3. Simple probability models for computing CLV and CE

Peter S. Fader and Bruce G.S. Hardie

INTRODUCTION

The past two decades have seen an evolution from transaction-oriented to customer-centric marketing strategies (Fader, 2012). As part of this shift, it is now common to think of a firm's customers as assets that generate cash flow, not just this period but in future periods as well. This is formalized in the notion of customer lifetime value (CLV), which Pfeifer, Haskins, and Conroy (2005) define as 'the present value of the future cash flows attributed to the customer relationship' (p. 17). Using the term 'customer equity' (CE) to denote the sum of the lifetime values of a firm's customers, Blattberg and Deighton (1996) suggested that a central goal for the marketing executive (and for the firm as a whole) should be to maximize CE.

At the heart of any attempt to measure CE is the calculation of CLV, which requires us to answer two questions: (1) How long will the customer remain 'alive'? and (2) What is the net cash flow per period while 'alive'? This requires the analyst to make multi-period forecasts of key behaviors (i.e., for each period, is the customer still 'alive' and, if so, how much do they spend?). In other words, this is a time-series modeling problem.

A popular classification scheme for time-series models is that proposed by Cox (1981), who distinguishes between observation-driven and parameter-driven models. Let Y_t denote the random variable(s) associated with the behavior(s) of interest at time t; this is assumed to be a function of parameter(s) θ_t. In an observation-driven model, θ_t is modeled as some function of past observed behavior: $\theta_t = f(y_{t-1}, y_{t-2}, \ldots)$. In a parameter-driven model, θ_t is a function of past values of the latent variable(s): $\theta_t = g(\theta_{t-1}, \theta_{t-2}, \ldots)$.

While observation-driven models are well suited for the task of making one-period-ahead forecasts, there are problems when using them to generate the multi-period forecasts required for the estimation of CLV and CE. Suppose we specify a model in which $\theta_t = f(y_{t-1}, y_{t-2}, \ldots)$ and have calibrated the parameters of $f(\cdot)$ using data on y_1, y_2, y_3. Standing at the end of period 3, we can predict period 4 behavior using y_2 and y_3. However, in order to predict period 5 behavior, we need y_3 and y_4, yet period 4 is

currently in the future. We therefore have to use the model to simulate period 4 behavior, and so on. As we move further away from the last period for which we have realizations of Y_t, we expect the forecasts of the behavior of interest to become increasingly unreliable as any prediction error propagates through the forecasts.[1]

Parameter-driven models do not suffer from this problem. The stochastic process underpinning $g(\cdot)$ is estimated using the observed data y_1, y_2, y_3 and no additional observations (actual or simulated) are required to generate forecasts of behavior regardless of the length of the forecast horizon.

The majority of models developed by researchers interested in CE are observation-driven models (e.g., Kumar and Petersen, 2012). In this chapter, we review the key parameter-driven models that have been developed by marketing scientists that can be used to generate the multi-period forecasts of key behaviors required for the calculation of CLV and CE. These models go under the label of 'probability models for customer-base analysis', and are specified, tested, and communicated following the well-established traditions of stochastic models of buyer behavior, which have been part of the marketing science literature from its very beginning (Fader, Hardie, and Sen, 2014).

Rather than try to tease out the effects of all the factors that may influence buyer behavior, a probability model completely embraces the notion of stochasticity, viewing the behavior of interest as the outcome of one or more probabilistic process(es). This does not imply that we think the individual is truly purchasing 'at random'. Instead it simply reflects our uncertainty regarding the factors that influence buying behavior; the probabilistic process captures the net effect of all the influences not explicitly considered in the model.

A probability model of buyer behavior typically has two components. First, the individual behavior of interest (e.g., length of time they remain a customer) is characterized in terms of a probability distribution (or several distributions in the case of more complex models), with the parameters of this distribution reflecting the individual's underlying behavioral propensities. Second, differences in these underlying behavioral propensities are captured by an additional probability distribution. These two components result in a mixture distribution, which characterizes the behavior of a randomly chosen individual. While such models do not give us any direct insights as to 'how' or 'why' consumers behave the way they do, they have proven capable of generating the accurate multi-period forecasts of buyer behavior needed for the calculation of CLV and CE.

CLASSIFYING ANALYSIS SETTINGS

Before reviewing the key models, we must first highlight some of the key types of firm–customer relationships, which will drive model formulations.

Consider, as a starting point, the following two statements regarding the size of a company's customer base:

- Based on numbers presented in a news release that reported Vodafone Group Plc's results for the six months ended 30 September 2013, we see that Vodafone UK had 11.3 million 'pay monthly' customers at the end of that period.
- In his 'Q3 2013 Earnings Conference Call', the CFO of Amazon made the comment that '[a]ctive customer accounts exceeded 224 million', where customers are considered active if they have placed an order during the preceding 12-month period.

At first glance, both statements seem like perfectly logical characterizations of the size of each firm's customer base. However, after careful consideration, we realize that only one of them (the first one) is a valid description. We can be 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'/prepaid tariffs), the customer must tell Vodafone when they wish to switch their mobile phone provider. Therefore, Vodafone knows for sure the time at which any customer 'dies'. Such a business is viewed as having a subscription-based or 'contractual' relationship with its customers.

In contrast, the Amazon statement defines an 'active' customer as someone who has 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 someone placed an order 364 days ago, they are still a customer? This 12-month cut-off point is completely arbitrary. If we were to change the cut-off point to six months, the apparent size of Amazon's customer base would be smaller, even though the true size would remain unchanged. In such a business setting, the 'death' of customers is unobserved by the firm (Schmittlein, Morrison, and Colombo, 1987). Customers do not formally end their relationship with the firm; instead they just 'silently attrite' (Mason, 2003). As such, the firm can never know for sure how many customers it has at any point in time. Reinartz and Kumar (2000) introduced the label 'non-contractual' to characterize such a business setting.

While ignored by a number of researchers, this contractual/non-contractual distinction is of fundamental importance for any researcher

developing models for estimating CLV and CE.[2] The key challenge of non-contractual 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.

A secondary distinction concerns the treatment of time. Consider the following four specific business settings: (1) airline lounges, (2) electrical utilities, (3) academic conferences, and (4) mail-order clothing companies. The first two are clearly contractual settings, while the last two are clearly non-contractual settings. 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 in any given year, and one either attends or does not attend. 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, a common characteristic of the second and fourth settings is that there are no constraints as to when the customer can purchase clothing or end his or 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 et al. (1987), as illustrated in Figure 3.1.[3] When developing a model for CLV and CE estimation, we must ask ourselves in which quadrant we are operating. While in certain circumstances a model developed for a discrete setting can be applied in a continuous setting (and

	Non-contractual	Contractual
Continuous	Grocery purchasing Doctor visits Hotel stays	Credit cards Utilities Continuity programs
Discrete	Conf. attendance Prescription refills Charity fund drives	Magazine subs Insurance policies 'Friends' schemes

Opportinities for transactions

Type of relationship with customers

Figure 3.1   Classifying customer bases

vice versa), the contractual/non-contractual divide is fundamental and the boundary cannot be crossed. We use this framework to guide our review of the literature.

