

Direct Marketing Profit Model

Bruce Lund, Marketing Associates, Detroit, Michigan and Wilmington, Delaware

ABSTRACT

A net lift model gives the expected incrementality (the incremental rate) of a prospect in a direct marketing or customer retention campaign. Typically, a net lift model would be used to score prospects in a campaign which included an offer of value to the prospect. This paper extends the net lift model concept by adding profit and cost components. This leads to a prediction of profit called the Profit Model. This model involves three component models. The first component model compares the buyers in treated group vs. the buyers in control to give the probability that a purchase was treated. The second component is a traditional in-market model which is built on the full control group. The third component model gives the probability that a treated buyer will use the offer versus buying without being aware of the offer. Costs and profits are externally provided. Then the three modeling components, costs, and profits are combined into a single formula in order to rank a prospect list in terms of expected profit per prospect. All data and models were processed using Base SAS® and SAS Stat®.

INTRODUCTION

The paper gives a methodology for ranking a direct marketing prospect file to receive a “private offer of value” (POV) for purchasing a target product so that expected profit is maximized. For example, the top decile in this ranking would give the highest expected profit among all possible selections of 10% of the prospects.

The offer is “private” because the offer is only available to the selected prospects via the direct marketing communication. The offer is “of value” because it is assumed in this application to be worth 2% to 4% of the purchase price of the targeted product and it is assumed that the price of the targeted product is, at a minimum, in the high hundreds of dollars. This product may be a consumer durable good such as an automobile, home appliance, or computer. But the targeted product could be a service such as internet, phone, and TV bundles.

This methodology will lead to what will be called in this paper the **Profit Model**.

Requirements for development of a Profit Model include:

- To have one (and ideally several more) historical large-scale direct marketing campaigns involving a POV with a large scale randomly selected control group.
- To be able to track purchases reliably for these historical treated and control groups. Furthermore, to have this tracking of purchases show the campaigns to have statistically significant incremental purchase rates as measured by subtracting the treated and control group buy-rates.
- To have at least 1,000 buyers from the treated group and a similar sized buyer group from control.
- To be able to reliably track the “redemption” of the POV from buyers in the treated group. Tracking redemptions is unimportant if all treated buyers use the POV. There are examples, however, where the usage rate is much lower.
- To have no “corruption” of the control group where a control group member can be given the POV.

A Profit Model can be developed for marketing campaigns where prospects are selected from large consumer databases (typically an in-house database) with abundant transactional data and demographics.

BACKGROUND: IN-MARKET MODEL, NET LIFT MODEL, AND PROFITS

An “in-market” or “response” model gives the likelihood that a prospect will buy a product within a specific period of time.

In contrast, a “net lift” model does not provide a probability value. Instead net lift gives the expected rate at which a prospect will be incremental due to the treatment. For example, a net lift value of 0.02 for each of a group of prospects means that 2% of such prospects will make an incremental purchase due to the treatment.

A net lift model gives rise to the expected profit per prospect once unit profit and unit cost are provided. Let “G” be unit profit and “C” be unit cost. Using a net lift model the estimated profit for prospect “j” is:

$$\text{Profit}_j = (\text{NetLift}_j * G) - C$$

e.g. If $\text{NetLift}_j = 0.1\%$, $G = \$3,000$, and $C = \$1$, then $\text{Profit}_j = \$2$

However, this is **over-simplified** because:

- As we discuss in a later section, C depends on “j”. Specifically, C depends on the likelihood by prospect “j” of buying and redeeming the offer.
- Naturally, G also depends on the prospect “j” provided there is more than one product that the prospect can buy (having different profits). A new model would be needed that predicts what specific product the prospect buys. This much greater level of complexity is not addressed in this paper.

DIRECT MARKETING COMMUNITY’S RECENT FOCUS ON NET LIFT MODELS

The Predictive Analytic World conferences have recently included speakers on this topic. Also SAS Institute offers a training course in Net Lift Modeling which was developed by Kim Larsen.

One of the earlier direct marketing papers on this topic was written by Victor S.Y. Lo in 2002. Lo combined the treated and control prospect lists into a single data set for logistic regression. Among the predictor variables was a “treatment indicator” which equaled “1” for treated prospects and “0” for control. Then interaction terms for the interaction of predictor variables and the treatment indicator variable were added to the original predictor list and entered into logistic regression.

