Analysing Efficacy of customer targeting methods by modelling quantile response rate decay

Abstract

This paper discusses and extends an approach to analysing efficacy of customer targeting methods (Courtheoux, 2004). It suggests fitting models to aggregated purchase incidence data obtained by sorting customers as to their predicted likelihood to purchase. Formulas for computing the optimal number of target customers and the maximum profitability are derived for three additional models. These formulas can be used to compare the efficacy of several customer targeting methods applied to several customer databases.

The initial study, which we try to extend, fitted an exponential function to five selected published decile tables from direct marketing modelling work (Magidson, 1993; Malthouse, 1993; Ratner, 2001; David Shepard Associates, 1999). Although the exponential function showed good fit for those selected decile tables, we argue that particular market situations or targeting methods could be better dealt with by other functional forms than the one described by the exponential model. For example while widely accepted and theory consistent predictive targeting models like logit and probit perform very well and often produce exponential decay in response rates by quantile, newer approaches can achieve better performance overall, like neural networks, or partially perform better only in the case of the first quantiles, like automatic tree classifiers (CHAID, CART). The latter by affecting the same likelihood score to all customers belonging to the same cluster provide less "granularity" in response rate decay. The number and nature of independent variables can also affect the shape of the response rate decay. In the direct marketing customer databases, targeting models that include more variables as for example FRAC-Frequency, Recency, Amount (of money), and Category (of product) (Kestnbaum, 1998) produce top quantiles that perform significantly better than the same quantiles for RFM-Recency, Frequency and Monetary models. All these aspects advocate the introduction of additional functional forms that may potentially better fit response rate decay by quantile in such cases. Therefore, in this paper, we introduce three additional functional forms and derive the formulas that compute the optimal number of customers to be targeted and the maximum profitability for each of those response rate decay functions. We use those functional forms and formulas to compare and analyse the efficacy of targeting methods applied to different datasets as well as different methods applied to the same customer data base. We also show that the additional models suggested by us fit

better to the datasets used by the original author then the exponential model. Changes in fitted model coefficients for situations that allow for a statistically reliable breakdown into groups smaller than deciles are also explored.

Key words: direct marketing, targeting efficacy, optimal file depth, response decay functions

Introduction and Objectives

The «gain chart» or "gains table" (also known as "ranking report") is a summary table that regroups for a given quantile level actual and predicted customer response indicators following the decreasing order of purchase probabilities that have been calculated using a predictive targeting model (adapted from Levin, Zahavi, 1998). Courtheoux (2004) suggests that by adjusting an equation to response probabilities from a decile table, that can be seen as a special kind of gain-chart, and by using formulas resulting through calculus there from to compute the financial gains of a direct marketing campaign, managers can approximate the optimum number of contacts to be targeted in order to maximise profits and infer additional profits that can result when investing in a better predictive targeting model.

Response rates in a decile table or more generally in a quantile table mark a systematic decay that can be captured and inferred by adjusting some well fitting decay function. The exponential function is a typical decay function and has been applied by Courtheoux (2004) to selected published decile tables from direct marketing campaigns. We argue that other decay functions can also be used. We introduce five additional functional forms and derive the formulas that compute the optimal number of customers to be targeted and the maximum profitability for each of those response rate decay functions.

The resulting approach can be used in managerial applications :

- to smoothen **quantile (decile) tables**
- to **interpolate them in order to find the optimal number of quantiles to be targeted (contacted)**
- to test potential financial efficacy of an improved targeting or scoring model or test several models.

As to Courtheoux (2004) decile tables, in general, and the five he mentions, in particular, have anomalies. For instance a lower ranked decile can have a higher response than its preceding adjacent decile. Another author (Hughes, 1995), in a RFM (Recency, Frequency and Monetary) segmentation context based on quintiles, calls euphemistically this kind of anomaly "cell personality". This is often due to insufficient number of customers per quantile. In such circumstances it is convenient to smoothen those fluctuations in order to obtain the expected monotone decay. Using finer quantiles than deciles can be considered when the number of contacts available in such a regroupment unit is sufficient to insure statistically significant decompositions.

The rest of the paper is organised as follows: In the first section we present a generalised version of the method and point out and correct some mathematical inconsistencies. In the second we apply the enhanced method to an artificially generated decile table and present the marketing analytic and financial indicators the method can produce. The third section extends the approach by introducing five additional response decay functions and by deriving formulas to compute cumulative response and profit for different file depths as well as formulas to find the optimal file depth and the maximum cumulative profit. The fourth section fits the five new models to five published decile tables from direct marketing modelling work and compares the results with those obtained with the exponential model. Finally some conclusions, limits of the approach and further research are discussed.

Generalising the method and correcting some mathematical inconsistencies

Curtheoux uses an exponential function in order to explain response rate decay by depth (proportion) of the customer file that is targeted. He uses the following formulation:

$$Y = k_1 + \exp(k_2 X + k_3) \qquad (1)$$

where

Y = response rate

X = depth (portion) of file

We suggest a simpler notation that keeps formulas more readable and allows for further extensions of the approach. It also conforms to notations usually adopted in well known texts dealing with Marketing Models like Lilien (1987), Lilien et al.(1992), Lilien & Rangaswami (2004). *Y* alias *f(x)* represents the response rate for the average customer representing a given file depth *x*.

$$y = f(x) = a\exp(-bx) + c \qquad (2)$$

The coefficients or parameters k1-3 in (1) or a-c in (2) determine the exact shape of the response decay curve by file depth. There is an upper bound *a* (or *exp(k3)*) a lower bound *c* (or *k1*) and a decay acceleration coefficient *b* (or *-k2*).