CLV REVISITED

As noted in the introduction, CLV is the present value of the future cash flows attributed to the customer relationship. Following Rosset et al. (2003), we can express this mathematically as:

$$E(CLV) = \int_0^\infty E[v(t)]S(t)d(t)dt, \qquad (3.1)$$

where, for $t \geq 0$ (with $t = 0$ representing 'now'), $E[v(t)]$ is the expected value (or net cash flow) of the customer at time t, $S(t)$ is the probability that the customer has remained alive to at least time t, and $d(t)$ is a discount factor that reflects the present value of money received at time t.

It is important to distinguish between (1) the expected lifetime value of an as-yet-to-be-acquired customer, (2) the expected lifetime value of a just-acquired customer, and (3) the expected *residual* lifetime value, $E(RLV)$, of an existing customer. (The difference between the first two quantities is simply the value of the first transaction that signals the start of the relationship.) We can express the notion of RLV mathematically as:

$$E(RLV) = \int_{t'}^\infty E[v(t)]S(t|t > t')d(t - t')dt, \qquad (3.2)$$

where t' is the 'age' of the customer at which time their residual lifetime value is computed.

Consider a contractual setting. Following Fader and Hardie (2010), if we assume a constant expected net cash flow per contract period (i.e., $E[v(t)] = \bar{v}$), we can express equation (3.1) as:

$$E(CLV) = \bar{v} \, DEL,$$

where the 'discounted expected lifetime' (DEL) of the customer is defined as:

$$DEL = \int_0^\infty S(t)d(t)dt. \qquad (3.3)$$

(With no discounting, we have $\int_0^\infty S(t)dt$, which is simply $E(T)$, the customer's expected lifetime.) Similarly, we can express equation (3.2) as:

$$E(RLV) = \bar{v}\, DERL,$$

where the 'discounted expected residual lifetime' (DERL) of the customer is defined as:

$$DERL = \int_{t'}^{\infty} S(t|t > t')\, d(t - t')\, dt$$

$$= \int_{t'}^{\infty} \frac{S(t)}{S(t')} d(t - t')\, dt. \tag{3.4}$$

At the heart of a probability model for a contractual setting is an expression for $S(t)$. For a given model specification, we can derive the corresponding expressions for DEL and DERL. We review the basic results in the next section.

Now consider a non-contractual setting. Following Fader, Hardie, and Lee (2005b), if we assume a constant net cash flow per transaction of $\bar{v}$, we have $v(t) = \bar{v}\, t(t)$, where $t(t)$ is the transaction rate at time t, and can express equation (3.1) as:

$$E(CLV) = \bar{v}\, DET,$$

where the 'discounted expected transactions' (DET) of the customer is:

$$DET = \int_{0}^{\infty} t(t)\, S(t)\, d(t)\, dt. \tag{3.5}$$

(For a just-acquired customer, DET measures the present value to the firm of all future transactions by the customer, with transactions at a future time point appropriately discounted to obtain their present values.) Similarly, we can express equation (3.2) as:

$$E(RLV) = \bar{v}\, DERT,$$

where the 'discounted expected residual transactions' (DERT) of the customer is:

$$DERT = \int_{t'}^{\infty} t(t)\, S(t|t > t')\, d(t - t')\, dt. \tag{3.6}$$

At the heart of a probability model for a non-contractual setting are expressions for $t(t)$ and $S(t)$. For a given model specification, we can derive the corresponding expressions for DET and DERT. We review the basic results in the 'Non-contractual Settings' section below.

CONTRACTUAL SETTINGS

Numerous researchers working in the areas of marketing, applied statistics, and data mining have developed a number of models that attempt to identify those customers with the greatest risk of churning in the next time period (Parr Rud, 2001; Lemmens and Croux, 2006; Neslin et al., 2006; Risselada, Verhoef, and Bijmolt, 2010; Berry and Linoff, 2011). 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. As such analyses are observation-driven models, they are of limited use to the analyst interested in estimating CLV and CE. For example, in order to predict churn in t period's time from now, we need to know the number of calls made to the call center over the next $t-1$ periods, which is an unknown.

If we consider a cohort of customers acquired at the same time, we typically observe that the cohort-level retention rates increase over time (e.g., Kumar and Reinartz, 2012, Figure 5.2). While it is tempting to tell a story of individual-level time dynamics (e.g., increasing loyalty as the customer gains more experience with the firm, and/or increasing switching costs with the passage of time), a far simpler story – and one consistent with the fundamental marketing concept of segmentation – is that of a sorting effect in a heterogeneous population.

A very simple parameter-driven model for discrete-time duration data is the beta-geometric (BG) model (Potter and Parker, 1964). In a contractual setting, it is formulated by assuming that (1) 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 a geometric distribution), and (2) individual differences in θ are characterized by a beta distribution with parameters γ and δ. The resulting survivor function is:

$$S(t|\gamma,\delta) = \frac{B(\gamma, \delta + t)}{B(\gamma, \delta)}, \quad t = 0, 1, 2, \ldots \qquad (3.7)$$

The corresponding cohort-level retention rate is:

$$r(t) = \frac{\delta + t - 1}{\gamma + \delta + t - 1},$$

which increases over time, even though each individual's unobserved (and unobservable) churn probability is assumed to be constant over

time. Despite what may seem to be overly simplistic assumptions, the analyses presented in Fader and Hardie (2007a) demonstrate that this two-parameter model (which can easily be implemented in a basic Excel spreadsheet) generates very accurate forecasts of retention. As such, we can confidently use it as the basis of attempts to compute CLV and CE in contractual settings, something explored in Fader and Hardie (2010) and extended in Braun and Schweidel (2011).

Substituting equation (3.7) in equation (3.3) and solving the integral (which becomes a sum in a discrete-time setting) gives us the following expression for the discounted expected lifetime of an as-yet-to-be-acquired customer:

$$DEL(\gamma, \delta, d) = {_2F_1}\left(1, \delta; \gamma + \delta; \frac{1}{1 + d}\right), \tag{3.8}$$

where ${_2F_1}(\cdot)$ is the Gaussian hypergeometric function and d is the appropriate discount factor used to represent the time value of money. (When there are k contract periods per year, an annual discount rate of $(100 \times r)$ percent maps to a period discount rate of $d = (1 + r)^{1/k} - 1$.) To compute DEL for a just-acquired customer, we simply subtract 1 from this quantity.

