Original Article

Deterministic and stochastic Customer Lifetime Value models. Evaluating the impact of ignored heterogeneity in non-contractual contexts

Received (in revised form): 31st July 2009

Mihai Calciu
is Associate Professor of Marketing at the Institute of Business Administration – European Institute of Direct Marketing, University of Sciences and Technologies, Lille, France. His preferred subjects are marketing decision support models and systems, interactive marketing, and customer relationship management.

ABSTRACT This article presents a panorama of Customer Lifetime Value (CLV) modelling and focuses on the two main categories of CLV calculation methods: the deterministic approach and the stochastic approach. The first adopts simplified calculations that ignore heterogeneity in customers' retention and/or churn rates within a cohort. It produces formulae that can be easily used by managers and solves a greater number of managerial problems. The second approach brings much more precision to CLV calculations by thoroughly considering retention and/or churn rate heterogeneity, but measures them as actuarial variables and not as responses to marketing effort. We derive formulae to compute expected residual transactions conditional on customer purchase profiles for deterministic models in non-contractual contexts, and suggest solutions to evaluate forecasting error as compared to stochastic models.

Journal of Targeting, Measurement and Analysis for Marketing (2009) **17,** 257–271. doi:10.1057/jt.2009.19; published online 9 November 2009

Keywords: lifetime value; probability models; retention model; migration model; Markov Chains

INTRODUCTION

The diffusion of interactive marketing and of database technologies increases the availability of customer transaction data. This leads, in most contexts, to the adoption of customer and customer portfolio evaluation methods that historically have emerged in the direct marketing and catalogue sales industry. Customer Lifetime Value (CLV) is the discounted (actualised) value of cash flows generated by a customer or a customer cohort during their 'life' with a firm. The term Net Present Value (NPV) has also been used in order to indicate that customer value is a financial measure evaluating an investment (acquisition costs) that produces future cash flows. These cash flows decline with time owing to customers' churn. At the level of an individual customer or of a cohort of customers, CLV is usually calculated as a sum of discounted gains or cash flows ignoring the initial investment or acquisition costs. Blattberg and Deighton[1] coined the term Customer Equity (CE) to designate the CLV from which acquisition costs are subtracted.

Correspondence: Mihai Calciu
Institut d'Administration des Entreprises,
104 av. du Peuple Belge, 59043 Lille, France

The same term has been used more recently in order to designate the NPV of the customer base. As a company's customer base is formed dynamically by a sequence of customer cohorts, in order to avoid confusion, Villanueva and Hanssens[2] suggest using the term Static Customer Equity at individual customer or cohort level and the term Dynamic Customer Equity (DCE) at customer base level. This last measure, according to Gupta *et al*,[3] is, under certain circumstances, a good proxy for the value of a company, as it accounts for both current and future relationships. In order to offer better definitions for the CLV and CE concepts at individual customer (or cohort) level, Pfeifer *et al*[4] introduce the distinction between the value of a just-acquired customer and the value of a new or existing one. If we note CLV as the value of a just-acquired customer (see Table 1), that is, a customer value calculation starting just after the acquisition period, the value of a new or existing customer also includes the margin (m) from the previous period (which for a new customer is the acquisition period). CE subtracts acquisition costs from the value of a new customer (these can be expressed as a fraction between prospect acquisition costs (A) and the acquisition rate (a) that could be achieved as a result of these spendings).

Before presenting CLV calculations and modelling approaches, it is necessary to distinguish between two types of temporal (dynamic) behaviour in a customer's relationship with a company. These have been named differently according to various authors: 'lost for good'/'always has share' (Jackson,[5]) 'retention model'/'migration model' (Dwyer[6]) or 'contractual'/'non-contractual' (Reinartz and Kumar[7]). The *retention model* considers that a person or company remains a customer as long as they generate transactions. The *migration model* considers that customers can reappear (turn up again) after some periods during which they did not make transactions, and traces their probability of 'reactivating'. Our study is organised as follows. After a presentation of wider CLV modelling approaches, two main categories of CLV calculation methods are analysed: the deterministic and the stochastic approach. In the final section, we develop formulae and apply them to investigate the impact of ignoring heterogeneity in non-contractual contexts using two datasets.

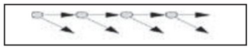
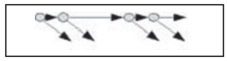
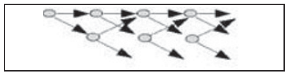
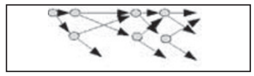
Table 1: Customer value in various client/prospect states (adapted from Pfeifer *et al*[4])

Measure of customer lifetime value	*Calculation*
Value of a just-acquired customer	CLV
Value of a new or existing customer	$m+\text{CLV}$
Customer Equity (CE)	$-A/a+m+\text{CLV}$

A REVIEW OF CLV MODELLING APPROACHES

Jain and Singh,[8] in one of the first and most frequently cited studies in of the CLV literature, identify three main research directions. The first is the development of CLV calculation models that focus on the revenue stream from customers, and on acquisition, retention and other marketing costs in order to facilitate calculations, resource allocation and optimisation of CLV. The second research stream, described as customer base analysis, concentrates on methods that analytically predict the probabilistic value of customer transactions from the existing customer base. The last direction uses analytical models in a normative way, and analyses CLV implications for managerial decisions. As in the meantime literature on the subject has largely increased and approaches have become more diverse, other classification criteria and perspectives have been added. Gupta *et al*,[9] for example, suggest a CLV analysis framework that takes as a starting point the customer acquisition, retention and expansion (cross- and up-selling) effects on customer value. Subsequently, interactions between these effects are analysed first at individual and/or cohort level and then at firm level in order to determine the customer base value of the company. By following this framework in order to structure modelling approaches, the authors also suggest a classification that distinguishes among classical

Table 2: Adequation between activities and/or industries and the typology of customer relationships, examples

Type of customer relationship	Contractual	Industry	Magazine subscriptions, Insurance policies, etc.	Credit cards, Mobile phone usage
		Dynamic		
	Non-contractual	Industry	Catalogue sales, Events attendance, Charity fund drives,	Retail purchases, Doctor visits, Hotel stays
		Dynamic		
			Discrete	Continuous
			Transactions occasions	

Source: Adapted from Fader and Hardie[10](p. 63).

