Modeling Customer Lifetime Value, Retention, and Churn

Herbert Castéran, Lars Meyer-Waarden, and Werner Reinartz

Contents

Abstract

Customers represent the most important assets of a firm. Customer lifetime value (CLV) allows assessing their current and future value in a customer base. The customer relationship management strategy and marketing resource allocation are

H. Castéran (✉)
Humanis Institute, EM Strasbourg Business School, Strasbourg, France
e-mail: herbert.casteran@em-strasbourg.eu

L. Meyer-Waarden
School of Management, CRM CNRS University Toulouse 1 Capitole, IAE Toulouse, Toulouse, France
e-mail: lars.meyer-waarden@iae-toulouse.fr

W. Reinartz
University of Cologne, Köln, Germany
e-mail: werner.reinartz@uni-koeln.de

C. Homburg et al. (eds), *Handbook of Market Research*,
https://doi.org/10.1007/978-3-319-57413-4_21

based on this metric. Managers therefore need to predict the retention but also the purchase behavior of their customers.

This chapter is a systematic review of the most common CLV, retention, and churn modeling approaches for customer-base analysis and gives practical recommendations for their applications. These comprise both the classes of deterministic and stochastic approaches and deal with both, contractual and noncontractual settings. Across those situations, the most common and most important approaches are then systematically structured, described, and evaluated. To this end, a review of the CLV, retention, as well as churn models and a taxonomy are done with their assumptions and weaknesses. Next, an empirical application of the stochastic "standard" Pareto/NBD, and the BG/NBD models, as well as an explanatory Pareto/NBD model with covariates to grocery retailing store loyalty program scanner data, is done. The models show their ability to reproduce the interindividual variations as well as forecasting validity.

Keywords

Customer lifetime value · Customer churn · Customer retention · NBD/Pareto model · BG/NBD model

Introduction

Customers represent assets and the cost of acquiring them relates to the cash flow they are expected to generate over time (Bolton et al. 2004; Tarasi et al. 2011). Customer retention and churn as well as customer lifetime value (CLV), retention, and churn measurement have become a powerful customer valuation metric (Glady et al. 2015; Gupta et al. 2004, 2006; Kumar and Reinartz 2006).

Customer retention refers to the ability of a company or product to retain its customers over some specified period. Defection or churn is the number of customers moving out of a cohort in a firm's database over a specific period of time. CLV is the value of individual customers, based on their past, present, and projected future cash flows (Gupta et al. 2004). To model CLV, it is important to measure customer retention and churn rates. CLV is an important concept on which the customer relationship management strategy; marketing resource allocation (to profitable customers), such as promotions; and the assessment of the marketing efficiency are based on Schulze et al. (2012). The CLV paradigm recognizes customers as the primary source of both current and future cash flows. According to this framework, the firm tries to maximize the net present value of both current and future customers (customer equity, Hogan et al. 2002), which represents a good proxy for the firm's value (Borle et al. 2008; Gupta et al. 2004), as well as an effective segmentation tool. Thus, CLV models offer a powerful means to maximize the return on marketing investments and guide allocations of the marketing budget (Blattberg and Deighton 1996; Reinartz et al. 2005).

A CLV model has prototypically three parameters: (1) margin (purchase baskets minus the costs including retention expenditure), (2) retention probability or lifetime duration, and (3) purchase frequency (Kumar 2007). One way of increasing CLV is to undertake marketing initiatives to reduce churn or the defection rate (and therefore increase the retention rate) of customers – which will have the impact of increase in the customer lifetime periods. Putting it another way, CLV analyses involve distinguishing active customers from defectors and then predicting their lifetime and future levels of transactions according to their observed past purchase behavior. Developing a valid measurement framework that adequately describes the process of birth, purchase activity, and defection is thus a crucial albeit not a trivial task, particularly due to the randomness of individual purchasing behavior and customer heterogeneity (Jain and Singh 2002; Reinartz and Kumar 2000). Whereas the analysis may be easier for contractual, "lost for good" relationships (e.g., subscription markets in which the inactivity date is known), it becomes particularly difficult for noncontractual relationships in which customers do not notify the firm when they disappear (Dwyer 1989; Jackson 1985); in this scenario, identifying active and inactive customers in the database at any given time requires systematic investigation (Schmittlein and Peterson 1994).

The objective of this chapter is to provide a systematic review of the most common retention and churn modeling approaches to model CLV. These comprise both the classes of deterministic and stochastic approaches and deal with both, always-a-share and lost-for-good situations. Across those situations, the most common and most important approaches are then systematically structured, described, and evaluated. To this end, first the retention models, their assumptions, and weaknesses are reviewed and thus a taxonomy is provided. After having presented the taxonomy of CLV, churn, and retention measurement models, this article shows in "A Taxonomy of Customer Lifetime Value Measurement Models" a practical application by using some of the presented stochastic models (Pareto/NBD, explanatory Pareto/NBD, BG/NBD) to model a customer base and the impact of a retail grocery loyalty program on customer churn, retention, and activity. The goal is to show how to implement, use, and interpret these sophisticated models by applying them on firms' frequently used grocery store loyalty program databases and panel data.

This article then concludes with a discussion, some limitations, and recommendations for future research directions.

A Taxonomy of Customer Lifetime Value Measurement Models

In practice, to choose an adequate CLV measurement model, one has to understand whether or not customer defection is observable. One thus has to differentiate between two types of market (Jackson 1985; Dwyer 1989), namely, contractual (*Lost-for-Good*) and non-contractual (*always-a-share*) markets.

In the first type of market, the customer enters into a contractual relationship with a firm (e.g., phone or insurance services, magazine subscriptions, etc.) and is consequently faced with a tangible cost of change. Defection is observable and occurs when consumers end their relationship with the firm. In this scenario, the seller can identify defection as soon as it occurs. This means it is easy to predict defection for modeling purposes and one has to adopt a simple retention model

(Ayache et al. 2006). The notion of *lost-for-good* merges in practice with the contractual situation since it considers that absence of transaction means the customer has become inactive. In the contractual approach, retention is in fact the most important aspect. Generally, it goes hand in hand with a more or less constant flow of income. The models are usually simple with a clear predominance of survival models.

In markets where the customer has no contractual relationship (typically consumer goods), the cost of switching is low and a buyer can simultaneously purchase from different suppliers (*always-a-share*). The supplier has no way of knowing if the customer has defected. The model therefore focuses on churn probability, customer migration, and customer "life span" (Berger and Nasr 1998). The longer the period of inactivity, the more likely it is that the customer has churned. Migration models more specifically cover this scenario.

Fader and Hardie (2009) add an additional distinction depending on whether the purchase occurs at a specific moment (discrete time) or whether it can occur at any time (continuous). This distinction mainly has technical consequences, which can also be computed more or less approximately in a relatively direct way by taking more or less extended periods of time into consideration. Fader and Hardie (2009) themselves admit that this distinction is less meaningful. However, the contractual/noncontractual distinction is conceptually and methodologically fundamental.

The vast majority of markets concern noncontractual markets (Allenby et al. 1999). Many researchers and business practitioners have attempted to develop forecasting systems in this context. Contributions fall into two main categories: purely descriptive approaches (deterministic) and stochastic approaches. Deterministic approaches are primarily based on calculations of actuarial values, reflecting financial flows without the inclusion of random factors or explanatory variables (e.g., expected individual cash flow models as applied by Berger and Nasr 1998). However, they fail to take interindividual heterogeneity into account. Calciu and Salerno (2002) highlighted the relations between these different attempts.

The following table provides an overall view of the models according to the nature of their affiliation with the company and the methodology (deterministic/stochastic) used. Some contributions may be found in two different scenarios, in that they include a comparison of several cases.

Other aspects of model characterization are also included: level of aggregation, inclusion of the competition, return on investment, and the capacity to optimize resource allocation. The nature of the model and the level of aggregation help to determine the model's sophistication and precision. Taking the competition into account is likely to affect the results of the models in that the long-term perspective is more complex when the competitive context is explicitly included. Finally, the capacity to determine return on investment or to optimize the distribution of marketing investment affects the model's operational nature.

Table 1 suggests several trends. The first is the increasing focus on stochastic models as compared to the deterministic models. Since 2005, eight new stochastic models have been presented against only two in the deterministic context. As already stated, probabilistic models are significantly more efficient than deterministic models. This tendency thus seems logical and desirable.

Table 1 Models of customer retention-churn modeling (Adapted from Villanueva and Hanssens 2007)

Authors	Level of analysis	Competition present	Return on investment	Allocation of resources
Deterministic models				
No application				
Rust et al. (2004)	Company	Yes	Yes	Yes
Blattberg et al. (2001)	Segment	No	Yes	No
Application to contractual cases				
Keane and Wang (1995)	Regions	No	No	No
Blattberg and Deighton (1996)	Company	No	Yes	No
Dwyer (1997)	Segment	No	No	No
Ryals (2005)	Individual	No	No	No
Wiesel et al. (2008)	Company	No	No	No
Application to noncontractual cases				
Dwyer (1997)	Segment	No	No	No
Berger and Nasr (1998)	Individual	No	No	No
Stauss and Friege (1999)	Individual	No	No	No
Berger and Nasr (1998)	Company	No	No	No
Gupta et al. (2002)	Company	No	No	No
Gupta and Lehman (2003)	Company	No	No	No
Stochastic models				
Application to contractual cases				
Bitran and Mondschein (1996)	Segment	No	No	No
Thomas et al. (2004)	Individual	No	Yes	No
Lewis (2005)	Individual	No	Yes	No
Villanueva et al. (2008)	Company	No	No	No
Application to noncontractual cases				
Schmittlein et al. (1987)	Individual	No	No	No
Reinartz and Kumar (2000)	Consumer	No	Yes	No
Pfeifer and Carraway (2000)	Segment	No	Yes	No
Libai et al. (2002)	Segment	No	Yes	Yes
Rust et al. (2004)	Company	Yes	Yes	Yes
Venkatesan and Kumar (2004)	Individual	No	Yes	Yes
Fader et al. (2005a)	Individual	No	No	No

(continued)

Table 1 (continued)

Authors	Level of analysis	Competition present	Return on investment	Allocation of resources
Reinartz et al. (2005)	Company	Yes	Yes	No
Villanueva et al. (2008)	Segment	No	Non	No
Simester et al. (2006)	Individual	No	Yes	No
Lewis (2006)	Individual	No	Yes	No
Castéran et al. (2007a, b)	Individual	No	No	No

The second underlying trend involves the increasing disaggregation of the models. From wholly aggregated models, one shifts to an analysis by company, then to one by segment, and, finally and increasingly often, to one by individual. While informational limitations may explain the inclusion of a company level, nothing, on the other hand, justifies grounding a marketing analysis on wholly aggregated models.