A white paper by Radcliffe and Surry “Real-World Uplift Modelling with Significance-Based Uplift Trees” (2011) includes a literature review of net lift modeling.

DESCRIPTION OF THE PROFIT MODEL

There are three component models:

1. Buyer Model

The analysis data set consists only of the **buyers** from the **treated and control** samples. Let target = 1 if the prospect was treated and target = 0 if not treated.

Notation:

- P_T is the probability that the buyer was treated (i.e. target = 1).
- $P_C = 1 - P_T$ is the probability that the buyer was in control.

In fitting the model to predict P_T and P_C it is required that treated and control records are weighted to account for differences in the sampling fractions to select the treated and control groups.

For example, suppose the eligible population for the campaign was 300,000 and that the treated group was sampled at 200,000 and control at 100,000. Then a weight variable (call it “W”) with $W = 1$ for treated and $W = 2$ for control should be included in the campaign data set.

With weighting, the mean of P_T would equal the percentage of “treatments” among the buyers from a population where the treatment had been randomly assigned to half of the population.

SAS PROC LOGISTIC can fit the Buyer model.¹ The weight W can be added to PROC LOGISTIC as shown:

¹ However, other techniques (e.g. decision trees) could be used.

PROC LOGISTIC DATA = BUYERS_ONLY DESCENDING; MODEL TARGET = <predictors>; WEIGHT W; ²

Campaign Incremental Units and the Buyer Model

The incremental units for a campaign are computed by: $\text{Size}_T * (\text{Sales}_T / \text{Size}_T - \text{Sales}_C / \text{Size}_C)$
(The suffices “T” and “C” refer to treated and control, “Size” is the total prospect count, “Sales” gives the buyer count.)

e.g. Incremental Units = $50,000 * (1,000 / 50,000 - 450 / 25,000) = 100$. We note that $W = 2$.

The calculation of campaign incremental units equals the incrementality implied by the Buyer model as shown: ³

- With weighting by W , the mean value of P_T from logistic regression is given by:
 $\bar{P}_T = (\text{Sales}_T) / (\text{Sales}_T + \text{Sales}_C * W)$ ⁴ and then $\bar{P}_C = 1 - \bar{P}_T$
- Let Lift = $(\bar{P}_T - \bar{P}_C) / \bar{P}_C$ and note that $\text{Size}_C * W$ is equal to Size_T
- Then, after algebraic manipulations, the incremental units for the campaign can be expressed as:

$$\text{Size}_T * (\text{Sales}_C / \text{Size}_C) * \text{Lift}$$

e.g. Incremental Units = $50,000 * (450 / 25,000) * (1,000 - 450 * 2) / (450 * 2) = 100$

2. In-Market Model

The analysis data set consists of the control group sample. Let target = 1 if the prospect buys and target = 0 if not.

Notation: S_C is the probability that the control group prospect buys the product.

SAS PROC LOGISTIC can be used to fit the In-Market Model. Weighting is not an issue.

If an all-purpose in-market model has already been developed on the population from which the POV campaigns are drawn, I believe this model could be used in place of a custom-fit model. Otherwise, a customized model would be developed.

The Buyer Model and the In-Market Model Combine to Form a Net Lift Model

The net lift (or expected incrementality) will be denoted by “NetLift” and is defined by:

$$\text{NetLift} = S_C * ((P_T - P_C) / P_C)$$

This NetLift really is a net lift model. That is, it does give the expected change in buy-rate due to making an offer to a prospect. This is because we can view $S_C * (P_T / P_C)$ as estimating the treated buy-rate S_T for this prospect. Here, (P_T / P_C) is an “expansion” factor that says that for every 1 control buyer, there are (P_T / P_C) treated buyers.

This net lift model is an example of a “probability decomposition model” (PDM) since it is the product of two factors with each involving a probability model. PDM’s are discussed further in a later section.

3. Redemption Model

The analysis data set consists only of the **buyers** from the **treated** samples. Let target = 1 if the prospect redeemed the offer when making the purchase and target = 0 if not.

