Behram Hansotia

is the former president and CEO of InfoWorks, a database marketing consultancy and an Omnicom company. Behram has a Masters degree in Systems Engineering and a PhD in Management Science from the University of Illinois at Urbana-Champaign. He has taught graduate database marketing courses at Northwestern University and Western Michigan University. He has published over 40 papers in Database Marketing and Operations Research journals and serves on the editorial board of *The Journal of Database Marketing & Customer Strategy Management*. Behram currently is an independent database marketing consultant and also serves on the board of directors of the Adult Learning Community of Southwest Michigan, a non-profit organisation dedicated to improving adult literacy.

Keywords: marketing by objectives, customer segmentation, hazard regression, customer equity, product propensity models, K-Means clustering

Behram Hansotia
226 North Street
South Haven
MI 49090, USA
Tel: +1 269 767 7235
E-mail: bhansotia@comcast.net

Marketing by objectives: Using segmentation based on purchase timing to enhance customer equity

Behram Hansotia

Received: 25 November 2008

Abstract

Enhancing customer equity is the primary challenge for all companies, but this responsibility is particularly critical for the Marketing department. Starting with a relatively simple three-dimensional value-based customer segmentation, we describe a goals-based approach to enhancing the value of the customer base. This requires setting marketing objectives for customers in each segment, namely the migration of customers from lower-value segments to higher-value segments. Several analytical tools for implementing our approach are discussed, including a purchase-timing model for the segmentation and clustering approaches, so marketers can develop greater insights into micro-segments characteristics and needs, as well as product propensity models for targeting the right products to different customers. We also discuss related applications of the segment transition matrix used to develop the marketing objectives. These include the development of *pro forma* income statements to justify possibly new marketing expenditures, the estimation of the lifetime value of a new customer and the value of the enterprise based on customer cash flows.

Journal of Direct, Data and Digital Marketing Practice (2009) **10**, 336–355. doi:10.1057/dddmp.2009.1

Introduction

Customer equity is the net present value (NPV) of customer-generated cash flows. In a previous paper,¹ we discussed some of the issues and challenges in developing individual-level estimates of customers' equities. In this paper, we discuss a systematic marketing approach to enhancing the value of the customer base. As discussed previously,^{2–3} enhancing customer equity results in increasing the value of the company, which in turn eventually results in increasing the company's market capitalisation. One could thus argue that increasing customer equity is the single most important responsibility of the marketing division.

For the sake of concreteness, the context of this paper is the retailing industry, and marketing contacts are primarily through the direct channels: e-mail when customers' e-mail addresses are available and regular mail when they are not. To implement our marketing strategy we discuss a value-based segmentation scheme that draws on a few basic customer behaviours related to their equity, namely the following:

- One-time buyers on average have lower value than multi-buyers (expected future cash flows of one-time buyers on average will be lower than those of multi-buyers).
- Single-product category buyers on average have lower value than multi-product category buyers (expected future cash flows of single-product category buyers on average is lower than those of multi-product buyers).
- Late customers on average have lower values than those who purchase 'on time' This requires estimating when a customer is likely to buy again and denoting the customer as late, or at-risk, if the purchase does not occur within that estimated time duration (late customers may never buy again, or may reduce their purchasing rate, resulting in lower expected future cash flows and lower equity than those customers who buy regularly, on time).
- Inactive customers have the lowest customer equity as they have not purchased in a very long time. For all practical purposes these may be considered as lapsed customers.

Our paper is organised as follows. We start with a general discussion of the segmentation scheme underlying our approach. The purchase-timing model, required for one of the segmentation's dimensions, is presented next. We follow this with a discussion of marketing objectives and performance goals for each segment. Next, we briefly discuss some tools for making customer communications more relevant and appealing, namely micro-segmentation and product propensity modeling. We then present testing in the context of developing optimal offers, so that we can meet our marketing objectives. This is followed by a discussion of a simulation method for developing *pro forma* income statements to justify the new marketing approach to senior management. As the customer income statement is closely related to estimating customer lifetime value and the valuation of the business, a discussion of these topics is also included. We next very briefly discuss the execution of marketing contacts, in the context of customised e-mails and customer performance tracking using segment transition matrices and customer value scorecards.⁴ We end this paper by briefly summarising the key ideas presented.

Defining segments

Some retailers sell to both consumers and businesses. If that is the case, we recommend that the segmentation be done separately for these two customer groups. In addition, if certain customers purchase only through direct channels, the web and catalogues, and others only

through stores, then consider them separately for the segmentation. For this situation, we would have four groups for segmentation: business direct buyers, business store buyers, consumer direct buyers and consumer store buyers. Customers who buy from both channels could be assigned to their dominant channel, or included in both groups.

For each of the groups, we propose a simple three-dimensional segmentation scheme for active customers, with the following basis dimensions:

- lifetime frequency of purchases;
- lifetime number of product category purchases;
- time since last purchase, or purchase recency.

Most companies use a simple definition of active customers based on the nature of the products they sell (durables vs. non-durables) and how customers pay for their purchases. One department store retailer, for example, denotes a customer as inactive if the recency of her purchase is greater than 36 months and if no credit card payment has been received in the last 8 months. Such *ad hoc* approaches typically work well in practice; however, a retailer could attempt to define a customer as inactive by building a purchase-timing model using hazard regression⁵ and defining a cut-off point for the purchase recency dimension, based on model score, to identify the inactives. This will result in different purchase recency definitions for inactives based on the customers' profiles and historic buying patterns. As we will need to build such a model to define late customers, we recommend this model-based approach to identifying active customers as well.