Two aspects have been relatively neglected to date, namely, inclusion of the competition and the way managers interpret models for resource allocation. Lack of information is frequently used to explain this shortcoming, but it nonetheless remains detrimental. This is especially true of the failure to include the competition insofar as its absence may substantially impact on conclusions and managerial implications (cf. Fudenberg and Tirole 2000). While the examination of optimized resource allocation remains fundamental, its absence does not, on the other hand, imply an analysis bias.

We present these approaches in more detail in a dual customer relations and methodology framework.

Retention Models for CLV Measurement

These models are divided between deterministic and probabilistic models. To determine CLV, customer retention and churn have to be modeled. Customer retention refers to the ability of a company or product to retain its customers over some specified period. It is measured in the following way (Gupta et al. 2004).

$$\text{Retention rate} = \begin{array}{l} n \text{ customers in cohort buying in } (t) \\ /n\,\text{customers in cohort buying in } (t-1) \times 100 \end{array} \tag{1}$$

The period t can refer to specific durations: months or years are the most frequently used. Customer defection or churn is the number of customers moving out of a cohort in a firm's database over a specific period of time. It is measured in the following way (Gupta et al. 2004):

$$\text{Churn rate} = 1 - \text{Retention rate} \tag{2}$$

Deterministic Models

Berger and Nasr (1998) provide the following general formula for the customer lifetime value (CLV):

$$\text{CLV} = \sum_{t=1}^{n} \pi(t) \frac{\rho^t}{(1+d)^t} \tag{3}$$

with $\pi(t)$ profit generated in period t, ρ the rate of retention, and d the discount rate. If one considers profit stability over time for an annual net gain h, then CLV is formulated as

$$\text{CLV} = h \frac{\rho^t}{(1+d)^t} \tag{4}$$

We have a monetary component h and an expected number of transactions (or products, *discounted expected transactions*). This expression has the advantage of being extremely simple: one just has to estimate the retention rate to obtain the CLV. On the other hand, this approach assumes that the retention rate is stable over time.

However, this assumption fails to take into account the customer base composed of different segments, over and above all considerations of variation in the retention srate at individual level. Imagine that a same cohort of customers is composed of p homogeneous segments, each with an annual retention rate assumed to be constant from 1 year to the next for purpose of simplicity, with ρ_i for each segment i. One also can reason in discrete time for greater simplicity, but the situation can easily be extrapolated to continuous time. Let us assume that by nature segment 1 has the highest retention rate. The average retention rate, for example, in the first year is equal to

$$\overline{r} = \frac{\sum_{i=1}^{p} n_i \rho_i}{\sum_{k=1}^{p} n_k} \tag{5}$$

with n_i the size of segment i. Traditionally, portfolio value is calculated on the basis of this average rate.

However, because of the retention dynamic, the probability of belonging to segment 1 will converge toward 1, and, at the same time, the average retention rate will also converge toward the retention rate of segment 1. In effect, according to Bayes' theorem, one gets the probability of customer c belonging to segment 1 active after t years, formulated as

$$\begin{aligned} \mathrm{P}(c \in S_1 | \text{ active after } t \text{ years}) &= \frac{\mathrm{P}(c \in S_1)\mathrm{P}\,(\text{active after } t \text{ years } | c \in S_1)}{\mathrm{P}\,(\text{active after } t \text{ years})} \\ &= \frac{p_1 \rho_1^t}{\sum_{i=1}^{n} p_i \rho_i^t} = \frac{1}{1 + \frac{p_2 \rho_2^t}{p_1 \rho_1^t} + \cdots + \frac{p_n \rho_n^t}{p_1 \rho_1^t}} \\ &= \frac{1}{1 + \frac{p_2}{p_1}\left(\frac{\rho_2}{\rho_1}\right)^t + \cdots + \frac{p_n}{p_1}\left(\frac{\rho_n}{\rho_1}\right)^t} \end{aligned} \tag{6}$$

However, since, by definition, $r_1 \geq r_i, \ \ \forall\ i \neq 1$ then $\forall i \neq 1, \lim_{t \to +\infty} \left(\frac{\rho_i}{\rho_1}\right)^t = 0$.

So the more time that passes (t becomes large), the higher the probability of belonging to segment 1, leaning toward a limit of 1. The average retention rate for a cohort thus converges toward the retention rate of segment 1. Variation in the retention rate is linked to the heterogeneous nature of the population. The use of an aggregate rate is not adapted for assessing the CLV. It can however be used by companies as a proxy for business health. Nowadays, adopting a stable retention rate represents a very particular case and is often inadequate.

Probabilistic Models

There are two types of probabilistic models: parametric and semi-parametric.

Parametric Models

In terms of parametric models, more elaborate models than the deterministic ones have been developed in the contractual framework. Thus, Fader and Hardie (2007b) used a survival function to obtain an expression such as (7)

$$\mathrm{E(CLV)} = h \frac{S(t)}{(1+d)^t} \tag{7}$$

considering time as discrete. The link with the preceding form is obvious apart from the fact that *S(t)* is the survival or retention function on date t and one can no longer speak about CLV but of expectancy of CLV. The authors assume that life span is given by a geometric distribution. The customer remains as such from one period to another with a probability *1-p*. In this context, $S(t) = (1 - p)^t$. Interindividual heterogeneity in terms of probability p is given by a beta distribution (with values between 0 and 1). One thus obtains the shifted beta-geometric model (sBG).

Naturally, other expressions of survival are possible, notably with the inclusion of explanatory variables and the shift to continuous time. Schweidel et al. (2008) thus included explanatory variables while retaining a formulation with latent traits in continuous time. They developed the formula

$$S(t) = \int S[t | \theta_i, X(t)] g(\theta_i).d\theta_i \tag{8}$$

with *X(t)* as all of the explanatory variables for t and θ_i a set of individual latent traits. $g(\theta_i)$ represents the distribution of θ_i. This formulation ensures the harmonious integration of latent traits and explanatory variables, giving us a mixed effects model with fixed and random components.

$g(\theta_i)$ is the distribution that can be used to measure interindividual heterogeneity. It generally involves a gamma distribution for reasons of flexibility and compatibility with most survival distributions. It is expressed as follows:

$$g(\theta_i \mid r, \alpha) = \frac{\alpha^r \theta_i^{r-1} e^{-\alpha\theta_i}}{\Gamma(r)} \tag{9}$$

One can express the survival function in the form of the hazard function. The hazard function measures the instantaneous risk of mortality.

$$S\,[t|\,\theta_i, X(t)] = e^{-\sum_{v=1}^{t} \int_{v-1}^{v} h[u|\theta_i, X(t)]du} \tag{10}$$

If one concentrates on the stochastic dimension, the basic hazard function h_0 can adopt the Weibull distribution:

$$h_0(t|\,\theta_i, c) = c\theta_i t^{c-1} \tag{11}$$

This formulation takes into account risk that evolves over time. Variation in the retention rate depends as much on heterogeneity (interindividual variations) as on intrinsic individual variations. If $c = 1$, one then shifts to the exponential-gamma (EG) model. Note that in continuous time, this model is the equivalent of the sBG model (Fader et al. 2003).

Semi-Parametric Models

The most famous representative of semi-parametric models is the Cox model, often called the proportional hazard model. It models a life span considered as a random variable with a probability density *f(t)* and a distribution function *F(t)*. The survival function is expressed as

$$S(t) = P(T \geq t) = 1 - F(t) \tag{12}$$

This function is of course monotonically decreasing.

The hazard function is written as

$$h(t) = \lim_{dt \to 0} \frac{P[(t \leq T < t + dt)(T \geq t)]}{dt} = \frac{f(t)}{S(t)} \tag{13}$$

Instead of taking the hazard function into consideration in a parametric way as in the preceding point, one estimates it following the Kaplan-Meier procedure. The cumulated hazard function is expressed as

$$H(t) = \int_0^t h\,(u).du = -\ln[S(t)] \tag{14}$$

The addition of explanatory variables in the form of an X matrix allows us to adopt a semi-parametric formulation:

$$h(t|\ X) = h_0(t)e^{X\beta} \tag{15}$$

$h_0(t)$ only depends on time. With Eq. 15, based on expression (11), the survival function becomes

$$S(t|X) = [S_0(t)]^{e^{X\beta}} \tag{16}$$

with the same formulation logic as the hazard function.

Migration Models for CLV Measurement

These models represent a generalization of retention models. The absence of transactions at any given moment does not mean that the customer has become inactive. This is typically the case in a noncontractual situation, in which customer inactivity cannot be observed.

The main idea is that customers go through different stages in their relationship with the brand with specific characteristics governing each stage. One therefore needs to describe the characteristics of these stages as well as the conditions for the transition from one stage to another.

Deterministic Models

Heuristics are frequently used to identify the situation of a customer in a deterministic context. The best known of these is the RFM segmentation (recency, frequency, and monetary value). Recency is the determinant factor to assess whether or not a customer is active. Customers are segmented on the basis of more or less valid thresholds. Traditionally, one distinguishes three levels per criterion R, F, and M, representing 27 segments. The more recently a customer has made a purchase, the greater his or her purchasing frequency, and the higher the average basket, the greater his or her supposed potential. This apparently logical hypothesis is, as noted earlier, qualified by observation of the behaviors of these different segments (e.g., Fader et al. 2005b).

At managerial level, a customer is traditionally considered as inactive beyond a certain length of time without arbitrarily fixed purchases. This method has been presented to us many times by firms that adopt a customer relation management approach. Schweidel et al. (2008) also noted its predominance in professional practice to determine whether or not a customer is active. Likewise, forecasts of future sales are made through a simple extrapolation of past sales.

