Stochastic Solutions

# Identifying who can be saved and who will be driven away by retention activity

## Using uplift modelling to reduce attrition in financial services

**Stochastic Solutions Limited** **T** +44 7713 787 602 **E** Nicholas.Radcliffe@StochasticSolutions.com http://StochasticSolutions.com

# Abstract

It has been repeatedly demonstrated that the very act of trying to 'save' some customers provokes them to leave. This is not hard to understand, for a key targeting criterion is usually estimated attrition probability, and this is highly correlated with customer dissatisfaction. Often, it is mainly lethargy that is preventing a dissatisfied customer from actually leaving. Interventions designed with the express purpose of reducing customer loss can provide an opportunity for such dissatisfaction to crystallise, provoking or bringing forward customer departures that might otherwise have been avoided, or at least delayed. This is especially true when intrusive contact mechanisms, such as outbound calling, are employed. Retention programmes can be made more effective and more profitable by switching the emphasis from customers with a high probability of leaving to those likely to react positively to retention activity. This paper discusses how targeting on the basis of such '*savability*' can be achieved, illustrating the effectiveness of the approach with a case study. Insofar as a paper can be summarised in a motto, this paper's is *"savability is the key to retention activity"*.

> *"Savability is the key to retention"*

## Management Implications

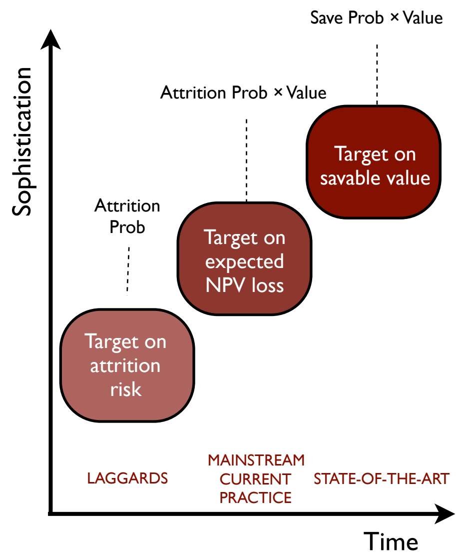
- Retention programmes can increase attrition as well as reduce it.
- Attrition risk alone is not always a good basis for intervention; neither is value-adjusted attrition risk.
- The customers at greatest risk of attrition are not necessarily the most cost-effective to save.
- Financial institutions not already using control groups in their retention programmes should adopt them as a matter of priority. Properly randomised control groups of adequate size provide the only proven and reliable way of assessing the true impact of retention activity.
- Successful retention programmes (i.e. ones that reduce attrition) can often be made more profitable by retargeting on the basis of *savability*. In general, this will remove from the target group a set of customers for whom the programme's impact is negative or marginal, reducing cost and—in some cases—increasing overall retention. It may also add a set of customers for whom the retention activity can be positive, but who are not identified by conventional targeting models.
- Even counterproductive retention programmes (ones that increase overall attrition) often have a positive impact for some segments of the population. Retargeting on the basis of savability may allow such counterproductive programmes to become effective and profitable.
- Conventional approaches to modelling attrition probability are structurally incapable of modelling savability directly. However, new modelling methods exist that are capable of predicting customer savability; the resulting models are known as *uplift models*.

# 1. Introduction

Over recent years, a number of factors have combined to increase the rate at which customers switch between financial institutions. These factors include regulatory changes, evolution of consumer attitudes, the rise of comparison sites on the web and the increasing use of the telephone and internet as primary channels of interaction for the customer. Against this backdrop, it is unsurprising that customer retention has assumed an increasing importance for most institutions.

In contrast to those in some other industries, retention efforts in the financial sector have always tended to be fairly targeted. When targeted retention efforts began, they tended to be driven primarily by predicted attrition risk, particularly at key times such as the end of a contractual lock-in or discount period. More recently, as institutions have focused more on customer profitability, it has become more common for modelled attrition risk to be weighted by some kind of customer value metric, so that in effect the key driver of how likely a customer is to be included in a retention programme has become *expected value loss* (probability of attrition multiplied by estimated net present value). The most advanced institutions are starting to adopt a different approach, based on modelling *savability*. It is this approach that is the focus of this paper.