The first dimension is straightforward and needs little explanation. It is basically a count of the purchase occasions.

The second dimension is based on the product hierarchy pyramid used by most retailers. At the bottom of the pyramid are the product stock keeping units, which generally number in the thousands. Near the top of the pyramid are typically the departments (eg home decorations), and at the next lower level there are generally 60–80 product categories. Though there is no hard and fast rule, we have found it useful to work at this level to define the second dimension. A useful empirical approach would be to employ different levels of the product pyramid and obtain a frequency distribution of the count of customers' product category purchases. The pyramid level that results in a quick drop-off in number of product categories purchased beyond two typically works well.

The third dimension, purchase recency, is needed to first identify active customers and further to identify, within the actives, customers who are late in their next purchase. A purchase-timing model helps us identify the late customers based on the historic purchase behaviours of customers with similar characteristics. In a following section, we will discuss this purchase-timing model and how it needs to be built separately for different customer groups.

The next step in identifying the segments is defining the class levels of each dimension. We recommend no more than three levels for each dimension, and typically we have found that two levels do an excellent job differentiating among customers while maintaining simplicity; a key requirement for selling any new approach to an organisation. We therefore recommend the following classes for each dimension:

1. Lifetime purchase frequency, two classes: one-time buyers and multi-buyers.
2. Lifetime product category purchases, two classes: single-product category buyers and multi-product category buyers.
3. Purchase recency, three classes: on-time buyers, late customers and inactive customers (where the i th customer is deemed on time (not yet late) if his purchase recency $< k_i$, late, if his purchase recency is between k_i and m_i and inactive, if his purchase recency $> m_i$, where k_i and m_i are based on the i th customer's model scores).

This results in eight segments of active customers:

1. One-time, single-product category buyers, late in their next purchase.
2. One-time, single-product category buyers, not yet late in their next purchase.
3. One-time, multi-product category buyers, late in their next purchase.
4. One-time, multi-product category buyers, not yet late in their next purchase.
5. Multi-, single-product category buyers, late in their next purchase.
6. Multi-, single-product category buyers, not yet late in their next purchase.
7. Multi-, multi-product category buyers, late in their next purchase.
8. Multi-, multi-product category buyers, not yet late in their next purchase.

There exists a natural value hierarchy among these segments, in terms of their expected value of customer equity. Segment 1 has the lowest value and segment 8 the highest.

Customers in segment 1, on average, have lower customer equity than customers in segment 2, etc. An additional purchase typically results in customers moving to a higher-value segment. If a late customer makes a purchase, as a result of our marketing contacts, she would migrate to the 'not-yet-late' segment, after that purchase. Likewise, if a one-time buyer makes a second purchase she becomes a multi-buyer and if a single-product category buyer makes a purchase in a different product category she becomes a multi-product category buyer. Of course, customers in segment 8 stay in segment 8 if they make an on-time purchase, but migrate to segment 7 if they do not.

Typically, most new customers make an initial trial purchase in a single-product category; as a result we have found that segments 3 and

4 tend to be quite small. If that is the case, for the sake of simplicity (and without much loss in accuracy) it makes sense to combine segments 1 and 3 and segments 2 and 4. This results in six active customer segments plus the segment of inactives:

1. One-time buyers, late in their next purchase (1TB, L).
2. One-time buyers, not yet late in next purchase (1TB, NYL).
3. Multi-, single-product category buyers, late in next purchase (MB, 1PC, L).
4. Multi-, single-product category buyers, not yet late in next purchase (MB, 1PC, NYL).
5. Multi-, multi-product category buyers, late in next purchase (MB, MPC, L).
6. Multi-, multi-product category buyers, not yet late in next purchase (MB, MPC, NYL).
7. Inactives (I).

Purchase-timing model

To identify the universe of active customers, we recommend building a next-purchase-timing model using hazard regression⁶ — more specifically an accelerated failure time regression model.⁷ Start with a random sample of customers of size say 10,000 whose first purchase occurred at least 5 years ago (if customers previously classified as inactive are archived, ensure that the sample contains at least 2,000 randomly selected inactives as well, whose first purchase also occurred more than 5 years ago).

Set the time origin for the sample to 5 years ago, say 1 October 2003. The study period for this sample is now 5 years and the dependent variable is the time to next purchase, measured from the last purchase (recency of last purchase as of 1 October 2003 plus the time to next purchase measured from 1 October 2003). If a customer buys again within the study period, his time to next purchase is noted as observed, and if he does not we set it to the elapsed time in months up to the end of the study period and denote that observation as censored.

The predictor variables for the model are constructed from purchases occurring prior to 1 October 2003 as well as any available customer demographics. Typical predictors constructed from purchase data include lifetime number of purchases, lifetime number of product category purchases, lifetime monetary amount of the purchases, the last observed time between purchases and the average time between purchases for multi-buyers. As the last two variables are not available for one-time buyers, we need a flag for multi-buyers to operationalise these variables. We illustrate the approach for one of these variables:

Let Y , the time to next purchase, be the dependent variable, t be the average time between purchases for multi-buyers and z be an indicator variable such that $z=1$, implies a multi-buyer for whom t can be computed and $z=0$, a one-time buyer, for whom t is not available.

The predictor t can then be incorporated into the model for both one-time and multi-buyers, as follows:

$$Y = a + b(zt)$$

Here t is the computed average time between purchases for multi-buyers, and is set to any arbitrary constant for one-time buyers, and a and b are the model parameters. In the above case, in which we have not shown any of the other predictor variables, the estimated time to next purchase for a multi-buyer is $a + b(t)$ and a for a one-time buyer.

