Available online at www.sciencedirect.com
ScienceDirect
Journal of Interactive Marketing 23 (2009) 61 – 69
JOURNAL OF INTERACTIVE MARKETING
www.elsevier.com/locate/intmar

Probability Models for Customer-Base Analysis

Peter S. Fader[a,*] & Bruce G.S. Hardie[b]

[a] 749 Huntsman Hall, Wharton School of the University of Pennsylvania, 3730 Walnut Street, Philadelphia, PA 19104-6340, USA
[b] London Business School, UK

Abstract

As more firms begin to collect (and seek value from) richer customer-level datasets, a focus on the emerging concept of *customer-base analysis* is becoming increasingly common and critical. Such analyses include forward-looking projections ranging from aggregate-level sales trajectories to individual-level conditional expectations (which, in turn, can be used to derive estimates of customer lifetime value). We provide an overview of a class of parsimonious models (called *probability models*) that are well-suited to meet these rising challenges. We first present a taxonomy that captures some of the key distinctions across different kinds of business settings and customer relationships, and identify some of the unique modeling and measurement issues that arise across them. We then provide deeper coverage of these modeling issues, first for *noncontractual* settings (i.e., situations in which customer "death" is unobservable), then *contractual* ones (i.e., situations in which customer "death" can be observed). We review recent literature in these areas, highlighting substantive insights that arise from the research as well as the methods used to capture them. We focus on practical applications that use appropriately chosen data summaries (such as *recency* and *frequency*) and rely on commonly available software packages (such as Microsoft Excel).

Keywords: Customer-base analysis; Customer lifetime value; CLV; Probability model; RFM

Introduction

In recent years, improvements in information technology have resulted in the increased availability of customer transaction data. This trend is closely linked to an ever-growing desire on the part of the marketing manager to use the firm's customer transaction databases to learn as much as possible about his customer base. These changes, in turn, have led to an overall shift in perspectives about how these bigger, better datasets will be used; specifically, many experts now talk about an evolution from *transaction-oriented to customer-centric* marketing strategies.

Initial analyses of these transaction databases are typically descriptive in nature. This includes basic summary statistics (e.g., average number of orders, average order size) as well as information on the distribution of behaviors across the customer base (e.g., the distribution of total spend per customer) from which concentration measures can be derived (e.g., determining the percentage of total sales accounted for by the top 20% of customers). Further analyses of the customer base may use various multivariate statistical methods and data mining tools to identify, say, the geodemographic characteristics of heavy buyers, or to determine which groups of products tend to be purchased together (i.e., a "market-basket" analysis). Whether we are talking about analyses performed in a spreadsheet or using a complicated data mining suite, the one thing they have in common is that they are descriptive in nature and focus on the past behavior of the firm's customer base.

The next step is to undertake customer-base analysis activities that are more forward-looking (or predictive) in nature. Given a customer transaction database, we are interested in making forecasts about the future purchasing by the firm's customers. These projections can range from aggregate sales trajectories (e.g., the total purchasing by the existing customer base for the next 52 weeks) to individual-level conditional expectations (i.e., the best guess about a particular customer's

* Corresponding author.
E-mail addresses: faderp@wharton.upenn.edu (P.S. Fader), bhardie@london.edu (B.G.S. Hardie).
URLs: http://www.petefader.com (P.S. Fader), http://www.brucehardie.com (B.G.S. Hardie).

1094-9968/$ - see front matter
doi:10.1016/j.intmar.2008.11.003

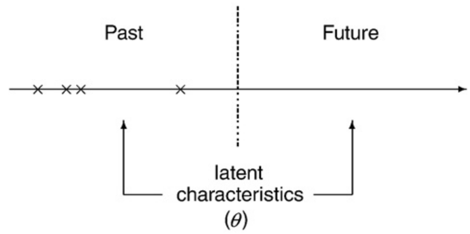

Fig. 1. A probability modeler's view of the transaction process.

future purchasing, given information about his past behavior). We feel that the ability to make such individual-level predictions lies at the heart of any serious attempt to compute the (residual) customer lifetime value of a firm's (existing) customers on a systematic basis. (See Blattberg, Malthouse, and Neslin (accepted for publication); Jain and Singh (2002); Kumar, Ramani, and Bohling (2004); Kumar et al. (accepted for publication); and Malthouse and Blattberg (2005) for other perspectives on the issue of computing customer lifetime value (CLV).)

In this paper we provide an overview of a class of parsimonious models (called probability models) that are being used to meet these rising challenges. While many of these models were first developed by marketing scientists years ago — often in different contexts for different analytical purposes — their usefulness for customer-base analysis is now becoming quite clear. In reviewing these models, we highlight some of the key differences (and similarities) across different kinds of business settings and customer relationships, and discuss some of the unique modeling and measurement issues that arise across them.

The probability modeling approach

Starting with Ehrenberg (1959), there has been a strong tradition of probability models in marketing. (Also see classic texts such as Ehrenberg (1988), Greene (1982) and Massy, Montgomery, and Morrison (1970). A probability modeler approaches the modeling problem with the mindset that observed behavior is the outcome of an underlying stochastic process. That is, we only have a "foggy window" as we attempt to see our customers' true behavioral tendencies, and therefore the past is not a perfect mirror for the future. For instance, if a customer made two purchases last year, is he necessarily a "two per year" buyer, or is there some chance that he might make three or four or perhaps even zero purchases next year? With this kind of uncertainty in mind, we wish to focus more on the latent process that drives these observable numbers, rather than the observables themselves.

As illustrated in Fig. 1, the transactions associated with a customer — those observed in the past and those from the as-yet-to-be-observed future — are a function of that customer's underlying behavioral characteristics (denoted by θ). That is, past$=f(\theta)$ and future$=f(\theta)$.

