Using Control Groups to Target on Predicted Lift:

Building and Assessing Uplift Models

Nicholas J. Radcliffe

Portrait Software
The Smith Centre
The Fairmile
Henley-on-Thames
Oxfordshire
RG9 6AB
UK

Department of Mathematics & Statistics
University of Edinburgh
The Kings Buildings
Mayfield Road
Edinburgh
EH9 3JZ
UK

Abstract

Various authors have independently proposed modelling the difference between the behaviour of a treated and a control population and using this as the basis for targeting direct marketing activity. We call such models Uplift Models. This paper reviews the motivation for such an approach and compares the various methodologies put forward. We present results from using uplift modelling in three real-world examples. We also introduce quality measures appropriate to assessing the performance of uplift models, for both binary outcomes (purchase, attrition, click, default) and continuous outcomes (spend, response size or value lost). Finally, we discuss some of the challenges faced when building uplift models and suggest some key challenges for future research.

1 Introduction

It is standard practice to employ control groups to allow post-campaign assessment of the incrementality (or *lift*) of marketing actions. There is also limited but growing recognition that campaigns should ideally be *targeted* so as to maximize predicted lift. This approach has been suggested, apparently independently, in at least five published papers.[1, 2, 3, 4, 5] Various modelling methods have been suggested in the papers, including paired regression models and decision trees with modified build algorithms. All of these approaches recognize that traditional "response" models actually predict either a conditional probability, such as

$$P(\text{purchase} \mid \text{treatment}), \tag{1}$$

(where P(A|B) denotes the probability of A given B) or sometimes, as in the case of attrition modelling, one such as

$$P(\text{attrition} \mid \text{no treatment}). \tag{2}$$

Clearly, the traditional approach does not model true *response* (i.e., the *change* in behaviour resulting from the action). As a result, all the papers referenced suggest methods for predicting *uplift*, which we define, for demand-generation applications, as

$$P(\text{purchase} \mid \text{treatment}) - P(\text{purchase} \mid \text{no treatment}). \tag{3}$$

This would appear to be the ideal basis for targeting in many circumstances.

This paper reviews all of the known published approaches and present results from a range of real-world problems. We also discuss the assessment of uplift models and introduce a family of quality measures (the Qini measures, Q, q_0 and Q_c) based on generalizations of the familiar Gini coefficient.

Real-world applications discussed include retention modelling from telecommunications operators (modelling "savability"), cross-selling in banking (modelling incremental account opening) and deep-selling in retail (modelling incremental revenue).

2 Motivation

A fundamental problem in motivating uplift modelling is that the traditional term used for the conventional approach is "response modelling", which sounds like exactly the same thing as uplift modelling. The very name "response" is strongly suggestive of the idea that a particular (marketing) action *caused* the "response". However, the typical practice in "response modelling" consists of

1. Choosing a target population;
2. Possibly holding back a randomized control group (to allow post-campaign assessment of "incrementality", or uplift);
3. Treating the target group (minus any control group);
4. Recording those people who take some desired action during an outcome period as "responders" (possibly with reference to a campaign code);
5. Building a "response model" on those customers subjected to the action to understand the variation in outcome ("response");
6. Possibly assessing the "incrementality" or uplift by comparing the overall level of the desired outcome ("response") in the treated and control groups.

The resulting model estimates a conditional probability, such as P (purchase | treatment) (equation 1), rather a change in probability resulting from an action, or (true) *response*,

$$\text{P (purchase | treatment)} - \text{P (purchase | no treatment)}. \qquad (3 \text{ bis})$$

This difference clearly lies behind references to "true response" and "true lift" in the titles of two of the publications.[1, 4]

Of course, response models are not necessarily used directly and naïvely: one particularly common approach is to weight a modelled "response" probability by some kind of value (such as purchase size, or a customer value). While sensible, this does not alter the underlying weakness of the traditional approach. Fundamentally, if we wish every unit of marketing spend to achieve the largest possible change in customer behaviour (however measured), then we need to model exactly that—the *change* in behaviour that results from our actions; a traditional "response" model (equation 1) does *not* do this.

3 Review of Approaches

There are two broad classes of approaches to building uplift models in the literature—broadly, those based on trees and those based on additive regression models.