Substituting equation (3.7) in equation (3.4) and solving the integral gives the following expression for the discounted expected residual lifetime of a customer who has renewed his or her contract $n - 1$ times, evaluated *just before* we find out whether they will make an nth contract renewal:

$$DERL(\gamma, \delta, d; n - 1 \text{ renewals})$$

$$= \left(\frac{\delta + n - 1}{\gamma + \delta + n - 1}\right) {_2F_1}\left(1, \delta + n; \gamma + \delta + n; \frac{1}{1 + d}\right). \tag{3.9}$$

The continuous-time analog of the BG is the Pareto distribution of the second kind (Pareto Type II), one characterization of which is a gamma mixture of exponentials. In other words, it is assumed that (1) the duration of the customer's relationship with the firm is characterized by an exponential distribution with rate parameter λ, and (2) individual differences in λ are characterized by a gamma distribution with shape parameter s and scale parameter β. The associated survivor function is:

$$S(t|s, \beta) = \left(\frac{\beta}{\beta + t}\right)^s, \quad t \geq 0. \tag{3.10}$$

The continuous-time analog of the churn rate is the hazard function. The hazard function of the Pareto Type II is:

$$h(t) = \frac{s}{\beta + t},$$

which 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, once again, the aggregate duration dependence is simply the result of a sorting effect in a heterogeneous population, as is the case with the BG model in discrete-time settings.

Substituting equation (3.10) in equations (3.3) and (3.4), and solving the integrals gives us:

$$DEL(s, \beta, d_c) = \beta^s d_c^{s-1} \Psi(s, s; \beta d_c) \qquad (3.11)$$

$$DERL(s, \beta, d_c; \text{tenure of } t')$$

$$= (\beta + t')^s d_c^{s-1} \Psi(s, s; (\beta + t')d_c) \qquad (3.12)$$

where $\Psi(\cdot)$ is the confluent hypergeometric function of the second kind and d_c is the continuously compounded discount rate. (A discrete-time discount rate of d is equivalent to a continuously compounded rate of $d_c = \ln(1+d)$. If the data are recorded in time units such that there are k periods per year ($k = 52$ if the data are recorded in weekly units of time), an annual discount rate of $(100 \times r)$ percent maps to a continuously compounded rate of $d_c = \ln(1+r)/k$.)

One stream of CE research has focused on the problem of valuing a customer base in contractual settings (e.g., Bauer, Hammerschmidt, and Braehler, 2003; Gupta and Lehmann, 2003; Gupta, Lehmann, and Stuart, 2004; Wiesel, Skiera, and Villaneuva, 2008; Libai, Muller, and Peres, 2009; Pfeifer, 2011). CLV and subsequent CE calculations are performed using a given constant retention rate (which implies the parameter-driven model underpinning this work is the geometric distribution for contract durations). The justification for such an assumption, if ever given, is that published retention rate numbers appear to be fairly constant – see, for example, Figure 3.2, which plots two years of quarterly annualized churn numbers from Vodafone Germany.

How do we reconcile such data with the above comments about increasing retention rates? The following comment on a professional society's membership patterns hints at the answer: 'I am happy to report that 41% of new members who joined in 2011 renewed their membership in 2012,
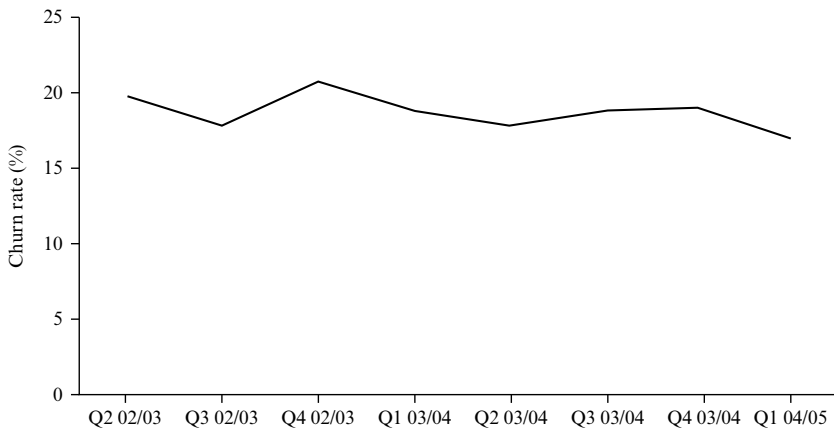
Figure 3.2   Annualized churn numbers for Vodafone Germany

and that ION has an overall retention of 78%' (Walter, 2013, p. 2). The aggregate retention rate numbers are a weighted average of the retention rates across all cohorts at any given time (i.e., a mix of 'young' and 'old' customers), and this weighted average will mask (and moderate) the within-cohort retention patterns, giving the analyst the impression that the retention dynamics are mild (and therefore possibly ignorable). This is illustrated in Figure 3.3, where the retention rate curve for each cohort is that associated with a beta-geometric distribution with parameters $\gamma = 0.764$ and $\delta = 1.296$.[4] Assuming the cohorts are of equal size and that the first cohort was acquired in period 0, the aggregate retention rate for any given period is simply the weighted average of each cohort-specific retention rate where the weights are the number of surviving cohort members at that point in time.

Does this mean we can 'assume away' the within-cohort retention dynamics as a minor source of noise when computing CE in contractual settings? The answer is a definite no. To illustrate, taking the aggregate retention rate in Figure 3.3 as a constant 0.833, Figure 3.4 compares the geometric survival curve given this constant retention rate with that implied by the common cohort-level retention curves plotted in Figure 3.3. Clearly any estimates of customer survival computed using the aggregate retention rate will be severely biased compared to those from a heterogeneous model, such as the BG, which allows for increasing cohort-level
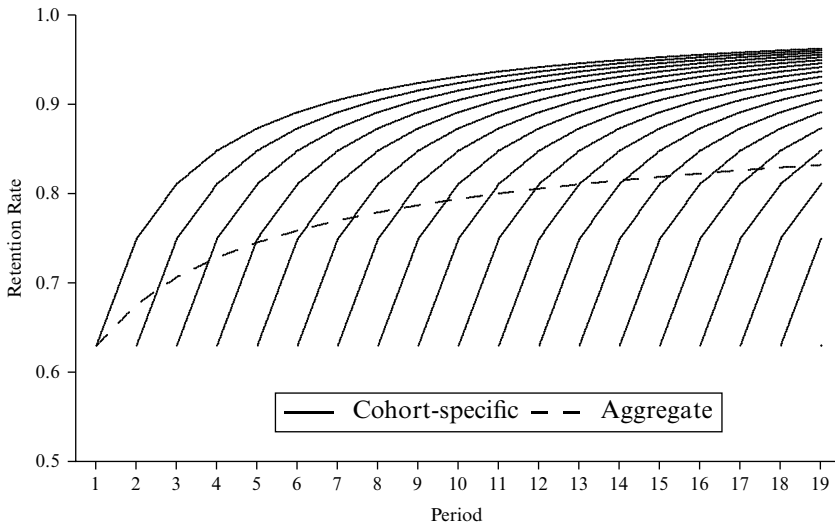
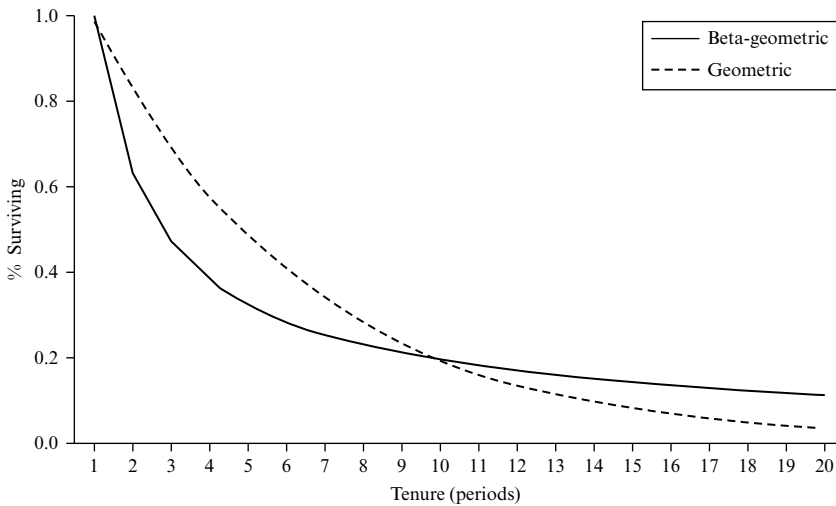*Figure 3.3   Aggregate vs cohort-level retention rates*