The starting point is to specify a mathematical model in which the observed behavior is a function of an individual's latent behavioral characteristics (i.e., past$=f(\theta)$). This is done by reflecting on what simple probability distribution (e.g., Poisson, binomial, exponential) can be used to characterize the observed behavior. (In many cases, including those to be discussed in this paper, observed behavior may be characterized using a combination of these basic probability distributions.) By definition, we do not observe an individual's latent characteristics (θ). Therefore, the next step is to make an assumption as to how these characteristics vary across the customer base by specifying a *mixing distribution* that captures the cross-sectional heterogeneity in θ. (The choice of distribution(s) is typically driven by the dual criteria of flexibility and mathematical convenience.) Combining this with the distribution for individual-level behavior gives us a *mixture model*, which characterizes the behavior of a randomly chosen customer.[1]

After fitting the mixture model to the data, a straight-forward application of Bayes' theorem enables us to make inferences about an individual's latent characteristics (θ) given his observed behavior. We can then make predictions regarding future behavior as a function of the inferred latent characteristics. Note that there is no attempt to explain the variation in θ as a function of covariates; we are, in most cases, content to capture the variation using probability distributions alone.

This two-step approach ($\theta=f$(past) and future$=f(\theta)$ can be contrasted with the single-step approach (future$=f$(past) associated with the use of regression models (and more sophisticated data mining procedures). Fader, Hardie, and Lee (2006) suggest that there are several advantages associated with the use of a formal probability model. First, there is no need to split the observed transaction data into two periods to create a dependent variable; we can use all of the data to make inferences about the customers' behavioral characteristics. Second, we can predict behavior over future time periods of any length; we can even derive an explicit expression for CLV over an infinite horizon (with discounting to acknowledge the lower present value of purchases that occur in the distant future).

While there is little that is truly new about these concepts, per se, their use in customer-base analysis exercises has been relatively limited — particularly when compared to the usage of regression-like methods. (For example, they receive no coverage in basic trade books such as Berry and Linoff (2004) and Parr Rud (2001).) As firms become more serious about becoming customer-centric, it is imperative for them to use the right methods — and to use them the right way, based on the different kinds of business settings in which they operate. We now explore some of these differences in more detail.

Classifying analysis settings

Before moving ahead to our review of specific probability models, we need to classify different kinds of firm–customer relationships, which will drive the choice of the probability

[1] This approach should also not be confused with work, such as Pfeifer and Carraway (2000), that uses Markov chains to characterize behavior. Such work does not account for heterogeneity in the underlying behavior characteristics, which can lead to misleading interferences about the nature of buying behavior (Massy et al. 1970).

distributions we use, as well as other tangible implementation details. Consider, as a starting point, the following two statements regarding the size of a company's customer base:

- Based on numbers presented in a January 2008 press release that reported Vodafone Group Plc's third quarter key performance indicators, we see that Vodafone UK has 7.3 million "pay monthly" customers.
- In his "Q4 2007 Financial Results Conference Call," the CFO of Amazon made the comment that "[a]ctive customer accounts [representing customers who ordered in the past year] exceeded 76 million, up 19%."

While both statements might seem perfectly logical at face value, more careful consideration suggests that only one of them (the first one) is a valid description of the true size of the firm's customer base. We can be reasonably confident about the number of "pay monthly" customers that Vodafone UK has at any single point in time. As these are contract customers (as opposed to customers on "pay as you talk"/pre-paid tariffs), the customer must tell Vodafone when he wishes to switch his mobile phone provider. Therefore, Vodafone knows for sure the time at which any customer formally becomes "inactive."

In contrast, the CFO of Amazon talks of "active customers" being those who have placed an order in the past year. Can we assume that someone is no longer a customer just because their last order was placed 366 days ago? Furthermore, can we necessarily assume that, just because a customer placed an order 364 days ago, he is still an active customer? We should not make such assumptions; this 12-month cut-off point is arbitrary. If we were to change the cut-off point to nine months, the apparent size of Amazon's customer base would be smaller, even though the true size would remain unchanged. This is an example of what Reinartz and Kumar (2000) call a noncontractual setting, a key characteristic of which is that the time at which a customer becomes inactive is unobserved by the firm; customers do not notify the firm "when they stop being a customer. Instead they just silently attrite" (Mason, 2003, p. 55). As such, any firm that has a noncontractual relationship with its customers can never know for such how many customers it has at any point in time.

This contractual/noncontractual distinction as to the type of relationship the firm has with its customers is of fundamental importance for any developer of models for customer-base analysis. The key challenge of noncontractual settings is how to differentiate those customers who have ended their relationship with the firm (without informing the firm) from those who are simply in the midst of a long hiatus between transactions. (We can never know for sure which of these two states a customer is in; however, we will be able to make probabilistic statements, as we will discuss shortly.)

To further refine our general classification of business settings that managers may encounter, consider the following four specific business settings: airport lounges (e.g., United's "Red Carpet Club"), electrical utilities, academic conferences, and mail-order clothing companies. The first two are clearly contractual settings since, in both cases, the time at which a customer becomes inactive is observed by the firm. (In the case of the airport lounge, the "notification" occurs by default when the customer fails to renew her membership.) The last two are clearly noncontractual settings; in the case of academic conferences, there is no reason for a past attendee to contact the conference organizers to inform them of the fact that he has no intention of attending any more conferences.

Reflecting on the first and third settings, we see that the transactions can only occur at discrete points in time. The conference occurs at a specific point in time, and one either attends or does not attend; if the conference is scheduled for June 20–23, one cannot attend it on May 30! Similarly, the airline lounge membership lapses at a specific point in time and the member either renews or does not renew. On the other hand, one characteristic shared by the second and fourth settings is that there are no constraints as to when the customer can purchase clothing or end her relationship with the firm. Thus we can talk of a second distinction: are the opportunities for transactions restricted to discrete points in time or can they occur at any point in time?

These two dimensions lead to a classification of customer bases, adapted from Schmittlein, Morrison, and Colombo (1987), as illustrated in Fig. 2.

When developing any model for customer-base analysis, we must ask ourselves for which quadrant it is designed. While in certain circumstances a model developed for a discrete setting can be applied in a continuous setting (and vice versa), the contractual/noncontractual divide is fundamental and the boundary cannot be crossed: it is completely inappropriate to apply a model developed for a contractual setting in a noncontractual setting (and vice versa). For the remainder of the paper, we use this framework to guide our review of the literature.