The model is estimated by the method of maximum likelihood, and can be developed in SAS⁸ using PROC LIFEREG. This procedure allows for a variety of underlying distributions for the dependent variable (time to next purchase) Y , including Weibull, exponential, gamma, log-logistic and log-normal. Note that all these distributions take on only positive values, as time cannot be negative. As a result, we have log-normal as a possible selection, but not the ubiquitous normal distribution. The distribution that fits the data best can be identified by developing competing models using each of the distributions (with all models based on the same set of statistically significant predictors, based on the Wald or the likelihood-ratio test) and picking the model with the highest value of the log-likelihood. (Note that all the log-likelihoods will be negative, so we pick the model with the lowest absolute value.)

After the best-fitting model is identified, we can use the OUTPUT statement in PROC LIFEREG to predict any percentile of Y , the predicted time to next purchase. (Unlike the standard multiple regression model, in which the expected value or the mean of Y is estimated, the accelerated failure time model allows us to estimate any percentile of Y given the distribution of Y and the values of the predictor variables.) Here, as we are interested in identifying active customers, we could define a customer as active if his purchase recency is less than, say, his 90th percentile of Y , $Y_{0.90}$ (the 90th percentile is of course arbitrary but it implies that there is only a 10 per cent chance for the customer to buy again beyond this point in time). To obtain predicted values for customers not included in the modeling sample, append these customers' observations (predictors) with their dependent variable set to missing. These observations will not be used in estimating the model, but predicted values will be generated for them by the OUTPUT statement. This way, $Y_{0.90}$ may be estimated for all customers. All customers whose recency of last purchase is less than $Y_{0.90}$ would now be deemed active.

A similar approach may be used to identify 'late' customers. A customer may be defined as late if his recency of last purchase falls between his 50th and 90th percentile of Y . That is, $Y_{0.50} < \text{recency of last purchase} < Y_{0.90}$. Again, $Y_{0.50}$ is arbitrary, and we are deeming the customer as late if his recency exceeds the median time to next purchase, but we could easily have set the cutoff at, say, $Y_{0.60}$.

A refinement that we have used in practice is to develop separate models for one-time buyers and multi-buyers and use these models' predictions to identify whether customers are late or not.

In the above formulation, we did not explicitly use the timing of customer marketing contacts, either before or during the study period, as predictor variables, as their exclusion is often standard practice. This is typically not a problem if all active customers are essentially receiving the same contacts. Of course, once a differentiated contact policy is implemented, it makes sense to recalibrate the model regularly as new data become available. Marketing contacts, however, can be incorporated into a model as time-varying predictors, but this would require using a different hazard regression model. The proportional hazard regression, first developed by Cox and Oaks,⁹ allows for the specific inclusion of time-varying predictors. In practice, however, the accelerated failure time model, as described above, works well and is easy to implement.

Setting segment performance goals and objectives

We recommend setting segment performance goals aligned with our overarching goal of enhancing customer equity. In a nutshell, our marketing programmes need to be geared towards migrating customers from lower-value segments to higher-value segments. For instance, this would require getting late customers to buy again, and converting one-time buyers into multi-buyers and single-product category buyers into multi-product category buyers. Each segment would thus have a well-defined goal. The goal for the highest value segment, namely 'Multi-, multi-product category buyers, not yet late in their next purchase', is to continuously recognise them and to ensure that they do not slip into the late category. The way to operationalise these goals is to set specific numerical objectives for customers' segment transition rates.

Segment transition matrix

The best way to set the transition rate objectives is to first develop a baseline segment transition matrix. This matrix shows segment members' transition rates before implementation of our new 'marketing by objectives' programme. If we define our time period as a month, the transition matrix shows the percentage of customers migrating from each segment at the end of a month to a different segment at the end of the following month, or remaining in the same segment. The (i,j) th cell, t_{ij} , of the transition matrix, $\mathbf{T}$, is essentially the conditional probability of migrating from state i in period $t-1$ to state j in period t . Here, the state of the system refers to the customer's segment. If X_t represents the state of the system in period t , then $t_{ij} = \text{Prob}(X_t = j | X_{t-1} = i)$. Note that as the t_{ij} are conditional probabilities, if the inactive customers are included, the states are mutually exclusive and collectively exhaustive, and hence the t_{ij} must sum to one in any row.

The segment 'One-time buyers not yet late in next purchase' may be considered as new customers, and this is the state customers enter the system when they make their first purchase. If late customers continue not

Table 1: Segment transition matrix

	Segment	1	2	3	4	5	6	7
T=	1	0.5			0.1		0.1	0.3
	2	0.5	0.3		0.1		0.1	
	3			0.4	0.2		0.2	0.2
	4			0.3	0.3		0.4	
	5					0.6	0.3	0.1
	6					0.3	0.7	
	7							1

Segment 1=(1TB, L) = Late, One-time buyers; Segment 2=(1TB, NYL) = Not-Yet-Late, One-time buyers; Segment 3=(MB, 1PC, L) = Late, Single product category, Multi-buyers; Segment 4=(MB, 1PC, NYL) = Not-Yet-Late, Single product category, multi-buyers; Segment 5=(MB, MPC, L) = Late, Multi product category, Multi-buyers; Segment 6=(MB, MPC, NYL) = Not-Yet-Late, Multi product category, multi-buyers; Segment 7=(I)=Inactives.

to make a purchase, they eventually enter the inactive state, and we can assume that for all practical purposes the inactives represent an absorbing state. That is, the probability of migrating out of the inactive state is zero. If the company occasionally tries to reactivate these customers (through marketing) and does successfully get a few of them to buy again, the reactivated customers should be considered as re-entering the system as new customers. Table 1 shows a seven-state transition matrix.

