Modelling Car Insurance Churn at AEGON

A Comparison of Approaches, from Parametric to Non-Parametric

Hanneke Tiktak
Sem Koelewijn
Sjoerd Bommer
Xingbao Fu

Erasmus School of Economics

March 22, 2024

Outline

Research question

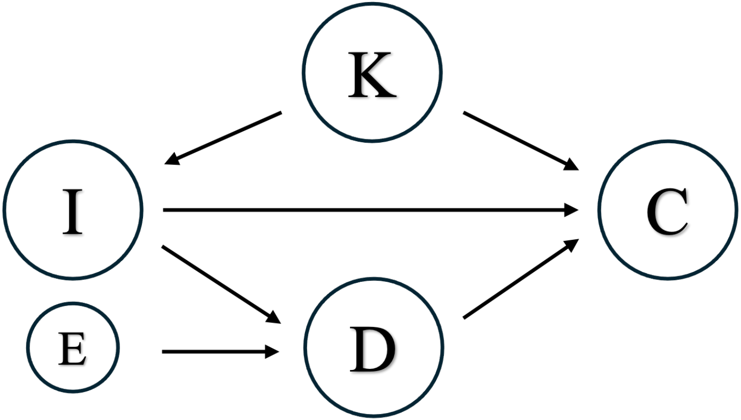
Research question and sub-questions

Does churning behaviour of customers change from year to year if a welcome discount or list price adjustment is applied, and if so, how?

1. Can different segments of customers be found, and is their churning behaviour different?
2. What is the effect of welcome discounts and list price adjustments on churning behaviour, and can this be determined causally?
3. From a parametric to a non-parametric approach, how well can future churn be predicted?

- Clustering
- Discrete-time Survival Model
- Hierarchical Bayesian Logistic Regression
- Random Survival Forest

Relevant variables



Directed acyclic graph of causal model

Investigating customer heterogeneity

Clustering: Relevance & Methods

- Discover segments of customers that are similar based on customer characteristics like age, accident free years, etc
- Examine the impact of discounts on similar customers who differ only in whether they received a discount or not
- 2 cluster methods: K-prototypes and DBSCAN
  - K-Prototypes: Combination of K-Means (numerical variables) and K-Modes (categorical variables). Within cluster distance of observations is minimized. Number of clusters is pre-divined
  - DBSCAN: Density based clustering. Observations are clustered based on a certain distribution. The number of clusters is not pre-divined
- Considered clusterings ranging from 2 to 7 clusters

- 3 validation criteria related to within cluster separation and between cluster distances
- Higher scores correspond to better clusterings (except for the Davies-Bouldin Score)
- Normalized for comparison
- Normalized Davies-Bouldin score flipped along 0.5 y-axis for visual comparison
- K-Prototypes-2 best on all 3 criteria, therefore preferred method

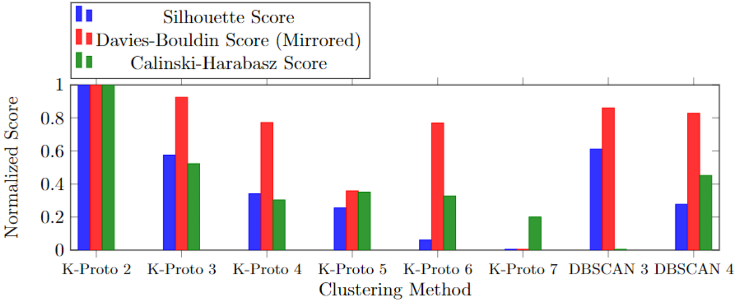
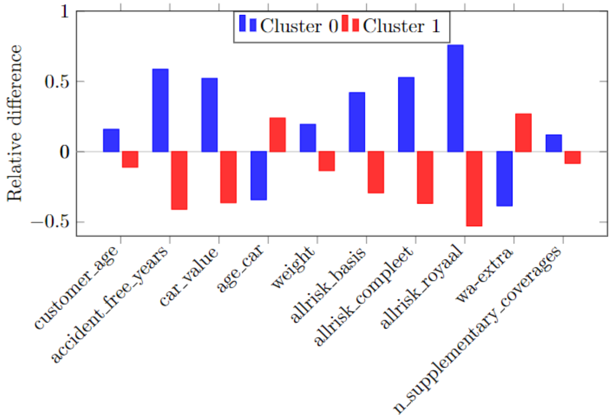


Figure: Normalized Clustering Validation Scores

Clustering: Interpretation

- Difference of the mean of the clusters compared to the overall mean for each numerical variable
- Cluster 0: Older customers characterized by an increased average of accident-free years and ownership of newer, higher-value vehicles, often insured extensively.
- Cluster 1: Relatively younger customers with fewer accident- free years on average, and possession of older, less valuable vehicles, frequently insured without extensive coverage.