Notation: P_R is the probability that the prospect redeems given that the prospect is a treated buyer.

² The weight statement has the effect of adjusting the model’s intercept term aside from random effects on the other coefficients.

³ It should be a minimum expectation of the Buyer model that its predictions P_T could be used to recapture the campaign’s incremental units.

⁴ This “equality” is sometimes only a close approximation.

SAS PROC LOGISTIC can be used to fit the Redemption Model. Weighting is not an issue.

The Redemption Model introduces the dependency of NetLift_j on cost at the prospect level.

4. A Formula for "M", Predicted Marketing Profit for a Prospect who is offered a POV.

The calculation of M requires external inputs:

- G for average unit profit
- V for POV coupon value

$$\text{Then: } M = \text{NetLift} * G - P_R * (S_C + \text{NetLift}) * V$$

M (for money) will denote the **Profit Model** prediction.

e.g. Let $S_C = 0.02$, $G = \$1,000$, $V = \$40$, $P_R = 0.55$, and $P_T = 0.51$, $P_C = 0.49$, so that $\text{NetLift} = 0.00082$, then:

$$M = 0.00082 * \$1,000 - 0.55 * (0.02 + 0.00082) * \$40 = \$0.37$$

The expected cost per prospect is $C = P_R * (S_C + \text{NetLift}) * V$. Here, S_C gives the control buy-rate and NetLift gives the increase (decrease) in the sales-rate due to the offer so that $S_C + \text{NetLift}$ is the treated group buy-rate. Then P_R gives the probability these treated sales will lead to a redemption. Multiplication by V gives the expected redemption cost. Optionally, communication unit costs could also be added to expected cost.

Usage of the Profit Model

"M" can not literally be used to predict profit when applied to a new campaign prospect list because the particular features of the campaign used to fit "M" will not be identically repeated. But the "ranking" of prospects by "M" can be used for list selection. The applicability of "M" to a new campaign depends on offering a similar POV to a similar universe of prospects, presumably drawn from an in-house transactional database with demographic overlays.

PROBABILITY DECOMPOSITION MODEL (PDM) OF JUN ZHONG

Roman Kubiak (2012) described the application of a net lift model for 1800Flowers.com. He utilized a PDM. Kubiak cites a presentation by Jun Zhong (2009) as the source of the technique.

The Zhong PDM decomposes the expected incrementality of a prospect into the product of two factors where each factor involves a probability.

- One probability is the probability of being in-market. This in-market model is built on the treated group.
- The other probability is the probability that a buyer was in the treated group. This model is built on the buyers from treated and control with the restriction that treated and control groups are of the same size or that weighting is applied to equalize the sizes of the treated and control groups.

Using the notation of this paper this net lift model is written as: $\text{NetLift}_2 = S_T * (2 - 1 / P_T)$ where S_T is an in-market probability which is fitted on the **treated** group. The suffix "2" distinguishes NetLift₂ from the NetLift discussed earlier in this paper.

NetLift = $S_C * ((P_T - P_C) / P_C)$ and $\text{NetLift}_2 = S_T * (2 - 1 / P_T)$ provide different values of the expected incrementality per prospect.

- The in-market models S_T and S_C will give different probabilities to a prospect due to (i) randomness in the sample selection; and (ii) different relationships between predictor variables and the target due to the effect of the offer on buying.
- The second factors of the two NetLifts utilize the same P_T but with very different formulas. In fact, when P_T is not equal to 0.50, then

$$\{ (P_T - P_C) / P_C \} / \{ (2 - 1 / P_T) \} = P_T / P_C \quad (\text{eq. A})$$

From (eq. A) it is seen that the values of $(P_T - P_C) / P_C$ and $(2 - 1 / P_T)$ are very different in the cases where P_T is not near 0.50^{5 6}

However, the values of S_T and S_C may compensate when P_T is not near 0.50 so that two NetLift values may be similar. Furthermore, in many practical applications P_T is concentrated near 0.50 since the overall treated buy-rate for the campaign is close to 0.50, and, in this case, a Buyer Model may have difficulty in distinguishing a treated buyer from a control buyer.

S_T or S_C?