In row 1, 50 per cent of the late one-time buyers continue to be late, 30 per cent do not buy again and become inactives, 10 per cent buy again but in the same single-product category they had purchased in previously, making them multi-buyers, and 10 per cent purchase again in either a new product category or in one of the previous product categories from their first purchase (Table 1).

If we assume that there is no seasonality, estimating the baseline transition matrix is relatively straightforward. Start with a large random sample of customers acquired at least 1 year ago, or acquired prior to October 2007 (we assume for the sake of concreteness that the current month is November 2008 and we have customer data up to October 2008). To this sample add all new customers acquired in October 2007, and score them with the timing model as of 31 October 2007. Each customer can now be either 'not yet late', late or inactive. Next place the not-yet-late and the late customers in their appropriate purchase-frequency category (one-time, or multi-buyer) and product-frequency category (single-, or multi-product). This will result in the six active customer segments discussed earlier and a segment of inactives. Note each customer's segment membership as of 31 October 2007. Next, take all customers active on 31 October 2007 and add to this group all new customers acquired in November 2007, and using the timing model assign them to their appropriate segment as of 30 November 2007, as done earlier. The transition rates from 31 October to 30 November can now be easily computed for each segment. Repeat this process for December and compute the transition matrix, across all states from November to December 2007. These computations are then repeated for every pair of successive months up to 31 October 2008 (assuming this is the last month for which we have customer data).

If there is no seasonality we can now assume that the transition matrix is stationary, that is, its values are invariant with respect to time and the differences observed among the 12 t_{ij} values are essentially due to random variations. An average baseline transition matrix may now be calculated by averaging the values of the t_{ij} across the 12 matrices.

The no-seasonality assumption of course is not very realistic, particularly in retailing. There may also exist underlying trends in the transition probabilities and even longer-term business cycles. These can be addressed through time series forecasting approaches such as ARIMA, exponential smoothing or decomposition techniques (see for example Nelson¹⁰ and Wheelwright and Makridakis¹¹). This would of course require computing the monthly transition matrices over at least 36–48 months. However, unless the business is very new or rapidly changing, it may be reasonable to initially focus only on seasonality and compute monthly seasonality indices so that the transition probabilities (we use probabilities interchangeably with rates) can be deseasonalised. The estimated baseline transition probability for any month is then the average transition probability (as computed above) times the seasonality index for that month.

The marketing objectives for the year can now be expressed in terms of improvements in the average transition rates over the average baseline transition probabilities. For example, if the average baseline transition probability of customers migrating from the segment 'One-time, single-product category buyers, late in next purchase' to the segment 'Multi-, single-product category buyers, not yet late in next purchase' is 0.2, we might set the objective for this transition rate at 0.22, a ten per cent improvement. At the end of the year, the 12 observed monthly transition rates for these segments would be averaged and compared to the objective of 0.22. If this average exceeded 0.22, then marketing would have beaten its objective of enticing one-time, single-product category buyers late in their next purchase to buy again at a faster rate under the new marketing programme. To ensure that progress is being made each month, the observed monthly segment transition rates can be compared to the estimated (under business-as-usual) transition rates, computed from the average baseline transition probabilities for that month and that month's seasonality index.

Implicit in our goals of improving the baseline transition probabilities is that marketing programmes should be entirely focused on customer retention (when the objective is to reduce the transition rates to the inactive customer segment or to increase the transition rate of late customers to the not-yet-late status), customer development (when the objective, for example, is to increase the transition rate from single-product category buyers to multi-product category buyers and of one-time buyers to multi-buyers) and customer recognition (when the objective is to increase the percentage of not yet late, multi-buyers who are also multi-product buyers). In addition to these customer-focused goals, to continuously enhance the value of the customer franchise, the

company must also have very specific customer-acquisition goals and marketing programmes to achieve them. The customer-acquisition goals should be operationalised through specific annual objectives in terms of numbers of new customers to be acquired each year at a maximum new customer acquisition cost. To optimise the long-term profitability of the customer-acquisition programme, we know that (at the margin) the acquisition cost of a new customer should never exceed her lifetime value, or the NPV of all cash flows attributable to that customer.

Tools for designing customer marketing programmes

Two analytic tools are particularly relevant in designing marketing programmes. The first is behavioural customer segmentation and the second is product propensity models. As our focus is on improving the transition rates by designing programmes and communications that are particularly relevant to customers in their current segment, we recommend developing a separate customer typology of micro-segments for the following customer groups:

- one-time buyers;
- multi-, single-product buyers;
- multi-, multi-product buyers.

The micro-segments should provide insights into customers' needs that could be exploited in designing relevant communications and offers. The above groups make sense from a data-density stand point. We would have less purchasing information on one-time buyers than on multi-, single-product buyers and multi-, multi-product buyers. This will allow us to use all information available on customers in each group without assuming a common distribution across customer groups or having to treat customer variables in a special way when certain variables do not apply to a specific customer group.

Behavioural customer segmentation

The objective here is to group customers on the basis of their key characteristics, relevant to marketing, such that customers within a segment are similar to each other and different from customers in other segments. The two major groups of clustering methods are based on whether a specific criterion (or dependent) variable is used to distinguish among the clusters defined by a set of independent variables, or whether no criterion is used. Scoring models, using regression analysis-type models, are an example of the former. Here the clusters are the resulting deciles direct marketers develop in a gains' table. Another example of criterion-based segmentation models is the family of tree-based methods (AID,¹² CHAID¹³ and CART¹⁴) based on the sequential splitting of the sample into groups that are maximally different on the given criterion, based on the values of the independent or basis variables.