Noncontractual settings

We begin in the upper-left quadrant of Fig. 2, the setting that has received the most attention by those developing and using probability models for customer-base analysis.

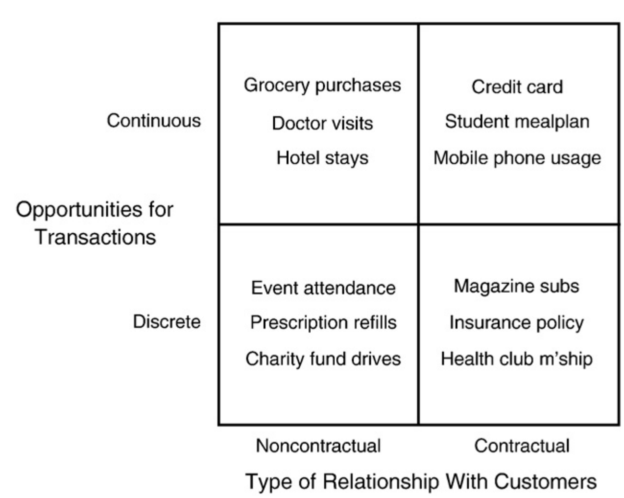

Fig. 2. Classifying customer bases.

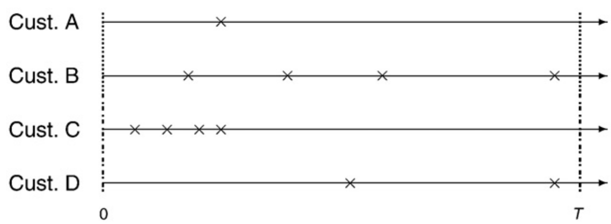

Fig. 3. Illustrative transaction histories.

To understand the objectives of these modeling exercises, let us consider the transaction histories for four customers as illustrated in Fig. 3 (where the occurrence of a transaction is denoted by ×).

As this is a noncontractual setting, we do not know whether or not any given customer is still "alive" at time T. Reflecting on these transaction patterns:

- While customers B and C have both made four purchases in the interval (0, T], the fact that a long time has passed since C's last transaction means that it is very likely that he is actually no longer an active customer. As a result, we would expect more purchases from B in the future.
- While customer A's last purchase occurred at the same time as that of customer C, the long hiatus could simply be a function of the fact she appears to be a light buyer. As a result, A is more likely to be "alive" at T than C. (Whether or not we can expect more purchases from A in the future is an open question: while C is less likely to be "alive," her purchase rate, conditional on being "alive," is likely to be higher than that of A.)
- Customers' B and D last purchases occurred at the same time, so they probably have an equally (high) likelihood of being active at time T. However we would expect fewer purchases in the future from D given his smaller number of observed purchases in the period (0, T].

The primary objectives of the probability models developed for this quadrant have been i) to predict which customers are most likely to be "alive" at T (given their individual purchase histories), and ii) to make predictions as to the amount of business we could expect from each of them in the future.

The standard probability model used to model repeat-buying behavior is the NBD (Ehrenberg 1959; Morrison and Schmittlein 1988), under which i) a customer is assumed to purchase "randomly" around his or her (time-invariant, or stationary) mean transaction rate (characterized by the Poisson distribution) and ii) customers are assumed to differ in their transaction rates (characterized by the gamma distribution). However, if we track the purchasing by a cohort of customers over time, we often observe a pattern of cumulative sales as illustrated in Fig. 4. Since a steady aggregate buying rate would be reflected by a straight line, we are clearly observing a form of nonstationarity (specifically a "slowing down") in the aggregate buying rate.

There are many processes that could give rise to such a pattern of aggregate buying behavior. In their seminal paper, Schmittlein, Morrison, and Colombo (1987) propose a "buy till

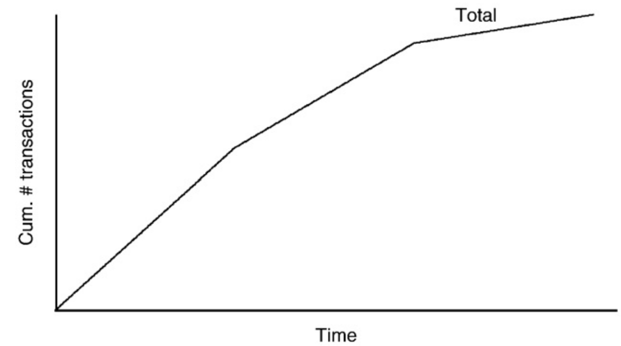

Fig. 4. Typical cumulative sales pattern for a cohort of customers in a noncontractual, continuous-time setting.

you die" explanation as illustrated in Fig. 5. Ignoring the effect of random purchasing around their means, individual customers purchase the product at steady but different underlying rates, as evidenced by the different sloped lines. At different points in time they become inactive or "die." (What lies behind this death? It could be a change in customer tastes, financial circumstances, and/or geographical location, the outcome of bad customer service experiences, or even physical death, to name but a few possible causes. But given the modeling objectives, why this death occurs is of little interest to us; our primary goal is to ensure that the phenomenon is captured by the model.)

More formally, it is assumed that a customer's relationship with the firm has two phases: he is "alive" for an unobserved period of time, then becomes permanently inactive. While "alive," the customer's purchasing is characterized by the NBD model. The customer's unobserved "lifetime" (after which he is viewed as being permanently inactive) is treated as if random, characterized by the exponential distribution, and heterogeneity in underlying dropout rates across customers is characterized by the gamma distribution. Noting that a gamma mixture of exponentials is also known as the Pareto (of the second kind) distribution, the resulting model of buyer behavior is called the Pareto/NBD.