Recency, Frequency and Monetary (RFM) methods, and stochastic (probability), econometric, persistence, computer science, and growth or diffusion models. By developing this same framework in which customer acquisition, retention and expansion through cross-selling and up-selling are seen as determinant of CE, Villanueva and Hanssens[2] classify CE models according to the kind of data sources they deal with, whether these are internal company databases, survey data, company reports, panel data or even data on managerial judgements obtained by decision calculus. The sub-area of CLV modelling approaches that we are interested in may be termed 'CLV calculation models' in the spirit of Jain and Singh.[8] As the diversity of CLV calculation models is rather wide, we adopt a progressive approach in dealing with these models. We start with a simple generic model and increase complexity gradually by adding dynamic customer behaviour contexts (contractual, non-contractual), transaction occasions' chronology aspects (discrete and continuous time), customer heterogeneity (deterministic and stochastic models) and various analysis levels (individual, cohort or customer base). By crossing CLV calculation models on two dimensions, several modelling contexts can be identified, as shown in Table 2.

The basic structural model (Jain and Singh[8]) outlines the financial dimension of CLV:

$$CLV = \sum_{t=1}^{T} \frac{M_t - C_t}{(1+d)^{t-0.5}} \qquad (1)$$

where i = the customer transaction period; M = margin obtained and C = customer incumbent costs in period t; and n = the number of total periods (years) of the expected lifetime of a customer.

The gain in each period is multiplied by a discount factor $1/(1+d)$ at a given power (here $t-0.5$, which tries to fit the mean time between revenue and spending flows). This model is general enough, M_t and C_t, as financial flows, can encapsulate various transactional dynamics which generate them. These flows can be the result of retention and survival rates or probabilities when the context is contractual, but can also result from customer reactivation when the customer relationship context is non-contractual. According to Gupta *et al*[9] CLV differs in two fundamental aspects from the purely financial NPV measure. First, it is estimated at an individual level or segment level, and second, it explicitly integrates possible future customer attrition.

DETERMINISTIC AND STOCHASTIC CLV MODELS

The focus of this study is deterministic and stochastic models belonging to the so-called 'buy till you die' framework (Schmittlein *et al*;[11] Fader and Hardie[10]), which is probably one of the most important research streams in CLV modelling at present. Both assume individually constant buy and die probabilities for customers. While deterministic models are suited for individual CLV calculations, stochastic models should be used when computing CLV at an aggregated level (customer cohorts or customer base), as they integrate heterogeneity of buy and/or die probabilities among individuals.

Deterministic retention models

More simple deterministic retention models are a good basis to introduce some fundamental aspects of CLV calculations. By ignoring the temporal shift between retention gains and spendings and regrouping them in order to form the net gain (g), which consists of the difference between margin (m) and cost (c), and remains constant for a customer over time the CLV calculation formula can become very concise. For a just-acquired customer it is:

$$CLV = \sum_{t=1}^{\infty} g \frac{r^t}{(1+d)^t} = g \frac{r}{1+d-r} \qquad (2)$$

where r is the retention rate and $r/(1+d-r)$, according to Gupta and Lehman,[12] is the 'margin multiple'. Gupta and Lehman[13] have shown that if gains increase with a constant rate q, the margin multiple becomes $r/[1+d-r(1+q)]$. By separating monetary flows from transaction flows, CLV can be expressed as the product of gains (here constant) and what has been termed by Calciu and Salerno[14] 'discounted expected number of transactions', or, more simply, 'discounted expected transactions' (DET). In other words, $CLV = g \times DET$. *DET* can be seen as the CLV for transactions producing a one-monetary unit gain ($g = 1$), and can be used as a proxy for CLV. *DET for a just-acquired customer* is the margin multiple $r/(1+d-r)$, as can be seen from formula (2). The DET for a new or existing customer is computed by adding an initial transaction (see Table 3). In order to predict future purchasing patterns by those customers listed in the firm's transaction database, and to generate estimates of their expected lifetime value, Fader *et al*[15] introduce the concept of 'as-yet-to-be-acquired' customers. A yet-to-be-acquired customer is seen as an existing customer who is going to be retained (or acquired), which means that the DET of such a customer is the DET for an existing customer multiplied by the retention rate (see Table 3).

Table 3: Discounted Expected Transactions (DET) calculations retention models

Measure of discounted expected transactions (DET)	*Calculation*
DET of a just-acquired customer	$DET_{\text{just acquired}} = r/(1+d-r)$
DET of a new or existing customer	$DET_{\text{new or existing}} = 1 + DET_{\text{just acquired}} = (1+d)/(1+d-r)$
DET of a yet-to-be-acquired customer (DERT)	$DET_{\text{yet-to-be-acquired}} = r\ DET_{\text{new or existing}} = r(1+d)/(1+d-r)$

The DET for a customer 'yet-to-be-acquired' at a given moment n, just before the decision as to whether to repurchase (or renew the contract), in fact represents the Discounted Expected Residual Transactions (DERT) from that moment on. It allows one to compute the Discounted Expected Residual Lifetime (DERL) value (Fader and Hardie[16]) of existing customers, for which it is a proxy. DERT is well adapted for situations in which the retention rates while constant individually are heterogeneous within a cohort, as it can be shown that in such situations aggregated retention rate increases over time. Deterministic retention models have been extended by Gupta *et al*[3] in order to measure the value of the customer base of a company or DCE, which is defined as the CLV of current and future customers. DCE calculated as the sum of all cohorts, taking into account new customer acquisitions accounted for on an annual basis, can be obtained using the following discrete time[17] formula:

$$DCE = \sum_{k=0}^{\infty} \frac{n_k}{(1+d)^k} \times \left(\sum_{t=k}^{\infty} m_{t-k} \frac{r^{t-k}}{(1+d)^{t-k}} - c_k \right) \quad (3)$$

The initial number of customers in each cohort is determined using a diffusion growth model.