3.1 Tree-based methods

The first two papers published on uplift modelling[1, 2] both came out in favour of tree-based approaches. There are three main classes of tree-based methods in common use for simple prediction—Classification and Regression Trees ("CART"),[6] Quinlan's C4.5 and predecessors[7, 8] and the AID[9]/CHAID[10] family. The first two are based on greedy, divisive methods that start with the whole population, consider a family of binary splits, assess the utility of each split using a quality measure (an "impurity" measure, such as variance, Gini or information gain), and then recurse. Various cross-validation methods are then commonly used to "right-size" the tree through pruning. AID and CHAID are slightly different, and result in non-binary trees, but again are fundamentally controlled by a single measure of split quality, in this case χ^2. Both Radcliffe & Surry[1] and Maxwell Chickering & Heckerman[2] proposed changes to the split criterion for the tree to focus on the difference in outcome between the treated and control populations. In the former case, the population size for the two resulting segments was also taken into account, while in the latter case this difference in outcome was used directly as the split criterion.

3.2 Regression-Based Methods

The remaining three papers focused primarily on regression-based approaches to modelling uplift.

The methods proposed by Hansotia & Rukstales[3] and Manahan[5] (and also discussed by all the other authors[1, 2, 4]) are equivalent, and consist of simply building two independent regression models, one for the treated population and one for the control population, and subtracting these. This obviously has a number of attractions, including the vast body of literature on and experience with regression, and its proven pedigree. Several of the authors, however, express a concern that because the two models are independent, there is no direct attempt to *fit* the difference in behaviour between the two populations: while it is clear that if the models were perfect, their difference would accurately predict uplift, it is much less clear what properties we should expect of the difference when the models are imperfect.

Lo[4] suggests a subtle variation on this regression theme.* Rather than simply fitting two independent regression models, he proposes building a single regression with an

* Lo's method is really a "meta-method" that can be applied to any modelling approach, but regression is the principal case he considers, together with briefer discussions of layered feed-forward neural networks ("multi-layer perceptions") and naïve Bayes models.

extended set of independent variables. If the original n independent variables are denoted $\boldsymbol{x} = (x_i)$ (for $i = 1, 2, \ldots, n$), Lo proposes regressing over $\{\boldsymbol{x}, t\boldsymbol{x}, t\}$ where for each customer

$$t = \begin{cases} 1, & \text{if customer is treated,} \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

This produces a model which separates out regression coefficients for the "main effect" (the coefficients of the x_i, the basic treatment effect (the coefficient for t) and for uplift effects (the coefficients for the interaction variables tx_i). The result is a model with the functional form $f(\boldsymbol{x}, t\boldsymbol{x}, t)$ which is scored by computing $f(\boldsymbol{x}, 1\boldsymbol{x}, 1) - f(\boldsymbol{x}, 0\boldsymbol{x}, 0)$ for each customer.

Lo argues against the more straightforward approach of simply subtracting two independent regression models, as proposed by other authors,[3,5] on the basis that "estimated lift can be sensitive to statistically insignificant differences in parameters of the treatment and control models".[11]

4 Measuring Performance

Before looking at the results of building uplift models, it is important to consider the question of quality measures for such models, a subject which does not appear to have been discussed in the literature.

In any specific targeting situation, there may be a clear goal such as maximizing the total profitability of a campaign, perhaps on the basis of a net present value model, or maximizing some non-monetary outcome such as the reduction in attrition achieved. However, just as quality measures such as Gini, R^2, the Kolmogorov-Smirnov statistic and occasionally even classification errors are useful for understanding the overall power of conventional models, it is desirable to have access to overall statistics that summarize the potency of an uplift model.

Many of the performance measures for traditional models depend fundamentally on a comparison of actual and predicted outcomes at the level of an individual. These are intrinsically unsuitable for generalization to the case of uplift because we can never have a definitive measure of uplift for an individual, since no one can simultaneously be treated and not treated. Uplift can therefore only be estimated on a per-segment basis, and the estimated uplift for an individual is generally different when estimated with respect to different segmentations. Partly for these reasons, we take the Gini coefficient as our starting point, and propose a family of measures under the umbrella name of "Qini" coefficients.

4.1 The Gini Coefficient for Conventional Models

There are many ways of defining Gini, but the most convenient starting point for this work is to define it with reference to the familiar gains chart, which is among the most common ways of assessing traditional models. Consider a demand generation

application in which the desired outcome is a fixed "purchase". The gains chart is constructed by first sorting the population, from "best" to "worst", by the score in question. The graph then shows the number of responses achieved (vertical axis) as a function of the number of people treated (horizontal axis). A perfect model assigns higher scores to all of the purchasers than any of the non-purchasers. Thus the perfect model first climbs at 45°, reflecting the fact that all purchases are assumed to be caused by treatment. After the last purchaser has been accounted for, the graph proceeds horizontally, as shown below. In contrast, random targeting results in a diagonal line from (0, 0) to (N, n) where N is the population size and n is the number of purchases achieved if everyone is targeted. Real models usually fall somewhere between these two, forming a broadly convex curve above the diagonal as shown, while a "reversed" model that tends to assign better scores to non-purchasers than purchases will fall below the diagonal (Figure 1).

