\ell t} = 1 \quad : \forall \ell$$

$$Z_{\ell t} \in \{0, 1\}$$

$$\sum_{\ell=1}^{L} \sum_{t=1}^{T} Z_{\ell t} \hat{r}_{\ell t} / L \leq \alpha$$

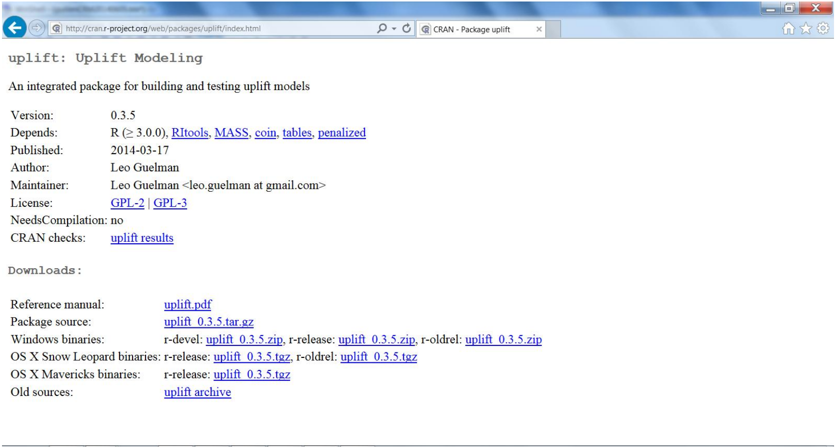
Empirical application: managerial implications



Conclusions

- We have presented an approach to estimate price elasticity functions which allows for heterogeneous causal effects as a result of rate change interventions
- The model can assist managers in selecting an optimal rate change level for each policyholder for the purpose of maximizing the overall profits for the company
- Valuable insights can be gained by knowing the current company's position of growth and profitability relative to the optimal values given by the efficient frontier
- The managerial decision is to determine in which direction the company should move towards the frontier, as each decision point places a different weight on each of these objectives.

An R programme package called **uplift**



uplift: Uplift Modeling

An integrated package for building and testing uplift models

Version:	0.3.5
Depends:	R (≥ 3.0.0), RItools, MASS, coin, tables, penalized
Published:	2014-03-17
Author:	Leo Guelman
Maintainer:	Leo Guelman <leo.guelman at gmail.com>
License:	GPL-2 | GPL-3
NeedsCompilation:	no
CRAN checks:	uplift results

Downloads:

Reference manual:	uplift.pdf
Package source:	uplift_0.3.5.tar.gz
Windows binaries:	r-devel: uplift_0.3.5.zip, r-release: uplift_0.3.5.zip, r-oldrel: uplift_0.3.5.zip
OS X Snow Leopard binaries:	r-release: uplift_0.3.5.tgz, r-oldrel: uplift_0.3.5.tgz
OS X Mavericks binaries:	r-release: uplift_0.3.5.tgz
Old sources:	uplift archive

R implementation: The **uplift** package in CRAN

The highlights:

- Implements various functions for training personalized treatment learning models (a.k.a., uplift)
- Currently 5 estimation methods are implemented
    - Causal conditional inference forests (`ccif`)
    - Uplift random forests (`upliftRF`)
    - Modified covariate method (`tian_transf`)
    - Modified outcome method (`rvtu`)
    - Uplift k-nearest neighbor (`upliftKNN`)
- Exploratory Data Analysis (EDA) tools designed for PTE models
- Functions for evaluating performance of PTE models
- Profiling results of PTE models
- PTE Monte Carlo simulations
- Package in continuous development

A case study: Direct mail campaign

```
> treat.form < − treatment   age+gender+withdrawals+deposit+credit_value+discounts
+ + transactions+bank_logs+accruals+charges+cash_total+loan_payment+e_trans
> cb < − checkBalance(treat.form , data = bankDM.train)
> Model.form < − response   trt(treatment)+age+gender+withdrawals+deposit+credit_value+discounts
+ + transactions+bank_logs+accruals+charges+cash_total+loan_payment+e_trans
> niv_res < − niv(Model.form, B = 100, nbins = 4, data = bankDM.train)
> eda < − explore(Model.form, data = bankDM.train)
> ### Causal conditional inference forests (ccif)
> set.seed(1)
> ccif_fit1 < − ccif(Model.form,
+ data = bankDM.train,
+ ntree = 1000,
+ split_method = ”Int”,
+ distribution = approximate (B=999),
+ verbose = TRUE)
> op < − par(mar = c(5, 10, 4, 2) + 0.1)
> varImportance(ccif_fit1, plotit = TRUE)
```