The approach, in which no criterion variable is used, is often described as descriptive or traditional clustering. Both the criterion-and

non-criterion-based methods can be further distinguished as *non-overlapping*, *overlapping* and *fuzzy* methods (see Wedel and Kamakura¹⁵ for an excellent discussion). Non-overlapping clustering methods are the most common type of traditional clustering, and may be developed using hierarchical or non-hierarchical methods. Both these approaches are available in SAS through PROC CLUSTER. In fuzzy clustering, customers have partial membership in more than one cluster. The best-known method for identifying these clusters is the latent class model, also known as the mixture model. We discuss below two of the better-known clustering techniques: *K*-Means for continuous variables and the more recent latent class modeling approach for categorical data.

***K*-Means clustering**

The *K*-Means approach to identifying non-overlapping descriptive clusters is quite popular with practitioners, as it can easily handle large data sets. It is not uncommon to develop the clusters on samples of 50,000–100,000 customers. *K*-Means is a non-hierarchical approach, with the researcher specifying an initial number of clusters and the clusters' centroids in terms of its basis variables. The algorithm then starts assigning individual customers to the clusters based on shortest distance. With the assignment of each observation to a cluster, the cluster's centroid is recalculated and the process continues until all customers have been assigned. Final clusters are identified by fine-tuning the procedure by initially eliminating outlier observations and reassigning the remaining observations all over again. To determine the right number of clusters, the entire procedure is repeated a number of times, starting with a different number of specified clusters. The best solution is then selected based on the cluster sizes and a few 'fit' statistics.

The final cluster solution may then be validated on a separate holdout sample, by starting with the cluster centroids from the calibrating sample and examining the final centroids derived from the validation sample. If the two sets of centroids are nearly the same, we should have reasonable confidence in the cluster solution.

Latent class model

In the latent class model, all the basis variables are categorical. This, however, is not a major issue for most practitioners, as any continuous variable could be divided into a few meaningful classes, or categories. The latent class model for clustering customers posits that an unobserved or latent categorical dimension underlies the set of categorical basis variables used to define the clusters, such that given the value of the latent variable, the basis variables are independent of each other. This assumption is often referred to as the assumption of local independence: the basis variables are only correlated because of their common cause: the underlying latent dimension. The latent class model may thus be viewed as the analogue of the factor analysis model for categorical variables.

Vermunt¹⁶ has developed the LEM programme that may be used to identify customer segments using the latent class modeling approach. LEM is a versatile programme for estimating log-linear path models with latent variables. The path models are used to specify log-linear models with latent variables, analogous to the LISERAL approach. Essentially the programme uses a system of simultaneous logit models to estimate the likelihood of a customer belonging to each class of the latent variable, or cluster, based on the values of the observed categorical basis variables. LEM estimates the model parameters through the method of maximum likelihood, and specifically the expectation-maximising, EM, algorithm (see Dempster, Laird and Lubin¹⁷). The logit models' coefficients help us understand the customer characteristics that differentiate the clusters, and as this approach results in the probabilities of customers' membership in each cluster, this method may be classified as a fuzzy clustering method, with customers assigned to specific segments based on their segment membership probabilities. For example, to assign each customer to a single segment, we could assign customers to the segment with the highest membership probability. Alternatively, we could define some segment membership probability cutoff (say, 0.35), and assign customers to multiple segments as long as their segment membership probability exceeded 0.35.

Basis variables for segmentation

One-time buyers have the least transaction information, and as a result only a few basis variables can be constructed, essentially variables that may be derived from their first order. These include the monetary size of the order, the number of product categories purchased and (0/1) indicators for the specific product categories purchased. In addition, depending on the merchandise sold, demographics could also be used, if the customers are consumers. If the market consists of businesses, it may be worthwhile to append such information as business size in terms of revenue and number of employees, as well as business type, using some combination of standard industry classification (SIC) codes to develop a more manageable number of business types. Multi-buyers have considerably more information, and their basis variables need to be constructed using information from all purchase occasions. Typical variables include customer tenure, number of purchases per month, average purchase amount (dollars), average number of product categories bought per purchase occasion for multi-product buyers, specific product categories purchased, etc.

As product categories are represented by indicators, we have found it most convenient to use the latent class modeling approach to segment the customer groups. These segments are used primarily to assist marketers in developing customised messages and imagery in their communications. The specific product categories that should be suggested for purchase in the communications can be determined through product propensity models, and the appropriate size of the incentive (if necessary) for purchase and the contact rate should be

based on the results of separate tests and a detailed economic analysis of test results. These topics are discussed in subsequent sections.

Product propensity models

We recommend building separate product propensity models for the same three customer groups used for the clustering. The product propensity models are essentially a system of logit models that estimate the likelihood of a customer buying specific product categories over some future period of time, say 3 months. If the most recent 3 months is the target window, then we define binary response variables corresponding to the purchase of each product category within the target window. If there are N product categories, then for each customer group there are N response variables (the log odds of buying in each product category) and N logit models.

To semi-automate the modeling process, we recommend predictor variables be the Recency, Frequency, Monetary (RFM) variables associated with each product category. This should result in N sets of RFM variables. If a customer has never purchased in a product category, Recency could be set to an arbitrary constant and handled via the missing indicator approach we described earlier in the section on the purchase-timing model. These models will enable us to estimate N scores for each customer, and the top three or four scores may be used to develop specific product recommendations, for each customer. Customers' top two or three micro-segment memberships (based on an appropriate probability cutoff) and product propensities may then be stored in a separate table that may be used for customer selection and for assembling the appropriate communication and product offer at time of contact.