## 1.1. Negative Effects

Accepted best practice for assessing the true impact of marketing actions requires the use of systematically randomised control groups. These provide a statistical baseline against which changes in behaviour resulting from interventions can be measured. As these have been increasingly employed by financial institutions running retention programmes, many have been shocked to observe that some customer segments are adversely affected by the retention activities employed. To be clear: the very act of trying to save some customers provokes them to leave. Far from being a mere statistical aberration, this is a real phenomenon. In the most extreme cases, the *overall* impact of the retention programme has been found to be counterproductive. There are a number of explanations for this.

> *"The very act of trying to save some customers provokes them to leave."*

1. Most customers at high risk of attrition are dissatisfied in some way, perhaps because of service, price or conditions. While some such customers will actively approach the institution to terminate, others will not, perhaps as a result of lethargy or resignation. In such cases, active contact with the customer, particularly through an interactive medium such as a phone call, can act as a catalyst, crystallising an attrition event that might otherwise have occurred much later or not at all.

2. Additionally, many customers are antagonised by what they feel to be intrusive contact mechanisms; indeed, we guess that relatively few customers are thrilled, on hearing their phone ring, to discover that the

caller is their bank or insurer. In some cases, particularly for customers who are already unhappy, such perceived intrusions may act not merely as a catalyst but as a constituent cause of attrition.

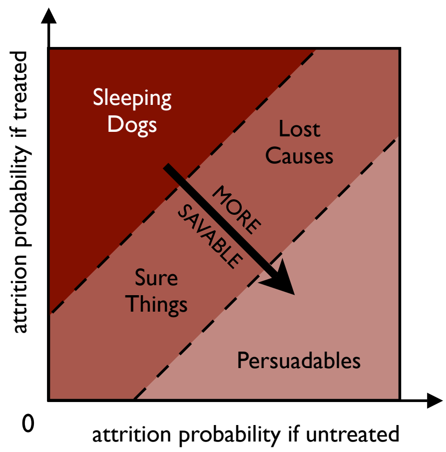
3. Finally, there is also simple awareness of the impending end of a lock-in period, a discount period or an annual renewal date. While most institutions have systems to alert them when customers approach the end of a lock-in period, the customers themselves generally do not. A retention action may inform or remind a customer that he or she has an opportunity to move, and even a fairly satisfied customer may at that point look around for other options. When this happens it is inevitable that a proportion of people will choose to move.

## 1.2. Optimal Targeting

As soon as it is accepted that interventions intended to save customers are a double-edged sword, it becomes obvious that targeting primarily on the basis of predicted attrition risk is a dangerous strategy, the more so if the correlation between attrition probability and dissatisfaction is accepted.

> *"The customers most at risk are not always the easiest to influence."*

The important corollary that this also brings into focus is that even among customers for whom the retention activity is neutral to positive, those at greatest risk of attrition are not necessarily those most positively affected by our actions. Indeed, given the correlation between dissatisfaction and attrition probability, there are some grounds for thinking that high-risk customers may be some of the hardest people to influence positively. A segmentation of customers by their different possible reactions to retention activity is shown in the schematic to the right.




**Sleeping Dogs** tend to be driven away by treatment

**Lost Causes** are likely to leave whether treated or not

**Sure Things** are likely to stay whether treated or not

**Persuadables** are made more likely to stay by treatment

The traditional approach is to model the probability of customer attrition. As the schematic illustrates, there are in fact two different attrition probabilities—the probability of attrition without intervention (horizontal axis) and the probability of attrition when the customer is subject to the retention activity in question (vertical axis). Which of these a particular institution models is normally determined by how much retention activity that institution undertakes. If the policy tends towards 'leaving no customer behind', it may be that most or all high-risk customers are included in retention programmes and this will tend to mean that the only attrition probability that can easily be modelled is the probability of attrition when treated, which we denote $p_A^T$. (Such models are similar to so-called "response" models, which are often used in marketing.[1]) In contrast, when there is less retention activity, the tendency will be to model attrition probability for untreated customers, which we denote $p_A^U$. (Such models are more like penetration models.[1]) In practice, the modelling populations are sometimes even mixed. Targeting on the basis of treated attrition probability, $p_A^T$, tends to focus attention on customers above some horizontal line on the schematic, typically taking in most of the Sleeping Dogs, and Lost Causes, but fewer of the Sure Things and Persuadables. Conversely, targeting on the basis of the untreated attrition probability, $p_A^U$, focuses attention on customers to the right of some vertical line on the schematic. This is better, in that it will tend to cause more of the Per-

[1] Shepard, D. (1999). "Penetration Models vs. Response Models." Direct. 1st August 1999. Available electronically at `http://directmag.com/mag/marketing_penetration_models_vs/`.

suadables to be targeted, but it will still capture many Lost Causes and a fair number of Sure Things and Sleeping Dogs.