The profit function is the margin times the expected response rate minus the cost per contact as in formula (3). It represents the profit obtained from an average customer.

$$p = M\, f(x) - C \qquad (3)$$

where

p = profit per contact

M = profit amount from an order

f(x) = response rate per contact of depth x

c = cost per contact

By putting (2) into (3) the profit function becomes

$$p \;=\; M(a\exp(-bx)+c)-C \qquad (4)$$

Since *f(x)* alias *y* gives the response rate per contact for any particular file depth x, but not the cumulative response rate, calculating total campaign profits requires basic calculus.

The **cumulative profit for the whole file *P***, when the profit is approximated by a continuous function as the one in formula (3), is given by the indefinite integral of that function.

$$P = \int (M\, f(x) - C)\, dx \qquad (5)$$

By putting (2) into (5) the cumulative profit becomes

$$\text{P}= \; x(M\,c-C)-\frac{M\,a}{b}\exp(-b\,x) \qquad (6)$$

The **cumulative profit for a given file depth D , *P(D),*** per average contact , is given by the definite integral of the profit function between a lower limit of zero and an upper limit of D.

$$P(D)=\int_0^D (M\, f(x)-C)\, dx \qquad (7)$$

From 7 it can be easily seen that the maximum cumulative profit is obtained where the response rate attains break-even. When the response rate falls under the break-even rate profits are negative and cumulative profit diminishes:

$$f(x)=C/M \qquad (8)$$

By putting (2) into (7) and calculating the definite integral between those limits we obtain

$$\text{P(D)} = \; D(M\,c-C)+\frac{M\,a(1-\exp(-b\,D))}{b} \qquad (9)$$

As the response rate diminishes with the file depth, cumulative profit increases as long as the margin times the response rate is bigger than the cost per contact and should be maximum at a depth where break-even is attained. Targeting contacts deeper in the list comes with financial losses. In mathematical terms the maximum value of the cumulative profit (which is the integral of the profit function) is attained at a depth where its derivative (meaning the profit function itself) equals zero.

By equaling (4) with zero and solving, the depth (*D**) for which the optimal cumulative profit is obtained is given by

$$D^* = x^* = \frac{-1}{b}\ln\left(\frac{C - M * c}{Ma}\right) \qquad (10)$$

where D* alias x* is the file depth that should be used in order to obtain the maximum cumulative profit. The optimal cumulative profit P(D*) can then be obtained by introducing the value of D* into 9.

The original paper appears to have a major inconsistency at this point. While the formula obtained for the optimal file depth is equivalent to the one presented in (10) the way it has been derived is somewhat tautologic. Instead of recognising that the derivative of the integral of a function is the function itself, Courtheoux computes the cumulative profit function as a definite integral of the profit function (see Courtheoux , 2004 formula 4) then the optimal file depth is obtained by solving the differential of that cumulative profit formula (see Courtheoux , 2004 formula 5) when it equals zero (see Courtheoux , 2004 formula 6) . This doesn't make much sense as this differential is necessarily the profit function itself that could have been solved directly. By comparing formulas 3 and 5 in the original paper it can be seen that the function entering the integral in the first formula is identical to the result of the differential of that integral in the second formula, although the components of that function have been written in a different order. Using the second derivative in order to verify that the optimum cumulative profit point is a maximum (see Courtheoux , 2004 formula 8) is also somewhat tautologic as by definition as long as the response rate is bigger than break-even (*C/M*) the cumulative profit is increasing and after break-even it decreases. This means that the cumulative profit function that is the integral of the profit function itself must be concave downward and has necessarily a negative second derivative.

Additional indicators of marketing efficacy result from integrating the response decay function in order to obtain the cumulative response R(D) at a given file depth (D) or the proportion of persons from the whole file that respond when only a part corresponding to the file depth has been contacted

$$R(D) = \int_0^D f(x)\,dx \qquad (11)$$

The cumulative response rate is then obtained by dividing the cumulative response by that depth (R(D)/D). The number of respondents is the cumulative response times the number of names in the file

$$B(D) = N\,R(D) \qquad (12)$$

The cumulative percentage of total potential buyers *F(D)* is *R(D)/R(1)* or

$$\mathrm{F(D)} = \int_0^D f(x)\,dx / \int_0^1 f(x)\,dx \qquad (13)$$

The application of this approach to real data will be illustrated in the next section.