Testing

The segmentation provides clues to the marketer regarding how they should speak to the customer (tone of voice, specific phrases, etc) and the product propensity scores help determine the specific products that should be offered. The latter information is particularly valuable in designing communications aimed at converting single-product category buyers into multi-product category buyers. The economic viability of different offers with financial implications, such as discounts, however should be tested separately for each segment and analysed prior to rollout to ensure that they make economic sense. Essentially, for each tested offer an expected profit after marketing cost is computed, and the offer with the highest expected value is assigned to the segment. Typically, the no-discount baseline offer is also included to determine whether the discount does in fact result in higher total net profits. The same approach may be used to determine whether multiple sequential contacts, with increasing urgency and increasing offer attractiveness, are needed, particularly to get customers in the most valuable segments that are late in their purchase to buy again.

Once testing is completed and the appropriate contact strategies identified, corresponding customer transition matrices can be developed for different contact strategies. These may then be used to set the marketing objectives, once budgetary approval is obtained from senior management.

Simulating customer contact plans and customer and enterprise value

In order to receive budget approval for the new marketing programmes, it may be worthwhile to present senior management with *pro forma* income statements under a business-as-usual scenario and under the new marketing programme, with differentiated contacts to each segment so as to maximise the chances of achieving segment objectives. The baseline transition matrix, along with historic monthly rates of new customer acquisition is used to develop the business-as-usual customer flows and segment counts over the 12-month budget cycle. Similarly, the transition matrix based on the test results and the customer performance objectives in terms of the desired transition probabilities, along with the new monthly customer acquisition objectives, are used to develop the monthly segment counts under the new marketing by objectives programme. These segment counts are next converted to monthly revenue, cost and cash flow streams, needed for the *pro forma* income statements, by using average segment values per customer.

Once a programme to carry out such a simulation is developed, it could be readily used to simulate the results of different spend levels on customer acquisition, development and retention, provided test results are available to drive these scenarios. The scenario that marketing has the most confidence in and that shows significant improvements in profits should then be presented to senior management in support of budgetary approval.

The above approach can easily be extended to estimating the lifetime value of a new customer, as well as the total value of the business under given assumptions of new customer acquisition.

Estimating lifetime value of a new customer

To estimate the lifetime value of a new customer, we assume that a new customer starts in the 'One-time, not-yet-late, buyer' segment ($i=2$), and we track the probabilistic flow of the customer, over the different segments over time, based on the transition probability matrix, $\mathbf{T}$. This is done by computing $\mathbf{T}^2$, $\mathbf{T}^3$, $\mathbf{T}^4$ etc. Here, $\mathbf{T}^n = \mathbf{T} \times \mathbf{T}^{(n-1)}$ refers to the n month transition probability matrix with elements $t_{ij}^n = \text{Prob}(X_{t+n}=j|X_t=i)$. As the 'inactive' state is an absorbing state with transition probability equal to 1, after a finite number of transitions the customer eventually lands in the inactive state. This implies that there exists an n such that $t_{2j}^n=0$ for all active states j and $t_{2k}^n=1$, for the inactive state k . To compute the lifetime value of a new customer, we need the second row of the n month

transition matrices $\mathbf{T}^n$ for $n=1, 2, 3 \dots$ and the net cash flows or profits associated with each state or segment (note that the cash flow associated with the inactive state is zero). This allows us to estimate the expected cash flow from a customer starting in state two for each of the subsequent n months. The lifetime value of a new customer may then be estimated as the discounted sum of these n expected cash flows.

Estimating the value of the enterprise

The above approach may be extended to estimate the value of the enterprise or business, if we define it as the NPV of cash flows generated from all current and future customers over, say, the next 10 years. The 10 years is arbitrary, but the contribution of distant cash flows of course diminishes over a long period of time. To estimate this value we need the following information:

1. The distribution and number of active customers in each of the segments as of the end of the last month.
2. The number of new customers we would be adding each year and each month. The annual figure could be set proportional to the previous year's total revenue, and this could be distributed over the 12 months based on historical new customer acquisition rates.
3. The acquisition cost of a new customer.
4. The net monthly cash flow of an active customer in each segment.

In Step 1, we estimate the total value of the current active customers. We would first estimate the number of customers in each segment for each of the subsequent 120 months by calculating $\mathbf{T}^n$ for $n=1, 2, 3 \dots, 120$ and applying the initial segment counts to each row of $\mathbf{T}^n$. Knowing the cash flow per customer in each segment, we can then estimate the total segment cash flow. These are summed over the segments to obtain the enterprise level cash flow for each of the 120 months. The discounted sum of these 120 cash flows represents the value of the current active customer base.

In Step 2, we estimate the values of new customers, over 120 months, 119 months (as each month the 10-year horizon gets shorter by 1 month) etc using just the second row ($i=2$; new customers) of $\mathbf{T}^n$. These values are then multiplied by the number of new customers we plan to add each month over the next 10 years. These values are then discounted back to the current time. For example, the value of customers acquired 1 month from now is discounted over 1 month, the value of customers acquired 2 months from now (and computed over 118 months) is discounted over 2 months, etc. These discounted values, generated by the new customers over the next 10 years, are then summed. The latter sum now provides an estimate of the total value generated by the new customers over the next 10 years. Finally, the sum of the values of the current customer base and the new customers provides an estimate of the value of all existing customers and the planned addition of new customers, or a customer

Table 2: One, two and three month transition matrices

	Segment	NYL	L	I
T=	NYL	0.5	0.5	0
	L	0.2	0.4	0.4
	I	0	0	1
T ² =	NYL	0.35	0.45	0.2
	L	0.18	0.26	0.56
	I	0	0	1
T ³ =	NYL	0.265	0.355	0.38
	L	0.142	0.194	0.664
	I	0	0	1

NYL=Not yet late; L=Late; I=Inactives.

value based estimate of the enterprise. We illustrate some of the key calculations through a simple example.