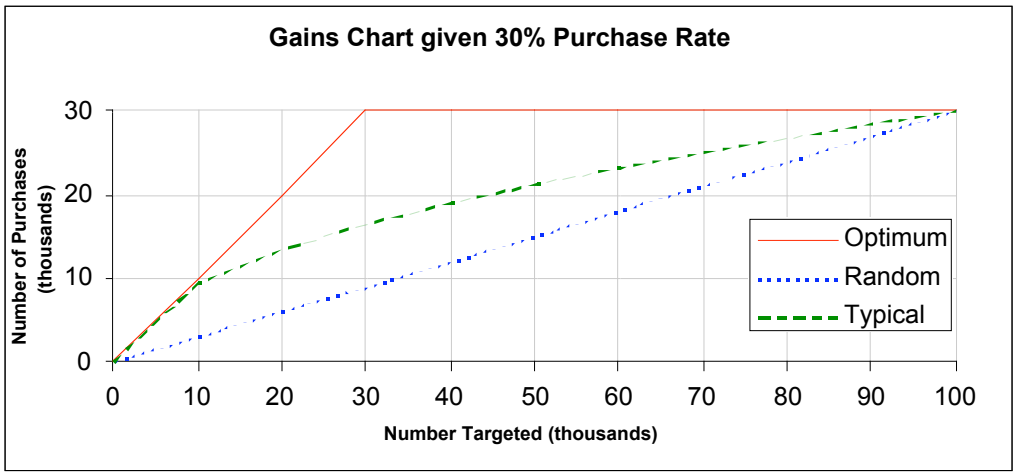


Figure 1: A traditional gains chart for a campaign to 100,000 people with an overall purchase rate of 30%. The vertical axis shows the number of purchases as a function of the number of people targeted.

The Gini coefficient is simply the ratio of the area above the diagonal of the actual curve to the corresponding area above the diagonal of the optimum curve. A perfect model therefore has a Gini of 1 (or 100%), a model of little or no predictive power will have a Gini close to zero, and an inverted model will have a negative Gini, which can be as low as −1 (or −100%). (The Gini coefficient is more commonly defined with reference to the "receiver operating characteristic" (ROC) curve, also known as the Gini curve, which plots the number of non-responders, rather than the number of people treated, on the horizontal axis. This can be shown to be equivalent, but is a less natural starting point for the case of uplift models.)

4.2 The Qini Coefficients Q and q_0 for Uplift Models

We now introduce the Qini coefficients, which are natural generalizations of the Gini coefficient to the case of uplift. Again, we start by drawing a graph, which we call either the "Gains Chart for Uplift" or, more simply, a Qini Curve. This is the same as a Gains Chart except that the vertical axis now shows the cumulative number of *incremental* sales achieved, or the uplift. This is estimated on a per-segment basis by

comparing the purchase rate in the treated group and the corresponding control group. The estimated number of incremental sales in a segment is given by

$$u = R_t - \frac{R_c N_t}{N_c} \tag{6}$$

where R_t and R_c are the number of purchases in the treated and control groups respectively (within the segment in question), and N_t and N_c are the corresponding total sizes of the treated and control groups within the segment.†

Were it not for the possibility of negative effects, we would then proceed as before. In reality, the potential for negative effects introduces significant complications. Suppose, for simplicity, that as well as a treated group of 100,000 with an overall purchase rate of 30% there is a control group of 100,000 with a purchase rate of 10%. The overall uplift is clearly twenty percentage points. It is certainly possible that the effect of the treatment was simply to persuade 20,000 people to buy who would not have done so otherwise. But is it also possible that up to 10,000 people who would have bought without the campaign (as estimated by the control group) were caused *not* to purchase by the campaign. These, of course, would have to be balanced by 10,000 more people who were persuaded to purchase who would not otherwise have done so. In this case, we would end up with a Qini curve as shown in Figure 2.