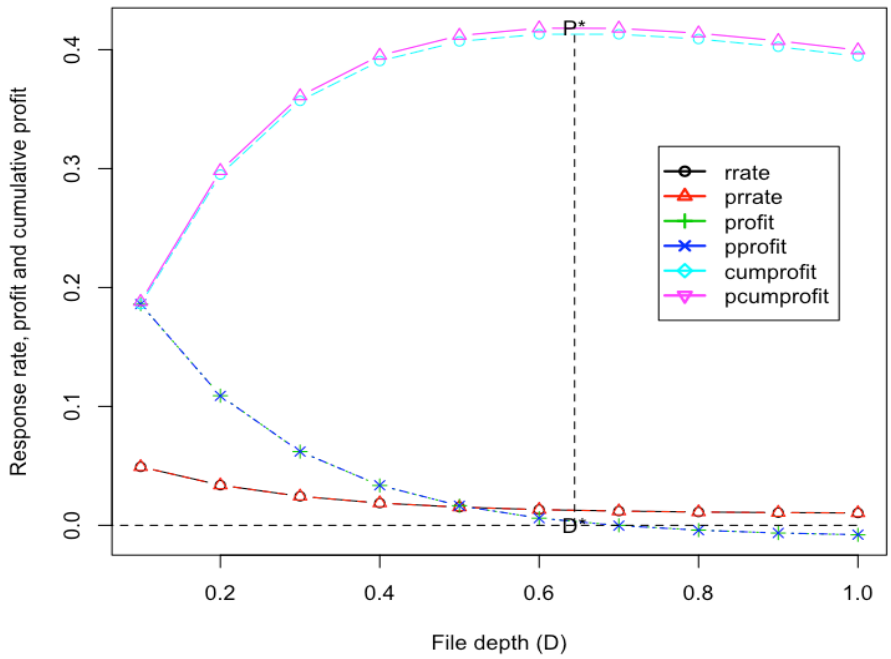
Enhanced method application

The application of the method to data from the original paper is illustrated in table 1 and figure 1. The response rate by decile or more generally by quantile is decreasing and is placed on the continuum of the file depth at the centre of each decile (quantile). By fitting the curve given in formula (2) to those data the coefficients of that model are obtained and used to compute the predicted response rate, profit and cumulative profit.

Table 1 - Actual and Predicted Response Rate and Profitability by decile

Decile	File depth	Actual Response Rate	Predicted Response Rate	Actual cumulative Profit	Predicted cumulative Profit
1	0.05	4.92	4.921	0.186	0.106
2	0.15	3.38	3.377	0.295	0.251
3	0.25	2.44	2.441	0.357	0.334
4	0.35	1.87	1.874	0.390	0.381
5	0.45	1.53	1.530	0.407	0.405
6	0.55	1.32	1.321	0.413	0.416
7	0.65	1.20	1.195	0.413	0.419
8	0.75	1.12	1.118	0.409	0.416
9	0.85	1.07	1.071	0.402	0.411
10	0.95	1.04	1.043	0.394	0.404

Figure 1 - Actual and Predicted Response Rate and Profitability by decile



* rrate=Response Rate, prrate = Predicted Response Rate, profit=Profit, pprofit=Predicted Profit, cumprofit=Cumulative Profit, pcumprofit=Predicted Cumulative Profit.

Curve fitting the decile (quantile) table – a two steps approach: As the response decay by quantile is non-linear, a non-linear estimation procedure needs to be used in order to determine the coefficients of the model. Usually these estimation procedures expect some initial (guess) values of the coefficients that will be calculated. As under given circumstances the response rate decay function (formula 2) can be linearised, we suggest the use of a two step curve fitting procedure. First use linear regression in order to compute initial parameter estimates which will serve as a start solution for the second step in order to produce final nonlinear estimates.

If one or two of the three parameters are known or can be approximated from data, the exponential model in formula (2) can be linearised by using logarithms. These parameters are usually approximated from upper and lower bound values observed in the data. The linearised exponential function is as follows:

$$\ln(y-c)=\ln(a)-b\ln(x) \qquad (14)$$

where the dependent variable y', a substitute for ln(y-c), is a linear function of the dependent variable x', a substitute for ln(x). The only parameter that needs to be approximated in advance is c, that is the lower bound (*L*) or minimum value and can be taken from existing data. By applying linear regression to the "canonic" formulation of the regression equation (see also tables 2 and 3) $y'=a'+b'x'$ parameters a' here *ln(a)* and b' here *-b* are obtained.

In the second step resulting initial parameters are computed $a = exp(a')$, $b=-b'$ and c = the minimum response rate in data (a= 0.06778895, b = 6.13821 and c = 0.0104). The final values of these parameters will be obtained through non-linear estimation that will further adjust parameters to real data. We used the nonlinear least-squares package (nls) in the statistical Software R in order to compute these estimates and obtain a= 0.05 , b= 5 and c = 0.01. From table 1 and especially from figure 1 it can be seen that quasi perfect fit between actual and predicted response rate decay data has been obtained[1].

For the linearisation of other response rate decay functions that we suggested in order to extend this approach see table 3 in the next section.

Compute and maximise cumulative profit and response: The cumulative profit for a given file depth P(D) can be approximated by applying formula (7) in general and (9) in particular.
The optimal file depth is attained at D^* by solving the profit function (3) in general terms and (4) in particular equaled to zero or by solving the response rate function f(x) equal to break-even. The resulting value of the optimal file depth D*=0.6444473, meaning that contacting 64.5% of the customers is optimal, is obtained by applying formula (10) and the maximum cumulative profit P(D*) of $0.4186425 per average customer results by introducing the optimal file depth (D*) in formula (9). By multiplying this amount with the number of customers (N=1,500,000, see Courtheoux, 2004) in the campaign database results $627963.8, the optimal profit of the campaign.