Although, as has been shown in this section determinist models produce very useful managerial applications, when computing CLV at aggregated cohort or multi-cohort level, they ignore heterogeneity of individual customer response probabilities. This aspect is discussed later in this article. There is a fundamental marketing belief that companies can influence customer response, meaning that they can 'control' customer acquisition and retention rates by varying marketing efforts. This implies that it is also possible to optimally allocate marketing retention and acquisition efforts. Formula 5 with retention rates expressed as a function of the retention budget ($r=f(R)$) becomes:

$$CLV = \left(m - \frac{R}{r} \right) \frac{r}{1+d-r} = \left(m - \frac{R}{f(R)} \right) \frac{f(R)}{1+d-f(R)} \quad (4)$$

The gain (g) is composed of the margin minus the average retention cost. It is expressed here as the fraction of spendings per targeted customer (R) and achieved retention rate (r). This formula helps to find optimal retention spendings, and is part of the elegant Blattberg and Deighton[1] model that also deals with optimal customer acquisition spendings using formula 5.

$$CE = -A/a + m + CLV \quad (5)$$

This model has recently been modified by Pfeifer.[18]

Deterministic migration models

The 'always a share' behaviour is the alternate scheme to 'lost for good' in what is known as a dichotomy. Here, customer value comes not only from surviving customers, but also from customers allowed to reactivate after a given number of inactive periods. Customers are considered 'lost for good' only after exceeding that number of successive periods of inactivity. By reducing the tolerated number of successive periods of inactivity to zero, the 'always a share' model reduces to the 'lost for good' model, which can be seen as a special case of this more general model. In non-contractual settings, firms cannot know when a customer becomes inactive. Intuitively they can apply rules (conventions) based on the RFM amount of past purchases in order to decide whether or not a customer is still active. By fixing RFM states, based on past behaviour, transition probabilities from one state to another can be computed and organised into a matrix of transition probabilities in order to form a Markov Chain. The migration scheme that controls the construction of the transition probabilities matrix if we use only purchase recency is shown in Figure 1. It indicates that a customer in recency t can either buy and pass in recency 1 or not buy and pass in an inferior recency state $t+1$.

A detailed discussion of the matrix approach applied to customer migration can be found in the study by Pfeifer and Carraway.[19] One can also find there a detailed migration scheme between RFM states and a graphical representation of transitions in and out of a given RFM state. A practical application is given further along in this article. The matrix approach based on Markov Chains in order to compute customer value has been used by Bitran and Mondschein[20] and Pfeifer and Carraway[19] using RFM variables in order to define transition states. Rust *et al*[21] have defined transition probabilities matrices between brands that vary over time

R(1) <------R(t)-------->R(t+1)

Figure 1: Migration scheme between Recency States.

Table 4: Discounted Expected Transactions calculations for migration models

Measure of discounted expected transactions (DET)	*Calculation*
DET of a just-acquired customer	$DET_{\text{just acquired}} = [(\boldsymbol{I} - \boldsymbol{P}/(1+d))^{-1} - \boldsymbol{I}]$
DET of a new or existing customer	$DET_{\text{new or existing}} = \mathbf{I} + DET_{\text{just acquired}} = (\boldsymbol{I} - \boldsymbol{P}/(1+d))^{-1}$
DET of a yet-to-be-acquired customer (DERT)	$DET_{\text{yet-to-be-acquired}} = \mathbf{P}\ DET_{\text{new or existing}} = \boldsymbol{P}(\boldsymbol{I} - \boldsymbol{P}/(1+d))^{-1}$

following a logit model. Using the analogy between the retention probability (r) and the transition probabilities matrix (P), it is easy to derive equivalent matrix formulae for DET, as shown in Table 4.

While the formula for new and existing customers has been suggested by Pfeifer and Carraway,[19] the other formulae are a contribution of our research. They are aimed to compare deterministic and stochastic approaches to CLV calculations.

STOCHASTIC CLV MODELS

Probability or stochastic models consider observed behaviour the result of an underlying stochastic process controlled by unobserved latent characteristics that vary across individuals. Gupta *et al*[9] attempt to find a simple paramorphic representation that describes and predicts observed behaviour instead of trying to explain differences between explained and observed behaviour as a function of covariables (as in regression models). They assume that a given behaviour varies across the population according to a probability law. Their main advantage is that they take into account and measure heterogeneity in observed behaviour. Their main disadvantage is that by measuring the characteristics of these probability laws, they assume that it is only these laws that control observed behaviour, and do not leave any place for marketing variables, as if the dynamic customer behaviour cannot be changed by marketing activities.[22] One of the most respected stochastic CLV models, and a forerunner of recent developments, is the Pareto/Negative Binomial Distribution (NBD) model developed by Schmittlein *et al.*[11] This is the benchmark model for non-contractual settings where transactions can occur at any moment over time. It is not well suited for purchase situations that occur at fixed time intervals, and even less suited for contractual settings.

The series of relatively recent developments undertaken by authors such as Fader and Hardie, which have begun with a modification of the Pareto/NBD the BG/NBD model (Fader *et al*[23]) that substantially simplifies parameter estimation, have led these authors to progressively cover all dynamic customer behaviour contexts presented in Table 2 with stochastic models matching each context and its corresponding industries. The names of these models, their authors and the years of publication are presented in Table 5. If we take only the prototype models (marked with bold characters) for each case in Table 5 and try to synthesise them, we could say that they use beta distributions for discrete time models and gamma distributions for continuous time models as mixing distributions that account for the heterogeneity in buy and/or die probabilities. For the simpler contractual case, the individual buying process representing the number of purchases during a given period[24] for a given buying probability[25] follows a shifted geometric distribution in the discrete time case and an exponential distribution in the continuous case. In the non-contractual settings, the buy and die probabilities are distinct, and the probability distributions that represent the buying and dying processes need to be integrated as sub-processes in order to represent the combined behaviour. It follows that for the discrete case the individual buying process follows a Bernoulli distribution and the dying process a geometric distribution, while in the continuous case the buying process is a Poisson distribution and the dying process is exponential. For most of these models, closed-form expressions have been derived for the CLV and for other useful proxy measures, such as DET and DERT. For example, the DERT for the stochastic migration model in discrete time the