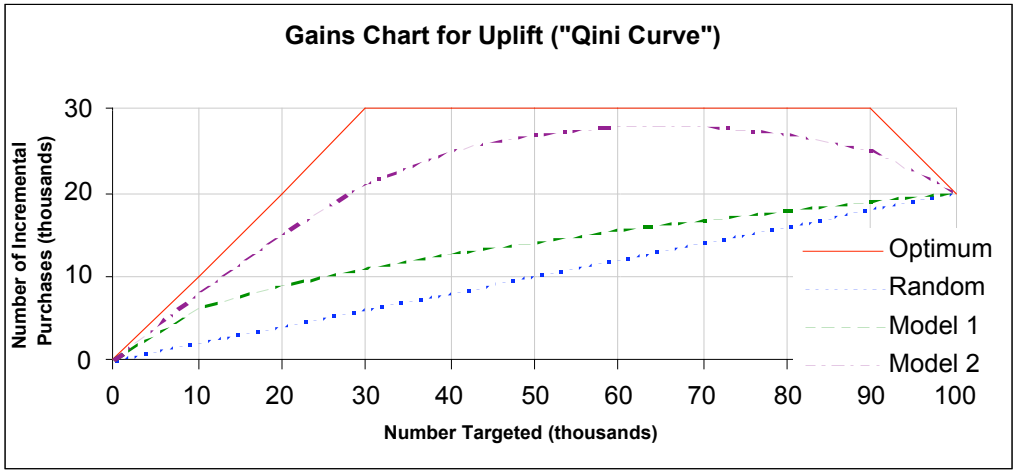


Figure 2: A gains chart for uplift ("Qini Curve"). The vertical axis now shows the number of incremental purchases as a function of the number targeted. Because of the possibility of negative effects, it is possible that targeting a smaller number of people may actually result in a larger number of purchases, as shown here. Model 1 ($Q \approx 21\%$, $q_0 \approx 32\%$) simply shows better-than-random targeting, while Model 2 ($Q \approx 76\%$, $q_0 \approx 114\%$) illustrates the possibility of achieving more incremental sales (and therefore more total sales) by targeting a smaller volume—in this case about 60% achieves the maximum uplift.

The extent to which it is theoretically possible to exceed the actual uplift observed is usually limited by the purchase rate in the control group. When there is a relatively

† There are some calculational subtleties with constructing the best Qini curve, which result largely from the fact that uplift estimates are not strictly additive. One consequence is that it is usually more accurate to estimate the cumulative uplift at each point from zero to that point directly, rather than accumulating a set of uplifts

low purchase rate in the control group, as here, the limit is that all "excess" positive effects must be balanced by negative effects, and the worst a treatment can do is to prevent the purchases that would otherwise have occurred, as quantified by the purchase rate in the control group: therefore, the uplift can never exceed the overall uplift by more than the purchase rate in the control group.

In rarer cases, the limiting factor is simply that the purchase rate can never exceed 100%. Thus, for example, if the treated and control purchase rates are 95% and 75% respectively, clearly the highest possible uplift is 100 – 75 = 25 percentage points.

Given these observations, we can now define the Qini value Q, for binary outcomes, in the same way as the Gini coefficient, i.e. as the ratio of the actual uplift gains curve above the diagonal to that of the optimum Qini Curve, shown as the solid red line in Figure 2. As with the Gini coefficient, this theoretically lies in the range [-1, 1], though because of the uplift can only ever be approximated actual calculations may occasionally lie slightly outside this range. Note also that whereas in the conventional setting it is clearly possible to order the customers in such a way as to achieve the optimal Gini (simply by sorting all the responders ahead of the non-responders), it is rarely clear whether such an "optimal" ordering actually exists for the case of uplift.

When the overall uplift is non-zero, it is also sometimes convenient to define the "little" q_0 value as the ratio of the area of the Qini curve above the diagonal to the area above the diagonal of the "zero downlift" optimum Qini, which is the maximum Qini curve that can be achieved without invoking negative effects. This measure in some ways behaves more like a conventional Gini coefficient, except that it is possible (and not uncommon) for it to exceed 100%.

The case of zero or near-zero overall uplift is also interesting because it emphasizes that, where treatment has negative effects on some portions of the population, even a campaign with no overall lift may contain segments in which the treatment is effective. This leads to optimum Qini curves such as that shown in Figure 3. As the overall uplift tends to zero, q_0 values tend to $\pm\infty$, so "big" Q values are much more useful in these cases.