Probabilistic Models

Two forms of approaches coexist. The first is in the form of Markov processes and the second in the form of combinations of models.

Markov Processes

In terms of migration models, the most widely used method is certainly that of Markov chains, also called Hidden Markov Models. Popularized by Pfeifer and Carraway in 2000, it has been the object of numerous extensions through the integration of sociodemographic or RFM variables. A customer is assumed to be in a certain relational situation with respect to the company, defined in advance. Naturally, these stages are never observed but remain latent which explains the term "Hidden Markov Models." One can subsequently calculate the probability of transition from one state to another. Thus, Pfeifer and Carraway (2000) identified five levels of customer relations, from the most recent customers to buyers that bought such a long time ago they are considered as "non" or former customers. The transition pattern can be expressed graphically as follows (Fig. 1):

In this framework, there is perfect sequentiality. At stage 5, customers are considered as definitively lost with no chance of reactivation. This hypothesis can easily be changed. These models may be likened to latent class models except that adherence to a segment in the framework of hidden Markov models is dynamic and follows a Markov process.

Adapting Kumar (2007), a customer's CLV may be expressed in the following way:

$$\mathrm{CLV} = \sum_{t=0}^{T} \frac{MM_t P_t}{(1+d)^t} \tag{17}$$

with MM_t the matrix of probability of transition from one state to another at t, d the rate of loss, and P_t the value generated by the customer on date t. Over time, the probability matrices merge with one another. Thus, if one starts from the probability of initial MM_0 transitions, one gets $t = 1 \ \ MM_1 = MM_0 \times MM_0 = (MM_0)^2$ and so $MM_t = (MM_0)^{t+1}$.

A specific application is that of Rust et al. (2004) with a brand change matrix. Combined with a logit model, this application demonstrates the flexibility and the potential of the Markov approach.

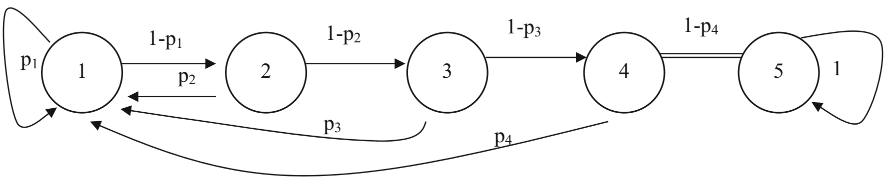

Fig. 1 Transition from one stage to another according to a Markov process

Combinations of Models

This approach was largely developed by Schweidel and Fader (2009). It considers that a stronger customer relationship with a brand (measured by the number of repeat purchases) is expressed through greater purchasing stability and so more stable interpurchase times. In short, it is the transcription of the transition from new customer to existing customer. There are thus two interpurchase periods that follow one another, the first characterized by an exponential distribution and the second by an Erlang 2 distribution. This distribution is a specific case of gamma distributions with a shape parameter equal to 2. The density is thus written as $\lambda^2\ xe^{-\lambda x}$.

The transition from one state to the other occurs after each purchase with probability p. One thus arrives at a transition which respects a geometric process.

All the parameters of the different models are assumed heterogeneous. The two parameters transcribing the rate of transactions (exponential distribution and Erlang 2) are themselves gamma distributed, and probability is distributed according to a beta law. This line of research is interesting for several reasons: it takes into account explanatory variables, can be generalized to a larger number of situations and other distributions, etc.

Continuous Mixed Models: The Family of NBD Models to Measure CLV

This term is rarely used but allows us to describe the underlying nature of these models. It involves estimating the different processes simultaneously: consumption, attrition, etc. To this end, each process is assumed to correspond to a specific law. Customer heterogeneity is also expressed by a distribution. All the consumption characteristics are considered to be governed by latent traits. Explanatory variables may be integrated, depending on the degree of sophistication of the different models. Fundamentally though, the introduction of explanatory variables is not in accordance with the philosophy of these models based on stochastic determinants.

The whole palette of statistical tools is used here.

The Pareto/NBD and BG/NBD Model

In this context, continuous mixed models are considered as one of the most promising lines of research, especially the negative binomial formalization (NBD) model. The Poisson distribution of data y is combined with a gamma distribution of purchasing frequency. This approach, developed by Ehrenberg (1959), has been extended by taking into account the inactivity factor: the Pareto/NBD model (Schmittlein et al. 1987; Morrison and Schmittlein 1988; Schmittlein and Peterson 1994; Abe 2009; Jerath et al. 2011) or betaPareto/NBD model-geometric/NBD model (BG/NBD by Fader et al. 2005a). Thus, consumer behavior is represented by a continuous representation that, in theory, takes all of the individual specificities into account.

However, continuous mixed models imply a total parametric specification (generally Poisson with a specific frequency parameter distribution) that is by nature restrictive and very often not very well adapted given the fundamental hypotheses of these distributions. Semi or nonparametric generalizations are naturally possible. However, their introduction requires a highly complex mathematical conceptualization process. In the same way, the introduction of explanatory variables is also possible but always at the price of a demanding mathematical formulation (Castéran et al. 2007a). Consequently, the operational and managerial scope of these models appears to be greatly reduced.

Finite mixed-effect models have been used for many years. The first principles were laid down by Newcomb (1886) and Pearson (1894). Finite mixed-effect models provide a specific case of latent class models (Baltagi 2003). They postulate the existence of latent classes within the population under study and a specific link between explained and explanatory variables within each of these classes. In this way, they underpin the existence of segments with specific behavioral patterns; the marketing implications are clearly apparent.

However, applications in a specifically marketing framework were initiated relatively late, mainly by Wedel et al. (1993). They provide a segmentation of the population beyond traditional behavioral segmentation. While apparently offering less detailed analysis than a continuous approach, segmentation does provide a clear interpretation of the results obtained as well as directly accessible managerial and operational implications. These implications are reinforced by the presence of explanatory variables. Each segment may be studied according to its own behavioral characteristics, which are explained by a set of variables. These explanatory variables help to determine the most effective marketing actions at the level of each segment. Finite mixed-effect models thus appear to be a promising alternative to continuous mixed models.

Nonetheless, to our best knowledge, the comparative efficiency of these models has only been demonstrated one time by Castéran et al. (2008). This comparison in terms of predictive validity between the finite mixed models and models of the NBD family (NBD simple, Pareto/NBD, BG/NBD) is worth exploring further.

The Explanatory Pareto/NBD and BG/NBD Model

The fact that all of these models are purely stochastic implies that they only have limited managerial potential. It is therefore important to reconcile the predictive validity of these purely stochastic models with an interpretative dimension resulting from the presence of explanatory variables. The introduction of explanatory variables within the Pareto/NBD model is a promising approach (Castéran et al. 2007b). This is done by the introduction of the explanatory variables in the gamma-gamma model by breaking down the variability of the scale parameter into two elements by distinguishing two components of parameter λ (purchasing frequency) and by using a regression with explanatory variables as well as a parameter λ_0:

$$\lambda = \lambda_0 e^{X_1 \beta} \tag{18}$$

with β the vector of coefficients and X_1 the individual characteristics and marketing actions. Parameter λ_0 is distributed according to a gamma law of parameters r (form) and α (scale). This parameter captures the residual heterogeneity not taken into account by the explanatory variables. Its density is expressed as

$$f(\lambda_0 | r, \alpha) = \frac{\alpha^r}{\Gamma(r)} {\lambda_0}^{r-1} e^{-\alpha\lambda_0} \tag{19}$$

with $\lambda_0 > 0$, $r > 0$ et $\alpha > 0$.

Then one has to adopt the same process for inactivity with regard to parameter μ. One notes the matrix of personal characteristics and marketing actions as X_2 (this can partially or entirely correspond to X_1). Inactivity thus becomes

$$\mu = \mu_0 e^{X_2 \gamma} \tag{20}$$

with parameter μ_0 following a gamma distribution of parameters s (form) and δ (scale),

$$f(\mu_0 | s, \delta) = \frac{\delta^s}{\Gamma(s)} {\mu_0}^{s-1} e^{-\delta\mu_0} \tag{21}$$

with $\mu_0 > 0$, $s > 0$ et $\delta > 0$.

In addition to the explanatory variables X (composed of X_1 and X_2), one needs three additional elements: number of purchases y made during the period [0, T], recency of the last purchase t_y (date of last purchase), and length of the period of estimation T. H is the combination of all three variables, $H = (y, t_y, T)$, and Θ the vector of all the coefficients, $\Theta = (r, \alpha, s, \beta, \gamma)$.

The limitation is due to the fact that one only deals with variables without a dynamic perspective, as they are constant over time.

Finally, Fader and Hardie (2007a) developed a general expression to introduce invariant explanatory variables over time within Pareto/NBD and BG/NBD models. The inclusion of these variables is conducted in a less complex way than the approach of Castéran et al. (2007b).

The Fig. 2 presents an overview of the CLV models.

The fundamental distinction is due to the nature of the relations between the customer and the company: is the customer's inactivity observed (contractual relations) or not? The second criterion comes from the type of model adopted: whole population, segment, or individual.

Finally, the last parameter is the distinction between continuous and discrete purchasing opportunities. However, this distinction is less crucial than the others insofar as certain discrete cases may be considered as continuous cases, while continuous cases can always be "discrete."

Casteran et al. (2007b) did not distinguish between variants with or without explanatory variables in this process. The presence of explanatory variables within purely stochastic formulations presents a methodological as well as a conceptual improvement.