Table 5: Stochastic Models for various dynamic customer behaviour contexts

Type of customer relationship	Contractual	Shifted Betageometric (**sBGD**-Fader and Hardie[33,34]) Beta-discrete-Weibull (BdWD – Fader and Hardie[33])	Exponential Gamma (**EGD** or Pareto Distribution of second kind)
	Non-contractual	Betageometric Betabinomial (**BG/BBD** – Fader, Hardie and Berger[15])	**Pareto/NBD** (Schmittlein *et al*[11]), Betageometric/NBD (Fader *et al*[23]), Modified Betageometric/NBD (Batislam *et al*[26])
		Discrete Transaction occasions	Continuous

beta-geometric/beta-binomial (BG/BB) distribution is given by the following formula:

$$\frac{B(\alpha+x+1,\beta+n-x)}{B(\alpha,\beta)}\frac{B(\gamma,\delta+n+1)}{B(\gamma,\delta)}$$
$$_2F_1(1,\delta+n+1,\gamma+\delta+n+1,1/(1+d))/L(\alpha,\beta,\gamma,\delta\,|x,n,m) \tag{6}$$

where (x, n, m) represent purchase history with x = number of purchases, m = recency and n = the current period; $(\alpha, \beta, \delta, \gamma)$ are the estimated beta distribution parameters defining the heterogeneous buying and dying probabilities; and $B(\cdot)$ is the Beta function, $_2F_1(\cdot)$ is the Gaussian hypergeometric function and $L(\cdot)$ is the Likelihood. This formula will be used later in this paper in a comparison with the deterministic migration model. Research on stochastic models for customer database analysis is very active. Existing models are modified or extended; for example Fader and Hardie (2005a) allow for the incorporation of time-invariant covariate effects into Pareto/NBD and BG/NBD models. The BG/NBD model developed in 2005 as an alternative to the Pareto/NBD model has itself been modified and extended by several researchers (Batislam *et al*;[26,27] Fader and Hardie;[16] Hoppe and Wagner;[28] Wagner and Hoppe[29]). New methods of carrying out parameter estimation and integration of stochastic models are being explored. Several researchers use Markov Chain Monte Carlo (MCMC) methods for this purpose. For example, Ma and Liu[30] and Abe[31] use the Gibbs sampler for parameter estimation, while Singh *et al*[32] attempt to build a generalised framework for estimating CLV when lifetimes are not observed using the Metropolis Hastings algorithm.

INVESTIGATING THE IMPACT OF HETEROGENEITY ON CLV

In their yet unpublished article entitled 'Customer-Base Valuation in a Contractual Setting:The Perils of Ignoring Heterogeneity', Fader and Hardie[33] show that models that assume constant retention at cohort or firm level are wrong, as typically retention rates that can be constant individually increase over time at aggregate cohort level. In an earlier version of the article, the authors warned managers and researchers not to use such models, and suggested that academics not teach formulae for computing CLV based on them. This is a rather strong point of view that seems to wipe out a substantial area of study that has been dominant in the early stages of research in CLV literature. We agree with all arguments showing that ignoring heterogeneity in customer response rates is producing erroneous results, and show how to extend the exploration of the precision gap between those incriminated models and correct stochastic models in a non-contractual setting. But by evaluating this precision gap, we do not go so far as to ban those homogeneous response models and calculation formulae, and seek in this way a means to retaining them, as we consider them more flexible, easier to use and understand, and better at integrating marketing mix.

To understand why deterministic models of CLV applied to aggregated customer behaviour can be wrong, let us consider the simplest case (see formula 2) of a deterministic retention model and compute CLV for a cohort composed at acquisition time of two equal-sized segments of customers with homogeneous but very different retention rates, $r_1 = 0.9$ and $r_2 = 0.3$. By ignoring this heterogeneity, we would consider the average retention rate of the cohort at acquisition time as $(r_1 + r_2)/2 = (0.9 + 0.3)/2 = 0.6$. By applying formula 2 to a 1-euro gain and a discount rate $d = 0$, we would obtain for a just-acquired customer a $CLV = 1 \times 0.6/(1 + 0 - 0.6) =$ 1.5 euros. This is very different from the 9-euros CLV that would be obtained if the cohort consisted only of individuals from segment 1, and also from the *0.4*-euro CLV that would be obtained if the cohort consisted only of individuals from segment 2.

The *1.5*-euros CLV computed by taking the mixed cohort's average retention rate is false, as over time the higher retention rate segment becomes quickly predominant in the cohort. The probability of belonging to this segment increases rapidly from 0.5 at acquisition time towards 1. This makes the mixed cohort retention rate evolve towards the upper limit of 0.9, so that expected transactions in later periods are close to those of segment 1. This fact can be demonstrated using Bayes' Theorem, according to which the probability of belonging to segment 1 after t periods is a conditional probability that can be computed in the following way:

$$P(\mathit{segment}\,1 \mid \mathit{repurchase}\; t\; \mathit{times}) = \frac{P(\mathit{repurchase}\; t\; \mathit{times} \mid \mathit{segment}\,1) \times P(\mathit{segment}\,1)}{P(\mathit{repurchase}\; t\; \mathit{times})} \tag{7}$$

If the probability that a customer belongs to segment 1 after one period is $0.9^1 \times 0.5/(0.9^1 \times 0.5 + 0.3^1 \times 0.5) = 0.75$, after 2 periods it becomes $0.9^2 \times 0.5/(0.9^2 \times 0.5 + 0.3^2 \times 0.5) = 0.9$ and rapidly approaches 1.

By applying this formula to compute probabilities of belonging to both segments for all future periods, and for simplicity reasons, using a zero discount rate we can compute the DET of a just-acquired customer as a proxy for CLV for the case where the heterogeneity of the cohort is taken into account and for the case where it is ignored.