Additional marketing efficacy indicators result from analysing the cumulative response rate. From (11) the cumulative response for the optimal file depth is R(D*)=0.01610622, meaning that 1.6% of all names or that 161062 persons in the file respond. The cumulative response rate is then R(D*)/D* = 0.2499230 or 2.5%. The potential proportion of responders in the whole file (when the file depth is 1 or 100%) is R(1)=0.01999354 or nearly 2%. The cumulative percentage of total potential buyers for the optimal depth is F(D*) = R(D*)/R(1) = 0.8055712, meaning that by contacting the optimal 64% of the file 80.5% of the potential respondents order.

1 This is probably due to the fact that these data taken from the original paper were created as an example using an exponential function having those parameters.

Extending the approach to other functions

Courtheoux (2004) fitted an exponential function (formula 1) to five selected published decile tables from direct marketing modelling work (Magidson, 1993; Malthouse, 2003; Ratner, 2001; David Shepard Associates, 1999). Although the exponential function showed good fit for those selected decile tables, we argue that particular market situations or targeting methods could be better dealt with by other functional forms than the one described by the exponential model. For example while widely accepted and theory consistent predictive targeting models like logit and probit perform very well and often produce exponential decay in response rates by quantile, newer approaches can achieve better performance overall, like neural networks, or partially perform better only in the case of the first quantiles, like automatic tree classifiers (CHAID, CART). The latter by affecting the same likelihood score to all customers belonging to the same cluster provide less "granularity" in response rate decay. The number and nature of independent variables can also affect the shape of the response rate decay. In direct marketing customer databases, targeting models that include more variables as for example FRAC-Frequency, Recency, Amount (of money), and Category (of product) (Kestnbaum, 1998) produce top quantiles that perform significantly better than the same quantiles for RFM-Recency, Frequency and Monetary models. All these aspects advocate the introduction of additional functional forms that may potentially better fit response rate decay by quantile in such cases. Therefore, in this paper, we introduce five additional functional forms and derive the formulas that compute the optimal number of customers to be targeted and the maximum profitability for each of those response rate decay functions.

At this stage we chose to select and adapt static response functions that are well known in marketing modelling literature. As most of these are typically growing response functions we modified them in order to obtain their symmetric decreasing functions using the reverse process by which the modified exponential function has been obtained from the exponential function (formula 2). As the values of $\exp(-bx)$ decrease from 1 to 0 for x increasing between 0 and the infinity, its extended formulation $a\exp(-bx)+c$ varies between an upper bound (U) value of $a + c$ and a lower bound (L) value of c, while a can be seen as the amplitude. Its symmetric function $a(1-\exp(-bx))+c$ is the modified exponential function, a static response function that is well known in marketing. "This functional form has been used in an analysis of the optimal number of sales-people (Buzzell 1964a), for the study of advertising expenditures in coupled markets (Shakun 1965), in a probe of advertising budget allocation across territories (Holthausen and Assmus 1982), in an investigation of the effects of channel effort (Rangan 1987), in an assessment of resource allocation rules (Mantrala, Sinha, and Zoltners 1992), and in the examination of the impact of booth space, booth location, and number of personnel at the booth on industrial trade show performance (Gopalakrishna and Lilien 1995)" (Fransens, p.110).

The reverse process by which to obtain symmetric decreasing functions from original increasing functions can be illustrated by what we call the modified ADBUDG function $a\left(1-\frac{x^b}{c^b+x^b}\right)+d$ that has been derived in a similar way from the classical increasing ADBUDG function $a\left(\frac{x^b}{c^b+x^b}\right)+d$.

The symmetric functions were rather easy to obtain from these two functions, Adbudg and Exponential, as they belong to the functional form $(U-L)*g(x)+L$, that consist of a sub-function *g(x)* varying between 0 and 1 that is premultiplied by what can be called amplitude (U-L) and has lower bound (L) added. More generally, other response functions like Logistic and Gomperz functions, be they S-shaped or simply diminishing returns functions, cannot be reduced to have upper and lower bounds limited between 0 and 1. In this case the symmetric function is *U+L-f(x)*. Table 2 shows the original growing response functions and their symmetric response decay functions together with their upper (U) and lower (L) bounds.

Table 2 - Obtaining symmetric response decay functions from well known response functions in marketing, using their upper and lower bounds

Response function	Symmetric function	Upper (U) and Lower (L) bounds
Modified Exponential $a\exp(-bx)+c$	Exponential $a\exp(-bx)+c$	U=a+c, L=c, a=(U-L)
Adbudg $a\left(\frac{x^b}{c^b+x^b}\right)+d$	Modified Adbudg $a\left(1-\frac{x^b}{c^b+x^b}\right)+d$	U=a+c, L=d, a=(U-L)

Logistic $\frac{a}{1+\exp(-(b+cx))}$	Modified Logistic $a+\frac{a}{1+\exp(-b)}-\frac{a}{1+\exp(-(b\cdots}$	U=a, L= $\frac{a}{1+\exp(-b)}$ $b=\ln\left(\frac{L}{U-L}\right)$
Gompertz ab^{cx} a>0,1>b>0,c<1	Modified Gompertz $a+ab-ab^{cx}$	U=a, L=ab, b = L/a

The resulting modified models can be considered as either "reverse-S" shaped models and/or diminishing loss decay models as their original counterparts were "S" shaped and/or diminishing return response models. Some simpler additional models like the Fractional Root model and the Semi-logarithmic model can be included as they can also represent diminishing loss decay. In marketing literature these last two models have been sometimes classified as "linear parameter non-linear" models in contrast to the previously evoked models, which are rather "intrinsic" non-linear. The fact that these two models have linear parameters makes them easier to estimate by using linear regression. They are simpler to estimate but also simpler in the sense that they are less flexible when adjusting to data.