Figure 3.4   Comparing the geometric survival curve (computed using the constant aggregate retention rate) with the cohort-level curve (from the BG model)

retention rates. This has obvious implications for estimates of CLV and CE. As we extend into the future (e.g., beyond period 20 in Figure 3.4), the gap between the two curves will persist, leading to a substantial underestimation of CLV (and CE) for the homogeneous model.

Consider a setting in which the net cash flow per period is $100 and the annual discount rate is 10 percent. Taking the aggregate retention rate as a constant 0.833, the expected residual lifetime value of a customer (computed at the end of the annual contract period) would be a constant $343, regardless of how many periods the individual has been a customer. In contrast, any calculation of the expected residual lifetime value of a customer that acknowledges the phenomenon of increasing cohort-level retention rates will see the estimate increase as we lengthen in tenure of the customer. Using equation (3.9) we find that the expected residual lifetime value is $288 for a customer approaching the end of their first contract period with the firm, $394 for a customer approaching the end of their second contract period with the firm, and $568 for a customer approaching the end of their fifth contract period with the firm. Fader and Hardie (2010) show that failing to account for cohort-level retention-rate dynamics will lead to systematically biased estimates of the residual value of the customer base (i.e., CE). Their analysis indicated that valuations performed using an aggregate retention rate will underestimate the true value of the customer base by a magnitude of 25–50 percent in standard settings.

Why then do we continue to see researchers performing CE calculations in contractual settings using a constant aggregate retention rate? Perhaps one reason is the belief that individual-level data or longitudinal data for each cohort are required in order to estimate the parameters of the model use to characterize contract durations (e.g., the BG distribution). It turns out that this need not be the case; Fader and Hardie (2007b), show that, under certain assumptions, all we require are data on the number of new subscribers and the total number of subscribers for each period. However, it is assumed that we have these data from the time the service of interest was launched on the market and that the reporting period is the same as the contract period. Fader, Hardie, and Liu (2012) show how to deal with the case where the reporting period spans multiple renewal periods and the data series is left censored (i.e., we do not have the data from the time the service was launched).

NON-CONTRACTUAL SETTINGS

The retention rate is defined as 'the ratio of customers retained to the number at risk' (Farris et al., 2010, p.156). As noted above, the defin-

ing characteristic of a non-contractual setting is that customer 'death' is unobserved (and unobservable). As such, we cannot compute a retention rate directly from customer purchasing data (i.e., without an explicit model of the latent attrition process). Two metrics we can compute in non-contractual settings are the repeat-buying rate and repurchase rate, defined as 'the percentage of brand customers in a given period who are also brand customers in the subsequent period' and 'the percentage of brand customers for a brand who repurchase that brand on their next purchase occasion' respectively (ibid., p. 48). The measures are frequently incorrectly referred to as retention rates; for example, the annual J.D. Power Customer Retention Study for automobiles talks of retention when they are actually computing a repurchase rate. Similarly, we can find a number of references to 'retention rates' in settings where the measure being discussed is clearly a repeat-buying rate (e.g., Ofek, 2002; Dev and Stroock, 2007). This has seen researchers unwittingly using these 'retention rate' numbers as if they were actually retention rates, using them in the basic formulas we associate with textbook discussions of CLV (Fader and Hardie, 2012). This leads to seriously biased estimates of CLV and CE.

If we track over time the repeat purchasing of a cohort of customers (acquired at the same time) in a non-contractual setting, we see the general pattern illustrated in Figure 3.5. Some individuals make frequent purchases; for others, there are long time intervals between their pur-
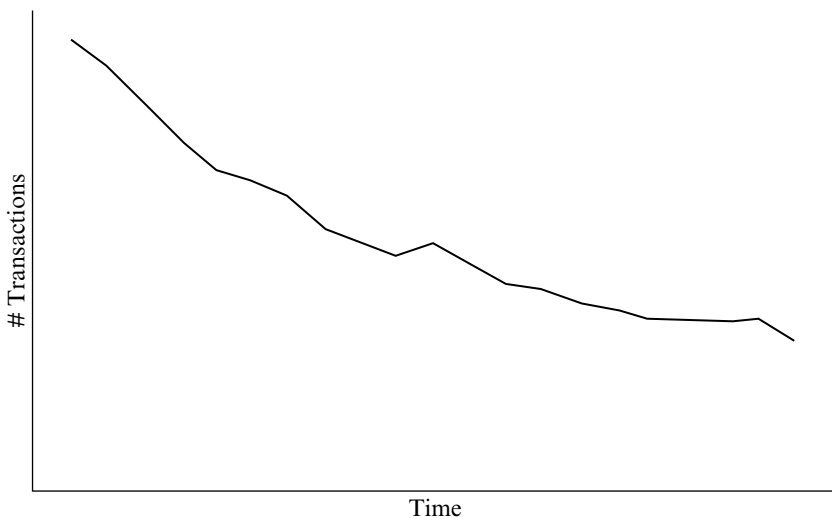

Figure 3.5   Stylized pattern of purchasing over time by a cohort of customers in a non-contractual setting

chases. (In contrast to a contractual setting, the fact that a customer has not made a purchase does not mean that they are 'dead'.)