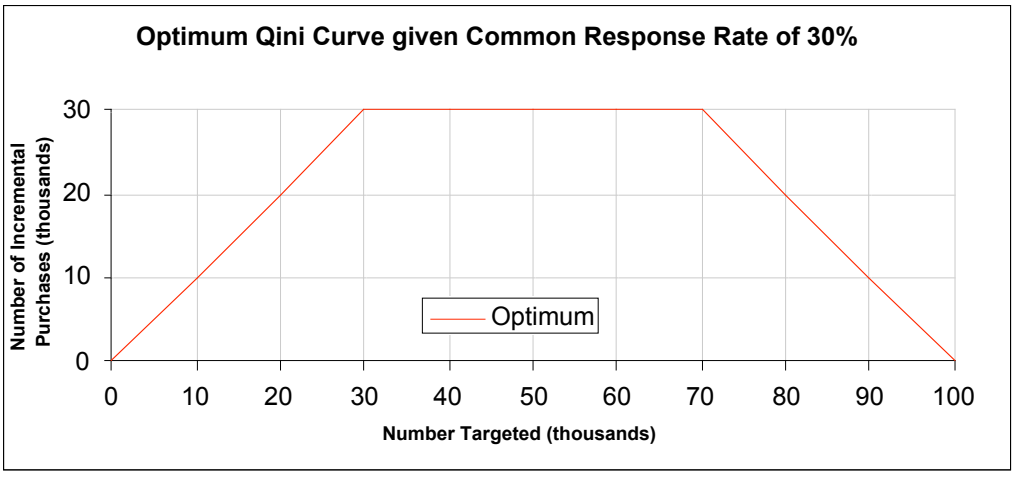


Figure 3: Optimum Qini Curve given Treated and Control Response Rates of 30%. It is possible for a treatment to have no net effect on a population because

positive and negative effects in different parts of the population cancel out. This curve shows the theoretical optimum Qini for a situation with a common "response" rate of 30%.

4.3 Continuous Outcomes

For the non-binary case, we again start by reviewing the construction of a Gains Chart and the computation of Gini from this, and then generalize. Again, without loss of generality, we will focus on the case of a purchase, but now we are interested in the total *value* of incremental purchases, rather than their number. The gains chart for a continuous outcome is the same as that for a binary outcome except that the vertical axis shows cumulative *value* rather than the cumulative count of purchases as function of the proportion of the population targeted. Since it is possible that the entire purchase value comes from a single customer, rather than rising at 45 degrees, when scaled "naturally", the optimal gains curve for a continuous outcome rises essentially vertically. Again, the Gini is simply calculated as the ratio of the areas above the diagonal of an actual model and the optimum model, and again lies in the range [-1, 1].

The problem we are then faced with in generalizing this case to handle incremental purchases is that there is no bounding ("optimal") Qini curve for the continuous case. This is because the availability of negative effects means that we could, in principle, have arbitrarily large positive and negative effects that cancel.

This lack of any well-defined "optimum" Qini curve does not affect q_0, which can be defined as before, as the ratio of areas above the diagonal of the actual uplift gains curve for a model and the "zero-downlift" Qini curve, which is now simply upper-triangular. Unfortunately, however, as before this is not satisfactory as the only "Qini" coefficient, because it is not well defined if the overall uplift is zero, and is extremely unstable if the overall uplift is small.

We therefore need to find a value similar to Q for the continuous case. We do this simply by dividing the area above the diagonal of our uplift gains curve (Qini curve) by half the square of the total number of customers, and we call the resulting value Q_c. Rather than a dimensionless quantity, this value is in the units of whatever the outcome is measured in, commonly money, which can be though of as *per head*. So if we have an overall uplift in revenue of $1,000,000 when we target 1,000,000 people, the zero-downlift Qini has a Q_c value of $1.00, or $1.00/head. There is clearly a degree of arbitrariness in the scaling of this quantity but the advantage in dividing by half the square of the population size is that the measure becomes independent of population size and is scaled in a way that is consistent with Q and q_0.