In order to initiate the two step parameter estimation process we suggested in the previous section, a starter estimation of the model parameters should be done using linear regression. In order to apply linear regression these non-linear models need to be transformed into linear models, that is "linearised". In the linearisation process upper (U) and/or lower (L) bound values will be taken from available data and by applying logarithmation a linear formulation of the models can be obtained as the ones shown in table 3. Using substitutions this linear formulation can be reduced to the canonic form needed by the linear regression method. Using the parameters of the canonic formulation that are output from linear regression the coefficients of the nonlinear function can be computed as in the last column of table 3. These coefficients will serve as a starting input into the nonlinear estimation algorithm that will be used.

Table 3 - Preparing the models for linear regression a first step in obtaining parameter estimates

Models	Nonlinear formulation	Linear formulation	Canonic	Coefficients
Exponential	$y=a e^{-bx}+c$	ln(y-c)=ln(a)-bx	y'=a'+b'x	a=exp(a'), b=-b',c=L
Logistic	$y=\frac{a}{1+e^{-(b+cx)}}$	ln(y / (a - y))=b + c x	y'=b+cx	a=U

Gompertz	$y=a\,b^{c^x}$	*ln(ln(a)-ln(y))=ln(-ln(b)) + ln(c) x*	y'=b'+c'x	a=U,b=exp(-exp(b')), c=exp(c')
Adbudg	$y=a\dfrac{x^b}{c^b+x^b}+d$	log((a+d-y)/(y -d))= b ln(c) - b ln(x)	y'=c'+b'x	a=U-L,b=-b', c=exp(c'/b),d=L
Modified exponential	$y=a(1-e^{-bx})+c$	ln(a + c – y) = ln(a) – b x	y'=a+b'x	a=exp(a'), b=-b, c=L
Modified logistic	$y=a+\dfrac{a}{1+\exp(-b)}-\dfrac{a}{1+\exp(-}$	$\ln\left(\dfrac{(1+\exp(-b))a}{-a+(1+\exp(-b))\,y}-1\right)=b+cx$	y'=b+cx	a=U, $\dfrac{a}{1+\exp(-b)}$
Modified gompertz	$a+ab-a\,b^{cx}$	*ln(ln(a)-ln(a+ab-y))=ln(-ln(b)) + ln(c) x*	y'=b'+c'x	a=U,b=exp(-exp(b')), c=exp(c')
Modified adbudg	$y=a\left(1-\dfrac{x^b}{c^b+x^b}\right)+d$	ln((a + d - y)/(y - d)) = - b ln(c) + b ln(x)	y'=c'+bx	a=U-L, c=exp(b/c'),d=L
Semi-logarithmic	$y=a+b\ln(x)$	y=a + b ln(x)	y=a+bx'	
Fractional Root	$y=a\,x^b+c$	ln(y-c)=ln(a)+b ln(x)	y'=a'+ bx'	a=exp(a')

After the best fit parameters are obtained and the most appropriate response rate decay model is be selected, formulas that compute the cumulative profit for given file depth can be used. We derived such formulas for all additional models that we have introduced as shown in table 4

Table 4 - Formulas to compute Cumulative Response (R) and Profit (P) at given file depth (D) using different response rate decay functions

Function	$R(D)=\int_0^D f(x)\,dx$	$P(D)=\int_0^D (M\,f(x)-C)\,dx$
Exponential	$\left(Dc+\dfrac{a(1-\exp(-bD))}{b}\right)$	$D(Mc-C)+\dfrac{M\,a(1-\exp(-bD))}{b}$
Modified Logistic	$a\left(D+\dfrac{1}{c}\ln\left(\dfrac{\exp(b)+1}{\exp(b)}\right)\right)$ + $D\dfrac{a\exp(b)}{\exp(b)+1}$ - $\dfrac{a}{c}\ln\left(\dfrac{\exp(b)\exp(cD)+1}{\exp(b)}\right)$	$M\,a\left(D+\dfrac{1}{c}\ln\left(\dfrac{\exp(b)+1}{\exp(b)}\right)\right)$ + $D\dfrac{M\,a\exp(b)-C\exp(b)-C}{\exp(b)+1}$ - $\dfrac{M\,a}{c}\ln\left(\dfrac{\exp(b)\exp(cD)+1}{\exp(b)}\right)$
Modified Gompertz	$D(a+ab)-\dfrac{a\,Ei\left(c^D\ln(b)\right)}{\ln(c)}$ *	-

Modified Adbudg	$D\,a\;{}_2F_1\left(\frac{1,1}{b};1+\frac{1}{b};-\left(\frac{D}{c}\right)^b\right)+D\,d$ *	-
Semi-logarithmic	$D(a-b)+b\,D\ln(D)$	$D\left(M(a-b)-C\right)+M\,b\,D\ln(D)$
Fractional Root	$Dc+a\frac{D^{b+1}}{b+1}$	$D(M\,c-C)+M\,a\frac{D^{b+1}}{b+1}$

* ${}_2F_1(.)$ Hypergeometric function, *Ei(.)* Exponential Integral

The integrals for the modified Gompertz and Adbudg models are more sophisticated. They need numeric calculations for the Exponential Integral function and for the Hypergeometric function respectively. Therefore for those models no closed form expression to compute the cumulative profit by file depth is available. Nevertheless the value of these cumulative profit functions or integrals can be easily approximated using the profit function itself.