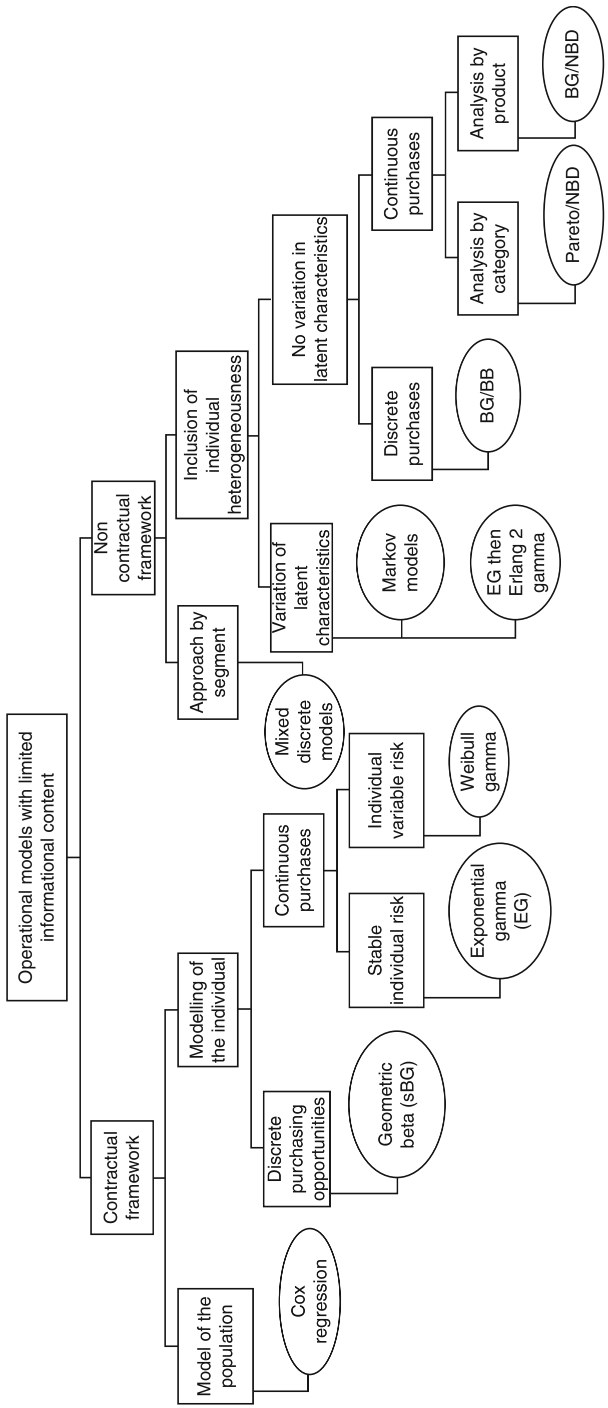

Fig. 2 Overview of the CLV measurement models

An Application of Stochastic Pareto/NBD and BG/NBD Models for Customer Base Management

After having presented the taxonomy of CLV, churn, and retention measurement models, this chapter shows a practical application by using some of the presented stochastic models (Pareto/NBD, BG/NBD) to model a customer base and the impact of a retail grocery loyalty program on customer churn, retention, and activity. The goal is to show how to implement, use, and interpret these sophisticated models by Customer base managementapplying them on firms' frequently used grocery store loyalty program databases and panel data.

Data and Methodology

The data for the practical application of customer base analysis come from a store loyalty program (LP) database of a large French grocery retailer (surface area 9,000 m^2).

An LP consists of integrated and ruled systems of marketing actions (based on collection and redemption rules) that aim to encourage repeat purchases and increase the cost of switching as well as retention and subsequently CLV by providing short- and long-term incentives (Meyer-Waarden 2007; Blattberg et al. 2008; Bijmolt et al. 2010) and enhance "true loyalty," that is, increase behavioral (e.g., cross purchases, repeat purchases, mean basket size) and attitudinal (relationship building through positives attitudes, trust, attachment; Morgan and Hunt 1994) loyalty.

Loyalty programs (LP) are vastly popular – 90% of Europeans and 90% of US shoppers own at least one loyalty card. In 2010, the number of LP memberships in the United States exceeded 2.1 billion memberships, growing by 16% from the previous year despite the worldwide recession (Hlavinka and Sullivan 2011). For example, research estimates that the UK pharmacy chain Boots invested 30 million British pounds in the launch of its Advantage Card LP (Temporal and Trott 2001), and the U.K. retailer Tesco has spent an estimated 60 million pounds to operate its Clubcard LP (Bijmolt et al. 2010).

The store's loyalty program, launched in 1994, is free and provides price discounts, points exchangeable for gifts, and purchase vouchers on a varying set of items. The value of points won increases linearly according to the amount customers spend. Cardholder purchases account for 70% of store revenues. In the analysis, cardholder information is used, identifiable on an individual basis, which includes the majority of customers. The data set contains also information about competing loyalty card memberships and household characteristics (e.g., age, revenue). Scanner data include the transaction details, such as date of purchase and amount paid. Because people often shop on a weekly basis, the daily purchases are aggregated by individuals into a weekly frequency. The transaction data pertain to 5,000 households over a period of 156 weeks (week 2/1998 to week 2/2001).

The LP data is matched with BehaviorScan panel test market data from Angers, France. Scanning technology provides exhaustive recording of the purchasing behavior (95% of fast-moving consumer goods sales in the area) of the panelist households, which are representative of the national population.

The implementation conditions of the models with LP data relate to two aspects: the consideration of new customers that have realized their first purchase, with certainty, and the duration of the estimation period. The date of first purchase in a noncontractual framework may be considered known, whether because of the nature of the data (e.g., Fader et al. 2005b) or the data processing method (Batislam et al. 2007). The Pareto/NBD, BG/NBD, and PDO models forecast all purchases by combining the number of customers at a given date with the unconditional expectancy of the number of purchases, according to the customer maturity level (Fader et al. 2005a). As stated higher, the first purchase is easy to identify in certain settings, when the entire customer history is available. Nevertheless, it is difficult, in most noncontractual relationships, to determine an exact date of the actual first purchase. (An exception are mail order business or e-commerce retail settings, where the shipping address of the customer is known and the purchase can be identified.) In grocery retailing, data from a loyalty program (LP) are left censored, and the first purchase cannot be characterized with certainty, because customers probably purchase before they enroll in a loyalty scheme. Therefore, it is unclear what types of customers are observed: “real” new customers, previously existing customers who have lately adopted the LP or customers who had lapsed or have low usage patterns. Batislam et al. (2007) and Jerath et al. (2011) both resort to drastic truncations to achieve a sample that consists entirely of new customers, who made no purchases in the first 13 months of their observations, whom they logically argue are genuine new customers. This approach means the loss of substantial information and raises questions regarding sample representativeness. For example, consumers who make their first purchases at different times, later or earlier, behave differently in their purchase frequency and loyalty in business settings (Schmittlein and Peterson 1994). The best customers often self-select as early adopters (Meyer-Waarden and Benavent 2009; Rogers 2003), so truncating samples could exclude insights into some of the firm’s best customers and earliest adopters. Another option to solve the left-censored data issue is to treat the first customer purchase observed in the LP database as the customer’s first actual buying act. This method creates the risk of combining different cohorts though, with different actual dates of first purchase; its consequences for models’ predictive validity have never been examined.

Both the Pareto/NBD and the BG/NBD models require tracking customer transactions, starting with their initial purchases, which raises the possibility of left-censored data, because one does not know when people became aware of the outlet’s loyalty program and if the first purchases recorded after enrollment are really their first transactions. In other words, households may have bought before adopting the loyalty scheme. Because the store does not have information about the initial purchases of cardholders, this methodological problem treated by left-filtering the panel customer records with transactions before October 14, 1998, which guarantees that the customers in the analysis are newcomers with known initial purchase times data (see Batislam et al. 2007). Thus, 5,000 households of a cohort of 997 new households is extracted that made their first purchases within the same 3-month period (October–January) and realized a total of 6,005 transactions. The panel data is left-filtered and aggregated on a weekly basis, and the final observation window covers 78 weeks, from October 14, 1998, to April 13, 2000. The estimation period is restrained to 26 weeks and the calibration period to 52 weeks (hold-out sample) to establish the predictive validity of the models.

Finally, we get a matrix with one row for each customer, and at least three columns:

1. Customer's frequency, y: number of repeated transactions made in the estimation period
2. Customer's recency, t_y: the time of their last transaction
3. Customer's total time observed in the estimation period, T

Other columns can be added, one for each explanatory variable. The explanatory variables have to be quantitative or dummy variables.

Estimation

Parameter estimation of the Pareto/NBD model is more complex (Fader et al. 2005a; Reinartz and Kumar 2003; Fader and Hardie 2013); in particular, the maximum likelihood estimation (MLE) approach to estimating key parameters used requires a numerical search algorithm that must evaluate the Gauss hypergeometric function (Schmittlein and Peterson 1994).

For the estimation, free statistical software *R* 3.3.1 (R Development Core Team 2010) is used. Our advice is to adopt a two-step approach. Firstly, we estimate the purely stochastic model parameters (Pareto/NBD or BG/NBD) without explanatory variables. Secondly, we incorporate the explanatory variables and launch a new estimation process based on the results of the first estimation.

Estimation of the Purely Stochastic Parameters

The easiest way for estimating simple Pareto/NBD and BG/NBD models is now to use a dedicated package BYTD (https://CRAN.R-project.org/package=BTYD) "Buy 'Til You Die Models" (Dziurzynski et al. 2014). The estimation of the parameters is made through, respectively, the functions *pnbd.EstimateParameters* (Pareto/NBD) and *bgnbd.EstimateParameters* (NG/NBD).

There are two major estimation issues. First, the procedures are time-consuming, especially with regard to the initial values. Second, depending on these values, non-convergence might be faced, which makes it even more difficult to find an operational set of initial values. Many tries can be required in order to assure a clear convergence. The best choice, even if convergence occurs, is to relaunch with several starting points in order to compare the results and the value of the log-likelihood.

A good starting point is to consider that the average purchase rate is the ratio r/α. It is not sufficient to determine the exact starting values of the parameters, but it can be useful that the initial values of r and α are chosen with respect to the average purchase rate of the dataset.

Since an optimal (in sense of log-likelihood maximization) set of parameters is found, the incorporation of explanatory variables can begin.