It turns out that, given the model assumptions, we do not require information on when each of the x transactions occurred (as illustrated in Fig. 3). The only customer-level information required to estimate the four model parameters and then make

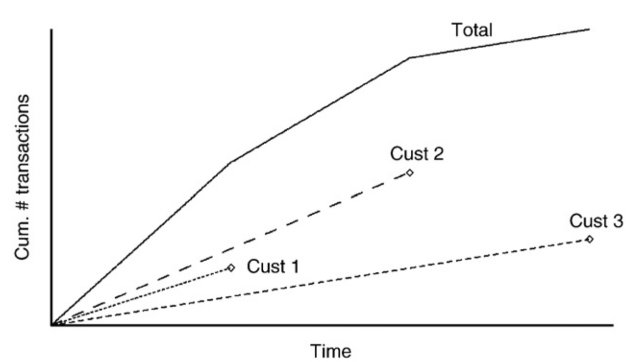

Fig. 5. Individual-level and aggregate view of the "buy till you die" purchasing scenario.

individual-level predictions are *recency* and *frequency*. This is a very important result that ties this model to the traditional direct marketing literature (e.g., David Shepard Associates 1999; Hughes 2006), and makes the data requirements for implementation relatively simple (Fader and Hardie 2005).

The notation used to represent this recency and frequency information is (x, t_x, T), where x is the number of transactions observed in the time interval $(0, T]$ and t_x $(0 < t_x \leq T)$ is the time of the last transaction. Schmittlein, Morrison, and Colombo (1987) derive expressions for (amongst other things) $P(\text{alive} \mid x, t_x, T)$, the probability that an individual with observed behavior (x, t_x, T) is still "alive" at time T, and $E[X(T, T+t) \mid x, t_x, T]$, the expected number of transactions in the future period $(T, T+t]$ for an individual with observed behavior (x, t_x, T).

Empirical validations of the model are presented in Schmittlein and Peterson (1994) and Fader, Hardie, and Lee (2005b), amongst others; its predictive performance is impressive. Recent applications of this model include the work of Reinartz and Kumar (2000, 2003) on customer profitability, Hopmann and Thede (2005) on "churn" prediction, and Wübben and Wangenheim (2008) and Zitzlsperger, Robbert, and Roth (2007) on managerial heuristics for customer-base analysis.

The basic Pareto/NBD model has been extended in a number of directions.

- Ma and Liu (2007) and Abe (2008) explore the use of MCMC methods for parameter estimation.
- Abe (2008) and Fader and Hardie (2007b) allow for the incorporation of time-invariant covariate effects.
- Schmittlein and Peterson (1994) and Fader, Hardie, and Lee (2005b) augment this model for the flow of transactions with a submodel for "monetary value" (i.e., average spend per transaction). In both cases, a customer's underlying spend per transaction is assumed to be independent of his transaction flow. Glady, Baesens, and Croux (2008) propose an extension in which this assumption of independence is relaxed.
- Fader, Hardie, and Lee (2005b) derive an expression for what they call "discounted expected residual transactions" which, when combined with the submodel for spend per transaction, allows us to estimate a customer's expected residual lifetime value conditional on his observed behavior. One of the key contributions of this work is that we only need to know three things about a customer's buying behavior in a given time period in order to compute their residual lifetime value: recency, frequency, and monetary value (i.e., RFM).
- In some situations we do not have access to recency and frequency data; for example, we may only have a series of cross-sectional summaries, such as those reported in the Tuscan Lifestyles case (Mason 2003). By only relying on data in histogram form, any individual-level information about each customer (e.g., recency and frequency) is lost, which may lead some to think that we cannot apply the Pareto/NBD model. Fader, Hardie, and Jerath (2007) show how the model parameters can still be estimated using such "repeated cross-sectional summary" data, despite this limitation.
- Under the Pareto/NBD model, the customer's unobserved "death" can occur at any point in time. Jerath, Fader, and Hardie (2007) present a variant in which it can only occur at discrete points in calendar time, which they call the *periodic death opportunity* (PDO) model. When the time period after which each customer makes his or her dropout decision (which is called periodicity) is very small, the PDO model converges to the Pareto/NBD. When the periodicity is longer than the calibration period, the dropout process is "shut off" and the PDO model converges to the basic (i.e., no death) NBD model.

Despite being published in 1987, the Pareto/NBD has seen relatively limited "real-world action," the major problem being perceived challenges with respect to parameter estimation. To address this problem, Fader, Hardie, and Lee (2005a) develop a variant of the Pareto/NBD model that they call the beta-geometric/NBD (BG/NBD). Changing the "death" story to one where a customer can become inactive after any transaction with probability p and where heterogeneity in dropout probabilities across customers is captured by a beta distribution (i.e., a beta-geometric death process) results is a model that is vastly easier to implement; for instance, its parameters can be obtained quite easily in Microsoft Excel. As the two models yield very similar results in a wide variety of purchasing environments, the BG/NBD can be viewed as an attractive alternative to the Pareto/NBD in most applications. The BG/NBD model has itself been modified and extended by several researchers (Batislam, Denizel, and Filiztekin, 2007; Batislam, Denizel, and Filiztekin, 2008; Fader and Hardie 2007b; Hoppe and Wagner 2007; Wagner and Hoppe 2008).

The "buy till you die" framework that is common to all these models is not the only way to capture the slowing down in aggregate purchasing illustrated in Fig. 4. While there are many other ways of capturing nonstationarity in buying rates, such as those used by Moe and Fader (2004) and Fader, Hardie, and Huang (2004) in different settings, they are typically more difficult to implement, generally requiring the full transaction history (i.e., cannot be estimated using only recency and frequency data). Furthermore, no one has yet derived expressions for quantities such as P(alive) and conditional expectations, which are central to any forward-looking customer-base analysis exercise.

Other researchers have sought to relax what they view as the overly restrictive assumption of exponentially distributed interpurchase times that corresponds to the Pareto/NBD's assumption of individual-level Poisson purchasing using, for example, an Erlang distribution (Wu and Chen 2000) or a generalized gamma distribution (Allenby, Leone, and Jen, 1999). Once again, such extensions are more difficult to implement, typically requiring the full transaction history, and there are no expressions for quantities such as P(alive) and conditional expectations for these alternative models.