While there are many non-stationary processes that could give rise to such a decline in repeat purchasing at the cohort level (e.g., Fader, Hardie, and Huang, 2004; Moe and Fader, 2004), one approach that has proven to be very popular is the latent attrition (or 'buy till you die') framework put forward by Schmittlein et al. (1987). A customer's relationship with a firm has two phases: they are 'alive' for an unobserved period of time, then 'dead'. Ignoring the effect of random purchasing around their means, individual customers purchase the product at steady but different underlying rates. At different unobservable points in time they 'die'.[5] This is illustrated in Figure 3.6, where each individual's purchasing while alive is represented by a dashed line, and the unobserved point at which they die by a diamond. The observed decline in repeat purchasing (the solid line) is due to the unobserved death of customers A, C, and D.

Schmittlein et al.'s (1987) operationalization of this framework is for a continuous-time setting and assumes that (1) while 'alive' the customer's purchasing is characterized by the negative binomial distribution (NBD) model (i.e., a gamma mixture of Poissons) with parameters r and α, and (2) the unobserved customer 'lifetimes' are treated as if random and are characterized by the Pareto Type II distribution (i.e., a gamma mixture of exponentials) with parameters s and β. Thus, the resulting four-parameter model of buyer behavior for non-contractual settings is called the Pareto/NBD.
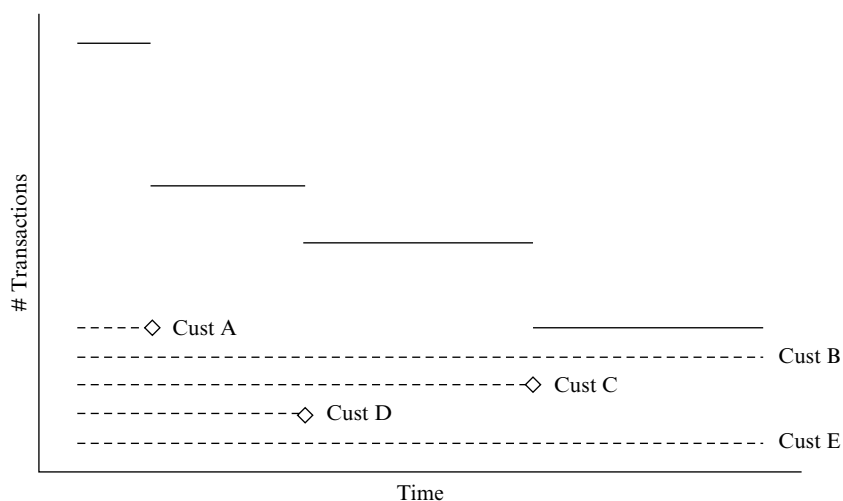

Figure 3.6   Visualizing the latent-attrition explanation for the decline in cohort-level purchasing in a non-contractual setting

Empirical validations of the model are presented in Schmittlein and Peterson (1994) and Fader et al. (2005b), amongst others. We find that its predictive performance is impressive, able to closely track the sales data in a hold-out period (at both the aggregate and the individual level). This gives us confidence in the Pareto/NBD model as the most appropriate starting point for CLV and CE calculations in a non-contractual setting. Two important early applications of the model are given by Reinartz and Kumar (2000, 2003), who used the Pareto/NBD model in their analysis of customer profitability.

The assumptions of the Pareto/NBD model mean that we do not need to know customers' entire purchase history; the only customer-level information required to estimate the four model parameters and then make individual-level predictions are 'recency' and 'frequency'. 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.

One of the key expressions derived by Schmittlein et al. (1987) is P (alive $| x, t_x, T)$, the probability that an individual with observed behavior (x, t_x, T) is still 'alive' at time T. This quantity is of limited value when computing CLV or CE, as we need to know the probability that the customer is alive at all future points. Of greater use is the expression they derive for the conditional expectation, $E(X(T, T + t) | x, t_x, T)$, which is the expected number of transactions in the future period $(T, T + t]$ for an individual with observed behavior (x, t_x, T). This quantity can be used to compute DERT in the following manner:

$$DERT = \sum_{i=1}^{\infty} \left(\frac{1}{1 + d}\right)^{i - 0.5} \{E(X(T, T + i)|x, t_x, T)$$

$$- E(X(T, T + i - 1)|x, t_x, T)\}. \tag{3.13}$$

A more elegant solution is to substitute the Pareto/NBD model expressions for $t(t)$ and $S(t)$ in equations (3.5) and (3.6) and solve the integrals, giving us (Fader et al., 2005b; Jerath, Fader, and Hardie, 2014):

$$DET(r, \alpha, s, \beta, d_c) = \frac{r}{\alpha} \beta^s d_c^{s-1} \Psi(s, s; \beta d_c), \tag{3.14}$$

$$DERT(r, \alpha, s, \beta, d_c; x, t_x, T) = \frac{\Gamma(r+x+1)}{\Gamma(r)} \frac{\alpha^r \beta^s d_c^{s-1}}{(\alpha + T)^{r+x+1}}$$

$$\times \frac{\Psi(s, s; (\beta + T)d_c)}{L(r, \alpha, s, \beta | x, t_x, T)}, \tag{3.15}$$

where $L(\cdot \mid \cdot)$ is the Pareto/NBD model likelihood function. (Note that the DET quantity is that for a just-acquired customer. To compute DET for an as-yet-to-be-acquired customer, we need to add 1 to this quantity (i.e., the purchase at time $t = 0$ that corresponds to the customer's first-ever purchase with the firm, and which starts the repeat purchase clock).)

To illustrate the importance of distinguishing between DET and DERT, let us consider an example where $r = 0.553$, $\alpha = 10.578$, $s = 0.606$, and $\beta = 11.669$.[6] Assuming an annual discount rate of 10 percent, DET = 4.8. The value of DERT obviously depends on the customer's exact purchase history. For the case of customers acquired one year ago, we evaluate equation (3.15) for all combinations of recency ranging from $t_x = 0$ to 52 weeks and frequency ranging from 0 to 10 transactions. For the case of customers acquired two years ago, we evaluate equation (3.15) for all combinations of recency ranging from $t_x = 0$ to 104 weeks and frequency ranging from 0 to 20 transactions. The corresponding contour (or 'iso-value') plots are presented in Figure 3.7. Each curve in the plots links customers with equivalent future value (in terms of discounted transactions) despite differences in their prior behavior. Depending on the age of the customer and their past purchasing behavior, their residual value ranges from almost zero to an order of magnitude greater than DET.