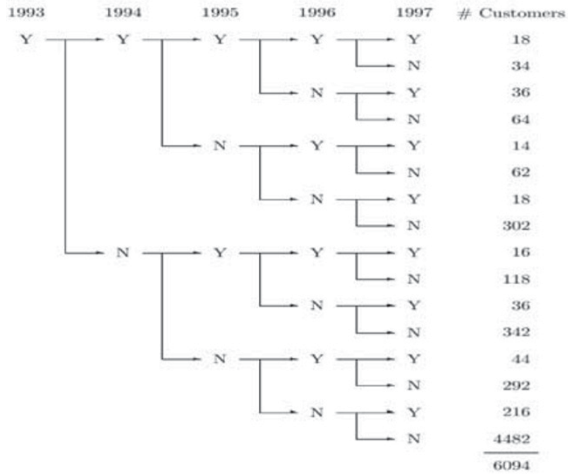
DET with heterogeneity taken into account $=$ $(0.5 \times 0.9 + 0.5 \times 0.3) + (0.75 \times 0.9^2 + 0.25 \times 0.3^2) +$ $(0.9 \times 0.9^3 + 0.1 \times 0.3^3) + \ldots + 1 \times 0.9^n = 4.71$

DET with heterogeneity ignored $= 0.6 + 0.6^2 +$ $0.6^3 + \ldots + 0.6^n = 1.5$

This shows clearly that by ignoring, heterogeneity CLV is systematically undervalued.

A thorough investigation of the impact of heterogeneity on CLV in contractual settings can be found in the study by Fader and Hardie.[33] They use the two datasets from a previous version of their paper introducing the Shifted Beta Geometric distribution model (Fader and Hardie[34]) in order to compare the latter with the fixed retention rate model CLV calculation given in formula 2. In order to extend the investigation of the impact of heterogeneity to non-contractual situations, we use the formula we have derived in Table 4 for computing DERT for a yet-to-be-acquired customer in order to produce lifetime value calculations for a customer dynamic behaviour where buying and 'death' probabilities are individually constant over time and 'wrongly' considered homogenous within a cohort. Additional matrix formulations for expected residual transactions at various discrete time periods are used in order to observe the increasing precision gap between this model that ignores heterogeneity and the corresponding stochastic model for non-contractual settings: the BG/BB model (Fader *et al*[15]). In our investigation we also use two datasets. The first is the Cruiseship dataset used in the paper introducing the BG/BB model (Fader *et al*[15]), consisting of cruise-line transactions for a cohort of 6094 customers over a period of 5 years. The second is a catalogue sales dataset that for confidentiality reasons we will call 'Brabant'. It consists of catalogue orders for a cohort of 36882 customers over seven seasons. As opposed to the first dataset in the second, the buying behaviour shows high seasonal effects with significantly more customers buying in autumn and fewer customers buying in spring. As the BG/BB model is not yet prepared to deal with such effects,[35] this opportunity is used in order to point out some limitations that these stochastic models can have in dealing with real-world situations (see Appendix).

a



b

	r3f3	r3f2	r3f1	r2f3	r2f2	r2f1	r1f3	r1f2	r1f1	r0
r3f3	0.35	0	0	0.65	0	0	0	0	0	0
r3f2	0.14	0	0	0	0.86	0	0	0	0	0
r3f1	0	0.13	0	0	0	0.87	0	0	0	0
r2f3	0	0	0	0	0	0	0	0	0	0
r2f2	0.36	0	0	0	0	0	0	0.64	0	0
r2f1	0	0.10	0	0	0	0	0	0	0.90	0
r1f3	0	0	0	0	0	0	0	0	0	0
r1f2	0	0	0	0	0	0	0	0	0	0
r1f1	0	0.06	0	0	0	0	0	0	0	0.94
r0	0	0	0	0	0	0	0	0	0	1.00

Figure 2: Building the transition matrix from observed buying behaviour (Cruiseship dataset).
Source: (**a**) Observed buying behaviour Fader *et al*, 2004, p. 10. (**b**) RF Transition Matrix (P).

Table 6: Expected transactions calculations for non-contractual settings (Cruiseship dataset)

Recency, Frequency states	*Expected transactions after 10 periods*		*Expected transactions (long term)*		*Discounted Expected Transactions (DET)*	
	determinist	*stochastic*	*determinist*	*stochastic*	*determinist*	*stochastic*
r1f3	0.53	2.05	0.53	3.57	0.51	1.77
r1f2	0.69	1.52	0.69	2.65	0.62	1.31
r1f1	0.44	0.97	0.44	1.68	0.38	0.83
r2f2	0.55	1.28	0.55	2.23	0.53	1.10
r2f1	0.25	0.85	0.25	1.47	0.22	0.73
r3f1	0.10	0.72	0.10	1.25	0.09	0.62
r4	0.00	0.28	0.00	0.48	0	0.24

The main contribution of this exercise is to adapt the homogeneous and constant transition probabilities matrix approach to make it comparable to the corresponding stochastic approach. We use the decision tree recording whether customers took cruises each year (Figure 2a) to compute the transition probabilities matrix (Figure 2b). By applying expected residual transactions calculations to both the deterministic migration model and the stochastic model, the results in Table 6 and Figure 3 are obtained. From Table 6 it can be seen that the model ignoring heterogeneity undervalues the long-term expected transactions and the DET, together with its proxy the CLV.

Figure 3 also shows that this undervaluation of the expected transactions is increasing with the number of transaction periods. The same result is obtained for the catalogue sales dataset. Models ignoring heterogeneity fail to recognise the sorting effect taking place in a heterogeneous cohort that makes individuals with low buying probability and higher probability of 'dying' leave the cohort earlier, so that over time the cohort progressively retains customers with higher repeat buy probabilities and lower probabilities of dying. Therefore, those models produce a downward-biased estimation of expected transactions and lifetime value.