All these models are for noncontractual settings where the transaction can occur at any point in time. As noted in Section 3, such an assumption does not always hold (e.g., an annual conference can only be attended at a discrete point in time). In other situations, the transaction can perhaps occur at any point in time but is treated as discrete by management. For example, a charity may record the behavior of each member of its supporter base in terms of whether or not they responded to the year *x* fund drive, even though the check could be received at any point in calendar time. Even for business settings that are truly noncontractual/continuous-time in nature, management may wish to discretize them for ease of summarization or data storage; this is particularly appropriate for very rare events. See, for example, Berger, Weinberg, and Hanna's (2003) characterization of the "repeat cruising" behavior of customers of a cruise ship company in terms of whether or not they make a repeat cruise in each of the four years following the year of their first-ever cruise with the company.

In these discrete-/discretized-time settings, a customer's transaction history can be expressed as a binary string, where 1 indicates that a transaction took place at the discrete point in time (or during the specified time interval), 0 otherwise. (See, for example, the summary of "repeat cruising" behavior from Berger, Weinberg, and Hanna's (2003) as reported in Fig. 6.) As in the case of the continuous-time setting, the challenge facing the modeler is determining whether a sequence of 0's reflects a "dead" customer or simply one that is in the middle of a long hiatus since the last transaction.

The natural starting point for the modeling of such data is to assume a Bernoulli purchasing process (as opposed to the Poisson purchasing process associated with most of the models for continuous-time settings discussed above). One such example is presented in Colombo and Jiang (1999). Another example is given in Fader, Hardie, and Berger (2004b): using the "buy till you die" framework, they assume Bernoulli purchasing with beta heterogeneity while "alive," with another Bernoulli death process (which leads to geometrically distributed lifetimes) with beta heterogeneity. The resulting model is called the BG/BB model. It is interesting to note that as the length of the discrete time period tends to zero, the beta-Bernoulli purchasing process tends to an NBD, while the beta-geometric time-to-death distribution tends to a Pareto (of the second kind) distribution. In order words, the BG/BB is a discrete-time analog of the Pareto/NBD model that tends to the Pareto/NBD when the length of the discrete time period tends to zero.

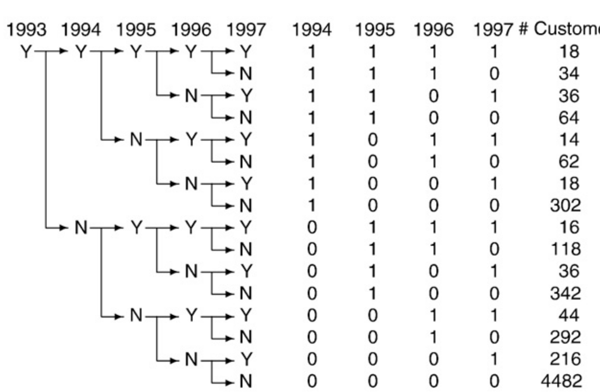

Fig. 6. Summarizing repeat purchasing for luxury cruises.

As with the Pareto/NBD and BG/NBD models, it turns out that the BG/BB model does not require information on when each of the *x* transactions occurred (i.e., the complete binary string as illustrated in Fig. 6); the only customer-level information required to estimate the four model parameters and then make individual-level predictions are *recency* and *frequency*. Fader, Hardie, and Huang (2004a); Fader, Hardie, and Berger (2004) derive expressions for (amongst other things) *P*(alive), the expected number of purchases in a future time period, and the "discounted expected residual transactions," all conditional on the customer's recency and frequency data.

The Bernoulli purchasing process assumes that the probability of making a purchase in one period is independent of whether or not a purchase was made in the preceding period. This assumption of a zero-order purchasing process at the individual level can be relaxed by assuming some type of first-order Markov process. An example of this is the Markov chain model of retail customer behavior at Merrill Lynch of Morrison et al. (1982) in which the "Brand Loyal Model" of Massy, Montgomery, and Morrison (1970) was extended by adding a (homogeneous) "exit parameter."

In contrast to the case of noncontractual, continuous-time settings, this setting (i.e., the lower-left quadrant of Fig. 2) has received little attention by model developers, and there are few published applications of these models. We find this surprising and feel that it is an area of research that deserves more attention.

Contractual settings

As previously noted, the defining characteristic of a contractual setting is that customer attrition is observed by the firm. The questions of managerial interest are typically either i) "which customers have the greatest risk of churning next period?" or ii) "how much longer can we expect this customer to stay with us, given information about his relationship with us to date?", the answer to which lies at the heart of any attempt to compute CLV in a contractual setting.

The first question is best answered using traditional regression-type models and other predictive data mining tools; see Berry and Linoff (2004); Blattberg, Kim, and Neslin (2008); Neslin et al. (2006); and Parr Rud (2001). A typical analysis exercise may use a logit model to predict churn where the independent variables include behavioral variables such as usage in the previous period, changes in usage over the last two periods, customer-initiated contacts with the company (e.g., contacting the call center), and marketing activity by the firm.

Such models are of limited value when faced with the second question, however. Standing at the end of period *t*, it would not be possible to predict contract renewal probabilities for period $t+3$ since we do not have the values of the independent variables for period $t+2$, let alone period $t+1$. Such a question is best addressed using duration-time models, which can better accommodate these longitudinal issues.

In this setting, the modeler is not faced by the challenge of trying to differentiate between a customer who has ended his or her relationship with the firm and one who is merely in the midst of a long hiatus between transactions. As such, there is a larger stock of existing probability models that the analyst can draw to model contract duration.[2] Most of these existing models are for settings where the event of interest (i.e., customer attrition) can occur at any point in time. When operating in discrete-time contractual settings (i.e., the lower-right quadrant of Fig. 2), there is a smaller stock of existing models.