In order to compute CLV or RLV using the Pareto/NBD model, we need to augment equations (3.14) and (3.15) with a model for spend per
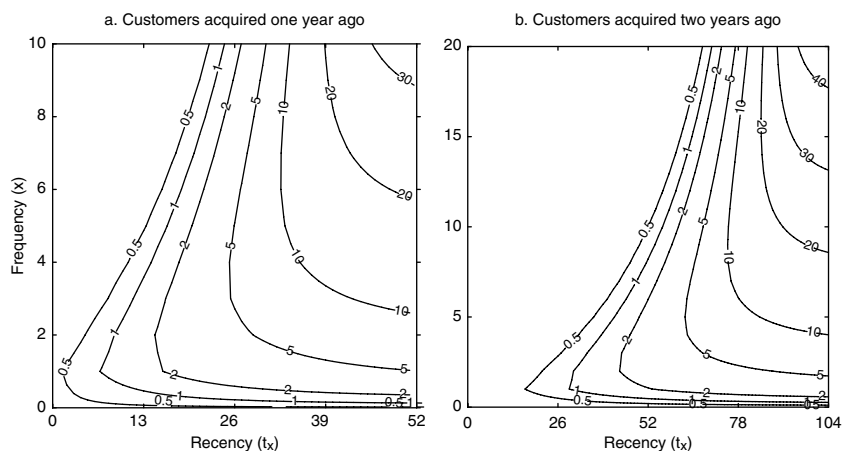

Figure 3.7a,b   DERT as a function of recency and frequency for customers acquired one and two years ago

transaction. Fader et al. (2005b) proposed the three-parameter 'gamma mixture of gammas' model for spend per transaction – also see Colombo and Jiang (1999) and Fader and Hardie (2013) – that can be used to estimate an individual customer's underlying mean transaction value given their average transaction value over the x observed transactions. This means we have closed-form expressions for the residual lifetime value of a customer conditional on the so-called RFM summary measures of their behavior: recency (the time of their last transaction), frequency (the number of transactions observed in the calibration period), and monetary value (average spend per transaction in the calibration period).

In some situations we do not have access to individual-customer-level 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) and Jerath et al. (2014) show how the model parameters can still be estimated using such 'repeated cross-sectional summary' data, despite this limitation. As such, we are able to compute CE in the absence of the raw individual-customer-level data (or the associated recency and frequency summaries).

Despite being published in 1987, the Pareto/NBD saw relatively limited 'real-world action' for a number of years, the major problem being perceived challenges with respect to parameter estimation. To address this problem, Fader, Hardie, and Lee (2005a) developed a variant of the Pareto/NBD model that they called the BG/NBD. Changing the 'death' story to one where a customer could only 'die' immediately after a transaction (as opposed to any point in continuous time as for the Pareto/NBD) and assuming that this time-to-death in 'transaction time' is characterized by the BG distribution with parameters a and b, resulted in a model that is vastly easier to implement. In fact, its key selling point is that its parameters can be obtained quite easily using 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, including as the basis for CLV and CE calculations (e.g., Lerner, Fader, and Hardie, 2012; Song, Kim, and Kim, 2013).

Note that the BG/NBD model expressions for DET and DERT are very messy. It makes more sense to compute DERT using the model-based conditional expectations, computed using:

$$E(X(T,T + t)|r, \alpha, a, b; x, t_x, T,) = \left(\frac{a+b+x-1}{a-1}\right)$$

$$\times \left[1 - \left(\frac{\alpha+T}{\alpha+T+t}\right)^{r+x} {}_2F_1\left(r+x, b+x; a+b+x-1; \frac{t}{\alpha+T+t}\right)\right]$$

$$\div \left[1+\delta_{x>0}\frac{a}{b+x-1}\left(\frac{\alpha+T}{\alpha+t_x}\right)^{r+x}\right], \tag{3.16}$$

in conjunction with equation (3.13). DET can be computed in a similar manner using the unconditional expectation formula:

$$DET = \sum_{t=1}^{\infty}\left(\frac{1}{1+d}\right)^{i-0.5}\{E(X(t)|r, \alpha, a, b)$$

$$- E(X(t-1)|r, \alpha, a, b)\}, \tag{3.17}$$

where the expected number of purchases in the interval $(0, t]$ is:

$$E(X(t)|r,\alpha,a,b) = \frac{a+b-1}{a-1}\left[1-\left(\frac{\alpha}{\alpha+t}\right)^r {}_2F_1\left(r, b; a+b-1; \frac{t}{\alpha+t}\right)\right]. \tag{3.18}$$

The Pareto/NBD and its extensions are for non-contractual settings where transactions can occur at any point in time. As reflected in Figure 3.1, 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 i fund drive, even though the check could be received at any point in calendar time. Even for business settings that are truly non-contractual/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 the customers of a cruise ship company in terms of whether or not they made 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. As in the case of the continuous-time setting, the

challenge facing the modeler is to determine whether a trailing sequence of 0s reflects a 'dead' customer or simply one that is in the middle of a long hiatus since the last transaction. Assuming the latent-attrition/'buy till you die' framework illustrated in Figure 3.6, Fader, Hardie, and Shang (2010) develop a model in which, while 'alive', the customer's purchasing is characterized by a Bernoulli process with beta heterogeneity (i.e., the BB model with parameters α and β), and the unobserved customer 'lifetimes' are treated as if random and characterized by the beta-geometric distribution (i.e., a beta mixture of geometrics with parameters γ and δ). The BB is the discrete-time analogue of the NBD, and the BG is the discrete-time analogue of the Pareto Type II. Thus, the resulting BG/BB model of buyer behavior is the discrete-time analogue of the Pareto/NBD, which tends to the Pareto/NBD when the length of the discrete time period tends to zero. 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 of purchases); the only customer-level information required to estimate the four model parameters and then make individual-level predictions are recency and frequency, denoted by (x, t_x, n) for a calibration period of length n discrete/discretized time-periods. As with the other models discussed in this review, the predictive performance of the BG/BB model is impressive, able to closely track the sales data in a holdout period (at both the aggregate and the individual level). As such, we have confidence in using it as the basis for CLV and CE calculations in discrete-time non-contractual settings. Such calculations would be based off the following expressions:

$$DET(\alpha, \beta, \gamma, \delta, d) = \left(\frac{\alpha}{\alpha + \beta}\right)\left(\frac{\delta}{\gamma + \delta}\right)\left(\frac{1}{1 + d}\right)$$

$$\times\ _2F_1(1, \delta + 1; \gamma + \delta + 1; \tfrac{1}{1 + d}) \tag{3.19}$$

$$DERT(\alpha, \beta, \gamma, \delta, d; x, t_x, n) = \frac{B(\alpha + x + 1, \beta + n - x)}{B(\alpha, \beta)} \frac{B(\gamma, \delta + n + 1)}{B(\gamma, \delta)(1 + d)}$$

$$\times \frac{_2F_1\left(1, \delta + n + 1; \gamma + \delta + n + 1; \dfrac{1}{1 + d}\right)}{L(\alpha, \beta, \gamma, \delta | x, t_x, n)} \tag{3.20}$$

where $L(\cdot|\cdot)$ is the BG/BB model likelihood function. (Note that the DET quantity is that for a just-acquired customer. To compute DET for an as-yet-to-be-acquired customer, we need to add 1 to this quantity (i.e.,

the purchase at time $t = 0$ that corresponds to the customer's first-ever purchase with the firm, and which starts the 'transaction opportunity' clock).)