I believe it is better to fit S_C for NetLift rather than S_T for the reasons given below:

- The model for S_T depends on a specific campaign while S_C is a generic model. Specifically, S_T may not generalize to the next campaign. If an all-purpose in-market model has already been developed on the population from which the POV campaigns are drawn, this model could be used.
- The predictor variables that reflect the effect of the treatment, when estimating S_T , may be overwhelmed by in-market propensity variables. So, of the two causes of buying (treatment and propensity), only propensity-related variables may appear in the model.

The Intuitive Appeal of $S_C * (P_T - P_C) / P_C$

When $P_T > 0.50$ the quantity $S_C * (P_T - P_C) / P_C$ gives the prospect's expected increase in buy-rate over the base control rate. This formula seems to me to be intuitively appealing as the correct formulation.

Only a campaign that has statistically significant incremental sales should be used to fit P_T . In this case most of the distribution of P_T will be greater than 0.50.

Of course, both NetLift methods face the same challenge in fitting P_T .

But No Conclusion: Any conclusion about which NetLift is more effective would depend on applying both approaches to actual campaigns and measuring their relative performances on the holdout groups. I have not performed this test.

VALIDATION OF A PROFIT MODEL

After fitting a profit model to a training data set drawn from a campaign, the model is validated on a holdout sample. If this is not practical due to (the usual) small sample sizes of buyers and redeemers, the validation can be performed on one or more other similar campaigns for which there was measurable incrementality.

The validation of a profit model involves both the treated and control groups. Here are the steps for validation:

1. Obtain "M" for the campaign's treated and control prospects.
2. Use PROC RANK to rank these prospects into "g" ranks based on "M". Twenty-five (or so) ranks are recommended, if sample size permits, so that the values of M do not vary much within the rank.
3. Based on post-campaign processing, find the actual treated buy-rates b_t , control buy-rates b_c , and redemptions within each rank.
4. Compute incremental sales within each rank $[= (b_t - b_c) * \text{treated_count}]$. Then compute actual campaign profit within rank by: Actual Profit = G * Incremental Sales - V * Redemptions
5. Create a cumulative Lift Chart or cumulative Gains Chart across the ranks and compute the cumulative actual profit down the ranks.

The modeler can determine the rank for which maximum actual profit occurs for the campaign. This rank (and the related value of M) suggests a stopping point for the selection of prospects for future campaigns.

⁵ $(P_T - P_C) / P_C > (2 - 1 / P_T)$ when $P_T > 0.5$. Also, as $P_T \rightarrow 1$, we have $(P_T - P_C) / P_C \rightarrow \infty$ while $(2 - 1 / P_T) \rightarrow 1$

⁶ If $G(P_T) = (P_T - P_C) / P_C$ and $H(P_T) = (2 - 1 / P_T)$ and $x = P_T - 0.5$, then $G(x) = -H(-x)$

SIMULATION EXAMPLE

A POV campaign with 1 million prospects was simulated. There was a random assignment of 50% to treated and 50% to control. The data set was also randomly split with 50% to training and 50% to holdout. So, there were 250,000 treated in training, 250,000 control in training, etc.

A variable called Y indicates a purchase (Y=1) or no purchase (Y=0). The control buy-rate was about 2% and treated buy-rate was about 2.3%. A target variable called target1 was defined for buyers. Target1=1 for treated and target1 = 0 for control. Target2 was created for treated buyers where target2=1 for redeemers and target2 = 0 for non-redeemers.

Three predictors were created: X1, X2, X3. Variable X1 measures “RFM” (recency, frequency, monetary value). Variable X2 captures a tendency to be influenced by the offer and Variable X3 captures a tendency to redeem the offer among treated buyers.

The code to generate the sample, fit the profit model, and measure the fit on a holdout sample is given at the end of the paper. The predictors X2 and X3 were designed to have weak relationships to their targets. This is very likely to be the case for a practical application.

The measurement of the fitted profit model on the holdout sample is shown below. Only five ranks were used in this example but, in practice, a larger number, depending on sample size, should be created.

The ranks are created by scoring the combined treated and control records from the holdout sample by the profit model and ranking by the predicted profit. The “predicted profit of a rank” is the sum of the predicted profits for the treated records in the rank. This gives column B. Actual profit was previously defined above in the “Validation of the Profit Model” (dot-point 4). This is column A.