Example

Assume that a company has developed a customer segmentation based only on the 'Time since last purchase' dimension, consisting of 'Not-yet-Late Customers (NYL)', 'Late Customers (L)' and 'Inactives (I)'. Suppose the average monthly cash flow of NYL, L and I customers is \$50.00, \$20.00 and \$0, respectively, and occurs at the end of each month. Further assume that at the end of the previous month the company has 50,000 NYL customers, 70,000 L customers and 200,000 I customers. It also plans to add 5,000 new customers at the start of each of the next 3 months at an acquisition cost of \$100 per new customer. If the company has the following transition matrix, **T**, estimate its total cash flow over the next quarter (Table 2).

The transition matrix **T**² displays the end states at the end of the second month and **T**³ the end states at the end of the third month. For example, a customer starting in state NYL at the end of the previous month has a 20 per cent chance of being inactive after 2 months and 38 per cent after 3 months.

Month-1 cash flow

At the start of month 1, there are 55,000 NYL customers (50,000+5,000 new customers) and 50 per cent of them, or 27,500, are still in this state at the end of month 1. The balance, 27,500, have now migrated to state L at the end of month 1.

At the start of month 1, there are also 70,000 L customers, of whom 20 per cent, or 14,000, make a purchase and hence end up in state NYL at the end of month 1, 40 per cent, or 28,000, remain in state L and the balance, 28,000, become inactive.

At the end of month 1, therefore, we have 27,500+14,000, or 41,500 customers in segment NYL, providing a cash flow of 41,500 × 50 or \$2,075,000 and 27,500+28,000, or 55,500 customers in segment L, creating a cash flow of 55,500 × 20, or \$1,110,000. The total cash flow

in month 1, therefore, is \$3,185,000 less \$500,000, the acquisition cost of new customers, or \$2,685,000.

Month-2 cash flow

The 55,000 NYL customers at the start of month 1 have migrated at the end of month 2 as follows: $55,000 \times 0.35$, or 19,250, are still NYL customers and $55,000 \times 0.45$, or 24,750, are L customers, and the balance, 11,000 customers, are now inactive.

The 70,000 L customers at the start of month 1 have now migrated at the end of month 2, as follows: $70,000 \times 0.18$, or 12,600, into NYL customers and $70,000 \times 0.26$, or 18,200 customers, remain in segment L, and the balance, 39,200, have become inactive.

In addition, the 5,000 new customers starting in segment NYL at the start of month 2 have now migrated by the end of month 2, as follows: $5,000 \times 0.5$, or 2,500, remain as NYL, and 2,500 are now L customers.

Hence, at the end of month 2, we have $19,250 + 12,600 + 2,500$, or 34,350 NYL customers, with a cash flow in month 2 of $34,350 \times 50$ or \$1,717,500, and $24,750 + 18,200 + 2,500$, or 45,450 L customers, contributing cash flow in month 2 of \$909,000. The total cash flow in month 2 therefore equals \$2,626,500 less \$500,000 (new customers) or \$2,126,500.

Month-3 cash flow

The 55,000 customers starting month 1 as NYL customers will end up at the end of month 3 as follows: $55,000 \times 0.265$, or 14,575, remain as NYL customers, and $55,000 \times 0.355$, or 19,525, are L customers (the balance are inactive).

The 70,000 customers starting month 1 as L customers will end up at the end of month 3 as follows: $70,000 \times 0.142$, or 9,940, as NYL customers, and $70,000 \times 0.194$, or 13,580, will remain as L customers. (the balance are inactive).

The 5,000 new customers starting as NYL customers at the start of month 2 will end up as $5,000 \times 0.35$ or 1,750, NYL customers, and $5,000 \times 0.45$, or 2,250 as L customers, at the end of month 2 (the balance are inactive).

The 5,000 new customers starting as NYL customers at the start of month 3 will end up as 2,500 NYL customers, and 2,500 L customers at the end of month 3.

At the end of month 3 we will therefore have $14,575 + 9,940 + 1,750 + 2,500$ or 28,765 NYL customers, with a cash flow of \$1,438,250 and $19,525 + 13,580 + 2,250 + 2,500$, or 37,855 L customers, with a cash flow of \$757,100, resulting in a total cash flow of \$2,195,350 less \$500,000, or \$1,695,350, in month 3.

Note that this business as described is non-sustainable; the cash flow in each month continues to decline. This is because customers are becoming inactive at a very high rate relative to the addition of new customers. Fundamental changes need to occur, namely the company needs to significantly slow down the rate at which customers are

becoming late and inactive. It may also need to increase the acquisition rate of new customers and also lower the acquisition cost per new customer. In this example, we included the ongoing marketing costs in the cash flows associated with each state. If these were broken out separately, and if we knew the relationship between the marketing expenditures and the transition probabilities, we could simulate the impact of different marketing spend levels on the company's cash flows. Such relationships can best be estimated through carefully designed tests.