REFLECTIONS

At the heart of any attempt to calculate CLV and CE is a model that can be used to make 'multi-period forecasts' of buyer behavior. We have identified a core set of simple probability models of buyer behavior well suited to this task; these models can be classified according to (1) whether they are designed to characterize buyer behavior in contractual versus non-contractual settings, and (2) whether the behavior of interest is best characterized as occurring in discrete or continuous time – see Figure 3.8 for a summary. We have deliberately limited our discussion to the core set of models for each quadrant. A discussion of the various extensions to these basic models can be found in Fader and Hardie (2009).

Whenever we are computing CLV (and therefore CE), it is crucially important that we distinguish between the expected lifetime value of a new customer and the expected residual lifetime value of an existing customer. Failure to make this distinction leads to severely biased estimates in the value of the firm's customer base. Making the (common) simplifying assumption of constant net cash flow per contract period or transaction,[7] the key quantities of interest to the analyst are DEL and DERL in a contractual setting, and DET and DERT in a non-contractual setting. We have presented the equations for computing these quantities for the five core models presented in Figure 3.8.

With all of these models, 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 resulting mixture models are 'simple' (Fader and Hardie, 2005): they are parsimonious (i.e., have very few parameters) and have limited data requirements (e.g., very simple data structures). With the exception of the Pareto/NBD, all the models can be implemented from scratch in an Excel spreadsheet; Matlab code[8] and an R package[9] are available for the Pareto/NBD model. Most importantly, the multi-period forecasting performance of these models has been established, which is fundamental to their use as the basis for any CLV and CE calculations.

Another thing these models have in common is the absence of any covariates that reflect the marketing activities of the firm. Any analyst using a with-covariates model will need to forecast the values of these

Figure 3.8   The core set of probability models for computing CLV by type of customer base

covariates far into the future for any true CLV or CE calculations. This clearly introduces a lot of additional noise into the exercise. The documented multi-period forecasting performance of these models indicates that the absence of marketing covariate effects is not a fundamental flaw. In fact, we view it as a positive characteristic. The only real downside is that the models cannot be used for any CE-based resource allocation exercise. It is worth noting that as covariates are incorporated, data structures and model estimation issues become more complex; if customers have been targeted with different marketing activities on the basis of their past behavior, we must also account for endogeneity (Schweidel and Knox, 2013).

Reflecting on the gap between what marketing academics create and what managers use, Urban and Karash (1971, p. 62) proposed an evolutionary approach to the development of marketing models: 'The introduction of models as an evolutionary development from simple to more complex but a related one would foster managerial acceptance, encourage an orderly development of data and analysis systems, and reduce risks of failure'. The models considered in the chapter represent an easy first step into the world of developing the models that can be used to make the multi-period forecasts of buyer behavior required for the calculation of CLV and CE.

NOTES

1. See the discussion of scoring models in Fader, Hardie and Lee (2005b) for further issues concerning the use of observation-driven model specifications.
2. This distinction is not the same as the 'lost-for-good' versus 'always-a-share' classification used elsewhere in the literature. See Fader and Hardie (2014) for a discussion of this matter.
3. Further dimensions of possible interest in contractual settings include whether usage while under contract is observed or unobserved, and whether the revenue associated with the contract is known in advance or not known in advance. These factors may be determined by technology and the firm's pricing policies (Ascarza, 2009).
4. These parameters come from fitting the BG model to the first five observations of the 'regular' dataset used in Fader and Hardie (2007a).
5. 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 the researcher focusing on CLV and CE, whose primary goal is to ensure that the phenomenon is captured in the model in order to make accurate forecasts.
6. These are the parameters associated with the CDNOW dataset as reported in Fader, Hardie, and Lee (2005a).
7. This assumption is not as simple as it first may seem. First, we are not assuming that $\bar{v}$ is constant across individuals. For example, the gamma-gamma spend model discussed above explicitly allows for heterogeneity between customers. Second, the assumption of stationarity can be relaxed without too much difficulty. For example, if we assume that $v(t)$ grows over time at a constant rate of $(100 \times g)$ percent per period, we can still use $\bar{v}$ and replace the discount rate d with $d' = (d - g)/(1 + g)$.
8. See http://brucehardie.com/notes/008/; last accessed 9 September 2014.
9. See http://cran.r-project.org/package=BTYD; last accessed 9 September 2014.

REFERENCES

Ascarza, E. (2009), 'Modeling customer behavior in contractual settings', unpublished PhD thesis, University of London.

Bauer, H.H., M. Hammerschmidt and M. Braehler (2003), 'The customer lifetime value concept and its contribution to corporate valuation', *Yearbook of Marketing and Consumer Research, Vol. 1*, Berlin: GfK-Nürnberg, e.V., Duncker & Humblot, pp. 47–67.

Berger, P.D., B. Weinberg and R.C. Hanna (2003), 'Customer lifetime value determination and strategic implications for a cruise-ship company', *Journal of Database Marketing & Customer Strategy Management*, **11**(1), 40–52.

Berry, M.J.A. and G.S. Linoff (2011), *Data Mining Techniques*, 3rd edition, Indianapolis, IN: Wiley Publishing, Inc.

Blattberg, R.C. and J. Deighton (1996), 'Manage marketing by the customer equity test', *Harvard Business Review*, **74**(4), 136–44.

Braun, M. and D.A. Schweidel (2011), 'Modeling customer lifetimes with multiple causes of churn', *Marketing Science*, **30**(5), 881–902.

Colombo, R. and W. Jiang (1999), 'A stochastic RFM model', *Journal of Interactive Marketing*, **13**(3), 2–12.

Cox, D.R. (1981), 'Statistical analysis of time series: some recent developments', *Scandinavian Journal of Statistics*, **8**(2), 93–108.

Dev, C.S. and L.M. Stroock (2007), 'Rosewood Hotels & Resorts: branding to increase customer profitability and lifetime value', *Harvard Business Publishing HBP No. 2088*, 15 June 2007.

Fader, P. (2012), *Customer Centricity: Focus on the Right Customers for Strategic Advantage*, 2nd edition, Philadelphia, PA: Wharton Digital Press.

Fader, P.S. and B.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**(4–5), 461–73.

Fader, P.S. and B.G.S. Hardie (2007a), 'How to project customer retention', *Journal of Interactive Marketing*, **21**(1), 76–90.

Fader, P.S. and B.G.S. Hardie (2007b), 'Fitting the sBG model to multi-cohort data', accessed 9 September 2014 at http://brucehardie.com/notes/017/.