Finally formulas to compute the optimal file depth for al functions are developed and listed in table 5.

Table 5 - Finding optimal file depth (D*) for different response rate decay functions

Function	D* -Optimal	D* in terms of break-even response rate (B), Upper- (U) and Lower (L) bounds
Exponential	$\frac{-1}{b}\ln\left(\frac{C/M-c}{a}\right)$	$\frac{-1}{b}\ln\left(\frac{B-L}{U}\right)$
Modified Adbudg	$\left(c^b\frac{(a+d-C/M)}{C/M-d}\right)^{(1/b)}$	$\left(c^b\frac{(U-B)}{B-L}\right)^{(1/b)}$
Modified Logistic	$\frac{-1}{c}\left(b+\ln\left(\frac{a}{a+\frac{a}{\exp(-b)+1}-C/M}-1\right)\right)$	$\frac{-1}{c}\left(b+\ln\left(\frac{B-L}{U+L-B}\right)\right)$
Modified Gompertz	$\frac{1}{\ln(c)}\ln\left(\frac{1}{\ln(b)}\ln\left(\frac{a+a*b-C/M}{a}\right)\right)$	$\frac{1}{\ln(c)}\ln\left(\frac{1}{\ln(b)}\ln\left(\frac{U+L-B}{U}\right)\right)$
Semi-logarithmic	$\exp\left(\frac{(C/M-a)}{b}\right)$	$\exp\left(\frac{(B-a)}{b}\right)$
Fractional Root	$\left(\frac{C/M-c}{a}\right)^{(\frac{1}{b})}$	$\left(\frac{B-L}{a}\right)^{(\frac{1}{b})}$

Optimal cumulative profit is obtained at a file depth were break-even is attained. Expressing the formulas for computing optimal file depth in terms of break-even response rate (*B*) and in terms of computed Upper (*U*) and Lower (*L*) bounds for the response decay functions makes interpretation of those formulas easier and can also help verify the consistency of the model parameters that were adjusted from data. For example if M=$25 and C=$1 the break-even response rate B = C/M = 1/25 = 0.04 must be bigger than the lower bound L in the exponential and modified logistic model, between the upper (U alias *a*+*d*) and lower bound (L alias d) for the modified Adbudg model, smaller than the sum of the upper and lower bound in the modified logistic and gompertz model. Looking for the estimated model parameters in table 6 shows that they all satisfy these conditions.

Fitting the models to several decile tables

Adjusting the response rate decay functions we introduced to published data from direct marketing campaigns (Malthouse, 2003) that have originally been used by Courtheoux shows that the new models fit to those data as well as the exponential model, as can be seen from table 6 and figure 2.

Table 6. - Response rate decay by decile, real data and fitted models

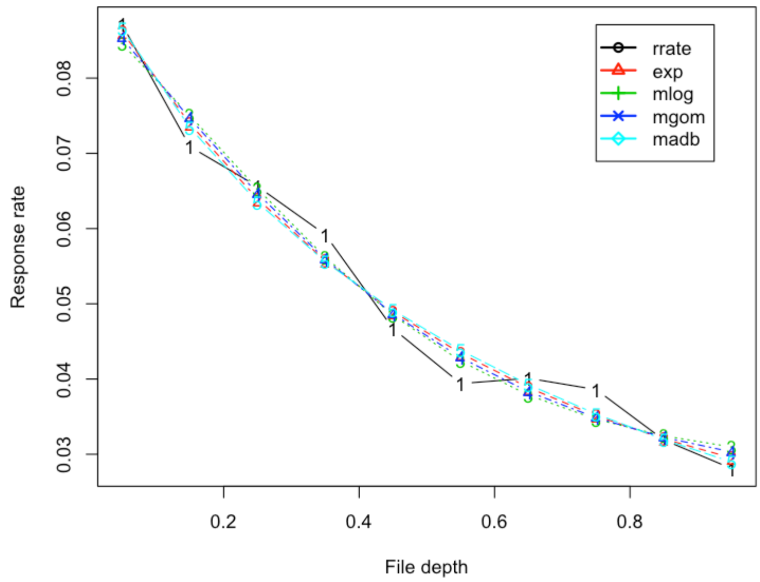
Decile	rrate	exp	mlog	mgom	madb
1	0.0871	0.08612666	0.08455444	0.08528013	0.08660732
2	0.0709	0.07394091	0.07502129	0.07464057	0.07341711
3	0.0656	0.06393177	0.06526565	0.06465161	0.06346995
4	0.0592	0.05571045	0.05619056	0.05591321	0.05559298
5	0.0467	0.04895761	0.04846752	0.04865183	0.04916641
6	0.0394	0.04341096	0.04237758	0.04283909	0.04380776
7	0.0401	0.03885505	0.03785779	0.03831056	0.03926284
8	0.0386	0.03511291	0.03465196	0.03485135	0.03535437
9	0.0319	0.03203919	0.03245050	0.03224649	0.03195420
10	0.0281	0.02951449	0.03097213	0.03030521	0.02896708
Fitted coefficients	a	0.07528	0.08897	0.09054	0.11505
	b	1.96768	-0.76896	0.27572	0.95011
	c	0.01790	4.41826	0.04018	0.68191
	d				-0.01957
Error SS		6.21E-005	7.379e-05	6.688e-05	6.163e-05

D*	0.5	0.62296	0.5985567	0.6095736	0.6326856
P(D*)	0.32375	0.33155	0.3301125	0.3299713*	0.3316051*
P(D*)/D*		0.532225993	0.551514167	0.541314945	0.52412304

- numerically calculated as no easy closed form expression available

All estimated model parameters in table 6 satisfy the condition that break-even response rate (C/M = 1/25 = 0.04) should be between the upper (U) and lower (L) bound levels as indicated in table 5.