A case study: Direct mail campaign

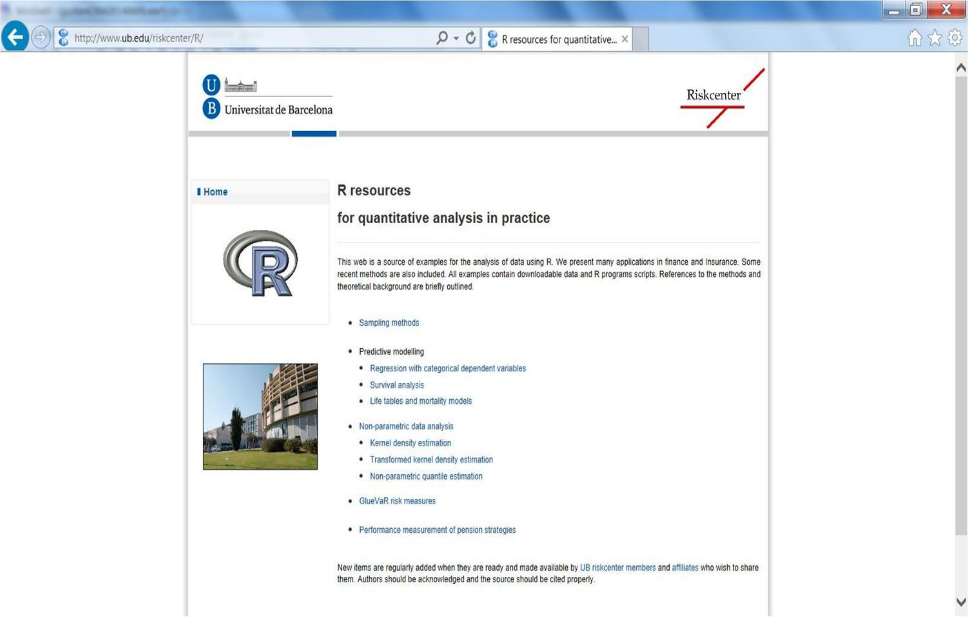
```
> ### Uplift random forests (upliftRF)
> set.seed(1)
> upliftRF_fit1 <- upliftRF(Model.form,
+ data = bankDM.train,
+ ntree = 1000,
+ interaction.depth = 3,
+ split_method = ”KL”,
+ minsplit = 50,
+ verbose = TRUE)
> ### Modified outcome method (mom)
> set.seed(1)
> bankDM.train.mom < − rvtu(Model.form, data = bankDM.train,
+ method = ”undersample”)
> Model.form.mom < − z ãge+gender+withdrawals+deposit+credit_value+discounts+transactions
+ +bank_logs+accruals+charges+cash_total+loan_payment+e_trans
> glm.fit1 < − glm(Model.form.mom, data = bankDM.train.mom, family = ”binomial”)
> ### Perform stepwise model selection by AIC
> glm.fit_step < − stepAIC(glm.fit1, direction = ”backward”, trace = 0)
```

A case study: Direct mail campaign

```
> ### Get predictions on test data
> pred_upliftRF < − predict(upliftRF_fit1, bankDM.test)
> pred_ccif < − predict(ccif_fit1, bankDM.test)
> pred_mom < − 2 * predict(glm.fit_step, bankDM.test) - 1
> ### Get uplift by decile
> ccif_perf < − performance(pred_ccif[, 1], pred_ccif[, 2],
+ bankDM.test$response, bankDM.test$treatment)
> upliftRF_perf < − performance(pred_upliftRF[, 1], pred_upliftRF[, 2],
+ bankDM.test$response, bankDM.test$treatment)
> mom_perf < − performance(pred_mom, rep(0, length(pred_mom)),
+ bankDM.test$response, bankDM.test$treatment)
> ### 1st to 3rd decile uplift
> Decile_3_perf < − data.frame(
+ ccif = (sum(ccif_perf[1:3, 4])  sum(ccif_perf[1:3, 2])) -
+ (sum(ccif_perf[1:3, 5])  sum(ccif_perf[1:3, 3])),
+ upliftRF = (sum(upliftRF_perf[1:3, 4])  sum(upliftRF_perf[1:3, 2])) -
+ (sum(upliftRF_perf[1:3, 5])  sum(upliftRF_perf[1:3, 3])),
+ mom = (sum(mom_perf[1:3, 4])  sum(mom_perf[1:3, 2])) -
+ (sum(mom_perf[1:3, 5])  sum(mom_perf[1:3, 3]))
+ )
```

A case study: Direct mail campaign

```
> Decile_3_perf
                                 ccif uplif tRF mom
                        1 0.3673691 0.3590956 0.3571356
> ### qini-coefficient
> qini < − data.frame(ccif = qini(ccif_perf, plotit = FALSE)$Qini,
+ upliftRF = qini(upliftRF_perf, plotit = FALSE)$Qini,
+ mom = qini(mom_perf, plotit = FALSE)$Qini)
> qini
                                 ccif upliftRF mom
                        1 0.02948474 0.03140241 0.02639283
```



R resources

for quantitative analysis in practice

This web is a source of examples for the analysis of data using R. We present many applications in finance and Insurance. Some recent methods are also included. All examples contain downloadable data and R programs scripts. References to the methods and theoretical background are briefly outlined.

- Sampling methods
- Predictive modelling
  - Regression with categorical dependent variables
  - Survival analysis
  - Life tables and mortality models
- Non-parametric data analysis
  - Kernel density estimation
  - Transformed kernel density estimation
  - Non-parametric quantile estimation
- GlueVaR risk measures
- Performance measurement of pension strategies

New items are regularly added when they are ready and made available by UB riskcenter members and affiliates who wish to share them. Authors should be acknowledged and the source should be cited properly.

www.ub.edu/riskcenter/R