Interpretable (causal) effects

Discrete-time survival model

Motivation:

- Reason for survival model: Right-censored "time-to-event" data.
    - A record of the time elapsed before the interesting event happens.
    - In this case, time is the year since the policy started, and the event is whether the customer churns.
    - Right censored means that some customers have not churned due to the limited data collection period, but in reality, they will churn eventually.
- Reason for discrete-time: Data is recorded yearly, and customers' information stays the same within the interval.
- High interpretability and easy to use.

Model Setting

The most important ideas of this model are the **hazard rate**, the **survival probability**, and the **probability of churn**. In a discrete-time model, the hazard rate is the conditional probability of a customer churn per year, which is based on the other variables. The survival probability is the unconditional probability of customers staying after interval s_i, and equation (3) shows the likelihood of churn.

$$h_{is} = \frac{1}{(1 + exp(-[\alpha_w^T Year_{is} + w_i(\beta_w^T Z_{is} + \gamma_w^T X_i) + (1 - w_i)(\beta^T Z_{is} + \gamma^T X_i)]))} \tag{1}$$

$$S_i(t) = \Pr(T_i \geq s_i) = \prod_{s=1}^{s_i}(1 - h_{is}) \tag{2}$$

$$\Pr(T_i = s_i) = h_{is_i} \prod_{s=1}^{s_i - 1}(1 - h_{is}) \tag{3}$$

Results of the Discrete-Time Survival Model

Results:

- A higher discount level causes a lower probability to churn in the year.
- People with welcome discount/LPA eligibility have lower hazard rates.
- Clusters calculated are not significant in this model.
- Welcome discount eligibility within the discount group. The positive average marginal effect is probably due to the selection bias of the discount group.

Variables	Estimator (SE)	Average Marginal Effect
size of WD	-0.479 (0.166)	-0.054 (0.019)
WD eligible	-0.105 (0.028)	-0.012 (0.003)
LPA eligible	-0.369 (0.030)	-0.042 (0.003)
cluster 1	-0.033 (0.033)	-0.004 (0.004)
cluster 1 (with WD)	0.007 (0.05)	-0.001 (0.006)
WD eligible (with WD)	1.046 (0.37)	0.117 (0.042)
LPA eligible (with WD)	-0.062 (0.04)	-0.008 (0.004)

Causal analysis

Due to the problem set-up, a standard instrumental variable approach to determine (local) average treatment effects cannot be taken. Bayesian regression techniques pose an attractive alternative.

$$\max_{\theta} p(y|\theta) \quad \rightarrow \quad p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} \propto p(y|\theta)p(\theta)$$

Advantages/drawbacks:

- More nuanced results: the parameter estimates are probability distributions.
- Prior distributions $p(\theta)$ can be set based on previous experience.
- Panel data can be incorporated more easily.
- Only data from new customers from 2021 and onwards can be used.

Hierarchical Bayesian Logistic model

$$
\begin{aligned}
c_{it} &\sim Bernoulli(p_{it}) \\
ln\left(\frac{p_{it}}{1-p_{it}}\right) &= \alpha_t + \phi_{E_i,t} + \beta_t' I_{it} + \gamma_t' K_{it} \\
\phi_{E_i,t} &= \alpha_{E_i,t} + \beta_{E_i,t}' I_{it}
\end{aligned} \tag{4}
$$

Parameter priors are set to be uninformative: normally distributed with a zero mean and large variance. The variance term is generated from a half-Cauchy distribution.

Parameters are assumed to be correlated over time. The correlation factor is assumed to be roughly uniformly distributed between -1 and 1.

Results:

- Only a few coefficients did not include zero in their smallest 95% density interval.
- The cluster label did not serve as a useful additional variable, either as a dummy or as a factor to split the data.
- Average predicted churn probabilities were higher for the list-price adjustment group, and lower for the welcome discount group (compared to the control group).