Executing customer selection and marketing contacts

The above essentially lays the stage for one-to-one communications, which are most easily executed in an e-marketing environment. Based on a customer's micro-segment membership, product propensities and optimal offer, a customised communication can be assembled. Appropriate salutations, text, digitised product photos and other graphics can be assembled from databases in real time based on pre-loaded customer selection and contact rules prior to executing a batch mailing. This is not trivial, and requires setting up libraries of appropriate text and digitised product photos, as well as tables that assign customers to micro-segments and that identify the product categories customers are most likely to buy. As many products fall within a product category, the most popular product pictures are typically selected for the customers' e-mails. The customer selection and e-mail construction software should have the flexibility for marketers to select specific pictures and text for each customer contact based on their micro-segment membership and optimal product propensities.

Tracking customer performance

Once the segment performance objectives are set, the micro-segments and the product propensity models developed and the optimal offers identified, the execution of the marketing programmes begins. Assuming that there are ongoing customer contacts each month, the monthly transition matrices need to be computed, and the observed transition rates and segment profits must be calculated and compared to the objectives. We recommend tracking mean segment profits in addition to the transition rates so that we can ensure that if we succeed in increasing the purchase rate, we have not also reduced the average monthly segment profits. If for two successive months marketers fail to achieve their objectives, we may need to test alternative marketing programmes or change our customer performance objectives. If there is an industry-wide cyclical economic downturn that we had not anticipated when setting the customer performance objectives, it might warrant changing the objectives. The latter situation would of course have to be discussed with senior management to get their buy-in.

Monthly customer value scorecards can be developed that show the baseline and observed transition rates, segment counts, mean segment marketing costs, revenues and profits, as well as total profits for each customer segment and the total customer base. The change in customer counts is due to the addition of new customers and the inactivation of a certain number of customers each month. In addition, the number of new customers added each month and the new customer acquisition costs should be tracked against the marketing objectives and budget. It may also be worthwhile to display the monthly trends in these figures and report cumulative year-to-date values to show how customer performance is tracking year-to-date against objectives. All this information can be developed as an application with a graphical user interface with different levels of accessibility for different levels of management. Invariably, there will be a variety of design issues that need to be addressed in a customised application.

Summary

In this paper, we described a methodology for managing the marketing function for a retailer, focused on increasing the value of its customer base. At the heart of our approach are some simple propositions, namely that customer value can be increased by targeting

- new customers and one-time buyers to buy again before they are ‘late’ in their next purchase;
- single-product category buyers to purchase products in additional categories, before becoming late;
- late customers with relevant messages and high-value offers to convince them to purchase again.

Our approach is highly data-intensive and requires ongoing monitoring of the customer base and the monthly dynamics in customer segment transitions. Such attention to detail, however, enables marketers to take timely action before customers are lost or become inactive. To ensure that the company is progressing on its goals for enhancing customer value, we recommended that specific objectives be set in improving the transition rates to higher-value segments and reducing the rate of transition to lower-value segments.

Finally, we discussed tools that marketers should deploy in order to sharpen their communications and make product offers that are both meaningful and relevant, namely customer micro-segmentation and product purchase propensity models. We also recommended that testing be used extensively to identify offers that are economically justifiable in improving the transition rates. Finally, we described how the segment transition matrix used in our approach could also be leveraged to estimate the lifetime value of a new customer as well as the value of the enterprise.

References

1. Hansotia, B. (2006) 'Issues and challenges in estimating customer equity', *Journal of Direct, Data and Digital Marketing Practice*, Vol. 7, No. 4, pp. 293–317.
2. Hansotia, B. (2004) 'Company activities for managing customer equity', *Journal of Database Marketing and Customer Strategy Management*, Vol. 11, No. 4, pp. 319–332.
3. Hansotia, B. (2004) 'Customer metrics and organizational alignment for maximizing customer equity', *Journal of Database Marketing and Customer Strategy Management*, Vol. 12, No. 1, pp. 9–20.
4. Hansotia, B. (2002–2003) 'The customer value scorecard', *Journal of Integrated Communications*, pp. 34–39, Northwestern University, Department of Integrated Marketing Communications.
5. Cox, D. R. and Oakes, D. (1984) *Analysis of Survival Data*, Chapman and Hall, London, UK.
6. Cox, D. R. and Oakes, D., see ref. 5 above.
7. Kalbfleisch, J. D. and Prentice, R. L. (1980) *The Statistical Analysis of Failure Time Data*, Wiley, New York, NY.
8. Allison, P. D. (1995) *Survival Analysis Using the SAS System: A Practical Guide*, SAS Institute, Cary, NC.
9. Cox, D. R. and Oakes, D., see ref. 5 above.
10. Nelson, C. R. (1973) *Applied Time Series Analysis for Managerial Forecasting*, Holden-Day, San Francisco, CA.
11. Wheelwright, S. C. and Makridakis, S. (1973) *Forecasting Methods for Management*, Wiley, New York, NY.
12. Assael, H. (1970) 'Segmenting markets by group purchasing behavior: An application of the aid technique', *Journal of Marketing Research*, Vol. 7, pp. 153–158.
13. Kass, G. (1980) 'An exploratory technique for investigating large quantities of categorical data', *Applied Statistics*, Vol. 29, pp. 119–127.
14. Breiman, L., Friedman, J. H., Olsen, R. A. and Stone, C. J. (1984) *Classification and Regression Trees*, Wadsworth International Group, Belmont, CA.
15. Wedel, M. and Kamakura, W. (2001) *Market Segmentation Conceptual and Methodological Foundations*, Second Edition, Kulwer Academic Publishers, Norwell, MA.
16. Vermunt, J. K. (1997) *Log-Linear Models for Event Histories*, Sage Publications, Inc., Thousand Oaks, CA.
17. Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977) 'Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion)', *Journal of the Royal Statistical Society, Ser. B*, Vol. 39, pp. 1–38.