The general model for discrete-time duration data is the shifted-beta-geometric (sBG) model (Kaplan 1982, Weinberg and Gladen 1986). In a customer-base analysis setting, it is formulated by assuming that i) at the end of each contract period, an individual remains a customer of the firm with constant retention probability $1-\theta$ (which is equivalent to assuming that the duration of the customer's relationship with the firm is characterized by the (shifted) geometric distribution), and ii) individual differences in θ are characterized by the beta distribution. Despite what may seem to be overly simplistic assumptions, the analyses presented in Fader and Hardie (2007a) demonstrate that the model generates very accurate forecasts of retention.

For any given cohort of customers, we almost always observe increasing retention rates over time (something that is ignored in most introductory discussions of CLV; e.g., Berger and Nasr 1998; Dwyer 1989). For example, "renewal rates at regional magazines vary; generally 30% of subscribers renew at the end of their original subscription, but that figure jumps to 50% for second-time renewals and all the way to 75% for longtime readers" (Fielding 2005, p. 9). This is often expressed as a decrease in churn propensity as a customer's tenure increases (e.g., Reichheld 1996; Hughes 2006). Many practitioners would explain this in terms of individual-level time dynamics (e.g., increasing loyalty as the customer gains more experience with the firm). We note that the aggregate retention rate associated with the sBG model is an increasing function of time, even though the individual-level retention probability is constant. In other words, the observed aggregate retention rate dynamics can be viewed as the result of a sorting effect in a heterogeneous population where individual customers exhibit no retention-rate dynamics.

Fader and Hardie (2009) derive an expression for what they call "discounted expected residual lifetime" which, when combined with an estimate of expected net cashflow per period, allows us to estimate a customer's expected residual lifetime value conditional on his length of tenure as a customer. They then use this to explore the task of computing the expected residual value of a customer base and retention elasticities (cf. Pfeifer and Farris 2004; Gupta accepted for publication; Gupta and Lehmann 2003; Gupta, Lehmann, and Stuart 2004).

Turning to the upper-right quadrant of Fig. 2, the continuous-time analog of the sBG is the exponential-gamma (or Pareto of the second kind) distribution, a very common probability model of duration times that has been used to model, amongst other things, new product trial (Hardie, Fader, and Wisniewski 1998), life expectancies of retail franchise stores (Dekimpe and Morrison 1991), and the duration of jobs, strikes, and wars (Morrison and Schmittlein 1980). Despite this tradition of using continuous-time models to study a wide variety of marketing (and other) phenomena, there has been limited interest in using them for forward-looking customer-base analysis exercises.

The continuous-time analog of the churn rate is the hazard function. The hazard function of the exponential-gamma model decreases as a function of time (i.e., it exhibits negative duration dependence), even though the hazard function associated with the individual-level exponential distribution is constant over time. Thus the aggregate duration dependence is simply the result of a sorting effect in a heterogeneous population, as was the case with the sBG model in discrete-time settings.

Many new analysts find this result uncomfortable, contending that the phenomenon of increasing retention rates (or decreasing churn rates) should be captured at the level of the individual customer. In order to accommodate this, we can replace the exponential distribution with the Weibull distribution, which not only captures negative duration dependence at the individual-customer-level but also allows for positive duration dependence (i.e., churn increasing as a subscriber's tenure increases). Unobserved heterogeneity can be captured using a gamma distribution, giving us the Weibull-gamma distribution (Morrison and Schmittlein 1980).

Schweidel, Fader, and Bradlow (2008) find that the basic exponential-gamma structure is often quite acceptable; that is, aggregate retention rate dynamics are simply the result of a sorting effect in a heterogeneous population. When a Weibull-based model provides a better fit, the finding is that individual subscribers exhibit positive duration dependence (i.e., an increasing likelihood of churn over time), even though the aggregate churn rates are decreasing over time. This finding is supported by the work of Jamal and Bucklin (2006).

Discussion

We have reviewed a set of probability models that can be used as a basis for forward-looking customer-base analysis exercises, structuring the review around the central distinction of contractual vs. noncontractual business settings and the secondary distinction of whether transactions can occur at any point in time or only at discrete points in time. The common thread linking these models is the approach taken in model development. Observed behavior is modeled as a function of an individual's latent behavioral characteristics using simple probability distributions (or combinations thereof), with other probability distributions used to characterize heterogeneity in these latent characteristics. The application of Bayes' theorem enables us to make inferences about an individual's latent characteristics and therefore compute quantities such as the P(alive), the expected number of purchases in a future time period, mean residual lifetime and (residual) CLV, all conditional on the customer's observed behavior.

These are all parsimonious models of buyer behavior. While it may be tempting to develop complex models that encompass the richness of a specific application setting, the philosophy

[2] Some researchers have even proposed the use of nonparametric methods to model contract duration and therefore CLV (e.g., Pfeifer and Bang, 2005).

underlying these models is an evolutionary approach to model building — one in which we start with the simplest reasonable representation of the behaviors of interest, with a view to creating an easy-to-implement model (Fader and Hardie 2005). Further texture is added to the model only if the end-use application really requires it (i.e., only when the end user finds that the basic structure is indeed too simple).

It is interesting to reflect on the fact that the building blocks of these models were developed at a time when data were scarce; the models were used to extract as much "value" from the small amounts of data at hand. How times have changed! The proverbial fire hose is invoked over and over again, and advances in computation are not enough to keep up with the ever-increasing torrent of data. The models presented here come into their own in such an environment as they are simple to implement and make use of easy-to-compute data summaries. A fifty-year tradition of probability models in marketing lives on.

Acknowledgment

Bruce G.S. Hardie acknowledges the support of the London Business School Centre for Marketing.

References

Abe, Makoto (2008), "Counting Your Customers One by One: A Hierarchical Bayes Extension to the Pareto/NBD Model," *Marketing Science*, doi:10.1287/mksc.1080.0383.

Allenby, Greg M., Robert P. Leone, and Lichung Jen (1999), "A Dynamic Model of Purchase Timing With Application to Direct Marketing," *Journal of the American Statistical Association*, 94 (June), 365–74.