Figure 2 - Response rate decay by decile, real data and fitted models

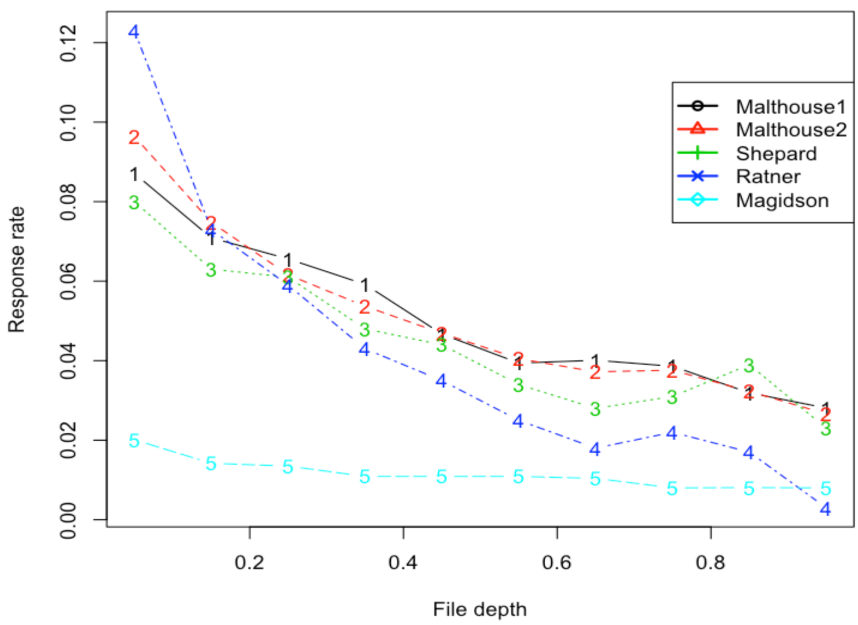


After the best fit parameters are obtained, the most appropriate response rate decay model can be selected and formulas that compute the cumulative profit for given file depth can be used. The optimal file depth (D*) computed using formulas in table 5 , is between 59.9% of customers for the modified logistic model and 63.3% for the modified adbudg model. Modified Adbudg being the model that fits best among available models, it is 63.3% of the customer list that should be selected. If the customer list contains N=1,000,000 names then the first 632,685 names in the sorted list should be contacted.

The maximum profit (P(D*)) computed essentially using formulas in table 4 is close to $0.33 per average customer for all models. The inferred profit of the campaign would be $N \times P(D^*)$ = USD 331,605 based upon the modified adbudg model.

The ratio between inferred optimum cumulative profit P(D*) and file depth D* is an inferred measure of customer targeting method efficacy. The modified logistic model gives the most optimistic ratio, followed in that order by the modified gompertz model, the exponential model and the modified adbudg model. After fitting the six suggested models to the five available decile tables (see table 7) this order remains virtually unchanged with the modified adbudg model at the end between the semi-logarithmic and the fractional root model.
The above analysis has been pursued to included all available decile tables whose graphical shape can be seen in figure 3.

Figure 3 - Five selected published decile tables from direct marketing modelling work



The five available decile tables were used to fit the six models that have been discussed previously (see table 7)

Table 7 - Error Sum of Squares from fitting six models to five decile tables

Model \ Data	Malthouse «Standard»	Malthouse « More »	Shepard	Ratner	Magidson
Exponential	6.207e-05	3.616e-05	0.0002046	0.0002597	7.605e-06
M. Logistic	7.379e-05	8.798e-05	**0.0001987**	0.0006768	1.011e-05
M. Gompertz	6.688e-05	5.937e-05	0.0002003	0.0004531	8.838e-06
M. Adbudg	**6.163e-05**	**1.916e-05**	0.0002073	0.0034430	4.598e-05

Semi-logarithmic	0.0001517	2.931e-05	0.0002546	0.0001435	**4.776e-06**
Fractional Root	0.0004452	0.0001859	0.0004343	**0.0001184**	3.004e-05

As can be seen from the table showing the Error Sum of Squares which indicates how well models fit to data, the Exponential Model is by no means the best fitting model to all situations. For each data set there are at least one or more models that outperform the exponential model.

The modified logistic model performs best in the Shepard dataset. By carefully looking at those data in figure 3, one can realise that in that case the response rate decay by decile takes the shape of a reversed « S » which is precisely the shape of the modified Logistic model, whose inflection point is exactly at half its upper bound (a/2) . It also involves a constant ratio of successive first differences of *1/f(x)*.