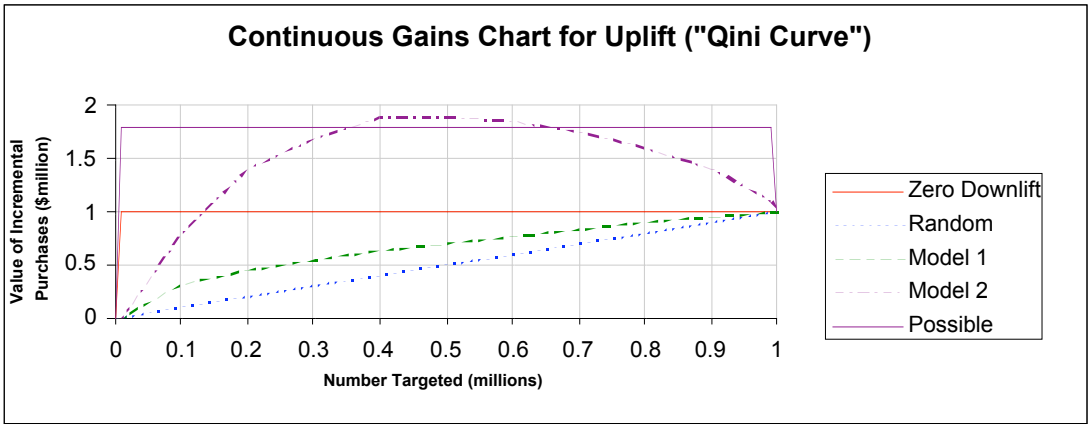


Figure 4: This illustrates several features of the Qini Curve for a continuous outcome. First, unlike in the continuous case, the zero-downlift Qini (red, mostly horizontal at y=1, Q_c=\$1.00, q_0=100%), can rise essentially vertically, because any customer can spend any amount. Secondly, there is no "optimum" Qini curve, because of the possibility of arbitrary sized cancelling positive and negative effects. So "Possible" (purple, mostly horizontal at y=1.8, Q_c≈\$2.60, q_0≈260%) is a possible curve, but equally there could be a higher one. Models 1 and 2 illustrate the kinds of results models can produce, in the first case (Q_c≈\$0.32, q_0≈32%) showing better-than random targeting, and in the second (Q_c≈\$1.96, q_0≈196%) showing a situation in which incremental (and thus total) revenue is maximized by targeting around 40% of the population.

5 Results

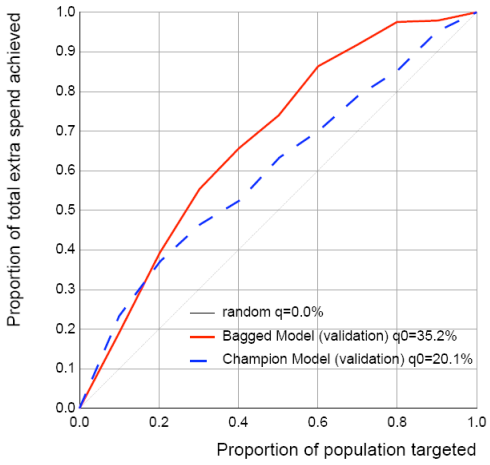
We present three examples of the application of uplift modelling to real-world, commercial problems. All of these problems were tackled using the various implementations of Radcliffe & Surry's approach[1] (as it developed), using commercial software developed by Quadstone and now marketed by Portrait Software.

5.1 Example 1: Deep-Selling

A retailer used a catalogue mailing to drive greater spend activity among active customers—an example of deep-selling.‡ The customers were selected on the basis of a conventional "response" model—the so-called "champion" model—built on data from customers mailed in a similar previous campaign. Approximately 100,000 of those ranked as likely high "responders" by the model were targeted, and approximately 50,000 were held back as a control group. On average, spend increased by $8 per head among those mailed. An uplift model was then built, using data from the same campaign as that used to construct the champion model. In contrast to the champion model, however, the uplift model used information about the control group as well as the mailed group.

Graph 1 shows the Qini curve for "uplift in spend".§ The blue line (dashed) shows the result of targeting with the "champion" model and the red line (solid) shows the effect of targeting with the uplift model.** Up to about 20%, the two models perform similarly. Thereafter, the models diverge. For example, if 50% are targeted, the uplift model manages to identify customers delivering about 16% more revenue than the champion model—approximately $11.84 against $10.24 per head. And if 80% are targeted, the uplift model manages to retain 97% of the incremental spend ($9.70 per head) against only 85% ($8.50 per head) for the champion model

Graph 1: Demand Generation



Clearly, the uplift model is significantly better at identifying customers for whom marketing spend generates a positive return.

‡ While *cross-selling* tries to sell new products to customers, and *up-selling* aims to drive customers to upgrade, *deep-selling* simply tries to increase the frequency or size of their transactions.

§ Because the results shown here are from real companies, some results have been systematically rescaled to protect client anonymity and confidentiality. In all cases, where this has been done, the scaling has been chosen to reduce the overall impact of claims made, never to exaggerate them.