Estimation of the Parameters Including Explanatory Variables

The first step is to declare the explanatory variables. If *mydata* is the name of the dataset, the vector of the explanatory variables for the purchasing frequency X_1 could be the following:

```
X1=cbind(mydata$LoyCard, mydata$HhSize, mydata$NumbCard,
mydata$DisStore, mydata$SeniorManager, mydata$Unemployment,
mydata$Income_6, mydata$Income6_12, mydata$Income12_18,
mydata$Income18_24, mydata$Income24_30, mydata$A30_39,
mydata$A40_49, mydata$A50+)
```

Our variables describe:

1. The characteristics of the individuals: the household size (*HhSize*), their age (the *A*… dummy variables), their net wages (the *Income*… dummy variables), and the professional occupation (*SeniorManager*, *Unemployment*)
2. The relationship to the store: the distance in kilometers from the store (*DisStore*), the owning of loyalty cards (*LoyCard*), and the total number of loyalty cards owned by the household (*NumbCard*)

```
X2=X1 #Explanatory variable vector for inactivity part
```

At the beginning, the explanatory variables for purchasing frequency and inactivity can be the same. Further, during the selection process, the two sets will become different.

In order to incorporate qualitative variables, we divide them into dummy variables (e.g., income or customer age). To avoid overidentification, one modality of each variable shall be excluded from the estimation process. Whatever the modality is, the exclusion of one modality per qualitative variable is mandatory.

The set of initial values (*b0* in the example) is determined on the basis of the first estimation with *pnbd.EstimateParameters* (Pareto/NBD). Those parameters are called *params* here.

```
b0<-c(params, rep(0, ncol(X1)+ncol(X2)))
```

The initial values for the explanatory variables are set to 0 in order to relaunch the estimation process at the same initial state as purely stochastic approaches. The reestimation process can now begin.

For the maximization of the log-likelihood function and estimation of the parameters, two functions are employed: *nlminb* and *optim*. The *nlminb* procedure is more flexible and presents fewer convergence problems. After estimating *nlminb*, the *optim* is used to compute the Hessian matrix for estimating the covariance matrix (standard error of the coefficients). In the following example however, we directly use the *optim* function.

The estimation of the Pareto/NBD model – an effort whose difficulty is frequently cited as a usage limitation – is considerably facilitated by the *gsl* package, with the expression *hyperg_2F1*, that enables the estimation of a Gaussian hypergeometric function to increase external validity.

The final log-likelihood of the explanatory Pareto/NBD model can be written as

$$
\begin{aligned}
LL(\Theta_i|H_i, X_i) = {} & \ln\Gamma(r+y) - \ln\Gamma(r) + r\ln\alpha + s\ln\delta + y\ln B \\
& + \ln[A_1A_2 + A_3A_0]
\end{aligned}
\tag{22}
$$

with

(i) $B = e^{X_1\beta}$ and $G = e^{X_2\gamma}$

(ii) Due to the presence of the Gaussian hypergeometric function and the form of the integrals, we must distinguish two cases: when $\alpha e^{X_2\gamma} \geq \delta e^{X_1\beta}$ and the opposite case. For each case, we note a different expression for A_0:

If $\alpha G \geq \delta B$,

$$
A_0 = \frac{\left(\frac{B}{G}\right)^{s+1}}{B} \text{ x} \left[\frac{{}_2F^{2}{}^{1}\left(\frac{s+1}{r+s+y}; r+s+y+1; \frac{\alpha - \delta\frac{B}{G}}{\alpha + t_yB}\right)}{\left(\alpha + t_yB\right)^{(r+s+y)}} - \frac{{}_2F^{2}{}^{1}\left(\frac{s+1}{r+s+y}; r+s+y+1; \frac{\alpha - \delta\frac{B}{G}}{\alpha + TB}\right)}{\left(\alpha + TB\right)^{(r+s+y)}} \right]
\tag{23}
$$

If $\alpha G \leq \delta B$,

$$
A_0 = \frac{\left(\frac{G}{B}\right)^{r+y}}{G} \text{ x} \left[\frac{{}_2F^{2}{}^{1}\left(\frac{r+y}{r+s+y}; r+s+y+1; \frac{\delta - \alpha\frac{G}{B}}{\delta + t_yG}\right)}{\left(\delta + t_yG\right)^{(r+s+y)}} - \frac{{}_2F^{2}{}^{1}\left(\frac{r+y}{r+s+y}; r+s+y+1; \frac{\delta - \alpha\frac{G}{B}}{\delta + TG}\right)}{\left(\delta + TG\right)^{(r+s+y)}} \right]
\tag{24}
$$

(iii) $A_1 = (TB + \alpha)^{-(r+y)}$ and $A_2 = (TG + \delta)^{-s}$

(iv) $A_3 = \frac{Gs}{r+s+y}$

Since optimization algorithms classically perform minimization, we use the negative form of the log-likelihood function.

```
library(gsl)
LL_Paretoexp <-function(p) {
# Parameters vector
     r<-p[1]
     alpha<-p[2]
     s<-p[3]
     delta<-p[4]
# Number of covariates
     nX1=ncol(X1)              # for purchasing frequency
     nX2=length(p)-4-nX1       # for inactivity process
# Coefficients of explanatory variables
     b1=p[5:(4+nX1)]           # for purchasing frequency
     g1=p[(5+nX1):length(p)]   # for inactivity process
# Regressions
     B<-exp(as.matrix(X1)%*%b1)
     G<-exp(as.matrix(X2)%*%g1)
#Meta-functions
     A1<-(B*T+alpha)^(-r-y)
     A2<-(G*T+delta)^(-s)
     A3<-G*s/(r+s+y)
# A0 expression
     coef1<-(B^s)/(G^(s+1))
     arg11<-hyperg_2F1(s+1, r+s+y, r+s+y+1, (alpha-delta*B/G)/
(alpha+t_y*B))/((alpha+t_y*B)^(r+s+y)) # t_y = ty
     arg12<-hyperg_2F1(s+1, r+s+y, r+s+y+1, (alpha-delta*B/G)/
(alpha+T*B))/((alpha+T*B)^(r+s+y))
     coef2<-((G/B)^(r+y))/G
     arg21<-hyperg_2F1(r+y, r+s+y, r+s+y+1, (delta-alpha*G/B)/
(delta+t_y*G))/((delta+t_y*G)^(r+s+y))
     arg22<-hyperg_2F1(r+y, r+s+y, r+s+y+1, (delta-alpha*G/B)/
(delta+T*G))/((delta+T*G)^(r+s+y))
     A0<-ifelse(alpha*G>delta*B, coef1*(arg11-arg12), coef2*
(arg21-arg22))
# Log-likelihood function
     -sum(lgamma(r+y)-lgamma(r)+ r*log(alpha)+ s*log(delta)+ y*log
(B)+ log(A1*A2+A3*A0))
}
```

The lower bounds are 10^{-3} for the stochastic parameters:

```
min<-c(rep(1e-3,4), rep(-Inf,length(b0)-4))
max<-rep(Inf, length(b0))
```

The selection of the explanatory variables represents a significant challenge. On the basis of the Hessian matrix, small set of variables is kept, though traditionally, modeling purchase behavior is quite complex and the identification of relevant variables very difficult (Ehrenberg 1988).

```
optimal<-optim(b0, fn=LL_Paretoexp, method="L-BFGS-B",
control=list(trace=6, REPORT=1), hessian=TRUE, lower=min,
upper=max)
optimal # Result of the estimation process
```

Let us remind that we get the standard errors by taking the square root of the diagonal elements of the covariance matrix. The covariance matrix is the inverse of the Hessian matrix obtained through the minimization of the negative log-likelihood: *optimal*. All results are presented in a table with the coefficients (*Coeffs*), the standard errors (*StdError*), and the t-values (*t*). The null hypothesis that the coefficients are not significantly different from 0 (β_i *or* $\gamma_i = 0$) is rejected at a 5% significance level if $t \in \,]-\infty; -1.96] \cup [1.96; +\infty[$.

We compute also the Bayesian information criterion *BIC*, a common indicator, expressed as $-2LL + kln(N)$, where k is the number of parameters to be estimated (*length(b0)*), N indicates the number of individuals (*length(y)*), and *LL* is the maximum log-likelihood value for the model (*-optimal$value*).

We recommend a step-by-step descending selection process by removing one variable at each step:

1. We launch the estimation with the whole set of potential covariates.
2. We remove the covariate with the closest to 0 t-value while the t-value belongs to]-1.96; 1.96[.
3. We relaunch the estimation process with the new set of variables.
4. We check for the improvement of the BIC value.
5. While we have t-values comprised between −1.96 (excluded) and 1.96 (excluded) and while the BIC value is improving, we return to step 2.

This selection process can be quite slow but allows an appropriate selection of the covariates.

```
# Computation of the standard errors
inverse<-solve(optimal$hessian)
result<-cbind(optimal$par, sqrt(diag(inverse)), optimal$par/sqrt
(diag(inverse)))
colnames(result)<-c("Coeffs", "StdError", "t")
rownames(result)<-c("r", "alpha", "s", "delta", colnames(X1),
colnames(X2))
print(result)
# Computation of the Bayesian Information Criterion
BIC<-optimal$value*2+length(b0)*log(length(y))
BIC
```

Results

The descriptive results offer a comprehensive overview of the data sets from the grocery sector which is compared with different data used in previous investigations. In addition, the parameter estimation and comparison of the different models is demonstrated.

Of the 997 total customers in the 26-week cohort, 46.3% are zero repurchasers (Means =1.69, SD = 3.59). The grocery data indicate that the median interpurchase times, even after excluding zero repurchase, is approximately 10.6 weeks, which is low compared with the other applications of the Pareto/NBD model, for which the median interpurchase time is 7 months (office supplies; Schmittlein and Peterson 1994), 17 weeks (catalogue sales; Reinartz and Kumar 2000), or 25 weeks (computer-related products; Reinartz and Kumar 2000). The grocery category features very short purchase cycles, because grocery items are not durable and require frequent replenishment. In addition, the number and heterogeneity of customers is higher in the grocery retail context. For example, the online CD customer base used by Fader et al. (2005a) includes a majority of customers (approximately 85%) who make zero (60%), one, or two repurchases. 46% of grocery retail customers are zero repurchasers, and customers with zero, one, or two repurchases constitute 80% of total grocery retail customers. In contrast, Batislam et al. (2007) find that approximately 40% of grocery retail customers are zero repurchasers, and customers with zero, one, or two repurchases make up around 65% of total grocery retail customers. Such high heterogeneity in grocery purchases decreases the precision of the models.