Batislam, Emine Persentili and Denizel Meltem (2007), "Empirical Validation and Comparison of Models for Customer-Based Analysis," *International Journal of Research in Marketing*, 24 (September), 201–9.

Batislam, E.P., M. Denizel, and A. Filiztekin (2008), "Formal Response to "Erratum on the MBG/NBD Model"," *International Journal of Research in Marketing*, 25 (September), 227.

Berger, Paul D. and Nada I. Nasr (1998), "Customer Lifetime Value: Marketing Models and Applications," *Journal of Interactive Marketing*, 12 (Winter), 17–30.

———, Bruce Weinberg, and Richard C. Hanna (2003), "Customer Lifetime Value Determination and Strategic Implications for a Cruise-Ship Company," *Journal of Database Marketing and Customer Strategy Management*, 11 (September), 40–52.

Berry, Michael J.A. and Gordon S. Linoff (2004), *Data Mining Techniques*, (2nd ed.). Indianapolis, IN: Wiley Publishing, Inc.

Blattberg, Robert C., Byung-Do Kim, and Scott A. Neslin (2008), *Database Marketing: Analyzing and Managing Customers*. New York: Springer.

———, Edward C. Malthouse, and Scott A. Neslin (accepted for publication), "Lifetime Value: Empirical Generalizations and Some Conceptual Questions," *Journal of Interactive Marketing*.

Colombo, Richard and Weina Jiang (1999), "A Stochastic RFM Model," *Journal of Interactive Marketing*, 13 (Summer), 2–12.

David Shepard Associates (1999), *The New Direct Marketing*, (3rd ed.). New York: McGraw-Hill.

Dekimpe, Marnik G., and Donald G. Morrison (1991), "A Modeling Framework for Analyzing Retail Store Durations," *Journal of Retailing*, 67 (Spring), 68–92.

Dwyer, F. Robert (1989), "Customer Lifetime Valuation to Support Marketing Decision Making," *Journal of Direct Marketing*, 3 (Autumn), 8–15.

Ehrenberg, A.S.C. (1959), "The Pattern of Consumer Purchases," *Applied Statistics*, 8 (March), 26–41.

——— (1988), *Repeat-Buying*, London: Charles Griffin and Company, Ltd.

Fader, Peter S. and Bruce G.S. Hardie (2005), "The Value of Simple Models in New Product Forecasting and Customer-Base Analysis," *Applied Stochastic Models in Business and Industry*, 21 (4), 461–73.

Fader, Peter S. and Bruce G.S. Hardie (2007a), "How to Project Customer Retention," *Journal of Interactive Marketing*, 21 (Winter), 76–90.

——— and Bruce G.S. Hardie (2007b), "Incorporating Time-Invariant Covariates into the Pareto/NBD and BG/NBD Models" (http://brucehardie.com/notes/019/).

——— and Bruce G.S. Hardie (2009), "Customer-Base Valuation in a Contractual Setting: The Perils of Ignoring Heterogeneity," *Marketing Science*, forthcoming.

———, Bruce G.S. Hardie, and Paul D. Berger (2004), "Customer-Base Analysis with Discrete-Time Transaction Data" (http://brucehardie.com/papers/020/).

———, Bruce G.S. Hardie, and Chun-Yao Huang (2004), "A Dynamic Changepoint Model for New Product Sales Forecasting," *Marketing Science*, 23 (Winter), 50–65.

———, Bruce G.S. Hardie, and Ka Lok Lee (2005a), "Counting Your Customers" the Easy Way: An Alternative to the Pareto/NBD Model," *Marketing Science*, 24 (Spring), 275–84.

———, Bruce G.S. Hardie, and Ka Lok Lee (2005b), "RFM and CLV: Using Iso-Value Curves for Customer Base Analysis," *Journal of Marketing Research*, 42 (November), 415–30.

———, Bruce G.S. Hardie, and Ka Lok Lee (2006), "More Than Meets The Eye," *Marketing Research*, 18 (Summer), 9–14.

———, Bruce G.S. Hardie, and Kinshuk Jerath (2007), "Estimating CLV Using Aggregated Data: The Tuscan Lifestyles Case Revisited," *Journal of Interactive Marketing*, 21 (Summer), 55–71.

Fielding, Michael (2005), "Get Circulation Going: DM Redesign Increases Renewal Rates for Magazines," *Marketing News*, 1 (September), 9–10.

Glady, Nicolas, Bart Baesens, and Christophe Croux (2008), "A Modified Pareto/NBD Approach for Predicting Customer Lifetime Value," *Expert Systems with Applications*, doi:10.1016/j.eswa.2007.12.049.

Greene, Jerome D. (2008), *Consumer Behavior Models for Non-Statisticians*. New York: Praeger.

Gupta, Sunil (accepted for publication), "Customer-Based Valuation," *Journal of Interactive Marketing*.

——— and Donald R. Lehmann (2003), "Customers as Assets," *Journal of Interactive Marketing*, 17 (Winter), 9–24.

———, Donald R. Lehmann, and Jennifer Ames Stuart (2004), "Valuing Customers," *Journal of Marketing Research*, 41 (February), 7–18.

Hardie, Bruce G.S., Peter S. Fader, and Michael Wisniewski (1998), "An Empirical Comparison of New Product Trial Forecasting Models," *Journal of Forecasting*, 17 (June–July), 209–29.

Hoppe, Daniel and Udo Wagner (2007), "Customer Base Analysis: The Case for a Central Variant of the Betageometric/NBD Model," *Marketing — Journal of Research and Management*, 3 (2), 75–90.

Hopmann, Jörg and Anke Thede (2005), "Applicability of Customer Churn Forecasts in a Non-Contractual Setting," in *Innovations in Classification, Data Science, and Information Systems (Proceedings of the 27th Annual Conference of the Gesellschaft fr Klassifikation e.V., Brandenburg University of Technology, Cottbus, March 12–14, 2003)*, Baier Daniel and Wernecke Klaus-Dieter, eds. Berlin: Springer-Verlag, 330–37.

Hughes Arthur M. (2006), *Strategic Database Marketing*. New York: McGraw-Hill.