Current mainstream practice weights these customers by some kind of value metric. This helps in the case of high-value Persuadables, but exacerbates the problem with high-value Sleeping Dogs.

Clearly a better approach is to model *savability* (the difference between the treated and untreated attrition probabilities), a quantity that increases in the direction of the arrow on the schematic. This allows targeting of the Persuadables without wasting money on the Sure Things and Lost Causes or worse, spending money to drive away the Sleeping Dogs.

# 2. Evidence and Measurement

## 2.1. Case Study: Insurance Retention

We will illustrate the effects we have discussed with reference to a retention campaign carried out by a European insurer. Overall, this campaign was successful and profitable before uplift modelling was introduced, but the impact of uplift modelling on both overall retention and campaign profitability was dramatic and positive. Key metrics for these two cases are shown in the table.

As can be seen from the figures, the overall impact of the retention campaign when everyone was targeted (the original strategy) was very positive, reducing the rate of customer loss from around 32% to 28% This campaign was very profitable. The natural assumption for the insurer was that the campaign was having the maximum impact that it could (since everyone was included), and that while there might be some financial benefit to reducing the targeting volume, this was likely to be small, and at the cost of somewhat higher attrition.

As we will see, despite the very healthy numbers from this campaign, the presence of hidden negative effects meant that it was actually possible to improve the situation a great deal.

**Key metrics from Insurance Retention Campaign**

| | |
|---|---|
| Attrition rate in control group | 32% |
| Attrition rate in treated group | 28% |
| Overall uplift (reduction in attrition; pc pt) | 4% |
| Size of control group | 100,000 |
| Size of treated group | 20,000 |
| Best targeting volume cutoff found | c. 70% |
| Overall uplift (reduction in attrition) at best cutoff (pc pt) | 5% |
| Proportion of population required to equal or exceed useful impact of targeting everyone | c. 40% |
| Approximate campaign profitability when targeting 1m customers | c. €0.5m |
| Approximate campaign profitability when 1m targeted using an uplift model | c. €2m |
| Approximate annual incremental financial impact of adopting uplift per 1m customers | c. €1.5m |

## 2.2. Uplift and Savability

If we want to target on the basis of savability, we need to be able to estimate this quantity for each customer.[2] We cannot measure savability at a customer level because we cannot simultaneously treat and not treat a customer. Estimating savability is a *modelling* problem.

We must first be very clear that conventional attrition models do *not* predict savability. A attrition model built on a historical population of customers who have been subject to retention activity can be used to estimate the probability that a customer will leave if subjected to that retention action:

$$p_A^T = \text{Prob (attrition | treatment)}.$$

[2] Radcliffe, N. J. (2007). "Using Control Groups to Target on Predicted Lift: Building and Assessing Uplift Models", Direct Marketing Analytics Journal, Direct Marketing Association, 2007.

Conversely, an attrition model built on a historical population of customers who have not been subject to retention activity allows us to estimate the probability that a customer will leave if *not* subjected to the retention activity in question:

$$p_A^U = \text{Prob (attrition} \mid \text{no treatment).}$$

Savability is the difference between these, so that if someone with an estimated attrition probability of 27% when treated and 32% when untreated has a savability of 5 percentage points (5pp). We define the uplift U (the net impact of the treatment) as

$$U = p_A^T - p_A^U = \text{Prob (attrition} \mid \text{treatment)} - \text{Prob (attrition} \mid \text{no treatment)}$$