The cumulative actual profit column (C) suggests to the marketer that next time a very similar campaign is run, the selection of prospects should stop after rank 3 where cumulative actual profit is maximized. Or, if the campaign list is volume-constrained, the prospects would be selected by rank until the targeted quantity was met.

Table 1: Validation Report for Profit Model – Holdout Sample

Rank for Variable M	Treated Prospects	Actual Profit (A)	Predicted Profit (B)	Actual Incremental Units	Predicted Incremental Units	Cum. Actual Increm. Units	Cum Actual Profit (C)
0	50,287	\$506,868	\$634,980	380	466	380	\$506,868
1	49,839	\$411,346	\$339,093	299	250	678	\$918,214
2	49,882	\$202,141	\$174,826	147	129	825	\$1,120,355
3	50,263	\$106,031	\$24,796	94	39	919	\$1,226,386
4	49,729	\$-1,782	\$-252,734	37	-133	956	\$1,224,605

In **Table 2** where this report is run on the full training sample (no ranks), the actual profit and incremental units are essentially equal to the predicted. Specifically, summing over treated prospects:

$$\sum \text{NetLift} = \sum S_C * ((P_T - P_C) / P_C) = \text{Actual Incremental Units (approx.)}$$

Table 2: Validation Report for Profit Model – Training Sample

Rank for Variable M	Treated Prospects	Actual Profit (A)	Predicted Profit (B)	Actual Incremental Units	Predicted Incremental Units
All	250,000	\$918,225	\$923,720	748	752

SAS CODE for simulation:

```
%let seed = 12345;
Data Test;
do i = 1 to 1000000;
  x1 = ranuni(&seed); x2 = ranuni(&seed); x3 = ranuni(&seed);
  if mod(i,4) in(0 2) then TC = "T"; else TC = "C"; /* 50% of obs are control (C) */
  if mod(i,4) in(0 1) then Train_Holdout = "Tr"; else Train_Holdout = "Ho";
/* Approx 2% of "C" are buyers. x1 is strongly associated with buying */
```

```

    if TC = "C" then do; if ranuni(&seed) < .04 * x1 then Y = 1; else Y = 0; end;
/* Approx 2.3% of "T" are buyers. x2 increases buy-rate while -0.05 scales buy-rate */
    else do; if ranuni(&seed) < .04 * (x1 + .25*x2 - .05) then Y = 1; else Y = 0; end;
    target1 = .; target2 = .;
    if TC = "T" & Y = 1 then target1 = 1; /* treated group buyers */
    else if TC = "C" & Y = 1 then target1 = 0; /* control group buyers */
    if TC = "T" & Y = 1 /* treated group redemptions */
    then do;
/* x3 increases redemptions, x1 reduces. Redemption rate is approx. 48% */
    if ranuni(&seed) >= .48 + .20*x3 - .10*x1 then target2 = 1; else target2 = 0;
    end;
    w_inmkt = (TC = "C" & Train_Holdout = "Tr"); /* weight for logistic */
    w_increm = (Y = 1 & Train_Holdout = "Tr"); /* weight for logistic */
    w_redeem = (TC = "T" & Y = 1 & Train_Holdout = "Tr"); /* weight for logistic */
    x1_decile = floor(round(100*x1,1)/10);
    x2_decile = floor(round(100*x2,1)/10);
    x3_decile = floor(round(100*x3,1)/10);
    output;
    end; run;

proc logistic data = Test descending; class x1_decile;
    model Y = x1_decile; output out = inmkt pred = SC;
    weight w_inmkt;
    title1 "Model for in-market of control group (SC)"; run;
proc logistic data = Test descending; class x1_decile x2_decile;
    model target1 = x1_decile x2_decile; output out = increm pred = PT;
    weight w_increm;
    title1 "Model for incrementality (PT)"; run;
proc logistic data = Test descending; class x1_decile x3_decile;
    model target2 = x1_decile x3_decile; output out = redeem pred = PR;
    weight w_redeem;
    title1 "Model for redemptions (PR)"; run;

%let V = 75; %let G = 1500;
data M; merge inmkt increm redeem;
    if Train_Holdout = "Ho"; /* <= Holdout sample */
    NetLift = ((2*PT - 1) / (1 - PT)) * SC;
    Redeem = (Y = 1 & target2 = 1); YT = (Y = 1 & TC = "T"); YC = (Y = 1 & TC = "C");
    TC_t = (TC = "T"); TC_c = (TC = "C");
    M = (NetLift * &G - PR * (SC + NetLift) * &V); /* M for all records */
    NetLiftT = NetLift * TC_t; /* NetLift for treated records only */
    MT = M * TC_t; /* M for treated records only */
    keep M MT TC YT YC TC_t TC_c Redeem NetLift NetLiftT PT SC PR; run;
proc rank data = M descending groups = 5 out = M_ranked; var M; ranks M_ranked; run;
/* In practice GROUPS >> 5 */
proc means data = M_ranked noprint; class M_ranked;
var M MT NetLiftT YT YC TC_t TC_c Redeem;
output out = gains sum = M MT NetLiftT YT YC TC_t TC_c Redeem; run;
data gains2; set gains; retain cum_actual_incremental cum_actual_profit 0;
    actual_incremental = TC_t * (YT/TC_t - YC/TC_c);
    actual_profit = &G * actual_incremental - &V * Redeem; /* for treated records */
    if _type_ = 1 then cum_actual_incremental + actual_incremental;
    if _type_ = 1 then cum_actual_profit + actual_profit;
    label actual_incremental = "Actual Incremental Units";
    label actual_profit = "Actual Profit";
    label cum_actual_incremental = "Cum. Actual Increm. Units";
    label cum_actual_profit = "Cum Actual Profit";
    label NetLiftT = "Predicted Incremental Units";
    label MT = "Predicted Profit";
    label TC_t = "Treated Prospects";
run;
proc print data = gains2(where = (_type_ = 1)) label noobs;