Prediction and model performance

Retention probability predictions and models performance

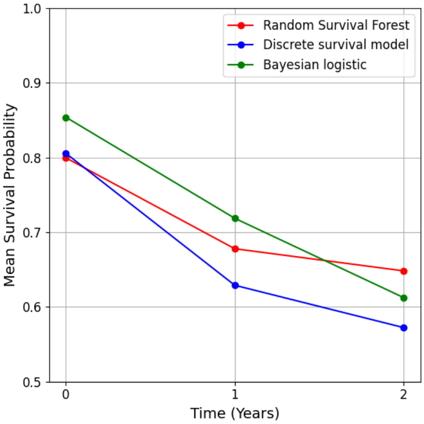
Idea: make model to compare with the survival discrete model for comparing **model performance** and make **survival predictions**

- **Random Survival Forest**
  - Extension of the random forest
  - Selects the splitting feature and value based on the **log-rank test**
  - Predicts the time until churn
  - High predictive performance
  - Handles **right censored time to event** data
  - capture **complex non linear relationships**
  - **Output:** feature importance, risk scores, and survival predictions

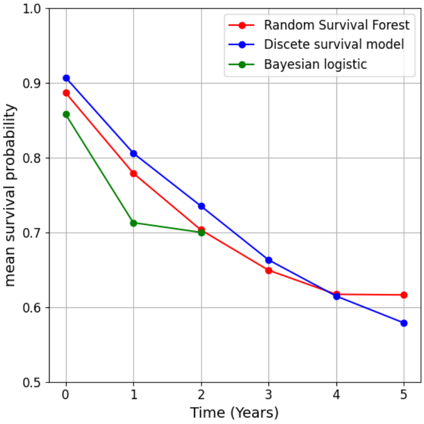
What we did:

- Computed the survival probabilities over time, for the **non-discount** and **discount** group, for the causal and survival models.
- Compared the mean survival function and predictive performance using **Brier score** and **AUC**.

Survival probabilities predictions



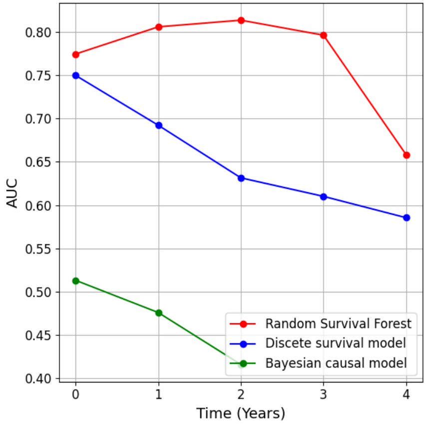
(a) welcome discount



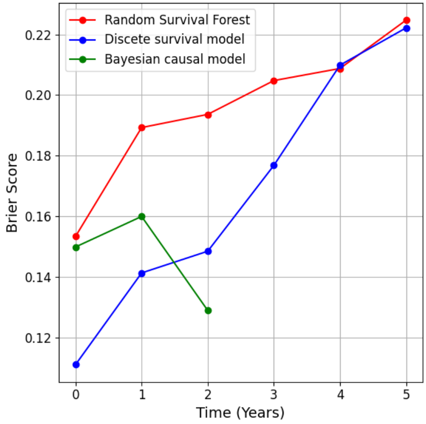
(b) no welcome discount

Comparison of different predicted survival probabilities

Model Performance



(a) AUC-score over time



(b) Brier-score over time

Comparison of AUC and Brier Score

Conclusion and outlook

Conclusions

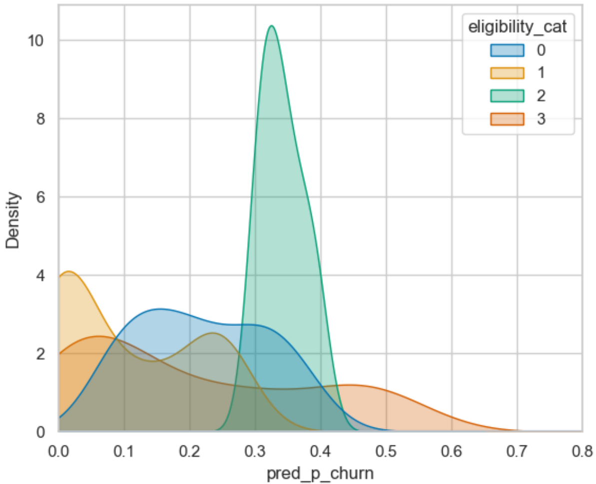
- Clusters did not have a significant impact on churning behaviour for all 3 models.
- For the discrete-time survival model, we see that people who received the discount actually have a higher chance of churn. This could be caused by:
  - the discount itself;
  - AEGON's selection mechanism.
- Causal modelling is possible, but the developed model could not produce useful results.
  - Assumptions might need to be simplified.
  - More information about the discount-assignment process is needed.
- Random survival forests predict the binary outcome better, but the discrete-time survival model provides the most accurate churning probabilities.