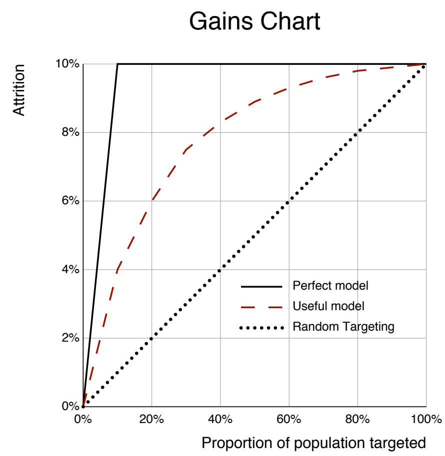
so the savability, S, is given by S = –U. In principle, uplift can be predicted simply by building two models, one on the treated population and one on the control population, and subtracting their predictions; unfortunately, this often works very badly in practice, so novel algorithmic approaches are required. These are beyond the scope of this paper, and some of the techniques are not public, but such methods are discussed in various papers.[3-8] We call models that predict this difference, U, *uplift models*.[2] They are also known variously as incremental models,[5] incremental impact models, net response models, lift models, true lift models,[6] true response models, differential response models[3] and proportional hazards models.[7]

## 2.3. Gains Charts

In order to understand uplift models we first need a means of measuring their effectiveness. We motivate our preferred method with reference to a familiar device—the gains chart.

Gains charts (and gains tables) are among the most common ways of understanding and quantifying the effectiveness of a marketing campaign or a predictive model. Suppose that we have an overall attrition rate of 10%. The gains chart then shows the cumulative losses as we run through our customer base, sorted from the highest to the lowest attrition scores (i.e. from those the model says are most at risk of leaving to those least at risk).

This is illustrated in the graph to the right for three different models.




The solid line shows the perfect model, which gives higher scores to all the 10% of customers who leave than to any of those who stay. Of course, such a perfect model can never be built in real-world situations.

The dashed line shows a more typical model, characterised by a curve bowed above the diagonal. The customers predicted to leave at a higher rate do indeed leave more often (as is indicated by the gradual decrease in slope from left to right) so that this model allows identification, for example, of 20% of customers who account for some 60% of the attrition.

The dotted line shows a completely useless model—one that has no ability to discriminate between those who leave and those who stay.

[3] Radcliffe N. J. & Surry, P. D. (1999). “Differential response analysis: Modeling true response by isolating the effect of a single action.” Proceedings of Credit Scoring and Credit Control VI. Credit Research Centre, University of Edinburgh Management School.

[4] Maxwell Chickering D. & Heckerman, D. (2000). “A decision-theoretic approach to targeted advertising.” Sixteenth Annual Conference on Uncertainty in Artificial Intelligence, (Stanford, CA).

[5] Hansotia B. & Rukstales, B. (2001). “Incremental value modeling.” DMA Research Council Journal, 1–11.

[6] Lo, V. S. Y.. (2002). “The true lift model”. ACM SIGKDD Explorations Newsletter. Vol. 4 No. 2, 78–86. 1

[7] Manahan, C. (2005) “A proportional hazards approach to campaign list selection”. SAS User Group International (SUGI) 30 Proceedings.

[8] Lo, V. S. Y. (2005). “Marketing Data Mining – New Opportunities”. Encyclopedia of Data Warehousing and Mining (ed. J. Wang). Idea Reference Group.

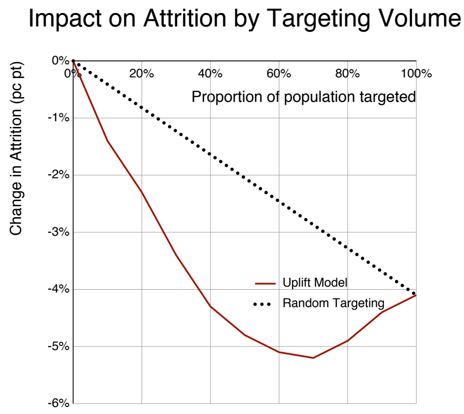
  

Thus on a gains chart, the more bowed is the curve above the diagonal, the more powerful is the model. In fact, there is a performance measure, known as the Gini coefficient,[9] that quantifies this as the ratio of the area above the diagonal for an actual model (such as the dashed line) to that for the optimal model (the solid line).

[9] Hand, D. (1997). Construction and Assessment of Classification Rules. John Wiley & Sons (Chichester).

## 2.4. Qini Graphs: Gains Charts for Uplift

Qini graphs and Qini coefficients are generalisations of the Gains Chart and the Gini coefficient to the case in which it is *uplift* that the model is supposed to predict.[2] The Qini graph to the right (upper graph) results from a model built on data from the insurer introduced in section 2.1.