The heterogeneity measures of stochastic models can also be used to evaluate the size of the downward bias induced by deterministic models. In discrete time models, the mixing distribution used to account for heterogeneity is

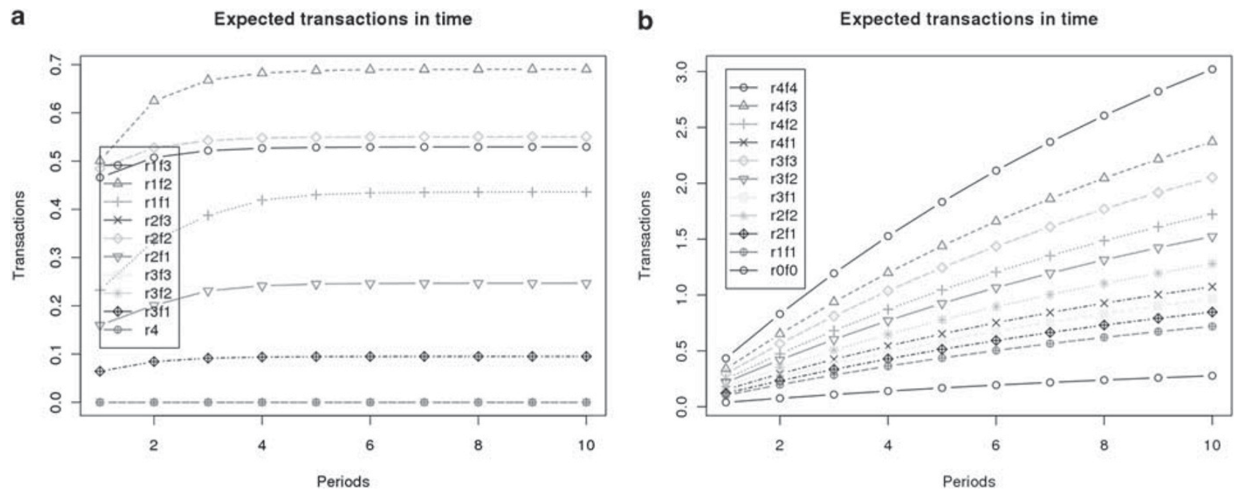


Figure 3: Evolution of residual expected transactions using determinist and stochastic models (Cruiseship dataset). (**a**) determinist migration model. (**b**) BG/BB Model.

the beta distribution. Its two parameters α and β can be characterised in terms of the mean $\mu = E(p) = \alpha/(\alpha+\beta)$ and the polarisation index $\phi = 1/(\alpha+\beta+1)$. The polarisation index is a measure of heterogeneity. When α, β tend towards zero, the polarisation index tends towards 1 and the beta distribution of the studied probability p becomes U-shaped. Its values are concentrated near $p=0$ and $p=1$, meaning high heterogeneity. Conversely, when these parameters are large (α and $\beta \to \infty$), the polarisation index tends towards zero and the beta distribution becomes a spike at its mean. In the case of the Cruiseship dataset, beta distributions parameters for the purchase probability are $\alpha = 0.657$, $\beta = 5.193$, indicating a mean buying probability of $\alpha/(\alpha+\beta) = 0.11$ and a polarisation $\phi = 1/(\alpha+\beta+1) = 0.146$. The corresponding beta distribution parameters for the probability of 'dying' $\gamma = 173.761$, $\delta = 1882.928$ indicate that as regards the dying process this cohort is highly homogeneous with a polarisation index near to zero and a spike at the mean probability of 0.084. In the case of the catalogue sales dataset, we have a higher mean buying probability of 0.26 and a higher polarisation index 0.216, indicating more heterogeneity. The customer death process is much less intense, with a mean probability of 0.02 but a higher polarisation (0.18) than in the cruiseship case. This explains why long-term expected transactions and the DET measure for CLV are significantly higher for the catalogue sales dataset. Further investigations of heterogeneity effects on long-term expected transactions and customer value can be made for each of the dynamic processes that are combined in the non-contractual settings: the buying process and the death process. A good illustration limited to only to the buying process is given in the study by Fader and Hardie (2006).

DISCUSSIONS AND FURTHER RESEARCH

The main purpose of this investigation of the impact of heterogeneity in non-contractual settings was to show that it was possible to evaluate this impact and eventually how this could be carried out. Obtaining a high degree of precision in such an exercise remains a future research objective. While the parameter estimation methodology from purchase history data is well defined for the stochastic migration model, this is not yet the case for the 'deterministic' Markov Chain model. From the few studies mentioned above that have applied Markov Chain transition matrix-based migration models, none seems to have suggested a particular methodology for inferring transition probabilities from real data.

A potential solution to the problem of truncated transition probabilities would be to infer from available data some causal relationships among purchase probabilities, recency and frequency.

In general, using deterministic CLV formulae on aggregated cohort data is error-prone and should be avoided. Nevertheless, there can be situations when a cohort's buying behaviour is homogeneous. As segmentation seeking groups of customers with a homogeneous behaviour is a fundamental marketing activity, there must be situations where cohorts are highly homogeneous. Even in such not particularly segmented data as the cruiseship dataset, the observed cohort was highly homogeneous with regard to the death rate, as indicated by the spike in the estimated beta distribution. In such situations deterministic models could be used without particular precautions. The use of stochastic models as precision benchmarks in CLV calculation has been questioned in a recent study by Wübben and Wangenheim (2008)[36] . These authors have compared the predictive performance of stochastic non-contractual models (Pareto/NBD and BG/NBD) to that of simple managerial heuristics based on three aspects: (1) distinguish active customers from inactive customers, (2) generate transaction forecasts for individual customers and determine future best customers, and (3) predict the purchase volume of the entire customer base. They found that the simple heuristics perform at least as well as the stochastic models with regard to all managerially relevant areas, except for predictions regarding future purchases at overall customer base level. The heuristics used were (1) the simple recency hiatus to distinguish between active and inactive customers, (2) the past 10 per cent best customers in a customer base will also be the future 10 per cent and (3) every customer continues to buy at his or her past mean purchase frequency. Although as compared to the simple heuristics, that extrapolates individual past purchases in the future, stochastic models give better predictions of the volume at overall customer base level (which is equivalent to CE), it is not sure whether stochastic models would have outperformed well tuned deterministic models under the same conditions. This doubt is implicitly expressed by Wübben and Wangenheim[36] when they suggest that 'it would be worthwhile to benchmark these models against the approach that Gupta *et al* (2004) use'. These deterministic models provide flexible and easy-to-use formulae that could be substituted with simple heuristics[36] and can be easily adapted by managers than, with the more sophisticated stochastic models. All this justifies efforts towards making stochastic and deterministic CLV calculation models comparable.