```

```

var M_ranked TC_t actual_profit MT actual_incremental NetLiftT cum_actual_incremental
cum_actual_profit;
title1 "HOLDOUT: Actual v. Predicted Profits - Actual v. Predicted Incremental Units";
format actual_profit MT actual_incremental NetLiftT cum_actual_incremental
cum_actual_profit 8.0; run;

/* Check Incremental Units in Training Sample */
/* Answer 750 */
proc freq data = test; tables Y * TC;
where Train_Holdout = "Tr";
run;
/* Check Incremental Units in Holdout Sample */
/* Answer 975 */
proc freq data = test; tables Y * TC;
where Train_Holdout = "Ho";
run;

```

REFERENCES

Kubiak, R. (2012). "Net Lift Model for Effective Direct Marketing Campaigns at 1800flowers.com", SAS Global Forum 2012, Paper 108-2012.

Lo, Victor S. Y. (2002). "The True Lift Model", ACM SIGKDD Explorations Newsletter, 4(2): 78-86

Radcliffe, N. J. and Surry, P. D. (2011). "Real-World Uplift Modelling with Significance-Based Uplift Trees", Stochastic Solutions White Paper, <http://stochasticsolutions.com/pdf/sig-based-up-trees.pdf>

Zhong, Jun (2009), VP Targeting and Analytics, Card Services Customer Marketing, Wells Fargo in the presentation: "Predictive Modeling & Today's Growing Data Challenges" at Predictive Analytics World in San Francisco, CA 2009.

ACKNOWLEDGMENTS

Joel Wesley of Marketing Associates read several drafts of this paper and provided helpful suggestions.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Bruce Lund
Marketing Associates, LLC
777 Woodward Ave, Suite 500
Detroit, MI, 48226
blund@marketingassociates.com

All code in this paper is provided by Marketing Associates, LLC. "as is" without warranty of any kind, either express or implied, including but not limited to the implied warranties of merchantability and fitness for a particular purpose. Recipients acknowledge and agree that Marketing Associates shall not be liable for any damages whatsoever arising out of their use of this material. In addition, Marketing Associates will provide no support for the materials contained herein.

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. © indicates USA registration. Other brand and product names are registered trademarks or trademarks of their respective companies.