Parameter Estimations

In order to show a practical application, we interpret the estimated coefficients. They seem coherent for the Pareto/NBD model, with signs in the correct direction (see Table 2).

Purchase frequency is positively influenced by a LP, which is coherent with existing literature (Meyer-Waarden 2007; Leenheer et al. 2007; Liu 2007). The professional occupation of the household members has a direct impact on the purchase activity (frequency) and retention, though a professional situation has the same positive impact on purchase frequency as does lower income. Furthermore,

Table 2 Regression coefficients of the Pareto/NBD model

Frequency regression (β)		Inactivity regression (γ)	
Loyalty card of the shop	+1.01	# of loyalty cards	−0.17
Senior manager	+0.30	Low wage (dummy)	+0.10
Low wage (dummy)	+0.29	Unemployment (dummy)	+1.10
Over 50 (dummy)	−0.73		

Notes: The insignificant coefficients ($p > 0.1$) are household size, profession (employee, worker), wages (1,000–2,000 €, > 2,000 €), and age (30–50 years, < 30 years) people older than 50 years of age are less mobile than younger people and display lower purchase frequencies. Younger shoppers are more likely to engage in smaller, more frequent fill-in trips than are older ones, probably because the former buyers have more disposable time but less income, which drives them to buy in smaller quantities at higher frequencies (Kahn and Schmittlein 1989; Bell et al. 1998). Financial instability of households (i.e., low wages, unemployment) has a negative impact on inactivity. Grocery patronage behavior depends on the level of education and income, which increase the chances that the consumers uses a more rational purchase process and thus attaches less importance to marketing variables (e.g., store advertisement, promotions, loyalty program rewards). Generally, the more education people possess, the less sensitive they are to a store's promotions or other marketing actions, and the less loyal they are, which means their defection probability is higher and retention is lower (Narasimhan 1984). Less educated households with lower incomes tend to remain loyal, because they experience more influence from store marketing variables. According to an alternative but not incompatible explanation, they also probably have higher switching costs related to mobility constraints (money, transports), which increases the utility of the closest and most familiar store.

Multiple LP memberships relate positively to inactivity, which is coherent with the results of Meyer-Waarden (2007) and may indicate a learning effect with regard to the use of loyalty schemes. Disloyal, opportunistic buyers who regularly shop in several stores and are members of different loyalty schemes (on average, European households possess three grocery retailing loyalty cards; ACNielsen 2005) are more experienced and have smaller switching costs. These purchasers join LP more readily and quickly (Meyer-Waarden 2007; Leenheer et al. 2007).

Table 3 provides the results of the gamma and beta distributions. The parameters for frequency do not vary significantly, despite the introduction of explanatory variables. However, the parameters for the inactivity or dropout rates vary strongly; the drastic growth of δ probably relates to the explanatory variables.

Table 3 Coefficients of the gamma/beta distributions

	Basic NBD	BG/NBD	Standard Pareto/NBD	Explanatory Pareto/NBD
Index of homogeneity in purchase rate: r	0.50	0.43	0.57	0.66
α	5.72	3.94	5.60	6.91
Average purchase rate: r/α	0.09	0.10	0.10	0.10
a		0.22		
b		1.14		
Average inactivity probability: $a/(a + b)$		0.16		
Index of homogeneity in inactivity rate: s			0.63	1.56
δ			30.16	107.55
Average inactivity rate: s/δ			0.02	0.01

The parameters r (which can be seen as purchase rates) and s (which can be seen as churn rates) increase in the explanatory model. Both provide an index of homogeneity (Schmittlein et al. 1987), and their increase denotes more significant homogeneity across customers in the explanatory model. For the explanatory formulation, gamma functions capture residual heterogeneity, not all the heterogeneity, as in the case of purely stochastic formulations.

Purchase Prediction Validity

Empirical analysis carried out for both the 26- and 52-week observation periods for the cohort relies on a popular criterion for adjustment, the Bayesian information criterion (BIC), whose values are based on a log scale. The expression is written as follows: $\mathrm{BIC} = -2LL + k\ln(n)$, where k the number of parameters and n the sample size.

The adjustment differences between the BG/NBD approach and the explanatory Pareto/NBD model are not very important, and the BIC is very close for both (see Table 4). If one considers the mean absolute percent error (MAPE) as an empirical criterion, the explanatory Pareto/NBD model has slightly worse results than either the standard or the BG/NBD model (15.5% vs. 12% and 10.5%; the basic NBD achieves the worst results at 38.1%). This result makes sense. According to Fader et al. (2005a), the BG/NBD forecasts are better when purchase frequency is high, as in the grocery retailing context, because of the differences among the model structures. Under the Pareto/NBD model, dropout occurs at any time – even before a customer has made a first purchase. However, under the BG/NBD, a customer cannot become inactive before making his or her first purchase. If buying rates are fairly high, BG/NBD and Pareto/NBD perform similarly well. However, in contexts in which purchase frequencies are low, the BG/NBD model suffers in comparison with the Pareto/NBD approach.

After having tested the robustness of the models, a more thorough investigation of their performance is completed. The accuracies of the different models are not similar (Fig. 3).

During the validation period, the BG/NBD model performs quite well, whereas the Pareto/NBD and explanatory Pareto/NBD formulations underestimate the weekly purchase frequency. The basic NBD model does not perform well at all. With the exception of the basic NBD model, the approaches converge to actual repeat purchases during the forecast period. Weekly sales rise during the first 14 weeks, due to new customers in the cohort and their repeat purchases. All models underestimate the peak in weekly actual purchases in the initial weeks, probably

Table 4 Log-likelihood and Bayesian information criterion

	Basic NBD	BG/NBD	Standard Pareto/NBD	Explanatory Pareto/NBD
Log-Likelihood	−4,954	−4,922	−4,935	−4,900
Bayesian information criterion	9,922	9,872	9,898	9,876

because they miss the increasing trend in repeat purchases due to promotions during the same period. Later in the observation period, all models (with the exception of the basic NBD) match the actual purchases.

The deviation of weekly estimates from actual purchases during the initial weeks leads to an underestimation of the cumulative repeat purchases in the initial weeks as well (see Fig. 4).

During the forecast period (52 weeks), the models underestimate actual purchases (Pareto/NBD model: −9%, explanatory Pareto/NBD model: –14%, BG/NBD

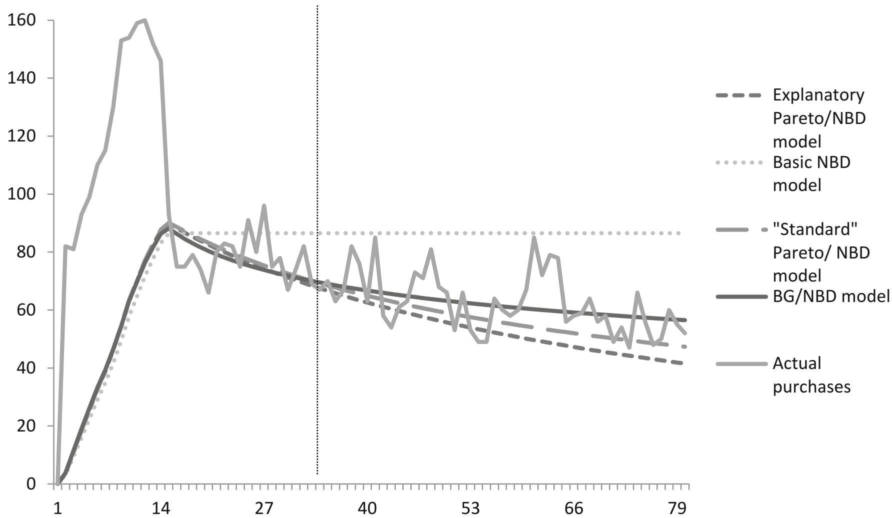

Fig. 3 Estimation of the weekly repeat purchases

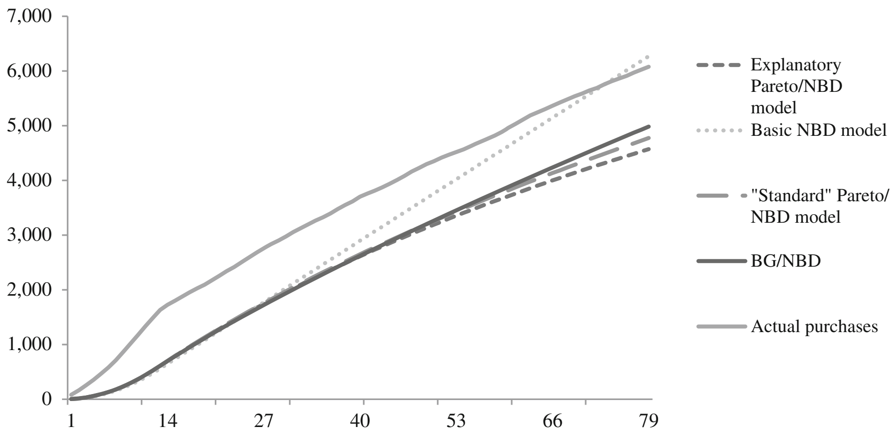

Fig. 4 Estimation of the cumulative repeat purchases

model: −2%). The fairly high purchase frequency rates in grocery shopping may explain the strong results derived from the BG/NBD model. In this case, the assumption that a customer is active until he or she makes a repeat purchase is not a problem. However, purchase frequency is not too high to affect the dropout time ("exhaustion" effect) of the BG/NBD model.

We measure individual-level performance according to the conditional expectations for the forecast period, depending on the number of repeat purchases in the observation period (Fig. 5). That is, for each value of x in the observation period, an average of the actual number of purchases in the forecast period is compared.

The forecasts of the BG/NBD and the standard Pareto/NBD models are very close and provide acceptable predictions of the expected number of transactions in the holdout period, consistent with the results of Fader et al. (2005b). The Pareto/NBD model offers slightly better predictions than the BG/NBD, but it is important to keep in mind that the number of heavy buyers is small. The explanatory model and the basic NBD model systematically overestimate the number of repeat purchases, especially for heavy customers.

Another way to assess the predictive validity of the models is to group customers on the basis of their recency and frequency characteristics. One can then compare the results with traditional recency/frequency (RF) segmentation analysis.