CONCLUSIONS

This study reviews CLV modelling approaches. It covers CLV computations at customer, cohort and company customer base level. The managerial problems dealt with include calculation aspects, CLV optimisation, and optimal balance between customer acquisition and retention spendings. In this study, several concepts and formulae have been newly organised and some new formulae have been derived. By introducing a clear distinction between CE at individual and cohort level as opposed to the customer base level, we help to disambiguate this term. The concept of expected residual transactions (Fader *et al*, 2005b)[37] after a given time period and for the long term is formalised for deterministic non-contractual settings, allowing for comparisons with corresponding deterministic and stochastic models.

The focus of the study is deterministic and stochastic models belonging to the so-called 'buy till you die' framework (Schmittlein *et al*, 1987; Fader and Hardie, 2009), which is an active research stream in CLV modelling at present. Both assume, in most cases, individually constant buy and die probabilities for customers. The deterministic models, ignoring heterogeneity in buy and die probabilities that normally characterises cohorts of customers, are less sophisticated, easier to develop and implement. Consequently, they offer solutions to a larger area of managerial problems. An overview of these solutions is given in the article in order to outline problems that cannot yet be solved with stochastic models. Deterministic models are also easier to understand by managers, as they are conceptually closer to simple heuristics.

While recognising the conceptual superiority and the quality of the measurement attained through stochastic models, we cannot overlook some difficulties in their estimation, and, at this stage, their limitations with regard to integration of marketing mix variables and sometimes their lack of flexibility when adapting to real-world situations. While some recent studies attempt to integrate stochastic models into more flexible frameworks, for example using MCMC methods for parameter estimation, these models still require extensive development in order to achieve the flexibility of deterministic models. This concern motivated us to find ways to explore the impact of heterogeneity in non-contractual settings. Using the Markov Chain approach to represent a deterministic customer migration model, we derive formulae for computing the residual expected transactions after a given time period and for the long term, which in their discounted form are proxies for residual lifetime value calculations. In this way, during each (discrete time) period of the lifetime of a cohort, the increasing precision gap between the deterministic migration model that ignores heterogeneity within a cohort and the corresponding BG/BB stochastic model can be evaluated. Illustrations have been given for two datasets. The observed deviations are rather high. This justifies banning the use of deterministic models in CLV calculations at aggregated levels, for example cohorts or company customer bases in non-contractual settings, as suggested by Fader and Hardie, who investigated the impact of heterogeneity on CLV in contractual settings. Nevertheless, we also see in the possibility of evaluating the precision gap a potential means to using deterministic models as a proxy for stochastic models in order to offer 'satisficing' solutions to managerial problems that cannot yet be efficiently solved with stochastic models.

REFERENCES AND NOTES

1 Blattberg, R.C. and Deighton, J. (1996) Manage marketing by the customer equity test. *Harvard Business Review* 74(4): 136–144.

2 Villanueva, J. and Hanssens, D. (2007) Customer equity: Measurement, management and research opportunities. *Foundations and Trends® in Marketing* 1(1): 1–95, http://dx.doi.org/10.1561/1700000002.

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

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

5 Jackson, B.B. (1985) *Winning and Keeping Industrial Customers: The Dynamics of Customer Relationships*. Lexington, MA: Lexington Books.

6 Dwyer, F.R. (1989) Customer lifetime valuation to support marketing decision making. *Journal of Direct Marketing* 9(1): 79–84.

7 Reinartz, W.J. and Kumar, V. (2000) On the profitability of long-life customers in a noncontractual setting: An empirical investigation and implications for marketing. *Journal of Marketing* 64(October): 17–35.

8 Jain, D. and Singh, S.S. (2002) Customer lifetime value research in marketing. A review and future directions. *Journal of Interactive Marketing* 16(2): 34–46.

9 Gupta, S. *et al* (2006) Modeling customer lifetime value. *Journal of Service Research* 9(2): 139–155.

10 Fader, P.S. and Hardie, B.G.S. (2009) Probability models for customer-base analysis. *Journal of Interactive Marketing* 23: 61–69.

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

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

13 Gupta, S. and Lehmann, D.R. (2005) *Managing Customers as Investments*. Philadelphia, PA: Wharton School Publishing.

14 Calciu, M. and Salerno, F. (2002) Customer value modeling: Synthesis and extension proposals. *Journal of Targeting, Measurement and Analysis for Marketing* 11(2): 124–147.

15 Fader, P.S., Hardie, B.G.S. and Berger, P.D. (2004) Customer-Base Analysis with Discrete-Time Transaction Data. unpublished working paper.

16 Fader, P.S. and Hardie, B.G.S. (2007a) Incorporating time-invariant covariates into the Pareto/NBD and BG/NBD models, http://brucehardie.com/notes/019.

17 Where *nk* is the number of customers in cohort *k*. The calculation for continuous time is:

$$DCE = \int_{k=0}^{\infty}\int_{t=k}^{\infty} n_k e^{-dk} m_{t-k} e^{-\frac{1+d-r}{r}(t-k)} \times dt\, dk - \int_{k=0}^{\infty} n_k e^{-dk} c_k\, dk$$

18 Pfeifer, P.E. (2005) The optimal ratio of acquisition and retention costs. *Journal of Targeting, Measurement and Analysis for Marketing* 13(2): 179–188.

19 Pfeifer, P.K. and Carraway, R. (2000) Modelling customer relationships as Markov chains. *Journal of Interactive Marketing* 14(2): 43–55.

20 Bitran, G. and Mondschein, S. (1996) Mailing decisions in the catalog sales industry. *Management Science* 42(9): 1364–1381.

21 Rust, R.T., Lemon, K. and Zeithaml, V. (2004) Return on marketing: Using customer equity to focus marketing strategy. *Journal of Marketing* 68(1): 109–126.

22 There are several studies that try to extend these models in order to integrate marketing covariates. Schweidel *et al*[24] use a mixed effects hazard model with both fixed and random components. Strictly speaking, such models exceed the definition that is commonly given to stochastic models, and should be classified as econometric models.

23 Fader, P.S., Hardie, B.G.S. and Lee, K.L. (2005a) Counting your customers' the easy way: An alternative to the Pareto/NBD Model. *Marketing Science* 24(Spring): 275–284.

24 Schweidel, D.A., Fader, P.S. and Bradlow, E.T. (2008) Understanding service retention within and across cohorts using limited information. *Journal of Marketing* 72(January): 82–94.