Which model to use?

- Trade off between explainability and prediction performance
- The causal model has the highest explainability, while the random forest survival model has the lowest.
- For now, we recommend the discrete-time survival model if explainable results are required, although it does not provide causal results.

Appendix

Bayesian churn posterior probability densities



Predictive posterior density of churn probabilities in the first year, by discount eligibility category

Appendix: DTS regression result

Variable	Coefficient	Std. Err.	z	P> $\|z\|$	[0.025	0.975]
total premium	0.5401	0.028	18.994	0.000	0.484	0.596
customer age	0.0008	0.001	0.946	0.344	-0.001	0.002
accident free years	-0.0177	0.001	-12.048	0.000	-0.021	-0.015
car value	0.0920	0.013	7.245	0.000	0.067	0.117
age car	0.0282	0.002	12.556	0.000	0.024	0.033
allrisk basis	-0.6659	0.053	-12.582	0.000	-0.770	-0.562
allrisk compleet	-1.0469	0.036	-29.483	0.000	-1.117	-0.977
allrisk royaal	-1.2335	0.047	-26.369	0.000	-1.325	-1.142
wa-extra	-0.5685	0.027	-20.818	0.000	-0.622	-0.515
n supplementary coverages	-0.0914	0.018	-5.004	0.000	-0.127	-0.056
years since policy started 0	-5.9432	0.498	-11.927	0.000	-6.920	-4.966
years since policy started 1	-5.6280	0.498	-11.293	0.000	-6.605	-4.651
years since policy started 2	-5.7822	0.499	-11.596	0.000	-6.759	-4.805

Variable	Coefficient	Std. Err.	z	P$>|z|$	[0.025	0.975]
years since policy started 3	-5.7536	0.500	-11.518	0.000	-6.733	-4.775
years since policy started 4	-5.8984	0.503	-11.728	0.000	-6.884	-4.913
years since policy started 5	-8.0449	0.871	-9.240	0.000	-9.751	-6.338
fuel type benzine	-0.1602	0.445	-0.360	0.719	-1.033	0.713
fuel type diesel	0.1275	0.446	0.286	0.775	-0.747	1.002
fuel type electro	-0.1662	0.451	-0.369	0.712	-1.049	0.717
fuel type gas	0.2361	0.478	0.493	0.622	-0.702	1.174
fuel type hybride	-0.1029	0.489	-0.210	0.833	-1.061	0.856
sales channel AEGON.nl	0.3485	0.039	8.851	0.000	0.271	0.426
sales channel Independer	0.6954	0.069	10.032	0.000	0.560	0.831
WD total premium	0.2516	0.041	6.149	0.000	0.171	0.332
WD accident free years	-0.0158	0.003	-5.946	0.000	-0.021	-0.011
WD car value	0.0614	0.025	2.492	0.013	0.013	0.110
WD customer age	-0.0022	0.001	-1.867	0.062	-0.004	0.000

Variable	Coefficient	Std. Err.	z	P> |z|	[0.025	0.975]
WD age car	0.0361	0.003	10.593	0.000	0.029	0.043
WD n supplementary coverages	-0.1182	0.022	-5.266	0.000	-0.162	-0.074
WD cluster 1.0	-0.0106	0.053	-0.200	0.842	-0.115	0.094
WD fuel type benzine	1.3689	0.640	2.140	0.032	0.115	2.622
WD fuel type diesel	1.7690	0.643	2.750	0.006	0.508	3.030
WD fuel type electro	1.6329	0.645	2.533	0.011	0.370	2.896
WD fuel type gas	0.7590	0.801	0.947	0.344	-0.811	2.330
WD fuel type hybride	1.4883	0.744	2.001	0.045	0.031	2.946
WD sales channel AEGON.nl	0.3759	0.040	9.399	0.000	0.298	0.454
WD sales channel Independer	0.5396	0.035	15.472	0.000	0.471	0.608
WD allrisk basis	-0.5454	0.062	-8.734	0.000	-0.668	-0.423
WD allrisk compleet	-0.5828	0.056	-10.353	0.000	-0.693	-0.472
WD allrisk royaal	-0.5862	0.076	-7.671	0.000	-0.736	-0.436
WD wa-extra	-0.3677	0.036	-10.118	0.000	-0.439	-0.296