Jain, Dipak and Siddharth S. Singh (2002), "Customer Lifetime Value Research in Marketing: A Review and Future Directions," *Journal of Interactive Marketing*, 16 (Spring), 34–46.

Jamal, Zainab and Randolph E. Bucklin (2006), "Improving the Diagnosis and Prediction of Customer Churn: A Heterogeneous Hazard Modeling Approach," *Journal of Interactive Marketing*, 20 (Summer/Autumn), 16–29.

Jerath, Kinshuk, Fader, Peter S., and Bruce G.S. Hardie (2007), "New Perspectives on Customer 'Death' Using a Generalization of the Pareto/NBD Model" (http://papers.ssrn.com/sol3/papers.cfm?abstract_id=995558).

Kaplan, Edward H. (1982), "Statistical Models and Mental Health: An Analysis of Records From a Mental Health Center," *M.S. Thesis, Department of Mathematics, Massachusetts Institute of Technology.*

Kumar, V., Girish Ramani, and Timothy Bohling (2004), "Customer Lifetime Value Approaches and Best Practice Applications," *Journal of Interactive Marketing*, 18 (Summer), 60–72.

———, Ilaria Dalla Pozza, J. Andrew Petersen, and Denish Shah (accepted for publication), "Reversing the Logic: The Path to Profitability through Relationship Marketing," *Journal of Interactive Marketing*.

Ma, Shao-Hui and Jin-Lan Liu (2007), "The MCMC Approach for Solving the Pareto/NBD Model and Possible Extensions," *Third International Conference on Natural Computation (ICNC 2007)*, 505–12.

Malthouse, Edward C. and Robert C. Blattberg (2005), "Can We Predict Customer Lifetime Value?" *Journal of Interactive Marketing*, 19 (Winter), 2–16.

Mason, Charlotte H. (2003), "Tuscan Lifestyles: Assessing Customer Lifetime Value," *Journal of Interactive Marketing*, 17, 54–60.

Massy, William F., David B. Montgomery, and Donald G. Morrison (1970), *Stochastic Models of Buying Behavior*, Cambridge. MA: The M.I.T. Press.

Moe, Wendy W., and Peter S. Fader (2004), "Capturing Evolving Visit Behavior in Clickstream Data," *Journal of Interactive Marketing*, 18 (1), 5–19.

Morrison, Donald G., and David C. Schmittlein (1980), "Jobs, Strikes, and Wars: Probability Models for Duration," *Organizational Behavior and Human Performance*, 25, 224–51.

——— and David C. Schmittlein (1988), "Generalizing the NBD Model for Customer Purchases: What Are the Implications and Is It Worth the Effort?" *Journal of Business and Economic Statistics*, 6 (April), 145–159.

———, Richard D.H. Chen, Sandra L. Karpis, and Kathryn E.A. Britney (1982), "Modelling Retail Customer Behavior at Merrill Lynch," *Marketing Science*, 1 (Spring), 123–41.

Neslin, Scott, Sunil Gupta, Wagner Kamakura, Junxiang Lu, and Charlotte H. Mason (2006), "Defection Detection: Measuring and Understanding the Predictive Accuracy of Customer Churn Models," *Journal of Marketing Research*, 46 (May), 204–11.

Parr Rud, Olivia (2001), *Data Mining Cookbook*. New York, NY: JohnWiley and Sons, Inc.

Pfeifer, Philip E. and Robert L. Carraway (2000), "Modeling Customer Relationships as Markov Chains," *Journal of Interactive Marketing*, 14 (Spring), 43–55.

——— and Paul W. Farris (2004), "The Elasticity of Customer Value to Retention: The Duration of a Customer Relationship," *Journal of Interactive Marketing*, 18 (Spring), 20–31.

——— and Heejung Bang (2005), "Non-Parametric Estimation of Mean Customer Lifetime Value," *Journal of Interactive Marketing*, 19 (Autumn), 48–66.

Reichheld, Frederick F. (1996), *The Loyalty Effect*. Boston, MA: Harvard Business School Press.

Reinartz, Werner and V. Kumar (2000), "On the Profitability of Long-Life Customers in a Non-Contractual Setting: An Empirical Investigation and Implications for Marketing," *Journal of Marketing*, 64 (October), 17–35.

——— and V. Kumar (2003), "The Impact of Customer Relationship Characteristics on Profitable Lifetime Duration," *Journal of Marketing*, 67 (January), 77–99.

Schmittlein, David C. and Robert A. Peterson (1994), "Customer Base Analysis: An Industrial Purchase Process Application," *Marketing Science*, 13 (Winter), 41–67.

——— and Robert A. Peterson (1987), "Counting Your Customers: Who They Are and What Will They Do Next?" *Management Science*, 33 (January), 1–24.

Schweidel, David A., Peter S. Fader, and Eric T. Bradlow (2008), "Understanding Service Retention Within and Across Cohorts Using Limited Information," *Journal of Marketing*, 72 (January), 82–94.

Wagner, Udo and Daniel, Hoppe (2008), "Erratum on the MBG/NBD Model," *International Journal of Research in Marketing*, 25 (September), 225–26.

Weinberg, Clarice Ring and Beth C. Gladen (1986), "The Beta-Geometric Distribution Applied to Comparative Fecundability Studies," *Biometrics*, 42 (September), 547–60.

Wu, Couchen and Hsiu-Li Chen (2000), "Counting Your Customers: Compounding Customer's In-Store Decisions, Interpurchase Time, and Repurchasing Behavior," *European Journal of Operational Research*, 127, 109–19.

Wübben, Markus and Florian Wangenheim (2008), "Instant Customer Base Analysis: Managerial Heuristics Often 'Get It Right'," *Journal of Marketing*, 72 (May), 82–93.

Zitzlsperger, David F.S., Thomas Robbert, and Stefan Roth (2007), "Forecasting Customer Buying Behaviour: The Impact of "One-time Buyer"," Proceedings of the ANZMAC 2007 Conference, University of Otago, Dunedin, New Zealand, (December 3–5).