Each of the customers is assigned to a RF segment in the following manner. The terciles for recency and frequency (the customers who made no repeat purchases are coded as R = F = 0) have to be determined. High recency means a low number of days since the last purchase, i.e., a recent repurchaser. At the opposite, a low recency characterizes an exceptional repurchase. In Table 5, the size of each RF group is shown.

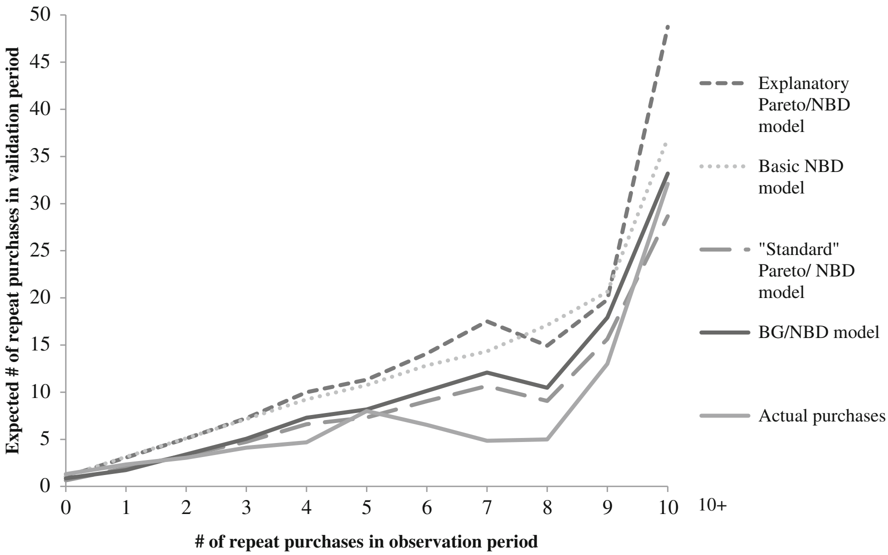

Fig. 5 Conditional expectations

Only 15% of the customers are frequent and recent repurchasers. On the other hand, the zero repurchasers during the estimation period represent almost half of the customers. In a traditional approach, managers in the retailing sector would assume that, after a half year of inactivity, a customer is inactive.

In fact, the average number of purchases made by those customers, during the following 52 weeks, is four times lower in average than the same number made by other segments. However, due to the size of this group, their contribution is really impressive: they represent 18.3% of the total of the purchases of the following 52 weeks, the second contribution of all the segments. This aspect is very interesting. It is taken into account by the models (even the contribution of the zero repurchaser is underestimated between 9.6% and 11.7% instead of 18.3%) (Fig. 6).

Table 5 Repartition of the customers between RF segmentation

Recency	Frequency of repeat purchases (estimation period: 26 weeks)	# of customers
No repeat purchase	0	**46.3%**
Low recency	1	3.0%
	2	0.5%
Total low recency		**3.5%**
Medium recency	1	14.1%
	2	5.6%
	3+	5.1%
Total medium recency		**24.9%**
High recency	1	5.2%
	2	4.9%
	3+	15.1%
Total high recency		**25.3%**

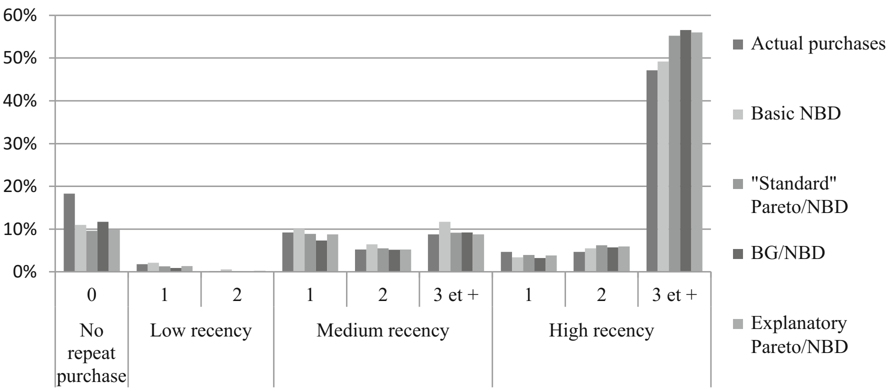

Fig. 6 Purchases by recency and frequency

Conclusion

We have seen an operational form of CLV, retention, and churn models, (namely, the Pareto/NBD, BG/NBD, and explanatory Pareto/NBD models) and their high degree of validity for customer base analysis and for forecasting a customer's future purchasing, conditional on his or her past buying behavior. The Pareto/NBD, BG/NBD, and explanatory Pareto/NBD models systematically outperform the basic NBD model, because it does not consider inactivity.

The Pareto/NBD and explanatory Pareto/NBD formulations underestimate weekly purchase frequency, whereas the BG/NBD model performs quite well. These results show that explanatory variables introduce more information and therefore generate a better forecast.

However, even if the predictive validity of the explanatory NBD/Pareto model is not necessarily better, its performance does not suffer in comparison with the Pareto/NBD and the BG/NBD models. Nevertheless, the advantages of the explanatory approach relate more to the opportunity to explain the impact of personal characteristics and the impact of marketing actions rather than the accuracy of the forecasts at an aggregate level. The ability of the explanatory Pareto/NBD model to predict future purchases is quite good. Even with a reduced set of explanatory variables, the explanatory Pareto/NBD model is as accurate as the standard formulation. Nevertheless, the results are not better than those of the BG/NBD approach. However, improvements are possible with other sets of variables (i.e., more marketing mix variables).

These CLV, retention, and churn models for customer base analysis can help managers understand why their marketing operations work, or do not work, and how and to which customer segments they should improve their efforts. The explanatory model approach represents a promising way to understand buyer behavior. The applications are broad, including segmentation, understanding customer life cycles, determining elasticities and elements that influence loyalty and purchase behavior, the possibility of analyzing marketing actions and personal characteristics, and a means to establish more valid customer CLV models to predict customer value.

Managers should be encouraged to use these models to determine their customer base analysis, CLV calculations, and resource allocations, using their often large longitudinal databases.

Further research should address underlying model assumptions that are unrealistic and not compatible with extant literature about purchasing behavior. For example, researchers could relax the Poisson distribution assumptions and perhaps use a Weibull distribution instead. The BG/NBD formulation suffers a major weakness because its underlying conditions (i.e., dropout rates independent of purchase frequencies) demand inactivity appear immediately after each repurchase act. This behavioral assumption is not compatible with purchasing behavior literature. In the same sense, the Pareto/NBD model supposes independence between purchase frequency and inactivity, which may be reasonable only in "always-a-share" markets (Reinartz and Kumar 2000). Some other authors also suppose a link between both variables (East et al. 2000).

Few current models explicitly incorporate competition, yet heightened competition can affect customer CLV in several ways – shortened expected lifetime, decreased prices, and increased acquisition costs. Panel data provide a promising source for some firms, and surveys can be very useful in capturing the effect of competition. Other empirical investigations should examine in which conditions (high/low purchase frequencies) and with which type of data (internal, panel) the different models perform best.

Compared with predicting purchase frequency and weekly repeat purchases, forecasts of individual purchases include more customer information and should provide higher accuracy in individual-level forecasts. However, it remains difficult to model individual purchase behavior, especially with regard to the highly heterogeneous purchase behavior encountered in grocery sales (Fader and Hardie 2013).

Finally, to allocate optimally, managers cannot simply measure CLV but instead must know how CLV reacts to changes in the marketing mix. Additional research should address this concern.

References

Abe, M. (2009). ‘counting your customers’ one by one: A hierarchical Bayes extension to the Pareto/NBD model. *Marketing Science, 28*(3), 541–553.

Allenby, G. M., Leone, R. P., & Jen, L. (1999). A dynamic model of purchase timing with application to direct marketing. *Journal of the American Statistical Association, 94*(446), 365–374.

Ayache, A., Calciu, M., & Salerno, F. (2006). Stochastic approach to customer equity and lifetime value calculations with applications to customer retention models and some extensions. *35th Conference of EMAC,* Athènes.

Baltagi, B. H. (2003). *A companion to theoretical econometrics*. Willinston: Blackwell.

Batislam, E., Denizel, M., & Filiztekin, A. (2007). Empirical validation and comparison of models for customer base analysis. *International Journal of Research in Marketing, 24*(3), 201–209.

Bell, D. R., Ho, T. H., & Tang, C. S. (1998). Determining where to shop: Fixed and variable costs of shopping. *Journal of Marketing Research, 35*(3), 352–369.

Berger P., & Nasr, N. (1998). Customer lifetime value: Marketing models and applications. *Journal of Interactive Marketing,* Winter 98, *12*(1), 17–30.

Bijmolt, T., Dorotic, M., & Verhoef, P. (2010). Loyalty programs: Generalizations on their adoption, effectiveness and design. *Foundations & Trends in Marketing, 5*(4), 197–258.

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

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

Blattberg, R. C., Getz, G., & Thomas, J. S. (2001). *Customer equity: Building and managing relationships as valued assets*. Boston: Harvard Business School Press.

Blattberg, R. C., Kim, B., & Neslin, S. A. (2008). *Database marketing. Analyzing and managing customers*. New York: Springer.

Bolton, R. N., Lemon, K. N., & Verhoef, P. C. (2004). The theoretical underpinnings of customer asset management: A framework and propositions for future research. *Journal of the Academy of Marketing Science, 32*(3), 271–292.