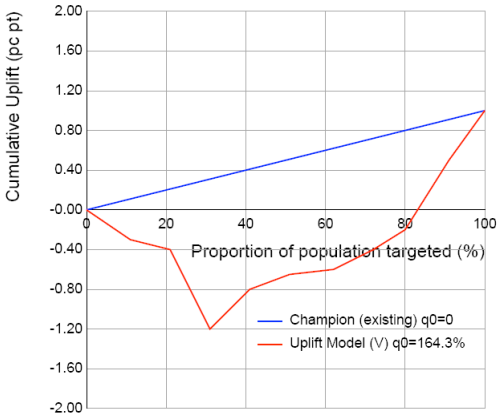
** both results are shown for validation (holdout) data.

5.2 Example 2: Retention

A mobile phone company was experiencing an annual churn (attrition) rate of approximately 9% in a segment. It targeted the entire segment with a retention offer, holding back only a control group. The net result was an *increase* in churn to 10% among the treated group, while the churn rate in the untreated group remained at 9%. Obviously, this is the exact opposite of the desired effect, but we have witnessed this phenomenon repeatedly. It seems that retention interventions often backfire because they variously remind customers of their ability to terminate, provide a catalyst to help overcome inertia and annoy customers through intrusiveness.

Clearly, one way to improve the situation is to stop this retention offer entirely. However, there was a strong belief within the business that the offer was valid and did work for some groups of customers. Also, no more successful approach had been identified. An uplift model was therefore built to try to identify a sub-segment within which the treatment was effective.

Graph 2: Churn Reduction



Graph 2 is a Qini curve, but the goal is now to achieve the greatest possible reduction, i.e. a negative increase in churn. As usual, the horizontal axis shows the proportion of the population targeted, while the vertical axis shows the resulting increase in total churn. The mailing was untargeted, so the diagonal line (blue) shows the effect of random targeting of customers. Of course, with such random targeting, extra customers are lost in proportion to the number treated. However, the red line (concave) shows the effect of targeting on the basis of the uplift model. The results are striking.

The model shows that retention activity was effective for about 30% of the customers: if only the 30% identified as "most savable" by the model are treated, churn across the *entire* segment falls by 1.2 percentage points, from 9% to 7.8%, i.e. over 13% fewer customers churn. Compared to targeting everyone, churn actually reduces from 10% to 7.8%—a proportionate fall of 22%. The segment contained approximately 1 million customers, so using industry-standard ARPU of $400/year, the financial impact of moving to uplift modelling, for one segment alone, is an improvement of around $8.8 million.

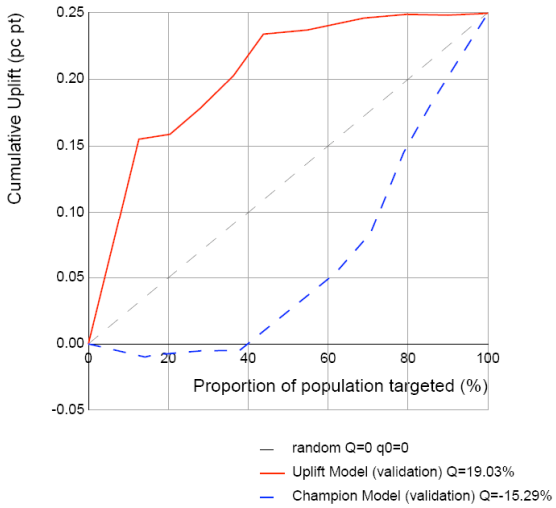
5.3 Example 3: Cross-selling High-value Products

We have also tackled a number of cross-sell targeting problems for banks. A typical scenario involves cross-selling activity aimed at increasing product holding. The value of many banking products is high, so that even an increase in product take-up as low as a tenth of a percentage point can provide a positive return on investment for mailings. However, we have shown that with appropriate targeting we can usually achieve between 80% and 110% of the same incremental sales while reducing mailing volumes by factors ranging from 30% to 80%. Because the banks in question have themselves attempted to model uplift, they typically have historical data allowing full longitudinal validation of results.

One of the complicating factors in these scenarios is that the uplift from the campaign is usually significantly smaller than the natural purchase rate of the product being promoted—typically by a factor of five to twenty. Thus it would not be unusual to see a purchase rate in the control group around 1%, and in the treated group of 1.1%. However, drivers of the base purchase rate are often quite different from those of the incremental purchases resulting from the campaign. Because of this, non-uplift approaches to targeting are often doomed to failure, and sometimes actually perform less well than random targeting.