The Qini graph is like a gains chart except that now the score used to sort the population along the horizontal axis is interpreted as an uplift score and the vertical axis shows cumulative uplift (usually in percentage points). This Qini graph ends at –4% because the overall impact of targeting everyone is to *reduce* attrition by 4 percentage points (from 32% to 28%). The diagonal shows the effect of random targeting, while the solid line shows the impact of targeting using an uplift model.

The Qini graph shows that by targeting the first 70% of the population by uplift, overall attrition can be reduced by around five percentage points instead of four. (This is *not* 5% of the 70%: it is 5% of the *total* population, i.e. if the right 70% are treated, a net 5% of the total population will be retained who would otherwise leave.) This shows that we can get a double win, increasing retention by one quarter while reducing the contact volume by around 30%.


The second graph shows campaign profitability at different targeting volumes. The original campaign was profitable, but nearly four times as much money can be made by targeting anywhere from 40% to 70% of the population. There are pros and cons of different cut-offs. The most obvious advantage of 40% is that it minimises marketing spend while returning the nearly the same absolute profit. This maximises the return on investment (i.e. the profit on each unit of marketing spend). The advantage of going up towards 70% is that there are more actual customer saves, as is shown by the Qini graph with no significant reduction in the profit generated by the campaign.

*Qini Graphs are generalisations of Gains Charts to the case of Uplift Models*

“*We actually save more customers by targeting the right 70%, than we do if we treat everyone*”

# 3. Predicting Savability with Uplift Models

## 3.1. Traditional Attrition Modelling

As noted above, there are two kinds of attrition model, the difference being whether or not the modelling population has been treated. The modelling procedure is the same for the two cases.

Attrition models are normally built by taking a historical sample of the relevant population of customers and using a fitting method (often logistic regression) to relate predictors to attrition outcome. Here, attrition outcome is a 0/1 indicator that might be measured by observing customers for some period. We might find that good predictors of attrition include change in renewal premium, whether or not the customer had a claim and a geodemographic code.[10] The resulting model might take the form of a scorecard that assigns a number of points for each range of these variables as in the table to the right.

[10] Geodemographic codes are mappings from a postal or zip code to a demographic banding, usually based on census data.

**A Possible Scorecard for Attrition Risk**

| | | | |
|---|---|---|---|
| **Premium increase** | ≤ €0 | €0.01 – €30 | >€30 |
| | -10 | +20 | +50 |
| **Had claim?** | no | yes | |
| | +10 | -50 | |
| **Geodem code** | A | B | C |
| | -10 | +40 | +10 |

A person with a premium increase of €10, who had not had a claim in the last year and who lives in an area with geodemographic code B would get a score of 20 + 10 + 40 = 70.

In some cases, it is then necessary to translate the score into an attrition probability by applying a simple transformation. In the case of a logistic regression, this would normally be something like

$$p = \frac{1}{1 + e^{-ks}}$$

where $s$ is the score and $k$ is a scaling constant. This will map the score into a probability between zero and one.

In constructing the modelling sample, it is of critical importance that an appropriate *observation window* be used. By this we mean that the predictors must measure the state of the customer at some time *before* the modelled outcome, in this case attrition. For example, if we use an outcome period of the second half of 2007 (the period during which we record which customers leave), it might be appropriate to measure predictors such as *whether the customer had a claim* during the first half of 2007. It is the very essence of all predictive modelling that we build the model by fitting outcomes in the recent past as function of data from the more distant past. We then make predictions by measuring the corresponding variables in the recent past, allowing us to predict outcomes in the future by applying our fitted function—in this case, the scorecard. In doing so, we make the fundamental assumption that the structural relationship between predictors and outcomes is relatively stable over time, i.e. that the past is a good guide to the future.