Borle, S., Singh, S., & Jain, D. C. (2008). Customer lifetime value measurement. *Management Science, 54*(1), 100–112.
Calciu, M., & Salerno, F. (2002). Customer value modelling: Synthesis and extension proposals. *Journal of Targeting, Measurement & Analysis for Marketing, 11*(2), 124–134.
Castéran, H., Meyer-Waarden, L., & Benavent, C. (2007a). *Empirical evaluation of NBD models for the estimation of life time value in the retailing sector*. XXIIIrd International Annual Conference of the French Marketing Association, Aix-les-Bains.
Castéran, H., Meyer-Waarden, L., & Benavent, C. (2007b). *Incorporation of covariates in the Pareto/NBD model: First formulations and comparison with others models in the retailing sector*. Third German French Austrian Conference, ESSEC, Paris.
Castéran, H., Meyer-Waarden, L., & Benavent, C. (2008). *Application of latent class models to purchases in the retailing sector and comparison with the Pareto/NBD formulation*. 7ème Congrès Tendances du Marketing Paris Venise (Venise).
Dwyer, F. R. (1997). Customer lifetime valuation to support marketing decision making. *Journal of Direct Marketing, 11*(4), 6–13.
Dwyer, R. (1989). Customer lifetime valuation to support marketing decision making. *Journal of Direct Marketing, 3*(4), 8–15.
Dziurzynski, L., Wadsworth, E., & McCarthy, D. (2014). BTYD: Implementing buy 'Til You Die Models. R package version 2.4. https://CRAN.R-project.org/package=BTYD
East, R., Hammond, K., Harris, P., & Lomax, W. (2000). First-store loyalty and retention. *Journal of Marketing Management, 16*(4), 307–325.
Ehrenberg, A. S. C. (1959). The pattern of consumer purchases. *Applied Statistics, 8*(1), 26–41.
Ehrenberg, A. S. C. (1988). *Repeat buying: Facts, theory and applications*. London: C. Griffin & Co..
Fader, P., & Hardie, B. (2007a). *Incorporating time-invariant covariates into the Pareto/NBD and BG/NBD models*. Working Paper, University of Pennsylvania and London Business School.
Fader, P. S., & Hardie, B. G. S. (2007b). How to project customer retention. *Journal of Interactive Marketing, 21*, 76–90.
Fader, P., & Hardie, B. (2009). Probability models for customer-base analysis. *Journal of Interactive Marketing, 23*(1), 61–69.
Fader, P., & Hardie, B. (2013). Overcoming the BG/NBD model's #NUM! error problem. http://brucehardie.com/notes/027/. Accessed 23 Feb 2015.
Fader, P. S., Hardie, B. G. S., & Zeithammer, R. (2003). Forecasting new product trial in a controlled test market environment. *Journal of Forecasting, 22*, 391–410.
Fader, P., Hardie, B., & Lok, L. K. (2005a). Counting your customers the easy way: An alternative to the Pareto/NBD model. *Marketing Science, 24*(2), 275–284.
Fader, P., Hardie, B., & Lok, L. K. (2005b). RFM and CLV: Using iso-value curves for customer base analysis. *Journal of Marketing Research, 42*(4), 415–430.
Fudenberg, D., & Tirole, J. (2000). Customer poaching and brand switching. *RAND Journal of Economics, 31*(4), 634–657.
Glady, N., Lemmens, A., & Croux, C. (2015). Unveiling the relationship between the transaction timing, spending and dropout behavior of customers. *International Journal of Research in Marketing, 32*(1), 78–93.
Gupta, S., & Lehmann, D. R. (2003). Customers as assets. *Journal of Interactive Marketing, 17*(1), 9–24.
Gupta, S., Lehmann, D. R., & Stuart, J. A. (2002). Valuing customers. *Journal of Marketing Research, XLI*, 7–18.
Gupta, S., Lehmann, D. R., & Stuart, J. A. (2004). Valuing customers. *Journal of Marketing Research, 41*(1), 7–18.
Gupta, S., Hanssens, D., Hardie, B., Kahn, W., Kumar, V., Lin, N., & Sriram, N. (2006). Modeling customer lifetime value. *Journal of Service Research, 9*(2), 139–155.
Hlavinka, K., & Sullivan, J. (2011). The art and science of building customer value. Colloquy.

Hogan, J. E., Lemon, K. N., & Rust, R. R. (2002). Customer equity management: Charting new directions for the future of marketing. *Journal of Service Research, 5*(1), 4–12.
Jackson, B. B. (1985). Build customer relationship that last. *Harvard Business Review, 63*, 120–128.
Jain, D., & Singh, S. (2002). Customer lifetime value research in marketing: A review and future directions. *Journal of Interactive Marketing, 16*(2), 34–46.
Jerath, K., Fader, P., & Hardie, B. (2011). New perspectives on customer death using a generalization of the Pareto/NBD model. *Marketing Science, 30*(5), 866–880.
Kahn, B. E., & Schmittlein, D. C. (1989). Shopping trip behavior: An empirical investigation. *Marketing Letters, 1*(4), 55–70.
Keane, T. J., & Wang, P. (1995). Applications for the lifetime value model in modern newspaper publishing. *Journal of Direct Marketing, 9*(2), 59–66.
Kumar, V. (2007). Customer lifetime value: The path to profitability. *Foundations & Trends in Marketing, 2*(1), 1–96.
Kumar, V., & Reinartz, W. (2006). *Customer relationship management: A databased approach*. Hoboken: Wiley.
Leenheer, J., Bijmolt, T. H. A., van Heerde, H. J., & Smidts, A. (2007). Do loyalty programs enhance behavioral loyalty? A market-wide analysis accounting for endogeneity. *International Journal of Research in Marketing, 24*(1), 31–47.
Lewis, M. (2005). Incorporating strategic consumer behavior into customer valuation. *Journal of Marketing, 69*, 230–238.
Lewis, M. (2006). Customer acquisition promotions and customer asset value. *Journal of Marketing Research, 43*(2), 195–203.
Libai, B., Narayandas, D., & Humby, C. (2002). Toward an individual customer profitability model: A segment-based approach. *Journal of Service Research, 5*(1), 69–76.
Liu, Y. (2007). The long-term impact of loyalty programs on consumer purchase behavior and loyalty. *Journal of Marketing, 71*(4), 19–35.
Meyer-Waarden, L. (2007). The effects of loyalty programs on customer lifetime duration and share of wallet. *Journal of Retailing, 83*(2), 223–236.
Meyer-Waarden, L., & Benavent, C. (2009). Retail loyalty program effects: Self-selection or purchase behavior change? *Journal of the Academy of Marketing Science, 3*(3), 345–358.
Morgan, R., & Hunt, S. (1994). The commitment-trust theory of relationship marketing. *Journal of Marketing, 58*(3), 20–38. doi:10.2307/1252308.
Morrison, D., & Schmittlein, D. (1988). Generalizing the NBD model for customer purchases: What are the implications and is it worth the effort? *Journal of Business and Economic Statistic, 6*, 145–159.
Narasimhan, C. (1984). A price discrimination theory of coupons. *Marketing Science*, Spring 84, *3*(2), 128–147.
Newcomb, S. (1886). A generalized theory of the combination of observations so as to obtain the best result. *American Journal of Mathematics, 8*, 343–366.
Pearson, K. (1894). Contributions to the mathematical theory of evolution. *Philosophical Transactions, A185*, 71–110.
Pfeifer, P. E., & Carraway, R. L. (2000). Modeling customer relationships as Markov chains. *Journal of Interactive Marketing, 14*(2), 43–55.
R Development Core Team. (2010). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.
Reinartz, W. J., & 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*(4), 17–35.
Reinartz, W., & Kumar, V. (2003). The impact of customer relationship characteristics on profitable lifetime duration. *Journal of Marketing, 67*(1), 77–99.
Reinartz, W. J., Thomas, J. S., & Kumar, V. (2005). Balancing acquisition and retention resources to maximize customer profitability. *Journal of Marketing, 69*, 63–79.

Rogers, E. M. (2003). *Diffusion of innovations*. New York: Free Press.
Rust, R. T., Lemon, K. A., & Zeithaml, V. A. (2004). Return on marketing: Using customer equity to focus marketing strategy. *Journal of Marketing, 68*(1), 109–127.
Ryals, L. (2005). Making customer relationship management work: The measurement and profitable management of customer relationships. *Journal of Marketing, 69*(4), 252–261.
Schmittlein, D., & Peterson, R. (1994). Customer case analysis: An industrial purchase process application. *Marketing Science, 13*(1), 41–67.
Schmittlein, D., Morrison, D., & Colombo, R. (1987). Counting your customers: Who are they and what will they do next? *Management Science, 33*(1), 1–24.
Schulze, C., Skiera, B., & Wiesel, T. (2012). Linking customer and financial metrics to shareholder value: The leverage effect in customer-based valuation. *Journal of Marketing, 76*(2), 17–32.
Schweidel, D. A., & Fader, P. S. (2009). Dynamic change points revisited. An evolving process model of new product sales. *International Journal of Research in Marketing, 26*(2), 119–124.
Schweidel, D. A., Fader, P. S., & Bradlow, E. T. (2008). Understanding service retention within and across cohorts using limited information. *Journal of Marketing, 72*, 82–94.
Simester, D. I., Sun, P., & Tsitsiklis, J. N. (2006). Dynamic catalog mailing policies. *Management Science, 52*(5), 683–696.
Stauss, B., & Friege, C. (1999). Regaining service customers: Costs and benefits of regain management. *Journal of Service Research, 1*(4), 347–361.
Tarasi, C., Bolton, R., Hutt, M., & Walker, B. (2011). Balancing risk and return in a customer portfolio. *Journal of Marketing, 75*(3), 1–17.
Temporal, P., & Trott, M. (2001). *Romancing the customer: Maximising brand value through powerful relationship management*. New York: Wiley.
Thomas, J., Blattberg, R., & Fox, E. (2004). Recapturing lost customers. *Journal of Marketing Research, 38*(2), 31–45.
Venkatesan, R., & Kumar, V. (2004). A customer lifetime value framework for customer selection and resource allocation strategy. *Journal of Marketing, 68*(4), 106–125.
Villanueva, J., & Hanssens, D. (2007). Customer equity: Measurement, management and research opportunities. *Foundations and Trends® in Marketing, 1*(1), 1–95.
Villanueva, J., Yoo, S., & Hanssens, D. M. (2008). The impact of marketing-induced versus word-of-mouth customer acquisition on customer equity growth. *Journal of Marketing Research, 45*(1), 48–59.
Wedel, M., Desarbo, W. S., Bult, J. R., & Ramaswamy, V. (1993). A latent class Poisson regression model for heterogeneous count data. *Journal of Applied Econometrics, 8*(4), 397–411.
Wiesel, T., Skiera, B., & Villanueva, J. (2008). Customer equity: An integral part of financial reporting. *Journal of Marketing, 72*, 1–14.