Graph 4 shows an example of one such campaign. Here, the net effect of the campaign was to increase the uptake of the product by a quarter of a percentage point. However, the uplift model shows that over 60% of the increase in sales comes from just 10% of the targeted population, 90% comes from 40% of the population and 99% comes from 70% of the population. Notice also that the "champion" model produced substantially worse results than random targeting—in fact, in this case, reversing the ranking from it would have been very much more effective than using its actual output. This suggests that this campaign is being effective in stimulating demand from the very people who tend not to purchase without intervention.

Graph 4: Banking Cross-Sell



6 Observations from Practical Examples

In practice, most of the real difficulties with uplift modelling derive from noise. Noise arises for two principal reasons. First, when building uplift models we are attempting to fit the *difference* between the behaviour of two populations. At the simplest level, when we do this, errors add. Secondly, while from a modelling perspective, we would ideally choose to have equal (and large) numbers in the treated and control populations, in practice, one is almost always much smaller than the other. This is because most targeting happens either in a trial setting or a roll-out setting. When a treatment is first being evaluated, the treatment volume is typically low to contain risk. Conversely, in roll-out situations, once a treatment is proven, the goal is usually to maximize its impact, and therefore the size of the control group tends to be limited. Unfortunately, it is the smaller of the two population that usually most strongly limits the performance of uplift models. The situation is also not helped by the fact that the uplift phenomenon being modelled is often small compared with the absolute outcome rates—for example, as quoted in section 5.3, we often see uplifts of a tenth of a percentage point on campaigns with an apparent "response" rate more like 1%.

We have therefore found it necessary to employ a wide variety of methods to control noise, including careful variable selection and binning methodologies, bagging,[12] stratified sampling and *k*-way cross-validation methods.

7 Conclusions and Further Research

Five researchers have independently proposed methods for modelling uplift to allow more appropriate targeting of marketing actions. In this paper, we have demonstrated three real-world examples in which such an approach has proved capable of delivering dramatic improvements in the profitability of marketing campaigns. We have also introduced a family of statistical measures appropriate to evaluating the performance of uplift models in ranking populations by uplift. These are the Qini measures Q, q_0 and Q_c. We have found these to be extremely useful in comparing and assessing uplift models.

Open research issues, as discussed by Lo,[11] include more complex treatment scenarios (where there are multiple treatments of treatment variables) and handling the challenges presented when the selection of the control group was not ("uniformly") random. Detailed comparative benchmarking of competing methods, while subject to all the usual difficulties of achieving fairness, would clearly be valuable. However, our experiences over several years strongly indicate that performance of uplift models on fabricated test data is often a particularly unreliable indicator of likely performance on real-world data. A significant challenge is therefore to find suitable data that can be made publicly available for benchmarking.

8 References

1. N. J. Radcliffe & P. D. Surry. "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 (1999)
2. D. Maxwell Chickering & D. Heckerman. "A decision-theoretic approach to targeted advertising." *Sixteenth Annual Conference on Uncertainty in Artificial Intelligence, Stanford, CA* (2000)
3. B. Hansotia & B. Rukstales. "Incremental value modeling." *DMA Research Council Journal*, 1–11. (2001)
4. V. S. Y. Lo. "The true lift model". *ACM SIGKDD Explorations Newsletter*. Vol. 4 No. 2, 78–86. 1 (2002)
5. C. Manahan. "A proportional hazards approach to campaign list selection". *SAS User Group International (SUGI) 30 Proceedings*. (2005)
6. L. Breiman, J. H. Freidman, R. A. Olshen & C. J. Stone. "Classification and Regression Trees". Wadsworth. (1984)
7. J. R. Quinlan. "Induction of decision trees." Machine Learning. Vol. 1, No 1, 81–106 (1986)
8. J. R. Quinlan. "C4.5: Programs for Machine Learning". Morgan Kaufmann, San Mateo, CA. (1993).
9. Kass, G. "An exploratory technique for investigating large quantities of categorical data". Applied Statistics, Vol. 29, 119-127 (1980)
10. J. Magidson. "The use of the new ordinal algorithm in CHAID to target profitable segments". The Journal of Database Marketing, Vol. 1, 29–48. (1993)
11. V. S. Y. Lo. "Marketing Data Mining – New Opportunities". *Encyclopedia of Data Warehousing and Mining* (ed. J. Wang). Idea Reference Group. (2005).
12. L. Breiman. "Bagging Predictors". Machine Learning. Vol. 24, No. 2, 123-140 (1996)