“*The fundamental assumption is that the past is a good guide to the future.*”

## 3.2. Modelling Savability with Uplift Models

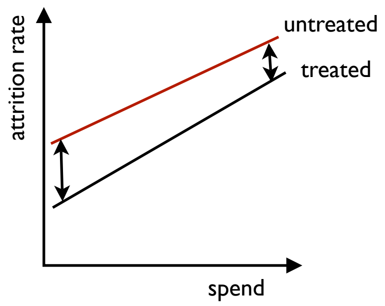
A customer's savability is defined simply as the reduction in attrition probability that results from treatment:

$$S = -U = p_A^U - p_A^T.$$

As noted in section 2.2, the complexity in modelling savability arises from the fact that we cannot simultaneously treat and not treat an individual customer. Conventional modelling is based on knowing the outcome to be modelled for each customer in some historical period (in this case, whether or not the customer left) and then finding correlations between those outcomes and the values of the predictor variables using a fitting or learning procedure. If our goal is to fit the difference in probability that results from our retention activity, conventional regression and similar approaches are of little help, because they all depend on having the known outcome for a historical population to learn from (or regress against).

Faced with this situation, the obvious approach is to build two models, one on the untreated population and the other on the treated population, and then subtract one from the other. In principle, this certainly provides a valid and unbiased estimate of the savability. Unfortunately, however, in many cases it does not work particularly well. We speculate and believe (though it is hard to prove) that there are two main reasons for this.

The first is that in practice the magnitude of the uplift is often small in comparison to the attrition rate. For example, for the insurer discussed above, the magnitude of the uplift, at 4 percentage points, is only around a seventh of the overall attrition rates (28% and 32% for treated and untreated populations respectively). This creates a significant problem with "signal-to-noise ratio", namely that in the main variations in outcome (as fitted by the two separate models) will tend to be much larger than the variations in uplift that we actually wish to estimate.




In this example, attrition increases with spend, but the effectiveness of the intervention decreases (as can be seen from the convergence of the lines). It may be unrealistic to expect a difference of two models to predict this variation in savability reliably.

The second problem is more profound: it is that the goal of the fitting procedure for the two component models (treated and untreated) is not necessarily strongly related to the fitting goal for uplift. For while it is the case that the difference between perfect treated and untreated models would, by definition, estimate uplift perfectly, there is no general reason to suppose that the main drivers of variation in uplift and those of variation in attrition will be the same. Indeed, it is not obvious that the factors controlling how likely someone is to leave should bear any particular relation to those governing how that person will respond to a given retention campaign.

So while subtracting two models is certainly a method worth trying, especially when the uplift is large, in general it is both theoretically and empirically better to use a dedicated uplift modelling technique than to model the two populations separately.

In order to apply uplift modelling successfully, just as with conventional modelling, it will be necessary to have a suitable historical sample to allow construction of a valid observation window. The added complication is that now there must be two historical populations, one of which has been subject to the retention activity in question and the other of which has not. In general, the two populations must be statistically equivalent in all respects except the treatment decision, and this is normally achieved by randomly withholding treatment from a random proportion of a target population. In some cases, it is possible to compensate if there is some bias in the allocation of customers to the treated and control groups, but this always significantly complicates the analysis.

## 4. Conclusion

As an increasing number of financial institutions have discovered, retention activity can have negative as well as positive effects. It follows from this that it is dangerous to target retention activity primarily on the basis of estimated attrition risk, whether or not this is weighted by customer value. We have shown an example of how a successful and profitable retention campaign can be radically improved by using uplift modelling to predict savability, leading to a 25% increase in the number of customers saved and a 300% increase in campaign profitability. We have also seen examples of retention campaigns that have had an overall negative impact, but which were positive for some segments, and in some cases have been able to extract a highly effective and profitable sub-campaign from within such value-destroying campaigns.

> “*Both successful and unsuccessful retention campaigns can be dramatically improved by uplift models.*”

The adoption of uplift modelling requires financial institutions to embrace fully the use of systematically randomised control groups, and commits them to using sophisticated, modern uplift modelling methods. These are more complex than traditional methods, but we believe that the demonstrable improvement in results that can be achieved more than justifies such such a transition. Once again, *savability is the key to retention activity*.

# Author

**Nicholas J. Radcliffe**

Nicholas Radcliffe is founder and director of Stochastic Solutions Limited, a consultancy that focuses on customer targeting and optimization problems. He was previously a founder and Technical Director of Quadstone Limited, where he led the development of their approach to uplift modelling and numerous client engagements.
Nicholas is also a Visiting Professor in the Department of Mathematics and Statistics at University of Edinburgh, where he works in the Operations Research group.

Nicholas.Radcliffe@StochasticSolutions.com