Fader, P.S. and B.G.S. Hardie (2009), 'Probability models for customer-base analysis', *Journal of Interactive Marketing*, **23**(1), 61–9.

Fader, P.S. and B.G.S. Hardie (2010), 'Customer-base valuation in a contractual setting: the perils of ignoring heterogeneity', *Marketing Science*, **29**(1), 85–93.

Fader, P.S. and B.G.S. Hardie (2012), 'Reconciling and clarifying CLV formulas', accessed 9 September 2014 at http://brucehardie.com/notes/024/.

Fader, P.S. and B.G.S. Hardie (2013), 'The gamma-gamma model of monetary value', accessed 9 September 2014 at http://brucehardie.com/notes/025/.

Fader, P.S. and B.G.S. Hardie (2014), 'The Pareto/NBD is not a "lost-for-good" model', accessed 7 September 2014 at http://brucehardie.com/notes/031/.

Fader, P.S., B.G.S. Hardie and C.-Y. Huang (2004), 'A dynamic changepoint model for new product sales forecasting', *Marketing Science*, **23**(1), 50–65.

Fader, P.S., B.G.S. Hardie and K. Jerath (2007), 'Estimating CLV using aggregated data: the Tuscan Lifestyles case revisited', *Journal of Interactive Marketing*, **21**(3), 55–71.

Fader, P.S., B.G.S. Hardie and K.L. Lee (2005a), '"Counting your customers" the easy way: an alternative to the Pareto/NBD model', *Marketing Science*, **24**(2), 275–84.

Fader, P.S., B.G.S. Hardie and K.L. Lee (2005b), 'RFM and CLV: using iso-value curves for customer base analysis', *Journal of Marketing Research*, **42**(4), 415–30.

Fader, P.S., B.G.S. Hardie and Y. Liu (2012), 'Bringing heterogeneity into customer equity calculations', presentation at the 34th ISMS Marketing Science Conference, 7–9 June.

Fader, P.S., B.G.S. Hardie and S. Sen (2014), 'Stochastic models of buyer behavior', in R. Winer and S. Neslin (eds), *The History of Marketing Science*, Singapore: World Scientific Publishing, pp. 165–205.

Fader, P.S., B.G.S. Hardie and J. Shang (2010), 'Customer-base analysis in a discrete-time non-contractual setting', *Marketing Science*, **29**(6), 1086–108.

Farris, P.W., N.T. Bendle, P.E. Pfeifer and D.J. Reibstein (2010), *Marketing Metrics*, 2nd edition, Upper Saddle River, NJ: Wharton School Publishing.

Gupta, S. and D.R. Lehmann (2003), 'Customers as assets', *Journal of Interactive Marketing*, **17**(1), 9–24.

Gupta, S., D.R. Lehmann and J.A. Stuart (2004), 'Valuing customers', *Journal of Marketing Research*, **41**(1), 7–18.

Jerath, K., P.S. Fader and B.G.S. Hardie (2014), 'Customer-base analysis using repeated cross-sectional summary (RCSS) data', accessed 9 September 2014 at http://brucehardie.com/papers/025/.

Kumar, V. and J. Andrew Petersen (2012), *Statistical Methods in Customer Relationship Management*, Chichester, UK: John Wiley & Sons, Ltd.

Kumar, V. and W. Reinartz (2012), *Customer Relationship Management*, 2nd edition, Heidelberg: Springer.

Lemmens, A. and C. Croux (2006), 'Bagging and boosting classification trees to predict churn', *Journal of Marketing Research*, **43**(2), 276–86.

Lerner, S.P., P.S. Fader and B.G.S. Hardie (2012), '"Buy till you die" to improve operational efficiency: realigning the sales force using the BG/NBD', presentation at the 23rd Advanced Research Techniques Forum, 24–27 June.

Libai, B., E. Muller and R. Peres (2009), 'The diffusion of services', *Journal of Marketing Research*, **46**(2), 163–75.

Mason, C.H. (2003), 'Tuscan Lifestyles: assessing customer lifetime value', *Journal of Interactive Marketing*, **17**(4), 54–60.

Moe, W.W. and P.S. Fader (2004), 'Capturing evolving visit behavior in clickstream data', *Journal of Interactive Marketing*, **18**(1), 5–19.

Neslin, S., S. Gupta, W. Kamakura, J. Lu and C.H. Mason (2006), 'Defection detection: measuring and understanding the predictive accuracy of customer churn models', *Journal of Marketing Research*, **46**(2), 204–11.

Ofek, E. (2002), 'Customer profitability and lifetime value', *Harvard Business School Note 9–503–01*, rev. 7 August 2002.

Parr Rud, O. (2001), *Data Mining Cookbook*, New York: John Wiley & Sons.

Pfeifer, P.E. (2011), 'On estimating current-customer equity using company summary data', *Journal of Interactive Marketing*, **25**(1), 1–14.

Pfeifer, P.E., M.E. Haskins and R.M. Conroy (2005), 'Customer lifetime value, customer profitability, and the treatment of acquisition spending', *Journal of Managerial Issues*, **17**(1), 11–25.

Potter, R.G. and M.P. Parker (1964), 'Predicting the time required to conceive', *Population Studies*, **18**(1), 99–116.

Reinartz, W. 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**(4), 17–35.

Reinartz, W. and V. Kumar (2003), 'The impact of customer relationship characteristics on profitable lifetime duration', *Journal of Marketing*, **67**(1), 77–99.

Risselada, H., P.C. Verhoef and T.H.A. Bijmolt (2010), 'Staying power of churn prediction models', *Journal of Interactive Marketing*, **24**(3), 198–208.

Rosset, S., E. Neumann, U. Eick and N. Vatnik (2003), 'Customer lifetime value models for decision support', *Data Mining and Knowledge Discovery*, **7**(3), 321–39.

Schmittlein, D.C. and R.A. Peterson (1994), 'Customer base analysis: an industrial purchase process application', *Marketing Science*, **13**(1), 41–67.

Schmittlein, D.C., D.G. Morrison and R. Colombo (1987), 'Counting your customers: who are they and what will they do next?', *Management Science*, **33**(1), 1–24.

Schweidel, D.A. and G. Knox (2013), 'Incorporating direct marketing activity into latent attrition models', *Marketing Science*, **32**(3), 471–87.

Song, T.H., J.Y. Kim and S.Y. Kim (2013), 'Optimality of customer relationship management: does profitability really matter?', *Asia Marketing Journal*, **15**(3), 141–57.

Urban G.L. and R. Karash (1971), 'Evolutionary model building', *Journal of Marketing Research*, **8**(1), 62–6.

Walter, T. (2013), 'Changing of the guard and thank you', *ION Newsletter*, Winter 2012–13, 2.

Wiesel, T., B. Skiera and J. Villanueva (2008), 'Customer equity: an integral part of financial reporting', *Journal of Marketing*, **72**(2), 1–14.