The modified Adbudg model, that can either be "reverse S-shaped" or have simply a decreasing loss decay, is the best fit for the first two decile tables Malthouse « Standard ». and « More ». These decile table resulted from the application of two different targeting methods. The second targeting method, being far better than the first, it captures significantly higher response rates in the first deciles. By applying formulas in table 4 and 5 for the modified Adbudg to the two Malthouse decile tables the superiority of the second method becomes obvious. For the better method optimal file depth is attained by contacting only 595870 persons in order to obtain a profit of $344487 instead of 632685 persons for a lower profit of $331605 for the other method. A more optimistic interpolation (approximation) is given by the Exponential model who indicates for the better method a optimal file depth of 565517 contacts with campaign profit $339083 and for the standard method a optimal file depth of 622958 with campaign profit $331554. This last result is almost identical to the one obtained by Courtheoux (2004).

The semi-logarithmic model fits best the Magidson decile table. All models fit rather well to this decile table as it starts with a strong decay between the first and the second decile and then takes a shape that seem reverse S-shaped.

The fractional root models adjusts best to the Ratner decile table. This decile table describes clearly diminishing loss decay and it also fits well with other similar functions like exponential and semi-logarithmic.

The relatively unsatisfactory fit of all selected models to the Shepard and Ratner decile tables, indicates that targeting methods often produce other response rate decay shapes that are nor diminishing loss curves nor "reverse S" shaped.

Conclusions

This paper discusses and extends an approach that uses curve fitting to infer response decay and calculus to analyse efficacy of customer targeting methods. The approach has been suggested by Courtheoux (2004). It originally consisted of adjusting (fitting) an exponential equation to represent response rate decay with file depth and was applied to five published decile tables from direct marketing modelling work. We give a more general formulation that can accommodate various response decay curves and several levels of response decay aggregation, not only decile tables but any quantile tables. We also discuss and correct some mathematical inconsistencies detected in the original paper.

We extend the approach by suggesting five additional response decay functions that are adapted from static response functions that are well known in marketing modelling literature.

These five functions together with the exponential function are then adjusted to the five decile tables using a two step curve fitting approach that uses linear and nonlinear regression. On the occasion linear formulas for all six functions are derived by using observed upper or lower bounds in data.

Calculus based formulas to compute cumulative response and profit for all these functions are derived, applied and tested using the same datasets (decile tables).

We give empirical evidence that the Exponential Model is by no means the best fitting model to all situations represented by available data. This confirms the need to extend the approach by including several response decay shapes (curves, models)

Also from empirical evidence we conclude that the suggested functions don't cover all response decay shapes resulting from direct marketing modelling work. This is particularly true for targeting methods that are capable to produce top quantiles that respond significantly better than the others. Therefore additional functional forms remain to be introduced by further research. A simple model that should give satisfactory results is the power series models as it can take many shapes.

Adjusting relatively simple models to aggregated response decay data and using calculus to infer marketing response and profitability measures is mainly a managerial artefact. In this sense it is similar to Little's subjective response based decision calculus and, like decision calculus itself, it can be accused, as Simon (1994) put it, of being not very scientific. As a managerial artefact it offers useful tools to perform sensitivity analysis to evaluate the impact of response decay function parameters on profitability and indirectly to evaluate the would be additional cost a company should be willing to pay for a better targeting method that would achieve response decay characterised by those parameters. A detailed illustration of such sensitivity analysis using the exponential function can be found in Courtheoux (2004).

As all the empirical work has been done using already aggregated data from published decile tables further research should apply the approach to disaggregated data from other customer files, from various direct or interactive marketing campaigns by using varying quantiles corresponding to various levels of aggregation and verify the precision that can be obtained by this curve fitting approach.

References

Courtheoux R. (2004), "Analysis of List Segmentation Efficacy and Campaign Optimization", Journal of Interactive Marketing, Vol. 18, No. 1, p. 71

David Shepard Associates. (1999). The New Direct Marketing. Boston: McGraw Hill.

Hughes, A.M. (1995) Making a database pay off using recency, frequency and monetary analysis. *Journal of Database Marketing*, Vol 3, 1, 1995, p.77-90.

Levin, N. and Zahavi, J. (1998), Continuous predictive modeling: a comparative analysis, Journal of Interactive Marketing, 12, 2, 5-22.

Lilien. G.L. (1987) "Analyse des Décisions Marketing avec LOTUS 1-2-3", (Traduit de l'américain par Pierre Y. Desmet), Economica, Paris. Lilien, G.L., Kotler, P., Moorthy, S.K. (1992) *Marketing Models*, Prentice Hall, Englewood Cliffs, New Jersey.

Lilien, G.L., and Rangaswamy, A. (2004) *Marketing Engineering: Computer-Assisted Marketing Analysis and Planning*, Revised Second Edition. Trafford Press, Victoria, BC, Canada.

Magidson, J. (1993). The use of the New Ordinal Algo- rithm in CHAID to target profitable segments. *Journal of Database Marketing*, 1(1), 36

Malthouse, E.C. (2003). Scoring models. In D. Iaco- bucci & B. Calder (Eds.), *Kellogg on Integrated Marketing* (pp. 227–249). Hoboken, NJ: Wiley.

Ratner, B. (2001). Finding the best variables for direct marketing models. DMA Research Council Journal, 31.

Simon, H. (1994), "Marketing Science's Pilgrimage to the Ivory Tower", en *Research Traditions in Marketing* G.Laurent, G.L. Lilien, B. Pras, Boston, Kluwer Academic Publishers, 27-43.