25 Which in contractual settings is equivalent to the $(1 - \text{die})$ probability.

26 Batislam, E.P., Denizel, M. and Filiztekin, A. (2007) Empirical validation and comparison of models for customer base analysis. *International Journal of Research in Marketing* 24(September): 201–209.

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

28 Hoppe, D. and Wagner, U. (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.

29 Wagner, U. and Hoppe, D. (2008) Erratum on the MBG/NBD Model. *International Journal of Research in Marketing* 25(September): 225–226.

30 Ma, S.-H. and Liu, J.-L. (2007) The MCMC approach for solving the Pareto/NBD Model and possible extensions, ICNC 2007. Third International Conference on Natural Computation, Vol. 2; 24–27 August, pp. 505–512.

31 Abe, M. (2009) Counting your customers' One by one: A hierarchical Bayes extension to the Pareto/NBD Model. *Marketing Science* 28(3): 541–553.

32 Singh, S.S., Borle, S. and Jain, D.C. (2007) A generalized framework for estimating customer lifetime value when customer lifetimes are not observed. (31 October). Available at SSRN: http://ssrn.com/abstract=1154709.

33 Fader, P.S. and Hardie, B.G.S. (2006) Customer-base valuation in a contractual setting: The perils of ignoring heterogeneity, http://brucehardie.com/papers/022/.

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

35 Extending the BG/BB model in order to integrate seasonality effects is a future research direction that we have begun to explore.

36 A more detailed discussion of managerial customer base analysis heuristics, and of the reasons why managers tend to use them instead of adopting more sophisticated academic methods, can be found in Wübben and Wangenheim (2008) Instant customer base analysis: Managerial heuristics often 'get it right'. *Journal of Marketing* 72(May): 82–93.

37 Fader, P.S., Hardie, B.G.S. and Lee, K.L. (2005b) RFM and CLV: Using iso-value curves for customer base analysis, *Journal of Marketing Research* 42(November): 415–430.

APPENDIX

Comparison of determinist and stochastic calculations of expected customer transactions for the Brabant dataset

See Table A1 and Figure A1.

Table A1: Expected transactions calculations for non-contractual settings (Brabant dataset)

Recency, Frequency states	*Expected transactions after 10 periods*		*Expected transactions (long term)*		*Discounted Expected Transactions (DET)*	
	determinist	*stochastic*	*determinist*	*stochastic*	*determinist*	*stochastic*
r1f5	3.62	6.9	3.95	8.13	2.9	7.41
r1f4	3.97	5.91	4.35	6.96	3.13	6.35
r1f3	4.14	4.91	4.62	5.79	3.17	5.28
r1f2	4.1	3.92	4.71	4.62	3.03	4.21
r1f1	3.63	2.93	4.44	3.45	2.62	3.15
r2f5	0	1.94	0	2.28	0	2.08
r2f4	2.97	5.7	3.24	6.71	2.39	6.12
r2f3	3.48	4.79	3.84	5.64	2.71	5.14
r2f2	3.68	3.84	4.18	4.52	2.77	4.13
r2f1	3.28	2.88	3.94	3.39	2.39	3.09
r3f5	0	1.91	0	2.25	0	2.05
r3f4	0	4.41	0	5.2	0	4.74
r3f3	1.69	3.66	1.86	4.31	1.33	3.93
r3f2	2.4	2.79	2.71	3.29	1.81	3
r3f1	2.56	1.87	3.05	2.2	1.87	2.01
r4f5	0	3.23	0	3.81	0	3.48
r4f4	0	2.62	0	3.09	0	2.82
r4f3	0	1.8	0	2.12	0	1.94
r4f2	1.31	2.29	1.47	2.69	1.01	2.46
r4f1	1.91	1.7	2.23	2	1.42	1.83
r5f5	0	1.52	0	1.79	0	1.63
r5f4	0	0.78	0	0.92	0	0.84
r5f3	0	0	0	0	0	0
r5f2	0	0	0	0	0	0
r5f1	0.87	0	1	0	0.66	0
r6	0	0	0	0	0	0

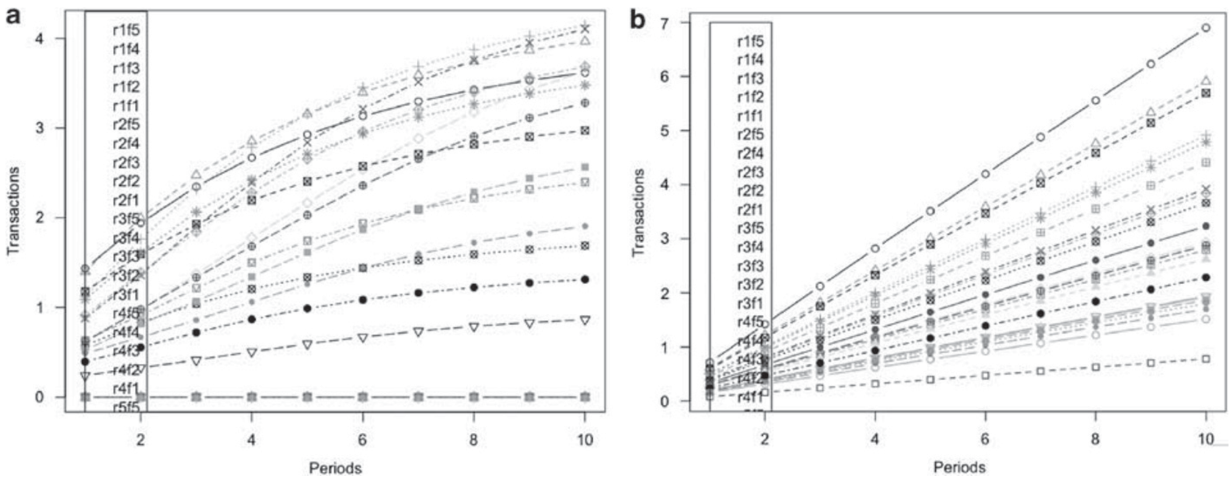


Figure A1: Evolution of residual expected transactions using determinist and stochastic models (Brabant dataset). (**a**) determinist migration model. (**b**) BG/BB Model.